1

Introduction

A key concern of geography and other disciplines which make use of spatially-referenced data is with differences between places. Whether the object of study is human populations or geomorphology, space is often of fundamental importance. We may be concerned with, for example, factors that effect unemployment or factors that influence soil erosion; traditionally, global methods have often been employed in quantitative analyses of datasets that represent such properties. The implicit assumption behind such methods is that properties do not vary as a function of space. In many cases, such approaches mask spatial variation and the data are under-used. The need for methods which do allow for spatial variation in the properties of interest has been recognised in many contexts. In geography and cognate disciplines, there is a large and growing body of research into local methods for spatial analysis, whereby differences between places are allowed for. This book is intended to introduce a range of such methods and their underlying concepts. Some widely-used methods are illustrated through worked examples and case studies to demonstrate their operation and potential benefits. To aid implementation of methods, relevant software packages are mentioned in the text. In addition, a summary list of selected software packages is provided in Appendix A.

The book is intended for researchers, postgraduate students, and professionals, although parts of the text may be appropriate in undergraduate contexts. Some prior knowledge of methods for spatial analysis, and of Geographical Information Systems (GISystems), is assumed. Background to some basic concepts in spatial data analysis, including elements of statistics and matrix algebra, is provided by O'Sullivan and Unwin (304), de Smith et al. (104), and Lloyd (245). There is a variety of published reviews of local models for spatial analysis (23), (127), (128), (367), but each has particular focuses. One concern here is to bring together discussion of techniques that could be termed 'local' into one book. A second concern is to discuss developments of (relatively) new techniques.

This chapter describes the remit of the book before introducing local models and methods. Then, the discussion moves on to issues of spatial dependence, spatial autocorrelation, and spatial scale. The concept of stationarity, which is key in the analysis of spatially or temporally referenced variables, is also outlined. Finally, key spatial data models are described, and the datasets used for illustrative purposes are detailed.

1.1   Remit of this book

The development and application of local models is a major research focus in a variety of disciplines in the sciences, social sciences, and the humanities. Recent work has been described in a number of journal papers and books. This book provides an overview of a range of different approaches that have been developed and employed within Geographical Information Science (GIScience). This book is not intended to be an introduction to spatial statistics in general. The aim is to start from first principles, to introduce users of GISystems to the principles and application of some widely used local models for the analysis of spatial data. The range of material covered is intended to be representative of methods being developed and employed in geography and cognate disciplines. Work is presented from a range of disciplines in an attempt to show that local models are important for all who make use of spatial data. Some of the techniques discussed are unlikely to enter widespread use in the GISystems community, and the main stress is on those approaches that are, or seem likely to be in the future, of most use to geographers. Some topics addressed, such as image processing, are not covered in detail. Rather, the principles of some key local approaches are outlined, and references to more detailed texts provided.

The applications used to illustrate the methods discussed are based on data representing various physical properties (e.g., precipitation and topography) and on human populations. These applications serve to highlight the breadth of potential uses of the methods and models discussed.

1.2   Local models and methods

Broadly, this chapter will stress the distinction between global and local methods. With a global model, the assumption is that variation is the same everywhere. However, it may be the case that a global model does not represent well variation at any individual location. Global methods make use of all available data, whereas local methods are often defined as those that make use of some subset of the data. But there are also approaches whereby the data are transformed in some way. For example, removal of a global trend (representing the spatially-varying mean of the property) may be conducted to remove large scale variation — the aim would be to obtain residuals from a regular trend (i.e., a relatively constant increase or decrease in values in some direction) across the region of interest, allowing a focus on local variation.

Local models have been used widely in some disciplines for several decades. For example, in image processing local filters have long been used to smooth or sharpen images. However, in geography a focus on the development of methods that account for local variation has been comparatively recent. Some methods, by definition, work locally. For example, many methods for analysing gridded data are always employed on a moving window basis (for example, methods for drainage network derivation and spatial filters). This book will discuss such techniques, although the main focus will be on reviewing models and methods of which there are global versions and local versions, the latter adapting in some way to local spatial variation.

That properties often vary spatially is recognised by Unwin and Unwin (367) who, in a review of the development of local statistics, outline some key concerns of geographical analysis. In particular, they note that:

1. Most spatial data exhibit spatial dependence (see Section 1.4).

2. Many analyses are subject to the modifiable areal unit problem (MAUP — results of an analysis depend on the division of space; see Section 6.4).

3. It is difficult to assume stationarity (see below) in any process observed over geographical space (for example, the mean and variance may vary markedly from place to place and thus the process can be called nonstationary).

The development of GISystems and the increased availability of spatial data has led to both the creation of problems and the development of solutions. Availability of datasets covering large areas (in particular, remotely-sensed images) increased the probability that regions with different properties would be encountered. As such, the need for local models that account for these differences increased. In addition, the capacity to collect data at very fine spatial resolutions meant that concern with spatial variation and its relation to spatial scale would increase (367).

Against this background, a key change in geography has been from a focus on *similarities* between places to *differences* across space (127). Fotheringham and colleagues (125), (128) include within this movement approaches for dissecting global statistics into their local components. Related concerns include concentration on local exceptions rather than the search for global regularities and production of local mappable statistics rather than global summaries.

Central to the theme of this section is the idea of spatial nonstationarity. That is, if the property of interest (for example, precipitation, elevation, or human population) varies from place to place, for some scale of analysis, then a nonstationary model is appropriate in the analysis of this property. Stationarity is discussed in more detail in Section 1.6. A model with constant parameters may not be appropriate in various situations. Fotheringham (125) gives three possible reasons:

1. There are spatial variations in observed relationships due to random sampling variations.

2. Some relationships are intrinsically different across space.

3. The model used to measure relationships is a gross misspecification of reality — relevant variables are missing, or the form of model adopted is inappropriate.

