Luc Anselin

Local Indicators of Spatial Association—LISA

The capabilities for visualization, rapid data retrieval, and manipulation in geographic information systems (GIS) have created the need for new techniques of exploratory data analysis that focus on the "spatial" aspects of the data. The identification of local patterns of spatial association is an important concern in this respect. In this paper, I outline a new general class of local indicators of spatial association (LISA) and show how they allow for the decomposition of global indicators, such as Moran's I, into the contribution of each observation. The LISA statistics serve two purposes. On one hand, they may be interpreted as indicators of local pockets of nonstationarity, or hot spots, similar to the G_i and G_i^* statistics of Getis and Ord (1992). On the other hand, they may be used to assess the influence of individual locations on the magnitude of the global statistic and to identify "outliers," as in Anselin's Moran scatterplot (1993a). An initial evaluation of the properties of a LISA statistic is carried out for the local Moran, which is applied in a study of the spatial pattern of conflict for African countries and in a number of Monte Carlo simulations.

1. INTRODUCTION

The increased availability of large spatially referenced data sets and the sophisticated capabilities for visualization, rapid data retrieval, and manipulation in geographic information systems (GIS) have created a demand for new techniques for spatial data analysis of both an exploratory and a confirmatory nature (Anselin and Getis 1992; Openshaw 1993). Although many methods are available in the toolbox of the geographical analyst, only few of those are appropriate to deal explicitly with the "spatial" aspects in these large data sets (Anselin 1993b).

In the analysis of spatial association, it has long been recognized that the as-

The research of which this paper is an outgrowth was supported in part by grants SES 88-10917 (to the National Center for Geographic Information and Analysis, NCGIA) and SES 89-21385 from the U.S. National Science Foundation, and by grant GA-AS 9212 from the Rockefeller Foundation. Earlier versions were presented at the NCGIA Workshop on Exploratory Spatial Data Analysis and GIS, Santa Barbara, Calif., February 25–27, 1993, and at the GISDATA Specialist Meeting on GIS and Spatial Analysis, Amsterdam, The Netherlands, December 1–5, 1993. The comments by Arthur Getis and two anonymous referees on an earlier draft are greatly appreciated.

Luc Anselin is research professor of regional science at the Regional Research Institute of West Virginia University, where he is also professor of economics, adjunct professor of geography, and adjunct professor of agricultural and resource economics.

Geographical Analysis, Vol. 27, No. 2 (April 1995) © Ohio State University Press
Submitted 1/94. Revised version accepted 6/94.

sumption of stationarity or structural stability over space may be highly unrealistic, especially when a large number of spatial observations are used. Spatial structural instability or spatial drift has been incorporated in a number of modeling approaches. For example, discrete spatial regimes are accounted for in spatial analysis of variance (Griffith 1978, 1992; Sokal et al. 1993), and in regression models with spatial structural change (Anselin 1988, 1990). Continuous variation over space is the basis for the spatial expansion paradigm (Casetti 1972, 1986; Jones and Casetti 1992) and spatial adaptive filtering (Foster and Gorr 1986; Gorr and Olligschlaeger 1994). In exploratory spatial data analysis (ESDA), the predominant approach to assess the degree of spatial association still ignores this potential instability, as it is based on global statistics such as Moran's I or Geary's c (as in Griffith 1993). A focus on local patterns of association (hot spots) and an allowance for local instabilities in overall spatial association has only recently been suggested as a more appropriate perspective, for example, in Getis and Ord (1992), Openshaw (1993), and Anselin (1993b). Examples of techniques that reflect this approach are the various geographical analysis machines developed by Openshaw and associates (for example, Openshaw, Brundson, and Charlton 1991; and Openshaw, Cross, and Charlton 1990), the distance-based statistics of Getis and Ord (1992) (see also Ord and Getis 1994), and the Moran scatterplot (Anselin 1993a). Also, a few approaches have been suggested that are based on a geostatistical perspective, such as the pocket plot of Cressie (1991) and the interactive spatial graphics of Haslett et al. (1991).

In the current paper, I elaborate upon this general idea and outline a class of *local indicators of spatial association* (LISA). These indicators allow for the decomposition of global indicators, such as Moran's I, into the contribution of each individual observation. I suggest that this class of indicators may become a useful addition to the toolbox of ESDA techniques in that two important interpretations are combined: the assessment of significant local spatial clustering around an individual location, similar to the interpretation of the G_i and G_i^* statistics of Getis and Ord (1992); and the indication of pockets of spatial nonstationarity, or the suggestion of outliers or spatial regimes, similar to the use of the *Moran scatterplot* of Anselin (1993a).

In the remainder of the paper, I first outline the general principles underlying a LISA statistic, and suggest how it may be interpreted. I next show how a number of familiar global spatial autocorrelation statistics may be expressed in the form of a LISA. As an example of a LISA, I examine the *local Moran* more closely, first empirically, comparing it to the G_i^* statistic and the Moran scatterplot in an analysis of spatial pattern of conflict between African nations in the period 1966–78. This is followed by a series of simple Monte Carlo experiments, to provide further insight into the properties of the local Moran, its interpretation, and the relation between global and local spatial association. I close with some concluding remarks on future research directions.

2. LOCAL INDICATORS OF SPATIAL ASSOCIATION

Definition

As an operational definition, I suggest that a *local indicator of spatial association* (LISA) is any statistic that satisfies the following two requirements:

a. the LISA for each observation gives an indication of the extent of significant spatial clustering of similar values around that observation;
b. the sum of LISAs for all observations is proportional to a global indicator of spatial association.

More formally, but still in general terms, I express a LISA for a variable y_i, observed at location i, as a statistic L_i, such that

$$L_i = f(y_i, y_{J_i}), \tag{1}$$

where f is a function (possibly including additional parameters), and the y_{J_i} are the values observed in the neighborhood J_i of i.

The values y used in the computation of the statistic may be the original (raw) observations, or, more appropriately, some standardization of these in order to avoid scale dependence of the local indicators, similar to the practice often taken for global indicators of spatial association. For example, in Moran's I, as well as in its local version discussed in the next section, the observations are taken as deviations from their mean.

The neighborhood J_i for each observation is defined in the usual fashion, and may be formalized by means of a spatial weights or contiguity matrix, **W**. The columns with nonzero elements in a given row of this matrix indicate the relevant *neighbors* for the observation that corresponds to the row, that is, the elements of J_i. Examples of criteria that could be used to define neighbors are first-order contiguity and critical distance thresholds. The spatial weights matrix may be row-standardized (such that its row elements sum to one) to facilitate interpretation of the statistics, but this is not required. However, when row standardization is carried out, the function $f(y_i, y_{J_i})$ typically corresponds to a form of weighted average of the values at all observations $j \in J_i$.

The L_i should be such that it is possible to infer the statistical significance of the pattern of spatial association at location i. More formally, this requires the operationalization of a statement such as

$$\text{Prob}\,[L_i > \delta_i] \leq \alpha_i, \tag{2}$$

where δ_i is a critical value, and α_i is a chosen significance or pseudo significance level, for example, as the result of a randomization test.

The second requirement of a LISA, that is, its relation to a global statistic, may be stated formally as

$$\sum_i L_i = \gamma\Lambda, \tag{3}$$

where Λ is a global indicator of spatial association and γ is a scale factor. In other words, the sum of the local indicators is proportional to a global indicator. For the latter, a statement such as

$$\text{Prob}\,[\Lambda > \delta] \leq \alpha, \tag{4}$$

indicates significant spatial association over the whole data set.

Identification of Local Spatial Clusters

Local spatial clusters, sometimes referred to as *hot spots*, may be identified as those locations or sets of contiguous locations for which the LISA is significant. Similar to the rationale behind the significance tests for the G_i and G_i^* statistics of Getis and Ord (1992), the general LISA can be used as the basis for a test on the null hypothesis of no local spatial association. However, in contrast to what holds for the G_i and G_i^* statistics, general results on the distribution of a generic LISA may be hard to obtain. This is similar to the problems encountered in

deriving distributions for global statistics, for which typically only approximate or asymptotic results are available.[1] An alternative is the use of a conditional randomization or permutation approach to yield empirical so-called pseudo significance levels (for example, as in Hubert 1987). The randomization is conditional in the sense that the value y_i at a location i is held fixed (that is, not used in the permutation) and the remaining values are randomly permuted over the locations in the data set. For each of these resampled data sets, the value of L_i can be computed. The resulting empirical distribution function provides the basis for a statement about the extremeness (or lack of extremeness) of the observed statistic, relative to (and conditional on) the values computed under the null hypothesis (the randomly permuted values). In practice, this is straightforward to implement, since for each location only as many values as there are in the neighborhood set need to be resampled. Note that this same approach can also easily be applied to the G_i and G_i^* statistics.

