Qualitative Geography

Perspectives on Spatial Data Analysis

A Stewart Fotheringham,
Chris Brunsdon & Martin Chalrton

Quantitative
Geography

QUANTITATIVE GEOGRAPHY

Perspectives on Spatial Data Analysis

- A. Stewart Fotheringham
- Chris Brunsdon
- Martin Charlton

SAGE Publications
Los Angeles • London • New Delhi • Singapore

First published 2000 , Reprinted 2002, 2004, 2005, 2007

SAGE Publications Ltd
1 Oliver's Yard, 55 City Road,
London EC1Y 1SP

SAGE Publications Inc
2455 Teller Road
Thousand Oaks, California 91320

SAGE Publications India Pvt Ltd
B1/I1 Mohan Cooperative Industrial Area
Mathura Road, New Delhi 110 044
India

SAGE Publications Asia-Pacific Pte Ltd
33 Pekin Street #02-01
Far East Square
Singapore 048763

British Library Cataloguing in Publication data

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

ISBN 978-0-7619-5947-2
ISBN 978-0-7619-5948-9 (pbk)

Library of Congress catalog record available

Typeset by Keytec Typesetting Ltd., Bridport, Dorset
Printed in Great Britain by The Cromwell Press Ltd, Trowbridge, Wiltshire

To our parents

Contents

Preface xi

1 Establishing the Boundaries 1
1.1 Setting the scene 1
1.2 What is quantitative geography? 4
1.3 Applications of quantitative geography 8
1.4 Recent developments in quantitative geography 10
1.5 Summary 13
Notes 14

2 Spatial Data 15
2.1 Introduction 15
2.2 Spatial data capture 16
2.3 Spatial objects 17
2.4 Location on the globe; location on a plane 18
2.5 Distance 20
2.6 Representing spatial data 21
2.7 Models for spatial data 22
2.7.1 The vector model 22
2.7.2 The raster model 23
2.8 Programming with spatial data 24
2.8.1 Point-in-polygon operations 24
2.8.2 The use of complex numbers to represent spatial data 25
2.9 Problems and opportunities 25
2.10 Summary 29
Notes 29

3 The Role of Geographical Information Systems 30
3.1 Introduction 30
3.2 Simple GIS-based spatial analysis 32
3.2.1 Feature selection by attribute 32
3.2.2 Feature selection by geometric intersection 33
3.2.3 Buffering features 34
3.2.4 Geometric intersection: union 35
3.2.5 Geometric intersection: intersection 37
3.2.6 Proximity 38
3.2.7 Contiguity 41

3.2.8 Interpolation and fields 42
3.2.9 Density functions 45
3.2.10 Analysis on networks 46
3.2.11 Query 49
3.3 Advanced GIS-based spatial analysis 49
3.3.1 Data-integration and management 51
3.3.2 Exploration 54
3.3.3 Post-modelling visualization 55
3.4 Problems 58
3.4.1 Error modelling 58
3.4.2 Cross-area aggregation 59
3.5 Linking higher-order types of spatial analysis to GIS 61
3.6 A role for GIS? 64
Notes 64

4 Exploring Spatial Data Visually 65
4.1 Introduction 65
4.2 Stem and leaf plots 66
4.3 Boxplots 68
4.4 Histograms 70
4.5 Density estimates 71
4.6 Maps 72
4.7 The scatterplot matrix 75
4.8 Linked plots 77
4.9 Parallel coordinate plots 80
4.10 RADVIZ 82
4.11 Projection pursuit 86
4.12 Summary 91
Notes 92

5 Local Analysis 93
5.1 Introduction 93
5.2 The nature of local variations in relationships 94
5.3 Measuring local relationships in univariate data 96
5.3.1 Local point pattern analysis 96
5.3.2 Other local measures of univariate spatial relationships 97
5.4 Measuring local relationships in multivariate data 102
5.4.1 Multilevel modelling 103
5.4.2 The spatial expansion method 106
5.4.3 Geographically weighted regression 107
5.5 An empirical comparison of the spatial expansion method and GWR 114
5.5.1 The data 114
5.5.2 Global regression model results 115
5.5.3 Spatial expansion method results 117
5.5.4 GWR results 118
5.6 Measuring local relationships in spatial interaction models 128
5.7 Summary 129
Notes 129

6 Point Pattern Analysis **130**
6.1 Introduction 130
6.2 Initial exploration 132
6.2.1 Scatterplots 132
6.2.2 Other exploratory plots 134
6.2.3 Non-graphical approaches to point pattern exploration 136
6.3 Modelling point patterns 138
6.4 First-order intensity analysis 144
6.4.1 Kernel density estimates 146
6.5 Second-order intensity analysis 149
6.6 Comparing distributions 154
6.6.1 Comparing kernel densities 155
6.6.2 Comparing *K* functions 157
6.6.3 Comparing a point pattern with a 'population at risk' 159
6.7 Conclusions 160
Notes 161

7 Spatial Regression and Geostatistical Models **162**
7.1 Introduction 162
7.2 Autoregressive models 166
7.2.1 Spatially autoregressive models 167
7.2.2 Spatial moving average models 169
7.3 Kriging 171
7.3.1 The statistical technique 171
7.3.2 A worked example 175
7.3.3 Trend surfaces from kriging residuals 176
7.4 Semi-parametric smoothing approaches 178
7.5 Conclusions 182

8 Statistical Inference for Spatial Data **184**
8.1 Introduction 184
8.2 Informal inference 185
8.2.1 Exploratory data analysis (EDA) and visualization 185
8.2.2 Data mining 187
8.2.2.1 Cluster analysis 188
8.2.2.2 Neural networks 190
8.3 Formal inference 193
8.3.1 Bayesian inference 193
8.3.2 Classical inference 198
8.3.3 Experimental and computational inference 201
8.3.3.1 Classical statistical inference and spatial autocorrelation 201
8.3.3.2 Experimental distributions and spatial autocorrelation 204
8.3.3.3 An empirical comparison of classical and experimental inference 206
8.3.4 Model building and model testing 210
8.4 Conclusions 211
Notes 211

