NR-FO-5740-E
1991

MINNESOTA EXTENSION SERVICE

UNIVERSITY OF MINNESOTA
NATURAL RESOURCES

INTRODUCTION to Data Analysis Using Geographic Information Systems

Daniel L. Falbo, LLoyd P. Queen, and Charles R. Blinn

Introduction

A Geographic Information System (GIS) is an automated information system that is able to compile, store, retrieve, analyze, and display mapped data. Only a decade ago this technology was limited to a relatively small number of colleges, universities, and local, state, and federal agencies. The two general types of users are systems users (who have hands-on use of the technology) and end users (who are users of the information generated by a GIS). Today, it is used by government officials, natural resource and social analysts, and many others. Its applications include environmental research and model building, urban demographic studies, and transportation analysis to mention only a few. While its use is expanding almost daily, its most important applications include those that support decision making.

Map data used by GIS are collected from existing maps, aerial photos, satellites, and other sources. A digitizer or similar device is used to convert compiled map data to a digital form in order to make it computer compatible. This transformation allows the storage, retrieval, and analysis of the mapped data to be performed by the computer. Maps produced by a GIS are typically displayed on computer monitors or are printed on paper. Unlike many other forms of computer graphics, such as computer-aided drafting (CAD) systems, a GIS displays actual geographic or mapped objects. GIS, however, is more than a mapping system. What sets it apart from even the most sophisticated mapping system is its power to analyze data and to present the results of that analysis as useful information to assist decision makers.

For a GIS to accurately represent occurrences on the earth's surface, data must be reliable, accurate, and pertinent. Because the success of the GIS and all decisions that are based on it ultimately rest on the integrity of the data, the GIS must be capable of compiling, updating, and maintaining its data. No matter how sophisticated the analytical tools, misused or questionable data will make the final output doubtful. The adage "garbage in — garbage out" certainly holds true in the world of GIS.

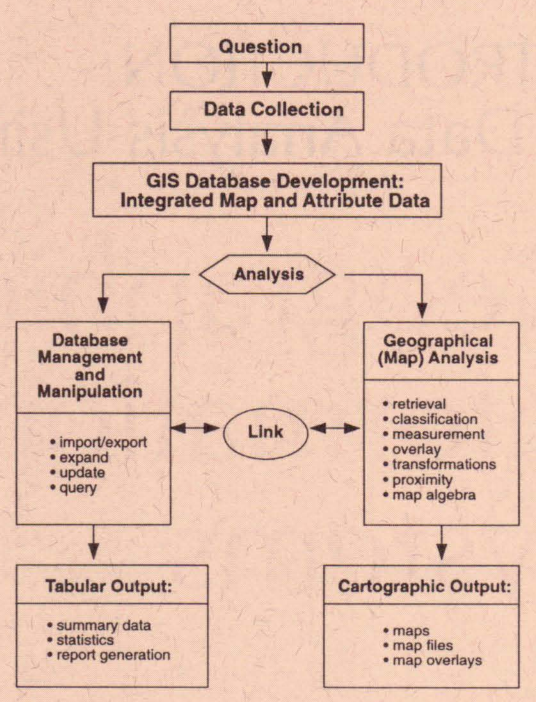
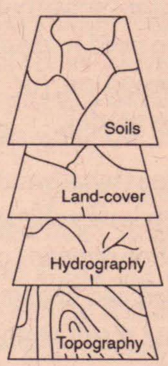


Figure 1. Logical flow chart of analytical operations within a GIS framework. This figure diagrams the logical structure of GIS operation. An initial geographic question drives the data collection. Commonly data collection is done by either manually digitizing data from appropriate maps or by importing tabular data which is referenced with georeferenced coordinates that define the geographic position of the tabular (attribute) data. The data is then converted to a digital format and is entered into a database where it is stored as integrated map and attribute data. The analytical process then manipulates both map and attribute data through the linkages that GIS establishes between them. Two types of output are common: tabular (summary data, statistics, reports) and cartographic (maps, map files, and map overlays).

