Landscape Ecology **13:** 15–25, 1998.

Multiscale sources of variation in ecological variables: modeling spatial dispersion, elaborating sampling designs

Claude Bellehumeur and Pierre Legendre
Département de sciences biologiques, Université de Montréal, C.P. 6128, succursale Centre-ville, Montréal (Québec) H3C 3J7, Canada

(Received 17 April 1997; accepted 8 August 1997)

Keywords: spatial pattern, variogram, correlogram, fractal, sampling design, analysis of variance

Abstract

Detection of structured spatial variation and identification of spatial scales are important aspects of ecological studies. Spatial structures can correspond to physical features of the environment or to intrinsic characteristics of ecological processes and phenomena. Spatial variability has been approached through several techniques such as classical analysis of variance, or the calculation of fractal dimensions, correlograms or variograms. Under certain assumptions, these techniques are all closely related to one another and represent equivalent tools to characterize spatial structures.

Our perception of ecological variables and processes depends on the scale at which variables are measured. We propose simple nested sampling designs enabling the detection of a wide range of spatial structures that show the relationships among nested spatial scales. When it is known that the phenomenon under study is structured as a nested series of spatial scales, this provides useful information to estimate suitable sampling intervals, which are essential to establish the relationships between spatial patterns and ecological phenomena. The use of nested sampling designs helps in choosing the most suitable solutions to reduce the amount of random variation resulting from a survey. These designs are obtained by increasing the sampling intensity to detect a wider spectrum of frequencies, or by revisiting the sampling technique to select more representative sampling units.

Introduction

Biological and ecological systems are maintained by a network of dependencies among physical, geological, environmental, and biotic processes of various origins. These natural processes, which operate simultaneously and interact at multiple scales, correspond to spatial (and/or temporal) patterns and structures that can be observed over a continuum of scales. When it is known that the phenomenon under study is structured as a nested series of spatial scales, this helps us understand and explain the mechanisms producing these patterns.

In a general review dealing with problems of patterns and scales in ecology, Levin (1992) mentioned that a fundamental problem of ecology is to relate broad-scale phenomena (such as climatic, geological, hydrological, etc.) to processes at the scale of ecological information, and to understand how information is transferred across scales. The scales involved in a process represent an important aspect to consider in order to understand ecological phenomena. Our perception depends on three main components of the sampling design:

1) The size of the sampling unit, which is the surface or volume support of any particular measurement: size of the vegetation quadrat, volume of the bucket of water in which a measurement is made, etc. Large sampling units filter out spatial variation occurring at scales finer than their size. On the other hand, small sampling units are less representative of the sampling stations; they introduce noise in the signal, and produce a larger estimated variance for the variable of interest (He et al. 1994; Bellehumeur et al. 1997).

2) The extent of the total area being sampled, or field size (length of the series, in time series analysis) determines the broadest spatial process that can be

detected (He et al. 1994). It also influences the relative importance of each of the various spatial scales detected. For example, climatic phenomena can contribute strongly to shape dispersion patterns in vegetation surveys covering very large areas, whereas biological phenomena become more important in surveys of small extent.

3) The sampling interval (lag, in time series) is the average distance between neighboring sampling units.

Several techniques have been developed for the description of spatial patterns of populations and ecosystems (Legendre and Fortin 1989; Levin 1992). The first studies concerned with spatial heterogeneity and spatial variability used classical analysis of variance to measure the relative importance of variance components corresponding to distances among sampling stations and to test the hypothesis that a phenomenon exhibits a spatially random pattern (Greig-Smith 1952; Platt and Filion 1973; Lewis 1978; Ludwig and Goodall 1978). Using techniques previously developed in geography, Sokal and Oden (1978a and 1978b) introduced, in population studies, the use of spatial correlograms based upon Moran's I and Geary's c coefficients. Several studies have also measured the Fourier spectrum of frequencies to characterize spatial variance components (Renshaw and Ford 1984). Burrough (1987), Legendre and Fortin (1989) and Rossi et al. (1992) introduced geostatistical concepts in ecology and used variograms as a tool for spatial characterization. Most applications of geostatistics have dealt with ore reserve estimation in the mining industry (David 1977; Journel and Huijbregts 1978), but these statistical techniques are now widely used in several other disciplines (Cressie 1991). The fractal dimension D is another commonly-used measure to study features of spatial dispersion (Milne 1991; Palmer 1988; Burrough 1981).

The purpose of this paper is to show that under certain assumptions, these techniques and concepts are all closely related and constitute equivalent tools to characterize spatial structures. We will suggest simple sampling designs enabling the detection of a wide range of spatial structures and showing the relationships between nested spatial scales.

As test cases, we will study: (1) the density of a tree species, *Macaranga lowii* (Euphorbiaceae), in 10 m $\times$ 10 m quadrats in a tropical rain forest plot of Malaysia; (2) counts of the bivalve mollusks *Macomona liliana* and *Austrovenus stutchburyi* in 10-cm-diameter sediment cores collected along six 150-m-long transects on the sandflat of Wiroa Island in the Manukau Harbour, North Island of New Zealand.

Materials

A tract of mapped forest, located at 102°18′ W and 2°55′ N, was established in the Pasoh Reserve, Negeri Sembilan, Malaysia, to monitor long-term changes in a primary forest (Kochummen et al. 1991). The forest tract under study is a rectangle 1 km long in the east-west direction and 0.5 km north-south, for a total of 50 ha. The survey enumerated all free-standing trees and shrubs at least 1 cm in diameter at breast height, positioning each one by geographic coordinates on a reference map, and identifying the species. We reorganized the data set in quadrat units and calculated tree densities (number of trees per square meter in each quadrat) corresponding to square quadrats of 100 m^2 (10 m $\times$ 10 m quadrat sizes). This data set contains exhaustive information about tree locations; so it allowed us to simulate various sampling designs. These data have been analyzed in more detail by He et al. (1994, 1996, 1997) and by Bellehumeur et al. (1997).