A complicating factor in the assessment of significance of LISAs is that the statistics for individual locations will tend to be correlated, as pointed out by Ord and Getis (1994) in the context of their G_i and G_i^* statistics. In general, whenever the neighborhood sets J_i and J_k of two locations i and k contain common elements, the corresponding L_i and L_k will be correlated. Due to this correlation, and the associated problem of multiple comparisons, the usual interpretation of significance will be flawed. Moreover, it is typically impossible to derive the exact marginal distribution of each statistic and the significance levels must be approximated by Bonferroni inequalities or following the approach suggested in Sidák (1967).[2] This means that when the overall significance associated with the multiple comparisons (correlated tests) is set to α, and there are m comparisons, then the individual significance α_i should be set to either α/m (Bonferroni) or $1 - (1 - \alpha)^{1/m}$ (Sidák). The latter procedure, which yields slightly sharper bounds, is suggested by Ord and Getis (1994), with $m = n$, that is, the number of observations.[3] Note that the use of Bonferroni bounds may be too conservative for the LISA of individual locations. For example, if m is indeed taken to equal the number of observations, then an overall significance of $\alpha = 0.05$ would imply individual levels of $\alpha_i = 0.0005$ in a data set with one hundred observations, possibly revealing only very few if any "significant" locations. However, since the correlation between individual statistics is due to the common elements in the neighborhood sets, only for a small number of locations k will the statistics actually be correlated with an individual L_i. For example, on a regular lattice using the queen criterion of contiguity, first-order neighbors (ignoring border and corner cells) will have four common elements in their neighborhood sets, second-order neighbors three, and higher-order neighbors none. Clearly, the number of common neighbors does not change with the number of observations, so that using the latter in the computation of the Bonferroni bounds may be overly conservative. Hence, while it is obvious that some correction to the individual significance levels is needed, the extent to which it is indeed necessary to take $m = n$ remains to be further investigated.

[1] With the exception of the results in Tiefelsdorf and Boots (1994), the general statement by Cliff and Ord (1981, p. 46) still holds: "except for very small lattices, exact evaluation of the distribution function is impractical and approximations must be found."

[2] An application of these procedures to the interpretation of the significance of a spatial correlogram was earlier suggested by Oden (1984).

[3] Note that the Sidák approach only holds when the statistics under consideration are multivariate normal, which is unlikely to be the case for the general class of LISA statistics [see also Savin (1980) for an extensive discussion of the relative merits of various notions of bounds].

Indication of Local Instability

The indication of local patterns of spatial association may be in line with a global indication, although this is not necessarily the case. In fact, it is quite possible that the local pattern is an aberration that the global indicator would not pick up, or it may be that a few local patterns run in the opposite direction of the global spatial trend. The second requirement in the definition of a LISA statistic is imposed to allow for the decomposition of a global statistic into its constituent parts. This is of interest to assess the extent to which the global statistic is representative of the average pattern of local association. If the underlying process is stable throughout the data, then one would expect the local indications to show little variation around their average. In other words, local values that are very different from the mean (or median) would indicate locations that contribute more than their expected share to the global statistic. These may be outliers or high leverage points and thus would invite closer scrutiny. This interpretation is roughly similar to the use of Cressie's (1991) pocket plots in geostatistics. By imposing the requirement that the L_i sum to a magnitude that is proportional to a global statistic, their distribution around the mean $\gamma\Lambda/n$ can be evaluated. Extreme L_i can be identified as outliers in this distribution, for example, as those values that are more than two standard deviations from the mean (the two-sigma rule) or more than 1.5 times the interquartile range larger than the third quartile (for example, in a box plot).

This second interpretation of the LISA statistics is similar to the use of a Moran scatterplot to identify outliers and leverage points for Moran's I (Anselin 1993a). In general, it may be more appropriate than the interpretation of locations as hot spots suggested in the previous section when an indicator of global spatial association is significant. In this respect, it is important to note that the Getis-Ord G_i and G_i^* statistics were suggested to detect significant spatial clustering at a local level when global statistics do not provide evidence of spatial association (Getis and Ord 1992, p. 201). Indeed, in their example, Getis and Ord find no global autocorrelation for SIDS cases in North Carolina counties, while several significant local clusters are indicated. However, the opposite case often occurs as well, that is, a strong and significant indication of global spatial association may hide totally random subsets, particularly in large data sets. For example, in an analysis of the 1930 elections in Weimar Germany, O'Loughlin, Flint, and Anselin (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 for six regional subsets. In such an instance, the distribution of the L_i statistic as indicator of local spatial clustering will be affected by the presence of global spatial association. However, the second interpretation of LISA statistics, as indications of outliers or leverage points in the computation of a global statistic is not affected. I return to this issue in section 5.

3. LISA FORM OF FAMILIAR SPATIAL AUTOCORRELATION STATISTICS

Local Gamma

A broad class of spatial association statistics may be based on the general index of matrix association or Γ index, originally outlined in Mantel (1967). The application of the Γ index to spatial autocorrelation in a wide range of contexts is described in a series of papers by Hubert and Golledge (for example, Hubert 1985; Hubert, Golledge, and Costanzo 1981; Hubert et al. 1985; Costanzo,

Hubert, and Golledge 1983).[4] Such an index consists of the sum of the cross products of the matching elements a_{ij} and b_{ij} in two matrices of similarity, say **A** and **B**, such that

$$\Gamma = \sum_i \sum_j a_{ij} b_{ij}. \tag{5}$$

Measures of spatial association are obtained by expressing spatial similarity in one matrix (for example, a contiguity or spatial weights matrix) and value similarity in the other. Different measures of value similarity yield different indices for spatial association. For example, using $a_{ij} = x_i x_j$ yields a Moran-like measure, setting $a_{ij} = (x_i - x_j)^2$ yields a Geary-like index, while taking $a_{ij} = |x_i - x_j|$ results in an indicator equivalent to the one suggested by Royaltey, Astrachan, and Sokal (1975) [see, for example, Anselin (1986) for details on the implementation].

Since the Γ index is a simple sum over the subscript i, a local Gamma index for a location i may be defined as

$$\Gamma_i = \sum_j a_{ij} b_{ij}. \tag{6}$$

Similar to what holds for the global Γ measure, different measures of value similarity will yield different indices of local spatial association. It is easy to see that the Γ_i statistics sum to the global measure Γ. It is possible that the distribution of the individual Γ_i can be approximated using the principles outlined by Mielke (1979) and Costanzo et al. (1983), though this is likely to be complex, and beyond the current scope. On the other hand, the implementation of a conditional permutation approach is straightforward. This allows the individual Γ_i to be interpreted as indicators of significant local spatial clusters. The second interpretation of the LISA statistic, as a diagnostic for outliers or leverage points can be carried out by comparing the distribution of the Γ_i to Γ/n.

Local Moran

As a special case of the local Gamma, a local Moran statistic for an observation i may be defined as

$$I_i = z_i \sum_j w_{ij} z_j, \tag{7}$$

where, analogous to the global Moran's I, the observations z_i, z_j are in deviations from the mean, and the summation over j is such that only neighboring values $j \in J_i$ are included. For ease of interpretation, the weights w_{ij} may be in row-standardized form, though this is not necessary, and by convention, $w_{ii} = 0$.

It can be easily seen that the corresponding global statistic is indeed the familiar Moran's I. The sum of local Morans is

$$\sum_i I_i = \sum_i z_i \sum_j w_{ij} z_j, \tag{8}$$

while Moran's I is

[4] Note that this statistic also forms the basis for the derivation of the distribution of Moran's I and Geary's c statistics in Cliff and Ord (1981, p. 23 and chapter 2). In Getis (1991), this index is applied to integrate spatial association statistics and spatial interaction models into a common framework.

$$I = (n/S_0) \sum_i \sum_j w_{ij} z_i z_j / \sum_i z_i^2, \tag{9}$$

or

$$I = \sum_i I_i / \left[S_0 \left(\sum_i z_i^2 / n \right) \right], \tag{10}$$

where $S_0 = \sum_i \sum_j w_{ij}$. Using the same notation as Cliff and Ord (1981, p. 45), and taking $m_2 = \sum_i z_i^2/n$ as the second moment (a consistent, but not unbiased estimate of the variance), the factor of proportionality between the sum of the local and the global Moran is, in the notation of (3),

$$\gamma = S_0 m_2. \tag{11}$$

Note that for a row-standardized spatial weights matrix, $S_0 = n$, so that $\gamma = \sum_i z_i^2$, and for standardized variables (that is, with the mean subtracted and divided by the standard deviation), $m_2 = 1$, so that $\gamma = S_0$. Also, the same type of results obtain if instead of (7) each local indicator is divided by m_2, which is a constant for all locations. In other words, the local Moran would then be computed as

$$I_i = (z_i/m_2) \sum_j w_{ij} z_j, \tag{12}$$

The moments for I_i under the null hypothesis of no spatial association can be derived using the principles outlined by Cliff and Ord (1981, pp. 42–46) and a reasoning similar to the one by Getis and Ord (1992, pp. 190–92). For example, for a randomization hypothesis, the expected value turns out to be

$$E[I_i] = -w_i/(n-1), \tag{13}$$

with w_i as the sum of the row elements, $\sum_j w_{ij}$, and the variance is found as

$$\begin{aligned} \text{Var}\,[I_i] &= w_{i(2)}(n - b_2)/(n-1) \\ &\quad + 2w_{i(kh)}(2b_2 - n)/(n-1)(n-2) - w_i^2/(n-1)^2, \end{aligned} \tag{14}$$

with $b_2 = m_4/m_2^2$, $m_4 = \sum_i z_i^4/n$ as the fourth moment, $w_{i(2)} = \sum_{j \neq i} w_{ij}^2$, and $2w_{i(kh)} = \sum_{k \neq i} \sum_{h \neq i} w_{ik} w_{ih}$. The details of the derivation are given in Appendix A.