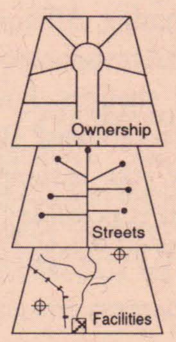
Use of GIS is driven by the need to answer geographical or spatial questions. The ensuing data collection, database analysis, and output are in response to those questions (Figure 1). The purpose of collecting data for a GIS is (a) to inventory a geographically defined area (for example, to locate all state-owned parcels within a particular area) or (b) to test hypotheses and build models. Though the data is initially collected for one of these reasons, data collected to answer one set of questions is frequently used in subsequent analyses to answer questions that were not anticipated at the time the data was collected. For example, after capturing detailed soils data to determine what crops can be grown, the data may be useful for answering questions about septic tank suitability.

The GIS database contains both map data (depicting location of geographical objects) and attribute data (describing physical characteristics of each object). Physical characteristics (such as timber species and tree diameter) and/or non-physical characteristics (such as estimated market value and management codes) are examples of attribute data that could be contained in a GIS used to analyze forestry problems. During a GIS analysis, site (map) data is linked with situation (attribute) data for each mapped timber stand. It is this link, which is automatically performed by the GIS software, that gives GIS its analytical power. The relationship between map data and its associated attribute data are shown in Figure 2 for typical natural resource and urban maps.

Maps → Attribute Database



Map Data	Attribute Data
Soils	• type • texture • pH • erosion potential
Land-cover	• dominant species • size • density
Hydrography	• lakes • streams • watersheds • gauging stations
Topography	• elevation • slope • aspect



Map Data	Attribute Data
Ownership	• address • size • zoning • legal description
Streets	• type • width / length • paving
Facilities	• powerlines • water • sewer • buried cables

Figure 2. Schematic representation of map data and associated attribute data. Examples are shown for typical natural resource and urban GIS applications. Data are geographically referenced on map layers, each with a primary theme. Specific details or attributes can then be identified for each feature of interest on the map. These attribute data or map key information are then stored in the GIS database.

In GIS terminology, the individual spatial phenomena or map themes are referred to as map layers. One layer can contain roads, another soils, and another can indicate land ownership. Each layer comprises all of the pertinent map and attribute data. Though other, non-layered approaches to GIS exist, the layered map model will be used throughout this paper to give examples of GIS analysis and use.

The next section of this publication, Capabilities of a GIS, examines spatial questions that drive a GIS and the basic requirements for the database. The final section, Analytical and Operational Functions, examines how GIS software links map and attribute data in order to analyze spatial problems.

Capabilities of a GIS

Spatial Questions That Drive an Analysis

GIS applications are the result of spatial questions. The following questions are commonly asked:

1. What exists at a particular location on the face of the earth (locational analysis)?
2. When is a specific spatial condition satisfied?
3. What has spatially changed over time?
4. What kind of pattern will emerge from geographical data?
5. What will happen if certain phenomena are entered into predetermined scenarios?

Locational Analysis

Mapped data primarily indicates where objects are located, but cannot explain why. For example, an aerial photo may show that corn is growing vigorously in certain sections of a field, but cannot explain why it does not grow well in other areas. GIS analysis, on the other hand, may show a connection between corn growth, soil type, and available water by simultaneously examining computerized crop, soil, and soil moisture maps.

Spatial analysts continually seek patterns in mapped data. Defining the distribution of a crop creates a geographical pattern. Investigating the commercial feasibility of a potential crop may require an evaluation of how production, transportation, and markets are related to each other. The GIS is able to store the necessary data and to study these complex relationships in a fast, flexible manner.

