1

Local Statistics and Local Models for Spatial Data

1.1 Introduction

Imagine reading a book on the climate of the United States which contained only data averaged across the whole country, such as mean annual rainfall, mean annual number of hours of sunshine, and so forth. Many would feel rather short-changed with such a lack of detail. We would suspect, quite rightly, that there is a great richness in the underlying data on which these averages have been calculated; we would probably want to see these data, preferably drawn on maps, in order to appreciate the spatial variations in climate that are hidden in the reported averages. Indeed, the averages we have been presented with may be practically useless in telling us anything about climate in any particular part of the United States. It is known, for instance, that parts of the north-western United States receive a great deal more precipitation than parts of the Southwest and that Florida receives more hours of sunshine in a year than New York. In fact, it might be the case that not a single weather station in the country has the characteristics depicted by the mean climatic statistics.

The average values in this scenario can be termed global observations: in the absence of any other information, they are assumed to represent the situation in every part of the study region. The individual data on which the averages are calculated can be termed local observations: they describe the situation at the local level.[1]

[1] There is at least one other slightly different definition of 'local' and 'global' in the literature. Thioulouse *et al*. (1995) define a local statistic as one which is calculated on pairs of points or areas which are adjacent and a global statistic as one calculated over all possible pairs of points or areas. Their use of the term 'local', however, is not the same as used throughout this book because it still produces a global model; it merely separates the model applications into different spatial regimes.

Only if there is little or no variation in the local observations do the global observations provide any reliable information on the local areas within the study area. As the spatial variation of the local observations increases, the reliability of the global observation as representative of local conditions decreases.

While the above scenario might appear rather ludicrous (surely no one would publish a book containing average climatic data without describing at least some of the local data?), consider a second scenario which is much more plausible and indeed describes a methodology which is exceedingly common in spatial analysis. Suppose we had data on house prices and their determinants across the whole of England and that we wanted to model house price as a function of these determinants (such models are often referred to as hedonic price models and an example of the calibration of these models is provided in Chapter 2). Typically, we might run a regression of house prices on a set of structural attributes of each house, such as the age and floor area of the house; a set of neighbourhood attributes, such as crime rate or unemployment rate; and a set of locational attributes, such as distance to a major road or to a certain school. The output from this regression would be a set of parameter estimates, each estimate reflecting the relationship between house price and a particular attribute of the house. It would be quite usual to publish the results of such an analysis in the form of a table describing the parameter estimates for each attribute and commenting on their sign and magnitude, possibly in relation to some *a priori* set of hypotheses. In fact this is the standard approach of the vast majority of empirical analyses of spatial data.

However, the parameter estimates in this second scenario are *global statistics* and are possibly just as inadequate at representing local conditions as are the average climatic data described above. Each parameter estimate describes the *average* relationship between house price and a particular attribute of the house across the study region (in this case, the whole of England). This average relationship might not be representative of the situation in any particular part of England and may hide some very interesting and important local differences in the determinants of house prices. For example, suppose one of the determinants of house prices in our model is the age of the house and the global parameter estimate is close to zero. Superficially this would be interpreted as indicating that house prices are relatively independent of the age of the property. However, it might well be that there are contrasting relationships in different parts of the study area which tend to cancel each other out in the calculation of the global parameter estimate. For example, in rural parts of England, old houses might have character and appeal, thus generating higher prices than newer houses, *ceteris paribus*, whereas in urban areas, older houses, built to low standards to house workers in rapidly expanding cities at the middle of the nineteenth century, might be in poor condition and have substantially lower prices than newer houses. This local variation in the relationship between house price and age of the house would be completely lost if all that is reported is the global parameter estimate. It would be far more informative to produce a set of *local statistics*, in this case local parameter estimates, and to map these than simply to rely on the assumption that a single global estimate will be an accurate representation of all parts of the study area.

The only difference between the examples of the US climate and English house prices presented above is that the first describes the representation of spatial data, whereas the second describes the representation of spatial relationships. It would seem that while we generally find it unhelpful to report solely global observations on spatial data, we are quite happy to accept global statements of spatial relationships. Indeed, as hinted at above, journals and textbooks in a variety of disciplines dealing with spatial data are filled with examples of global forms of spatial analysis. Local forms of spatial analysis or spatial models are very rare exceptions to the overwhelming tide of global forms of analysis that dominates the literature.

In this book, through a series of examples and discussions, we hope to convince the reader of the value of local forms of spatial analysis and spatial modelling, and in particular, the value of one form of local modelling which we term *Geographically Weighted Regression (GWR)*. We hope to show that in many instances undertaking a global spatial analysis or calibrating a global spatial model can be as misleading as describing precipitation rates across the USA with a single value.

1.2 Local Aspatial Statistical Methods

Spatial data contain both attribute and locational information: aspatial data contain only attribute information. For instance, data on the manufacturing output of firms graphed against the number of their employees are aspatial, whereas the numbers of people suffering from a certain type of disease in different parts of a country are spatial. Unemployment rates measured for one location over different time periods are aspatial but unemployment rates at different locations are spatial and the spatial component of the data might be very useful in understanding why the rates vary. The difference between aspatial and spatial data is important because many statistical techniques developed for aspatial data are not valid for spatial data. The latter have unique properties and problems that necessitate a different set of statistical techniques and modelling approaches (for more on this, see Fotheringham *et al.* 2000, particularly Chapter 2). This is also true in local analysis.

There is a growing literature and an expanding array of techniques for examining local relationships in aspatial data. For example, there are techniques such as the use of spline functions (Wahba 1990; Friedman 1991; Green and Silverman 1994); LOWESS regression (Cleveland 1979); kernel regression (Cleveland and Devlin 1988; Wand and Jones 1995; Fan and Gijbels 1996; Thorsnes and McMillen 1998); and variable parameter models in the econometric literature (Maddala 1977; Johnson and Kau 1980; Raj and Ullah 1981; Kmenta 1986; Casetti 1997) that are applicable to the local analysis of aspatial data. Good general discussions of local regression techniques for aspatial data are given by Hardle (1990), Barnett *et al.* (1991), Loader (1999) and Fox (2000a; 2000b).

The basic problem that local statistics attempt to solve is shown in Figure 1.1. Here there is a relationship between two aspatial variables, Y and X, which needs to be determined from the observed data. A global linear regression model, for example, would produce a relationship such as that depicted by line A; although the model gives a reasonable fit to the data, it clearly misses some important local variations in the relationship between Y and X. Here, notice, 'local' means in terms of attribute space, in this case that of the X variable, rather than geographical

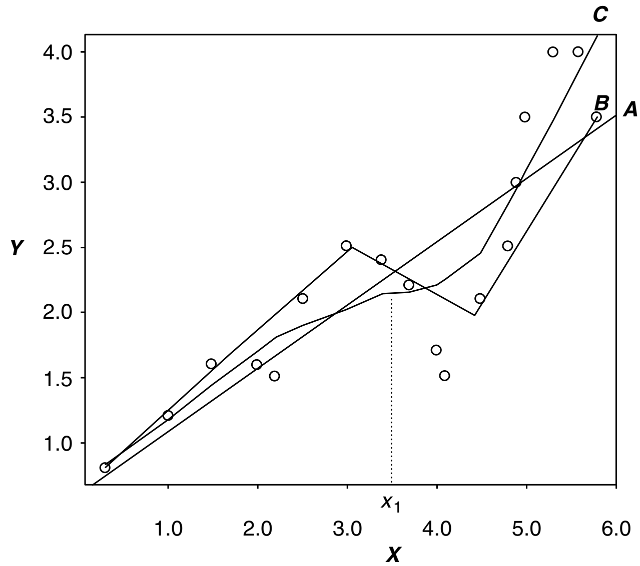

Figure 1.1 *Global and local aspatial relationships*

space.[2] A local technique, such as a linear spline function, depicted by line *B*, would give a more accurate picture of the relationship between *Y* and *X*. This would be obtained by essentially running four separate regressions over different ranges of the *X* variable with the constraint that the end points of the local regression lines meet at what are known as 'knots'.[3] Finally, a very localised technique such as LOWESS regression would yield line *C* where the relationship between *Y* and *X* is evaluated at a large number of points along the *X* axis and the data points are weighted according to their 'distance' from each of these regression points.[4] For example, suppose the regression point were at x_1. Then the data points for the regression of *Y* on *X* would be weighted according to their distance from the point x_1 with points closer to x_1 being weighted more heavily than points further away. This weighted regression yields a local estimate of the slope parameter for the relationship between *Y* and *X*. The regression point is moved along the *X* axis in small intervals until a line such as that in *C* can be constructed from the set of local parameter estimates.

