Vegetatio **83**: 209–222, 1989.

Spatial autocorrelation and sampling design in plant ecology

Marie-Josée Fortin[1], Pierre Drapeau[2] & Pierre Legendre[2]
[1]*Department of Ecology and Evolution, State University of New York, Stony Brook, New York 11794-5245, USA;* [2]*Département de sciences biologiques, Université de Montréal, C.P. 6128, Succursale A, Montréal, Québec, Canada H3C 317*

Accepted 6.3.1989

Keywords: Kriging, Pattern analysis, Reliability, Sampling theory

Abstract

Using spatial analysis methods such as spatial autocorrelation coefficients (Moran's I and Geary's c) and kriging, we compare the capacity of different sampling designs and sample sizes to detect the spatial structure of a sugar-maple (*Acer saccharum* L.) tree density data set gathered from a secondary growth forest of southwestern Québec. Three different types of subsampling designs (random, systematic and systematic-cluster) with small sample sizes (50 and 64 points), obtained from this larger data set (200 points), are evaluated. The sensitivity of the spatial methods in the detection and the reconstruction of spatial patterns following the application of the various subsampling designs is discussed. We find that the type of sampling design plays an important role in the capacity of autocorrelation coefficients to detect significant spatial autocorrelation, and in the ability to accurately reconstruct spatial patterns by kriging. Sampling designs that contain varying sampling steps, like random and systematic-cluster designs, seem more capable of detecting spatial structures than a systematic design.

Abbreviations: UPGMA = Unweighted Pair-Group Method using Arithmetic Averages.

Introduction

The spatial component is of prime importance in the planning of any field ecological study. Ecologists have to answer the following questions: At what scale is the system going to be studied? What quadrat size should be used? How far apart should the sampling stations be located? The answers to these questions depend to a large extent on the purpose of the study, and to the knowledge that can be acquired during pre-sampling campaigns (pilot studies) about the spatial distribution of the variables of interest.

Ecologists are interested in adequate descriptions of spatial distributions for three different reasons. One is that spatial heterogeneity plays a central role, implicitly or explicitly stated, in most ecological theories. This point has been made in some detail in a companion paper by Legendre & Fortin (1989), and also by Legendre *et al.* (1989). A second reason is that modelling requires that the dependent variable's range of variability be sampled adequately, which can only be done from knowledge of the spatial distribution of the values of that variable; for instance, a model's dependent variable distribution area can be stratified into

geographic zones that are homogeneous in terms of the values of that variable, following which each zone can be sampled with equal intensity. Furthermore, proper descriptions of the spatial structures of the variables to be included in a model can help formulate or support hypotheses concerning possible causal mechanisms determining the dependent variable's spatial distribution. A third reason is that when the purpose of the study is to estimate population parameters, such as the mean, the variance, or the total amount of resource represented by the variable in a given area, stratified sampling can be cheaper than either random or systematic sampling to attain a given level of precision of the estimation. Notice however that random or systematic sampling, even though possibly more expensive to carry out, are still adequate to achieve an unbiased estimation of the parameters of interest, because in random or systematic sampling designs each point in the geographic distribution area has *a priori* the same probability of being included in the sample. A good way of stratifying is to divide the area into geographic zones where the variance, or the coefficient of variation of the variable of interest, are approximately equal; following that, sampling is carried out more intensively in the zones where the values are more variable, and less so in more homogeneous ones. For all these purposes, a variable's spatial structure has to be described adequately, which can be done using structure functions such as correlograms and variograms, maps, and other spatial analysis methods described for instance by Upton & Fingleton (1985), Legendre & Fortin (1989), and other authors.

Ecologists may also be interested in spatial structures because they intend to carry out tests of classical inferential statistics, based upon data obtained through some geographic distribution area. The problem, described in some detail by Legendre & Fortin (1989), is that classical hypothesis testing makes the assumption that the observations (elements) are independent from one another, while this condition is not met by data that are autocorrelated through space. Indeed, the very existence of a spatial structure in the sampling area implies that any ecological phenomenon found at a given sampling point may have an influence on other points located close by, or even some distance away. Thus, even though the observations are independent from the point of view of the probability of any particular geographic location to be sampled, their values at neighbouring points may not be independent from one another, which violates the assumption of independence of the observations. Contrary to the advocacies of several authors (for instance Cochran 1977; Green 1979; Scherrer 1982), sampling designs cannot alleviate this problem. In this context, spatial analysis becomes a prerequisite to hypothesis testing for data gathered through space. It may be used to demonstrate that there is no significant spatial autocorrelation at the given sampling scale, in which case classical statistical tests of hypothesis can be used; or it may indicate that other methods should be used for testing.

In order to adequately plan a pre-sampling study, or a full-scale sampling program from which spatial structures will be analyzed, it is essential to know how the properties of different sampling designs can affect the sensitivity of spatial analyses in detecting significant spatial autocorrelation or in reconstructing spatial patterns by mapping. Different sampling designs have different properties, advantages and disadvantages. As mentioned above, simple random sampling allows one to obtain a representative sample of the statistical population, when the purpose is to estimate population parameters (Cochran 1977; Scherrer 1982). The major disadvantage of simple random sampling is that it is often very difficult to carry out in the field because of difficulties in positioning the sampling stations. Systematic sampling, on the contrary, is much easier to conduct in the field; furthermore, it is regarded as appropriate for detecting spatial or temporal autocorrelation, if and only if the sampling step (interval between successive samples) is right (Cochran 1977). These two sampling designs do not necessitate prior knowledge of the variable subjected to sampling and of the properties of the area under study. Other types of design such as stratified sampling, regression sampling, cluster

sampling, usually require previous knowledge of the behaviour of the variable in the study area, which can only be obtained from pre-sampling (pilot) studies; Legendre *et al.* (1989) discuss methods for designing full-scale sampling programs from the results of such preliminary samplings.

There are two types of considerations that may influence the choice of the sampling design: statistical and nonstatistical. The former include the accuracy of the estimation and the type of statistics that have to be used afterwards, while the later include financial, time, equipment and personnel constraints, as well as data processing limitations (McCall 1982). Unfortunately, in a majority of instances, the nonstatistical considerations will determine the sampling design and the sample size. Systematic sampling is often preferred to simple random sampling because of its simplicity in the field (Scherrer 1984). Sample size is often limited by financial and time constraints. However, there is a lower limit to the sample size which is dictated by the subsequent statistical or numerical analyses. Usually, the number of samples must be an inverse function of the homogeneity of the area under study.

