Theor Appl Genet (1994) 88:33–41

P. Monestiez · M. Goulard · G. Charmet

Geostatistics for spatial genetic structures: study of wild populations of perennial ryegrass

Received: 7 June 1993 / Accepted: 16 June 1993

Abstract Methods based on geostatistics were applied to quantitative traits of agricultural interest measured on a collection of 547 wild populations of perennial ryegrass in France. The mathematical background of these methods, which resembles spatial autocorrelation analysis, is briefly described. When a single variable is studied, the spatial structure analysis is similar to spatial autocorrelation analysis, and a spatial prediction method, called "kriging", gives a filtered map of the spatial pattern over all the sampled area. When complex interactions of agronomic traits with different evaluation sites define a multivariate structure for the spatial analysis, geostatistical methods allow the spatial variations to be broken down into two main spatial structures with ranges of 120 km and 300 km, respectively. The predicted maps that corresponded to each range were interpreted as a result of the isolation-by-distance model and as a consequence of selection by environmental factors. Practical collecting methodology for breeders may be derived from such spatial structures.

Key words Perennial ryegrass · Population genetics · Geostatistics · Spatial autocorrelation

Introduction

In nature many phenomena show variations that are not randomly distributed in space but are spatially arranged or structured. Examples of these are common in the earth sciences. Geologists were the first to develop statistical methods, called geostatistics, adapted to this kind of variable (Matheron 1965). Geostatistical analysis, as simply described in Webster and Oliver (1990) consists of two main steps:

- modeling and identification of the spatial structure of the variable using variograms or spatial covariance functions.
- linear estimation or prediction of the variable everywhere in the studied space to obtain a cartography. This spatial prediction method, called "kriging", also gives a map of prediction variance. More complete technical descriptions of the methods and recent developments in geostatistics may be found in Journel and Huijbregets (1978) and in Cressie (1986, 1991).

Sokal and Oden (1978a, b) introduced similar approaches in biology, an example being the spatial autocorrelation analysis. In the field of population biology or evolution science, spatial autocorrelations may have different origins:

- a trait may be differentiated in response to an environmental gradient, leading to a clinal variation.
- in the same way, if environmental conditions vary patchily, such patches may also be found in the spatial distribution of adaptative traits.
- for selectively neutral traits, such as the allelic frequencies of isozymes, a limited gene flow, founder effects and isolation-by-distance generally lead to a decrease of genetic identity with distance between populations; Sokal and Wartenberg (1983) even demonstrated that a few generations of isolation allow a spatial structure of allelic frequency to be created.

Spatial structure analysis has been extensively applied in biology, mostly through the use of the spatial autocorrelation method (Sokal and Oden 1978a, b). This method leads to a graphic representation of a coefficient of genetic identity as a function of geographical distance either between individuals within a population or between populations. It has been applied to both morphological traits (Epperson and Clegg 1986) and to allelic frequencies of isozymes (Waser 1987; Dewey and Heywood 1988; Epperson and Allard 1989; Perry and Knowles 1991).

Other theoretical (Barbujani 1987) or simulation studies (Sokal et al. 1989; Epperson 1990) have been devoted to the problem of spatial structuring in population genetics.

Communicated by P. M. A. Tigerstedt

P. Monestiez (✉) · M. Goulard
INRA* Biométrie, 84140 Montfavet, France

G. Charmet
INRA Amélioration des Plantes, 63039 Clermont-Ferrand, France

* Institut National de la Recherche Agronomique

In the study reported here geostatistical methods were applied to quantitative traits of agricultural interest measured on a collection of 547 wild populations of perennial ryegrass from France. Adaptative traits like seasonal growth or reproductive characteristics are not uniformly distributed over space. The aim of our study was to determine the parameters of this underlying spatial structure and to deduce some of its implications in the spatial distributions of the ryegrass populations.

Materials and methods

Materials and experimental design

Seed samples of wild populations of perennial ryegrass were collected in France during the summers of 1983 and 1984. The framework was a cooperative program between private companies and INRA. Details of the collection design and evaluation procedure have been described in Charmet et al. (1990). Briefly, 226 populations were studied from 1984 to 1986 and another 321 populations from 1985 to 1987. All were evaluated in spaced plant nurseries, with three replicated blocks of 10 plants of every population at each of the six evaluation sites (Fig. 1). Ten traits of agricultural interest were scored on a single plant basis according to a 1–9 visual scale, except for heading date, which was scored in days from January 1st.

Since the empirical distributions of most traits fit the Gaussian distribution quite well, a variance analysis was performed on the original variables using the following model with fixed effects.

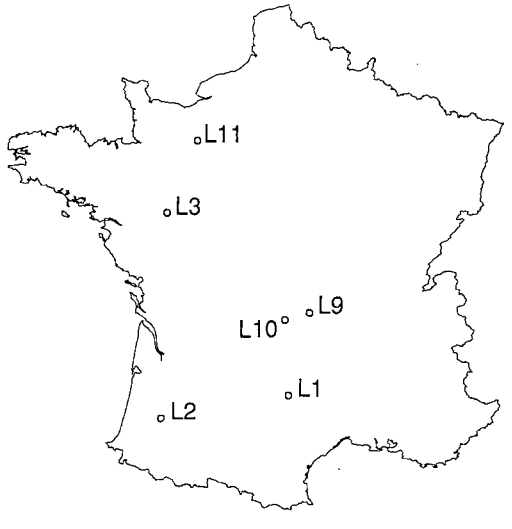
$$X = \mu + YE + LO + YE.LO + YE.LO.BL + PO + LO.PO + \varepsilon \quad (1)$$

where μ is the overall mean, *YE* the main year effect, *LO* the main evaluation site location effect, *YE.LO* the year × evaluation site interaction, *YE.LO.BL* the block effect (hierarchized in year and evaluation site location), *PO* the main population effect, *LO.PO* the evaluation site × population interaction and ε the residual error (i.e., mostly intra-population error).

The results of these analyses are described in Charmet et al. (1990). Most factors are highly significant. However, the variation associated with population main effect and population × evaluation site interaction is used to separate the traits associated with reproductive development from the growth or susceptibility traits. The first group generally has high ratios of main population effect over population × evaluation site interaction variation, and therefore interactions may be neglected. The second group shows higher degrees of interaction, which cannot be discarded in further analysis.

