Plant Ecology **130:** 89–98, 1997.

Variance and spatial scales in a tropical rain forest: changing the size of sampling units

Claude Bellehumeur[1], Pierre Legendre[1] & Denis Marcotte[2]
[1]*Département de sciences biologiques, Université de Montréal, C.P. 6128, succursale Centre-ville, Montréal, Québec, Canada H3C 3J7;* [2]*Département de génie minéral, École Polytechnique de Montréal, C.P. 6079, succursale Centre-ville, Montréal, Québec, Canada H3C 3A7*

Received 15 December 1995; accepted in revised form 30 October 1996

Key words: Autocorrelation, Geostatistics, Sampling design, Spatial scale, Size of sampling units, Tropical rain forest

Abstract

The size of a sampling unit has a critical effect on our perception of ecological phenomena; it influences the variance and correlation structure estimates of the data. Classical statistical theory works well to predict the changes in variance when there is no autocorrelation structure, but it is not applicable when the data are spatially autocorrelated. Geostatistical theory, on the other hand, uses analytical relationships to predict the variance and autocorrelation structure that would be observed if a survey was conducted using sampling units of a different size. To test the geostatistical predictions, we used information about individual tree locations in the tropical rain forest of the Pasoh Reserve, Malaysia. This allowed us to simulate and compare various sampling designs. The original data were reorganised into three artificial data sets, computing tree densities (number of trees per square meter in each quadrat) corresponding to three quadrat sizes (5×5, 10×10 and 20×20 m^2). Based upon the 5×5 m^2 data set, the spatial structure was modelled using a random component (nugget effect) plus an exponential model for the spatially structured component. Using the within-quadrat variances inferred from the variogram model, the change of support relationships predicted the spatial autocorrelation structure and new variances corresponding to 10×10 m^2 and 20×20 m^2 quadrats. The theoretical and empirical results agreed closely, while the classical approach would have largely underestimated the variance. As quadrat size increases, the range of the autocorrelation model increases, while the variance and proportion of noise in the data decrease. Large quadrats filter out the spatial variation occurring at scales smaller than the size of their sampling units, thus increasing the proportion of spatially structured component with range larger than the size of the sampling units.

Introduction

The concept of spatial scale is of fundamental importance in ecological modelling. A model cannot account for every detail of a system, and in return it may introduce false characteristics into our representation of the system. Determination of the appropriate spatial scale improves the quality of a model (Matheron 1984). Over the past few years, considerable efforts have been expanded to determine the scales at which the most significant variations of ecological variables occur, and to quantify these variations (Levin 1992). Most variables in natural environments display spatial structures such as gradients, patches, trends or other complex patterns. These structures can exist at many scales; they correspond to physical features of the environment, or they may result from ecological processes.

Several quantitative techniques have been brought to the attention of ecologists to assess the spatial variations of ecological variables: autocorrelograms (Legendre & Fortin 1989), variograms (Rossi et al. 1992), spectral analysis (Kenkel 1988), analysis of variance (Ludwig & Goodall 1978), fractal analysis (Palmer 1988), aggregation indices (Taylor et al. 1988). All these techniques depend on features of the sampling programme, and particularly on the size of

the sampling units. The size of the sampling units is an important component of the scale of an investigation and may critically influence our perception of ecological phenomena. Changing the size of the sampling units induces changes in the variance, the spatial autocorrelation structure, the indices of aggregation based on the variance-to-mean-ratio, and so on (He et al. 1994; Levin 1989).

Relating the size of the sampling units to the variance of plant densities has been extensively used in plant community ecology to empirically estimate the spatial scale of phenomena, or the size of patches (Carlile et al. 1989; Greig-Smith 1952; Ludwig & Goodall 1978; Marceau et al. 1994). These empirical studies do not provide a general framework to make predictions from a single sampling design; these methodologies require specific surveys to construct empirical relationships.

Classical statistical and geostatistical theories provide analytical solutions to predict the change in variance associated with different sizes of sampling units. Classical statistical theory works well to predict these changes when the sampling units are independent of one another. This is rarely the case in the natural sciences, however. In the present study, our primary objective is to present an analytical approach enabling the prediction of statistical parameters and features of the spatial autocorrelation structure that would be observed if a survey had been designed using different sizes for the sampling units; this may help plan future surveys. This analytical study complements the previous empirical studies which attempted to relate spatial scales to the size of the sampling units.

This method is already known in geostatistics as a change of support transformation (Journel & Huijbregts 1978). We will study a rain forest plot of Malaysia as a test case, using variable 'tree density' measured for different quadrat sizes. The change of support methodology is very general and the conclusions drawn from this test case are applicable to a wide range of studies dealing with additive variables that are spatially autocorrelated. Values of an additive variable can be added or averaged linearly in space to create larger quadrats.

Change of sampling unit size transformations can be used to compare different surveys where variables are estimated using different measurement devices. Buckland & Elston (1993) pointed out that the rapid growth of geographic information systems gives access to a wide range of variables recorded at different levels of resolution. For example, field work performed at small scale (i.e., small quadrat size) is essential to understand mechanisms and phenomena, and to estimate the activity or contribution of these phenomena at some local scale. Global environmental and ecological studies attempt to verify the applicability of ecological models over a wide range of spatial scales. Airborne or shipborne remote sensing involves data representing a large-scale perception of phenomena. A change in the sampling unit size is required when comparing or jointly using different types of data representing various scales of investigation.

