4. SPATIAL ANALYSIS: METHODS AND PROBLEMS IN LAND USE MANAGEMENT

Michael F. Goodchild

INTRODUCTION

Management of land use is an essentially spatial activity, since the use of land is intimately tied to the land's location, to the properties of the site, and to the site's geographic situation. To manage land wisely we need to know as much as possible about the land itself, including information on prior uses, drainage, soils, slope, and many other site factors; and also about its situation with respect to the properties of neighbouring land and details of nearby relevant activities. This is ideal territory for geographic information systems, with their ability to store representations of the Earth's surface and its characteristics as digitized maps and images. It is also prime territory for spatial analysis, the subject of this chapter.

The word *analysis* connotes transformation or manipulation in the interests of exposing things that are not otherwise apparent. We analyze data to find patterns or anomalies, using tools that make such patterns and anomalies conspicuous, or help us to search for and identify them in systematic ways. If this process is speculative we use the term *exploratory*. Analysis is also used to determine whether data match some prior theory or expectation, in which case the term *confirmatory* is appropriate (Haining, 1990). But in many cases it may be difficult to see the difference between the two approaches. Moreover, the eye and the brain are enormously powerful mechanisms, especially in the case of spatial information portrayed visually, in the form of a map or picture, because they are excellent detectors of patterns of various kinds. In this context analysis is best seen as a collaboration, between tools for manipulation and transformation, and the human eye and brain. Analysis tools can augment in tasks where the eye and brain are incapable or less efficient, such as making precise measurements or conducting rigorous statistical tests; and they can also provide an essential check in tasks where the eye and brain are easily misled — optical illusions are excellent examples of how the eye and brain can be fooled.

This notion of collaboration between tools and human intuition is important, because it is easy to be seduced by technology into believing that human intuition is inherently suspect, and only tools are capable of reaching useful conclusions. In this chapter I take the view that looking at a map, using the power of the eye and brain, is just as much spatial analysis as using the most sophisticated and complex statistical tools.

Spatial analysis can be defined as analysis that requires knowledge of site and geographic situation, in other words, that requires knowledge of the locations of the objects of analysis on the Earth's surface. Another way of expressing that idea is to say that the results of spatial analysis are *not* invariant under changes in the locations of the objects of analysis (Figure 1). By contrast, the results of the vast majority of methods of statistical analysis, such as regression, *are* invariant if the objects are moved around. It follows that spatial analysis is impossible if locations are not known; or in a computational

