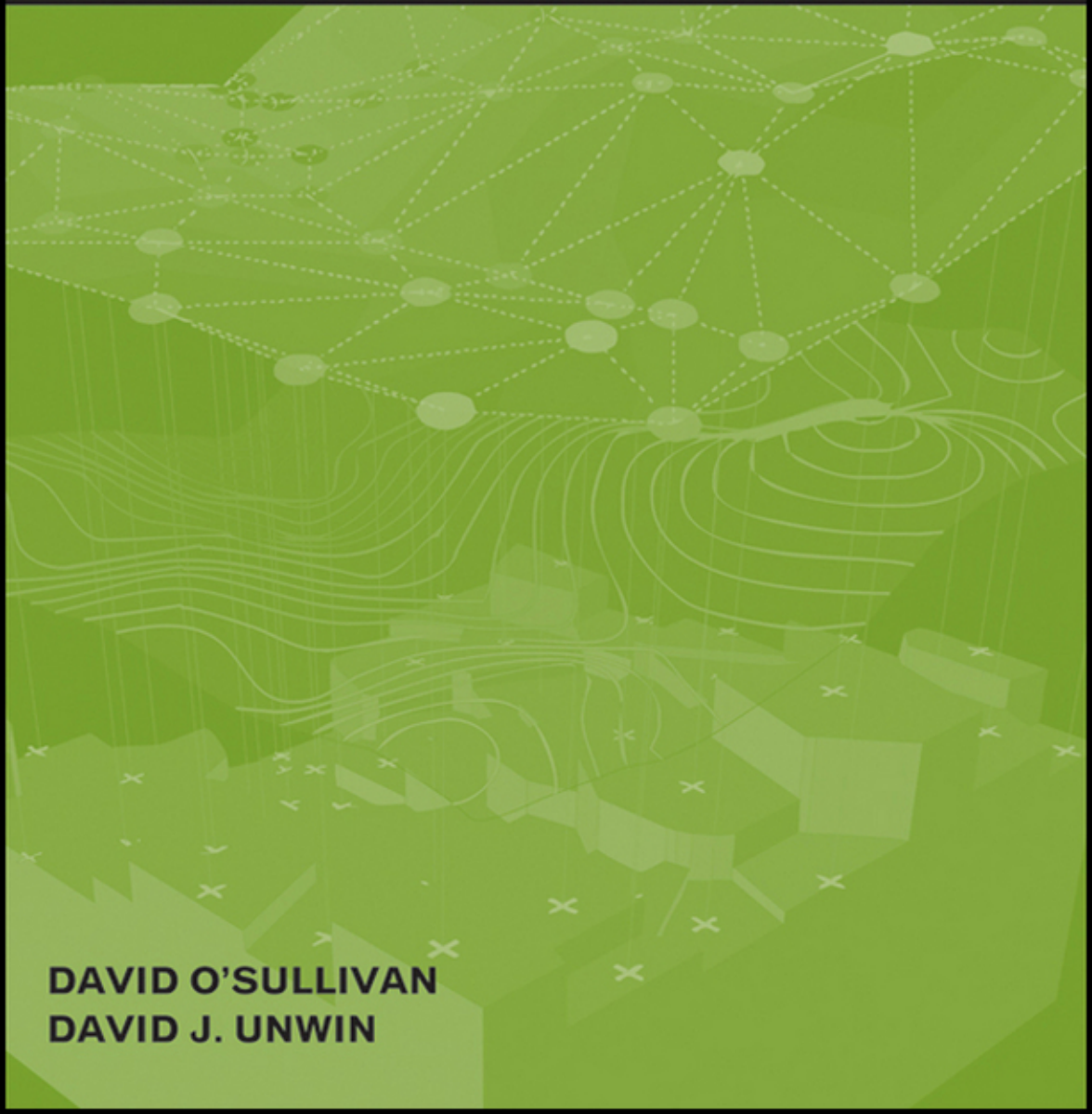


GEOGRAPHIC INFORMATION ANALYSIS

second edition

The background of the cover is a vibrant green. It features a complex network of white dashed lines connecting circular nodes, overlaid on a topographic map with contour lines. At the bottom, there are 3D-style geometric shapes, possibly representing buildings or terrain, with small white 'x' marks scattered across them.

DAVID O'SULLIVAN
DAVID J. UNWIN

GEOGRAPHIC INFORMATION ANALYSIS

GEOGRAPHIC INFORMATION ANALYSIS

Second Edition

David O'Sullivan and David J. Unwin

WILEY

This book is printed on acid-free paper. ☺

Copyright © 2010 by John Wiley & Sons, Inc. All rights reserved

Published by John Wiley & Sons, Inc., Hoboken, New Jersey

Published simultaneously in Canada

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning, or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 646-8600, or on the web at www.copyright.com. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at www.wiley.com/go/permissions.