9 Spatial Modelling and the Evolution of *Spatial* Theory 213
9.1 Introduction 213
9.2 Spatial interaction as social physics (1860–1970) 215
9.3 Spatial interaction as statistical mechanics (1970–80) 217
9.4 Spatial interaction as aspatial information processing (1980–90) 222
9.5 Spatial interaction as spatial information processing (1990 onwards) 225
9.6 Summary 234
Notes 235

10 Challenges in Spatial Data Analysis 236
10.1 Retrospect 236
10.2 Current challenges 237
10.2.1 The modifiable areal unit problem 237
10.2.2 Spatial non-stationarity 240
10.2.3 Alternative inferential frameworks (Bayes, MCMC) 242
10.2.4 Geometry 243
10.2.5 Basic spatial concepts: proximity and accessibility 244
10.2.6 Merging space and time 245
10.3 Training people to think spatially 246
10.3.1 Teaching quantitative geography 246
10.3.2 Software 247
10.4 Summary 247

Bibliography 249

Index 267

Preface

One of the more puzzling paradoxes that will face those who come to review the development of geography will be why, at the end of the twentieth century, much of geography turned its back on quantitative spatial data analysis just as many other disciplines came to recognize its importance. At a time when geography should have been meeting the rapidly growing demand for spatial data analysts, the majority of its graduates were, at best, non-quantitative and, in quite a few cases, were actively anti-quantitative.

A commonly expressed reason for the negative attitude by many geographers towards one of the discipline's basic elements is a disillusionment on their part with the positivist philosophical underpinnings of much of the early work in quantitative geography. Another, less frequently stated reason, is that spatial data analysis and spatial modelling are perceived to be relatively difficult, not only by students, but also by many academic geographers who typically have non-quantitative backgrounds. Unfortunately, this perception has deterred many researchers from appreciating the nature of the debates which have emerged and which will continue to emerge within modern quantitative geography. This becomes clear in continuing criticisms of quantitative geography which pertain to methodologies that have been surpassed by developments within the field.

This book attempts to redress the rather antiquated view of quantitative geography held by many of those outside the area. Despite what is sometimes perceived from the outside as a relatively static research area, there have in fact been a large number of major intellectual changes within the past decade in quantitative geography. These are often not simply the development of new techniques, which is inevitably happening, but reflect philosophical changes in the way quantitative geography is approached. It is fair to say that some debate has accompanied these changes; one purpose of this book is to describe these developments and to review some of the concomitant issues. In this way, the book portrays quantitative geography as a vibrant and intellectually exciting part of the discipline in which many new developments are taking place and many more await discovery.

This text is, therefore, not intended as a recipe book of quantitative techniques; nor is it meant to be a comprehensive review of *all* of quantitative geography. Rather, it is our aim to provide a statement on the vitality of modern quantitative geography. As such, it provides examples of how quantitative geography, as currently applied, differs from that of twenty, and even ten, years ago. Perhaps the most important role of this book is to provide examples of recent research in quantitative geography where the emphasis has been on the development of techniques explicitly for *spatial* data analysis. What makes the methods of modern quantitative geography different from many of their predecessors is that they have been developed with the recognition that spatial data have unique properties and that these properties make the use of methods borrowed from aspatial disciplines highly questionable. As such, this book acknowledges what the authors see as a turning point in the development of quantitative geography. It is written at a period when quantitative geography has reached a stage of maturity in which its practitioners are no longer primarily importers of other disciplines' techniques but are mainly exporters of novel ideas about the analysis of spatial data.

We hope that by advertising some of the recent developments in spatial analysis and modelling we might foster a greater interest in, and appreciation for, modern quantitative geography. This is particularly the case, for example, when considering developments in visualization, exploratory data analysis, spatial statistical inference and GIS-based forms of spatial analysis.

Inevitably, this book will find its main audience among established quantitative geographers who wish to keep abreast of the rapid developments in the field. However, we hope that it also finds a broader readership, particularly among researchers in related fields who increasingly recognize the need for specialized techniques in spatial data analysis. It should also be useful to non-quantitative geographers who would like to understand some of the current issues and debates in quantitative geography. Perhaps rather ambitiously, we also hope that the book will be read by those geography students who would like to have a better understanding of what quantitative geography can offer as a contribution to making informed decisions related to career paths.

Given the diversity of our intended audience, we recognise that it will be impossible to satisfy every level of readership on every page. Those who are not quantitatively trained are advised to skim some of the more mathematical treatises while those who are quantitatively trained are asked to be tolerant through some of the more descriptive sections.

Finally, the authors would like to express their gratitude to Ann Rooke for her help with some of the figures and to Robert Rojek at Sage Publications for his enthusiasm, encouragement and patience. We are also extremely grateful for the comments of Dave Unwin and Mike Goodchild on earlier drafts of the book. Any remaining errors are, of course, the sole responsibility of the authors.

1 Establishing the Boundaries

1.1 Setting the scene

For many reasons, it is often difficult to write anything definitive about academic trends. Some trends are so short lived that they have relatively little impact; some are cyclical so that their impact at the time of writing is different from that at the time of reading; and some trends exhibit marked variations across countries in both their intensity and their timing so that any comments have limited spatial application. These caveats aside, it is fair to say that quantitative geography generally experienced a 'downturn' in its popularity between the early 1980s and the mid-1990s (Johnston, 1997; Graham, 1997). The reasons for this are difficult to separate and probably include a mix of the following:

1 A disillusionment with the positivist philosophical underpinnings of much of the original research in quantitative geography and the concomitant growth of many new paradigms in human geography, such as Marxism, post-modernism, structuralism and humanism, which have attracted adherents united often in their anti-quantitative sentiments. This disillusionment is very much a phenomenon of human geography: there appears to be no equivalent in physical geography where quantitative methods are generally viewed as an essential component of research. The demise of quantitative human geography has therefore inevitably led to an unfortunate widening of the gap between human and physical geographers because of the lack of any common language or philosophy. As Graf (1998, p. 2) notes:

> While their human geographer colleagues have been engaged in an ongoing debate driven first by Marxism, and then more recently by post-structuralism, post-modernism, and a host of other isms, physical geographers are perplexed, and not sure what all the fuss is about. . . . They do not perceive a need to develop a post-modern climatology, for example, and they suspect . . . that some isms are fundamentally anti-scientific.

