40

The future of GIS and spatial analysis

M F GOODCHILD AND P A LONGLEY

The chapter explores factors affecting spatial analysis, in theory and practice, and their likely impacts. A model is presented of the traditional role of spatial analysis, and is examined from the perspectives of increasing costs of data, the increased sharing of data between investigators across a wide range of disciplines, the emergence of new techniques for analysis, and new computer architectures. Practical problems are identified that continue to face investigators using GIS to support spatial analysis, including accuracy, the technical problems of integration, and the averaging of different feature geometries. The chapter ends with a speculation on spatial analysis of the future.

1 INTRODUCTION

GIS and spatial analysis have enjoyed a long and productive relationship over the past decades (for reviews see Fotheringham and Rogerson 1994; Goodchild 1988; Goodchild et al 1992). GIS has been seen as the key to implementing methods of spatial analysis, making them more accessible to a broader range of users, and hopefully more widely used in making effective decisions and in supporting scientific research. It has been argued (e.g. Goodchild 1988) that in this sense the relationship between spatial analysis and GIS is analogous to that between statistics and the statistical packages. Much has been written about the need to extend the range of spatial analytic functions available in GIS, and about the competition for the attention of GIS developers between spatial analysis and other GIS uses, many of which are more powerful and better able to command funding. Specialised GIS packages directed specifically at spatial analysis have emerged (e.g. Idrisi; see also Bailey and Gatrell 1995). Openshaw and Alvanides (Chapter 18) have set out some of the ways in which developments in computation may feed through to enhanced GIS-based spatial analysis. Finally, Anselin (Chapter 17), Getis (Chapter 16) and others have discussed the ways in which implementation of spatial analysis methods in GIS is leading to a new, exploratory emphasis.

The purpose of this chapter is to explore new directions that have emerged recently, or are currently emerging, in the general area of GIS and spatial analysis, and to take a broad perspective on their practical implications for GIS-based spatial analysis. In the next section, we argue that in the past the interaction between GIS and spatial analysis has followed a very clearly and narrowly defined path, one that has more to do with the world of spatial analysis prior to the advent of GIS than with making the most of both fields – the path is, in other words, a legacy of prior conditions and an earlier era (see also Openshaw and Alvanides, Chapter 18). The following section expands on some of the themes of the introduction to this volume by identifying a number of trends, some related to GIS but some more broadly based, that have changed the context of GIS and spatial analysis over the past few years, and continue to do so at an increasing rate. The third section identifies some of the consequences of these trends, and the problems that are arising in the development of a new approach to spatial analysis. The chapter concludes with some comments about the complexity of the interactions between analysis, data and tools, and speculation on what the future may hold, and what forms of spatial analysis it is likely to favour.

2 TRADITIONS IN SPATIAL ANALYSIS

2.1 The linear project design

In the best of all possible worlds, a scientific research project (the term 'research' will be interpreted broadly to include both scientific and decision-making activities) begins with clearly stated objectives. Some decision must be made, some question of scientific or social concern must be resolved by resorting to experiment or real-world evidence. A research design is developed to resolve the problem, data are collected, analyses are performed, and the results are interpreted and reported. Although this implies a strictly linear sequence of events, the most robust research designs also include feedbacks and checks in order to ensure that the principles of good scientific research are not overly compromised in practical implementation. This simple, essentially linear, structure with recursive feedbacks underlies generations of student dissertations, government reports, and research papers. It is exemplified by the classic social survey research design illustrated in Figure 1. The sequential events in this design together constitute a holistic research project, and the feedbacks are all internal to the research design. Thus once the project has been initiated, the availability of existing data has no further influence upon problem definition; methods of analysis that are consistent with the type, quality, and amount of data to be collected are identified at the design stage (and the data collection method changed if no suitable analytical method exists); the sample design is not guided by considerations and priorities that lie outside the remit of the research; and so on.

In this simple, sequential world, the selection of methods of analysis can be reduced to a few simple rules (in the context of statistical analysis, see for example Levine 1981: chapter 17; Marascuilo and Levin 1983: inside cover; Siegel 1956: inside cover). Choice of analytic method depends on the type of inference to be drawn (e.g. whether two samples are drawn from the same, unknown population, or whether two variables are correlated), and on the characteristics of the available data (e.g. scale of measurement – nominal, ordinal, interval, or ratio: Wrigley 1985). Inference about, and exploration of, the research problem will take place in what is loosely described as the confirmatory (hypothesis testing and inference seeking) and exploratory (pattern or anomaly seeking) stages of the research.

In contrast to this coherent research design, the terms 'data-driven' and 'technique-driven' are highly pejorative in research generally, as are such phrases as 'a technique in search of a problem' – in this ideal world, the statement of the problem strictly precedes the collection of data and the performance of analysis.

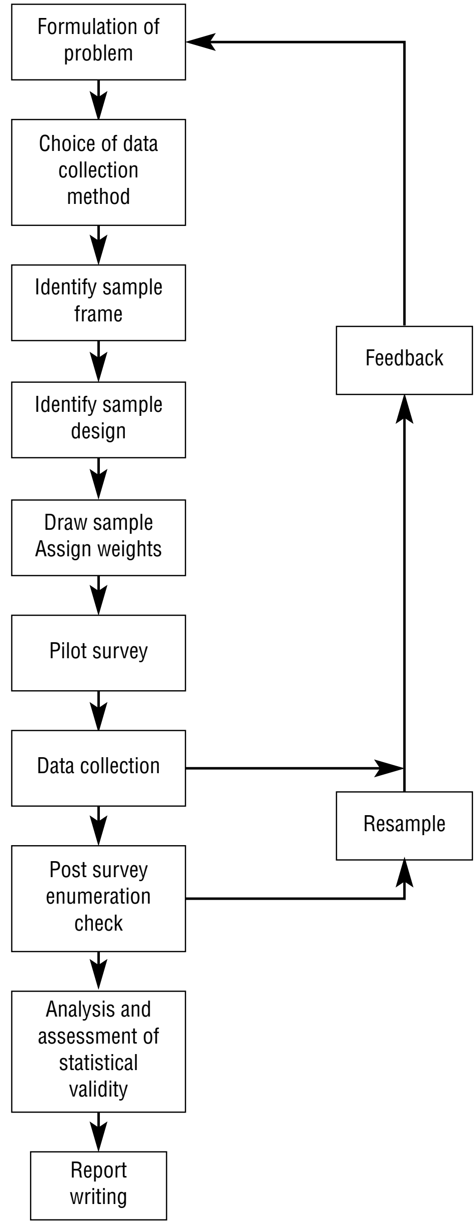

Fig 1. Sequential stages of a typical social survey research design.