[2] For something of a hybrid application of local modelling the reader is referred to McMillan (1996) in which land values in Chicago are regressed on distance to various features within the city. Although this is essentially an aspatial model because the local regressions are calibrated only in attribute space and not in geographical space, the use of distance as an independent variable does allow a spatial interpretation of the results to be made. As such, McMillan's application can be thought of as 'semi-spatial'.

[3] Although a linear spline function is depicted in this example, cubic spline functions are often used in curve fitting exercises. The linear spline is shown here to distinguish it from the LOWESS fit.

[4] The terms LOWESS and LOESS are used interchangeably in the literature; use is based on personal preference.

The difference between applying local techniques to aspatial data and to spatial data is that the relationship between Y and X, as shown in Figure 1.1 might vary depending on the location at which the regression is undertaken. That is, instead of simply having the problem of fitting a non-linear function to a set of data, this non-linear function itself might vary over space as shown for two locations in Figure 1.2. Consequently, local statistical analyses for spatial data have to cope

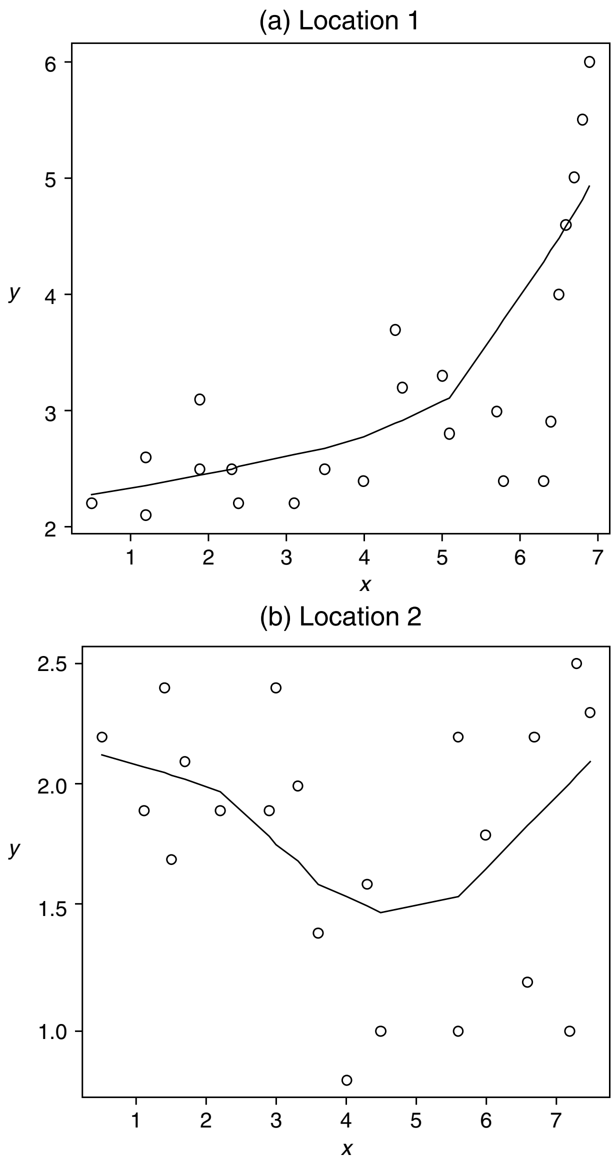

Figure 1.2 *Local relationships in attribute space for two geographical locations*

with two potential types of local variation: the local relationship being measured in attribute space and the local relationship being measured in geographical space. Compounding the problem of measuring spatial variations in relationships is the fact that the relationships in geographical space can vary in two dimensions rather than just in one. That is, local variations in attribute space, such as those shown in Figure 1.1, take place along a line and the dependency between relationships is easier to establish than in the two-dimensional equivalent of geographical space.

Because local statistical techniques for aspatial data are already fairly well established and because such techniques do not always translate easily to spatial data, the remainder of this book concentrates almost exclusively on the local analysis of spatial data. Henceforth, any discussion of local analysis is assumed to refer to spatial data unless otherwise stated.

1.3 Local versus Global Spatial Statistics

Local statistics are treated here as spatial disaggregations of global statistics. For instance, the mean rainfall across the USA is a global statistic; the measured rainfall in each of the recording stations, i.e. the data from which the mean is calculated, represent the local statistics. A model calibrated with data equally weighted from across a study region is a global model that yields global parameter estimates. A model calibrated with spatially limited sets of data is a local model that yields local parameter estimates. Local and global statistics differ in several respects as shown in Table 1.1.

Global statistics are typically single-valued: examples include a mean value, a standard deviation and a measure of the spatial autocorrelation in a data set. Local statistics are multi-valued: different values of the statistic can occur in different locations within the study region. Each local statistic is a measure of the attribute or the relationship being examined *in the vicinity of* a location within the study

Table 1.1 *Differences between local and global statistics*

Global	Local
Summarise data for whole region	Local disaggregations of global statistics
Single-valued statistic	Multi-valued statistic
Non-mappable	Mappable
GIS – unfriendly	GIS – friendly
Aspatial or spatially limited	Spatial
Emphasise similarities across space	Emphasise differences across space
Search for regularities or ‘laws’	Search for exceptions or local ‘hot-spots’
Example:	*Example:*
Classic Regression	*Geographically Weighted Regression (GWR)*

region: as this location changes, the local statistic can take on different values.[5] Consequently, global statistics are unmappable or 'GIS-unfriendly', meaning they are not conducive to being analysed within a Geographic Information System (GIS) because they consist of a single value. Local statistics, on the other hand, can be mapped and further examined within a GIS. For instance, it is possible to produce a map of local parameter estimates showing how a relationship varies over space and then to investigate the spatial pattern of the local estimates to establish some understanding of possible causes of this pattern. Indeed, given that very large numbers of local parameter estimates can be produced, it is almost essential to map them in order to make some sense of the pattern they display. Local statistics are therefore spatial statistics whereas global statistics are aspatial or spatially limited.

By their nature, local statistics emphasise differences across space whereas global statistics emphasise similarities across space.[6] Global statistics lead one into thinking that all parts of the study region can be accurately represented by a single value whereas local statistics can show the falsity of this assumption by depicting what is actually happening in different parts of the region. Consequently, local statistics are useful in searching for exceptions or what are known as local 'hot spots' in the data. This use places them in the realm of exploratory spatial data analysis where the emphasis is on developing hypotheses from the data, as opposed to the more traditional confirmatory types of analysis in which the data are used to test *a priori* hypotheses (Unwin and Unwin 1998; Fotheringham *et al*. 2000). It also suggests the techniques are not rooted fully in the positivist school of thought where the search for global models and 'laws' is important. However, this issue is not as clear-cut as it might seem because local statistics can also play an important role in confirmatory analyses as well as in building more accurate global models, a point expanded upon below.

