INTRODUCTION

What distinguishes a GIS from other types of information systems are its spatial analysis functions. These functions use the spatial and non-spatial attribute data in the GIS data base to answer questions about the real world.

The GIS data base is a model of the real world that can be used to mimic certain aspects of reality. To mimic behaviour, a model must represent certain entities (i.e. things) and the relationships among them (i.e. the rules that govern how they interact). The entities might be the names of individuals and a list of properties. The relationships might include ownership, mortgage, and mortgage. Together these entities and relationships provide a model for land ownership. A model may be represented in words, mathematical equations, or as a set of spatial relations displayed as a map or stored in the computer hardware and software of a GIS.

It is important to recognize that models are designed to mimic only selected aspects of reality. In general, the more factors that a model takes into account, the more complex it becomes, and the more expensive it is to use and maintain. A more complex model may or may not provide "better" answers; it depends on the questions to be addressed.

Models are used when it is more convenient or it is not possible to collect the information directly. It is more convenient to read road distances off a map than to go out and measure them, and it is not possible to measure the height a forest will reach in 100 years time. The value in using a model is that it can be tested and manipulated more conveniently, at a faster (or slower) rate, and less expensively than the conditions it mimics. In many cases, the model is used to repeatedly perform analyses that test alternative scenarios, such as the effectiveness of an emergency evacuation plan in response to different kinds of events. A model is used to answer questions about what exists now or existed at some point in the past. Perhaps most importantly, it can be used to predict what will happen or has happened in another location or at another point in time. A GIS provides these capabilities by means of its analysis functions.

An important GIS application is predicting the consequences of proposed activities. They may involve large areas, e.g. when a reservoir is formed behind a dam, or they may involve relatively small areas, such as the effect on traffic flow of closing a city street. This ability to model what will occur provides the opportunity to select the "best" alternative. But what is "best"? A GIS can use spatial analysis functions to provide part of the answer to "What is best?". The part it cannot answer is the human value judgments that define the wants, goals, and values of the organization or society that are using the information. These judgments define the allowable trade-offs among alternatives. A failure to consider the relevant value judgments in a GIS analysis can make the procedure a theoretical exercise instead of a practical tool. Or worse yet, the analysis results may be misleading.

To develop the best answers from the information available requires a systematic framing of the questions to be addressed. There is often a strong tendency to begin the analysis with only a general idea of the questions to be answered and the data needed. This is a particularly common tendency whenever a new computer-based system, such as a GIS, is introduced.
In order to find useful answers, one must ask the right questions. Working backwards through an analysis is often a good approach. That is, begin at the end. Assuming that good answers have been produced, what questions would they have answered, what concerns would they have addressed, and what data, including judgments, would have been used in the analysis that produced those answers. This approach to the design of an analysis ensures that the effort is focussed on answering the appropriate questions. The following framework and examples illustrate this approach.

The answers provided by a GIS can be categorized into three types as illustrated in Figure 7.1:

1. a presentation of the current data, i.e. the data in the database such as a map of the city streets.
2. a pattern in the current data, such as all houses valued at over $100,000. and;
3. a prediction of what the data could be at a different time or place. For example, predicting the services that would be lost in the event of an earthquake. This type of analysis might be used to develop emergency response plans.

The types of questions to be answered can also be characterized by three categories:

1. What are the data? i.e. what is the information currently stored in the database. For example, what is the name and address of the owner of a specified property?
2. What is the pattern in the data? This type of question is a search for entities that possess a specified set of characteristics. For example, plotting a map of all lots with houses valued at over $100,000 would be defining a pattern in the data that may not be obvious when all the data are viewed together -- the pattern for this type of house.
3. What could the data be? This type of question implies that a predictive model will be used. The model may be as simple as predicting that a field will produce the same crop next year as this year. It may be as complex as predicting the change in stream flow after a forest has been removed from a watershed.

The functions used to produce these answers can similarly be categorized by the types of answers they provide:

1. storage and retrieval functions,
2. constrained query functions, and
3. modelling functions.

Figure 7.1 Categorizing Questions, Functions, and Answers in a GIS Analysis.
These categories of questions, functions, and answers are not mutually exclusive. A given answer, function, and question will have aspects of each category. This is represented conceptually by their position in the circles. The following examples illustrate this idea. The letter representing each example is shown in the Figure.

A. Retrieving the street map for an area primarily involves the retrieval of existing information. The map exists in the GIS and need only be recalled and output.

B. Retrieving those lots with houses valued over $100,000 is an example of the second category of use. Here the value of each lot must be considered and only those satisfying the constraint are accepted.

C. Determining the optimum routing of a powerline is an example of modelling. Multiple layers of information are used together to weigh different alternatives in order to optimize the design.

The strategy for undertaking a specific GIS analysis will depend on the answers that are to be provided. By defining the most important answers to be provided, an appropriate set of questions and analysis methods can be specified.

The remainder of this chapter presents the different functions used in a GIS to manipulate and analyze geographic information. The art and science of using a GIS is to know how to combine the analysis functions available on a particular system to provide the required information using the available data. The purpose of first presenting this framework of answers, questions, and functions is to highlight the fact that individual analysis functions must be used in the context of a complete analysis strategy. One of the unwanted characteristics of computer technology is the ability to produce incorrect information at a rapid rate, and with all the apparent authenticity of the "real McCoy." The quality of the information produced from a GIS depends on the intelligent use of a systematic analysis approach.

ORGANIZING GEOGRAPHIC DATA FOR ANALYSIS

Geographic information are organized within a GIS so as to optimize the convenience and efficiency with which they can be used. The form of organization chosen will be influenced by the types of data to be used, the types of analyses to be performed, and the methods used to encode the data. Methods used to represent geographic data within a GIS were discussed in Chapter 6. In this section, the logical organization of the data is addressed.

On a paper map, geographic information is usually organized as a set of themes, such as roads, streams, land cover types, and political boundaries. They are often thought of as map layers, and each layer may actually have been plotted separately in the process of compiling the final map. To cover a large area, such as a country, it may be necessary to use several map sheets, and so a convenient system is used to divide the coverage area into individual map sheets. The level of detail used to present the geographic information is chosen according to the information needs that were specified and the limitations of the storage medium. In the case of a paper map, the map itself is both the means of storing the geographic information, as well as the form of presentation. Symbols, colours, line widths, and other map elements are selected to suit the need for visual analysis of the map data. The double service of the paper map as both a storage and presentation medium requires that trade-offs be made between the amount and accuracy of information that can be shown and the need for the map to be legible.

In a computer-based GIS, these organizational considerations are handled somewhat
differently. Since the data storage and presentation (or output) are separate, the level of detail at which the geographic information can be stored is limited by the storage capacity of the hardware and the method used by the software to represent the data. The legibility of an output product like a map can be controlled by selecting the scale, amount, and level of detail of information, and symbols at the time when the map is to be plotted. (One can present information at a less detailed level than it was stored but cannot present more detailed information than exists in the database. For this reason, information need only be entered once, at the finest level of detail that will be required. However, in the case of paper maps, separate maps are needed to show information at different levels of detail and different scales.)

Large coverage areas are subdivided into smaller units for more efficient storage, in a manner similar to the map sheet concept. Each unit is commonly stored as a separate set of data files. Unlike the paper map, a GIS

![GIS Data Layers Commonly Used in Natural Resource GIS Applications. (Courtesy of ESRI, Redlands, California.)](image)
can provide sophisticated functions to ensure that adjacent units match precisely along their borders. In addition, many systems hide these subdivisions from the user, presenting a seamless coverage of the entire area as if it were a single very large map. The different types of thematic information, represented as different map layers in paper maps, are treated as different data layers in a GIS.

**DATA LAYERS**

A data layer consists of a set of logically related geographic features and their attributes. The features to be grouped in a single data layer are chosen for the convenience of the users. The organizing principal may be to group similar feature types. For example, the data may be organized thematically, i.e., by the type of geographic features they represent. For example, roads and railways might be combined as a single transportation data layer and streams and lakes as a hydrology data layer. Figure 7.2 illustrates the organization of data layers for a natural resource application, and Figure 7.3 illustrates an urban application. The organization of the data layers will also depend on the restrictions imposed by the GIS software used. It may be necessary or more convenient to store point, line, and area features in separate data layers.

**PARTITIONING THE COVERAGE AREA**

When a GIS must handle large amounts of spatial data, the coverage area may be subdivided into smaller units termed tiles, as shown in Figure 7.4. The allowable shapes and sizes of tiles depend on the restrictions

![Figure 7.3 GIS Data Layers Commonly Used in Urban GIS Applications. (Courtesy of ESRI, Redlands, California.)](image)
of the software. In general, tile boundaries should be chosen that will be stable for the life of the database and that will enhance the performance and use of the system. A grid defined by latitude and longitude or UTM coordinates is commonly used. Tiles may also be organized by interest areas. Each tile might represent a different ranger district in a national forest. Searches are faster when the areas to be retrieved correspond to the tile structure used. If users most frequently access data by 7.5 quadrangle map sheets, then tiles representing one or even multiples of map sheets might be suitable.

In some systems the user must directly create and manage the tiles as separate coverage areas, re-assembling the files for adjacent areas when needed. A more sophisticated approach is to provide special purpose software to automatically create and manage the tiling so that the data are re-assembled as needed without operator interaction. The automated management of tiling is a database operation that provides a service analogous to a map library. (For this reason the software is often termed map library software.) In addition to managing the partitioning and re-assembling of data layers, map library functions may include checks on data consistency, control of access and updating, and storage of repeatedly used map output formats.

The way that a map library is implemented will affect the overall performance of the system. The ease with which users can access and analyze their data, and the maintenance of the database. Once the map library has been set up it is usually difficult to change. For this reason, the design of a map library should be given careful consideration by experienced personnel. There should be a systematic design stage, followed by a pilot implementation and evaluation. Interviews can be used to define the coverage areas
and the types and level of detail of data required by the different user groups. Other important factors in developing the design are the frequency with which the data will be used and the volume of data to be stored and accessed. If there are different user groups that have competing needs, it may be necessary to provide them with separate map libraries in order to optimize system performance.

The partitioning of data layers into tiles can increase system performance by providing efficient retrieval of a subset of data. However, when an entire coverage area must be retrieved, tiling will slow the storage and retrieval operations because additional steps are needed to re-assemble the data. For smaller data volumes, the overhead of partitioning the data layers may not be justified and the other data base functions provided by map library software, such as consistency checks and access control, may not be needed. However, for large data volumes, there is no choice but to use tiles. At some point the size of the file needed to store a data layer will exceed the file size limit of the system and the area will have to be subdivided into units that can be stored in smaller files.

A CLASSIFICATION OF GIS ANALYSIS FUNCTIONS

The development of GIS techniques has provided a constantly growing number of ever more sophisticated analysis functions. A description of even the most common functions would quickly overwhelm the uninitiated. The approach used here is to group the functions into four major categories, each with several subdivisions. Figure 7.5 presents a classification of GIS analysis functions. The first level of classification contains four groups: 1. Maintenance and Analysis of the Spatial Data. 2. Maintenance and Analysis of the Attribute Data. 3. Integrated Analysis of Spatial and Attribute Data, and 4. Output Formatting. Each major group is further subdivided into types of functions. The distinctions among these categories are somewhat artificial and not clear-cut, but they do provide a useful framework.

The way that a GIS function is implemented depends on such factors as the data model (e.g., raster versus vector), the hardware, and performance criteria (e.g., how fast it must run, what options must be provided). These details are important and require considerable expertise to properly evaluate. However, this level of detail is not needed to understand the types of analysis functions that a GIS can provide, how they are used, and why they are valuable. The remainder of this chapter discusses GIS analysis and manipulation functions in this context.