In practice, it may be difficult to distinguish between these reasons, but the methods described in this book constitute ways to explore these issues and, hopefully, to enhance our understanding of spatial processes. Fotheringham and Brunsdon (127) divide local methods into those approaches for analysis of univariate data, methods for analysis of multivariate data, and methods for analysis of movement patterns (spatial interaction models). The first two areas are concerns within this book, but the latter is largely outside its remit (although spatial interaction modelling approaches may be based on regression, a topic which is a concern here).

In the last decade, several important developments have taken place in quantitative geography and in allied disciplines. Such developments include methods for exploring local spatial autocorrelation (see Section 4.4) and methods for exploring variation in spatial relations between multiple variables (geographically weighted regression, discussed in Section 5.8, is an example of this). That is, models have been developed to allow for differences in properties at different locations. For example, the relationship between two properties may be markedly different in one region than in another and a local model that allows for these differences may be more appropriate than a model for which the parameters are fixed. In other areas major developments have taken place. For example, wavelets provide a powerful means of decomposing and analysing imagery (see Section 3.6). Such methods are receiving widespread attention, and there is a large range of sophisticated software packages to implement such methods. This book is intended to bring together discussions of such methods and to provide pointers to material about these methods which will enable their exploration further. Allied to these developments is a number of important summaries of recent developments, written by various authors, which are cited in the text.

1.3   What is local?

The term local can have a multitude of meanings in different contexts. In physical geography, for example, a local space may be some area over which a particular process has an obvious effect. A watershed might also

be considered, in some sense, a local space. In geomorphology, a landscape may be classified into discrete spatial areas which could be regarded as local spaces. In socioeconomic contexts, a local space may be the neighbourhood which an individual is familiar with, or the wider set of areas with which they interact on a regular basis. In terms of spatial data analysis, a local space is often expressed in terms of distance from some point or area (the locality or neighbourhood of that point or area).

A study area can only be local in the context of a global dataset, or a larger subset of the dataset. Of course, a locally-based approach is not necessarily beneficial. If a dataset is transformed or partitioned this may, for example, provide a better model fit but the results may not be meaningful or interpretable. Application of local models may be more problematic than the application of global models because of the additional complexity — factors such as the size of a moving window or the type of transform applied may have a major impact on the results obtained from an analysis. Indeed, the division of geographical space is important in any analyses of spatial data. For example, the statistics computed from an image are a function of the spatial resolution of the image (as discussed below). Similarly, results from analyses based on one set of administrative zones will be different than those obtained when another set of zones is used. Unwin and Unwin (367) outline the need to (i) define which areas to include in an operation, and (ii) decide how to treat non-zero entries. That is, which data are included in the analysis, and how much influence (weight) should each observation have? An example of the latter is the weight assigned to an observation using a spatial kernel, as discussed in, for example, Sections 2.2, 2.4, 5.8, and 8.10.2, as well as throughout Chapter 4.

1.4 Spatial dependence and autocorrelation

The core principle behind many local methods is the concept of spatial dependence. That is, objects close together in space tend to be more similar than objects which are farther apart. This principle was termed the "First Law of Geography" (as outlined by Tobler (361)). In cases where data values are not spatially dependent many forms of geographical analysis are pointless. Figure 1.1 shows synthetic examples of strong and weak spatial dependence (note that the data are treated as ratio variables in this example). The term spatial autocorrelation refers to the correlation of a variable with itself and where neighbouring values tend to be similar this is termed positive spatial autocorrelation — the data are spatially dependent. Where neighbouring values tend to be dissimilar this is termed negative spatial autocorrelation. This topic is developed further in Section 4.3.

A   B

22 black cells = 'high' values
21 grey cells = 'medium' values
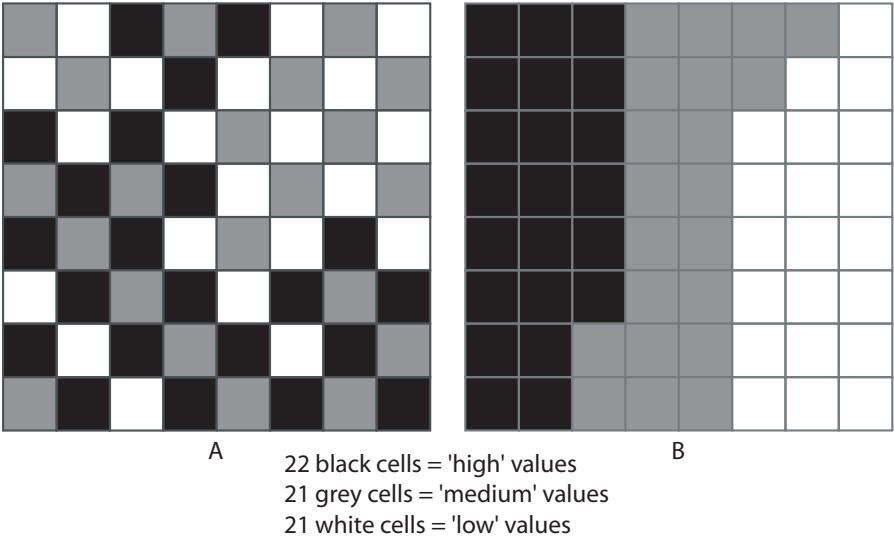21 white cells = 'low' values

FIGURE 1.1: (A) Weak spatial dependence and (B) Strong spatial dependence.

