ILL record updated to IN PROCESS
Record 9 of 28

ILL pe

CAN YOU SUPPLY? YES NO

COND FUTUREDEATE

:ILL: 8005616  :Borrower: GZM  :ReqDate: 20030725  :NeedBefore: 20030908

:Status: IN PROCESS 20030731  :RecDate:  :RenewalReq:

:OCLC: 36261688  :Source: OCLCILL  :DueDate:  :NewDueDate:

:Lender: IAY,IUL,*CGU,EEM,UIU

:CALLNO:

:TITLE: Spatial analytical perspectives on GIS /


:SERIES: GISDATA ; 4

:ARTICLE: NEED CHAPTER/ESSAY ONLY: Anselin, Luc: 'The Moran Scatterplot as
an ESDA Tool to Assess Local Instability in Spatial Association'

:VOL:

:NO:


:PATRON: <TN:26679>

:SHIP TO: ILL Borrowing

231 Memorial Library

Univ of Wisconsin Libraries

728 State St.

Madison, WI 53706-1494

:BILL TO: same FEIN 39-6006-492

:SHIP VIA: Best way

:MAXCOST: WILS or $25ifm

:COPYRT COMPLIANCE: CCL

:fax: 608 262-4649  ARIEL: 128.104.63.200  NOTE: NEW ARIEL IP ADDRESS

:E-MAIL: gzmill@library.wisc.edu

:BILLING NOTES: IFM preferred  CIC

:BORROWING NOTES: Non-CIC libraries: deduct via ifm.

:AFFILIATION: CIC

:LENDING CHARGES:

:SHIPPED:

:SHIP INSURANCE:

:RETURN TO:

:RETURN VIA:

Please report all Ariel
transmission problems
within 48 hours of receipt.
Thank you.
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 technology 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 so-called 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 component 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, $W_y$) 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 (coordinates 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 non-geographer’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 evidenced 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).

### 8.3 Local instability in spatial data

Exploratory data analytic approaches the data without explicit consideration of the so-called resistant extreme or atypical observations (for example, 1983) and, in a spatial context, the EDA are particularly geared towards the initial exploration of data matrices, or for post-mode of fit and of significance. As pointed out in Anselin and colleagues, the exploratory spatial data analysis explicitly deals with the spatial association and spatial heterogeneity of data. This is not very meaningful in terms of small data sets and mininonly limited advantage (if any) of a GIS. More specifically, spatial data analysis can be divided into two main categories: spatial data analysis and spatial data analysis of the Weimarer data.

**Note:** The focus of this chapter is on the analysis of spatial data and the implications of this analysis for understanding the spatial patterns of conflict. Future research should focus on the development of more sophisticated tools for the analysis of spatial data and the development of new methods for the analysis of spatial patterns of conflict. Future research should also focus on the development of new techniques for the analysis of spatial data and the implications of this analysis for understanding the spatial patterns of conflict. Future research should also focus on the development of new techniques for the analysis of spatial data and the implications of this analysis for understanding the spatial patterns of conflict.
tion coefficients (Cliff sed measures of spatial ords, they indicate the it is true for the whole , especially when very ial association between ury, spatial regimes or y the degree of spatial on the interpretation of of the spatially lagged ion readily allows for be used in isolation, in on the data in a system scatter-plot brushing cient characteristics of me methods that have This is followed by an n of its properties and atial pattern of conflict e statistical study of ation, area, topology, attention. This is well (1979) First Law of near things are more nal, observations must points, lines or areal s of a GIS. absolute location (co- rearrangement, topology), be carried out. In fact, atial dependence and w of Geography. This r, in other words, will From a geographical ption, and it conflicts s. The dependence in nt review from a non-ally important spatial ws from the intrinsic tionarity) may be evi- ficients (see Anselin, 1, 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, \( W y \)) 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 \( G_i^* \) 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 \( G_i^* \) statistic). This statistic may be computed for many different distance bands, for example, as \( G_i^*(d) = \sum_j w_{ij}(d)y_j / \sum_j y_j \), where \( w_{ij}(d) \) is a binary matrix with \( w_{ij} = 1 \) when \( i \) and \( j \) are within a distance \( d \) from each other and zero otherwise. Getis and Ord derive the moments for the \( G_i \) and \( G_i^* \) statistics under the assumption of normality, which allows the indication of significant local spatial association for each observation. The \( G_i \) and \( G_i^* \) 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 contaguity 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 \( G_i^* \) 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 autor interaction with the user. The search for algorithms for pattern recognition space–time as well as multi to the indication of so-call such clusters are truly sig

#### 8.3.2 Indicators based on

##### 8.3.2.1 Pocket plot

The pocket plot is a device of non-stationarity in the lattice, the residual contrib for different lags. For each described by a box plot, vertex (which is the expect for which the residual contrib sense, they are the coun

##### 8.3.2.2 Interactive spatial

Though not specifically dynamic graphics tools & techniques, include the variogram additional view of the data a map. This adds a means to statistics and also allows rather, pairs of locations) approach, as implemented, combination of indicators (stability) with a map and interactive manner.

