CHAPTER EIGHT
The Moran scatterplot as an ESDA tool to assess local instability in spatial association

LUC ANSELIN

8.1
Introduction

As large spatial databases become increasingly available to researchers in the social and physical sciences, new tools are needed for the analysis of this information that match the sophistication in storage, retrieval and display provided by the rapidly evolving tech-nology of geographic information systems (GIS). In many instances, the context is *data rich* but *theory poor* (Openshaw, 1991, 1993) and techniques are needed to 'let the data speak for themselves' (Gould, 1981), that is, to aid in discovering patterns, and to suggest potential relationships and hypotheses. A large battery of such methods now exists, following the pioneering ideas of Tukey (1977) on *exploratory data analysis* (EDA), which stress the interaction between the individual and the data by means of summarising displays, innovative graphics and other highly computational tools (see for example, the overview in Cleveland and McGill, 1988). EDA techniques such as box plots, Chernoff faces, Tukey star diagrams, and scatter-plot matrices are commonly used in studies that combine GIS and spatial analysis, for example, as illustrated in the applications of a socalled archaeologist's workbench in Farley *et al.* (1990) and Williams *et al.* (1990). However, such applications are *aspatial* in that they ignore the special characteristics of spatial data, such as spatial dependence and spatial heterogeneity (Anselin, 1990). As is well known, such properties will affect the validity of standard statistical techniques, and a special set of spatial statistical methods or spatial econometric methods are needed (for overviews, see Cliff and Ord, 1973, 1981; Anselin, 1988a; Cressie, 1991; Haining, 1990).

Methods of exploratory data analysis that take into account the spatial aspects of the data, that is, *exploratory spatial data analysis* (ESDA) are by no means as accepted as standard EDA tools, although they are often suggested as being an important part of the integration of spatial analysis and GIS (for example, Anselin and Getis, 1992; Bailey, 1992; Goodchild *et al.*, 1992; Fotheringham and Rogerson, 1993). An important com-ponent of such an ESDA is to measure the spatial association between observations for one or several variables. As argued in Anselin and Getis (1992) and illustrated in Anselin *et al.* (1993b), such measures can easily be incorporated in a framework that combines spatial analysis with a geographic information system. Most indices of spatial association, such as Moran's I and Geary's c spatial autocorrelation coefficients (Cliff and Ord, 1973, 1981), the variogram (Cressie, 1991), and generalised measures of spatial autocorrelation (Hubert, 1987) are global in nature. In other words, they indicate the presence or absence of a stable pattern of spatial dependence that is true for the whole data set. In practice, such a viewpoint may not be very realistic, especially when very large data sets are analysed. In these instances, the degree of spatial association between observations may show instability in the form of local non-stationarity, spatial regimes or spatial drift (for example, Anselin, 1990).

In this chapter, I suggest a simple tool to visualise and identify the degree of spatial instability in spatial association by means of Moran's I. It is based on the interpretation of this statistic as a regression coefficient in a bivariate regression of the spatially lagged variable (say, Wy) on the original variable (y). Such an interpretation readily allows for the use of a scatterplot for easy visualisation. This scatterplot may be used in isolation, in the traditional fashion, or may be integrated as an additional *view* on the data in a system of dynamic or interactive graphics, to allow for so-called scatterplot brushing (Monmonier, 1989; Haslett *et al.,* 1990, 1991; Unwin, 1993).

In the remainder of the chapter, I first briefly discuss the salient characteristics of techniques for exploratory *spatial* data analysis. Next, I review some methods that have been suggested to deal with local instability in spatial association. This is followed by an outline of the ideas behind the Moran scatter plot and a discussion of its properties and potential use. The technique is illustrated with an analysis of the spatial pattern of conflict between African countries.

8.2
Exploratory spatial data analysis

Broadly speaking, spatial data analysis can be defined as the statistical study of phenomena that manifest themselves in space. As a result, location, area, topology, spatial arrangement, distance and interaction become the focus of attention. This is well recognised in geography, for example, as expressed in Tobler's (1979) *First Law of Geography,* in which 'everything is related to everything else, but near things are more related than distant things'. In order to make this concept operational, observations must be referenced in space, that is, their locations must be specified as points, lines or areal units. The spatial referencing of observations is the salient feature of a GIS.

The important role of location for spatial data, both in terms of absolute location (co-ordinates in a space) as well as in terms of relative location (spatial arrangement, topology), has major implications for the way in which statistical analysis may be carried out. In fact, location leads to two different types of so-called spatial effects: *spatial dependence* and *spatial heterogeneity*. The former results directly from the First Law of Geography. This law will tend to result in observations that are spatially clustered, or, in other words, will yield samples of geographical data that will not be independent. From a geographical perspective, this spatial dependence is the rule rather than the exception, and it conflicts with the usual assumption of independent observations in statistics. The dependence in spatial data is often referred to as spatial autocorrelation (for a recent review from a nongeographer's perspective, see Legendre, 1993). The second, but equally important spatial effect is related to spatial (or regional) differentiation which follows from the intrinsic uniqueness of each location. Such spatial heterogeneity (or, non-stationarity) may be evi-denced in spatial regimes for variables, functional forms or model coefficients (see Anselin, 1988a, Chapter 9, for a review, and, more recently, Dutilleul and Legendre, 1993).