Spatial dependence is often accounted for explicitly through the use of geographical weighting functions. This can be illustrated using the example of spatial interpolation. Suppose that there is the need to predict the value of some property at a location where no data are available. One sensible way to proceed is to take a weighted average of the observations surrounding the location at which we wish to make a prediction. That is, observations close to the prediction location will be given larger weights (more influence) than observations that are more distant from the prediction location. A procedure which operates on this principle is inverse distance weighting (IDW; see Section 2.2). The spatial structure of a process, and thus the degree of spatial dependence, may vary from place to place. Standard IDW accounts for local variation using a geographical weighting scheme of a given form, but it may be that a different weighting function should be employed in different areas. This book introduces methods that allow for variation in spatial structure. A property may be spatially structured at one scale, but spatially unstructured at another scale. For example, an image may appear 'noisy' at a fine spatial scale and structured at a coarser spatial scale, and this structure may vary from place to place.

1.5   Spatial scale

The concept of scale is central to all disciplines concerned with the spatial arrangement of properties. The term has been defined in many different ways. For example, a map is defined by its scale (we can talk of a large scale map or a small scale map). In the context of this discussion, and in common with the account of Atkinson and Tate (29), scale is taken to refer to the size or extent of a process, phenomenon, or investigation. Bian (41) uses the term 'operational scale' to refer to the scale over which a process operates. The availability of a wide range and type of data sources for locations around the globe means that users of spatial data are faced with working with multiscale representations — for example, a user may have, for one region, several remotely-sensed images that have different spatial resolutions. Users of such data usually have little choice about the scale of measurement (i.e., in this context, spatial resolution). As such, it is necessary to develop ways to work at a range of spatial scales (152).

Atkinson and Tate (29) state that spatial scale comprises (i) scales of spatial measurement, and (ii) scale of spatial variation (see Section 7.14 for relevant discussion). There are two scales of measurement — the support (geometrical size, shape, and orientation of the measurement units) and the spatial coverage of the sample (29). A variety of approaches exist for characterising scales of spatial variation and some approaches are mentioned below. There are many reviews of spatial scale problems; Atkinson and Tate (29) provide such a review in a geostatistical framework (see Chapter 7 for more information on related approaches).

The scale of spatial variation may change with location. That is, the dominant scale of spatial variation at one location may be quite different from that at another location. Hence, there is a need for approaches that allow variation with respect to (i) spatial scale, and (ii) spatial location. Many spatial processes, both (for example) physical and socioeconomic, may appear homogeneous at one scale and heterogeneous at another (Lam (228) discusses this idea with respect to ecological diversity). Clearly, locally-adaptive approaches are only necessary if the property of interest is spatially heterogenous at the scale of measurement. This book outlines a variety of approaches that allow exploration of local differences in scales of spatial variation.

Clearly, the shape and size of the area over which a property is recorded affects directly the results obtained through analyses of those data. Each level of a hierarchy of data presented at different spatial scales has unique properties that are not necessarily a simple sum of the component (disaggregated) parts (41). As Lam (228) states, the spatial resolution of an image changes fundamental biophysical relationships (known as the ecological fallacy), and the same is true in other contexts. That is, spatial models are frequently scale

dependent — models that are applicable at one scale may not be appropriate at another scale (41). The modifiable areal unit problem (as defined on page 3 and discussed further in Section 6.4) reflects the fact that areal units can be changed and observed spatial variation altered. For example, the degree of spatial dependence is likely to change as the areal units are changed (see, for example, Lloyd (242)).

In the context of physical geography, Atkinson and Tate (29) note that nearly all environmental processes are scale dependent. So, the observed spatial variation is likely to vary at different scales of measurement. This means that there is a need to construct a sampling strategy that enables identification of spatial variation of interest. To facilitate acquisition of suitable data and integration of data at different spatial scales or different variables, the scaling properties of spatial variables should be used (29). However, spatial dependence may be unknown or may differ markedly in form from place to place, and there is the added problem that patterns at a given scale may be a function of interactions amongst lower-level systems (29). This has been referred to as the dichotomy of scale. As such, it is often necessary to downscale (starting at a coarse resolution relative to the spatial scale of interest — an increase in the spatial resolution) or upscale (starting with fine resolution components and constructing outputs over a coarser resolution — a decrease in the spatial resolution).

Many different methods have been developed to allow analyses of scales of spatial variation. These include fractal analysis, analysis of spatial structure using variograms, and wavelets (228). Throughout this book, spatial scale is a central concern. Chapter 7, in particular, deals with the characterisation of dominant scales of spatial variation while methods such as geographically weighted regression (Chapter 5), kernel estimation, and the K function (Chapter 8), for example, allow exploration of scales of spatial variation. In Section 6.4, some methods for changing from one set of areal units to another are discussed.

1.5.1   Spatial scale in geographical applications

There have been many published studies which detail attempts to characterise spatial variation in some geographical property. In a social context, an individual's perception of an area is a function of their knowledge of the neighbourhood and such perceptions have, therefore, inherent scales (149). Likewise, to model appropriately some physical process it is necessary to obtain measurements that capture spatial variation at the scale of interest. For example, if the sample spacing is larger than the scale of spatial variation that is of interest, then models derived from these data may not be fit for the task in hand (149).

As noted previously, the concern here is with approaches that enable exploration of local differences in scales of spatial variation. A relevant study is that by Lloyd and Shuttleworth (253), who show that the relations between

mean commuting distance (as represented in the 1991 Northern Ireland Census of Population) and other variables differ markedly from place to place. In addition, the size of the areas over which these relations were assessed (that is, the size of the spatial kernel (see Section 2.4)) was varied. This study demonstrated regional variation in the spatial scale of the relations between these variables (see Section 5.8 for more details of this kind of approach).

