2

Concepts for Modelling Spatial and Spatial-Temporal Data: An Introduction to "Spatial Thinking"

2.1 Introduction

A "spatial" database or a "spatially-referenced" database is one where each individual datum in that database is attached to a location. The implication is that we know the places on the earth's surface to which the data refer. When spatial data are also recorded for a sequence of times (instants in time or time periods, for example over a week, a month or a year), we refer to a "spatial-temporal" database. Knowledge of the spatial or spatial-temporal location of each datum is part of the information content of that database and is important both for the statistical analysis of that data and the interpretation of results.

Many areas of the social, economic and public health sciences analyse spatial and spatial-temporal data. Through organizational and technological advances it has become possible to access ever larger volumes of such data and to be ever more precise in the geographical assignment process and ever more prompt in the temporal updating process. In this chapter we describe the key concepts that inform the analysis of spatial and spatial temporal data – sometimes referred to as "spatial and spatial-temporal thinking" or "spatial thinking" for short.

Core geographical concepts that are fundamental to this mode of thinking include: the location, distribution and patterning of objects and events (attributes) in geographical space and space-time; places and regions and their particular attributes; the interdependencies that exist between places and regions and the various types of spatial-temporal interaction processes between places at different times; the geographical extent of an area or domain of study; and the scale (or level of generalization) at which we observe that domain. Clifford et al. (2009) discusses core geographical concepts which ground the academic discipline of geography, and these concepts are also important to spatial thinking, which is about developing a geographical "habit of mind" when scientifically approaching certain types of problems and issues.[1] We note here that although this book does not deal with Geographical Information Systems (GISs), such systems have an important contribution to make to the modelling activities that this book deals with. So in the appendix to this chapter we briefly discuss them.

Attributes are observed and measured, and our aims include one or more of: providing a *description* of the spatial or spatial-temporal variation of attributes; *explaining* the observed variation in one or more attributes by identifying spatial associations and developing theory; and making *predictions* or *forecasts* about future spatial patterns with the aim of

[1] See: https://blogs.esri.com/esri/gisedcom/2013/05/24/a-working-definition-of-spatial-thinking/.

informing *policy*. Our focus is the study of patterns and processes in geographical space, arriving at conclusions based on scientific principles. In this book, to achieve these aims, we apply statistical theory and method. In the previous chapter we considered some of the special challenges and opportunities that the statistical modelling of spatial and spatial-temporal data gives rise to. In this chapter we focus on the geographical and spatial concepts that are frequently called upon and which inform how statistical method is applied. Taken together, Chapters 1 and 2 are designed to promote a *fusing of two reasoning processes* – the spatial thinking and the statistical thinking. Both are essential for the effective implementation, by social, economic and public health scientists, amongst others, of statistical techniques applied to spatial and spatial-temporal data (Haining, 2014 and Waller, 2014).

2.2 Mapping Data and Why It Matters

Knowing and recording where objects are located or when and where events happen in geographical space-time is important for research in many academic disciplines as well as for those concerned with more immediate, practical concerns such as the design and delivery of public or private sector policies. We use the term "object", in the geographical information science sense, to refer to an entity (a house, a road, a city) that endures through some extended, though not necessarily long, period of time (Warboys and Hornsby, 2004). In the short to medium term at least, the location of an object is fixed, for example the location of a business, a retail outlet, a landfill site, a hospital or clinic, an archaeological artefact, a transit terminal, a line of communication (road, railway, river). Recording the geographical location of objects in particular classes at a point in time means such data can be organized and visualized in map form. Mapping becomes a starting point for interrogating the data both in terms of characterising and making comparisons between places.

We use the term "event", again drawing on the geographical information science literature, to refer to an entity that happens and is then gone (Warboys and Hornsby, 2004).[2] For example, an event can be a street robbery or a burglary that occurred on a particular day at a particular location within a city; a new case of malaria in a region; or an individual diagnosed as suffering from a chronic respiratory condition or some limiting long-term illness. Recording and mapping the space-time location of events (new cases of a disease, instances of household burglary) may provide an important starting point for developing a better understanding of the underlying processes responsible for the observed outcomes. Combining real-time event data, such as data on disease occurrences, and current object data, such as the location of clinics and other support services, as well as areas with a known increased risk of an outbreak, provides a starting point for designing response strategies in the case of a new outbreak. Combining event data on street crime with object data on the location of transport routes, transit terminals and certain types of retail outlets may suggest strategies for crime prevention and/or responding to local crime hotspots.

The study of the attributes that attach to objects or events and which describe their properties is an aspect, to a greater or lesser degree, of many academic fields within the arts, humanities and the social, economic and public health sciences. In the environmental sciences, mapping the physical, environmental and human activity of an

[2] Warboys and Hornsby (2004) refer to objects as "continuant" entities and events as "occurrent" entities. Events and processes, they argue, essentially "speak to the same idea".

area helps in meeting environmental challenges and threats. The study of particular places involves not just cataloguing the attributes of different geographically defined areas, it also involves developing an awareness of the role different processes, operating at different geographical scales, have in shaping the characteristics of each place and how those characteristics may in turn impact on or modify larger scale processes. Thinking spatially means thinking through the implications of where objects and events are located, why objects are located where they are, why events happen where they do and why the changes we observe over time (perhaps in response to a common stimulus such as a national economic recession, a regional crime wave or a threat to health) may not be the same everywhere. In the fields of policy-making and policy evaluation, to think spatially means to think about how the impacts associated with (say) a national policy vary between areas as well as thinking about the need for geographic targeting (distinguishing between areas in terms of whether or not they should receive financial support) or geographic tailoring (making adjustments to policies at the local level to reflect local conditions). Thinking spatially, as opposed to non-spatially, usually involves presenting data in particular ways, asking different kinds of questions and using different kinds of methods to answer those questions. For a wide range of examples, see Lawson et al., 2016; Beck et al., 2006; Clarke, 1997; Cromley and McLafferty, 2012; Cuzick and Elliott, 1992; Gamarnikow and Green, 1999; Martin et al., 2015; Tita and Radi, 2010; and Weisburd et al., 2012.