Exploratory data analysis may be considered as *data-driven* analysis, in that it approaches the data without many preconceived ideas, theories or hypotheses. The focus is on generating insight into patterns and associations, and on describing the data by means of so-called resistant methods, that is, methods that are not (or are less) sensitive to *extreme* or *atypical* observations (for a more detailed discussion, see Tukey, 1977; Good, 1983; and, in a spatial context, Haining, 1990, Chapter 2). None of the traditional tools of EDA are especially geared to dealing with spatial data. Moreover, many EDA techniques suggested for the initial exploration of correlation between variables, such as scatter-plot matrices, or for post-model diagnostics, such as added variable plots, generate measures of fit and of significance that become invalid in the presence of spatial dependence, as pointed out in Anselin and Getis (1992).

Exploratory *spatial* data analysis (or, spatial exploratory data analysis) should focus explicitly on the spatial aspects of the data, in the sense of spatial dependence (spatial association) and spatial heterogeneity. In other words, these techniques should aim to describe spatial distributions, discover patterns of spatial association (spatial clustering), suggest different spatial regimes or other forms of spatial instability (non-stationarity), and identify atypical observations (outliers). In a general sense, all currently available indicators of spatial autocorrelation could thus be considered as part of ESDA. However, this is not very meaningful in terms of the link between ESDA and GIS. In fact, many of the *old* techniques of spatial data analysis were developed in an era of scarce computing power, small data sets and minimal computer graphics, and their current implementations take only limited advantage (if at all) of the data storage, retrieval and visualisation capabilities of a GIS. More specifically, such methods tend to summarise a complete spatial distribution into a single number, such as Moran's I coefficient of spatial autocorrelation (Moran, 1948). While this may have been useful in an analysis of small data sets, such as the classic 26 Irish counties in Cliff and Ord (1973), it is not very meaningful (or may even be misleading) in an analysis of spatial association in hundreds or thousands of spatial units. The degree of non-stationarity (spatial instability) in large spatial data sets is likely to be such that several regimes of spatial association would be present. For example, in an analysis of the Weimar elections in 1930 in Germany, O'Loughlin *et al.* (1994) found that a highly significant Moran's *I* at the level of 921 electoral districts in effect hides several distinct local patterns of spatial clustering and complete spatial randomness. Therefore, the sole emphasis on *global* measures of spatial association as the type of spatial statistics needed in a GIS (for example, as in Griffith, 1993) is misplaced, even though the computation of such a statistic may be implemented with currently available GIS software in a fairly straightforward manner (for example, Ding and Fotheringham, 1992). Instead, the focus of ESDA techniques used in conjunction with a GIS should be on measuring and displaying *local* patterns of spatial association, on indicating local non-stationarity, on discovering *islands* of spatial heterogeneity and so on. A few methods that have been suggested to accomplish this goal are reviewed next.

8.3
Local instability in spatial association

Measures of spatial association can be broadly classified into two groups, based on the way in which spatial interaction is conceptualised. In one approach, more commonly found in geography, the interaction is seen as a covariation between neighbouring observations. I will refer to this as the *neighbourhood view* of spatial association. Neighbours are typically defined as spatial units that have a common boundary or that are within a given critical distance of each other, although more complex definitions are possible as well (see Anselin, 1988a, for a review). The neighbourhood or contiguity structure of a data set is formalised in a spatial weights matrix W, with elements w_{ij}=0 when i and j are not neighbours, and non-zero otherwise (typically, w_{ii} is assumed to be zero). In a general sense, in the neighbourhood view of spatial association, indicators are computed based on functions of the values observed at each location and the weighted average (or, spatial lag, Wy) of observations at neighbouring locations. In other words, these measures tend to deal with covariation or correlation between neighbouring values, but no interaction occurs with locations further away, that is, interaction takes the form of a step function. In the other approach, based on geostatistics, the spatial interaction is conceptualised as a continuous function of a distance metric. I will refer to this as the *distance view* of spatial association. The indicator of choice is the variogram or semi-variogram, which is based on the (squared) difference between values observed at a given distance apart (for a detailed overview, see Cressie, 1991).

The indicators of spatial association from either view that are most relevant for an exploratory approach to spatial data analysis are those that show local patterns and allow for local instabilities. Four particular strands of research are interesting in this respect. I will briefly review them next (for a more extensive review, see Anselin, 1994).

8.3.1
Indicators based on the neighbourhood view of spatial association

8.3.1.1
G statistics

In a recent article, Getis and Ord (1992) suggest two statistics to measure the degree of local spatial association for each observation in a data set. Their G_i and statistics consist of the ratio of the sum of values in neighbouring locations, defined by a given distance band, to the sum over all observations (excluding the value at i for the G_i statistics, but including it for the statistic). This statistic may be computed for many different distance bands, for example, as where $w_{ij}(d)$ is a binary matrix with w_{ij}=1when i and j are within a distance d from each other and zero otherwise. Getis and Ord derive the moments for the G_i and statistics under the assumption of normality, which allows the indication of significant local spatial association for each observation. The G_i and statistics can be easily implemented and visualised in an integrated GIS-ESDA framework, as illustrated in Ding and Fotheringham (1992) and Anselin *et al.* (1993b).