In plant ecology, there is a long tradition of studies on the importance of sample size, sampling design, and their effects on the gathering of data (Greig-Smith 1952, 1964, 1979). The sensitivity and robustness of multivariate ordination techniques, under different sampling designs, have been tested by Mohler (1981, 1983), Minchin (1987) and Podani (1984, 1987). These studies have shown that the performance of ordination techniques varies under different sampling patterns (Minchin 1987), and that the 'optimal' sample size varies with the method used (Podani (1987) and Podani (1984, 1987). These studies have shown that the performance of ordination techniques varies under different sampling patpling designs, have been conducted (McBratney *et al.* 1981; Webster & Burgess 1984; McBratney & Webster 1986). These studies, however, were concerned mostly with the behaviour of kriging, which is a mapping interpolation method, under types of sampling designs that are more common in soil sciences than in plant ecology (McBratney *et al.* 1981).

This paper describes the effect of the number of samples and of the sampling design (i.e., relative position of the samples) in terms of the ability of various designs to identify spatial structures and to produce accurate maps, using a data set of sugar-maple tree densities from a secondary growth forest in southwestern Québec. It is a companion to the paper by Legendre & Fortin (1989) describing methods that can be used to study the spatial structure of biological populations and communities. In this study, we will subsample this real vegetation data set, simulating various sampling designs. This will allow to evaluate the effect of these various sampling designs and intensities on the estimation of spatial structures, on the one hand, and on the other the sensitivity of different methods used in the spatial analysis of ecological structures. The interest of this study lies in that it is based upon real vegetation data; its weakness, compared to Monte-Carlo simulation studies, resides in the fact that only one real extensive data set is available and will be subjected to the various subsampling designs and intensities.

Spatial analysis methods

Spatial autocorrelation coefficients have been introduced by Moran (1950) and Geary (1954). Geographers (Cliff & Ord 1981) have used them to analyze epidemiological data; population geneticists (Sokal & Menozzi 1982), to study gene frequencies and gene flow; recently they have been applied to the study of ecological data (Jumars 1978; Sokal 1979; Gloaguen & Gauthier 1981; Bouxin & Gauthier 1982; Sakai & Oden 1983; Fortin 1985; Sokal & Thomson 1987; Legendre & Troussellier 1988; Legendre & Fortin 1989).

Moran's I and Geary's c coefficients are used to measure the degree of spatial autocorrelation displayed by a quantitative variable, and to test the null hypothesis (H_0) that there is no significant spatial autocorrelation (positive: aggregation;

negative: segregation). Since these coefficients compare values for pairs of points, the set of available point pairs is divided into a number of distance classes. This number of classes is left to the user. Like Pearson's correlation coefficient, Moran's I is based on the computation of cross-products of centered data. Geary's c is a distance-type coefficient, summing squared differences between adjacent pairs of values. Spatial autocorrelation analysis should not be performed with fewer than ca 30 localities, because the number of pairs of localities in each distance class would then become too small to produce significant results (Cliff & Ord 1981; Legendre & Fortin 1989). Formulas for computing the coefficients as well as the standard error of the estimated statistics can be found in Cliff & Ord (1981), in Sokal & Oden (1978) and in Legendre & Legendre (1984).

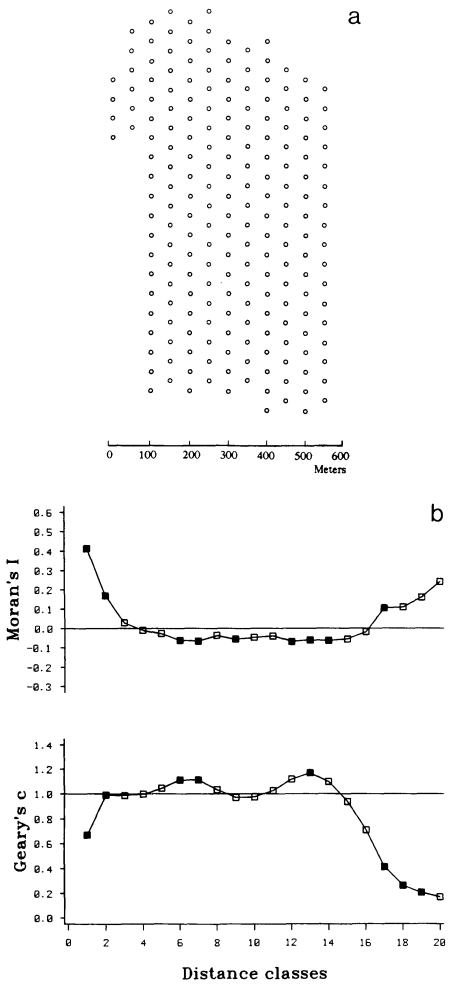
A correlogram is a plot of autocorrelation coefficient values in ordinate, against distance classes in abscissa. Correlograms provide evidence for the autocorrelation intensity, the size of the zone of influence and the type of spatial pattern of the variable under study. The shape of a correlogram gives indications about the spatial pattern of the variable, as well as about the underlying generating process (Sokal 1979; Legendre & Fortin 1989). Inference about the underlying generating process can be made from the shape of the correlogram only when the correlogram is globally significant; Oden (1984) and Legendre & Fortin (1989) show how to compute such a global test, whose aim is to correct for simultaneous multiple testing.

Spatial ecological structures can also be analyzed with the help of density contour maps. In the present study we produced interpolated maps by kriging, which is a geostatistical method developed by mining engineers (Matheron 1973; David 1977; Journel & Huijbregts 1978). Kriging is a method of interpolating that makes use of the spatial autocorrelation structure of the variable. It is used in soil science (McBratney & Webster 1986; Burrough 1987), in forestry (Bouchon 1974; Marbeau 1976; Fortin 1985; Legendre & Fortin 1989) and in other fields. Kriging uses a structure function, called a semi-variogram (or simply a variogram), to give weights to the various data points located in the vicinity of each point to be estimated. Kriging assumes that the data are stationary, or in other words, that they contain no significant trend (Journel & Huijbregts 1978; Legendre & Fortin 1989). However, when the data are not stationary, kriging can be used with a relaxed assumption, the intrinsic hypothesis, which implies that increments between all pairs of points located a given distance d apart have a mean of zero and a finite variance, that remains the same in the various parts of the area under study. When the data are very non-stationary, other forms of kriging can be used, that base these weights on so-called 'intrinsic random functions of order k' instead of the variogram. Depending of the number of neighbouring points used for the estimation (interpolation), kriging is said to be local (estimation based on a few neighbouring points) or global (estimation based on all data points). The semi-variance function used to compute a variogram is closely related to Geary's c coefficient (Legendre & Fortin 1989), but contrary to Geary's c it cannot be tested for significance.