A test for significant local spatial association may be based on these moments, although the exact distribution of such a statistic is still unknown. This is further explored in section 5. Alternatively, a conditional randomization approach may be taken, as outlined earlier. Given the structure of the statistic in (12), it follows that only the quantity $\sum_j w_{ij} z_j$ needs to be computed for each permutation (since the z_i/m_2 remains constant). Note that the randomization method applied to (12) will yield the same empirical reference distribution as when applied to the Getis and Ord G_i and G_i^* statistics. Hence, inference based on this nonparametric approach will be identical for the two statistics. This easily follows from considering which elements in the statistics change for each permutation of the data. For example, the G_i statistic for an observation i is defined as

$$G_i = \sum_j w_{ij}(d)\, z_j / \sum_j z_j, \tag{15}$$

where $w_{ij}(d)$ are the elements in a distance-based weights matrix [for details, see Getis and Ord (1992)]. The only aspect of equation (15) that changes with each permutation is the numerator, since the denominator does not depend on the spatial allocation of observations. Clearly, this is the same term as the varying part of the numerator in (12). In other words, the pseudo significance levels (that is, the inference) generated with a permutation approach applied to the I_i statistic will be identical to that for a G_i or G_i^* statistic.[5]

The interpretation of the local Moran as an indicator of local instability follows easily from the relation between local and global statistics expressed in equation (11). Specifically, the average of the I_i will equal the global I, up to a factor of proportionality. Extreme contributions may thus be identified by means of simple rules, such as the two-sigma rule, or by identifying outliers in a box plot. Note that this notion of extremeness does not imply that the corresponding I_i are significant in the sense outlined earlier, but only indicates the importance of observation i in determining the global statistic. This similarity to the identification of outliers, leverage and influence points in the Moran scatterplot (Anselin 1993a) will be further examined in the empirical illustration.

Local Geary

Using the same principles as before, a local Geary statistic for each observation i may be defined as

$$c_i = \sum_j w_{ij}(z_i - z_j)^2, \tag{16}$$

or as

$$c_i = (1/m_2) \sum_j w_{ij}(z_i - z_j)^2, \tag{17}$$

using the same notation as before. Using expression (17) (without loss of generality), the summation of the c_i over all observations yields

$$\sum_i c_i = n \left[\sum_i \sum_j w_{ij}(z_i - z_j)^2 / \sum_i z_i^2 \right]. \tag{18}$$

In comparison, Geary's familiar c statistic is

$$c = [(n-1)/2S_0] \left[\sum_i \sum_j w_{ij}(z_i - z_j)^2 / \sum_i z_i^2 \right]. \tag{19}$$

Thus, the factor of proportionality between the sum of the local and the global Geary statistic is, in the notation of (3),

$$\gamma = 2nS_0/(n-1). \tag{20}$$

Clearly, for row-standardized weights, since $S_0 = n$, this factor becomes

[5] See also Ord and Getis (1994) for a discussion of the relationship between their statistics and Moran's I.

$2n^2/(n-1)$. The c_i statistic is interpreted in the same way as the local Gamma and the local Moran.

4. ILLUSTRATION: SPATIAL PATTERNS OF CONFLICT IN AFRICA

A geographical perspective has received much interest in recent years in the analysis of international interactions in general, and of international conflict in particular [see, for example, the review by Diehl (1992)]. Measures of spatial association, such as Moran's I, have been applied to quantitative indices for various types of conflicts and cooperation between nation-states, such as those contained in the COPDAB data base (Azar 1980). For such indices of international conflict and cooperation, 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 cooperation was confirmed in a study of the interactions among forty-two African nations, over the period 1966–78, reported in a series of papers by O'Loughlin and Anselin (O'Loughlin and Anselin 1991, 1992; 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, by a Γ index of spatial association, and by the estimates in a mixed regressive, spatial autoregressive model), as well as of spatial heterogeneity in the form of two distinct spatial regimes (as indicated by Getis-Ord G_i^* statistics and the results of a spatial Chow test on the stability of regression coefficients). This phenomenon is thus particularly suited to illustrate the LISA statistics suggested in this paper. The illustration focuses on the two interpretations of the LISA statistics, as indicators of local spatial clusters and as diagnostics for local instability. It is approached from the perspective of exploratory spatial data analysis and the substantive interpretation of the models is not considered here [see O'Loughlin and Anselin (1992) for a more extensive discussion].

The spatial pattern of the index for total conflict is illustrated in the quartile map in Figure 1, with the darkest shade corresponding to the highest quartile [for details on the data sources, see Anselin and O'Loughlin (1992)]. The suggestion of spatial clustering of similar values that follows from a visual inspection of this map is confirmed by a strong positive and significant Moran's I of 0.417, with an associated standard normal z-value of 4.35 ($p < 0.001$), and a Geary c index of 0.584, with associated standard normal z-value of -2.90 ($p < 0.002$).[6] These statistics are computed for a row-standardized spatial weights matrix based on first-order contiguity (common border), given the importance of borders in the study of international conflict (Diehl 1992).

Identification of Local Spatial Clusters

I first focus on a comparison of the identification of local spatial clusters provided by the Getis-Ord G_i^* statistic (as a standardized z-value) and the local Moran I_i indicator presented in equation (12). Note that the former, while being a statistic for local spatial association, is not a LISA in the terminology of section 2, since its individual components are not related to a global statistic of spatial association. This requirement is not needed for the identification of significant local spatial clusters, but it is important for the second interpretation of a LISA, as a diagnostic of local instability in measures of global spatial

[6] All computations were carried out with the *SpaceStat* software for spatial data analysis (Anselin 1992); the map was created with the *Idrisi* software (Eastman 1992), using the *SpaceStat-Idrisi* interface; other graphics were produced by means of the *SPlus* statistical software.

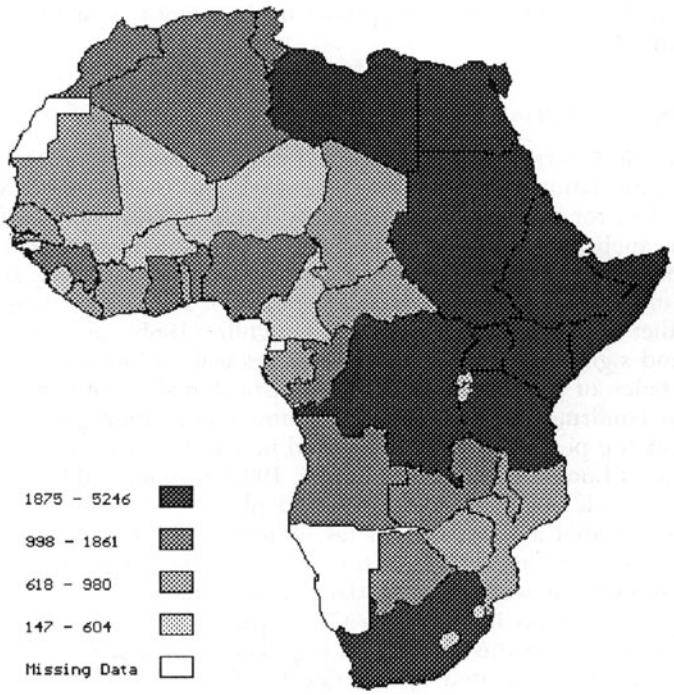

FIG. 1. Total Conflict Index for African Countries (1966–78)

association (for example, in the presence of significant global association), which is discussed in the next section.

Using the same row-standardized weights matrix as for the global measures given earlier, the results for the indicators of local spatial association are reported in the third and fifth columns of Table 1, for each of the forty-two countries in the example. The standardized z-value for I_i, computed by subtracting the expected value (13) and dividing by the standard deviation [the square root of (14)], is listed in the sixth column. Two indications of significance are given, one based on an approximation by the normal distribution, p_n (in the seventh column of Table 1) and one derived from conditional randomization, using a sample of 10,000 permutations, p_r (in the last column of Table 1).[7] As mentioned earlier, the pseudo significance obtained by means of a conditional randomization procedure is identical for G_i^* and I_i. While this may suggest that the normal approximation shown to hold for G_i^* (listed in column four) and assessed in detail in Ord and Getis (1994) may be valid for the I_i statistic as well, this has not been demonstrated. In fact, evidence from some initial Monte Carlo experiments in section 5 seems to indicate otherwise.