In this methodology-oriented paper, the use of geostatistics on one trait of each category is illustrated: heading date as an example of a "stable" trait and summer growth as an example of "interactive" one. Although summer growth does not have the lowest ratio of population effect/interaction, it was chosen because its variation is thought to be related to adaptative characteristics such as summer dormancy, and thus, to have a more pronounced spatial structure in relation to ecological factors.

Fig. 1 Map of the evaluation sites. *L1* Rodez, *L2* Mont-de-Marsan, *L3* Angers, *L9* Clermont-Ferrand, *L10* Bourg-Lastic, *L11* le Pin-au-Haras



Univariate geostatistical methods

Stochastic modelling is used to describe the spatial data. Suppose that the studied variable is a realization of a random field and that the observed data are a spatial sample on some sites of this realization. Let $X(y)$ denote the random field at location y. The data are given by the $X(y_i)$'s where the y_i's, for $i = 1, \ldots, n$, denote the location of the n sampled sites.

In order to carry out estimation from a single realization of the random field, the stationarity of order two is assumed, i.e., the first two moments are supposed to be invariant by translation. In mathematical terms, we have:

$$\mathsf{E}[X(y)] = m \text{ where } m \text{ is independent of } y \quad (2)$$

$$\mathsf{E}[(X(y+h)-m)(X(y)-m)] = C(h) \quad (3)$$

$$\frac{1}{2}\mathsf{E}[(X(y+h)-X(y))^2] = G(h) \quad (4)$$

Equations 3 and 4, where $y + h$ is the translation of y and E denotes the expectation, define the covariance function $C(h)$ and the variogram $G(h)$ respectively. The two expectations depend only on the vector h. If the distribution is Gaussian, these two functions and the mean m give a complete characterization of the random field. Also, it follows directly that:

$$G(h) = C(0) - C(h) = \sigma^2 - C(h) \quad (5)$$

To describe the spatial distribution of the variable, one of the two functions is estimated from the set of data. For example, the variogram may be estimated by:

$$\begin{cases} g^*(h) = \dfrac{1}{W(h)} \sum\limits_{i<l} w(y_i - y_l, h)(x(y_i) - x(y_l))^2 \\ \text{with} \quad W(h) = \sum\limits_{i<l} w(y_i - y_l, h) \end{cases} \quad (6)$$

where $w(u, h)$ is a function of the proximity between h and u. For example $w(u, h) = 1$ if the distance between h and u is less than some ε; otherwise, it is zero. For the isotropic case, h is just replaced by d and w is a function both of d and the distance between the two points. The properties of the estimator are given in the Gaussian case in Journel and Huijbregts (1978). Note that $g^*(h)$ is an unbiased estimate of $G(h)$ even if m is unknown.

Structural analysis consists in describing and modeling the estimated variogram $g^*(h)$. Variograms are characteristic of the relations of dependence existing between sites. Classically, three different items are described by such functions:

- First, the discontinuity at the origin. The variogram equals zero at the origin by definition. For a distance close to zero, a significant value usually called "nugget effect" is often observed. This initial variance can be seen as measurement error and microscale variation.
- Second, the variogram usually increases with increasing distances. The shorter the distances between the sites, the more dependent the observations.
- Third, the "sill" is the steady value reached by the function. When the distance between sites is larger than a certain value, the variogram function becomes constant and the sites should be considered to be independent. This distance value, called the range, is an

essential parameter in spatial description. The value of the sill is close to that of the global population variance. When the distance values are highest, random fluctuations may appear on the estimated variogram because the accuracy in the estimation is at its lowest.

The modelling of the variogram leads to the second phase, which consists of the spatial prediction – or spatial interpolation – of the variable between the observed sites. Suppose we want to estimate the value $X(y_0)$ of the random field at a site y_0. Let our estimate be a linear expression – or a weighted average – of the observed values:

$$X_0^* = \sum_{i=1}^{n} \lambda_i X(y_i), \tag{7}$$

where the λ_i are coefficients or weights chosen so that the error of prediction is minimal. The weights take into account the known spatial dependences expressed in the variogram $G(h)$ and the geometric relationships among the observed sites. This prediction, known as "kriging", is the best linear unbiased prediction based on the observed values.

Multivariate spatial analysis

It is assumed here that a data table was assembled such that the rows represent populations collected over the 547 sites and the columns the observations of a variable at the six evaluation sites. For example, the observations of "summer growth" in the six evaluation sites are considered to be six different variables. Each variable under study is taken as a realization of some random field. These random fields are denoted by $X^j, j = 1, \ldots, p$, where p is the number of observed variables. $X^j(y)$, the value of the random field at site of coordinates y, is a random variable. Each realization is a surface that is only partially sampled at n sites, the value of the random variables being $X^j(y_i), i = 1, \ldots, n$. When the random fields are grouped into a vector, where each component is a random field, and if it is assumed that this vector random field is stationary up to the second order, then:

$$\mathrm{E}[X^j(y)] = m^j \text{ with } m^j \text{ independent of } y \tag{8}$$

$$\mathrm{E}[(X^j(y+h) - m^j)(X^k(y) - m^k)] = C^{jk}(h) \tag{9}$$

$$\frac{1}{2}\mathrm{E}[(X^j(y+h) - X^j(y))(X^k(y+h) - X^k(y))] = G^{jk}(h) \tag{10}$$

The last two expectations depend only on the vector h, where $y + h$ is the translation of y. C^{jk} is called the cross-covariance function and G^{jk} is the cross-variogram. When $j = k$, the two functions are the covariance and variogram functions, respectively, introduced in the previous section. As in one dimension, they tell us about the multidimensional spatial structure. It may be noted that the equality:

$$G^{jj}(h) = C^{jj}(0) - C^{jj}(h) \tag{11}$$

is not true for $j \neq k$ because:

$$G^{jk}(h) = C^{jk}(0) - \frac{1}{2}(C^{jk}(h) + C^{jk}(-h)) \tag{12}$$

Moreover, the cross-variogram may exist when the cross-covariance does not exist. From now on, only the cross-variogram will be used even if it means limiting the observations to symmetrical relationships. When the dependence relation is isotropic, the cross-variogram depends on the distance d between y and $y + h$.