Variograms and kriged maps were obtained using the GEOSTAT computer package (Geostat Systems International Inc., 4385 St-Hubert, Suite 1, Montréal, Québec, Canada H2J 2X1). The spatial correlograms were computed using the 'R package' (Legendre 1985).

Materials and methods

Data presented in this paper were gathered during a multidisciplinary ecological study of the terrestrial ecosystem of the Municipalité Régionale de Comté du Haut-Saint-Laurent (Bouchard *et al.* 1985) in southwestern Québec. In an area of approximately 0.5 km^2, a systematic sampling design was used to survey 200 vegetation quadrats (Fig. 1a) each 10 by 20 m in size. The quadrats were placed at 50 m intervals along staggered rows separated also by 50 m. Trees with more than 5 cm diameter at breast height (DBH),



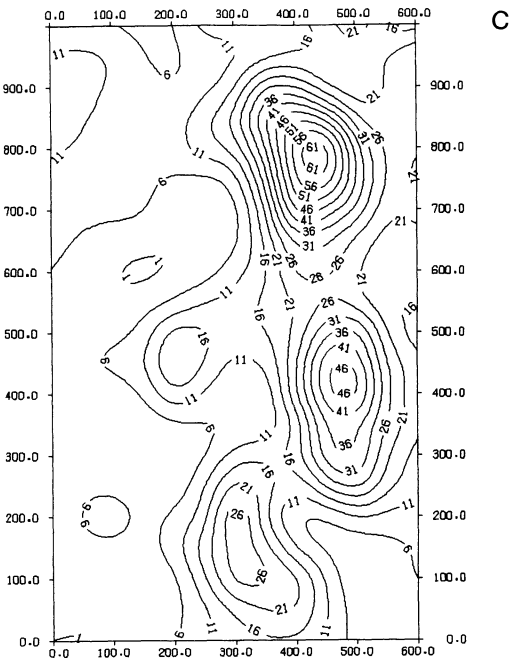
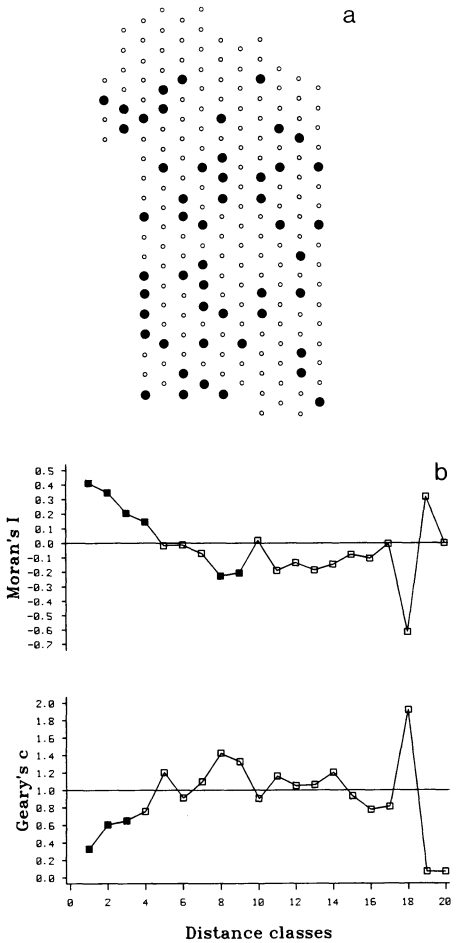


Fig. 1 a. Position of the 200 vegetation quadrats, systematically sampled in southwestern Québec during the summer of 1983. b. Spatial correlograms of the sugar-maple densities (*Acer saccharum* L.). Abscissa: distance classes; the width of each distance class is 57 m. Ordinate: Moran's *I* and Geary's *c* statistics. Black squares represent significant values at the $\alpha = 5\%$ level, before applying the Bonferroni correction to verify the overall significance of the correlograms; white squares are non-significant values. c. Contour map of sugar-maple densities obtained by kriging. Abscissa and ordinate are in m. From Fortin (1985).

identified a species level, were tallied in classes of 5 cm. In this study, we use only the sugar-maple (*Acer saccharum* L.) tree densities, sampled in these 200 quadrats.

This sugar-maple tree density data set has been used as the reference in this study. Subsamples were drawn from the set of 200 quadrats, in order to examine if the spatial analysis methods manage to identify or reconstruct spatial structures correctly using fewer data points. Since the reference locations of the quadrats follow a systematic sampling design and the smallest distance among quadrats is 50 m, so the smallest distance available in the subsamples is also 50 m. In fact, since the smallest sample size recommended for correlogram analysis is around 30, and we want to study the behaviour of the spatial analysis methods when given less than 100 observations, subsample sizes of 50 and 64 quadrats were used. These two subsample sizes were the ones that we could fit onto the reference grid (Fig. 1a) using systematic sampling. Since we were not able to construct other replicates of the systematic subsample, we replicated none of the three subsampling designs (below) and were left with a single study of the behaviour of the spatial analysis methods, which of course forbids any statistical testing of the observed differences.

Three subsampling designs were used to compare the ability of the methods to detect spatial patterns. The simple random sampling and systematic sampling designs were used because they



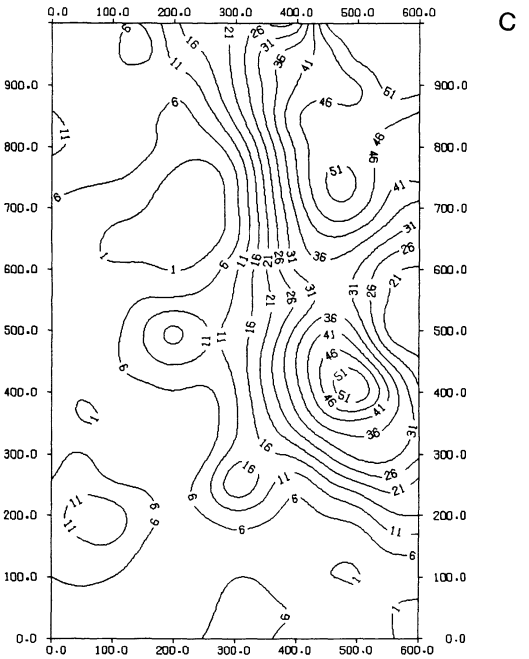
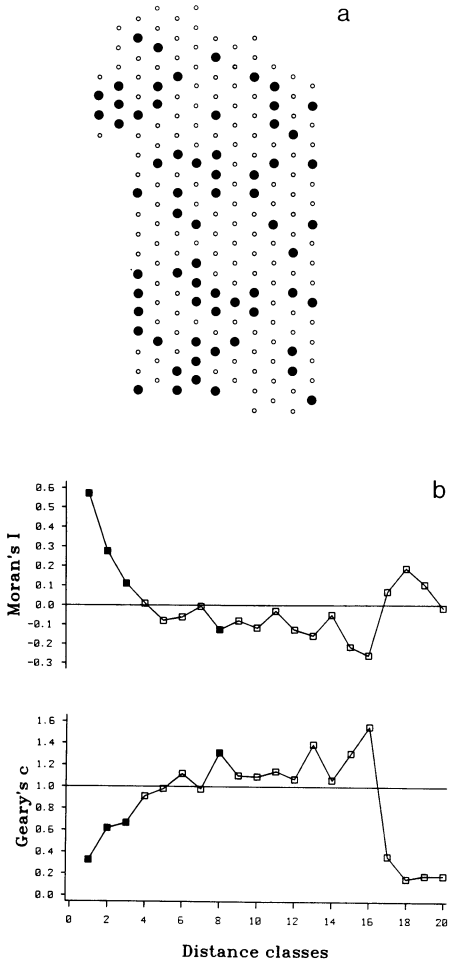