2 The seemingly never-ending desire for some new paradigm or, in less polite terms, 'bandwagon' to act as a cornerstone of geographical research. The methodology of quantitative geography, had, for some, run its course by 1980 and it was time to try something new. While it is a strength of geography that the discipline quickly absorbs new trends and research paradigms, it is also a

considerable weakness. The observations of de Leeuw (1994) on social sciences in general are apposite here. In adding to Newton's famous phrase concerning the cumulative nature of research that we stand 'on the shoulders of giants', de Leeuw (p. 13) comments:

> It also means . . . that we stand on top of a lot of miscellaneous stuff put together by thousands of midgets. . . . This is one of the peculiar things about the social sciences. They do not seem to accumulate knowledge, there are very few giants, and every once in a while the midgets destroy the heaps.

3 A line of research that appears to be better accepted in human geography than in some related disciplines is one that is critical of existing paradigms. As quantitative geography was a well-established paradigm, it became, inevitably, a focal point for criticism. Unfortunately, much of this criticism originated from individuals who had little or no understanding of quantitative geography. As Gould (1984, p. 26) notes:

> few of those who reacted against the later mathematical methodologies knew what they were really dealing with, if for no other reason than they had little or no mathematics as a linguistic key to gain entry to a different framework, and no thoughtful experience into the actual employment of such techniques to judge in an informed and reasoned way. Furthermore, by associating mathematics with the devil incarnate, they evinced little desire to comprehend. As a result, they constantly appeared to be *against* something, but could seldom articulate their reasons except in distressingly emotional terms.

4 As part of the broader 'information revolution' which has taken place in society, the growth of geographical information systems (GIS), or what is becoming known as geographical information science (GISc), from the mid-1980s onwards has had some negative impacts on quantitative studies within geography. Interestingly, these negative impacts appear to have resulted from two quite different perceptions of GISc. To some, GISc is seen either as the equivalent of quantitative geography, which it most certainly is not, or as the academic equivalent of a Trojan horse with which quantitative geographers are attempting to reimpose their ideas into the geography curriculum (Johnston, 1997; Taylor and Johnston, 1995). To others, particularly in the USA where geography has long been under threat as an academic discipline, GISc has tended to displace quantitative geography as the paramount area in which students are provided with all-important job-related skills (Miyares and McGlade, 1994; Gober et al., 1995).[1]

5 Quantitative geography is relatively 'difficult' or, perhaps more importantly, is perceived to be relatively difficult both by many academic geographers, who typically have limited quantitative and scientific backgrounds, and by many students. This affects the popularity of quantitative geography in several ways.

It is perceived by many students to be easier to study other types of geography and their exposure to quantitative methodology often extends little beyond a mandatory introductory course. It deters established non-quantitative researchers from understanding the nature of the debates that have emerged and which will continue to emerge within quantitative geography. It also makes it tempting to dismiss the whole field of quantitative geography summarily through criticisms that have limited validity rather than trying to understand it. As Robinson (1998, p. 9) states:

> It can be argued that much of the antipathy towards quantitative methods still rests upon criticisms based on consideration of quantitative work carried out in the 1950s and 1960s rather than upon attempts to examine the more complete range of quantitative work performed during the last two decades.

The relative difficulty of the subject matter might also have encouraged some researchers to 'jump ship' from quantitative geography (for some interesting anecdotes along these lines, see Billinge et al., 1984) as they struggled to keep up with the development of an increasingly wide array of techniques and methods. As Hepple (1998) notes:

> I am inclined to the view that some geographers lost interest in quantitative work when it became too mathematically demanding, and the 'hunter-gatherer' phase of locating the latest option in SPSS or some other package dried up.

This book is written in response to several of the issues raised in the above discussion. Despite being perceived from the outside as a relatively static research area, quantitative geography has witnessed a number of profound changes in the way it is approached. One purpose of this book is to describe not only some of these developments but also the debates surrounding them. In this way, we hope to present a view of quantitative geography as a vibrant, intellectually exciting, area in which many new developments are taking place, and in which many more await discovery.

A second reason for writing the book is that we hope to demonstrate that because of the changes taking place and that have taken place within the subject, several of the well-oiled criticisms traditionally levelled at quantitative geography no longer apply. For instance, the overly simplistic depictions of many that quantitative geographers search for global laws, and that individuals' actions can be modelled without understanding their cognitive and behavioural processes, have rather limited applicability. For those who insist on 'pigeon-holing' everything, modern quantitative geography, with its emphasis on issues such as local relationships, exploratory analysis and individuals' spatial cognitive processes, must be a difficult area to classify.

A third reason is the hope that some of the changes taking place in quantitative geography might make it more appealing to students and by advertising the existence of these developments, we might foster a greater interest in and appreciation for what modern quantitative geography has to offer. This is particularly the case, for example, in the subsequent discussions on topics such as visualization, exploratory data analysis, local forms of analysis, experimental significance testing and GIS-based forms of spatial analysis.