In December 1993, three sites were selected on the Wiroa Island sandflat, on the northeastern side of Manukau Harbour, within the security zone of the Auckland Airport, New Zealand. This sandflat has an extensive mid-tide area 1 to 2 km wide; sediment is composed of well-sorted fine sand and is bivalve dominated. At each site, two directions were selected, one parallel and the other perpendicular to the main hydrological gradient (tide). Transects 150 m long were established in each of these two directions, crossing at their mid-point. Along each transect, 61 sediment cores, 10 cm in diameter, were obtained following two sampling intervals: (1) every five meters, that is, at positions 0 m, 5 m, . . . , 145 m, 150 m along the transect; and (2) at one meter after each of previous sampling units, that is, at positions 1 m, 6 m, . . . , 146 m along the transect. The sediment was sieved and the bivalve mollusks found therein were counted (specimens larger than 4 mm of species *Macomona liliana* and *Austrovenus stutchburyi*). To study the spatial dispersion of these mollusks, the three transects parallel to the shoreline on the one hand, and the three perpendicular transects on the other (tide movement direction), were combined, making up two distinct data sets representing the main hydrological directions. These data have been analyzed in more detail by Hewitt et al. (1997).

Measures of spatial dispersion and spatial heterogeneity

Ecologists conducting surveys in geographic space are often interested in measuring the spatial heterogeneity of the variables of interest (Dutilleul and Legendre 1993). Spatial heterogeneity indicates spatially structured variation, whereby we can expect zones of aggregation of high values and other zones of low values. Heterogeneity contrasts with the concept of homogeneity, which indicates that there is no local differentiation, that zones of low and high abundance cannot be identified, and that there is no tendency for neighboring areas to be more similar than distant areas. Notice that homogeneity does not mean 'no variation'. Important variation is possible, but no spatial organization of values can be recognized.

As pointed out by Dutilleul (1993), while nature appears clearly heterogeneous, the scales at which spatial heterogeneity is identified vary widely. As mentioned above, our perception of spatial scales depends on the sampling design (i.e., the size of the sampling units, the extent of the study area, and the sampling interval). Even if the true spatial distribution can be called heterogeneous, the information available may look like noise on a map, leading to the perception of a homogeneous phenomenon. The search for structures revealing spatial heterogeneity has led to the development of several mathematical techniques and sampling designs.

Classical analysis of variance

Surveys can be hierarchically organized, suggesting the application of classical analysis of variance using a nested (or hierarchic) design to estimate the components of variance associated with different spatial scales. Several scientific disciplines such as aquatic biology (Platt and Filion 1973; Lewis 1978; Pinel-Alloul et al. 1988; Troussellier et al. 1989), vegetation science (Greig-Smith 1952; Ludwig and Goodall 1978), geology (Garrett 1983), and pedology (Nortcliff 1978; Oliver and Webster 1986), have developed and used this type of approach. Nested surveys are constructed in such a way that the total population is subdivided into primary sampling units (lots, sites, or stations), which in turn are subdivided into secondary sampling units in which several determinations are made, representing the tertiary sampling units, and so on. Generally, the stations (primary sampling units) are separated by a distance chosen to represent a broad spatial scale of variation of the study variable; this distance, and the variation expected from it in the response variable, are assumed to be known *a priori*. The secondary level could represent a collection of replicate sampling units at the same location, to verify the representativeness of these sampling units, and the third level could represent replications of measurements to verify the precision of measurement devices.

The underlying model of nested variation is based on the notion that the spatial components of the population of interest are divided into distinct organization levels (Sokal and Rohlf, 1995, chap. 10; Oliver and Webster, 1986). Observations are viewed as the result of the nested contributions of the various levels. For *k* levels (random factors), the nested analysis of variance model is:

$$Z_{ij\ldots\,k} = \mu + A_i + B_{ij} + \ldots + \epsilon_{ij\ldots k} \qquad (1)$$

where $Z_{ij\ldots\,k}$ is the observed value of the *k*th unit in the *j*th class of level 2 and in the *i*th class of level 1. The observed value $Z_{ij\ldots\,k}$ is the result of deviations from a general population mean μ, by amounts related to the spatial components of variability *A*, *B* and ϵ. A_i is the deviation (random effect) between μ and class *i*, representing values separated, for example, by the inter-site distances; B_{ij} is the deviation (random effect again) from the mean μ_i of the mean of the *j*th subclass within the previous class of higher level. The quantity $\epsilon_{ij\ldots\,k}$, which is a random error term, represents the deviation of the observed value from its class mean at the last level of subdivision. The model assumes that *A*, *B* and ϵ are normally distributed, independent and unbiased with means zero and variances $\sigma^2_A,\ \sigma^2_{B\in A}, \ldots\ \sigma^2_\epsilon$. The symbol $\sigma^2_{B\in A}$ indicates that the variance is of level *B* within level *A*. The total variance is viewed as containing the components:

$$\sigma^2_Z = \sigma^2_A + \sigma^2_{B\in A} + \ldots + \sigma^2_\epsilon \qquad (2)$$

Traditionally, nested sampling schemes have been balanced, which means that the number of sampling units in each group of any level of the hierarchy is the same, although it may differ from level to level. Sample size at least doubles for each additional level. If the survey aims at characterizing the spatial dispersion over a wide range of spatial scales, the number of sampling units would become very large. In order to reduce the number of sampling units without an important loss of precision, Gower (1962)

introduced the notion of unbalanced sampling design, where economy can be achieved by replicating at only a portion of the sampling sites, at one or more levels (Oliver and Webster, 1986). This approach has been widely used in geochemical surveys (Garrett, 1983), in pedology (Nortcliff 1978) and in ecology (Shaw and Mitchell-Olds 1993). However, even with an unbalanced design, it remains difficult to design a survey exceeding three or four levels for the regional characterization of a phenomenon. For example, Nortcliff (1978) used a four-level sampling design characterizing spatial scales from 5 m to 500 m. The third and fourth levels (scales from 200 to 500 m) were defined by only four pairs of points which do not cover exhaustively the study area. So, generally the efficiency of the technique relies on an *a priori* knowledge of the spatial scales of the phenomenon to be studied. To reveal a wider range of spatial scales, surveys can be designed to allow the computation of spatial (or temporal) structure functions, such as variograms and autocorrelograms, where a sequence of data is compared to itself for various lags (geographic distances or time classes).