These statistics are particularly useful in the detection of potential non-stationarities, for example, when the spatial clustering of like values is concentrated in one subregion of the data. Their interpretation differs from that of other measures of spatial association (such as Moran's I) in that positive association means clustering of high values and negative association clustering of low values (and not the contiguity of opposite magnitudes). A slightly different form was recently suggested in Ord and Getis (1995), where the distributional characteristics are discussed in detail. An extension of the idea behind the G_i and statistics to a general class of local indicators of spatial association (LISA) is presented in Anselin (1995).

8.3.1.2
Geographical analysis machines

In the various *geographical analysis machines* developed by Openshaw and associates (for example, as described in Openshaw, 1993; Openshaw *et al.*, 1990, 1991), the focus is on the efficient and automatic search for patterns in a spatial data base, with little interaction with the user and limited capability in terms of visualisation or statistical inference. The search for indications of association is based on computationally intensive algorithms for pattern recognition, such as neural networks, and is applied to spatial, space-time as well as multivariate association. This approach is particularly well suited to the indication of so-called *hot spots* or spatial clusters, although the extent to which such clusters are truly significant is sometimes unclear.

8.3.2
Indicators based on the distance view of spatial association

8.3.2.1
Pocket plot

The pocket plot is a device suggested in Cressie (1991) as a way to identify local *pockets* of non-stationarity in the variogram. When observations are given on a regular grid or lattice, the residual contribution of each row or column to the variogram can be computed for different lags. For each row or column, the distribution of these residuals can be described by a box plot, which indicates whether the central tendency is different from zero (which is the expected value) and also allows outliers (that is, distance lags for which the residual contribution of the row or column is extreme) to be identified. In a sense, they are the counterpart of the local indicators of spatial association in the neighbourhood view, and can be readily visualised in a linked map.

8.3.2.2
Interactive spatial graphics

Though not specifically intended to measure local spatial association, the interactive dynamic graphics tools developed by Haslett (1993) (Haslett *et al.,* 1990, 1991; Unwin, 1993), include the variogram (in the form of a semi-variogram or variogram cloud) as an additional *view* of the data, in addition to more traditional views, such as a histogram and a map. This adds a measure of spatial association to the otherwise mostly descriptive statistics and also allows the assessment of the extent to which particular locations (or, rather, pairs of locations) drive the overall measure of association. In other words, their approach, as implemented in the SPIDER-REGARD software packages allows for the combination of indicators of spatial association and spatial heterogeneity (nonstationarity) with a map view and non-spatial descriptive statistics, in a highly visual and interactive manner.

8.4
The Moran scatterplot

8.4.1
Principle

Moran's well known *I* statistic (Moran, 1948; Cliff and Ord, 1971, 1981) gives a formal indication of the degree of linear association between a vector of observed values y and a weighted average of the neighbouring values, or spatial lag, Wy. The linear association between y and Wy underlies the specification of spatial autoregressive processes, which are typically used to express the generating mechanism behind the spatial dependence. Formally, Moran's I can be expressed in matrix notation as:

$$I = (N/S_0)y'Wy/y'y$$

where N stands for the number of observations, S_0 is the sum of all elements in the spatial weights matrix ($S_o = \ _i \ _j w_{ij}$), y are the observations in deviations from the mean, and Wy is the associated spatial lag. When the spatial weights matrix is row-standardized such that the elements in each row sum to 1, this expression simplifies to:

$$I = y'Wy/y'y$$

since in this case, $S_0=N$.

Since the y are in deviations from their mean, I is formally equivalent to a regression coefficient in a regression of Wy on y (but not of y on Wy, which would be a more natural way to specify the spatial process). The interpretation of Moran's I as a regression coefficient provides a way to visualise the linear association between y and Wy in the form of a bivariate scatterplot of Wy against y (and not of y against Wy, which would be the usual form). I will refer to this as a *Moran scatterplot*. The Moran scatterplot can be augmented with a linear regression (as a linear smoother of the scatterplot) which has Moran's I as slope, and which can be used to indicate the degree of fit, the presence of outliers, of leverage points, and so on, in the usual fashion. It is important to note that the regression of Wy on y conforms to all the classical assumptions in regression analysis, and thus can be subjected to all the standard diagnostics for model fit (for example, Belsley *et al.*, 1980). The slope in this regression is a legitimate estimate for Moran's I, but its significance (using the standard t-test for the regression) is not appropriate. The interpretation of Moran's I in this manner clearly illustrates the way in which the statistic summarises the overall pattern of linear association, in the sense that a lack of fit would indicate important local pockets of non-stationarity.

The interpretation of Moran's I as a bivariate regression coefficient is perfectly general, and in fact applies to any statistic that can be expressed as a ratio of a quadratic form and its sum of squares. An example of this is the familiar Durbin-Watson statistic for serial correlation in time series, which takes the form that is, the coefficient in a regression of Ae on e. In spatial analysis, the same approach can be taken for Moran's I on regression residuals and the Lagrange multiplier statistics for spatial dependence in Anselin (1988b).

8.4.2
Implementation

The implementation of a Moran scatterplot is straightforward, since most statistical and many GIS software packages include a scatterplot function and an associated linear regression smoother and indication of fit. The only complicating factor is the construction of the spatial lag, Wy. In order to accomplish this, the information on the spatial arrangement of the observations, for example, as contained in a GIS, must be taken to construct a spatial weights matrix. A number of approaches to carry this out with current software are outlined in Anselin *et al.* (1993a). Once a spatial weights matrix is available, a spatially lagged variable can be computed easily (Anselin, 1992; Anselin and Hudak, 1992).