Study site and variables

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 and 0.5 km wide (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. This data set is almost unique in that all individual trees are identified, sized, and geographically located. We reorganised the data into quadrats and calculated tree densities (number of trees per m^2 in each quadrat) corresponding to 5×5, 10×10, and 20×20 m^2 quadrats.

Methods

Classical statistical theory

Classical statistical theory attempts to predict the change in variance due to different sampling unit sizes for additive variables such as tree density. Classical theory states that given a population and a group of *n* independent sampling units drawn from that population of mean μ and variance σ^2, the mean value of this sample ($\overline{x}$) is itself a random variable of mean μ and variance σ^2/n. Classical statistical theory suggests (assuming no spatial structure) that the variance of the quadrats should decrease linearly with the number of sampling units in a quadrat or in a composite sample:

$$\text{Var}(V_c|A) = \text{Var}(v|A)/N, \quad (1)$$

where Var ($V_c \mid A$) is the variance of the composite samples V_c in an area A, Var ($v \mid A$) is the original

variance of the sampling units v in the same area, and N is the number of sampling units in a composite. A composite sample is formed by combining several adjacent sampling units.

This relationship (Equation (1)) is valid only for homogeneous areas where sampling units are independent of each other. When the process is complicated by patterns of spatial heterogeneity, Equation (1) is no longer valid. Levin (1989) and Wiens (1989) present some empirical results showing the complexity of such heterogeneous processes, where non-linear relationships occur between the variance and the size of the sampling units.

Most of the classical sampling textbooks mention that when a finite population is sampled using random sampling with replacement, the n observations are independent and identically distributed, so that the usual central limit theorem applies (e.g. Cochran 1977; Thompson 1992). Classical sampling theory is of limited help to analyse and characterise the spatial variability of natural processes. The random sampling method does not consider the structured spatial component of the variance. It assumes instead that the random component accounts for all the variance. Ecologists, on the other hand, are directly interested in the spatially-structured component of variation and have now developed methods to explicitly measure and map this component, independently of the non-spatially-structured random component (Borcard et al. 1992; Borcard & Legendre 1994).

Performing random sampling over spatial (or temporal) series does not insure that the sampling units are independent, because the values are part of a physical spatial (or temporal) process where the similarities between values depend expressly on the distance among locations. So, it is necessary to consider the spatial structure of phenomena in order to correctly predict the effect of a change of sampling unit size, or 'compositing' (i.e., creating composite samples by combining several sampling units into a single sample: Isaaks & Srivastava 1989, p. 341).

Spatial structure

Spatial continuity among observations of a given variable may be characterised by a variogram, which reveals the random and the structured aspects of the spatial dispersion. The variogram has been widely used to describe the spatial structure of ecological variables (Legendre & Fortin 1989; Rossi et al. 1992). The traditional estimator of the variogram is defined as (Journel & Huijbregts 1978; Cressie 1993):

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

where $Z(\mathbf{x})$ and $Z(\mathbf{x}+\mathbf{h})$ are measurements of a given variable at locations $\mathbf{x}$ and $\mathbf{x}+\mathbf{h}$, separated by the vector of directional distance $\mathbf{h}$, and $N(\mathbf{h})$ is the number of pairs of samples considered in the given distance class. This calculation is repeated for different values of $\mathbf{h}$ and provides the empirical variogram, which is a plot of the values of $\gamma^*(\mathbf{h})$ as a function of distance $\mathbf{h}$. The rate of increase of $\gamma^*(\mathbf{h})$ allows us to characterise 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 discontinuity at the origin (non-zero intercept) is called the nugget effect, which is a random component; the geostatistical origin of the method transpires in that name. It corresponds to local variations occurring at scales smaller than the sampling interval, including sampling error, fine-scale spatial variability, and measurement error. The variance component due to the nugget effect is called C_0, while the spatially structured component is C_1.

Change of sampling unit size operations

Problems of change of support have received attention in the geostatistical literature dealing with ore reserve estimation (e.g. Journel & Huijbregts 1978, pp. 61-94; Isaaks & Srivastava 1992, Chapter 19). Geostatisticians want to estimate the grade of large blocks from small drill core data. Equation (1) cannot be used here because the data are usually autocorrelated. In the field of vegetation science, Ver Hoef et al. (1993) have developed a relationship to express the variogram of aggregated (blocked) contiguous quadrats (large composite quadrats) as a function of an unaggregated variogram (smallest quadrats). The technique requires measurement of contiguous quadrats along transects or surfaces.

The additivity property of variances in nested designs implies:

$$\text{Var}(v|A) = \text{Var}(v|V) + \text{Var}(V|A), \quad (3)$$

where Var $(v \mid A)$ is the variance of a small sampling unit v in area A, Var $(V \mid A)$ is the variance of a large sampling unit V in area A, and Var $(v \mid V)$ is the variance of a small sampling unit v in the large sampling

unit V. This relationship shows that the variance of sampling units v in a certain area A can be expressed as the variance of sampling units v in blocks V, plus the variance of block values V within the area A. Journel & Huijbregts (1978, pp. 66–67) show that the variance Var $(v \mid V)$ is related to the variogram:

$$\text{Var}(v|V) = \overline{\gamma}(V, V) - \overline{\gamma}(v, v), \tag{4}$$

where $\overline{\gamma}$ (V,V) is the average point variogram value calculated over all possible distance vectors **h** contained in V, and similarly for $\overline{\gamma}$ (v,v). It represents the within-surface variance. Equations (3) and (4) allow one to calculate the variance corresponding to a new sampling unit size V, if we know the spatial autocorrelation structure for a point support.

