

Spatial analysis of sample data

Introduction

Point pattern processes generate spatial distributions of point events, which can then be evaluated for their spatial structure using the methods presented in Chapter 4 ('point pattern analysis'). The x - y coordinates of all events, whether individuals, objects, or categorized entities, in a given study area are required for this analysis. In ecological studies, a number of point pattern processes take place, a common one being seed or spore dispersal. Many ecological and environmental processes, however, generate spatially continuous data (e.g. soil moisture, temperature, etc.). The processes that generate such quantitative continuous variables are called surface pattern processes and the spatial statistics that analyse them 'surface pattern' methods or 'area pattern' methods.

There are grey zones, however, in the dichotomy between point pattern and surface pattern. Indeed, point data can be transformed into surface data by summing the number of events per sampling unit creating density values (see, for example, Gaucherel *et al.* 2007). Discrete qualitative point data can thus be converted into continuous quantitative data and so the spatial structure of point event data can be analysed using surface pattern methods. When the entire study area is surveyed by using contiguous sampling units, these quantitative data represent the entire population of data in the study area and can be analysed using the spatial methods for contiguous sample unit data presented in Chapter 5. As the sampling effort required to census an entire area is usually high, much of our understanding of natural complexity is based on sample data. The term 'sample data'

refers to the fact that, within a study area, not all the area was studied explicitly, but only a subset of it. The aim of this chapter is to present a wide range of spatial statistics that explore, characterize, quantify, interpolate and test spatial patterns from sample data, which are, by their nature, incomplete. Although most surface pattern methods were developed to analyse sample data, they can of course be used with statistical 'population data' from a complete census.

Having sample data implies that a variable is measured using sampling units distributed according to some sampling design (Chapter 1). In any sampling design, researchers make several implicitly or explicitly assumptions about the process under investigation. The most important underlying principle shared by most spatial statistical methods is the notion that nearby values of a variable are more likely to be similar than distant ones ('Tobler's first law of geography', Tobler 1970). Then an often-forgotten limitation of statistical inference is that the smallest spatial scale at which a variable is measured (gathering either quantitative or qualitative data) is the smallest grain size and that no information below this resolution can be obtained once the data are recorded. Hence the measured value is assumed to represent the entire sampling unit as if there was no spatial heterogeneity within sampling units. Consequently the spatial layout of the sampling design used (Chapter 1) can bias our ability to identify spatial pattern due to several confounding effects. The potential biases in evaluating spatial pattern will be also determined by the spatial statistics themselves, given their assumptions and their sensitivities to extreme values, as will be discussed in more detail in this chapter.

6.1 Join count statistics

The characterization and testing of the spatial association, or spatial aggregation, of categorical data (i.e. nominal, qualitative, categorical, binary variables) can be quantified using join count statistics (Moran 1948; Cliff & Ord 1981). As the name suggests, join count statistics test whether or not the co-occurrence of categorical attributes at adjacent sampling locations can be accounted for by randomness alone. In essence, it is testing whether the categories at adjacent locations match or not.

Join count statistics were first developed in human geography to test whether adjacent counties, which are areas delimited arbitrarily (Figure 6.1a), showed significant spatial aggregation of incidence of contagious diseases such as measles or chickenpox. In such analyses, each county is given a binary nominal value (black indicating disease incidence, white indicating none) assuming that the nominal attribute prevailed for the entire area of the administrative polygon. With contiguous polygons, the 'join' adjacency between counties is defined as cases where the polygons share part of a boundary (Figure 6.1b). Hence initially, join count statistics were developed to treat aggregation of match or mismatch between the attributes of adjacent polygons only for first nearest neighbours.

There are three join count statistics for binary data; if the categories are labelled black and white, there are two counts for adjacent pairs that match, both black, J_{BB} , and both white, J_{WW} , and one count of the mismatched pairs, J_{BW} . The statistics J_{BB} and J_{WW} assess

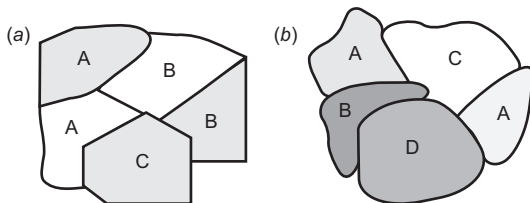


Figure 6.1 Spatial adjacency of polygons. (a) Arbitrary political boundaries of counties make it possible to have adjacent counties with the same category (here either white or grey). (b) Natural forest stands have, by definition, different types of adjacent forest surrounding them (different grey shades).

the presence of positive spatial association between adjacent units, while J_{BW} assesses spatial repulsion (or negative spatial association). The total number of joins (J) among sampling units is determined by the spatial arrangement of the sampling units, as well as the connectivity algorithm used, and it is the sum of the three join count statistics:

$$J = J_{BB} + J_{WW} + J_{BW}. \quad (6.1)$$

Therefore, only two of the statistics are independent, and any third one can be derived from the other two.

The observed join count statistic, J_{BB} , counts the number of joins in adjacent regions (or cells) having the same category (black-black):

$$J_{BB} = \frac{1}{2} \left(\sum_{\substack{i=1 \\ i \neq j}}^n \sum_{\substack{j=1 \\ j \neq i}}^n \delta_{ij} x_i x_j \right). \quad (6.2)$$

where i and j are the two sampling units compared, x_i is the attribute of a sampling unit (black = 1 and white = 0) and δ_{ij} is the entry in the connectivity matrix indicating the adjacency of sampling units i and j : 1, when i and j are adjacent; 0, otherwise. The statistic J_{BW} counts the pairs of adjacent sampling units with unlike categories (black-white), and is computed as follows:

$$J_{BW} = \frac{1}{2} \left(\sum_{\substack{i=1 \\ i \neq j}}^n \sum_{\substack{j=1 \\ j \neq i}}^n \delta_{ij} (x_i - x_j)^2 \right). \quad (6.3)$$

and J_{WW} can be computed using the two previous statistics:

$$J_{WW} = \frac{1}{2} \left(\sum_{\substack{i=1 \\ i \neq j}}^n \sum_{\substack{j=1 \\ j \neq i}}^n \delta_{ij} \right) - (J_{BB} + J_{BW}). \quad (6.4)$$

The null hypothesis of complete spatial randomness (CSR) of categories can then be tested, assuming

stationarity, by computing the expected values of joins based on the proportion of each category and the number of joins (links) in the study area. Then, using the observed and expected values, z values can be calculated for comparison with the standard normal distribution, $N(0, 1)$. Approximate normality is assured when the number of sampling units is 20 or more and the probability of neither category is less than 0.2 (Cliff & Ord 1973). One of two assumptions is required to establish the probability of each category:

- (1) the assignment of the categories (here black and white) are assumed to be independent for each sampling unit (free-sampling assumption, with replacement) such that $p = q = 0.5$; where p is the probability that the category is black and q is the probability the category is white;

or

- (2) the categories are assumed to be dependent (i.e. the category in one county affects the category type in other counties, non-free-sampling assumption, without replacement) such that $p = n_B/n$ and $q = n_W/n = (1 - p)$, where n_B is the number of black sampling units, n_W is the number of white units and n is the total number of sampling units.

Mathematical details about and equations for the expected values and variances for these statistics are presented in Cliff & Ord (1973, 1981) and in Sokal & Oden (1978).

The join count statistics can also be used with lattice data, such as those from adjacent square sampling units as illustrated in Figure 6.2. With lattice data, and as presented in Chapter 3, three types of neighbourhood rules can be used to determine which sampling units are joined to one another, based on the chessboard moves of the rook, the bishop and the queen (Figure 3.11). When sampling units or point locations are not spatially contiguous, adjacency can be established using the centroid of each sampling unit as nodes and then connect them using any of the network algorithms presented in Chapter 3 (nearest neighbour, Minimum Spanning Tree, Gabriel graph, Delaunay triangulation being common choices). By building a topological network of connections, join count statistics can be computed using higher orders

of neighbours in addition to the usual first order: second order, third order, and so on. Instead of building such topological neighbour lists, one can compute Euclidean distances between sampling units and then create distance classes, as described and illustrated in Chapter 3.

Using simulated data with known spatial patterns (random, uniform or regular, and patchy or clumped), Figure 6.2 illustrates the effects of the three most commonly used connectivity algorithms for lattice data (rook, bishop, and queen) on the detection of significant spatial associations by join count statistics. In this example, there are 100 sampling units (50 black and 50 white) and the total number of joins varies according to the connectivity criteria used (180, 162 and 342, respectively). For randomly distributed categorical data, the join count statistics that use the rook directional joins indicate that the pattern is significantly different from random, while for the two other connectivity algorithms the statistics are not significant, as expected. For uniformly distributed data, both the rook and bishop connectivity algorithms result in significant statistics compared to the null hypothesis of no spatial association. However, with the queen algorithm, that null hypothesis was not rejected. Finally, for patchy data, all three connectivity definitions produced join count statistics that were significant. These examples highlight how sensitive join count statistics may be for detecting (or not) significant spatial aggregation when different adjacency rules are used.

6.1.1 Join count statistics for k -categories

Join count statistics have been extended beyond the binary case (with $k = 2$ categories) to the more general analysis of the spatial association of k -categories where $k > 2$ (Cliff & Ord 1973, 1981; Sokal & Oden 1978). Also, join count statistics can be computed for joins beyond the first neighbour to several spatial distances defined either in terms of neighbour links, using neighbour network topology, or measuring distances among sampling units as d -neighbours (topological connectivity or Euclidean distance as distance classes,




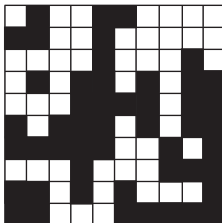
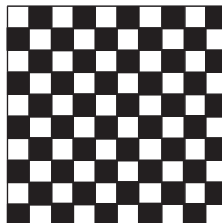
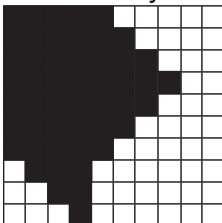
		Type of Connectivity		
		Rook	Bishop	Queen
				
J		180	162	342
		Observed values (z values)		
	J_{BB}	52 (z = 2.06)	35 (z = -1.24)	87 (z = 0.39)
	J_{WW}	54 (z = 2.61)	39 (z = -0.26)	93 (z = 1.39)
	J_{WB}	74 (z = -2.54)	88 (z = 0.98)	162 (z = -1.19)
	J_{BB}	0 (z = -12.32)	81 (z = 10.03)	81 (z = -0.60)
	J_{WW}	0 (z = -12.32)	81 (z = 10.03)	81 (z = -0.60)
	J_{WB}	180 (z = 13.40)	0 (z = -12.98)	180 (z = 0.81)
	J_{BB}	82 (z = 10.36)	76 (z = 8.80)	158 (z = 12.20)
	J_{WW}	75 (z = 8.42)	63 (z = 5.61)	138 (z = 8.87)
	J_{WB}	23 (z = -10.21)	23 (z = -9.33)	46 (z = -14.11)
		Expected values		
J_{BB}		44.545	40.090	84.636
J_{WW}		44.545	40.090	84.636
J_{WB}		90.909	81.818	172.727

Figure 6.2 Join count statistics (J_{BB} , J_{WW} and J_{WB}). The expected values are based on the number of joins (rook = 180, bishop = 162, and queen = 342) and the number of black and white values ($n_B = 50$; $n_W = 50$). The observed values of the three statistics (and corresponding z values) are computed according to the three chess moves (rook, bishop, and queen) based on binary data from a 10×10 lattice from three types of spatial patterns (random, uniform, and patchy). This example stresses how the numbers of joins (i.e. the connectivity rule) affect the observed statistics and their significance. For example, in the case of the random pattern, the observed statistics based on the rook move are significant (values $> | -1.96 |$), while these statistics are not significant for the bishop and the queen moves.

Chapter 3). Combining these two extensions, Eq. (6.2) can be rewritten as

$$J_{rr}(d) = \frac{1}{2} \left(\sum_{\substack{i=1 \\ i \neq j}}^n \sum_{\substack{j=1 \\ j \neq i}}^n \delta_{ij}(d) x_{ri} x_{rj} \right), \quad (6.5)$$

and Eq. (6.3) can be rewritten as

$$J_{rs}(d) = \frac{1}{2} \left(\sum_{\substack{i=1 \\ i \neq j}}^n \sum_{\substack{j=1 \\ j \neq i}}^n \delta_{ij}(d) x_{ri} x_{sj} \right), \quad (6.6)$$

where r indicates one category and s another, (d) indicates the distance class either in terms of d -neighbours or d -Euclidean distance classes at which the sampling units are to be considered connected (1) or not (0); and x_{ri} is an indicator function that takes the value 1 when unit i belongs to category r , and 0 otherwise. Epperson (2003) presented a mathematical description of these k -category join count statistics and applied them to genetic data.

Up to this point in the chapter, the spatial units we have considered were either arbitrarily determined a priori, as in the case of political counties, or as regularly or irregularly spaced sampling units. The implication of this is that adjacent counties (Figure 6.1a), or sampling units (Figure 6.2), can belong to the same category. There are cases, however, where this is not so. For example, while using forest inventory maps based on photographic interpretation, the delineation of each forest stand implies that it is different from neighbouring stands because it is defined by boundaries with different vegetation (Figure 6.1b). Lowell (1997) showed that with such forest inventory data, there are no joins of the same category (J_{rr}) but he demonstrated, based on simulations, that the join count statistic (J_{rs}) is robust and provides unbiased results when there are five or more categories.

There are, however, circumstances where the join count statistics are not robust. This occurs when there is an obvious lack of stationarity over the entire study area (see Chapter 1), for example in the presence of a gradient or trend, giving a mean that is not constant

(first-order heterogeneity). The likelihood of encountering data with first-order heterogeneity is growing as ecologists are using remotely sensed data covering large areas. To address the issue of first-order heterogeneity over a study area, and violation of the stationarity assumption, Kabos & Csillag (2002) developed an H Moran statistic (or heterogeneous Moran) to analyse lattice data. This statistic, H Moran, assesses the probability of each category at each sample location given the values of neighbouring locations. By doing so, the significance of the estimated statistic at each sampling location is not influenced by the lack of stationarity over the entire study area.

6.2 Global spatial statistics

6.2.1 Spatial covariance

Ecologists are well-accustomed to the notion of covariance and correlation between two variables, x and y , where one is testing whether the two variables covary positively, negatively, or not at all (i.e. testing the null hypothesis of no relationship between two variables resulting in a correlation not different from 0). Linear correlation between quantitative variables can be estimated using Pearson's product-moment correlation coefficient. This coefficient is the standardized sample covariance between two variables (e.g. x and y) which measures the deviations of the variables from their own means \bar{x} and \bar{y} ,

$$\hat{\rho}(x, y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (6.7)$$

and which estimates their correlation,

$$\rho(x, y) = \frac{\text{Cov}(x, y)}{\sqrt{\text{Var}(x)\text{Var}(y)}}. \quad (6.8)$$

Stemming from this notion of covariation between two variables, the estimates of the covariance or the correlation of a single variable with itself (giving the prefix 'auto-') can be computed for all pairs of sampling

units that are separated by the same given spatial lag (spatial autocorrelation) or temporal interval (temporal autocorrelation). The spatial autocovariance, $C(d)$, of the variable x can be estimated as the product of the deviation of the value of the variable x at the location i with the expected value ($E(x_i)$) with the deviation of the value at the location $i + d$, where d is a given distance:

$$C(d) = E\{[x_i - E(x_i)][x_{i+d} - E(x_i)]\}. \quad (6.9)$$

The spatial autocorrelation, $\rho(d)$, of the variable x for distance class d , is the autocovariance at d divided by the variance, which is $C(d)$ for $d = 0$:

$$\rho(d) = \frac{C(d)}{C(0)}, \quad (6.10)$$

with

$$C(0) = \text{Var}(x) = \sigma^2. \quad (6.11)$$

The estimation of spatial structure of a variable can be computed using various spatial statistics derived from Pearson's product-moment correlation (see Eq. (6.8)) as presented in this section. These spatial statistics are based on the assumption of stationarity. We refer to these spatial statistics as 'global' because they estimate the intensity of spatial dependence for the entire study area and summarize it with a single value.

Spatial pattern represents the lack of independence among values of a variable due to location and is the result of endogenous and exogenous processes (Chapters 1 and 2). It is generally not possible to disentangle the relative contribution of these two kinds of process to the spatial pattern, based simply on the quantification provided by spatial statistics. Indeed, spatial statistics estimate the spatial dependence in the data regardless of its origin, whether induced spatial dependence or inherent spatial autocorrelation. The only way to determine the relative contribution of endogenous and exogenous processes is to use prior knowledge of the potential processes and of factors acting on the ecological data, and an appropriate experimental design (see Chapters 1 and 8). Hence, although we are going to rely on the term 'spatial autocorrelation' in this section, keep in mind

that spatial autocorrelation coefficients cannot discriminate between the spatial structure induced by spatial dependence and that inherent to the variable itself.

6.2.2 Spatial autocorrelation coefficients for one variable

6.2.2.1 Moran's I coefficient

Moran (1948) proposed a coefficient of spatial autocorrelation, Moran's I , stemming from the Pearson correlation coefficient, which can be computed for each distance class, d as follows:

$$I(d) = \left(\frac{n}{W(d)} \right) \frac{\sum_{i=1}^n \sum_{j=1, j \neq i}^n w_{ij}(d)(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}, \quad (6.12)$$

where $w_{ij}(d)$ is the distance class connectivity matrix (also called the weight matrix) that indicates whether a pair of sampling locations, i and j , are in distance class d (Figures 3.12 and 3.13); x_i and x_j are the values of the variable x at sampling location i and j ; and $W(d)$ is the sum of $w_{ij}(d)$, here being the number of pairs of sampling locations in the distance class. As with Pearson's correlation coefficient, positive autocorrelation is indicated by positive values (usually ranging from 0 to 1), negative autocorrelation by negative values (usually ranging from 0 to -1) and the expected value for the absence of spatial autocorrelation is close to 0: $E(I) = -(n - 1)^{-1}$ (Cliff & Ord 1973, 1981). When there are too few pairs of sampling locations in distance class d , and the spatial layout of the data looks non-stationary, the estimated value is unstable and can fall outside the expected bounded range of -1 to $+1$. This occurs most often at the largest distances where there are the fewest pairs contributing to the index.

When estimating spatial autocorrelation using Eq. (6.12), the researcher needs to understand that the

estimation of spatial autocorrelation is computed by first summing the covariation between sampling units a given distance apart, d . Then, this sum of covariations from each sampling location of the entire study at a given distance class d , is divided by the actual number of pairs of locations, $W(d)$, in that distance class, d . Hence, the spatial autocorrelation coefficient, for a distance class d , is the average value of spatial autocorrelation at that distance (in all directions) for the entire study area: a *global average isotropic* estimated value of spatial autocorrelation.