Note that the two statistics measure different concepts of spatial association. For the G_i^* statistic, a positive value indicates a spatial clustering of high values, and a negative value a spatial clustering of low values, while for the I_i, a positive value indicates spatial clustering of *similar* values (either high or low), and negative values a clustering of *dissimilar* values (for example, a location with high values surrounded by neighbors with low values), as in the interpretation of the

[7] More precisely, the sample consists of the original observed value of the statistic and the values computed for 9,999 conditionally randomized data sets.

TABLE 1
Measures of Local Spatial Association

Id	Country	G_i^*	p	I_i	$z(I_i)$	p_n	p_r
1	Gambia	−0.984	0.1626	0.375	0.428	0.3342	0.4727
2	Mali	−1.699	0.0447	0.464	1.482	0.0692	0.0456
3	Senegal	−1.463	0.0717	0.257	0.623	0.2667	0.0270
4	Benin	−1.301	0.0966	0.194	0.484	0.3142	0.0612
5	Mauritania	−0.605	0.2726	0.097	0.269	0.3940	0.4111
6	Niger	−1.049	0.1471	0.231	0.774	0.2193	0.2404
7	Ivory Coast	−1.417	0.0782	0.290	0.788	0.2154	0.0611
8	Guinea	−1.449	0.0737	0.183	0.519	0.3020	0.0365
9	Burkina Faso	−1.751	0.0400	0.508	1.479	0.0695	0.0339
10	Liberia	−1.041	0.1490	0.186	0.398	0.3452	0.1333
11	Sierra Leone	−0.870	0.1921	0.265	0.444	0.3286	0.4006
12	Ghana	−1.103	0.1351	0.148	0.326	0.3721	0.0885
13	Togo	−0.991	0.1610	0.219	0.462	0.3219	0.1894
14	Cameroon	−1.133	0.1285	0.259	0.711	0.2387	0.1706
15	Nigeria	−1.173	0.1205	0.114	0.306	0.3798	0.0851
16	Gabon	−0.789	0.2150	0.204	0.349	0.3634	0.3139
17	CAR	1.174	0.1203	−0.442	−1.046	0.1477	0.0613
18	Chad	0.463	0.3218	−0.105	−0.225	0.4111	0.2125
19	Congo	−0.203	0.4198	0.011	0.079	0.4684	0.4734
20	Zaire	2.023	0.0216	0.710	2.591	0.0048	0.0404
21	Angola	1.235	0.1085	0.118	0.270	0.3936	0.0999
22	Uganda	3.336	**0.0004**	1.943	4.928	**0.0000**	0.0031
23	Kenya	3.503	**0.0002**	1.197	3.060	**0.0011**	0.0016
24	Tanzania	1.098	0.1360	0.272	0.973	0.1652	0.1898
25	Burundi	0.774	0.2194	−0.484	−0.872	0.1915	0.1040
26	Rwanda	1.457	0.0725	−0.752	−1.613	0.0534	0.0285
27	Somalia	1.183	0.1184	0.453	0.731	0.2324	0.1266
28	Ethiopia	2.627	0.0043	0.725	1.422	0.0775	0.0090
29	Zambia	0.753	0.2258	0.042	0.219	0.4134	0.1934
30	Zimbabwe	−0.200	0.4209	−0.010	0.033	0.4868	0.4041
31	Malawi	0.212	0.4161	−0.229	−0.388	0.3490	0.2088
32	Mozambique	−0.288	0.3868	0.017	0.114	0.4545	0.4728
33	South Africa	−0.868	0.1927	−0.183	−0.480	0.3156	0.1435
34	Lesotho	−0.298	0.3827	−0.419	−0.423	0.3361	0.2341
35	Botswana	0.041	0.4837	−0.004	0.039	0.4845	0.3691
36	Swaziland	−0.659	0.2548	0.017	0.063	0.4749	0.4128
37	Morocco	0.022	0.4913	−0.097	−0.111	0.4557	0.4995
38	Algeria	−0.363	0.3583	−0.010	0.040	0.4841	0.4139
39	Tunesia	0.579	0.2813	0.005	0.046	0.4818	0.1804
40	Libya	2.553	0.0053	0.804	2.300	0.0107	0.0133
41	Sudan	4.039	**0.0000**	2.988	9.898	**0.0000**	**0.0003**
42	Egypt	4.421	**0.0000**	6.947	10.679	**0.0000**	0.0058

global Moran's I. This explains the sign differences between the values in the third and fifth columns of Table 1 (for example, for the first sixteen countries in the table). Following the suggestion by Ord and Getis (1994), a Bonferroni bounds procedure is used to assess significance. With an overall α level of 0.05, the individual significance levels for each observation should be taken as 0.05/42, or 0.0012.[8] Given this conservative procedure, the normal approximation for both the G_i^* and the I_i show the same four countries to exhibit local

[8] For $\alpha = 0.10$, the corresponding individual significance level is 0.0024. Since normality was not demonstrated, the original Bonferroni bounds were used, rather than the slightly sharper Sidák procedure suggested in Ord and Getis (1994). This does not affect the interpretation of the results in Table 1, since the difference between the two only appears at the fifth significant digit. For example, for $\alpha = 0.05$, the Bonferroni bound is 0.001190, while the Sidák bounds are 0.001221.

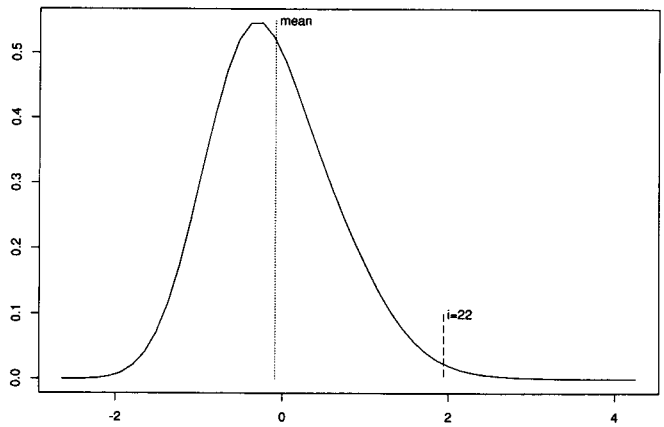

FIG. 2. Density of Randomized Local Moran for Uganda ($i = 22$)

spatial clustering (with the significance levels in bold type in Table 1). They are Uganda (22), Kenya (23), Sudan (41), and Egypt (42), which themselves form a cluster in the northeast of Africa, part of the so-called Shatterbelt.[9] This spatial clustering (or spatial autocorrelation) of both the G_i^* and the LISA statistics is a result of the way they are constructed, and should be kept in mind when visually interpreting a map of LISAs (or G_i^*).

The conditional randomization approach provides a still more conservative picture of (pseudo) significant local spatial clustering, with only Sudan meeting the Bonferroni bound for an overall $\alpha = 0.05$. For this country, two out of the 9,999 statistics computed from the randomized samples exceed the observed one, clearly labeling the latter as "extreme."[10] Of the other three previously significant countries, only Kenya comes close to the threshold (with a pseudo significance of 0.0016), but both Uganda (0.0031) and Egypt (0.0058) fall short of even the bounds for an overall $\alpha = 0.10$.

Some insight into the reasons for the differences in interpretation between the normal approximation and the randomization strategy can be gained from Figure 2, which shows the empirical distribution of the I_i for the 10,000 samples used in the computation of the pseudo significance for Uganda (22). This country was chosen since it has different significance indications between the two criteria, and it is not a boundary or corner location (it has five neighbors, which is about average for the sample). The density function in Figure 2 is smoothed, using a smoothing parameter of twice the interquartile distance. The sample average and the observed value are indicated on the figure (the latter with the label "$i = 22$"). The density under the curve for values larger than 1.943 (the observed value) is 0.0031, indicating its extremeness (but not significance according to the Bonferroni criterion). The distribution is clearly non-normal, and heavily skewed to the right (skewness is 0.7997). Its average of −0.0904 is smaller than the expected value under the null hypothesis for observation 22, which is −0.0244. In addition, its standard deviation of 0.6340 is more than 1.5 times the value that would be expected under the theoretical null distribution, or 0.3991 [the square root of expression (14)].

[9]The identification numbers in parentheses correspond to the labels in the Moran scatterplot of Figure 3.

[10]The Bonferroni bound for an overall significance level of $\alpha = 0.01$ would be 0.0002.

