**Spatial Data Analysis**
Theory and Practice

*Spatial Data Analysis: Theory and Practice* provides a broad-ranging treatment of the field of spatial data analysis. It begins with an overview of spatial data analysis and the importance of location (place, context and space) in scientific and policy-related research. Covering fundamental problems concerning how attributes in geographical space are represented to the latest methods of exploratory spatial data analysis and spatial modelling, it is designed to take the reader through the key areas that underpin the analysis of spatial data, providing a platform from which to view and critically appreciate many of the key areas of the field. Parts of the text are accessible to undergraduate and master's level students, but it also contains sufficient challenging material that it will be of interest to geographers, social scientists and economists, environmental scientists and statisticians, whose research takes them into the area of spatial analysis.

ROBERT HAINING is Professor of Human Geography at the University of Cambridge. He has published extensively in the field of spatial data analysis, with particular reference to applications in the areas of economic geography, medical geography and the geography of crime. His previous book, *Spatial Data Analysis in the Social and Environmental Sciences* (Cambridge University Press, 1993) was well received and cited internationally.

# Spatial Data Analysis

## Theory and Practice

ROBERT HAINING
University of Cambridge

CAMBRIDGE UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK
40 West 20th Street, New York, NY 10011-4211, USA
477 Williamstown Road, Port Melbourne, VIC 3207, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain
Dock House, The Waterfront, Cape Town 8001, South Africa

http://www.cambridge.org



First published 2003

Printed in the United Kingdom at the University Press, Cambridge

*Typeface* Lexicon No. 2 10/14 pt and Lexicon No. 1 *System* LaTeX 2ε [TB]

*A catalogue record for this book is available from the British Library*

*Library of Congress Cataloguing in Publication data*

Haining, Robert P.
Spatial data analysis: theory and practice / Robert Haining.
p. cm.
Includes bibliographical references and index.
ISBN 0-521-77319-9 – ISBN 0-521-77437-3 (pbk.)
1. Spatial analysis (Statistics) 2. Geology – Statistical methods – Data processing. I. Title.
QA278.2 .H345 2003
001.4′22–dc 2002031242

ISBN 0 521 77437 3 paperback

To my wife, Rachel, and our children,
Celia, Sarah and Mark

# *Contents*

# *Preface*

Interest in analysing spatial data has grown considerably in the scientific research community. This reflects the existence of well-formulated questions or hypothesis in which location plays a role, of spatial data of sufficient quality, of appropriate statistical methodology.

In writing this book I have drawn on a number of scientific and also policy-related fields to illustrate the scale of interest – actual and potential – in analysing spatial data. In seeking to provide this overview of the field I have given a prominent place to two fields of research: Geographic Information Science (GISc) and applied spatial statistics.

It is important as part of the process of understanding the results of spatial data analysis to define the relationship between geographic reality and how that reality is captured in a digital database in the form of a data matrix containing both attribute data and data on locations. The usefulness of operations on that data matrix – revising or improving an initial representation (e.g. spatial smoothing), testing hypotheses (e.g. does this map pattern contain spatial clusters of events?) or fitting models (e.g. to explain offence patterns or health outcomes in terms of socio-economic covariates) – will depend on how well the reality that is being represented has been captured in the data matrix. Awareness of this link is important and insights can be drawn from the GISc literature.

I have drawn on developments in spatial statistics which can be applied to data collected from continuous surfaces and from regions partitioned into sub-areas (e.g. a city divided into wards or enumeration districts). In covering this material I have attempted to draw out the important ideas whilst directing the reader to specialist sources and original papers. This book is not an exhaustive treatment of all areas of spatial statistics (it does not cover point processes), nor of all areas of spatial analysis (it does not include cartographic modelling).

Implementing a programme of spatial data analysis is greatly assisted if supporting software is available. Geographic information systems (GIS) software are now widely used to handle spatial data and there is a growing quantity of software some of it linked to GIS for implementing spatial statistical methods. The appendix directs the reader to some relevant software.

### Readership

This book brings together techniques and models for analysing spatial data in a way that I hope is accessible to a wide readership, whilst still being of interest to the research community.