MAINTENANCE AND ANALYSIS OF THE SPATIAL DATA

Maintenance and analysis functions are used to transform spatial data files, edit them, and assess their accuracy. They are primarily concerned with the spatial data and require little if any reference to the associated non-spatial attribute information. The approaches used to provide these functions differ among GISes. All GISes require the capability to transform source data into the data structure used within the system and to edit those files once they have been created. In addition, the data may need to be transformed so that different data layers for the same area are properly registered to each other or to a selected geographic coordinate system.

It may also be necessary to assemble files for adjacent areas into a single file (termed mosaicing). In order to improve storage efficiency, it may be desirable to reduce the quantity of data used to store the information. Coordinate thinning is a procedure that reduces the number of coordinate pairs
Figure 7.5 A Classification of GIS Analysis Functions.
used to define boundaries. It may also be necessary to reconcile lines that represent the same boundary in different data layers but do not coincide. A particular GIS may include all of these functions or depend on the user to supply data that have been suitably pre-processed.

**FORMAT TRANSFORMATIONS**

Data may be supplied to a GIS in the form of lists of points that were generated from a digitizer. They may be input as a digital file of gridded elevation values or as base maps in Digital Line Graph (DLG) format. These files must be transformed into the data structure and file format used internally by the GIS. The transformation procedure may be fast and straightforward in cases where little additional processing is needed. A raster file that is input to a raster-based GIS may require virtually no re-formatting. The internal files may only differ from the original data by the addition of some information (often termed header information or simply a header) to identify the name, origin, size, and other parameters used by the system.

In the case of topologically structured vector-based systems, it is usually necessary to “build” or create the topology from the coordinate data. This procedure is critical, for without it the topology of the overlay is not available to the system. It is also relatively time-consuming, taking minutes to hours depending on the number of map elements (i.e., the number of points, lines, and polygons) and the capabilities of the hardware and software.

The format transformation procedure can become a very costly and time-consuming operation when the data are not collected in a form well-suited to the GIS. For example, map information is often digitized for automated drafting applications in a format that is not topologically structured. When these files are later used as input to a GIS, they are often difficult or impossible to transform into the topological structure needed by the GIS. Such problems as polygons that don’t close and lines that don’t meet may require considerable editing. It may, in fact, be less expensive to re-digitize the entire map. For this reason, the cost of format transformation can significantly affect the cost of using digital data sets and will influence the source data selected for input to the GIS.

**GEOMETRIC TRANSFORMATION**

Geometric transformations are used to assign ground coordinates to a map or data layer within the GIS or to adjust one data layer so it can be correctly overlayed on another of the same area. The procedure used to accomplish this correction is termed registration (i.e., the different data layers are registered to a common coordinate system or to one data layer that is used as a standard). Data layers for the same area must all be registered so that the same location in each overlay has the same map coordinates. The data may not have been precisely registered during data entry because the digitizing was inaccurate, there were inaccuracies in the source maps, or different map projections were used.

Two approaches are used in registration: the adjustment of absolute positions and the adjustment of relative position. The term relative position refers to the location of features relative to other features. The term absolute position refers to the location of features in relation to a geographic coordinate system.

**Registration by Relative Position**

In this procedure, one data layer, termed the slave, is registered to a second data layer, termed the master. The relative position of the slave data layer is adjusted by choosing features that can be easily and precisely identified on both the data layers to be registered. A road intersection, the confluence of two streams, or a small island are
some examples of features that can be used for registration.

The locations of these features are then input to the GIS, a procedure that is usually done graphically and interactively. One commonly used procedure is to position a cursor on a display of the data layers or to use the cursor of a digitizer to identify the registration points. After the corresponding points have been entered for both data layers, the GIS then calculates a mathematical function to transform the coordinates of the slave into coordinates that more closely fit the master data layer. The operator is then usually presented with statistics indicating the quality of the registration, i.e. how well the data layers match. If the operator chooses to proceed, then the data layer is processed and new coordinate values are assigned to the features in the slave data layer.

This type of registration operation is often termed rubber sheeting. (The procedure is analogous to stretching one data layer, as if it were a rubber sheet, to fit another.) It is an empirical solution that makes few assumptions about the coordinate systems being used in the two data layers. The accuracy of this registration is predicted from the position errors of the registration points. It is a somewhat biased prediction because the transformation was calculated using these same points. Other areas of the map may not be as accurately registered. (In some systems, a set of points that was not used to calculate the transformation function can be used for accuracy assessment.) If the registration points can be accurately identified on both maps and the points are well distributed throughout the map, then this method of assessing registration accuracy does provide a reasonable prediction of the registration for the data layer.

Registration by Absolute Position

The other approach used to register data layers is to correct the absolute position of each data layer. In this case, each data layer is separately registered to the same geographic coordinate system (such as UTM coordinates). They should then be registered to each other. When data layers are registered by relative position, position errors in the master data layer are propagated to the slave data layers. The advantage of registering each data layer by absolute position is that this type of error propagation does not occur. Also, the accuracy of positions as represented on a data layer (i.e. the digital map) can be directly assessed with reference to ground coordinates. The disadvantage is that the small position errors that occur in each data layer will be independent, and so boundaries that should precisely overlay may be slightly misaligned. These discrepancies can be reconciled using an additional processing step termed conflation (discussed below).

TRANSFORMATIONS BETWEEN GEOMETRIC PROJECTIONS

The earth has a spherical shape. In order to uniquely reference locations on the earth’s surface, the system of latitude and longitude coordinates was developed. As shown in Figure 7.6, this system is based on the angles formed by a line drawn from the center of the earth to a point on the surface. (A more detailed explanation is given in the next section.)

The spherical surface of the earth can be easily represented on a spherical map, such as a globe, using this coordinate system. However, a globe is not as convenient to use as a flat 2-dimensional map. A map projection is a mathematical transformation that is used to represent a spherical surface on a flat map (Figure 7.7). The transformation assigns to each location on the spherical surface a unique location on the 2-dimensional map. However, this transformation cannot be done without some distortion. Map projections differ in the degree of distortion that
is introduced in the representation of area, shape, distance, and direction. The trade-offs in the degree and types of distortions that will be accepted should be considered in selecting a map projection. However, a more important selection criterion may be to use the projection commonly accepted for the discipline or application. Wherever possible, geographic information should be stored in the GIS so that it can be output in a form that is familiar to the user, i.e. a map should use the commonly accepted map projection for the discipline.

The data layers to be used together in a GIS should all be represented using the same coordinate system. A GIS commonly
Figure 7.1 Map Projections. A map projection is used to represent a spherical surface on a 2-dimensional map. (Adapted from maps provided by S. Prashker, Carleton University, Ottawa, Ontario.)
supports several projections and has software to transform data from one projection to another. The map projections most commonly used for mapping at scales of 1:500,000 or larger in North America is the UTM (Universal Transverse Mercator) projection. In the United States, the State Plane coordinate system is also used for regional scale mapping. For maps of continental extent, the Albers, Lambert's Azimuthal, and Polyconic projections are commonly used. Map projections tend to be standardized in a given field either by conscious decision or by tradition. It is important to ensure that the GIS can use the map projections in which the input data will be provided and map outputs are to be produced. Where more than one projection is to be used, then appropriate projection transformations should be provided as well.

The Latitude/Longitude System

Figure 7.6 illustrates the principles of the latitude/longitude system of geographic coordinates. Lines of longitude are drawn from the north pole to the south pole. The line of longitude passing through the Greenwich Observatory in England has the value of 0°. Moving west, the value of any line of longitude is the horizontal angle formed between the line drawn from that point to the center of the earth and a line drawn from the center of the earth to a point along the 0° line of longitude. Since these values are west of 0° longitude, they are termed west longitude values. In the diagram, point X is located at 45° west longitude. Similarly, lines of longitude east of 0° longitude are termed east longitude values. The two sets of longitude values meet at 180° longitude on the opposite side of the earth from 0°.

The lines of latitude are drawn perpendicular to the lines of longitude. In this case the lines of latitude are referenced to the equator designated as 0° latitude. The latitude of any point is defined as the vertical angle formed by a line drawn from the point to the earth's center and a line from the equator to the earth's center. Latitudes in the northern hemisphere are termed north latitudes, those in the southern hemisphere are south latitudes. In Figure 7.6, point X is located at 55° north latitude and 45° west longitude.

CONFLATION

Conflation is the procedure of reconciling the positions of corresponding features in different data layers. For example, consider two forest cover maps for the same area mapped in different years. Ideally, when these two maps are digitized and registered, features that had the same geographic position should precisely coincide when the two maps are overlayed. In practice, they may not precisely overlay because small errors were introduced during the input operation, the source maps were slightly different, the position of such features as streams had actually shifted slightly over the intervening years, or for other reasons.

Conflation functions are used to reconcile these differences so that the corresponding features overlay precisely. This is important when data from several data layers are used in an analysis. If the boundaries are slightly in error, "new" polygons, often termed slivers, may be created that do not represent information about the area but are rather inaccuracies in the mapping.

A manual procedure to minimize these errors is to re-draw the source maps using feature boundaries in one of the overlays as a standard or template. The representation of roads, streams, and lakes are commonly reconciled in this way. The template is drawn on a map that becomes the basemap which is then used to guide the redrafting of these features on all the maps. Templates are also used to standardize the shape of features with geographic positions that change but that are more conveniently
handled as having a fixed location over time. The position of the shoreline of a water reservoir will change over the year as water accumulates and is then released. If each data layer uses the position of the shoreline at the time of mapping, there will be discrepancies in its position on different data layers. To standardize the shoreline position, a standard template can be used to represent the shoreline at the same location on each data layer instead of using the actual shoreline position at the time of mapping. All mapping is then extended to the shoreline of the template. In this way, the representation of the reservoir is made to be consistent for all data layers.

Computerized techniques can be used to perform a similar function. The procedure usually requires the operator to identify corresponding features in the two maps or data layers that should have the same location. As more points are identified, one or both maps are adjusted. In some cases this procedure can be interactive so that the best match can be developed by adding points where discrepancies exist. Computerized conflation is a relatively recent development. It is an important function in applications where integrated analyses of data sets from diverse sources will be required. However, it is generally much less expensive to reconcile maps manually in the map preparation stage, before they are digitized. Reconciling maps that are already in the GIS data base incurs the added expenses of the computer system, the more highly trained and higher-paid operator, and the need to re-process the edited files.

**EDGE MATCHING**

Edge matching, illustrated in Figure 7.8, is a procedure to adjust the position of features that extend across map sheet boundaries. The coverage area stored in a GIS is usually larger than a single map sheet. Data entered from separate map sheets are usually organized within the GIS so as to present the data to the user in the form of a continuous geographic coverage. Theoretically, the data from adjacent map sheets should precisely meet at the map edges. However, in practice, features that extend across map boundaries often do not align perfectly. Minor errors can be caused by such factors as errors in the original mapping, differences in the dates of mapping, paper shrinkage in the

![Diagram showing edge matching](image)

**Figure 7.8** Edge Matching. Edge matching is used to reconcile the position of features that extend onto an adjacent map but are not correctly aligned at the map boundary.
source maps, or errors in the digitizing process. The discrepancies may be difficult to correct because the features along the map edge may not be shifted to the same degree or in the same direction. Some GISes provide software to reconcile these differences by making adjustments to the position of features in one or both maps. Edge matching must be done for the geographic information from several adjacent maps to be represented as a single continuous data layer.

Software differ considerably in the degree of automation provided. Beard and Chrisman (1986) review the approaches used in edge matching and provide an example of a highly automated procedure.