The differences between the empirical density in Figure 2 and (i) the theoretical moment and (ii) an approximation of the null distribution by a normal raise two important issues. First, the normal may not be an appropriate approximation, and higher order moments may have to be used, as in the approximation to the global Γ statistics in Costanzo, Hubert, and Golledge (1983). However, it may also be that the sample size and/or the number of neighbors in this example (*respectively, forty-two and five*) are too small for a valid approximation by the normal.[11] Secondly, and more importantly, the moments under the null hypothesis are derived assuming that each value is equally likely at any location, which is inappropriate in the presence of global spatial association. In other words, the theoretical moments in (13) and (14) do not reflect the latter. This is appropriate when the objective is to detect local spatial clusters in the absence of global spatial association [for example, as was the stated goal in Getis and Ord (1992)], but is not correct when global spatial association is present (as is the case in the example considered here). While the *z*-values for both G_i^* and I_i would suffer from this problem, the conditional randomization strategy does not, since it treats the observations *as if* they were spatially uncorrelated. This issue is revisited in section 5.

Indication of Local Instability

The second interpretation of a LISA is as a diagnostic for outliers with respect to a measure of global association, in this example Moran's *I*. The I_i statistics are compared to the insights provided by the Moran scatterplot, suggested by Anselin (1993a) as a device to achieve a similar objective, that is, to visualize local instability in spatial autocorrelation. Note that the Moran scatterplot is not a LISA in the sense of this paper, since no indication of significant local spatial clustering is obtained. The principle behind the interpretation of the Moran scatterplot is that many statistics for global association are of the form $x'\mathbf{A}x/x'x$, where x is a vector of observations (in deviations from the mean) and $\mathbf{A}$ is a matrix of known elements. In the case of Moran's *I*, the $\mathbf{A}$ is the row-standardized spatial weights matrix $\mathbf{W}$. Given this form for the statistic, it may be visualized as the slope of a linear regression of $\mathbf{W}x$ on x [see also Anselin (1980) for the interpretation of Moran's *I* as a regression coefficient]. A scatterplot of $\mathbf{W}x$ on x [similar to a spatial lag scatterplot in geostatistics, for example, as in Cressie (1991)], with the linear regression line superimposed, provides insight into the extent to which individual $(\mathbf{W}x_i, x_i)$ pairs influence the global measure, exert leverage, or may be interpreted as outliers, based on the extensive set of standard regression diagnostics (for example, Cook 1977; Hoaglin and Welsch 1978; Belsley, Kuh, and Welsch 1980).

The Moran scatterplot for the African conflict data is given as Figure 3, with the individual countries labeled as in Table 1. The $(\mathbf{W}x_i, x_i)$ pairs are given for standardized values, so that "outliers" may be easily visualized as points further than two units away from the origin. In Figure 3, both Sudan (41) and Egypt (42) have values for total conflict that are more than two standard deviations higher than the mean (on the horizontal axis of Figure 3), while Egypt also has values for the spatial lag that are twice the mean (vertical axis of Figure 3). The use of standardized values also allows the Moran scatterplots for different variables to be comparable. The four quadrants in Figure 3 correspond to the four types of spatial association. The lower left and upper right quadrants indicate spatial clustering of similar values: low values (that is, less than the mean) in

[11] See Getis and Ord (1992, pp. 191–92) for the importance of both sample size and the number of neighbors for the normal approximation of the G_i and G_i^* statistics.

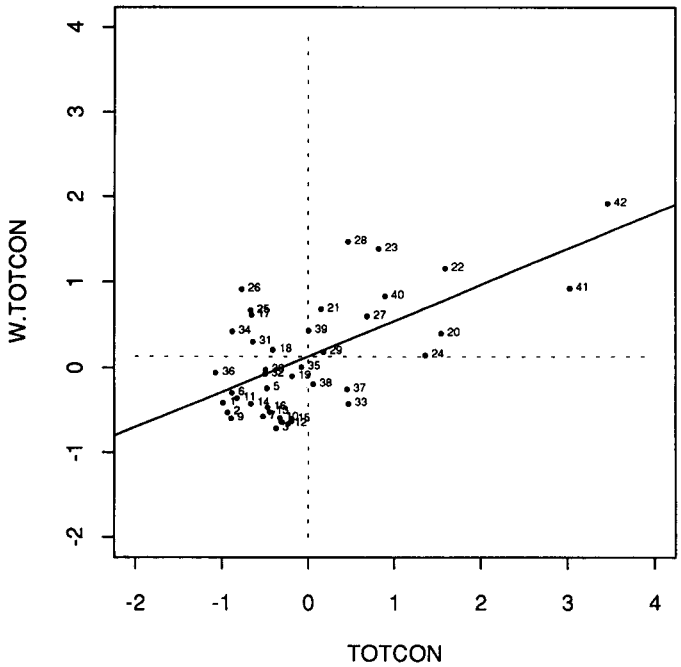

FIG. 3. Moran Scatterplot for Total Conflict ($I = 0.417$)

the lower left and high values in the upper right. Stated differently, the lower left pairs would correspond to negative values of the G_i and G_i^*, and the upper right pairs to positive values. With the I_i statistics, no distinction is possible between the two forms of association since both result in a positive sign. The upper left and lower right quadrants of Figure 3 indicate spatial association of dissimilar values: low values surrounded by high neighboring values for the former, and high values surrounded by low values for the latter. These correspond to I_i statistics with a negative sign. Since they are not cross-product statistics, the G_i and G_i^* statistics do not capture this form of spatial association.

While the overall pattern of spatial association is clearly positive, as indicated by the slope of the regression line (Moran's I), eleven observations show association between dissimilar values: eight in the upper left quadrant, also shown as light islands within the darkest clusters of Figure 1; and three in the lower right quadrant (Algeria, 38, Morocco, 37, and South Africa, 33), surrounded by countries in the first and second quartile in Figure 1. This may indicate the existence of different regimes of spatial association.

The application of regression diagnostics for leverage to the scatterplot suggests that two observations deserve closer scrutiny. The highly significant local spatial association for Sudan (41) and Egypt (42) finds a match with the indication of leverage provided by the diagonal elements of the hat matrix. These are respectively 0.247 (for Sudan) and 0.316 (for Egypt), both distinctly larger than the usual cutoff of $2k/n$ (where k is the number of explanatory variables in the regression, or 2 in this example), or 0.095.[12] The third largest hat value of 0.085

[12]The diagonal elements of the hat matrix $H = X(X'X)^{-1}X'$, with X as the matrix of observations on the explanatory variables in a regression, are well known indicators of leverage. See, for example, Hoaglin and Welsch (1978).

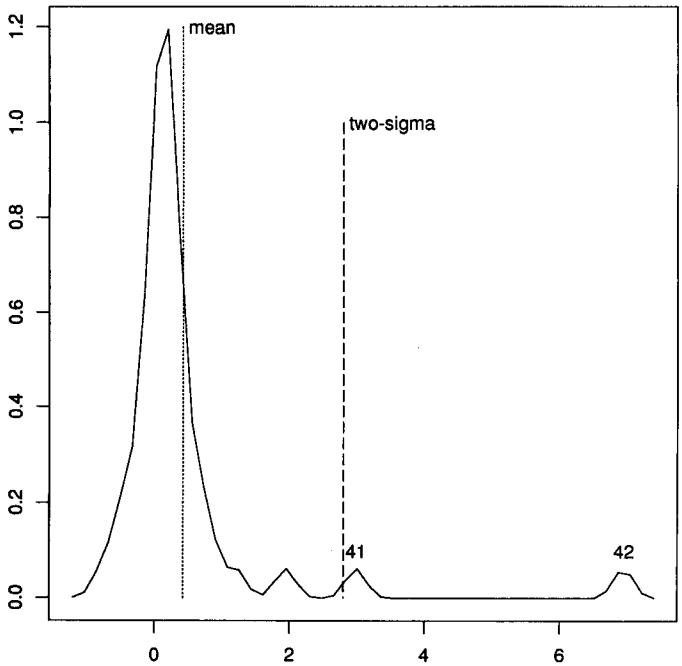

FIG. 4. Local Moran Outliers

for Uganda, 22, does not exceed this threshold. One may be tempted to conclude that the elimination of Egypt and Sudan from the sample would void the indication of spatial association, but this is not the case. Without both countries, Moran's I drops to 0.254, but its associated z-value is 2.53 (using the randomization null hypothesis), which is still highly significant at $p < 0.006$.

The distribution of I_i statistics for the sample can similarly be exploited to provide an indication of outliers or leverage points. In Figure 4, this is illustrated by means of a simple two-sigma rule. The mean of the distribution of the I_i is Moran's I, or 0.417, and twice the standard deviation from the mean corresponds to the value of 2.798. Clearly, this is exceeded by both Sudan (41), with a value of 2.988 for I_i, and by Egypt (42), with a value of 6.947. While this is obviously not a test in a strict sense, it provides useful insight into the special nature of these two observations. All four indicators are in agreement in this respect, that is, the G_i^* and I_i as measures of local spatial clusters, and the Moran scatterplot and I_i as indicators of outliers. The substantive interpretation of the special nature of these observations is beyond the scope of the exploratory data analysis. The role of the latter is to point them out and by doing so to aid in the suggestion of possible explanations or hypotheses. Alternatively, the indication of "strange" observations may point to data quality problems, such as coding mistakes, or, in the case of spatial analysis, problems with the choice of the spatial weights matrix.