2.2 Spatial analysis

Spatial analysis, or spatial data analysis, is a well-defined subset of the methods of analysis available to a project. One might define spatial analysis as a set of methods useful when the data are spatial, in other words when the data are referenced to a 2-dimensional frame. More narrowly, the Earth's surface provides a particular instance of such a frame, the geographical frame, with its peculiar properties of curvature. This definition of spatial analysis is arguably too broad, because in basing the definition on the properties of data it does not address the question of whether the 2-dimensional frame actually matters – could the same results have been obtained if the frame were distorted in some way, or if objects were repositioned in the frame? More precisely, then, spatial analysis can be defined as that subset of analytic techniques whose results depend on the frame, or will change if the frame changes, or if objects are repositioned within it. To distinguish analytic methods from more mundane operations, they might be defined as methods for processing data with the objective of solving some scientific or decision-making problem.

Methods of spatial analysis have accumulated in a literature that spans many decades, indeed centuries (see Getis, Chapter 16). They have been invented in many disciplines, including mathematics, particularly geometry; statistics, particularly spatial statistics and statistical geometry; and in geography and other Earth sciences. Compendia have been published (among others, see Bailey and Gatrell 1995; Berry and Marble 1968; Haining 1990; Taylor 1977; Unwin 1981), and various approaches proposed for structuring this body of technique. Spatial analytic techniques may also be classified into those which are confirmatory and those which are exploratory. Choice of analytical method also relates to data characteristics – documented since Chorley and Haggett's (1965) analogies between (respectively) nominal, ordinal, interval and ratio data and point, line, area, and surface objects (see also Chrisman 1997; Martin 1996).

2.3 The well-informed analyst

Traditionally, the responsibilities of the inventor of a technique ended when the technique had been tested and described. Even the testing of a technique can be suspect in an academic world that often values theory over empiricism, and is suspicious of empirical results that cannot be demonstrated to be generally true. The advent of the digital computer changed this world fundamentally because it became possible for a scientist to perform a method of analysis automatically, without taking personal responsibility for every aspect of the performance. It was now possible using the 'black box' of the computer to perform an analysis that one did not know everything about – that one could not perform by hand. Methods emerged, beginning in the 1970s and particularly in the area of multivariate statistics, that would be impossibly impractical to perform by hand. Pedagogically, a fundamental shift became possible in how analysis was taught – that one might learn about a technique by studying the nature of its response to particular inputs, rather than by studying the procedure which generated the response. But there is a fundamental difference between these two positions: between whether one understands the results of a principal components analysis, for example, as the extraction of eigenvalues from a specific matrix, or the generation of statistics that broadly indicate some concept of 'relative importance' without presuming any understanding of what eigenvalues are and *how* they formalise the structure in data.

Exactly where this change occurred is open to debate, of course. It may have occurred when students were no longer required to perform statistical analyses by hand before being let loose on computer packages; or when FORTRAN appeared, making it necessary to understand less about how instructions were actually carried out; or when the growth of the scientific enterprise had reached such a level that potential replication of every result was a practical impossibility.

Of course the digital computers that were introduced to the scientific community beginning in the late 1950s produced rapid change in the labour demands of many statistical methods. The intricate calculations of factor analysis (Harman 1976) could be performed by a fully automatic machine, provided the researcher could command sufficient computer time, and provided labour was available to punch the necessary cards. Computers and the brains of young humans are in some ways similar: both begin 'hard wired' with the primitive elements of reasoning (e.g. binary processing in computers, linguistic abilities in infants) and both can build enormously complex structures out of simpler ones, apparently ad infinitum (see Fischer, Chapter 19, for a broader discussion of computer 'reasoning'). What began in the 1960s as a set of uncoordinated efforts by individual scientists writing their own programs had developed by the 1990s into a complex of enormously sophisticated tools, each integrating a large number of methods into an easy-to-use whole.

If software packages and user-friendly computer environments have made aspiring spatial analysts less aware of the computational and statistical context to inference, then the opposite is true to some extent of exploratory analysis, where the innovation of computer graphics and windows, icons, mice, and pointers (WIMPs) has created a more intuitive context to the interrogation of spatial data (see Anselin, Chapter 17; Kraak, Chapter 11). Indeed one of the criticisms of GIS-based graphics developments from the spatial analysis community has been that the computer graphics medium has been allowed to dominate the spatial analysis message, by analogy to 'data-led' thinking as described in section 2.1.

2.4 Extending the functions of analytic software

Although they show clear evidence of their roots, the packages used by the scientists of the 1990s are different in fundamental respects from the programs of the 1960s. Besides implementing large numbers of statistical methods, today's packages also provide support for the maintenance of data and the creation of information. There will be tools for documenting datasets, and describing their properties, such as accuracy and history. Other tools will support the sharing of data, in the form of format converters or interfaces to the Internet. In short, the functions of today's digital computers in supporting research go far beyond those of a simple calculating machine, carrying out well-defined methods of analysis. The same digital computer may now be involved in:

- the selection and formulation of a problem, by providing access to automated library catalogues and on-line literature;
- the collection of data through support for real-time data acquisition;
- management of data, performance of analysis, visualisation of results, writing of conclusions;
- even publication through access to the Internet and the World Wide Web.

