

## Conceptual and mathematical relationships among methods for spatial analysis

Mark R. T. Dale, Philip Dixon, Marie-Josée Fortin, Pierre Legendre, Donald E. Myers and Michael S. Rosenberg

Dale, M. R. T., Dixon, P., Fortin, M.-J., Legendre, P., Myers, D. E. and Rosenberg, M. S. 2002. Conceptual and mathematical relationships among methods for spatial analysis. – *Ecography* 25: 558–577.

A large number of methods for the analysis of the spatial structure of natural phenomena (for example, the clumping or overdispersion of tree stems, the positions of veins of ore in a rock formation, the arrangement of habitat patches in a landscape, and so on) have been developed in a wide range of scientific fields. This paper reviews many of the methods and describes the relationships among them, both mathematically, using the cross-product as a unifying principle, and conceptually, based on the form of a moving window or template used in calculation. The relationships among these methods suggest that while no single method can reveal all the important characteristics of spatial data, the results of different analyses are not expected to be completely independent of each other.

*M. R. T. Dale (mark.dale@ualberta.ca), Dept of Biological Sciences, Univ. of Alberta, Edmonton, AB, Canada T6G 2E9. – P. Dixon, Dept of Statistics, Iowa State Univ., Ames, IA 50011-1210, USA. – M.-J. Fortin, Dept of Zoology, Univ. of Toronto, Toronto, ON, Canada M5S 3G5. – P. Legendre, Dept de Sciences Biologiques, Univ. de Montréal, C.P. 6128 succ. Centre-Ville, Montréal, QC, Canada H3C 3J7. – D. E. Myers, Dept of Mathematics, Univ. of Arizona, P.O. Box 210089, Tucson, AZ 85721, USA. – M. S. Rosenberg, Dept of Biology, Arizona State Univ., Tempe, AZ 85287, USA.*

Most systems in the natural world are not spatially homogeneous but exhibit some kind of spatial structure. Ecologists and other scientists have become both increasingly aware of the importance of the spatial components of the phenomena they study, and increasingly sophisticated in their ability to quantify it and to include it in their understanding of ecological processes (Legendre and Legendre 1998, Liebhold and Gurevitch 2002). Partly because the study of spatial structure has arisen more-or-less independently in various branches of science (e.g., geology, geography, ecology, hydrology, engineering) and with somewhat different motivations and for different applications, a great variety of methods have been proposed in the past decades (Perry et al. 2002).

The motivations that have given rise to the development of these methods include the estimation of ore

reserves (mining), the detection of the clumping of individual organisms (ecology), and the search for unifying concepts of the spatial structure of natural objects. The wide range of methods also reflects the diversity of data that are used for analysis: the mapped locations of objects in a plane (point pattern process), mapped objects with an associated characteristic (a marked point process); spatially dispersed samples either regularly or irregularly arranged; transects of contiguous units recording the abundances of different species; grids of units, each with quantitative or qualitative characteristic; and so on (see Perry et al. 2002). The methods can therefore be classified by the kinds of data to which they can be applied. For example, some can only be applied to data that are from contiguous sampling units; some apply to very sparse samples from an area; some require a complete map of all the points in

---

Accepted 19 February 2002

Copyright © ECOGRAPHY 2002  
ISSN 0906-7590

a plane; and others can be applied to the characteristics of a very incomplete sample of individuals.

Within the broad range of the methods that will be discussed here, there are some concepts and properties that are relevant to almost all of them. One of these is spatial autocorrelation, which arises when the process generating the variable of interest is such that the values of samples that are close together have a tendency to be more similar (for positive autocorrelation) than those randomly placed in the study area. Processes such as growth and reproduction generate spatial autocorrelation in species and so autocorrelation is a general property of ecological data. As we will describe below, there are a number of different ways of measuring autocorrelation and how its strength varies with distance. A second general concept is isotropy, which is the property that the characteristics of the pattern are the same in any direction; whereas anisotropy refers to the case where the characteristics are different, depending on direction (for example, oblong rather than circular patches due to wind or water flow). A related concept is that of stationarity, which is that the underlying characteristics of the pattern, such as the mean and variance of a variable, are constant over the area under study (Legendre and Legendre 1998). Most of the methods are affected by the edges of the study area and the fact that the pattern beyond the edge is unknown. The correction of this edge effect in some methods has received considerable attention (see Cressie 1991).

Another factor that contributes to the breadth of methods is the relationship of a particular method to tests of statistical significance. Some methods have been specifically designed to provide strict statistical tests, whereas others are meant to be descriptive, or purely exploratory, providing the opportunity for the development of hypotheses concerning the relationship between spatial pattern and biological processes which can then be tested in other ways. In some cases, where the same data are used many times in the same analysis, correct statistical tests may be unattainable. On the other hand, the characteristics of the spatial structure revealed by these analyses may be included in the evaluation or adjustment of standard statistical test, which might otherwise be rendered invalid by the spatial autocorrelation in the data (Legendre et al. 2002). In some instances, the evaluation of the statistical significance of detected characteristics can be assessed using randomization or Monte Carlo techniques (cf. Manly 1997).

The understanding that the characteristics may vary in space leads to the distinction between local spatial statistics that quantify the pattern relative to particular nearby locations, as opposed to global spatial statistics that summarize the pattern's characteristics over the entire study area. A large proportion of the methods that will be discussed here can exist in two versions, one local and one global.

This paper will describe the relationships, conceptual and mathematical, among the wide range of spatial statistics available to analyze spatial pattern. We do not intend to provide a complete review of all methods, their purposes and interpretation; that would require a work of text-book length. We cannot even include all the general approaches that have been or can be used, but we will try to provide as broad a range as possible. For example, many methods that can deal with univariate and bivariate data also have extensions for multivariate data. Usually, these multivariate extensions will not be discussed explicitly in order to limit the length of this description. Similarly, extensions of methods to three dimensions will not be fully described and discussed.

Our aim, here, is to show different ways in which the methods relate to each other, which we can show either informally based on several of the methods' conceptual characteristics or formally by showing mathematical equivalency or similarity. Relationships can be based on theoretical grounds, empirical calculations, or conceptual affinity. Equations will be denoted by numbers in brackets, as is customary. Less formal relationships between methods, where described in the text, will be indicated by "relationship" numbers in curly brackets, thus: {R0}. Some of the methods are related by the motivation for their use, some by the conceptual bases on which they were developed, and some have close mathematical relationships. We present some of the relationships formally, following the example of Getis' (1991) cross-product approach to the unification of these statistical methods. That approach expresses each technique as the sum of the products of the data and the values of a weighting function particular to the method. We provide a conceptual analog of the cross-product approach to unification by considering the characteristics of the "window" or "template" used in calculation for each method. Finally, we will illustrate our perception of the relationships among the methods pictorially with an ordination diagram based on the features of those window templates and other characteristics of the methods.

## Methods

### Variance:mean ratio

The simplest and oldest measures of "spatial pattern", and the ones most frequently cited in introductory ecology textbooks, are based on the counts of individuals in some kind of sampling units such as quadrats. In many instances, the aim is to distinguish among three categories of spatial point patterns: random; underdispersed or clumped; and overdispersed or "regular", as illustrated in many textbooks (e.g. Dale 1999: Fig. 1.9). Many of these measures are based on the relationship

of the sample mean to the sample variance for the entire study area.

Given a set of  $n$  variates  $x_i$ , representing the counts of individuals in sampling units, the mean is the first moment,  $m_1$ :

$$m_1 = \sum_{i=1}^n x_i/n \quad (1)$$

The variance,  $s^2$  is:

$$s^2 = \left( \sum_{i=1}^n x_i^2 - nm_1^2 \right) / (n-1). \quad (2)$$

For large sample size, the divisor is sometimes replaced by  $n$ . Using the notation of the second moment:

$$s^2 \approx m_2 - m_1^2 \quad \text{where } m_2 = \sum_{i=1}^n x_i^2/n. \quad (3)$$

The simple variance to mean ratio is:

$$D = s^2/m_1. \quad (4)$$

It is sometimes suggested that as a statistical test of randomness,  $(n-1)D$  can be compared to the  $\chi^2$ -distribution on  $n-1$  degrees of freedom because if the points are random, the counts come from a Poisson distribution for which the variance equals the mean. In the presence of spatial correlation, the sample variance is not an unbiased estimator of the variance, but the sample mean is an unbiased estimator of the mean. While it is true that if the points are randomly arranged, the distribution is Poisson and if the distribution is Poisson, then the variance equals the mean, the reverse is not true. It is possible to have a distribution that is not Poisson for which the variance equals the mean (Hurlbert 1990) and it is possible for the distribution to be Poisson when the points are not randomly arranged, at least for one size of quadrat (Dale 1999).

Nevertheless, a number of indices of spatial pattern have been based on this ratio. For example, David and Moore's index of crowding,  $C_{DM}$ , is:

$$C_{DM} = s^2/m_1 - 1 = D - 1. \quad (\text{David and Moore 1954}). \quad (5)$$

A closely related index of the points' aggregation is Morisita's  $I_s$ , (Morisita 1959) where  $N$  objects are distributed among the  $n$  sampling units ( $N = \sum x_i$ ):

$$I_s = \frac{N}{N-1} \left[ \frac{s^2}{m_1} + m_1 - 1 \right] / m_1 = \frac{N}{m_1(N-1)} (C + m_1) \approx \frac{D}{m_1} + 1 \quad (\text{Hurlbert 1990}). \quad (6)$$

The perceived dispersion of a point pattern may depend greatly on the scale of the study and the size of the sample unit used. If a single grove of trees is studied, the stems may be seen as overdispersed, but when several groves are included, the trees may appear to be clumped (Dale 1999: Fig. 1.10). Also, in many applications, the principal interest may not be merely to determine which of the three categories (random, under- and overdispersed) a point pattern falls for a particular scale of study. Frequently, if the points are overdispersed we may want to know the average spacing between the points. If the points are underdispersed, forming clumps of higher density separated by gaps of lower density, we may want to know the average sizes of the patches and gaps and whether there is a single scale of clumping or several. For those kinds of questions, the spatial locations of the sampling units must be somehow included as information in the analysis.

### Block and quadrat variance methods

In the next family of methods, the spatial locations of the sample units are included in the analysis and it requires that the data be collected as a complete census in strings or grids of contiguous quadrats. The data can be counts of individuals (or such) or records of density such as estimates of cover. We will begin with methods that apply to data collected in one-dimensional strings. These can be viewed as extensions of the variance:mean ratio by using a range of sizes of units on which the values are calculated and by calculating the variance not based on all units at once but only on pairs of adjacent units.

"Lacunarity analysis" uses the mean and variances of groupings of  $r$  adjacent sampling units (whether the data are counts or density measures), based on the first and second moments for windows of size  $r$  (Plotnick et al. 1993, 1996). This way of calculating lacunarity can be thought of as a one-part gliding window that includes  $r$  units for  $r = 1, r = 2$ , and so on. It is placed at the first possible position of the string of data and the total in that window is calculated and recorded to contribute to the sum and the sum of squares of the totals. The window is then moved one position along and the process is repeated (Fig. 1). This progression is continued until the last possible position for the window is reached.