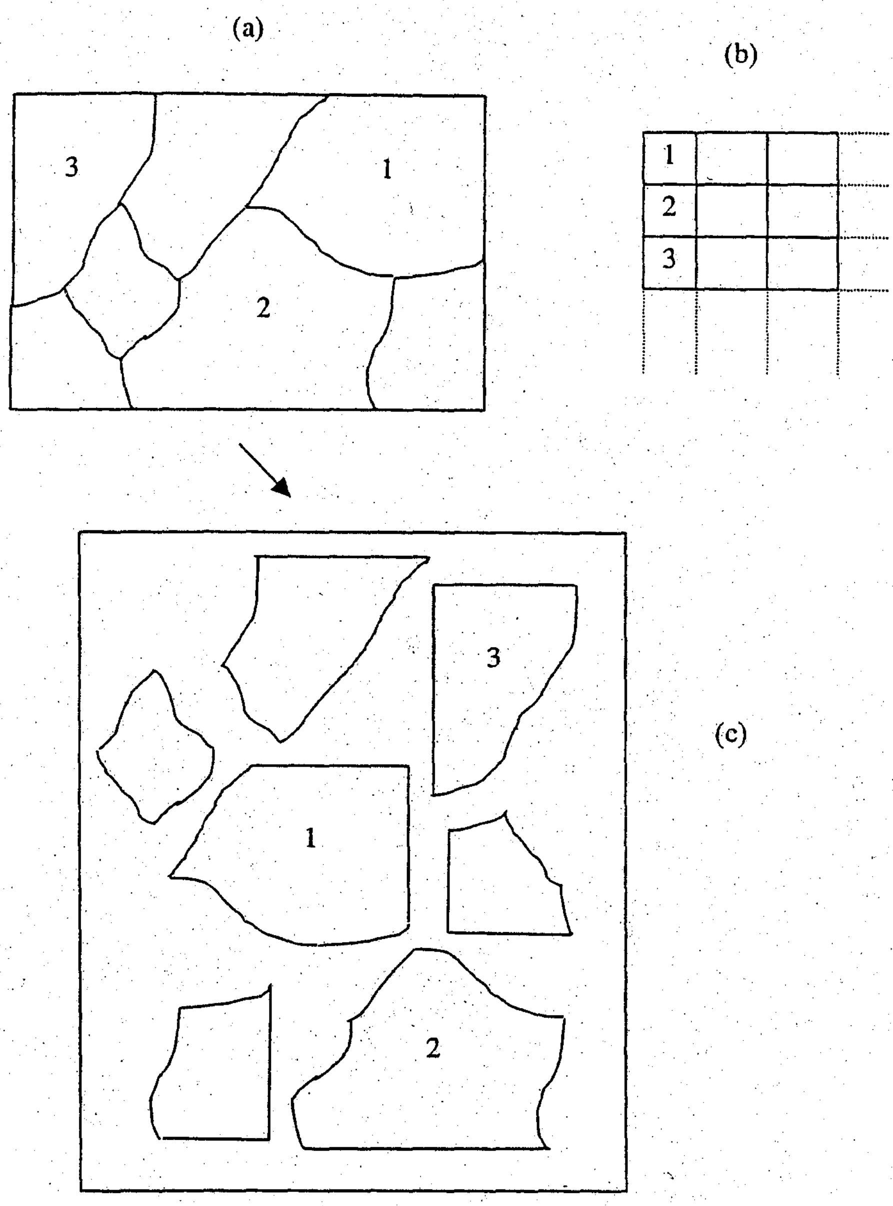

Figure 1. A map showing a number of areas (a), and a table of the attributes associated with each area (b). In (c) the areas have been relocated. Many methods of analysis will not be affected by the relocation, but methods of spatial analysis will give different results.

environment, if locations are not stored in the database, and accessed by the tools. It also follows from this and the prior discussion that the preparation and examination of a map is a form of spatial analysis.

Locations affect the results of spatial analysis in different ways. In some cases, space is reduced to the distances between pairs of objects, so the results change only if the distances change, and not if the entire set is rotated or inverted. For example, it might be that land use management practices would be unaffected if the study area was rotated on the Earth's surface so the bottom of the map became North, rather than the top. In other cases space is represented only by adjacencies between objects, so changes that affect locations but not adjacencies will not affect the results of analysis. For example, the management of a parcel may depend on what happens in neighboring parcels, but not on how far away they are.

GIS AND SPATIAL ANALYSIS

GIS and spatial analysis have a longstanding association, and spatial analysis has often been seen as the ultimate objective of representing space in a GIS. Much has been written about how implementation of methods of spatial analysis as GIS tools has made them more accessible and popular (Goodchild, 1987; Goodchild *et al.*, 1992), and more recently it has become clear that the availability of GIS has had significant impact on the development of spatial analysis (Goodchild and Longley, 1999). Today the two fields are inseparable.

One way to think about the relationship between GIS and spatial analysis is to compare it to the relationship between the statistical packages and statistics. Many statistical methods were developed long before the advent of digital computers, and the availability of the personal computer was an enormous boon, making it far easier to do the calculations and table look-ups necessary for statistical analysis. GIS has made it possible for far more people to make use of spatial analysis, and has greatly increased awareness of the power of spatial analytic techniques and the value of examining site and situation in applications like land use management.

In addition to implementing specific techniques, GIS provides many capabilities that support the user or researcher in doing spatial analysis. Tools are provided for digitizing and editing data, for obtaining data from remote sites, changing map projections to make data compatible, visual display through advanced cartographic techniques, export of data to be shared with others, and much more. Because a GIS is built around a common database, it enjoys substantial economies of scale to the software developer, by providing a comprehensive platform to which new methods of spatial analysis can be added cheaply and with little need for new expertise on the part of the user.

Nevertheless it is clear that GIS is much bigger than spatial analysis, and that many applications of GIS never make use of its spatial analytic capabilities. Some of the largest applications of GIS are in utility companies, and in other areas where large numbers of geographically dispersed facilities must be inventoried and managed. There may be no reason in such applications for analysis, or manipulation and transformation in order to reveal things that are not otherwise apparent; instead, the system may do no more than faithfully reproduce the facts that were put into it. Location in such systems is no more than another convenient organizing dimension, to add to names, times, social security numbers, or license numbers and to be used whenever location is important.

This wider applicability of GIS has an unfortunate consequence, however, since the commercial software industry is driven largely by the size of each application market, and designs its products accordingly. Much of the effort to enhance GIS software in recent years has gone into making the software easier to use, and into inventory applications, rather than into implementing additional analysis techniques. The best record in this regard is that of software products specifically targeted to niche markets, such as the academic and research sectors, where analysis is in high demand.