The map is an essential presentational tool for any science that works with data with a spatial or geographical reference (or "geo-reference"). The geo-reference of a single piece of data identifies the location on the earth's surface to which the datum refers. The location may be a point co-ordinate that defines the position of the datum on east-west (x) and north-south (y) axes, it may be a line such as a transport route or it may be an area/polygon such as a census tract or region. The top left-hand diagram in Figure 2.1 shows a dot map recording the location of some event – a household burglary for example or a case of street robbery. All the events have been *binned* (aggregated) into a single time interval. Such data may be spatially binned, that is aggregated into small spatial units as shown on the top right-hand diagram in Figure 2.1, where the greyscale goes from white (0 cases) to black (5 or more cases). Much of the economic, social and demographic data that social scientists work with (and which we will work with in this book) are reported through irregular spatial bins, such as census tracts. On the other hand, much environmental data, particularly remotely sensed data, are reported using pixels as shown on the top right-hand diagram in Figure 2.1. How much spatial detail is lost by binning depends on the size of the bins (map scale or resolution).

These data can be refined by examining the time (*t*; a point in time or interval of time) when each event occurred. Such data can be represented by a space-time plot in which the location of any event is captured by x and y co-ordinates (two horizontal axes) and time, *t*, is captured by a third (vertical) axis. The data can then be aggregated or "binned" into space-time units as shown in the bottom diagram in Figure 2.1, where a time has been attached to each of the events shown in the top diagram in Figure 2.1. Each bin records the number of cases in that space-time bin. In the process of binning, one should also bear in mind that in some situations, the exact time of events may not be known. For example, residential burglaries often occur when the occupiers are away for a period of time. A patient may be diagnosed sometime after they first contracted a disease. The significance of this uncertainty depends on the interval of time used for the purposes of data analysis and the length of time of the uncertainty. With the demand for greater temporal precision to meet the operational needs of, say, police forces or health services, some adjustment to

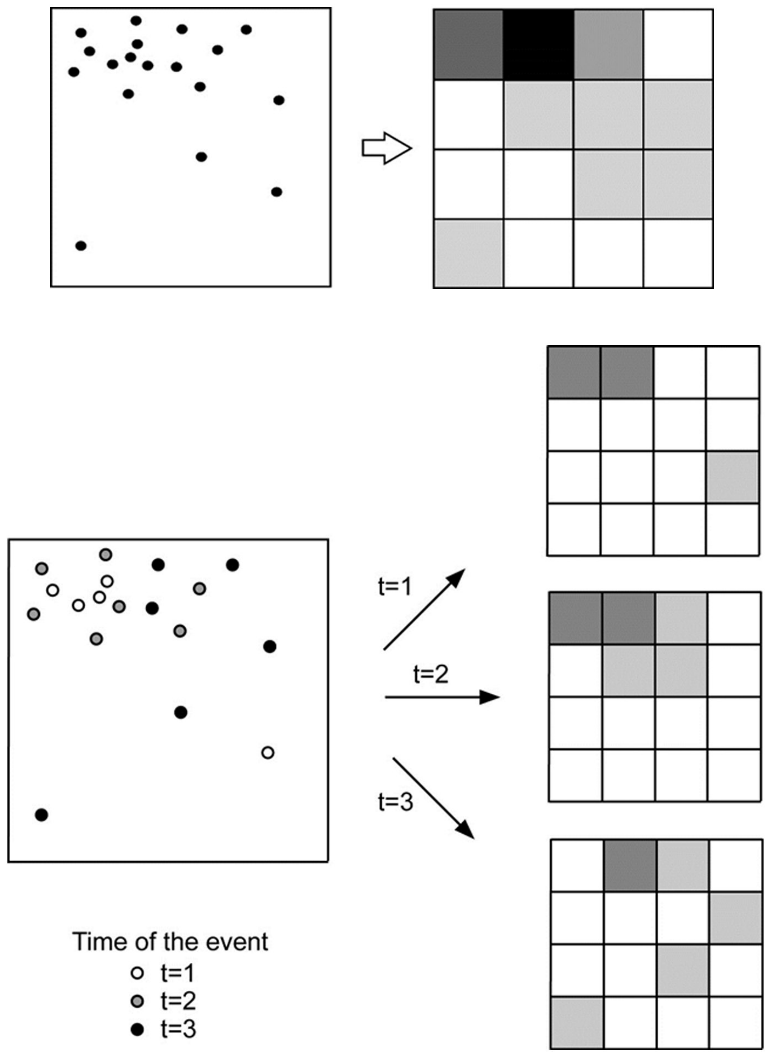

FIGURE 2.1
Top: event data showing location binned into small, regular spatial units. Bottom: if the time of the event is recorded then event data can be binned into space-time bins.

the binning process to allow for this form of uncertainty may be necessary. In the case of crime recording, see for example Ratcliffe and McCullagh, 1998 and Ratcliffe, 2002.

A purely spatial analysis examines each one of the (horizontal) slices through the space-time data. A space-time analysis, considering all the space-time data together, allows us to show change over time and to suggest the presence of space-time interaction, where events occurring at a particular space-time bin are correlated with events occurring in nearby space-time bins – a topic that we shall examine closely in Part III of the book. The bottom diagram in Figure 2.1 might suggest some localized diffusion process associated with the event, an observation that would not be evident from the top diagram in Figure 2.1 where the binning of events in time is too coarse.