The measure of lacunarity for windows of size  $r$  is:

$$\Lambda(r) = m_2(r)/(m_1(r))^2. \quad (7)$$

This measure is closely related to the variance:mean ratio and to Morisita's index because

$$\Lambda(r) = (m_2(r) - m_1(r)^2 + m_1(r)^2)/(m_1(r))^2 \approx s^2(r)/m_1(r)^2 + 1. \quad (8)$$

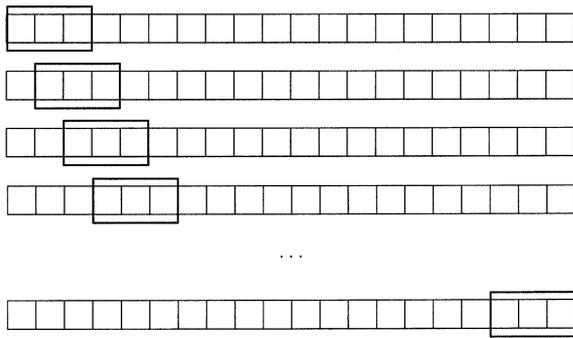


Fig. 1. The calculation of lacunarity for a string of contiguous quadrats: the smaller squares are the quadrats and the larger rectangle is a moving window of size 3. The mean and variance is calculated from the values at each possible position of the window and for a range of window sizes.

A major difference, however, is that lacunarity is usually calculated on a moving window, so that the same datum may be counted in several overlapping windows; whereas the other measures are more usually applied to data from non-overlapping sample units.

The relationships among this first grouping of methods are illustrated schematically in Fig. 2.

The next methods can be thought of as following procedures similar to that used in lacunarity analysis, except that the template is a window consisting of two parts rather than one, and the variance is calculated from the differences between the two halves of the template:

One-part window (lacunarity)  $\Rightarrow$  Two-part windows. {R1}

The first of these is called “two term local quadrat variance”, TTLQV (Hill 1973), in which, as in lacunarity, the window changes size with increasing values of  $b$ . We use “ $b$ ” here, rather than “ $r$ ” which we used for the window size in lacunarity, because the block size affects both the size of the window and the distance

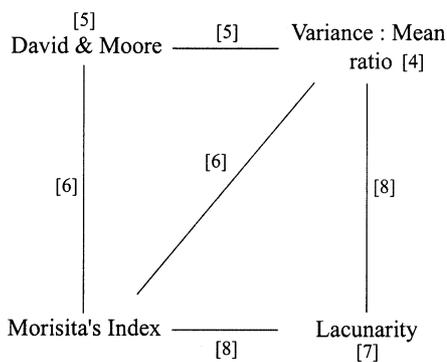


Fig. 2. The relationships among the methods that employ the variance to mean ratio. The numbers in square brackets refer to equations.

between the two parts of the template. Elsewhere, we will use “ $d$ ” where only the distance is affected. The variance in TTLQV is:

$$V_2(b) = \sum_{i=1}^{n+1-2b} \left( \sum_{j=1}^{i+b-1} x_j - \sum_{j=i+b}^{i+2b-1} x_j \right)^2 / 2b(n+1-2b). \quad (9)$$

This variance is calculated for a range of block sizes and when plotted, peaks in the variance are interpreted as being indicative of scales of pattern in the data (Hill 1973, Dale 1999). This method is often used with density data, but it can be used for presence/absence data, or for counts.

An alternative to having a two-part window of which the window size increases is to have a two-part template for which only the spacing changes with each half containing only a single original sample unit; this method is known as paired quadrat variance, PQV (Ludwig and Goodall 1978):

Two-part window: changing size (TTLQV)  $\Rightarrow$  changing spacing (PQV). {R2}

Its equation is:

$$V_p(d) = \sum_{i=1}^{n-d} (x_i - x_{i+d})^2 / 2(n-d). \quad (10)$$

As with TTLQV, peaks in the plot of  $V_p$  as a function of  $d$  are interpreted as indicating scales of pattern in the data (cf. Ludwig and Reynolds 1988). ver Hoef et al. (1993) provide an equation that shows a close, but not simple, relationship between TTLQV and PQV:

$$V_2(b) = V_p(b) + \frac{1}{b} \sum_{i=1}^{b-1} \{2(i-b)V_p(i) + (b-i) \times [V_p(b+i) + V_p(b-i)]\}. \quad (11)$$

Both TTQLV and PQV can be extended to a three-part window form, called “three term local quadrat variance”, 3TLQV (Hill 1973), and “triplet quadrat variance”, tQV (Dale 1999):

Two-part window (TTLQV, PQV)  $\Rightarrow$  Three-part window (3TLQV, tQV). {R3}

The equation for 3TLQV is:

$$V_3(b) = \sum_{i=1}^{n+1-3b} \left( \sum_{j=1}^{i+b-1} x_j - 2 \sum_{j=i+b}^{i+2b-1} x_j + \sum_{j=i+2b}^{i+3b-1} x_j \right)^2 / 8b(n+1-3b). \quad (12)$$

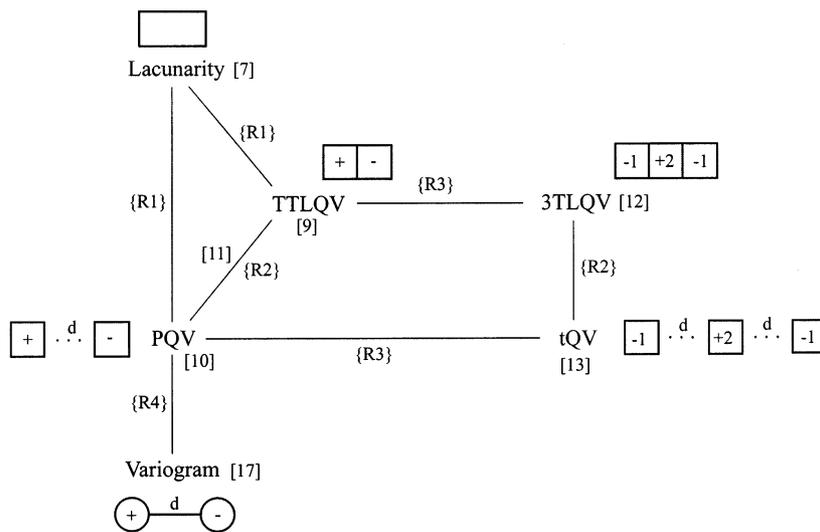


Fig. 3. The relationships among some methods that employ a moving window and calculate a variance. The numbers in square brackets refer to equations and those in curly brackets refer to less formal relationships described in the text.

For tQV, it is:

$$V_t(d) = \sum_{i=1}^{n-2d} (x_i - 2x_{i+d} + x_{i+2d})^2 / 4(n-2d). \quad (13)$$

In both these methods, peaks in the variance are considered to be indicative of scales of pattern in the data, as in the previous two methods. The two-part window methods can filter out the addition of a constant and the three-part window methods can filter out a linear trend. Therefore, 3TTLQV and tQV are less sensitive to trends in the data (Dale 1999).

### Measures of spatial autocorrelation

The concepts of autocorrelation and autocovariance are derived from the familiar statistical concepts of covariance and correlation. For two variables,  $x$  and  $y$ , their covariance is related to the expected value of their product:

$$\begin{aligned} \text{Cov}(x, y) &= E(x - E(x)) \times E(y - E(y)) \\ &= E(xy) - E(x) \times E(y). \end{aligned} \quad (14)$$

Their correlation is:

$$\rho(x, y) = \frac{\text{Cov}(x, y)}{\sqrt{\text{Var}(x)\text{Var}(y)}}. \quad (15)$$

Autocovariance and autocorrelation are simply measures of the covariance and correlation of the values of a single variable for all pairs of points separated by a given spatial lag.

The quadrat variance methods, just described, come from quantitative plant ecology and spatial autocorrelation functions from statistical geography. In geostatistics

similar techniques have been developed under different names (Matheron 1962, Rossi et al. 1992). One of the most commonly used geostatistical techniques is the calculation of a sample variogram, which quantifies autocorrelation over a range of lags,  $d$ , by estimating what is sometimes called the semivariance,  $\gamma(d)$ . For the general case, where  $d_{ij}$  is the distance between the two samples,  $i$  and  $j$  with values  $x_i$  and  $x_j$  (whether counts or other kind of measure), let  $w_{ij}(d)$  be a distance indicator function or an element of a distance weight matrix: it is 1 if  $d_{ij}$  is in distance class  $d$  and 0 otherwise.  $W(d)$  is the sum of the  $w_{ij}(d)$ . The omnidirectional sample variogram, which is an estimate of  $\gamma(d)$ , is calculated as:

$$\hat{\gamma}(d) = \sum \sum w_{ij}(d)(x_i - x_j)^2 / 2W(d). \quad (16)$$

For a transect of  $n$  contiguous or equally spaced quadrats, this is the same as:

$$\hat{\gamma}(d) = \sum_{i=1}^{n-d} (x_i - x_{i+d})^2 / 2(n-d). \quad (17)$$

Under these conditions, the latter is identical to the calculation for PQV:

Equation [10] (PQV)  $\Leftrightarrow$  eq. [17] (Sample Variogram).  
{R4}

One conceptual difference is that PQV was initially designed for strings of contiguous quadrats (Ludwig and Goodall 1978), whereas the variogram is often used for spaced samples (cf. Rossi et al. 1992). Figure 3 gives a schematic illustration of the relationships among the group of methods just described.

Other measures of autocorrelation are also used, for example (in the same notation) Moran's index of autocorrelation (Moran 1950) is:

$$I(d) = \frac{\sum \sum w_{ij}(d)(x_i - \bar{x})(x_j - \bar{x})/W(d)}{\sum (x_i - \bar{x})^2/n} \quad (18)$$

(Legendre and Legendre 1998). Geary's measure (Geary 1954) is:

$$c(d) = \frac{\sum \sum w_{ij}(d)(x_i - \bar{x})^2/2W(d)}{\sum (x_i - \bar{x})^2/(n-1)}. \quad (19)$$

Only the denominator of this formula makes it different from the equation for the variogram, (16), and therefore there is also a close relationship with the quadrat method called PQV.