The computer is no longer part of the research environment – we are rapidly approaching a world in which the computer *is* the research environment.

These trends are all echoed strongly in GIS. Although a particular scientist might use a GIS in ways that are more analogous to the early days of statistical computing, by performing a single buffering operation, for example, scientific applications are much more likely to include integration of many GIS functions. Today's scientist or decision-maker is likely to see a GIS as an environment for research, rather than as a means of automating analysis. The GIS is likely to be involved in the project from beginning to end, and to be integrated with other tools and environments when these are needed. GIS will be used for collecting, assembling, verifying, and editing the data; performing some of the analyses required by the project; and presenting and interpreting the results. Moreover, much GIS use may not be tied to a specific project – GIS finds extensive use in the collection of data for purposes that may be generic, or not well-defined, or may be justified in anticipation of future demand. Even though these may not be 'spatial analysis' in the sense of the earlier discussion, analysis may still be necessary as part of the data production process – for example, when a soil scientist must analyse data to produce a soil map.

2.5 When to choose GIS

A related issue is the extent to which GIS remains a separately identifiable technology, and in what senses the 'GIS environment' is distinctive. The general drift of many of the chapters in this 'Technical issues' Part of this book is that GIS is increasingly becoming both a background technology (more akin to wordprocessing than, say, spatial interaction modelling), and a technology that can be broken up and packaged as niche products (Elshaw Thrall and Thrall, Chapter 23). And yet the various discussions in the 'Principles' Part of this book document the important agenda for spatial analysis set in the environment of GIS, and why GIS-based spatial analysis is likely to remain a distinctive area of activity for the foreseeable future.

If GIS has multiple roles in support of science and problem-solving, then one might not be surprised to find that the choice between GIS alternatives is complex and often daunting. The many GIS packages offer a wide range of combinations of analysis functions, housekeeping support, different ways of representing the same phenomena, variable levels of sophistication in visual display, and performance. In addition, choice is often driven by: the available hardware, since not all GIS run on all platforms; the format in which the necessary data have been supplied; the personal preferences and background of the user; and so forth. Even the extensive and frequently updated comparative surveys published by groups such as GIS World Inc. can be of little help to the uninitiated user.

The existence of other classes of analytic software complicates the scene still further. Under what circumstances is a problem better solved using a package that identifies itself as a GIS, or using a statistical package, or a mathematical package, or a scientific visualisation package? Under what circumstances is it better to fit the square peg of a real problem into the round GIS hole? GIS are distinguished by their ability to handle data referenced to a 2-dimensional frame, but such capabilities also exist to a more limited extent in many other types of software environment. For example, it is possible to store a map in a spreadsheet array, and with a little ingenuity to produce a passable 'map' output; and many statistical packages support data in the form of images.

Under what circumstances, then, is an analyst likely to choose a GIS? The following conditions are suggested, although the list is certainly not complete, and the items are not intended to be mutually exclusive:

- when the data are geographically referenced;
- when geographical references are essential to the analysis;
- when the data include a range of vector data types (support for vector analysis among non-GIS packages appears to be much less common than support for raster analysis);
- when topology – representation of the connections between objects – is important to the analysis;
- when the curvature of the Earth's surface is important to the analysis, requiring support for projections and for methods of spatial analysis on curved surfaces;
- when the volume of data is large, since alternatives like spreadsheets tend to work only for small datasets;
- when data must be integrated from a variety of sources, requiring extensive support for reformatting, resampling, and other forms of format change;
- when geographical objects under analysis have large numbers of attributes, requiring support from integrated database management systems, since many alternatives lack such integration;
- when the background of the investigator is in geography, or a discipline with strong interest in geographical data;
- when the project involves several disciplines, and must therefore transcend the software traditions and preferences of each;
- when visual display is important, and when the results must be presented to varied audiences;
- when the results of the analysis are likely to be used as input by other projects, or when the data are being extensively shared.

3 ELEMENTS OF A NEW PERSPECTIVE

This section reviews some of the changes that are altering the context and face of spatial analysis using GIS. Some are driven by technological change, and others by larger trends affecting society at the turn of the millennium.

3.1 The costs of data creation

The collection of geographical data can be extremely labour-intensive. Early topographic mapping required the map-maker to walk large parts of the ground being mapped; soil mapping requires the exhausting work of digging soil pits, followed often by laborious chemical analysis; census data collection requires personal visits to a substantial proportion of (sometimes all) household respondents; and forest mapping requires 'operational cruise', the intensive observation of conditions along transects. Although many new methods of geographical data creation have replaced the human observer on the ground with various forms of automated sensing, there is no alternative in those areas that require the presence of expert interpreters in the field.

Many of the remaining stages of geographical data creation are also highly labour-intensive. There is still no alternative to manual digitising in cases where the source document is complex, compromised, or difficult to interpret. The processes of error detection and correction are difficult if not impossible to automate, and the methods of cartographic generalisation used by expert cartographers have proven very difficult to formalise and replace. In short, despite much technical progress over the past few decades, geographical data creation remains an expensive process that is far from fully automated.

Labour costs continue to rise at a time when the resources available to government, the traditional source of geographical data, continue to shrink (see Elshaw Thrall and Thrall, Chapter 23). Many geographical datasets are collected for purposes which may be far from immediate, and it is difficult therefore to convince taxpayers that they represent an essential investment of public funds, especially in peacetime. Governments in financial straits call for evidence of need: for example, census organisations are under continual pressure to demonstrate that their costly operations do not replicate information that is available elsewhere; and many governments have moved their mapping operations onto a semi-commercial basis in order to allow demand to be expressed through willingness to pay (see Rhind, Chapter 56). To date, the US Federal mapping agencies have resisted the trend, but internationally there is more and more evidence of the emergence of a market in geographical information.