Satisfying a Spatial Condition

Frequently a GIS user wants to discover whether the mapped data will meet certain conditions. Suppose someone wants to know where to situate new groundwater wells. The wells need to be located within 10 miles of a particular subdivision and need to occur along or adjacent to pipelines already serving that location. The general spatial conditions would be used to define a broad area in which the wells could be situated. Then, the specific location of each new well would be based, in part, on locational analyses performed using well-siting criteria.

Temporal Analysis

Both our perception of the world and the world itself are constantly changing. The growth of suburban cities and the effects on metropolitan land-use have changed dramatically in the past decade. Knowledge that is restricted to a single point in time may not be meaningful for answering some questions. For example, an analyst might look at the relationship between changing land-use practices and zoning law changes over many years. By storing and comparing maps of various dates, GIS can perform temporal analyses.

Emerging Patterns

GIS users seek spatial patterns. In other words, they want to know if two or more things vary similarly in space. They may want to know, for example, if there are a proportionately larger number of traffic fatalities along roads with posted speeds of 65 m.p.h. compared to those posted at 55 m.p.h. If there are, other questions such as, "Which stretches of roads are the most hazardous?" may follow.

Evaluating Different Scenarios

"Scenario building" is the result of "What would happen if . . . ." type questions. For example, what would happen to coastal areas if global temperatures increased, the icecaps partially melted, and sea levels were elevated? Here, the user employs a model designed to forecast and map the potential impact of climate changes on sea level in coastal areas. Application of such a model allows the user to construct a hypothetical situation and forecast the outcome.

Sometimes a change in assumptions results in a forecast that is substantially different than expected. For this reason, it is important to test a model's assumptions through a process known as "sensitivity analysis." Sensitivity analysis can be used to determine how much the outcome depends on (or is sensitive to) assumptions used by the model.

In the above example, the outcome (what would happen to coastal areas) partially depends on a change in sea elevation. Change in sea elevation depends on the degree to which icecaps are melted and changes in rainfall and evaporation which may result from a change in global temperature. If subsequent analyses are performed with the model assuming different feasible changes in melting, rainfall, and evaporation for a range of global temperature changes, how different are the projected effects on costal areas? By testing different scenarios that vary your assumptions and evaluating the response, it is possible to determine how sensitive the model is to changes in assumptions.

Basic Database Requirements

A GIS must allow the operator to: (1) incorporate (import) data from outside sources, (2) easily update and alter data, and (3) ask data-related questions of (or query) the database. The database management system (DBMS) software that is a part of a typical

GIS provides these capabilities. Also, commercially available database management software programs can be programmed to perform these same tasks outside of the GIS.

Importing and Expanding the Database

Data imported into a GIS often comes in the form of standard ASCII (American Standard Code for Information Interchange) files. ASCII is a standardized code that can be read by nearly all computer systems. A GIS may also be capable of importing data files that are in other formats.

Another method of expanding the database is to manually enter data. Every DBMS has the capacity to create new geographical objects of interest or records. This technique is commonly used to add a relatively small amount of new objects (a few printed pages) to the database.

The DBMS also allows for error checking as new records are created or existing ones are updated. Not all errors can be eliminated in this way, however, so care must be taken when collecting, automating, and changing the database.

A GIS must also provide the ability to create data files that can be exported to other systems. During the exporting process, data files are written in a common format (e.g., ASCII) to a file that can then be imported by other systems

Updating Attributes

Another common task is updating or editing the database. Since no user can foresee all future data needs and applications, a GIS must provide ways to easily modify, refine, or correct the database. Attribute data are seldom static. Therefore, maintaining the currency of the data depends on updating capability.

Query the Database

Manipulating the database to answer specific data-related questions is accomplished through a process known as database analysis. Tabular output is the result of a database analysis query. Those output products might become part of a summary report or they might be imported into spreadsheet software where further analyses might be performed.

To query the database, logical expressions that impose limits or conditions on the database search are defined. These logical expressions specify which geographical objects are to be included in the analysis and/or how that data is to be analyzed. A subset of the database is produced. Some logical expressions are simple and require only one condition while others are very complex and contain multiple conditions.