There are also clear and direct relationships between measures of autocorrelation and autocovariance. Let  $C(d)$  be the autocovariance for units at distance  $d$ :

$$C(d) = E[(x_i - E(x_i))(x_{i+d} - E(x_{i+d}))], \quad (20)$$

which, for equally spaced data on a transect, can be estimated by

$$\hat{C}(d) = \frac{1}{n-d} \sum_{i=1}^{n-d} (x_i - \bar{x})(x_{i+d} - \bar{x}). \quad (21)$$

Clearly, there is a close relationship between Moran's measure and the sample covariance, and between Geary's measure and some form of the sample variogram {R5}. Note that:

$$C(0) = \text{Var}(x) = \sigma^2. \quad (22)$$

Then, assuming second order stationarity (mean and variance of the random function describing the underlying process are constant with respect location):

$$\gamma(d) = C(0) - C(d) \quad (\text{ver Hoef et al. 1993}). \quad (23)$$

Let  $\rho(d)$  be the autocorrelation at distance  $d$ , assuming the same stationarity:

$$\rho(d) = C(d)/C(0) = 1 - \gamma(d)/C(0) \quad (\text{Legendre and Legendre 1998}) \quad (24)$$

and for Geary's  $c$ :

$$E(c(d)) = 1 - C(d)/C(0) = 1 - \rho(d) \quad (\text{Legendre and Legendre 1998}). \quad (25)$$

## Neighbour networks

The above measures of autocorrelation such as Geary's and Moran's can be estimated using not only physical distance, but also for the values, counts or other measures, at pairs of points that are defined as neighbours by a network of lines joining them:

Autocorrelation with physical distance

⇔ autocorrelation on network. {R6}

There are a number of such networks to choose from, the nearest neighbours, the  $k$ th nearest neighbours, the Gabriel graph (Gabriel and Sokal 1969), and so on (Fig. 4). Of particular interest for our purposes, here, is the Delaunay network (Fig. 4c); it is formed by the rule that the lines of the triangle ABC are in the network provided that the circle circumscribing the triangle

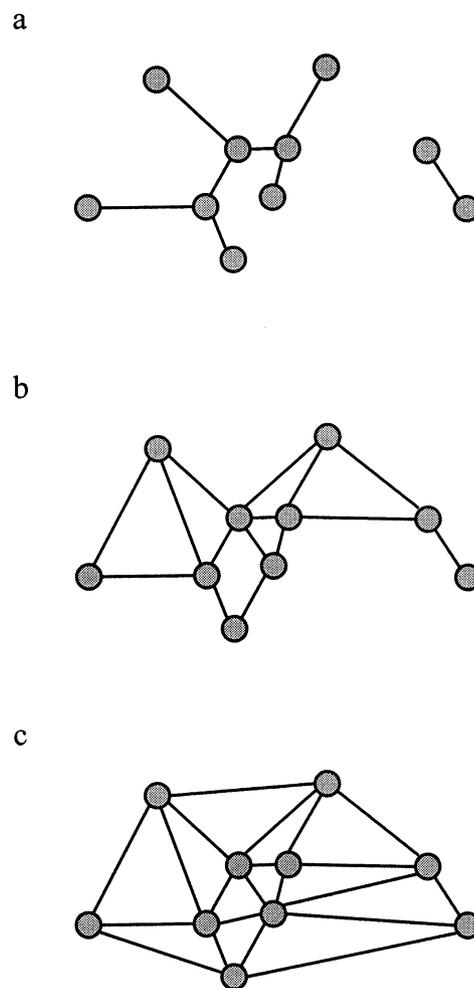


Fig. 4. Three examples of neighbour networks: the nearest neighbour network, the Gabriel graph, and the Delaunay triangulation.

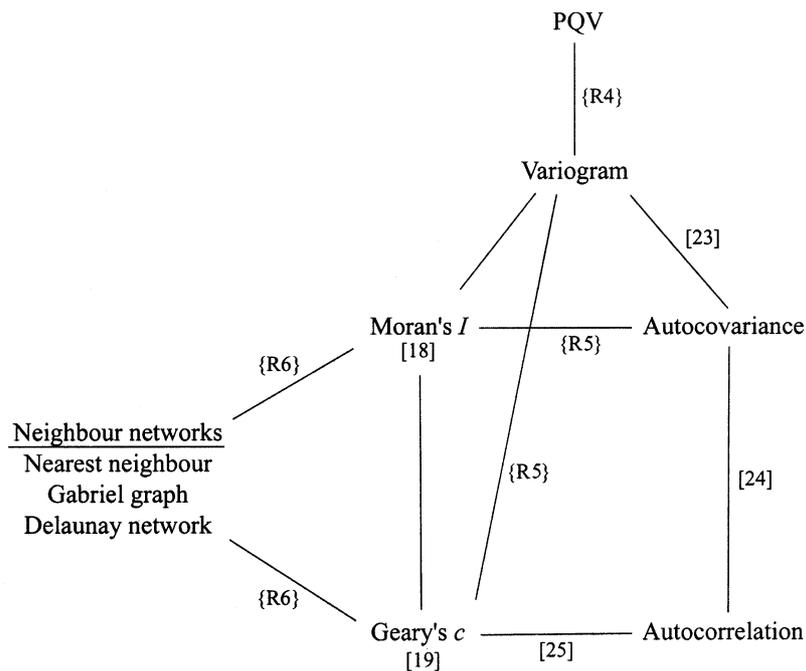


Fig. 5. The relationships among methods that focus on autocorrelation and autocovariance. The numbers in square brackets refer to equations and those in curly brackets refer to less formal relationships described in the text.

ABC (its “circumcircle”) contains no other points. It is the mathematical dual of the familiar Dirichlet (or Thiessen or Voronoi) tessellation (Okabe et al. 1992). For a good discussion of these networks, see Legendre and Legendre (1998).

Figure 5 summarizes some of the relationships among the methods described in this section.

### Spectral analysis and related techniques

Spectral analysis is a technique that examines periodicity in the spatial pattern of density data by fitting sine and cosine functions to the data and determines which frequencies or wavelengths best fit the data (Ripley 1978). Usually the data to which this analysis is applied are measures of some kind in continuous or evenly spaced series. One technique for this kind of analysis is the Fourier transform, which decomposes the “signal” into combinations of sine waves of various frequencies and positions (see Legendre and Legendre 1998). This method has been applied to two-dimensional ecological data by Renshaw and Ford (1984). Although originally developed for the analysis of continuous signals, spectral analysis can also be applied to point pattern data; see Mugglestone and Renshaw (1996).

A closely related technique (one of a family of transforms) is the use of the Walsh transform which decomposes the signal into combinations of square waves of various frequencies and positions (see Ripley 1978). There is an obvious relationship between the two approaches:

Fourier analysis (sine/cosine waves)

⇔ Walsh transform (square waves). {R7}

### Wavelets

Wavelet analysis is an approach to analyzing spatial data, related to spectral analysis, that uses a finite template or wavelet rather than sine and cosine functions, applied over the length of the data sequence. The analysis proceeds by providing measures of how well the wavelet template, of different sizes and at different positions, matches the data. (As with spectral analysis, the data are typically a set of measured values in a continuous or evenly spaced series.) The wavelet transform,  $T$ , is a function of the wavelet size and position:

$$T(b, u_i) = \frac{1}{b} \sum_{j=1}^n y(u_j)g((u_j - u_i)/b) \quad (26)$$

where  $b$  is a measure of the wavelet’s relative width (Fig. 6),  $y(u_j)$  is the density at  $u_j$ , and  $g$  is some windowing function or wavelet. This is like calculating the inner product of  $y(u)$  with a sequence of functions localized in size and position (Daubechies 1993).  $T(b, u_i)$  takes large positive values when the match between the data centred at  $u_i$  and the wavelet template is very good and large negative values when the match is very bad. The wavelet given in eq. (26) is a discrete form of a continuous wavelet transform (i.e. with summation rather than integration).

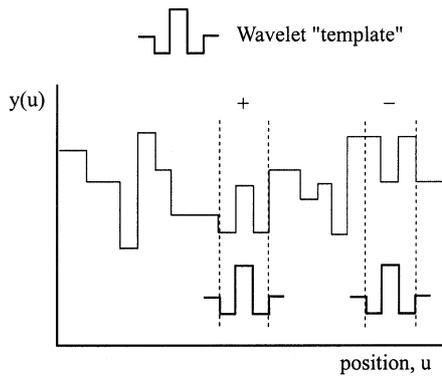


Fig. 6. An illustration of the basic concept of wavelet analysis. The wavelet template, over a range of sizes and positions, is compared to the data. Sizes and positions that match very well produce a large positive score; those that match very badly produce a large negative score.

Different functions can be used, but the “Mexican Hat” template (Fig. 7) is frequently used. For  $b = 1$  its general form is:

$$g_M(u) = \frac{2}{3^{0.5}} \pi^{-0.25} (1 - 4u^2) e^{-2u^2} \quad (27)$$

The wavelet variance is:

$$V_w(b) = \sum_{i=1}^n T^2(b, u_i) / n. \quad (28)$$

Three other wavelets are shown in Fig. 7: the Haar, the French Top Hat (FTH), and the Morlet. It is obvious

that the wavelet variance based on the Haar wavelet is equivalent to TTQLV and that based on the French Top Hat wavelet is equivalent to 3TLQV (Dale and Mah 1998):

$$\begin{aligned} \text{Haar wavelet} &\Leftrightarrow \text{TTQLV}; \text{ FTH wavelet} \\ &\Leftrightarrow \text{3TLQV}. \end{aligned} \quad \{\text{R8}\}$$

Both are also related to square wave spectral analysis using the Walsh transform:

$$\begin{aligned} \text{Haar wavelet (TTLQV)}; \text{ FTH wavelet (3TLQV)} \\ &\Leftrightarrow \text{Walsh transform}. \end{aligned} \quad \{\text{R9}\}$$

Because both are given in their discrete form here, the two wavelets in application to continuous data would produce the continuous equivalents of TTLQV and 3TLQV.

We can also use the wavelet approach to perform the equivalent of spectral analysis by using the following function:

$$g_s(u) = \begin{cases} \sin(\pi u) & \text{if } -1 \leq u \leq 1 \\ 0 & \text{otherwise} \end{cases} \quad (29)$$

If the elaboration of the sombrero wavelet into the Morlet was continued indefinitely, the resulting very long wavelet would also produce something very much like Fourier analysis:

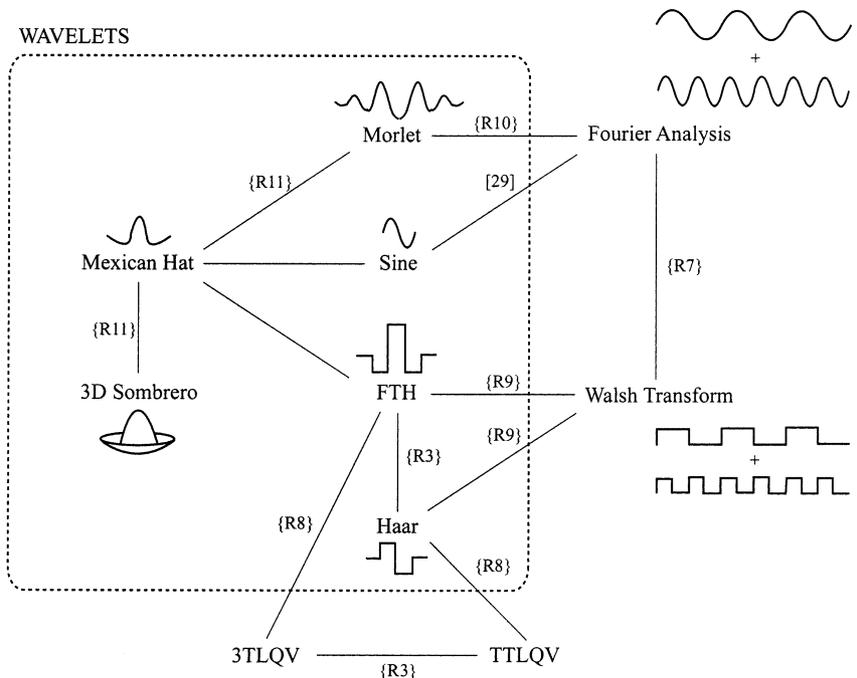


Fig. 7. The relationships among a variety of wavelet based methods of analysis. The numbers in square brackets refer to equations and those in curly brackets refer to less formal relationships described in the text.