When v is used to compute an empirical variogram, a regularised form of variogram is estimated. We must deduce a point model γ(**h**) (i.e., $v = 0$) from a regularised model γ_v(**h**), using the following formula from Journel & Huijbregts (1978, p. 78):

$$\gamma_v(\mathbf{h}) = \overline{\gamma}(v, v_h) - \overline{\gamma}(v, v). \tag{5}$$

γ_v(**h**) is the variogram defined for a sampling unit of size v, while $\overline{\gamma}$ $(v,\ v_h)$ represents the average value of the variogram, where one extremity of vector **h** describes a point within sampling unit v and the other extremity describes a point within another sampling unit v of the same size located at distance **h**. If $\gamma_v(\infty) = C_{1v}$, which is the sill value or the variance component of the spatial structure for v, then:

$$C_{1v} = C_{1p} - \overline{\gamma}(v, v), \tag{6}$$

where C_{1p} is the sill value for a point support. This relationship leads to:

$$C_{1p} = C_{1v}/(1 - F), \tag{7}$$

where F is equal to $\overline{\gamma}_1$ (v,v), representing the mean variogram value for a point variogram model with a sill equal to 1. Then, F can be computed from only the knowledge of the type of model and its range (Journel & Huijbregts 1978, p. 109). This correction only concerns the spatially-structured part of the variance. The variance component ascribed to random variation and modelled by a nugget effect (C_0) follows the classical relationship (Equation (1)).

The range of the spatial structure is also affected by the size of the sampling units. For example, the range of a spatial structure estimated from a sampling unit of size $l \times l = v$, is $a_p + l$, where a_p is the practical range that would be measured if the support was a point. Therefore, changing the size of the sampling unit produces changes in the overall variance and in the parameters characterising the variogram: nugget effect, relative nugget effect, structured variance component, and range. Analytical relationships are developed to predict the variance and the autocorrelation structure corresponding to change of sampling unit size.

In practice, if the data are defined for v, it is necessary to deduce first an approximate point model γ(**h**) which is coherent with the empirical variogram γ_v(**h**). Obtaining the point variogram from a regularised variogram is, strictly speaking, impossible as it requires knowledge of the point scale structure, which is not available. Moreover, the model corresponding to a sampling unit of size v is not exactly equivalent to the model associated to a sampling unit of size v'. However, the main features (sill and range) of the new model $\gamma_{v'}(\mathbf{h})$ can be deduced from the model $\gamma_v(\mathbf{h})$. The following rules provide acceptable approximations to obtain the point variogram and the regularised variograms $\gamma_{v'}(\mathbf{h})$ corresponding to new sampling units of size v':

(1) For the structured part of $\gamma_v(\mathbf{h})$, the point variogram is approximated by a variogram of the same type with a practical range of $a_p = a_v - l$ and a sill $C_{1p} = C_{1v}/(1 - F)$.

(2) The above-defined point variogram is used, assuming that $\gamma_{v'}(\mathbf{h})$ is of the same type with a range $a_p + l'$ and a sill $C_{1v'} = C_{1p} - \overline{\gamma}(v', v')$.

(3) The nugget effect component, corresponding to v', is computed as $C_{0v'} = C_{0v}/v'$, where C_{0v} is the nugget effect corresponding to v. This random component is added to the spatially-structured model $\gamma_{v'}(\mathbf{h})$ defined in (2).

So, a point variogram model $\gamma(\mathbf{h})$ is defined from the model fitted to the empirical values ($\gamma_v(\mathbf{h})$), and the various parameters (range, sill, etc.) of the point model are obtained from the equations above. Once the point model $\gamma(\mathbf{h})$ and its parameters have been found, another expression $\gamma_{v'}(\mathbf{h})$ can be deduced for another sampling unit size v'. Knowledge of the point model allows the calculation of mean variogram values ($\overline{\gamma}(v,v)$) for any size of sampling units.

Results

Histograms and summary statistics

Histograms and summary statistics of density values for quadrats of 5×5, 10×10, and 20×20 m^2, show that as size increases, extreme values disappear because they are diluted and combined into larger quadrats (Figure 1). The mean remains constant but the variance decreases. The variance-to-mean-ratio, which is the basic statistic for a large family of aggregation indices (Taylor's *b*, Morisita and Lloyd indices, among others) decreases as size increases (Patil & Stiteler 1974). This dependence of the variance-to-mean-ratio on quadrat size has been pointed out by Sawyer (1989). Equation 1 shows that, for additive variables such as density, and assuming no spatial structure in the data, the variance of quadrats should decrease linearly with the number of sampling units in a quadrat or in a composite sample. Our empirical results show a significant departure from classical theory predictions. Considering the empirical counts in 5×5 m^2 quadrats as our basis for calculations (variance $= 0.0610$), we would expect, from classical theory, variances of 0.0153 and 0.00381 for quadrats of 10×10 and 20×20 m^2 respectively. These results are much smaller than the empirical variances of 0.0275 and 0.0161 (Figures 1b, 1c).

