J Geograph Syst (2001) 3:369–385

Journal of Geographical Systems

Developing a bivariate spatial association measure: An integration of Pearson's *r* and Moran's *I*

Sang-Il Lee

Department of Geography, The Ohio State University, 1036 Derby Hall, 154 North Oval Mall, Columbus, OH 43210-1361, USA (e-mail: slee@geography.ohio-state.edu)

Received: 07 November 2000 / Accepted: 02 August 2001

Abstract. This research is concerned with developing a bivariate spatial association measure or spatial correlation coefficient, which is intended to capture spatial association among observations in terms of their point-to-point relationships across two spatial patterns. The need for parameterization of the *bivariate spatial dependence* is precipitated by the realization that aspatial bivariate association measures, such as Pearson's correlation coefficient, do not recognize spatial distributional aspects of data sets. This study devises an *L* statistic by integrating Pearson's *r* as an aspatial *bivariate association* measure and Moran's *I* as a univariate *spatial association* measure. The concept of a spatial smoothing scalar (SSS) plays a pivotal role in this task.

Key words: Spatial association, spatial correlation, Moran's *I*, spatial smoothing, SDA

JEL classification: C13, C15, C49

1 Introduction

The identification and measurement of the spatial clustering of a geographical variable have been a focal issue in both confirmatory and exploratory spatial data analyses. Global measures of spatial autocorrelation, such as Moran's *I*, provide summary statistics for overall spatial clustering (Moran 1948; Geary 1954; Cliff and Ord 1981; Goodchild 1986; Griffith 1987; Odland 1988). Corresponding local indices, such as local Moran's I_i, allow researchers to explore local variations in spatial dependence by measuring each area's relative contribution to the corre-

An earlier version of this paper was presented at the 2001 meeting of the Association of American Geographers, winning the Best Student Paper Award of the AAG's Spatial Analysis and Modeling (SAM) Specialty Group. The author wishes to thank Michael Tiefelsdorf and Lawrence A. Brown at the Ohio State University for their invaluable comments on earlier drafts.

sponding global measure (Getis and Ord 1992; Ord and Getis 1995; Anselin 1995, 1996). These efforts are part of a broader endeavor to *spatialize* general statistics by recognizing that regular statistical assumptions seldom hold for spatial data. For example, data points in geographically referenced data sets are not independent from one another due to spatial autocorrelation (spatial dependence), and spatial distributions often display significant local variations resulting in the presence of discrete spatial regimes within a study area (spatial heterogeneity or nonstationarity) (Anselin 1988, 1990; Anselin and Griffith 1988; Haining 1990; Anselin and Getis 1992; Getis and Ord 1996; Goodchild 1996; Fotheringham 1997; Fischer 1999).

Following on this *spatial turn* or 'renaissance of spatial analysis' (Unwin 1996, p. 541), the present study is concerned with developing a *bivariate spatial association* measure. Whereas univariate spatial association measures focus on the spatial clustering of observations in terms of a single variable, a bivariate spatial association measure captures the relationship between two variables, taking the topological relationship among observations into account. In other words, it parameterizes the *bivariate spatial dependence*. The need for a bivariate spatial association measure reflects that aspatial measures, such as Pearson's correlation coefficient (r), do not recognize the spatial distributional aspects of data sets (Haining 1990, 1991). For example, one can generate $n!$ different pairs of spatial patterns from two variables consisting of n observations with different values; the Pearson's rs will be identical among pairs, but the degree of visual correspondence will vary. What differentiates the pairs of spatial patterns with an identical Pearson's r is the *spatial dependence* of bivariate correspondence. With respect to this, Hubert et al. (1985) make a distinction between 'point-to-point association' (the relationship *within* a pair at each location) and 'spatial association' (the relationship *between* distinct pairs across locations). This conceptual decomposition of 'association' should be statistically reconciled by an integrative measure. In short, a bivariate spatial association measure needs to capture the *spatial co-patterning* by calibrating both *numerical co-varying* ('point-to-point association') and *spatial clustering* ('spatial association').

In this paper, I first clarify the need for a bivariate spatial association measure, which revolves around defining the concept of bivariate spatial dependence. I demonstrate that a bivariate spatial association measure should contain information on the univariate spatial association of both variables in its equation. Second, I formulate the concept of a *spatial smoothing scalar* (SSS) as representative of the univariate spatial association by decomposing Moran's I. SSS is an element of the Moran's I equation and is defined as the degree of smoothing when a variable is transformed to its spatial lag. Third, I develop a bivariate spatial association measure (L) that is presented as a product of three elements: SSSs of two variables and Pearson's r between spatial lags of the variables. Fourth, I illustrate the computational process and usefulness of the measure with a hypothetical data set. A significance testing procedure is also discussed.

2 Parameterization of the bivariate spatial dependence

The concept of spatial dependence points to the propensity for nearby locations to influence each other and to possess similar attributes (Anselin

1988; Anselin and Griffith 1988; Anselin and Getis 1992). At the heart of problems that spatial dependence may cause lies the loss of information that each observation carries. When spatial dependence is present, the information from observations is less than would have been obtained from independent observations, because a certain amount of the information carried by each observation is duplicated by other observations in the cluster (Haining 1990, p. 40–41; Anselin 1990). This loss of information invalidates most statistical tests, because it lowers the effective number of degrees of freedom (Goodchild 1996, p. 244). For example, in the context of the OLS regression, the presence of spatial autocorrelation causes misleading significance tests and measures of fit (Anselin and Griffith 1988, p. 16; Fotheringham and Rogerson 1993, p. 11). In the same vein, the significance testing for Pearson's correlation coefficient may be flawed when similar associations are spatially clustered since the degree of freedom cannot be calibrated by $n - 2$ (Bivand 1980; Richardson and Hémon 1981; Clifford and Richardson 1985; Clifford et al. 1989; Haining 1991; Dutilleul 1993).

A numeric vector with n data points with different values can generate $n!$ different permutations or arrangements, each of which has a distinct order of data points. When referenced by spatial locations, different orders of a numeric vector result in different spatial patterns with different degrees of the *univariate spatial dependence* or spatial clustering. To illustrate, I generate three different spatial patterns from a numeric vector on a hypothetical space consisting of 37 hexagons (Fig. 1). Since the spatial patterns are three out of all possible 37!/(7!17!13!) geographical variables, they share the same numerical properties: a mean of 1.838 and a variance of 0.514. Differences in the univariate spatial dependence or spatial clustering among the three patterns are parameterized by Moran's I.