The extent to which global estimates of relationships can present very misleading interpretations of local relationships is shown in Figure 1.3, a spatial example of Simpson's Paradox (Simpson 1951).[7] Simpson's paradox refers to the reversal of results when groups of data are analysed separately and then combined. In the spatial example presented in Figure 1.3, data are plotted showing the relationship between the price of a house and the population density of the area in which the house is located. In Figure 1.3(a) data from more than one location are aggregated

---

[5] This is the case even for statistics which measure the degree to which observations vary, such as a standard deviation, or the degree to which they covary, such as a covariance. A standard deviation presents a global average degree of variation in the data; it supplies no information on whether the degree of variation in the data varies spatially. For example, in some parts of the region, the data could be very stable, whereas in other parts, the data might vary wildly. A similar statement can be made for covariance. The traditional measure of covariance is a global statistic because it measures the degree of covariance between two variables averaged over a region. One could produce a local covariance measure that describes how the covariance between two variables differs across the region; in some areas, the two variables might exhibit considerable covariation, whereas in others the covariance might be negligible.

[6] Again, this statement is true even with statistics that measure the degree to which data vary over space. In such cases it is the degree of variation that is measured globally and we are led into thinking that this degree of variation is constant over space when in fact it might not be.

[7] For an example of Simpson's Paradox in aspatial data, see Appleton *et al*. (1996).

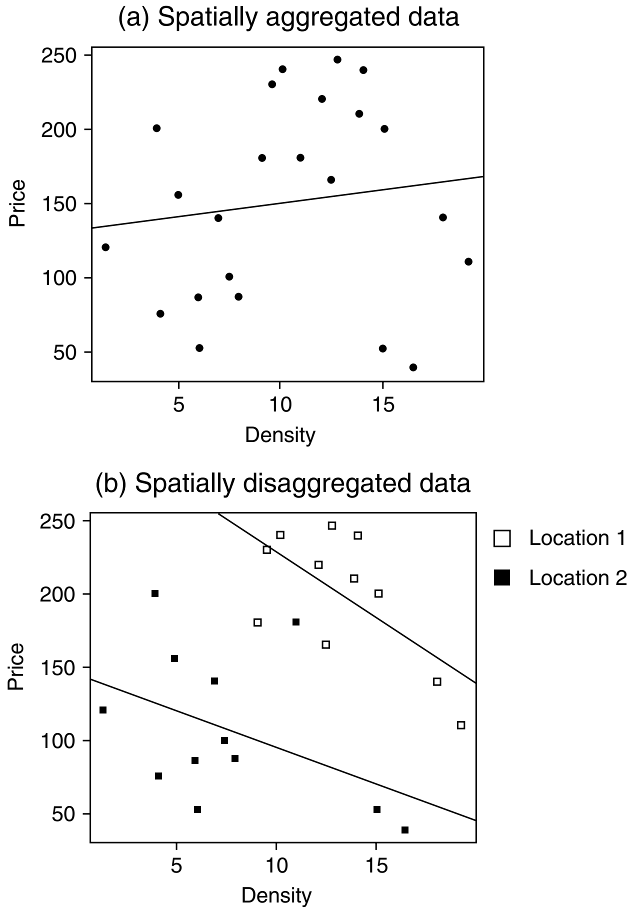

Figure 1.3 *A spatial example of Simpson's Paradox*

and the relationship, shown by the included linear regression line, is a positive one which suggests that house prices rise with increasing population density. However, in Figure 1.3(b) the data are separated by location and in both locations the relationship between house price and population density is a negative one. That is, for both individual locations, there is a negative relationship between house price and density but when the data from the two locations are aggregated, the relationship appears to be a positive one. Simpson's Paradox highlights the dangers of analysing aggregate data sets. Whilst it is normally demonstrated in aspatial data sets where the aggregation is over population subgroups, the paradox applies equally to spatial data where the aggregation is over locations.

1.4 Spatial Non-stationarity

Social scientists have long been faced with a difficult question and a potential dilemma: are there any 'laws' that govern social processes, and if there are not, does a quantitative approach have any validity? The problem is more clearly seen as two sub-problems. The first is that models in social sciences are not perfectly accurate. There is always some degree of error (sometimes quite large) indicating that a model has not captured fully the process it is being used to examine. We continually strive to produce more accurate models but the goal of a perfect model is elusive. The second is that the results derived from one system can rarely, if ever, be replicated exactly in another. An explanatory variable might be highly relevant in one application but seemingly irrelevant in another; parameters describing the same relationship might be negative in some applications but positive in others; and the same model might replicate data accurately in one system but not in another. These issues set social science apart from other sciences where the goal of attaining a global statement of relationships is a more realistic one. Physical processes tend to be *stationary* whereas social processes are often not. For instance, in physics, the famous relationship relating energy and mass, $E = mc^2$, is held to be the same no matter where the measurement takes place: there is not a separate relationship depending on which country or city you are in.[8] Social processes, on the other hand, appear to be *non-stationary*: the measurement of a relationship depends in part on where the measurement is taken. In the case of spatial processes, we refer to this as *spatial non-stationarity*. In essence, the process we are trying to investigate might not be constant over space. Clearly, any relationship that is not stationary over space will not be represented particularly well by a global statistic and, indeed, this global value may be very misleading locally. It is therefore useful to speculate on why relationships might vary over space; in the absence of a reason to suspect that they do vary, there is little or no need to develop local statistical methods.

There are several reasons why we might expect measurements of relationships to vary over space. An obvious one relates to sampling variation. Suppose we were to take spatial subsets of a data set and then calibrate a model separately with each of the subsets. We would not expect the parameter estimates obtained in such calibrations to be *exactly* the same: variations would exist because of the different samples of data used. This variation is relatively uninteresting in that it relates to a statistical artefact and not to any underlying spatial process, but it does need to be accounted for in order to identify more substantive causes of spatial non-stationarity.

A second possible cause of observed spatial non-stationarity in relationships is that, for whatever reasons, some relationships are intrinsically different across space. Perhaps, for example, there are spatial variations in people's attitudes or preferences or there are different administrative, political or other contextual issues

---

[8] Even with this equation there is a controversy over whether the speed of light is actually a constant everywhere. However, the argument is only about extreme conditions not met in any practical circumstances and the argument has far from universal acceptance.

that produce different responses to the same stimuli over space. Contextual effects appear to be well documented, for example, in studies of voting behaviour as evidenced by, *inter alia*, Cox (1969); Agnew (1996) and Pattie and Johnston (2000). The idea that human behaviour can vary intrinsically over space is consistent with post-modernist beliefs on the importance of place and locality as frames for understanding such behaviour (Thrift 1983). Within this framework the identification of local variations in relationships would be a useful precursor to more intensive studies that might highlight why such differences occur.

A third possible cause of observed spatial non-stationarity is that the model from which the relationships are estimated is a gross misspecification of reality and that one or more relevant variables are either omitted from the model or are represented by an incorrect functional form. This view, more in line with the positivist school of thought and very much in line with that in econometrics, runs counter to that discussed above: it assumes a global statement of behaviour can be made but that the structure of the model is not sufficiently well formed to allow this global statement to be made. Within this framework, mapping local statistics is useful in order to understand more clearly the nature of the model misspecification. The spatial pattern of the measured relationship can provide a good clue as to what attribute(s) might have been omitted from the model and what might therefore be added to the global model to improve its accuracy. For example, if the local parameter estimates for a particular relationship tend to have different signs for rural and urban areas, this would suggest the addition of some variable denoting the 'urban-ness' or the 'rural-ness' of an area. In this sense, local analysis can be seen as a model-building procedure in which the ultimate goal is to produce a global model that exhibits no significant spatial non-stationarity. In such instances, the role of local modelling is essentially that of a diagnostic tool which is used to indicate a problem with the global model; only when there is no significant spatial variation in measured relationships can the global model be accepted.