**EDITING FUNCTIONS**

Editing functions are used to add, delete, and change the geographic position of features. The sophistication of the editing software can greatly affect the speed and accuracy with which these essential functions can be done.

**Slivers or splinters** are thin polygons that are often created during digitizing and overlay operations. In Figure 7.9 polygons A and B share a common boundary. If the boundary between these polygons was mistakenly digitized twice a sliver would be created as shown. Polygons C and D represent features in two different data layers that share a common boundary. For example, a forest stand and an adjacent agricultural field will share a common boundary. If the boundary is not coded with precisely the same geographic coordinates in the two data layers, then, as shown in the Figure, a sliver may be created if they are used together in an overlay operation. Some software packages provide automated detection and correction of slivers or splinters. These functions can substantially reduce the editing time needed to remove slivers, but some features are correctly represented as long thin shapes. As a result some operator supervision is required to check that only the slivers are removed, not the valid features.

Often a digitized line is short by a few millimeters and does not quite reach the feature to which it is connected. Automated **line-snapping** can correct this type of error by connecting lines to a node if they end within a specified distance. Other editing aids will search for such inconsistencies as lines that are left dangling (i.e. are not connected to another element) or polygons that do not close. In some cases the program will aid the operator by moving the cursor to each inconsistency and waiting for the operator to correct the problem.

![Figure 7.9 Slivers. Slivers may be created during digitizing and overlay operations.](image)
LINE COORDINATE THINNING

This function is used to reduce the quantity of coordinate data that must be stored by the GIS. Often, more coordinates are entered than are actually needed to define a line or a polygon. This usually occurs during digitizing or scanning operations. Coordinate thinning, by reducing the number of coordinate points, reduces the size of the data file, thereby reducing the volume of data to be stored and processed in the GIS. Reducing the data volume will tend to improve system performance.

The thinning function reviews all the coordinate data in a file, identifies and then removes unnecessary coordinates. The degree of coordinate thinning is controlled by the operator. As the level of thinning is increased, fewer coordinates are stored. This results in a less detailed but more compact representation of lines (see Figure 7.10). In practice, the number of coordinate points can usually be significantly reduced without a perceived loss of detail.

MAINTENANCE AND ANALYSIS OF NON-SPATIAL ATTRIBUTE DATA

This group of functions is used to edit, check, and analyze the non-spatial attribute data. Many GIS analyses can be performed using these attribute functions alone. For example, in a vector-based GIS, the area and perimeter of polygons are commonly stored in the attribute file, along with the class, and other characteristics. To produce a table of areas for all polygons of a certain class, the data can be retrieved from the attribute file, without reference to the spatial data.

In a simple raster-based GIS, non-spatial attribute data may be embedded in the spatial data file. For example, a legend, the latitude and longitude coordinates of the corners of the data layer, and a title for the data layer might be attached to the beginning or end of the file.

Figure 7.10 Line Coordinate Thinning. Coordinate thinning reduces the number of coordinate pairs used to store a line segment within the GIS.

In more sophisticated systems, the attributes are stored separately from the spatial data, often in a separate data base system. The greater complexity of this type of GIS also can provide greater analysis flexibility and the power and capacity to handle large data sets.

ATTRIBUTE EDITING FUNCTIONS

Editing functions allow the attributes to be retrieved, examined, and changed. New
attributes can be added or old ones deleted. In some systems, it is difficult to add new categories of attributes once the data base has been defined. This is usually related to restrictions of the data base software. Where the attributes are not well-known in advance, as is the case with many natural resource applications, such restrictions may be an important system limitation. It is often valuable to import data from other sources. For example, in municipal applications, files containing ownership data keyed to addresses might be added to the attribute file for a street map so that the owner's name and address becomes an attribute of the spatial data. Many systems provide a function to match corresponding records in the two attribute data sets using a common data field. In this case the address. This capability is termed file matching or address matching.

**ATTRIBUTE QUERY FUNCTIONS**

Query functions retrieve records in the attribute data base according to conditions specified by the operator. Depending on the software, query functions can be restricted to very simple retrievals such as finding the class assigned to a selected polygon. A full-featured GIS will have a data base that supports more complex queries that involve selective searches of all the attributes for one or more data layers and the generation of a report that tabulates the results.

Figure 7.11 illustrates the use of an attribute query to generate a report of forest area by dominant species. The attributes for the forest cover data layer are stored as two tables. They can be queried together, as shown, to generate a report of the total area of forest more than 30 years old. The stand number, a data field common to both tables, is used as the link between the tables. (The procedure of retrieving information from the two tables by means of a common data field is termed a relational join operation.) The

**Figure 7.11** An Attribute Query Used to Produce a Forest Cover Summary Report. The figure illustrates the use of an attribute query to generate a report of forest cover by species for stands older than 30 years. The data are retrieved from two attributes tables by using a common data field, the stand number, to identify corresponding records in the two tables. This procedure is termed a relational join.

Information is summarized by dominant species in the report. Note that to generate the report, it was only necessary to query the attribute data files, the spatial data were not needed.

Attribute analyses can be very powerful and efficient because complex spatial operations are not used. A full-featured data base provides commands for an operator to perform a wide range of queries and to output
the results in the desired report format. In
many operational applications, the attribute
entry and retrieval functions are the prin-
cipal day-to-day operations. Such questions
as how large an area was planted to a spe-
cific crop, who is the owner of a specified
property, and what is the property value,
can be answered by operations on the attri-
bute data base alone.

INTEGRATED ANALYSIS OF
SPATIAL AND ATTRIBUTE DATA

The power of a GIS lies in its ability to
analyze spatial and attribute data together.
It is these capabilities that most distinguish a
GIS from automated mapping and comput-
eraided drafting systems. The range of analysis
procedures in this group of functions is very
large. They have been subdivided into four
categories: retrieval/classification/measurement,
overlay, neighbourhood, and connectivity or
network functions.

A specific GIS may implement a function
as a single computer task or by linking sev-
eral tasks together. The way a specific func-
tion is implemented affects the ease and flex-
bility of operation, the storage requirements
(e.g. for intermediate files), and the level of
performance. Though important for system
evaluations and comparisons, these consid-
erations are too detailed to be considered
here. For the purpose of understanding how
GIS analysis and manipulation functions are
used, the method of implementation is of
secondary consideration. The emphasis in
this section is on the conceptual procedures
used to generate different types of informa-
tion from the GIS data base.

RETRIEVAL, CLASSIFICATION,
AND MEASUREMENT FUNCTIONS

In this set of functions, spatial and attribute
data are retrieved, but only the attribute
data are modified or created. No changes
are made to the location of spatial elements
and no new spatial elements are created.

Retrieval Operations

Retrieval operations on the spatial and attri-
butte data involve the selective search,
manipulation, and output of data without the
need to modify the geographic location of
features or to create new spatial entities.
These operations work with the spatial ele-
ments as they were entered in the data
base. The production of a city map showing
buildings classified by their age is an exam-
ple of this type of operation (see
Figure 7.12).

Classification and Generalization

The sets of elements (the land parcels with
buildings of a certain age range) retrieved
in the previous example (Figure 7.12) could
be assigned class names such as 'pre-1900',
'1900–1930', '1931–1950', and 'post-
1950'. These class names could be stored
as attributes of the buildings in the data
base. This new class designation could then
be used to select these buildings for further
analyses.

This procedure of identifying a set of
features as belonging to a group is termed
classification. Some form of classification
function is provided in every GIS. In the case
of a single data layer, classification may
involve the assignment of a class name to
each polygon as an attribute. The classifica-
tion may be for land cover, and so the class
names might be 'forest land', 'agricultural
land', 'urban areas', and so on. In this case,
the classification process involves looking at
the attributes for a single data layer and
assigning an additional attribute, the new
class name. In a raster-based GIS, numerical
values are often used to indicate classes.
A cell might be assigned the value 1 to indi-
cate agricultural land, 2 for forest land, and
so on. The classification process would then
involve assigning numerical values to the
cells (sometimes termed recoding) and
writing these new values into a new data
layer.
Classification is important because it defines patterns. One of the important functions of a GIS is to assist in recognizing new patterns. These patterns might be areas of the city with the highest crime rate, areas of forest land suitable for timber harvest, or areas of agricultural land most likely to be converted to residential development.

Classification is done using single data layers, as illustrated in the building age classification, as well as with multiple data layers as part of an overlay operation (overlay operations are discussed later in this chapter). For example, a desirable site for a cottage might be a forested area, with well-drained soils, a southern exposure, and a non-agricultural land use zone. Each of these factors might be represented as separate data layers in the GIS. An overlay analysis could be used to identify the areas meeting these criteria and to assign them the class name “good cottage areas”.

Generalization, also called map dissolve, is the process of making a classification less detailed by combining classes (see Figure 7.13). Generalization is often used to reduce the level of classification detail to make an underlying pattern more apparent.

Measurement Functions

Every GIS provides some measurement functions. Spatial measurements include distances between points, lengths of lines, perimeters and areas of polygons, and the size of a group of cells with the same class. Sample applications are: finding all forest areas larger than 200 sq km that are potentially suitable for use as a conservation area or locating airports less than 10 km apart that might be unnecessary. Where digital terrain data are used, 3-dimensional measurements are often needed for engineering applications, such as calculating the amount of cut and fill material involved in road construction. (Measurements along a network, such as along a road system, are a special case discussed in the section on network functions.)

In a vector-based GIS, spatial elements can vary in size and shape. New attributes can be calculated for the spatial elements...
of an overlay. The calculation of areas and centroids (the center point of a polygon) is commonly done automatically as part of the polygon creation process. Measures of shape, narrowest and broadest distance across a polygon, the length and sinuosity of a line are other useful measurement functions. In a raster-based GIS, these types of functions become neighbourhood operations because they involve the identification of connected cells. As a result, the software algorithms are different and the strategy for using them is conceptually different as well.

**OVERLAY OPERATIONS**

Arithmetic and logical GIS overlay operations are part of all GIS software packages. Arithmetic overlay includes such operations as addition, subtraction, division, and multiplication of each value in a data layer by the value in the corresponding location in a second data layer. A logical overlay involves finding those areas where a specified set of conditions occur (or do not occur) together.

For example, desirable areas for cottages might be defined as those areas that have a forest vegetation cover, have well-drained soils, and have a south-facing exposure. If vegetation, soils, and exposure are represented as separate data layers in the GIS, then a logical overlay operation could be used to identify the locations where these conditions occur together.

The flexibility provided to the operator and the level of performance of overlay operations vary widely among GISes. One of the major factors affecting the performance of these functions is the data model being used. Raster and vector models differ significantly in the way arithmetic and logical operations are implemented. Overlay operations are usually performed more efficiently in raster-based systems. Because this has been a critical issue, the difference in approach will be illustrated by the following examples.

Figure 7.14 illustrates an arithmetic function, multiplication, applied to rain gauge data. The example shows a raster and a vector implementation. The procedure is being used to convert the data from units of inches to units in millimeters by multiplying each rain gauge value by 25.4 mm/inch.

In the raster case, the rain gauge data for the five locations are entered directly into a data layer (the input data layer in the Figure). The cells with no rain gauge data are shown to be blank in the figure for clarity.
Figure 7.14 An Arithmetic Operation on a Single Data Layer in the Raster and Vector Domains.

however each cell would be assigned a value (commonly these blanks would be zeroes). The multiplication operation is performed on every cell of the input data layer and the result is written to the corresponding cells of the output data layer. The operation is performed as many times as there are cells, regardless of the fact that most cells are blank.