Limit of Liability/Disclaimer of Warranty: While the Publisher and the author have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The advice and strategies contained herein may not be suitable for your situation. You should consult with a professional where appropriate. Neither the Publisher nor the author shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

For general information about our other products and services, please contact our Customer Care Department within the United States at (800) 762-2974, outside the United States at (317) 572-3993 or fax (317) 572-4002.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books. For more information about Wiley products, visit our web site at www.wiley.com.

Library of Congress Cataloging-in-Publication Data:

O'Sullivan, David, 1966-

Geographic information analysis / David O'Sullivan, David Unwin. – 2nd ed.
p. cm.

Includes bibliographical references and index.

ISBN 978-0-470-28857-3 (cloth)

1. Geographic information systems. 2. Spatial analysis (Statistics)

I. Unwin, D. (David John) II. Title.

G70.212.O88 2010

910.285-dc22

2009042594

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

David O'Sullivan: for Mum and Dad
David Unwin: for Polly, winner RHS Gold Medal Chelsea
Flower Show 2009

Contents

Preface to the Second Edition	xi
Acknowledgments	xv
Preface to the First Edition	xvii
1 Geographic Information Analysis and Spatial Data	1
Chapter Objectives	1
1.1 Introduction	2
1.2 Spatial Data Types	5
1.3 Some Complications	10
1.4 Scales for Attribute Description	18
1.5 GIS and Spatial Data Manipulation	24
1.6 The Road Ahead	28
Chapter Review	29
References	29
2 The Pitfalls and Potential of Spatial Data	33
Chapter Objectives	33
2.1 Introduction	33
2.2 The Bad News: The Pitfalls of Spatial Data	34
2.3 The Good News: The Potential of Spatial Data	41
Chapter Review	52
References	53
3 Fundamentals—Mapping It Out	55
Chapter Objectives	55
3.1 Introduction: The Cartographic Tradition	56
3.2 Geovisualization and Analysis	58
3.3 The Graphic Variables of Jacques Bertin	60
3.4 New Graphic Variables	63
3.5 Issues in Geovisualization	65
3.6 Mapping and Exploring Points	66
3.7 Mapping and Exploring Areas	72
3.8 Mapping and Exploring Fields	80

3.9	The Spatialization of Nonspatial Data	84
3.10	Conclusion	86
	Chapter Review	87
	References	87
4	Fundamentals—Maps as Outcomes of Processes	93
	Chapter Objectives	93
4.1	Introduction: Maps and Processes	94
4.2	Processes and the Patterns They Make	95
4.3	Predicting the Pattern Generated by a Process	100
4.4	More Definitions	106
4.5	Stochastic Processes in Lines, Areas, and Fields	108
4.6	Conclusions	116
	Chapter Review	117
	References	118
5	Point Pattern Analysis	121
	Chapter Objectives	121
5.1	Introduction	121
5.2	Describing a Point Pattern	123
5.3	Assessing Point Patterns Statistically	139
5.4	Monte Carlo Testing	148
5.5	Conclusions	152
	Chapter Review	154
	References	154
6	Practical Point Pattern Analysis	157
	Chapter Objectives	157
6.1	Introduction: Problems of Spatial Statistical Analysis	158
6.2	Alternatives to Classical Statistical Inference	161
6.3	Alternatives to IRP/CSR	162
6.4	Point Pattern Analysis in the Real World	166
6.5	Dealing with Inhomogeneity	168
6.6	Focused Approaches	172
6.7	Cluster Detection: Scan Statistics	173
6.8	Using Density and Distance: Proximity Polygons	177
6.9	A Note on Distance Matrices and Point Pattern Analysis	180
	Chapter Review	182
	References	183
7	Area Objects and Spatial Autocorrelation	187
	Chapter Objectives	187
7.1	Introduction: Area Objects Revisited	188
7.2	Types of Area Objects	188