5. MONTE CARLO EVIDENCE: GLOBAL AND LOCAL SPATIAL ASSOCIATION

Two issues raised by the results of the empirical illustration in the previous section are revisited here by means of some initial Monte Carlo experiments. The first pertains to the distribution of the local Moran I_i statistic under the

null hypothesis of no (global) spatial autocorrelation. The second issue is the distribution of the local statistic when global spatial autocorrelation is present, and its implication for the assessment of significance. This may also have relevance for the distribution of the G_i and G_i^* statistics in this situation, since their distribution under the null is also based on the absence of global association. As pointed out earlier, it is quite common to study local association in the presence of global association, for example, this is the case in the illustration presented in the previous section. This second issue also has relevance for the assessment of outliers or local instability, that is, the second interpretation of the local Moran. It is well known that many spatial processes that produce spatially autocorrelated patterns also generate spatial heterogeneity. For example, this is the case for the familiar spatial autoregressive process (Anselin 1990). The spatial heterogeneity indicated by LISAs, based on a null hypothesis of no spatial association may therefore be a natural characteristic of the spatial process, and not an indication of local pockets of nonstationarity.

Two sets of experiments were carried out, one based on the same spatial weights matrix as for the African example (with $n = 42$), the other on the weights matrix for a 9 by 9 regular grid, using the queen notion of contiguity (with $n = 81$). Both weights matrices were used in row-standardized form. For each of these configurations, 10,000 random samples were generated with increasing degrees of spatial autocorrelation, constructed by means of a simple spatial autoregressive transformation. More formally, given a vector ε of randomly generated standard normal variates, a spatially autocorrelated landscape was generated as a vector y:

$$y = (\mathbf{I} - \rho\mathbf{W})^{-1}\varepsilon, \tag{21}$$

where ρ is the autoregressive parameter, taking values of 0.0, 0.3, 0.6, and 0.9, and $\mathbf{I}$ is a n by n identity matrix. While the resulting samples will be spatially autocorrelated for nonzero values of ρ, there is no one-to-one match between the value of ρ and the global Moran's I. As is well known, the latter is capable of detecting many different forms of spatial association, and is not linked to a specific spatial process as the sole alternative hypothesis.

Distribution of the Local Moran under the Null Hypothesis

The distribution of the standardized z-values that correspond to the I_i statistic was considered in detail for two selected observations, the location corresponding to Uganda, $i = 22$, for the African weights matrix, and the location corresponding to the central cell, $i = 41$, for the regular lattice. Not only are the dimensions of the data sets different in the two examples ($n = 42$ and $n = 81$), but also the number of neighbors differ for the observations under consideration, as they are respectively 5 and 8. The moments of each distribution for the z-values, based on the 10,000 replications, are given in the first row of Table 2. While the mean and standard deviation are roughly in accordance with those for a standard normal distribution, the kurtosis and to a lesser extent the skewness are not. This is further illustrated by the density graph in Figure 5 (for $n = 81$), which clearly shows the leptokurtic nature of the distribution and the associated thicker tails (compared to a normal density). The density graph for the African case is very similar and is not shown. Instead, a quantile-quantile plot for the African example is given in Figure 6, to further illustrate the lack of normality. While there is general agreement in the central section of the two distributions (total agreement would be shown as a perfect

TABLE 2
Moments of Local Moran with Global Spatial Autocorrelation[a]

	$n = 42$				$n = 81$			
ρ	Mean	St.Dev.	Skew	Kurtosis	Mean	St.Dev.	Skew	Kurtosis
0.0	0.0032	0.9895	−0.2599	7.993	0.0236	1.0356	−0.1073	7.711
0.3	0.2491	1.0730	0.7417	7.635	0.2666	1.1733	0.9320	7.853
0.6	0.5833	1.2144	1.4748	7.454	0.6057	1.3958	1.7475	8.673
0.9	1.0782	1.3465	1.5357	5.850	0.8961	1.4690	2.4073	11.114

a. z-values for local Moran; 10,000 replications, using observation 22 for $n = 42$ and observation 41 for $n = 81$.

linear fit), at the tails, that is, where it matters in terms of significance, this clearly is not the case. A more rigorous assessment of the distribution, based on an asymptotic chi-squared test constructed around the third and fourth moments (Kiefer and Salmon 1983) strongly rejects the null hypothesis of normality in both cases.

This more extensive assessment confirms (in a controlled setting) the earlier suggestion implied by the discrepancy between the significance levels under the normal approximation and the conditional randomization in Table 1. Note that the African example in Table 1 exhibited significant global spatial autocorrelation, while the simulations here do not (by design). Further results are needed to see whether larger sample sizes or higher numbers of neighbors are needed before normality is obtained. However, from the initial impressions gained here it would seem that the normal approximation may be inappropriate, and that higher moments (given the values for skewness and kurtosis in Table 2) would be needed in order to obtain a better approximation [for example, as in Costanzo, Hubert, and Golledge (1983) for the Γ statistic].

The implications of these results for inference in practice are that even when no global spatial autocorrelation is present, the significance levels indicated by a normal approximation will result in an over-rejection of the null hypothesis for a given α_i Type I error. Clearly, a more conservative approach is warranted, although the exact nature of the corrections to the α_i awaits further investigation. In the meantime, a conditional randomization approach provides a useful alternative.

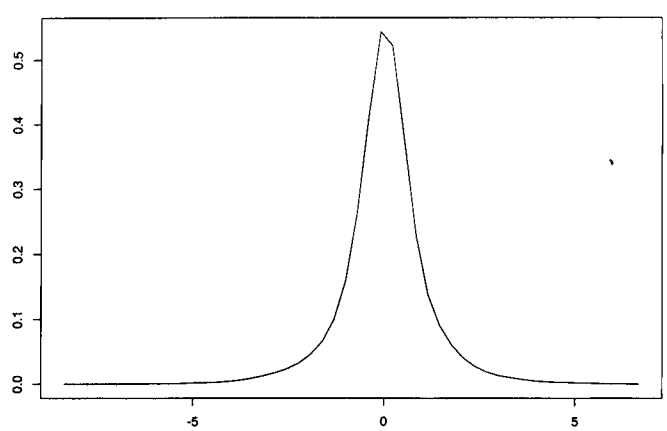

FIG. 5. Density of z-value for Local Moran ($n = 81$; 10,000 Replications)

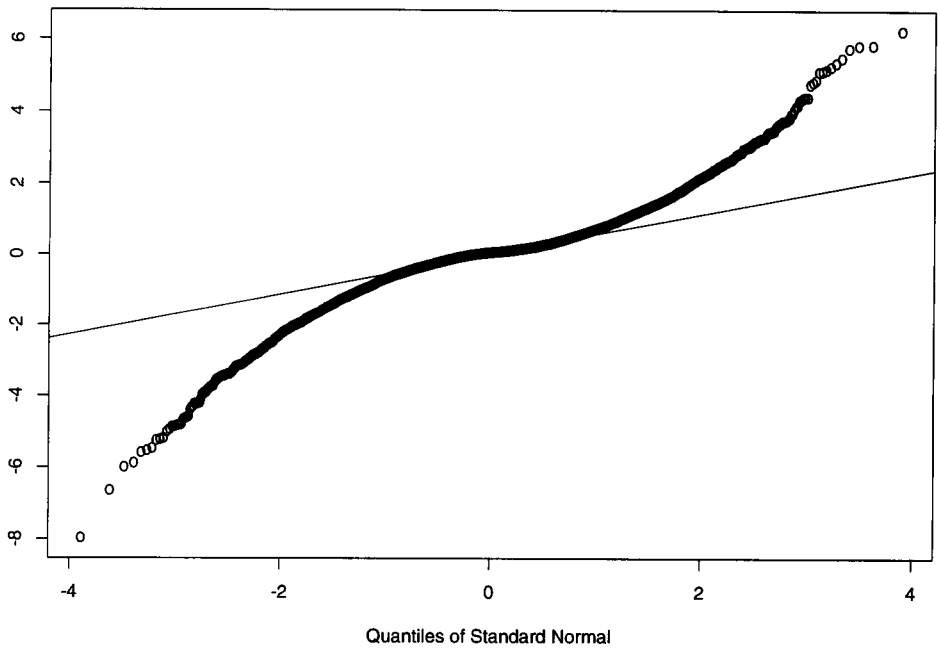

FIG. 6. Quantiles of z-values of Local Moran against the Normal Distribution ($n = 42$; 10,000 Replications)

Distribution of the Local Moran in the Presence of Global Spatial Autocorrelation

The presence of global spatial autocorrelation has a strong influence on the moments of the distribution of the local Moran, as indicated by the results in Table 2. Both mean and standard deviation increase with spatial autocorrelation, but the most significant effect seems to be on the skewness of the distribution. This is further illustrated by the box plots in Figure 7 (for the case with $n = 42$; the results for the larger sample size are similar). As ρ increases, the distribution becomes more and more asymmetric around the median, while both the interquartile range and the median itself increase as well. Clearly, in the presence of global spatial autocorrelation, the moments indicated by the expressions (13) and (14) become inappropriate estimates of the moments of the actual distribution. The same problem would seem to also affect the distribution for the Getis and Ord G_i and G_i^* statistics, since they are derived in a similar manner. Consequently, inference for tests on local spatial clusters that ignores this effect is likely to be misleading. The magnitude of the error cannot be derived from the initial Monte Carlo results reported here, and further investigation is needed, both empirical and analytical. In practice, inference based on the pseudo significance levels indicated by a conditional randomization approach seems to be the only viable alternative.