In the bivariate context, $n!$ different pairs can be drawn from *two* numeric vectors, when elements in each variable are all different (note that the corresponding data points are bound in a permutation process). The $n!$ different pairs are identical to one another in terms of the point-to-point association, e.g. Pearson's r. Since data points are spatially indexed, however, different pairs are characterized by different degrees of the *bivariate spatial dependence*, thus different levels of *spatial co-patterning* are revealed. To illustrate, three patterns in Fig. 1 are now seen as different variables, and the three pairs, A-B, B-C, and C-A, show identical relationships in terms of Pearson's r (0.422). The association of A-B, however, shows a higher level of bivariate spatial dependence or spatial co-patterning than those of B-C and C-A: the association of A and B displays the highest level of spatial clustering of hexagons sharing the same values between the two maps.

Having realized that a pair of variables under investigation represents only a particular case of all possible bivariate spatial associations, one may wish to devise a measure that effectively differentiates the associations by integrating the two concepts of 'association'. The importance of a conceptual disintegration and computational reintegration of 'point-to-point association' and 'spatial association' can be clarified by two conceptual illustrations. In the first example, a bivariate spatial association is seen as a Pearson's r between two sets of local Moran's I_is. This illustrates that locations with an identical value might be differently recognized in measuring a bivariate association if their relations with neighboring locations are different.

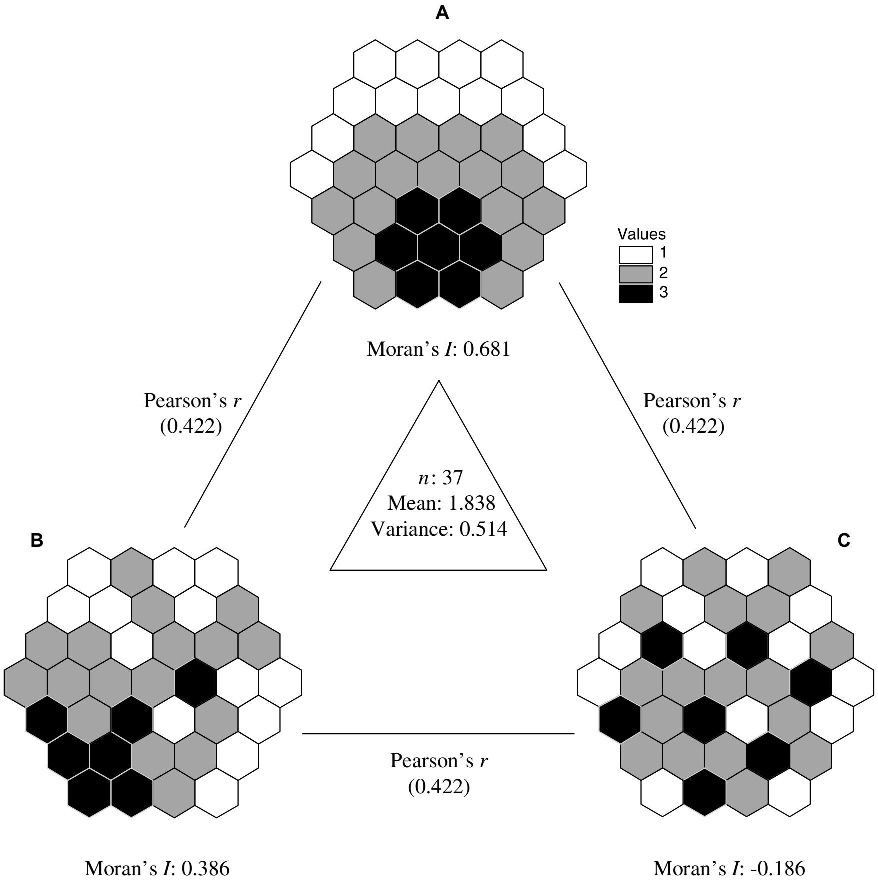


Fig. 1. Three spatial realizations of a hypothetical numeric vector

The second conceptual illustration is provided by the global Moran's I of what can be termed *local Pearson's* r_i. A local Pearson's r_i captures the degree of numerical correspondence between two values at a location, and is simply calculated by multiplying two z-scores of the values, each of which is standardized by the mean and standard deviation of each variable. The mean value of local Pearson's r_is is nothing but a *global* Pearson's r. When local Pearson's r_is are mapped, a global Moran's I captures the degree of spatial dependence of point-to-point associations across locations. This provides a conceptual foundation for a bivariate spatial association measure and suggests that an integration of Moran's I as a univariate *spatial association* measure and Pearson's r as an aspatial *bivariate association* measure may lead to a feasible measure. As can be seen from Fig. 1, the level of bivariate spatial dependence is determined by the level of univariate spatial dependence of variables involved when the point-to-point association is held constant. This further suggests that a bivariate spatial association measure should be a composite of three elements: *univariate spatial associations of two variables and their point-to-point association in a certain form.*

Although the need for a bivariate spatial association measure has long been recognized, the only comprehensive attempt to devise a parametric bivariate spatial association measure is Wartenberg's work (1985), which proposed a matrix algebraic form for the *bivariate* Moran's *I* intended to provide an alternative correlation matrix for a spatial principal components analysis. His measure has drawbacks, however, which will be discussed subsequently.

3 Decomposition of Moran's *I* and formulation of spatial smoothing scalar (SSS)

As demonstrated in the previous section, a certain form of measuring the univariate spatial association should be derived to construct a bivariate spatial association measure. I illustrate that it can be done by decomposing the Moran's *I* equation. Since both Pearson's *r* and Moran's *I* are variants of Mantel's general cross-product association measure (Mantel 1967; Hubert et al. 1981, 1985; Hubert and Golledge 1982), I begin with clarifying computational similarities between Pearson's correlation coefficient and Moran's *I*. Doing this involves: (i) rewriting the equation for Moran's *I* using the concept of *spatial lag*; (ii) decomposing the equation into two parts, Pearson's *r* between a variable and its spatial lag, and a spatial smoothing scalar (SSS) of the variable; and (iii) proposing SSS as representative of the univariate spatial association.