In the vector case, the rain gauge locations are represented as points in the input data layer. The attributes of the points are stored separately in an attribute table, as illustrated. The attribute data consist of only the five rain gauge values and the multiplication procedure can be done using the attribute data alone. In the Figure, a second column containing the calculated millimeter values is shown to have been added to the attribute table. Only five multiplications were needed. For sparse data, this operation on one data layer is much more efficiently processed in the vector domain than in the raster domain.

Figure 7.15 illustrates an arithmetic function of two data layers as implemented in the raster domain. The value at each location in Input Data Layer A is to be added to the value in the corresponding location in Input Data Layer B. The result is to be written to the corresponding cell of the output data layer. Because the map area is regularly subdivided into cells and the cells
of each data layer are in perfect registration, there is no uncertainty in the location of boundaries since they are the cell boundaries and always coincide. The value assigned to a cell represents the condition for all points within the cell, i.e., the cell is the smallest unit of division.

Figure 7.16 illustrates the same operation in the vector domain. Here areas are represented as polygons. Each polygon can differ in size and shape and so the boundaries of the polygons in one data layer usually will not coincide with the boundaries of polygons in the other data layer. The polygon values are stored separately in attribute tables, as shown. The value assigned to a polygon represents the condition for every location within the polygon, as for a raster cell. However, the size of the polygon is variable. In order to add the two polygon data layers, new polygons must first be created in the output data layer. The area where polygons A and D overlap must be defined as a separate new polygon and then assigned the new attribute value. This process of subdividing polygons is termed clipping. Although there will be fewer arithmetic operations, because there are fewer polygons than cells, clipping makes overlay operations considerably more complex in the vector domain than in the raster domain. When a large number of irregularly shaped polygons are involved, the vector overlay procedure generally requires significantly more processing time than raster overlay.

In the raster domain, the data file consists of an ordered list of values. The spatial locations to which these values refer (i.e., the row and column positions) are determined by

![Figure 7.16](image)

Figure 7.16 Arithmetic Operations on Two Data Layers in the Vector Domain. To perform an overlay operation in the vector domain, polygons A and D in the input data layers must be subdivided or clipped to create the new boundaries for the output polygons (F,G,H,I). Then the addition operation is performed using the values from the attribute table. The clipping operation is complex and can require considerable processing time when there are large numbers of irregularly shaped polygons.
the position of the value within the file. As a result, overlay processing involves retrieving and comparing the data in corresponding positions in the data files. There is no need to calculate intersections of boundaries or make any modifications to the feature boundaries because each spatial element is a single cell of standard size.

The regular subdivision of space makes overlay operations easy to implement in the raster domain. However, sparse data sets require as much processing as densely populated ones because the operation is performed on every cell regardless of the data it contains. In the vector domain only the data of interest are processed. The more sparse the data set, the faster the processing. However, the overlay operation is much more complex to begin with. Where operations are performed on the attribute data for a single data layer, the storage of attributes separate from the spatial data is advantageous. A single data layer can have an attribute table assigning multiple attributes to each spatial element. Operations on these attributes are effectively independent of the spatial data.

In many GISes a hybrid approach is used that takes advantage of the capabilities of both data models. A vector-based system may implement some functions in the raster domain by performing a vector-to-raster conversion on the input data, doing the processing as a raster operation, and converting the raster result back to a vector file. In a raster-based GIS, data compression techniques can be used in effect to create elements that represent contiguous areas (a connected group of cells) that have the same value. (See the discussion of data compression in Chapter 6.) There is a trend to introduce more integrated raster-vector processing in GIS because the two data models offer different advantages and because many digital data sets are available in only one of the two formats.

**NEIGHBOURHOOD OPERATIONS**

Neighbourhood operations evaluate the characteristics of the area surrounding a specified location. Counting the number of residential dwellings within a 5 km radius of a fire station (Figure 7.17) is an example of a neighbourhood operation. Every neighbourhood function requires the specification of at least three basic parameters: one or more target locations, a specification of the neighbourhood around each target, and a function to be performed on the elements within the neighbourhood. In this example, the target is the fire station, the specified neighbourhood is the area within a 5 km radius, and the function is to count the number of elements that are residential buildings.

Virtually all GIS software packages provide some form of neighbourhood operations. They vary in the flexibility and sophistication with which the three basic parameters can be specified and in the specialized operations provided. The most common types are the search function (which includes the generation of summary statistics), topographic functions, and interpolation.

Neighbourhood operations are particularly valuable in evaluating the character of a local area, for example to find all residential areas that are in the vicinity of a school.

![Figure 7.17 Defining a Search Area With a 5 km Radius.](image-url)
a park, and a shopping area. Another example is the assessment of the quality of an area as wildlife habitat. Wildlife require a specific combination of vegetation and terrain types with access to water, food, and shelter within a limited size of neighbourhood.

Many neighbourhood operations must be implemented using some regular division of the study area. For this reason, a raster model is commonly used. In some vector-based systems, the geographic data is converted into a raster form for the analysis and then converted back to vector format. Other vector systems generate the neighbourhoods directly and intersect them with the geographic data as an overlay operation.

Search

The search function is one of the most common of the neighbourhood operations. This function assigns a value to each target feature (such as the fire stations in the previous example) according to some characteristic of its neighbourhood. (The search function, as presented here, is defined in a general way. GISes will differ in the names used to refer to this function, and several individual procedures may be needed to provide some of the search operations discussed.)

The three basic parameters to be defined in a neighbourhood search are the targets, the neighbourhood, and the function to be applied to the neighbourhood to generate the neighbourhood value. The target elements and the elements of the neighbourhood are commonly stored in one or more data layers. In the previous example, the locations of the fire stations and residential buildings could be in a single data layer of the buildings in the municipality or in separate data layers, one containing the locations of all emergency facilities and the other containing the data on housing.

The search area is most commonly square, rectangular, or circular with a size selected by the operator (e.g. a circular area with a 5 km radius). The search functions are usually predefined and the operator selects one of those offered. Search functions are of two types: those that operate on numerical data and those that operate on thematic data. Typical numerical functions are the total, average, maximum, minimum, and measures of diversity (see Table 7.1). In each case, the function is applied to the corresponding neighbourhood for each target.

<table>
<thead>
<tr>
<th>Table 7.1 Neighbourhood Functions.</th>
</tr>
</thead>
<tbody>
<tr>
<td>FUNCTION</td>
</tr>
<tr>
<td>AVERAGE</td>
</tr>
<tr>
<td>DIVERSITY</td>
</tr>
<tr>
<td>MAJORITY</td>
</tr>
<tr>
<td>MAXIMUM, MINIMUM</td>
</tr>
<tr>
<td>TOTAL</td>
</tr>
</tbody>
</table>

The data layer used to define the neighbourhood may be different from the data layer to which the function is applied. For example, consider an analysis of the average dollar value of residential buildings within 5 km of the fire stations in a metropolitan area. The target locations (fire stations) may be identified on a data layer of emergency service facilities, the residential buildings identified on a second data layer, and the values of each residential building stored in a separate attribute file. The neighbourhood would be defined using the first data layer but would be applied to the data of the
second data layer — to identify the residential buildings within the 5 km search radius. The values of these residences would then be retrieved from the attribute file and averaged. The result of this analysis could be provided in the form of a map showing the fire stations and their 5 km surrounding neighbourhoods shaded according to the average value of the residences in their neighbourhoods. The results could also be output as a table listing each fire station and the average value of the residences in each neighbourhood.

Neighbourhood functions using thematic data commonly use the following operators: the majority (also termed the mode), the maximum, the minimum, and diversity measures (such as the number of different classes in the neighbourhood).

In more sophisticated implementations of the search function, more flexibility in the specification of the three basic control parameters is provided. Instead of using only predefined functions, the operator may be able to enter an equation. The selection of the target points or target areas might be defined using a query such as "select hospitals that have more than 500 beds and are less than 5 years old". This type of query might use data from both the attribute and spatial data bases to generate a list of locations that would then be used by the search routine.

In many applications, it is useful to specify a neighbourhood area that may be different for each target and may not be a regular shape. The neighbourhood might be a political region like a county or state. The neighbourhood might also be generated by other GIS functions, such as an overlay operation. The application of a function to a user-defined neighbourhood is termed region of interest processing. The interactive definition of a search area is often termed windowing, and the search area itself is termed a window.

Consider the example in Figure 7.18 in which the boundaries from the county data layer are used to define a processing window for analyzing land uses in County 25. Note that the land use polygons and county polygons do not have the same boundaries. In effect, the county polygon has been used like a "cookie-cutter" to extract the corresponding portions of the land use polygons. Then the area for each type of land use in

![Figure 7.18 An Overlay Operation Used to Define a Region of Interest. A land use report for County 25 is generated by overlaying the County and Land Use data layers.](image-url)
the county has been totalled and reported in tabular form.

The polygons used as windows can be quite complex. A network function might be used to define areas within a 20 minute ambulance ride from a hospital, as shown in Figure 7.19. In this case, the window might be quite irregular because it would depend on the traffic capacity and speed of travel along the surrounding streets. Once the area has been defined, it could be used to further analyze that service area, such as to retrieve information about the residents of the area. In this example, a report is shown of the population age structure within the 20 minute service area.

In some cases, particularly in a raster-based GIS, search functions are applied to every location (i.e. every cell is a target). The processing is then in effect like moving a window the size of the neighbourhood through the data layer, cell-by-cell. At each step the neighbourhood function is evaluated and the result assigned to the corresponding position in the output data layer, as illustrated in Figure 7.20. Here the objective was to find all locations surrounded by at least a 3 km by 3 km area of forested land. Since each cell represents a 1 km by 1 km area, a 3 cell by 3 cell window was used. This window was applied at each cell location in the input data layer. (No values are calculated for the cells at the edge of the data layer because some of the cells of the neighbourhood would be off the map.) The function applied to each neighbourhood is to count the number of non-forested cells (shown here in black) and assign that value to the cell in the output data layer that has the same position as the center cell of the window.

Line-in-Polygon and Point-in-Polygon Operations

In a vector-based GIS, the identification of points and lines contained within a polygon area is a specialized search function. In a raster-based GIS, it is essentially an overlay operation, with the polygons in one data layer and the points and/or lines in a second data layer (see the section on overlay operations earlier in this chapter). Quadtree-based systems also deal with this type of function in a specialized way. Using a quadtree data structure, this type of search operation can be implemented very efficiently, usually with performance levels significantly higher than vector- or raster-based systems can provide.

Figure 7.21 is a map showing the major highways (lines) crossing a metropolitan area

![AMBULANCE SERVICE AREA](image)

**Figure 7.19** Twenty Minute Ambulance Service Area. To define the areas that can be reached by ambulance within twenty minutes, a network function is used to evaluate travel times along the street network.
and the location of three repair depots (points P1, P2, and P3). A simple point-in-polygon operation might be to find all repair depots in the metropolitan area. This type of search would require that the coordinates of all repair depots be evaluated and those that fall within the polygon representing the city limit be reported.

A search for all highways that cross the city is an example of a line-in-polygon operation. At first glance, it would appear that this search could be done as a point-in-polygon operation by searching for points that comprise the line and fall inside the boundary of the polygon. However, in a vector-based GIS, lines or arcs are stored as a series of points connected by straight-line segments. The line itself is only created when needed, such as when a map is produced. If a straight line crosses a polygon it may be represented by a single line segment with both endpoints outside the polygon.

In Figure 7.21, the black dots indicate the points in the line for which coordinates are actually recorded. Notice that Highway 80 has no coordinate points within the city limit, which is the search polygon. To identify all highways crossing the city, a search for coordinates that fall within the polygon would miss Highway 80. For this reason more sophisticated algorithms are used that can properly handle these cases.

A more challenging example is the map enquiry system used by the US Geological
sample points are commonly organized as a grid of points, essentially a raster form of organization.