For example, using a database that contains data about the location (township, range, and section), ownership, and size of lakes, the user could ask the simple question: "Where are all the lakes in my county?" In searching for the answer, the DBMS creates a data subset that is presented in tabular format (rows and columns of information). The subset created to answer this question meets a single condition: the condition that the data apply to lakes in that particular county. In contrast, the question "Where are the private lakes in section 34 of township 61 range 19 that are larger than 15 acres?" contains multiple conditions. As a result, the database creates a subset of the data which meets all of these conditions.

Another type of question that can be asked using logical expressions requires that mathematical analysis be performed. These questions prompt the DBMS to perform such functions as calculating population density, generating descriptive statistical summaries, and translating between measurement systems (e.g., converting area defined in square meters into acres or square miles). These types of operations are common when the database contains numeric data and a statistical analysis of the data is required.

Analytical and Operational Functions

Both traditional DBMS software and GIS support database analysis, but a GIS also supports map analysis. It is useful to think of GIS map analysis in a layered-model context. The layered GIS model is analogous to transparent maps that can be accurately stacked upon one another. Typically, each layer contains only one mapped theme. Traditionally, map analysis of multiple overlays had to be performed manually.

When attempting to answer a geographic question, the user determines which phenomena are to be examined and how the analysis will proceed. A GIS provides a set of "tools" or computer programs that allow the user to perform a specific set of operations on map and attribute data. These tools, which are in the form of operating commands, permit spatial inquiry, manipulation, and analysis. Examples of some of these operations, which give the user the power to analyze the map layers, are the focus of this section.

Functional Tools for Map Analysis

Tools that manipulate attribute data may employ logical expressions similar to those described in the database analysis discussion presented above. In terms of map data analysis, and with the exception of those tools used specifically for overlaying (or merging) data from multiple map layers, operations performed on map data can be performed on either a single layer or upon multiple layers.

The functional tools available for map analysis can be grouped into six categories: (1) projection and spatial transformation utilities, (2) spatial retrieval, classification, and measurement functions, (3) logical and visual overlaying capabilities, (4) proximity and network functions, (5) map algebra utilities, and (6) output generation.

While the specific names for the various analytical tools may vary from one GIS to another, the operations they perform are similar. The tools may be used in various combinations and sequences to accomplish the desired task. The selection and sequencing of tools should be determined according to the specific need at hand. Some of the spatial analysis operations that can be performed using these tools are illustrated in Figures 3 - 7.

Projection and Spatial Transformation Utilities

Various pre-processing operations may need to be performed to remove errors that may occur as observations are made, maps are compiled, and/or as layers are encoded into the database. A common source of error is in the registration of multiple layers. Registration involves the systematic adjustment of a map layer so that it can be accurately laid over another layer of the same area. Data development may not consistently use the same units of measure during data collection. Because of differences in data sources, errors during data input, and because some layers may be in one coordinate system [such as Latitude/Longitude or Universal Transverse Mercator (UTM) units] while some may be in another, registration is important. A GIS is able to reconcile these irregularities by mathematically converting the map data from one coordinate and projection system to another (Figure 3).

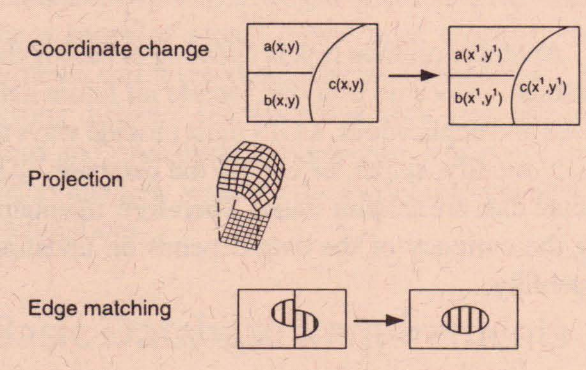


Figure 3. Projection and spatial transformation utilities allow a GIS to reconcile irregularities between map layers. Map data can be mathematically converted from one coordinate and projection system to another.