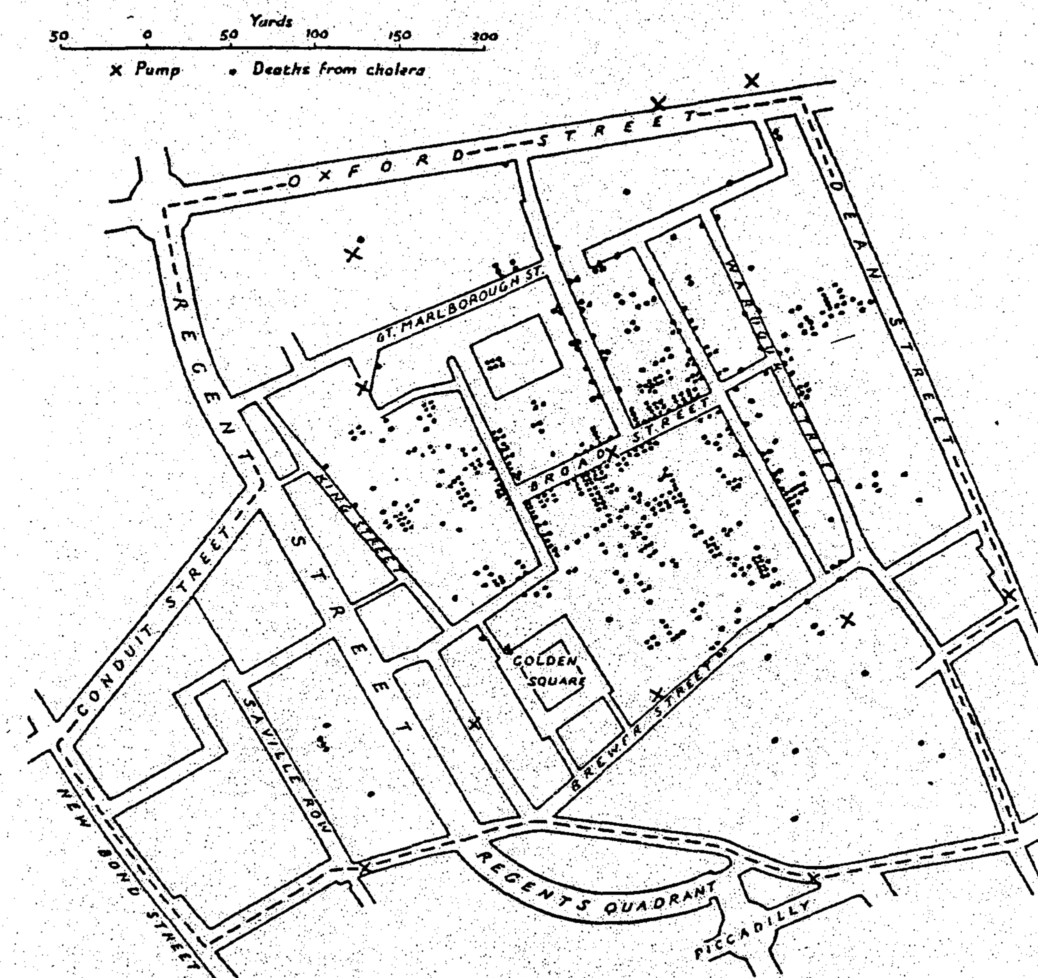

Figure 2. The Snow map of cholera incidence in the area of Broad Street, London, in 1854. The contaminated water pump is located at the center of the map, just to the right of the D in BROAD STREET. Source: Figure 1, page 174 in Gilbert (1958), reproduced by permission.

There is abundant recent literature on GIS and spatial analysis. Bailey and Gatrell (1995) and Haining (1990) have written comprehensive texts, and there are several recent collections of papers on the state of the art (Fischer *et al.*, 1996; Fotheringham and Rogerson, 1994; Longley and Batty, 1996). Two journals (the *Journal of Geographical Systems* and *Geographical Analysis*) focus on spatial analysis, and other GIS journals often carry articles (see particularly the *International Journal of Geographical Information Science* and *Transactions in GIS*). GIS trade magazines (*GeoWorld* and *GeoInfoSystems*) also frequently carry interesting articles on spatial analysis.

THE VALUE OF SPATIAL ANALYSIS

One of the classic demonstrations of the power of spatial analysis comes from epidemiology, and the work of Dr. John Snow on cholera in the mid 19th century. Snow suspected that cholera was caused by drinking polluted water, a theory that was sharply at odds with the prevailing wisdom, that the disease was passed between victims through the air. A simple map of the locations of deaths in the Soho district of London, and their situation with respect to a pump in Broad Street, proved to be the smoking gun, and Snow was able to confirm the hypothesis by removing the pump's handle (Figure 2; Gilbert, 1958).