1.6 Stationarity

The concept of stationarity is central in the analysis of spatial or temporal variation. In order to utilise the literature on local models for spatial analysis, an understanding of the key concepts is essential. The term stationarity is often taken to refer to the outcome of some process that has similar properties at all locations in the region of interest — it is a stationary process. In other words, the statistical properties (e.g., mean and variance) of the variable or variables do not change over the area of interest. A stationary model has the same parameters at all locations, whereas with a nonstationary model the parameters are allowed to vary locally. So, the focus of this book is on nonstationary models. There is little point in employing a nonstationary model if it offers, for example, no increased ability to characterise spatial variation or to map accurately a particular property. As such, it would be useful to be able to test for stationarity. However, testing for stationarity is not strictly possible, as discussed in Section 7.2. In the case of spatial prediction, for example, the performance of a stationary and a nonstationary model could be compared through assessment of the accuracy of predictions and thus the utility of a nonstationary approach considered.

1.7 Spatial data models

This book discusses local models that can be used in the analysis of properties that are represented by different kinds of data models. The key data models (or data types) are defined below. However, many analytical tools can be applied to properties represented using a range of different data models.

1.7.1 Grid data

Many operations used in the analysis of grid (or raster) data are, by definition, local. In particular, there is a wide range of methods used to analyse image

data that are based on the idea of a moving window. Some key classes of operations are outlined in Chapter 3. Given the importance of remotely-sensed imagery in many applications areas, grid operations are a particular concern in this book.

1.7.2   Areal data

A frequent concern with areal data (e.g., areas dominated by a certain soil type, or population counts over particular zones) is to ascertain the neighbours of an area. That is, with what other areas does a particular area share boundaries, and what are the properties of these areas? Analysis of areal data is discussed in Chapters 4 and 5, while reassigning values across different areal units is discussed in Chapters 6 and 7. The centroids of areas may, in some contexts, be analysed in the same way as geostatistical data, as outlined next.

1.7.3   Geostatistical data

A typical geostatistical problem is where there are samples at discrete locations and there is a need to predict the value of the property at other, unsampled, locations. An example is an airborne pollutant sampled at a set of measurement stations. The basis of geostatistical analysis is the characterisation of spatial variation, and this information can be used to inform spatial prediction or spatial simulation. Parts of Chapters 4, 5, 6, and 7 discuss methods for the analysis of these kinds of data (Chapter 7 is concerned with geostatistical methods specifically).

1.7.4   Point patterns

Most of the models described in this book are applied to examine spatial variation in the values of properties. With point pattern analysis the concern is usually to analyse the spatial configuration of the data (events), rather than the values attached to them. For example, the concern may be to assess the spatial distribution of disease cases with respect to the total population. The population density is greater in urban areas than elsewhere and the population at risk is spatially varying — this must, therefore, be taken into account. The focus of Chapter 8 is on methods for assessing local variations in event intensity and spatial structure.

1.8   Datasets used for illustrative purposes

A variety of applications are mentioned to help explain different techniques, and specific case studies are also given to illustrate some of the methods. The application of a range of techniques is illustrated, throughout the book, using six main sets of data with supplementary datasets used in particular contexts. These are (with data model in parenthesis):

1. Monthly precipitation in Great Britain in 1999 (geostatistical).

2. A digital elevation model (DEM) of Great Britain (grid).

3. Positions of minerals in a slab of granite (point pattern).

4. A Landsat Thematic Mapper (TM) image and vector field boundaries for a region in south eastern Turkey (grid, areal).

5. A digital orthophoto of Newark, New Jersey (grid).

6. Population of Northern Ireland in 2001 (two datasets; areal).

The datasets are described below. In addition, two other areal datasets used in particular contexts are detailed in Section 1.8.7.

1.8.1   Monthly precipitation in Great Britain in 1999

The data are ground data measured across Great Britain under the auspices of the UK Meteorological Office as part of the national rain gauge network. The data were obtained from the British Atmospheric Data Centre (BADC) Web site*. Daily and monthly data for July 1999 were obtained and combined into a single monthly dataset. Only data at locations at which measurements were made for every day of the month of July were used. The locations of observations made during July 1999 are shown in Figure 1.2, and summary statistics are as follows: number of observations = 3037, mean = 38.727 mm, standard deviation = 37.157 mm, skewness = 2.269, minimum = 0.0 mm, and maximum = 319.0 mm. The smallest values were two zeros and the next smallest value was 0.5 mm. Elevation measurements are also available for each of the monitoring stations and, since precipitation and elevation tend to be related over periods of weeks or more, these data are used to demonstrate the application of multivariate techniques.

In parts of the book, a subset of the data is used to illustrate the application of individual techniques. Two subsets were extracted: one containing five observations (used to illustrate spatial prediction) and one containing 17

---

*www.badc.rl.ac.uk

Precipitation observations

N

0   100   200   Kilometres

FIGURE 1.2: Measurements of precipitation for July 1999.

observations (used to illustrate local regression techniques). The full dataset is then used to demonstrate differences in results obtained using alternative approaches. The data are used in Chapters 5, 6, and 7.

1.8.2 A digital elevation model (DEM) of Great Britain

The relevant section of the global 30 arc-second (GTOPO 30) digital elevation model (DEM)[†] was used. After conversion from geographic coordinates to British National Grid using a nearest-neighbour algorithm (see Mather (264) for a summary), the spatial resolution of the DEM (Figure 1.3) was 661.1 metres. The data are used in Chapter 3 and to inform analyses in Chapters 5, 6, and 7.