7.3	Geometric Properties of Areas	191
7.4	Measuring Spatial Autocorrelation	199
7.5	An Example: Tuberculosis in Auckland, 2001–2006	206
7.6	Other Approaches	210
	Chapter Review	212
	References	213
8	Local Statistics	215
	Chapter Objectives	215
8.1	Introduction: Think Geographically, Measure Locally	216
8.2	Defining the Local: Spatial Structure (Again)	218
8.3	An Example: The Getis-Ord G_i and G_i^* Statistics	219
8.4	Inference with Local Statistics	223
8.5	Other Local Statistics	226
8.6	Conclusions: Seeing the World Locally	234
	Chapter Review	235
	References	236
9	Describing and Analyzing Fields	239
	Chapter Objectives	239
9.1	Introduction: Scalar and Vector Fields Revisited	240
9.2	Modeling and Storing Field Data	243
9.3	Spatial Interpolation	250
9.4	Derived Measures on Surfaces	263
9.5	Map Algebra	270
9.6	Conclusions	273
	Chapter Review	274
	References	275
10	Knowing the Unknowable: The Statistics of Fields	277
	Chapter Objectives	277
10.1	Introduction	278
10.2	Regression on Spatial Coordinates: Trend Surface Analysis	279
10.3	The Square Root Differences Cloud and the (Semi-) Variogram	287
10.4	A Statistical Approach to Interpolation: Kriging	293
10.5	Conclusions	311
	Chapter Review	312
	References	313
11	Putting Maps Together—Map Overlay	315
	Chapter Objectives	315
11.1	Introduction	316

11.2	Boolean Map Overlay and Sieve Mapping	319
11.3	A General Model for Alternatives to Boolean Overlay	326
11.4	Indexed Overlay and Weighted Linear Combination	328
11.5	Weights of Evidence	331
11.6	Model-Driven Overlay Using Regression	334
11.7	Conclusions	336
	Chapter Review	337
	References	337
12	New Approaches to Spatial Analysis	341
	Chapter Objectives	341
12.1	The Changing Technological Environment	342
12.2	The Changing Scientific Environment	345
12.3	Geocomputation	346
12.4	Spatial Models	355
12.5	The Grid and the Cloud: Supercomputing for Dummies	363
12.6	Conclusions: Neogeographic Information Analysis?	365
	Chapter Review	367
	References	368
Appendix A:	Notation, Matrices, and Matrix Mathematics	373
A.1	Introduction	373
A.2	Some Preliminary Notes on Notation	373
A.3	Matrix Basics and Notation	376
A.4	Simple Matrix Mathematics	379
A.5	Solving Simultaneous Equations Using Matrices	384
A.6	Matrices, Vectors, and Geometry	389
A.7	Eigenvectors and Eigenvalues	391
	Reference	393
Index		395

Preface to the Second Edition

The first edition of this text (O’Sullivan and Unwin, 2003) was written in the first two years of the twenty-first century, but in its basic framework it relied on two key ideas that have a long history in geographic information science: cartography and statistics. Perceptive (and aged?) readers will perhaps have noted sections that owe their origins to *Introductory Spatial Analysis* (Unwin, 1981), a little book one of us wrote almost 30 years ago. The first key idea was the use of a framework for describing geographic objects by their dimension of length into points, lines, areas, and continuous surfaces (fields); the second was to regard mapped distributions as realizations of some spatial stochastic process.

Like any overarching framework, it’s not perfect. For example, heavy reliance on fixed geometric entities and on close attention to statistical hypothesis testing may seem dated in the light of developments in spatial representation and statistical inference. In developing this second edition, we thought for some time about moving with the times and adopting one of a variety of alternative frameworks. Our eventual decision to stick to the original blueprint was not taken lightly, but we are sure we have done the right thing. Since the first edition appeared, we have taught classes with curricula built on this framework at senior undergraduate and beginning graduate levels in the United States, the United Kingdom, New Zealand, and globally over the Internet, and we have found it to be pedagogically clear and resilient. For many students, either spatial data or statistical reasoning (and, not infrequently, both) are new or fairly new concepts, and it is important that we provide a “way in” to more advanced topics for that large segment of readers. For those few readers happily familiar with both topics, we hope that the book is broad enough in its coverage and makes enough nods in the direction of more advanced material to remain useful.

CHANGES

In spite of broad continuities, a chapter-by-chapter comparison will show substantial updates in our treatment of point pattern analysis, spatial

autocorrelation, kriging, and regression with spatial data, and we have made a number of larger changes, some of which are of emphasis and some of which are more substantial. Those familiar with the first edition will notice that formal hypothesis testing has receded further into the background in favor of greater emphasis on Monte Carlo/randomization approaches that, in most practical work, using the usually messy data that we have to handle, seem to us to offer ways around many of the well-known problems related to spatial data. Even in the seven years since the first edition was published, computing power has made this approach more practical and easier to implement. Second, readers will find that throughout the revised text there is a greater emphasis on essentially *local* descriptions. We believe this also reflects an important methodological change within the science, arguably made possible by access to today's computing environments.