Today such simple problems of infectious disease transmission have long since been worked out, but spatial analysis remains a powerful tool in health management (Gatrell and Senior, 1999). In a GIS, it allows us to address many different forms of hypothesis:

1) What patterns are evident in the distribution of a phenomenon? What patterns exist, for example, in the use of land in a study area? Do particular uses cluster in certain areas, and what pairs of uses tend to be neighbors?
2) How do different phenomena at a location influence each other? How is land use affected by or correlated with soil type, rainfall, or slope?
3) How is a phenomenon influenced by the same phenomenon at other locations? Do certain land uses tend to disperse themselves, or to cluster?
4) How is a phenomenon influenced by other phenomena in its neighborhood? Is land use influenced by distance to the nearest urban area, for example?
5) What changes are occurring on the landscape, and how are they driven by processes of different types?
6) How well do the phenomena in an area fit the predictions of a specific model? What patterns are predicted to emerge in the future? For example, what trends are apparent in the land uses in an area, and what future patterns will they lead to?

For example, Plate 4 shows a section of a soil map from an area of Ohio. Although complex, the patterns of soils shown on the map are far from random, and can reveal much useful information about soil and its uses. Techniques of spatial analysis can be used to examine spatial forms and relationships, using properties such as shape, adjacency, or proximity; or to compare soil types with other variables for the same area, represented on other maps, such as slope, aspect, vegetation cover, or land use.

FOUR PROBLEMS

Four general problems underlie spatial analysis, and pervade its applications. Each poses difficulties, and requires an approach that is distinctly adapted to the spatial context.

Spatial Dependence

In traditional statistical analysis it is assumed that the objects of interest are drawn randomly and independently from some parent distribution; hypotheses about that parent distribution can then be tested based on the sample, and conclusions can be generalized from the sample to the parent. Unfortunately this simple approach often breaks down in space.

Consider, for example, an area that is divided into 100 parcels of land, each with an associated land use class (Figure 3). We might be interested in asking whether the land

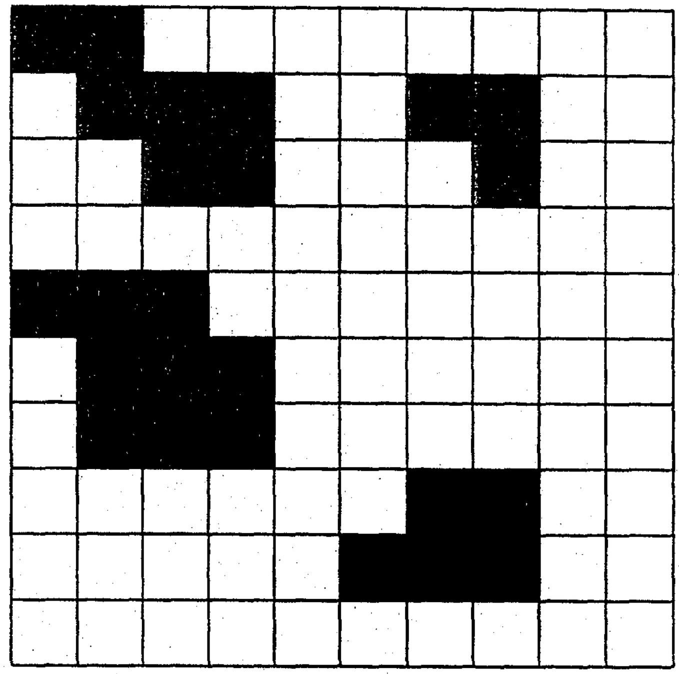

Figure 3. An illustration of spatial dependence. Three values are shown on the map, distributed over the 100 parcels, but their distributions show strong clustering. Any statistical test that assumed that parcel values had been drawn randomly and independently from a parent distribution would clearly be false.

use in each parcel is related to the underlying soil type. Conventional statistical thinking might lead us to devise an analysis of variance, comparing the variability of soils under one class of land use with the variability under other classes, and the variability as a whole. For example, we might observe that variability under the agricultural land use was less than the variability overall, and use an F test to confirm that the difference could not have occurred by chance in a sample of this size. But implicit in the test is the assumption that the parcels were drawn randomly and independently from some parent distribution, and this is patently false given the strong spatial patterns evident in Figure 3. Instead of a random sample we have drawn all of the cases in the area, and since adjacent parcels may well be underlain by the same soil type the independence assumption is clearly not tenable.