The cross-variograms can be used to obtain the spatial relationships. They can be estimated by:

$$g^{jk}(h) = \frac{1}{W(h)} \sum_{i<l} w(y_i - y_l, h)(x^j(y_i) - x^j(y_l))(x^k(y_i) - x^k(y_l)) \tag{13}$$

where $w(u, h)$ and $W(h)$ are defined as in Eq. 6. The structural analysis of this cross-variogram is more difficult to perform than in the one-dimensional case because of the number of estimated curves to interpret simultaneously.

A model of the cross-variograms seems to be a convenient manner by which to summarize the multivariate spatial dependences. This model must satisfy some strong constraints in order to be a cross-variogram model. The linear coregionalization model for which

$$G^{jk}(h) = S_1^{jk} g_1(h) + \ldots + S_r^{jk} g_r(h) \tag{14}$$

where the functions g_t are variograms and the matrices S_t of elements S_t^{jk} are positive definite, is an adequate model that is often used in soil science (Goulard and Voltz 1992; Wackernagel 1988). The basic physical idea is that the variables under study are generated by different processes acting additively with various specific spatial structures. The matrices tell us about the relationships between variables for the specific underlying processes. The variogram function g_t can be chosen from the physical model of the phenomenon or by inspection of the behaviour of the variograms. The fit of the model is then done by a least-squares procedure. Afterwards, each matrix is analyzed as a variance-covariance matrix of variables.

As for the one-dimensional case, the cross-variograms lead to cokriging, which is the multivariate version of kriging. One component of the vector random field is predicted at a location using the observed values of all components at observed sites. If it is assumed that the component X^j for $j = 1, \ldots, p$ at sites y_i for $i = 1, \ldots, n_j$ is known, cokriging a component X^l at site y_0 consists in searching for the best unbiased linear estimator:

$$X^l(y_0)^* = \sum_{j=1}^{p} \sum_{i=1}^{n_j} \lambda_{ij} X^j(y_i) \tag{15}$$

The scalars λ_{ij} are chosen so that they satisfy specific linear constraints, so that the prediction is unbiased and the variance of prediction error minimal.

The model may also be viewed as a decomposition of each variable under study on r specific multivariate spatial structures. The structures are themselves decomposed on p spatial components, which are all mutually independent:

$$X^j(y) = \sum_{t=1}^{r} \sum_{k=1}^{p} A_t^{jk} Z_t^k(y) \tag{16}$$

where the Z_t^k independent random fields are obtained by using a special case of cokriging (Goulard and Voltz 1992) and the A_t^{jk} matrices result from a decomposition of the covariance matrix S_t^{jk} in factors.

$$S_t^{jk} = \sum_{l=1}^{p} A_t^{jl} A_t^{kl} \tag{17}$$

The factors are worked out using an inertia criterion as in principal component analysis. An effective summary of the multivariate covariance structure is to describe the two or three main spatial components that are associated to the factors with larger inertia. A map can be obtained for every spatial component Z_t^k related to each factor or at least to the principal ones. This has been called in geostatistics the kriging analysis.

As in principal component analysis, the relationship between original variables and spatial components can be described for each spatial structure of variogram g_t using the correlation circle in the first two principal-factor planes.

Results

Variance analysis and year interaction

Since the samples of populations studied in 1984–1986 and 1985–1987 respectively, are not randomly located, the year effect and all interactions with year may produce a spatial structure that is artificial.

Such a structure would have shown discrepancies between predicted population means across the limits of the two sets of samples (border effect). This can be seen of Fig. 2a where the symbol sizes are proportional to on the mean values of summer growth at the evaluation site Rodez L1 (i.e., the whole model of Eq. 1 estimated for $LO = 1$) for the samples sown in 1984 (empty squares) and 1985 (filled squares), respectively:

the former has on average greater values of summer growth score, and a marked border effect can be observed. This effect appears to be fully corrected for on Fig. 2b, which presents the estimates $\mu + PO + LO.PO$ from the analysis of variance of summer growth for $LO = 1$ (Rodez).

To verify that the year correction does not add any artificial spatial structure, the kriging analysis was carried out on the data obtained from the 1984 and 1985 nurseries, considering only pairs of population sites from within a common evaluation test. All of the results were similar to those obtained from the full pair set: considering pairs of populations with a different evaluation time did not alter the spatial structures. The correction of year effect through the analysis of variance can thus be considered as being appropriate for the following spatial analyses.

For heading date the $LO.PO$ interaction is not significant and pooled with the error so the correction of year effect, which was done by estimation of $\mu + PO$, seems also to be satisfactory, as shown in Fig. 2c. in order to take into account the correction for year main effects and interaction effects involving year, we used the estimated effects $\mu + PO$ (for heading date) and $\mu + PO + LO.PO$ (for summer growth) in the geostatistical analyses.

Spatial analysis of heading date

Because of the absence of evaluation site × population interaction, the population means can be analysed as an univariate spatial random field. Isotropy of the random field is assumed. The empirical variogram is presented in Fig. 3. In accordance with Eq. 6, the estimation of $g^*(h)$ has been performed for a series of h regularly spaced every 10 km and for a function $w(u, h)$; which equals 1 when the discrepancy between u and h is lower than 5 km and 0 elsewhere.

A spherical model with a nugget effect was fitted to the empirical variogram:

$$g(h) = \begin{cases} 10. + 32.\left(\frac{3}{2}\frac{h}{120} - \frac{1}{2}\left(\frac{h}{120}\right)^3\right) & \text{if } h < 120 \\ 42. & \text{if } h \geq 120 \end{cases} \quad (18)$$

where the nugget effect, i.e., the variance of the difference between adjacent populations ($h = 0$), equals 10 and where the range of the variogram is 120 km. The nugget effect represents the lower limit of the variance of heading date difference between adjacent populations. This variance is only 10 (i.e., a standard deviation of 3.2 days), about 25% of the variance of difference between samples that are more than 120 km apart.