1.8.3 Positions of minerals in a slab of granite

A regular grid with a 2 mm spacing was placed over a granite slab and the presence of quartz, feldspars, and hornblende was recorded at each node of the grid. The dominant mineral in each grid cell was recorded. Feldspar and mafic minerals such as hornblende occur in clusters and give an indication of magma chamber crystal settling. Therefore, the degree of clustering or dispersion aids interpretation of a rock section. The slab is illustrated in Figure 1.4. The dataset comprised 1326 point locations (37 columns by 36 rows, but there were no observations at 6 locations since none of the three selected minerals were present at those 6 locations). The data cover an arbitrarily selected part of the surface of a granite slab and, therefore, the study region is an arbitrary rectangle. The locations identified as comprising mafic minerals are shown in Figure 1.5. The data are used in Chapter 8 to illustrate methods for the analysis of point patterns[‡].

1.8.4 Landcover in Turkey

The dataset is a Landsat Thematic Mapper (TM) image for 3rd September 1999. It covers an area in the south-eastern coastal region of Turkey called Cukurova Deltas. The area has three deltas that are formed by the rivers Seyhan, Ceyhan, and Berdan. The study area lies in the centre of this region, and covers an area of approximately 19.5 km by 15 km (29,250 hectares). The image was geometrically corrected and geocoded to the Universal Transverse Mercator (UTM) coordinate system using 1:25,000 scale topographic maps. The image was then spatially resampled to a spatial resolution of 25 m using a nearest-neighbour algorithm. The dataset is described in more detail in Berberoglu et al. (39). In that paper, the focus was on classification

---

[†]edcdaac.usgs.gov/gtopo30/gtopo30.html

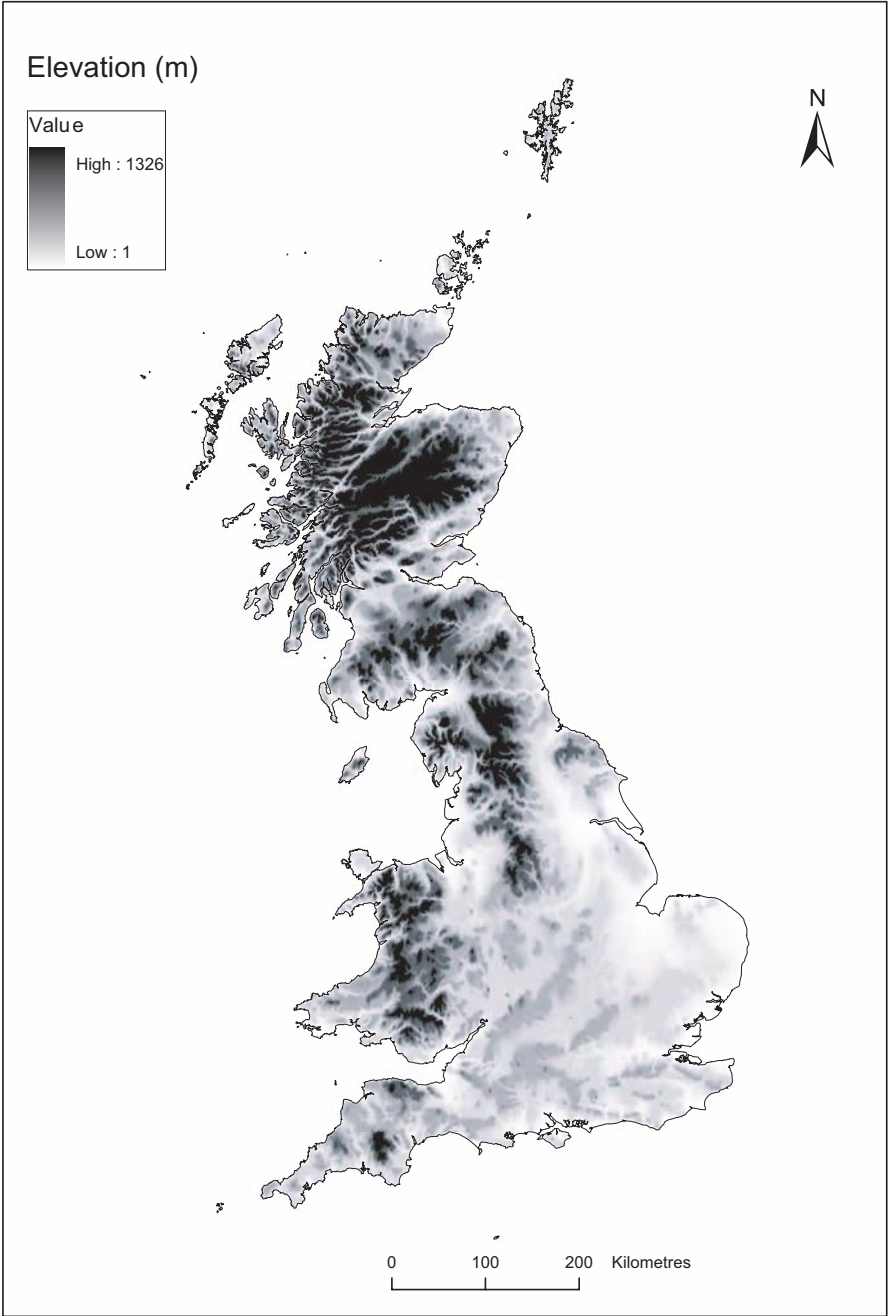[‡]The data were provided by Dr. Alastair Ruffell of Queen's University, Belfast.

FIGURE 1.3: DEM of Great Britain; spatial resolution of 661.1 m.