Spatial structure of tree density

Empirical variograms of the tree density variable corresponding to the 5×5, 10×10 and 20×20 m^2 quadrat sizes, for the north-south and east-west directions, show well-defined sills (Figure 2). They are largely isotropic, i.e., $\gamma(\mathbf{h})$ does not depend on the direction of $\mathbf{h}$. Exponential models with nugget effect provided good fit to the empirical variograms:

$$\gamma(\mathbf{h}) = C_0 + C_1(1 - \exp(-\mathbf{h}/a)), \qquad (8)$$

where C_0 is the nugget effect, C_1 is the variability due to the structure in the exponential model and a is a shape parameter related to the practical range (Table 1). The exponential model reaches its sill $(C_0 + C_1)$ asymptotically. The ratio of the nugget effect to the sill, called the relative nugget effect $RelC_0$, represents the random component proportion in the spatial variance. The range of a model with a sill is the distance where the spatial influence disappears, i.e. where $\gamma(\mathbf{h})$ ceases to increase. The practical range of an exponential model is defined as $3a$, the distance at which the variogram is 95% of C_1.

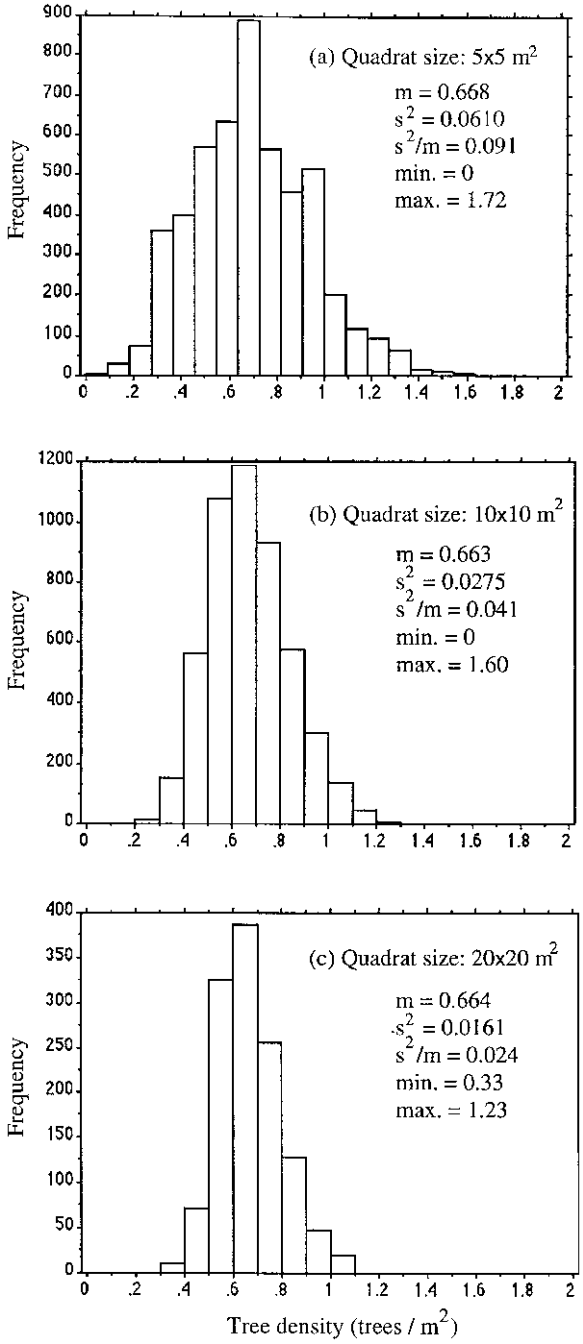

Figure 1. Histograms of tree density values for (a) 5×5 m^2, (b) 10×10 m^2, and (c) 20×20 m^2 quadrats.

As quadrat size increases, sill values decrease and ranges increase. The most important effect is the decrease in relative nugget effect. For the 5×5 m^2 quadrats, the proportion of random variation is very