Mexican Hat wavelet  $\Rightarrow$  Morlet wavelet

$\Rightarrow$  Fourier analysis. {R10}

In any case, wavelet variance analysis can be modified to give a wavelet covariance for bivariate data but we will not describe this feature in detail. Wavelet analysis can also be extended to data from two-dimensional samples such as densities measured on a plane, for example the amount of vegetation cover in a grassland (Csillag and Kabos 1996). One wavelet for such an analysis would be the function created by rotating a slightly modified version of the Mexican Hat wavelet about its centre, which would then strongly resemble a true three-dimensional sombrero:

Mexican Hat wavelet  $\Rightarrow$  three-dimensional

“sombrero” wavelet. {R11}

Simple rotation produces an isotropic wavelet and that isotropy would have to be considered in interpreting any analysis that used it. Figure 7 illustrates the relationships among the methods just described.

### Fractal dimension

The term “fractal” was introduced to describe phenomena that are continuous but not differentiable, so that they seem to have a fractional rather than integer dimension (Mandelbrot 1982). For example, the length of a coastline, which will seem to become longer, the smaller the unit used to measure it, has a fractal dimension between 1 and 2. There is a great deal of interest in the potential usefulness in applying the concepts of fractals to the study of spatial structures in ecology (Milne 1988, Palmer 1988, Kenkel and Walker 1993). For example, it has been suggested that fractal dimension can be used as a measure of habitat complexity and that knowing the fractal dimension of a habitat can facilitate predicting the frequency distribution of organisms by size class (for a review of fractals in ecology, see Kenkel and Walker 1993.) The data for fractal dimension analysis are often the outlines or surfaces of objects such as islands of habitat patches or tree branches (Kenkel and Walker 1993).

The fractal dimension of a spatial structure can be calculated in several ways, in part depending on the kind of data (cf. Stoyan and Stoyan 1994). The first way is to calculate it from the slope of the log-variogram based on the assumption that the variogram is an isotropic power function: if slope is:

$$m(d) = [\log(\hat{\gamma}(2d)) - \log(\hat{\gamma}(d))]/\log(2) \quad (30)$$

then the fractal dimension is:

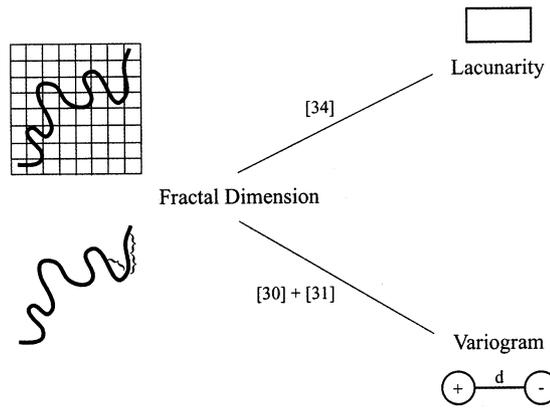


Fig. 8. The relationships between fractal dimension (box counting and divider methods) and other methods described in the text. The numbers in square brackets refer to equations.

$$D(d) = [4 - m(d)]/2. \quad (31)$$

Another method for calculating the fractal dimension of a complicated curve such as a coastline is the “dividers” method. For a given divider length  $d$ , the length of the curve is then  $L_d$ , consisting of  $N_d$  straight line segments each of length  $d$  (Fig. 8). When the log of  $L_d$  is plotted as a function of the log of  $d$ , the fractal dimension is estimated from the slope:

$$D = 1 - m. \quad (32)$$

The third method considered here is the “box counting” method. A grid of square boxes with sides of length  $r$  is superimposed on the curve and the number of these,  $B_r$ , which contain any part of the curve is counted (Fig. 8). Again, a range of values of  $r$  is used and the fractal dimension is estimated from the slope of  $\log(B_r)$  as a function of  $\log(r)$ ; here

$$D = -m. \quad (33)$$

The box counting method can also be applied to point patterns. Using the same system of grids of different unit sizes, for each size, the relative dispersion of the number of points per unit is calculated as the standard deviation over the mean:

$$R_r = s_r/m_{1,r} \cong \sqrt{\Lambda(r)}. \quad (34)$$

The fractal dimension is estimated from the slope of the log-log plot:

$$D = 1 + m. \quad (35)$$

The relative dispersion is only approximately the square root of lacunarity, as indicated in eq. (34), because the units from which it is calculated do not overlap,

whereas in the calculation of lacunarity, the “gliding boxes” in which counts are made do overlap. The relationships of this group of methods are summarized in Fig. 8.

### Run length and join counts

In one-dimensional data, consisting of 1's and 0's (presence vs absence; or black vs white), a commonly used approach is to define a “run” as a sequence of 1's with 0's at both ends (or vice versa). The number of runs observed,  $R$ , can be compared with the expected number of runs based on the null hypothesis of random arrangement, using a  $t$ -test. The expected value is  $M$  with standard deviation  $s$ :

$$M = 1 + \frac{2n_1n_2}{n_1 + n_2} \quad (36)$$

$$s = \sqrt{\frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)}} \quad (37)$$

where  $n_1$  and  $n_2$  are defined as the numbers of 0's and 1's (Knight 1974). The test statistic

$$t = (R - M)/s \quad (38)$$

is compared with the  $t$  distribution on  $n_1 + n_2 - 1$  degrees of freedom.

Dai and van der Maarel (1997) suggested using the runs of presences, the 1's, as an approach to detecting patch size. The frequency of runs of a given length are compared with the expected value, in their work, using randomizations. Run lengths that are much more common than expected are interpreted as being common patch sizes. This method clearly can only detect the patches of first order pattern (i.e. it will not detect clumps of patches) and will work best if the variance of the run lengths is small.

In two-dimensional data, consisting of a grid (or lattice) of black and white squares or of 1's and 0's, a related approach is to count the number of white-white or black-black joins, and compare that number with the expected value:

Runs test in 1 dimension

$$\Leftrightarrow \text{join counts in 2 dimensions.} \quad \{\text{R12}\}$$

Technically, let  $w_{ij}$  be a neighbour weighting function (or the set of matrix elements) that takes value 1 if squares  $i$  and  $j$  share a boundary and 0 otherwise. Where the observation  $x_i$  is 0 or 1, the count of black-black (1-1) joins can be calculated as

$$J_{BB} = \sum \sum w_{ij} Y_{ij} \quad \text{where } Y_{ij} = x_i x_j, \quad (39)$$

$$J_{WW} = \sum \sum w_{ij} Y_{ij} \quad \text{where } Y_{ij} = (1 - x_i)(1 - x_j). \quad (40)$$

Using the Kronecker delta function,  $\delta$ , we can also define the count of BW or 1-0 joins:

$$J_{BW} = \sum \sum w_{ij} Y_{ij}^* \quad \text{where } Y_{ij}^* = 1 - \delta(x_i, x_j). \quad (41)$$

The observed values can be evaluated using known formulae or by randomization (cf. Pielou 1977). In one dimension, let  $q_1$  be the number of runs of 1's and let  $q_0$  be the number of runs of 0's. The number of BW or 0-1 joins is:

$$J_{BW} = J_{01} = q_1 + q_0 - 1. \quad (42)$$

The same approach of counting joins of units of like or unlike labels (such as black and white) can be used for irregularly placed points that are connected by one of the several possible neighbour networks mentioned above:

Join counts in a grid

$$\Leftrightarrow \text{Join counts on a neighbour network.} \quad \{\text{R13}\}$$

A large number of like-like joins would lead to a high degree of spatial autocorrelation among first order neighbours:

$$\text{Join counts} \Leftrightarrow \text{spatial autocorrelation.} \quad \{\text{R14}\}$$

The concept of run length might be useful for this kind of data, but it is not clear how best to proceed. Figure 9 illustrates some of the relationships among methods described in this section.

### Second order point pattern analysis for mapped data

The next group of methods are used for point pattern analysis; that is, for analysing the mapped positions of objects in the plane, such as the stems of trees, and assume a complete census of the objects of interest in the area under study. One of the most commonly used methods is called Ripley's  $K$  (Ripley 1976). The calculation for a given radius,  $t$ , is:

$$\hat{K}(t) = A \sum_i^n \sum_{j \neq i}^n w_{ij}(t) \times q_{ij} / n^2, \quad (43)$$

where  $A$  is the area of the plot,  $w_{ij}$  is 1 if  $d_{ij} < t$  and 0 otherwise, and  $q_{ij}$  is a weighting factor for edge correction. This weight  $q_{ij}$  is 1 if the circle centred on  $i$  with

radius  $d_{ij}$  is completely within the study plot, otherwise it is the reciprocal of the proportion of that circle's circumference that is in the plot (Diggle 1983). It is usual to plot a corrected version of the estimated K function against the radius.

$$\hat{L}(t) = t - \sqrt{\hat{K}(t)/\pi}. \quad (44)$$

Values greater than 0 indicate overdispersion and negative values indicate clumping.

This approach can be easily modified for bivariate data:

$$\hat{K}_{12}(t) = A \sum_i^{n_1} \sum_{j \neq i}^{n_2} w_{ij}(t) \times q_{ij}/n_1 n_2 \quad \text{and} \quad (45)$$

$$\hat{K}_{21}(t) = A \sum_i^{n_1} \sum_{j \neq i}^{n_2} w_{ji}(t) \times q_{ji}/n_1 n_2. \quad (45)$$

$$\hat{L}_{12}(t) = t - \sqrt{(n_2 \hat{K}_{12}(t) + n_1 \hat{K}_{21}(t))/\pi(n_1 + n_2)} \quad (46)$$

(cf. Upton and Fingleton 1985, Andersen 1992). Values greater than 0 indicate segregation and values below 0 indicated aggregation of the different kinds of points.