Pearson's correlation coefficient (*r*) for variables *X* and *Y* is computed by:

$$r_{X,Y} = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2}\sqrt{\sum_i (y_i - \bar{y})^2}} \tag{1}$$

and, Moran's *I* is given by:

$$I_X = \frac{n}{\sum_i \sum_j c_{ij}} \cdot \frac{\sum_i \sum_j c_{ij}(x_i - \bar{x})\left(x_j - \bar{x}\right)}{\sum_i (x_i - \bar{x})^2} \tag{2}$$

where c_{ij} is an element of a binary connectivity matrix (**C**) whose elements have a value of 1 for contiguous spatial units and 0 for the others. Following Griffith (1995) and Tiefelsdorf et al. (1999), I define **W** as a row-standardized version of **C** (each element is divided by its row-sum). When **W** is applied, (2) can be simplified to:

$$I_X = \frac{\sum_i \sum_j w_{ij}(x_i - \bar{x})\left(x_j - \bar{x}\right)}{\sum_i (x_i - \bar{x})^2} \tag{3}$$

The homology between Moran's *I* and Pearson's *r* is more obvious when the former is rewritten by utilizing the concept of a spatial lag (SL), which is composed of weighted averages of neighbors defined by the spatial weights matrix (Anselin 1988; Anselin and Smirnov 1996), and is given as:

$$\tilde{x}_i = \sum\nolimits_j w_{ij}x_j \tag{4}$$

where $\tilde{x}_i$ is an element of a spatial lag vector of X ($\tilde{X}$). By applying (4) to (3) and restating the denominator of (3), we have

$$I_X = \frac{\sum_i (x_i - \bar{x})(\tilde{x}_i - \bar{x})}{\sqrt{\sum_i (x_i - \bar{x})^2}\sqrt{\sum_i (x_i - \bar{x})^2}} \tag{5}$$

Since Moran's I measures the relationship between observations and their neighbors, it may be useful to address the substantive meaning of Moran's I by comparing (5) and the equation for Pearson's r between a variable (X) and its SL ($\tilde{X}$), which is given by:

$$r_{X,\tilde{X}} = \frac{\sum_i (x_i - \bar{x})(\tilde{x}_i - \bar{\tilde{x}})}{\sqrt{\sum_i (x_i - \bar{x})^2}\sqrt{\sum_i (\tilde{x}_i - \bar{\tilde{x}})^2}} \tag{6}$$

where $\bar{\tilde{x}}$ denotes the mean of the SL vector of X. Here, a comparison of (5) and (6) provides an important insight into a practical understanding of Moran's I. A major difference occurs in the right side of the denominator. Since elements in SL can be seen as smoothed values of the original ones, the variance of SL (given by the right side of the denominator in (6)) is always smaller than that of the original values (given by the right side of the denominator in (5)). In addition, the numerators in (5) and (6) are identical because the difference between them is zero. Now, (5) can be rewritten in terms of (6):

$$I_X = \sqrt{\frac{\sum_i (\tilde{x}_i - \bar{\tilde{x}})^2}{\sum_i (x_i - \bar{x})^2}} \cdot r_{X,\tilde{X}} \tag{7}$$

From (7), Moran's I is seen as a Pearson's r between a variable and its SL scaled by the square root of the ratio of the SL's variance to the original variable's variance (or the ratio of the SL's standard deviation to the original variable's standard deviation). This derivation corresponds to a well-known finding that Moran's I is a regression coefficient when a variable's SL is regressed on the original variable (Anselin 1995; Griffith and Amrhein 1997). By utilizing the general relationship between a regression coefficient in a bivariate regression and Pearson's r between two variables, (7) is easily proved.

The ratio of standard deviations can further be decomposed:

$$I_X = \sqrt{\frac{\sum_i (\tilde{x}_i - \bar{x})^2}{\sum_i (x_i - \bar{x})^2}} \cdot \underbrace{\sqrt{\frac{\sum_i (\tilde{x}_i - \bar{\tilde{x}})^2}{\sum_i (\tilde{x}_i - \bar{x})^2}}}_{\mathrm{A} \cong 1} \cdot r_{X,\tilde{X}} \cong \sqrt{\mathrm{SSS}_X} \cdot r_{X,\tilde{X}} \tag{8}$$

Since the means of the original variable and its SL are expected to be very similar, a part of A in (8) is approximately 1. Further, there is virtually no relation between A and $r_{X,\tilde{X}}$, thus Moran's I, allowing it to be regarded as a random noise. Next, a ratio of two total sums of squares is defined as a *spatial smoothing scalar* (SSS), which is approximately identical to the variance ratio in (7), and is similar to what has been conceptualized as a *variance reducing factor* in general smoothing techniques (Loader 1999: 7). The SSS can be formulated in a general form:

$$\mathrm{SSS}_X = \frac{n}{\sum_i \left(\sum_j v_{ij}\right)^2} \cdot \frac{\sum_i \left(\sum_j v_{ij}(x_j - \bar{x})\right)^2}{\sum_i (x_i - \bar{x})^2} \tag{9}$$

where v_{ij} is an element in a general spatial weights matrix **V**. Note that, when **W** is applied, (9) is reduced to the equation defined in (8). SSS is the degree of smoothing of a geographical variable or spatial pattern when its observations are represented by their corresponding elements in a SL in accordance with a particular smoothing method.

Although an intensive investigation of relationships between the SSS and spatial autocorrelation is beyond the scope of this paper, an initial observation suggests that the SSS reveals substantive information about the spatial clustering of a variable. If a variable is more spatially clustered, its SSS is larger, because variance of the original vector is less reduced when it is transformed to its SL. For example, the SSSs for the three spatial patterns shown earlier are respectively 0.649, 0.418, and 0.175 (Fig. 2). The value of 0.649 for pattern A indicates that the variance of A's SL is approximately 64.5% of that of the original A.

The decomposition of Moran's *I* provides new insights into the univariate spatial association. First, the SSS itself can be seen as a direction-free univariate spatial association measure that theoretically ranges from 0 to 1.