Numerous map projections have been developed to represent the earth's three-dimensional surface on a two-dimensional map. Usually, a GIS user has no control over the projection of existing maps. Some maps may be in a cylindrical Mercator projection, others in a conic Albers, and still others in a planar projection. A typical GIS will have a set of programs which contain the algorithms needed to convert data from one map projection to another (Figure 3).

Another problem commonly arises where the area of interest is on adjoining maps. Here the location of some geographical objects may need to be modified to create a "seamless" layer between maps. This process requires that objects spanning two adjoining maps be "edge matched" to reconcile shape and location discrepancies (Figure 3). Often the margin of error is distributed between the adjoining polygons or lines. For example, a road layer may span several maps. Assume that one road is disjointed as it crosses into an adjoining sheet. The operator can reconcile the difference and make the road continuous by editing both parts of the study area simultaneously and removing the error (bringing the separate road segments together at a mutual point).

Spatial Retrieval, Classification, and Measurement Functions

In the previous discussion of database analysis, a GIS was asked to locate lakes according to user-specified criteria and the retrieved data was presented in a tabular format. In contrast, data may be retrieved according to its location on a map layer or its spatial relationship to other mapped data. This process is known as spatial retrieval (Figure 4). Spatial retrieval produces a spatial representation (map). For example, polygons representing lakes between 0 and 10 acres could be coded in one color, those between 11 and 20 in another color, and so forth, and the resulting color map could show the distribution of the various-sized lakes. The polygons could be presented in a spatial pattern corresponding to the actual location and configuration of the lakes in the real world. This spatial representation would likely be more meaningful to the user than the same information presented in a tabular format.

When map data are graphically displayed on a computer monitor, it is possible to select a specified area and examine it visually or analytically in more detail. This is known as "windowing" or "clipping" (Figure 4). This may be done by physically indicating the desired location or by entering coordinate

