as a sterile, atheoretical technology, as feared by some in geography, may not enter into the archaeological debate on the use of GIS but should be borne in mind nonetheless (Openshaw 1991; Taylor and Overton 1991).

Despite the many demonstrable advantages of GIS functionality for archaeological analysis, there are real limitations because of the quality of archaeological data and the current inability of GIS to handle imprecise spatial information and dynamic temporal data. Defining the spatial and temporal nature of an archaeological feature lies at the core of archaeology's main analytical axes, those of space and time. There is considerable fuzziness in defining archaeological site boundaries, especially at the regional level. There are also current limitations in incorporating information that is either poorly defined spatially or has no spatial description at all. For example, social and economic ideas of political, religious, and ritual influences, all of which were probably important factors acting on the visible landscape in prehistoric times, are important considerations. Similarly the dependence on multilayer GIS coverages to handle the temporal nature of data reflects current limitations in archaeology itself. The categorical classification of time by archaeologists and within GIS is unsatisfactory. Temporal GIS, and the ability to deal with temporal fuzziness would greatly enhance the potential of GIS for archaeological analysis. There are obvious concerns, therefore, regarding how archaeological data and ideas can be expressed in a system that demands spatial and temporal exactitude. Similarly, how human perception and cognitive space is handled within a GIS are important considerations for archaeologists. In much the same way as an understanding of aboriginal dreamscapes contributes to an understanding of aboriginal culture and its organization of the landscape, so archaeologists will need to consider how GIS can be used in modeling prehistoric cognitive landscapes. It is within these areas of research rather than in the repetition of existing applications that the full potential of GIS in archaeological analysis is to be found.

Acknowledgments

The authors are indebted to Joel Halverson for technical assistance. Gary Lock would like to thank the organizers of the Santa Barbara conference for funds towards expenses.

14

Geographic Information Systems and Spatial Analysis in the Social Sciences

Michael F. Goodchild

Although geographers have used them for centuries, the idea that they might be part of the formal or informal process of data analysis and scientific inference is still novel in many areas of the social sciences. In economics, for example, the dominant paradigm emphasizes search for theories and principles that apply uniformly within human society or its economies, and if geographic areas such as nation-states are used as units of analysis, their role is more or less that of statistical samples, each equally representative of a hypothetical population of all possible nation-states. Within this paradigm (and I do not wish to suggest that it is the only paradigm within economics or that it is necessarily inappropriate), a list of nation-states, ordered alphabetically, is more precise and useful for analysis than a map, since maps tend to lump data values into coarse classes of color or shade and make it difficult to identify the attributes of small nation-states like Luxembourg or St. Lucia. Tables, on the other hand, present each nation state's value clearly and precisely.

The chapters in this book represent a striking departure from this orthodoxy, since each describes research that depends in some way on a belief in the importance of spatial context or in looking at data from a spatial perspective. This might mean merely looking at a map, despite the disadvantages already noted. Or it might mean direct reference to such primitive spatial concepts as adjacency, proximity, direction, or coincidence in space. We might choose to call this examination of data in spatial context GIS or spatial analysis, and the methods used might be mathematically sophisticated, computationally intensive, or simply intuitive—the chapters in this book cover a wide range of options. Taken together, however, it is clear that they represent a striking departure from the aspatial ways of thinking described earlier. It seems appropriate in this concluding chapter to attempt a summary of the main features and assumptions of spatial analysis and its limitations and impediments. One of the aims of GIS is to make spatial analysis easier, more flexible, and more powerful. This chapter reviews some of these efforts, particularly those that are likely to be of use to social science.

For the purposes of this discussion, spatial analysis is defined as "a set of techniques whose results are dependent on the locations of the objects of analysis" (Goodchild 1987). This excludes the conventional application of many basic statistical techniques such as regression or the calculation of a mean, since one can freely relocate the objects of analysis without affecting their attributes. I assume that the definition includes descriptive and intuitive techniques, such as simple map display and the analysis carried out by the eye-brain when it receives a visual stimulus, as well as the more sophisticated applications of spatial statistics and operations research.