The use of the fitted variogram allows the variable heading date to be interpolated everywhere by kriging, and then a map with curves of isovalue can be obtained from the sampled sites. The method filters the nugget effect term and shows only

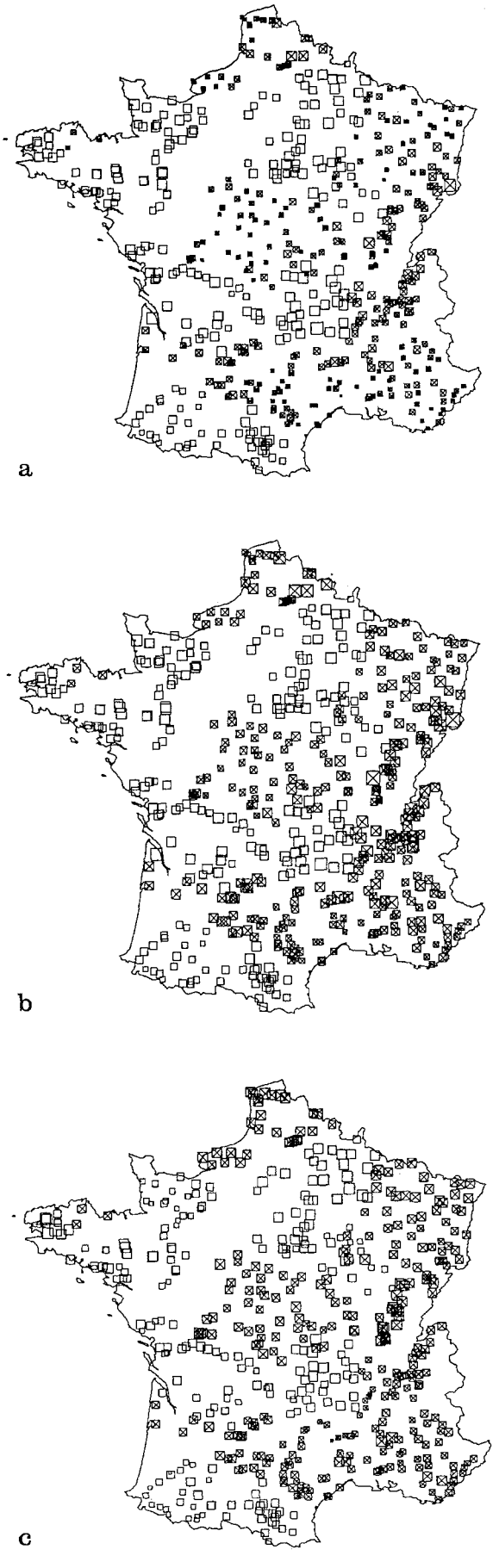


Fig. 2 **a** Map of "summer growth" measured at evaluation site *L1* for 547 collected populations. Markers are located at collection sites. Marker size is proportional to the variable. Populations evaluated in 1984 are depicted by *empty squares* and populations evaluated in 1985 by *filled squares*. **b** Map of "summer growth" after correction of the year effect. **c** Map of "heading date" averaged over the evaluation sites after correction of the year effect

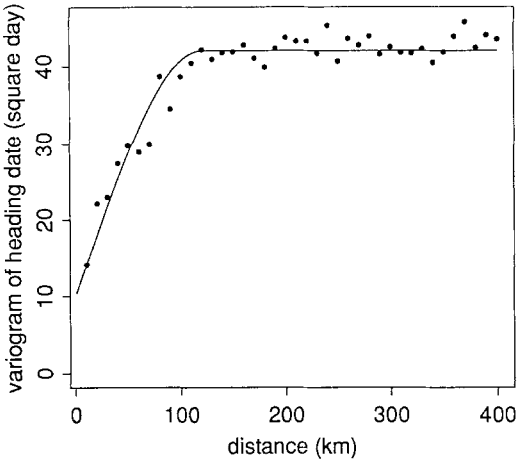


Fig. 3 Empirical variogram (*dots*) of "heading date" and fitted model according to equation 18 (*solid line*)

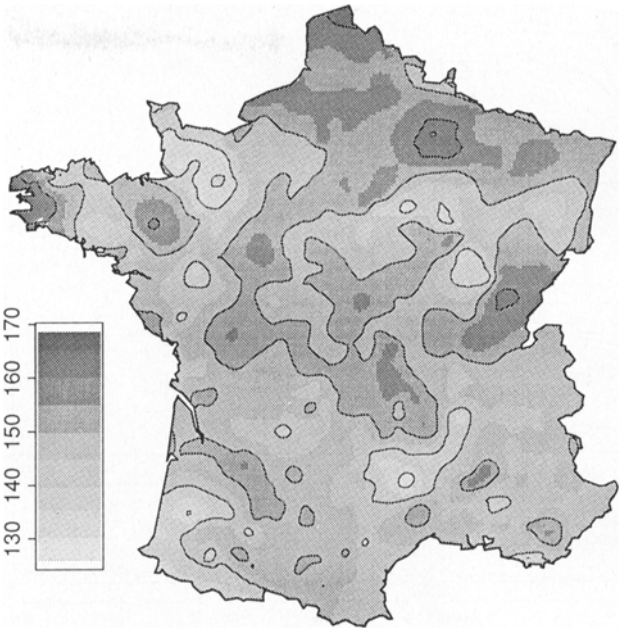


Fig. 4 Interpolated map resulting from the kriging of "heading date"

smoother spatial variations with a range of 120 km. The resulting map in Fig. 4 has to be compared with the primary data map of Fig. 2c. The two maps feature similar spatial variations, but local erratic variations have been removed from the kriged one. There is no obvious simple relation between the spatial variation of Fig. 4 and known climatic or environmental factors. The only pattern that may be interpreted is a North-South trend corresponding to a very long range variation not modeled by the variogram, but whose interpretation might be a climatic adaptation of population. This must not be confused with a direct climatic effect because all heading dates were measured at the same set of evaluation sites and there is no significant $LO.PO$ interaction.

Spatial analysis of summer growth

The variable "summer growth" is characterized by six variables L1–3 and L9–11, one for each evaluation site. Each variable represents the estimated effect $\mu + PO + LO.PO$ from model 1. The variable maps (not shown except for L1 in Fig. 2b) show significant spatial patterns with many variations from one map to another, so a multivariate geostatistical approach is used. Figure 5 displays the variograms of the six variables as well as the cross-variograms for every pair of variables. Isotropy of all the variables is assumed, so isotropic variograms and cross-variograms are employed. On this display the functions plotted are normalized using inverse variance of variables [i.e., $G_{ij}(h)$ is transformed to $G_{ij}(h)/(\sigma_i \sigma_j)$].