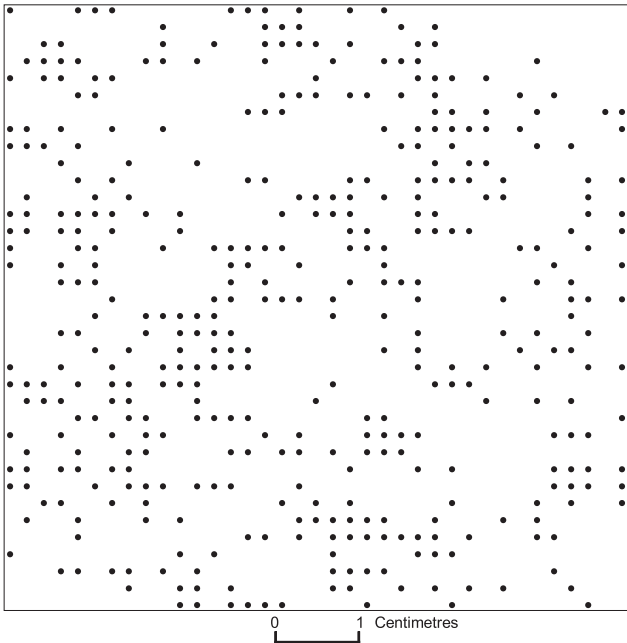FIGURE 1.4: Slab of granite, width = 76mm.

FIGURE 1.5: Mafic mineral locations in a slab of granite.

of land covers. In the book, the first principal component (PC1) of six wavebands (bands 1–5 and 7; Figure 1.6) is used to illustrate a variety of local image processing procedures. A related dataset is vector field boundary data digitised from Government Irrigation Department (DSI) 1:5000 scale maps. The image of PC1 is used in Chapter 3. The vector boundary data are used in Chapter 4, in conjunction with the values from the image, to illustrate measures of spatial autocorrelation[§].
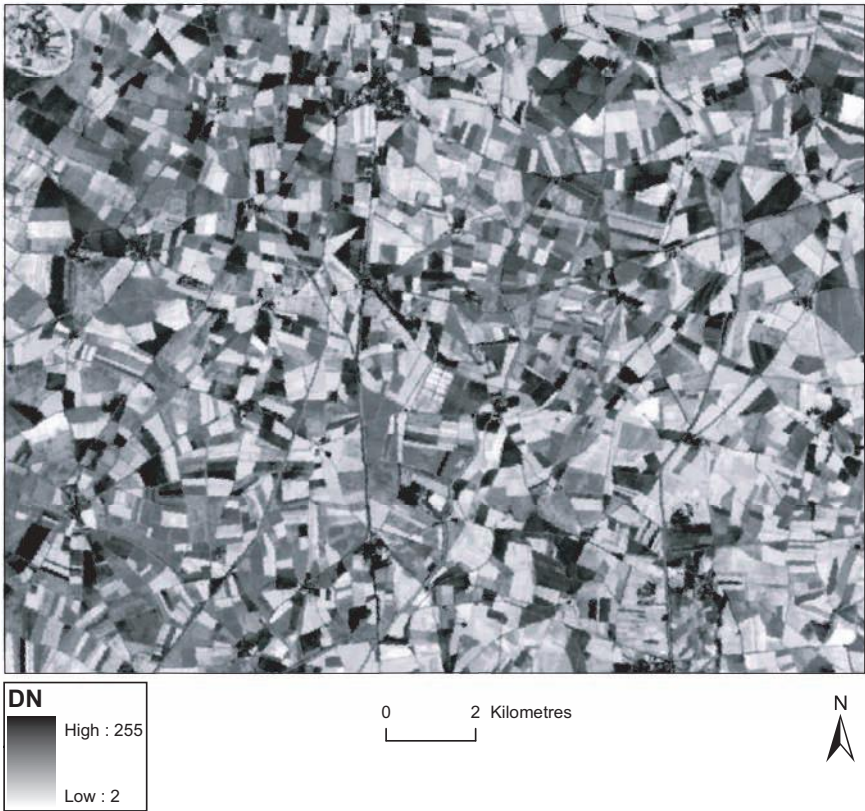
FIGURE 1.6: First principal component of six wavebands of a Landsat TM image. DN is digital number.

[§]Dr. Suha Berberoglu, of the University of Cukurova, provided access to the processed data.

1.8.5 Digital orthophoto of Newark, New Jersey

A digital orthophoto quadrangle (DOQ) ¶ of part of Newark, New Jersey, was acquired and a subset was extracted for the illustration of the discrete wavelet transform in Chapter 3. The image subset is shown in Figure 1.7.
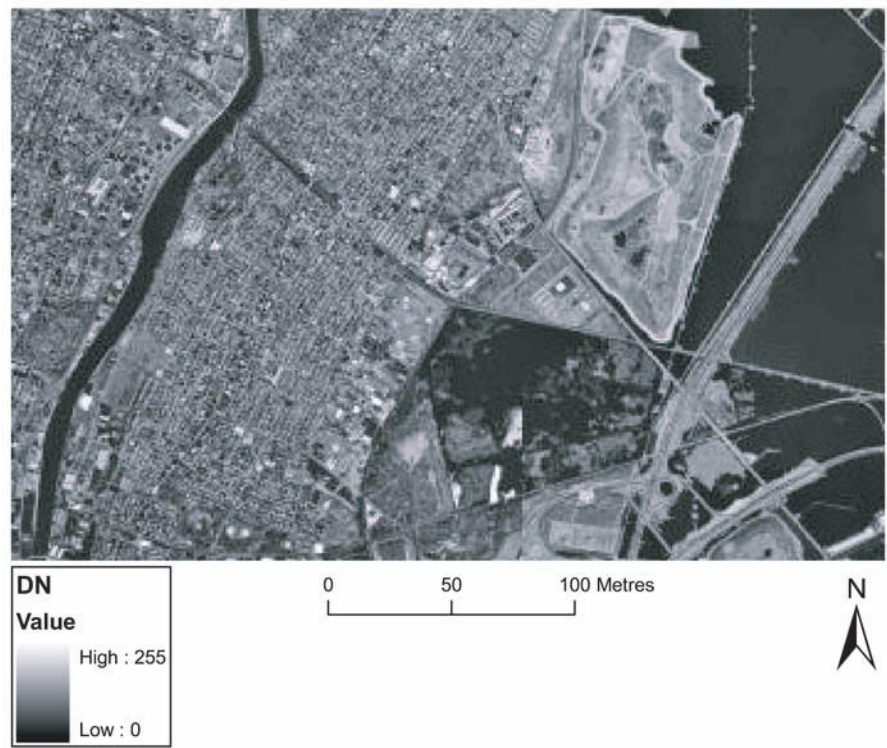