The two most substantial changes follow from and relate back to these changes of emphasis, in that we have added entirely new chapters (Chapters 3 and 8) on *geovisualization* and *local statistics*. In fact, a chapter on maps and mapping was written for the first edition but was not included. This we justified to ourselves by noting the need to keep the length of the book down and by considering that most of our readers would be familiar with some of the central concepts of the art and science of cartography. Subsequent experiences teaching courses on geographic information analysis to what one of us has called "accidental geographers" (Unwin, 2005) have shown that this omission was a mistake. By *accidental geographer*, we mean those new to the analysis of spatial data, whose understanding of geographic science is based largely on the operations made possible by geographic information system (GIS) software. Cartography, or, if you prefer, geovisualization, has added relevance for three reasons. First, even with the enormous range of statistical methods that are available, some form of mapping remains perhaps the major analytical strategy used. Second, an emphasis on local statistics that are then mapped has increased the need for understanding basic cartographic principles. Third, as a walk around almost any GIS trade exhibition will show, otherwise sophisticated GIS users continue to make quite basic cartographic errors. Our new Chapter 3 bears little resemblance to the one originally drafted. The new materials rely heavily on the use of an Internet search engine to find and critique examples, something the senior author was taught almost half a century ago in a student class on map appreciation. That said, we have also tried to locate much of the chapter in the long, and regrettably sometimes neglected, cartographic tradition.

The second major addition is a chapter on local statistics. Again, this is not without its organizational problems, since, as we have discovered, a considerable proportion of materials originally developed in different contexts can plausibly be brought into this framework. Examples include all the materials

associated with concepts of distance, adjacency, and neighborhood that go into the definition of geographic structure (**W**) matrices; estimation of the mean height of a field from control point data (spatial *interpolation*); identification of local peaks in the estimated intensity of a point process (*clustering*); and the identification of groups of similar zones by decomposition of a global Moran's *I* measure of spatial autocorrelation (*Moran scatterplot*). Readers will doubtless find other examples, a sign of the centrality of the concept in much spatial analysis. This chapter provides a more explicit treatment of various local indicators of spatial association and allows us to include an introduction to the ideas behind geographically weighted regression (GWR). Although kernel density estimation (KDE) might easily be placed in this same chapter, we believe that it is most often used in a geovisualization context, and we have moved it to Chapter 3 from its original home with materials on point pattern analysis in Chapter 5. We recognize that these changes make it necessary for the reader from time to time to refer back to previous materials, and we have attempted to signal when this is wise by use of boxed thought exercises.

These additions have been balanced by the removal of some materials. First, for entirely pragmatic reasons, we have removed almost all the text on the analysis of line objects. Although it dealt with some of the basic ideas, neither of us was happy with the original chapter, which for reasons of length did not, and could not, reflect the increasing importance of network analysis in almost every branch of science. As readers of that chapter would have recognized, when dealing with linear objects we struggled to maintain our basic stochastic process approach. Somebody, somewhere, someday will write what is necessary—a major text book on geographic information analysis in a network representation of geography—but the task is well beyond what can be covered in a single chapter of the present book. A chapter on multivariate statistics, which sat a little uncomfortably in the first edition on the pretext of treating *n*-dimensional data as spatial, has been omitted. We have retained some of that material in the new chapter on geovisualization under the heading of “spatialization.” In addition to these larger-scale adjustments, we have removed the extended treatment of the joins count approach to characterizing spatial autocorrelation, which, although pedagogically useful, seemed increasingly irrelevant to contemporary practice. Finally, in the interests of keeping the size of the book manageable, we have dropped an appendix introducing basic statistical concepts, assuming that readers can work from one of the many fine introductory text books available.

Since the publication of the first edition, much has changed and the general field has grown enormously, with developments in computing, statistics, and geographic information science. In updating the materials, we have tried as best we can to reflect this new work and the increasingly

“location-aware” scientific and social environment in which it is placed, but we are aware of numerous things that we have omitted. If you look for something and are disappointed, we can only apologize.

SOFTWARE

One major change that we have tried to reflect is the increasing gap between methods used by academic spatial analysts and the functionality embedded in most commercial GIS. It is true that, if you know what you are doing and don't always rely on default settings, many of the methods we describe can be used within such a system, but such use is not ideal. Over the past decade, it has become increasingly obvious that most of today's leading researchers have developed their work in the public domain R programming environment (see Ihaka and Gentleman, 1996). Readers looking to implement the methods we describe should note that almost all of them, and many more, have been implemented in this environment (Baddeley and Turner, 2005; Bivand et al., 2008). Readers wishing to develop new and innovative approaches to geographic information analysis would be well advised to join this community of scholars.