Note that, in some circumstances, the spatial layout of the data looks non-stationary as the estimated value is based on fewer pairs and can reflect an extreme spatial pattern of high or low spatial autocorrelation resulting in estimated values higher than 1 (or lower than -1). Furthermore, by computing the deviation of each value from the arithmetic mean of the variable (\bar{x}), the estimation of spatial autocorrelation can be biased when the data are not normally distributed. Indeed, a skewed distribution resulting from the presence of a few outliers (either extremely low or high values) may bias the estimate of the arithmetic mean, \bar{x} , which in turn will result in an under- or over-estimation of the degree of spatial autocorrelation because the deviations are computed using a biased \bar{x} . This is why geostatisticians do not use a spatial statistic based on deviations from \bar{x} as the Moran's I coefficient does, but rather one based on deviations between adjacent sampling locations (see Geary's c and variography below). Moran's I is favoured, however, by ecologists because it is easier to relate to the familiar notion of correlation and it is more intuitive to interpret, given that it behaves like a Pearson's correlation coefficient.

6.2.2.2 Geary's c coefficient

To avoid measures of spatial pattern based on deviations from the arithmetic mean, Geary (1954) proposed another spatial autocorrelation coefficient, Geary's c , that measures the difference between values of a variable at nearby locations, so that the degree of spatial autocorrelation is based on differences at a given d distance class:

$$c(d) = \left(\frac{n-1}{2W(d)} \right) \frac{\sum_{i=1}^n \sum_{j=1, j \neq i}^n w_{ij}(d)(x_i - x_j)^2}{\sum_{i=1}^n (x_i - \bar{x})^2}. \quad (6.13)$$

Geary's c behaves like a distance measure and varies from 0 (indicating the highest value of positive autocorrelation) to 2 and greater (strong negative autocorrelation). The expected value, $E(c)$, is 1, indicating the absence of spatial autocorrelation. Note that when distance d is zero (i.e. correlation of the variable with itself), the value of Geary's c is 0. As for Moran's I , estimated values of Geary's c based on too few pairs of sampling locations will result in strange values, often greater than 2. Also, the estimated values of spatial autocorrelation with Geary's c will be biased in the presence of skewed data (as for Moran's I , but for another reason) because the differences between adjacent locations are squared (see Eq. 6.13). The squared difference between an outlier with an extreme value and other values will have more effect on the coefficient and may distort the estimation of spatial autocorrelation.

6.2.2.3 Significance of coefficient values

Significance testing allows us to determine which spatial autocorrelation coefficient values can be interpreted to understand the spatial structure of the data. The significance of each value of spatial autocorrelation estimated at a given distance class can be tested using either a randomization procedure or a normal distribution approximation test. Mathematical details about the respective variance equations used to test the significance of these coefficients can be found in Cliff & Ord (1973, 1981). The significance testing of the coefficients estimated at several distance classes is, however, problematic due to the lack of independence in the data. Indeed, the same data are used to estimate the coefficient values at different distance classes (Fortin & Dale 2009). This problem is inherent to all multiple testing analyses, such as the quadrat variance methods

presented in Chapter 5. Several procedures have been proposed to address this issue. The most widely used because of its simplicity and robustness (Oden 1984) is the Bonferroni correction, which adjusts the probability level at which we test significance, by dividing α by the number of distance classes, k :

$$\alpha' = \frac{\alpha}{k}. \quad (6.14)$$

For example, when using a probability level of $\alpha = 0.05$ and a number of distance classes of $k = 10$, $\alpha' = 0.005$. Thus in order to consider all the coefficient values as significant, given the number of distance classes (here 10), the probability value of each coefficient needs to be smaller than $\alpha' = 0.005$. The Bonferroni-corrected level value is directly related to the number of distance classes selected; however, because the selection of the number of distance classes is more or less arbitrary, this arbitrary number of distance classes can affect our ability to detect a significant spatially autocorrelated structure. To diminish this effect, we can use a progressive (sequential) Bonferroni correction (Legendre & Legendre 2012), where the Bonferroni-corrected level is computed for each distance class separately. This is done by using for each distance class the number of tests actually performed up to that distance class. For example, using the same values as above ($\alpha = 0.05$ and $k = 10$), the progressive Bonferroni levels are: $\alpha'(d=1) = 0.05$, $\alpha'(d=2) = 0.025$ and so on up to $\alpha'(d=10) = 0.005$ (the Bonferroni corrected level α). Equation (6.14) can therefore be rewritten as

$$\alpha'(d) = \frac{\alpha}{d}, \quad (6.15)$$

where d is the distance class of interest ($d = 1$ to 10).

6.2.2.4 Spatial correlograms

The characterization of spatial structure includes its intensity (magnitude or degree), its spatial range, and its shape (isotropic or not). Determining the pattern's characteristics can be facilitated by plotting the values of spatial autocorrelation against the distance d . This plot is called a spatial correlogram (see, for example, Figure 6.3), where the significance of each distance's coefficient can be indicated by a specific symbol, such

as solid circles for significant values ($\alpha = 0.05$) and open circles for the nonsignificant. When the distance d is zero (i.e. correlation of the variable with itself), the value of Moran's I is 1 (whereas Geary's c is 0). Given that most ecological data show some degree of positive autocorrelation, at short distances, the values of autocorrelation are generally positive. In fact, most of the time, the strongest spatial autocorrelation is in the first distance class. As the effective strength of the process decreases with distance (e.g. in the case of seed dispersal, most of the seeds are found close to the seed tree), the values of spatial autocorrelation also decrease with distance. The way in which this happens and how it shows up in the spatial correlogram can be used to characterize the spatial pattern: a trend in the correlogram, from positive through zero to negative spatial autocorrelation with increasing distance, is indicative of a gradient in the data. A plot (Figure 6.3b) showing a levelling off of values around the expected value of no autocorrelation ($E(I)$) is characteristic of the absence, or non-detection, of spatial pattern. When the values at short distances are positive and then, at a given distance, show somewhat regular fluctuation around the expected value, this can be interpreted as an indication of patchiness (Figure 6.3c–f). Repeated alternation of values, from positive to negative, is indicative of patch structure in the study area (Figure 6.3d, e). The distance at which the value of autocorrelation reaches, or crosses, the expected value is considered the 'spatial range', the 'zone of influence' or the 'patch size' of the spatial pattern under study, depending on the phenomenon being studied.

To interpret the shape of the correlogram as indicating the characteristics of the spatial pattern (magnitude, spatial range, shape, etc.), the significance of the entire correlogram needs to be considered. This can be determined using either the Bonferroni correction, where at least one value needs to be smaller than the corrected level α , or the progressive Bonferroni correction. In the former correction, the values estimated at large distances should be omitted because these estimates are based on very few pairs of observations. The use of the progressive Bonferroni correction avoids the arbitrary selection of non-interpretable values such as these (see Figure 6.3). Consequently,

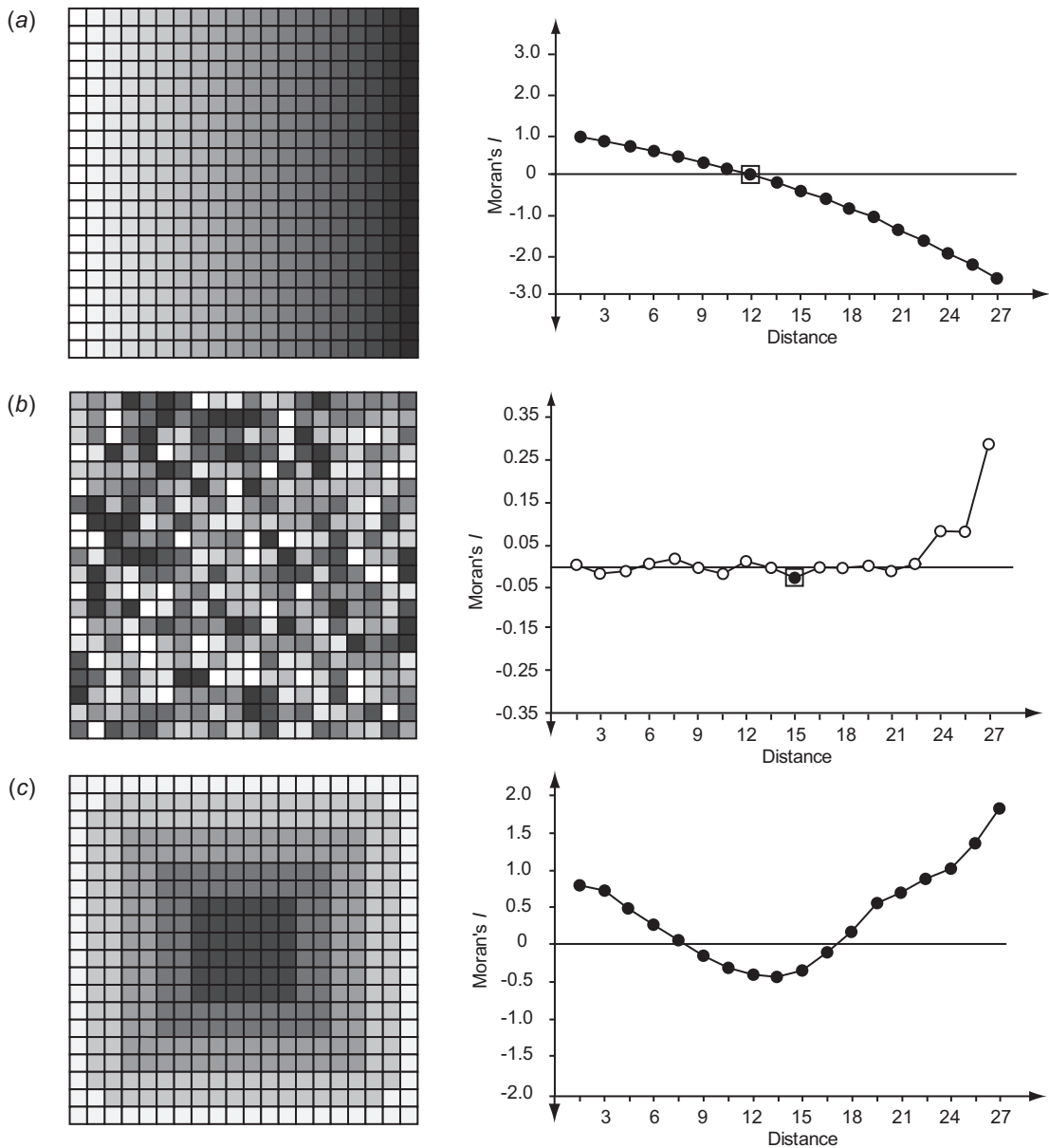


Figure 6.3 Moran's I spatial correlograms corresponding to simulated spatial patterns (a 20×20 lattice where the values increase from 0 to 10 from white to black). (a) Gradient: the correlogram shows the corresponding characteristic trend of significant positive values at short distances to negative ones at large distances. (b) Random: the values are oscillating along the zero value (i.e. the absence of significant spatial autocorrelation). (c) One big patch: the values are all significant and positive at short and large distances, while negative at intermediate. The spatial range (zone of influence, patch size) is around 7.5 units, a distance at which the sign of the values changes from positive to negative.

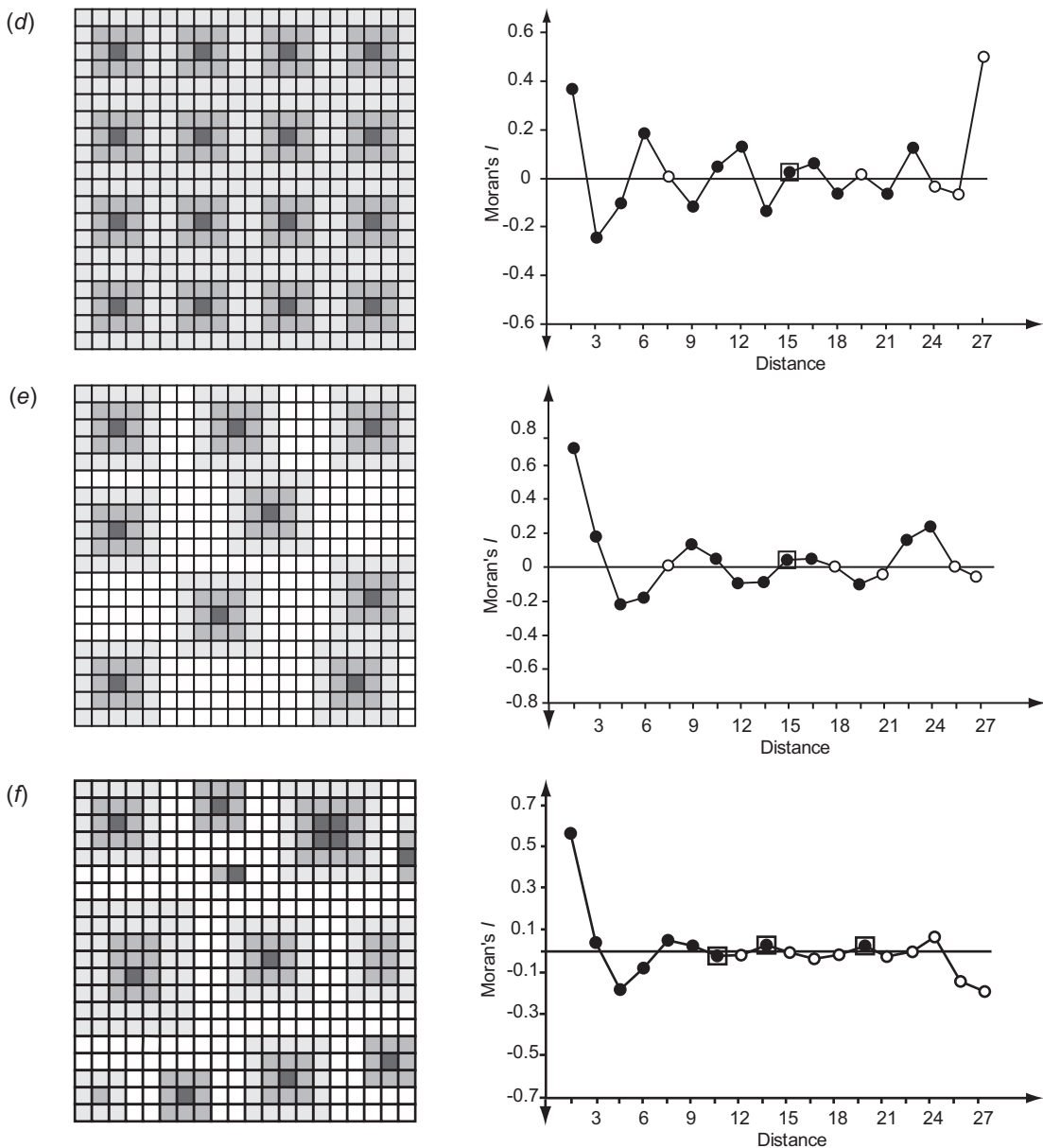


Figure 6.3 (cont.) (d) Sixteen patches: a first change of sign from positive to negative values occurs around 2.0 units, which corresponds to the spatial range of the patches. Then the correlogram repeats this oscillation in decreasing amplitude with distance revealing a repetitive spatial pattern of patches. Note here that the patches all have the same size and distance among them. (e) Nine patches: as described previously, the sign change of the values indicates that the patch size is around 3.0 units and that there is a repetitive pattern of patches. Note here that the patches have the same size but not the same distance among them. (f) Twelve patches: the sign change of the values indicates that the patch size is around 2.5 units. Here the repetitive pattern of patches is not detected by the correlogram as both the patch size and distance among the patches vary. Solid circles indicate significant coefficient values at $\alpha = 0.05$; open circles indicate non-significant coefficient values; open squares indicate coefficient values that are non-significant after progressive Bonferroni correction.

the interpretation of the spatial pattern is facilitated if nonsignificant values at the progressive corrected $\alpha'(d)$ levels are not considered.

Another approach that minimizes the effect of multiple tests is to compute a partial spatial correlogram instead of a spatial correlogram. The technique is to correct the autocorrelation at lag d for the lack of independence caused by observations of the variable that intervene between the pairs of values separated by d . This is the spatial equivalent of the partial time series analysis using Durbin's autoregressive procedure. The mathematical details are briefly presented in Cliff & Ord (1981).

The interpretation of the Geary's c correlogram (Figure 6.4) differs from that of the Moran's I correlogram in that the values at short distances are close to zero and then, at a given distance, show fluctuation around the expected value which is 1. Again, somewhat regular fluctuations in the correlogram can be interpreted as indicating patchiness in the data (Figure 6.4c).

6.2.2.5 Anisotropic spatial correlograms

Up to this point, we have considered the case where the values of spatial autocorrelation are estimated by comparing only the distance between sampling locations. This therefore assumes implicitly that the process of interest is isotropic, that is the same in all directions (see Chapters 1 and 3). This is why these spatial correlograms are often referred to as 'omnidirectional', 'all directions' or simply 'one-dimensional' (Figure 6.5a). There are, however, many ecological processes (e.g. species responses to environmental conditions, downwind seed dispersal, etc.) that can generate spatial patterns that vary with direction, such as along a gradient, resulting in elongated patches. These spatial patterns are called anisotropic because the pattern varies with direction (see Chapter 1). To determine the degree of anisotropy in a spatial pattern, spatial autocorrelation can be estimated by considering both the distances among the sampling locations (using a distance class matrix) and their relative direction using an angle class matrix (Oden & Sokal 1986). For the distance class matrix, we can select the number of angles, directions,

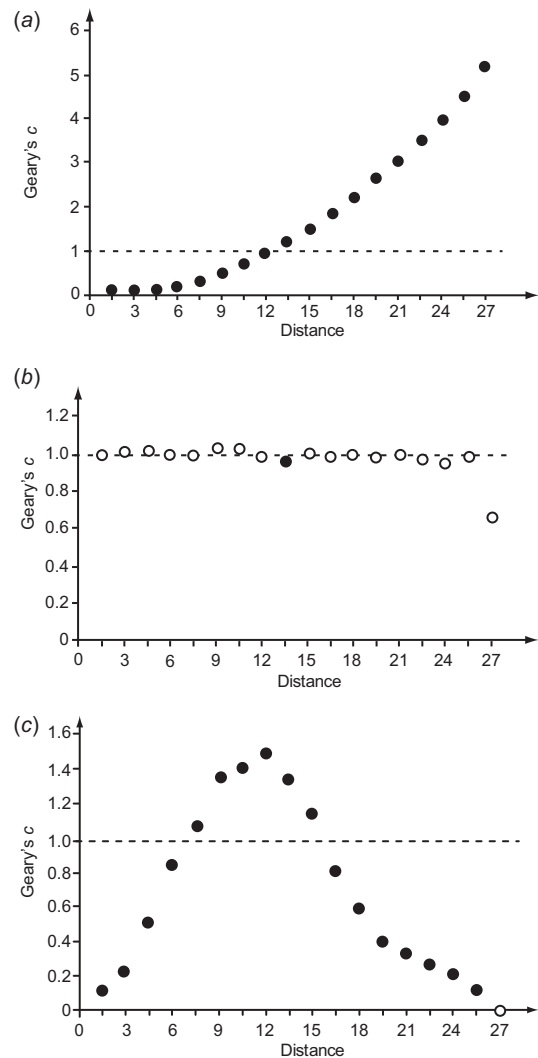


Figure 6.4 Geary's c spatial correlograms based on the same simulated spatial patterns from Figure 6.3. (a) Gradient: the correlogram shows the corresponding characteristic trend of significant positive spatial autocorrelation values (so between 0 and 1) at short distances to negative spatial autocorrelation values (so larger than 1) ones at large distances. (b) Random: the values are oscillating along the expected value of 1. (c) One big patch: the values are all significant and positive (between 0 and 1) at short and large distances, while negative at intermediate (greater than 1). Solid circles indicate significant coefficient values at $\alpha = 0.05$; open circles indicate nonsignificant coefficient values.