Parts of this book have been tried out on year 2 geography undergraduates at the University of Cambridge in an eight-hour lecture course that introduced them to certain areas of geographic information science and methods of spatial analysis. The parts used are chapters 1, 2, sections 3.1, 3.2.1, 3.2.3, 3.2.4(a) from chapter 3, selected sections from chapter 4 (e.g. detecting errors and outliers, areal interpolation problems), selected sections from chapter 7 (section 7.1.3, map smoothing) and some selected examples on modelling and mapping output using the normal linear regression model. In associated practicals simple methods for hot spot detection are applied (the first part of section 7.3.1(a)) together with logistic regression for modelling (along the lines of section 11.2.2(a)).

Parts of the book have been tried out on postgraduate students on a one year M.Phil. in Geographic Information Systems and Remote Sensing at Cambridge. One 16-hour course was on general methods of spatial analysis but particularly for data from continuous surfaces. In addition to some of the foundation material covered in chapters 1 to 4 there was an extended treatment of the material in section 4.4.2 with particular reference to kriging with Gaussian data (including estimation and modelling of the semi-variogram taken from chapter 10 and the references therein). A second 16-hour course dealt with exploratory spatial data analysis and spatial modelling with reference to the analysis of crime and health data. This focused on area data. The material in chapter 7 was included with an introduction provided by the conceptual frameworks described in chapter 5. The part of the course on modelling took selected material from chapter 9 and drew on examples referred to in that chapter and chapter 11.

# *Acknowledgements*

This book has taken shape over the last two years at the University of Cambridge but has its roots in teaching and research that go back over many years most significantly to my time at the University of Sheffield. In one sense at least the book dates back to the early 1970s and a one-off lecture given by Michael Dacey at Northwestern University on spatial autocorrelation. That lecture was my introduction to the problems of analysing spatial data. Michael Goodchild invited me to spend some time at the NCGIA in Santa Barbara in the later 1980s and this too proved very formative.

I am grateful to friends and colleagues over the years with whom I have worked. The University of Sheffield had the foresight in the mid 1990s to invest in a research centre – the Sheffield Centre for Geographic Information and Spatial Analysis. This opened up opportunities for me to work on a range of different problems both theoretical and applied and fostered numerous collaborations both within the University and with local agencies. I would like to thank in particular Max Craglia, Ian Masser and Steve Wise in working with me to establish SCGISA and with whom I have undertaken many projects and had many interesting discussions.

I have had the benefit of working with many excellent researchers and in particular I would like to acknowledge Judith Bush, Paul Brindley, Vania Ceccato, Sue Collins, Andrew Costello, Young-Hoon Kim, Jingsheng Ma, Xiaoming Ning, Paola Signoretta and Dawn Thompson. At Cambridge I am working with Jane Law and together we are learning to apply Bayesian methodology to crime and health data, using the WinBUGS program. The examples in chapters 10 and 11 owe a great deal to her hard work. Jane and I are also working to encourage interest in these methods in agencies in the Cambridge region.

Sections on error propagation, missing-data estimation and spatial sampling have benefited from research collaborations with Giuseppe Arbia, Bob

Bennett, Dan Griffith, Luis Flores and Jinfeng Wang. Some of the visualization material and exploratory data analysis has benefited from two ESRC projects undertaken with Steve Wise. I have worked with Max Craglia on several projects with strong policy dimensions, notably two recent Home Office projects. Some of the material on spatial modelling has benefited from collaboration with Eric Sheppard and Paul Plummer on modelling price variation in interdependent markets. My interest in applications of spatial analysis methods to problems in the areas of health studies have been stimulated by projects with Marcus Blake, Judith Bush and Dawn Thompson; also with David Hall and Ravi Maheswaran at Sheffield and recently with Andy Cliff with whom I have done some work on American measles data. Martin Kulldorff has kindly given me advice on the use of his scan test. My more recent interest in the application of spatial analysis methods to data in criminology, as well as drawing my attention to relevant literature, owe much to advice from Tony Bottoms, Andrew Costello and Paul Wiles. Thanks also to the many people who have drawn my attention to a wide range of relevant literature – apologies for not including them all by name.

Parts of this book have been tested on undergraduate and postgraduate students at the University of Cambridge. My thanks to them for sitting through the 'first draft'. My thanks also to three anonymous readers who saw the first part of this book and made many excellent suggestions.