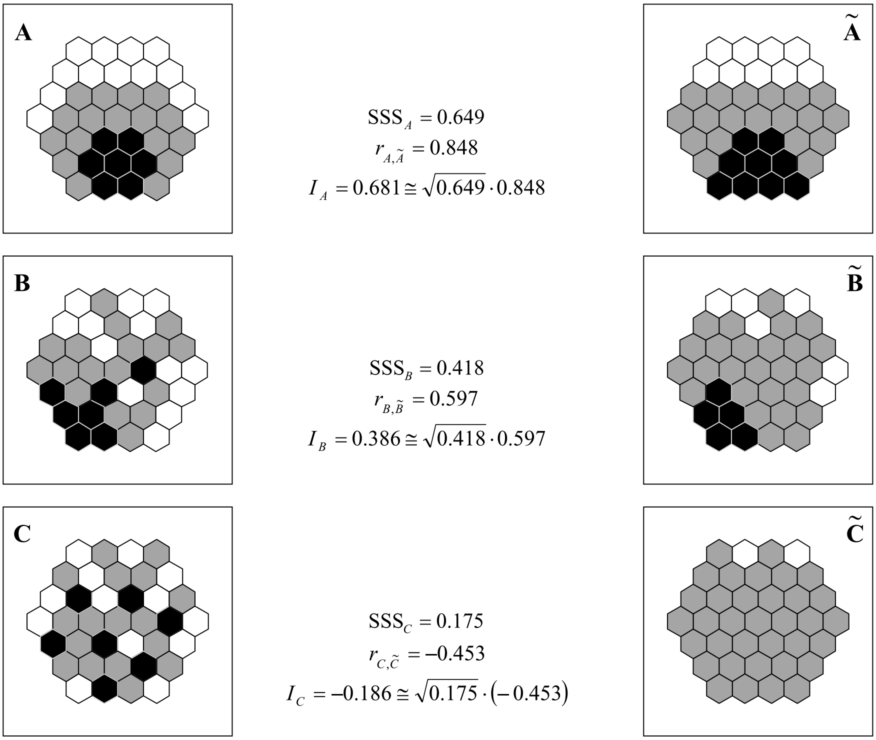

A
$\tilde{A}$
$\mathrm{SSS}_A = 0.649$
$r_{A,\tilde{A}} = 0.848$
$I_A = 0.681 \cong \sqrt{0.649} \cdot 0.848$
B
$\tilde{B}$
$\mathrm{SSS}_B = 0.418$
$r_{B,\tilde{B}} = 0.597$
$I_B = 0.386 \cong \sqrt{0.418} \cdot 0.597$
C
$\tilde{C}$
$\mathrm{SSS}_C = 0.175$
$r_{C,\tilde{C}} = -0.453$
$I_C = -0.186 \cong \sqrt{0.175} \cdot (-0.453)$

Fig. 2. The relationship between the spatial smoothing scalar (SSS) and Moran's *I*

Second, the SSS is a crucial element in the Moran's I equation. The other element, Pearson's r between a variable and its SL, remains a measure of point-to-point association in the sense that very different associations between an area and its neighbors could result in very similar or even identical contributions to Pearson's $r_{X,\tilde{X}}$. For example, if two observations have the same value and their neighbors' (weighted) means are the same, their spatial lag elements will be identical; thus their contributions to Pearson's r between a variable and its SL are identical. However, a neighbors' mean does not take *variance among neighbors* into account: one observation could be surrounded by homogeneous neighbors; the other could be connected to neighbors which are very different from one another. The concept of a *spatially varying variance* or *local instability in variance* is as important as that of a *spatially varying average* in defining a univariate spatial pattern.

As a conclusion, the SSS should be utilized as a representative of the univariate spatial association to obtain an equation for the bivariate spatial association. In other words, the SSSs of two variables should define the bivariate spatial association along with a certain form of Pearson's correlation coefficient between them, allowing the former to spatialize the latter.

4 A bivariate spatial association measure (L)

4.1 Criteria for a bivariate spatial association measure and critiques on Wartenberg's formulation

By reference to findings in the previous section, two criteria can be suggested for developing a bivariate spatial association measure. First, the measure should conform to Pearson's r between two variables in terms of direction and magnitude to a certain extent. Although the measure has an exclusive interest in the *spatial* association among observations, it should retain the direction and magnitude of a point-to-point association between two variables, which requires the inclusion of a certain form of Pearson's correlation between two variables. Second, a bivariate spatial association measure should reflect the degrees of spatial autocorrelation for both variables under investigation. In other words, it should respond to the collective effect of the SSSs of the variables.

The most comprehensive attempt to develop a bivariate spatial association measure by extending Moran's I is Wartenberg's work (1985). He developed a bivariate Moran's I following Mantel's formulations.

$$\mathbf{I} = \frac{\mathbf{Z}^T\mathbf{C}\mathbf{Z}}{\mathbf{1}^T\mathbf{C}\mathbf{1}} \tag{10}$$

where $\mathbf{I}$ is a variable-by-variable Moran correlation matrix, $\mathbf{Z}$ is a case-by-variable matrix whose elements are z-scored, $\mathbf{C}$ is a case-by-case binary connectivity matrix, and $\mathbf{1}$ is a case-by-1 column matrix with all elements being 1s. The diagonal values of $\mathbf{I}$ are Moran's I coefficients for the variables, with each off-diagonal element being a bivariate Moran's I (Griffith (1993, 1995) terms it Cross-MC (Moran coefficient)), which is similar to the cross-correlation approach in geostatistics (Isaaks and Srivastava 1989).

By decomposing the matrix and applying a row-standardized spatial weights matrix, one can write an equation for an off-diagonal element in the matrix **I** between two variables, X and Y:

$$I_{X,Y} = \frac{\sum_i (x_i - \bar{x})(\tilde{y}_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2}\sqrt{\sum_i (y_i - \bar{y})^2}} \cong \sqrt{\mathrm{SSS}_Y} \cdot r_{X,\tilde{Y}} \tag{11}$$

A comparison of (11) and (5) reveals that Wartenberg's bivariate spatial association measure, or bivariate Moran's I, captures a bivariate association between X and the SL of Y, and the association is scaled by square root of the SSS for Y.

Using Wartenberg's formula as a bivariate spatial association measure has two obvious disadvantages that violates the two criteria established before. First, it is conceptually untenable to allow a bivariate spatial association measure to be primarily calibrated by the relationship between a variable and the other variable's SL. Moreover, a bivariate spatial association measure should incorporate both SSSs of two variables in the equation, not just the SSS of a variable. (11) also implies that $I_{X,Y}$ and $I_{Y,X}$ may be different when a row-standardized spatial weights matrix is involved, which nullifies much of Wartenberg's attempt to formulate a spatial principal component analysis using an **I** matrix in (10).

Second, Wartenberg's equation is vulnerable to a reverse of the direction of association. For example, when an area i with a higher-than-average value for both X and Y are surrounded by lower-than-average values, the numerator value in (11) could be given a negative value, because the SL of Y for the area is negative (the right part of the numerator), with the left part being necessarily positive. A simulation observed that most of the associations with negatively autocorrelated Y vectors were assigned negative bivariate association indices. In conclusion, Cross-MC should not be used as a bivariate spatial association measure.