Within the domain of geographical data the pressures of increased labour costs favour data that can be collected and processed automatically. Given a choice between the labour-intensive production of vector topographic data, and the semi-automated generation of such raster products as digital elevation models and digital orthophotos, economic pressures can lead only in one direction. It is easy to imagine a user trading off the ability to identify features by name against the order of magnitude lower cost, and thus greater potential update frequency, of raster data.

The broader context to these changes is that we now live in a digital world, in which far more data are collected about us, in computer readable form, than ever before. This is what has been termed the 'information economy', in which government no longer has a monopoly in the supply of geographical information, and in which information has become both a tradeable commodity and a strategic organisational resource. Global trends such as deregulation and privatisation, allied to the increasing competitive edge of consumer-led markets, are multiplying the potential number of sources of information, yet at the expense of system-wide standardisation (Rhind, Chapter 56). Such data are not ideally suited to the linear research design set out in section 2.1, yet (in socioeconomic research at least) they frequently are far richer in detail than anything that has been collected hitherto. A clear challenge to spatial analysis is therefore to reconcile diverse datasets with different data structures or spatial referencing systems, and to gauge how representative they are with reference to existing (more limited or less frequently updated) public sources. A good example of this concerns the development of geodemographic indicators which have traditionally been derived from census data. These are typically updated only every ten years and are frequently reliant upon very indirect indicators of likely consumer behaviours (Longley and Clarke 1995). 'Lifestyles' approaches based upon questionnaire returns from a range of self-selecting respondents (Birkin 1995) offer the prospect of 'freshening up' and in time replacing conventional census-based geodemographics, although thorny issues of representativeness and bias must be grappled with before credible 'data fusion' can be deemed to have taken place.

Of course, the principle of information commerce is alien to the scientific community, which is likely to resist strongly any attempt to charge for data that is of interest to science, even peripherally. But here too there are pressures to make better use of the resources invested in scientific data collection. Research funding agencies increasingly require evidence that data collected for a project have been disseminated, or made accessible to others, while recognising the need to protect the interests of the collector.

Trends such as these, while they may be eminently rational to dispensers of public funds, nevertheless fly directly in the face of the traditional model of science presented earlier. For example, the best-known definition of the discipline of geography is that it is 'concerned to provide accurate, orderly, and rational description and interpretation of the variable character of the Earth surface' (Hartshorne 1959). As a general rule, commercial datasets are not accurate (they provide little indication of the sources of unknown errors in data collection or the ways in which they are likely to operate in analysis); they are orderly only in a minimalist sense (for example, satellites provide frequently-updated *coverage* information yet cannot comprehensively measure land *use*; 'lifestyles' data do not provide information about all groups in society); and they are not rational in that they separate still further the analyst from the context to the research problem and lead to data- or machine-led thinking. How can projects fail to be driven by data, if data are forced to obey the economic laws of supply and demand? Where in traditional science are the rules and standards that allow scientists to trade off economic cost against scientific truth? It seems that economic necessity has forced the practice of science to move well beyond the traditions that are reflected in accepted scientific methodologies and philosophies of science.

3.2 The life of a dataset

In the traditional model presented earlier data were collected or created to solve a particular problem, and had no use afterwards except perhaps to historians of science. But many types of geographical data are collected and maintained for generic purposes, and may be used many times by completely unrelated projects. For other types, the creation of data is itself a form of science, involving the field skills of a soil scientist, for example, or a biologist. Thus a dataset can be simultaneously the output of one person's science, and the input to another's. This is to conceive of spatial analysis within GIS as the process of building 'models of models' – whereby the outcome of a 'higher level' spatial analysis is dependent upon data inputs which are themselves a previously modelled version of reality. These relationships have become further complicated by the rise of multidisciplinary science, which combines the strengths and expertise of many different sciences, and partitions the work among them. Once again, the linear model of science is found wanting, unable to reflect the complex relationships between projects, datasets, and analytic techniques that exist in modern science. The notion that data are somehow subsidiary to problems, methods and results is challenged, and traditional dicta about not including technical detail in scientific reports may be counterproductive.

In truth, of course, this is nothing new in the sense that most spatial analysis in the socioeconomic realm has been based upon crude surrogate data, obtained for inappropriate areal units in obsolete time periods. Thirty-five years ago we were all 'information poor' and the limited data-handling capabilities of early spatial analysis methods reflected this. Arguably, it was this impoverishment that was the root cause of the failure of many such methods to generate detailed insights into the functioning of social systems (see Openshaw and Alvanides, Chapter 18). The potential to build detailed data-rich depictions of reality within GIS will make some problems more transparent, yet others will likely be further obscured. From a pessimist's standpoint, data-rich modelling within GIS represents a return to the shifting sands of naive empiricism. For the optimist, sensitive honing of such data to context allows data-rich models to shed light upon a wider range of social and economic research problems.

In this new world, a given set of data is likely to fall into many different hands during its life. It may be assembled from a mixture of field and remote sensing sources, interpreted by a specialist, catalogued by an archivist or librarian, used by scientists and problem-solvers, and passed between its custodians using a range of technologies (Figure 2). It is quite possible in today's world that the various creators and users of data share little in the way of common disciplinary background, leaving the dataset open to misunderstanding and misinterpretation. Recent interest in metadata, or ways of describing the contents of datasets, is directed at reducing some of these problems, but the easy access to data provided by the Internet and various geographical data archives has tended to make the problem of inappropriate use or application worse.