### Mark correlation function

The next methods are designed to investigate the interactions of neighbouring trees in a forest and appear in

the works of Penttinen et al. (1992), Gavrikov and Stoyan (1995) and Stoyan and Penttinen (2000). Consider two infinitely small circles of radius  $da$  and  $db$ , with the distance between their centres being  $t$ . The probability that both contain points is  $P(t)$ :

$$P(t) = \lambda^2 g(t) da db \quad (47)$$

where  $g(t)$  is the pair correlation function and  $\lambda$  is the per area unit density. This is clearly related to Ripley's K function because, where  $u$  is the variable of integration:

$$K(t) = \int_0^t g(u) 2\pi u du. \quad (48)$$

This approach can be modified to take account of a quantitative characteristic associated with the points,  $m_i$ , for example the diameter of a tree. Where  $k_m(t)$  is a mark covariance function,  $M(t)$  is the mean of the product of the marks at distance  $t$ :

$$M(t) = \lambda^2 g(t) k_m(t) da db. \quad (49)$$

If  $\mu$  is the mean value of the  $m_i$  then if  $k_m(t) > \mu^2$  indicates a positive correlation of the marks at distance  $t$ . We can also define the cumulative function:

$$K_m(t) = \int_0^t g(u) k_m(u) 2\pi u du. \quad (50)$$

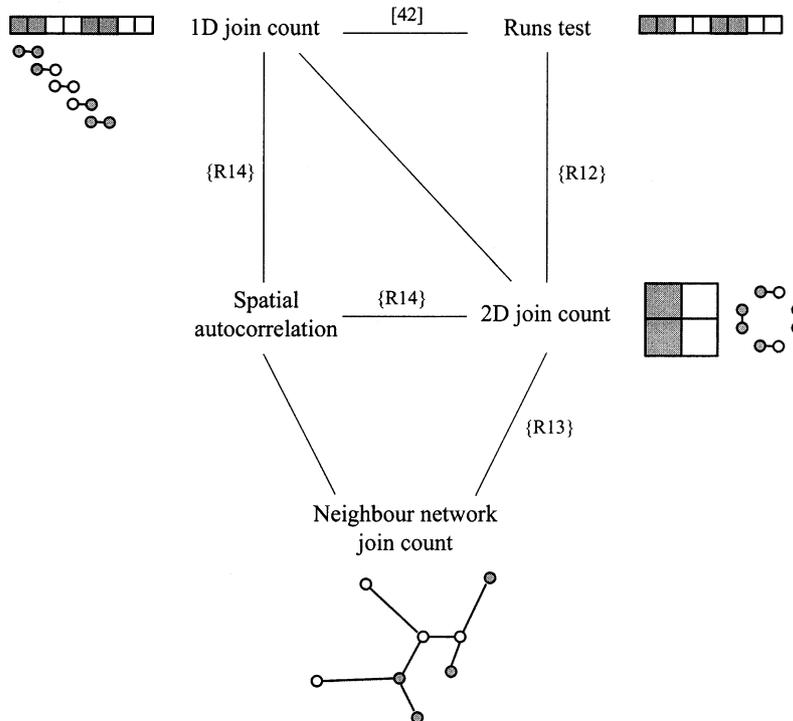


Fig. 9. Methods related to join counts. The numbers in square brackets refer to equations and those in curly brackets refer to less formal relationships described in the text.

If the values of the marks are all 1, this function reduces to Ripley's K:

Cumulative pair correlation function

$$\Rightarrow \text{Ripley's K.} \quad \{R15\}$$

We can also define the mark correlation function as:

$$\kappa_m(t) = k_m(t)/\mu^2. \quad (51)$$

With a map of  $n$  trees, giving their positions and their marks, where  $d_{ij}$  is the Euclidean distance between trees  $i$  and  $j$ , the functions can be estimated by:

$$\hat{g}(t) = \sum_{i=1}^n \sum_{j \neq i}^n f(d_{ij} - t) / \hat{\lambda}^2 2\pi r s(d_{ij}). \quad (52)$$

The edge correction function is  $s(d_{ij})$  and  $f(x)$  is a kernel function such as:

$$f(x) = \begin{cases} 3 \left(1 - \frac{x^2}{\delta^2}\right) / 4\delta & \text{if } -\delta < x < \delta \\ 0 & \text{otherwise} \end{cases}. \quad (53)$$

The suggested value of  $\delta$  is  $0.2/\lambda$ . When  $g(t)$  is plotted as a function of  $t$ , values of greater than 1 are interpreted as indicating a cluster process at that scale and values less than 1 an inhibition process.

$$\hat{\kappa}_m(t) = \frac{1}{\hat{\lambda}^2 \mu^2} \sum_{i=1}^n \sum_{j \neq i}^n f(d_{ij} - t) m_i m_j / 2\pi t \hat{g}(t) s(d_{ij}). \quad (54)$$

The cumulative function  $K_m(t)$  is estimated by:

$$\hat{K}_m(t) = \sum_i \sum_{j | d_{ij} \leq t} \frac{m_i m_j}{\lambda^2 s(d_{ij})}. \quad (55)$$

$$\hat{L}_m(t) = t - \sqrt{\hat{K}_m(t) / \pi \mu^2}. \quad (56)$$

When  $L$  is plotted as a function of  $t$ , large positive values indicate overdispersion of the marks and large negative values indicate their aggregation. The parallels with the interpretation of Ripley's K function are obvious {R16}. Figure 10 shows the relationships among these methods.

## LISAs

In some applications, it may be useful and interesting to evaluate how the strength of spatial autocorrelation varies with location within the study area. This can be accomplished using a "Local Index of Spatial Association" or "LISA". Both Moran's coefficient and Geary's can be calculated at each site,  $i$ , separately to give indices of local association or autocorrelation (Anselin 1995, Ord and Getis 1995):

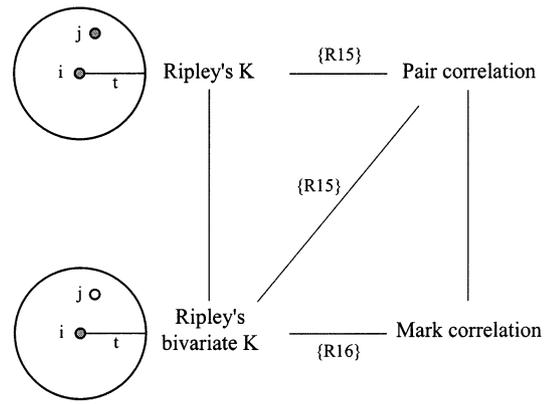


Fig. 10. The relationships with Ripley's K function. The numbers in square brackets refer to equations and those in curly brackets refer to less formal relationships described in the text.

$$I_i(d) = \frac{\sum_{i \neq j} w_{ij}(d)(x_i - \bar{x})(x_j - \bar{x}) / W_i(d)}{\sum (x_j - \bar{x})^2 / n}, \quad (57)$$

where the  $x$ 's may now be counts and the rest of the notation following that in eqs (18) and (19) with the obvious modifications. Geary's local measure is:

$$c_i(d) = \frac{\sum_{i \neq j} w_{ij}(d)(x_i - x_j)^2 / 2W_i(d)}{\sum (x_j - \bar{x})^2 / (n - 1)}. \quad (58)$$

Similarly, other measures can be considered in a local form; for example, Ripley's K for the  $i$ th point is:

$$K_i(t) = A \sum_{j \neq i}^n w_{ij}(t) \times q_{ij} / n. \quad (59)$$

$$\hat{L}_i(t) = t - \sqrt{\hat{K}_i(t) / \pi}. \quad (60)$$

Getis and Franklin (1987) suggested creating contour maps for various values of  $t$ , based on the Ripley scores of the individual points. This can be done for uni- or bivariate data.

## Circumcircle methods

Expanding on the idea of counting points in circles for completely mapped point data, we can consider ways of locating the circles, other than centering them on single points in the pattern, as in Ripley's K. Each trio of points in a mapped pattern defines a triangle and each triangle has associated with it a circle that goes through all three points, the circumcircle (Dale and Powell 2001). The relationship with Ripley's approach is based on counting points in circles:

Circle on 1 point (Ripley's)

⇔Circle on 3 points (circumcircle). {R17}

There is also a relationship, based on the use of the circumcircle with the definition of one of the neighbour networks described above, the Delaunay triangulation:

Counts in circumcircles

⇔empty circumcircles define network. {R18}

Let  $A$  be the total area of the mapped plot and let  $n$  be the number of points in it, giving an average point density of  $\lambda = n/A$ . For the  $k$ -th circle, let  $n_k$  be the observed number of points within it (excluding the three that define its triangle). Let  $a_k$  be the area of the circle that is within the sample plot; the expected number of points within the circle is  $e_k = (n - 3)a_k/A$ , based on a hypothesis of random point positions. The observed and expected numbers of plants can be compared using the Freeman-Tukey standardized residual,

$$z_k = \sqrt{n_k} + \sqrt{n_k + 1} - \sqrt{4e_k + 1}. \quad (61)$$

The value of  $z$  can be considered as a measure of density in the circle relative to the overall density in the plot. For example, those with values less than  $-1.96$  can be considered to be indicative of gaps and those with values greater than  $1.96$  show patches. The motivation is related to the use of runs tests to look for common patch and gap sizes in transect data (Dai and van der Maarel 1997):

Runs test for patch sizes

⇔Standardized residual to detect patches. {R19}

To distinguish among the hierarchy of overlapping patches or gaps detected by high or low values of  $z$ , we can define the "best" patches and "best" gaps as those that provide the greatest contrast with their surrounding. To find these, count the number of points in a ring of width  $(\sqrt{2} - 1)r_k$  around circle  $k$ . Let that number be  $p_k$  and its expected value  $e_k$ . The Freeman-Tukey standardized residual, for the outer ring is then:

$$\zeta_k = \sqrt{p_k} + \sqrt{p_k + 1} - \sqrt{4e_k + 1}. \quad (62)$$

The inner residual,  $z_k$ , can then be combined with the outer residual,  $\zeta_k$ , to produce a measure of the contrast between the inner circle and the outer ring:

$$Z_k = (z_k - \zeta_k)/\sqrt{2}. \quad (63)$$

The double circle template from which  $Z$  is calculated is essentially a wavelet, clearly related to the French

Top Hat but in one more dimension, the "boater" wavelet (Dale and Powell 2001):

FTH wavelet⇔boater wavelet. {R20}

The value of  $Z$  for a given circle measures how well the data match the shape of the template. Following the procedures used in wavelet analysis, we can plot the average  $Z^2$  as a function of the circle radius. Peaks in this graph will reflect the sizes of patches and gaps in the pattern.

For some applications, it is desirable to make the results of analysis spatially explicit as with LISA's. Here, the  $z$  or  $Z$  score of each circle in a particular size class could be associated with the centre of the circle. In that way, a contour map of the scores could be produced for each of several size classes or scales. The conceptual similarity with some of the LISA approaches is obvious:

LISA score⇔circumcircle  $Z$  score. {R21}

### Cluster detection

Fotheringham and Zhan (1996) discuss three methods of detecting clusters of "diseased" points in a point pattern. The two methods that are most comparable to the circumcircle method are one which counts the points in circles of a range of sizes centred on a regular grid superimposed on the data map and a second which uses randomly placed circles of randomly chosen radii (Fig. 11). The most significant circles are drawn onto the map for the purposes of visualizing the incidence of the disease. This approach has many similarities with the circumcircle method described above:

Circles on grid (cluster detection)

⇔circles on triplets of points. {R22}

Figure 11 shows some of the connections among the methods described in this section of the paper.