Geostatistics and autocorrelation structure

Students of spatial structures belonging to several disciplines are using correlograms based upon Moran's *I* or Geary's *c* coefficients of spatial autocorrelation (Moran 1950; Geary 1954; Cliff and Ord 1981). On the other hand, the French school of geostatistics, led by Matheron (1965), developed the theory of regionalized variables which provides a means for analyzing the spatial variability of a variable distributed across a geographic area. A regionalized variable is an attribute that takes a value at every point of the study area. These values are considered to be realizations of a random function whose spatial distribution is characterized by the variogram. The variogram reveals the randomness (irregularity of values from location to location) and the structured aspects of the spatial dispersion. The empirical (semi-) variogram is defined as:

$$\gamma^*(\mathbf{h}) = (2N(\mathbf{h}))^{-1}\Sigma[\mathrm{z}(x) - \mathrm{z}(x+\mathbf{h})]^2 \quad (3)$$

where z(*x*) and z(*x*+**h**) are measurements of a given random variable at locations *x* and *x*+**h**, separated by the directional distance **h**, and *N*(**h**) is the number of pairs of sampling units considered in the given distance class. This calculation is repeated for different values of **h** and provides the empirical variogram which is a plot of the empirical values of variance γ^*(**h**) as a function of distance **h**. The rate of increase of γ^*(**h**) allows one to characterize the continuity of the variable. Generally, the variogram tends to level off at a *sill* equal to the variance of the variable. The distance at which this occurs is referred to as the *range* of the variable. The range is the distance over which the sampling units are not spatially correlated any longer. The discontinuity at the origin (non-zero intercept) is called the *nugget effect*. It corresponds to the local variation occurring at scales finer than the sampling interval, including sampling error.

A variogram function is a theoretical model chosen to represent the spatial structure of a phenomenon and to perform estimations of unknown values. This inference is valid only if some hypothesis of stationarity can be fulfilled, representing conditions of application of the model. If the population mean and the spatial covariance are assumed to be constant over the study area, the variable under study is said to be second-order stationary (Journel and Huijbregts 1978, pp. 32–34). In other words: (1) the first moment (mathematical expectation) E[Z(*x*)] = *m* exists and does not depend on the sampling point locations; it is constant and finite for all *x* inside the study area; and (2) for each pair of random variables Z(*x*) and Z(*x*+**h**), the second moment (spatial covariance):

$$\mathrm{C}(\mathbf{h}) = \mathrm{E}[\mathrm{Z}(x+\mathbf{h}) \cdot \mathrm{Z}(x)] - m^2 \quad (4)$$

exists and depends only on **h** and on the orientation of the distance vectors, but not on their positions in the study area.

Stationarity of the covariance implies stationarity of the variance and of the variogram, so that:

$$\mathrm{Var}[\mathrm{Z}(x)] = \mathrm{E}[\mathrm{Z}(x) - m]^2 = \mathrm{C}(0)$$
$$\gamma(\mathbf{h}) = 1/2\mathrm{E}[(\mathrm{Z}(x+\mathbf{h}) - \mathrm{Z}(x))^2] = \mathrm{C}(0) - \mathrm{C}(\mathbf{h})$$

are the same anywhere on the surface ('Var' is the variance of classical statistics). Under the hypothesis of second-order stationarity, the variogram and the covariance are related and can define the correlogram ρ(**h**) (Journel and Huijbregts 1978; Rossi et al. 1992):

$$\begin{aligned} \gamma(\mathbf{h}) &= \mathrm{C}(0) - \mathrm{C}(\mathbf{h}) = \sigma^2 - \mathrm{C}(\mathbf{h}) \\ \rho(\mathbf{h}) &= 1 - (\gamma(\mathbf{h})/\sigma^2) \end{aligned} \quad (5)$$

So, ρ(**h**) = 1 – Geary's *c* coefficient of spatial autocorrelation (Geary 1954). The variogram, covariance, Geary's *c* coefficient and correlogram represent equivalent tools to characterize spatial structures under the hypothesis of second-order stationarity.

The presence of a sill implies stationarity of the covariance, i.e., the covariance exists and depends only on vector **h** (second-order stationarity hypothesis). Models without a sill correspond to random functions that are said to be intrinsic only. Their *a priori* variance and covariance are not defined. The increments (z(x) – z(x+**h**)) have a finite variance which does not depend on locations but only on **h**. If only the intrinsic hypothesis is satisfied, the variance is an increasing function of the size of the study area and the variogram is a monotonically increasing function of **h**. The intrinsic hypothesis is less restrictive than the second-order stationarity conditions; on the other hand, this hypothesis is always satisfied if the second-order stationarity conditions are satisfied. Even if interpolation or estimation are not the main purposes of a study, the shape of the variogram characterizes the spatial dispersion of a phenomenon and allows one to assess the appropriateness of the stationarity hypothesis. Models without sills may correspond to fractal dispersion (next section).

