Copyrighted Materials
Copyright © 2010 Oxford University Press Retrieved from www.knovel.com

1

Introduction

1.1 Spatial data analysis

Geographical or spatial data play a vital role in many parts of daily life. Either directly, as in the use of a map for navigating around a city, or indirectly, where we use resources like water or gas, we are dependent on information about where things are located and about the attributes of those things. Making use of spatial data requires a whole set of approaches to extract information from those data and make them useful. Geographical information systems (GIS) play a key role in this context. GIS provide a means of generating, modifying, managing, analysing, and visualizing spatial data. The key contribution of GIS, above and beyond functions provided by other forms of software such as cartographic mapping or computer-aided design packages, is in the analysis of spatial data. Broadly, analysis is concerned with breaking apart a problem with the aim of finding a solution to this problem. In terms of this book, the aim is to introduce some ideas and methods that may be useful in the analysis of spatial data. For example, approaches are presented for:

- summarizing a set of values (e.g. what is the mean average of all values?)
- identifying overlaps between different features (e.g. what areas with pollution above some critical threshold are located in areas that have a population of greater than a particular size?)
- finding the shortest route between one place and another through a network (e.g. a road network)
- identifying clustering in point events such as cases of some disease (e.g. where are disease incidence rates highest?)
- exploring spatial patterning in variables (a variable being a quantity that may vary across samples; precipitation amount or elevation, for example, can be considered variables) (e.g. does the concentration of some pollutant vary spatially and where are values largest?)

- exploring how two or more variables are related at different places (e.g. does the relationship between altitude and snowfall vary from place to place?)
- estimating the values of some property (e.g. precipitation amount) at locations where there are no samples available (necessary as a prior stage to many other procedures)
- assessing the construction costs of alternative routes for a new road.

There are many other kinds of approaches covered in this book, but many are based on common concepts such as measurement of distances or differences in properties in different areas. These fundamental concepts are described along with some particularly widely used approaches and the selected approaches are illustrated with example applications. The emphasis throughout is on *education* rather than simply *training*, based on the conviction that users of spatial data analysis tools should know something about how the approaches work rather than simply how to apply them. Appendix G details some spatial analysis tasks and the sections of the book that contain relevant material. The remainder of this chapter sets out the purpose of the book and its contents.

1.2 Purpose of the book

The aim of this book is to introduce a set of key ideas or frameworks that will give the reader knowledge of the kinds of problems that can be tackled using widely available tools for the analysis of spatial data. Another key concern is that readers *understand* how the methods work, therefore a large majority of methods introduced are demonstrated through small case studies. The book includes detailed coverage of a relatively limited number of basic key methods for the analysis of spatial data. These are intended to illustrate the workings of particular methods as well as to demonstrate key concepts that will support understanding of other approaches.

This is not an introduction to GIS. Readers who wish to know about, for example, data models, databases, or visualization will find brief introductions in this book, although they should consult one of a range of textbooks such as those by Heywood *et al.* (2006) or Longley *et al.* (2005a) for more in-depth accounts. A full description of some key GIS algorithms (put simply, sets of instructions that are worked through to achieve some particular objective) is provided by Wise (2002). In this book, little prior knowledge is assumed of readers. However, it is expected that readers will have some basic knowledge of GIS principles. The book is also not an introduction to statistics, although some key ideas are discussed. The book by Rogerson (2006) provides a good introduction to statistics for geographers. Only a very limited prior knowledge of statistics is required to make full use of the present book and it is an aim in this book that no terms likely to cause confusion will be dropped in without explanation. While it is assumed that most readers will work through the book systematically, it is designed so

that it is possible to dip in to a particular topic and the discussions about methods are, in the majority of cases, fairly self-contained.

Many other books provide introductions to spatial data analysis (or geospatial analysis, as it sometimes called). The book by O'Sullivan and Unwin (2002) provides an excellent account of some key ideas as well as more advanced material. Lloyd (2006), De Smith *et al.* (2007), and Chang (2008) provide other detailed accounts. Lee and Wong (2000) also provide clear introductions to core principles and in-depth accounts of more complicated ideas. The book by Cressie (1993), which focuses on spatial statistics, is encyclopaedic in its coverage, extremely well regarded, and has become something of a standard work. Other works, such as the book by Griffith (1988), seek to disseminate research findings for those already familiar with the principles of spatial data analysis. The present book is intended to have a rather different focus to any of these books. The specific intention is to communicate some ideas and concepts that are central to the analysis of spatial data without discussing alternative approaches in great depth. It is hoped that the book will provide material to develop the reader's conceptual understanding such that books dealing with a broader range of methods can then be encountered with greater confidence. By keeping things to the point, but addressing key issues, I hope that this book will build knowledge and interest and encourage enthusiasm for learning more about spatial data and their analysis.