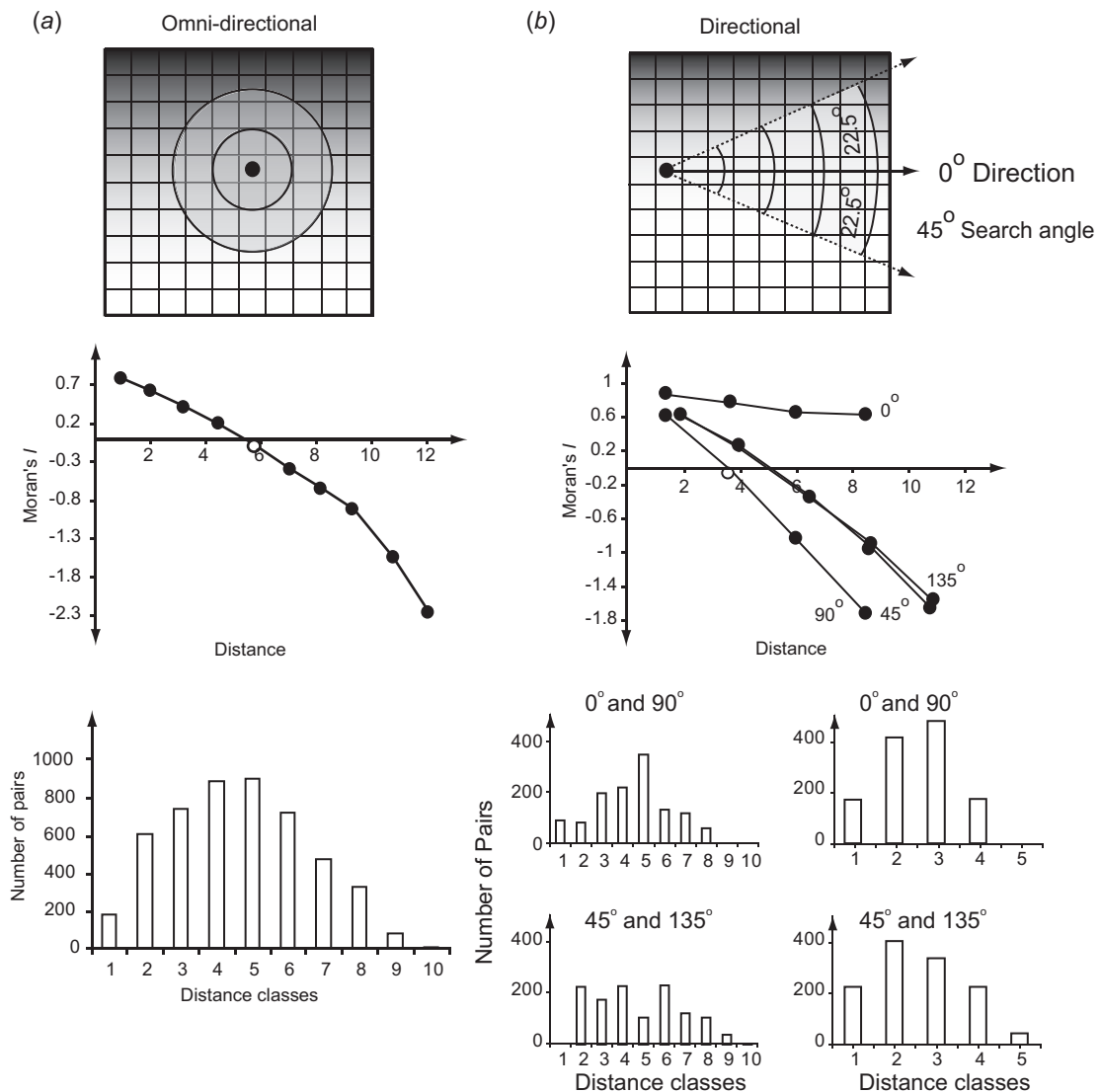


Figure 6.5 Moran's I spatial correlograms. (a) Omnidirectional correlogram based on 10 isotropic equidistant classes. The trend in the 10×10 lattice data is detected by the characteristic shape of the correlogram. The histogram of the pairs of sampling locations per distance class shows the edge effects at short and large distances (Chapter 3), but, except for the last distance class, there are more than 20 pairs. (b) Directional correlograms computed in four directions (0° , 45° , 90° and 135°) using a search angle of 45° (22.5° on each side). The directional correlograms are based on five equidistant classes; because with 10 equidistant classes, the number of pairs of locations per distance and angle class is too low. With fewer classes, the number of pairs per distance and angle class increases making the coefficient estimates and the comparisons more reliable. Here, the four directional correlograms allow us to detect anisotropy because the values of Moran's I do not all overlap one another (although the 45° and 135° ones are identical) and the correlograms have different spatial ranges (0° : no spatial range; 45° and 135° : around 5.0 units; and 90° : 3.8 units).

in which we want to compute the degree of spatial autocorrelation. The angle class matrix indicates, therefore, which sampling locations are in a given orientation with respect to each other given a specified search angle (the angular separation, a bit like the distance lag). Hence, for example, we could be interested in determining the spatial structure related to wind direction using search sectors 45° wide, centred on selected search directions of interest (i.e. 22.5° on each side of the selected direction, Figure 6.5*b*). Spatial autocorrelation can be computed in several directions and the resulting spatial correlograms are called ‘directional’ correlograms (Figure 6.5*b*). The detection of spatial anisotropy is then established by comparing how well the shapes and spatial ranges of the directional correlograms coincide. For example, in Figure 6.5*b*, while two directional correlograms (45° and 135°) coincide perfectly, both in shape and range, this is not true for all four directional correlograms because the shapes and spatial ranges differ (e.g. the 0° correlogram does not have a spatial range). The estimation of spatial autocorrelation based on both distance and direction is computed with fewer pairs of locations than in the omnidirectional case, and so we recommend using fewer distance classes (Figure 6.5*b*) and only four search direction classes (usually 0° , 45° , 90° and 135°).

6.2.3 Variography

Parallel to the development of spatial statistics by human geographers, mining engineers created a family of spatial statistics, known as geostatistics (Matheron 1970). Geostatistics use the spatial structure estimated from sampled data by computing the spatial variance (described in this section) to predict values at unsampled locations by modelling the spatial structure using techniques known as Kriging (Section 6.7.4). In this section, we provide a brief overview of the notions related to the estimation of this spatial variance, referred to as variography. More information about the wide range of geostatistical techniques can be found in numerous textbooks (among others, Journel & Huijbregts 1978; Isaaks & Srivastava 1989; Cressie 1993; Govaerts 1997; Schabenberger & Gotway 2005; Diggle & Justiniano 2010; Dutilleul 2011; Chilès & Delfiner 2012).

Geostatistics are based on ‘regionalized variable theory’, which assumes that the value of a variable z at a given location s is a particular realization of a random function $Z(x)$. The value, $z(x)$, is composed of three components:

$$z(s) = m(s) + \varepsilon(s) + \varepsilon, \quad (6.16)$$

where $m(s)$ is the deterministic structural function of the variable at location s ; $\varepsilon(s)$ is the spatially dependent residual from $m(s)$, that is the spatial variance component; and ε is the spatially independent normally distributed residual component. With stationarity, $m(s)$ is the average value of the variable within the study area, and the estimation of the spatial structure of a variable from sampled data is based on the intrinsic hypothesis of stationarity. That is, the expected difference between the values at two sampling locations at a given h distance apart is 0, and that the variance of this difference varies only according to h . Given that it is rare to have a process that is truly stationary over the entire study area, the spatial variance can be estimated only when quasi-stationarity (also called weak stationarity) prevails. Weak stationarity requires that the first-order moment, $E(Z(s))$, and second-order moment, $\text{Var}(Z(s))$, can be assumed to be stationary only within the range of a relatively small neighbourhood (i.e. within an isotropic moving window).

The notation common in geostatistics differs slightly from the notation in spatial statistics. For example, in geostatistics, the spatial lag parameter is ‘ h ’ (when in bold, \mathbf{h} denotes a vector of both distance and direction) rather than ‘ d ’. Hence, Eq. (6.10) for spatial autocorrelation can be rewritten using h instead of d :

$$\rho(h) = \frac{C(h)}{\sigma^2}. \quad (6.17)$$

Then, there is a direct relationship between spatial covariance, $C(h)$, and spatial variance, $\gamma(h)$, where the spatial variance is the variance, σ^2 , of the variable minus the autocovariance function:

$$\gamma(h) = \sigma^2 - C(h). \quad (6.18)$$

When the assumption of stationarity is met, spatial autocovariance and spatial variance are mirror images

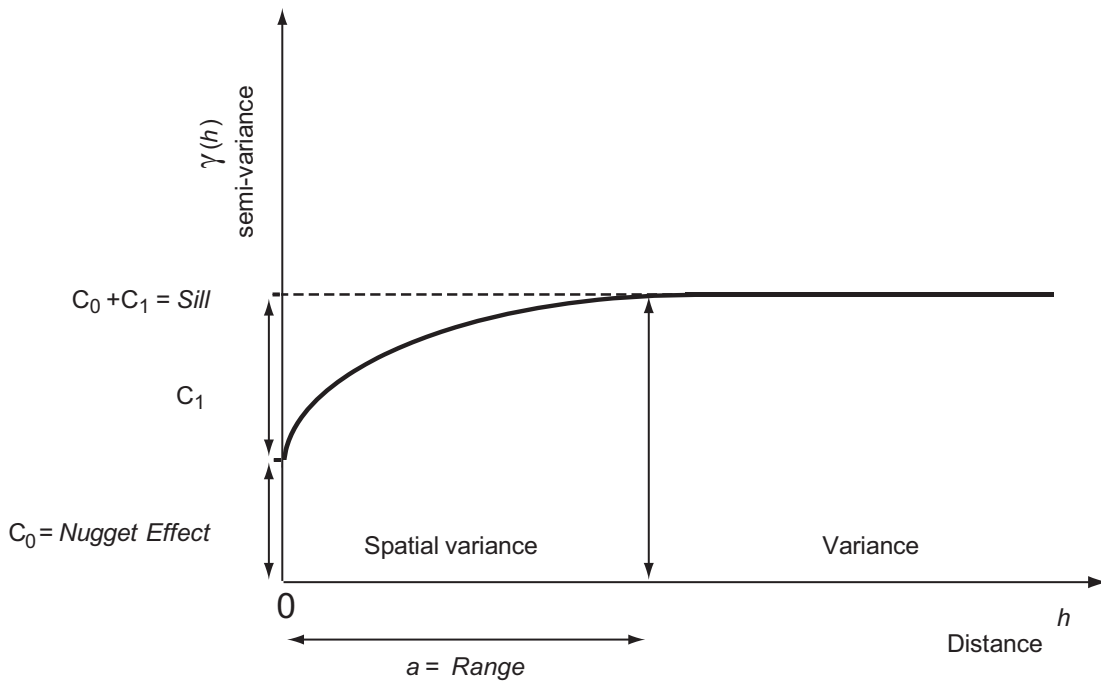


Figure 6.6 Experimental variogram (dashed line) and corresponding theoretical spherical variogram (solid line) and its parameters: the range, the nugget and the sill (h indicates the spatial lag; $\gamma(h)$ is the semi-variance).

of one another. Similarly, the relationship between spatial autocorrelation and spatial variance is

$$1 - \rho(h) = \frac{\gamma(h)}{\sigma^2}. \quad (6.19)$$

The spatial variance of the quantitative variable, z , is then estimated by the semi-variance function, $\hat{\gamma}(h)$:

$$\hat{\gamma}(h) = \frac{1}{2n(h)} \sum_{i=1}^{n(h)} (z(s_i) - z(s_i + h))^2, \quad (6.20)$$

where z is the value of the variable at the i th sampling location, s_i , and $n(h)$ is the number of pairs of sampling locations located at distance h from one another. Some geostatistics books use \mathbf{h} instead of h to indicate the more general case where the spatial variance is computed according to both lag distance, h , and direction, θ . Since the summation ranges from 1 to n (the

number of sampling locations), each pair of sampling locations is considered twice in the calculation. This is why the value is divided by 2 and the function is called the semi-variance. Technically, the plot of the semi-variance values against the spatial lag h is a semi-variogram. However, for simplicity, it is usually referred to as a 'variogram'. When a variogram is computed from sample data, it is called an experimental variogram (also called a sample or observed variogram; Figure 6.6). Then a model variogram or theoretical variogram can be modelled to fit the observed experimental variogram.

The equation for the semi-variance function (Eq. (6.20)) is quite comparable to the one of Geary's c (Eq. (6.13)) except that it lacks the division by the standard deviation in the denominator, which standardizes the spatial autocorrelation value. Hence, the semi-variance function is in the same units as the analysed data, and is not bounded as the values of Geary's c (0 to 2). At short distance lags, the values of

semi-variance are also small (close to zero) indicating that the spatial structure is at its strongest intensity. As the distance lags increase, the semi-variance values rise to level off at a plateau called the sill.

As for spatial correlogram, the shape of the variograms can provide insights about the spatial pattern studied. We computed experimental variograms using the same simulated data from Figure 6.3 as illustrated in Figure 6.7. These variograms are comparable to the spatial correlograms based on Geary's c (Figure 6.4) and are mirror images to those based on Moran's I (Figure 6.3).

Three key parameters are usually estimated from an experimental variogram to fit a theoretical variogram (Figures 6.6, 6.7 and 6.8): the nugget effect, C_0 , the spatial range, a , and the sill, C_1 . The nugget is the intercept at the origin that is greater than zero. Theoretically, at $h = 0$, the semi-variance is also equal to 0. However, based on the shape of the experimental variogram, it can be unrealistic sometimes to force the theoretical variogram to go through 0. The nugget parameter is therefore used to account for the observed variability at short distances due to local random effects or measurement errors (e.g. accuracy of measurements, inappropriate sampling unit size, etc.). The spatial range indicates the distance up to which the spatial structure varies. In other words, the range indicates the maximal distance at which the variable is spatially autocorrelated. Beyond the range, the distance among sampling locations does not affect the spatial structure of the data and the semi-variance values level off, forming the sill (e.g. variance of the variable). One implication of a variogram that reaches a constant sill value at some finite range is that pairs of samples separated by distances greater than that range are effectively independent of each other. There are cases where the experimental variogram does not have a sill. This also implies that the range is undetermined (Figure 6.8a) or that there is none. This can happen when the extent of the study area is smaller than the spatial pattern (i.e. the spatial range) of the variable of interest or if differences in the underlying processes continue to accumulate (see Bell *et al.* 1993a, b).

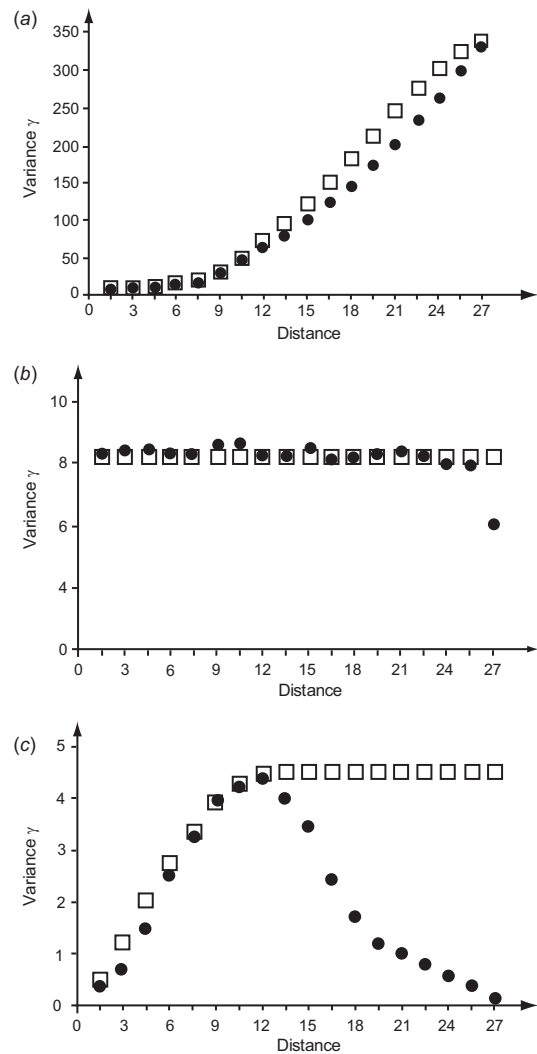


Figure 6.7 Experimental variogram (solid circles) and corresponding theoretical variogram (open squares) based on the same simulated spatial patterns from Figure 6.3. (a) Gradient: the variogram shows the corresponding characteristic trend where values increase as distance increases. (b) Random: the values are oscillating around a value of 8.2. (c) One big patch: the experimental values increase up to a distance of 12 units, the range, and then decrease at larger distances. The values of theoretical model, here the spherical model, are well fitted up to the range.

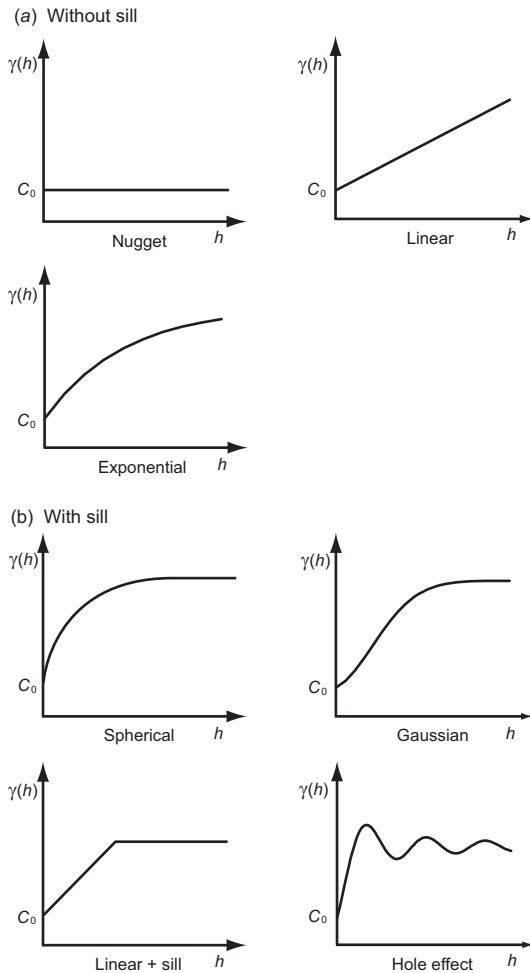


Figure 6.8 Theoretical variograms: (a) without a sill: showing the pure nugget effect, linear and exponential models; (b) with a sill: showing spherical, Gaussian and nested models (linear and a sill), and the hole effect. See text for the equations that describe these models.