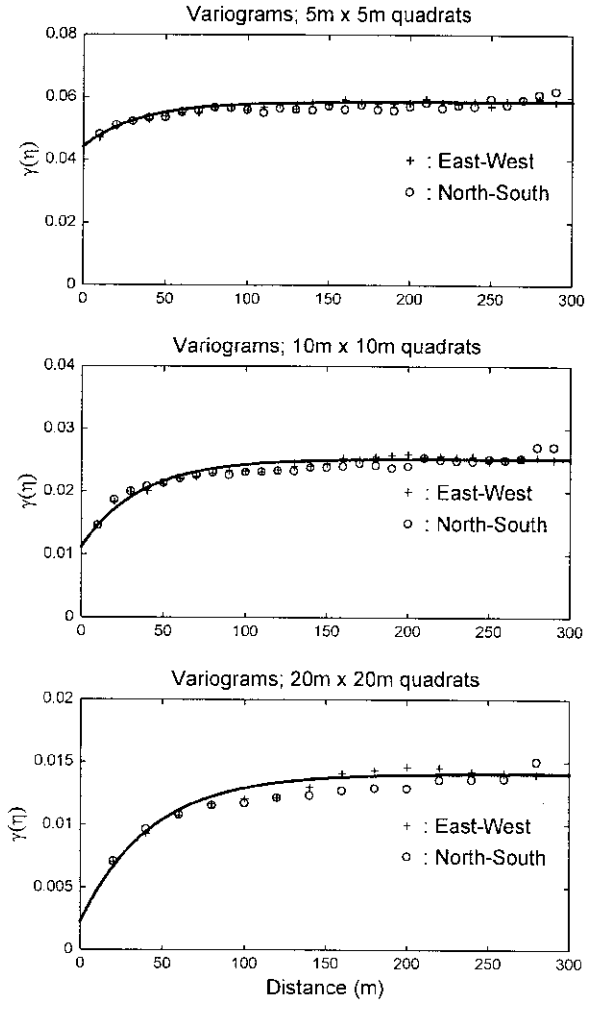

Figure 2. Directional variograms of the tree density variable for 5×5 m^2, 10×10 m^2 and 20×20 m^2 quadrats.

Table 1. Parameters of exponential variogram models for 5×5 m^2, 10×10 m^2 and 20×20 m^2 quadrat sizes.

Quadrat size (m^2)	C_0	C_1	Sill	a (m)	$RelC_0$
5×5	0.0446	0.0151	0.0597	104	0.75
10×10	0.0113	0.0139	0.0253	110	0.45
20×20	0.00245	0.0118	0.0145	129	0.17

C_0 is the nugget effect, C_1 is the variance component associated to the structured spatial scale, the sill is $C_0 + C_1$, a is the practical range and $RelC_0$ is the relative nugget effect $(C_0/(C_0 + C_1))$.

high (75%), and the process does not seem very strongly spatially structured. On the other hand, for 20×20 m^2 quadrats, the process displays an important spatially-structured component accounting for 83% of the spatial variance. Large quadrats have filtered out the spatial variation occurring at scales smaller than their sampling unit size, thus increasing the proportion of spatially-structured components with ranges greater than the size of the sampling units.

Empirical verification of the change of sampling unit size relationships

The previous sections have shown empirically that changing the size of the sampling units extensively modifies the variance estimate, as well as our perception of the spatial autocorrelation structure of data. We will now check whether the geostatistical relationships provide an analytical solution allowing us to predict our empirically obtained results.

For the 5×5 m^2 quadrat size, the practical range is 3×34.67 m $= 104$ m (Table 1). Given that for an exponential model, the practical range equals $3a$, then the parameter a_p of a point model is equal to $(3a - l)/3 = (104 - 5)/3 = 33.0$. Estimating the point sill value of the structured component requires the evaluation of the within-quadrat variance $\overline{\gamma}(v, v)$. The mean value $\overline{\gamma}(v, v)$ can be calculated numerically from function $\gamma(\mathbf{h})$ by discretising sampling unit v into a finite number of points or by generating random lags within v, and calculating the average variogram values for lags contained in v (Table 2). This calculation can also be done by evaluating the integral of the variogram function. For this purpose, Journel & Huijbregts (1978, pp. 108–123) developed a series of auxiliary functions giving a precalculated mean value $\overline{\gamma}(v, v)$ corresponding to simple geometries of v which are frequently found in practice. Tables and graphs giving $\overline{\gamma}(v, v)$ from these auxiliary functions are presented in Journel & Huijbregts (1978, pp. 125–147) for some variogram models.

Using the mean variogram values for the 5×5 m^2 quadrat size (Table 2), the point sill value of the structured component is given by formula (7) as:

$$C_{1p} = C_{1v}/(1 - F),$$

$$C_{1p} = 0.0151/(1 - 0.0755),$$

$$C_{1p} = 0.0163.$$

The theoretical point support variogram is an exponential model:

$$\gamma(\mathbf{h}) = 0.0163(1 - \exp(-\mathbf{h}/33)). \quad (9)$$

Table 2. Mean variogram values of various quadrat size.

Quadrat size v (m^2)	$\overline{\gamma}(v,v)$	F
5×5	0.00123	0.0755
10×10	0.00234	0.143
20×20	0.00429	0.263

$\overline{\gamma}(v,v)$ is the average point variogram value calculated for a quadrat v, representing the within-surface variance. The calculation uses the following exponential model [$\gamma(\mathbf{h}) = 0.0163(1 - \exp(-\mathbf{h}/33))$] as a theoretical point variogram. F is the mean variogram value for the point variogram model with a sill equal to 1.

In the previous section, the exponential model was shown to be an appropriate model of spatial structures (Figure 2). Such a point variogram model could be deduced for any other quadrat size, as long as the quadrats are not too large relative to the range of the point variogram. From this theoretical point model, it is possible to calculate the variance of any given sampling unit size in the whole area and to find an appropriate variogram model describing the spatial structure features for various quadrat sizes.