A space-time multivariate dataset, where many attributes are recorded, can be thought of as a "space-time-attribute" data cube in which one axis identifies location ("where"), the

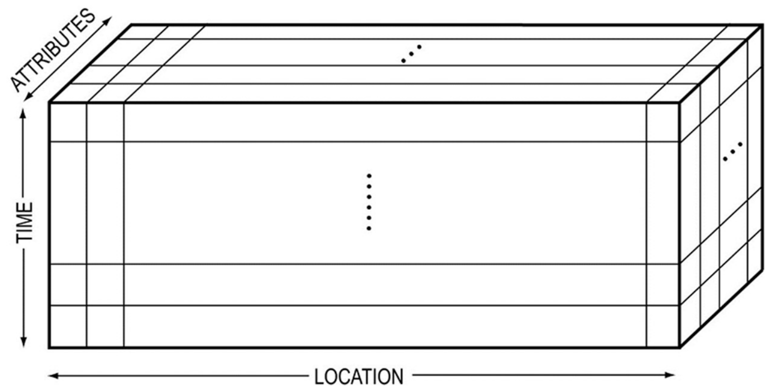

FIGURE 2.2
The space-time-attribute data cube. Each small "box" inside the cube identifies an attribute value associated with a single area at a point in, or during an interval of, time.

second axis identifies time ("when") and the third axis specifies the set of attributes that are recorded for each space-time unit of observation ("what"). Each individual bin records the value of a specified attribute in a specified space-time interval, such as the number of cases of a particular crime or disease occurring in a specified area during some interval of time (see Figure 2.2). The representations in Figure 2.1 retain the spatial and temporal relationships between the data points for the purpose of visualising the data. By contrast, the representation given by Figure 2.2 does not retain the spatial relationships between the spatial bins, and if these are important, as they usually are, we need another vehicle for storing this information. In Chapter 4 we discuss the connectivity or weights matrix, ***W***, used for this purpose amongst others.

Maps are central to the methods of this book. A map can be defined as a symbolic depiction of the arrangement or distribution of entities (objects or events) and their attribute values across some area. The emphasis of a map is the spatial relationships between those entities and their attribute values. A map may be printed or it may be held in digital form as in a Geographical Information System (GIS), where the data are held as a series of layers where each layer can be independently switched "on" or "off", giving rise to the possibility of multiple different views. If values are recorded over time so that each layer refers to a different time window, displaying such data in time order can be used to show, for example, urban expansion (or contraction), network evolution or the spread of some infectious disease.

From the perspective of statistical analysis, we can think of a map as both a presentation and a scientific visualization tool. As a presentation tool the map is used to provide the reader with a geographical summary or display, either of the original data or of important results arising from some analysis of that data. As a visualization tool the map is used in exploratory data analysis (see Chapters 6 and 13) in which the data analyst displays and experiments with their data in map form in order to gain insights into data characteristics and data relationships. A single map referring to a defined interval of time or snapshot gives us a picture of the "here and now" of what is happening and can suggest associations between attributes. Maps over time reveal how that "here and now" changes and may indicate the presence of space-time interaction. Maps that show movement between nodes, for example if the data refer to flows of goods or people or information between urban areas over a period of time, indicate how places interact with one another economically or socially.

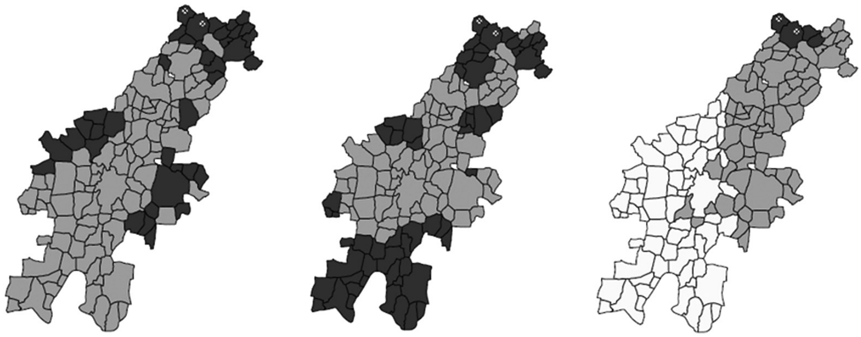

FIGURE 2.3
New cases of malaria reported at the village level for Karnataka, India. From left to right: August–September 2014 (warm-wet season); October–November 2014 (warm-dry season); December 2014–January 2015 (cool-dry season). On each map, shading goes from dark (high risk) to light (low risk). Villages identified as malaria "hotspots" are marked (⊗) (Shekhar et al., 2017). For scale: From North East to South West is approximately 80km.

A sequence of maps covering a series of time points or intervals allows us the possibility of investigating process. If geographical data are available over time then not only can we observe changes over time, we can observe the way a process unfolds in space and time – for example, whether events occurring close together in space also occur close together in time – and vice versa. This is important in many areas of research, including studying the spread of an infectious disease (see Figure 2.3); how the local geography of crime risk varies over time; the geography of the take-up over time of an innovation or new technology; the impact over space and time of a new targeted policy (for illustration see: Bowers and Johnson, 2005; Gao et al., 2013; Grubesic and Mack, 2008; Hagerstrand, 1967; Newton and Felson, 2015; and Qi and Du, 2013).