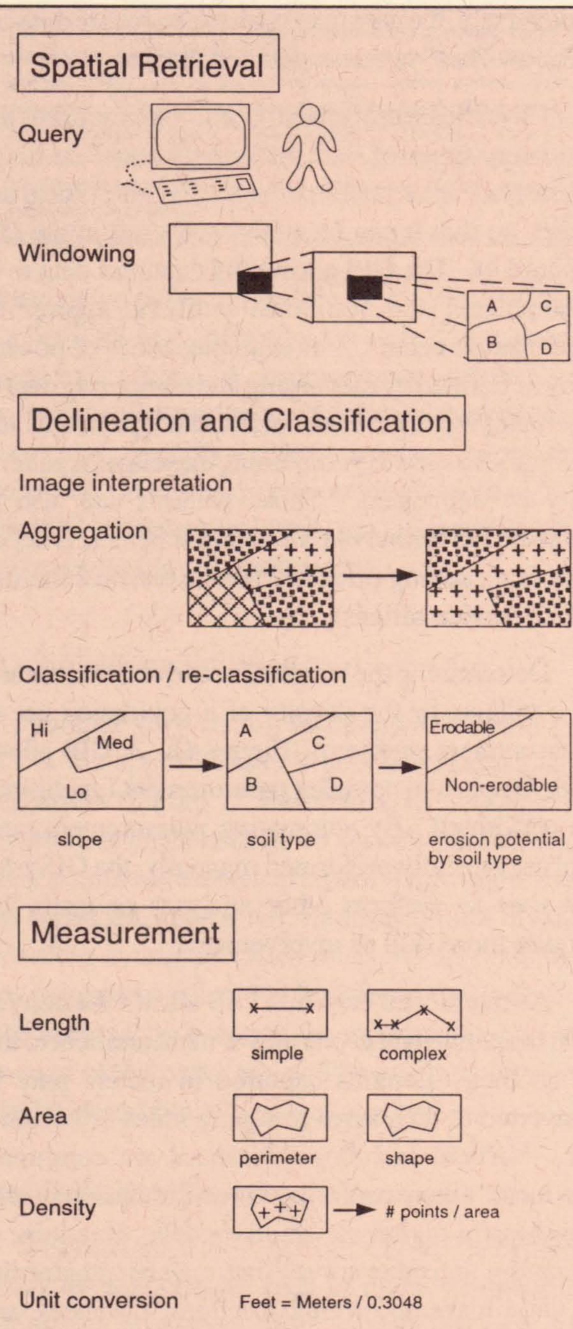


Figure 4. Spatial retrieval, delineation and classification, and measurement are separate functions, but are commonly used together. For example, a GIS user interested in legislative reapportionment may want to retrieve a district map, classify zones according to party affiliation, and determine the size of the area under study.

values that delineate the area of interest. The elements (polygons, lines, or points) found within the windowed area can be extracted and set aside on a new layer, allowing faster analysis of the smaller, selected data set. During windowing, the scale of the display can be changed to allow a more detailed display of the area of interest (provided that detailed data for the area of interest is present in the database). Conversely, the user may wish to generalize data and window "out" to a less-detailed display.

Classification of spatial phenomena requires that the many types of data commonly found in maps, aerial photos, or satellite imagery be interpreted and coded so that it can be stored and used in the GIS (Figure 4). The GIS allows the captured data to be manipulated and combined until the appropriate aggregation occurs. For example, zones of poverty may be defined by classifying and combining data on unemployment, government assistance payments, and income statistics. By combining these layers, patterns may be delineated. Those patterns can also be subjected to further classification schemes, if necessary (e.g., setting priorities for the establishment of social service offices).

Determining the length of a road or river, the area of a village, or the density of a population are all measurement problems (Figure 4). A GIS allows both simple and complex measurement functions to be performed. By automating measurement tasks that are typically performed manually, the GIS frees the user to perform other analyses or tasks that require more skill or involvement.

As mentioned earlier, a GIS allows the conversion of spatial data in one unit of measurement to that of another. Lengths captured in meters may be converted to feet, acres to square miles or hectares, etc. Vertical and slope distances are commonly produced when exploring three-dimensional relationships in digital elevation models. Bearings, or azimuths, and other survey data may be transformed to yield maps specifying locations, distances, and precise angles.

The GIS user may need to know how many times a spatial phenomenon occurs. For example, someone studying contaminated groundwater well sites may want to know the location of wells that become contaminated and may also want to know the number of contaminated sites. In this case, the user will want map data that can be used to indicate both spatial patterns and frequency of occurrence.

In certain spatial analyses, the user may want to find the best division of space based on spatial and non-spatial attributes. Attribute data within the study area are ranked, correlated, and then used to assist in the spatial ordering and delineation of areas. Legislative reapportionment is an example where spatial attributes (location and population) may be weighted with a non-spatial attribute (political bias) to calculate areal divisions.

Logical and Visual Overlaying Capabilities

GIS is probably best known for its ability to build map layers and evaluate the relationships between them. The relationships between map layers can be assessed from both mathematical (logical) and graphic (visual) perspectives.

Logical overlays superimpose layers using logical or spatial functions and store the results in the GIS database as new layers of data. This mathematical approach examines the quantitative association between the phenomena of interest. This relationship is determined by combining various data layers to create a composite data set (Figure 5). Because the layers have already been registered, they can be accurately placed over one another. In a typical GIS, the analysis might require that data from existing layers be combined to create new data or map layers. For example, streams (lines) and deer sightings

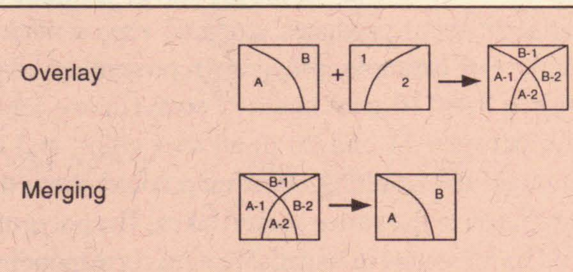


Figure 5. The overlay function allows the user to "stack" map layers on one another, showing spatial relationships between the layers. Merging generalizes classes within map layers by combining attributes to reveal new map features.