An alternative form of representation is the Triangulated Irregular Network or TIN used in vector-based systems. In a TIN, a network of triangular facets is generated by the GIS from a set of elevation sample points that can be irregularly distributed. These facets can then be manipulated as polygons and the elevation, slope, aspect, and other parameters can be assigned to the facets as polygon attributes. Raster and vector data models and TINs are discussed in Chapter 6.

Topography can be used to study data other than elevation. Any characteristic that has a continuously changing value over an area can be represented as a surface. In geology, aeromagnetic data and geochemical data are often represented as a surface. The noise levels in the vicinity of an airport, the income levels of neighbourhoods in a city, or the levels of pollution in a lake can also be represented in this way.

Topographic functions are used to calculate values that describe the topography at a specific geographic location (e.g. the elevation at the location) or in the vicinity of the location (e.g. the slope of the area immediately surrounding the location). Most topographic functions use a neighbourhood to characterize the local terrain. The two most commonly used terrain parameters are the slope and aspect, which are calculated using the elevation data of the neighbouring points.

Slope is defined as the rate of change of elevation. Aspect is the direction that a surface faces. Conceptually, the calculation of the slope and aspect at a point can be thought of as fitting a plane to the elevation values of the neighbouring points. The slope of this plane and the direction it faces are the slope and aspect of the point. As shown in Figure 7.22, the slope may be calculated in the X-direction, the Y-direction, or in the direction of maximum slope. The maximum slope is termed the gradient.

Topographic Functions

Topography refers to the surface characteristics, i.e. the relief, of an area. The topography of a land area refers to the hills, valleys, and plains of which it is comprised. The topography is thus defined by the elevation of each location within the area. The topography of a land surface can be represented in a GIS by digital elevation data. This data set consists of the elevation of a large number of sample points distributed throughout the area being represented. These
Figure 7.22 Measurement of Slope. The slope at a point is the angle measured from the horizontal to a plane tangent to the surface at that point. The value of the slope will depend on the direction in which it is measured. Slope is commonly measured in the direction of the coordinate axes e.g. in the X-direction and Y-directions. The slope measured in the direction at which it is a maximum is termed the gradient.

Slope is usually measured in degrees of arc or as a percentage (the change in elevation divided by the corresponding horizontal distance). Aspect is defined by the horizontal and vertical angles that the surface faces. The horizontal angle is usually measured in degrees of azimuth, the angle formed by moving clockwise from north, as shown in Figure 7.23. The vertical angle or angle of elevation, the positive angle measured from the horizontal to a line drawn perpendicular to the surface, is sometimes used as well. This angle is equal to 90° minus the gradient.

Although commonly used in the analysis of elevation data, slope and aspect can be usefully applied to other data sets as well. Slope measurements are commonly used in the analysis of gravity and aeromagnetic data in geology. In an urban setting, slope values could be calculated for land costs. High values of slope would then indicate areas where land costs change abruptly with distance. Such areas might represent zones of potential social conflict, or they might also indicate areas with good investment potential.

Figure 7.23 Measurement of Aspect. There are two components in the measurement of aspect. The horizontal angle or azimuth is the aspect direction. It is the angle formed by moving clockwise from north to the direction of maximum slope. The vertical angle or elevation angle is measured from the horizontal to a line drawn perpendicular to the surface.
Other important functions used in topographic analyses are illumination, viewshed modelling, and perspective view generation. These are discussed in subsequent sections.

**Thiessen Polygons**

Thiessen or voronoi polygons define individual areas of influence around each of a set of points. Data from rain gauges are commonly analyzed in this way. It is an approach to extending point information which assumes that the "best" information for locations with no observations is the value at the closest point with a known value. Thiessen polygons are commonly used in the analysis of climatic data, such as rain gauge data. In the absence of a local observation, the data from the nearest weather station are used.

Thiessen polygons are constructed around a set of points in such a way that the polygon boundaries are equidistant from the neighbouring points. In other words, each location within a polygon is closer to its contained point than to any other point.

Figure 7.24 illustrates the use of Thiessen polygons to analyze rain gauge data. The rain gauge locations are represented in the GIS as points. Thiessen polygons are then generated around each point and the rainfall value for the rain gauge is assigned to its surrounding polygon. The rainfall at all locations within each Thiessen polygon is considered to be that of the contained rain gauge station. The amount of rain falling on each polygon can then be calculated as the amount recorded by the rain gauge multiplied by the area of the polygon. For the entire study area the total rainfall would be estimated by totalling the rainfall calculated for each Thiessen polygon.

Thiessen polygons, in effect, are used to predict the values at surrounding points from a single point observation. The method has a number of limitations. The division of a region into Thiessen polygons is completely dependent on the location of the observation points. This can produce polygons with shapes that are quite unrelated to the phenomenon being mapped. The position of rain gauges may produce long thin polygons, a pattern in which rainfall would not normally occur. The value assigned to each polygon is estimated from a sample of one, the observation point. Estimates of error cannot be calculated from a single sample. Finally, Thiessen polygons do not assume that points close together are more similar than points far apart, an assumption that is usually appropriate in geographic analyses (Burrough 1986).

**Interpolation**

Interpolation is the procedure of predicting unknown values using the known values at neighbouring locations. The neighbouring points may be regularly or irregularly spaced. Figure 7.25 is a simple example of this function, in this case presented using raster data layers (the function can also be implemented using TINs in a vector-based system). The stippled cells contain the known values. A simple linear function, derived by analyzing the known points, has been used to generate the missing values.

Interpolation programs employ a range of methods to predict unknown values, including polynomial regression, Fourier series, splines, moving averages, and kriging.

The quality of the interpolation results depends on the accuracy, number, and distribution of the known points used in the calculation and on how well the mathematical function correctly models the phenomenon. Interpolation assumes that the phenomenon being predicted (e.g. terrain elevation) is closely approximated by the mathematical function used. The unknown values are then calculated according to this function. The choice of an appropriate model is therefore essential in order to obtain reasonable results.
**Figure 7.24** Thiessen Polygons Used to Analyze Rain Gauge Data. A Thiessen polygon is constructed around each rain gauge location. The rainfall value for every location within the polygon is considered to be that of the contained rain gauge station.

**Figure 7.25** Interpolation. Interpolation is the procedure of predicting unknown values using the known values at neighbouring locations, usually by means of a mathematical function.
The best results are obtained when the mathematical function behaves in a manner similar to the phenomenon. For example, air temperatures over a uniform landscape would be expected to have gradual changes. An algorithm that interpolates temperature values by averaging the changes smoothly over distance might be appropriate. However, for elevation data in rough terrain, such an algorithm would tend to smooth abrupt elevation changes, such as steep cliffs and ridges, features that should be retained. Alternative functions that preserve abrupt changes can be used to overcome these types of problems. Of course, more complex algorithms tend to exact higher processing costs.

Many different mathematical models are used for interpolation. Some of the more common ones used for elevation data are the average of the neighbouring values or a weighted average where the weights are inversely proportional to the distance. More complex operations, such as kriging, take into account the trend of the known values in the vicinity of each point to be predicted.

A detailed discussion of interpolation is beyond the scope of this work. It is important to realize that in calculating missing values from neighbouring points, the data are assumed to behave in a spatially predictable manner over the map area. As the assumptions of the model are more severely violated, the interpolation results will become less accurate. For this reason, the quality of the interpolation result should be reported as well as the results.

Contour Generation

Contour lines are used to portray surface relief as a set of lines that connect points of the same value. In a topographic map, perhaps the most familiar application, contour lines connect points with the same elevation value. Contours are routinely used to portray a wide range of spatial data sets that can be represented as a surface: crime rates, housing values, geochemistry, wildlife population counts, and climatic data are a few examples. Although contour generation is used to produce output products, it has been included here because of the importance of interpolation in its implementation.

The process of generating contours is a more involved procedure than connecting data points with the same value. It also requires that predictions be made for missing values. Contouring functions make use of interpolation routines to generate those missing data points. Figure 7.26 shows how a single set of data points can produce two different patterns of contour lines. Although both results are reasonable, they differ in the way the elevation values have been interpolated at key locations to generate the contours. The topography that can be inferred from the 30 m data points is ambiguous. As a result, there is more than one reasonable solution. In Result A, no additional 30 m data points were produced. As shown in the perspective view, the contours portray a surface with a north-south ridge. In Result B, additional 30 m points were generated, and the resulting contours portray an area with two hills separated by an east-west gap.

Software packages will differ in the way they handle ambiguous conditions such as this. Often the results are judged by comparison with the way a cartographer would have drawn the contours, with the one that more closely matches the hand drawn version being deemed the "better" solution. The evaluation of contouring software must, therefore, take into account the expectations of the users. Users that have become accustomed to a particular style of contouring may judge a different, though equally correct style, to be unacceptable.

CONNECTIVITY FUNCTIONS

The distinguishing feature of connectivity operations is that they use functions that accumulate values over the area being traversed. That is, they require that one or more attributes be evaluated and a running
Figure 7.26 Different Interpolation Results Produced From the Same Set of Data Points. There can be more than one solution to the interpolation of a set of values. The figure illustrates two different results that could reasonably be obtained from the same set of data points. Each result is shown as a contour map and in perspective view. Though both solutions are technically correct, only one solution correctly depicts the actual landscape which these data describe.
total of the results be retained in a step-by-step fashion. Each step represents a movement in space, such as a 100 m segment along a street. The running total may be quantitative, such as the accumulated distance travelled or the accumulated travel time. The running total can also be qualitative, such as whether a point is or is not still visible. Every connectivity function must include the following:

1. a specification of the way spatial elements (such as roads) are interconnected;
2. a set of rules that specify the allowed movement along these interconnections;
3. a unit of measurement.

Consider, for example, a connectivity function for travelling along city streets. A street map could be used to define the way the elements (the streets) are interconnected. The rules for movement might include observing one-way streets and speed limits. The measurement unit might be distance, e.g. find the travel distance between two points along the street network. To find the distance along the street network, the route must first be defined and then the measurement of each segment progressively totalled. This is a considerably more complex problem than calculating the straight-line distance between two points, which can be done as a single calculation using the coordinates of the start and destination points.

Connectivity functions provide considerable flexibility in defining units of measure. Often the unit of measurement is not distance. In the travel example, the travel time might be of greater interest. Depending on the sophistication of the software package, such factors as the traffic flow at different times of day and automated route-finding might also be provided. Then an automated route selection procedure could map out the optimum route (i.e. the route with the least travel time) to go from a given starting location to any specified destination.

Software packages vary considerably in the connectivity functions they provide and the algorithms used to implement them. Both vector- and raster-based methods are used. The approach used in a specific software package depends on the data model best suited to the problem and the format in which the data are stored in the GIS. Conversions between raster and vector data structures are sometimes used when a function is more easily implemented in a data structure other than the one used for storage. In the following discussion, connectivity functions have been grouped into the following categories: contiguity, proximity, network, spread, stream, and intervisibility.

Contiguity Measures

Contiguity measures evaluate characteristics of spatial units that are connected. A contiguous area consists of a group of spatial units that share one or more specified characteristics and form a unit, as illustrated in Figure 7.27. The contiguous areas may be specified to have unbroken adjacency, i.e. no gaps are permitted. In other cases, gaps may be allowed, e.g. when a green space is considered to be contiguous even though it is crossed by a road.

The definition of unbroken may change with the application. A corn field containing an unplanted area may still be considered one contiguous unit even though the field area would include both the planted and unplanted portions. In other cases, only the planted part might be considered the contiguous area.