Fig. 2 a. Position of the 50-point random subsample. b. Spatial correlograms of the sugar-maple densities (*Acer saccharum* L.). Abscissa: distance classes; the width of each distance class is 51 m. Ordinate: Moran's *I* and Geary's *c* statistics. Symbols as in Fig. 1 c. Contour map of sugar-maple densities obtained by kriging. Abscissa and ordinate are in m. From Fortin (1985).

are often favoured by ecologists, not requiring any previous knowledge about the spatial distribution of the data. Systematic-cluster sampling design was also used because it allows different sampling steps to be present in the same data set (see below). The simple random samples of 50 and 64 quadrats were drawn at random from the reference set of 200 quadrats using subroutine GGSRS of the ISML subroutine package (Figs 2a and 3a). The systematic and systematic-cluster samples were designed by hand to fit the map and are shown in Figs 4a, 5a, 6a and 7a. With such small subsample sizes, some methods such as two-dimensional spectral analyses (Renshaw & Ford 1984; Legendre & Fortin 1989) could not be used.

From the sugar-maple reference data set of 200 samples, both Moran's *I* and Geary's *c* coefficients were computed; 20 distance classes, each 57 m wide, were used to construct the correlograms (Fig. 1b). These two correlograms were used as references for comparisons with the correlograms obtained from the subsamples (computed also with 20 distance classes). The reference data set (200 quadrats) was also used to compute 273 interpolated values by local kriging (13 columns × 21 rows, 50 m apart) under a spherical model, basing the interpolation at each point on the 25 neighbouring points (see Journel & Huijbregts 1978 for details); this produced the reference contour map of sugar-maple densities (Fig. 1c). The maps obtained from the different subsamples, also using 273 interpolated values, were then compared to this reference map.



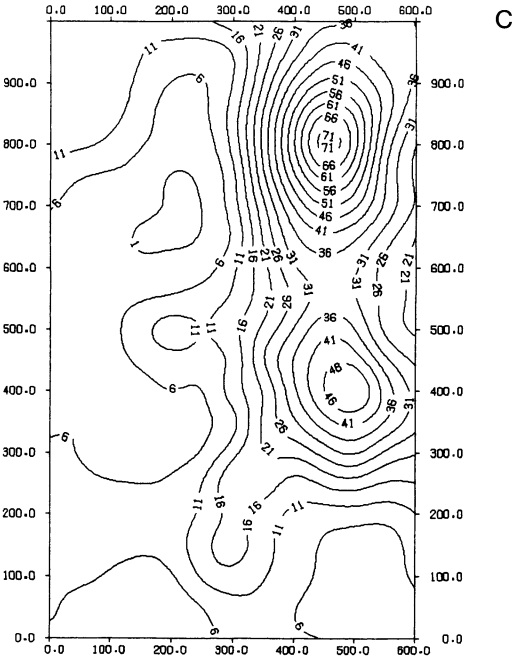
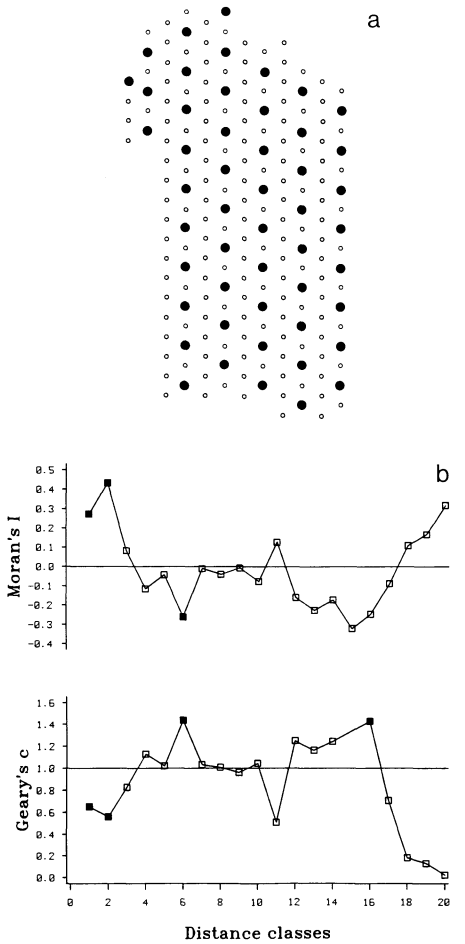


Fig. 3 a. Position of the 64-point random subsample. b. Spatial correlograms of the sugar-maple densities (*Acer saccharum* L.). Abscissa: distance classes; the width of each distance class is 55 m. Ordinate: Moran's *I* and Geary's *c* statistics. Symbols as in Fig. 1 c. Contour map of sugar-maple densities obtained by kriging. Abscissa and ordinate are in m. From Fortin (1985).

Results and discussion

The spatial correlograms computed for the random and systematic subsamples are presented in Figs 2b, 3b, 4b and 5b. Most Moran's *I* correlograms were globally significant, since they possessed at least one value that was significant at the Bonferroni-corrected level $\alpha' = 5\%/$number of simultaneous tests (Legendre & Fortin 1989; Oden 1984). The only one that did not pass this global and rather stringent test of significance was the correlogram resulting from 64 systematic points. The only Geary's *c* correlogram that was globally significant (Bonferroni-corrected test) was from the 64-point random subsample, the others (except Fig. 5b) showing significance for individual values only. The correlogram obtained from the 64-point systematic subsampling was almost a perfect example of a random pattern, demonstrated by an absence of significant autocorrelation at any distance class. Only the correlograms (Moran's *I* and Geary's *c*) computed from 64 random points (Fig. 3b) had shapes similar to the reference correlograms (Fig. 1b; discussion below), although almost all subsample correlograms had their highest coefficient values in the first distance class.