4.2 A bivariate spatial association measure (L)

A bivariate spatial association measure (L) is defined as:

$$L_{X,Y} = \frac{n}{\sum_i \left(\sum_j v_{ij}\right)^2} \cdot \frac{\sum_i \left[\left(\sum_j v_{ij}(x_j - \bar{x})\right) \cdot \left(\sum_j v_{ij}(y_j - \bar{y})\right)\right]}{\sqrt{\sum_i (x_i - \bar{x})^2}\sqrt{\sum_i (y_i - \bar{y})^2}} \tag{12}$$

and, when a row-standardized spatial weights matrix (**W**) is applied, (12) is simplified to:

$$L_{X,Y} = \frac{\sum_i \left[\left(\sum_j w_{ij}(x_j - \bar{x})\right) \cdot \left(\sum_j w_{ij}(y_j - \bar{y})\right)\right]}{\sqrt{\sum_i (x_i - \bar{x})^2}\sqrt{\sum_i (y_i - \bar{y})^2}} \tag{13}$$

Further, when the SL operation is introduced, (13) is transformed to:

$$L_{X,Y} = \frac{\sum_i (\tilde{x}_i - \bar{x})(\tilde{y}_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2}\sqrt{\sum_i (y_i - \bar{y})^2}} \tag{14}$$

Note that $\tilde{x}_i$ and $\tilde{y}_i$ are elements for a location i in X's and Y's SL vectors ($\tilde{X}$ and $\tilde{Y}$). To decompose (14) as undertaken for the Moran's I equation, it is compared to an equation for Pearson's r between SLs, which is given as:

$$r_{\tilde{X},\tilde{Y}} = \frac{\sum_i (\tilde{x}_i - \bar{\bar{x}})(\tilde{y}_i - \bar{\bar{y}})}{\sqrt{\sum_i (\tilde{x}_i - \bar{\bar{x}})^2}\sqrt{\sum_i (\tilde{y}_i - \bar{\bar{y}})^2}} \tag{15}$$

Note that $\bar{\bar{x}}$ and $\bar{\bar{y}}$ are mean values of the SL vectors. By utilizing (15), (14) can be rewritten as:

$$L_{X,Y} = \sqrt{\frac{\sum_i (\tilde{x}_i - \bar{x})^2}{\sum_i (x_i - \bar{x})^2}} \cdot \sqrt{\frac{\sum_i (\tilde{y}_i - \bar{y})^2}{\sum_i (y_i - \bar{y})^2}} \cdot \underbrace{\sqrt{\frac{\sum_i (\tilde{x}_i - \bar{\bar{x}})^2}{\sum_i (\tilde{x}_i - \bar{x})^2}} \cdot \sqrt{\frac{\sum_i (\tilde{y}_i - \bar{\bar{y}})^2}{\sum_i (\tilde{y}_i - \bar{y})^2}} \cdot r_{\tilde{X},\tilde{Y}}}_{\mathrm{A}\cong 1}$$
$$+ \underbrace{\frac{(\bar{\bar{x}} - \bar{x}) \cdot \sum_i (\tilde{y}_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2} \cdot \sqrt{\sum_i (y_i - \bar{y})^2}}}_{\mathrm{B}\cong 0} \tag{16}$$

As in (8), the element of A is approximately 1, and the element of B will be zero when either variable's mean is identical to one of its SL, which is very likely. Then L is redefined as:

$$L_{X,Y} \equiv \sqrt{\mathrm{SSS}_X} \cdot \sqrt{\mathrm{SSS}_Y} \cdot r_{\tilde{X},\tilde{Y}} \equiv \sqrt{\mathrm{BSSS}_{X,Y}} \cdot r_{\tilde{X},\tilde{Y}} \tag{17}$$

Now, L between two variables is calculated by multiplying Pearson's correlation coefficient between their SL vectors by the square root of the product of their SSSs. The product of the SSSs may be termed the *bivariate spatial smoothing scalar* (BSSS), differentiating it from the SSS or *univariate spatial smoothing scalar* (USSS).

Further, a matrix algebraic form for L is provided, when variables are z-transformed:

$$\mathbf{L} = \frac{\mathbf{Z}^T(\mathbf{V}^T\mathbf{V})\mathbf{Z}}{\mathbf{1}^T(\mathbf{V}^T\mathbf{V})\mathbf{1}} \tag{18}$$

where $\mathbf{L}$ is a variable-by-variable bivariate spatial association matrix, $\mathbf{Z}$ is an area-by-variable (z-scored) data matrix, and $\mathbf{V}$ is an area-by-area general spatial weight matrix. Note that, when $\mathbf{W}$ is applied, the denominator is reduced to n. A spatial correlation matrix driven by (18) can be furthered to calibrate a spatial principal components analysis as seen from Wartenberg's attempt (1985).

In addition, it should be noted that the diagonal elements in matrix $\mathbf{L}$ have a particular meaning. From Eq. (14), a diagonal element can be written as:

$$L_{X,X} = \frac{\sum_i (\tilde{x}_i - \bar{x})^2}{\sum_i (x_i - \bar{x})^2} \tag{19}$$

where $L_{X,X}$ is simply the SSS of X defined in (8) and (9). (19) allows a transformation of (17):

$$L_{X,Y} \equiv \sqrt{L_{X,X}} \cdot \sqrt{L_{Y,Y}} \cdot r_{\tilde{X},\tilde{Y}} \tag{20}$$

A higher value in the diagonal of the matrix **L** implies a higher Moran's *I* for the variable, and results in a higher *L* index between the variable and other variables, all other conditions being constant.

In summary, the *L* index as a bivariate spatial association measure is largely determined by Pearson's *r* between two SL vectors, which generates a smoothed version of Pearson's correlation coefficient between the original variables. Pearson's *r* between SLs, then, is scaled by a square root of BSSS (or a product of univariate SSSs) of the variables, which suggests that *L* captures not only the bivariate 'point-to-point association' between two variables, but also the univariate spatial autocorrelation.

4.3 An illustration with a hypothetical data set

For the purpose of illustration, the three different spatial patterns, A, B, and C in Fig. 1 and 2 are utilized (Table 1). A$'$, B$'$, and C$'$ are spatially rotated versions of those patterns, such that the univariate spatial dependence of the original patterns remain unchanged in terms of SSS and Moran's *I*. From Table 1, four things should be acknowledged.

First, the sign in Pearson's *r* between two variables remains unchanged in *L* as long as the sign of Pearson's *r* between their SLs is given accordingly. The only exception is found in the association of A-C$'$, where Pearson's *r* between the two patterns is positive (0.107), but one between their SLs is negative (−0.240). One way of dealing with this problem may be to apply the spatial moving average operation where the weighted mean of neighbors for an area is computed with the area itself being included. This means that the spatial weights matrix **W** as a row-standardized version of **C** is replaced by a matrix of a row-standardized version of a modified **C**, where $c_{ii} = 1$.

