GEOGRAPHIC INFORMATION SYSTEMS

Michael F. Goodchild

A geographic information system is designed to capture, store, display, communicate, transform, analyze, and archive georeferenced information, that is, information tied to specific locations on the Earth's surface. Geographic information systems enhance and to some extent replace the traditional role played by maps, but are also capable of handling information in the form of satellite images of the Earth's surface, as well as information from surveys and administrative records that have been georeferenced. They are increasingly used in the social sciences to support research based on cross-sectional data, or studies for which geographic location and context are important and useful.

I. Introduction

The origins of geographic information systems can be traced to the mid 1960s. Early computers were designed primarily for numerical processing, following the lead of Babbage and others. But by 1965 other applications had begun to appear, supported in part by the development of specialized peripherals, notably pen plotters and map digitizers. About this time the Canada Land Inventory faced a massive problem of map data processing: how to take the very large number of maps created to document Canada's underutilized land resource, and to produce tables of the amounts of land available for various types of development and use. Measurement of area from maps had always been time consuming, tedious, and inaccurate when performed by hand. But if the maps could be converted to digital form, simple algorithms would allow areas to be measured and tabulated electronically. The Canada Geographic Information System (CGIS) was thus a response to a well-defined need.

By the 1980s, commercial GISs had begun to appear, offering a wide range of functions that in various ways were too complex, tedious, inaccurate, or expensive for humans to perform by hand. These included simple measurement of area and length, transformations needed to alter data formats, simple statistical analyses such as the calculation of means and standard deviations, and a host of more complex and sophisticated methods generally termed spatial analysis. In addition, GISs were provided with advanced capabilities for data display, including mapping and various forms of data visualization. The scientific community was quick to recognize the potential of GIS, and through the 1980s and 1990s GIS emerged as an indispensable tool for research in any discipline dealing with the surface of the Earth, or the near-surface. In the social sciences some of the first applications were in archaeology (*), but political scientists, criminologists, demographers, and epidemiologists have also been prominent early adopters. For extensive bibliographies covering applications of GIS and spatial analysis in the social sciences, as well as information on sources of tools, data, and other resources see the web site of the Center for Spatially Integrated Social Science (www.csiss.org).

In recent years GIS has undergone significant transformation, as applications have emerged that go well beyond the early notion of a digital assistant performing tasks that humans find difficult. The advent of the Internet and the WWW had by 1995 induced a sharp change of perspective, in which GIS was viewed as a means for sharing information between people, in addition to its more traditional role. Many web sites were created offering to supply visitors with geographic data sets, or to create maps on demand, or to perform simple GIS services to user specifications, using data provided by the user or by the site. Maps are compelling ways of presenting information, and spatial analysis has been reinterpreted in recent years as a set of methods by which one person adds value to information, by making visible what might otherwise be invisible to another person, thus strengthening the message. For a comprehensive survey of Internet GIS see the introductory GIS text by Longley *et al.* (2001).

Most recently, advances in technology have brought the promise of GIS that is no longer confined to the office, but carried into the field in the form of portable and wearable devices. Wireless communication is available to download and upload data to and from Internet sites, and sufficient power is now available in portable devices to support virtually any GIS operation. The advent of field GIS offers to revolutionize the nature and practice of field work, in social surveys and other field-based social science.

II. Representation

At the heart of a GIS is a system of representation, by which features in the real world are coded in the binary alphabet of the digital computer. GIS representations typically include three aspects of real-world features: their locations on the Earth’s surface, using a convenient coordinate system such as latitude and longitude; their attributes, or the things that are known about them; and any relationships of importance between them. Examples of relationships include adjacency, such as the adjacency that might exist between two neighborhoods; and connectivity, such as the connections that might exist between parts of a street network.

Attributes provide much of the richness of a GIS representation, especially in the social sciences. Reporting zones such as census tracts might carry large numbers of descriptive attributes, created from the summary tables provided by the Census, such as average income or percent unemployed. Individual points representing the locations of individuals in a sample survey might carry as attributes the information collected in the survey from each individual.

Underlying the representation of geographic variation are two distinct conceptualizations. In the first, the features on the Earth's surface are discrete objects, much as a tabletop might be littered by books, pens, or coffee mugs. Discrete objects can overlap, and empty space can exist between them. This *discrete object* view is particularly appropriate in the representation of moving or persistent objects, such as individual people or vehicles. Objects can be represented as points, lines, or areas depending on their size in relation to the geographic extent of the representation. Areas are represented as sequences of points connected by straight lines (polygons), and lines similarly (*polylines*).