The first part of the chapter discusses the concepts that underlie spatial analysis and its implementation in GIS. The second part reviews widely recognized impediments to spatial analysis and some current research efforts to remove them. Although GIS is often seen as a tool for support of analysis and modeling in any science that concerns itself with distribution and differentiation over the surface of the earth, the emphasis in this chapter is strictly on social science applications.

Concepts of Spatial Analysis

Space as an Organizing Frame

Perhaps the simplest function of a map is as a way of organizing data. Just as a person's home street address is often used as a form of identification, so locating an artifact on a plan of a dig provides a way of coping with large amounts of information that might seem impossibly confusing if presented in the form of a table. In and of itself, spatial organization of data seems very unlikely to lead to any kind of insight. But once data is organized spatially (as described, for example, in Chapter 11), it becomes easy to display, and to gain insight through more sophisticated concepts of spatial analysis.

Spatial Anomalies

Other things being equal, we might expect geographic space to be uniform and geographic distributions to be similarly uniform in space. Random processes distribute objects over space with uniform density, as when raindrops fall on an area of dry pavement. But the human eye is remarkably efficient at scanning otherwise random patterns and detecting anomalies such as the clusters, or areas of anomalously high density, that often occur in the distribution of cases of a disease like cancer. Spatial anomalies can take the form of abnormally high or low densities or clusters of objects having similar values.

In practice, the human ability to detect pattern can be almost too powerful. People have spun elaborate stories around imagined patterns in the configuration of stars in the night sky and searched fruitlessly for explanations for apparent spatial clusters of certain diseases. So in addition to providing the ability to display data in ways that invite the eye to find patterns, a GIS might also usefully help the observer by providing access to objective tests of pattern based on sound statistical principles. In this way, GIS can provide tools that improve on or assist the observer's skills of human intuition. For an example of the use of GIS to search for clusters in patterns of disease, see Openshaw and coworkers (1987).

Spatial Coincidence

Insight comes not so much from the observation of an anomaly as from the recognition that it coincides spatially with some other factor and the implication that this factor is somehow involved in causing the anomaly. As soon as a cancer cluster is observed on a map, the observer begins a rapid and instinctive process of searching his or her experience for factors that might be more or less unique to that place. Is there not a high-voltage transformer in that area, or a factory that emits a large amount of unpleasant smoke? Each of us carries vast stores of such information about familiar neighborhoods, and we are powerfully equipped to retrieve and associate it given a suitable display of data. The map itself may offer clues if the disease locations have been plotted on top of a standard city street map. A GIS raises this potential to an even higher plane by automating the process of assembling and displaying the data and by allowing the user to overlay layers of information for areas that may be totally unfamiliar. At the same time, it may make it too easy for the observer to ignore the fact that spatial coincidence is not necessarily an indication of cause and thus to become too easily diverted by spurious geographic associations.

Spatial Proximity

The causal factor for a cancer cluster might be spatially coincident with the cluster, or it might occur some distance away if we are willing to assume that the cluster and the cause are somehow linked over space. They might be linked through the atmospheric movement of pollution if the cluster is downwind of the source of pollution. Because disease occurrences are typically recorded at place of residence, the linkage between pattern of disease and causal factor might be the journey to work, with a cluster spread over the dormitory neighborhoods near a particular contaminated workplace.

One of the best-known examples of this type of spatial thinking is the work of Dr. James Snow on a cholera outbreak in the Soho district of London in the 1800s (Gilbert 1958). By making a map, Dr. Snow was able to show that deaths in the outbreak clustered around a particular public water pump, thus implying that contaminated water from the pump was the source of the infection. In this case the link between cluster and cause was the collection of water; Dr. Snow was able to show conclusively that the outbreak was confined to those who habitually drew water from that pump and approximated by those who lived closer to that pump than to any other.

Spatial Dependence

Closely related to the concept of spatial proximity is spatial dependence, or the tendency for "things that are close together in space to be more alike than more distant things" (Tobler, 1970). The presence of spatial dependence is not in itself of great interest, since virtually all geographically distributed phenomena display it to some degree and it is impossible to imagine a world in which spatial dependence is absent; the slightest movement in any direction would produce a complete change of scene. But some phenomena display greater spatial dependence than others, and the distance over which dependence is observed varies also. It is sometimes interesting to compare phenomena according to their patterns of spatial dependence or to observe how spatial dependence changes through time.