Ultimately, we hope that this book finds readership not just amongst established quantitative geographers who wish to keep abreast of the rapid developments in the field. It should also be useful to quantitative researchers in related disciplines who are increasingly recognizing the need for specialized techniques for handling spatial data. We hope it might also be of some use to non-quantitative geographers who would like to understand some of the current issues and debates in quantitative geography. Finally, it may be of assistance to students who would like to have a better understanding of what quantitative geography can offer, prior to making informed, rather than prejudicial, decisions related to career paths. Given the diversity of our intended audience, we recognize that it will be impossible to satisfy every level of readership on every page. Those who are not quantitatively trained are advised to skim some of the more mathematical sections whilst those who are quantitatively trained are asked to be tolerant through some of the more descriptive sections.

1.2 What is quantitative geography?

Quantitative geography consists of one or more of the following activities: the analysis of numerical spatial data; the development of spatial theory; and the construction and testing of mathematical models of spatial processes. The goal of all these activities is to add to our understanding of spatial processes. This can be done directly, as in the case of spatial choice modelling (Chapter 9) where mathematical models are derived based on theories of how individuals make choices from a set of spatial alternatives. Or, it can be done indirectly, as in the analysis of spatial point patterns (Chapter 6), from which a spatial process might be inferred.

It would perhaps be difficult to claim that the field of quantitative geography is sustained by any deep-rooted philosophical stance or any political agenda. For most of its practitioners, the use of quantitative techniques stems from a simple belief that in many situations, numerical data analysis or quantitative theoretical reasoning provides an efficient and generally reliable means of obtaining knowledge about spatial processes. Whilst it is recognized that various criticisms can be levelled at this approach (and quantitative researchers are often their own sternest critics), it is also recognized that no alternative approach is free of criticism and none comes close to providing the level of information on spatial processes obtained from the quantitative analysis of spatial data. The objective of most studies in quantitative geography is therefore not to produce a flawless piece of research (since in most cases, especially when dealing with social science data, this is impossible), but rather it is to maximize knowledge on spatial processes with the minimum of error. The appropriate question to ask of quantitative research therefore is 'How useful is it?' and not 'Is it completely free of error?'. This does not mean that error is to be ignored. Indeed, the ability to assess error is an important part of many quantitative studies and is obviously a necessary component in determining the utility of an analysis. It does imply, however, that studies can be useful even though they might be subject to criticism.

It might be tempting to label all quantitative geographers as positivists or naturalists (Graham, 1997) but this disguises some important differences in philosophy across the protagonists of quantitative geography. For example, just as some quantitative geographers believe in a 'geography is physics' approach (naturalism) which involves a search for global 'laws' and global relationships, others recognize that there are possibly no such entities. They concentrate on examining variations in relationships over space through what are known as 'local' forms of analysis (Fotheringham, 1998; Fotheringham and Brunsdon, 1999; see also Chapters 5 and 6). This division of belief is perhaps quite strongly correlated with subject matter. Quantitative physical geographers, because their investigations are more likely to involve predictable processes, tend to adopt a naturalist viewpoint more frequently than their human geography counterparts. In human geography, where the subject matter is typically clouded by human idiosyncrasies, measurement problems and uncertainty, the search is not generally for hard evidence that global 'laws' of human behaviour exist. Rather, the emphasis of quantitative analysis in human geography is to accrue sufficient evidence which makes the adoption of a particular line of thought compelling. As Bradley and Schaefer (1998, p. 71) note in discussing differences between social and natural scientists:

> the social scientist is more like Sherlock Holmes, carefully gathering data to investigate unique events over which he had no control. Visions of a positive social science and a 'social physics' are unattainable, because so many social phenomena do not satisfy the assumptions of empirical science. This does not mean that scientific techniques, such as careful observation, measurement, and inference ought to be rejected in the social sciences. Rather, the social scientist must be constantly vigilant about whether the situation being studied can be modeled to fit the assumptions of science without grossly misrepresenting it. . . . Thus, the standard of persuasiveness in the social sciences is different from that of the natural sciences. The standard is the compelling explanation that takes all of the data into account and explicitly involves interpretation rather than controlled experiment. The goals of investigation are also different – the creation of such compelling explanations rather than the formation of nomothetic laws.

As well as being less concerned with the search for global laws than some might imagine, quantitative geography is not as sterile as some would argue in terms of understanding and modelling human feelings and psychological processes (Graham, 1997). Current research, for example, in spatial interaction modelling emphasizes the psychological and cognitive processes underlying spatial choice and how we think about space (see Chapter 9). Other research provides information on issues such as the effects of race on shopping patterns (Fotheringham and Trew, 1993) and gender on migration (Atkins and Fotheringham, 1999). There appears to be a strong undercurrent of thought amongst those who are not fully aware of the nuances of current quantitative geography that it is deficient in its treatment of human influences on spatial behaviour and spatial processes. While there is some validity in this view, quantitative geographers increasingly recognize that spatial patterns resulting from human decisions need to account for aspects of human decision-making processes. This is exemplified by the current interest in spatial information processing strategies and the linking of spatial cognition with spatial choice (see Chapter 9). It should also be borne in mind that the actions of humans in aggregate often result from two types of determinants which mitigate against the need to consider every aspect of human behaviour. There are those, such as the deterrence of distance in spatial movement, which can be quantified and applied to groups of similar individuals; and those, such as shopping at a store because of knowing someone who works there, which are highly idiosyncratic and very difficult to quantify. One of the strengths of a quantitative approach is that it enables the measurement of the determinants that can be measured (and in many cases these provide very useful and very practical information for real-world decision making) whilst recognizing that for various reasons, these measurements might be subject to some uncertainty. This recognition of the role of uncertainty is often more important in the applications of quantitative techniques to human geography than to physical geography and makes the former in some ways more challenging and at the same time more receptive to innovative ideas about how to handle this uncertainty.