If we leave out the nugget effect, two kinds of spatial structures appear for summer growth in evaluation site variograms (diagonal plots), one with the 120-km range and the second with a 300-km range. The two structures seem to be additively mixed in various proportions depending on the evaluation sites. Most of the cross-variograms are quite flat (non-diagonal plots). Some cross-variograms, however, show significant interrelationships for distant pairs of populations, with also a maximum reached for distances of 120 km or 300 km.

The presence of significant covariances between spatial structures estimated from different evaluation sites requires a global fitting to be carried out on the whole set of empirical cross-variograms. A simple model was chosen with three additive spatial structures: a structured one that is a spherical model with a 120-km range, a second spherical model with a 300-km range and another that is spatially erratic (nugget effect). The model was fitted using a least-squares method. Results of the global fitting are illustrated on Fig. 5 by solid lines in the graphics.

In terms of variance decomposition, the parts of the two spherical structures are globally balanced and explain from 40% (evaluation site L11) to 80% (evaluation site L9) of the total variation (average 60%). The erratic term represents 40% of the total variation (Fig. 6). Subsequently, the decomposition in factors of the two spatial structures and the nugget effect, using a principal component analysis carried out on each S_t, was performed. Each spatial structure observed from six evaluation sites may be summarized by a smaller number of factors.

Figure 7 shows for each structure, i.e., 120 km, 300 km and erratic term, the relative contributions of the original variables, L1–3 and L9–11, to the first two factors. As for a principal component analysis, correlations between factors and variables are plotted inside a unit circle in the first factorial plane. The information contained in the whole scatter plot of Fig. 5 is summarized in these correlation plots. For example, the variables L9 and L10 are independent on the 300-km-range structure (orthogonal vector on the 300-km correlation plot) and positively correlated on the 120-km-range structure (acute angle between vectors). This is to represent that cross-variogram L9–L10 mainly features a positive dependence of the 120-km range, although both variograms L9 and L10 feature a 300-km-range structure. One can also note that variables L1–L2 are negatively correlated on the 120-km-range structure (vector in opposition) and

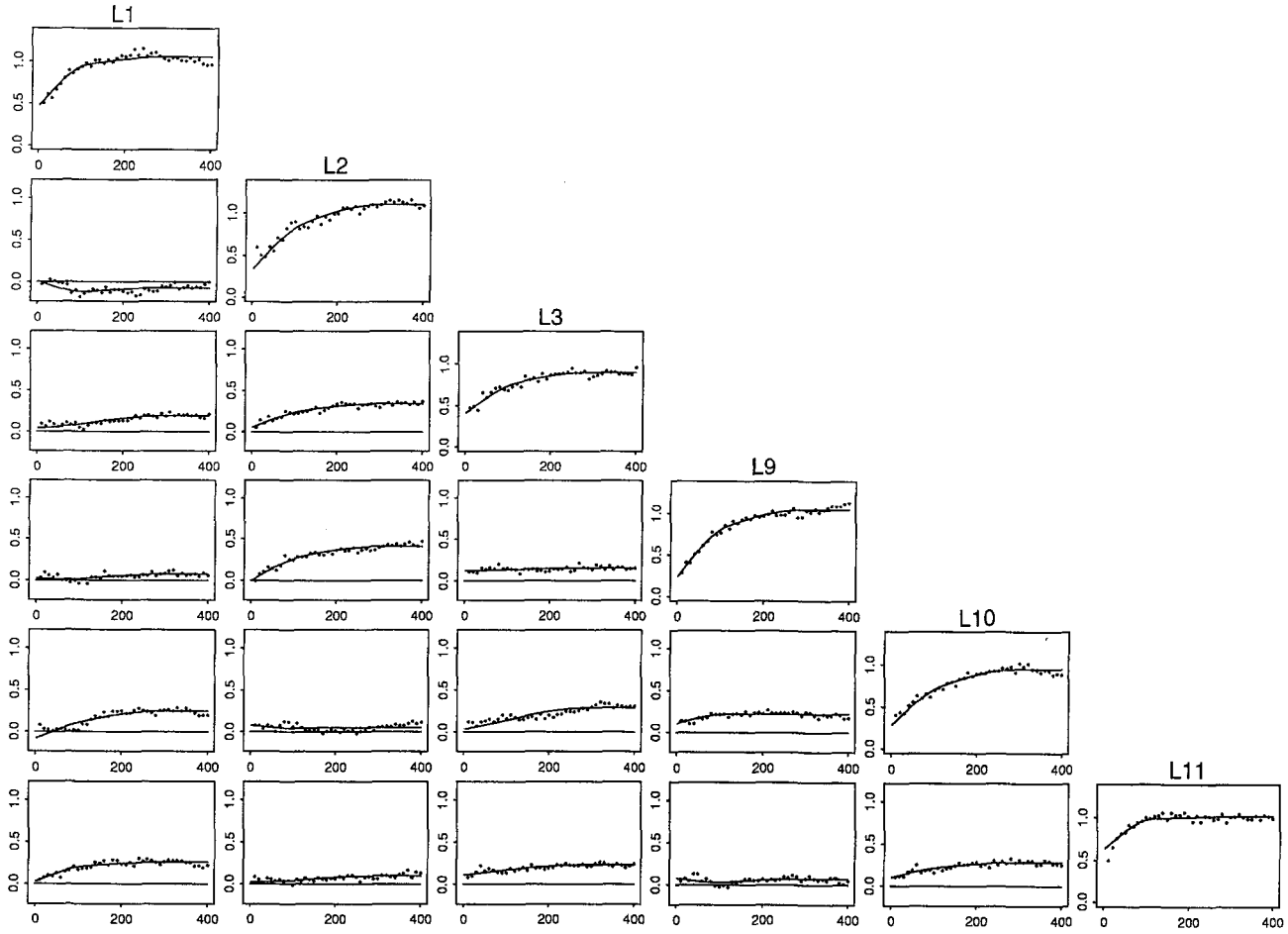


Fig. 5 Variograms (*diagonal plots*) and cross-variograms (*non-diagonal plots*) of the six variables "summer growth" evaluated at sites *L1–3* and *L9–11*. The functions are normalized using an inverse variance of variable. *Dots* represent empirical values and *solid lines* the fitted coregionalization model. Distance unit on *X axis* is kilometers

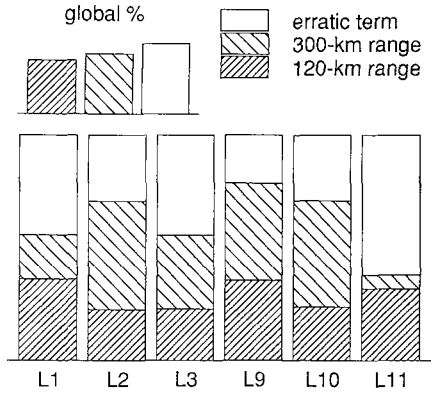


Fig. 6 Decomposition for each variable of the variation over the three spatial structures and global percentage for total variation

positively correlated on the 300-km-range structure. This second correlation tends to compensate the first one for increasing distance.