8.4.3
Interpretation

An effective interpretation of a Moran scatterplot should centre on the extent to which the linear regression line reflects the overall pattern of association between Wy and y. In other words, the indication of observations that do not follow the overall trend represents useful information on local instability or non-stationarity. Three aspects in particular merit some attention.

8.4.3.1
Pockets of positive and negative association

Since the variables are taken as deviations from their means, the scatter plot is centred on 0, 0. The four quadrants in the scatter-plot box thus represent different types of association between the value at a given location (y_i) and its spatial lag, that is, the weighted average of the values in the surrounding locations (wy_i). The upper right and lower left quadrants represent positive spatial association, in the sense that a location is surrounded by similar valued locations. For the upper right this is association between high values (above the mean), while for the lower left quadrant this is association between low values (below the mean). Note that these two quadrants correspond to the notions of positive (high-high) and negative (low-low) spatial association of the Getis-Ord (1992) statistic. In other words, an examination of the relative densities of these two quadrants provides an indication of the extent to which the global measure of spatial association is determined by (dominated by) patterns of association between high or low values, similar to a pattern of significant positive and negative G_i^* statistics. Clearly, the substantive interpretation of such a pattern should be of interest, but it may also indicate a poor choice of the spatial weights matrix.

The upper left and lower right quadrants correspond to negative association, that is, low values are surrounded by high values (upper left) and high values are surrounded by low values (lower right). Again, the relative densities of these quadrants indicate which of these patterns of negative spatial association (in the traditional sense) dominate.

It is highly unlikely that a positive (negative) Moran's I is obtained by observations that are only in the lower left and upper right (upper left and lower right) quadrants. However, it is important to note the extent of *deviant* association and the degree to which these points influence the slope of the regression line (Moran's I). In some instances, a considerable mix of the two types of association for a given Moran's I may indicate the presence of different spatial regimes or local non-stationarity. It also indicates that the global indicator of spatial association may be a poor measure of the actual dependence in the process at hand.

8.4.3.2
Outliers and leverage points

Points in the scatterplot that are *extreme* with respect to the central tendency reflected by the regression slope may be outliers in the sense that they do not follow the same process of spatial dependence as the bulk of the other observations. They could thus be considered pockets of local non-stationarity, especially if they correspond to spatially contiguous locations or boundary points. The presence of outliers may also point to problems with the specification of the spatial weights matrix or with the spatial scale at which the observations are recorded. An intuitive indication of outliers can be based on the normalised residuals from the regression of Wy on y.

Similarly, observations that exert a large influence or leverage on the regression slope are of interest, again, particularly if they are spatially clustered or correspond to boundary points. The latter case provides a way to assess the influence of boundary values on the global measure of spatial association. A number of measures of leverage or influence, such as the diagonal elements of the hat matrix of Hoaglin and Welsch (1978), and Cook's (1977) measure of influence have been suggested in the literature and most statistical packages contain ways to implement them.

8.4.3.3
Spatial regimes

Useful insight into the extent to which the linear regression is a proper approximation to the pattern of spatial dependence in the data may be given by a robust local regression or scatterplot smoother, such as a LOWESS (locally weighted scatterplot smoother, Cleveland, 1979). A distinct non-linearity, alternating patterns of positive and negative association, or clearly different slopes in the smoother all indicate the inappropriateness of a single global measure for the spatial association in the data. When the distinct slopes may be associated with spatially clustered observations, they may indicate the presence of different spatial regimes, or spatial heterogeneity.

8.5
Illustration: spatial patterns of conflict in Africa

A geographical perspective has been increasingly applied in recent years to the analysis of international interactions in general, and international conflict in particular (for a review, see, for example, the collection of papers in Ward, 1992, and in particular Diehl, 1992). Measures of spatial association, such as Moran's *I,* have been applied to quantitative indices for various types of conflicts and co-operation between nation states, such as those contained in the COPDAB database (Azar, 1980). For such indices of international conflict and co-operation, both O'Loughlin (1986) and Kirby and Ward (1987) found significant patterns of spatial association indicated by Moran's *I*. The importance of spatial effects in the statistical analysis of conflict and co-operation was confirmed in a study of the interactions between 42 African nations, over the period 1966-78, reported in a series of papers by O'Loughlin and Anselin (1991, 1992) and Anselin and O'Loughlin (1990, 1992). For an index of total conflict in particular, there was strong evidence of both positive spatial autocorrelation (as indicated by Moran's *I* and the estimates in a mixed regressive, spatial autoregressive model) as well as spatial heterogeneity in the form of two distinct spatial regimes (as indicated by Getis-Ord statistics and the results of a spatial Chow test on the stability of regression coefficients). This phenomenon is thus particularly suited for an application of the Moran scatterplot as an exploratory device.