Commonly used measures of contiguity are the size of the contiguous area and the shortest and longest straight-line distances across the area. A common application of these measures is to identify areas of terrain with specified size and shape characteristics. For example, a search for a land unit to be used as a park might be specified as a contiguous land unit of forest having a minimum
area of 1000 square km with no section narrower than 20 km.

Proximity

Proximity is a measure of the distance between features. It is most commonly measured in units of length but can be measured in other units, such as travel time or noise levels. Four parameters must be specified to measure proximity: the target locations (e.g. a road, a hospital, a park), a unit of measure (e.g. distance in metres, travel time in minutes), a function to calculate proximity (e.g. straight-line distance, travel time), and the area to be analyzed.

Figure 7.28 illustrates one application of a proximity function. In this example, the 300 ft buffer zone drawn around the dirt roads defines the forest area where logging is not permitted. This type of proximity analysis is often called buffer zone generation. A buffer zone is an area of a specified width drawn around one or more map elements. A report can be generated of the forest stands within the buffer zone and the area unavailable for harvest by overlaying the buffer zones on a forest cover map, as illustrated in Figure 7.29. The results of this analysis can be reported in tabular form as shown in Table 7.2.

More complex proximity analyses may require values to be calculated for a large number of point locations and may also involve overlay operations with multiple data layers. For example, the noise level from an airport would be expected to decrease with distance. This could be analyzed using a proximity function to calculate the distance of each location from the sound source. A mathematical model of sound propagation would be used to calculate the reduction in sound level for each increment of distance. Then, using the mathematical model together with the proximity data, a separate data layer could be generated showing the expected sound levels at each geographic location.

Figure 7.30 illustrates the results of this type of GIS analysis for noise levels. The map shows the expected noise levels for the area surrounding a proposed airport. The outermost contour represents the 65 LDN noise level. The inner contours represent higher noise levels. This map was produced by computing the noise levels expected when each type of aircraft landed. The

---

Figure 7.27 Contiguous Areas. The area of 1s would be a single contiguous area if corner-to-corner adjacencies are considered contiguous. If contiguity were limited to edge-to-edge connections, there would be two contiguous areas of 1s. The park, field, and bird sanctuary are grouped as a single contiguous green space area. In this case, the road crossing the area is treated as an allowable gap.
Figure 7.28 Buffer Zone Generation. A 300 ft buffer zone around the dirt roads defines the forest areas where logging is not permitted. (Courtesy of ESRI, Redlands, California.)
results of this analysis were overlayed with the land use data layer to assess the area of residential land that would be affected.

By evaluating different landing patterns and aircraft types, an aircraft traffic schedule could be developed that would minimize the amount of residential land exposed to high noise levels.

Most GIS packages include some form of proximity function. They differ widely in the sophistication of the analyses that can be performed. One fundamental distinction among proximity functions is whether they can generate an accumulation surface (see discussion of spread functions). An accumulation surface is generated by moving outward incrementally from a target. At each step, a function (such as a travel time calculation) is applied for that geographic location and added to a running total. The value in the output data layer for that geographic location is set to the current total. Then processing moves to the next increment where the process is repeated. The values in the output data layer are thus totals, each of which have been accumulated over the previous incremental steps. This type of proximity function is sufficiently different that it is discussed later as the Spread function.

Network Functions

A network is a set of interconnected linear features that form a pattern or framework.

<table>
<thead>
<tr>
<th>MANAGEMENT UNIT</th>
<th>STAND NUMBER</th>
<th>STAND AREA (acres)</th>
<th>BUFFER AREA (acres)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>327.7</td>
<td>84.8</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>92.1</td>
<td>16.9</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>42.9</td>
<td>2.8</td>
</tr>
<tr>
<td>TOTAL</td>
<td></td>
<td>462.7</td>
<td>106.5</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>19.7</td>
<td>0.0</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>97.1</td>
<td>43.4</td>
</tr>
<tr>
<td>2</td>
<td>12</td>
<td>87.1</td>
<td>24.7</td>
</tr>
<tr>
<td>2</td>
<td>23</td>
<td>69.5</td>
<td>0.9</td>
</tr>
<tr>
<td>2</td>
<td>24</td>
<td>57.0</td>
<td>21.4</td>
</tr>
<tr>
<td>TOTAL</td>
<td></td>
<td>330.4</td>
<td>90.4</td>
</tr>
</tbody>
</table>
Networks are commonly used for moving resources from one location to another. A city’s streets, a grid of power transmission lines, an airline’s service routes, or the streams of a drainage basin are familiar examples of networks.

A GIS is used to perform three principal types of network analyses: prediction of network loading, route optimization, and resource allocation. The transport of water and sediment in a river system can be predicted using a network model. When
several storms occur in a region the effects of the increased stream flow can be complex. By correctly predicting the flows of water through the stream network, the magnitude and location of a flood can be predicted in advance so that emergency services can be prepared. Route optimization applications range from emergency routing of ambulances, fire, and police vehicles to airline scheduling and the routing of bus services, mail delivery, and municipal garbage collection. A common resource allocation application is the division of a metropolitan area into zones that can be efficiently serviced by individual police and fire stations. For example, an area to be policed might be divided into zones that can be patrolled in equal amounts of time. In the upper diagram of Figure 7.31 the solid line defines an optimum route through a network of streets. In the lower diagram the street network has been divided into two zones each of which can be patrolled in the same amount of time.

Networks have unique properties that require special analysis functions. The resources to be transported are usually dispersed throughout the network. For example, traffic is dispersed throughout city streets, power is dispersed throughout the power grid, water and sediment are dispersed throughout the stream network. The links within a network have characteristics that determine the type of resources and the conditions under which they can be transported. For example, some city streets are one way, some are closed to trucks, and streets have different speed limits and capacities — as we are reminded when caught in rush-hour traffic.

Network analyses usually involve four components:

1. a set of resources (such as goods to be delivered);
2. one or more locations where the resources are located (such as the warehouse where the goods are stored);
3. an objective, to deliver the resources to a set of destinations (such as the location of the customers who will receive the goods) or to provide a minimum level of service to an area (such as a police patrol zone);
4. a set of constraints that places limits on how the objective can be met (such as the maximum speed that vehicles can travel).

Figure 7.31 Network Functions for Routing and Resource Allocation. Network functions can be used to optimize vehicle routing and to divide an area into service districts to optimize the allocation of resources.
The network functions of a GIS are used to simulate the behavior of networks that would be too difficult, expensive, or impossible to measure. In a network model, the elements of the actual network are represented by a set of rules (such as the permitted direction of travel along a street) and mathematical relationships (such as the power loss along a transmission line as a function of distance). The more sophisticated the network functions and the representation of the network, the more closely the behavior of the model can be made to mimic reality.

There is considerable variability in the network capabilities of commercially available GIS software. They differ in the size and complexity of network model that can be defined, the level of performance, and the degree of interactive control.

### Spread Functions

The spread function is a very general yet powerful operation that can be used to analyze a wide range of phenomena. It can be used to evaluate transportation time or cost over a complex surface. It can also be used to define drainage basins (e.g. by spreading out from a point and allowing movement only to adjacent cells with the same or higher elevation).

The spread function has characteristics of both network and proximity functions. A spread function evaluates phenomena that accumulate with distance. Its operation can be thought of as moving step-by-step outward in all directions from one or more starting points and calculating a variable, such as travel time, at each successive step.

The distinguishing feature of a spread function is that a running total is kept of the function being evaluated. It is the value of the running total at each location that is written to the output file. (The output is sometimes termed an accumulation surface or friction surface.) In a simple case, the accumulated value may be the straight-line distance from the starting point. (This is the case in a proximity analysis, which is, in effect, a special case of the more general spread function.)

In more complex applications, the accumulated value may represent travel time and take into account multiple constraint factors. Because values are accumulated incrementally over relatively small spatial units, constraint factors with irregular spatial distributions can be accommodated. For example, movement may be constrained by partial barriers that reduce the rate of movement or by absolute barriers that stop.

In Figure 7.32, a spread function has been used to evaluate the travel distance from a target cell. The distance was accumulated by stepping outward one unit in every direction. All the cells adjacent and in line with the target were assigned the value 1, indicating 1 unit of distance (e.g. cell A). Cells diagonally adjacent to the target were assigned a value of 1.4 (the diagonal distance between the centers of the two cells, e.g. cell B). Cell C can be reached by moving diagonally across one cell and up one cell for a total distance of 2.4 units. It can also be reached by moving up two cells and across one cell for a total distance of 3 units.

The spread function assigns the shorter...
distance in the case of multiple routings. This procedure is used to calculate a value for each cell.

Rather than writing in all the values for each cell, the spread function results will be represented by contours. In Figure 7.33, the contours represent distances in km away from the starting point at A. The shortest distance from location A to location B is given by the dashed straight line connecting them. In this case, the function is essentially the same as the proximity function discussed previously.

Figure 7.34 shows the effect of an absolute barrier, i.e. one that does not allow any movement across it. A lake would be an absolute barrier to truck travel, for example. As before, a spread function has been used to generate the travel distances. Now the shortest travel distance is not the straight line connecting locations A and B because the route must go around the obstacle. The spread function evaluated the distances incrementally and accumulated the distance to go around the obstacle. This type of analysis involving obstacles cannot be accommodated by proximity functions.

In Figure 7.35, the effect of a partial barrier is illustrated. A partial barrier, such as rough terrain, impedes progress but does not stop travel. Instead of measuring distance, the spread function is used here to

---

**Figure 7.33** Travel Zones Defined Using a Spread Function. Equidistant travel zones in 1 km increments from the target (A) are indicated by the concentric rings. The shortest travel distance from A to B is shown by the dashed line.

**Figure 7.34** The Effect of an Absolute Barrier on Travel Zones Defined Using a Spread Function. Travel zones in 1 km increments from the target (A) are defined by the travel distance contours. The shortest travel distance from A to B must follow a route around the barrier as traced by the dashed line.

**Figure 7.35** The Effect of a Partial Barrier on Travel Time Defined Using a Spread Function. The travel time from the target (A) to any location on the map is defined by the travel time contours. The label on each line indicates travel time in minutes. The rate of travel is 5 km/hr except through the partial barrier where the rate of travel is reduced.
evaluate travel time. The fastest travel time from Location A to B is 90 minutes. There are two routes with equal travel times: a longer route around the obstacle at a faster rate of travel and a shorter but slower route through the obstacle.

One of the advantages of a spread function is that irregularly distributed factors can be accommodated. In Figure 7.36, a land use map is shown on the left. The walking time is to be mapped from a designated point to any location in the map area. First a travel time data layer in raster format is generated. Based on testing, previous experience, or other information, the travel time to traverse one cell is determined for each land use type. A new data layer, shown on the right in Figure 7.36, is produced by assigning to each cell the travel time for its land use type. Roadway cells have the lowest values because they can be traversed the most quickly. Other types of terrain have higher values, indicating the slower rate of travel across them. Once the travel time data layer has been generated, the spread function can be used to evaluate the travel time from one or more target locations to any point on the map. The two input data layers and resulting travel time contours are shown graphically in Figure 7.37.

The calculation used to generate the travel time map is illustrated in the simplified example shown in Figure 7.38. Here only two classes are used. An additional simplifying assumption is made about the calculation of distances: they are calculated from cell edges instead of the more usual cell centers. The travel time data layer defines the time to traverse each cell. The more general term friction surface is often used to describe this type of data layer (the value in each cell represents the degree to which movement across it is retarded). The second data layer identifies the location of the start points (in this case only one point is used). Multiple starting locations are often used, e.g. to calculate the minimum travel time from any of several emergency facilities.