2.3 Thinking Spatially

2.3.1 Explaining Spatial Variation

To "think spatially" is not only to be concerned with the study of places. Maps reveal more than that the numbers of events or the presence/absence of objects vary from place to place. Our attention may be drawn to spatial or geographical structures and relationships in one or more attributes, which may be important in terms of the causes and consequences of the spatial variation that we observe.[3]

(i) *Variation in attribute values relative to significant fixed points or areas.* Census tracts within a city will vary in terms of their distance from the city centre. Levels of anti-social behaviour by census tract might show a declining trend with increasing distance from the centre. Street robberies may be observed to cluster around bars. Vandalism hot spots are often found at or near public transit stops. Certain health indicators for

[3] Maps of several attributes may draw attention to relationships between different attributes. Scatterplots and added variable plots are also useful, but such tools may need to be adapted if we want to explore the relationship between attributes across a set of spatial units (see Chapter 3 and Chapter 6).

a group of settlements might correlate with their distance and direction from possible sources of environmental risk. Archaeological artefacts may be found in specific locations, such as close to and with views of coastal areas that imply the sea may have some symbolic or practical significance. For a range of examples see: Ceccato, 2013; Elliot et al., 2009; Fisher et al., 1997; Gallup et al., 1999; and Venables, 1999.

(ii) *Spatial gradients in attribute values between adjacent areas.* The close juxtaposition of neighbourhoods that show a marked contrast in terms of their affluence or social cohesion may have implications for a range of social, economic and health-related outcomes. Consider the following thought experiment. Two communities are identical in terms of their internal or "place" characteristics, but they differ in terms of the characteristics of their immediate neighbouring communities. Outcomes might differ between the two areas because of interaction effects taking place across their respective boundaries – interaction that is triggered by the attribute gradient. There is evidence, for example, that communities that present very similar socio-economic characteristics may differ in terms of health outcomes or crime rates depending on the social and economic characteristics of adjacent communities. For some examples see: Block, 1979; Bowers and Hirschfield, 1999; Dow et al., 1982; Gatrell, 1997; and MacLeod et al., 1999.

(iii) *Macro-spatial configuration.* The spatial structure of a society and its economy may have implications for social and economic outcomes. The presence of significant income inequality in a society may have implications that are influenced by the geography of that inequality. Consider the following thought experiment. Imagine two cities identical in all important respects and in particular with equivalent numbers of people living in different income bands. In one of the two cities a significant majority of the population living in poverty live in one large ghetto, whilst in the other city the same population is more dispersed, perhaps as a consequence of the city's public housing policy. Both the city level health profiles and crime rates in those two cities as well as their micro-geographies may be influenced by the difference in the macro-spatial configuration of the population living in poverty. Recent literature in economics has drawn attention to the benefits of spatial clustering of economic activity. Geographical proximity may confer benefits on firms that cluster in the form of positive externalities that derive from their geographical context. These externalities include benefits in terms of factor conditions for firms within the cluster (e.g. labour availability), demand conditions (e.g. access to markets), firm strategy (e.g. knowledge acquisition) and the presence of related and supporting industries (e.g. proximity to supply industries). For some examples see: Kahn et al., 2006; Krugman, 1996, 1998; Porter, 1998; Sampson et al., 1997; Snow and Moss, 2014; Sparks et al., 2009; Szwarcwald et al., 2000; and Wilson, 1997.

Recognizing spatial relationships in the study of events is also important. That is because the processes that underlie the events we observe are not constrained by the spatial units in terms of which the data are reported. People move about in their daily lives and over their life course, and they have information about what is happening elsewhere – never more so than in an age of social media. Environmental influences can be carried from place to place by movements of air and water. The spread of influences from one place to another are sometimes referred to as "spatial spill-over (or spillover) effects" (see Sections 1.3.2.3 and 1.4.4). When considering data reported over a set of well-defined spatial units, for example local government areas or health districts, spillover effects are important because decisions made (e.g. resources allocated and outcomes arising) in one area may have consequences for other areas. But spillover effects also need to be considered when the spatial units are artificial constructions, such as census tracts. The outcomes we observe in one place may in part be a consequence of circumstances elsewhere. How a spillover effect impacts on a study depends on the relationship between the scale at which the spillover process is operating in the real world and the size or scale of the spatial units through which outcomes are observed. If spatial spillover processes in the real world are highly localized but the spatial units are large, then the contribution of any spillover effect may result in an internal multiplier effect with numbers of cases (where cases are present) inflated by the localized spillover effect. So a map of case counts might show large numbers of areas with zero or small counts, whilst there are other areas with large counts. If spillover processes operate at a scale that is larger than the spatial units of analysis, this may then lead to spatial correlation effects between adjacent units. As a result, areas with high (low) case counts tend to be located close to other areas also with high (low) case counts. Spillover effects may operate at many scales, so that both characteristics might be evident in a dataset. For some examples, see Brett and Pinkse, 2000; Case et al., 1993; Brueckner, 2003; Hanes, 2002; Hassett and Mathur, 2015; and Revelli, 2002.

The differences that are observed from place to place will be a consequence of many different factors. Differences, for example in health outcomes and crime rates, reflect differences in the composition of the population. With all other things being equal, a place with a larger proportion of older people will tend to show worse morbidity statistics than places with a younger population, but an area with a larger proportion of younger people (especially young males) may have higher rates of certain types of offending. Because people move around in the course of a day, a week, a year and over the course of their lives, morbidity differences in the case of chronic conditions reported by small area are likely to depend on the life histories of the population living in the area. In the case of infectious diseases, on the other hand, morbidity differences will depend on recent movement behaviours in the population (for work or for leisure).

Differences between places will reflect the presence or absence of other attributes – the presence of subway stations or bars in a neighbourhood on criminal assaults and anti-social behaviour. Differences between places may also reflect group or ecological scale effects associated with the population. For example, places with higher levels of social cohesion tend to experience lower rates of crime. Differences in population health profiles may reflect differences in area level environmental characteristics, such as the amount and location of green space in an urban area, levels of air pollution or natural radiation.