To some extent the above comments can be made about the use of quantitative methods in other disciplines. What distinguishes quantitative geography from, say, econometrics or quantitative sociology, or, for that matter, physics, engineering or operations research, is its predominant focus on spatial data. Spatial data are those which combine attribute information with locational information (see Chapter 2). For example, a list of soil chemistry properties or unemployment figures is aspatial unless the locations for which the data apply are also given. As described in Chapter 2, spatial data often have special properties and need to be analysed in different ways from aspatial data. Indeed, the focus of this book centres on this very point. Until relatively recently, the complexities of spatial data were often ignored and spatial data were analysed with techniques derived for aspatial data, a classic case of this being regression analysis (see Chapters 5 and 7). What we concentrate on in this book are those areas where techniques and methodologies are being developed *explicitly* for spatial data. Hence, topics such as log–linear modelling and various categorical data approaches, which have been applied to spatial data but which have not been developed with spatial data explicitly in mind, are not covered in this text. The increasing recognition that 'spatial is special' reflects the maturing of quantitative geography from being predominantly a user of other disciplines' techniques to being an exporter of ideas about the analysis of spatial data.

The definition of quantitative geography at the beginning of this section encompasses a great variety of approaches to the subject. Some of these approaches conflict with one another and, where they do, debates arise. We will try to give a flavour of some of these debates in subsequent chapters but the debates have not been particularly acrimonious and generally a *laissez-faire* attitude prevails in which different approaches are more often seen as complementary rather than as contradictory. For example, quantitative geography encompasses both empirical and theoretical research. Advances in theory are typically very difficult to accomplish but are clearly essential for the progression of the subject matter. Obviously, any theoretical development needs to be subject to intense empirical examination, particularly in the social sciences where the general acceptance of theoretical ideas usually takes hold slowly. Typically in geography, as in other disciplines, empirical research has depended on theoretical ideas for its guidance and the dependency is still very much in this direction. However, with the advent of new ideas and techniques in exploratory spatial data analysis (see Chapter 4), empirical research is increasingly being used to guide theoretical development to form a more equal symbiosis. The last decade has probably seen a gradual decline in purely theoretical research in quantitative geography and more of an emphasis on empirical research. To a large part, this change has been brought about by the enormous advances in computational power available to most researchers which has certainly boosted empirical investigations, often computationally very intensive, of large spatial data sets (Fotheringham, 1998; 1999a). However, while computationally intensive methods are revolutionizing some areas of quantitative geography and have made the calibration of theoretical models easier, it has also been argued that in some cases computational power is being relied upon too heavily (Fotheringham, 1998). The 'solutions' to geographical problems found in this way may have limited applicability and may be obtained at the expense of the deeper understanding that comes from theoretical reasoning.

Another division within quantitative geographical research is that between research which is centred on the statistical analysis of spatial data and research focused on mathematical modelling. However, the distinction between what constitutes statistical as opposed to mathematical research can sometimes be blurred and it is perhaps not a particularly important one to make here. A model might, for example, be developed from mathematical principles and then be calibrated by statistical methods. Typically, areas such as the analysis of point patterns (Chapter 6), spatial regression concepts (Chapters 5 and 7) and various descriptive measures of spatial data such as spatial autocorrelation (Chapter 8) are thought of as 'statistical' whereas topics such as spatial interaction modelling (Chapter 9) and location–allocation modelling (Ghosh and Rushton, 1987; Fotheringham et al., 1995) are thought of as 'mathematical'. Statistical methods have come to dominate quantitative geography, particularly in the social sciences, because of the need to account for errors and uncertainty in both data collection and model formulation. Indeed the term 'spatial analysis' is sometimes used as a synonym for quantitative geography although to some the term implies only stochastic forms of analysis rather than deterministic forms of spatial modelling. It is worth noting that a different usage of the term 'spatial analysis' appears to have become commonplace in the GIS field where software systems are advertised as having a suite of data manipulation routines for 'spatial analysis'. However, these routines typically perform geometrical operations such as buffering, point-in-polygon, overlaying and cookie-cutting which form an extremely minor part of what is typically thought of as 'spatial analysis' by quantitative geographers (see Chapter 3).

1.3 Applications of quantitative geography

A major goal of geographical research, whether it be quantitative or qualitative, empirical or theoretical, humanistic or positivist, is to generate knowledge about the processes influencing the spatial patterns, both human and physical, that we observe on the earth's surface. Typically, and particularly so in human geography, acceptance of such knowledge does not come quickly; rather it emerges after a long series of tests to which an idea or a hypothesis is subjected. The advantages of quantitative analysis in this framework are fourfold.

First, quantitative methods allow the reduction of large data sets to a smaller amount of more meaningful information. This is important in analysing the increasingly large spatial data sets obtained from a variety of sources such as satellite imagery, census counts, local government, market research firms and various land surveys. Many spatial data sets can now be obtained very easily over the World Wide Web (e.g. see the plethora of sites supplying spatial data given at *http://www.clark.net/pub/lschank/web/census.html* and in particular the sites of the US Census, *http://www.census.gov*, the US National Imagery and Mapping Agency, *http://www.nima.mil*, and the US Geological Survey, *http://www.usgs.gov*). Summary statistics and a wider body of data reduction techniques (see Chapter 4 for some examples of the latter) are often needed to make sense of these very large, multidimensional data sets.

Secondly, an increasing role for quantitative analysis is in *exploratory data analysis* which consists of a set of techniques to explore data (and also model outputs) in order to suggest hypotheses or to examine the presence of outliers (see Chapter 4). Increasingly we recognize the need to visualize data and trends prior to performing some type of formal analysis. It could be, for example, that there are some errors in the data which only become clear once the data are displayed in some way. It could also be that visualizing the data allows us to check assumptions and to suggest ways in which relationships should be modelled in subsequent stages of the analysis.