The apparently ubiquitous presence of spatial dependence in geographic phenomena has often been called the First Law of Geography, and attributed to Waldo Tobler: "all things are alike, but nearby things are more alike than distant things." Several courses of action are open to the analyst:

1) Avoid inferential tests completely, using only descriptive statistics of the pattern observed in a given area. No formal basis for generalization beyond the given area is available, and such descriptive case studies are generally regarded as having less value to science.
2) Discard a large proportion of the cases, retaining only those that are sufficiently far apart to make the independence assumption tenable. This is very wasteful of data, and reduces the analyst's ability to observe patterns and obtain results.

3) Model the spatial dependence explicitly, using advanced methods specially adapted to this case. A large number of such techniques have been developed, and reviews are available (Cliff and Ord, 1981; Anselin, 1988; Haining, 1990). Unfortunately these methods are inevitably more complex than those they replace, and less accessible to analysts with limited statistical understanding.

Spatial Heterogeneity

The second problem also arises because conventional statistical practice fails in the spatial context. If a study makes use of a sample, one normally assumes that if the study were conducted again using a different sample the results would be similar; and the precise distribution of results from repeated samples can be explored using standard statistical methods. It is tempting to equate the concept of a sample with the definition of a geographic study area, and to think that if the area were to change the effect on the results would be akin to that of taking a different sample. But spatial processes are affected by many factors, and if any of those vary on a scale comparable to the dimensions of the study area, then changing the boundaries has a direct and explicit effect on the results that is not analogous to sampling effects. For example, it is tempting to believe that if one takes a single 1:24,000 quadrangle as a study area, that one is taking approximately a 1/50,000 sample of the contiguous United States. But that is clearly absurd, because it ignores the regional variation of landscapes across the U.S. In a most extreme case, one would be suggesting that the quadrangle that lies entirely within the boundaries of the Great Salt Lake is a representative sample of the contiguous U.S.

An analysis of the upper 50 parcels of Figure 3 would give very different results from an analysis of the lower 50 parcels, or of all 100 parcels. But this serves to illustrate an important point, that spatial dependence and spatial heterogeneity are largely indistinguishable in practice, although very different in concept. For example, the area of gray parcels in the upper left may have been caused by favorable factors that vary slowly over the map, an instance of spatial heterogeneity. Alternatively, it may have been caused by a tendency for the use of one parcel to depend on the uses of neighboring parcels. Once the gray use was established, it spread to neighboring parcels to form a cluster, an instance of spatial dependence.

Spatial heterogeneity can be addressed by appropriate sampling designs that select samples from across the range of conditions represented in the population. It can be addressed by local area analysis, and this topic is discussed at greater length in the next section. It can also be addressed by interactive techniques that give the analyst direct control of the study area boundaries, by limiting analysis to a window of user-defined dimensions that is swept interactively over the space.

Effects of Scale

The term *scale* is used very loosely, to refer to two distinct properties of a data set: the *resolution* or level of detail, and the *extent* of the study area. Both can be expressed in terms of length, and the ratio is a dimensionless number. Unfortunately both are often not explicitly stated, although resolution is often embedded in the definitions of geographic phenomena. For example, the land use *urban* may make sense at a resolution of 10 km, where all of the Los Angeles basin appears as a uniformly developed blur; but at resolutions of 10 m or less the concept of a uniform urban class breaks down into surfaces of concrete, asphalt, grass, *etc.*

The Earth's surface is almost infinitely complex, so any representation in the form of a map, image, or GIS database must be an approximation or generalization, and often this is achieved by removing detail. A raster data set, such as a remotely sensed image, does this explicitly by using a constant cell size, ensuring that objects much smaller than a single cell will be invisible. However vector data sets do not have explicitly defined resolution, making it very difficult to take this factor into account in analysis. Actual resolution may be expressed indirectly through the existence of a *minimum mapping unit*, or minimum threshold for the area of a represented object, but the rules for aggregation of smaller areas are not often made explicit.

Selective Testing

Finally, conventional statistical practice assumes that the analyst plays no explicit part in the selection of data or the specification of tests; normally, all of the available data are analyzed. Spatial data sets are often enormous, however, and it is tempting to apply tests selectively, perhaps in areas where effects are anticipated. Testing for cancer clusters is a perfect example, because the initiative for testing comes from the prior observation that rates in an area look suspicious. But in principle, such prior knowledge should not exist, or should not play any part in the design of the experiment; if it does, the assumptions of statistical analysis will be invalid. Cancer cluster testing is an enormously difficult problem, despite its apparent conceptual simplicity, because it is virtually impossible to conduct such tests without implicating the data in the selection of study sites. The power of maps to communicate spatial data works against the need for objectivity in this instance. An analysis of the shaded area in Figure 4 would show that the density of points within it is anomalously high. But it is virtually impossible not to implicate the data in the selection of the test area, since the eye conducts a very large number of such tests informally as soon as the data are viewed in map form.