#### 8.4 The Moran scatter

##### 8.4.1 Principle

Moran’s well known \( I \) st: indication of the degree c weighted average of the between y and Wy under are typically used to es Formally, Moran’s \( I \) can

\[
I = \frac{(N/S_0)y^T W y}{y^T y}
\]

where \( N \) stands for the n
complex definitions are


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,


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


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, \( W_y \). The linear association


where \( N \) stands for the number of observations, \( S_0 \) is the sum of all elements in the spatial
weights matrix \( S_0 = \sum \sum w_{ij} \), \( y \) are the observations in deviations from the mean, and \( W y \) 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:

\[ I = y' W y / y'y \]

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 \( W y \) on \( y \) (but not of \( y \) on \( W y \), 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 \( W y \) in the form of a bivariate scatterplot of \( W y \) against \( y \) (and not of \( y \) against \( W y \), 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 \( W y \) 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 \( e^T A e / e^T e \), that is, the coefficient in a regression of \( e^T A e \) 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, \( W y \). 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 \( W y \) and \( y \). In other words, the indication of observations that do not follow the overall trend represents useful information on local attention.

#### 8.4.3.1 Pockets of \( p \)

Since the variables are 0.0, the four quadra between the value of \( p \) and the values in the represent positive \( p \) valued locations. For the mean, while for the mean. Note that the negative (low-)words, an examination of the relationships to the \( x \) by (dominated by) a of significant position such a pattern should weights matrix.

The upper left a low values are supra low values (lower right) these patterns of \( p \).

It is highly unlikely that are only in the We have these points influencing considerable mix of presence of diffuse global indicator of the process at hand.

#### 8.4.3.2 Outliers analysis

Points in the scatter diagram the regression slope of spatial depend considered pocket: contiguous locational problems with the which the observation the normalised residuals.

Similarly, observation of interest, again points. The latter global measure of such as the diags Cook’s (1977) in statistical package.
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) and its spatial lag, that is, the weighted average of the values in the surrounding locations (Wy). 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 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 G̃_i 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 a strong 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 G̃_i 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:
proper approximation to robust local regression or scatterplot smoother, positive and negative; the inappropriateness then the distinct slopes indicate the presence.

...years to the analysis in particular (for and in particular Diehl, have been applied to between nation states. For such indices of and Kirby and Ward by Moran's $I$. The and co-operation was, particularly, there indicated by Moran's $I$ del) as well as spatial by Getis-Ord $G_i^l$ regression coefficients). The Moran scatterplot as in the quintile map in the interpretation, see 2). The suggestion of ap is confirmed by a ed standard normal $z$- software for spatial weights matrix based at each country has at used in Anselin and equals the distance statistic (using a er shade corresponds toopia, and Somalia, conflict indices. The ion, and again results and in West Africa:

**Figure 8.1** Total conflict index, Africa 1966–78.

**Figure 8.2** Significant $G_i^l$ statistics for Conflict Index ($p = 0.01$).
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 8.3 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 \( G^*_p \) 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 \( G^*_p \) 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 \( G^*_p \) 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, 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 by the bulk of the other countries.

![Figure 8.3 Moran scatterplot](image.png)

**Table 8.1 Country labels**

<table>
<thead>
<tr>
<th>Label</th>
<th>Country</th>
<th>Label</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Gambia</td>
<td>22</td>
<td>Uganda</td>
</tr>
<tr>
<td>2</td>
<td>Mali</td>
<td>23</td>
<td>Kenya</td>
</tr>
<tr>
<td>3</td>
<td>Senegal</td>
<td>24</td>
<td>Tanzania</td>
</tr>
<tr>
<td>4</td>
<td>Benin</td>
<td>25</td>
<td>Burundi</td>
</tr>
<tr>
<td>5</td>
<td>Mauritania</td>
<td>26</td>
<td>Rwanda</td>
</tr>
<tr>
<td>6</td>
<td>Niger</td>
<td>27</td>
<td>Somalia</td>
</tr>
<tr>
<td>7</td>
<td>Ivory Coast</td>
<td>28</td>
<td>Ethiopia</td>
</tr>
<tr>
<td>8</td>
<td>Guinea</td>
<td>29</td>
<td>Zambia</td>
</tr>
<tr>
<td>9</td>
<td>Burkina Faso</td>
<td>30</td>
<td>Zimbabwe</td>
</tr>
<tr>
<td>10</td>
<td>Liberta</td>
<td>31</td>
<td>Malawi</td>
</tr>
<tr>
<td>11</td>
<td>Sierra Leone</td>
<td>32</td>
<td>Mozambique</td>
</tr>
<tr>
<td>12</td>
<td>Ghana</td>
<td>33</td>
<td>South Africa</td>
</tr>
<tr>
<td>13</td>
<td>Togo</td>
<td>34</td>
<td>Lesotho</td>
</tr>
<tr>
<td>14</td>
<td>Cameroon</td>
<td>35</td>
<td>Botswana</td>
</tr>
<tr>
<td>15</td>
<td>Nigeria</td>
<td>36</td>
<td>Swaziland</td>
</tr>
<tr>
<td>16</td>
<td>Gabon</td>
<td>38</td>
<td>Morocco</td>
</tr>
<tr>
<td>17</td>
<td>CAR</td>
<td>39</td>
<td>Algeria</td>
</tr>
<tr>
<td>18</td>
<td>Chad</td>
<td>40</td>
<td>Tunisia</td>
</tr>
<tr>
<td>19</td>
<td>Congo</td>
<td>41</td>
<td>Libya</td>
</tr>
<tr>
<td>20</td>
<td>Zaire</td>
<td>42</td>
<td>Sudan</td>
</tr>
<tr>
<td>21</td>
<td>Angola</td>
<td>43</td>
<td>Egypt</td>
</tr>
</tbody>
</table>