On the one hand, spatial dependence is an interesting property of geographically distributed phenomena that can provide useful insights, as Arnold and Appelbaum show in Chapter 3. On the other hand, many statistical tests assume that observations are drawn randomly and independently, and this assumption is clearly invalid for much geographic data. In much of the literature of spatial analysis, we find spatial dependence being viewed as a problem to be treated and removed rather than as a useful source of insight, since its presence causes such problems for conventional statistical reasoning.

Consider the following example. Two points are antipodal if a line between them passes exactly through the center of the earth. The antipodal point of Santa Barbara, California, at longitude 120° west and latitude 34° north, is a point at latitude 34° south and 60° east, located south of Mauritius in the Indian Ocean. Roughly one-third of the earth's surface is land and a smaller proportion is antipodal land, in the sense that pairs of antipodal points are both on land. If land is randomly distributed over the earth's surface, one would expect this proportion to be roughly one third times one-third, or one-ninth. In fact, the proportion is very much smaller, the only substantial areas of antipodal land being in southern South America and northern China. This is surprising until one realizes that the distribution of land displays a very high degree of spatial dependence, since it is caused by the movement of very large continental masses. Spatial dependence and its effects are important issues in reasoning about spatial patterns and their meaning.

Spatial Heterogeneity

Anselin (1989) argues that two properties make spatial data "special"—spatial dependence and spatial heterogeneity. Both to some degree question the assumptions of traditional statistical inference. Because of spatial dependence, the number of truly independent observations in a given sample taken over geographical space is normally lower than one might expect—spatial dependence lowers the effective number of degrees of freedom in a test. Spatial heterogeneity, on the other hand, invalidates the assumption that all cases in a sample are drawn randomly from the same population.

Consider, for instance, a sample of early settlement sites in a small region of South America. It is likely that one or more gradients exist within the region—for example, the settlements might be dispersed over an area sloping upward away from the shore of a lake. The study is confined to the area covered by a single air photo and includes an examination of every known site within the area. These are treated as a sample and subjected to standard statistical tests.

Unfortunately, there is an inherent conflict in this example between actual practice and the assumptions of the test. Rather than a sample drawn randomly from a population, we have analyzed all of the cases that occur within a defined area, and the area is differentiated by a gradient. It is not clear what population is being described when inferences are drawn from analysis of the sample. Is it the set of all early South American settlements, the ones in a region of which this study area is typical, or the settlements that might have occurred had history repeated itself? If it is not possible to identify the population, then what is the point of statistical inference?

In practice, spatial heterogeneity is most problematic when one considers the effects of the choice of study-area boundary on the results of the analysis. If the air photo had been positioned further upslope, the mix of settlements in the sample would have been different and the conclusions would have been affected. Thus, one consequence of spatial heterogeneity—the tendency for conditions, particularly the parameters of models, to vary geographically—is an unwanted dependence of the results of investigation on the definition of the study area. In traditional statistical thinking, the choice of study area merely defines one random sample from a homogeneous population rather than another. For geographical phenomena, however, it is necessary to think much more carefully about the effects of the choice of study area and about the nature of the population that is the subject of statistical inference.

Issues in Spatial Analysis

Spatial analysis, or more generally the examination of data in their spatial context, seems to offer the potential for powerful new insights into the processes that occur on the geographical landscape, and there is ample evidence of this in the pages of this book. In recent years, the growth in sophistication and widespread availability of GIS has led to a revival of interest in many of the methods of spatial analysis that can be found in a rich literature dating back over many decades (Goodchild et al. 1992; Heywood 1990). In this climate of enthusiasm, it is easy to forget that spatial analysis was the subject of a series of powerful critiques, beginning in the 1970s (Gregory 1978; Johnston 1987), and that these extended to spatial thinking in general. The previous two sections have already identified ways in which spatial thinking may be more complex or problematic than first appearances would suggest. This section reviews some of the other impediments to spatial analysis, and this leads to the subsequent section on recent efforts to remove or reduce these impediments through the use of GIS.