TRENDS IN SPATIAL ANALYSIS

This section focuses on three areas where there has been rapid development of new techniques in recent years, driven in part by the power of GIS and the ready availability of computing. All three address the previous four concerns to some degree, although they fall far short of solving them.

Local Area Analysis

Ideally, one would like to be believe that the purpose of spatial analysis is to discover laws and principles that apply uniformly to all of the Earth's surface. The laws of motion and thermodynamics are such laws. But in practice the laws of interest to land use managers, or to other disciplines that deal with the Earth's surface, are likely to be partial at best, meaning that there will always be other factors influencing outcomes. If so, then

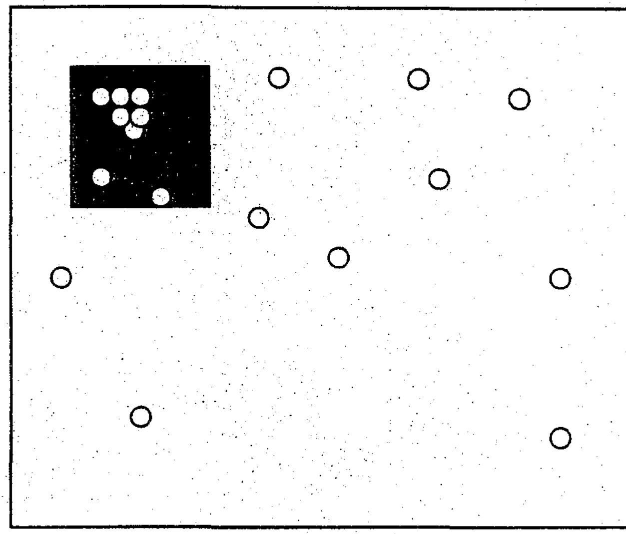

Figure 4. The problem of selective testing. An analysis of the shaded area would show that the density of points within it is anomalously high. But it is virtually impossible not to implicate the data in the selection of the test area, since the eye conducts a very large number of such tests informally as soon as the data are viewed in map form.

a good strategy is to try to control those other factors by studying only small areas, on the grounds that many of the factors will vary slowly. For example, there is little value in including Antarctica in a model of agricultural land use, in the interests of finding a model that fits the entire Earth; and models that ignore the Desert Southwest may be more useful in the Farm Belt than models that include it.

Local area analysis takes this argument to justify a somewhat heretical approach to spatial analysis that focuses directly on the local, exploring variation in conditions from one area to another through explicit analysis. Its position is thus sharply different from the traditional one, since it in effect abandons search for general laws, and deals explicitly with spatial heterogeneity. Instead of a single value of a model parameter, this approach will fit values everywhere based on local data, and produce a map of the spatial variation of the parameter. For example, instead of a single measure of the relationship between soil type and land use, one would obtain a map of the degree to which soil type predicted land use, and associated maps of exactly how.

Fotheringham and his colleagues (Fotheringham *et al.*, 1998) have been among the strongest proponents of this approach (and see also Anselin, 1995; Getis and Ord, 1996). Geographically weighted regression creates maps of regression parameters by fitting models locally at every point, weighting each item of data by its distance from the point. Similar approaches might be used to modify each of the standard methods of analysis to provide local focus.

Visual Methods

I argued at the outset that the act of examining a map with the eye and brain should be regarded as a form of spatial analysis. Visual display of data has always been an important part of scientific research, and the visual nature of GIS makes the technology particularly appealing. Maps are fascinating to many people, and the combination of maps and computers explains at least part of the success of GIS.

Exploratory data analysis (EDA) emerged in the 1970s (Tukey, 1977) as a set of methods for examining statistical data, with the objective of revealing what might otherwise go unnoticed. New methods of visual display were developed that proved much more effective than simple tables at extracting meaning from data. In the late 1980s a similar trend took hold in GIS, with the development of methods of exploratory spatial data analysis (ESDA). Anselin (1995) traces the origins of both EDA and ESDA to new computational capabilities for interactive manipulation of data; and the desire to involve the eye and brain more directly in the analysis task. He distinguishes between *spatialized* EDA, in which methods of EDA are given a spatial dimension, and true ESDA.

