Learning Computer-Assisted Map Analysis

by Joseph K. Berry*

"Old-fashioned math and statistics can go a long way toward helping us understand GIS"

Note: This paper was first published as part of a three-part series on GIS in the Journal of Forestry October 1986.

Geographic Information Systems (GIS) technology is expanding the computer revolution by integrating spatial information with the research, planning and management of forestlands. The teaching of GIS technology poses problems in the classroom, and innovative ways of learning to apply the technology are being developed.

Unlike most other disciplines, GIS technology was born from specialized applications. A comprehensive theory tying these applications together is only now emerging. In one sense GIS technology is similar to conventional map processing, involving traditional maps and drafting aids—pens, run-on shading, rulers, planimeters, dot grids, and acetate sheets for light-table overlays. In another sense these systems provide advanced analytical capabilities, enabling land managers to address complex issues in entirely new ways.

As essential as computer-assisted map analysis has become, the technology is difficult to teach. Practical experience is required as well as theory, yet very few classrooms can provide extensive hands-on learning. Most GIS require expensive and specialized hardware, and even where equipment is available, instructors are faced with teaching the procedures of a system not designed for classroom use. Consequently, the approach most commonly used relies on case studies and selected literature.

Mathematics provides a useful starting point. In fact, GIS theory is based on a mathematical framework of primitive map-analysis operations analogous to those of traditional statistics and algebra. The teacher presents basic data characteristics and map processing as in a math course. Lectures and exercises provide a general toolbox for map analysis that embodies fundamental concepts and stimulates creative application. This toolbox is as flexible as conventional mathematics in expression relationships among variables—but with GIS, the variables are entire maps.

The map analysis toolbox helps resource define and evaluate spatial considerations in land management. For example, forest managers can characterize timber supply by considering the relative skidding and log-hauling accessibility of harvest parcels. Wildlife can consider such factors as proximity to roads and relative housing density in order to map human activity and incorporate this information into conventional habitat maps. Forest planners can assess the visual aesthetics of alternative sites for a facility or clearcut.

The Fundamentals

The main purpose of a geographic information system is to process spatial information. The data structure can be conceptualized as a set of "floating maps" with common registration, allowing the user to "look" down and across a stack of maps. The spatial relationships of the data can be summarized (database inquiries) or manipulated (analytic processing). Such systems can be formally characterized as "internally referenced, automated, spatial information systems …designed for data mapping, management, and analysis."

All GIS contain hardware and software for data input, storage, processing, and display of computerized maps. The processing functions of these systems can be grouped into four categories: computer mapping, spatial database management, spatial statistics, and cartyographic modeling.

Computer mapping—also termed automated cartography, computer mapping involves the preparation of map products. The focus of these operations is the input and display of computerized maps.

Spatial Database Management—these procedures focus on the storage component of GIS, efficiently organizing and searching large set of data for frequency statistics and coincidence among variables. The database allows rapid updating and examining of mapped information. For example, a spatial database can be searched for areas of silty-loam soil, moderate slope, and ponderosa pine forest cover. A summary table or a map of the results can then be produced.

These mapping and database capabilities have proven to be the backbone of current GIS applications. Aside from the significant advantages of processing speed and ability to handle tremendous volumes of data, such uses are similar to those of manual techniques. Here is where the parallels to mathematics and traditional statistics may be drawn.

Because of those parallels, the generalized GIS structure provides a framework for discussing the various data types and storage procedures involved in computer mapping and data management. It also provides a foundation for advanced analytic operations.

Spatial statistics—the dominant feature of GIS technology is that spatial information is represented numerically rather than in an analog fashion, as in the inked lines of a map. Because of the analog nature of the map sheets, manual analytic techniques are limited to nonquantitative processing. Digital representation, on the other hand, has the potential for quantitative as well as qualitative processing.

GIS have stimulated the development of spatial statistics, a discipline that seeks to characterize the geographic distribution or pattern of mapped data. Spatial statistics differs from traditional statistics by describing the more refined spatial variation in the data, rather than producing typical responses assumed to be uniformally distributed in space.

An example of spatial analysis is shown in figures 1 and 2. Figure 1 depicts density mapping of a microorganism determined from laboratory analysis of surface water samples. Figure 2 depicts the natural extension to multivariate statistics for spatial coincidence between two microorganisms in the samples.