Generally for our subsamples, Moran's *I* detected significant spatial autocorrelation more efficiently than Geary's *c*. Considering the fact that the variable's spatial distribution is non-stationary in our data – Fig. 1c shows bumps in



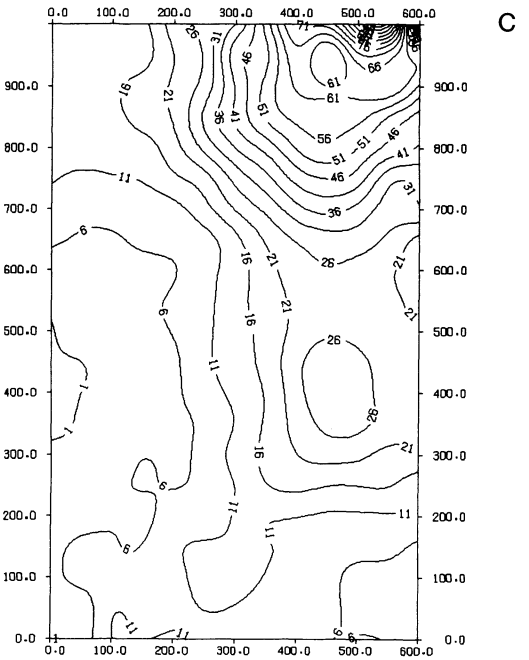


Fig. 4 a. Position of the 50-point systematic subsample. b. Spatial correlograms of the sugar-maple densities (*Acer saccharum* L.). Abscissa: distance classes; the width of each distance class is 55 m. Ordinate: Moran's *I* and Geary's *c* statistics. Symbols as in Fig. 1 c. Contour map of sugar-maple densities obtained by kriging. Abscissa and ordinate are in m. From Fortin (1985).

its right-hand part, and a relatively flat area on the left, this result may indicate a greater power of Moran's *I* test of significance to detect the presence of autocorrelation when the condition of stationarity, or the intrinsic hypothesis (which is a relaxed form of the stationarity hypothesis), is violated; this result should be checked by Monte-Carlo simulations.

Since the variograms and Geary's *c* correlograms are two distance-type coefficients, the variograms are not be presented here, while the interpolated maps produced by local kriging are. As in the spatial correlograms, the interpolated maps derived from the random subsampling designs (50 and 64 points) brought out the most important features of the spatial structure (Figs 2c and 3c), since the three high-density patches were in approximately the same position as on the reference map (Fig. 1c). With systematic subsampling (Figs 4 and 5), only the 50-point subsample detected a spatial structure (Fig. 4c); this pattern is somewhat distorted compared to the reference map. As it was the case for the spatial correlograms, the 64-point systematic subsample led to a flat variogram displaying no spatial structure, so that only a flat map could have been produced by kriging.

In both the spatial correlograms and the kriging methods, it is not the number of points that seems to make the difference, but rather their relative location in space. Kriging was very good at reconstructing maps of spatially autocorrelated variables even when the variogram, like Geary's *c* correlogram, displayed only weak evidence of spatial

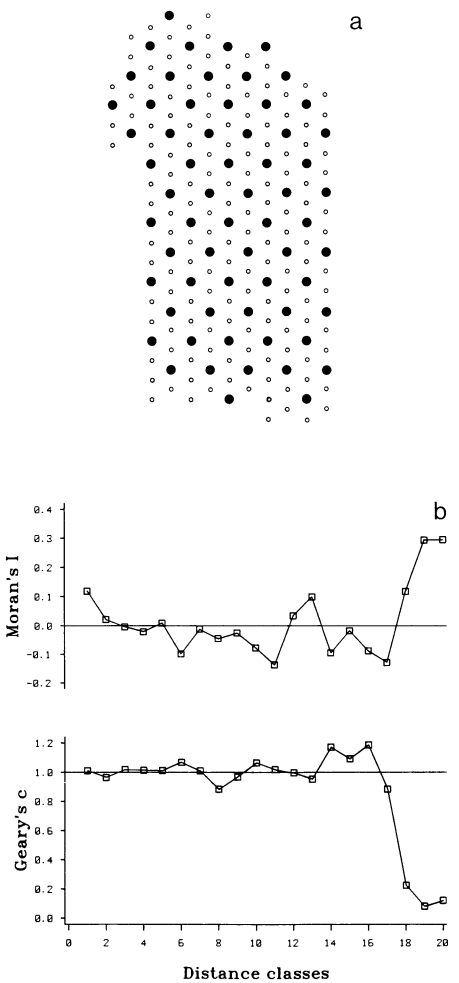
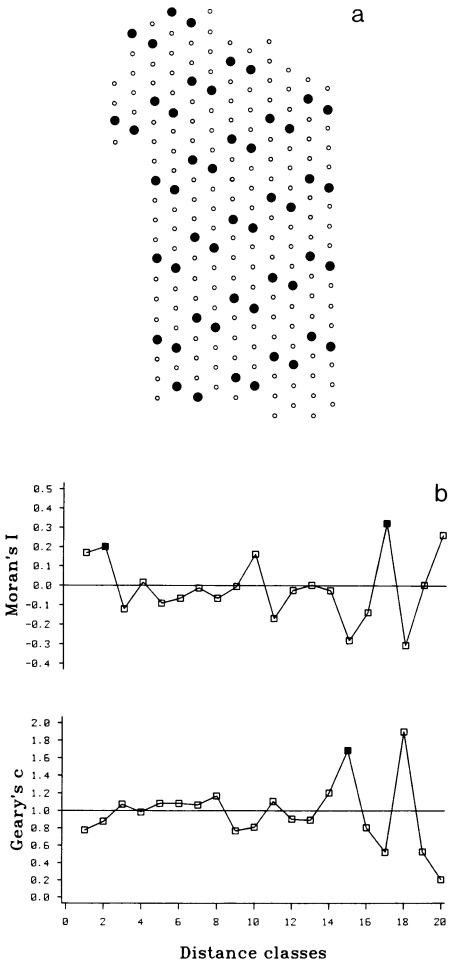


Fig. 5 a. Position of the 64-point systematic subsample. b. Spatial correlograms of the sugar-maple densities (*Acer saccharum* L.). Abscissa: distance classes; the width of each distance class is 55 m. Ordinate: Moran's *I* and Geary's *c* statistics. Symbols as in Fig. 1.