Determination of the experimental variogram is only the first step in a series leading to the prediction of values at unsampled locations using Kriging (i.e. a family of spatial interpolation techniques as presented in Section 6.7). The next step involves fitting the best (derived analytically) theoretical variogram to the

experimental variogram. The most commonly used unbounded theoretical variograms (without a sill) are (Figure 6.8a):

- the nugget model:

$$\gamma(h) = C_0, \quad (6.21)$$

- the linear model:

$$\gamma(h) = C_0 + bh, \quad (6.22)$$

where b is the parameter of the slope,

- the exponential model:

$$\gamma(h) = C_0 + C_1 \left\{ 1 - \exp \left[-\frac{h}{a} \right] \right\}. \quad (6.23)$$

Bounded variograms (with a sill, Figure 6.8b) include:

- the spherical model:

$$\gamma(h) = C_0 + C_1 \left\{ \frac{3h}{2a} - \frac{1}{2} \left(\frac{h}{a} \right)^3 \right\}, \quad (6.24)$$

$$\gamma(h) = C_0 + C_1, \text{ for } h \geq a,$$

- the Gaussian model:

$$\gamma(h) = C_0 + C_1 \left\{ 1 - \exp \left(-3 \frac{h^2}{a^2} \right) \right\}, \quad (6.25)$$

- the linear model with a sill:

$$\gamma(h) = C_0 + bh \text{ for } 0 < h < a, \quad (6.26)$$

$$\gamma(h) = C_0 + C_1 \text{ for } h \geq a,$$

- the hole-effect model:

$$\gamma(h) = C_0 + C_1 \left\{ 1 - \frac{a \sin(h/a)}{h} \right\}. \quad (6.27)$$

Given that the strongest spatial variance usually occurs at short distances (i.e. up to the spatial range), it is important to get the best possible fit of the theoretical variogram model at the shorter distances (as illustrated in Figure 6.7). Adjustments of the theoretical variogram, and its parameters, are needed for the next step, which is spatial interpolation of the data using Kriging techniques (see Section 6.7). Inappropriate parameter estimations can result in very different (and possibly very misleading) predicted values. For example, when the spatial range is too small, the spatial structure will

be modelled for only small distances generating small patches, and when the spatial range is too large, the resulting spatial structure will be overly smoothed.

Selection of the best variogram model, and its parameters, used to be determined 'by eye' based on experience. The reliability of these parameter estimates can be evaluated using cross-validation techniques that require withholding some sampling locations and then comparing the variogram models' performance in predicting these values using Kriging. Nowadays, generalized least-squares and maximum likelihood measures (e.g. Akaike's information criteria) can be used to select the parameters, making the process more objective (Cressie 1993; Goovaerts 1997; Chilès & Delfiner 2012).

As for Geary's c , the semi-variance is affected by outlier values from skewed data because the estimation is based on the squared differences among the values of a given variable. This is why geostatisticians recommend that the data should be:

- (1) plotted at different spatial lags to identify outliers using h scattergrams (Rossi *et al.* 1992); and,
- (2) transformed, as required, to reduce the amount of skewness in the data.

Furthermore, as for spatial correlograms, the experimental variogram is a plot of semi-variance versus spatial lag h , where the number of pairs of distances decreases as distance increases. It is a rule of thumb to interpret only the first two-thirds of a variogram in order to determine the spatial structure of the sampled data. As mentioned above, \mathbf{h} denotes a vector including both spatial lag and direction. Anisotropic pattern can therefore be determined by computing directional variograms (Figure 6.9). In geostatistics, three types of anisotropy are distinguished:

- (1) the ranges differ but the sills are the same (geometric anisotropy; Figure 6.9a),
- (2) the ranges are the same but the sills differ (zonal anisotropy; Figure 6.9b), and
- (3) both the ranges and the sills differ.

Geometric anisotropy can be modelled, using a linear relationship, to predict the values of a variable at unsampled locations. It is not as straightforward for the two other types of anisotropy where more than one variogram model is needed to model the spatial

structure. This can be achieved by determining different models as a function of distance, as illustrated in Figure 6.9, where up to the range a spherical model is used and beyond the range, the sill is used.

Ecological data are rarely stationary and often show some trend or gradient called 'drift' in geostatistics. The presence of drift in the data implies that $m(s)$ is not a good estimate of the average value of the variable over the entire study area. Drift can be detected by comparing the values of $m(s)$ computed using only the values at the sampling locations at the beginning of vector \mathbf{h} (i.e. at $z(s_i)$, called 'head' locations) with those computed with sampling locations at the end of vector \mathbf{h} (i.e. at $z(s_i + \mathbf{h})$, called 'tail' locations). When these average values differ, the covariance is said to be non-ergodic (where ergodic covariance implies that $m = m_{-h}(\text{head}) = m_{+h}(\text{tail})$; Isaaks & Srivastava 1989; Rossi *et al.* 1992). In the presence of drift, the generalized random intrinsic functions of order k should be used to estimate the spatial variance, rather than the experimental variogram (see more advanced geostatistical textbooks for mathematical details: Journel & Huijbregts 1978; Goovaerts 1997; Chilès & Delfiner 2012; Webster & Oliver 2001, 2007).

Ecological data are often essentially binary (e.g. presence : absence) or based on threshold values to indicate state responses that are of interest. The spatial variance of qualitative or categorical data can be estimated using indicator functions, where k is the threshold value at which continuous data are considered as the response of interest (e.g. presence) and are set to 1 or otherwise are set to 0. Continuous variables can therefore be transformed into binary indicator variables $i(s; z)$, where the average of the indicator variable is the proportion of the variable in the study area (Rossi *et al.* 1992). The semi-variance for an indicator function is computed as:

$$\hat{\gamma}_{i(s;z)}(h) = \frac{1}{2n(h)} \sum_{i=1}^{n(h)} (i(s_i; z) - i(s_i + h; z))^2. \quad (6.28)$$

The indicator function can be extended to include several thresholds corresponding to multiple-state

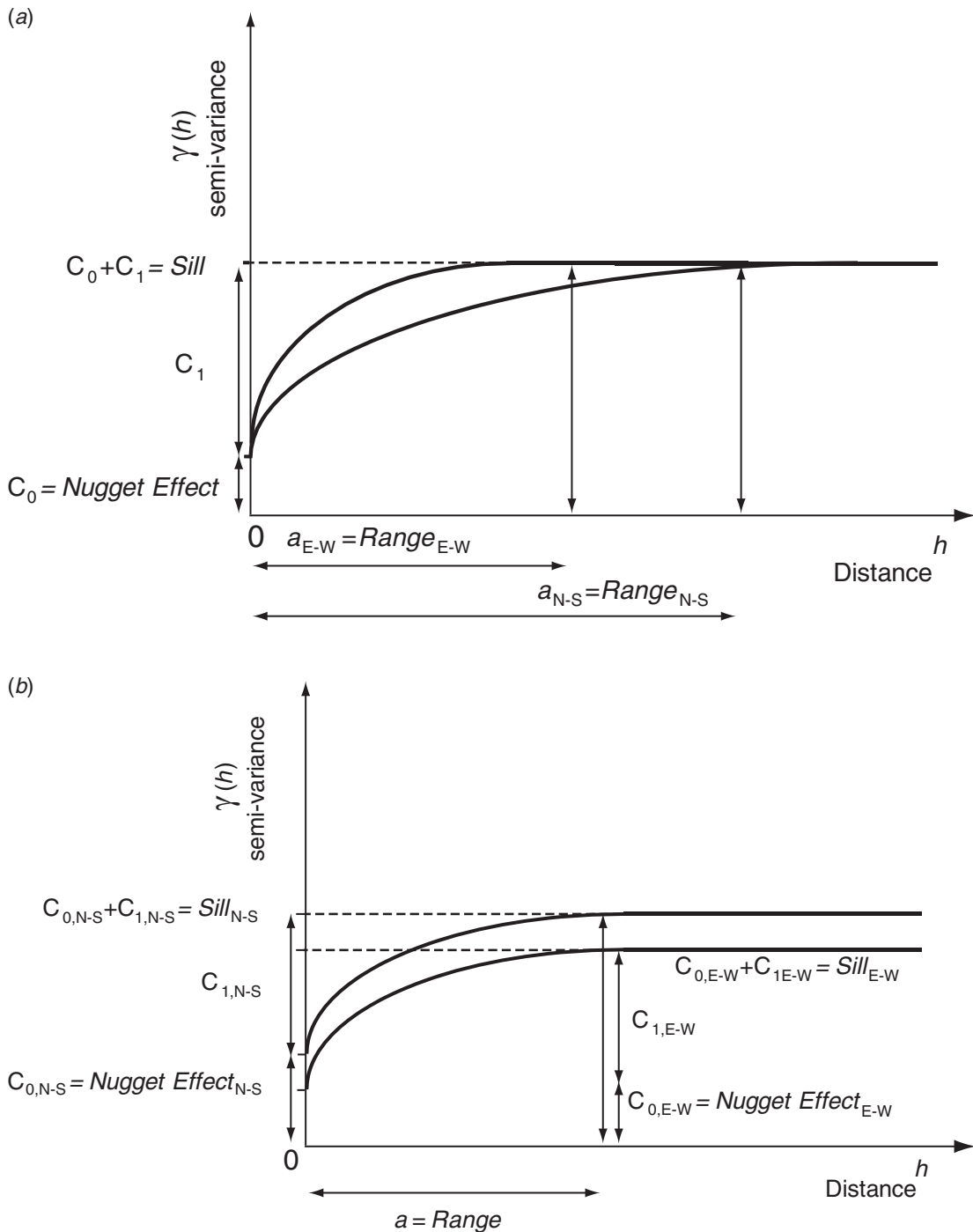


Figure 6.9 Directional variograms (here in the north-south, N-S, and east-west, E-W, directions) allowing the identification of anisotropic spatial variance in the data. (a) Geometric anisotropy is depicted as the range values differ but the sill values are the same. (b) Zonal anisotropy is depicted as the range values are the same but the sill values differ.

variables, or to explore the sensitivity of the selected thresholds (Webster & Oliver 2007; Polakowska *et al.* 2012).

To conclude this brief overview of variography, it is important to stress that variograms aim to determine the parameters used to model spatial pattern. Hence, to maintain the same measurement units as the original data, semi-variance values are not standardized. Furthermore, while no significance tests have been developed for semi-variance function values, randomization and bootstrap procedures can be used to assess their significance. Finally, there is a much wider range of geostatistical techniques available than those presented here, but further details are beyond the scope of this book. We refer the reader to more advanced textbooks for greater detail (among others, Deutsch & Journel 1992; Cressie 1993; Goovaerts 1997; Chilès & Delfiner 2012).

6.2.4 Fractal dimension

When dealing with landscape features and entities that can be delineated, such as rivers, islands, patches, and so on, we may be interested in characterizing and quantifying their spatial structure. The most fundamental property of these landscape entities is their geometrical dimension. In fact, objects can lie either in Euclidean space, with an integer number of dimensions (0 for points, 1 for lines, 2 for surfaces, and 3 for volumes), or in a fractal dimension (fractal for fractional), introducing a new way of characterizing the occupancy of space by objects (between 0 and 1 for clusters of points, 1 and 2 for curves, 2 and 3 for surfaces, and 3 and 4 for volumes). The fractal dimension is therefore a mathematical coefficient that measures the fractal geometry (i.e. non-integer dimension) of objects in physical space (Mandelbrot 1983). In theory, fractal objects can have several interesting properties, such as having the same fractal dimension at all scales of observation: the property of self-similarity. Furthermore, the fractal dimension is independent of units of measurement (whether centimetres or kilometres), allowing comparisons among objects, organisms, community structures, and so on. The fractal dimension can also be seen as a measure of the degree of occupancy of a low-dimensional object

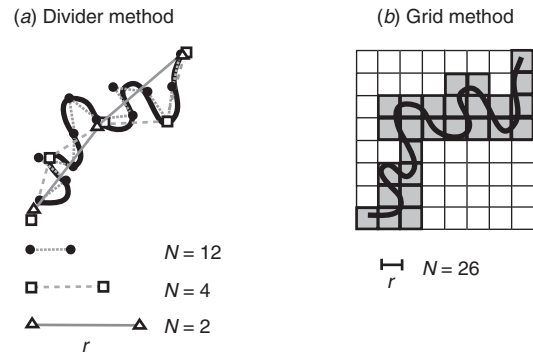


Figure 6.10 Fractal dimension templates. (a) Divider method: three sizes of sticks, r , and the respective number of sticks needed to measure the shaped line. (b) Grid method: the grey shaded boxes indicate those that include the shaped line and their number.

in a higher-dimensional space. The concept of fractal dimension has since been extended to include quantification of the spatial structure of continuous data using statistical fractals. Ecologists have been exploring the potential usefulness of the concept of fractal dimension for spatial characterization of landscape patches (Gustafson 1998), for modelling of an organism's geometric growth such as exhibited by plants and shells, as well as for modelling of landscape spatial structures (Milne 1992; Palmer 1992; Kenkel & Walker 1993; Hargrove *et al.* 2002). Whatever the application, fractals can be seen at least as a good null model of spatial structure with which to compare the spatial characteristics of ecologically important systems over a range of scales (Halley *et al.* 2004).

One of the simplest ways of computing fractal dimension is to measure the length of the object using several different sizes of a divider (e.g. a ruler). The fractal dimension of an object, D , can be estimated using a simple relationship between the number of units, N , needed to measure either the length (divider dimension, Figure 6.10a) or area occupied (grid dimension, Figure 6.10b), and the size of the 'divider' or the 'box', r , used. Using the box method:

$$N(r) \propto r^{-D}, \quad (6.29)$$

so that the slope of the log of $N(r)$ against the log of r , gives an estimate of $-D$, the fractal dimension.

Similarly, the statistical fractal dimension of ecological patterns from continuous data can be computed from the slope of the variogram in a log-log plot (assuming that the variogram is isotropic, linear, and without a sill):

$$2\hat{\gamma}(h) = h^{4-2D}, \quad (6.30)$$

where the slope is $(4-2D)$. Consequently, there is a direct relationship between both spatial variance (variogram) and fractal dimension as defined in Eq. (6.30), as well as between spatial autocorrelation (correlogram) and fractal dimension. Figure 6.11 shows this, by comparing the fractal dimensions estimated from theoretical linear variograms. We recommend plotting the variogram before estimating the fractal dimension with continuous data to evaluate the relevance of using this method.

When examining the fractal properties of ecological objects, which we do not really expect to be self-similar at all scales, what is perhaps most interesting are the scales at which the fractal dimension changes (Burrough 1981; Allen & Hoekstra 1992). This is because at these critical scales, ecological processes that act upon the variable of interest may also be changing (Dungan *et al.* 2002). Thus fractal dimensions may be used to distinguish the range of influence of a process, implying different levels of organization of organisms in space; furthermore, in moving from one scale to another, we may be sure that we are changing between levels of organization. Therefore a shift in the fractal dimension could be interpreted as a change in ecological processes, reflected by a reorganization in the structural scale of the parameters measured from the data (Figure 6.12).

6.3 Sampling design effects on the estimation of spatial pattern

As presented in Chapter 1, regardless of the type of spatial analysis used, the sampling design (including the three components of sample size, sampling unit size, and their layout in space) determines the power

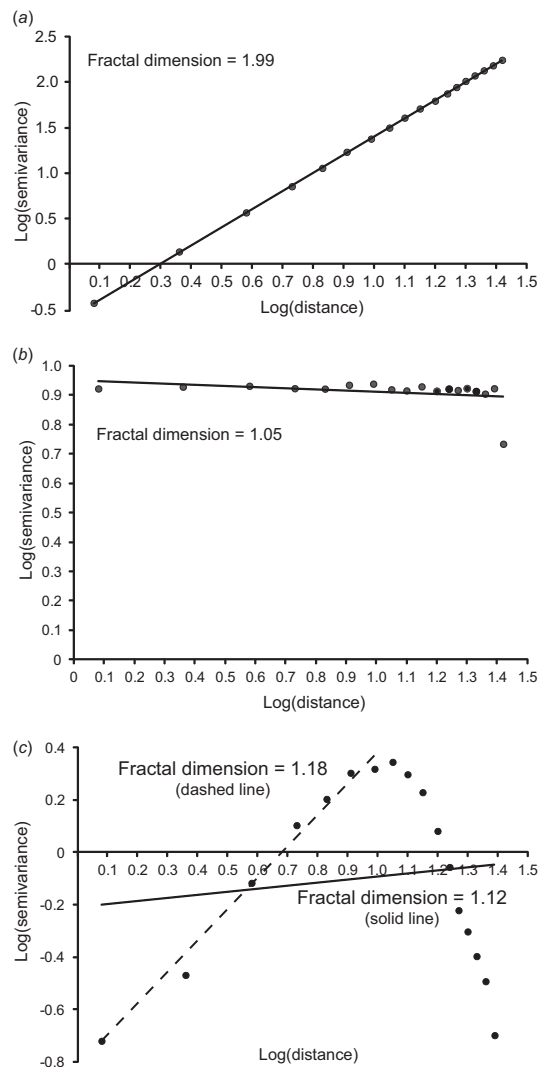


Figure 6.11 Variogram plots on the log scale (distance and semi-variance) based on the same simulated spatial patterns from Figure 6.3 (see Figure 6.7) such that the slope of the linear regression is the fractal dimension. (a) The gradient data have a fractal dimension of 1.99. (b) The random data have a fractal dimension of 1.05. (c) The single patch data have a fractal dimension of 1.12 based on the 17 coefficient values or a fractal dimension of 1.18 based on the eight first coefficient values which is the spatial range of the patch (dashed line).

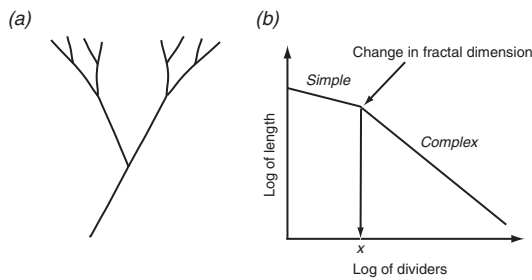


Figure 6.12 Change in fractal dimension values as divider size increases. In (a), the shape of a patch is smooth at short distances (straight lines) as indicated by the middle slope (i.e. low fractal dimension), and complex (peninsulas and bays) at intermediate distances as indicated by the steeper slope (i.e. higher fractal dimension). (b) The scale at which the change in fractal dimension (i.e. pattern) occurs can be used as an indicator of change in ecological processes: here from man-made edges to patch shape matching geomorphological feature such as a relief. (Adapted from Allen & Hoekstra 1992.)

to detect a significant signal in the sample data. Furthermore, determination of one of the components of the sampling design affects the other two (Dungan *et al.* 2002). The first component to be determined is usually the sample size because it is directly related to our sampling effort or ability (usually limited by cost and time). In time series analysis, it is recommended that at least 100 data points be sampled in order to detect temporal patterns. In geostatistics, the records of at least 100 sampling locations are also recommended. In ecology, it is quite rare to have such large sample sizes when field work is involved. In fact, when a spatial signal is strong, fewer sampling locations, say around 50, can detect it. Moreover, if the spacing among sampling units is such that it is within the spatial range of the pattern, then even fewer sampling locations, say between 20 and 30, may be able to capture the essence of the spatial pattern. We stress that the spatial pattern needs to be clear in order to detect it with so few sampling locations (Fortin 1999a). This does not hold when the pattern is weak to start with, as illustrated by Figure 6.13: the spatial pattern is detected only with 84 sampling locations and not when the sample size is reduced to 42 or 24, while

keeping the sampling unit size constant. This holds true even when different equidistant classes are used (here 17.4 m and 34.8 m; Figure 6.13). Therefore, failure to detect significant spatial pattern does not necessarily imply that there is no spatial pattern and a revised sampling design may be able to detect one, if present (Fortin *et al.* 1989; Fortin 1999a).