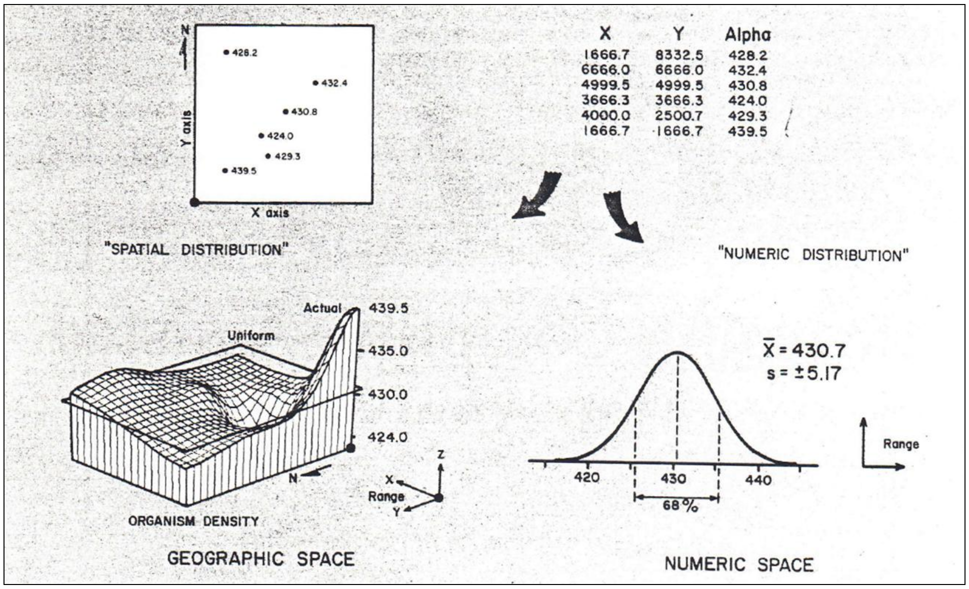

Figure 1. Spatial Statistics. Whereas traditional statistics identifies the typical response and assumes this estimate to be uniformally distributed in space, spatial statistics seeks to characterize the geographic distribution (pattern) of mapped data. The tabular data identify the location and population density of microorganisms sampled in a lake. Traditional statistical analysis shows an average density of about 430, assumed to be uniformally distributed throughout the lake. Spatial statistics incorporates locational information in mapping variations in the data. The pattern contains considerable variation from the average (shown as a plane) in the northwest portion of the lake.

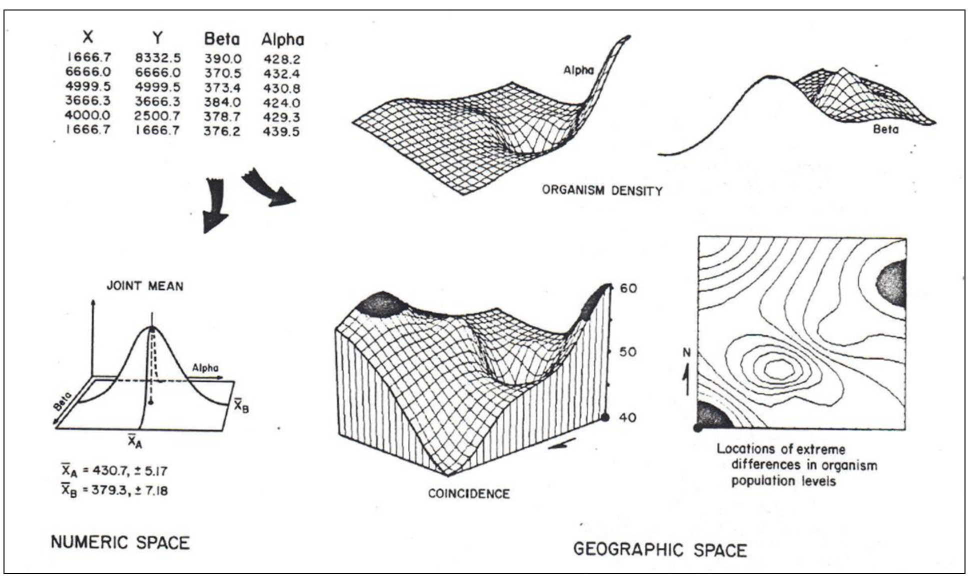

Figure 2. Multivariate Statistics. Maps characterizing spatial variation among two or more variables can be compared and locations of unusual coincidence identified. Traditional statistical analysis identifies the typical paired responses of two microorganism populations. This information assumes that the joint condition is uniformally distributed in space and does not identify where atypical joint occurrences might be found. Spatial statistics compares the two maps of population densities to identify two areas of unusually large differences in the populations.