Alternatively, it might not be possible to reduce or remove the misspecification problem with the global model by the addition of one or more variables: for example, it might be impossible to collect data on such variables. In such a case, local modelling then serves the purpose of allowing these otherwise omitted effects to be included in the model through locally varying parameter estimates.

The above discussion on the possible causes of spatial non-stationarity raises an interesting and, as yet unsolved puzzle in spatial analysis. If we do observe spatial variations in relationships, are they due simply to model misspecification or are they due to intrinsically different local spatial behaviour? In a nutshell, can all contextual effects be removed by a better specification of our models (Hauser 1970; Casetti 1997)? Is the role of place simply a surrogate for individual-level effects which we cannot recognise or measure? If the nature of the misspecification could be identified and corrected, would the local variations in relationships disappear? We can only speculate on whether, if one were to achieve such a state, all significant spatial variations in local relationships would be eliminated (see also Jones and Hanham 1995 for a useful discussion on this and the role of local analysis in both realist and positivist schools of thought). We can never be completely confident that our models are correct specifications of reality because of our lack of theoretical understanding of the processes governing human spatial behaviour. In some ways, this is a chicken-and-egg dilemma. We can never completely test theories of spatial behaviour because of model misspecification, but model misspecification is the product of inadequate spatial theory.

However, the picture is not so bleak: in specific applications of any form of spatial model, we can ask whether the current form of the model we are using produces significant local variations in any of the relationships in which we are interested. If the answer is 'yes', then an examination of the nature of the spatial variation can suggest to us a more accurate model specification or the nature of some intrinsic variation in spatial behaviour. In either case, our knowledge of the system under investigation will be improved, in some cases dramatically.

Given the potential importance of local statistics and local models to the understanding of spatial processes, it is surprising that local forms of spatial analysis are not more frequently encountered. However, there have been some notable contributions to the literature on spatially varying parameter models that we now describe. These developments can be divided into three categories: those that are focussed on local statistics for univariate spatial data, including the analysis of point patterns; those that are focussed on more complex multivariate spatial data; and those that are focussed on spatial patterns of movement. We now describe some of the literature on local models and local statistics prior to a full description in Chapter 2 of one local modelling technique, Geographically Weighted Regression, that forms the focus of this book.

1.5 Examples of Local Univariate Methods for Spatial Data Analysis

Four types of local univariate analysis for spatial data can be identified. These are: local forms of point pattern analysis; local graphical analysis; local filters; and local measures of spatial dependency.

1.5.1 Local Forms of Point Pattern Analysis

Many data, such as the locations of various facilities, or the incidence of a particular disease, consist of a set of geocoded points that make up a spatial point pattern. The analysis of spatial point patterns has long been an important concern in geographical enquiry (*inter alia*, Getis and Boots 1978; Boots and Getis 1988). Traditionally, most methods of spatial point pattern analysis, such as quadrat analysis and neighbour statistics, have involved the calculation of a global statistic that describes aspects of the whole point pattern (*inter alia* Dacey 1960; King 1961; Tinkler 1971; Boots and Getis 1988). From this global analysis, a judgement would be reached as to whether the overall pattern of points was clustered, dispersed or random. Clearly, such analyses are potentially flawed because interesting spatial variations in the point pattern might be subsumed in the calculation of the average or global statistic. In many instances, particularly in the study of disease, such an approach would appear to be contrary to the purpose of the study, which is to identify any interesting local clusters of disease incidence (see, for example, Lin and Zeng 1999). Typically, we are not interested in some general statistic referring to the whole point pattern: it is more useful to be able to identify particular parts of the study region in which there is a raised incidence of the disease. Consequently, there has been a growing interest in developing local forms of point pattern analysis.

One of the first of these was the Geographical Analysis Machine (GAM) developed by Openshaw *et al.* (1987) and updated by Fotheringham and Zhan (1996). As Fotheringham and Brunsdon (1999) note, the basic components of a GAM are:

1. a method for defining sub-regions of the data;
2. a means of describing the point pattern within each of these sub-regions;
3. a procedure for assessing the statistical significance of the observed point pattern within each sub-region considered independently of the rest of the data;
4. a procedure for displaying the sub-regions in which there are significant patterns as defined in 3.

The basic idea outlined in Fotheringham and Zhan (1996) demonstrates the emphasis of this type of technique on identifying interesting local parts of the data set rather than simply providing a global average statistic. Within the study region containing a spatial point pattern, random selection is made initially of a location, and then of a radius of a circle to be centred at that location. Within this random circle, the number of points is counted and this observed value compared with an expected value based on an assumption about the process generating the point pattern (usually that it is random). The population-at-risk within each circle is then used as a basis for generating an expected number of points which is compared to the observed number. The circle can then be drawn on a map if it contains a statistically interesting count (that is, a much higher or lower observed count of points than expected). The process is repeated many times so that a map is produced which contains a set of circles centred on parts of the region where interesting clusters of points appear to be located. The GAM and similar statistics are a subset of a much broader class of statistics known as 'Scan Statistics' of which there are several notable spatial applications, particularly in the identification of disease clusters (*inter alia* Kulldorf and Nagarwalla 1995; Hjalmars *et al.* 1996; Kulldorf 1997; Kulldorf *et al.* 1997; Gangnon and Clayton 2001).[9]

1.5.2 Local Graphical Analysis

One of the by-products of the enormous increases in computer power that have taken place is the rise of techniques for visualising data (Fotheringham, 1999a; Fotheringham *et al.* 2000, Chapter 4). Within spatial data analysis, exploratory

[9] At the time of writing, software for calculating spatial, temporal and space-time scan statistics can be downloaded from http://dcp.nci.nih.gov/bb/SaTScan.html

graphical techniques which emphasise the local nature of relationships have become popular. For example, using software such as MANET (Unwin *et al*. 1996), or XLispstat (Tierney 1990; Brunsdon and Charlton 1996), it is possible to link maps of spatial data with other non-cartographical representations (such as scatterplots or dotplots). Selecting an object on one representation highlights the corresponding object on the other (an early example of this is Monmonier 1969). For example, if a scatterplot reveals a number of outlying observations, selecting these points will highlight their locations on a map. Similarly, selecting a set of points or zones on a map will highlight the corresponding points on a scatterplot. In this way, the spatial distribution of an attribute for a locally selected region can be compared to the distribution of the same attribute across the study area as a whole. Using techniques of this sort, combined with a degree of numerical pre-processing, it is possible to carry out a wide range of exploratory tasks on spatial data which are essentially local. For example, one can identify local clusters in data and investigate whether these are also associated with spatial clusters. Equally, one can also identify spatial outliers, cases that are *locally* unusual even if not atypical for the data set as a whole. More complex graphical techniques for depicting local relationships in univariate data sets include the spatially lagged scatterplot (Cressie 1984), the variogram cloud plot (Haslett *et al*. 1991) and the Moran scatterplot (Anselin 1996).