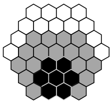
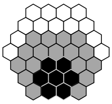
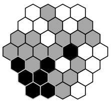
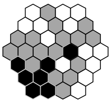
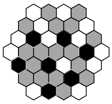
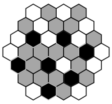
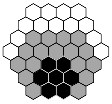
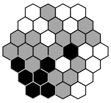
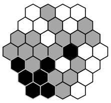
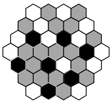
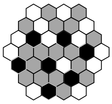
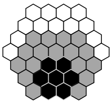
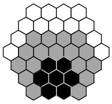
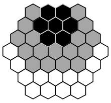
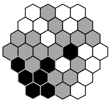
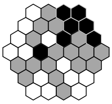
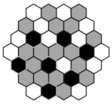
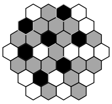
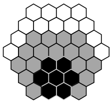
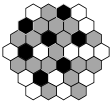
Second, as seen in equation (19), *L* between two identical patterns does not yield a value of 1, and the value changes between pairs of variables (compare A-A, B-B, and C-C in Table 1). This provides a crucial insight into the comparison between two spatial patterns. That is, the bivariate spatial dependence between identical patterns is completely determined by the univariate spatial dependence of the pattern.

Third, *L* differentiates different spatial associations with an identical Pearson correlation coefficient. A-B, B-C, and C-A are identical in terms of Pearson's *r* (0.422); however, they have *L*s respectively of 0.327, 0.154, and 0.214 (Table 1). This implies that *L* is largely determined by the SSSs of the two variables involved when Pearson's *r* is identical. Since a negative *L* indicates a spatial discrepancy, a poorer spatial co-patterning should be given a negative value with a larger amount. This is well illustrated by a comparison between B-B$'$ and C-C$'$: Pearson's correlation coefficients are identical (−0.051), but the spatial discrepancy is much more obvious in B-B$'$, which is reflected in *L* values (−0.162 and −0.024).

Fourth, *L* differentiates different spatial associations with identical SSSs but different Pearson's *r*, which can easily be acknowledged by comparing A-A and A-A$'$, B-B and B-B$'$, and C-C and C-C$'$ in Table 1.

Comparing *L* values among different spatial patterns, one may recognize that the *L* effectively measures similarity/dissimilarity among variables in terms of bivariate associations and their spatial clustering. In computation, the numerical point-to-point association is calibrated largely by Pearson's *r*

Table 1. L with different bivariate spatial associations

Association	Pattern		SSS		Correlation		$L_{X,Y}$ [e]
	X	Y	X[a]	Y[b]	$r_{\tilde{X},\tilde{Y}}$ [c]	$r_{X,Y}$ [d]	
A-A			0.649	0.649	1.000	1.000	0.649
B-B			0.418	0.418	1.000	1.000	0.418
C-C			0.175	0.175	1.000	1.000	0.175
A-B			0.649	0.418	0.628	0.422	0.327
B-C			0.418	0.175	0.577	0.422	0.154
C-A			0.175	0.649	0.634	0.422	0.214
A-A'			0.649	0.649	-0.800	-0.472	-0.512
B-B'			0.418	0.418	-0.388	-0.051	-0.162
C-C'			0.175	0.175	-0.185	-0.051	-0.024
A-C'			0.649	0.175	-0.240	0.107	-0.074

According to equation (17), $\mathrm{e} \cong \sqrt{\mathrm{a}} \cdot \sqrt{\mathrm{b}} \cdot \mathrm{c}$.

between SL vectors, and the spatial association is recognized by the BSSS. Thus, the two elements collectively capture the spatial co-patterning and parameterize the bivariate spatial dependence.

In order to evaluate the usefulness of the L index with real data sets, it was applied to an empirical study on spatio-temporal shifts in the female proportion in the labor force between 1970 and 1990 at the U.S. county level (Brown et al. 2000). The study compared 1970 and 1990 patterns in terms of the female proportion in the labor force, not only for the overall U.S., but also for the Ohio River Valley (ORV) region as a microcosm. Pearson's correlation coefficients between the 1970 and 1990 maps are almost identical between the overall U.S. and ORV (0.701 for the US and 0.705 for the ORV). L statistics, however, provides a value of 0.358 for the US and 0.251 for the ORV. The difference indicates that the US displays a much higher level of bivariate spatial dependence between 1970 and 1990 maps than ORV does. In other words, spatial clustering of temporal continuities in counties is more prevalent for the overall US than the ORV region.

4.4 A note on the significance testing of L

It has been recognized that Moran's I and Geary's c are special cases of Mantel's (1967) generalized cross-product association measure (Cliff and Ord 1981; Hubert et al. 1981), and the associated generalized significance testing method can be used for deriving distributional properties of those indices (Cliff and Ord 1981; Upton and Fingleton 1985). Practically, when two matrices in Mantel's equation being properly defined for the measures, equations for the first two moments of Mantel's statistic (Mantel 1967; Cliff and Ord 1981, p. 23, Eq. 1.44–1.46) bring exactly the same set of values as one computed from commonly used equations based on the randomization assumption (see Cliff and Ord 1981, p. 21, Eq. 1.37, 1.39 and 1.42).

By extending the Mantel's generalized significance testing method as presented by Heo and Gabriel (1998), equations for the expected value and variance can be derived. The full discussion of the method is beyond the scope of the current paper, and will be presented elsewhere (Lee 2001). Only the equation for the expected value for L with a $\mathbf{W}$ is given here as:

$$\mathrm{E}(L) = \frac{\mathrm{tr}\left(\mathbf{W}^T\mathbf{W}\right) - 1}{n - 1} \cdot r_{X,Y} \tag{21}$$

The equation gives an expected value of 0.0840 for the three associations among A, B, and C patterns (note that they are derived from the same numeric vector and the Pearson's correlation coefficients among them are identical).

Mantel's test corresponds to a bound or conditional permutation approach, which is to conduct a large number of permutations where two values for each observation are tied to each other. All of the different permutations, then, have the same point-to-point association (Pearson's r) but different BSSSs. Fig. 3 displays a simulated distribution of 10,000 different L measures. However, it should be noted, as Boots and Tiefelsdorf (2000) demonstrate, that other tessellation systems rather than hexagons may result in different distributional characteristics. The mean of 10,000 permutations in Fig. 3 is 0.0836, which is well approximated by (21).