### SADIE

SADIE refers to a class of methods known as Spatial Analysis by Distance Indices (Perry 1995, 1996, 1999). Given a number of individuals in each of several quadrats, we could calculate the total distance that individuals would have to move in order to get them all in one quadrat, the "distance to crowding", as one index of their spatial arrangement (Fig. 12). More usefully, given the set of counts, we could characterize the pattern by calculating the total movement necessary to get the same number of points (the mean) in each

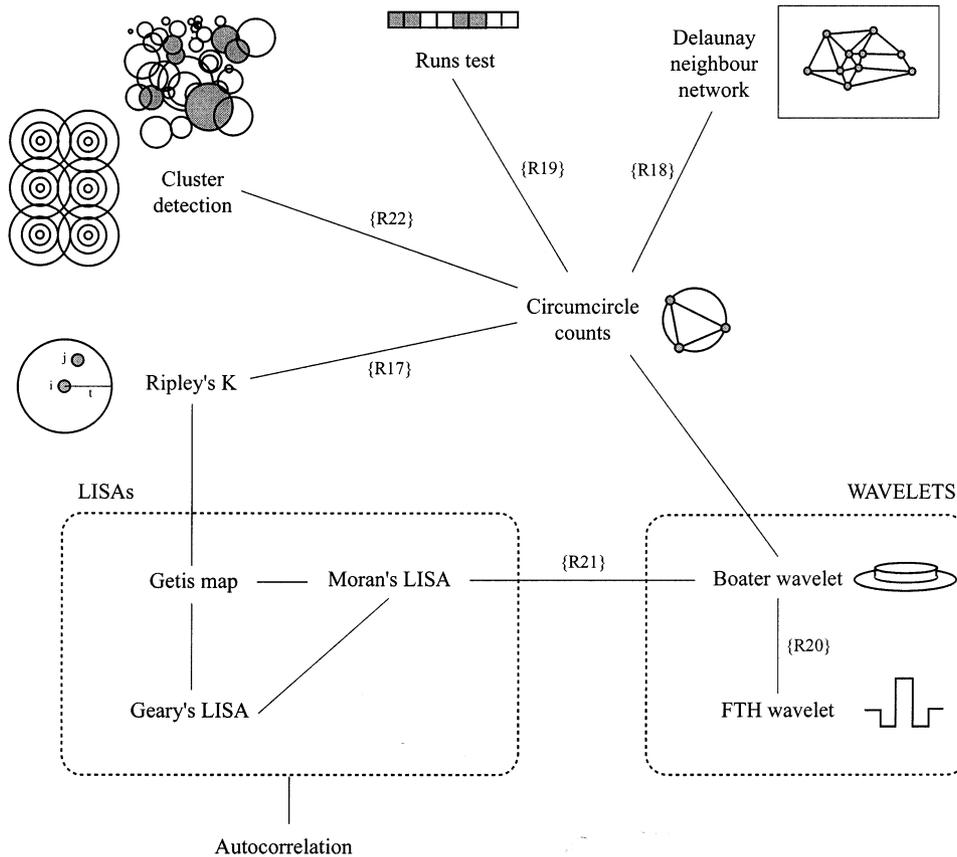


Fig. 11. The relationships with methods based on counts in circles. The numbers in square brackets refer to equations and those in curly brackets refer to less formal relationships described in the text.

quadrat, “distance to regularity” (Fig. 13). This approach is a spatially explicit revision of the variance:mean measures of quadrat counts that either maximizes or minimizes the variance of the numbers in the sample units:

SADIE  $\Leftrightarrow$  maximize or minimize variance ratio. {R23}

Suppose there are  $N$  individuals among  $n$  units in a set of quadrats, either in a grid or possibly irregularly placed. For each unit, there is the  $x, y$  coordinate and a count of individuals in it,  $c$ . Let the mean count be  $m = N/n$ . Let us consider the flow of numbers from the  $p$  units that have counts greater than  $m$ , (origins or sources) to the  $q$  units that have counts less than  $m$  (destinations or sinks). There are  $pq$  pairs of these units and we can consider the flow values from source units,  $i$ , to sink units,  $j$ ,  $v_{ij}$ . The conditions on  $v_{ij}$ , which are non-negative, are:

$$\sum_{j=1}^q v_{ij} = c_i - m \quad \text{and} \quad (64)$$

$$\sum_{i=1}^p v_{ij} = m - c_j. \quad (65)$$

Now consider the total flow distance, where  $d_{ij}$  is the distance between the units:

$$D = \sum_{i=1}^p \sum_{j=1}^q v_{ij} d_{ij}. \quad (66)$$

The transportation algorithm used in SADIE works to generate the minimum value of  $D$  as a unique solution, which is the distance to regularity. The test of significance is by a procedure that randomizes the allocation

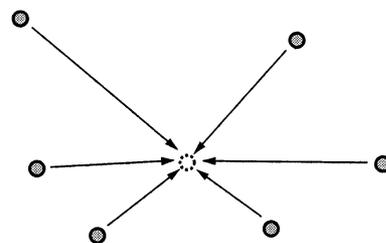


Fig. 12. SADIE: distance to crowding. This technique measures the total distance that all points would have to be moved to be perfectly clustered.

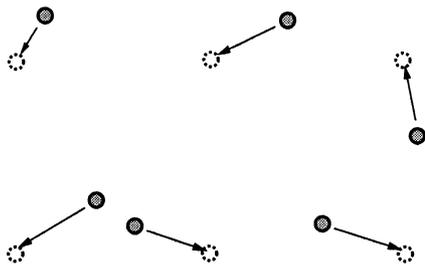


Fig. 13. SADIE: distance to regularity, a measure of the total distance all points would have to be moved to end up with maximum overdispersion.

of counts to sample units. Spatially explicit results can be obtained by plotting arrows on a map of the grid, showing the movement of numbers from the sources to the sinks (Fig. 13).

This description of the SADIE technique has been phrased in terms of counts in a set of sample units, but a similar approach can be used for point pattern data. In it, the points are nominally moved to positions that make the point pattern completely regular and the index used is the minimum total distance of movement required.

The SADIE approach can also be extended to detect clusters and gaps in count data. Using the notation of earlier in this section, the average outflow distance from the  $i$ -th unit is:

$$u_i = \frac{\sum_{j=1}^q v_{ij} d_{ij}}{\sum_{j=1}^q v_{ij}} \quad (67)$$

The average inflow distance to the  $j$ -th unit is:

$$v_j = - \frac{\sum_{i=1}^p v_{ij} d_{ij}}{\sum_{i=1}^p v_{ij}} \quad (68)$$

To evaluate the observed values, they are compared with the average values derived from randomizations:  $\bar{u}_i$  is the average absolute value of  $u$  observed at position  $i$  over the randomizations;  $\bar{u}_c$  is the average absolute value of  $u$  observed for count  $c$  as it is moved around in the randomizations; and  $\bar{u}_o$  is the overall average absolute values of  $u$  over all counts and positions. The observed and expected are compared by calculating the ratio:

$$z_i = u_i / (\bar{u}_i \bar{u}_c / \bar{u}_o) \quad (69)$$

Values of 1.5 or greater are considered to be indicative of patches, which will be “hot spots” of high outflow sources. A similar procedure is used to evaluate  $z_j$  to find gaps, which will be “cold spots” of high inflow sinks. The indices can be mapped on the sample units and the map completed by interpolation. Because of an obvious choice of colours, these are referred to as “red-blue” plots (Perry 1999). They are clearly similar to the spatially explicit results of the “boater” wavelet approach:

SADIE red-blue plots  $\Leftrightarrow$  boater wavelet scores. {R24}

The relationships among the methods of this section are outlined in Fig. 14.

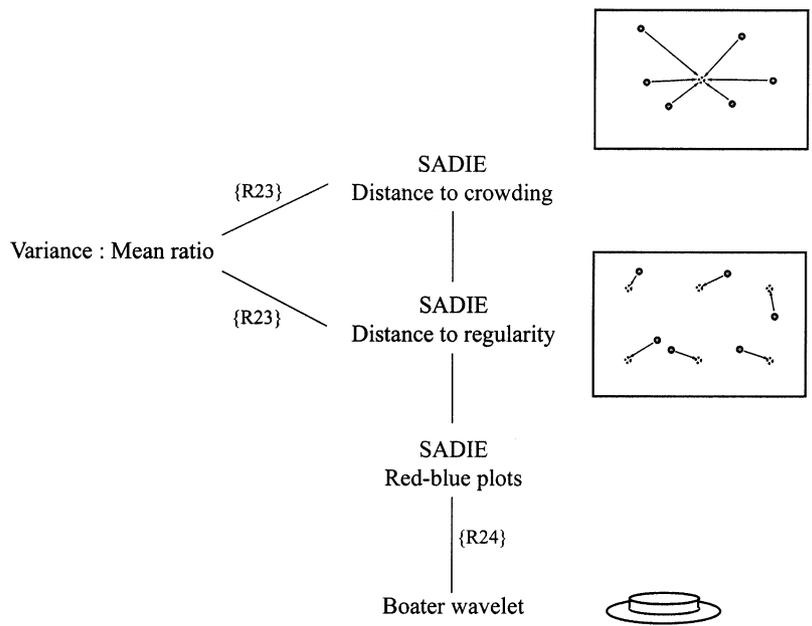


Fig. 14. The relationships of methods to the SADIE approach. The numbers in square brackets refer to equations and those in curly brackets refer to less formal relationships described in the text.

## Mantel test

The Mantel (1967) test is a widely used method for assessing the relationship between two distance matrices, where distance may be of physical location or a measure of some other kind of dissimilarity. The simple Mantel test (Mantel 1967, Mantel and Valand 1970) is a procedure to test the hypothesis that the distances (or similarities) among objects in a matrix **A** are linearly independent of the distances (or similarities) among the same objects in another matrix **B**. This test may be used to evaluate the hypothesis that the process that generated the first set of distances is independent of the process that generated the second set. The original Mantel (1967) statistic is simply the sum of the products of the corresponding distances in **A** and **B**. For symmetric matrices, it is customary to use only the distances in the upper (or lower) triangular portions of the two matrices, excluding the diagonal, which are 0 in distance or 1 in similarity matrices. The Mantel statistic can also be obtained by computing the element-by-element product:

$$\mathbf{A} \cdot \mathbf{B} = \mathbf{C}. \quad (70)$$

The statistic is the sum of the elements in matrix **C** (or, more often, in the upper or lower triangular portion). In recent years, it has become customary to compute a standardized Mantel statistic  $r_M$ , instead of the original statistic:  $r_M$  is the simple linear correlation coefficient computed between the two sets of distances. The advantage is that the statistic now takes values between  $-1$  and  $+1$ . The Mantel statistic can be tested either by randomization, or through a normal approximation when the number of observations  $n$  is large.