There are two ways to revise the sampling design: (1) by changing the spacing among sampling units and (2) by changing the size of the sampling unit. The spacing of the sample locations can be arranged to maximize, as much as possible given the sample size, the number of locations within the spatial range of the variable (see Figure 1.10). Spatial layouts with more than one spatial lag seem to be more efficient at capturing spatial pattern with relatively small sample sizes (Fortin *et al.* 1989; Webster & Oliver 2007). Interestingly, a random sampling design is able to detect significant spatial pattern because there is a wide range of spatial distances among the sampling locations (Fortin *et al.* 1989).

The other aspect to consider is whether the choice of sampling unit size is the most appropriate one. For example, when spatial autocorrelation at the first distance class is not significant (high nugget effect in the variogram) or negative, this can indicate that the sampling unit size is larger than the spatial range of the spatial pattern, or that it includes more than one spatial pattern (Fortin 1999a). In such circumstances, a smaller sampling unit size should be used. The question is: how small? When no prior knowledge is available about a system, we recommend using the smallest sampling unit size that is large enough to include more than one object of interest (as presented in Chapter 1). Then, with this relatively fine spatial resolution, the data can be aggregated into a coarser resolution (Figure 6.14; Qi & Wu 1996; Fortin 1999a; Hui 2009). By doing so, the spatial domain of a pattern, and the underlying process (or processes), can be identified, as well as the size or distance at which it changes. This is comparable to searching for a change in fractal dimension value (Figure 6.12). Indeed, for a range of sampling unit sizes, estimation of spatial autocorrelation will be comparable up to a given size, at which the intensity of spatial autocorrelation will change (Fortin 1999a). For example, in Figure 6.14,

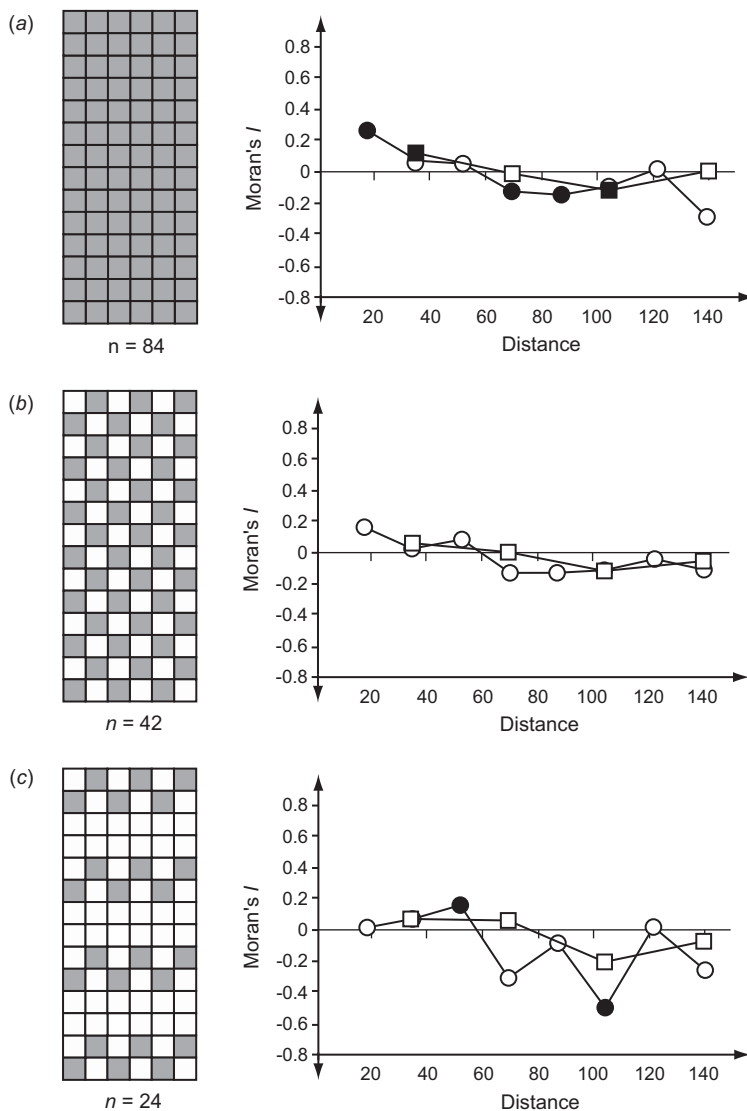


Figure 6.13 Sample size effect on the estimation of spatial autocorrelation. Moran's I omnidirectional correlograms of sassafras abundance data using (a) 84 sampling locations (indicated in grey), (b) 42 sampling locations (indicated in grey) and (c) 24 sampling locations (indicated in grey). Solid symbols indicate significant values at 0.05; open symbols indicate nonsignificant values. Correlograms were computed using eight equidistant classes (distance interval = 17.4 m; circles) and four equidistant classes (distance interval = 34.8 m; squares). Significant values only occur when 84 sampling locations with eight equidistant classes are used for the analysis. All other subsampling combinations ($n = 42$ or $n = 24$) and numbers of equidistant classes do not detect the patchy structure of the sassafras data.

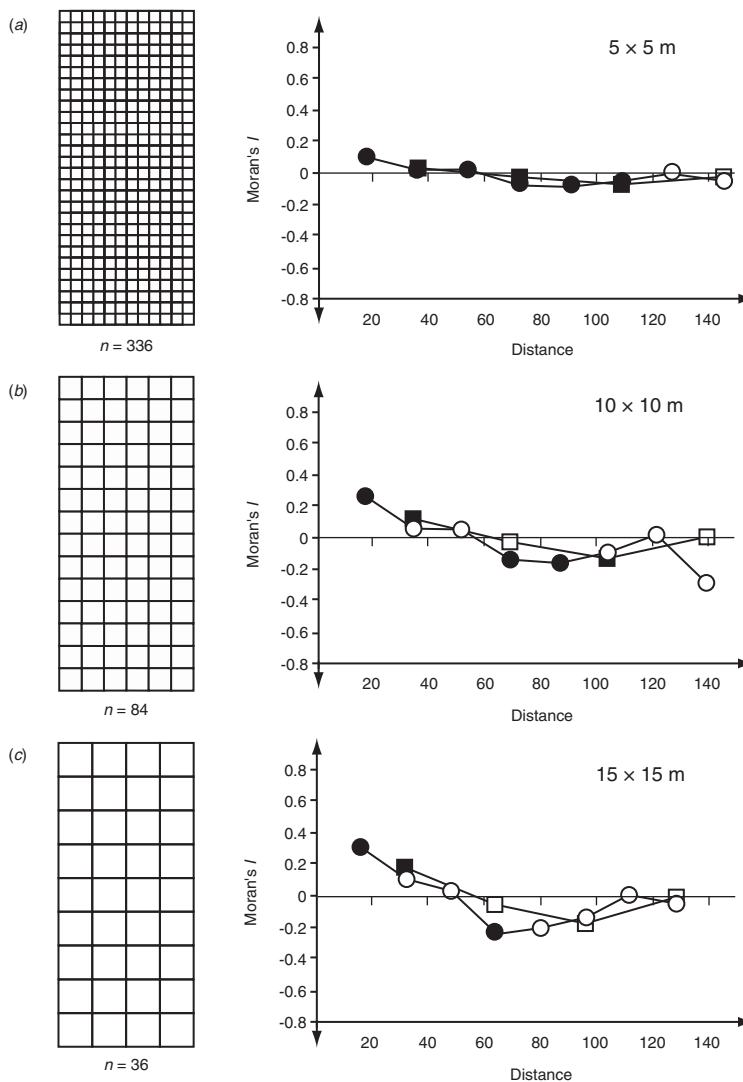


Figure 6.14 Sampling unit size effect on the estimation of spatial autocorrelation. Moran's I omnidirectional correlograms of sassafras abundance data using (a) a 5 × 5 m sampling unit size ($n = 336$), (b) a 10 × 10 m sampling unit size ($n = 84$), (c) a 15 × 15 m sampling unit size ($n = 36$) and (d) a 20 × 20 m sampling unit size ($n = 21$). Solid symbols indicate significant values at 0.05; open symbols indicate nonsignificant values. Correlograms were computed using eight (circles) and four (squares) equidistant classes. Overall, with two equidistant intervals, the estimation of spatial autocorrelation increases (from 0.113 to 0.250 to 0.303) with increasing sample unit size (5, 10 and 15 m, respectively) but decreases to 0.154 at the larger sampling unit size (20 m). Comparison among these correlograms can be used to select the optimal sampling unit size needed to analyse the spatial structure. Here, the sampling unit of 15 m estimated the highest values of spatial autocorrelation.

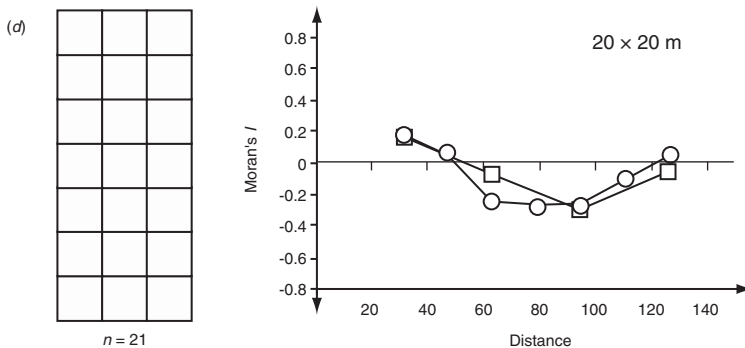


Figure 6.14 (cont.)

as the sampling unit size increases from 5×5 m to 10×10 m, the autocorrelation value in the first distance class increases from 0.113 (significant) to 0.250 (significant); then at a sampling unit size of 15×15 m, it reaches its maximum value of 0.303 (significant), and begins to decrease thereafter to 0.154 with a 20×20 m sampling unit (not significant). Although the values in the first distance class using 5×5 m and 20×20 m units are similar, the reasons for obtaining these values are different: the 5×5 m unit is too small and includes only one or two tree stems, so several sample units are empty, resulting in a weak spatial autocorrelation structure because it is similar to a random spacing level. With the 20×20 m unit, the sampling unit size is large enough to include tree stems but may also include more than one process. It may also be larger than the spatial range, such that there is random noise as well. By comparing the results from Figure 6.13 with those of Figure 6.14, using more or less comparable sample sizes of 42 and 36, we see that the spatial autocorrelation estimated was strongest with 36 sampling locations. Thus, this sampling unit size was more appropriate because it could contain more than one tree. For comparable sample sizes, 24 and 21, respectively, the spatial pattern was too weak to be detected despite the fact that 21 units sampled the entire area.

The fact that values of spatial autocorrelation, or spatial variance, change with the sample unit size is well known in spatial analysis (Wu & Qi 2000; Dungan *et al.* 2002) and in plant ecology (Greig-Smith 1961).

It is referred to as the 'modifiable area unit problem' (MAUP; Openshaw 1984) or as the 'change of support' in geostatistics (Journel & Huijbregts 1978; Dungan *et al.* 2002). This is related to the 'ecological fallacy' problem that inferences about individual objects can be made based on aggregated information (e.g. using data recorded at the sampling unit level to make conclusions about individual tree stems while no information about individual trees was in fact recorded). These are inherent problems for any kind of spatial data and they are the subject of active research in geography, (especially with the advent of geographical information systems; O'Sullivan & Unwin 2003, 2010), in satellite image analysis (Woodcock & Strahler 1987; Arbia *et al.* 1996; Bradshaw & Fortin 2000; Dungan 2001), in geostatistics (Cressie 1996; Bellehumeur & Legendre 1997; Bellehumeur *et al.* 1997), as well as in landscape ecology (Wu *et al.* 2002).

6.4 Spatial relationship between two variables

So far, we have presented spatial methods that quantify the spatial structure of only one variable at a time. There are obvious cases where it is of interest to analyse the spatial interaction between two variables, as presented in Chapter 4 (e.g. bivariate Ripley's K), or to predict the spatial pattern of one variable based on the

known correlation between two variables (e.g. co-Kriging, as presented in Section 6.7). To do so, spatial autocorrelation coefficients can be modified to estimate cross-correlation between two variables:

$$I_{xy}(d) = \left(\frac{n}{W(d)} \right) \frac{\sum_{i=1}^n \sum_{\substack{j=1 \\ i \neq j}}^n w_{ij}(d)(x_i - \bar{x})(y_j - \bar{y})}{\left(\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \right) \left(\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2} \right)}. \quad (6.31)$$

Similarly, the semi-variance function can be modified to estimate the cross-covariance function:

$$\hat{\gamma}_{uv}(\mathbf{h}) = \frac{1}{2n(\mathbf{h})} \sum_{i=1}^{n(\mathbf{h})} \left(z_u(s_i) - z_u(s_i + \mathbf{h}) \right) \left(z_v(s_i) - z_v(s_i + \mathbf{h}) \right). \quad (6.32)$$

While the mathematics is quite straightforward, very few software packages offer the option of computing cross correlation. The cross-variance function, being a step in co-Kriging analysis, is more widely available in geostatistical packages.

6.5 Local spatial statistics

As more and more ecologists engage in ecological studies with large spatial extents, the likelihood that their data sets violate the assumption of stationarity is very high. When using global spatial statistics with these data sets, local and small areas of spatial heterogeneity are masked by the fact that the statistics are summaries for the entire study area, with a single average value of spatial autocorrelation, or a series of average values for different distances as in a correlogram or variogram. In such cases, a global assessment of spatial dependence can be misleading because average values provide information about neither the range of variability in intensity of spatial dependence nor the exact localization of patterns. For example, in a sloping study area, tree abundance may vary from the top of the slope (say high) to the bottom (say

moderately low), with some small openings here and there, creating localized patches with lower tree abundance. While global statistics may detect the large-scale trend in abundance values, they may miss local patterns because, by their algorithm, they lump all the local deviations together by summing and then averaging. While the average value of spatial dependence is meaningful where only one process occurs (either induced or inherent dependence), it is misleading when several processes act at various intensities in different parts of the study area.

These limitations and the misapplication of global spatial statistics have been acknowledged for at least two decades (Fortin 1992; Getis & Ord 1992; Anselin 1995; Sokal *et al.* 1998; Fotheringham *et al.* 2000; Boots 2002; Wulder *et al.* 2007). Also, the recognition that local estimation of the intensity of spatial autocorrelation may reveal interesting insights into local processes resulted in the development of local spatial statistics (Getis & Ord 1992; Anselin 1995; Kabos & Csillag 2002; Boots 2003; Goovaerts & Jacquez 2005). Anselin (1995) proposed 'LISA' (local indicator of spatial association) as an acronym for these local spatial statistics. Here, we will use the acronym LISA in an inclusive way to refer to local statistics quantifying spatial dependence in general.

As presented in Section 6.2 and illustrated in Figures 6.3 and 6.4, the calculation of global spatial statistics, such as Moran's I and Geary's c , requires a computation at each sampling location, comparing with the mean or with neighbouring locations. Then, these deviations, which are local deviations, are lumped together. Hence, in essence, these global spatial statistics are averages of the local spatial variations over a study area. It is therefore not too surprising that the first local spatial statistics proposed were modified global spatial statistics, calculated at each sampling location i based only on its j local neighbours, and not accumulated into an overall average. Consequently, the local Moran's I_i is calculated as

$$I_i(d) = \frac{(x_i - \bar{x})}{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \sum_{\substack{j=1 \\ j \neq i}}^n w_{ij}(d)(x_j - \bar{x}), \quad (6.33)$$

where $w_{ij}(d)$ is the weight matrix given a local neighbourhood search of radius d . Here the weight matrix can be either binary, stressing only the connectivity among sampling locations (see Figure 3.12), or made up of actual weights (e.g. inverse distance weighting function), to emphasize the local neighbourhood effect on local spatial pattern.

Under the assumption of completely uniform randomization (i.e. all randomizations are equally likely), the expected value of I_i is

$$E(I_i) = \frac{-1}{n-1} \sum_{j=1}^n w_{ij}. \quad (6.34)$$

Consequently, unlike global Moran's I , which has the same expected value for the entire study area and the different distance classes, the expected value of local Moran's I varies for each sampling location because it is proportional to the number of neighbours. The significance of local I_i can be tested by comparison to the standard normal distribution once I_i is transformed by subtracting the expected value and dividing by the square root of the variance, call it $z(I_i)$:

$$z(I_i) = \frac{[I_i - E(I_i)]}{\sqrt{\text{Var}(I_i)}}. \quad (6.35)$$

The required value of $\text{Var}(I_i)$ for this equation can be derived by assuming either complete randomization or conditional randomization, and the mathematical details of these equations can be found in Boots (2002). As for global Moran's I , local Moran's I_i can be computed with different neighbour search distances, d . Any type of Bonferroni correction that adjusts for both the number of sampling locations and the number of neighbour search distances, which results in a very large number of multiple comparisons, would be too conservative, and only rarely would I_i be considered significant. In general, we recommend using local spatial statistics as spatial exploratory tools to detect localized spatial structures, especially when a lack of stationarity is suspected.

As with global Moran's I , the local version computes the deviations from the average value of the variable, x , over the entire study area. Positive I_i values occur when the value at location i is similar to those of its

neighbours in their deviation from the average (\bar{x}). In other words, positive values of I_i indicate that the values in the vicinity of location i and at location i are either all larger (positive deviation) or smaller (negative deviation) than the average. Negative values of I_i also indicate that deviation from the average is either larger or smaller than the average but where the value at location i is of a different sign from its neighbours. When the value of I_i is close to zero, the deviation from the average is small and no local spatial structure can be detected. This can occur either because there is no spatial pattern or because there is a subtle pattern that we cannot detect because the local values are too similar to the overall average. Figure 6.15a shows significant local I_i at $\alpha = 0.05$ based on the same simulated data used to compute the omnidirectional correlograms in Figure 6.3d (16 regularly spaced patches of equal size), Figure 6.3e (9 irregularly spaced patches of equal size) and Figure 6.3f (12 regularly spaced patches of different size). These maps of I_i help to identify the position, size, shape and layout of local spatial structures. For example, in the case of the 16 regularly spaced patches (Figure 6.3d), the centroid of each patch is identified as having the most positive spatial autocorrelation with its neighbours (Figure 6.15a). The spatial maps of local I_i values for the two other cases (Figure 6.15a) are not as informative, however, because we cannot discriminate between clusters of high and low spatial structure as significant because positive values of I_i (as indicated by +) can result from either clusters of high or low values. Indeed, without looking at the maps of the raw data (Figure 6.3e, f), we cannot determine which clusters have high or low values. Other local spatial statistics may be more appropriate in such cases, as described below.