Thirdly, quantitative analysis allows us to examine the role of randomness in generating observed spatial patterns of data and to test hypotheses about such patterns. In spatial analysis we typically, although not always, deal with a sample of observations from a larger population and we wish to make some inference about the population from the sample. Statistical analysis will allow such an inference to be made (see also Chapter 8). For instance, suppose we want to investigate the possible linkage between the location of a nuclear power station and nearby incidences of childhood leukaemia. We could use statistical techniques to inform us of the probability that such a spatial cluster of the disease could have arisen by chance. Clearly, if the probability is extremely low then our suspicions of a causal linkage to the nuclear power station are increased. The statistical test would not provide us with a definite answer – we would just have a better basis on which to judge the reliability of our conclusion. Arguably, the use of such techniques provides us with information on spatial patterns and trends in a less tendentious manner than other techniques. For example, leaving inferences to the discretion of an individual after he or she has been presented with rather nebulous evidence is clearly open to a great deal of subjectivity. How the evidence is viewed is likely to vary across individuals. Similarly, the results from quantitative analyses are likely to be more robust than, for example, studies that elicit large amounts of non-quantitative information from a very small number of individuals.

Fourthly, the mathematical modelling of spatial processes is useful in a number of ways. The calibration of spatial models provides information on the determinants of those processes through the estimates of the models' parameters (see Chapters 5 and 9 for examples). They also provide a framework in which predictions can be made of the spatial impacts of various actions such as the building of a new shopping development on traffic patterns or the building of a seawall on coastal erosion. Finally, models can be used normatively to generate expected values under different scenarios against which reality can be compared.

In summary, the quantitative analysis of spatial data provides a robust testing ground for ideas about spatial processes. Particularly in the social sciences, ideas become accepted only very gradually and have to be subject to fairly rigorous critical examination. Quantitative spatial analysis provides the means for strong evidence to be provided either in support of or against these ideas. This is as true in many other disciplines as it is in geography because it is increasingly recognized that most data are spatial in that they refer to attributes in specific locations. Consequently, the special problems and challenges that spatial data pose for quantitative analysis (see Chapters 2 and 10) are increasingly seen as relevant in a variety of subject areas beyond geography (Grunsky and Agterberg, 1992; Goovaerts, 1992; 1999; Cressie, 1993; Krugman, 1996; Anselin and Rey, 1997). Examples include economics, which is increasingly recognizing that many of its applications are spatial; archaeology, where settlement data or the location of artefacts clearly have spatial properties; epidemiology, where space plays an important role in the study of morbidity and mortality rates; political science, where voting patterns often exhibit strong spatial patterns; geology and soil science, where inferences need to be made about data values that lie between sampled points; health care services, where patients' residential locations are important in understanding hospital rationalization decisions; and marketing, where knowledge of the locations of potential customers is vital to understanding store location. For these reasons, quantitative geographers have skills which are much in demand in the real world and are much sought after to provide inputs into informed decision making.

1.4 Recent developments in quantitative geography

The basic reason for writing this book is that quantitative geography has undergone many changes in the last 20 years and particularly in the last decade. These changes have, in some cases, involved fundamental shifts in the way quantitative geographers view the world. However, few outside the area appear to be aware of these changes. Instead they tend to view and to criticize the field for what it was rather than for what it is. While the subsequent chapters will give a much greater feel for the recent changes that have taken place in quantitative geography, this section gives a flavour of some of these discussions.

The development and maturation of GIS has had an effect on quantitative geography, not always in a positive way as noted above and as also commented on by Fotheringham (1999b). In terms of the development of quantitative methods for spatial data, however, the ability to apply such methods within GIS, or at least link the outcome of such methods with GIS, leads to an increase in the potential for gaining new insights (see Chapter 3). As Fotheringham (1999b, p. 23) notes:

> I would argue that it is not necessary to use a GIS to undertake spatial modelling and integrating the two will not necessarily lead to any greater insights into the problem at hand. However, for certain *aspects of the modelling procedure*, integration will have a reasonably high probability of producing insights that would otherwise be missed if the spatial models were not integrated within the GIS.

It is argued that these 'certain aspects of the modelling procedure' for which integration within GIS will be especially beneficial are exploratory techniques (see Chapter 4). Exploratory techniques are used to examine data for accuracy and robustness and to suggest hypotheses which may be tested in a later confirmatory stage. This typical usage can be classified as *pre-modelling exploration*. However, exploratory techniques are not confined to data issues and another use, termed *post-modelling exploration*, is to examine model accuracy and robustness. One relatively simple example of post-modelling exploration with which many readers will already be familiar is the mapping of the residuals from a model in order to provide improved understanding of why the model fails to replicate the data exactly. Clearly, this is a situation where an interactive mapping system would be useful: not only could the map of residuals be viewed but also it could be interrogated. Zones containing interesting residuals could be highlighted to show various attributes of the zone which might be relevant in understanding the performance of the model. Similarly, an aspatial distribution of residuals in one window could be brushed and the brushed values highlighted on a map in a linked window to explore spatial aspects of model performance. Xia and Fotheringham (1993) provide a demonstration of the exploratory use of linked windows in Arc/Info. Further examples of the power of interactive visualization for spatial data are provided in Chapter 4 and by Anselin (1998), Brunsdon and Charlton (1996) and Haslett et al. (1990; 1991).

Current research on visualization in spatial data sets is focused on the need for visualization tools for higher-dimensional spatial data sets (Fotheringham, 1999c). Most visualization techniques have been developed for simple univariate or bivariate data sets (extensions of some of these techniques can be made to visualize trivariate data). However, most spatial data sets have many attributes and hence these relatively simple visualization techniques are inadequate to examine the complexities within such data sets. Relatively few techniques have been developed for more realistic and more frequently encountered hypervariate (having more than three dimensions) data sets (Cleveland, 1993) and the development of such techniques is therefore becoming of greater concern to quantitative geographers. Some examples of visualization techniques for higher-dimensional spatial data sets are provided in Chapter 4.