FIGURE 1.7: Digital orthophoto of Newark, New Jersey. Image courtesy of the US Geological Survey.

1.8.6 Population of Northern Ireland in 2001

The datasets detailed in this section represent the human population of an area. The data are population counts made as a part of the 2001 Census of Population of Northern Ireland. Of the two datasets used in the book, the first had counts within zones called Output Areas. There are 5022

¶See http://online.wr.usgs.gov/ngpo/doq/

Output Areas, with populations ranging between 109 and 2582 and a mean average population of 336. Population counts are shown in Figure 1.8. Population densities, as shown in Figure 1.9, are a sensible way of visualising such data and urban areas like Belfast obviously become more visible using such an approach. The data are used in Chapter 6 to illustrate areal interpolation from zone (that is, Output Area) centroids to a regular grid. The centroids are population-weighted, and the large majority of centroids were positioned using household counts (using COMPASS [COMputerised Point Address Service], a database of spatially-referenced postal addresses) with some manual adjustments where centroids fell outside of their Output Area because the zone was unusually shaped (for example, a crescent shape). These data are available through the Northern Ireland Statistics and Research Agency[‖].
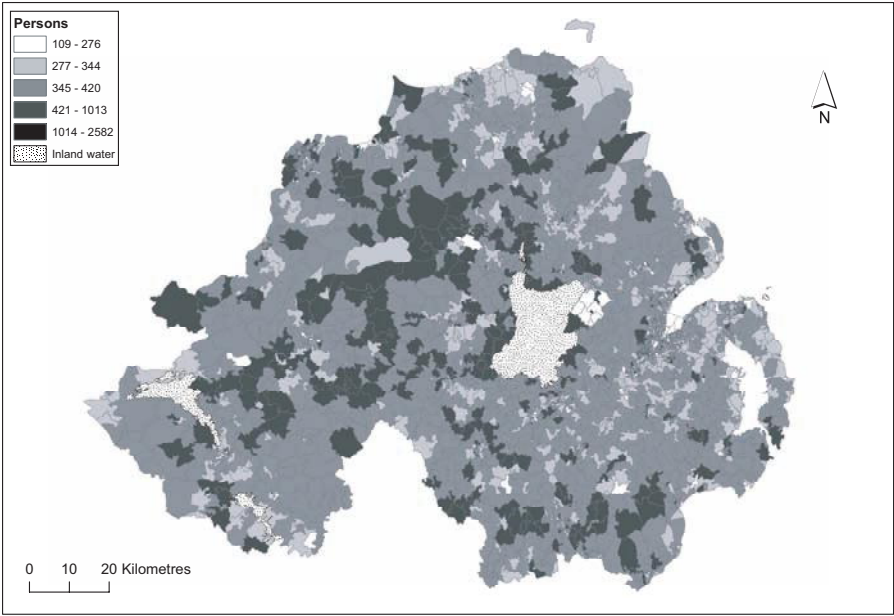

FIGURE 1.8: Population of Northern Ireland in 2001 by Output Areas. Northern Ireland Census of Population data — © Crown Copyright. Reproduced under the terms of the Click-Use Licence.

The second population dataset utilised in this book relates to the percentages of people in Northern Ireland who, in 2001, were Catholic or non-
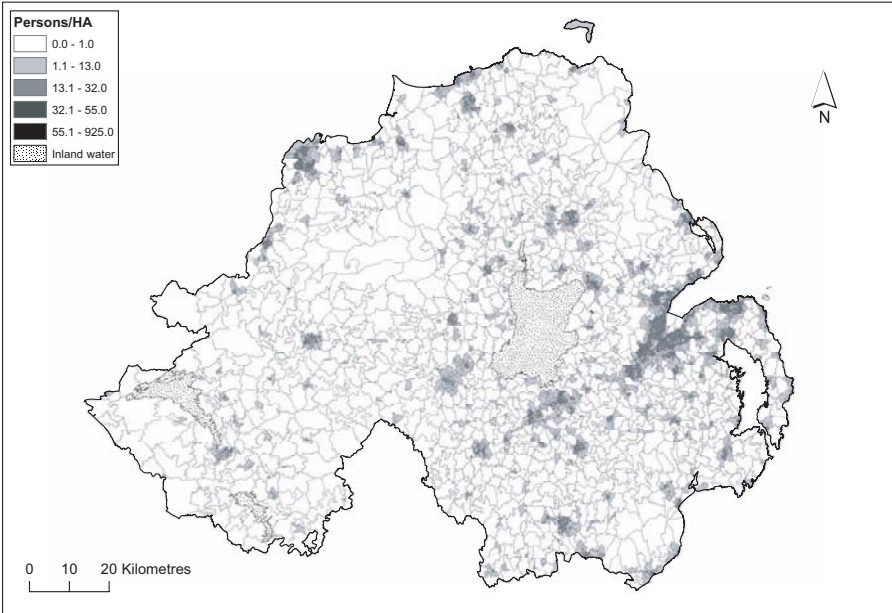
[‖]www.nisra.gov.uk

FIGURE 1.9: Population of Northern Ireland per hectare in 2001 by Output Areas. Northern Ireland Census of Population data — © Crown Copyright. Reproduced under the terms of the Click-Use Licence.