The spatial pattern of the index for total conflict is illustrated in the quintile map in Figure 8.1 (for details on the data sources and the substantive interpretation, see O'Loughlin and Anselin, 1992, and Anselin and O'Loughlin, 1992). The suggestion of spatial clustering that follows from a visual inspection of this map is confirmed by astrong positive and significant Moran's *I* of 0.555, with an associated standard normal z-value of 6.99 (all computations were carried out with the *SpaceStat* software for spatial data analysis, Anselin, 1992). This statistic is computed for a spatial weights matrix based on distance contiguity, using the smallest distance cut-off such that each country has at least one neighbour (the distance cut-off is different from the one used in Anselin and O'Loughlin, 1992, hence the slightly different results; it roughly equals the distance between the centroids of Egypt and Sudan).

The countries with significant values for the Getis-Ord statistic (using a significance level of p=0.01) are depicted in Figure 8.2. The darker shade corresponds with strong positive spatial association for Egypt, Sudan, Ethiopia, and Somalia, indicating a spatial cluster of nations in Northeast Africa with high conflict indices. The lighter shade on the map corresponds with strong negative association, and again results in a spatial cluster of nations, but now with low conflict indices and in West Africa:

Mali, Burkina Faso, Liberia, Ivory Coast, Benin, Togo and Ghana. This confirms the earlier suggestion of two spatial regimes found in Anselin and O'Loughlin (1992).

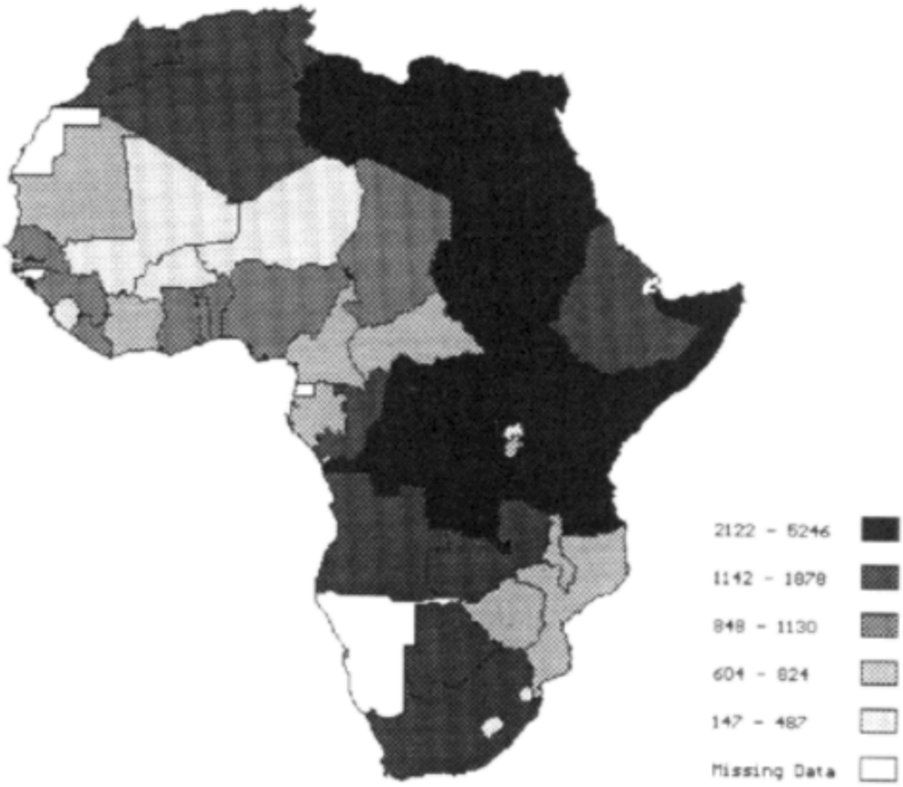

Figure 8.1 Total conflict index, Africa 1966–78.

Figure 8.3 is the Moran scatterplot for this data, with a linear smoother superimposed. The country labels are listed in Table 8.1. More than half of the associations fall in the lower left quadrant (25 out of 42), indicating the dominance of spatial contiguity of low values for the conflict index (negative spatial association in the terminology of the Getis-Ord statistics). Of the 11 points in the upper right quadrant (high-high association), four stand out, corresponding to Egypt (43), Sudan (42), Somalia (27) and Ethiopia (28), the same four nations that obtained a highly significant positive statistic. While the overall tendency portrayed in the scatterplot is one of positive association, six countries show the opposite: low values surrounded by high values for Burundi (25) and Rwanda (26), and high values surrounded by low values for Zaire (20), Angola (21), Zambia (29) and South Africa (33). Except for the latter, which is a special case, this pattern suggests a spatial cluster of nations with negative spatial autocorrelation (in the traditional sense) in West Africa, around and south of the equator. Note that neither the global Moran's I nor the G_i^* statistics are able to provide an indication of this phenomenon.