Global Geary's c can also be modified to obtain a local spatial statistic, the local c_i :

$$c_i(d) = \frac{1}{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \sum_{\substack{j=1 \\ j \neq i}}^n w_{ij}(d) (x_i - x_j)^2. \quad (6.36)$$

The difference between the local I_i and the local c_i is in the numerator, where for local I_i it is the

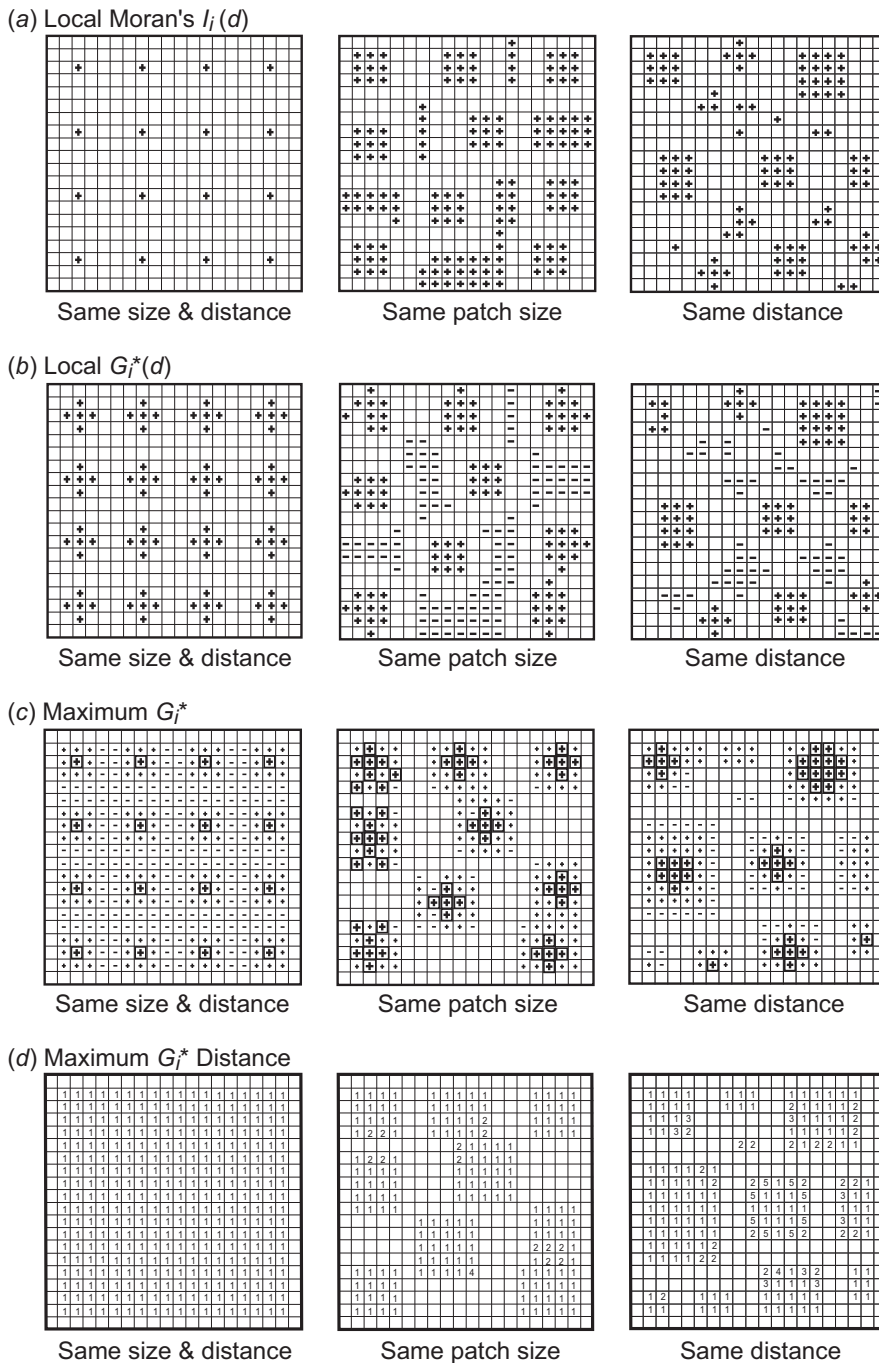


Figure 6.15 LISA (Local Indicators of Spatial Association) using the same simulated data shown in Figure 6.3: 16 patches (Figure 6.3d); nine patches (Figure 6.3e) and 12 patches (Figure 6.3f). (a) Local Moran's I_i estimated using a distance interval class of 1.5 units (+ indicates significant values at 0.05, positive or negative). (b) Local G_i^* estimated using a distance interval class of 1.5 units (+ indicates positive and - indicates negative significant values at 0.05). (c) Maximum local G_i^* estimated using five distance classes of 1.5 units each (+ in a square indicates positive and - negative significant values at 0.05; + indicates positive values but nonsignificant). (d) Maximum local G_i^* distance (number indicates the distance class at which the maximum local G_i^* was estimated). See text for more detail.

deviation from the value at location i and the overall average of the variable, while for local c_i it is the value of the variable at location i . Similarly, under the assumption of complete randomness, the expected value of c_i is proportional to the number of local neighbours:

$$E(c_i) = \frac{2n}{n-1} \sum_{j=1}^n w_{ij}. \quad (6.37)$$

The equations for $\text{Var}(c_i)$ under complete and conditional randomness can be found in Boots (2002). Here, positive values of local Geary's c_i correspond to cases where the value at location i is similar to its neighbours, while negative values indicate a difference in sign from its neighbours.

Getis & Ord (1992) proposed two new local spatial statistics: local G_i , in which the value at location i is excluded from the computation; and local G_i^* in which the value at location i is included. Local G_i is computed as follows:

$$G_i(d) = \frac{\sum_{\substack{j=1 \\ j \neq i}}^n w_{ij}(d)x_j}{\sum_{\substack{j=1 \\ j \neq i}}^n x_j}, \quad (6.38)$$

where the expected value, under the assumption of complete randomness, depends on the number of local neighbours:

$$E(G_i) = \frac{1}{n-1} \sum_{\substack{j=1 \\ j \neq i}}^n w_{ij}. \quad (6.39)$$

Similarly, the local G_i^* is computed as:

$$G_i^*(d) = \frac{\sum_{j=1}^n w_{ij}(d)x_j}{\sum_{j=1}^n x_j}, \quad (6.40)$$

where the expected value, under the assumption of complete randomness, also depends on the number of local neighbours:

$$E(G_j^*) = \frac{1}{n} \sum_{j=1}^n w_{ij}. \quad (6.41)$$

The expected variances of the two local Getis statistics are given in Boots (2002).

These two local Getis statistics are, in essence, the ratio of local averages (around the location i and at the location i) over the global average of the variable of interest for the entire study area, (i.e. local spatial moving averages). These statistics detect clusters of either high or low values, which are often referred to as 'hot spots' or 'cold spots', respectively. As with local Moran's I_i , the local G statistics cannot differentiate between cases where there is an absence of spatial structure and cases where the local average equals the global average.

As the number of neighbours increases with neighbour search distance, the G statistics are asymptotically normally distributed and can be standardized to facilitate interpretation (Getis & Ord 1996): positive G values indicate clusters with high values (hot spots), and negative ones indicate clusters with low values (cold spots). Figure 6.15b illustrates how, unlike local Moran's I_i , G_i^* can discriminate between locations of significant hot spots, indicated by +, and cold spots, indicated by -, in the three different spatial arrangements of patches.

Wulder & Boots (1998) proposed that the G_i^* statistic could be computed using different neighbour search distances, d , identifying:

- (1) the maximum G_i^* value calculated at each location i regardless of d , and
- (2) the distance d at which the maximum G_i^* value occurred.

Maps of these two values provide information about (1) the maximum intensity of spatial dependence at a location i , and (2) its local spatial extent. Following Wulder & Boots' (1998) suggestion, we mapped the maximum G_i^* (Figure 6.15c) and the maximum G_i^* distance (Figure 6.15d) based on the three spatial arrangements of patches. The maximum G_i^* map

allows us to identify the centre of the patches, as indicated by positive values (the + in squares), while the maximum G_i^* distance map shows the local spatial range (extent) of each patch.

Despite the fact that local statistics, G_i and G_i^* , are more informative than local Moran's I_b , they are still estimated relative to the global average of the data and are therefore sensitive to the presence of overall global spatial structure in the data (e.g. a trend). The H Moran statistic proposed by Kabos & Csillag (2002) addressed this issue by computing a local statistic for qualitative data that is not affected by the presence of a global pattern both in its computation and in its significance tests. In the spirit of assessing significance locally, Boots (2003) proposed a local join count approach that he called the local indicator for categorical data (LICD). In short, the LICD assesses the significance of the local spatial arrangement (configuration) of categorical data by adjusting for the proportion (composition) of each category within a local neighbourhood search window. So far, there are no equivalent local statistics for quantitative data that can account for the presence of global patterns in the data and detect local spatial patterns. This may be another area for further work on the development of local spatial statistics for ecological applications.

6.6 Spatial scan statistics

In the same way as clusters of high and low values of spatially explicit local statistics, such as the Getis statistics just described, can be used to identify 'hot spots' and 'cold spots' of any spatial characteristic, there is a set of statistical techniques designed to detect spatial clusters of events or of similar values. These are generally referred to as 'scan statistics' (also known as Kulldorff scan statistics), because they scan the data for clusters, and usually involve the calculation of associated probabilities, the values of which can be used to indicate the most unexpected clusters (Glaz *et al.* 2001). As a simple example, consider a circle of area a placed at random in a large sample plane of area A in which there are n point-events in total. The probability that the circle contains m or more events is

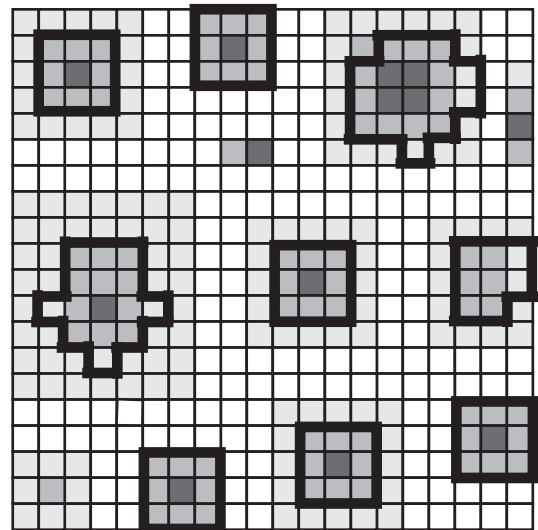


Figure 6.16 Scan statistic based on the same simulated data shown in Figure 6.3f. The grey tone corresponds to the scan statistic values: darker grey indicates higher probability of being a significant spatial cluster. The bold lines indicate local limit of each patch based on the CSR test at $\alpha = 0.05$.

$$P_a(m) = \sum_{i=m}^n \binom{n}{i} \left(\frac{a}{A}\right)^i \left(1 - \frac{a}{A}\right)^{n-i}. \quad (6.42)$$

By considering a large number of randomly placed circles of different sizes, we could claim that those circles with the highest density of events, m/a , and the lowest probability $P_a(m)$ were the best indicators of the location of clusters of the events of interest. Of course, that claim is over-simplistic because the calculated probabilities are not correct because of lack of independence among the circles and among the counts within circles. This example does, however, provide the basic concept of the use of a simple spatial scan statistic to detect significant clusters of events by comparison with complete spatial randomness (CSR) as the null hypothesis, which can be tested using a Poisson randomization procedure (Figure 6.16).

Scan statistics can be formulated in purely spatial, purely temporal, or combined spatio-temporal versions, depending on whether the clusters being sought

exist in space, in time or in both. Typical examples to which scan statistics are applied include examining health data for spatial clusters of a particular type of cancer, scanning time series records for clumps of rare events, or studying epidemiological data for groupings of disease onset in time and space (Grillet *et al.* 2010a). The reliability of the probabilistic calculations that determines how unexpected the detected clusters actually are depends on the assumption of process stationarity, as do so many of the techniques we describe.

The key to understanding scan statistics is, of course, the word ‘scan’; this fits quite naturally with our theme of using particular templates for the calculation of exploratory or analytical statistics (see Table 4.2). Imagine standing on a boulder in the middle of a flat plain, using binoculars to search through a full 360° circle and all the way from the boulder to the horizon for aggregations of the antelope which dot the plain. The field of view in the binoculars is the template, the number counted is the raw statistic, and the number corrected for distance and for comparison with the overall density is the scan statistic of interest. The probabilities of the values of this statistic are treated as a ‘discrepancy score’ (Agarwal *et al.* 2006) and examined for surprising deviations. The direction with the greatest positive discrepancy score gives the location of the hottest of ‘hot spots’ for groupings of antelope. In general, spatial scan statistics are designed to detect and evaluate clusters, whether the clusters are related to disease, resources, or critical areas for monitoring or policy implementation (cf. Patil & Taillie 2004). Spatial scans make use of localized windows which sweep over the entire study area, with a probability or likelihood ratio associated with each location; and there may be an obvious interaction between the shape of the template used and the shape of the spatial clusters detected (cf. Duczmal *et al.* 2006). The second key to understanding the interpretation of scan statistics is to remind ourselves that low probability occurrences are still possible, and given enough positions of a scan template, some will have apparently significant results, even when the system is truly random. The larger the number of template positions used, the more likely it is that some will appear

to be significant. (Some coincidences are just coincidental!) Instead of relying on the raw probabilities for each template and position combination, the maximum of a likelihood ratio is used (Kulldorff 1997) and in many cases, the number of events of an ‘at risk’ population must be considered as well as the number of those in the category of interest.

The simplest example of how scan statistics work is to consider a number of events occurring in a single dimension, whether it represents time or one spatial dimension, and then to look for unexpected clusters of events. As a concrete example, consider a line transect 1000 m in length, set out through sparse desert vegetation, which is used to record the positions at which shrubs and the line intersect. Although the transect intersects only 20 shrubs in its entire length, in one 50 m section it intersects five. The question then is whether this local density is consistent with a null hypothesis of uniform and independent occurrence, or whether there is reason to suspect that the five intersections represent a cluster. We need to calculate $p(5/30 | 50/1000)$, the probability that any 50 m section of 1000 m contains five or more independent events; this is the additive inverse of the probability of encountering four or fewer (cf. formula provided by Glaz *et al.* 2001, p. 12):

$$\begin{aligned} p(5/30 | 50/1000) &= 1 - \sum_{k=0}^4 \binom{30}{k} \left(\frac{50}{1000}\right)^k \left(\frac{950}{1000}\right)^{30-k} \\ &= 1 - 0.98436 = 0.0156. \end{aligned} \quad (6.43)$$

This probability calculation shows that it is quite improbable that an independent arrangement would result in one-sixth of the plants being found in one-twentieth of the transect length, suggesting the existence of a cluster. A similar technique can detect the clusters of species boundaries in transects through zoned plant communities (cf. Dale 1988), which we will address in Chapter 9. Here, the template for calculation is a simple line segment of a given length, say 50 m, that can be visualized as sliding up and down the entire 1000 m of the line transect. One of the problems for this approach is that the scanning template or moving window may identify a number of potential clusters that overlap, and it may be necessary to

determine which of these has the lowest probability associated with it. On the other hand, one could use non-overlapping positions of the template so that this situation is avoided, but then finding the exact position of the cluster becomes highly unlikely.

The calculations used to illustrate the one-dimensional case were relatively easy to carry out because the numbers were small, but that may not always be the case. When the numbers involved are large, these probability calculations become unwieldy and much of the work in developing scan statistics has been in providing useful and accurate approximations (cf. Glaz *et al.* 2001; Agarwal *et al.* 2006), as well as determining the power of these tests.

We used a one-dimensional example to introduce the technique, but a large proportion of the applications of spatial scan statistics are two-dimensional, dealing with geographic data such as locations of disease or other kinds of critical events. For example, in forest ecology, spatial scan statistics have been used to detect 'hot spots' of forest degradation in the United States of America (Riitters & Coulston 2005). Where our one-dimensional scan template was just a line segment, in two dimensions a circle, ellipse, square or rectangle (cf. Glaz *et al.* 2001, p. 65) will be used. The isodiametric forms (circle or square) are most usual, and of course match best when the phenomenon being sought is isotropic, or at least when there is no good reason to expect otherwise. In some instances, however, when anisotropic structures are known or expected, the use of anisodiametric templates such as ellipses of different sizes can prove useful for mildly irregular clusters, although for highly irregular clusters a simulated annealing method may be preferable (Duczmal *et al.* 2006).

If we can concentrate on the use of circles as templates for two-dimensional scanning, it becomes obvious that the technique is closely related to others we have already described such as Ripley's *K*-function analysis, Openshaw's 'geographical analysis machine' (Openshaw 1987), and the circumcircle methods, because they also use counts of events in circles as their basis. Both Openshaw's method and circumcircles are close relatives of the scan technique, both using circles of different size, with positions

determined by a grid in the first case and by the positions of the events themselves in the second. In either case, the calculations are also similar, resulting in a statistic that compares the observed counts with the expected in some standardized way.

Applications of scan statistics in three spatial dimensions are not common (two spatial dimensions and one temporal being much more usual), but they do exist. One of the most commonly cited applications concerns the clustering of galaxies in space, for which the three dimensions are reduced to two by using scan statistics on astronomical photographs (Glaz *et al.* 2001, p.79), as is also true of the star cluster example, originally by R. A. Fisher, given in Application 5.3 and Figure 5.1 of Glaz *et al.* (2001).

Scan statistics are often applied in the spatio-temporal analyses of epidemiological data (Grillet *et al.* 2010a; Read *et al.* 2011), but scan statistics appear only rarely in ecological studies. There is clearly the potential for greater use of these techniques in ecology, particularly as the use of large geographic databases continues to grow. More work may need to be done on developing those approaches that are best suited to ecological spatial data and on making them available in a readily applied format.

6.7 Interpolation and spatial models

In most spatial studies, we collect data to obtain information about a target population over the study area of interest. In ecological studies, such information could be related to species abundance, species behaviour, and so on. In a spatial context, we are often interested in the estimation and prediction of a variable at unsampled locations within the given study area. This can be achieved by modelling spatial pattern using spatial interpolation techniques which assume that the process that generated the spatial pattern is stationary. As in any regression and modelling context, a spatial model attempts to summarize the spatial pattern using as few parameters as possible. In fitting these parameters to the data, the spatial interpolation techniques are modelling the main spatial signal and trying to minimize the error resulting in smoothed

estimated values at locations not sampled. Several interpolation methods are available, each having various advantages and limitations in the way the spatial structure is modelled. Interpolation techniques can be classified according to the following broad properties.