(points) can be combined with forest stands (polygons) to create a new database that reveals the spatial relationships between all three in the form of habitat quality.

The merging of data allows the user to take a complex data layer and to dissolve lines between shared attributes (Figure 5). The result is a more general data layer. Merging is the reverse of the process of combining various data layers or classes of attributes, as described above. As an example, if an ownership layer contains data for state government, local government, nonindustrial private owners, and private industry, a merge could be performed to reveal the spatial relationship between public and private ownerships.

Visual overlays, on the other hand, allow the user to graphically view spatial relationships between the various layers instead of seeking specific mathematical relationships. The overlays are graphically presented on a map or computer monitor. Visually displaying these overlays does not create a new layer in the database; it simply provides visual cues of the relationships between the layers.

Proximity and Network Functions

These types of analyses consider predefined areas around a geographical object or the connectivity of phenomena (Figure 6). For example, assume that the user had two layers, one of roads and another of timber stands with average stand diameter as an attribute. Furthermore, assume that the user wanted to know how much merchantable timber (trees of a specified diameter) was located within 500 feet of any road. By setting the proper parameters that considered merchantability and road location, a buffer zone would be created measuring 500 feet on either side of each road. All appropriate map and attribute data pertaining to the timber resources within the buffer would be generated as a separate layer available for observation and analysis.

Another type of proximity analysis involves the network function of a GIS (Figure 6). Networks are commonly established to evaluate options for the purpose of route optimization and resource allocation. Specifically, this means locating the best route between two points or the selection of service zones in a network (e.g., pizza delivery areas, fire service zones, mail routes). A common network function is the routing of emergency vehicles on road systems. In this case, the GIS analyzes distance factors, road speed, and other transportation variables (e.g., flow of traffic, traffic control measures) to generate alternative routes between two points.

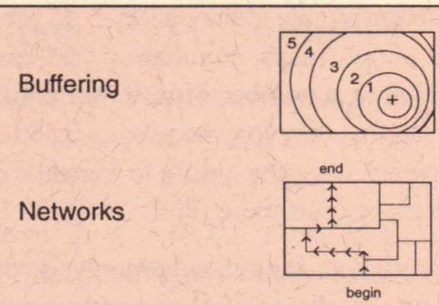


Figure 6. The buffer function examines an area which surrounds an object of interest. This function is used to create zones and to determine routes within zones. For example, a zone can be created based on specified distances from map features (e.g., the area within five miles of a road). Network functions examine the movement of objects along an interconnected pathway (e.g., traffic flow along a map of highway segments).

Map Algebra Utilities

Another very useful, quantitative capability of GIS is the application of algebraic expressions to map layers (Figure 7). Referred to as map algebra, this process enables users to specify mathematical relationships between map layers. Thus, entire maps can be added, subtracted, multiplied, and divided, according to user-specified rules. As an example, a new map can be generated by determining the difference in elevation between the topographic map and the map of the water table. The resulting layer can then be stored for subsequent analyses.

Output Generation

The final set of tools we consider here provide the ability to create output such as maps, geographical summaries or reports, and geographical base files (files containing both the digital map and attribute data). Output can be either hard copy, digital files, or displayed on a computer monitor. The sophistication of the GIS software and the output capabilities of the hardware/software system dictate the quality and variety of options available to the operator. In most cases, the GIS allows maps, summaries, and base files to be written in a number of different digital export formats so that they may be used by another GIS. Most systems have the ability to translate or directly import and export these files.