The Pearson's correlation coefficient of 0.422 is significant at the 99% confidence level (p-value = 0.009) according to the regular t-test, which means that all the three pairs are significantly correlated in terms of the point-to-point association. Fig. 3 provides a platform on which to conduct a *pseudo*-significance test for the measure. The three associations (A-B, B-C, and C-A)

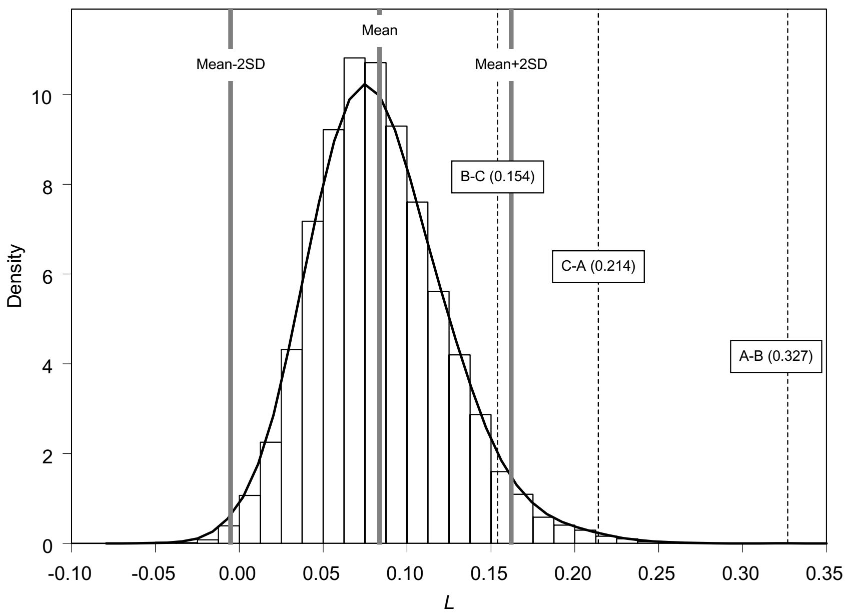


Fig. 3. Distributional properties of L based on a bound permutation approach ($n = 37$, 10,000 permutations)

are given two-tailed p-values of 0.0002, 0.0946, and 0.0080 respectively by 10,000 permutations. This indicates that B-C association is not significant at the 95% confidence level, while the other associations are significant at the 99% confidence level. (It is generally known that 1000 permutations is a reasonable minimum for a test at the 95% confidence level and 5000 at the 99% confidence level (Manly 1997: 83)).

5 Conclusions

This paper developed a bivariate spatial association measure (L) by reference to Moran's I and Pearson's r. A well-designed bivariate spatial association measure should capture the spatial co-patterning by collectively gauging the point-to-point association between two variables and the topological relationship among spatial entities. The concept of a spatial smoothing scalar (SSS) was formulated by decomposing the Moran's I equation. A bivariate spatial association index (L) was defined as an adjusted Pearson's r between variables' spatial lags drawn from the original variables scaled by the square root of the bivariate spatial smoothing scalar, which is the product of univariate spatial smoothing scalars.

The L index makes several contributions to a substantive spatial data analysis:

First, the index provides a complementary measure to Pearson's correlation coefficient, as it effectively captures how much bivariate associations are spatially clustered. In other words, the measure can be used to parameterize the bivariate spatial dependence.

Second, a *local* bivariate spatial association measure can be easily derived from the index as:

$$L_i = \frac{n \cdot \left[\left(\sum_i w_{ij}\left(x_j - \bar{x}\right)\right)\left(\sum_i w_{ij}\left(y_j - \bar{y}\right)\right)\right]}{\sqrt{\sum_i (x_i - \bar{x})^2}\sqrt{\sum_i (y_i - \bar{y})^2}} = \frac{n \cdot (\tilde{x}_i - \bar{x})(\tilde{y}_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2}\sqrt{\sum_i (y_i - \bar{y})^2}} \tag{22}$$

A local L_i first indicates the relative contribution an individual area makes to the global L, and also captures an observation's association with its neighbors in terms of the point-to-point association between the two variables. A spatial distribution of local L_is may allow researchers to explore a *bivariate spatial heterogeneity* in the sense that it may reveal the local instability in relationships between two variables.

Third, the L index can be employed to measure spatial segregation or dissimilarity (e.g., Morrill 1991; Wong 1993; Wardorf 1993; Chakravorty 1996). Since indices of segregation or dissimilarity measure the extent to which a spatial distribution of a racial/ethnic group is correspondent to that of the other group, the bivariate spatial association measure may provide a new insight into the understanding of the relative degree of spatial exclusion between racial/ethnic groups.

Fourth, a matrix of indices for a set of variables can be used to spatialize other multivariate statistical procedures, such as principal component analysis (Wartenberg 1985; Griffith and Amrhein 1997, Chap. 6).

This study will be further developed by subsequent research. A corresponding local version and associated graphic and mapping techniques, similar to a Moran scatterplot (Anselin 1996), will be elaborated. In addition, important inferential properties for the global measure, including significance testing, will be examined, as for Moran's I (Terui and Kikuchi 1994; Tiefelsdorf and Boots 1995, 1997; Hepple 1998; Tiefelsdorf 1998).

References

Anselin L (1988) *Spatial econometrics: methods and models*. Kluwer Academic Publishers, Boston

Anselin L (1990) What is special about spatial data? Alternative perspectives on spatial data analysis. In: Griffith DA (ed) *Spatial statistics, past, present and future*. Institute of Mathematical Geography, Ann Arbor, MI, pp 63–77

Anselin L (1995) Local indicators of spatial association: LISA. *Geographical Analysis* 27: 93–115

Anselin L (1996) The Moran scatterplot as an ESDA tools to assess local instability in spatial association. In: Fischer M, Scholten H, Unwin D (eds) *Spatial analytical perspectives on GIS*. Taylor & Francis, London, pp 111–125

Anselin L, Getis A (1992) Spatial statistical analysis and geographic information systems. *The Annals of Regional Science* 16: 19–33

Anselin L, Griffith DA (1988) Do spatial effects really matter in regression analysis? *Papers of the Regional Science Association* 65: 11–34

Anselin L, Smirnov O (1996) Efficient algorithms for constructing proper higher order spatial lag operators. *Journal of Regional Science* 36: 67–89

Bivand RS (1980) A Monte Carlo study of correlation coefficient estimation with spatially autocorrelated observations. *Quaestiones Geographicae* 6: 5–10

Boots B, Tiefelsdorf M (2000) Global and local spatial autocorrelation in bounded regular tessellations. *Journal of Geographical Systems* 2: 319–348