Fractal dimension and variogram

The fractal dimension of transects and surfaces can be related to the variogram (Bolviken et al. 1992; Carr and Benzer 1991). For a fractal transect, the variogram follows the equation (Mandelbrot 1983, p. 353):

$$\gamma(\mathbf{h}) = K\ \mathbf{h}^{2H} \qquad (6)$$

corresponding to a power model without a sill. The fractal dimension of a transect is given by:

$$D = 2 - H \qquad (7)$$

(that of a surface would be given by D = 3–H). This result allows one to calculate the fractal dimension D for a real data set from the log-log plot of the variogram:

$$log(\gamma(\mathbf{h})) = \alpha + \beta \log(\mathbf{h}) \qquad (8)$$

The slope β (= 2H) of the equation is then equal to 4 - 2D. So, a fractal transect represents a spatial dispersion characterized and modeled by a power model variogram (eq. 6). For a transect, if a variable has its values randomly distributed in space (no autocorrelation), D equals 2. Burrough (1981) and Palmer (1988) have calculated fractal dimensions from empirical variograms for transects in one dimension. He et al. (1994) have calculated the fractal dimensions of tree density, species richness and Shannon diversity in a tropical forest plot (i.e. a surface).

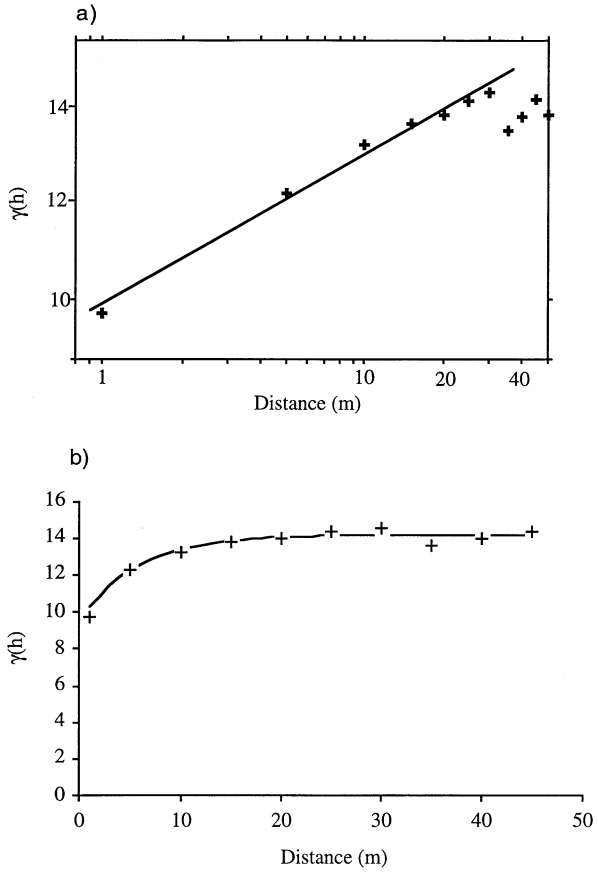

Figure 1. Empirical variogram values for the density of *Macomona liliana* along the transects perpendicular to the shoreline of the Wiroa Island sandflat: a) variogram values plotted as a log-log graph; the power function logγ(**h**) = α + βlog(**h**) is fitted to the variogram between 1 m and 20 m (β = 0.12). b) Empirical variogram values and exponential variogram model.

The variogram of the density of *Macomona liliana* along the transects perpendicular to the shoreline of the Wiroa sandflat is given as a log-log plot of γ(**h**) *versus* **h** (Fig. 1a). A straight line can be fitted to the empirical values between 1 m and 20 m, with a slope corresponding to a fractal dimension D = 1.94. This high value, close to 2, indicates a rugged surface, with variation mostly occurring at fine scale.

Computation of the fractal dimension corresponds to the fit of a variogram power model to the spatial dispersion of the variable of interest. The power model, which describes a self-similar fractal, corresponds to a phenomenon with an unlimited capacity for spatial dispersion and with an undefined *a priori* variance. So

the fractal nature or the self-similarity of the phenomenon is only defined locally, near the origin where the variogram is linear in the log-log plot. It is bounded by a lower and an upper scale of self-similarity which are 1 and 20 m in the present case.

Figure 1b shows that an exponential model can also be fitted to the empirical variogram values of the density of *Macomona liliana* along the transects perpendicular to the shoreline, between 1 m and 45 m. The exponential model has a sill and a defined and finite *a priori* variance. The model has a range of 17 m and a sill of 14.2. The nugget effect (random variation) represents 66.9% of the spatial variation (= 9.5/14.2). The structured spatial variation accounts for 33.1% [= (14.2–9.5)/14.2] of the total spatial variation. This example shows that both the exponential and fractal models can describe the same reality. From a methodological point of view, examination of the empirical variogram is, however, much more informative than the simple calculation of a fractal dimension.

Analysis of variance and variogram

Greig-Smith (1952) and Ludwig and Goodall (1978) proposed the paired-quadrat variance method which calculates the variance between quadrats at particular spacings. These authors pointed out that the variance estimates are not derived from independent data, so the usual tests of significance cannot be applied. But the technique can be useful for descriptive purposes and to calculate variance components corresponding to particular spatial scales. The paired-quadrat variance estimate of Ludwig and Goodall (1978) exactly corresponds to the variogram (equation 3).

The link between a variogram and the classical analysis of variance applied to nested surveys was pointed out by Miesch (1975) and Ver Hoef et al. (1993). Miesch (1975) used the additive scheme of equation (2), where a component $\sigma^2_{\mathbf{h}}$ is a measure of the variance at scales finer than **h**, and a component $\sigma^2_{2\mathbf{h}}$ is a measure of the variance at scales between **h** and 2**h**. On the other hand, variogram values measured at distances **h** and 2**h** represent all the variance at scales finer than **h** and 2**h** respectively. Considering the *k*th level of a nested design, the relationship between the variogram and the variance components is given by:

$$\gamma(k \cdot \mathbf{h}) = \sigma^2_{\mathbf{h}} + \sigma^2_{2\mathbf{h}} + \ldots + \sigma^2_{k\mathbf{h}} \qquad (9)$$

In the field of vegetation science, grids of continuous quadrats have been studied by agglomeration, whereby groups of adjacent quadrats are combined hierarchically into blocks of different sizes (Greig-Smith 1952). This agglomerative technique uses the formalism of nested ANOVA. Sums of squared differences between adjacent blocks (for blocks containing *r* quadrats) are calculated and plotted against block size. A peak in this plot characterizes the patch size and the scale of pattern. Ver Hoef et al. (1993) have demonstrated that the method corresponds to the first lag of a variogram calculated for a given block size. Gardner (1997) pointed out that the main problems with this agglomerative approach are the inefficiency of the sampling technique and the limited extent of scales that can be detected by this method.