Another recent and potentially powerful movement within quantitative geography is that in which the focus of attention is on identifying and understanding *differences* across space rather than *similarities.* The movement encompasses the dissection of global statistics into their local constituents; the concentration on local exceptions rather than the search for global regularities; and the production of local or mappable statistics rather than on 'whole-map' values. This trend is important not only because it brings issues of space to the fore in analytical methods, but also because it refutes the rather naive criticism that quantitative geography is unduly concerned with the search for global generalities and 'laws'. Quantitative geographers are increasingly concerned with the development of techniques aimed at the local rather than the global (Anselin, 1995; Unwin, 1996; Fotheringham, 1997a). This shift in emphasis also reflects the increasing availability of large and complex spatial data sets in which local variations in relationships are likely to be more prevalent.

The development of local statistics in geography is based on the idea that when analysing spatial data, it might be incorrect to assume that the results obtained from the whole data set represent the situation in all parts of the study area. Interesting insights might be obtained from investigating spatial variations in the results. Simply reporting one 'average' set of results and ignoring any possible spatial variations in those results is equivalent to reporting a mean value of a spatial distribution without seeing a map of the data. It is therefore surprising that local statistics have not been the subject of much investigation until recently. The importance of the emphasis on 'local' instead of 'global' is presented in Chapter 5 which also includes a detailed description of several examples of locally based spatial analysis. The chapter concentrates on geographically weighted regression, an explicitly spatial technique derived for producing local estimates of regression parameters, which can be used to produce parameter maps from which a 'geography of spatial relationships' can be examined (Brunsdon et al., 1996; 1998a; Fotheringham et al., 1996; 1997a; 1997b; Fotheringham et al., 1998).

Another development in quantitative geography that explicitly recognizes the special problems inherent in spatial data analysis is that of spatial regression models (Ord, 1975; Anselin, 1988). The fact that spatial data typically are positively spatially autocorrelated, that is high values cluster near other high values and low values cluster near other low values, violates an assumption of the classical regression model that the data consist of independent observations. This creates a problem in assessing statistical significance and model calibration: in essence, the errors in the regression model can no longer be assumed to have zero covariance with each other. To counter this problem, Anselin (1988) has suggested two alternative models, a spatial lag model in which the dependent variable exhibits positive spatial autocorrelation and a spatial error model in which the errors in the regression are spatially autocorrelated. These models are described in Chapter 7.

The term 'geocomputation' has been coined to describe techniques, primarily quantitative, within geography that have been developed to take advantage of the recent massive increases in computer power and data (Openshaw and Openshaw, 1997; Openshaw and Abrahart, 1996; Openshaw et al., 1999; Fotheringham, 1998; 1999a; Longley et al., 1998). The term 'computation' carries two definitions. In the broader sense it refers to the use of a computer and therefore any type of analysis, be it quantitative or otherwise, could be described as 'computational' if it were undertaken on a computer. In the narrower and perhaps more prevalent use, computation refers to the act of counting, calculating, reckoning or estimating – all terms that invoke quantitative analysis. The term 'geocomputation' therefore refers to the computer-assisted quantitative analysis of spatial data *in which the computer plays a pivotal role* (Fotheringham, 1998). This definition is meant to exclude fairly routine analyses of spatial data with standard statistical packages (for instance, running a regression program in SAS or SPSS). Under this definition of geocomputational analysis, the use of the computer *drives* the form of analysis undertaken rather than being a convenient vehicle for the application of techniques developed independently of computers. Geocomputational techniques are therefore those that have been developed *with the computer in mind* and which exploit the large increases in computer power that have been, and still are being, achieved.

A simple example, which is developed in Chapter 8 in a discussion of statistical inference, serves to distinguish the two types of computer usage. Consider a spatial autocorrelation coefficient, Moran's I, being calculated for a variable x distributed across n spatial units. Essentially, spatial autocorrelation describes how an attribute is distributed over space – to what extent the value of the attribute in one zone depends on the values of the attribute in neighbouring zones (Cliff and Ord, 1973; 1981; Odland, 1988; Goodchild, 1986). To assess the significance of the autocorrelation coefficient one could apply the standard formula for a t-statistic calculating the standard error of Moran's I from one of two possible theoretical formulae (see Cliff and Ord, 1981; Odland, 1988; Goodchild, 1986; or Chapter 8 for these formulae and examples of their application). Such a procedure is not geocomputational because the computer is simply used to speed up the calculation of a standard error from a theoretical equation. An alternative, geocomputational, technique would be to derive an estimate of the standard error of the autocorrelation coefficient by experimental methods. One such method would be to permute randomly the x variable across the spatial zones and to calculate an autocorrelation coefficient for each permutation. With a sufficiently large number of such autocorrelation coefficients (there is no reason why millions could not be computed but thousands or even hundreds are generally sufficient), an experimental distribution can be produced which allows statistical inferences to be made on the observed autocorrelation coefficient. An example of this type of geocomputational application is given in Chapter 8.

The use of computational power to replace an assumed theoretical distribution has the advantage of avoiding the assumptions underlying the theoretical distribution which may not be met, particularly with spatial data. Consequently, the use of experimental significance testing procedures neutralizes the criticism that hypothesis testing in quantitative geography is overly reliant on questionable assumptions about theoretical distributions. Another criticism of quantitative geography, addressed in Chapter 9, is the assumption that spatial behaviour results from individuals behaving in a rational manner and armed with total knowledge. Perhaps the classic case of this kind of assumption is in spatial interaction modelling where the early forms of what are known as 'gravity models' were taken from a physical analogy to gravitational attraction between two planetary bodies. In Chapter 9 we attempt to show how far we have come since this analogy was made over 100 years ago (although quantitative geography is still criticized for it!). Newer forms of spatial interaction models, based on sub-optimal choices, limited information, spatial cognition and more realistic types of spatial decision-making processes, are described.