My thanks to Phil Stickler in the Cartography Laboratory at the University of Cambridge for drawing up the figures. Thanks also to the editorial and production guidance of Tracey Sanderson, Carol Miller and Anne Rix at Cambridge University Press.

Thanks to the following for allowing me to use their data in the examples: Dawn Thompson and Sheffield Health for the breast cancer screening data; South Yorkshire Police for several crime data sets, including the burglary and victimized offender data sets; Sheffield children's services unit for the data on children excluded from school; James Reid for the updated ward boundary data for Cambridgeshire; Sara Godward of the Cancer Intelligence Unit for the Cambridgeshire lung cancer data, Andy Cliff for the US measles data.

Finally my thanks to my mother and father who have given me such encouragement over the years. This book is dedicated in particular to Rachel, my wife and 'best friend', for all her support and not least her willingness and enthusiasm to upsticks and try something new and different on occasions too numerous to count.

## Copyright acknowledgements

Some figures in this book display boundary material which is copyright of the Crown, the ED-LINE consortium. Ordnance Survey data supplied by EDINA Digimap (a JISC supplied service) were also used. Some of the figures in this book are reproduced with the kind permission of the original publishers. These are as follows and with full references in the bibliography.

Figure 3.5: Kluwer Academic Publishers. From *Geo-ENV II Geostatistics for Environmental Applications*, edited by J. Gomez-Hernandez, A. Soares and R. Frodevaux (1998) Savelieva et al. Conditional stochastic co-simulations of the Chernobyl fallout, fig. 14, p. 463.

Figure 4.3: Pion Limited, London. From M. Tranmer and D.G. Steel (1998) Investigating the ecological fallacy. *Environment and Planning*, A, **30**, fig. 1, p. 827 and fig. 2, p. 830.

Figure 6.2: Taylor and Francis Limited, London. From M. Craglia, R. Haining and P. Wiles (2000) A comparative evaluation of approaches to urban crime pattern analysis. *Urban Studies* **37**, fig. 5, p. 725.

Figure 6.3: Oxford University Press. From *Journal of Public Health Medicine* (1994) R. Haining, S. Wise and M. Blake. Constructing regions for small area health analysis, fig. 3, p. 433.

Figure 6.4: Kluwer Academic Publishers. From *Mathematical Geology*, 20 (1989) M. Oliver and R. Webster. A geostatistical basis for spatial weighting in multivariate classification, fig. 3, pp. 15–35.

Figure 6.7: Kluwer Academic Publishers. From G. Verly et al. (1984) *Geostatistics for Natural Resources Characterization.* N. Cressie towards resistant geostatistics, fig. 8, p. 33.

Figures 6.8 to 6.12: Springer-Verlag. From *Journal of Geographical Systems*, **2** (2000) R. Haining et al. Providing scientific visualization for spatial data analysis, pp. 121–40.

Figure 7.1: John Wiley and Sons Limited. From *Statistics in Medicine* (1999) K. Kafadar. Simultaneous smoothing and adjusting mortality rates in US counties. Figs. 2(a)–2(d). Thanks also to Dr Kafadar for providing the original digital version of these figures.

Figure 7.3 and 7.7: Routledge. From *GIS and Health*, edited by A. Gatrell and M. Loytonen. M. Kulldorff. Statistical methods for spatial epidemiology, figs. 4.1 and 4.2.

Figure 7.8: John Wiley and Sons Limited. From *Statistics in Medicine* (1988) R. Stone. Investigations of excess environmental risks around putative sources. Fig. 3.

Figures 8.1 and 8.2: Ohio State University Press, Columbus, Ohio. From *Geographical Analysis* (1991) R. Haining. Bivarate correlation with spatial data Figs. 2, 3 and 5

Figure 8.3: International Biometric Society. From *Biometrics* (1997). A. Cerioli Modified test of independence in $2 \times 2$ tables with spatial data, Fig. 2, pp. 619–28.

Figure 10.1: Ohio State University Press, Columbus, Ohio. From *Geographical Analysis* (1994) D. Griffith et al. Heterogeneity of attribute sampling error in spatial data sets, Fig. 1, p. 31a.

Figure 11.2: Taylor and Francis Limited, London. From *Urban Studies* (2001) M. Craglia et al. Modelling high intensity crime areas in English cities p. 1931.