To illustrate some of the ideas described above, the reader is referred again to the first three examples in Section 1.3.2 where regression is used to model the number of deaths due to stroke in Sheffield, the number of cases of sexual assault by small areas in Stockholm and the spatial spillover effects in a retail price setting. In each case, spatial thinking, as defined here, has either played a role in model specification (including a spatially-lagged covariate to measure the spatial aspects of NO_x exposure in modelling stroke deaths; including covariates to measure accessibility in order to explain variation in the number of cases of sexual assault) or underpinning the purpose of the analysis (testing for price competition effects between retail sites).

2.3.2 Spatial Interpolation and Small Area Estimation

The previous examples of spatial thinking refer to how this mode of thinking may help us to better *understand or explain* what we observe in a response or outcome variable. There is a further dimension to what we mean by the phrase "spatial thinking" that stems from the observation that data values near together in geographical space tend to be more alike than data values that are further apart – the property of spatial dependency we have already discussed in Chapter 1. This property of spatial data is the basis for a number of statistical methods, including spatial interpolation and small area estimation, which broadly speaking are methods that enable us to tackle specific types of data challenges.

Spatial interpolation methods are used to estimate (or predict) *data values* at unsampled locations on a continuous surface from which a map of spatial variation of the attribute can be constructed. These methods can also be used to estimate data values which are missing in the case of an area partitioned into small spatial units such as census tracts. Spatial interpolation methods exploit the property of spatial dependence for the purpose of estimation (or prediction).

By way of illustration, Figure 2.4 shows six point sites on a continuous surface where sample values of the attribute y have been measured. These observed attribute values are labelled $y_1,\ldots,y_6$. The problem is to estimate the unknown value at s (y_s) and attach a measure of error to the estimate. It is assumed that because values in geographical space are correlated, if the sample points are close enough to s, then as we have discussed in Chapter 1, their observed values carry information about the value at s. The closer a sample point is to s, the more information its observed value is expected to carry about y_s.

Let $\hat{y}_s$ denote the estimated value of y_s. Let $\psi_1,\ldots,\psi_6$ denote six weights assigned to the six sample point values $y_1,\ldots,y_6$. Then, for this sample, let:

$$\hat{y}_s = \sum_{i=1}^{6} \psi_i y_i \qquad \sum_{i=1}^{6} \psi_i = 1.0$$

There are many ways in which the weights can be determined. In geostatistics, the weights, $\psi_1,\ldots,\psi_6$, are estimated using a valid spatial correlation function or variogram that describes the spatial dependency in the set of sample values $y_1,\ldots,y_6$, particularly over short distances. The weights also adjust downwards to lower the contribution of the observed values that come from a spatial cluster of sample points – for example, locations 4, 5 and 6 in Figure 2.4 (Isaaks and Srivastava, 1989). This is referred to as cluster downweighting. This sketch is an example of a spatial interpolation method, known as kriging, grounded in spatial thinking. Spatial interpolation is carried out using the locational information contained in the database. It is also an example of how spatial dependence can be exploited to borrow information, in this case, to estimate an unknown data value.

Suppose we now replace the sample points by a set of polygons that partition the area. That is, the data now refer to attribute values for a set of polygons such as census tracts. Some polygons have data values whilst others, for no special reason (perhaps the data were just lost, or never recorded) do not.[4] Subject to finding a satisfactory set of weights, this is a similar problem to that illustrated in Figure 2.4. So, spatial interpolation methods can be used to obtain estimates of the "missing values" (see for example Haining, 2003, p.164–174). Whilst we will not be covering geostatistics in this book, we shall return in Chapter 17 to look at some recent developments in geostatistics that are of interest when tackling problems where data values are attached to small, irregularly shaped polygons.

Small area estimation methods share common ground with spatial interpolation methods. These methods are employed to improve the precision of each small area parameter

[4] To be precise, we say the missing data values are "missing at random" as opposed to missing because they were deleted, perhaps for confidentiality reasons, or suppressed for some reason, perhaps because their values were large or small.

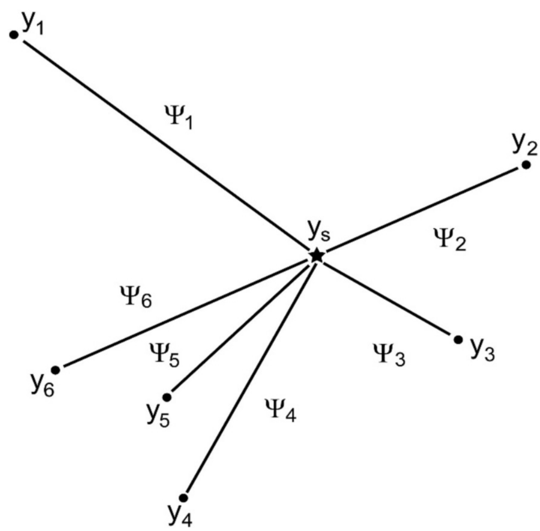

FIGURE 2.4
The weights, $\psi_1,\ldots,\psi_6$, are used to estimate, y_s, the data value at location s. Each of the observed values at locations 1 to 6, $(y_1,\ldots,y_6)$, are weighted by their corresponding $\psi_i \geq 0$, with the weights reducing in size as distance to location s increases.

estimate (e.g. the underlying disease or crime risk in each area) and/or the predicted outcome values for areas that have no data. We gave an example of this in Section 1.3.1 in Chapter 1. If an estimate for an area is produced using only the data points falling within that area, and if that area only has a small number of data points, the resulting estimate is likely to be highly unreliable (e.g. with a very wide or very narrow uncertainty interval). This can arise, for example, if the sampling scheme is not spatially stratified, so some tracts end up with few or no samples in them. In the extreme case where there are no sample values in a census tract, no parameter estimate can be obtained. It can also arise if the event is rare (e.g. a rare disease) and/or the at-risk population in each area is small. The underlying problem here is data sparsity (Section 1.2.3).