For the 120-km-range structure, the first two factors seem to be adequate to represent the main part of the variance (75%). The first one is related to the opposition between evaluation sites L1–L2 and the second between evaluation sites L10–L3. For the 300-km-range structure, the first factor collects the same part of variance and may be sufficient to summarize this spatial structure. If the nugget effect is considered, the percentage of variance decreases slowly along the six factors, which is characteristic of an absence of multivariate structure. We can assume that all of the nugget effects at the six evaluation sites are independent and cannot be summarized by one or two factors.

As explained before, spatial components related to factors of the principal component analysis can be mapped using cokriging. This was done for components related to the first two factors of the 120-km-range structure and for the first factor of the 300-km-range structure. Corresponding maps are shown in Fig. 8a–c. In order to illustrate a component of the nugget effect, the first one was mapped on ryegrass collection sites in Fig. 8d. In this late case, a smoothed continuous map is not possible because interpolation of the erratic term is meaningless. The first two maps reinforce the interpretation of the spherical structure of the 120-km range as easily distinguishable patches. The map of the spherical structure of the 300-km range features spatial variations that are related to a spatial division in homogeneous climatic regions. The map of the

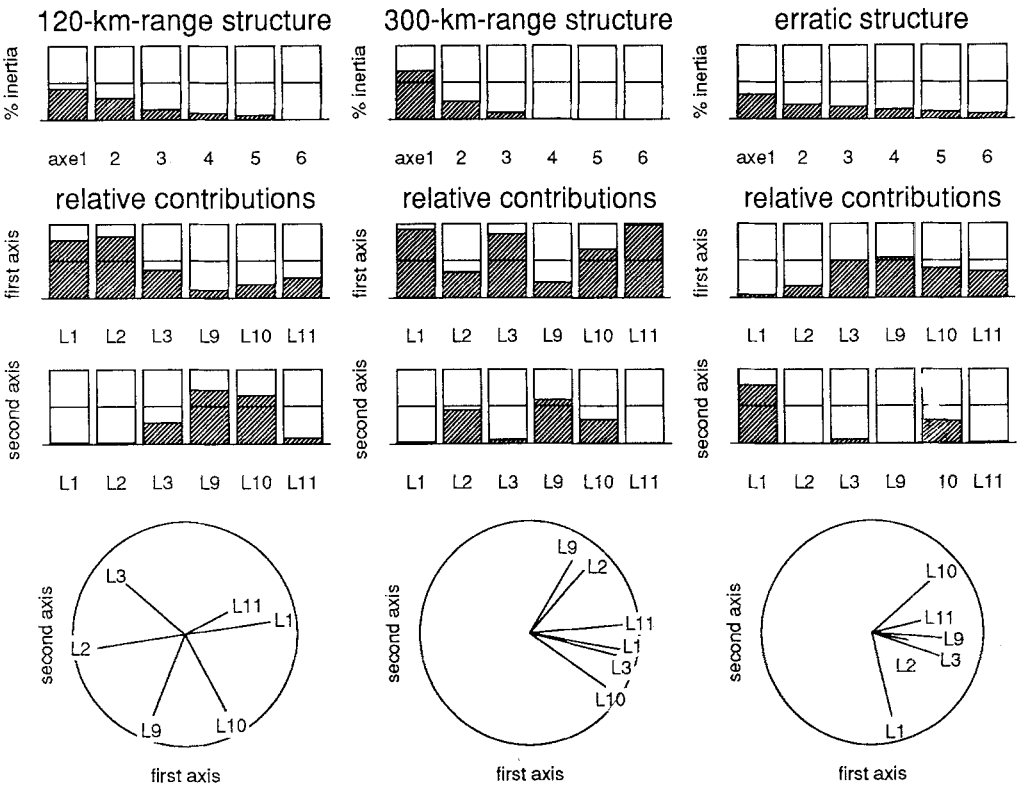


Fig. 7 Factorial decomposition of the variation of each spatial structure. *Left* 120-km-range structure, *middle* 300-km-range structure, *right* erratic spatial structure. *Upper* Inertia of the principal components and then relative contributions of the variables to the first two axes, *lower* correlation plots on the first-two-factor plane

nugget effect cannot be interpreted from a spatial point of view because variations come from experimental errors and very local interactions between collection site and test site. Any other test site should give another independent component for the multivariable nugget effect.

Discussion

Clear spatial patterns of variation have been found for the two analysed variables: a single structure with a 120-km-range for heading date and a more complex one for summer growth with two spatial structures of 120-km and 300-km range.

The value of the range of the variogram is equivalent to the value of the distance showing the first zero value in autocorrelation analysis. This value was interpreted by Sokal (1979) as indicating the path size; i.e., the diameter of homogeneous area in the surface of the variable studied. Such homogeneous areas whose size fit the range of the variogram can be observed in Figs. 4 and 8.

To explain homogeneous patches two hypotheses have been proposed by Sokal (1979, 1986) and Sokal and Jacquez (1991).

1) Selection: If the variable under study is affected by a selection pressure of environmental origin, then the patches of the variable surface will reflect the patches of this environmental factor. Although difficult to prove without additional experimental evidence, this hypothesis may be suspected whenever the map of the variable resembles that of any environmental factor – for example, temperature or rainfall – that may be candidate for a selective agent.