These issues are particularly prominent in the case of data quality, and the ability of the user of a dataset to understand its limitations, and the uncertainty that exists about the real phenomena the data are intended to represent. To take a simple example, suppose information on the geodetic datum underlying a particular dataset – potentially a very significant component of its metadata – were lost in transmission between source and user; alternatively, suppose that the user simply assumed the wrong datum, or was unaware of its significance. This loss of metadata, or specification of the data content, is equivalent in every respect to an actual loss of accuracy equal to the difference between the true datum and the datum assumed by the user, which can be several hundreds of metres. This is perhaps an obvious example, but what, say, is the magnitude of error associated with soil profile delineation? What is the magnitude of likely ecological fallacy associated

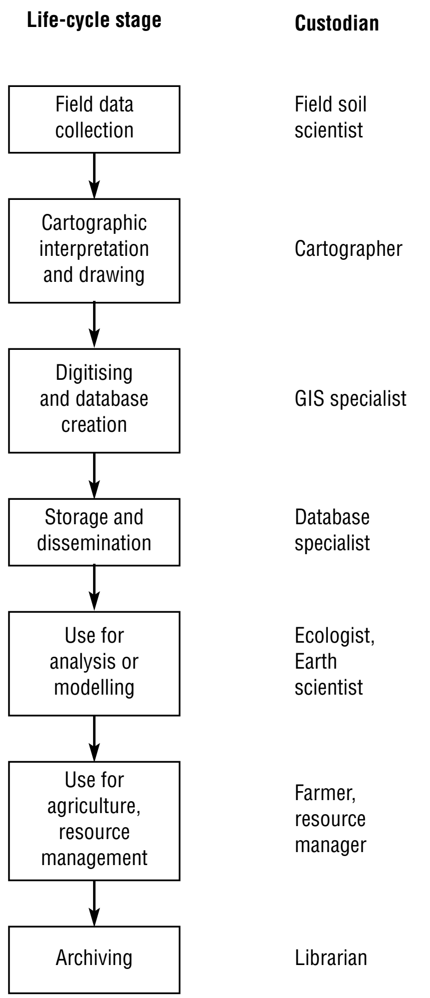

Fig 2. The life-cycle of a soil database: an example of the complex patterns of custodianship and transfer now common for many types of spatial data.

with comparison of a geodemographic classifier with the results of a survey? In short, the quality of a dataset to a user is a function of the difference between its contents and the user's understanding of its meaning, not the creator's.

3.3 Data sharing

In this new world of shared data, the term metadata has come to function as the equivalent of documentation, cataloguing, handling instructions, and production control. The US Federal Geographic Data Committee's Content Standard for Digital Geospatial Metadata (FGDC 1994) has been very influential in providing a standard, which has been emulated frequently (see Salgé, Chapter 50). If the custodian of a large collection of geographical datasets provides metadata in this form, it is possible for others to search its records for those that match their needs. The FGDC's National Geospatial Data Clearinghouse (*http://www.fgdc.gov*) is one such directory (see also the Alexandria Digital Library project to provide distributed library services for geographically referenced datasets: Smith et al 1996; and see *http://alexandria.sdc.ucsb.edu*).

The user of a traditional library will rarely know the exact subject of a search – instead, library search has an essential fuzziness, which is supported by the traditional library in several essential ways. By assigning similar call numbers to books on similar subjects, and shelving by call number, the traditional library is able to provide an environment that allows the user to browse the collection in a chosen area. But this support is missing when the records of a metadata file are searched using simple Boolean methods. It would make better sense to model the search process as one of finding the best fit between a metadata record representing the user's ideal, and metadata records representing the datasets available. It is very unlikely, after all, that data exist that perfectly match the needs of a given problem, especially in the ideal world of problem-solving represented earlier. This is especially true when the object of search is to find data covering a particular location, as it often is in the GIS context. In such cases, it seems very unlikely that there will be an exact fit between the area requested by the user, and the area covered by a data set in an archive.

3.4 New techniques for analysis

Several chapters in this section have focused on new methods of spatial analysis, particularly new methods that have emerged in the data-rich computational environment now available to scientists. These include neural nets (Fischer, Chapter 19), new methods of optimisation such as simulated annealing and genetic techniques, and computationally intensive simulation. The term *geocomputation* (Openshaw and Alvanides, Chapter 18) has been suggested. Anselin (Chapter 17) and others have extended the principles of exploratory data analysis (Tukey 1970) to spatial data.

In science generally, the combination of vast new sources of data and high-speed computation have led to an interest in methods of *data mining*, which implies the ability to dredge data at very high speed in a search for patterns of scientific interest. In a geographical context, the vague notion of 'scientific interest' might suggest the need for methods to detect features or measurements that are inconsistent with their surroundings, in apparent violation of Tobler's 'first law of geography' (Tobler 1970; see also Johnston, Chapter 3). Linearities in images are of potential interest in geological prospecting; and one can imagine circumstances in which atmospheric scientists might want to search large numbers of images for patterns consistent with weather events. Such techniques of pattern recognition were pioneered many years ago in particle physics, to search vast numbers of bubble-chamber photographs for the tracks characteristic of rare new particles.

One might argue that such techniques represent a renewal of interest in inductive science – the search for regularities or patterns in the world that would then stimulate new explanatory theories (see Fischer, Chapter 19). Inductivism has fallen out of fashion in recent decades, at least in disciplines that focus on geographical data, leading one to ask whether a renewal of interest represents a fundamental shift in science, or merely a response to the opportunities for data-led thinking offered by more powerful technology. On this issue the jury is clearly still 'out' – geocomputation has not yet provided the kinds of new insights that might support a broad shift to inductivism.

3.5 New computer architectures

The communication technologies that have emerged in the past decade have allowed a fundamental change in the architecture of computing systems. Instead of the early mainframes and later stand-alone desktop systems, today's computers are linked with high-speed networks that allow data, software, and storage capacity located in widely scattered systems to be integrated into functioning wholes. Data can now be 'served' from central sites on demand, avoiding the need to disseminate many copies, with subsequent confusion when updates are needed. Coleman (Chapter 22) reviews the architectural alternatives now common in computing systems, and their technical impacts on GIS.

The new approaches to computing that are possible in this interconnected environment are having a profound effect on spatial analysis. Because it is no longer possible to assume a lifetime association between a user and a particular system design, there are mounting pressures for standards and interoperability between systems to counter the high costs of retraining of staff and reformatting of data.