Traditional and spatial statistics complement each other for decision-making. In both statistical approaches, the nature of the data is critical. For analyses similar to those in Figures 1 and 2, thematic values (“what” information) must identify variables that form continuous gradients in both numeric and geographic space.

Many forms of mapped data exist, including digital maps which, coupled with traditional mapping considerations—scale, projection, registration, resolution—can help foresters understand the potential and pitfalls of GIS applications. The quantitative nature of digital maps provides the foundation for a mathematical framework for analyzing maps.

Just as spatial statistics has been developed by extending concepts of conventional statistics, a spatial mathematics has evolved. This “map algebra” uses sequential processing of mathematical primitives to perform complex map analyses. It is similar to traditional algebra, in which primitive operations (add, subtract, exponentiation) are logically sequenced on variables to form equations; but in map algebra, entire maps composed of thousands of number represent the variables.

Most traditional mathematical capabilities plus an extensive set of advanced map-processing primitives emerge. Transpose, inverse, and diagonalize are examples of new primitives based on the nature of matrix algebra. Within map analysis, the spatial coincidence and juxtapositioning of values among and within maps create new operators such as masking, proximity, and optimal paths.

This set of map-analysis operators can be flexibly combined through a processing structure similar to conventional mathematics. The logical sequence involves retrieval of one or maps from the database, processing that data as specified by the user, creation of a new map containing the processing results, and storage of the new map for subsequent processing.

The cyclical processing is similar to “evaluating nested parentheticals” in traditional algebra. Values for the “known” variables are first defined, and then they are manipulated by performing the primitive operations in the order prescribed by the equation.

For example, in the equation $A = (B + C) / D$ the variables B and C are first defined and then added, with the sum stored as an intermediate solution. This intermediate value, in turn, is retrieved and divided by the variable D to derive the value of the unknown variable A. This same processing structure provides the framework for computer-assisted map analysis, but the variables are represented as spatially registered maps. The nuber contained in the solution map (in effect solving for A) are a function of the input maps and the primitive operations performed.

Cartographic modeling—This mathematical structure forms a conceptual framework easily adapted to a variety of applications in a familiar and intuitive manner. For example:

$$\%Change = \frac{\text{(new value – old value)}}{\text{old value}} \times 100$$

This equation the percent change in value for a parcel of land. In a similar manner, a map of percent change in land value for an entire town may be expressed in such GIS commands as:

```
COMPUTE NEWVALUE.MAP MINUS OLDVALUE.MAPFOR DIFFERENCE.MAP
COMPUTE DIFFERENCE.MAP TIMES 100 DIVIDED BY OLDVALE.MAP
   FOR PERCENTCHANGE.MAP
```

Within this model, data for current and past land values are collected and encoded as computerized maps. These data are evaluated as shown above to form a solution map of percent change. The simple model might be extended to provide coincidence statistics, such as

```
CROSSTAB ZONING.MAPWITH PERCENTCHANGE.MAP
```

for a table summarizing the spatial relationship between the zoning map and change in land value. Such a table would show which zones experienced the greatest increase in market value.

The basic model might also be extended to include such geographic searches as

```
RENUMBER PERCENTCHANGE.MAP FOR BIG_CHANGES.MAP
  ASSIGNING 0 TO -20 THRU 20 ASSIGNING 1 TO 20 THRU 100
```

for a map search that isolates those areas that experienced more than a +20 percent change in market value.

Analytic Toolbox

In traditional statistics and mathematics, pencil and paper are all this is needed to complete exercises. Use of a pocket calculator or computer enhances this experience by considering larger, more realistic sets of numbers. However, the tremendous volume of data involved in even the simplest map-processing task requires a computer for solution. Also, many of the advanced operations, such as effective distance measures and optimal paths, are so analytically complex that computer processing is essential.