The second is the *field* view. In this conceptualization, geographic variation is characterized by the continuous variation of a number of variables, each a function of position. Elevation is of this nature, since it can be determined at every point on the surface of the planet, and so is population density. Note that a field of population density is a generalization of a discrete object view, in which each person is a point, surrounded by empty space. This process of generalization is termed *density estimation* (*), and is an important function of a GIS.

Somewhat confusingly, representations of fields must also be built from points, polylines, and polygons, using one of six recognized methods. A field can be represented by values at sample points, distributed either irregularly, or regularly on a grid. It may be represented by a set of non-overlapping polygons that collectively exhaust the space, each having the mean value or the integral of the field over the area as an attribute (*e.g.*, mean population density, or total population). Polylines can also be used, if the isolines (contours) of the field variable are digitized. Finally, a field can be represented as a mesh of non-overlapping triangles, with values of the field associated with each triangle vertex, and linear variation within each triangle, in the model known as the Triangulated Irregular Network (TIN).

Social scientists are likely to use both discrete object and field conceptualizations. Individual crimes, archaeological artifacts, or deaths from a disease are viewed as discrete objects. But aggregate variables such as those available in census summary statistics are more likely to be viewed in terms of fields that have been averaged within reporting zones, which must behave according to the rules of a polygon-based field representation: polygons may not overlap, and must collectively exhaust the space. Other phenomena likely to be conceptualized as fields include crowding, ambient noise, or slope.

III. Georeferencing

A. Measuring location

A system for accurately identifying location on the surface of the Earth is an essential component of any GIS representation. The Meridian Convention of * established latitude and longitude as the universal standard for georeferencing, based on measurements from the Greenwich Meridian and the Equator. Unfortunately the Earth is not a perfect sphere, and has been approximated by a variety of mathematical functions over time and in different parts of the world, each of which potentially leads to a slightly different latitude and longitude. Currently the system or *datum* of choice in North America is the North American Datum of 1983 (NAD83), but other datums may be encountered, such as the earlier NAD27, and the datums used in other countries. All of this means that it is impossible to determine location exactly, and variations of as much as 200m on the ground may exist between determinations of latitude and longitude using different datums. Modern GIS software makes it possible to convert easily from one datum to another, but nevertheless social scientists will occasionally encounter datum differences. In addition to latitude and longitude, georeferencing often makes use of methods for projecting the Earth's curved surface onto a plane, and associated planar coordinate systems. These include the Universal Transverse Mercator (UTM) system, the NATO standard widely used by national mapping agencies. UTM consists of 60 distinct map projections and coordinate systems, each designed to provide accuracy within a six-degree zone of longitude (for example, Zone 11 runs from 120 West to 114 West). U.S. users of GIS may also encounter the State Plane Coordinate systems, adopted by each state for high-accuracy survey; UK users may be familiar with the National Grid; and many other countries also have national grids.

The use of map projections and the practice of flattening the Earth were essential in the era of paper maps, but are somewhat paradoxical in a technology based on digital representation, since nothing in a digital computer requires flattening. But paper maps are still a very important product of GIS, as well as an important source of input. Moreover most social science research is conducted over small areas, where the distortions introduced by flattening the Earth are small, and where the benefits of being able to work in a simple rectangular coordinate system are compelling. Distance and area are comparatively difficult to compute from latitude and longitude, and much easier from planar coordinates expressed in meters.

B. Street addresses and placenames

Although latitude and longitude and planar coordinate systems are convenient, they are essentially inaccessible to the average person, who is more likely to remember and report location in more familiar ways, such as through the use of placenames or street addresses. GIS techniques have been developed for easy recognition and conversion, in effect providing an interface between the vague world of human discourse and the precise world of georeferencing. These processes are known by a variety of terms, including geocoding and address matching.

Consider, for example, a survey conducted with a sample of households, in which each household is identified by street address. In order to map the results of the survey, it is necessary to identify the coordinates corresponding to each address. This can be done by making use of one of a number of available *street centerline* data sets, which represent streets as polylines, and include the ranges of addresses on each side of each street segment (the stretch of a street between two adjacent intersections). The first street centerline data sets for the U.S. were developed in conjunction with the census, and distributed free to interested users. The TIGER (Topologically Integrated Geographic Encoding and Referencing) data set, which was first developed for the 1980 census, stimulated a substantial industry concerned with adding value and exploiting the applications of TIGER.

Unfortunately geocoding is not as simple as it sounds. Specific addresses must be interpolated between the addresses of the endpoints of each segment (*e.g.*, 951 would be placed halfway between 901 and 999 in the 900 block), which is a dubious assumption in rural areas, where street addresses may not even exist, and in condominium complexes and townhouse developments. In Japan, houses are often numbered by date of construction rather than order along a street. Spelling variations, the lack of a standard syntax, and other problems typically result in success rates of less than 80% for automated geocoding, and require expensive human intervention, even in urban areas of the U.S.