The proprietary GIS that once dominated the industry attempted to provide a full range of GIS services in one homogeneous environment. Data were stored in proprietary formats, often kept secret by vendors to maintain market position, but making it difficult for others to expand the capabilities of the system by programming extra modules. The 'open GIS' movement (Buehler and McKee 1996 and see *http://www.ogis.org*) mirrors efforts in other areas of the electronic data processing world to promote interoperability, open standards and formats, and easy exchange from one system to another. While such ideas were often regarded as counter to the commercial interests of vendors, there is now widespread acceptance in the industry that they represent the way of the future.

The implications of open systems for spatial analysis are likely to be profound. First, they offer the potential of a uniform working environment, in which knowledge of one system is readily transferable to another. To make this work, however, it will be necessary to achieve a uniform view, and its acceptance across a heterogeneous user community. There is no prospect of interoperability and open systems without agreement on the fundamental data models, terminology, and objectives of GIS-based analysis. Thus much effort will be needed on the part of the inventors and implementors of spatial analysis to develop this uniform view.

Second, the possibility of easy sharing of data across systems gives even greater momentum to efforts to make geographical information more shareable, and even greater demands on the existence and effectiveness of metadata.

Third, interoperability is likely to create an environment in which it is much easier to implement methods of spatial analysis in GIS. Traditionally, vendors of monolithic systems have added functions when market demand appears to justify the development costs. It has been impossible, in a world of proprietary systems, for third parties to add significant functionality. Thus expansion of spatial analytic capabilities has been slow, and has tended to reflect the needs of the commercial market, rather than those of science and problem-solving, when these diverge. In a world of open systems it will be much easier to add functions, and the new environment will encourage the emergence of small companies offering specialised functionality in niche markets.

Finally, new interoperable approaches to software will encourage the modularisation of code (Sondheim et al, Chapter 24). It is already possible in some mainstream software environments to launch one specialised application within another – for example, to apply spreadsheet functions to information in a word processing package. This ‘plug and play’ environment offers enormous scope to GIS, since it will lead ultimately to a greater integration of GIS functions, and map and imagery data in general, into mainstream electronic data processing applications.

The scientific world has grown used to a more or less complete separation between data, and the functions that operate on and manipulate data. Functions are part of ‘analysis’, which plays a role in the traditional approach to problem-solving outlined earlier that is clearly distinct from that of data. But it has already been argued that in a world of extensive data sharing and interaction between disciplines it is impossible to think of data in isolation from its description, or metadata, which allows the meaning of information to be shared.

In the abstract world of object-oriented methods, it is argued that the meaning of data lies ultimately in the operations that can be performed. If datasets exist in two systems, and pairs of functions exist in both systems that produce the same answers, then the two datasets are the same in information content, irrespective of their specific formats and arrangements of bits. It makes sense, then, to *encapsulate* methods with data. When more than one method is available to perform a given function, it makes sense for the choice to be made by the person best able to do so, and for the method thereafter to travel with the data. For example, a climatologist might encapsulate an appropriate method for spatial interpolation with a set of point weather records, because the climatologist is arguably better able to select the best method of spatial interpolation, given his or her knowledge of atmospheric processes.

In future, and especially given the current trend in computing to object-oriented methods, it is likely that the distinction between data and methods will become increasingly blurred. Commonly used techniques of spatial analysis, such as spatial interpolation, may become encapsulated with data in an extension of the concept of metadata to include methods. Of course this assumes that methods are capable of running in a wide variety of host systems, which takes the discussion back to the issue of interoperability introduced earlier.

4 SPATIAL ANALYSIS IN PRACTICE

At this stage, it seems useful to introduce a discussion of the practical problems which face the users of today’s GIS. While it is now possible to undertake a wide range of forms of spatial analysis, and to integrate data from a range of sources that would have seemed inconceivable as little as five years ago, there continue to be abundant limitations that impede the complete fulfilment of the technology’s promise. The following subsections discuss several of these enduring impediments.

4.1 Absolute and relative position

First, and perhaps foremost, are problems of varying data quality. In science generally it is common to express quality in terms such as ‘accurate to plus or minus one per cent’. But while such methods are useful for many types of data, they are less so when the data are geographical. The individual items of information in a geographical dataset are typically the result of a long and complex series of processing and interpretation steps. They bear little relationship to the independent measurements of traditional error analysis. Section 2 dealt at length with the data quality issue, and the theme is taken up again in the context of decision-making by Hunter (Chapter 45), and those discussions will not be repeated here. Instead, the following discussion is limited to the particular problems encountered when merging datasets.

While projections and geodetic datums are commonly well-documented for the datasets produced by government agencies, the individual scientist digitising a map may well not be in a position to identify either. The idea that lack of specification could contribute to uncertainty was discussed earlier, and its effects will be immediately apparent if a dataset is merged with one based on another projection or datum. In practice, therefore, users of GIS frequently encounter the need for methods of *conflation*, a topic discussed in detail below.

The individual items of information in a geographical dataset often share lineage, in the sense that more than one item is affected by the same error. This happens, for example, when a map or photograph is registered poorly – all of the data derived from it will have the same error. One indicator of shared lineage, then, is the persistence of error – all points derived from or dependent on the same misregistration will be displaced by the same or a similar amount. Because neighbouring points are more likely to share lineage than distant points, errors tend to show strong positive spatial autocorrelation (Goodchild and Gopal 1989).

Rubber-sheeting is the term used to describe methods for removing such errors on the assumption that strong spatial autocorrelations exist. If errors tend to be spatially autocorrelated up to a distance of x, say, then rubber-sheeting will be successful at removing them, at least partially, provided control points can be found that are spaced less than x apart. For the same reason, the shapes of features that are less than x across will tend to have little distortion, while very large shapes may be badly distorted. The results of calculating areas, or other geometric operations that rely only on relative position, will be accurate as long as the areas are small, but will grow rapidly with feature size. Thus it is important for the user of a GIS to know which operations depend on *relative* position, and over what distance; and where *absolute* position is important (of course the term absolute simply means relative to the Earth frame, defined by the Equator and the Greenwich meridian, or relative over a very long distance).