1.5.3 Local Filters

A number of techniques exist in image-processing that can be considered as ‘local’. The data for an image is usually presented as a regular array of intensity values each value referring to a single cell of known area (or a pixel). In order to determine which pixels are likely to represent edges in the image, a high-pass filter can be applied; this acts to increase high-intensity values, and decrease low ones. To remove isolated high values, a low-pass filter can be employed; its action is to make the values in nearby pixels more similar. Other filters may be applied to enhance the values of linear objects in the image; these are known as directional filters. Such filters are usually a square array of weights, often 3×3 pixels. The output from a filter is a weighted mean value of the pixel at its centre and its immediate neighbours; the filter is applied to each pixel in the input image to produce an output image. The reader is referred to Lilliesand and Kiefer (1995) for further information on the use of filters in image processing.

These filtering techniques have also been applied to raster GIS data (i.e., data stored as a regular lattice). Tomlin (1990) proposed a wide variety of functions that can be applied to local neighbourhoods in such data. Examples of these include the ‘focalmean’, the ‘focalmedian’ and the ‘focalvariety’. The focalmean function provides a weighted mean of the values in the raster which are immediate neighbours of the central one; in this way both high-pass and low-pass filters can be applied to raster GIS data. The focalmedian will return the median of the nine values in the surrounding 3×3 matrix (in some implementations the filter size can be varied). If the values in the raster are categorical (for example, they may represent land uses), then focalvariety will count the number of different values in the 3×3 matrix.

Some early examples of the use of filters for spatial analysis are contained in Schmid and MacCannell (1955) and Unwin (1981). More sophisticated examples are given by Cheng *et al*. (1996) who use variable window sizes and shapes for the local filtering of geochemical images. Rushton *et al*. (1995) apply a spatial filter to student enrolment projections.[10] A similar technique, popular in fields such as geodesy and meteorology, is that of optimal interpolation in which data weighted by spatial proximity are used to estimate unknown values (Liu and Gauthier 1990; Daley 1991, Reynolds and Smith 1995). The technique is also known as objective analysis (Cressman 1959).

1.5.4 Local Measures of Spatial Dependency

Spatial dependency is the extent to which the value of an attribute in one location depends on the values of the attribute in nearby locations. Although statistics for measuring the degree of spatial dependency in a data set have been formulated for almost three decades (*inter alia* Cliff and Ord 1972; Haining 1979), until very recently these statistics were only applied globally. Typically a single statistical measure is calculated which describes an overall degree of spatial dependency across the whole data set. Recently, however, local statistics for this purpose have been developed by Getis and Ord (1992), Ord and Getis (1995; 2001), Anselin (1995; 1998) and Rogerson (1999). Getis and Ord (1992), for example, develop a global measure of spatial association inherent within a data set that measures the way in which values of an attribute are clustered in space. A local variation of this global statistic is then formulated to depict trends in the data around each point in space. There are two variants of this localised value depending on whether or not the calculation includes the point i, around which the clustering is measured, although both are equivalent to spatially moving averages (Ord and Getis 2001). The local spatial association statistic allows that different trends in the distribution of one variable might exist over space. In some parts of the study area, for example, high values might be clustered; in other parts there might be a mix of high and low values. Such differences would not be apparent in the calculation of a single global statistic. In their empirical example, Getis and Ord (1992) find several significant local clusters of sudden infant death syndrome in North Carolina although the global statistic fails to identify any significant clustering.

Another local statistic for measuring spatial dependency is a local variant of the classic measure of spatial autocorrelation, Moran's I (Anselin 1995). When spatial data are distributed so that high values are generally located in close proximity to other high values and low values are generally located near to other low values, the data are said to exhibit positive spatial autocorrelation. When the data are

[10] At the time of writing, details of the application of spatial filters to health data, plus a downloadable copy of software for this purpose, DMAP, are provided by Rushton and his colleagues at http://www.uiowa.edu/%7Egeog/health/index11.html

distributed such that high and low values are generally located near each other, the data are said to exhibit negative spatial autocorrelation. However, it is possible that within the same data set, different degrees of spatial autocorrelation could be present; both positive and negative spatial autocorrelation could even exist within the same data set. Global measures of spatial autocorrelation would fail to pick up these different degrees of spatial dependency within the data. A global statistic might therefore misleadingly indicate that there is no spatial autocorrelation in a data set, when in fact there is strong positive autocorrelation in one part of the region and strong negative autocorrelation in another. The development of a localised version of spatial autocorrelation allows spatial variations in the spatial arrangement of data to be examined. Anselin (1995) presents an application of the localised Moran's I statistic to the spatial distribution of conflict in Africa and Sokal *et al.* (1998) demonstrate its use on a set of simulated data sets. Other studies of local Moran's I include those of Bao and Henry (1996), Tiefelsdorf and Boots (1997), and Tiefelsdorf (1998). Rosenberg (2000) provides a partially local measure of spatial autocorrelation through a directionally varying Moran's I coefficient and Brunsdon *et al.* (1998) describe a different method of estimating local spatial autocorrelation through Geographically Weighted Regression.

Finally, Rogerson (1999) derives a local version of the chi-square goodness-of-fit test and applies this to the problem of identifying relevant spatial clustering. This local statistic is related to Oden's (1995) modification of Moran's I that accounts for spatial variations in population density and is a special case of a test suggested by Tango (1995). The local statistic incorporates a spatially weighted measure of the degree of dissimilarity across regions.

1.6 Examples of Local Multivariate Methods for Spatial Data Analysis

The local univariate statistical methods described above are of limited use in the large and complex spatial data sets that are increasingly available. There is a need to understand local variations in more complex multivariate relationships (see, for example, the attempts by Ver Hoef and Cressie, 1993 and Majure and Cressie, 1997 to extend some of the local visual techniques described above to the multivariate case). Consequently, several attempts have been made to produce localised versions of traditionally global multivariate techniques. Perhaps the greatest challenge, given its widespread use, has been to produce local versions of regression analysis. The subject matter of this book, Geographically Weighted Regression, is one response to this challenge but there have been others. Here we describe five of these: the spatial expansion method; spatially adaptive filtering; multilevel modelling; random coefficient models; and spatial regression models. We leave the description of GWR to Chapter 2. Each of the five techniques described below has limited application to the analysis of spatially non-stationary multivariate relationships for reasons we now explain.

1.6.1 The Spatial Expansion Method

The Expansion Method (Casetti 1972; 1997; Jones and Casetti 1992) recognises explicitly that the parameters in a regression model can be functions of the context in which the regression model is calibrated. It allows the parameter estimates to vary locally by making the parameters functions of other attributes. If the parameters are functions of location (that is, if the relationships depicted by the parameter estimates are assumed to vary over space), a *spatial expansion model* results in which *trends* in parameter estimates over space can be measured (Brown and Jones 1985; Brown and Kodras 1987; Brown and Goetz 1987; Fotheringham and Pitts 1995; Eldridge and Jones 1991). Initially, suppose a global model is proposed such as:

$$y_i = \alpha + \beta x_{i1} + \ldots \tau x_{im} + \varepsilon_i \tag{1.1}$$

where y represents a dependent variable, the xs are independent variables, $\alpha, \beta, \ldots \tau$ represent parameters to be estimated, ε represents an error term and i represents a point in space at which observations on the ys and xs are recorded. This global model can be expanded by allowing each of the parameters to be functions of other variables. While most applications of the expansion method (see Jones and Casetti 1992) have undertaken aspatial expansions, Brown and Jones (1985), Eldridge and Jones (1991) and McMillen (1996) show that it is relatively straightforward to allow the parameters to vary over geographic space so that, for example:

$$\alpha_i = \alpha_0 + \alpha_1 u_i + \alpha_2 v_i \tag{1.2}$$

$$\beta_i = \beta_0 + \beta_1 u_i + \beta_2 v_i \tag{1.3}$$

and

$$\tau_i = \tau_0 + \tau_1 u_i + \tau_2 v_i \tag{1.4}$$

where u_i and v_i represent the spatial coordinates of location i. Equations (1.2)–(1.4) represent very simple linear expansions of the global parameters over space but more complex, non-linear, expansions can easily be accommodated.