David O'Sullivan
University of Auckland
Te Whare Wananga o Tamaki Makaurau

David Unwin
Birbeck, University of London

Matariki 2009

REFERENCES

- Baddeley, A. and Turner, R. (2005) Spatstat: an R package for analyzing spatial point patterns. *Journal of Statistical Software*, 12: 1–42.
- Bivand, R. S., Pebesma, E. J., and Gomez-Rubio, V. (2008) *Applied Spatial Data Analysis with R* (New York: Springer).
- Ihaka, R. and Gentleman, R. (1996). R: a language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, 5: 299–314.
- O'Sullivan, D. and Unwin, D. J. (2003) *Geographic Information Analysis* (Hoboken, NJ: Wiley).
- Unwin, D. J. (1981) *Introductory Spatial Analysis* (London: Methuen).
- Unwin, D. J. (2005) Fiddling on a different planet. *Geoforum*, 36: 681–684.

Acknowledgments

DAVID O'SULLIVAN

Students and colleagues in courses at Penn State (GEOG 454, 455, 586) and Auckland (GEOG 318, 771) have helped to refine materials in this edition in many ways, large and small. I am grateful for the supportive input (in a variety of forms) of David Dibiase, Mark Gahegan, Frank Hardisty, Jim Harper, and George Perry. I also acknowledge a University of Auckland Study Leave award and the Spatial Literacy in Teaching (SpLinT) consortium (Universities of Leicester and Nottingham, and University College London), particularly Paul Longley and Nick Tate, for assisting with a Fellowship award in 2008. Study Leave and the SpLinT Fellowship enabled Dave and I to work together on the book for a pleasant and productive couple of days and helped to shape this edition.

I also thank Gill for her constant forbearance and support. Finally, the Saturday morning exploits of Fintan and Malachy, on cricket and football pitches, not to mention farflung virtual LEGO™ galaxies, have reminded me throughout of where the real center of things lies.

DAVE UNWIN

The Washington “Space cadets” fired my graduate student enthusiasm for spatial analysis, and it has been a pleasure and a privilege eventually to meet several of them. The Departments of Geography at Aberystwyth, Leicester, and Birkbeck London have provided supportive bases and knowledgeable colleagues. Particular thanks for numerous discussions go to my former colleagues at the Leicester Midlands Regional Research Laboratory: Alan Strachan, Mike Worboys, David Maguire, Jo Wood, Jason Dykes, and, more recently, Pete Fisher. All teachers should learn from their students, and it is a pleasure to acknowledge the contributions to my thought process made, sometimes unwittingly, sometimes directly, by graduate classes in the Universities of Leicester, Birkbeck/University College London, Waikato (New Zealand), Canterbury (New Zealand), and Redlands (the United

States). A final and particularly formative experience has been that of teaching and tutoring classes of “accidental geographers” over the Internet, which has confirmed my convictions about the value of using geography in science. Like my colleague, I am grateful to the SpLinT Initiative for its 2008 support.

The errors that remain, are, of course, our own.

David O’Sullivan
Tamaki-makau-rau, Aotearoa

Dave Unwin
Maidwell, England

Preface to the First Edition

Like Topsy, this book “jes growed” out of a little book one of us wrote in the period from 1979 to 1981 (*Introductory Spatial Analysis*, London: Methuen). Although that was fully a decade after the appearance of the first commercial geographical information systems (GIS) and more or less coincided with the advent of the first microcomputers, that book’s heritage was deep in the quantitative geography of the 1960s, and the methods discussed used nothing more sophisticated than a hand calculator. Attempts to produce a second edition from 1983 onward were waylaid by other projects—almost invariably projects related to the contemporary rapid developments in GISs. At the same time, computers became available to almost everyone in the developed world, and in research and commerce many people discovered the potential of the geography they could do with GIS software. By the late 1990s, it was apparent that only a completely new text would do, and it was at this point that the two of us embarked on the joint project that resulted in the present book.

The materials we have included have evolved over a long period of time and have been tried and tested in senior undergraduate and postgraduate courses we have taught at universities in Leicester, London (Birkbeck and University Colleges), Pennsylvania (Penn State), Waikato, Canterbury (New Zealand), and elsewhere. We are passionate about the usefulness of the concepts and techniques we present in almost any work with geographic data and we can only hope that we have managed to communicate this to our readers. We also hope that reservations expressed throughout concerning the overzealous or simpleminded application of these ideas do not undermine our essential enthusiasm. Many of our reservations arise from a single source, namely the limitations of digital representations of external reality possible with current (and perhaps future?) technology. We feel that if GIS is to be used effectively as a tool supportive of numerous approaches to geography and is not to be presented as a one-size-fits-all “answer” to every geographical question, it is appropriate to reveal such uncertainty, even to newcomers to the field.