When two datasets are merged that share no common lineage (for example, they have not been subject to the same misregistration), then the relative positions of objects inherit the absolute positional errors of both, even over the shortest distances. While the shapes of objects in each dataset may be accurate, the relative locations of pairs of neighbouring objects may be wildly inaccurate when drawn from different datasets. The anecdotal history of GIS is full of such examples – datasets which were perfectly adequate for one application, but failed completely when an application required that they be merged with some new dataset that had no common lineage. For example, merging GPS measurements of point positions with streets derived from the US Bureau of the Census TIGER (Topologically Integrated Geographic Encoding and Referencing) files may lead to surprises where points appear on the wrong sides of streets. If the absolute positional accuracy of a dataset is 50 metres, as it is with parts of TIGER, then such surprises will be common for points located less than 50 metres from the nearest street. In a similar vein but in the context of the fragmented data holdings of UK local authorities, Martin et al (1994) describe the problems and mismatches inherent in matching individual and household information with property gazetteers.

4.2 Semantic integration

Some of the most challenging problems in GIS practice occur in the area of semantic integration, where integration relies on an understanding of meaning. Such problems can occur between geographical jurisdictions, if definitions of feature types, or classifications, or methods of measurement vary between them. It is common, for example, for schemes of vegetation classification to vary from one country to another, making it difficult to produce horizontally merged data (Mounsey 1991). 'Vertical' integration can also be problematic, as for example in merging the information on land classification maps produced by different agencies, or different individuals (Edwards and Lowell 1996).

While some of these problems may disappear with more enlightened standards, others merely reflect positions that are eminently reasonable. The problems of management of ecosystems in Florida are clearly different from those of Montana, and it is reasonable that standards adopted by the two states should be different (see Fisher, Chapter 13). Even if it were possible to standardise for the entire US, one would be no further ahead in standardising between the US and other countries. Instead, it seems a more reasonable approach is to achieve interoperability without standardisation, by more intelligent approaches to system design.

4.3 Conflation

Conflation appears to be the term of choice in the GIS community for functions that attempt to overcome differences between datasets, or to merge their contents. Conflation attempts to replace two or more versions of the same information with a single version that reflects the pooling of the sources; it may help to think of it as a process of weighted averaging. The complementary term '*concatenation*' refers to the integration of the sources, so that the contents of both are accessible in the product. The polygon overlay operation familiar to many GIS users is thus a form of concatenation.

Two distinct forms of conflation can be identified, depending on the context:

1 conflation of feature geometry and topology, and concatenation of feature attributes;
2 conflation of geometry, topology, and attributes.

As an example of the first case, suppose information is available on the railroad network at two scales, 1:100 000 and 1:2 million. The set of attributes available is richer at the 1:2 million scale, but the geometry and topology are more accurate at 1:100 000. Thus it would be desirable to combine the two, thereby discarding the coarser geometry and topology.

As an example of the second case, consider a situation in which soils have been mapped for two adjacent counties, by two different teams of scientists. At the common border there is an obvious problem, because although the county boundary was defined by a process that was in no way dependent on soils, the border nevertheless appears in the combined map. Thus it would be desirable to 'average' the data at and near the boundary by combining the information from both maps in compatible fashion. As these two examples illustrate, the need for conflation occurs both horizontally, in the form of edge matching, and 'vertically'. A further example of the second case is provided by spatially extensive property valuation exercises, such as that which accompanied the introduction of the UK Council Tax (Longley et al 1994): surveyors were each individually responsible for allotted areas and conflation of estimates around the area boundaries was used to enhance consistency.

4.4 Perfect positioning

Some of the problems of conflation, and of relative and absolute positional accuracy, might be expected to dissipate as measurement of position becomes more and more accurate, leading eventually to 'perfect' positioning. Unfortunately, there are good reasons to anticipate that this happy state will never be reached. Although the positions of the Greenwich meridian and various geodetic control points have been established by fixing monuments, fundamental uncertainty will continue to be created by seismic motions, continental drift, and the wobbling of the Earth's axis. Any mathematical representation of the Earth's shape must be an approximation, and different approximations have been adopted for different purposes. Moreover, there will always be a legacy of earlier, less accurate measurements to deal with. Thus it seems GIS will always have to deal with uncertainty of position, and with the distinctions between relative and absolute accuracy, and their complex implications for analysis.

Instead, strategies must be found for overcoming the inevitable differences between databases, either prior to analysis or in some cases 'on the fly'. Consider, for example, the problems caused by use of different map databases for vehicle routing. Systems are already available on an experimental basis that broadcast information on street congestion and road maintenance to vehicles, which are equipped with map databases and systems to display such information for the driver. In a world of many competing vendors, such systems will have to overcome problems of mismatch between different databases, in terms both of position and of attributes. For example, two databases may disagree over the exact location of 100 Main Street, or whether there *is* a 100 Main Street, with potentially disastrous consequences for emergency vehicles, and expensive consequences for deliveries (see Cova, Chapter 60). Recent trends suggest that the prospects for central standardisation of street naming by a single authority are diminishing, rather than growing.

5 CONCLUSION

The prospects for spatial analysis have never been better. Data are available in unprecedented volume, and are easily accessed over today's communication networks. More methods of spatial analysis are implemented in today's GIS than ever before, and GIS has made methods of analysis that were previously locked in obscure journals easy and straightforward to use. Nevertheless, today's environment for spatial analysis raises many issues, not the least of which is the ability of users to understand and to interpret correctly. Questions are being raised about the deeper implications of spatial analysis, and the development of databases that verge on invasion of individual privacy (Curry, Chapter 55). Our expectations may be unreasonable given the inevitable problems of spatial data quality.

