Introduction

THE EDITORS

It is clear from the discussion of GIS principles thus far that we now live in a data-rich world in which a vast and increasing array of geographical phenomena are represented in digital form. GIS-based data models are by definition selective abstractions and the data used to build them are error prone, yet they can lay the foundation for legitimate context-sensitive inputs to generalisable analysis and forecasting. The contributions to this section set out to identify how GIS allows us to summarise the properties of spatial distributions, inductively solve spatial problems, and contribute towards spatial decision-making.

The established paradigm for quantitative description and generalisation about geographical phenomena has been to use spatial statistics. What makes spatial statistics distinct from its parent discipline is its concern with observations which are located near to one another in space and which, as a consequence, tend to share similar attribute values – in Anselin's words (Chapter 17): 'the phenomenon where locational similarity (observations in spatial proximity) is matched by value similarity (correlation)'. Spatial and geostatisticians have developed a range of specialised methods and techniques for dealing with such cases. The emergence and chronological development of spatial statistics in the pre-GIS era is the first theme considered in the contribution by Art Getis (Chapter 16).

An important emergent debate within GIS has been the continuing relevance of spatial statistical approaches. Briefly, the key arguments may be summarised as: first, spatial statistics developed in what all of the contributors to this section would recognise as a 'data-poor' era, in which statistics were based upon few (by present-day standards) observations; second, this paucity of data made computation a fairly straightforward procedure; and, third, the geography of areal units was fixed (and usually coarse), and not itself subjectable to the range of transformations and sensitivity analyses that have been outlined in previous contributions to this Principles part of the book. There is evident consensus among the contributors to this section that spatial analysis has received far too little attention in the development of GIS, yet they have some different views as to whether and how spatial statisticians can continue to contribute practical spatial analysis skills to GIS. On the one hand, Getis (Chapter 16) and Luc Anselin (Chapter 17) set out some of the enduring contributions that spatial statistics is making to GIS, especially in the area of exploratory spatial data analysis (ESDA). On the other hand, the views of Stan Openshaw and Seraphim Alvanides (Chapter 18) and Manfred Fischer (Chapter 19) lean towards the view that the changes associated with the development of GIS require a more root-and-branch reappraisal of the practice of spatial analysis in GIS. Richard Church (Chapter 20) presents a review of the ways in which GIS is being applied to locational analysis problems in GIS.

The review by Getis (Chapter 16) charts the considerable progress that has been made in developing ESDA spatial statistics through the media of GIS. However, he laments the relative lack of progress in developing spatial hypothesis testing within GIS. By implication, he seems to sound a warning that the media of GIS are in danger of overwhelming the message of spatial statistical analysis as conventionally understood. Indeed, developments in scientific visualisation and ESDA appear to have contributed little to our incomplete theoretical and statistical understandings of the ways that observations should be differentially weighted across space (i.e. the effects of distance) and the effects of boundaries and edges on the results of spatial statistical analysis.

To dwell upon such (possibly unresolvable) statistical issues might be seen as admonishing failure, when the media of GIS have been demonstrably effective in exploring locational scenarios and visualising spatial outcomes. Data exploration within GIS is the domain of ESDA techniques – defined by Anselin (1994) as being used 'to describe and visualise spatial distributions, identify atypical locations (spatial outliers), discover patterns of spatial association (spatial clusters) and suggest different spatial regimes and other forms of spatial instability or spatial non-stationarity'. Anselin's contribution to this volume develops the views that GIS has become data rich but theory poor, and that ESDA statistics can be used to structure, visualise, and explain a wide array of geographical data. His comprehensive review builds upon Getis' conceptions of spatial autocorrelation (i.e. geostatistical and spatial weights formulations), and in this context goes on to identify important domains of ESDA as pertaining to identification of local patterns of spatial association within global patterns.

For Openshaw and Alvanides, by contrast, the dominant impression is that we are seemingly unable to structure and analyse the vast quantities of spatial data that are now available, and that this failure reflects our continued adherence to the 'pre-GIS' spatial statistical analysis paradigm. Given that the development of GIS and of modern databases was substantially technology-led, it is at least intuitively plausible that analysis might be developed through the same guiding force. Thus for Openshaw and Alvanides (Chapter 18) the way forward lies through broad-based techniques of 'geocomputation' – that is, 'the adoption of a large-scale computationally-intensive approach to the problems of physical and human geography in particular, and the geosciences in general'. This kind of approach has undoubtedly had a profound impact through demonstration and exploration of modifiable areal unit effects, and the geocomputational paradigm has clear application in allowing spatial analysts to create zone designs that satisfy particular constellations of constraints. Building upon this, Openshaw and Alvanides see generic solutions emerging from the use of 'intelligent' pattern-seeking techniques which might become integral to GIS. The implication is that new computational techniques may be used to search for new theories and to generate new knowledge using applied, problem-solving approaches.