Although it was not planned this way, on reflection we progress from carefully developed basics spelled out in full and very much grounded in the

intellectual tradition of *Introductory Spatial Analysis*, to more discursive accounts of more recent computationally intensive procedures. The early material emphasizes the importance of fundamental concepts and problems common to any attempt to apply statistical methods to spatial data and should provide a firm grounding for further study of the more advanced approaches discussed in more general terms in later chapters. The vintage of some of the references we provide is indicative of the fact that at least some of the intellectual roots of what is now called *geographical information science* are firmly embedded in the geography of the 1960s and range far and wide across the concerns of a variety of disciplines. Recent years have seen massive technical innovations in the analysis of geographical data, and we hope that we have been able in the text and in the suggested reading to capture some of the excitement this creates.

Two issues that we have struggled with throughout are the use of mathematics and notation. These are linked, and care is required with both. Mathematically, we have tried to be as rigorous as possible, consistent with our intended audience. Experience suggests that students who find their way to GIS analysis and wish to explore some aspects in more depth come from a wide range of backgrounds with an extraordinary variety of prior experience of mathematics. As a result, our “rigor” is often a matter merely of adopting formal notations. With the exception of a single “it can be shown” in Chapter 9, we have managed to avoid use of the calculus, but matrix and vector notation is used throughout and beyond Chapter 5 is more or less essential to a complete understanding of everything that is going on. Appendix B provides a guide to matrices and vectors that should be sufficient for most readers. If this book is used as a course text, we strongly recommend that instructors take time to cover the contents of this appendix at appropriate points prior to the introduction of the relevant materials in the main text. We assume that readers have a basic grounding in statistics, but to be on the safe side we have included a similar appendix outlining the major statistical ideas on which we draw, and similar comments apply.

Poor notation has a tremendous potential to confuse, and spatial analysis is a field blessed (perhaps cursed) by an array of variables. Absolute consistency is hard to maintain and is probably an overrated virtue in any case. We have tried hard to be as consistent and explicit as possible throughout. Perhaps the most jarring moment in this respect is the introduction in Chapters 8 and 9 of a third locational coordinate, denoted by the letter z . This leads to some awkwardness and a slight notational shift when we deal with regression on spatial coordinates in Section 9.3. On balance we prefer to use (x,y,z) and put up with accusations of inconsistency than to have too many pages bristling with subscripts (a flip through the pages should reassure the less easily intimidated that many subscripts remain). This

pragmatic approach should serve as a reminder that, like its predecessor, this book is about the practical analysis of geographic information rather than being a treatise on spatial statistics. First and foremost, this is a geography book!

No book of this length covering so much ground could ever be the unaided work of just two people. Over many years one of us has benefited from contacts with colleagues in education and the geographic information industry far too numerous to mention specifically. To all he is grateful for advice, for discussion, and for good-natured argument. It is a testament to the open and constructive atmosphere in this rapidly developing field that the younger half of this partnership has already benefited from numerous similar contacts, which are also difficult to enumerate individually. Suffice it to say that supportive environments in University College London's Centre for Advanced Spatial Analysis and in the Penn State Geography Department have helped enormously. As usual, the mistakes that remain are our own.

David O'Sullivan
The Pennsylvania State University
(St Kieran's Day, 2002)

Dave Unwin
London, England
(St Valentines' Day, 2002)

GEOGRAPHIC INFORMATION ANALYSIS

Chapter 1

Geographic Information Analysis and Spatial Data

CHAPTER OBJECTIVES

In this first chapter, we:

- Define *geographic information analysis* as it is meant in this book
- Distinguish geographic information analysis from *GIS-based spatial data manipulation* while relating the two
- Review the *entity-attribute model* of spatial data as consisting of *points, lines, areas, and fields*, with associated *nominal, ordinal, interval, or ratio* data
- Note some of the complications in this view, especially *multiple representation* at different scales, *time*, objects with *uncertain boundaries*, objects that are *fuzzy*, and objects that may be *fractal*
- Review *spatial data manipulation operations* and emphasize their importance
- Examine the various *transformations* between representations, noting their utility for geographic information analysis

After reading this chapter, you should be able to:

- List four different approaches to spatial analysis and differentiate between them
- Give reasons why modern methods of spatial analysis are not well represented in the tool kits provided by the typical GIS
- Distinguish between spatial objects and spatial fields and discuss why the vector-versus-raster debate in GIS is really about how we choose to represent these entity types