Form and Process

Geographical information is primarily cross-sectional—that is, it represents a "snapshot" at a point in time. Mechanisms of cause and effect, on the other hand, are strongly temporal, and it is consequently difficult or impossible to deduce cause and effect from patterns observed in geographical data. Spatial analysis is primarily analysis of form, whereas understanding requires analysis of process.

There are two simple but distinct ways of producing a cluster of points in space. One is by contagion—the presence of one case or carrier makes other cases in the vicinity more likely. Alternatively, a cluster can arise because the density or likelihood of occurrence varies geographically in response to some causal factor. Although a number of tests for clustering in a spatial point pattern have been devised (Ripley 1981), none of them is capable of distinguishing between contagion and varying density as the causal mechanism. On the other hand, it is easy to resolve them if data is available on the temporal sequence of occurrence.

Although the clustering of cholera cases around the Broad Street pump strongly suggested water from the pump as the cause of the disease, many other mechanisms could explain the observed geographic pattern equally well. Transmission by human contact would also lead to geographic clustering, as would a causal factor in the environment. It was only when Dr. Snow put a lock on the pump and the outbreak came to a rapid end that the drinking water hypothesis was truly confirmed. In such cases, it is probably better to see spatial analysis as a source of possible hypotheses about cause rather than as a means of confirmation.

The Ecological Fallacy

The geographical landscape is almost unbelievably complex; the more closely one looks at most phenomena, the more detail one sees, apparently ad infinitum. In mapping, this is expressed as the relationship between scale and generalization—a map at a scale of 1:24,000 shows more phenomena and more detail than a map at a scale of 1:100,000. Fractals (Mandelbrot 1982) provide perhaps the only theoretical framework for modeling the effects of scale and generalization on the natural landscape.

A similar need for generalization occurs in dealing with social data. Although data are often collected on individuals, for reasons of practical necessity as well as protection of privacy it is more often distributed, mapped, and analyzed on the basis of reporting zones such as precincts, wards, census tracts, counties, or nation-states. Choropleth maps present the characteristics of each reporting zone as if they applied uniformly to the zone's entire area, inviting the map reader to assume that the zone is more homogeneous than it really is.

Because much data are available in that form, much spatial analysis in social science uses information aggregated by reporting zone. For example, one might compare the level of unemployment in each county to the percentage of people in the county who have not completed high school. Robinson (1950) was among the first to draw attention to the fact that a positive correlation between these two aggregate measures is not necessarily evidence that individuals without high school graduation are more likely to be unemployed. The correlation indicates that unemployed and people without high school education are likely to be found in the same areas, but it does not indicate that they are necessarily the same people. It is an ecological fallacy to make a false conclusion about individuals from an observation about spatially aggregated data.

Modifiable Areal Units

Closely related to the ecological fallacy is what is often called the modifiable areal unit problem (MAUP). While the results of aggregate analysis can be falsely imputed to individuals or lower levels of aggregation, they are also strongly dependent on the particular choice of reporting zones, such that it may be possible to manipulate the outcome of the analysis by manipulating the zones. Openshaw (1983) was the first to draw attention to this problem in spatial analysis, and Openshaw and Taylor (1979) provide a striking example. Using data on the percent over 65 and percent registered Republican voters in Iowa counties, they were able to produce virtually any result from perfect positive to perfect negative correlation by manipulating the reporting zone boundaries. Fotheringham and Wong (1991) have made a series of detailed simulations of the effects of zone boundaries on a variety of forms of spatial analysis.

Recent Developments in GIS

Because of the valuable applications of GIS in the management of utility company operations, land records, and natural resources, the growth of interest in it during the past two decades has been driven at least as much by its use for information management and inventory as its capabilities for spatial analysis. Maguire (1991) argues that spatial analysis is only one of three traditions of GIS. But considering only its scientific applications, it is often argued that GIS is a set of computerized tools for implementing the techniques of spatial analysis, just as statistical packages are tools for implementing statistical analysis.

