11

Summary

11.1 Review of key concepts

The book has introduced a range of general concepts and specific approaches for spatial data analysis. From data models to aspatial statistics, moving windows, geographical weights, and spatial autocorrelation to methods for overlay of vector features, local regression, analysis of point patterns, and spatial interpolation, the focus has been broad. Chapter 2 introduced a range of core ideas, including data models, databases, spatial scale, spatial data collection, data errors, visualization, and simple data queries. In Chapter 3, some basic statistical approaches were detailed to give background for the spatial statistical methods presented later in the book. Chapter 4 was concerned initially with measurement of distances and areas. The following sections dealt with moving windows, geographical weights, the concepts of spatial dependence and autocorrelation, and the ecological fallacy, and the modifiable areal unit problem. In Chapter 5, overlay of different features and identification of areas fulfilling a range of criteria were the focus. Chapter 6 dealt with the characterization and analysis of networks. Chapter 7 outlined some approaches for analysing point patterns. In Chapter 8, various methods for exploring spatial patterning in variables were detailed. In particular, approaches for characterizing spatial structure in single variables (i.e. the degree to which values close together in space are similar) and methods for analysing geographical variations in the relationships between different variables were illustrated. Chapter 9 presented methods for spatial interpolation and Chapter 10 was concerned with the analysis of grids and surfaces.

While the focus is quite broad there are links running through the book that relate to the general approaches used. For example, the methods outlined in Chapters 7, 8, 9 and 10 make use of geographical weighting schemes. Spatial data analysis encompasses a very large array of methodological frameworks developed to overcome a huge variety of different problems. The range of approaches introduced has been from methods simply for exploring spatial patterns, to methods for assessing how meaningful (in some sense) those patterns might be, through to methods for finding optimal solutions to problems (e.g. the shortest path or the most suitable area) and methods

for estimating data values at locations where there is no sample. This book could have been organized in many different ways and the balance of coverage of different methods could have been altered markedly. Nonetheless, it is hoped that it covers in reasonable detail a sufficient variety of kinds of approaches, if not specific methods, to demonstrate clearly the diversity of approaches.

There is very much more that could be said about the issues raised in this book and some suggestions as to how to proceed and further develop knowledge of spatial data analysis are given in Section 11.4. Appendix G provides a list of some problems and the corresponding solutions detailed in this book.

11.2 Other issues

There are many issues touched on within this book that could have been developed substantially, and references to additional sources are given to allow readers to expand their knowledge. There are also many issues that could have been discussed, but which were considered outside the remit of the book. However, this book is intended as a starting point that provides pointers to other sources where necessary and the coverage of material within the book is necessarily focused on particular issues.

Many specific issues that some readers might like to have seen covered are omitted, and this is necessarily the case because of limitations of space and because the key aim of this book is to introduce in a focused way a relatively limited array of key concepts and methods. As an example, the book discusses connectivity in terms of identifying neighbouring areas. An important area of research, with many applications, concerns connectivity, as well as the form of landscape areas. McGarigal and Marks (1995) present a software package, Fragstats, which is designed to quantify landscape structure (e.g. by size or shape of landscape patches or their density over an area). Such approaches are not yet a standard part of GIS software and are, therefore, not discussed in this book. A more general omission is geographic data mining. The theme of geographic data mining (Shekar *et al.*, 2003; Miller, 2008) is an important one. Geographic knowledge discovery, with, as a core component, geographic data mining is based on the belief that there is new and useful knowledge to be extracted from the vast array of geographic data sources now available (Miller, 2008). The development and use of dynamic models is central to many applications, but this is another topic excluded from this book. It is hoped that the sources cited will provide the necessary information where some topics are covered only briefly in this book or omitted completely.

11.3 Problems

This book presents a variety of solutions to problems, but it only makes passing reference to other core issues that have an impact on spatial data and their analysis. There are widely used alternatives to most of the methods detailed this book and the

selection of appropriate methods is often not straightforward. Since there are many alternative approaches for the spatial analyst to choose from, often experience is important. As an example, there are numerous methods for spatial interpolation (the focus of Chapter 9) and choice of specific method, interpolation neighbourhood (use all data or a local subset), and use of exact or approximate interpolation, are all likely to be important. Guidance has been offered in the text, although this can be no substitute for experimentation and direct assessment of the differences in results obtained using different combinations of approaches. Descriptions of other methods are provided in the publications referenced in Section 11.4.

Expensive technology and sophisticated-looking software cannot disguise the fact that spatial data analysis is based on models that may be poor abstractions of reality. These limitations and the various kinds of errors that affect any data source must be considered. The propagation of errors from one stage of processing to another (as referred to in Section 2.9.1) is the subject of much research (see Burrough and McDonnell, 1998) and all users of (spatial) data are obliged to consider the quality of their data and possible impacts on analyses that are based on these data.

11.4 Where next?

GIS and spatial data analysis are practical topics. If the tools offered are not used then there is no point in them. To begin to develop an understanding of how and why particular approaches are employed and to become aware of their shortcomings, it is necessary to apply the methods. In short, experimentation is a vital part of the learning process. The book website provides details of some software packages that implement the methods described in the book. Commercial packages like ArcGIS™ include all of the basic analytical functions that most users are likely to need. Indeed almost all of the case studies in this book can be directly replicated using ArcGIS™ and its associated extensions. Some very extensive and powerful packages are completely free and there is, therefore, not necessarily any financial barrier to sophisticated spatial data analysis. For example, the GIS GRASS (Neteler and Mitásová, 2007) and the R programming environment[1] (Bivand *et al.*, 2008) and associated routines offer sophisticated functionality at no cost.

In terms of reading material, there are several book chapters and full books that may make sensible next steps. Brief summaries of methods for the analysis of spatial data which provide pointers to more extensive material are provided by Anselin (2005), Getis (2005), Fischer (2005), and Charlton (2008). The book by O'Sullivan and Unwin (2002) expands on some of the issues discussed here. Bailey and Gatrell (1995), Lloyd (2006), and De Smith *et al.* (2007) provide further accounts of methods discussed in this book. Many books have been written for particular audiences. A good example is the book by Plane and Rogerson (1994), which deals with the analysis of data about human

1 http://www.r-project.org/

populations. For more specific material, see the further reading sections at the end of each chapter.

There are many other issues that are the focus of research efforts. Looking at key journals such as the *International Journal of Geographical Information Science*, *Transactions in GIS*, *International Journal of Remote Sensing*, *Computers and Geosciences*, and others will reveal the range and extent of research that is ongoing. As with any well-developed discipline, it is easy to become lost in the detail of a plethora of diverse applications. However, there are plenty of introductory starting points to guide your efforts, as the further reading lists in this book suggest. The book website provides specific material to support use of this book as well as links to other relevant websites and information about other material that may be useful.

Summary and conclusions

This book has merely scraped the surface of spatial data analysis. Hopefully, it will have successfully introduced some new ideas that will aid understanding of other issues discussed in the literature. It is also hoped that it has assisted more in developing understanding than in generating confusion. The analysis of spatial data is a commercially and academically highly significant field. It is hoped that this book will have introduced some approaches and concepts to a broader readership and illustrated something of the breadth of tools available and the problems that they can be used to solve.