- Differentiate between point, line, and area objects and give examples of each
- List the fundamental data properties that characterize a field
- Provide examples of real-world entities that do not fit easily into this scheme
- Maintain a clear distinction between a real-world entity, its representation in a digital database, and its display on a map
- Differentiate between nominal, ordinal, interval, and ratio attribute data and give examples of each
- Give examples of at least 12 resulting types of spatial data
- List some of the basic geometrical data manipulations available in the typical GIS
- Outline methods by which the representations of entities can be transformed and explain why this is useful for geographic information analysis

1.1. INTRODUCTION

Geographic information analysis is not an established discipline. In fact, it is a rather new concept. To define what we mean by this term, it is necessary first to define a much older term—*spatial analysis*—and then to describe how we see the relationship between the two. Of course, a succinct definition of spatial analysis is not straightforward either. The term comes up in various contexts. At least four broad areas are identifiable in the literature, each using the term in different ways:

1. *Spatial data manipulation*, usually in a geographic information system (GIS), is often referred to as *spatial analysis*, particularly in GIS companies' promotional material. Your GIS manuals will give you a good sense of the scope of these techniques, as will the texts by Tomlin (1990) and Mitchell (1999).
2. *Spatial data analysis* is descriptive and exploratory. These are important first steps in all spatial analysis, and often are all that can be done with very large and complex data sets. Books by geographers such as Unwin (1982), Bailey and Gatrell (1995), and Fotheringham et al. (1999) are very much in this tradition.
3. *Spatial statistical analysis* employs statistical methods to interrogate spatial data to determine whether or not the data can be represented by a statistical model. The geography texts cited above touch on these issues, and there are a small number of texts by statisticians interested in the analysis of spatial data, notably those by Ripley (1981, 1988), Diggle (1983), and Cressie (1991).

4. *Spatial modeling* involves constructing models to predict spatial outcomes. In human geography, models are used to predict flows of people and goods between places or to optimize the location of facilities (Wilson, 2000), whereas in environmental science, models may attempt to simulate the dynamics of natural processes (Ford, 1999). Modeling techniques are a natural extension of spatial analysis but are beyond the scope of this book.

In practice, it is often difficult to distinguish between these approaches, and most serious research will involve all four. First, data are collected, visualized, and described. Then exploratory techniques might raise questions and suggest theories about the phenomena of interest. These theories are then subjected to statistical testing using spatial statistical techniques. Theories of what is going on might then be the basis for computer models of the phenomena, and their results, in turn, may be subjected to more statistical investigation and analysis.

It is impossible to consider geographic information without considering the technology that is increasingly its home: geographical information systems (GISs). Although GISs are not ubiquitous in the way that (say) word processors are, they have infiltrated more and more businesses, government agencies, and other decision-making organizations. Even if this is the first time you've read a geography textbook, chances are that you will have already used a GIS without knowing it, perhaps when you used a website to generate a map of a holiday destination or to find driving directions to get you there.

In the above list, current GISs typically include item 1 as standard (since a GIS without these functions would be just a plain old IS!) and have some simple data analysis capabilities, especially exploratory analysis using maps (item 2). GISs have recently begun to incorporate some of the statistical methods of item 3 and only rarely include the capability to build spatial models and determine their likely outcomes (item 4). In fact, it can be hard to extend GIS to perform such analysis, which is why many geographic information analysts use other software environments for work that would be classified as belonging to items 3 and 4. In this book, we focus mostly on items 2 and 3. In practice, you will find that, in spite of rapid advances in the available tools, statistical testing of spatial data remains relatively rare. Statistical methods are well worked out and understood for some types of data but less so for many others. As this book unfolds, you should begin to understand why this is so.

If spatial analysis is so necessary—even worth writing a book about—then why isn't it a standard part of the GIS toolkit? We suggest a number of reasons, among them the following:

- *The GIS view of spatial data and that of spatial analysis are different.* The spatial analysis view of spatial data is more concerned with *processes* and *patterns* than it is with database management and manipulation, whereas the basic requirement for a *spatial database* is far more important to most large GIS buyers (government agencies, utilities) than the ability to perform complex and (sometimes) obscure spatial analysis.
- *Spatial analysis is not widely understood.* Spatial analysis is not obvious or especially easy, although we aim to address that issue in this book. The apparent difficulty means that it is difficult to convince software vendors to include spatial analysis tools as standard products. Spatial analysis tools are a possible addition to GIS that is frequently left out. This rationale has become less significant in recent years as software engineering methods enable GIS vendors to supply “extensions” that can be sold separately to those users who want them. At the same time, third-party vendors can supply add-on components more easily than previously, and open source software has become an increasingly important alternative in some quarters.
- *The spatial analysis perspective can sometimes obscure the advantages of GIS.* By applying spatial analysis techniques, we often raise awkward questions: “It looks like there’s a pattern, but is it significant? Maybe not.” This is a hard capability to sell!