The capabilities of GIS for spatial analysis have grown substantially over the past few years. Versions 6.1 and 7.0 of ARC/Info (developed and distributed by Environmental Systems Research Institute, Redlands, CA) added significant new functions for spatial analysis, particularly in the types of network analysis illustrated by Ruggles and Church in their chapter, the measurement of spatial autocorrelation, and the modeling of spatial interactions. IDRISI (developed and distributed by The IDRISI Project, Clark University, Worcester, MA) has added new capabilities for analysis in support of decisionmaking and the management of error and uncertainty in spatial data. Other GIS with significant spatial analysis capabilities include TransCAD and GisPlus (Caliper Corporation, Newton, MA), and MapInfo (MapInfo Corporation, Troy, NY).

The next sections review efforts that are in place or under way to address the impediments to spatial analysis identified earlier.

Spatial Heterogeneity

Traditional uses of statistical methods have emphasized the power of tests to confirm or deny hypotheses by following rigorous procedures. In the early days of computing, many of these procedures were implemented on mainframes, taking over much of the tedium of mechanical calculations. Beginning in the 1970s, however, with the introduction of interactive computing, new methods of analysis based more on exploration than confirmation began to emerge. Interactive techniques, which now dominate computing, are ideally suited to dealing with problems of spatial heterogeneity. Rather than analyzing all data within a study area, a user of an interactive system can explore the characteristics of data using user-defined windows. In the example of South American settlement discussed earlier, an interactive user could "brush" the map with a window, determining the characteristics of the settlements lying within the window's current position. Results then become a function of the (user-controlled) window position rather than a function of the (uncontrolled) study area boundary. Monmonier (1989) was among the first to draw attention to the potential of geographical brushing.

A related development is the use of logically linked windows in packages designed for advanced spatial analysis, such as REGARD (Haslett et al. 1991), or the work of MacDougall (1992). Here the user is able to display several representations of data simultaneously, such as a map and a histogram or a map and a scatterplot, in separate windows on the screen. Moreover, the windows are linked, so that cases highlighted in one window are automatically highlighted in all other windows. For example, pointing to an outlying case on a scatterplot causes its position to be highlighted on the linked map, and "lassoing" a group of cases on a map causes them to be highlighted on the linked histogram.

Spatial Dependence

Several GIS, including IDRISI and ARC/Info, now offer the ability to compute indices of spatial dependence. The Geary and Moran indices of spatial autocorrelation are perhaps the most widely used, although the variogram, which computes spatial dependence as a function of distance, may provide more insight for suitable kinds of data (Goodchild 1986). Unfortunately, although it has been relatively easy to add the ability to compute such indices, it has proven much more difficult to provide the user with help in interpretation or insight into the consequences of spatial dependence. A wide range of models of spatially dependent data have been proposed, as well as modifications to statistical tests when spatial dependence is present, but these remain largely inaccessible to the nonspecialist user. Anselin's SpaceStat (1992) includes many of these techniques, but it must be linked to a GIS through a suitable data-transfer mechanism.

Animation

As noted earlier, the time dimension to data is often essential if insight is to be gained into causal mechanisms. Maps, on the other hand, are static, as are many geographical data. Recently, efforts have been made to introduce the time dimension into spatial databases (Langran 1992), and to develop GIS that allow animation (Buttenfield et al. 1991). An animated display of the temporal sequence of deaths in the Snow cholera data, for example, could have led to much richer insight into the transmission mechanism and would presumably have eliminated the possibility of direct transmission between individuals. Unfortunately, many existing GIS use designs that have been optimized for the display of static data and give very poor performance in animations.

Reporting Zone Effects

The literature on the ecological fallacy and MAUP gives conspicuously little help in finding solutions to either of these problems. No simple techniques can be applied to aggregate data to determine whether a fallacious inference is being made, and there are no methods for removing the effects of zone boundaries on the results of analysis. In general, the less aggregated the data the lower the chance of an ecological fallacy. Similarly, lower levels of aggregation are observed to produce a smaller range of results when reporting zones are manipulated.