Placenames also provide the basis for a second method of geocoding, although at much coarser scale. An index providing the basis for translation between placenames and coordinates is known as a *gazetteer*, and in recent years there has been much interest in these data sets, in conjunction with information retrieval. Suppose, for example, that one wants to conduct a study of a particular city, and to collect and assess any data that may be available. Many large archives of GIS data, such as the Geography Network (www.geographynetwork.com) developed and maintained by Environmental Systems Research Institute (ESRI, a major vendor of GIS software) or the Alexandria Digital Library (www.alexandria.ucsb.edu), an online map and imagery library developed by the University of California, Santa Barbara, allow users to search for data within the archive by starting with a placename, and using the services of a gazetteer to translate it into a latitude and longitude reference. This reference is then used, along with other user-supplied criteria, to search the archive for suitable data, which can then be retrieved, examined, and downloaded. Other services based on gazetteers have also emerged; *geoparsing*, for example, allows large masses of text to be searched for placenames, which are then used to establish geographic context.

IV. Visualization

GIS is an inherently visual technology, inviting its users to take advantage of the power and effectiveness of data when rendered visually (Tufte, *). Maps are the traditional way of visualizing geographic information, and GIS owes much to the legacy of cartography, the science and art of map-making, and to successful efforts by cartographers to systematize the discipline (*e.g.*, Robinson *). Summary or aggregate data associated with polygons is often displayed in the form of *choropleth* maps, using shading and other forms of polygon fill to distinguish values of the variable of interest. Point data are typically displayed as symbols, again with color or symbol size used to distinguish attribute values. Today's commercial GIS software supports a vast array of possible mapping techniques, including contour or *isopleth* maps of fields, and cosmetic features such as legends, north arrows, annotation, and scale bars.

It is important, however, to recognize the fundamental differences between GIS displays and paper maps, and the advantages that the digital technology provides over traditional methods. First, GIS has changed map-making from an expensive and slow process carried out by a few highly trained cartographers, to a fast and cheap process available to all. Today, anyone armed with a computer, data, and simple software can produce compelling maps (and also misleading maps; Monmonier, *).

Second, GIS displays are inherently dynamic and interactive, whereas paper maps are essentially immutable once created. GIS displays can protray changes through time, or allow users to zoom and pan to expose new areas or greater detail. More than one display can be created simultaneously on a single screen. Maps can also be displayed beside other forms of presentation, such as tables, and tables and maps can be linked in interesting ways (*e.g.*, clicking on a polygon in a map display can highlight the corresponding row in a table). The term *exploratory spatial data analysis* has been coined to describe the interactive exploration of GIS data through maps and other forms of presentation.

V. Spatial analysis

While the display of geographic information in the form of maps can be powerful, the true power of GIS lies in its ability to analyze, either inductively in searching for patterns and anomalies, or deductively in attempts to confirm or deny hypotheses based on theory. The techniques of analysis available in GIS are collectively described as *spatial analysis*, reflecting the importance of location. More precisely, spatial analysis can be defined as a set of techniques whose results depend on the locations of the objects of analysis. This test of locational dependence clearly distinguishes techniques of spatial analysis from more familiar statistical techniques, such as regression, that are invariant under relocation of the objects of analysis. Thus GIS can be understood as a technology implementing methods of spatial analysis, just as the familiar statistical packages implement methods of statistical analysis, or word processors implement the process of writing.

Over the past five or six decades a vast array of methods of spatial analysis have been devised, for the detection of patterns and anomalies, and the testing of hypotheses. Many texts, such as that by Bailey and Gatrell (*), organize spatial analysis according to the types of data for which they are designed: techniques for the analysis of point patterns, or polygon data, for example. Longley *et al*. (2001*) use a somewhat different approach based on the objectives of analysis, and that approach is followed here in a brief review. Interested readers are referred to the more detailed discussion in that source.

A. Query

Interactive displays allow users to determine answers to simple queries, such as “What are the attributes of this object?”, or “Where are the objects with this attribute value?” Some queries are best answered by interacting with a *map* view, by pointing to objects of interest. Other queries are better answered by interacting with a *table* view, by searching the table for objects whose attributes satisfy particular requirements. A *histogram* view is useful to find objects whose attribute values lie within ranges of interest, and a *scatterplot* view allows objects to be selected based on comparisons of pairs of attributes. Finally a *catalog* view allows the user to explore the contents of the many data sets that might comprise a complete GIS project.