A closer look at the fit of the linear smoother is provided in Table 8.2, where the three most extreme observations are listed according to the normed residuals (outliers), the diagonal element in the hat matrix (leverage) and Cook's distance (influence). For comparison purposes, the three most extreme (most significant) z-values corresponding to the G_i^* statistics are listed as well. While the linear regression has an acceptable R^2 of 0.574, the results in Table 8.2 indicate the presence of outliers that strongly influence the regression slope. Specifically,

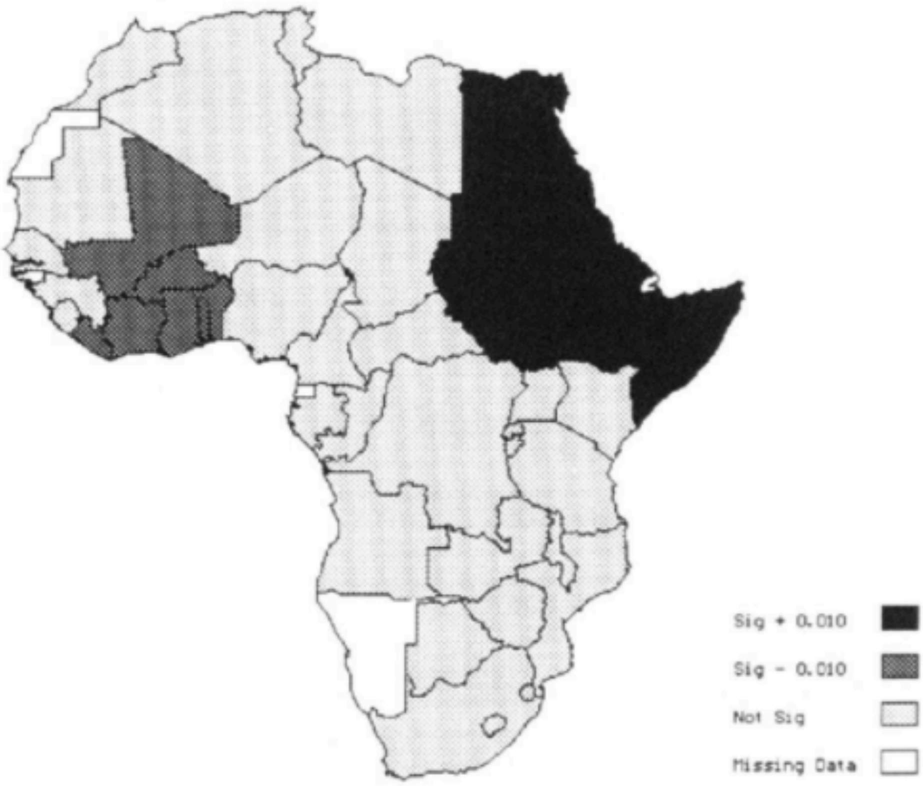

Figure 8.2 Significant G_i^* statistics for Conflict Index (p=0.01).

Egypt (43) has the most extreme value for three of the four indicators, and the second highest for the fourth (outlier). In other words, its unusually strong pattern of spatial association with its neighbours pulls the regression line (Moran's I) upwards, providing a stronger indication of positive spatial association than warranted

Table 8.1 Country labels

Label	Country	Label	Country
1	Gambia	22	Uganda
2	Mali	23	Kenya
3	Senegal	24	Tanzania
4	Benin	25	Burundi
5	Mauritania	26	Rwanda
6	Niger	27	Somalia
7	Ivory Coast	28	Ethiopia
8	Guinea	29	Zambia
9	Burkina Faso	30	Zimbabwe
10	Liberia	31	Malawi
11	Sierra Leone	32	Mozambique
12	Ghana	33	South Africa
13	Togo	34	Lesotho