Test case 1: Nested spatial structures of tree density in the Pasoh forest

Figure 2 shows the empirical variograms for the north-south and east-west directions and the fitted model describing tree density for the species *Macaranga lowii* measured in 10 m × 10 m quadrats in the Pasoh forest. These empirical variograms show abrupt changes in their slopes at a distance of approximately 50 m, revealing the existence of more than one scale of variation. The empirical variogram exhibits a superposition of different scales of variation, called "intermeshed" or "nested structures", each scale adding its own contribution. The resulting model can be adequately represented by combining simple variogram models with different ranges and with their own structured variance components, each one characterizing the variability at a particular scale.

The variogram model of tree density in the east-west direction has three components indicated by the discontinuities in the shape of the empirical variogram. Its parameters are given in Table 1. The global model reaches a sill at a distance of approximately 500 m. The variance of the data is equal to this sill value (1.46). The nugget effect (model 0 in Table 1), corresponding to the amount of fine-scale random variation in the data, represents more than 37% of the spatial variation (= 0.55/1.46). The nugget effect reflects micro-structures at scales finer than the sampling interval and an integration of the measurement errors.

The structured spatial variation exhibits two structures corresponding to a pronounced increase at small

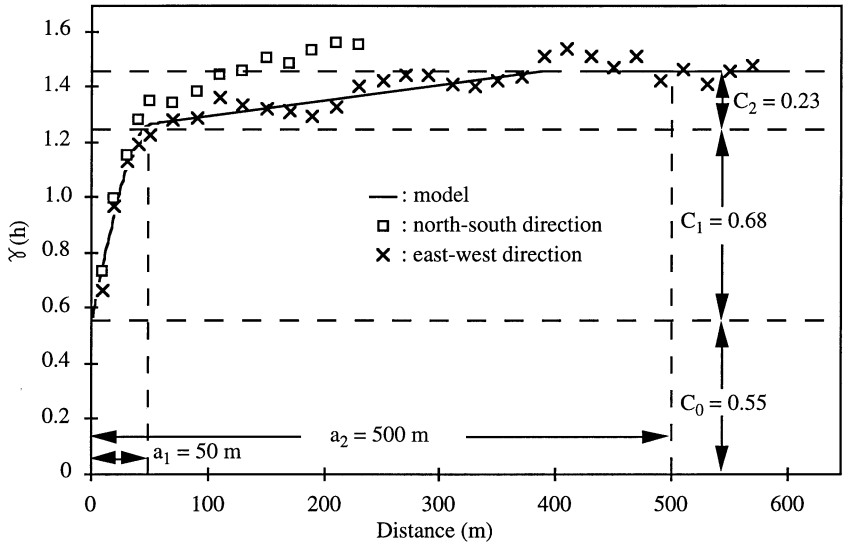

Figure 2. Empirical directional variograms (north-south and east-west) for density values of the tree species *Macaranga lowii* in 10 m × 10 m quadrats, with a 10 m sampling interval; a three-component model is fitted to the east-west variogram (Tab. 1).

Table 1. Parameters of nested variogram models for tree density (east-west direction) in 10 m × 10 m quadrats (visually adjusted parameters). C_0 is the nugget effect, C_n is the variance component associated to a given structured spatial scale and a_n is the practical range of the model.

Model	Parameters	
0. Nugget effect	C_0 = 0.55	
1. Spherical	C_1 = 0.68	a_1 = 50 m
2. Spherical	C_2 = 0.23	a_2 = 500 m

lag distances (0–50 m) and a long-range structure (0–500 m). The global variogram is modeled by two spherical models (models 1 and 2 in Table 1). The most important part of the structured spatial variation is related to a local scale where the zone of influence (range) is approximately 50 m. This structure accounts for 47% (= 0.68/1.46) of the total spatial variation. The long-range component is 500 m long and it accounts for 16% of the total spatial variation. It would be possible to add a periodic component modeled by a function such as:

$$\gamma(\mathbf{h}) = C[1 - \cos(2\pi \cdot \mathbf{h}/a)] \qquad (10)$$

where *a* is the period of the model. However, as indicated by the amplitude of cycles, this structure would account for less than 5% of the spatial variance and it is difficult to differentiate it from random sampling fluctuations in the variogram computation.

In the north-south direction, the short-range structure is recognized, but the long-range structure reaches higher variogram values than in the east-west direction, corresponding to a slight zonal anisotropy. This type of pattern expresses different degrees of overall spatial variability depending on the direction.

The nested spatial structures can indicate environmental, biological, or behavioral influences that are shaping the spatial patterns of tree density. Environmental, biological, or other variables are needed to relate ecological phenomena to these spatial scales of variation; see for instance Borcard and Legendre (1994) for a method to investigate this question. The spatial pattern features indicate that for this tract of tropical rain forest, 50 ha in area, an important proportion of the spatial variation of tree density occurs at scales finer than 50 m. So, the spatial analysis provides the scale of investigation to understand the underlying phenomena producing the spatial pattern of tree species. In the next section, we will suggest simple sampling designs enabling the detection of a wide range of spatial structures and showing the relationships among nested spatial scales.