As we have seen in Sections 1.4.2 and 1.4.3.3, Bayesian hierarchical models tackle the above problem by embedding the spatial thinking that parameters are *spatially dependent*. Such a modelling assumption allows Bayesian hierarchical models to utilise ("borrow") the data from an area's neighbours in order to improve the estimation of its own parameter. In Chapter 8 we shall see the spatial models used for this borrowing process. In addition to improving the precision of small area parameter estimates, Bayesian hierarchical modelling provides a robust methodology for fitting models to small area data with the goal of explaining spatial variation (see Section 1.4.3).

2.4 Thinking Spatially and Temporally

2.4.1 Explaining Space-Time Variation

By the term "spatial-temporal process" we mean a process where spatial-temporal relationships, between people or places, for example, enter explicitly into the process. All such processes involve some form of interaction between entities (people, businesses, political and other institutions) in both space *and* time. The interaction process itself may be highly complex, and those which involve interactions between individuals might not be observed (or even observable), although with increased usage of location-aware technologies such as mobile phones, the possibilities for studying interaction effects and identifying network linkages amongst individuals, using big data algorithms, are increasing. The term "spatial-temporal interaction process" describes a broad class of dynamic flows from one location to another along a network and where the flows might comprise physical items (goods), people, information or ideas. The network's character might be physical and fixed like a transport link or trade route. Or its character may be various, intangible and fluid like connections between people as they communicate with each other or move around in the course of their lives. These interactions in space and time are likely to underlie the variation in attribute values that we observe in our data.[5] In Part III we look closely at how to model spatial-temporal data and how to handle the presence of what is termed "space-time interaction" in a dataset. We briefly describe some generic examples of space-time processes that may underlie this data property, which appears as clusters of cases in particular places at particular times.

(i) *Diffusion processes.* These are processes where the attribute of interest, such as information, a rumour, a disease or a new technology, is adopted, or taken up, or otherwise spreads through a population of individuals (e.g. firms or people) who for the purposes of analysis or because of data limitations are assumed to have fixed locations. By "fixed" we mean, for example, the map of adoption identifies the residential location of the adopting (and non-adopting) individuals, or the location of the adopting (and non-adopting) businesses. At any point in time it can be specified which individuals have the attribute and which do not. The spread of an infectious disease is modelled as a diffusion process where infected individuals mix with individuals from the susceptible population who, dependent on the infection risk, may then become infected. The spatial distribution of the individuals that are susceptible and the extent of mixing can have important implications for the spread of the disease (speed, spatial extent, size of the epidemic). In the case of an outbreak of rioting or civil unrest, new participants may be drawn in by copycat behaviour through inter-personal contact or first-hand experience. The spread of cultural traits may be the product of a diffusion process or it may be the product of a dispersal process, a type of process that we now turn to.

(ii) *Dispersal processes.* In contrast to a diffusion process, a dispersal process involves the re-location or spatial distribution of individuals, including people, animals and plants. The geographical spread of cultural or physical traits may be the consequence of individuals dispersing over a region carrying those traits with them and bringing them to the newly settled area. Other examples include: seed dispersal from parent plants where the spatial distribution of the attribute (the number of new plants in an area) depends on the scale and intensity of the dispersal

[5] Depending on the scale of the study, the spatial locations of the interactions might differ significantly from the spatial locations of where attributes are recorded. In the case of chronic diseases with long latency periods, individuals exposed to an environmental risk at one time in their lives might have moved to a new residential location before they are diagnosed with the condition. Similar problems arise when mapping cases of an infectious disease, although on a shorter time scale. Other examples of where the critical interactions may have happened at a different location from where an outcome is measured include voting preferences, crime victimization, social attitudes and life expectancy.

process as well as environmental attributes, such as the capacity of an area to support new arrivals; the spread of a human population over a region as a result of a migration process, with people populating a previously uninhabited or sparsely populated area; or enforced migration of an ethnically defined sub-population due to war or persecution. Various "push" and "pull" factors may determine the migrant distribution, including the distance travelled in the course of migration. Factors also include the existence of any "intervening opportunities", as well as any resistance encountered in newly settled areas.

(iii) *Interaction processes involving exchange and transfer between entities with a fixed location.* Urban and regional economies are bound together by processes of commodity exchange and income transfer. Income earned in one place may be spent elsewhere, thereby binding together the economic fortunes of different places. These transfers will be reinforced by intra-area and inter-area economic multiplier processes. Such spillover effects may be reflected in the spatial structuring of social and economic attributes such as per capita income or economic growth rates.

(iv) *Interaction processes involving action and reaction between agents.* In a market economy, the determination of prices at a set of retail outlets may involve a process of action and reaction amongst retailers. For example, a price adjustment by one retailer may, depending on the effect that price change has on levels of demand at other retail sites, lead to a price response by other retailers. The adjustment may depend on spatial proximity, with retailers nearby responding more rapidly to meet the price change than those further away. A pattern of prices may develop across the set of retail sites that reflect these competitive interactions as retailers seek to defend their market share and profitability.

(v) *Interaction processes involving spatial convergence.* The necessary condition for the occurrence of an offence, following the arguments of Routine Activity Theory, is the convergence in space and time of a motivated offender, a victim or "suitable target" (a person or a house) in the absence of a capable guardian. The underlying interactions involve potentially three groups of individuals, since the definition of a capable guardian includes the presence of other people who by their mere presence discourage the criminal act.[6] Different crimes, since they may involve different individual elements (offender, victims, capable guardians), have different space-time interaction signatures. In some cases we observe intense concentrations of crime events known as crime "hotspots".