This procedure was originally designed by Mantel (1967) to relate a matrix of spatial distance measures and a matrix of temporal distances in a generalized regression approach. The general procedure, now known as the Mantel test in the biological and environmental sciences, includes any analysis relating two distance matrices or, more generally, two resemblance or proximity matrices (Fortin and Gurevitch 2001). Indices of spatial autocorrelation such as Moran's *I* and Geary's *c* coefficients may be obtained as special cases of the Mantel test (Anselin 1995):

$$\text{Mantel test} \Rightarrow \text{Geary's } c; \text{ Moran's } I. \quad \{\text{R25}\}$$

Given matrices **A**, of squared Euclidean distances among sites for a single variable, and **B**, the binary matrix of weights for a given distance class, the sum of terms in the upper triangle of  $\mathbf{C} = \mathbf{A} \cdot \mathbf{B}$  gives a form of the sample variogram:

$$\text{Mantel test} \Rightarrow \text{sample variogram}. \quad \{\text{R26}\}$$

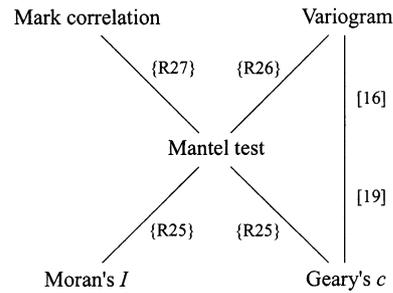


Fig. 15. Relationships of the Mantel test. The numbers in curly brackets refer to informal relationships described in the text.

There is also a close correspondence with the mark correlation functions defined above because both approaches take into account distance and similarity characteristics:

$$\text{Mantel test} \Leftrightarrow \text{mark correlation functions}. \quad \{\text{R27}\}$$

This relationship is conceptual, rather than based on mathematical theory. The relationships of the Mantel test described here are shown in Fig. 15.

## A cross-product approach

In attempting to provide mathematical unification of the range of methods described in this paper, Getis (1991) has provided valuable guidance by showing that many of the methods can be expressed as a cross-product of the form  $\sum \sum w_{ij} Y_{ij}$ . For example, the joint-count statistics in eqs (39)–(41) are already in that form. In this section, we will make explicit the application of this approach to some of the methods already described.

In general, we will follow the notation already introduced, with the  $x_i$  being the data and  $W(h)$  being  $\sum \sum w_{ij}(h)$ . For the quadrat variance methods, we can introduce a block sum function:

$$X_i(h) = \sum_{k=i}^{i+h-1} x_k. \quad (71)$$

$$\text{Defining } Y_{ij}(x) = (x_i - x_j)^2 \text{ and} \quad (72)$$

$$w_{ij}(b) = \begin{cases} 1 & j = i + b \\ 0 & j \neq i + b \end{cases}, \quad (73)$$

then TTLQV and PQV (the sample variogram) are:

$$V_2(b) = \sum_i \sum_j w_{ij}(b) Y_{ij}(X) / 2bW(b); \quad (74)$$

$$V_p(b) = \sum_i \sum_j w_{ij}(b) Y_{ij}(x) / 2W(d). \quad (75)$$

$$\text{Redefining } Y_{ijk}(x) \text{ as } (x_i - 2x_j + x_k)^2 \text{ and} \quad (76)$$

$$w_{ijk}(b) = \begin{cases} 1 & j = i + b; \quad k = i + 2b \\ 0 & \text{otherwise} \end{cases}, \quad (77)$$

then 3TLQV and tQV are:

$$V_3(b) = \sum_i \sum_j \sum_k w_{ijk}(b) Y_{ijk}(X) / 8bW(b); \quad (78)$$

$$V_t(d) = \sum_i \sum_j \sum_k w_{ijk}(d) Y_{ijk}(x) / 4W(d). \quad (79)$$

Getis (1991) shows how some forms of the sample variogram, Moran's I, and Geary's c can all be expressed in this way. Analysis using wavelets also involves the use of a cross-product (eq. 26), as does the Mantel test (previous section). Ripley's K function is a cross-product (Getis 1991), and the mark-correlation approach is also clearly of that form (eq. 49). Fractal analysis does not fit this form directly itself, but the fractal dimension is often derived from the calculation of a slope in a log-log plot of a technique that can be expressed as a cross-product.

The various versions of the circumcircle approach can be included in this form by defining:

$$w_{ijkl} = \begin{cases} 1 & l \text{ is in circle } ijk \\ 0 & \text{otherwise} \end{cases}. \quad (80)$$

In spite of the structural similarity of eq. (66), the SADIE methods do not fit well into the cross-product form. On the whole, however, the cross-product concept makes an important contribution to unifying a range of methods used in spatial analysis. We should note, in passing, that some other spatial methods, which we have not included in this paper can also be represented in a cross product form.

The mathematical approach just described using the cross-product construction has a more intuitive analog in describing the methods by the characteristics of the template or window that depicts the calculation. For example, Fig. 1 shows the calculation of lacunarity using a simple one-part window, Fig. 3 shows the two-part window for TTLQV and the three-part window for 3TLQV, and Fig. 7 shows the multipart templates used in the Fourier and Walsh transforms. We can, therefore use the number of parts (one, two, three or more) of the template used to estimate the spatial structure as a criterion for characterizing the methods. In addition to lacunarity, other examples of spatial statistics having one-part templates or windows are the

aggregation indices, the Fractal Dimension based on Box-Counting (Fig. 8) and Ripley's K function (Fig. 10). Many of spatial methods use two-part templates: e.g. Join Count (Fig. 9), Boater Wavelet (Fig. 11), and Mantel test. There are also several three-part templates (as in 3TLQV, tQV, FTH Wavelet; Figs 3 and 7) and a few multipart templates, as in Spectral Analysis, e.g. the Morlet Wavelet (Fig. 7). In some cases such as wavelets, the number of parts is determined by the number of positive and negative parts of a windowing function.

The second criterion for classifying the methods based on the window used is the "positioning" criterion, determined by whether the positions of the windows used in calculation are determined by the positions of the data. An example of a "position-dependent" method is Ripley's K statistic that uses circles centred on each point of the data set unlike cluster detection methods in which the circles occur on a grid independent of the data (Fig. 10). Other position-dependent statistics are the circumcircle methods and neighbour network algorithms (Gabriel, Delaunay, etc.) that connect points with according to their spatial arrangement. Join Count, sample variogram, Moran's I, Geary's c, Mantel test, and Fractal Dimension have position-independent templates, where the spatial structure is computed either a relative position of the sampled points or in agglomerating them into sampling units (e.g., quadrats, regions).

The third criterion, the "gliding" criterion, considers whether the calculation is based on gliding and overlapping windows (as in lacunarity, illustrated in Fig. 1, TTLQV etc.), or stepping and non-overlapping windows (e.g. Join Count, Fractal Dimension based on Box-Counting, Fig. 8).

The fourth criterion is the shape of the window used to characterize or compute spatial structure. For example, in network algorithms, the shape of the template is a simple link that connects sampled points to characterize the spatial pattern (Fig. 9). For some other spatial statistics the shape of the window used to compute the spatial structure is a circle (e.g., Ripley's K, cluster detection, Fig. 11), a square or rectangle (e.g. Lacunarity, TTLQV, 3TLQV, Fig. 3), or a curve (e.g., some wavelets, spectral analysis, Fig. 7).

Describing the methods based on the characteristics of the template or window used in calculation provides a visual representation of the cross-product generalization, and helps clarify the relationships among the various methods. To round out this paper, we will describe six other criteria that can be used to characterize the methods, in addition to the four based on the "window", which will be used in an ordination to create a summary diagram depicting some aspects of the relatedness among the various methods described.

## A pictorial summary

Following the four criteria described in the previous section, the fifth criterion is the “data type”: discrete data, such as categorical and counts from mapped coordinates (e.g., Ripley’s K, Join Count, SADIE); continuous numerical data, such as measurements (e.g., some forms the sample variogram, Moran’s I, Geary’s c, LISA); or both (e.g., Mantel test).

The sixth criterion is the “distance” criterion that characterizes the spatial lag used to compute spatial structure. Spatial distance lag can either be in terms of nearest neighbours (e.g., Join Count), Euclidean distance (e.g., SADIE), both (e.g., Mantel test) or neither (e.g., aggregation indices).

The seventh criterion is “significance”, indicating whether the spatial analysis was first developed to assess whether the spatial pattern identified was significantly different from random. The significance test could be achieved either by using probabilistic or randomized distribution (e.g., aggregation indices, Ripley’s K, Join Count, Moran’s I). Some spatial statistics were developed only to describe the spatial structure (e.g. TTLQV, 3TLQV, Fractal dimension) but randomization or Monte Carlo techniques can be used also to assess whether the spatial pattern is significantly different from random (cf. Manly 1997).

The eighth criterion is “directionality”, determined by whether the method was first developed for characterizing spatial pattern regardless of the direction (e.g., neighbour networks, SADIE) or according to direction (e.g., sample variogram, Moran’s I).

The ninth criterion is “stationarity”. This criterion creates a dichotomy between the spatial statistics more sensitive to departure from stationarity (Cressie 1991) (e.g., sample variogram, Geary’s c, etc.) and those that are less sensitive to departures from stationarity (e.g., Network algorithms, Lacunarity, 3TLQV, Wavelets).

The last criterion is that of “scale” which separates the spatial statistics that provide information about the spatial scale of a pattern (e.g., Ripley’s K, 3TLQV, Wavelets, etc.) from those that do not (e.g., aggregation indices, Network algorithms). These ten criteria were used in a Principal Coordinates Analysis (PCoA) of the spatial statistics using Gower similarity coefficient (see Legendre and Legendre 1998) by assigning values as follows: 1) Number of template parts: 1 = one; 2 = two; 3 = three; 4 = > 3; 2) Positioning: 1 = dependent on data; 2 = independent; 3) Gliding criterion: 1 = gliding; 2 = stepping; 3 = either; 4) Shape: 1 = link; 2 = square/rectangular; 3 = circle/ellipse; 4 = all; 5) Data: 1 = continuous (measurement); 2 = discrete (categorical or count); 3 = both; 6) Distance: 1 = adjacency or link based; 2 = Euclidean; 3 = both; 7) Significance: 1 = yes

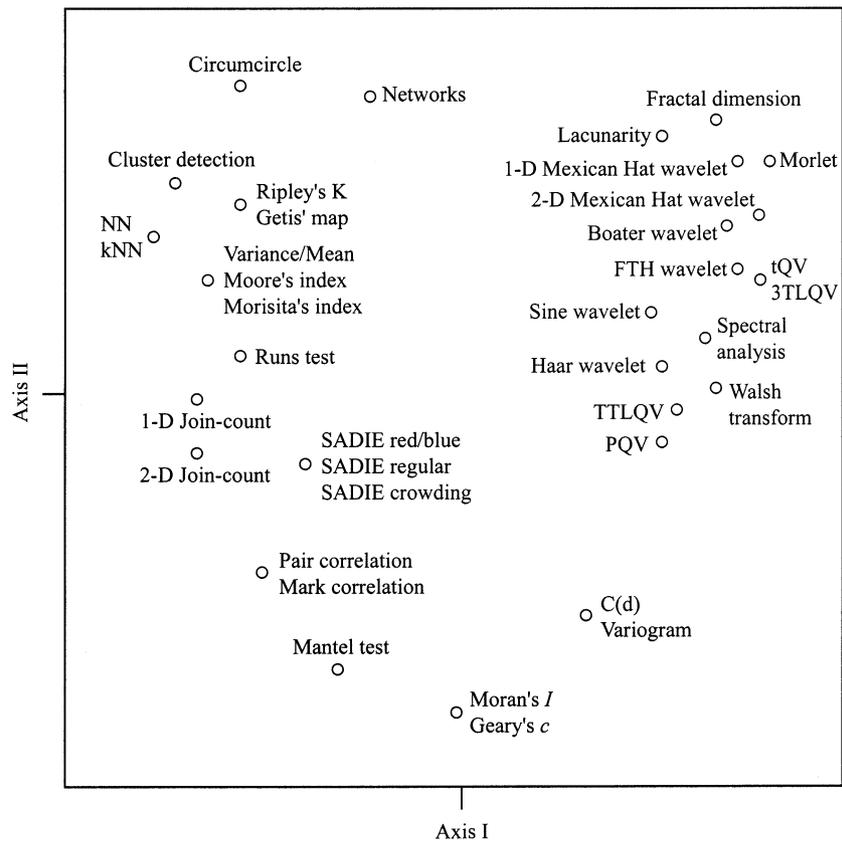


Fig. 16. The relationships of the methods described in this paper as arranged in ordination space, based on ten criteria described in the text. The first axis accounts for 37% of the variance and the second for 18%.