Test case 2: Detection of spatial scales, and design of a mollusk sampling program

Variograms provide methodological information that may be used at an early stage in ecological surveys. They enable the estimation of suitable sampling intervals for detecting specific components with particular spatial continuities. The sampling interval has a determining influence on the detection of specific structures and phenomena with a particular continuity. In the vegetation data of the previous test case, for example, a sampling interval of 50 m would not have enabled the detection of the fine-scale structures which have been revealed by a sampling interval of 10 m (compare Fig. 3 to Fig. 2). Using a sampling interval of 50 m, the fine-scale structure (range = 50 m) appears as a pure nugget effect (Fig. 3). Thus such sampling schemes cannot allow for the distinction between fine (range = 50 m) and broad (range = 500 m) spatial scales. A pure nugget effect represents the perception of a high degree of homogeneity of the variable of interest: no aggregation, no cycle, no spatial trend, etc. There is no local differentiation and zones of low and high abundances cannot be identified. As in classical statistical theory, at every location in the study area, the best estimator of the abundance is the mean abundance. It should be noted that the most important part of the spatial variation occurs at scales finer than 50 m. To understand this spatial variation, the sampling scheme must allow the identification of local biological and environmental factors controlling the spatial distribution of trees.

Our perception of the random component is closely related to the scale of observation. The random component represents variation occurring at scales finer than the sampling interval and includes: (1) the variance due to high frequencies, which cannot be detected using the sampling interval that has been used in the study (or in the variogram calculation); and (2) the variance due to the survey method itself (counting, sampling unit processing, etc.).

The detection of fine spatial scales and the reduction of the random component may be achieved in two ways: (1) by increasing the sampling density in order to detect a wider spectrum of frequencies; or (2) by revisiting the sampling technique in order to collect observations more representative of each sampling location. The first method should always be considered in designing pilot studies, prior to a large survey effort. It will be effective only if a part of the random component is due to phenomena with fine spatial structures, which should be detected by a denser sampling grid. If the sampling units are collected at regular intervals, it is good practice to collect at least some sampling units at a smaller interval, which will allow the detection of nested scales if present. The basic sampling interval for the density measurements of the bivalve *Austrovenus* along the six transects of the Wiroa Island sandflat survey was 5 m, but additional sampling units have been collected 1 m from this first set of sampling units. The variogram shows that a fine-scale variance component is indeed present in the direction perpendicular to the shoreline (Fig. 4). This component could not have been detected with the sampling interval of 5 m only. The 1 m sampling interval allows the assignment of approximately one third of the spatial variance to a fine spatial scale of less than 5 m. So, to detect the main components of the spatially-structured variability, a broad-scale sampling program should include a small sampling interval. In the direction parallel to the shoreline, no spatially-structured component is detected by the sampling design used in this study; the variogram only displays a nugget effect (Fig. 4). Likewise, Oliver and Webster (1986) and Fortin et al. (1989) have recommended the use of nested sampling designs and analysis of variance, in pilot studies, to avoid the problem of failing to detect phenomena occurring at fine spatial scales.

For studies dealing with variables where measurement error is anticipated, or where representative sampling units are difficult to collect, taking replicate sampling units may aid in determining if it is worth the effort to conduct an additional sampling campaign to detect finer-scale phenomena. The goal is to verify whether the residual variance (nugget effect) is larger than the variance due to analysis and sampling units processing. The combined sampling and analytical variance is calculated as follows:

$$s^2_{sa} = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{1}{n_i}\sum_{j=1}^{n_i}(x_{ij} - \overline{x}_i)^2\right] \qquad (11)$$

where N is the number of sites where replicates were collected, n_i is the number of replicates at each site i, $\overline{x}_i$ is the average value of replicates at site i and x_{ij} is the value of a replicate at site i. Obtaining a value of $s^2{}_{sa}$ smaller than the nugget effect allows the detection of phenomena whose extent is finer than the sampling interval. On the other hand, equality of these variances signifies that taking additional sampling units on a denser grid would not allow new high-frequency

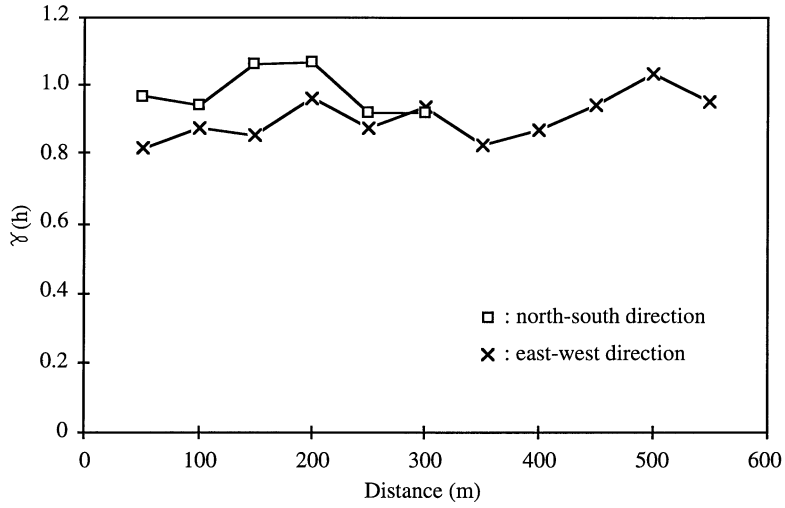

Figure 3. Empirical directional variograms of density values of the tree species *Macaranga lowii* for 10 m × 10 m quadrats, with a 50 m sampling interval.

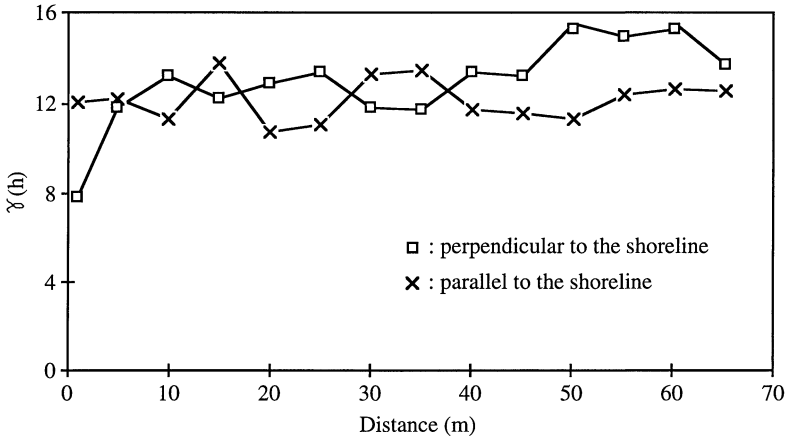

Figure 4. Empirical variograms of the abundance of the bivalve *Austrovenus stutchburyi* along transects parallel and perpendicular to the shoreline of the Wiroa Island sandflat.