The point variogram model is obtained by assuming that it is of the same type as the one corresponding to support *v*. We verify the adequacy of this approximation by comparing theoretical regularised variograms (for 5×5 m^2, 10×10 m^2 and 20×20 m^2 quadrats) obtained from the point variogram model (Equation 9) and Equation (5), with variograms obtained under the assumption that the point variogram model is of the same type as the one corresponding to sampling unit size v. These models were also compared with the empirical variograms modelled from the original data. The point variogram model deduced for variable 'tree density' is shown in Figure 3a. The regularised variograms for supports of size 5×5 m^2, 10×10 m^2 and 20×20 m^2 are shown in Figures 3b, 3c, and 3d. These regularised variograms were calculated numerically from equation (5), by stochastic integration where 100 000 random lags were generated within v. The above approximation leads to models (Figure 3b, c, d, dashed) showing a good fit to the regularised variograms (full lines). Departure from the theoretical model is more important near the origin, at a scale smaller than the sampling units, although departure from the whole curve increases with the size of the sampling units. It could be explained by a slight lack of stationarity in the data indicated by an increase in variogram values beyond 250 m in the north–south direction. This increase corresponds to another structure with a range that cannot be evaluated from the available data; it actually represents a trend in the data.

The variogram models have two components each: a random and a spatially structured component (Figure 2). The change in the random component due to a change of support follows the classical relationship and is defined by Equation (1). So, knowledge of the variance corresponding to a given sampling unit size allows the estimation of the random component corresponding to any other sampling unit size. The random component for the 5×5 m^2 quadrats is 0.0446. Therefore, for 10×10 m^2 quadrats, the random component should be $0.0446/4 = 0.0112$, and for 20×20 m^2 quadrats, $0.0446/16 = 0.00279$.

On the other hand, the effect of a change of sampling unit size operation on the spatially-structured component of variance depends on the spatial autocorrelation structure (point model variogram) and is given by Equation (6). For example, for 10×10 m^2 quadrats, using parameters given in Table 2, the structured variance component for the 10×10 m^2 quadrat size is:

$$C_1(10 \times 10) = C_{1p} - \overline{\gamma}(10,10),$$

$$C_1(10 \times 10) = 0.0163 - 0.00234,$$

$$C_1(10 \times 10) = 0.0140.$$

This analytical solution gives an overall variance of 0.0252 for 10×10 m^2 quadrats ($C_{0(10\times10)} + C_{1(10\times10)}$). The empirical value is 0.0275, while the classical approach would have given $0.0610/4 = 0.0153$. Thus the analytical solution is closer to the empirical value than the classical relationship. The slight underestimation may be due to a long-range spatial structure in the north-south direction which is not modelled, considering the size of this structure compared with the size of the study area. Table 3 allows comparison of the empirical variogram parameters to the parameters inferred from the theoretical point model (Equation (9)); the theoretical models are very close to the empirical results. These results show that in practice, from an empirical variogram, we can deduce a theoretical point model which enables the estimation of appropriate models corresponding to given sampling unit sizes.

Discussion

The change of sampling unit size operation uses both the among-sampling-unit and within-sampling-unit variances to compute the statistical parameters of