The third data layer shows the cumulative travel time to any point from the target. It is produced by accumulating the travel times for each cell while moving outward from the starting location. Where a cell can have two values, the smaller value is selected. The computer implementation is somewhat different in order to improve

![Figure 7.36 Generation of a Travel-Time Data Layer. The time to traverse cells of each land use class is assigned to a travel time data layer. Cropland is assigned a travel time of 6 minutes per cell, the roadway is 2 minutes per cell, and the rangeland is traversed at 4 minutes per cell. The quarry cannot be traversed and is assigned a +.](image-url)
processing efficiency, but the principle is the same.

One application of this type of analysis is in predicting the time needed to traverse terrain with variable conditions. It is an application often encountered in military operations, and has been termed *terrain trafficability*. The trafficability, or ease and speed of movement, will vary with the type of ground cover, topography, mode of transport, and season of travel. The analysis is complex because the route taken to reach a given destination is not necessarily a straight line and the travel time to cross each unit of terrain may change suddenly along a route as a result of changes in land cover or other factors. The spread function is able to incorporate these diverse and irregularly distributed constraint factors.

Spread functions have been implemented in research GIS software packages. The spread function was originally developed by Dana Tomlin and is included in the Map Analysis Package, available from the Harvard Graphics Laboratory, Harvard University. Although some features of the spread function are provided in commercial GIS software, the more general version of the spread function as described here is to the author’s knowledge available only in non-commercial research software. In part, this may be a result of the difficulty of implementing the task for large numbers of cells. The function requires that the area be regularly subdivided into relatively small terrain units to provide the cells over which values are progressively accumulated. For this reason the spread function is implemented using raster-based techniques.
What makes the spread function particularly important is that it can incorporate information about the variability of conditions and tally the cumulative effects of this variability. Spread functions are a valuable and very flexible GIS capability, particularly for cartographic modelling. As users become more involved with cartographic modelling and more familiar with contiguity functions, the demand for general spread function capabilities may lead to their wider availability.
Seek or Stream Functions

The seek function (also termed a stream function) performs a directed search outward in a step-by-step manner from a start location using a specified decision rule. The procedure is repeated until any further movement would violate the decision rule. The output from the seek operation is a trace of the one or more paths taken in moving from the start point(s) until the function stops. For example, a seek function can be applied to a digital elevation model to trace the path of water flow. The rule used might be to move from a start location to the adjacent point with the lowest elevation. This operation would be repeated until a location is reached where all the adjacent points have higher elevations (a local depression) or the edge of the study area is reached.

This procedure could be used to evaluate erosion hazard. The seek operation could be applied using a regular grid of start points. The number of times each location or cell is traversed would be a measure of the potential flow of water over each unit of terrain. This value could be used as a measure of erosion potential from surface runoff. Other factors that affect erosion potential, such as vegetation cover and soil type, could also be incorporated into the analysis.

The seek function can be used with the spread function to provide an automated route selection capability. Figure 7.39 illustrates this concept using the travel time overlay developed in the previous figure. The spread function was used to generate a data layer in which the value at each location is the shortest travel time from the start point. The fastest route from any point back to the start can be defined using a seek function with the decision rule to seek the adjacent cell having the lowest value. Since the travel time data layer was constructed outward from the start point, the seek function must end at that same point, and in the process trace the quickest route.

Intervisibility Functions

Intervisibility functions, also termed viewshed modelling or viewshed mapping, is also a cumulative type of operation. The viewshed is the area that can be "seen" (i.e. is in direct line-of-sight) from the specified target locations. Intervisibility functions can be used to map the area visible from a scenic lookout, map the area that can be detected by a radar antenna, or assess how effectively a road will be hidden from view. It is valuable for such diverse applications as landscape planning, military planning, and communications.

Intervisibility functions use digital elevation data to define the surrounding topography. Depending on the sophistication of the software, additional data can be included in the analysis such as the heights of individual features (e.g. buildings or transmission towers) or the heights of different land cover classes. These features may be represented in separate data layers or as a list of point locations and corresponding heights.
Figure 7.40 The Concept of Intervisibility. An intervisibility analysis identifies locations that are within the unobstructed line-of-sight of a viewing position. Areas that are screened from view are shaded in the diagram. (Adapted from an illustration by D. Tomlin, Ohio State University, Columbus, Ohio.)

Figure 7.41 Viewing Parameters for Intervisibility Analysis. The viewing parameters used in an intervisibility analysis are: the 3-dimensional position of the viewer, the horizontal and vertical angles of view, the viewing direction, and the maximum viewable line-of-sight distance.
The concept of intervisibility analysis is illustrated in Figure 7.40. The areas that are hidden from the viewer are shown stippled. The data used in the analysis are the viewing parameters: 3-dimensional data for the landscape to be analyzed, and any specific targets to be considered. Figure 7.41 illustrates the viewing parameters needed. They are the maximum viewable line-of-sight distance, the 3-dimensional location of the viewing position within the landscape, the vertical and horizontal angle of view, and the viewing direction (usually specified as degrees of azimuth, which is the angle in degrees measured clockwise from north).

In Figure 7.42, the components of an intervisibility or viewshed analysis are shown using separate data layers for each component. The target data layer is needed if the visibility of specified targets is to be assessed. Often the purpose of the analysis is to generate a viewshed map, in which case a target data layer would not be required.

The intervisibility function is a powerful tool for planning the siting of features in a

---

**Figure 7.42 Data Sets for Intervisibility Analysis.** Conceptually, an intervisibility analysis requires three sets of spatial data: a surface topography and a surface features dataset to define the landscape, and a set of viewing positions from which intervisibility is to be assessed. A set of targets may be included as a fourth data set. In which case the output of the analysis could be a tabulation of the targets visible from each viewing position. The more common output is a map showing the areas visible from each viewing position.
landscape. It is a tool that lends itself well to a trial-and-error analysis procedure in which the placement of objects is progressively refined by re-evaluating the viewshed as the location of objects is changed. The siting of such facilities as microwave and radar systems that require an unobstructed line-of-sight can be optimized in this way. In landscape planning, there is often the need to design facilities that are hidden from view, e.g. minimizing the roadway or cleared land that can be seen from a scenic lookout. Alternative locations for the lookout and the facilities can be evaluated by generating a viewshed map using the scenic lookouts as the viewing positions. Successive modifications to the plan could be assessed by generating revised viewshed maps.

As computer processing power becomes less expensive, interactive viewshed modelling will probably become more readily available. Two similar and more widely available functions are perspective view generation and illumination mapping (used to generate a shaded relief image). They are computationally similar to intervisibility analysis since they involve tracing rays of light, and can be implemented using the more general intervisibility function. However, they are usually discussed as separate functions, and that convention is followed here.

**Illumination**

Illumination functions portray the effect of shining a light onto a 3-dimensional surface. The three sets of factors that control this function are the nature and position of the illumination source, the topography and reflectance of the surface, and the position and direction from which the model is viewed. Software differ in the parameters under operator control. For example, the viewing position and surface reflectance are often fixed parameters.

Figure 7.43 is the output from one of the more common types of illumination functions. It is termed a shaded relief image or shaded relief model and was produced from digital elevation data. The landscape is represented as if it were composed of a material of uniform reflectance illuminated from the left of the image. The position of the illumination source was chosen to provide sufficient shadow for the relief to be easily perceived.

The human mind can perceive shape from an image much better than from data points plotted on a map. Similarly, the relationship between the magnitude of a variable and its spatial location is easily understood as a 3-dimensional surface. The surface need not even represent elevation. It could represent any data that behave as a more or less continuous surface, such as gravity or magnetic
field data which are measurements commonly used in geology. In viewing the shaded relief image of these data, the geologist can perceive the spatial distribution of a set of measurements as a more easily interpreted landscape.

Illumination functions are also used to assess natural conditions. Growing conditions are affected by the quantity and direction of sunlight. Erosion potential and vegetation regeneration rates can be influenced by solar illumination conditions. By modelling illumination conditions, this factor can be included in planning activities.

Shaded relief images provide a rendering of the landscape that can add the surface information to thematic maps or digital images like satellite imagery. The process of applying another data set over a shaded relief image is termed draping. It is as if the map were placed or draped over a 3-dimensional model of the terrain. Draping functions are available in several commercial GIS packages. These models are usually portrayed as if viewed from the side and are termed perspective views, as discussed in the subsequent section.

The results of an illumination analysis are usually presented in the form of an image, although tabulations are sometimes used (e.g. the average illumination for each forest stand may be used as a stand attribute). A photographic image can best reproduce the subtle grey tones of a shaded relief image so that it appears 3-dimensional. While some types of plotters can produce a range of grey tones, they cannot provide the fine gradations of a photographic image. Instead, plotter outputs are usually presented as mesh diagrams viewed from an oblique position, i.e. a perspective view.

**Perspective View**

A surface portrayed from a viewing position other than vertical is termed a perspective view. Perspective views are primarily a presentation tool. They are useful in showing the 3-dimensional context of features on a surface, such as a natural landscape. Whereas the vertical view tends to flatten the perceived relief, in a perspective view the relief can be exaggerated to emphasize surface features. Perspective views are commonly generated as photographic outputs or plotted as mesh diagrams. In a mesh diagram, the topography is represented as if a grid of regularly spaced lines had been draped over it. A line-shaded image uses parallel lines with variable width to provide a perspective rendering, as shown in Figure 7.44. In a similar way, thematic maps or satellite imagery can be draped over a shaded relief model to give a 3-dimensional perspective view of the landscape. Plate 13 is a perspective view generated in this way from a satellite image and digital elevation data. By generating a series of perspective views like this one, a motion picture flying sequence can be produced. Perhaps the most sophisticated use of computer-generated perspective views has been in the production of flying sequences for the commercial film industry.

**OUTPUT FORMATTING**

Output formatting is the preparation of analysis results for output. In the case of tabular data summaries, the preparation is generally incorporated into the analysis function itself and the output file need only be sent to the printer. Outputs in the form of maps are generated in hardcopy formats by such devices as pen plotters, electrostatic plotters, and photographic devices. Map-like outputs are also displayed as electronic images (also termed softcopy) on monochrome or colour monitors.

The software functions provided to create these types of output vary widely in flexibility and ease of use. The simplest approach has been to provide one or more standard presentation formats. The operator may be restricted in the placement of titles, legend blocks, and
other annotation. More sophisticated systems provide a range of digital cartographic functions, such as the generation of coordinate grids, a wide selection of text fonts, line weights and colors, the definition of symbols, and even the automated placement of text labels within the map. Some of the more common types of output formatting functions are discussed in the following sections.

**MAP ANNOTATION**

Titles, legends, scale bars, and north arrows are perhaps the most common form of annotation. In its simplest form the title block and legend blocks have fixed positions on the map and the operator can only enter the text and legend symbols. More flexible implementations allow the operator to select the placement and size of these map features. These types of map annotation are placed either outside the map boundary or they overwrite a portion of the map information. They are not generally interspersed with the map information itself.

**TEXT LABELS**

Text labels (also termed name labels) are placed within the map area and interspersed with the map information. They may be placed next to a point location (city names), along a linear feature (the name of a river placed along the curvilinear river edge), or within a polygon (the name of a country). Text labels form an important part of a map. Not only do they provide the name of the geographic feature, they can also be used to show the orientation of the feature, its relative size, and even its class. This is done through the font, size, spacing, and placement of the label. Labels also affect the appearance of the map—whether it looks cluttered or is clearly legible. Some of the
general principles used in map label design and placement are as follows (Jimhof 1975):

1. The names should be legible and located close to the feature they describe.

2. The association between the name and the object it identifies should be easily recognized.

3. Labels should not overlap and the covering or concealing of map information should be minimized.

4. The format and positioning of a name label should directly assist in showing relative importance, territorial extent, connections, and in distinguishing among groups of map features. For example, the name of an area feature should span the entire area and conform to the general shape of the element.