Multiscale Methods

Concern over the effects of scale, and the failure of many analysts and techniques to make them explicit, led to the formulation of two related problems: the ecological fallacy (Robinson, 1950), and the modifiable areal unit problem (MAUP; Openshaw, 1983). The ecological fallacy refers to the temptation to infer conclusions at one level of resolution from conclusions at a coarser level, and originated in the question of whether individual behavior could be deduced from the behavior of aggregates. King (1997) has shown that the problem is soluble under certain conditions, but these are not often found in reality. The MAUP refers to the analysis of aggregated data, and the dependence of results on the specific configuration of zones used as the basis of aggregation. If the zones vary in size then there is clearly overlap between the two problems.

Recent research has focused on methods that formalize scale explicitly, and examine its effects. Fractals are one formal framework for examining scale effects, and it and others are reviewed by Goodchild and Quattrochi (1997). Multiscale representations have been researched in GIS, and several software packages now offer support for multiscale databases and manipulations of scale. It is clear that in future scale will be much more likely to be treated and controlled explicitly in spatial analysis, and that tools to support this will continue to improve.

THE INTERNET

The advent of the Internet and the World Wide Web have had profound effects on the development of GIS and spatial analysis. It is now possible to assemble data from widely scattered sources using search engines, file transfer protocols, clearinghouses, and a host of other mechanisms, to perform analyses using integrated software, and to share results with colleagues and collaborators around the world. The Internet has driven the costs of dissemination almost to zero, provoking major restructuring of traditional data production organizations. The U.S. Geological Survey, the nation's primary source of spatial data, now sees its role increasingly in terms of coordination, partnership, and the setting of standards, through the mechanism of the National Spatial Data Infrastructure (NRC, 1993). The WWW is chaotic and unmanaged, creating the impression that society is drowning in a sea of unstructured and possibly unreliable information.

Four issues seem to be of major concern with respect to spatial analysis and GIS (Goodchild, 1998). First, successful use of the Internet and WWW for data acquisition requires mechanisms for searching over largely uncatalogued information spaces. Major catalogues exist, including the National Geospatial Data Clearinghouse (http://www.fgdc.gov) and the Alexandria Digital Library (http://www.alexandria.ucsb.edu), but the vast majority of data is scattered in hard-to-find pages. Plate 5 illustrates the Alexandria Digital Library's user interface, showing the ability to display the contents and details of digital entries. A recent report of the National Research Council (NRC, 1999) defined the concept of a *distributed geolibrary* as a vision for future dissemination of geospatial data, and identified many of its components and services.

Second, the potential user of data must have methods for assessing fitness for use. Geospatial data sets tend to be large, so there can be substantial cost in time and effort in retrieving a data set from a remote site, only to find that it is unsuitable because of scale, quality, or other concerns. The widely used Content Standards for Digital Geospatial Metadata (http://www.fdgc.gov) attempt to resolve this by providing digital documentation, and the research community is actively extending its capabilities.

Third, within the client–server architecture of the WWW there exists the possibility of processing at either end of the communication chain, in other words, in the user's system with the user's software, or at the source of the data. A distinction can be drawn between geographic information *systems*, integrated software existing in the user's system and processing data that is locally resident but may have been obtained from a remote source; and geographic information *services*, provided by software resident on a remote server in response to requests received from the user. In the GIServices model no large data sets need to be transferred; and the monolithic approach to integrated GIS software can be replaced by one of distributed, modular services. For example, several WWW sites offer geocoding services, or the ability to convert street addresses into coordinates, a function that used to be exclusively associated with GISystems. Exactly what the future will hold in redefining the balance between GISystems and GIServices remains to be seen.

Finally, the Internet and WWW have drawn attention to the increasing costs associated with lack of interoperability between software packages, data, and languages. Experience with spatial analysis using one GIS product may be of little help in using another, and it can be very difficult to transfer data from one product's format to another's. Much effort has gone into addressing the lack of interoperability in recent years, led by the Open GIS Consortium (http://www.opengis.org), and interoperability is a major focus of research (Goodchild *et al.*, 1999).

CONCLUSION

Geographic information systems have made spatial analysis readily accessible to researchers and decision-makers, and have led to a very rapid growth in the use of a spatial perspective. Fields such as land use management are innately spatial, and have benefited enormously from the widespread adoption of GIS.

At the same time, these changes have had interesting effects on the previously obscure field of spatial analysis. They have led to a shift away from the notion of the scientist as neutral exploiter of objective tools, to an increased interest in the human factors of analysis: visualization, intuition, the power of the eye and brain, and user interface design.

Exploration is regarded as a legitimate pursuit, as a precursor or even an alternative to model-based confirmation.