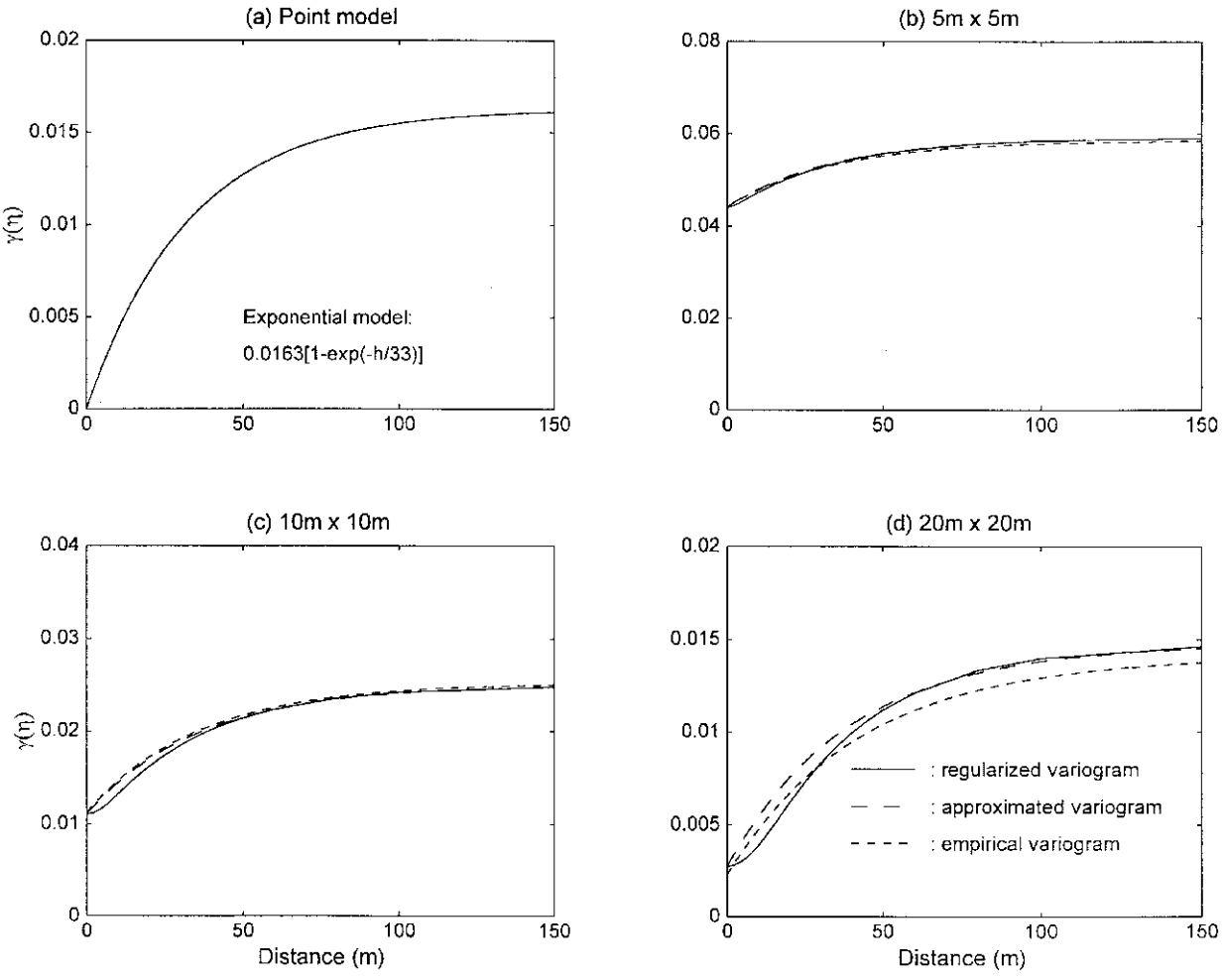

Figure 3. (a) Point model variogram deduced for the tree density variable. (b) to (d): regularised variograms (full lines) for sampling units of size 5×5 m^2, 10×10 m^2 and 20×20 m^2, calculated numerically from Equation (5), models derived from approximations (dashed), and empirical models (points).

Table 3. Results of the change of support transformation.

Quadrat size (m^2)	C_0(emp)	C_0(inf)	C_1(emp)	C_1(inf)	Var($v \mid A$)	s^2
5×5	0.0446	0.0446	0.0151	0.0151	0.0597	0.0610
10×10	0.0113	0.0112	0.0139	0.0140	0.0252	0.0275
20×20	0.00245	0.00279	0.0118	0.0120	0.0148	0.0161

C_0(emp) is the empirical nugget effect, C_0(inf) is the nugget effect inferred from the empirical value of the 5×5 quadrat size, C_1(emp) is the empirical structured variance component, C_1(inf) is the structured variance component inferred from the theoretical model where $C_{1v} = C_{1p} - \overline{\gamma}(v, v)$, Var$(v|A)$ is the variance of a unit v in the study area A, given by the analytical relationship, and s^2 is the empirical variance. The number of significant digits has been kept the same from row to row in the table.

a distribution and the features of the spatial structure that would be observed if a survey had been conducted using different sizes for the sampling units. The calculation of these variance components is based on a variogram model which expresses the variance of a spatial process as a function of geographical distances. The variance is broken up into its random and spatially structured components of variation.

Greig-Smith (1952) and subsequent authors (among others, Ludwig & Goodall 1978) developed methods based on quadrat-count analysis. The peaks in a plot of mean square against support size are considered to be associated with a given patch size, or the scale of an ecological phenomenon. Change of sampling unit size relationships lead to very different conclusions about the determination of spatial scales. Parameters of the variograms for different quadrat sizes show that phenomena spread out in space do not have discrete spatial scales, but a continuum of spatial structures whose perception depends on the size of the sampling units.

A change of sampling unit size operation involves the following three steps: (1) An empirical variogram $\gamma^*(\mathbf{h})$ is calculated from the data, and a variogram model $\gamma_v(\mathbf{h})$ is derived, corresponding to a regularised form of variogram for a given sampling unit size v. (2) A point model $\gamma(\mathbf{h})$ (i.e., $v = 0$) is deduced from the regularised model $\gamma_v(\mathbf{h})$, using Equation (7): $C_{1p} = C_{1v}/(1-F)$. The variance component ascribed to random variation follows the classical relationship (Equation (1)). The range of the point model is deduced from the relationship: $a_v = a_p + l$, where a_p is the practical range of the point model. (3) Once the point model $\gamma(\mathbf{h})$ and its parameters have been found, an expression $\gamma_{v'}(\mathbf{h})$ can be derived for another sampling unit size v'. Knowledge of the point model allows one to calculate the mean variogram values $\overline{\gamma}(v, v)$ for any size of sampling units.

The geostatistical predictions were verified using an exhaustive data set including tree density in the tropical rain forest of the Pasoh Reserve. This data set allowed us to confirm that change of sampling unit relationships agree closely with the empirical results. We have shown five key results: (1) As the sampling unit size increases, extreme values disappear from the distribution because they are diluted and combined into larger quadrats; consequently, the variance decreases while the mean remains constant. (2) With the increase in quadrat size, the range of autocorrelation increases, while the variance and the proportion of noise in the data decrease. (3) For a homogeneous area, the reduction in variance associated with a certain increase in sampling unit size is more important than for a heterogeneous area where spatial autocorrelation is present. (4) Large quadrats filter out the spatial variation occurring at scales smaller than their sampling unit size, thus increasing the proportion of the spatially-structured component with range greater than the size of the sampling units. (5) From an empirical variogram, we can deduce a theoretical point model that enables the estimation of appropriate models corresponding to various supports.