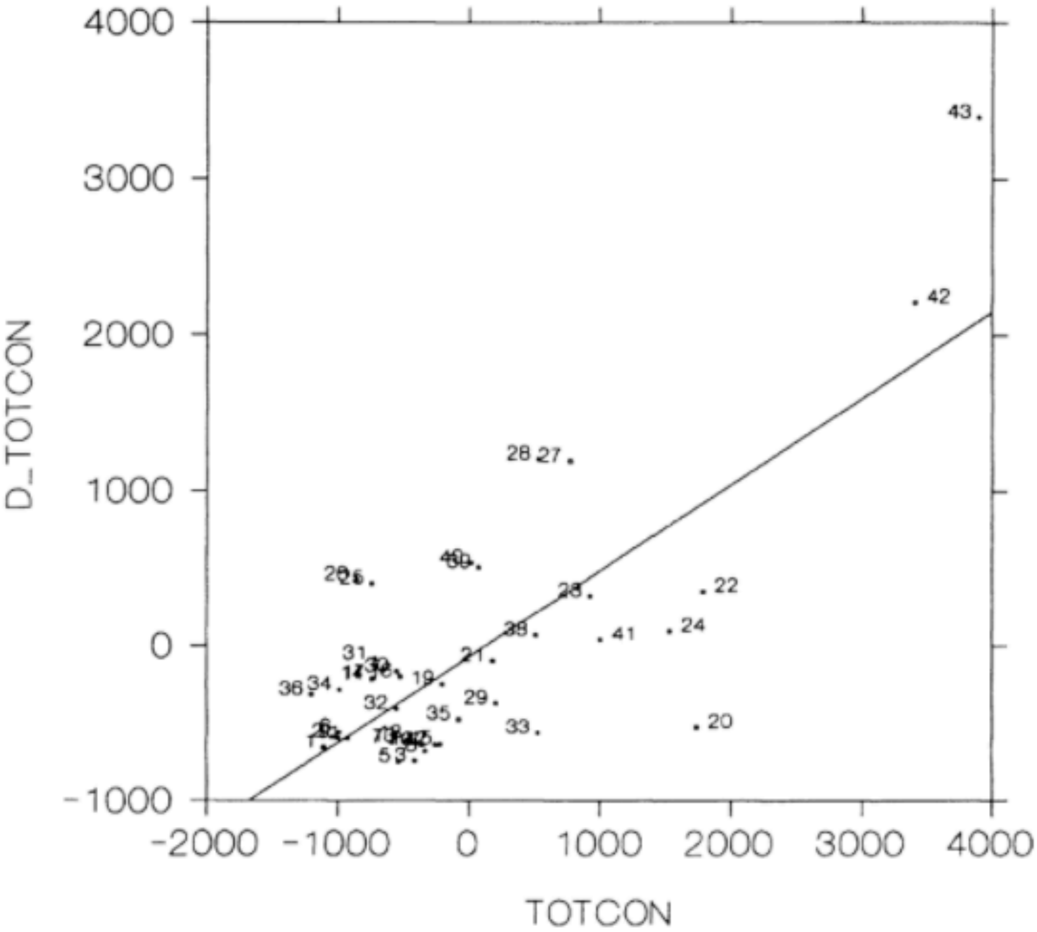

Figure 8.3 Moran scatterplot with linear smoother.

Label	Country	Label	Country
14	Cameroon	35	Botswana
15	Nigeria	36	Swaziland
16	Gabon	38	Morocco
17	CAR	39	Algeria
18	Chad	40	Tunisia
19	Congo	41	Libya
20	Zaire	42	Sudan
21	Angola	43	Egypt

by the bulk of the other observations. A similar effect is exerted by Sudan (42). When both countries are removed from the analysis, Moran's *I* drops. It should be noted that both Egypt and Sudan are boundary observations. Moreover, due to the construction of the spatial weights matrix, Egypt only has one neighbour (Sudan), while Sudan only has two (Egypt and Ethiopia). Since these countries all have high values for the conflict index, the global measure of spatial association is unduly affected. In other words, a careful analysis of the outliers and leverage points in the Moran scatter plot may indicate problems with the specification of the spatial weights matrix, as is the case in this example.

Finally, in Figure 8.4, a LOWESS smoother is superimposed on the Moran scatterplot. Again, the strong influence of Egypt (43) and Sudan (42) on the slope of the line is made clear. Also, the dip in the curve indicates a shift from positive to negative association which points to spatial heterogeneity.

8.6 Conclusions

The interpretation of Moran's I as a bivariate regression coefficient and the associated Moran scatter plot suggested in this chapter turn out to be useful devices in exploratory spatial data analysis. In particular, a careful analysis of the Moran scatterplot may help to gain insight into at least six important aspects of spatial association:

Table 8.2 Indicators of extreme observations

Index	Country label	Country name	Value
Outlier	20	Zaire	0.119
(normed residual)	43	Egypt	0.110
	28	Sudan	0.082
Leverage	43	Egypt	0.412
(hat matrix)	42	Sudan	0.174
	28	Ethiopia	0.052
Influence	43	Egypt	7.023
(Cook's distance)	28	Ethiopia	0.189
	42	Sudan	0.136
Spatial association	43	Egypt	4.70
(G_i^* z-value)	42	Sudan	4.17
	28	Ethiopia	2.56

Decomposing spatial association into its four components: low-low and high-high positive association and low-high and high-low negative association,

Identifying observations that are outliers relative to the global measure of spatial autocorrelation given by Moran's I,

Discovering different spatial regimes in the degree (slope) of spatial association,

Finding observations that exert a large influence (leverage) on the Moran coefficient,

Indicating the leverage and influence of observations that suffer from boundary effects,

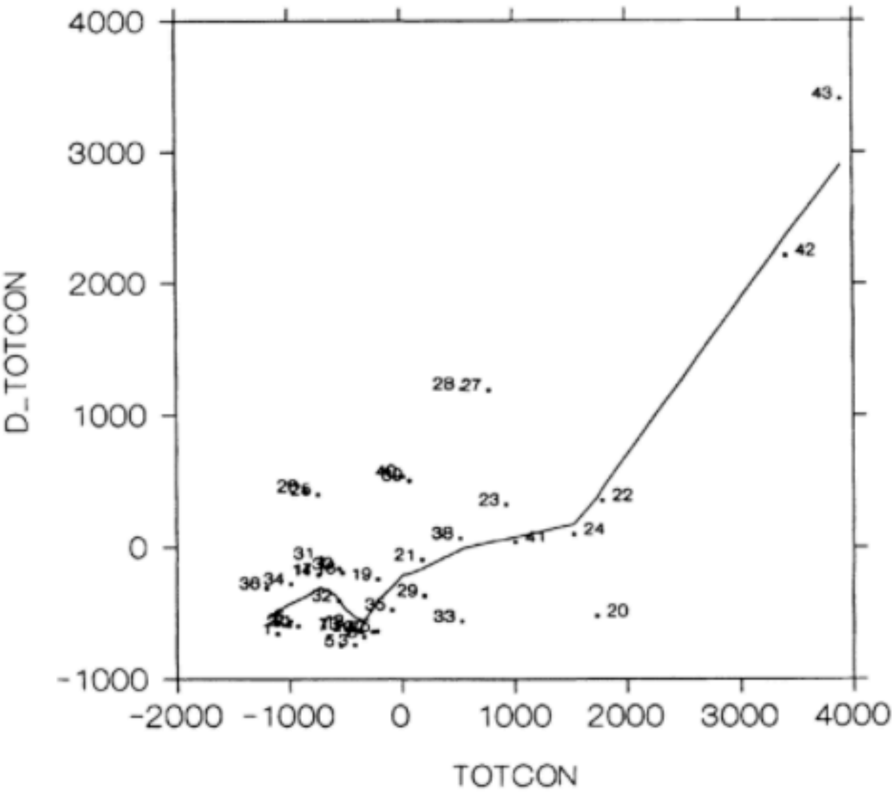

Figure 8.4 Moran scatterplot with LOWESS smoother.