More recently, the WWW and Internet have begun to engender their own changes, with an increased emphasis on collaboration and sharing, and spatial analysis as a method of communication of knowledge from computer to individual, and from one individual to another. It has become essential to be able to describe the contents of a data set, and other properties such as its quality, in formal structures known as metadata to allow potential users to assess the value of the data to their applications. Today, spatial analysis with GIS is much more than the ability to perform specific functions on data resident in a computer. Effective analysis requires the ability to find and obtain data in usable form, and to communicate effectively the knowledge gained from analysis to others. In many ways the old model of scientific research — problem formulation, experimental design, data collection, analysis, and publication — is being replaced by a much more fluid model that is more attuned to the technology of the times, and more powerful as a result.

REFERENCES

1. L. Anselin (1988) *Spatial Econometrics: Methods and Models*. (Dordrecht: Kluwer).
2. L. Anselin (1995) *Geog. Anal.*, **27**, 93–115.
3. T.C. Bailey and A.C. Gatrell (1995) *Interactive Spatial Data Analysis*. (Harlow, UK: Longman Scientific and Technical).
4. Cliff and J.K. Ord (1981) *Spatial Processes, Models and Applications*. (Pioneer, London).
5. M. Fischer, H.J. Scholten and D. Unwin (editors) (1996) *Spatial Analytical Perspectives on GIS*. (London: Taylor and Francis).
6. A.S. Fotheringham, M.E. Charlton and C. Brunsdon (1998) *Environ. Planning A*, **30**, 1905–1927.
7. A.S. Fotheringham and P.A. Rogerson (editors) (1994) *Spatial Analysis and GIS*. (London: Taylor and Francis).
8. C. Gatrell and M. Senior (1999) Health and health care applications, in *Geographical Information Systems: Principles, Techniques, Management and Applications*, edited by P.A. Longley, M.F. Goodchild, D.J. Maguire and D.W. Rhind, pp. 925–938. (New York: Wiley).
9. Getis and J.K. Ord (1996) Local spatial statistics: an overview, in *Spatial Analysis: Modelling in a GIS Environment*, edited by P.A. Longley and M. Batty. (Cambridge, UK: GeoInformation International).
10. E.W. Gilbert (1958) *Geog. J.*, **124**, 174.
11. M.F. Goodchild (1987) *Int. J. Geog. Inf. Syst.*, **1**, 327–334.
12. M.F. Goodchild (1998) Different data sources and diverse data structures: metadata and other solutions, in *Geocomputation: A Primer*, edited by P.A. Longley, S.M. Brooks, R. McDonnell and W. Macmillan, pp. 61–74. (London: Wiley).
13. M.F. Goodchild, M.J. Egenhofer, R. Fegeas and C.A. Kottman (editors) (1999) *Interoperating Geographic Information Systems*. (Norwell, MA: Kluwer Academic Publishers).
14. M.F. Goodchild, R.P. Haining, S. Wise and 12 others (1992) *Int. J. Geog. Inf. Syst.*, **6**, 407–424.
15. M.F. Goodchild and P.A. Longley (1999) The future of GIS and spatial analysis, in *Geographical Information Systems: Principles, Techniques, Management and Applications*, edited by P.A. Longley, M.F. Goodchild, D.J. Maguire and D.W. Rhind, pp. 567–580. (New York: Wiley).
16. M.F. Goodchild and D.A. Quattrochi (editors) (1997) *Scale in Remote Sensing and GIS*. (Boca Raton: CRC Press).

17. R.P. Haining (1990) *Spatial Data Analysis in the Social and Environmental Sciences.* (New York: Cambridge University Press).
18. G. King (1997) *A Solution to the Ecological Inference Problem: Reconstructing Individual Behavior from Aggregate Data.* (Princeton: Princeton University Press).
19. P.A. Longley and M. Batty (editors) (1996) *Spatial Analysis: Modelling in a GIS Environment.* (Cambridge, UK: GeoInformation International).
20. National Research Council (1993) *Toward a Coordinated Spatial Data Infrastructure for the Nation.* (Washington, DC: National Academy Press).
21. National Research Council (1999) *Distributed Geolibraries: Spatial Data Resources.* (Washington, DC: National Academy Press).
22. S. Openshaw (1983) *The Modifiable Areal Unit Problem.* (Norwich: GeoBooks).
23. W.S. Robinson (1950) *Am. Sociol. Rev.*, **15**, 351–357 (1950).
24. J.W. Tukey (1977) *Exploratory Data Analysis.* (Reading, MA: Addison-Wesley).