Generating graphic output, commonly in the form of maps, requires that a GIS have a wide variety of symbols and format options. Most offer numerous line, polygon, and point symbols to represent geographical phenomena as well as text options for labeling and annotating output. Since these utilities allow maps to be produced at various page sizes and map scales, output can be custom-designed to a format that is most appropriate for the situation.

Geographical and tabular summaries are common types of GIS output. Such summaries differ from those created by a traditional database query because they are based on map analysis. For example, upon completion of a windowing of clipping function, summary statistics presented in tabular form for the clipped area are often necessary to complement the analysis. A GIS will usually offer this type of report-generating capability and it is especially useful during complex spatial analysis or database documentation.

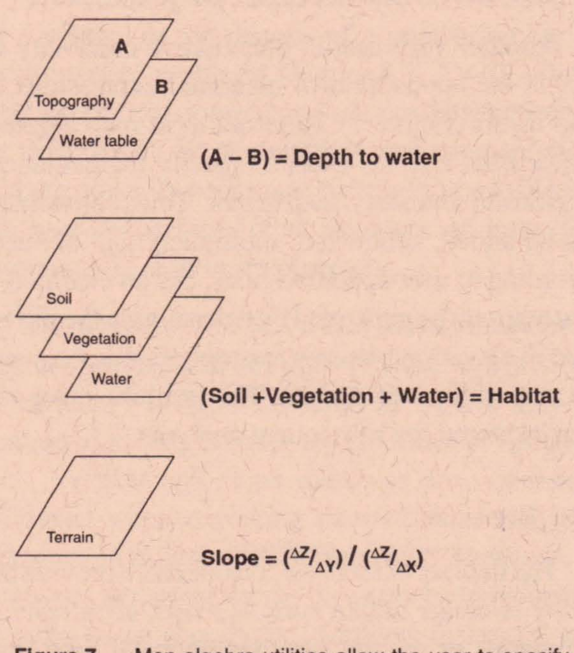


Figure 7. Map algebra utilities allow the user to specify mathematical relationships between map layers. Entire maps can be added, subtracted, multiplied, and divided according to user-specified rules. For example, a new map can be generated by determining the difference in elevation between a topographic map and the corresponding map of the water table.

Use of reference maps is an important part of the output (mapping) process. It is common for a GIS to contain a library of basic, often-used map and attribute data of the study area for creating these simple reference maps. Such base files could include county, section, parcel, or zip code boundaries, major transportation routes, or hydrological features. When included in output maps, these cartographic features provide a useful frame of reference for the map user.

Conclusions and Recommendations

As an integrated approach to managing and analyzing map and attribute data, geographic information systems are becoming increasingly common. GIS applications begin with a spatial question and proceed to a data collection phase during which both map and attribute data are collected. The initial spatial question drives the data collection and the model-building process. In the analysis phase which follows, both map and attribute data are manipulated via the analytical tools provided by the GIS software.

While the review of both database and map analysis were separated for purposes of discussion, they are closely linked in practice. Indeed, the power of GIS-based investigations comes largely from the ability to simultaneously access and process both map and attribute data. The tools presented in this paper for performing these analyses are commonly used and represent broad categories of functionality. However, this discussion does not constitute a complete or comprehensive listing of GIS tools.

The increasing influence and use of GIS can be largely attributed to its ability to support decision making. Toward this end, a GIS needs reliable data and must be able to analyze and synthesize this data

Editor: Richard Sherman
Graphic Designer: Michael Mechavich

Printed on recycled paper with agribased inks

Issued in furtherance of cooperative extension work in agriculture and home economics, acts of May 8 and June 30, 1914, in cooperation with the U. S. Department of Agriculture, Patrick J. Borich, Dean and Director of Minnesota Extension Service, University of Minnesota, St. Paul, Minnesota 55108. The University of Minnesota, including the Minnesota Extension Service, is committed to the policy that all persons shall have equal access to its programs, facilities, and employment without regard to race, religion, color, sex, national origin, handicap, age, veteran status, or sexual orientation.