Evidence of Outliers in the Presence of Global Spatial Autocorrelation

A final issue to be examined is how the magnitude of global spatial autocorrelation affects the distribution of the I_i around the sample mean (the global Moran's I), which is used to detect outliers. In contrast to the earlier experi-

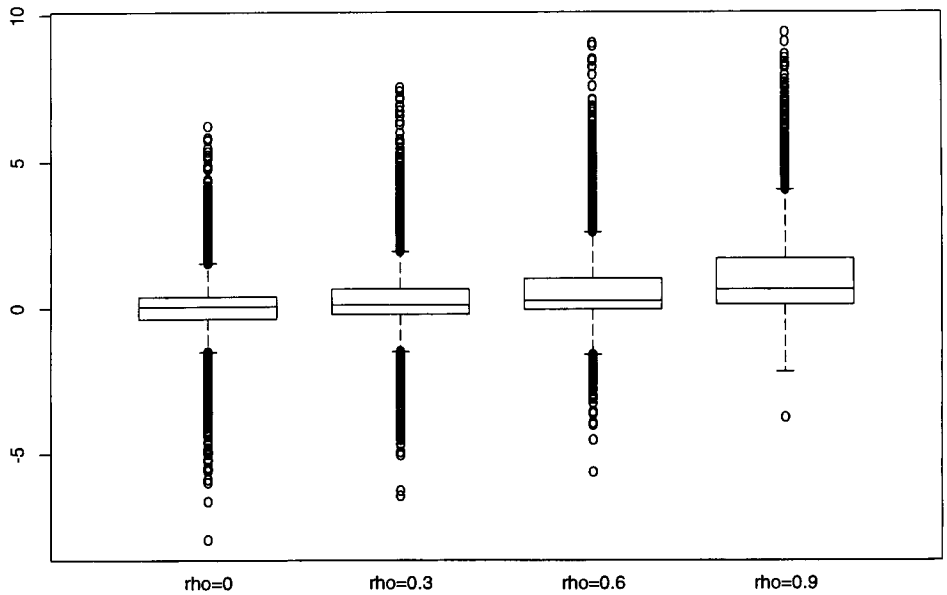

FIG. 7. Box Plots of Local Moran z-value with Spatial Autocorrelation ($n = 42$; 10,000 Replications)

ments, the focus is not on I_i for an individual location, but on how the spread of the statistics in each sample is affected by the strength of global spatial autocorrelation. In Table 3, the average over the 10,000 replications of twice the standard deviation around the mean in each replication is listed, as well as the average (over the 10,000 replications) number of outliers indicated by using the two sigma rule. With increasing global spatial autocorrelation, both the spread and the number of "outliers" increases. This implies that in the presence of a high degree of spatial autocorrelation, several extreme values of the I_i statistic are to be expected as a "normal" result of the heterogeneity induced by a spatial autoregressive process. In practice, this is not much different from the usual treatment of outliers, and without further evidence, it is not possible to state in a rigorous manner which extreme values are to be expected and which are unusual observations. However, as an exploratory device, the lack of symmetry of the distribution of the I_i around the global I, and/or the presence of very large values provides insight into the stability of the indication of global spatial association over the sample.

TABLE 3
Two-Sigma Rule with Global Spatial Autocorrelation[a]

	$n = 42$		$n = 81$	
ρ	2σ	Outliers	2σ	Outliers
0.0	1.0199	3	0.7538	5
0.3	1.1171	3	0.8675	6
0.6	1.3519	4	1.1280	7
0.9	1.7112	5	1.7017	9

a. 2σ computed as average 2σ over 10,000 replications; outliers are median number of observations more than 2σ from the mean in each sample.

6. CONCLUSION

The general class of local indicators of spatial association suggested in this paper serves two main purposes. Firstly, the LISA generalize the idea underlying the Getis and Ord G_i and G_i^* statistics to a broad class of measures of local spatial association. Secondly, by directly linking the local indicators to a global measure of spatial association, the decomposition of the latter into its observation-specific components becomes straightforward, thus enabling the assessment of influential observations and outliers. It is this dual property that distinguishes the class of LISA from existing techniques, such as the G_i and G_i^* statistics and the Moran scatterplot. The LISA presented here are easy to implement and lend themselves readily to visualization. They thus serve a useful purpose in an exploratory analysis of spatial data, potentially indicating local spatial clusters and forming the basis for a sensitivity analysis (outliers). While the former is more appropriate when no global spatial autocorrelation is present, the latter is particularly useful when there is spatial autocorrelation in the data.

A number of issues remain to be investigated further. The illustration in this paper primarily pertained to the local Moran I_i indices, but the extension to the wider class of LISA statistics can be carried out in a straightforward way. From both the empirical example and the initial simulation experiments, it follows that the null distribution of the local Moran cannot be effectively approximated by the normal, at least not for the small sample sizes employed here. Also, it seems that higher moments may be necessary in order to obtain a better approximation. Furthermore, the uncritical use of the null distribution in the presence of global spatial autocorrelation will give incorrect significance levels. The problem also pertains to the G_i and G_i^* statistics and would suggest that a test for global spatial autocorrelation should precede the assessment of significant local spatial clusters. However, such a two-pronged strategy raises the issue of pretesting and multiple comparisons, and would require an adjustment of the significance levels to reflect this. This further complicates the determination of a proper significance level for an individual LISA, given the built-in correlatedness of measures for adjoining locations. It is clear that some type of bounds procedure is needed, but which degree of correction is sufficient still remains to be addressed.

Finally, the conditional randomization approach suggested here seems to provide a reliable basis for inference for the LISA, both in the absence and in the presence of global spatial autocorrelation.

LITERATURE CITED

Anselin, L. (1980). *Estimation Methods for Spatial Autoregressive Structures*. Ithaca, N.Y.: Regional Science Dissertation and Monograph Series.

——— (1986). "MicroQAP, a Microcomputer Implementation of Generalized Measures of Spatial Association." Department of Geography, University of California, Santa Barbara, Calif.

——— (1988). *Spatial Econometrics: Methods and Models*. Dordrecht: Kluwer Academic.

——— (1990). "Spatial Dependence and Spatial Structural Instability in Applied Regression Analysis." *Journal of Regional Science* 30, 185–207.

——— (1992). *SpaceStat: A Program for the Analysis of Spatial Data*. National Center for Geographic Information and Analysis, University of California, Santa Barbara, Calif.

——— (1993a). "The Moran Scatterplot as an ESDA Tool to Assess Local Instability in Spatial Association." Paper presented at the GISDATA Specialist Meeting on GIS and Spatial Analysis, Amsterdam, The Netherlands, December 1-5 (West Virginia University, Regional Research Institute, Research Paper 9330).

——— (1993b). "Exploratory Spatial Data Analysis and Geographic Information Systems." Paper pre-

sented at the DOSES/Eurostat Workshop on New Tools for Spatial Analysis, Lisbon, Portugal, November 18-20 (West Virginia University, Regional Research Institute, Research Paper 9329).

Anselin, L., and A. Getis (1992). "Spatial Statistical Analysis and Geographic Information Systems." *The Annals of Regional Science* 26, 19–33.

Anselin, L., and J. O'Loughlin (1990). "Spatial Econometric Models of International Conflicts." In *Dynamics and Conflict in Regional Structural Change*, edited by M. Chatterji and R. Kuenne, pp. 325–45. London: Macmillan.

——— (1992). "Geography of International Conflict and Cooperation: Spatial Dependence and Regional Context in Africa." In *The New Geopolitics*, edited by M. Ward, pp. 39–75. Philadelphia, Penn.: Gordon and Breach.

Azar, E. (1980). "The Conflict and Peace Data Bank (COPDAB) Project." *Journal of Conflict Resolution* 24, 143–52.

Belsley, D, E. Kuh, and R. Welsch (1980). *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. New York: Wiley.

Casetti, E. (1972). "Generating Models by the Expansion Method: Applications to Geographical Research." *Geographical Analysis* 4, 81–91.

——— (1986). "The Dual Expansion Method: An Application for Evaluating the Effects of Population Growth on Development." *IEEE Transactions on Systems, Man and Cybernetics* SMC-16, 29–39.

Cliff, A., and J. K. Ord (1981). *Spatial Processes: Moels and Applications*. London: Pion.

Costanzo, C. M., L. J. Hubert, and R. G. Golledge (1983). "A Higher Moment for Spatial Statistics." *Geographical Analysis* 15, 347–51.