1.5 Summary

There are at least two constraints to undertaking quantitative empirical research within geography. One is our limited ability to think about how spatial processes operate and to produce insights that lead to improved forms of spatial models. The other is the restricted set of tools we have to test and refine these models. These tools might be used for data collection (e.g. GPS receivers, weather stations, stream gauges) or for data display and analysis (GIS, computers). In the early stages of computer use, it was relatively easy to derive models that could not be implemented because of the lack of computer power. This was an era when the second constraint was more binding than the first: the level of technology lagged behind our ability to think spatially. We are now no longer in this era. We are now in a situation where the critical constraint is more likely to be our ability to derive new ways of modelling spatial processes and analysing spatial data. The increases in computer power within the last 20 years have been so enormous that the technological constraint is much less binding than it once was. The challenge is now to make full use of the technology to improve our understanding of spatial processes. In many instances the change is so profound that it can alter our whole way of thinking about issues: the development of experimental significance testing procedures and the subsequent decline in the reliance on theoretical distributions is a case in point. The movement from global modelling to local modelling is another.

This book provides a statement on the vitality of modern quantitative geography. It does not, however, attempt to cover all the facets of the subject area. Instead, it concentrates on examples of how quantitative geography differs from the possibly widespread perceptions of it outside the field. In doing so, it provides examples of the research frontier across a broad spectrum of applications where techniques have been developed *explicitly with spatial data in mind*. The book acknowledges a turning point in the development of quantitative geography: it is written at a period when quantitative geographers have matured from being primarily importers of other disciplines' techniques to being primarily exporters of novel ideas and insights about the analysis of spatial data.

Notes

1. The diffusion of both quantitative geography and GISc has been less extensive within the UK where, mainly because of the traditionally more selective nature of university education, geography students have enjoyed relatively good prospects of employment without necessarily having many specific skills. However, this situation is changing very rapidly.

2 Spatial Data

2.1 Introduction

A glance at the shelves of almost any university library will reveal a plethora of books concerned with geographical information systems (e.g. Burrough, 1986; Huxhold, 1991; Laurini and Thompson, 1992; Rhind et al., 1991; Haines-Young et al., 1993; Bonham-Carter, 1994; Martin, 1996; DeMers, 1997; Chrisman, 1997; Heywood et al., 1998).[1] Fundamental to the operation of GIS are spatial data. Although geographers have been using (and abusing) spatial data long before the mid-1980s, there has been a marked diffusion of interest in spatial data handling since then, and an increasing appreciation of the opportunities offered by, and the problems associated with, such data. Given that spatial data are so pervasive, we need to be aware of the nature of spatial data and their interaction with quantitative geography. Indeed, a number of articles in the magazine *GIS Europe* (Gould, 1996) explored briefly the notion that 'spatial is special'. Others also have begun to realize that there are special problems in analysing spatial data (Berry, 1998). This chapter explores some of the issues.

Spatial data comprise observations of some phenomenon that possesses a spatial reference. The spatial reference may be explicit, as in an address or a grid reference, or it may by implicit, as in a pixel in the middle of a satellite image. One form of spatial reference known to almost everyone in the developed world is the address of one's home although few individuals will be able to quote a map reference of their home. However, we normally convert the former into the latter to carry out any processing of such data. This chapter concerns itself first with the nature of spatial data, then with an examination of the opportunities that arise in the analysis of spatial data, and then with a consideration of the problems that confront the would-be spatial data analyst.

Spatial data are not new. Ptolemy was experimenting with spatial data in second-century Egypt when he was attempting to map his world. The early astronomers who were attempting to map the heavens were using spatial data. Attempts at global exploration by various civilizations required knowledge of locations and the means of getting from one place to another. However, the computational facilities at their disposal were rather primitive compared with the desktop computer and the proliferation of software in the last 20 years. The so-called 'GIS Revolution' has led to a more explicit interest in the handling and analysis of spatial data, an interest that has diffused widely outside geography. Geographers may lay first claim to an interest in spatial data, but they have been joined by mathematicians, physicists, geomaticists, biologists, botanists, archaeologists and architects, to name but a few.

Spatial data may mean different things to different users. John Snow's attempt to postulate a particular water pump in Soho (Gilbert, 1958) as the source of contaminated water leading to a cholera outbreak was an early attempt at spatial analysis. He integrated three spatial data sets in a single map: the locations of the streets in Soho, the locations of cholera cases, and the locations of water pumps. It is not known whether he thought of his exercise as 'spatial data analysis', or 'spatial data handling'. Almost every airline's in-flight magazine contains a map of the airline's routes – again, this displays spatial data. Readers of this book will at some time in their life ask another person for an address; here is a different form of spatial data – the address relates to some location on the earth's surface. The UK Post Office postcode system (Raper et al., 1992), originally developed for the automated handling of mail, is another form of spatial data; most people in the UK know their postcode. In the USA, the zip code is used by the US Postal Service in a similar manner. Clearly, spatial data are more common than we might realize yet not all users of spatial data think of their objects of interest as inherently spatial. Anyone who has stepped into a taxi will find in the driver someone who has had to demonstrate a wide familiarity with spatial data (in knowing the locations of streets and landmarks), and a high degree of proficiency in spatial data manipulation (the ability to work out the best route), perhaps far in advance of the capability of any current GIS.

2.2 Spatial data capture

Spatial data arise when we attempt to sample information from the real world. The nature of the sampling is such that we are interested in not only the variation of some phenomenon, but also the location of that variation. We need therefore to sample not just the nature of the phenomenon of interest, but also its location. There is a wide variety of techniques, both manual and automatic, for doing this.

Digitizing is a process that involves the transfer of locational information about features on paper maps into some computer-processable form. This involves a device known as a digitizer, or digitizing tablet. The map is fixed to the surface of the digitizer, and a cursor is moved across the map by hand. The cursor has a pair of cross-hairs in a transparent window which aids precise positioning, and one or more buttons to transfer locational information to a computer. The cross-hairs are positioned above the point whose position is to be digitized, and one of the buttons pressed to send some measurements describing its location on the surface of the digitizer to a computer. Lines are digitized by digitizing points a short distance apart; curves are approximated by a linked series of short line segments. Where a line changes direction, the operator must digitize sufficient points for the changes to be captured with the desired degree of accuracy. Deciding what is 'sufficient' is