Fischer (Chapter 19) takes a wider perspective on the emergence of what he terms 'computational intelligence' (CI) technologies in relation to classical spatial statistics. He shares many of the expressed doubts of Openshaw and Alvanides that conventional spatial statistics can be adapted to accommodate the richness of the GIS environment, and instead advocates a CI paradigm (involving artificial life, evolutionary computation, and neural networks) based essentially upon geocomputation. He develops an extended case study involving the use of a neural net model for satellite image classification, and shows how spatial analysis proceeds through model specification, estimation, and testing phases. His exposition is lucid and non-technical, and he is at pains to dispel suspicion about the 'mystique and metaphorical jargon' hitherto associated with CI techniques. Whether the widest GIS audience will share his confidence that use of the 'universal language of mathematics' alone will dispel such scepticism is a moot point: indeed, this raises important issues about the gulf between 'machine-intelligent' spatial analysis of digital abstractions and scientific theory and reasoning as conventionally understood.

We have seen in the first section to this part of the book how ontology ('meta-theory') prescribes particular detailed approaches to analysis, and that no stage in scientific reasoning can be considered in isolation. The emphasis in much of the second part was on demonstrating how digital data provide only imperfect, incomplete, and error-prone representations of reality. Together this would seem to require that choice of spatial analytical method is rational, informed, and sensitive to context, rather than being data led in a naive empiricist way. Just as Goodchild and Longley (Chapter 40) argue later in the Technical issues section, that volume of data is not a substitute for scientific rigour, so the paradigm of geocomputation needs to demonstrate why and in what circumstances spatial analysts should have 'confidence' (both broadly and narrowly defined) in its substantive findings. Openshaw and Alvanides cite a number of high-profile and celebrated case studies which have adopted what has come to be described as a geocomputational approach (notably in the identification of clusters of diseases), yet none appears wholly to have withstood scientific scrutiny: as such, the jury must still be out regarding the advisability of wholesale reliance upon geocomputational approaches. It is beyond doubt that the world has never been as data rich but, in the realm of spatial analysis, there have been a number of false dawns before – as our comments in the Introduction on the coverage of artificial intelligence in the first edition of this book testifies. Thus while Openshaw and Alvanides conjecture that ‘it may be possible to compute our way out of the data swamp’, it remains a moot point as to how and why we may have become lured into it in the first place. It remains to be seen whether and how far the protagonists of computational approaches are able to rebuff concerns that they are indulging in uninformed pattern-seeking empiricism in the absence of clear theoretical guidance as conventionally understood. Elsewhere, Openshaw and Openshaw (1997) have begun to demystify new ways of viewing our digital world although, at the other end of the spectrum, Curry (1995: 82) has concluded that ‘to develop an understanding of the data adequate to a resolution of the problems which arise in the production of a GIS would very likely render those systems irrelevant’. There is still some way to go, both with regard to demystifying technique and to understanding data.

An oft-rehearsed but nevertheless resonant theme running through all of these contributions concerns the integration of spatial analytic functionality into proprietary GIS. Neither spatial statistical models, nor geocomputational methods, nor refined ESDA techniques form part of proprietary GIS. This is in part because of user ignorance about the range of simplifying assumptions that routine usage brings, and in part because (as a consequence) vendors are unlikely to prioritise functionality for which there are no strong user demands. Given the very small likelihood of fully integrated spatial analytical GIS in the foreseeable future, Getis, Openshaw and Alvanides, and Anselin each explore a number of options for the close and loose coupling of GIS to specialist spatial analysis packages, as well as the potential role of the Internet as a platform for integrating GIS and spatial analysis.

Some of the earliest applications of spatial analysis involved the calculation of statistical moments and distributions for the classic locational models of geography, models which were used in the pre-GIS era to identify the best location for industrial and service facilities. The enduring relevance of locational modelling to GIS is the focus of the contribution by Church (Chapter 20). A conventional facet to this problem has been the use of GIS to locate single facilities with respect to spatial patterns of demand and, through use of the overlay model, to identify corridors linking the different sites involved in activities in the most cost-efficient way. In recent years, progress has been made towards solving multiple-site location problems, most notably instances in which sites may need to be relocated in response to very short-term changes in demand (as in the relocation of ‘on call’ ambulances to cover for vehicles which are already attending emergencies). This echoes Getis’ sentiment of continuity of approach in spatial analysis, but what has changed here with the innovation of GIS is the richness of the data which can be brought to bear on site location problems, the computational support for new and complex location–allocation algorithms, and the visual quality of the computer environment for data exploration, investigation of scenarios and decision support. Location modelling also provides a good exemplar of the wider problems that remain on the spatial analysis agenda, namely: the compatibility of data structures/data quality issues; the representation of spatial patterns of demand, and the screening process used to identify sites; the size and scale of the elemental units used to specify location–allocation problems; and the ways in which errors are created and propagated in formulation of problems. GIS is clearly having far-reaching impacts upon the specification, estimation, and testing of spatial relationships, and the Applications part of this book (in Volume 2) provides evidence of the practical relevance of these techniques (e.g. Cova, Chapter 60; Gatrell and Senior, Chapter 66).

References

Anselin L 1994a Exploratory spatial data analysis and geographic information systems. In Painho M (ed.) *New tools for spatial analysis*. Luxembourg, Eurostat: 45–54

Curry M R 1995a GIS and the inevitability of ethical inconsistency. In Pickles J (ed.) *Ground truth: the social implications of geographic information systems*. New York, Guilford Press

Openshaw S, Openshaw C A 1997 *Artificial intelligence in geography*. Chichester, John Wiley & Sons