- Global: a single interpolation function is used to interpolate the values for the entire study area. The resulting map of the interpolated data is usually a smooth surface (e.g. trend surface analysis). Change in one value will affect the function and thus the predicted values.
- Local: the interpolation function is applied locally for a limited number of locations. The resulting map of the interpolated data is still smooth but includes both global and local patterns. Change in one value will affect only neighbouring locations (e.g. proximity polygons, inverse distance weighting, Kriging).
- Approximate: at the sampling locations the predicted values will not be the same as the observed ones (e.g. trend surface analysis).
- Exact: at the sampling locations the predicted values will be exactly the observed values (e.g. proximity polygons, inverse distance weighting, Kriging).

Other properties can also be used to characterize these interpolation techniques: deterministic (i.e. there is only one possible predicted value at a given location – all methods except Kriging, which can be stochastic when used in a conditional annealing procedure), point interpolators (again all methods except the proximity polygons method), and areal (only the proximity polygons and Kriging). Here, we provide only a brief overview of these methods as there are several textbooks that present them in great detail (Journel & Huijbregts 1978; Isaaks & Srivastava 1989; Haining 1990, 2003; Cressie 1993; Goovaerts 1997; Webster & Oliver 2007; O'Sullivan & Unwin 2010; Chilès & Delfiner 2012; among others).

There are also spatial models (moving average, MA; simultaneous autoregressive, SAR; conditional autoregressive, CAR) that simulate data with a known degree of spatial dependence. These spatial autoregression models are mostly used to simulate data rather than to interpolate and therefore we will present them in Chapter 8.

6.7.1 Proximity polygons

Probably every day without realizing it we interpolate sampled data assuming that the sampled quantity does not vary over the entire sampling unit. This concept can be extended to point location data (which we refer to as point events in Chapter 4) where polygons (also called Dirichlet, Thiessen or Voronoi; Chapter 3) are defined based on the x - y coordinates of each point given their proximity to each other (Okabe *et al.* 1992, 2000). The notion of proximity refers here both to the spatial arrangement of all neighbouring points and to how these points can potentially interact with one another. The characteristic at the x - y coordinate is then assigned to the whole area of the polygon; when the characteristic is qualitative, this creates a categorical attribute; when it is quantitative, it is a numerical value that is assumed to apply to the entire polygon. Therefore, these Voronoi polygons can be used to define the spatial area of influence around each point (Figure 6.17). This simple technique results in abrupt changes in characteristics from one polygon to another.

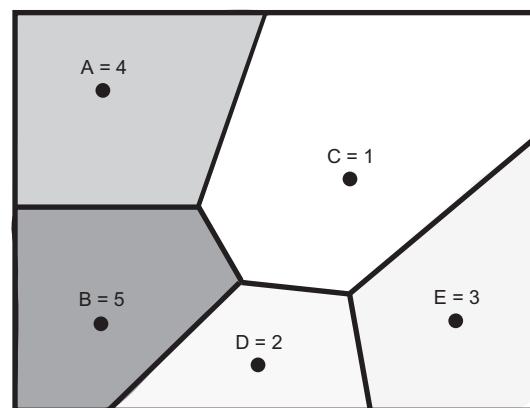


Figure 6.17 Proximity polygon interpolation using Voronoi polygons. The five sampling locations (A–E) were first linked using a Delaunay tessellation algorithm and then the Voronoi polygons (solid lines) were determined. Each polygon indicates the zone of influence of each sampling location and the value at the sampling location is assigned to the entire polygon (here indicated by a different grey tone corresponding to the sampling value).

6.7.2 Trend surface analysis

In an aspatial context, and when data for a sampled quantitative variable are available, the best predictor at an unsampled location is the average value of the variable over the entire sample. When data and knowledge of the relationship between an independent variable and the variable of interest exist, the most commonly used interpolation method is regression. In a spatial context, the x - y coordinates of sampling units can be used as independent variables in a regression. Hence, when the spatial pattern is a linear trend, values can be interpolated using multiple linear regression (here without interaction term between x and y):

$$\hat{z}(\mathbf{s}_0) = b_0 + b_1x + b_2y, \quad (6.44)$$

where $\hat{z}(\mathbf{s}_0)$ is the predicted value at location \mathbf{s}_0 , b_0 is the intercept, and b_1 and b_2 are coefficients of the slope of the surface (Figure 6.18a). This multiple regression approach is simple and useful because a general equation can be used over the entire study area to model large-scale spatial patterns.

When the overall spatial pattern of the study area is a nonlinear trend, the values can be approximated using polynomial regression of various orders such as quadratic (x^2 , or y^2), cubic (x^3 , or y^3) or higher orders. Hence when the pattern is a relatively smooth, monotonic, curved surface (e.g. hill or valley shape), a second-order polynomial (quadratic) can be used to interpolate the data (Figure 6.18b):

$$\hat{z}(\mathbf{s}_0) = b_0 + b_1x + b_2y + b_3x^2 + b_4xy + b_5y^2. \quad (6.45)$$

When the surface has a saddle, a third-order polynomial (cubic) may be appropriate (Figure 6.18c):

$$\begin{aligned} \hat{z}(\mathbf{s}_0) = & b_0 + b_1x + b_2y + b_3x^2 + b_4xy + b_5y^2 \\ & + b_6y^3 + b_7y^2 + b_8y^2 + b_9y^3. \end{aligned} \quad (6.46)$$

We could continue to fit the sample data using higher and higher order polynomials (e.g. Figure 6.18d). This is not recommended, however, because the gain in simplicity acquired with polynomial regression is lost due to the necessity of estimating several regression coefficients. More importantly, by improving the fit in some

areas, others may start to fit less well, becoming distorted. In fact, trend surface analysis is a global interpolator and should not be used to model local spatial pattern. So, in a nutshell, the advantages of using a trend surface analysis method to interpolate data are:

- (1) no a priori knowledge of the spatial pattern is needed because the interpolation is based on empirical data,
- (2) both multiple and polynomial regression are available in most statistical software packages which also offer the possibility of testing the significance of regression coefficients using F -test. This can help evaluate which order of polynomial is best (linear, quadratic, or higher).

The disadvantages of the trend surface analysis are, as mentioned, that it should not be used when there are several small patches and that it is not an exact interpolator but rather an approximate one, such that at the sampled locations the predicted values are not equal to the observed ones.

6.7.3 Inverse distance weighting

In the spirit of trend surface analysis, linear interpolation can be used to interpolate data based on the data from sampled locations given a restricted neighbourhood search area. The underlying premise is that nearby locations are more likely to have similar values and the linear interpolator weights the interpolated data, $\hat{z}(\mathbf{s}_0)$ at unsampled locations \mathbf{s}_0 , according to the proximity of known sampled data as follows:

$$\hat{z}(\mathbf{s}_0) = \sum_{j=1}^m (w_j z(\mathbf{s}_j)). \quad (6.47)$$

Here $z(\mathbf{s}_j)$ is the value of variable z at the sampled location j , m is the number of neighbouring sampling locations based on some definition such as being within a search radius, and w_j is the weight according to the distance among the unsampled and sampled locations such that:

$$\sum_{j=1}^m w_j = 1. \quad (6.48)$$

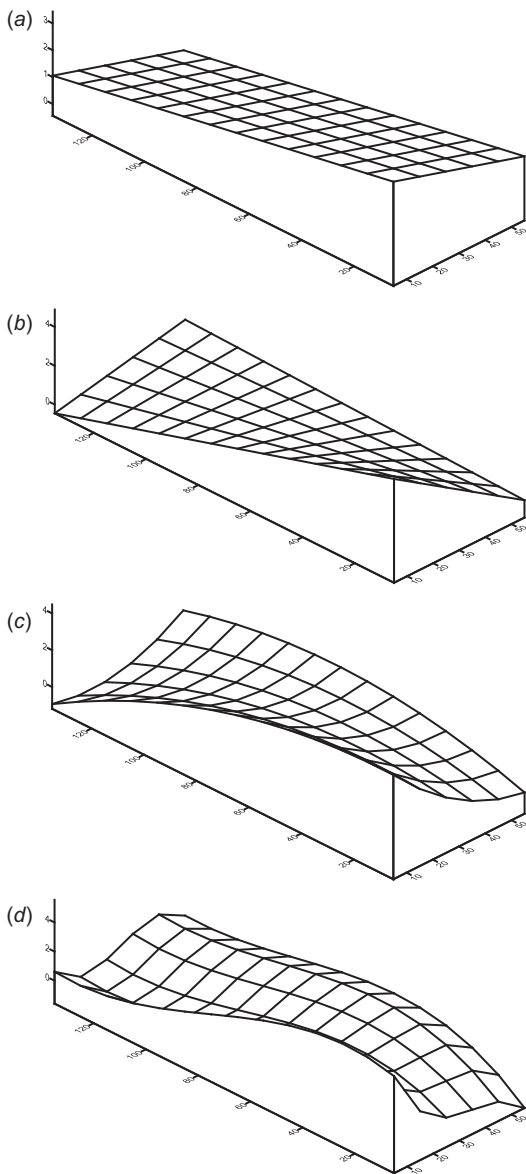


Figure 6.18 Trend surface analysis: (a) first-order polynomial (linear surface), (b) second-order polynomial (nonlinear surface: valley type), (c) third-order polynomial (nonlinear surface: saddle surface), and (d) fourth-order polynomial (nonlinear surface: wavy surface).

The most common form of the inverse distance weight is

$$\hat{z}(\mathbf{s}_0) = \sum_{j=1}^m \left(d_{ij}^{-k} z(\mathbf{s}_j) \right) \div \sum_{j=1}^m d_{ij}^{-k}, \quad (6.49)$$

where k is a real value from 0 to 1, and d_{ij} is the distance between the unsampled location i and sampled location j . More weight can be put on nearby locations by varying the value of the exponent k . When the distance between the sampled and the unsampled locations is zero, the interpolated value will be the observed one. The advantage in this type of linear interpolator is that it is weighted locally around each sampling location therefore preserving more of the complexity of local spatial patterns (Figure 6.19) than trend surface analysis. It is also very easy to use and does not require prior knowledge about the data. When a map of the study area is needed for illustration purposes alone, this linear interpolation technique is quite useful. It does not, however, provide any information about the discrepancies between the interpolated values and the 'real' spatial pattern at the unsampled locations. This is why Kriging is often preferred over the inverse distance weighting method.

6.7.4 Kriging

Stemming from the two previous interpolation methods, Krige (1966) proposed an interpolation method using a system of linear equations, based on prior knowledge of the degree of spatial dependence in the data. This spatial interpolation technique is, in essence, a weighted moving average technique and it is called Kriging. Kriging is one of the geostatistical techniques (Journel & Huijbregts 1978) and it uses the parameters (spatial range, nugget and sill) estimated by the experimental variogram as described in Section 6.2.3.

Having its origin in applications to mining questions, Kriging was first developed to address specific needs in the spatial prediction of ore resources, such

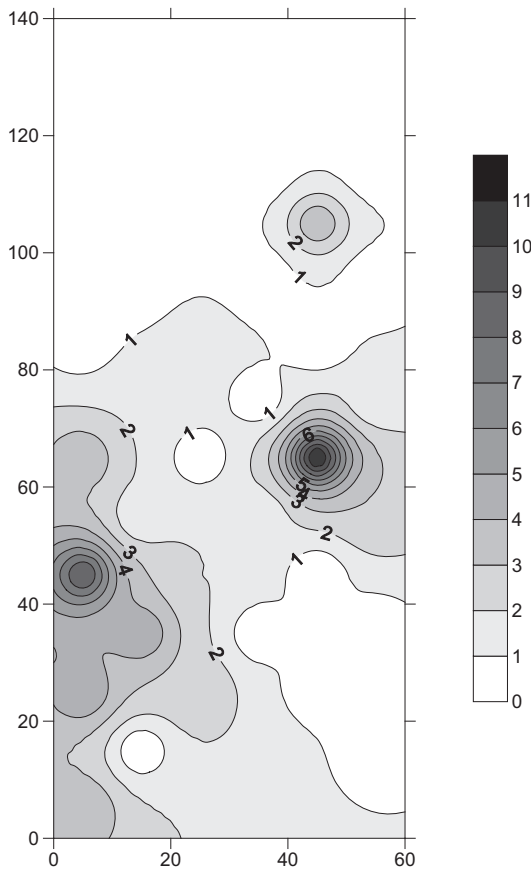


Figure 6.19 Inverse weighted distance. Interpolated values of sassafras abundance data ($n = 84$) based on 42 sampling locations (as shown in Figure 6.13b). The estimated values are illustrated using contour lines.

as the spatial interpolating from either punctual (point) or block (area) samples over a two-dimensional region, as well as predicting the values of the ore for a given volume (three-dimensional). This last feature made this technique known quite rapidly in oceanography (see Simard *et al.* 2003 for a recent example) and meteorology, but it took longer to become widespread in ecology (Legendre & Fortin 1989; Rossi *et al.* 1992).

Kriging is a set of linear regressions that determine the best combination of weights to interpolate the data, as in the inverse weight distance method, by minimizing

the variance as derived from the spatial covariance in the data. Here, the weights, w_i , are based on the spatial parameters of the variogram model, which are derived from an experimental variogram such that sampling locations within the spatial range (distance) have more influence on the predicted value. To solve this system of linear algebraic equations, the sum of the weights is constrained to equal 1, such that there are more equations than unknown parameters to estimate:

$$\hat{z}(\mathbf{s}_0) = \sum_{j=1}^m (w_j z(\mathbf{s}_j)), \text{ with } \sum_{j=1}^m w_j = 1, \quad (6.50)$$

where m is the number of neighbouring sampling locations within a given search neighbourhood. To minimize the estimator error, the Lagrangian multiplier (λ) is added as a constant (for details see Journel & Huijbregts 1978):

$$\sigma_E^2(\mathbf{s}_0) = \sum_{j=1}^m (w_j \gamma(\mathbf{s}_j, \mathbf{s}_0)) + \lambda, \quad (6.51)$$

The estimation error is also called the Kriging variance or Kriging error. Equation (6.50) can be written in matrix notation:

$$\mathbf{C}\mathbf{w} = \mathbf{c}, \quad (6.52)$$

where the predicted value at unsampled location \mathbf{s}_0 , is a vector \mathbf{c} that is obtained by multiplying the variance-covariance matrix, \mathbf{C} , between known locations, i and j , as estimated by the theoretical variogram model selected; by the vector of weights, \mathbf{w} , which are to be determined; \mathbf{c} is the vector of covariances between the sampled locations i and the unsampled location 0. The covariance matrix and vector values are given by the variogram model, where the vector of weights is to be estimated:

$$\begin{bmatrix} \gamma(d_{11}) & \gamma(d_{12}) & \cdots & \gamma(d_{1m}) & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \gamma(d_{m1}) & \gamma(d_{m2}) & \cdots & \gamma(d_{mm}) & 1 \\ 1 & 1 & \cdots & 1 & 0 \end{bmatrix} \begin{bmatrix} w_1 \\ \vdots \\ w_m \\ \lambda \end{bmatrix} = \begin{bmatrix} \gamma(d_{10}) \\ \vdots \\ \gamma(d_{m0}) \\ 1 \end{bmatrix}. \quad (6.53)$$

This is achieved by multiplying both sides of Eq. (6.52) with the inverse covariance matrix, \mathbf{C}^{-1} , such that the vector \mathbf{w} can be solved:

$$\begin{aligned}
\mathbf{C}\mathbf{w} &= \mathbf{c} \\
\mathbf{C}^{-1}\mathbf{C}\mathbf{w} &= \mathbf{C}^{-1}\mathbf{c} \\
\mathbf{I}\mathbf{w} &= \mathbf{C}^{-1}\mathbf{c} \\
\mathbf{w} &= \mathbf{C}^{-1}\mathbf{c},
\end{aligned}
\tag{6.54}$$

where $(\mathbf{C}^{-1}\mathbf{C})$ is the identity matrix \mathbf{I} (equivalent to multiplication by 1). The determination of weights is therefore related to both the variogram model and the number of sampling locations considered. Most of the geostatistical software packages offer two types of rules to determine the search neighbourhood, that is advice on how to select the number of sampled locations to consider in an interpolation procedure. The first method is to define a search distance. Given that most of the spatial dependence occurs within the spatial range, the search distance should not exceed the spatial range of the variogram. There are cases, however, when the spatial layout of the sampling locations is such that there are very few locations within the range. It is then recommended that the search radius be increased until a minimum number of locations is reached. Usually, a minimum of 15–20 neighbouring points is used. It is important to realize, however, that in these circumstances beyond the spatial range, the values at sampled locations contribute little to the Kriged values and the resulting Kriged values will show a smooth spatial pattern. Also, since the best estimation of the weights necessitates the inversion of the covariance matrix at each interpolated location, it is faster to perform if this matrix is not too large. The search neighbourhood does not need to be isotropic and can be more elliptic or even a volume in the case of three-dimensional Kriging.

Kriging has some of the same features as trend surface analysis in that only one model is used for the entire study area. It is also similar to the inverse weighted distance method in that the interpolation is performed locally. Thus, given that the weights are proportional both to the spatial variance of the data and to the distance among sampling locations, Kriging is an exact interpolator because the Kriged values at sampling locations are equal to the observed ones. That is not to say that the selected variogram model and its parameters are the best ones. This is why cross-validation was proposed in order to evaluate

the overall robustness of the model. Cross-validation consists of removing each sampled location one at a time and then Kriging at that location; and then comparing how reliably this Kriged value matches the observed one. This procedure was very important early on when maximum likelihood methods did not exist to facilitate model selection. The effectiveness of the Kriging depends on how well the selected model fits the data.

Another way to evaluate the plausibility of Kriged values is to map both the Kriged values and their associated estimation error (Figure 6.20). Indeed, given that the Kriging variance is in the same units as the Kriged data, areas where the errors are higher than others can be identified. These areas of high variance can be due to:

- (1) too few sampled locations within the spatial range in those areas, or
- (2) the selection of an inappropriate variogram model or wrong parameter values (e.g. Figure 6.21).

Note that the Kriged errors are relative to the variogram model used. If the selected theoretical model seems to be the best one, these areas of high errors can be used to detect sectors requiring more sampling effort. This procedure is often used to determine the optimal spatial sampling design (Webster & Oliver 2001).