8.6 Conclusions

The interpretation of Moran scatter plot spatial data analysis allows for a better understanding of spatial patterns in conflict, as demonstrated by the results of this study.
Ghana. This confirms the findings of O’Loughlin (1992).

More than half of the cases show positive spatial association (the upper right quadrant), with Egypt (43), Sudan (42), and Ethiopia (35, 29) having highly significant positive associations. The scatterplot is one of many visualizations that can help interpret data.

Table 8.2 lists the three largest residuals (outliers), the distance (influence). For example, case 1 strongly influences the regression by three of the four factors. It shows unusually high conflict levels compared to other countries with negative spatial clustering patterns. The bulk of the observations are not significantly affected by this case. 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. 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

<table>
<thead>
<tr>
<th>Index</th>
<th>Country label</th>
<th>Country name</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outlier (normed residual)</td>
<td>20</td>
<td>Zaire</td>
<td>0.119</td>
</tr>
<tr>
<td></td>
<td>43</td>
<td>Egypt</td>
<td>0.110</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>Sudan</td>
<td>0.082</td>
</tr>
<tr>
<td>Leverage (hat matrix)</td>
<td>43</td>
<td>Egypt</td>
<td>0.412</td>
</tr>
<tr>
<td></td>
<td>42</td>
<td>Sudan</td>
<td>0.174</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>Ethiopia</td>
<td>0.052</td>
</tr>
<tr>
<td>Influence (Cook’s distance)</td>
<td>43</td>
<td>Egypt</td>
<td>7.023</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>Sudan</td>
<td>0.189</td>
</tr>
<tr>
<td></td>
<td>42</td>
<td>Ethiopia</td>
<td>0.136</td>
</tr>
<tr>
<td>Spatial association ($G^*_z$-value)</td>
<td>43</td>
<td>Egypt</td>
<td>4.70</td>
</tr>
<tr>
<td></td>
<td>42</td>
<td>Sudan</td>
<td>4.17</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>Ethiopia</td>
<td>2.56</td>
</tr>
</tbody>
</table>

The indication of extreme observations obtained from the Getis-Ord index can be supplemented with local indicators of spatial association. The Moran scatterplot modeling strategy, esp. with LOWESS scatterplot, is already part of the framework, and it is important to take advantage of additional tools that can provide additional support for the analysis.

Acknowledgements

The research described in this article was supported by grants SES 89-21385 and SES 89-21386 from the National Center for Geographic Information and Analysis (NCGIA). It was presented at the NCGIA Workshop on GIS and Spatial Analysis, February 25–27, 1993.

- Indicating the leverage and influence of observations that suffer from boundary effects,
- 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 Geis-Ord $G_m$ 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 $W*$ 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-AS 9212 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.

REFERENCES


for Geographic Information and Analysis, University of California.
nia, and spatial data analysis in project. Journal of Conflict
ation systems: A review of the dings of the Third European
, Utrecht: EGIS Foundation.
ostics: Identifying Influential
thing scatterplots, Journal of
for Statistics, Pacific Grove,
Pion.
tions, London: Pion.
gression, Technometrics, 19,
ey, the empirical literature, in PA: Gordon and Breach.
spatial analysis and GIS, tainst heteroscedasticity: an
ist's workbench: Integrating n, Green, F. and Zubrow, E.
: Taylor & Francis.
 spatial analytical problems, -19.
y use of distance statistics, Philosophy of Science, 50,
GIS and spatial analysis – Association of American
verted to GIS functions? Systems, Spatial Modelling
sial Sciences, Cambridge:
and interactive graphics. ad GIS, Santa Barbara, CA:
91). Dynamic graphics for anomalies, The American
tive statistical tool for the
THE MORAN SCATTERPLOT AS AN ESDA TOOL TO ASSESS LOCAL INSTABILITY