Classroom needs are not being ignored. Software and materials supporting instruction in computer-assisted map analysis are appearing on the scene. Among these is the *Map Analysis Package (MAP)*, a widely distributed system developed at Yale in 1976 for mainframe computers. Over 90 universities, many with natural-resource programs, have acquired these materials.

The *Professional Map Analysis Package (pMAP)* is a commercially developed implementation of MAP for PCs (Cooney and Tucker 1986). The *Academic Map Analysis Package (aMAP)* is a special educational version of pMAP available from Yale for classroom use. All of the student exercises can be performed on a standard IBM PC or similar system.

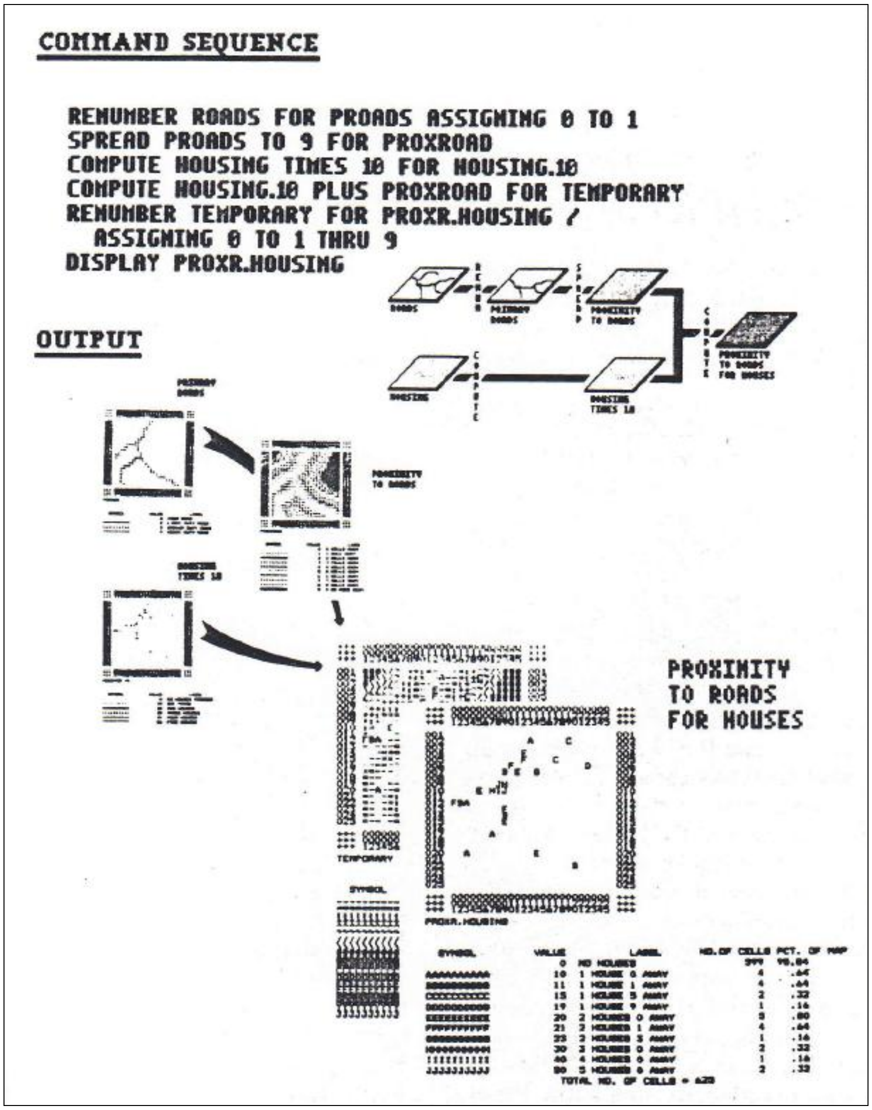

Figure 3. Student Exercise. Proximity to major roads is first determined and then combined with housing information. Two-digit codes identify the number of houses (first digit; tens) and the distance to the nearest major road (second digit; ones).

**Joseph K. Berry is associate professor and associate dean, School of Forestry and Environmental Studies, Yale University, New Haven, CT.*

Part III

GIS