B. Measurement

Earlier discussion of the origins of GIS emphasized the importance of area measurement in the development of CGIS. Today, many other simple measurements are supported by GIS, including distance, length, terrain slope and aspect, and polygon shape. Measurements are typically returned as additional attributes of objects, and can then be summarized, or used as input to more complex forms of analysis.

C. Transformation

Many techniques of spatial analysis exist for the purpose of transforming objects, creating new objects with new attributes or relationships. The *buffer* operation creates new polygons containing areas lying within a specified distance of existing objects, and is used in the analysis of spatial proximity. The *point in polygon* operation determines which of a set of polygons contains each of a set of points, and is used to summarize point data by area, in the analysis of crime or disease. *Polygon overlay* determines the areas of overlap between polygons, and is often used by social scientists to estimate summary statistics for new areas that do not coincide with reporting zones (*e.g.*, to estimate populations of communities whose boundaries do not respect census zone boundaries). *Density estimation* also falls into this category, since it transforms point data sets into representations of continuous fields.

D. Summary statistics

Search for pattern is often conducted by computing statistics that summarize various interesting properties of GIS data sets. The *center* of a point data set is a useful two-dimensional equivalent to the mean, and *dispersion* a useful equivalent to the standard deviation. Measures of *spatial dependence* are used to determine the degree of order in the spatial arrangement of high and low values of an attribute. For example, rates of unemployment by census tract might be highly clustered, with adjacent tracts tending to have similarly high or similarly low values; or they might be arranged essentially independently; or adjacent tracts might be found to have values that are more different than expected in a random arrangement.

E. Optimization

A large number of methods have been devised to search for solutions that optimize specific objectives. These include methods for finding point locations for services such as libraries or retail stores; for finding optimum routes through street networks that minimize time or cost; for locating power lines or highways across terrain; or for planning optimum arrangements of land use. These methods are often embedded in *spatial-decision support systems*, and underpinned by GIS software.

F. Hypothesis testing

The sixth class consists of methods that apply the concepts of statistical inference, in reasoning from a sample to the characteristics of some larger population. Inference is well established in science, and it is common to subject numerical results to significance tests, in order to determine whether differences or effects could have arisen by chance because of limited sample size, or are truly indicative of effects in the population as a whole.

It is tempting to adopt statistical inference in dealing with geographic information, but several problems stand in the way. First, geographic data sets are often formed from *all* of the information available in an area of interest, and it is therefore difficult to believe that the data are representative of some larger universe, and that results can be generalized. Instead, one tends to believe in *spatial heterogeneity*, or the variation of conditions from place to place; in this context, it is difficult to regard a study area as representative of any larger area. Second, geographic data sets typically exhibit spatial dependence, which means that one object's attributes are unlikely to be truly independent of the attributes of its neighboring objects. The endemic presence of spatial dependence in geographic data has been called the First Law of Geography, and is often attributed to Waldo Tobler (*).

There are several possible solutions to this dilemma. First, objects might be chosen sufficiently far apart, allowing the assumption of independence to be acceptable, but this results in discarding data. Second, one might limit analysis to description of the data and area of study, and avoid any suggestion of inference about larger areas or other data sets, but this flies in the face of scientific tradition and the norms of peer review. Third, one might assume that the universe consists of all possible spatial arrangements of the data, in a form of randomization, particularly if the actual spatial arrangement of the data is the issue of interest. But this approach, while attractive, does not support inference in areas not covered by the data.

VI. Issues

As will be obvious from the previous section, the use of GIS raises numerous issues concerning the nature of geographic information, and inference from cross-sectional data. It is generally accepted that cross-sectional data cannot be used to confirm hypotheses about process, but it can certainly be used to reject certain false hypotheses, and to explore data in the interests of hypothesis generation. Although GIS has evolved from the static view inherent in paper maps, there is much interest in adding dynamics, and in developing methods of spatio-temporal analysis.

Uncertainty is a pervasive issue in GIS. It is impossible to measure location on the Earth’s surface exactly, and other forms of uncertainty are common also. For example, summary statistics for reporting zones are means or totals, and clearly cannot be assumed to apply uniformly within zones, despite efforts to ensure that census tracts are approximately homogeneous in socioeconomic characteristics. Results of analysis of aggregated data are dependent on the boundaries used to aggregate (the *modifiable areal unit problem*; *), and inferences from aggregated data regarding individuals are subject to the *ecological fallacy* (*).

Nevertheless, the outcomes of the widespread adoption of GIS in the social sciences over the past two decades are impressive. It is clear that GIS has brought new power to the analysis of cross-sectional data, and the integration of diverse data sets. It has also shifted the ground of social science to some degree, by increasing emphasis on local data, geographic variation, and highly disaggregated analysis, in contrast to the pervasive nomethetic approach of earlier decades.