The GIS can offer two forms of assistance in helping to reduce the impact of reporting zone effects. First, the power of modern computers means that it is almost always practical to process, analyze, and display data at the lowest possible levels of aggregation. In the past, the labor-intensive manual methods used to handle and map data meant that aggregation was often essential. In today's computer-rich research environment, volume of data has virtually no effect on cost. Martin and Bracken (1991) provide some striking illustrations of the insights gained by processing U.K. census data at the lowest available level of aggregation.

Second, while they cannot be removed, reporting zone effects can at least be explored by using the capability of GIS to repeatedly reaggregate and reanalyze data. Although this can be done only at levels of aggregation above the lowest available, the results may still be useful in indicating the importance of reporting zone effects in a given analysis.

Conclusion

In this concluding chapter I have tried to identify the consistent themes of spatial analysis that run through many of the earlier chapters of this book. The GIS can bring useful concepts and techniques to bear on data and can lead to interesting insights. On the other hand, while the eye-brain combination is an extraordinarily powerful processor of spatial information, it is easily misled, particularly when it is aided by statistical techniques whose assumptions do not match the nature of geographical data. As Anselin argues, spatial data are special, and their power to yield insights must be balanced by an awareness of the issues identified in this chapter.

Many techniques of spatial analysis are already implemented and available in packages like ARC/Info or IDRISI. Newer techniques, and particularly those which exploit the capabilities of interactive computing for exploration of data and animation, are available to date only in specialized research packages. But many development projects are under way, many of them in connection with NCGIA's Research Initiative 14 on GIS and Spatial Analysis (Fotheringham and Rogerson 1992), and it is likely that the availability of interesting and powerful techniques will improve rapidly in the next few years.

References

Abasolo, J.A.
1974 Carta arqueológica de la provincia de Burgos. I Partidos judiciales de Belorado y Miranda de Ebro. *Studia Arqueológica* No. 33, Valladolid.

Abasolo, J.A. and I. Ruiz Velez
1977 *Carta arqueológica de la provincia de Burgos.* Partido Judicial de Burgos. Burgos.

Acheson, S.R.
1991 In the wake of the ya'aats' xaatgaay ("iron people"): A study of changing settlement strategies among the Kunghit Haida. Unpublished Ph.D. dissertation. Oxford, University of Oxford.

Adams, E.C.
1985 *Annual Report of Test Excavations at 5MT765, Sand Canyon Pueblo and Archaeological Survey in T36N, R18W, Sections 12 and 24, and T36N, R16W, Sections 29 and 3.* Crow Canyon Archaeological Center. Submitted to Bureau of Land Management, San Juan Resource Area Office, Durngo, Colorado.

Adams, E.C.
1986 Report to the National Geographic Society on excavations at Sand Canyon Pueblo, an Anasazi ceremonial center in Southwestern Colorado. Ms. on file, Crow Canyon Archaeological Center, Cortez, Colorado.

Adler, M.A.
1994 Population aggregation and the Anasazi social landscape: A view from the Four Corners. In *The Ancient Southwestern Community: Models and Methods for the Study of Prehistoric Social Organization,* edited by W.H. Wills and R.D. Leonard, pp. 85–101. University of New Mexico Press, Albuquerque.

Adler, M.A.
1992 The upland survey. In *The Sand Canyon Archaeological Project: A Progress Report,* edited by W.D. Lipe, pp. 11–23. Occasional Paper 2. Crow Canyon Archaeological Center, Cortez, Colorado.

Adler, M.A. and M.D. Varien
1991 The changing face of the community in the Mesa Verde region, A.D. 1000–1300. Paper presented at the Third Anasazi Symposium, October, 1991, Mesa Verde National Park, Mesa Verde, Colorado.

Aldenderfer, M.
1990 The analytical engine: Computer simulation and archaeological research. In *Archaeological Method and Theory,* vol. 3, Schiffer, M. (ed.), pp. 195–248. University of Arizona Press, Tucson.

Aldenderfer, M.
1992 The state of the art: GIS and anthropological research. *Anthropology Newsletter* 33(5):14.

Allen, K.M., S.W. Green, and E.B.W. Zubrow (editors)
1990 *Interpreting Space: GIS and Archaeology.* Taylor and Francis, London.