Suggesting problems with the specification of the spatial weights matrix.

The indication of extreme observations in terms of spatial association is similar to that obtained from the Getis-Ord G_i^* statistics. However, the Moran scatterplot provides additional information as well and thus should be considered as a useful complement to local indicators of spatial association.

The Moran scatterplot can easily be incorporated in an integrated ESDA-GIS modelling strategy, especially one based on interactive dynamic graphics. Since a scatterplot is already part of the usual views of the data implemented in this approach, it takes little additional effort to include a special scatterplot that relates the Wy to y. Such a framework would provide a powerful tool for the exploratory analysis of spatial dependence.

Acknowledgements

The research of which this chapter is an outgrowth was supported in part by grants SES 89–21385 and SES 88–10917 (from the National Center for Geographic Information and Analysis, NCGIA) from the US National Science Foundation, and by grant GA-AS9212 from the Rockefeller Foundation. An initial version was presented at the NCGIA *Workshop on Exploratory Spatial Data Analysis and GIS,* Santa Barbara, CA, February 25–27, 1993.

CLIFF, A. and **ORD, J.K.** (1981). *Spatial Processes, Models and Applications,* London: Pion.

COOK, R. (1977). Detection of influential observations in linear regression, *Technometrics,* **19,** 15–18.

CRESSIE, N. (1991). *Statistics for Spatial Data,* New York: John Wiley.

DIEHL, P. (1992). Geography and war: A review and assessment of the empirical literature, in Ward, M. (Ed.) *The New Geopolitics,* pp. 121–37, Philadelphia, PA: Gordon and Breach.

DING, Y. and **FOTHERINGHAM, A.S.** (1992). The integration of spatial analysis and GIS, *Computers, Environment and Urban Systems,* **16,** 3–19.

DUTILLEUL, P. and **LEGENDRE, P.** (1993). Spatial heterogeneity against heteroscedasticity: an ecological paradigm versus a statistical concept, *Oikos,* **66,** 152–71.

FARLEY, J.A., LIMP, W.F. and **LOCKHART, J.** (1990). The archaeologist's workbench: Integrating GIS, remote sensing, EDA and database management, in Allen, K., Green, F. and Zubrow, E. (Eds), *Interpreting Space: GIS and Archaeology,* 141–64, London: Taylor & Francis.

FOTHERINGHAM, A.S. and **ROGERSON, P.** (1993). GIS and spatial analytical problems, *International Journal of Geographical Information Systems,* **7,** 3–19.

GETIS, A. and **ORD, K.** (1992). The analysis of spatial association by use of distance statistics, *Geographical Analysis,* **24,** 189–206.

GOOD, I.J. (1983). The philosophy of exploratory data analysis, *Philosophy of Science,* **50,** 283–95.

GOODCHILD, M.F., HAINING, R.P., WISE, S. *et al.* (1992). Integrating GIS and spatial analysis Problems and possibilities. *International Journal of Geographical Information Systems,* **6,** 407–23.

GOULD, P. (1981). Letting the data speak for themselves, *Annals Association of American Geographers,* **71,** 166–76.

GRIFFITH, D.A. (1993). Which spatial statistics techniques should be converted to GIS functions? in Fischer, M.M. and Nijkamp, P. (Eds) *Geographic Information Systems, Spatial Modelling and Policy Evaluation,* pp. 101–114, Berlin: Springer-Verlag.

HAINING, R. (1990). *Spatial Data Analysis in the Social and Environmental Sciences,* Cambridge: University Press.

HASLETT, J. (1993). Exploratory analysis of spatial data: Networks and interactive graphics. *Position Papers, Workshop on Exploratory Spatial Data Analysis and GIS,* Santa Barbara, CA: National Center for Geographic Information and Analysis.

HASLETT, J., BRADLEY, R., CRAIG, P., UNWIN, A. and **WILLS, C.** (1991). Dynamic graphics for exploring spatial data with applications to locating global and local anomalies, *The American Statistician,* **45,** 234–42.

HASLETT, J., WILLS, G. and **UNWIN, A.** (1990). SPIDER—An interactive statistical tool for the analysis of spatially distributed data, *International Journal of Geographical Information Systems,* **4,** 285–96.

HOAGLIN, D. and **WELSCH, R.** (1978). The hat matrix in regression and ANOVA, *The American Statistician,* **32,** 17–22.

HUBERT, L.R. (1987). *Assignment Methods in Combinatorial Data Analysis,* New York: Marcel Dekker.

KIRBY, A. and **WARD, M.** (1987). The spatial analysis of peace and war, *Comparative Political Studies,* **20,** 293–313.