In practice, it is not possible to infer properties and features for sampling unit sizes smaller than the smallest available sampling unit without introducing unverifiable hypotheses. The variogram calculated from large sampling units does not contain information about structures smaller than the size of the actual sampling units. Large quadrats do filter out possibly interesting short-range spatial scales.

On the other hand, small sampling units introduce a larger variance and a smaller degree of precision than larger units. Several authors have discussed the problem of determining the sampling variance for any given sampling unit size (general review in Kratochvil et al. 1984), which is related to the expected fluctuations of the estimation error due to a lack of representativeness of that unit size. Considering surfaces, this variance can be expressed as a relationship of the form:

$$s^2 \propto d^2/A, \tag{10}$$

where d is the linear size of the object of interest (average diameter of trees for example) and A is the surface area of the sampling unit. The sampling variance, in that case, is proportional to the square (or the cube in the case of a volume) of the linear size of the object of interest in the sampling unit, and inversely proportional to the area (or volume) of the sampling unit. So, measuring large sampling units decreases the estimation error.

David (1977, p. 337) proposes the following example. Let us count the number of objects in sampling units that contain on average three objects each, distributed among the units following a Poisson process (random distribution). The variance being equal to the mean, the standard deviation is $\sqrt{3} = 1.73$ or, in relative values (coefficient of variation), $(1.73/3) \times 100 = 58\%$. Now, let us use larger sampling units containing on average ten objects each. The standard deviation is $\sqrt{10} = 3.16$ and its relative value to the mean is $3.16/10 \times 100 = 32\%$, which is smaller than 58%. Obviously, reality displays a higher degree

of complexity; but these relationships clearly show that random fluctuations become less important, and the relative precision becomes better, when increasing the size of the sampling units.

Acknowledgements

We are grateful to Dr Jay Ver Hoef for extremely useful comments on this manuscript. The large-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. 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 programme, 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. OGP0007738 to P. Legendre.

References

Borcard, D., Legendre, P. & Drapeau, P. 1992. Partialling out the spatial component of ecological variation. Ecology 73: 1045–1055.

Borcard, D. & Legendre P. 1994. Environmental control and spatial structure in ecological communities, with an example on Oribatid mites (Acari, Oribatei). Environ. and Ecol. Statistics 1: 37–53.

Buckland, S. T., & Elston, D. A. 1993. Empirical models for the spatial distribution of wildlife. J. Appl. Ecol. 30: 478–495.

Carlile, D. W., Skalski, J. R.,. Barker, J. E., Thomas, J. M. & Cullinan, V. I. 1989. Determination of ecological scale. Landscape Ecol. 2: 203–213.

Cochran, W. G. 1977. Sampling techniques. 3rd edition. John Wiley and Sons, New York, USA.

Cressie, N. A. C. 1993. Statistics for spatial data. Revised edition. John Wiley & Sons. New York, USA.

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

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

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

Isaaks, E. & Srivastava, R. M. 1989. An introduction to applied geostatistics. Oxford University Press, New York, USA.

Journel, A. G. & Huijbregts, C. J. 1978. Mining geostatistics. Academic Press, London, England.

Kenkel, N. C. 1988. Spectral analysis of hummock-hollow pattern in a weakly minerotrophic mire. Vegetatio 78: 45–52.

Kochummen, K. M.,. LaFrankie, J. V. & Manokaran, N. 1991. Floristic composition of Pasoh forest Reserve, a lowland rain forest in Peninsular Malaysia. J. Trop. For. Sci. 3: 1–13.

Kratochvil, B., Wallace, D. & Taylor, J. K. 1984. Sampling for chemical analysis. Anal. Chem. 56: 113R–129R.

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

Levin, S. 1989. Challenges in the development of a theory of ecosystem structure and function. pp. 242–255. In: Roughgarden, J., May, R. M. & Levin, S. A. (eds), Perspectives in Ecological Theory. Princeton University Press, Princeton, New Jersey, USA.

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

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

Marceau, D. J., Gratton, D. J., Fournier, R. A. & Fortin, J-P. 1994. Remote sensing and the measurement of geographical entities in a forested environment. 2. The optimal spatial resolution. Remote Sensing Envir. 49: 105–117.

Matheron, G. 1984. Estimating and choosing. Springer-Verlag, Berlin.

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

Patil, G. P. & Stiteler, W. M. 1974. Concepts of aggregation and their quantification: a critical review with some new results and applications. Res. Pop. Ecol. 15: 238–254.

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

Sawyer, A. J. 1989. Inconstancy of Taylor's b: simulated sampling with different quadrat sizes and spatial distribution. Res. Popu. Ecol. 31: 11–24.

Taylor, L. R., Perry, J. N., Woiwod, I. P. & Taylor, R. A. J. 1988. Specificity of the spatial power-law exponent in ecology and agriculture. Nature 332: 721–722.

Thompson, S. K. 1992. Sampling. John Wiley and Sons. New York, USA.

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

Wiens, J. A. 1989. Spatial scaling in ecology. Functional Ecol. 3: 385–397.