Many different types of spatial-temporal processes have been studied by social, political and economic scientists over many years, and the interested reader can use the previously cited literature in this chapter together with the following small sample to explore this literature and other related areas of research. Archaeology (Hodder and Orton, 1976); anthropology (Alexander and Maschner, 1996; Relethford, 2008); behavioural ecology (Fayet et al., 2014); criminology (Felson and Cohen, 1980; Skogan, 1990; Bursik and Grasmick, 1993; Braga and Weisburd, 2010); economics (Besley and Case, 1995; Besley et al., 1997; Brown and Rork, 2005; Florkowski and Sarmiento, 2005; Garrett and Marsh, 2002; Haining, 1984; Kalnins, 2003; Patton and McErlean, 2003); epidemiology and public health (Bailey, 1967; Gorman et al., 2005); veterinary epidemiology (Ward and Carpenter, 2000); geography (Haining, 1987;

[6] This idea underlies forms of urban design that encourage "eyes on the street" as a way of reducing urban crime. See for example Jacobs (1961).

Sheppard et al., 1992); political science (Bailey and Rom, 2004; Cho, 2003; Gleditsch and Ward, 2006; Salehyan and Gleditsch, 2006; Williams and Whitten, 2014); sociology (Land et al., 1991; Morenoff, 2003; Papachristos et al., 2013; Tolnay, 1995; Tolnay et al., 1996); and transport research (Eckley and Curtin, 2013).

Some processes evolve over time to a final state, and it is this we may be principally interested in (for example, as in the case of the outcome of a political election by constituency). In other situations, we are interested in the way processes develop or evolve over time as well as the final state (as in the case of the spread of an infectious disease or the emergence of cultural or physical traits in a population). However, many of the outcomes that are observed and analysed in the social, economic and health sciences are the result of processes that are in a constant state of flux involving a wide range of different types of interactions between people and places and with policy shifts originating in both the public and private sectors, adding further complexity.

2.4.2 Estimating Parameters for Spatial-Temporal Units

When data are available in time as well as in space, the presence of positive temporal autocorrelation in the observed values provides a further basis for information sharing when estimating the parameters associated with each space-time unit. For example, when estimating (say) the disease risk for each small area, the numbers of disease cases observed at time point t may give us some information about the risk of disease at time point $t - 1$ or $t + 1$. We will see eventually that "temporal neighbours" of a time point t include time points that are both before *and* after t (assuming both are in the study period), that is $t - 1$ *and* $t + 1$. This definition of temporal neighbours raises an important point: instead of our usual thinking that time is "unidirectional" (time goes forward), for *information borrowing*, it is "bi-directional", meaning that we borrow information both from the "past" (e.g. $t - 1$) and from the "future" (e.g. $t + 1$). If we are willing to assume that the levels of disease risk between year $t - 1$ and year t are similar and, likewise, the levels of disease risk between year t and year $t + 1$ are similar, then when estimating the disease risk in year t, why not borrow information from both $t - 1$ and $t + 1$? We shall discuss these ideas further in Chapter 12 (Section 12.3). When modelling spatial-temporal data, the idea of spatial information sharing (Section 2.3.2) and the idea of temporal information sharing can be combined, so that when estimating the level of disease risk at one space-time unit we can borrow information from its *temporal neighbours*, its *spatial neighbours*, as well as its *temporal neighbours' spatial neighbours*. Chapter 15 provides more detail.

2.5 Concluding Remarks

"Spatial thinking" means recognizing that where entities (objects and events) and their attributes are located in space, and when they occur in time, are not just a set of co-ordinates but, rather, are important elements of the information contained in a spatial database. This information can be used both to address certain types of questions (small area estimation, spatial interpolation), to study patterns, associations and processes in space and time and to make forecasts and support policy-making. In the next chapter we consider the nature of the data in a spatial or spatial-temporal database and the implications for the application of statistical methodology.

2.6 Exercises

Exercise 2.1. With reference to any area of study with which you are familiar, describe a research question where interest focuses on:

- Spatial variation in an attribute relative to some significant fixed point(s) or area(s) (2.3.1(i)).
- How spatial gradients in attribute values between adjacent areas might influence outcomes (2.3.1(ii)).
- How the macro-spatial configuration of a society might impact on certain outcomes (2.3.1(iii)).

Give your reasons for thinking these associations might exist.

Exercise 2.2. The dataset, `malaria_individual_cases.csv`, contains individual level data on new cases of malaria reported across 139 villages in Karnataka, India. Each case is georeferenced to one of the villages and is allocated to one of three time windows: August–September 2014, October–November 2014 and December 2014–January 2015. Use a software of your choice to aggregate the individual cases to a case count (i.e. number of malaria cases) for each village-time-window. In Section 13.3, we will consider methods to visualise the resulting set of spatial-temporal count data.

Exercise 2.3. Give examples, from your field of study, where (a) spatial interpolation (applied to point or area data) and (b) small area estimation techniques would be useful.