variation to be detected; the variation due to analysis and sampling is masking these frequencies. In such a case, it is preferable to improve the processing of sampling units or the analytical techniques, rather than to collect additional sampling units on a finer grid. This approach requires the collection of replicates at some sampling sites of the survey area (at least 30 to 50). We do not have such a data set to illustrate the method; we have merely suggested this simple technique to verify the occurrence of variation at short spatial scales. Classical analysis of variance is probably inappropriate to compare s^2_{sa} to C_0 because the sampling units are spatially autocorrelated, but this technique allows nevertheless a useful evaluation of the corresponding variance ratio.

Discussion

Many environmental variables can be described by a spatial model that presents both a structured and a random component. This description of spatial heterogeneity contrasts with more traditional ways of considering spatial processes, where a homogeneous dispersion of variables in space is assumed. The hypothesis of homogeneous dispersion implies the independence

of sampling units and does not allow for the search of local differentiation.

Several techniques have been used to detect spatially-structured variation. The classical analysis of variance, coupled with nested survey designs, is historically one of the first approaches used to detect spatial heterogeneity. The total variance is partitioned into variance components associated with distance classes. This approach, very useful for small surveys and preliminary investigations, is difficult to implement effectively in large surveys and for more than three or four levels of replication. Therefore, we generally have few spatial variance measurements and the efficiency of the technique depends on the investigator's *a priori* knowledge of the spatial scales of variation of the phenomenon. To reveal a wider range of spatial scales, surveys can be designed to use similarity association measures (variograms, autocorrelation functions) where a data sequence is compared to itself for various lags (geographic distances, or time distance classes). These techniques are easy to adapt to regular or stratified random sampling designs. Using such regular sampling schemes, we have shown that it is useful to collect sampling units at two different scales of observations, making possible the detection of nested structures and providing solutions to reduce the amount of random variation in the survey.

It has been shown that under the hypothesis of second-order stationarity, the variogram, spatial covariance and correlogram represent equivalent tools to characterize spatial structures. Moreover, variogram and classical analysis of variance are closely linked, since the cumulative variance components of a nested scheme form the variogram. The variogram reflects the variance components acting at different scales, enabling the extent of the zone of influence of a phenomenon to be characterized and, therefore, providing useful information about the nature of the phenomenon. The recognition of scales of variation may have an impact on the sampling methodology used in surveys, suggesting that sampling unit spacing should be chosen according to the extent of the component of interest. Variograms also enable the proportion of the structured spatial variance resulting from a specific sampling interval to be estimated.

Acknowledgments

The broad-scale forest plot at the Pasoh Forest Reserve is an ongoing project of the Malaysian Government, directed by the Forest Research Institute of Malaysia through its Director-General, Datuk Dr. Salleh Nor, under the leadership of N. Manokaran, Peter S. Ashton, and Stephen P. Hubbell. Mr. K. Kochummen supervised the initial identification of trees while a Senior Fellow at the Smithsonian Tropical Research Institute. Supplemental funds from the following sources are gratefully acknowledged: National Science Foundation (USA) BSR Grant No. INT-84-12201 to Harvard University through Peter S. Ashton and Stephen P. Hubbell; Conservation, Food and Health Foundation, Inc. (USA); United Nations, through the Man and the Biosphere program, UNESCO-MAB grants nos. 217.651.5, 217.652.5, 243.027.6, 213.164.4, and also UNESCO-ROSTSEA grant no. 243.170.6; NSERC grant No. OGP7738 to P. Legendre.

The December 1993 benthos sampling on the Wiroa Island sandflat, New Zealand, has been conducted, under the direction of Drs. Rick D. Pridmore and Simon F. Thrush, by a scientific team belonging to the Ecosystems division of the National Institute of Water and Atmospheric Research (NIWA) of New Zealand, prior to an international workshop held and financed by that Institute in January and February 1994. Travel support to P. Legendre, as well as access to this data base, are gratefully acknowledged.

References

Bellehumeur, C., P. Legendre and D. Marcotte. 1997. Variance and spatial scales in a tropical rain forest: changing the size of sampling units. Plant Ecology 130: 89–98.

Bolviken, B., P.R. Stokke, J. Feder, T. Jossang. 1992. The fractal nature of geochemical landscapes. J. Geochem. Explor. 43: 91–109.

Borcard, D. and P. Legendre. 1994. Environmental control and spatial structure in ecological communities: an example using oribatid mites (Acari, Oribatei). Environmental and Ecological Statistics 1: 37–61.

Burrough, P.A. 1981. Fractal dimensions of landscapes and other environmental data. Nature 294: 240–242.

Burrough, P.A. 1987. Spatial aspects of ecological data. *In* Data Analysis in Community and Landscape Ecology. pp. 213–251. Edited by R.H.G. Jongman, C.J.F. ter Braak and O.F.R. van Tongeren. Pudoc, Wageningen.

Carr, J.R. and W.B. Benzer. 1991. On the practice of estimating fractal dimension. Math. Geol. 23: 945–958.

Cliff, A.D. and J.K. Ord. 1981. Spatial processes: models and applications. Pion, London.

Cressie, N.A.C. 1991. Statistics for spatial data. John Wiley & Sons, New York.

David, M. 1977. Geostatistical ore reserve estimation. Elsevier, Amsterdam.

Dutilleul, P. 1993. Spatial heterogeneity and the design of ecological field experiments. Ecology 74: 1646–1658.

Dutilleul, P. and P. Legendre. 1993. Spatial heterogeneity against heterocedasticity: an ecological paradigm versus a statistical concept. Oikos 66: 152–171.