2) Isolation-by-distance: This is caused by restricted gene flow and is the main force tending to produce homogeneity of allelic frequency and, therefore, homogeneity of any genetically controlled trait, including the quantitative, polygenic ones. Gene flow in plant populations is mainly due to seed and pollen transport. Consequently, the average diameter of ge-netically homogeneous patches is partly determined by the average lifetime migration distance (Sokal 1979, 1986). Since this migration distance is the same for all genes, variograms of all variables are expected to be similar, as has been demonstrated in a simulation study by Sokal and Wartenberg (1983). However, their surface patterns are likely to be globally uncorrelated, since the initial allelic frequencies are entirely due to stochastic events or, possibly, to local microenvironmental selection. A common diffusion pattern will thus lead to different, uncorrelated maps for different variables, which, however, will bear a similar spatial structure.

The erratic component (nugget effect) may cover several sources of variation: (1) an experimental error sensu stricto caused, for instance, by uncontrolled field heterogeneity or unadequate eye assessment; (2) spatial variations on a scale below that of the study (selection pressure caused by microen-

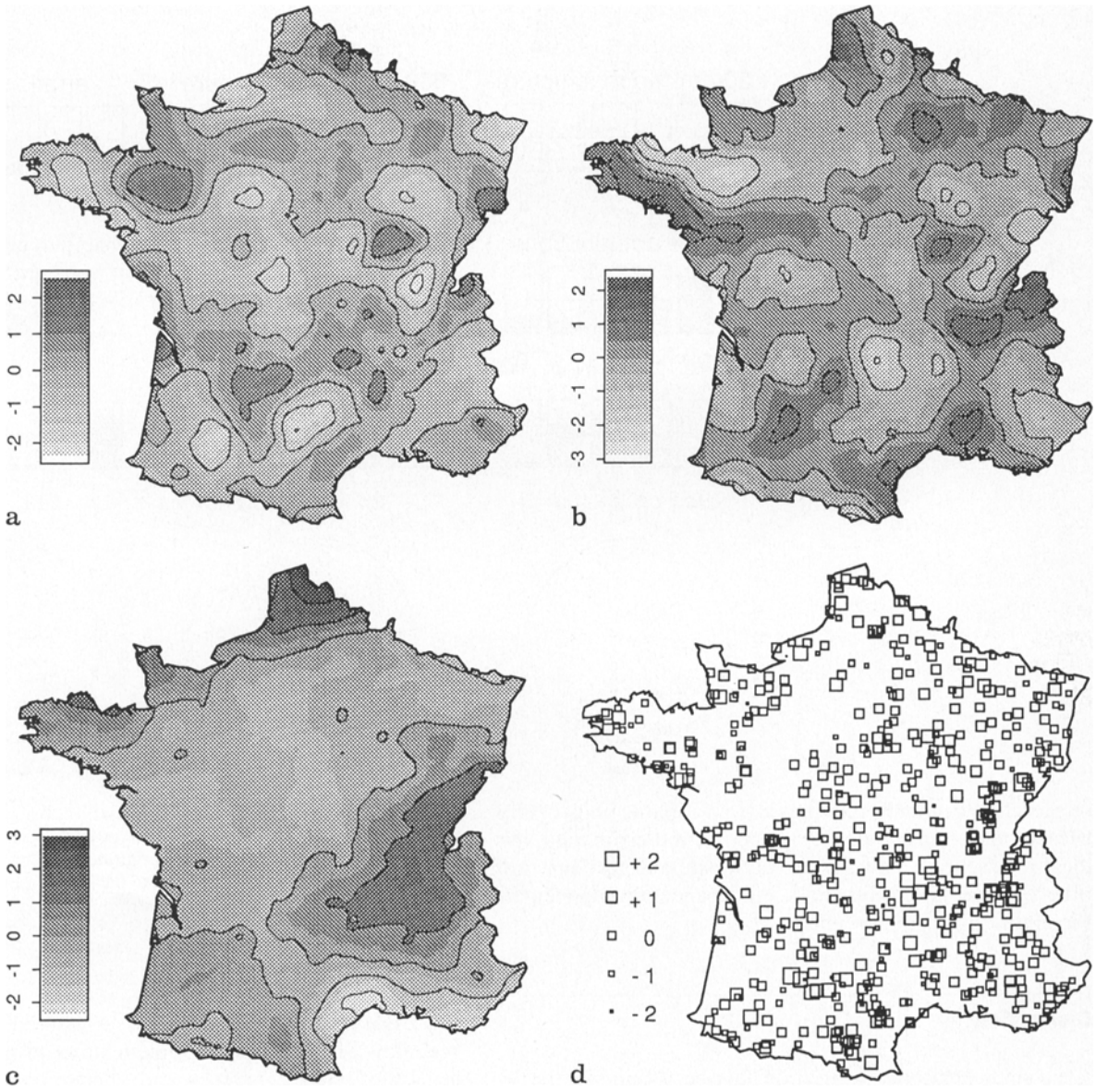


Fig. 8 **a** Map of the first component of the 120-km-range structure, **b** map of the second component of the 120-km-range structure, **c** map of the first component of the 300-km-range structure, **d** map of the first component of the erratic spatial structure

vironmental factors could particularly be incriminated) (3) a temporary "white noise" caused by stochastic events such as genetic drift caused by population bottleneck, extinction or colonization events, which may remain visible before gene flow between populations smooths allelic frequency differences.

Interpretation of the ryegrass population spatial structure

Using this theoretical background and reviewing the results of the geostatistical methods, we are able to formulate interesting interpretations on the case study. Briefly, one can deduce that the spatial structure of the 120-km range results from the gene flow, while the spatial structure of the 300-km range is caused by selection. The erratic component reflects more locally restricted variations and experimental errors.

As a matter of fact, the 120-km-range structure has been found in all of the variables analyzed: mean heading date and summer growth at most evaluation sites, and is even clear on some cross-variograms. This structure, which is the only one for heading date, overlaps with a 300-km-range spatial structure for summer growth. The kriging analysis has been powerful in separating the two structures. The comparison of kriged maps of heading date and the cokriged maps of the first two principal components of the 120-km-range structure of summer growth allows us to completely discard the selection hypothesis, since these three maps are uncorrelated and no association with any environmental factor is obvious. With the conclusion of isolation-by-distance cause for the 120-km structure, the first two corresponding components of summer growth, which are logically independent from each other,

might correspond to different gene pools. Each pool governs a particular adaptation to drought conditions but has the same spatial diffusion pattern.