As noted previously, it is assumed that readers will have some background knowledge of GIS. Reference is made to key principles such as data models and databases with the expectation that these will be familiar. However, an attempt is made to introduce all potentially new topics. The book only briefly discusses key topics such as data input, visualization, errors and error propagation (transfer of errors from one processing stage to another). Instead, the focus is directly on the analysis of spatial data and, where appropriate, other sources of information are suggested. The further reading section at the end of this chapter provides some starting points for major issues that are not discussed in depth here and each chapter has a further reading section at its end.

1.3 Key concepts

A book of this length can only provide an in-depth account of a limited range of methods and ideas; it necessarily skirts over many major issues. Nonetheless, the aim has been to provide sufficient background to some key concepts and some specific approaches that readers will be able to explore other methods in an informed way. The next three chapters seek to outline a set of basic principles, an understanding of which is necessary to make use of the rest of the text. For readers new to the topic of spatial data analysis these chapters may be more challenging than most of the rest, but it is hoped they will allow development of a knowledge base appropriate for making use of the rest of the book.

1.4 Structure of the book

Chapters 2, 3, and 4 introduce a range of key concepts that provide the foundations for the rest of the book. Chapter 2 describes data models (ways of representing real-world objects or features), data management, spatial scale, data collection, data errors, visualization, and querying. Chapter 3 introduces some key statistical concepts and methods. In Chapter 4, some methods for the analysis of spatial data are outlined. More specifically, Chapter 4 deals with measuring distances and areas, moving windows (a key concept in spatial data analysis, which allows assessment of differences between geographical areas), geographical weights, and a variety of additional core issues in the handling and analysis of spatial data. Chapter 5 is concerned with overlaps between features (e.g. do areas with particular characteristics overlap?) while Chapter 6 is concerned with links between component parts of networks (e.g. a road network) to address questions like what is the shortest path between locations A and B? Chapter 7 introduces some methods for the analysis of point patterns (geographically located sets of point events such as cases of a disease). Chapter 8 is concerned with the analysis of spatial patterning in single variables and in relations between multiple variables. In simple terms, it presents methods for exploring how values vary geographically and how the relationships between values vary (the example given in Section 8.7 considers the relationship between altitude and snowfall). Chapter 9 outlines some methods for generating surfaces from point data and for transferring values between different sets of zones. In Chapter 10, the focus is on the analysis of grids and surfaces (both literally, as in topography, and in terms of other properties, such as precipitation, that can be treated as surfaces). Finally, Chapter 11 pulls together some key themes addressed in the book and suggests some ways forward for those who would like to know more about the topics addressed in the book.

Appendices A to F include short outlines of particular topics to support discussion in the main body of the text and these are referred to where relevant. Appendix G includes a table that details some common kinds of problems that are often encountered in spatial data analysis. The relevant sections of the book that offer solutions to these problems are detailed. The table is intended as a means of quickly identifying sections of the text that are relevant for specific applications.

All of the substantive chapters include worked examples using either synthetic or real-world data. In addition, Chapters 5 to 10 include case studies at their conclusions with the data on which these studies are based being provided on the book website along with guidance on how some key methods are implemented in popular GIS packages. All of the synthetic data are also provided on the book website. It is hoped that the text, example applications, and data will, in conjunction, allow readers to develop a firm understanding of the key ideas and techniques described in the book.

Further reading

Each chapter details some texts that provide more detail on the topics discussed. In this case, some introductions to GIS are suggested. General introductions with descriptions of data models, data input procedures, and spatial data management include the books by **Burrough and McDonnell (1998)**, **Longley *et al.* (2005a)**, and **Heywood *et al.* (2006)**. The books edited by **Longley *et al.* (2005b)** and **Wilson and Fotheringham (2008)** provide detailed accounts of key issues and concepts. There are also several general introductions to GIS written for specific disciplines or research areas. These include books for geoscientists (**Bonham-Carter, 1994**), social scientists (**Martin, 1996; Steinberg and Steinberg, 2006**) and archaeologists (**Conolly and Lake, 2006**).

➜ The following chapter introduces some key concepts in GIS and is the first of three chapters that outline some fundamental principles on which the rest of the book builds.