Catholic (i.e., mostly Protestant) by community background ('religion or religion brought up in'). The percentages were converted to log-ratios given by ln(CathCB/NonCathCB), where CathCB is the percentage of Catholics by community background and NonCathCB is the percentage of non-Catholics. The mapped values are shown in Figure 1.10. The rationale behind this approach is detailed by Lloyd (242). The data are for 1km grid squares, one of several sets of zones for which counts were provided from the Census (see Shuttleworth and Lloyd (331) for more details).

1.8.7 Other datasets

Two other datasets are used to illustrate particular methods. Both of these datasets are well-known and have been described in detail elsewhere. The first dataset represents residential and vehicle thefts, mean household income and mean housing value in Columbus, Ohio, in 1980. The data are presented by Anselin (9), and the zones used to report counts are shown in Figure 5.4 (page 112). The second dataset comprises births and cases of sudden infant death syndrome (SIDS) in North Carolina. The data were used by Symons et al. (355) and Bivand (45) provides a description. These data are used in Section 8.13 to demonstrate a method for detection of clusters of events.
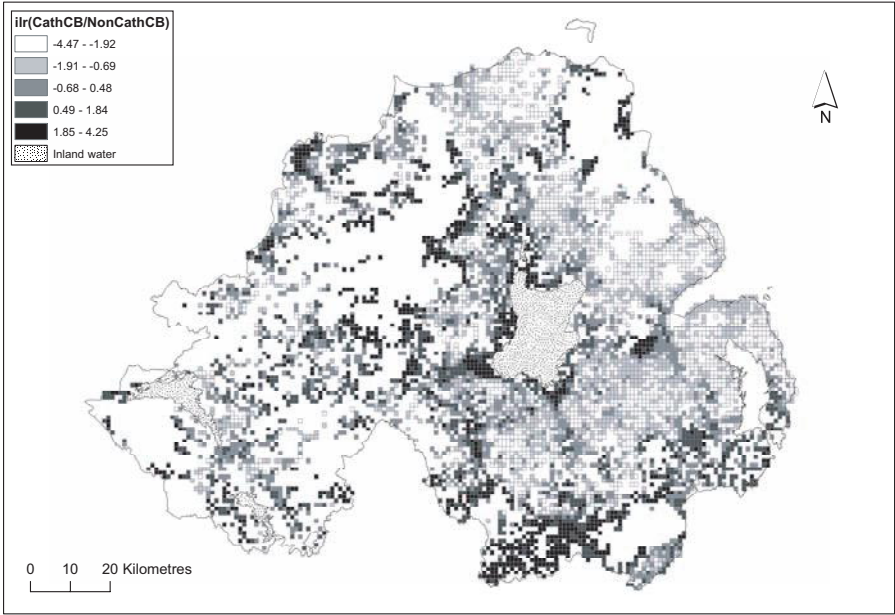
FIGURE 1.10: Log-ratio for Catholics/Non-Catholics in Northern Ireland in 2001 by 1km grid squares. Northern Ireland Census of Population data — © Crown Copyright.

1.9 A note on notation

Some symbols are used in the text to mean different things, but consistency has been the aim between chapters where possible where this does not conflict with well-known use of symbols in particular situations. Examples of symbols defined differently include the use of h to mean a scaling filter in the context of wavelets (Chapter 3), and as a separation distance in the case of geostatistics (Chapter 7). In Chapter 5 spatial coordinates are given with u, v as x is used to denote an independent variable. Elsewhere the more conventional x, y is used. Two forms of notation are use to represent location i. These are the subscript i (e.g., z_i) and the vector notation $\mathbf{s}_i$ (e.g., $z(\mathbf{s}_i)$). The selection of one or the other is based on clarity and convenience. It is hoped that readers will find the meaning to be clear in each case following definitions given in individual chapters, although in most cases notation is consistent between chapters.

1.10   Overview

At a very broad level the book discusses methods that can be used to analyse data in two key ways. That is, when the concern is with the analysis of (i) spatial variation in the properties of observations of one or more variables, and (ii) spatial variation in the configuration of observations (for example, are observations more clustered in some areas than in others?). Most of the book focuses on (i) but (ii), in the form of point pattern analysis, is discussed in Chapter 8.

Chapter 2 discusses some ways of adapting to local variation. In Chapter 3, the focus is on local models for analysing spatial variation in single variables on grids, Chapter 4 is concerned with analysis of single variables represented in a variety of ways, while in Chapter 5 the concern is with local models that can be used to explore spatial variation in multivariate relations between variables. Chapter 6 outlines some methods for the prediction of the values of properties at unsampled locations; techniques which enable transfer of values between different zonal systems and from zones to points are also discussed. Chapter 7 illustrates geostatistical methods for analysing spatial structure and for spatial prediction. Chapter 8 is concerned with the analysis of spatial point patterns. Chapter 9 summarises the main issues raised in the previous chapters and brings together some key issues explored in this book.