Despite this focus, don’t underestimate the importance of the *spatial data manipulation* functions provided by GIS such as buffering, point-in-polygon queries, and so on. These are essential precursors to generating questions and formulating hypotheses. To reinforce their importance, we review these topics in Section 1.5 and consider how they might benefit from a more statistical approach. More generally, the way spatial data are stored—or *how geographical phenomena are represented in GIS*—is becoming increasingly important for analysis. We therefore spend some time on this issue in Sections 1.2 and 1.3.

For all of these reasons, we use the broader term *geographic information analysis* for the material we cover. A working definition of this term is that it is concerned with investigating the *patterns* that arise as a result of *processes* that may be operating in space. Techniques and methods to enable the representation, description, measurement, comparison, and generation of spatial patterns are central to the study of geographic information analysis. Of course, at this point our definition isn’t very useful, since it raises the question of what we mean by *pattern* and *process*. For now, we will accept whatever intuitive notion you have about the meaning of the key terms. As we work through the concepts of point pattern analysis in Chapters 4 and 5, it

will become clearer what is meant by both terms. For now, we will concentrate on the general spatial data types you can expect to encounter.

1.2. SPATIAL DATA TYPES

Thought Exercise: Representation

Throughout this book, you will find thought exercises to help you follow the text in a more hands-on way. Usually, we ask you to do something and use the results to draw some conclusions. You should find that these exercises help you remember what we've said. This first exercise is concerned with how we represent geography in a digital computer:

1. Assume that you are working for a road maintenance agency. Your responsibilities extend to the roads over a county-sized area. Your GIS is required to support operations such as surface renewal, avoiding clashes with other agencies—utility companies, for example—that also dig holes in the roads and make improvements to the road structure.

Think about and write down how you would record the geometry of the network of roads in your database. What road attributes would you collect?

2. Imagine that you are working for a bus company in the same area. Now the GIS must support operations such as time-tabling, predicting the demand for existing and potential new bus routes, and optimizing where stops are placed.

How would the recording of the geometry of the road network and its attributes differ from your suggestions in step 1 above?

What simple conclusion can we draw from this? It should be clear that how we represent the same geographic entities differs according to the purpose of the representation. This is obvious, but it can easily be forgotten.

Quite apart from the technical issues involved, social critiques of geographic information analysis often hinge on the fact that analysis frequently confines itself to those aspects of the world that can be easily represented digitally (see Fisher and Unwin, 2005).

When you think of the world in map form, how do you view it? In the early GIS literature, a distinction was often made between two kinds of system characterized by how the geography is represented digitally:

1. One type of system provides a *vector* view, which records locational (x, y) coordinates of the features that make up a map. In the vector view, we list features and represent each as a point, line, or area *object*. Vector GIS originated in the use of computers to draw maps based on digital data and were particularly valued when computer memory was an expensive commodity. Although the fit is inexact, the vector model is closest to an *object* view of the world, where space is thought of as an empty container occupied by different sorts of objects.
2. Contrasted with vector systems are *raster* systems. Instead of starting with objects on the ground, a grid of small units, called *pixels*, of the Earth's surface is defined. For each pixel, the value, or presence or absence of something of interest, is then recorded. Thus, we divide a map into a set of identical, discrete elements and list the contents of each. Because every location in space has a value (even if it is zero or null), a raster approach generally uses more computer memory than a vector one. Raster GIS originated mostly in image processing, where data from remote sensing platforms are often encountered.

In this section, we hope to convince you that at a higher level of abstraction the vector/raster distinction isn't very useful, and that it obscures a more important division between what we call an *object* and a *field* view of the world.

The Object View

In the object view, we consider the world as a series of *entities* located in space. Entities are (usually) real: you can touch them, stand in them, perhaps even move them around. An *object* is a digital representation of all or part of an entity. Objects may be classified into different object types—for example, *point objects*, *line objects*, and *area objects*—and in specific applications, these types are *instantiated* by specific objects. For example, in an environmental GIS, woods and fields might be instances of area objects. In the object view of the world, places can be occupied by any number of objects. A house can exist in a census tract, which may also contain lampposts, bus stops, road segments, parks, and so on.

Because it is also possible to associate *behavior* with objects, the object view has advantages when well-defined objects change over time—for example, the changing data for a census area object over a series of population censuses. Note that we have said nothing about *object orientation* in the