Brown LA, Lee S-I, Moore J, Lobao L (2000) Continuity amidst economic restructuring: the US gender division of labor, 1970–1990. Presented in the 29th International Geographical Congress, Seoul, Korea

Chakravorty S (1996) A measurement of spatial disparity: the case of income inequality. *Urban Studies* 33: 1671–1686

Cliff AD, Ord JK (1981) *Spatial processes: models & applications*. Pion Limited, London

Clifford P, Richardson S (1985) Testing the association between two spatial processes. *Statistics & Decisions* (Supplement Issue) 2: 155–160

Clifford P, Richardson S, Hémon (1989) Assessing the significance of the correlation between two spatial processes. *Biometrics* 45: 123–134

Dutilleul P (1993) Modifying the *t* test for assessing the correlation between two spatial processes. *Biometrics* 49: 305–314

Fischer MM (1999) Spatial analysis: retrospect and prospect. In: Longley PA, Goodchild MF, Maguire DJ, Rhind DW (eds) *Geographical information systems, vol. 1: principles and technical issues*. Wiley, New York, pp 283–292

Fotheringham AS (1997) Trend in quantitative methods I: stressing the local. *Progress in Human Geography* 21: 88–96

Fotheringham AS, Rogerson PA (1993) GIS and spatial analytical problems. *International Journal of Geographical Information Systems* 7: 3–19

Geary RC (1954) The contiguity ratio and statistical mapping. *Incorporated Statistician* 5: 115–145

Getis A, Ord JK (1992) The analysis of spatial association by use of distance statistics. *Geographical Analysis* 24: 189–206

Getis A, Ord JK (1996) Local spatial statistics: an overview. In: Longley P, Batty M (eds) *Spatial analysis: modelling in a GIS environment*. GeoInformation International, Cambridge, pp 261–277

Goodchild MF (1986) *Spatial autocorrelation*. Concepts and Techniques in Modern Geography 47, Geo Books, Norwich

Goodchild MF (1996) Geographic information systems and spatial analysis in the social science. In: Aldenderfer M, Maschner HDG (eds) *Anthropology, space, and geographic information systems*. Oxford University Press, Oxford, pp 241–250

Griffith DA (1987) *Spatial autocorrelation: a primer*. Association of American Geographers, Washington DC

Griffith DA (1993) Which spatial statistics techniques should be converted to GIS functions? In: Fischer M, Nijkamp P (eds) *Geographic information systems, spatial modelling and policy evaluation*. Springer, Berlin, Heidelberg New York, pp 101–114

Griffith DA (1995) The general linear model and spatial autoregressive models. In: Anselin L, Florax RJGM (eds) *New directions in spatial econometrics*. Springer, Berlin Heidelberg New York, pp 273–295

Griffith DA, Amrhein CG (1997) *Multivariate statistical analysis for geographers*. Prentice Hall, Upper Saddle River

Haining R (1990) *Spatial data analysis in the social and environmental sciences*. Cambridge University Press, New York

Haining R (1991) Bivariate correlation with spatial data. *Geographical Analysis* 23: 210–227

Heo M, Gabriel KR (1998) A permutation test of association between configurations by means of the RV coefficient. *Communications in Statistics: Simulation and Computation* 27: 843–856

Hepple LW (1998) Exact testing for spatial autocorrelation among regression residuals. *Environment and Planning A* 30: 85–108

Hubert LJ, Golledge RG (1982) Measuring association between spatially defined variables: Tjøstheim's index and some extensions. *Geographical Analysis* 14: 273–278

Hubert LJ, Golledge RG, Costanzo CM (1981) Generalized procedures for evaluating spatial autocorrelation. *Geographical Analysis* 13: 273–278

Hubert LJ, Golledge RG, Costanzo CM, Gale N (1985) Measuring association between spatially defined variables: an alternative procedure. *Geographical Analysis* 17: 36–46.

Isaaks EH, Srivastava RM (1989) *An introduction to applied geostatistics*. Oxford University Press, New York

Lee S-I (2001) A generalized significance testing method for spatial association measures: an extension of Mantel Test. *Environment and Planning A* (on review)

Loader C (1999) *Local regression and likelihood*. Springer, Berlin Heidelberg New York

Manly BFJ (1997) *Randomization, bootstrap and Monte Carlo methods in biology*, 2nd Ed. Chapman & Hall, New York

Mantel N (1967) The detection of disease clustering and a generalized regression approach. *Cancer Research* 27: 209–220

Moran PAP (1948) The interpretation of statistical maps. *Journal of the Royal Statistical Society* Series B (Methodological) 10: 243–251

Odland J (1988) *Spatial autocorrelation*. SAGE Publications, Newbury Park

Ord JK, Getis A (1995) Local spatial autocorrelation statistics: distributional issues and an application. *Geographical Analysis* 27: 286–306

Richardson S, Hémon D (1981) On the variance of the sample correlation between two independent lattice processes. *Journal of Applied Probability* 18: 943–948.

Terui N, Kikuchi M (1994) The size adjusted critical region of Moran *I*: test statistics for spatial autocorrelation and its application to geographical areas. *Geographical Analysis* 26: 213–227

Tiefelsdorf M (1998) Some practical applications of Moran's *I*'s exact conditional distribution. *Papers in Regional Science* 77: 101–129

Tiefelsdorf M (2000) *Modelling spatial processes: the identification and analysis of spatial relationships in regression residuals by means of Moran's I*. Springer, Berlin Heidelberg New York

Tiefelsdorf M, Boots B (1995) The exact distribution of Moran's *I*. *Environment and Planning A* 27: 985–999

Tiefelsdorf M, Boots B (1997) A note on the extremities of local Moran's I_is and their impact on global Moran's *I*. *Geographical Analysis* 29: 248–257

Tiefelsdorf M, Griffith DA, Boots B (1999) A variance-stabilizing coding scheme for spatial link matrices. *Environment and Planning A* 31: 165–180

Unwin DJ (1996) GIS, spatial analysis and spatial statistics. *Progress in Human Geography* 20: 540–551

Upton GJG, Fingleton B (1985) *Spatial data analysis by example: volume 1 point pattern and quantitative data*. Wiley, New York

Wardorf BS (1993) Segregation in urban space: a new measurement approach. *Urban Studies* 30: 1151–1164

Wartenberg D (1985) Multivariate spatial correlation: a method for exploratory geographical analysis. *Geographical Analysis* 17: 263–283

Wong DWS (1993) Spatial indexes of segregation. *Urban Studies* 30: 559–572