Fortin, M.-J., P. Drapeau and P. Legendre. 1989. Spatial autocorrelation and sampling design in plant ecology. Vegetatio 83: 209–222.

Gardner, R.H. 1997. Pattern, process and the analysis of spatial scales. *In* Ecological Scale: Theory and Applications. Edited by D.L. Peterson and V.T. Parker. Columbia University Press, New York.

Garrett, R.G. 1983. Sampling methodology. *In* Handbook of Exploration Geochemistry. Vol. 2. pp. 83–110. Edited by R.J. Howarth. Elsevier, Amsterdam.

Geary, R.C. 1954. The contiguity ratio and statistical mapping. The Incorporated Statistician 5: 115–145.

Gower, J.C. 1962. Variance component estimation for unbalanced hierarchical classification. Biometrics 18: 537–542.

Greig-Smith, P. 1952. The use of random and contiguous quadrats in the study of the structure of plant communities. Ann. Bot. New Series 16: 293–316.

He, F., P. Legendre, C. Bellehumeur, J.V. LaFrankie. 1994. Diversity pattern and spatial scale: a study of a tropical rain forest of Malaysia. Environmental and Ecological Statistics 1: 265–286.

He, F., P. Legendre and J.V. LaFrankie. 1996. Spatial pattern of diversity in a tropical rain forest of Malaysia. Journal of Biogeography 23: 57–74.

He, F., P. Legendre and J.V. LaFrankie. 1997. Distribution patterns of tree species in a Malaysian tropical rain forest. J. Veget. Sci. 8: 105–114.

Hewitt, J.E., P. Legendre, B.H. McArdle, S.F. Thrush, C. Bellehumeur and S.M. Lawrie. 1997. Identifying relationships between adult and juvenile bivalves at different spatial scales. J. Exp. Mar. Biol. Ecol. (in press).

Journel, A.G. and Ch.J. Huijbregts. 1978. Mining geostatistics. Academic Press, London.

Kochummen, K. M., J.V. LaFrankie and N. Manokaran. 1991. Floristic composition of Pasoh forest Reserve, a lowland rain forest in Peninsular Malaysia. Journal of Tropical Forest Science 3: 1–13.

Legendre, P. and M.-J. Fortin. 1989. Spatial pattern and ecological analysis. Vegetatio 80: 107–138.

Levin, S.A. 1992. The problem of pattern and scales in ecology. Ecology 73: 1943–1967.

Lewis, W.M., Jr. 1978. Comparison of temporal and spatial variation in the zooplancton of a lake by means of variance components. Ecology 59: 666–671.

Ludwig, J.A. and D.W. Goodall. 1978. A comparison of paired- with blocked-quadrat variance methods for the analysis of spatial pattern. Vegetatio 38: 49–59.

Mandelbrot, B.B. 1983. The fractal geometry of nature. Freeman, San Francisco.

Matheron, G. 1965. Les variables régionalisées et leur estimation; une application de la théorie des fonctions aléatoires aux sciences de la nature. Masson, Paris.

Miesch, A.T. 1975. Variograms and variance components in geochemistry and ore evaluation. Geological Society of America, Memoirs, 142: 333–340.

Milne, B.T. 1991. Heterogeneity as a multiscale characteristic of landscapes. *In* Ecological Heterogeneity. pp. 69–84. Edited by J. Kolasa and S.T.A. Pickett. Springer-Verlag, New York.

Moran, P.A.P. 1950. Notes on continuous stochastic phenomena. Biometrika 37: 17–23.

Nortcliff, S. 1978. Soil variability and reconnaissance soil mapping: a statistical study in Norfolk. J. Soil Sci. 29: 403–418.

Oliver, M.A. and R. Webster. 1986. Combining nested and linear sampling for determining the scale and form of spatial variation of regionalized variables. Geogr. Analysis 18: 227–242.

Palmer, M.W. 1988. Fractal geometry: a tool for describing spatial patterns of plant communities. Vegetatio 75: 91–102.

Pinel-Alloul, B., J.A. Downing, M. Pérusse and G. Codin-Blumer. 1988. Spatial heterogeneity in freshwater zooplankton: variation with body size, depth, and scale. Ecology 69: 1393–1400.

Platt, T. and C. Filion. 1973. Spatial variability of the productivity: biomass ratio for phytoplankton in small marine basin. Limnol. Oceanogr. 18: 743–749.

Renshaw, E. and E.D. Ford. 1984. The description of spatial pattern using two-dimensional spectral analysis. Vegetatio 56: 75–85.

Rossi, R.E., D.J. Mulla, A.G. Journel and E.H. Franz. 1992. Geostatistical tools for modeling and interpreting ecological spatial dependence. Ecol. Monogr. 62: 277–314.

Shaw, R.G. and T. Mitchell-Olds. 1993. Anova for unbalanced data: an overview. Ecology 74: 1638–1645.

Sokal, R.R. and N.L. Oden. 1978a. Spatial autocorrelation in biology, 1. Methodology. Biological Journal of the Linnean Society 10: 199–228.

Sokal, R.R. and N.L. Oden. 1978b. Spatial autocorrelation in biology, 2. Some biological implications and four applications of evolutionary and ecological interest. Biological Journal of the Linnean Society 10: 229–249.

Sokal, R.R. and F.J. Rohlf. 1995. Biometry, 3rd ed. W.H. Freeman and Co., San Francisco.

Troussellier, M., A. Maul and B. Baleux. 1989. Stratégies d'échantillonnage. *In* Microorganismes dans les Écosystèmes Océaniques. pp. 27–62. Edited by M. Bianchi, D. Marty, J.-C. Bertrand, P. Caumette and M. Gauthier. Masson, Paris.

Ver Hoef, J.M, N.A.C. Cressie and D.C. Glenn-Lewin. 1993. Spatial models for spatial statistics: some unification. J. Veg. Sci. 4: 441–452.