autocorrelation. In our simple random subsamples, the closest points were 50 m part, while in the systematic subsamples the smallest distance between points was 100 m. The simple random samples seem to carry more information about the spatial structure than the systematic subsamples, because they are more likely to contain different lag steps which may reflect several different harmonics of the spatial pattern; this is not the case with systematic sampling. So with an aggregated spatial pattern, the same number of points can lead to a better analysis or reconstruction of the spatial structure when they are not evenly spaced, as it was the case in our random subsamples; if the sampling step of a systematic subsample is too large to detect the spatial pattern, or if the location of the samples is not in phase compared to the existing spatial structure, the analysis can miss the spatial structure completely. Because the 50 points in our systematic subsampling design are in phase with the spatial structure, the methods were able to detect significant spatial autocorrelation and to reconstruct a meaningful (although distorted) map by kriging, while the 64-point systematic subsampling did not lead to the same result, despite the fact that it contained more data points.

Following these considerations, we decided to try a systematic-cluster sampling design with 50 and 64 points (Figs 6a and 7a). This new sampling design contained clusters of two samples, located 50 m apart; the clusters themselves were spaced 100 m from one another. The idea behind this was to capture different lag harmonics of a spatial structure, when no prior information was available about it, without the difficulties involved in implementing a random sampling design. In the same way, Oliver & Webster (1986) suggested an unbalanced nested design to capture the spatial variation at different scales of observation. Spatial autocorrelation coefficients and kriging were computed for these new subsamples (Figs 6b and 7b). The only correlogram with overall significance (Bonferroni-corrected test) was Moran's *I* for 64 points, which had the same general shape and intensity as the reference correlogram (Fig. 1b). The contouring map interpolated by kriging from the 64 points gave a map more similar to the reference than the one from 50 points (Fig. 6c). Both the 50-point and 64-point interpolated maps (Figs 6c and 7c) were rather good when compared to the reference map (Fig. 1c), and represented the sugar-maple spatial structure far better than the map obtained after systematic subsampling (Fig. 4c).

UPGMA classification of the Moran's *I* correlograms was performed, as suggested by Sokal (1986), to measure similarity among the sub-



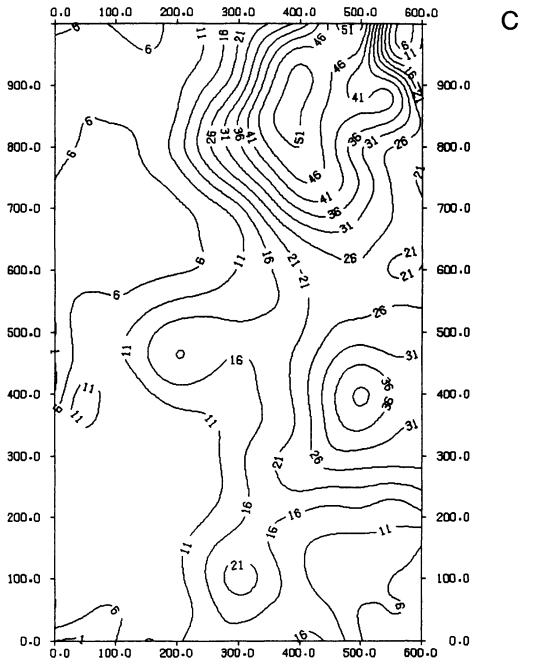
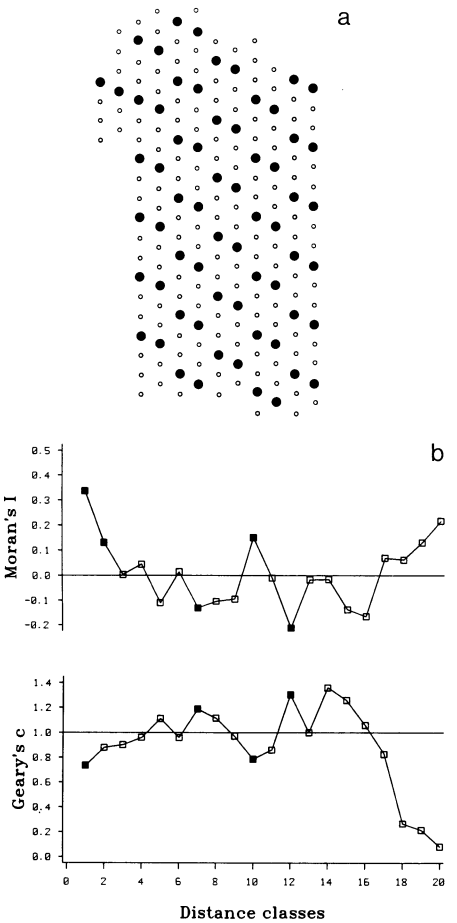


Fig. 6 a. Position of the 50-point systematic-cluster subsample. b. Spatial correlograms of the sugar-maple densities (*Acer saccharum* L.). Abscissa: distance classes; the width of each distance class is 53 m. Ordinate: Moran's *I* and Geary's *c* statistics. Symbols as in Fig. 1 c. Contour map of sugar-maple densities obtained by kriging. Abscissa and ordinate are in m. From Fortin (1985)

sample correlograms, the reference correlogram, and a flat correlogram in which all Moran's *I* values are equal to 0.0. The UPGMA classification (Fig. 8) was based upon a Manhattan distance matrix computed among correlogram value vectors. Two distinct groups of correlograms were found: a first group with the reference correlogram (200 points), the 64-point systematic-cluster sampling and the 64-point random sampling correlograms; and a second group with the flat correlogram and the 64-point systematic design. The 50-point systematic-cluster, the 50-point systematic and the 50-point random sampling correlograms did not form clusters. This classification showed that the 64-point random and systematic-cluster designs were the subsampling designs most efficient in reproducing the spatial structure of the 200-point reference data set, while the 64-point systematic design was the worst. Shape differences between the last three correlograms explain why they did not cluster with the reference correlogram or with the flat correlogram. In fact, both the 50-point systematic-cluster and the 50-point systematic correlograms had lower values of autocorrelation in the first than in the second distance class, while all other correlograms had higher values in the first distance class. The 50-point random sampling correlogram differed from all others in that it contained the longest sequence of significant positive values for Moran's *I* coefficient.