Cook, R. (1977). "Detection of Influential Observations in Linear Regression." *Technometrics* 19, 15–18.

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

Diehl, P. (1992). "Geography and War: A Review and Assessment of the Empirical Literature." In *The New Geopolitics*, edited by M. Ward, pp. 121–37. Philadelphia, Penn.: Gordon and Breach.

Eastman, R. (1992). *IDRISI Version 4.0*. Worcester, Mass.: Clark University Graduate School of Geography.

Foster, S. A., and W. Gorr (1986). "An Adaptive Filter for Estimating Spatially-Varying Parameters: Application to Modeling Police Hours in Response to Calls for Service." *Management Science* 32, 878–89.

Getis, A. (1991). "Spatial Interaction and Spatial Autocorrelation: A Cross-Product Approach." *Environment and Planning* A 23, 1269–77.

Getis, A., and K. Ord (1992). "The Analysis of Spatial Association by Use of Distance Statistics." *Geographical Analysis* 24, 189–206.

Gorr, W., and A. Olligschlaeger (1994). "Weighted Spatial Adaptive Filtering: Monte Carlo Studies and Application to Illicit Drug Market Modeling." *Geographical Analysis* 26, 67–87.

Griffith, D. A. (1978). "A Spatially Adjusted ANOVA Model." *Geographical Analysis* 10, 296–301.

——— (1992). "A Spatially Adjusted N-Way ANOVA Model." *Regional Science and Urban Economics* 22, 347–69.

——— (1993). "Which Spatial Statistics Techniques Should Be Converted to GIS Functions? In *Geographic Information Systems, Spatial Modelling and Policy Evaluation*, edited by M. M. Fischer and P. Nijkamp, pp. 101–14. Berlin: Springer Verlag.

Haslett, J., R. Bradley, P. Craig, A. Unwin, and C. Wills (1991). "Dynamic Graphics for Exploring Spatial Data with Applications to Locating Global and Local Anomalies." *The American Statistician* 45, 234–42.

Hoaglin, D., and R. Welsch (1978). "The Hat Matrix in Regression and ANOVA." *The American Statistician* 32, 17–22.

Hubert, L. J. (1985). "Combinatorial Data Analysis: Association and Partial Association." *Psychometrika* 50, 449–67.

——— (1987). *Assignment Methods in Combinatorial Data Analysis*. New York: Marcel Dekker.

Hubert, L. J., R. Golledge, and C. M. Costanzo (1981). "Generalized Procedures for Evaluating Spatial Autocorrelation." *Geographical Analysis* 13, 224–33.

Hubert, L. J., R. Golledge, C. M. Costanzo, and N. Gale (1985). "Measuring Association between Spatially Defined Variables: An Alternative Procedure." *Geographical Analysis* 17, 36–46.

Jones, J. P., and E. Casetti (1992). *Applications of the Expansion Method*. London: Routledge.

Kiefer, N., and M. Salmon (1983). "Testing Normality in Econometric Models." *Economics Letters* 11, 123–8.

Kirby, A., and M. Ward (1987). "The Spatial Analysis of Peace and War." *Comparative Political Studies* 20, 293–313.

Mantel, N. (1967). "The Detection of Disease Clustering and a Generalized Regression Approach." *Cancer Research* 27, 209–20.

Mielke, P. W. (1979). "On Asymptotic Non-Normality of Null Distributions of MRPP Statistics." *Communications Statistical Theory and Methods A* 8, 1541–50.

Oden, N. L. (1984). "Assessing the Significance of a Spatial Correlogram." *Geographical Analysis* 16, 1–16.

O'Loughlin, J. (1986). "Spatial Models of International Conflicts: Extending Current Theories of War Behavior." *Annals, Association of American Geographers* 76, 63–80.

O'Loughlin, J. and L. Anselin (1991). "Bringing Geography Back to the Study of International Relations: Dependence and Regional Context in Africa, 1966–1978." *International Interactions* 17, 29–61.

——— (1992). "Geography of International Conflict and Cooperation: Theory and Methods." In *The New Geopolitics*, edited by M. Ward, pp. 11–38. Philadelphia, Penn.: Gordon and Breach.

O'Loughlin, J., C. Flint, and L. Anselin (1994). "The Political Geography of the Nazi Vote: Context, Confession, and Class in the Reichstag Election of 1930." Annals, Association of American Geographers 84, 351–80.

Openshaw, S. (1993). "Some Suggestions concerning the Development of Artificial Intelligence Tools for Spatial Modelling and Analysis in GIS." In *Geographic Information Systems, Spatial Modelling and Policy Evaluation*, edited by M. M. Fischer and P. Nijkamp, pp. 17–33. Berlin: Springer Verlag.

Openshaw, S., C. Brundson, and M. Charlton (1991). "A Spatial Analysis Toolkit for GIS." *EGIS '91, Proceedings of the Second European Conference on Geographical Information Systems*, pp. 788–96. Utrecht: EGIS Foundation.

Openshaw, S., A. Cross, and M. Charlton (1990). "Building a Prototype Geographical Correlates Exploration Machine." *International Journal of Geographical Information Systems* 4, 297–311.

Ord, J. K., and A. Getis (1994). "Distributional Issues concerning Distance Statistics." Working paper.

Royaltey, H., E. Astrachan, and R. Sokal (1975). "Tests for Patterns in Geographic Variation." *Geographical Analysis* 7, 369–96.

Savin, N. E. (1980). "The Bonferroni and the Scheffé Multiple Comparison Procedures." *Review of Economic Studies* 67, 255–73.

Sidák, Z. (1967). "Rectangular Confidence Regions for the Means of Multivariate Normal Distributions." *Journal of the American Statistical Association* 62, 626–33.

Sokal, R., N. Oden, B. Thomson, and J. Kim (1993). "Testing for Regional Differences in Means: Distinguishing Inherent from Spurious Spatial Autocorrelation by Restricted Randomization." *Geographical Analysis* 25, 199–210.

Tiefelsdorf, M., and B. Boots (1994). "The Exact Distribution of Moran's *I*." *Environment and Planning A* (forthcoming).

APPENDIX A

The moments of the local Moran statistic can be derived using the results in Cliff and Ord (1981, pp. 42–46). Using (12), the expected value of I_i under the randomization hypothesis is

$$\mathrm{E}[I_i] = \left(\sum_j w_{ij}/m_2 \right) \mathrm{E}[z_i z_j].$$

The value of the expectations term is

$$\mathrm{E}[z_i z_j] = -m_2/(n-1),$$

based on equation (2.37) of Cliff and Ord (1981, p. 45). Consequently, the expected value of I_i becomes

$$\mathrm{E}[I_i] = -w_i/(n-1),$$

with w_i as the sum of the row elements, $\sum_j w_{ij}$. Obviously, in the case a row-standardized weights matrix is used, this sum will be one.

To obtain the second moment, the following expression must be evaluated:

$$\mathrm{E}[I_i^2] = (1/m_2^2)\mathrm{E}\left[z_i^2\left(\sum_j w_{ij}z_j\right)^2\right]$$

or

$$\mathrm{E}[I_i^2] = (1/m_2^2)\mathrm{E}\left[z_i^2\left(\sum_{j\neq i} w_{ij}^2 z_j^2 + \sum_{k\neq i}\sum_{h\neq i} w_{ik}w_{ih}z_k z_h\right)\right],$$

for which the following results are important, based on equation (2.39) of Cliff and Ord (1981, p. 46):

$$\mathrm{E}[z_i^2 z_j^2] = (nm_2^2 - m_4)/(n-1);$$

$$\mathrm{E}[z_i^2 z_k z_h] = (2m_4 - nm_2^2)/(n-1)\,(n-2)$$

with $m_4 = \sum_i z_j^4/n$ as the fourth moment. The first weights term in the expectation consists of the sum of all weights squared, or, $w_{i(2)} = \sum_{j\neq i} w_{ij}^2$, and the second is twice the sum of the cross products (avoiding identical subscripts), or, $2w_{i(kh)} = \sum_{k\neq i}\sum_{h\neq i} w_{ik}w_{ih}$. After combining terms, the second moment is found as

$$\mathrm{E}[I_i^2] = (1/m_2^2)\,[w_{i(2)}(nm_2^2 - m_4)/(n-1)$$

$$+2w_{i(kh)}(2m_4 - nm_2^2)/(n-1)\,(n-2)],$$

which simplifies somewhat after using $b_2 = m_4/m_2^2$, to

$$\mathrm{E}[I_i^2] = w_{i(2)}(n-b_2)/(n-1) + 2w_{i(kh)}(2b_2 - n)/(n-1)(n-2).$$

Consequently, the variance of I_i is

$$\mathrm{Var}[I_i] = w_{i(2)}(n-b_2)/(n-1) + 2w_{i(kh)}(2b_2 - n)/(n-1)(n-2)$$

$$-w_i^2/(n-1)^2.$$