Once a suitable form for the expansion has been chosen, the original parameters in the global model are replaced with their expansions. For instance, if it is assumed that parameter variation over space can be captured by the simple linear expansions in equations (1.2)–(1.4), the expanded model would be:

$$\begin{aligned} y_i = {} & \alpha_0 + \alpha_1 u_i + \alpha_2 v_i + \beta_0 x_{i1} + \beta_1 u_i x_{i1} + \beta_2 v_i x_{i1} \\ & + \ldots \tau_0 x_{im} + \tau_1 u_i x_{im} + \tau_2 v_i x_{im} + \varepsilon_i \end{aligned} \tag{1.5}$$

This model can then be calibrated by ordinary least squares regression to produce estimates of the parameters which are then fed back in to equations (1.2)–(1.4) to obtain spatially varying parameter estimates. These estimates, being specific to location i, can then be mapped to display spatial variations in the relationships represented by the parameters.

The expansion method has been very important in promoting awareness of spatial non-stationarity. However, it does have some limitations. One is that the technique is restricted to displaying trends in relationships over space with the complexity of the measured trends being dependent upon the complexity of the expansion equations. Consequently, the distributions of the spatially varying parameter estimates obtained through the expansion method might obscure important local variations to the broad trends represented by the expansion equations. A second limitation is that the form of the expansion equations needs to be assumed *a priori*, although more flexible functional forms than those shown above could be used. A third is that the expansion equations are assumed to be deterministic in order to remove problems of estimation in the terminal model.

1.6.2 Spatially Adaptive Filtering

Another approach to regression modelling that allows coefficients to vary locally is that of adaptive filtering (Widrow and Hoff 1960; Trigg and Leach 1968). When applied to multivariate time series data this method is used to compensate for drift of regression parameters over time. Essentially, this works on a ‘predictor-corrector’ basis. Suppose a model is assumed of the form

$$y_t = \sum_j x_{tj}\beta_{tj} + \varepsilon_t \tag{1.6}$$

where t is an index of discrete time points. When a new multivariate observation occurs at time t, the existing regression coefficients, $\hat{\boldsymbol{\beta}}_{t-1}$, are used to predict the dependent variable. However, if the prediction does not perform well, the values of the regression coefficient are ‘adjusted’ to improve the estimate. The adjusted coefficients are referred to as $\hat{\boldsymbol{\beta}}_t$. The degree of adjustment applied has to be ‘damped’ in some way to avoid problems of overcompensation. That is, in most cases a set of estimates of the β_j values could be found which gave a perfect prediction, but which also fluctuate wildly and do not give a good indication of the true values of $\boldsymbol{\beta}$ at time t. A typical damping approach is to use an update rule of the form

$$\hat{\beta}_{jt} = \hat{\beta}_{jt-1} + |\hat{\beta}_{jt-1}|\alpha_j(y_t - \hat{y}_t)/|\hat{y}_t| \tag{1.7}$$

where $\hat{\beta}_{jt}$ is the jth element of $\hat{\boldsymbol{\beta}}_t$, $\hat{y}_t$ is the predicted value of y_t based on $\hat{\boldsymbol{\beta}}_{t-1}$ and α_j is a damping factor controlling the extent to which the correction is applied for coefficient j.

Foster and Gorr (1986) and Gorr and Olligschlaeger (1994) suggest applying adaptive filtering ideas to spatial data to investigate the ‘drift’ of regression parameters. With spatial data the predictor-corrector approach then becomes iterative. With time series data, $\hat{\boldsymbol{\beta}}_{t-1}$ is simply updated in terms of its nearest *temporal* neighbour at time t; a given case has a unique neighbour and the flow of updating is only one way. However, when dealing with a spatial arrangement of data, zones (or points) typically do not have unique neighbours and the coefficient estimates have to be updated several times. In addition, the flow of updating is two-way between a pair of neighbouring zones which requires the process to iterate between coefficient estimates until some form of convergence is achieved. If convergence does occur, then the result should be a unique estimate of the regression coefficient vector $\boldsymbol{\beta}$ for each case. The fact that the casewise correction procedure is damped and based on incremental corrections applied between adjacent zones, suggests that some degree of spatial smoothing of the estimates of the individual elements of $\boldsymbol{\beta}$ must take place. Thus, the method tends to produce models in which regression parameters slowly 'drift' across geographical space. Local and regional effects may be investigated by mapping the coefficient estimates.

1.6.3 Multilevel Modelling

The typical spatial application of multilevel modelling attempts to separate the effects of personal characteristics and place characteristics (contextual effects) on behaviour (Goldstein 1987; Jones 1991a; 1991b; Duncan and Jones 2000). It is claimed that modelling spatial behaviour solely at the individual level is prone to what is known as the atomistic fallacy, missing the context in which individual behaviour occurs (Alker 1969). Equivalently, modelling behaviour solely at the aggregate or contextual level is prone to the ecological fallacy, that the results might not apply to individual behaviour (Robinson 1950). Multilevel modelling tries to avoid both these problems by combining an individual-level model representing disaggregate behaviour with a macro-level model representing contextual variations in behaviour. The resulting model has the form

$$y_{ij} = \alpha_j + \beta_j x_{ij} + \varepsilon_{ij} \tag{1.8}$$

where y_{ij} represents the behaviour of individual i living in place j; x_{ij} is the ith observation of attribute x at place j; and α_j and β_j are place-specific parameters where

$$\alpha_j = \alpha + \mu_j^{\alpha} \tag{1.9}$$

and

$$\beta_j = \beta + \mu_j^{\beta} \tag{1.10}$$

Each place-specific parameter is therefore viewed as consisting of an average value plus a random component. Substituting (1.9) and (1.10) into (1.8) yields the multilevel model,

$$y_{ij} = \alpha + \beta x_{ij} + (\varepsilon_{ij} + \mu_j^{\alpha} + \mu_j^{\beta} x_{ij}) \tag{1.11}$$

This model cannot be calibrated by OLS regression unless μ_j^α and μ_j^β are zero so that specialised software is needed such as HLM (Bryk *et al.* 1986), Mln (Rasbash and Woodhouse 1995) or MlwiN (Goldstein *et al.* 1998). Place-specific parameter estimates can be obtained by estimating separate variance effects and substituting these into equations (1.9) and (1.10).

Several refinements to the basic multilevel model described above are possible and are probably necessary for the accurate estimation of individual and contextual effects, making the modelling framework highly complex. These include adding place attributes in the specifications for α_j and β_j; extending the number of levels in the hierarchy beyond two (Jones *et al.* 1996); and the development of cross-classified multilevel models where each lower unit can nest into more than one higher order unit (Goldstein 1994). Although there are numerous examples of the application of multilevel modelling to spatial data, including those of Congdon (1995), Charnock (1996), Jones (1997), Verheij (1997), Duncan *et al.* (1996), Jones and Bullen (1993), Smit (1997), Duncan (1997), Reijneveld (1998) and Duncan and Jones (2000), Duncan and Jones (2000: 298) offer the following warning:

> Importantly, there are also other, more general, issues surrounding the use of multilevel models that require careful consideration. Until recently, researchers in many disciplines seem to have been carried away in an enthusiastic rush to use the new technique and such issues have tended to be ignored.