It would be interesting to compare statistically



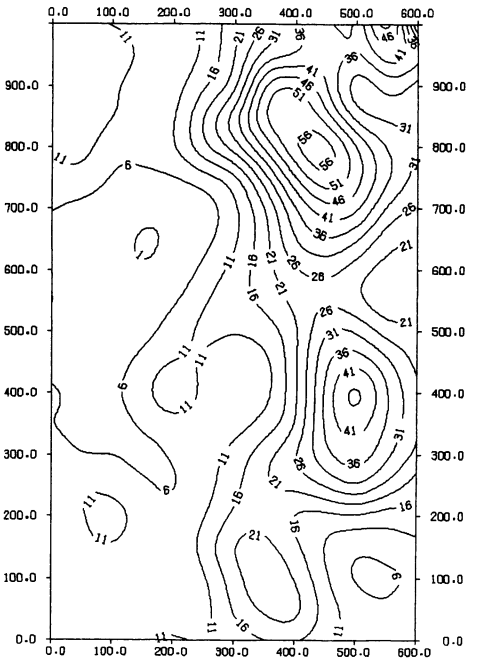


Fig. 7 a. Position of the 64-point systematic-cluster subsample. b. Spatial correlograms of the sugar-maple densities (*Acer saccharum* L.). Abscissa: distance classes; the width of each distance class is 53 m. Ordinate: Moran's *I* and Geary's *c* statistics. Symbols as in Fig. 1 c. Contour map of sugar-maple densities obtained by kriging. Abscissa and ordinate are in m. From Fortin (1985)

the interpolated values of the various subsampling designs to the reference interpolated values. This might be done by comparing the 273 interpolated values on the various maps obtained by kriging, since these 273 locations are the same on all maps. However, since we had only one replicate of each subsampling, confidence intervals cannot be computed; true testing would have required more subsamples for each design. So, the comparison was done by computing Spearman's correlation coefficients between the 273 interpolated values of each subsample map and those of the 200-point reference map. Spearman's coefficient was used here only as a measure of the resemblance between sets of interpolated points; since each subsample is drawn from the full set of 200 points and thus is not independent from it, these coefficients were not tested for significance. The results (Table 1) show that there are three different qualities of reconstruction, differing both by the sampling design and by the number of samples. As mentioned above, only a flat map could have been produced by kriging for the 64-point systematic sampling, so this case is excluded from the comparison. The worst recon-

Table 1. Spearman's *r* for pairwise comparisons between the interpolated values for each of the subsample maps and the interpolated values for the reference map.

Subsample	Spearman's *r*
64 systematic-cluster	0.8950
64 random	0.8761
50 systematic-cluster	0.8349
50 random	0.8200
50 systematic	0.7825

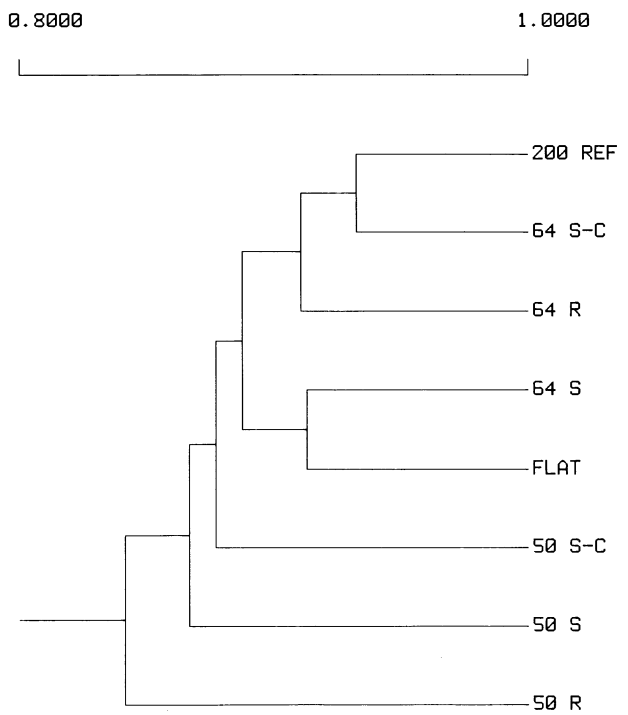


Fig. 8. Dendrogram of the UPGMA classification of the various Moran's *I* correlograms: the reference correlogram (200 REF), the different subsamples' correlograms (R = random, S = systematic, S-C = systematic-cluster), and a random structure correlogram (FLAT).

struction was from the 50-point systematic subsample; next come the reconstructions from the two other 50-point subsamples, random and systematic-cluster; the best reconstructions were from the random and systematic-cluster 64-point subsamples. These results agreed with the findings of the comparison of correlagrams, for the best types of sampling designs. On the other hand, kriging did better than spatial autocorrelation coefficients with the 50-point random and systematic-cluster subsamples, since one can recognize the major features of Fig. 1c on Figs 2c and 6c; so, kriging seems to be less affected than spatial autocorrelation coefficients by small sample sizes.

Conclusion

The first conclusion that can be drawn from our subsampling experiments is that the type of sampling design is very important for the accuracy of the detection of spatial patterns both by spatial autocorrelation coefficients and by kriging, and that sample size can be critical for spatial autocorrelation coefficients. We have shown in particular that sampling designs that draw information at several spatial scales, such as our random or systematic-cluster designs, can bring out more information about the spatial structure than a systematic design. The problem with a systematic design may be the inadequacy of the sampling step, or the fact that the samples are out of phase with the existing spatial structure. In any case, when no previous knowledge of the spatial structure is available, a sampling design using several different sampling steps is to be recommended. This conclusion has also been reached by Boehm (1967), Podani (1984) and Oliver & Webster (1986).

Our second conclusion is that Moran's *I* is more sensitive and efficient at detecting spatial autocorrelation than is Geary's *c*, at least with non-stationary data. Indeed, Moran's *I* correlograms displayed a significant spatial structure for most of the subsamples (except in one of the systematic designs), while Geary's *c* correlograms failed to do so in most instances. This result should be checked by Monte-Carlo simulations.

Acknowledgements

This is publication No. 706 in Ecology and Evolution from the State University of New York at Stony Brook, and contribution No. 353 from the Groupe d'Écologie des Eaux douces, Université de Montréal. We are thankful to Junhyong Kim and to Cajo J.F. ter Braak for helpful comments. Dr. Michel David, École polytechnique de Montréal, gave us instructions for and access to his GEOSTAT computer package, that we used for kriging. Alain Vaudor, computer analyst, developed part of the programs of 'The R Package

for Multivariate Data Analysis' during and for the present study. This study was supported by NSERC grant No. A7738 to P. Legendre, and by a NSERC scholarship to M.-J. Fortin.

References

Boehm, B.W. 1967. Tabular representation of multivariate functions – with applications to topographic modelling. Report RM-4636-PR, Rand Corporation, Santa Monica, California.