Building from this system of linear equations, several Kriging variants have been developed to account for the particular characteristics of environmental data such that the property of unbiased predictors can hold (Goovaerts 1997; Webster & Oliver 2007). First, interpolation can be performed for specific x - y coordinates (*punctual* Kriging) or for an area (*blocked* Kriging). When the mean is known, that is when second-order stationarity applies, this system of linear equations is referred to as *simple* Kriging. In most cases, the mean is not known and only the weak stationarity assumption prevails, thus *ordinary* Kriging should be used. However, especially with environmental data, there could be a large-scale trend as well as local spatial patterns in the data. This implies that there is a 'drift' in the mean and that the simple system of linear equations cannot be used. *Universal* Kriging was developed to model such

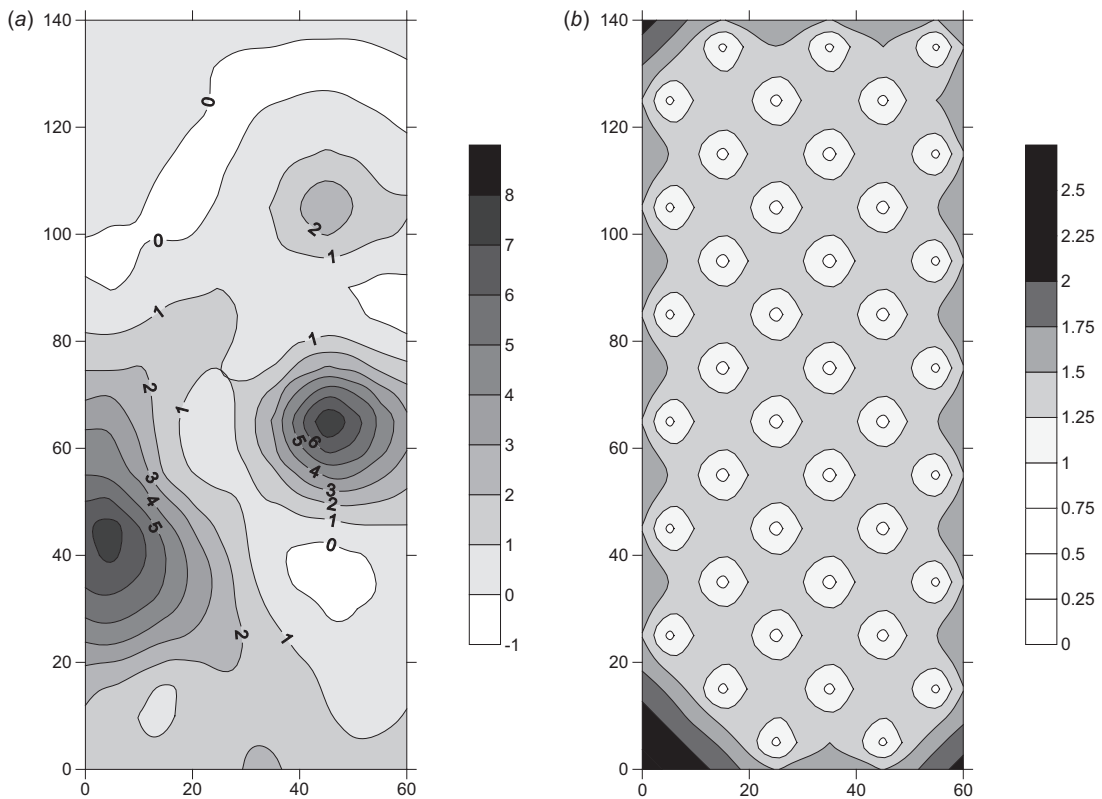


Figure 6.20 Kriged values of sassafras abundance data ($n = 84$) based on 42 sampling locations (as shown in Figure 6.13b). The estimated values are illustrated using contour lines. (a) An isotropic spherical model using a range of 32 m. (b) The associated Kriging errors with the Kriged values based on the isotropic spherical model. The Kriged errors are higher at the unsampled locations than at the sampled ones, but given the uniform spatial sampling design, the errors are comparable in term of values.

large-scale trends and then the residuals are Kriged after the trend is removed. As with trend surface analysis, the trend may not necessarily be linear, so an 'intrinsic random function of order k ' should be used. These functions are the equivalent of the k -order polynomial functions in trend surface analysis and account for nonlinear trends in the data before Kriging. When the spatial pattern shows nonlinearity that cannot be fitted by a polynomial function, *non-linear* Kriging can be used.

As described earlier in this chapter, spatial pattern can be anisotropic where the spatial variance changes according to direction. When anisotropy is such that the sill is the same and only the range varies (i.e. geometric anisotropy), the distance matrix can be adjusted directly to account for it. When the anisotropy is 'zonal' (i.e. the range stays the same but the sill differs) then the adjustment needs to be made in the covariance matrix by adding more terms in a nested way (see Deutsch & Journel 1992). A nested procedure

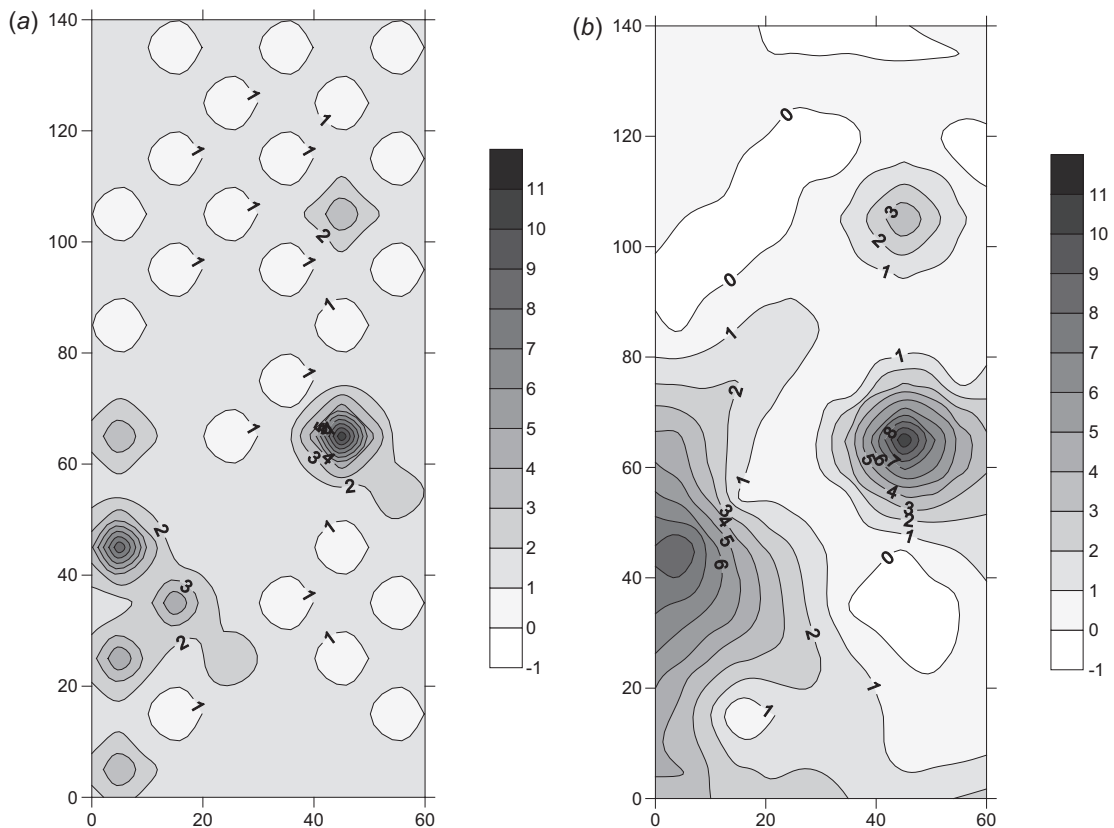


Figure 6.21 Kriged values of sassafras abundance data ($n = 84$) based on 42 sampling locations (as shown in Figure 6.13b). The estimated values are illustrated using contour lines. (a) An isotropic spherical model using a range of 10 m which is too small (compared to the fixed 32 m range of Figure 6.20) and create small patches around each sampled locations. (b) An isotropic spherical model using a range of 60 m which is too large (compared to the fixed 32 m range of Figure 6.20) creating too smoothed values between sampled locations.

can also be implemented by using different variogram models as a function of distance. Stemming from this nested property of adding more terms in the system of linear equations, *stratified* Kriging offers a way to interpolate over regions that have different spatial variances due to some change in strata types. For example, in forested landscapes, it is inappropriate to Krige deciduous and coniferous stands using the same variogram. When information about the forest-stand type is available, stratified Kriging, using different

variograms for each forest type, should be used (Wal-
lerman *et al.* 2002).

Too often in ecological and environmental studies, the variable of interest, z_1 , is too costly to sample. If, however, knowledge of how this costly variable correlates with another variable, z_2 , that can be sampled more easily (more cheaply or available from other sources), *co-Kriging* can be used to interpolate z_1 given the spatial variance of z_2 as estimated by a cross-variogram between the two variables (for

example see Hudak *et al.* 2002). This is an appealing method, but it assumes that the linear relationship between the two variables holds even at locations where z_2 is sampled but z_1 is not. Often, the resulting Kriged map of z_1 looks like a mirror image of z_2 , which may or may not reflect the real spatial pattern of z_1 .

Sometimes, in ecology, we are interested in mapping the spatial structure of an assemblage of species as a community rather than individually. This can be done using *multivariate* Kriging (see the textbook by Wackernagel (2010) on the subject). Finally, given measurement errors, thresholding responses of variables, or availability of only presence : absence data, we may want to use our quantitative variables as qualitative ones but would still like to determine their spatial structure and Krige them. In such circumstances, an indicator variogram can be estimated (as presented above) and *indicator* Kriging can be used (for examples see Todd *et al.* 2003; Polakowska *et al.* 2012).

Here we presented a succinct overview of the dynamic field of geostatistics. There are many more geostatistical methods that are used and these topics are treated in advanced geostatistical textbooks. Before concluding on the subject of geostatistics, however, there is one more aspect of geostatistics that is very useful to ecological studies and deserves to be mentioned: stochastic simulation based on conditional annealing (Deutsch & Journel 1992). Spatial stochastic simulations are used more and more to generate a series of spatial data that have a given degree of spatial dependence in order to evaluate whether or not observed sample data show significant spatial patterns (Fortin *et al.* 2003; Fortin *et al.* 2012a, b). Here, the parameters of the variogram model are derived from an experimental variogram and can be used to generate stochastic simulations having the same degree of spatial variance as the observed data. This approach was first proposed in geostatistics to generate maps having more spatial variability than the Kriged ones and hence looking more realistic in comparison to the observed map. Such simulated data are generated by an iteration

process where the values at the sampled locations are kept as anchor locations from which the annealing algorithm iteratively spreads data values around them while ensuring that the overall degree of spatial variance is maintained (i.e. the range, nugget and sill values). This spatial stochastic simulation approach, based on a theoretical variogram, is computer intensive but allows us to address significance testing of spatially autocorrelated data.

6.8 Concluding remarks

The first issue to be decided is whether the analysis of the data should be spatially global, summarizing the spatial structure for the entire study area, or local, measuring the spatial structure in neighbourhoods around each sampling location. The choice between these two levels of analysis should be guided by both the goal(s) of the study and knowledge (or lack of knowledge) about the stationarity of the processes of interest. When no prior information is available, we recommend that more than one method is performed and that their results are compared to identify whether or not the stationarity property holds.

Then, while using global spatial statistics, Moran's I is usually favoured, if only because of its direct correspondence in meaning with Pearson's linear correlation. Users should be aware of its sensitivity to extreme outliers, which influence the value of the average used to estimate the spatial deviation. This is why most would prefer Geary's c or semi-variance γ because outlier values affect only the spatial deviations (differences) computed with them. Unfortunately, these differences will have more weight in the final measure because they are squared.

Similarly, as local spatial statistics exist, it is important to realize that some are affected by the presence of a global pattern which may result in biased estimation of the spatial pattern. Two newly developed methods, H Moran and LICD, are to be favoured when a large trend in the data is present. In the absence of overall structure, local G_i^* statistics in their standardized versions are easier to interpret as indicating local areas of

high (hot spots) or low (cold spots) values. Both global and local spatial statistics provide information only about the spatial structure.

Interpolation techniques can be used for illustration purposes or for informal evaluation in mapping spatial pattern, where the simplest method (inverse weighted distance) provides good results. Such maps can also be modified by using various smoothing algorithms, such as a spline, to give a more visually pleasing or more intuitive product. On the other hand, when information about the actual values at unsampled locations, as well as an estimate of the associated errors, is needed, then one of the various Kriging techniques should be used. There is no magical recipe, however, for which methods, and their respective parameter values, should be favoured under all circumstances. The ability to perform a meaningful interpolation using Kriging comes with experience where the general rule of thumb is to capture the intensity and range of the spatial variance at short distances as the first priority. Last, keep in mind that Kriging errors associated with Kriged values are a function of the theoretical model selected and the parameter values provided, not of the data themselves. The results depend in a crucial way on the appropriateness of the model on which Kriging is based.

Chapters 5 and 6 are organized by the distinction between sample data and data from complete sets of contiguous sample units. While the distinction is clear, and has clear consequences, the question may arise of how to proceed when the data from an array of contiguous units are incomplete, whether by structural constraints or by error. We will answer this general question by a particular example and particular solutions for the problem it exemplifies, with the understanding that the solutions can be extended to greater generality. Researchers are using video recordings of transects across coral reefs to obtain species densities. The aim is to determine the density and distribution patterns of a cold-water coral *Lophelia pertusa* and of the mega-fauna associated with it, also elucidating any correlations among the species recorded. They obtained the densities of several faunal groups along seven video transects. If the data sets were uninterrupted, the transects could be divided into virtual

quadrats, giving series of counts of individual organisms, and methods such as 3TLQC could be used for the analysis. Here, because of poor visibility, detection can be impossible in some sections, and therefore the density measurements have many breaks along the transects. This problem necessitates a switch from 'population spatial statistics' to 'sampling spatial statistics' and the question then arises of how best to analyse these data to achieve the research aims.

In technical terms, the situation is one of 'censored' series data, with individual observations or strings of observations missing. Focusing on count data, the situation is different from having 'error' quadrats within well-defined patches or gaps of density, with stray 0 counts in the middle of patches of high density, or occasional occupied quadrats in otherwise consistent gaps. We know that wavelet analysis can deal quite well with this structure of errors (Dale & Mah 1998).

Error quadrats in count data

See Table 6.1(a). Here, we believe, the censored data include regions where it is known that detection is impossible, producing strings of quadrats that are not 'errors' but known blanks. In that case, the blanks may be associated with particular environmental conditions, and thus may be confounded with the organisms' distributions. The situation is rather like attempting to carry out a spatial analysis of a forested ecosystem in which much of the area is actually occupied by lakes (cf. James *et al.* 2011).

Censored data

See Table 6.1(b). If the strings of missing data are much shorter than the overall scales of pattern in the data, the usual methods of pattern analysis, especially wavelets, will probably be adequate, but if the strings of missing data are relatively long, the real pattern may be hard to detect.

Relative length of missing data

See Table 6.1(c). The long series of blanks may obliterate smaller scales of patterns or may distort the scale

Table 6.1 Dealing with incomplete transect data. The 0's are absences and the 8's represent unspecified non-zero data. The first row in (a) is without errors, and the italic font in the second row indicates the error quadrats, with presence where there should be absence and *vice versa*. In (b), we have censored data with missing observations that are known to be missing, not like the errors that are undetected, indicated by 'x'. (c) If the missing data are in short blocks, any one may be contained within a single patch or single gap; if they are long, a single block of missing observations may include parts of both patches and gaps, as shown. One approach to dealing with the missing observations is shown in (d) where the missing blocks (x's) are replaced by randomly chosen subseries of the original data, indicated by underlining. In (e), missing quantitative data (indicated by x) may be replaced by randomly chosen individual values, shown in *Italic*, rather than subseries of original data.

(a) *Error Quadrats in Count Data*

Without: 0000000088888888880000000000888888888800000000000
 With: 000800000888888808088000080000808888088888800008000080

(b) *Censored Data*

00000xxx0888888888xxx00000008888x8x8x8880000xx000xxxx

(c) *Relative Length of Missing Data*

Short blocks: 00000xxx0888x8888xxx0x0000008888x8x8x880000xx000xx8x
 Long blocks: 00008888000088880000888800008888000088880000888800008888
 0000888xxxxxx88000088880000888800xxxxxx0008888000xxxxxx

(d) Replace missing blocks with randomly chosen subseries (underlined)

0000888xxxxxx888000088880000888800xxxxxxx0008888000xxxxxxx
 00008888880000888000088880000888800008888000088880000888800

(e) For quantitative data replace missing blocks with randomly chosen observations

0131868xxxxxx86833101688630036883100xxxxxxx0036886100xxxxxx
 01318680186301086833101688630036883100013686110036886100866083


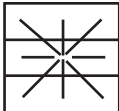

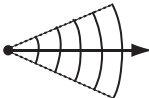
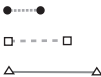


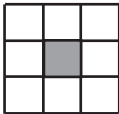
detected for the larger ones. One obvious step in an exploratory analysis would be to do a spatial analysis of the data blanks themselves, as if they were data, in order to determine the distribution and scales of these sections of turbidity.

Given the concerns about missing data, standard analyses can proceed provided they use only those locations where there is some confidence in the counts recorded, and that any zeros are real absences and not artifacts of poor visibility. For a single variable (counts of any one species), acceptable methods would include Moran's *I* to detect spatial autocorrelation at different scales. Because the quadrats are virtual, being imposed on a continuous record of data, researchers need to be aware of the potential effects of the choice

of quadrat size on the spatial structure detected in this way (see Fortin & Dale 2005, section 3.3.4). To examine the relationship between two kinds of organisms in the data, the cross-correlation approach can be used (Fortin & Dale 2005, section 3.3.5), whether based on Moran's *I* or on another autocorrelation statistic.

A second approach to understanding the spatial structure of the data. See Table 6.1(c) is to use methods designed for spatially complete data sets, replacing the censored section of the data series for any single species with randomly chosen sub-series of uninterrupted data or with random observations drawn from the frequency distribution of the observed values, especially where the data take several values. In either case, a large number of realizations should be produced and

Table 6.2 Summary of the spatial analysis methods presented in Chapter 6

Spatial analysis method	Template
Join count statistics (Topology: network)	
Join count statistics (Lattice: chess moves)	
Global spatial statistics (Isotropic spatial lag)	
Global spatial statistics (Anisotropic spatial lag)	
Fractal dimension (Dividers)	
Fractal dimension (Boxes)	
Local spatial statistics (Topology: network)	
Local spatial statistics (Lattice: chess moves)	

analysed, for example with wavelet analysis, and the common features of these analyses determined. To examine the correlation or co-variation of different species, wavelet covariance analysis (Fortin & Dale 2005, section 2.6) can be used with same data ‘fill’ technique to determine scales of positive and negative correlation between species.

The conclusion to be drawn from our discussion of this example is that the censored nature of the data presents a problem, certainly, but it is not insurmountable. Both standard methods for sample data, and methods designed for spatially complete data sets assisted by numerical approaches can be used to provide useful insights.

Turning back now to specific sample data, as sketched in this chapter, several spatial statistics can estimate spatial dependence for such data (Table 6.2). They share a common root in the determination of spatial covariance among the values of variable(s) of interest at different sampling locations. Hence the question becomes: which one is best for the purpose?

Finally, the unresolved issue is that of significance testing while estimating spatial dependence at several distances based on the same data. Although spatial statistics traditionally have significance tests, and even progressive Bonferroni corrections, applied to them, these do not fully account for the dependence of values from one distance class to another. We will revisit this problem in Chapter 8.