One particular problem with the application of multilevel modelling to spatial processes is that it relies on an *a priori* definition of a discrete set of spatial units at each level of the hierarchy. While this may not be an issue in many aspatial applications, such as the definition of what constitutes the sets of public and private transportation options, or what constitutes the sets of brands of decaffeinated and regular coffees, it can pose a serious problem in many spatial contexts. The definition of discrete spatial units in which spatial behaviour is modified by certain attributes of those units obviously depends critically on the units themselves being accurately identified. It also implies that the nature of whatever spatial process is being modelled is discontinuous. That is, it is assumed that the process is modified in exactly the same way throughout a particular spatial unit but that the process is modified in a different way as soon as the boundary of that spatial unit is reached. Most spatial processes do not operate in this way because the effects of space are continuous. Hence, imposing a discrete set of boundaries on most spatial processes is unrealistic. One exception would be where administrative boundaries enclose regions in which a policy that affects the behaviour of individuals is applied evenly throughout the region and where such policies vary from region to region. Duncan and Jones (2000) and de Leeuw and Kreft (1995) raise other issues with the application of multilevel models. Consequently, the application of the multilevel modelling framework to most spatial processes appears limited and the application to continuous processes awaits development (see Langford *et al.*, 1999, for a start in this direction).

1.6.4 Random Coefficient Models

While many local statistical techniques assume that the local relationships being measured vary smoothly, an alternative approach allows coefficients to vary *randomly* for each case (Rao 1965; Hildreth and Houck 1968; Raj and Ullah 1981; Swamy *et al.* 1988a; 1988b; 1989). While most applications of random coefficients models have been aspatial, Swamy (1971) presents an early example of parameter estimates that are allowed to vary cross-sectionally. To see how the random coefficients approach is applied, consider a study of house sales specifying a regression model where the dependent variable is house price, and the independent variables are characteristics of the houses. The classical linear regression approach would assume that the regression coefficient for a given variable would be the same for all cases – so that, say, the presence of a second bathroom had an identical effect on house price for any house in the study. The form this model would take, in matrix algebra, is the familiar

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \tag{1.12}$$

where $\boldsymbol{X}$ is a matrix of predictor variables (in this case house characteristics), $\boldsymbol{\beta}$ is a vector of regression coefficients, $\boldsymbol{y}$ is a vector of response variables and $\boldsymbol{\varepsilon}$ is a vector of independent random error terms with distribution $N(0,\sigma^2)$. In terms of individual cases, the model can be written as

$$y_i = \sum_j x_{ij}\beta_j + \varepsilon_i \tag{1.13}$$

In random coefficients modelling, the parameters of this model are not assumed to be constant over space but are assumed to vary from case to case, and are drawn from some random distribution, typically the normal. The model can then be written as

$$y_i = \sum_j x_{ij}\beta_{ij} + \varepsilon_i \tag{1.14}$$

where β_{ij} is now a random variable. For each variable j there are i draws of the random regression coefficient from some distribution. Assuming this distribution to be normal,

$$\beta_{ij} \sim N(\beta_j,\sigma_j^2). \tag{1.15}$$

Calibrating a random coefficients model is then a task of estimating the parameters of the distributions from which casewise parameters are drawn – in this case $\{\beta_j,\sigma_j^2\}$ for all j and σ^2, the error term variance. Then, using Bayes' theorem, it is possible to estimate the value of the regression coefficient actually drawn for each case. A further, non-parametric, extension of the technique is to drop the assumption that the coefficients are drawn from a pre-specified distribution and to estimate the distribution itself from the data (Aitkin 1997).

An alternative approach is used in Besag (1986). Here it is assumed that local observations are governed by local parameters, but are independent of one another given these parameters. However, in a Bayesian framework it is also assumed that the parameters have a prior distribution which does exhibit spatial autocorrelation. Using a technique referred to as Iterated Conditional Modes (ICM), local parameter estimates are obtained. As with the Aitkin approach, coefficients are thought of as random, but here the interpretation is Bayesian: randomness is interpreted as beliefs about the coefficient values prior to data collection and analysis. The spatial autocorrelation in the prior distribution represents the belief that nearby coefficients are likely to have similar values.

The random coefficient modelling approach is not intrinsically spatial – local regression coefficients are drawn independently from some univariate distribution and no attention is paid to the *location* to which the parameters refer. Locations that are in close proximity to each other can have regression coefficients drawn from very different-looking distributions. However, once the local parameter estimates are obtained, they can be mapped and their spatial pattern explored. In this way, local variability of certain types of models can be considered. Brunsdon, Aitkin *et al.* (1999), for example, provide an empirical comparison of the application of GWR and the random coefficients model to a data set in which the spatial distribution of a health variable is related to the spatial variability of a set of socioeconomic indicators.

1.6.5 Spatial Regression Models

Recognising that spatial data are not generally independent, so that statistical inference in ordinary regression models applied to spatial data is suspect, a number of attempts have been made to provide a regression framework in which spatial dependency is taken into account. These approaches may generally be described as spatial regression models. While such models are generally not thought of as local models, they do recognise the local nature of spatial data, for instance by relaxing the assumption that the error terms for each observation are independent. In particular, if each observation is associated with a location in space, it is assumed that the error terms for observations in close spatial proximity to one another are correlated. The vector of error terms, $\boldsymbol{\varepsilon}$, is assumed to have a multivariate Gaussian distribution with a zero mean and a variance–covariance matrix having non-zero terms away from the leading diagonal. This implies that although any given ε_i will have a *marginal* distribution centred on zero, its *conditional* distribution will depend on the values of the error terms for surrounding observations. For example, if nearby error terms tend to be positively correlated, then given a set of positive error terms one would expect the error term of another observation close to these to be positive also. That is, its conditional distribution would be centred on some positive quantity rather than zero. Although the output from such models still consists of a set of global parameter estimates, local relationships are incorporated into the modelling framework through the covariance structure of the error terms. In this sense, these models can be thought of as 'semi-local' rather than fully local.

There are a number of examples of models that can be classed as 'semi-local'. Perhaps the oldest such technique is that of Kriging (Krige 1966). Here, it is assumed that the spatial data are a set of measurements taken at n points. Suppose one of the data is a dependent variable and the others are predictor variables. Then one could fit an OLS regression model, but this would ignore the spatial arrangement of the locations at which the data are measured. An alternative is to assume that the covariance between any two error terms will be a function of the distance between them. That is, if $\boldsymbol{C}$ is the variance–covariance matrix for the n error terms, and $\boldsymbol{D}$ is the distance matrix for the sampling points, then

$$C_{ij} = f(d_{ij}) \tag{1.16}$$

where f is some distance-decay function and d_{ij} is an element of $\boldsymbol{D}$. There are a number of restrictions on the possible functional form of f, mainly due to the fact that $\boldsymbol{C}$ must be positive definite in order for the model to be well-defined. Typical functions might be the exponential

$$C_{ij} = \sigma^2 \exp(-3d_{ij}/k) \tag{1.17}$$

or the Gaussian

$$C_{ij} = \sigma^2 \exp(-3{d_{ij}}^2/k^2) \tag{1.18}$$

where the parameter σ^2 determines the level of variation of the error terms and k determines the spatial scale over which notable covariance between pairs of measurements occurs. Essentially, k controls the degree of locality in the model with small values of k suggesting that correlation only occurs between very close point pairs and large values of k suggesting that such effects exist on a larger spatial scale.