(parametric or randomization tests); 2 = no; 8) Directionality: 1 = yes; 2 = no; 9) Stationarity: 1 = more sensitive; 2 = less sensitive; 10) Scale: 1 = yes; 2 = no.

The results of this ordination are displayed in Fig. 16. There is a clear similarity between the PCoA arrangement of methods and some of the relationships portrayed in the various figures of relationships among methods illustrating the earlier parts of this paper. This pictorial summary provides the reader with one of many possible overviews of the relationships, both conceptual and mathematical, among the broad range of methods described here.

## Conclusions

In spite of the diversity of the backgrounds and motivations that gave rise to the methods described here, there are some obvious conceptual themes and mathematical similarities that tie them together. One is the use of a moving window or template function with which calculations are made; the preceding section described many of the methods in those terms. More formally, many of the methods can be united by expressing them as a cross product of weights and data. While we do not expect that any one method can reveal all the important features of any data set, we must also be aware that the results of different analyses may not be fully independent of each other (Legendre and Fortin 1989, Perry et al. 2002). With that in mind, future work may be to develop sequences of methods to be applied in a given order to answer specified questions about spatial characteristics. While the current paper has attempted to explore and reveal some of the relationships among the methods described, there remains more to be done in providing a full understanding of the relationships of their results.

*Acknowledgements* – The work reported in this paper was conducted as part of the Working Group “Integrating the Statistical Modeling of Spatial Data in Ecology” supported by the National Center for Ecological Analysis and Synthesis (NCEAS), a Center funded by NSF (Grant # DEB-94-21535), the Univ. of California at Santa Barbara, and the State of California and by a grant from the Natural Sciences and Engineering Council of Canada to the first author.

## References

Andersen, M. 1992. Spatial analysis of two species interactions. – *Oecologia* 91: 134–140.  
Anselin, L. 1995. Local indicators of spatial association – LISA. – *Geogr. Anal.* 27: 93–115.  
Cressie, N. A. 1991. *Statistics for spatial data*. – Wiley.  
Csillag, F. and Kabos, S. 1996. Hierarchical decomposition of variance with applications in environmental mapping based on satellite images. – *Math. Geol.* 28: 385–405.  
Dai, X. and van der Maarel, E. 1997. Transect-based patch size frequency analysis. – *J. Veg. Sci.* 8: 865–872.  
Dale, M. R. T. 1999. *Spatial pattern analysis in plant ecology*. – Cambridge Univ. Press.

Dale, M. R. T. and Mah, M. 1998. The use of wavelets for spatial pattern analysis in ecology. – *J. Veg. Sci.* 9: 805–814.  
Dale, M. R. T. and Powell, R. D. 2001. A new method for characterizing point patterns in plant ecology. – *J. Veg. Sci.* 12: 597–608.  
Daubechies, I. 1993. Different perspectives on wavelets. – *Am. Math. Soc.*  
David, F. N. and Moore, P. G. 1954. Notes on contagious distributions in plant populations. – *Ann. Bot. (Lond.)* 18: 47–53.  
Diggle, P. J. 1983. *Statistical analysis of spatial point patterns*. – Academic Press.  
Fortin, M. J. and Gurevitch, J. 2001. Mantel tests: spatial structure in field experiments. – In: Scheiner, S. M. and Gurevitch, J. (eds), *Design and analysis of ecological experiments*, 2nd ed. Oxford Univ. Press, pp. 308–326.  
Fotheringham, A. S. and Zhan, F. B. 1996. A comparison of three exploratory methods for cluster detection in spatial point processes. – *Geogr. Anal.* 28: 200–218.  
Gabriel, K. R. and Sokal, R. R. 1969. A new statistical approach to geographic variation analysis. – *Syst. Ecol.* 18: 259–270.  
Gavrikov, V. and Stoyan, D. 1995. The use of marked point processes in ecological and environmental forest studies. – *Environ. Ecol. Stat.* 2: 331–344.  
Geary, R. C. 1954. The contiguity ratio and statistical mapping. – *Incorp. Stat.* 5: 115–145.  
Getis, A. 1991. Spatial interaction and spatial autocorrelation: a cross product approach. – *Environ. Plann. A* 23: 1269–1277.  
Getis, A. and Franklin, J. 1987. Second-order neighborhood analysis of mapped point patterns. – *Ecology* 68: 473–477.  
Hill, M. O. 1973. The intensity of spatial pattern in plant communities. – *J. Ecol.* 61: 225–235.  
Hurlbert, S. H. 1990. Spatial distribution of the montane unicorn. – *Oikos* 58: 257–271.  
Kenkel, N. C. and Walker, D. J. 1993. Fractals and ecology. – *Abstr. Bot.* 17: 53–70.  
Knight, W. 1974. A run-like statistic for ecological transects. – *Biometrics* 30: 553–555.  
Legendre, P. and Fortin, M. J. 1989. Spatial pattern and ecological analysis. – *Vegetatio* 80: 107–138.  
Legendre, P. and Legendre, L. 1998. *Numerical ecology*, 2nd English ed. – Elsevier.  
Legendre, P. et al. 2002. The consequences of spatial structure for the design and analysis of ecological field surveys. – *Ecography* 25: 601–615.  
Liebhold, A. M. and Gurevitch, J. 2002. Integrating the statistical analysis of spatial data in ecology. – *Ecography* 25: 553–557.  
Ludwig, J. A. and Goodall, D. W. 1978. A comparison of paired with blocked quadrat variance methods for the analysis of spatial pattern. – *Vegetatio* 38: 49–59.  
Ludwig, J. A. and Reynolds, J. F. 1988. *Statistical ecology*. – Wiley.  
Mandelbrot, B. B. 1982. *The fractal geometry of nature*. – Freeman.  
Manly, B. F. J. 1997. *Randomization, bootstrap and Monte Carlo methods in biology*. – Chapman and Hall.  
Mantel, N. 1967. The detection of disease clustering and a generalized regression approach. – *Cancer Res.* 27: 209–220.  
Mantel, N. and Valand, R. S. 1970. A technique of nonparametric multivariate analysis. – *Biometrics* 26: 547–558.  
Matheron, G. 1962. *Traité de géostatistique appliquée*. – Editions Techniq, Paris.  
Milne, B. T. 1988. Measuring the fractal geometry of landscapes. – *Appl. Math. Comput.* 27: 67–79.  
Moran, P. A. P. 1950. Notes on continuous stochastic phenomena. – *Biometrika* 37: 17–37.

- Morisita, M. 1959. Measuring the dispersion of individuals and the analysis of distributional patterns. – *Mem. Fac. Sci. Kyushu U. Ser. E (Biol.)* 1: 187–197.
- Mugglestone, M. and Renshaw, E. 1996. A practical guide to the spectral analysis of spatial point processes. – *Comp. Stat. Data Anal.* 21: 43–65.
- Okabe, A., Boots, B. and Sugihara, K. 1992. *Spatial tessellations: concepts and applications of Voronoi diagrams.* – Wiley.
- Ord, J. K. and Getis, A. 1995. Local spatial autocorrelation statistics: distributional issues and an application. – *Geogr. Anal.* 27: 286–306.
- Palmer, M. W. 1988. Fractal geometry: a tool for describing spatial patterns of plant communities. – *Vegetatio* 75: 91–102.
- Penttinen, A. K., Stoyan, D. and Henttonen, H. M. 1992. Marked point processes in forest statistics. – *For. Sci.* 38: 806–824.
- Perry, J. N. 1995. Spatial analysis of distance indices. – *J. Anim. Ecol.* 64: 303–314.
- Perry, J. N. 1996. Simulating spatial patterns of counts in agriculture and ecology. – *Comp. Electron. Agricult.* 15: 93–109.
- Perry, J. N. 1999. Red-blue plots for detecting clusters in count data. – *Ecol. Lett.* 2: 106–113.
- Perry, J. N. et al. 2002. Illustrations and guidelines for selecting statistical methods for quantifying spatial pattern in ecological data. – *Ecography* 25: 578–600.
- Pielou, E. C. 1977. *Mathematical ecology.* – Wiley.
- Plotnick, R. E., Gardner, R. H. and O'Neill, R. V. 1993. Lacunarity indices as measures of landscape texture. – *Landscape Ecol.* 8: 201–211.
- Plotnick, R. E. et al. 1996. Lacunarity analysis: a general technique for the analysis of spatial patterns. – *Physical Rev. E* 53: 5461–5468.
- Renshaw, E. and Ford, E. D. 1984. The description of spatial pattern using two-dimensional spectral analysis. – *Vegetatio* 56: 75–85.
- Ripley, B. D. 1976. The second-order analysis of stationary point processes. – *J. Appl. Probabil.* 13: 255–266.
- Ripley, B. D. 1978. Spectral analysis and the analysis of pattern in plant communities. – *J. Ecol.* 66: 965–981.
- Rossi, R. E. et al. 1992. Geostatistical tools for modeling and interpreting ecological spatial dependence. – *Ecol. Monogr.* 62: 277–314.
- Stoyan, D. and Stoyan, H. 1994. *Fractals, random shapes and point fields.* – Wiley.
- Stoyan, D. and Penttinen, A. 2000. Recent applications of point process methods in forestry statistics. – *Stat. Sci.* 15: 61–78.
- Upton, G. J. G. and Fingleton, B. 1985. *Spatial data analysis by example. Vol. I. Point pattern and quantitative data.* – Wiley.
- van Hoef, J. M., Cressie, N. A. C. and Glenn-Lewin, D. C. 1993. Spatial models for spatial statistics: some unification. – *J. Veg. Sci.* 4: 441–452.