Most GIS software has some text labelling capability. More limited implementations severely restrict the size and orientation of the labels. More comprehensive text labelling software can allow the operator to position labels interactively while viewing an image of the output map. Interactive scaling of text size, the automated retrieval of labels from the database, and even automated label placement may also be provided. Systems designed originally as digital cartographic systems to which GIS functions have been added have tended to provide more sophisticated capabilities, such as automated label placement. However, over the past few years the cartographic quality of GIS systems has improved considerably.

TEXTURE PATTERNS AND LINE STYLES

The selection of line widths and colours is dependent on the output device. Most devices can generate texture patterns. Line widths and colours are used to portray attributes of the line. Lines that represent such features as highways, railways, or political boundaries are commonly distinguished in this way. Line types, such as dashed lines or dotted lines, are also used to distinguish elements. Some systems provide for user-specified dash patterns. In a similar way, patterns (including solid colours) can be used to distinguish different types of areas. The patterns generally include different patterns of cross-hatching, shading, and colours. Software differ in the amount of effort needed to select these drawing parameters. In some cases, standard drawing parameters can be saved and applied to other maps containing the same types of elements. Otherwise, the definition of the drawing parameters may have to be done separately for each map.

GRAPHIC SYMBOLS

Graphic symbols are used to represent map objects. The symbols used to designate a
city, a mountain peak, a bridge are common examples. Some systems provide a standard set of symbols but do not allow the operator to create symbols. Others provide the capability to create symbols and store them within the GIS so they can be recalled as needed, termed a symbol library. Some systems enable symbols to be assigned according to a user-specified attribute. In this way the appropriate symbol can be automatically plotted.

CARTOGRAPHIC MODELLING: A GIS ANALYSIS PROCEDURE

The previous sections presented an overview of the analysis functions available in geographic information systems. The key to using these functions effectively is to use a systematic approach in defining the information needed and in designing the analysis procedure to meet them. Cartographic modelling is one procedure that has been used for predictive modelling using a GIS.

The term cartographic modelling was coined by Tomlin (1983) to mean the use of basic GIS manipulation functions in a logical sequence to solve complex spatial problems. It was developed to model land use planning alternatives, an application that requires the integrated analysis of multiple geographically distributed factors.

The cartographic modelling concept is illustrated first using a contrived example and then an actual application. Figure 7.45 diagrams a land use planning application of cartographic modelling to site a roadway in a hypothetical National Park. The design of a cartographic modelling procedure is best approached by working backwards from the required final result. In the Figure, the final result, a map of the final route location, is generated by Procedure 6, at the extreme right. It is an output formatting procedure that plots the route selected in the previous step.

Procedure 5, the immediately preceding step, is the selection of the final route from several proposed alternatives. The selection is guided by an evaluation of the quantifiable costs and benefits of the competing routes, as well as by the consideration of qualitative factors, such as aesthetics, public sentiment, and so on. In this example, the competing concerns are to minimize construction costs, to minimize the visibility of the road from scenic lookouts, and to minimize the loss of those land use types most critical to wildlife.

Trade-offs must often be made between qualitative and quantitative objectives. Is it worth an additional $10,000 to minimize the visibility of a roadway from scenic lookouts? Is the value of preserving an additional 15 sq km of wetland worth the cost of constructing an additional kilometer of roadway? The cartographic modelling process provides a systematic means to explicitly identify these issues and provide information to support the decision. However, it does not automatically provide the decision: subjective value judgments must still be made.

Procedure 4 is the process of generating the alternative routes. The inputs used are a map showing the relative cost of road construction in the study area, a map of the areas visible from the scenic lookouts, a map of the land cover types, the location of the start and end points for the road, and judgments that reflect the design objectives. The ideal solution might be a road with the minimum construction cost, that is not visible from the lookouts, and does not disturb any critical wildlife habitat. Usually the ideal solution is unattainable, in which case trade-offs must be made among the competing objectives. An iterative process is used to develop route locations that satisfy these objectives to different degrees. The relative importance of these objectives is a judgment made by the participants in the planning process to guide the search for alternative routes. Of the many alternatives considered in this step those considered to be the “best” would be passed to the final route selection process (Procedure 5).
Figure 7.45 Cartographic Modelling for Route Selection. The flow chart illustrates the use of cartographic modelling to define an optimum route location.
Procedure 3 is the generation of the road construction cost data layer that shows the cost of roadway construction at each location in the study area. The factors taken into account are the slope of the terrain, the type of soil material, and the type of land cover. Rules are used to calculate the construction costs for each terrain condition. These values are then used to produce the construction cost data layer used as one of the inputs to Procedure 4.

Procedure 2 is a viewed analysis to identify the park areas that are visible from the scenic lookout locations. The inputs to this procedure are the lookout locations, the average heights of the different land cover types, the land cover data layer, and digital elevation data for the study area. The output from the viewed analysis is a data layer showing those areas visible from the lookouts and those that are hidden.

Procedure 1 is the generation of a slope data layer that is calculated from the digital elevation data.

The process of working backwards through the analysis ensures that all data that will be needed are identified, data that
Figure 7.17 Environmental Constraint: Map for Transmission Line Routing. Environmental constraint categories were produced using the data from ten resource data sets weighted according to priorities expressed by regulatory agencies and the public. (Courtesy of the Environmental Planning Branch, Ontario Hydro, Toronto, Ontario)

will not be used are not collected, and that steps where value judgments must be made are explicitly identified. It is a systematic approach that can be applied to a wide range of planning activities. This approach also facilitates the documentation of how design decisions were reached, making it easier to refine the analysis and to scrutinize the process.

**A TRANSMISSION LINE PLANNING EXAMPLE**

Figures 7.46 to 7.49 illustrate several steps in the cartographic modelling procedure used by Ontario Hydro to select a right-of-way for a power transmission line. The planning process required that engineering, environmental, and social factors be considered and that the design trade-offs be justified before internal reviews as well as before government regulatory agencies and public hearings. It is required that the criteria used to select a route location be documented and that outside concerns be addressed.

To support this type of planning activity, Ontario Hydro used an in-house GIS. The source data consist of existing geographic
information in map and tabular form, as well as original data collected by field surveys and analysis of aerial photography and satellite imagery. Digitizing facilities are available to convert non-digital data sets for input to the GIS. An image analysis system is used for reconnaissance level land cover and land use mapping. The remote sensing results are output in a digital form that can be input directly to the GIS data base.

Figure 7.46 is a map produced from one of the GIS data layers used in the route location study. It is a map of agricultural land use systems produced by the Ontario Ministry of Food and Agriculture. Some 23 classes are shown in this map. The GIS was used to group these detailed land cover and land use classes into constraint categories appropriate for the route selection analysis. Other data sets used in the study were agricultural capability, current land use, forestry capability, mineral potential, recreation, hydrology, heritage features, and human settlements.

Figure 7.47 is a constraint map that takes into account all the environmental concerns. It is not a simple addition of the constraint
ratings for the individual resources. The constraints were weighted according to the priorities expressed by the participating advisory and regulatory groups. A total of nine constraint categories were defined. These were combined into five categories to produce a more legible map for illustration. However, the nine categories were retained in the database. In the Figure, the more severe constraint classes are darker.

Figure 7.48 shows the alternative routes proposed for the transmission line. They were developed using maps of the individual resources and derived maps produced by integrated GIS analyses, such as the constraint map. The routes were actually drawn on an airphoto mosaic and then digitized for input to the GIS. These proposed routes were then analyzed to determine the quantity of various resources that would be affected by each right-of-way. This was done by generating buffer zones (the shaded areas in the figure) around the transmission line routes (shown as heavy black lines).
Table 7.3 Tabulation of Constraints for A Proposed Transmission Line Route. (Courtesy of the Environmental Planning Branch, Ontario Hydro. Toronto, Ontario.)

<table>
<thead>
<tr>
<th>AFFECTED AREA (ha)</th>
<th>AFFECTED LENGTH (km)</th>
<th>% OF ROUTE AREA</th>
<th>CONSTRAINTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>.32</td>
<td>.2</td>
<td>major non-urban settlement</td>
</tr>
<tr>
<td>6</td>
<td>.79</td>
<td>.5</td>
<td>military base, communication towers</td>
</tr>
<tr>
<td>26</td>
<td>3.47</td>
<td>2.2</td>
<td>proposed settlement or airport</td>
</tr>
<tr>
<td>2</td>
<td>.32</td>
<td>.2</td>
<td>wetland, waterfowl nesting/staging</td>
</tr>
<tr>
<td>84</td>
<td>11.96</td>
<td>7.0</td>
<td>deer yards</td>
</tr>
<tr>
<td>6</td>
<td>.79</td>
<td>.5</td>
<td>classes 1,2,3 wetland habitat</td>
</tr>
<tr>
<td>43</td>
<td>5.69</td>
<td>3.8</td>
<td>sensitive woodland wildlife habitat</td>
</tr>
<tr>
<td>6</td>
<td>.79</td>
<td>.5</td>
<td>class 1-4 productive wildlife habitat</td>
</tr>
<tr>
<td>461</td>
<td>60.67</td>
<td>38.4</td>
<td>class 1-3 forest land</td>
</tr>
<tr>
<td>127</td>
<td>18.91</td>
<td>10.7</td>
<td>class 4-5 forest land</td>
</tr>
<tr>
<td>63</td>
<td>8.37</td>
<td>5.3</td>
<td>principal fruit &amp; vegetable producing areas</td>
</tr>
<tr>
<td>687</td>
<td>90.53</td>
<td>57.3</td>
<td>class 1 agricultural soil</td>
</tr>
<tr>
<td>201</td>
<td>28.54</td>
<td>16.8</td>
<td>class 2-3 agricultural soil</td>
</tr>
<tr>
<td>22</td>
<td>3.00</td>
<td>1.9</td>
<td>existing surface and mineral extraction</td>
</tr>
<tr>
<td>480</td>
<td>63.20</td>
<td>40.0</td>
<td>potential surface and mineral extraction</td>
</tr>
<tr>
<td>22</td>
<td>3.00</td>
<td>1.9</td>
<td>class 4 wetlands</td>
</tr>
<tr>
<td>109</td>
<td>14.38</td>
<td>9.1</td>
<td>cold water fish, areas sensitive to erosion effects</td>
</tr>
<tr>
<td>2</td>
<td>.32</td>
<td>.2</td>
<td>warm water fish, areas sensitive to erosion effects</td>
</tr>
<tr>
<td>383</td>
<td>50.40</td>
<td>31.9</td>
<td>cold water fish, areas with low sensitivity to erosion effects</td>
</tr>
<tr>
<td>20</td>
<td>2.99</td>
<td>1.7</td>
<td>areas with risk of erosion effects for which no fish information is available</td>
</tr>
</tbody>
</table>

An overlay operation was then used to generate quantitative information for the areas that would be affected by each of the routes. Figure 7.49 illustrates this procedure graphically using one of the resource maps. The agricultural resource types are shown only in the buffer zone areas. The detailed constraint data were analyzed in a similar manner. Table 7.3 is a tabulation of these constraints for areas within one of the proposed transmission line routes. In fact, it was the tabular data summaries more than the buffer zone maps that were used to compare the alternative routes.

By using the map and tabular information produced in this way, alternative route locations could be developed through a process that was documented, could be scrutinized by internal and outside reviewers, and was defensible. By means of this systematic analysis procedure, the logic used to reach the design alternatives could be reviewed and refined throughout the process. And effective use was made of the analytical capabilities of the GIS.

REFERENCES