Calibrating such a model is typically treated as a two-stage problem. First, one has to estimate σ^2 and k, and once this has been done, the regression model itself is calibrated using the formula

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{y} \tag{1.19}$$

where $\hat{\boldsymbol{\beta}}$ is a vector of estimated regression coefficients, $\boldsymbol{X}$ is a matrix of independent variables and $\boldsymbol{X}^{\mathrm{T}}$ is its transpose, $\boldsymbol{y}$ is a vector of dependent variables. $\boldsymbol{C}$ is the covariance matrix from the error term estimated using the parameter estimates described above. However, this approach is not without its shortcomings. In particular, it is assumed here that $\boldsymbol{C}$ is known exactly, whereas in reality it is itself estimated from the data. It is also assumed that $\boldsymbol{C}$ is constant over space, which differentiates much of Kriging from GWR.[11] Further, the estimation of the semi-variogram, from which estimates of σ^2 and k are derived, is controversial: a good discussion of the estimation procedure is given in Bailey and Gatrell (1995) which reveals it to be something of a 'black art'.[12] Although useful results can be obtained from this approach, caution should be exercised when drawing formal statistical inferences.

Recently, these objections have been addressed to some extent by Diggle *et al.* (1998). Here, the analytically awkward form of the likelihood function for $\boldsymbol{\beta}$, σ^2 and k is dealt with in a Bayesian context. In particular, drawings from the posterior probability function for these unknown parameters are simulated using Monte Carlo Markov Chain (MCMC) techniques (Besag and Green 1993).

The last quarter century has also seen the growth of other kinds of spatial regression models, particularly those applied to zonal data such as states, counties or electoral wards. As with Kriging, one would expect that ordinary linear regression models applied to data aggregated in this way would fail to encapsulate any spatial interactions taking place because local relationships in the error terms are not represented in the simple non-spatial model. In spatial regression models, zonal proximities are taken as surrogates for local relationships and are typically measured by a *contiguity matrix* – an n by n matrix whose (i,j)th element is one if zones i and j are contiguous, and zero otherwise. Clearly this matrix is symmetrical and encapsulates the *relative* spatial arrangement of the zones. Note that this approach does not take into account the size or shape or absolute location of the zones; the information is solely topological. In most applications the contiguity matrix is standardised so that the rows sum to one and referred to as $\boldsymbol{W}$. A number of such spatial regression models exist: for example, the spatial autoregressive model of Ord (1975):

$$\boldsymbol{y} = \mu\mathbf{1} + \rho\boldsymbol{W}(\boldsymbol{y} - \mu\mathbf{1}) + \boldsymbol{\varepsilon} \tag{1.20}$$

where μ is an overall mean level of the random variate $\boldsymbol{y}$ multiplied by $\mathbf{1}$, a vector of ones; $\boldsymbol{\varepsilon}$ is a vector of independent normal error terms; and ρ is a coefficient determining the degree of spatial dependency of the model. The model can be extended so that the error vector $\boldsymbol{\varepsilon}$ also exhibits spatial autocorrelation. In this case, the coefficient ρ does not determine the distance decay rate of the spatial autocorrelation, but the degree to which the values at individual locations depend on their neighbours. A problem with this type of modelling is that neighbourhood influence is not calibrated in terms of the data but is generally prescribed by the specification

---

[11] Note that block Kriging is an attempt to overcome the problem of having to assume $\boldsymbol{C}$ is a constant over space by obtaining estimates of $\boldsymbol{C}$ separately for different spatial units. However, this is akin to running separate regressions in different spatial units and suffers from the same problems as this procedure, namely that the processes being modelled are likely to be continuous and not discrete; the processes are likely to vary within spatial units as well as between them and that there is usually no *a priori* justification for the selection of the spatial units in which the process is assumed to be stationary.

[12] Further discussion of the estimation procedure followed in Kriging, along with a worked example, can be found in Fotheringham *et al.* (2000: 171–8).

of $\boldsymbol{W}$. Similar arguments may also be applied to pre-whitening procedures (Kendall and Ord 1973). With this approach, spatial filters to remove autocorrelation effects are applied to all variables (dependent and independent) prior to fitting a regression model. Again, however, the model being calibrated is a global one. Autoregressive models were also considered in a more general form by Besag (1974) for spatial data arranged on a regular lattice.

As mentioned above, spatial regression models are really mixed models in the sense that although they recognise the impact of local relationships between data, such relationships are almost always measured with a global autocorrelation statistic and the output of the model is a set of global parameter estimates. Brunsdon *et al.* (1998) provide an interesting example where GWR is applied to a spatially autoregressive model such as that in equation (1.20) so that the output from the model is a locally varying set of parameter estimates which includes a locally varying autocorrelation coefficient. GWR applied to spatially autoregressive models is therefore an alternative, and perhaps simpler, method of deriving local measures of spatial autocorrelation.

1.7 Examples of Local Methods for Spatial Flow Modelling

Although not directly applicable to what follows in this book because the disaggregations are for discrete points rather than for a continuous space, it is still worthwhile mentioning the literature on local models of spatial interaction because of its relatively long history. It was recognised quite early that global calibrations of spatial interaction models hid large amounts of spatial information on interaction behaviour and that localised parameters yielded much more useful information (Linneman 1966; Greenwood and Sweetland 1972; Gould 1975). This was very obvious when distance-decay parameters were estimated separately for each origin in a system instead of a single global estimate being provided (see Fotheringham 1981, for further discussion). The origin-specific parameter estimates could then be mapped to provide visual evidence of spatial variations and spatial patterns in their values.

Interestingly, it was the consistent, but counter-intuitive, spatial patterns of origin-specific distance-decay parameters in these local studies that led to the realisation that the global models were gross misspecifications of reality (Fotheringham 1981; 1984; 1986; Meyer and Eagle 1982; Fotheringham and O'Kelly 1989). This, in turn, led to the development of the competing destinations framework from principles of spatial information processing and a new set of spatial interaction models from which more accurate parameter estimates can be obtained (Fotheringham 1984; 1991). It is worth stressing that the global model misspecification only came to light through local parameter estimates being obtained and then mapped. The diagnostic spatial pattern of the local distance-decay parameter estimates would be completely missed in the calibration of a global model.

1.8 Summary

Interest in local forms of spatial analysis and spatial modelling is not new. The recognition that the calibration of global models produces parameter estimates which represent an 'average' type of behaviour, and are therefore of very limited use when behaviour does vary over space, dates back at least to Linneman's calibration of origin-specific models of international trade flows (Linneman 1966). Johnston (1973) also provides an early example of local analysis in the context of voting behaviour. However, as Fotheringham (1997) notes, the current high level of interest in the 'local' rather than the 'global' and the emergence of a battery of techniques for local modelling is notable for several reasons.[13]

Among these are that it refutes the criticism that those adopting a quantitative approach to investigate spatial processes are only concerned with the search for broad generalisations and have little interest in identifying local exceptions, an observation also made by Jones and Hanham (1995). Local forms of spatial analysis also provide a linkage between the outputs of spatial techniques and the powerful visual display capabilities of GIS and some statistical graphics packages. Perhaps most importantly though, they provide much more information on spatial relationships as an aid to both model development and the better understanding of spatial processes. Local statistics and local models provide us with the equivalent of a microscope or a telescope; they are tools with which we can see so much more detail. Without them, the picture presented by global statistics is one of uniformity and lack of variation over space; with them, we are able to see the spatial patterns of relationships that are masked by the global statistics.

To this point, we have discussed several types of local analytical techniques for spatial data. However, none of these is without problems. In the remainder of this book we turn our attention to a very general local modelling technique developed for spatial data termed Geographical Weighted Regression. In the next chapter we describe the development of GWR through an empirical example of house price determinants across London. We also describe another local modelling technique, moving-window regression, in Chapter 2 because this is shown to be a stepping stone towards GWR. Indeed, moving-window regression is a rudimentary form of GWR.

---

[13] As an example of the high level of interest in local techniques, see the two special issues of *Geographical and Environmental Modelling* devoted to this topic, details of which can be found in Fotheringham (1999b) and Flowerdew (2001).