Postmodern scientific discourse has fragmented, and with regard to GIS there is diversity not just in the sources of digital geographical information, but also increasingly (and especially with respect to human systems) in its interpretive meaning. There is a need to communicate clear interpretive conceptions of the rich but widely-distributed and piecemeal data holdings of networked GIS. We need now to think about spatial analysis not just in terms of outcomes, but also in terms of inputs. Metadata will come to fill a crucial role in the comparative assessment *between* different datasets just as, in a previous era, exploratory data analysis allowed *within* dataset assessment to take place. Such assessment and interpretation will become essential in an era in which the relative importance of conventional, governmental data providers is set to diminish. It seems clear that tomorrow's science will be increasingly driven by complex interactions, as data become increasingly commodified, technology increasingly indispensible to science, and conclusions increasingly consensual. New philosophies of science that reflect today's realities are already overdue.

These changes are profound and far-reaching, but they provide grounds for cautious optimism about the future of GIS-based spatial analysis. The established self-perception of rigour among spatial analysts has hitherto been to some extent misplaced, in that data quality, resolution, and richness have not always been commensurate with the sophistication of spatial analytic methods. However nostalgically we may at times now view it, the linear project design was by no means a panacea in practice.

If science and problem-solving are to be constrained by these new realities, then what kinds of spatial analysis are most likely to dominate in the coming years? The points raised in this chapter's discussion suggest that the future environment will favour the following:

1. data whose meanings are clearly understood, making it easier for multidisciplinary teams to collaborate;
2. data which are routinely collected in the day-to-day functioning of society and the everyday interactions between humans and computers;
3. data with widespread use, generating demands that can justify the costs of creation and maintenance;
4. data with commercial as well as scientific and problem-solving value, allowing costs to be shared across many sectors;
5. methods of analysis with commercial applications, making it more likely that such methods will be implemented in widely available form;
6. methods implemented using general standards, allowing them to be linked to other methods using common standards and protocols.

References

Bailey T C, Gatrell A C 1995 *Interactive spatial data analysis*. New York, John Wiley & Sons Inc.

Berry B J L, Marble D F (eds) 1968 *Spatial analysis: a reader in statistical geography*. Englewood Cliffs, Prentice-Hall

Birkin M 1995 Customer targeting, geodemographics and lifestyle approaches. In Longley P, Clarke G (eds) *GIS for business and service planning.* Cambridge (UK), GeoInformation International

Buehler K, McKee L (eds) 1996 *The Open GIS guide*. Wayland, The Open GIS Consortium Inc.

Chorley R J, Haggett P 1965 Trend surface models in geographical research. *Transactions of the Institute of British Geographers* 37: 47–67

Chrisman N R 1997 *Exploring geographic information systems*. New York, John Wiley & Sons Inc.

Edwards G, Lowell K E 1996 Modeling uncertainty in photointerpreted boundaries. *Photogrammetric Engineering and Remote Sensing* 62: 377–91

FGDC 1994 *Content standards for digital geospatial metadata*. Washington DC, Federal Geographic Data Committee, Department of the Interior. *http://www.fgdc.gov*

Fotheringham A S, Rogerson P A (eds) 1994 *Spatial analysis and GIS*. London, Taylor and Francis

Goodchild M F 1988 A spatial analytic perspective on geographical information systems. *International Journal of Geographical Information Systems* 1: 327–34

Goodchild M F, Gopal S 1989 *Accuracy of spatial databases*. London, Taylor and Francis

Goodchild M F, Haining R P, Wise S 1992 Integrating GIS and spatial analysis: problems and possibilities. *International Journal of Geographical Information Systems* 6: 407–23

Haining R P 1990 *Spatial data analysis in the social and environmental sciences*. Cambridge (UK), Cambridge University Press

Harman H H 1976 *Modern factor analysis.* Chicago, University of Chicago Press

Hartshorne R 1959 *Perspective on the nature of geography.* Chicago, Rand-McNally/London, John Murray

Levine G 1981 *Introductory statistics for psychology: the logic and the methods*. New York, Academic Press

Longley P, Clarke G 1995 Applied geographical information systems: developments and prospects. In Longley P, Clarke G (eds) *GIS for business and service planning*. Cambridge (UK), GeoInformation International: 3–9

Longley P, Higgs G, Martin D 1994 The predictive use of GIS to model property valuations. *International Journal of Geographical Information Systems* 8: 217–35

Marascuilo L A, Levin J R 1983 *Multivariate statistics in the social sciences: a researcher's guide*. Monterey, Brooks/Cole

Martin D 1996 *Geographic information systems: socioeconomic applications*. London, Routledge

Martin D, Longley P, Higgs G 1994 The use of GIS in the analysis of diverse urban databases. *Computers, Environment, and Urban Systems* 18: 55–66

Mounsey H 1991 Multisource, multinational environmental GIS: lessons learned from CORINE. In Maguire D J, Goodchild M F, Rhind D W (eds) *Geographical information systems: principles and applications.* Harlow, Longman/New York, John Wiley & Sons Inc. Vol. 2: 185–200

Siegel S 1956 *Nonparametric statistics for the behavioral sciences*. New York, McGraw-Hill

Smith T R, Andresen D, Carver L, Dolin R, and others 1996 A digital library for geographically referenced materials. *Computer* 29(5): 54, 29(7): 14

Taylor P J 1977 *Quantitative methods in geography: an introduction to spatial analysis.* Boston, Houghton Mifflin

Tobler W R 1970 A computer movie simulating urban growth in the Detroit region. *Economic Geography* supplement 46: 234–40

Tukey J W 1970 *Exploratory data analysis*. Reading (USA), Addison-Wesley

Unwin D J 1981a *Introductory spatial analysis*. London, Methuen

Wrigley N 1985 *Categorical data analysis for geographers and environmental scientists*. Harlow and New York, Longman