Bouchard, A., Bergeron, Y., Camiré, C., Gangloff, P. & Gariépy, M. 1985. Proposition d'une méthodologie d'inventaire et de cartographie écologique: le cas de la MRC du Haut-Saint-Laurent. Cah. Géogr. Qué. 29: 79–95.

Bouchon, J. 1974. Utilization of regionalized variables in forest inventories. IUFRO and SAF Meeting, June 20-26, Syracuse, New York.

Bouxin, G. & Gauthier, N. 1982. Pattern analysis in Belgian limestone grasslands. Vegetatio 49: 65–83.

Burrough, P.A. 1987. Spatial aspects of ecological data. In: Jongman, R.H.G., ter Braak, C.J.F. & van Tongeren, O.F.R. (eds), Data analysis in community and landscape ecology, pp. 213–251. Centre for Agricultural Publishing and Documentation, Wageningen.

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

Cochran, W.G. 1977. Sampling techniques, 3rd ed. John Wiley & Sons, New York.

David, M. 1977. Geostatistical ore reserve estimation. Developments in Geomathematics, 2. Elsevier, Amsterdam.

Fortin, M.-J. 1985. Analyse spatiale de la répartition des phénomènes écologiques: méthodes d'analyse spatiale, théorie de l'échantillonnage. Mémoire de Maîtrise ès Sciences, Université de Montréal.

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

Gloaguen, J.C. & Gauthier, N. 1981. Pattern development of the vegetation during colonization of a burnt heathland in Brittany (France). Vegetatio 46: 167–176.

Green, R.H. 1979. Sampling design and statistical methods for environmental biologists. John Wiley & Sons, New York.

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

Greig-Smith, P. 1964. Quantitative plant ecology, 2nd ed. Butterworth, London.

Greig-Smith, P. 1979. Pattern in vegetation. J. Ecol. 67: 755–779.

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

Jumars, P.A. 1978. Spatial autocorrelation with RUM (Remote Underwater Manipulator): vertical and horizontal structure of a bathyal community. Deep-Sea Research 25: 589–604.

Legendre, L. & Legendre, P. 1984. Écologie numérique. 2ième ed. Tome 2: La structure des données écologiques. Masson, Paris et les Presses de l'Université du Québec.

Legendre, P. 1985. The R package for multivariate data analysis. Département de sciences biologiques, Université de Montréal.

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

Legendre, P. & Troussellier, M. 1988. Aquatic heterotrophic bacteria: modeling in the presence of spatial autocorrelation. Limnol. Oceanogr. 33: 1055–1067.

Legendre, P., Troussellier, M., Jarry, V. & Fortin, M.-J. 1989. Design for simultaneous sampling of ecological variables: from concepts to numerical solutions. Oikos (in press).

Marbeau, J.-P. 1976. Géostatique forestière, état actuel et développements nouveaux, pour l'aménagement en forêt tropicale. Thèse de Doctorat, École Nationale Supérieure des Mines de Paris, Centre de Géostatique et de Morphologie Mathématique, Fontainebleau.

Matheron, G. 1973. The intrinsic random functions and their applications. Adv. Appl. Prob. 5: 439–468.

McBratney, A.B. & Webster, R. 1986. Choosing functions for semi-variograms of soil properties and fitting them to sampling estimates. J. Soil Sci. 37: 617–639.

McBratney, A.B., Webster, R. & Burgess, T.M. 1981. The design of optimal sampling schemes for local estimation and mapping of regionalized variables. I. Theory and methods. Comp. Geosci. 7: 331–334.

McCall Jr., C.H. 1982. Sampling and statistics handbook for research. Iowa State Univ. Press, Ames, Iowa.

Minchin, P.R. 1987. An evaluation of the relative robustness of techniques for ecological ordination. Vegetatio 69: 89–107.

Mohler, C.L. 1981. Effects of sample distribution along gradients on eigenvector ordination. Vegetatio 45: 141–145.

Mohler, C.L. 1983. Effect of sampling pattern on estimation of species distribution along gradients. Vegetatio 54: 97–102.

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

Podani, J. 1984. Analysis of mapped and simulated vegetation patterns by means of computerized sampling techniques. Acta Bot. Hung. 30: 403–425.

Podani, J. 1987. Computerized sampling in vegetation studies. Coenoses 2: 9–18.

Oden, N.L. 1984. Assessing the significance of a spatial correlogram. Geogr. Anal. 16: 1–16.

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

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

Sakai, A.K. & Oden, N.L. 1983. Spatial pattern of sex expression in silver maple (*Acer saccharinum* L.): Morisita's index and spatial autocorrelation. Am. Nat. 122: 489–508.

Scherrer, B. 1982. Techniques de sondage en écologie. In: Frontier, S. (ed.), Stratégies d'échantillonnage en écologie. Collection d'Écologie, 17, pp. 63–162. Masson, Paris et les Presses de l'Université Laval, Québec.

Scherrer, B. 1984. Biostatistique. Gaëtan Morin Editeur, Chicoutimi, Québec.

Sokal, R.R. 1979. Ecological parameters inferred from spatial correlograms. In: Patil, G.P. & Rosenzweig, M.L. (eds), Contemporary quantitative ecology and related ecometrics. Statistical Ecology Series, Vol. 12, pp. 167–196. Int. Co-operat. Publi. House, Fairland, M.D.

Sokal, R.R. 1986. Spatial data analysis and historical processes. In: Diday, E. *et al.* (eds), Data analysis and informatics, IV. Proceedings of the Fourth International Symposium on Data Analysis and Informatics, pp. 29–43. Versailles, France, 1985. North-Holland, Amsterdam.

Sokal, R.R. & Menozzi, P. 1982. Spatial autocorrelation of HLA frequencies in Europe support demic diffusion of early farmers. Am. Nat. 119: 1–17.

Sokal, R.R. & Oden, N.L. 1978. Spatial autocorrelation in biology. 1. Methodology. Biol. J. Linnean Soc. 10: 199–228.

Sokal, R.R. & Thomson, J.D. 1987. Applications of spatial autocorrelation in ecology. In: Legendre, P. & Legendre L. (eds), Developments in numerical ecology. NATO ASI Series, Vol. G 14, pp. 431–466. Springer-Verlag, Berlin.

Upton, G.J.G. & Fingleton, B. 1985. Spatial data analysis by example. Vol. 1: Point pattern and quantitative data. John Wiley & Sons, Chichester.

Webster, R. & Burgess, T.M. 1984. Sampling and bulking strategies for estimating soil properties in small regions. J. Soil. Sci. 35: 127–140.