Since the patch size associated to gene flow is unique,the selection hypothesis must be put forward to explain the 300-km-range structure of summer growth. Indeed, the map of the first principal component of this structure shows clear relation ships with maps of climatic factors such as summer water deficit (Bessemoulin 1969), which are very likely to exert a selection pressure on the summer growth ability of ryegrass populations.

The erratic component associated to the nugget effect may be a simple experimental error. It is remarkable, however, that this effect accounts for more than 50% of the variation at evaluation site L11, which is located in Normandy and is the most suitable climate for rye-grass growth in the summer. This is probably not by chance: in such favorable conditions, the genetic adaptation associated with the two spatial structures of the 120-km and 300-km range may be "buffered", thereby allowing microscale variation to become proportionally larger. These microscale variations may be due to microenvironmental conditions, as demonstrated in the grass *Anthoxanthum odoratum* by Jain and Bradshaw (1966).

The average distance of gene flow inferred from our results in perennial ryegrass could be compared with other estimates obtained from spatial autocorrelation studies in plant populations: Jensen (1986) found about 700 km in the oak *Quercus ellipsoidalis*, and Sokal et al. (1986b) also reported a distance of 600–700 km in a study of *Populus deltoides* in eastern USA. The average distance found in ryegrass is lower than that found in the tree species. This might be accounted for by differences in the sizes or shapes of the pollens, which would affect their dispersion ability.

Consequences on collecting methodology for breeding

Assuming that observed spatial patterns are mixtures of isolation-by-distance processes and selection processes, we propose that the following points should be considered when collecting natural populations of ryegrass. (1) The selection structure has only one dimension, and a simple use of the map of Fig. 8c may help the plant breeder to identify the geographic area where natural selection has operated in the direction desired with regards to its breeding objectives. (2) As homogeneous patches of the first two components of the isolation-by-distance structure are supposed to have independent gene pools, collecting populations in every patch will ensure that the genetic diversity is maximal. (3) One should, however, keep in mind that an erratic, unpredictable effect still remains. On an average, it represents 40% of the variation of summer growth. Therefore, plant breeders should be advised to collect (or conserve) several populations in every combination of patches (selection and gene-flow) in order to represent the within-patch-component, possibly of microenvironmental origin.

Geostatistic methods have proved to be powerful in analysing a complex spatial structure of several agricultural traits observed on a large collection of natural populations of perennial ryegrass. Further studies would be needed to confirm the hypotheses on isolation-by-distance patches. Geostatistics or spatial autocorrelation on neutral markers such as isozyme frequencies, for example, may be employed.

References

Barbujani G (1987) Autocorrelation of gene frequencies under isolation by distance. Genetics 117:777–782

Bessemoulin M (1969) Atlas climatique de la France. Ministere des transports, Direction de la meteorologie nationale, Paris

Charmet G, Balfourier F, Bion A (1990) Agronomic evaluation of a collection of French perennial ryegrass populations: multivariate classification using genotype × environment in interaction. Agronomie 10:807–823

Cressie N (1986) Kriging non-stationary data. J Am Stat Assoc 81:625–634

Cressie N (1991) Statistics for spatial data. Wiley, New York

Dewey SE, Heywood JS (1988) Spatial genetic structure in a population of *Psychotria nervosa*. I. Distribution of genotypes. Evolution 42:834–838

Epperson BK (1990) Spatial autocorrelation of genotypes under directional selection. Genetics 124:757–771

Epperson BK, Allard RW (1989) Spatial autocorrelation analysis of the distribution of genotypes within populations of Lodgepole Pine. Genetics 121:369–377

Epperson BK, Clegg MT (1986) Spatial autocorrelation analysis of flower color polymorphisms within substructured populations of morning glory (*Ipomea purpurea*). Am Nat 128:840–858

Goulard M, Voltz M (1992) Linear coregionalization model: tools for estimation and choice of cross-variogram matrix. Math Geol 24: 269–286

Jain SK, Bradshaw AD (1966) Evolutionary divergence among adjacent plant populations. I. The evidence and its theoretical analysis. Heredity 21:407–441

Jensen RJ (1986) Geographic spatial autocorrelation in *Quercus ellipsoidalis*. Bull Torrey Bot Club 113:431–439

Journel AG, Huijbregts CJ (1978) Mining geostatistics. Academic Press, London

Matheron G (1965) Les variables régionalisées et leur estimation. Masson, Paris

Perry DJ, Knowles D (1991) Spatial genetic structure within three sugar maple (*Acer saccharum* Marsh.) stands. Heredity 66:137–142

Sokal RR (1979) Ecological parameters inferred from spatial autocorrelograms. In: Patil GP, Rosenszweig ML (eds) Contemporary quantitative ecology and related ecometrics. Int Cooperative Publ House, Fairland, Md., pp 167–196

Sokal RR (1986) Spatial data analysis and historical processes. In: Diday E et al. (eds) Data analysis and informatics IV. North-Holland, Amsterdam, pp 29–43

Sokal RR, Jacquez GM (1991) Testing inferences about microevolutionary processes by means of spatial autocorrelation analysis. Evolution 45:152–168

Sokal RR, Oden NL (1978a) Spatial autocorrelation in biology. 1. Methodology. Biol J Linn Soc 10:199–228

Sokal RR, Oden NL (1978b) Spatial autocorrelation in biology. 2. Some biological implications and four applications of evolutionary and ecological interest. Biol J Linn Soc 10:229–249

Sokal RR, Wartenberg DE (1983) A test of spatial autocorrelation analysis using an isolation-by-distance model. Genetics 105:219–237

Sokal RR, Crovello TJ, Unnash RS (1986) Geographic variation of vegetative characters in *Populus deltoides*. Syst Bot 11:419–432

Sokal RR, Jacquez GM, Wooten MC (1989) Spatial autocorrelation. Analysis of migration and selection. Genetics 121:845–855

Wackernagel H (1988) Geostatistical techniques for interpreting multivariate spatial information. In: Chung CF et al. (eds) Quantitative analysis of mineral and energy resources. Proc NATO Conf. Reidel, Dordrecht, pp 393–409

Waser NM (1987) Spatial genetic structure in a population of the montane perennial plant *Delphinium nelsonii*. Heredity 58:249–256

Webster R, Oliver MA (1990) Statistical methods in soil and land resource survey. Oxford University Press, Oxford