Exercise 2.4. From your field of study, give an example of the following:

- A diffusion process (2.4.1(i))
- A dispersal process (2.4.1(ii))
- An interaction process (between places) involving some physical exchange or transfer (2.4.1(iii))
- An interaction process involving action and reaction amongst agents at different locations or in different areas (2.4.1(iv))
- An interaction process involving the convergence in geographical space of different actors or agents (2.4.1(v))

Exercise 2.5. Suppose the values observed at the six locations in Figure 2.4 are $y_1 = 20$, $y_2 = 2$, $y_3 = 4$, $y_4 = 10$, $y_5 = 6$ and $y_6 = 5$. The coordinates of these six locations are (–10, 10), (6, 4), (3, 3), (–3, –6), (–4, –4) and (–8,–4) respectively. Explore the impacts on y_s, the estimated value for location s, which is located at (0, 0) from the following three different ways to define the weights, $\psi_1,\ldots,\psi_6$:

(a) Equal weights assigned to all six points.
(b) Equal weights to the nearest four neighbours defined by Euclidean distance.
(c) Weights are defined using an exponential decay function, $\exp(-0.5 \cdot d_i)$ where d_i is the Euclidean distance between point i and point s. Note that one needs to modify the values from the exponential decay function to satisfy the condition that $\sum_{i=1}^{6} \psi_i = 1$. What is the implication if $\sum_{i=1}^{6} \psi_i \neq 1$?

Compute y_s using the above three sets of weights and explain why the resulting values for y_s are different.

Exercise 2.6. From Exercise 2.5, consider again defining the weights using the exponential decay function, which, as we shall see in Section 4.3, can be written more generally as $\exp(-\lambda \cdot d_{ij})$, where d_{ij} is a distance measure between two points i and j and λ is a positive-valued parameter. In Exercise 2.5, λ was fixed at 0.5, but in practice the value of λ is unknown but is typically estimated using data. Would it be reasonable to just calculate a value for y_s by fixing this unknown parameter to a single value (say 0.5, a point estimate from some software)? Or, would it be more reasonable to take both the point estimate (0.5) and the associated interval estimate (often reported as a 95% confidence interval between, say, 0.2 and 0.8) into the estimation of y_s? Explain why the latter is more appropriate than the former. How can we incorporate the uncertainty associated with the point estimate (as quantified by the interval estimate) into the estimation of y_s? (Hint: see Section 5.2 in terms of how to express uncertainty of a parameter using a suitable probability distribution; once a plausible probability distribution is formed based on the point and interval estimates for λ, we can perform a Monte Carlo simulation to examine the impact on y_s – see Exercise 8.7 in Chapter 8 and Exercise 16.8 in Chapter 16 for the idea of Monte Carlo simulation.)

Exercise 2.7. Review, descriptively and comparatively, the principles behind information borrowing in space and in space-time.

Exercise 2.8. With reference to some specific scientific research question within your field of study, sketch how you think the functionality of a GIS might support your work (see Appendix).

Appendix: Geographic Information Systems

Scientific reasoning depends on the existence of good quality data; clearly formulated hypotheses that can be empirically tested; a rigorous methodology which allows data-based conclusions to be drawn about these hypotheses; and an enabling technology that allows practical and precise implementation of the methodology. The scientific approach to spatial thinking requires for its development (Goodchild and Haining, 2004):

- The existence of good-quality, spatially-referenced, geocoded data
- Spatial questions that are clearly formulated as spatial hypotheses
- A rigorous methodology to enable us to compare observed data with the expectations derived from our spatial hypotheses
- An enabling technology that allows us to handle practically and precisely spatially referenced data and implement the methodology

A Geographic Information System (GIS) is an important part of the enabling technology for spatial thinking. There are several definitions of what is meant by a GIS, each highlighting the different ways in which a GIS can support working with spatially referenced data:

- GIS as a toolbox: "a powerful set of tools for collecting, storing, retrieving at will, transforming and displaying spatial data from the real world for a particular set of purposes" (Burrough and McDonnell, 2000, p.11)

- GIS as a database management system: "any… computer based set of procedures used to store and manipulate geographically referenced data" (Burrough and McDonnell, 2000, p.11)
- GIS as a spatial decision support system: "a decision support system involving the integration of spatially referenced data in a problem solving environment" (Cowen, 1988)

A GIS is part of a "constellation of computer technologies for capturing and processing geographic data" which includes the global positioning system (GPS), satellite data collection systems and digital scanners (Cromley and McLafferty, 2012, p.16). Physically, a GIS comprises computer hardware (e.g. networked computers; large format scanners and printers), software (e.g. that enables data input and output, storage and database management) and an organizational structure that includes skilled people who are able to operate the system. But growth of the internet means users of GIS do not need to have their own in-house, physical GIS. Distributed GIS services have made it possible for many more users to take advantage of GIS capability. Books by Maheswaren and Craglia (2004) and Cromley and McLafferty (2012) discuss how GIS has contributed to research and practice in public health.

GIS functionality falls into the following broad categories (Cromley and McLafferty, 2012, p.30):

(i) Measurement (e.g. distance, length, perimeter, area, centroid, buffering, volume, shape)
(ii) Topology (e.g. adjacency, polygon overlay, point and line in polygon, dissolve, merge)
(iii) Network and location analysis (e.g. connectivity, shortest path, routing, service areas, location-allocation modelling, accessibility modelling)
(iv) Surface analysis (e.g. slope, aspect, filtering, line of sight, viewsheds, contours, watersheds)
(v) Statistical analysis (e.g. spatial sampling, spatial weights, exploratory data analysis, nearest neighbour analysis, spatial autocorrelation, spatial interpolation, geostatistics, trend surface analysis)

At the time of writing there is an entry in Wikipedia (last edited 28th December 2018), "List of geographic information systems software", which gives both open-source and commercial GIS products. Notable in the former category are GRASS GIS and QGIS. Notable in the latter category are ERDAS IMAGINE, ESRI (which includes ArcMap, ArcGIS, ArcIMS), Intergraph and Mapinfo. However, there are many more – see https://en.wikipedia.org/wiki/List_of_geographic_information_systems_software.