COENOSES 6 (3): 151-168, 1991

SPATIAL INTERPOLATION METHODS FOR INTERPRETATION OF ORDINATION DIAGRAMS

Mike Hauser and Laco Mucina, Unit of Population Biology, Inst. of Plant Physiology, University of Vienna, Althanstrasse 14, A-1091 Wien, Austria

Keywords: Canonical analysis, Kriging, Moving weighted averages, Polynomial fitting, Trend surface analysis

Abstract. Spatial interpolation methods are very popular in geosciences. Among these, trend surface analysis (TSA), inverse-distance interpolation, splines, and kriging are the most commonly used. Although vegetation science handles many spatial aspects of ecological data, these methods have experienced less appreciation. TSA and kriging interpolations are discussed in the present paper. Detailed formulations of the techniques are given, and their advantages and flaws briefly discussed. Two example data sets are used to elucidate the applications of these methods in the interpretation of ordination diagrams.

Introduction

The complexity of ecological phenomena has attracted numerous multivariate techniques for handling of data. In vegetation science these methods have been used mainly to define vegetation types (Mucina and Dale 1989), variational trends in space, eg. in biogeography (Birks 1987, Legendre 1990) or in time (Austin 1977, see also Miles et al. 1988 for references). The application of numerical techniques and computers not only allowed large data sets to be treated (van der Maarel et al. 1987) or enhanced the handling procedures themselves, but aided the emergence of a body of theory (van der Maarel 1989, Pickett and Kolasa 1989).

Recent ecological literature witnesses a rapid development of new multivariate techniques. This is demonstrated by a number of recent milestone textbooks or proceedings (Legendre and Legendre 1983, 1987, Pielou 1984, Digby and Kempton 1987, Jongman et al. 1987, Ludwing and Reynolds 1988). It seems to us that the techniques are being developed (or improved) more rapidly than they can be tested or fully grasped, especially in the field of the interpretation of results. Incidentally, this has always been the weakest point of numerical data analysis in vegetation science and ecology. A short review (see also van der Maarel 1980, Mucina and van der Maarel 1989) of the interpretational tools used in the ordination of vegetation demonstrates this point.

The widespread indirect ordination methods used in vegetation science include principal components analysis (cf. Orlóci 1978), correspondence analysis and its derivates (Hill 1974, Jongman et al. 1987), and principal coordinate analysis (eg. Pielou 1984). As a rule the vegetation scientist is interested in looking for (and explaining) various data structures plotted on an optically comprehensible surface or in few-dimensional spaces being combinations of ordination axes with high information content. All of the methods mentioned above uses eigenanalysis as the computational background. Thus items such as the eigenvalues (variation explained by an axis) and eigenvectors (variable weights) can be directly utilized in the interpretation (Nichols 1977, Orlóci 1978, Pielou 1984). Most often a pragmatical approach is used - the obtained structures are tested for predictivity against an external variable, that means some data staying outside the ordination analysis itself gets compared with the ordination results (Feoli 1983). Van Groenewoud (1965) used simple correlations of axes with an external variable.

The superimposition of the values of an external variable on the ordination plane is common. These variables are metric when one looks for trends, eg. content of skeleton (Mucina 1982), percentage occurrence of certain species groups (Feoli-Chiapella and Feoli 1977), geographic coordinates (Mucina 1989); ordinal, eg. Ellenberg's (1979) indication values (Persson 1981); or can represent nominal classes, particularly when looking for clusters, eg. by explaining syntaxonomic structures (for many papers see Mucina and Dale 1989).

The analysis of the metric variables on ordination planes was done mainly intuitively - by drawing hand-fitted isolines (van der Maarel 1969, Mucina 1982). In earth sciences a family of interpolation methods (including polynomial trend surfaces, Dirichlet tessellation, splines, kriging, etc.) is in use which might be applied also in the interpretation of ordination diagrams (eg. Dargie 1984). Following a paper by Gittins (1968) and later applications by Mucina and Poláčik (1982) and reviews by Mucina et al. (1988), Fortin et al. (1989), Legendre and Fortin (1989) the ecological field is very well aware of the these methods. However, understanding and application of these techniques is still low.

Some applications of interpolation on real (field) surfaces were reviewed by Legendre and Fortin (1989). This chapter attempts to explore the possibilities of their

application on image surfaces (ordination diagrams, two-dimensional graphs) for the purposes of interpretation of numerical data analysis in vegetation science.

Methods

Trend Surface Analysis (TSA)

Is based on the method of fitting a regression line on observed point data. TSA is the special case when a surface (name!) is fitted to points in euclidian three dimensional space, where the first two dimensions (X, Y) are the independent variables (position variables, explanatory variables), the third dimension (Z) is the dependent one.

In practice this is achieved by constructing polynomials using the least square method to minimalize the sum of squares between the observed values and the fitted surface. That polynomial function works as the corresponding function between the coordinates and the observed values.

The effect is that the information of the whole sample gets divided into two parts:

a) the long-range change of the spatial variation of Z (systematic part),

b) the short-range change (random component), that is represented by the deviations from the regression (error part), namely ϵ.

The basic model is additive:

$$Z_i = f(X_i, Y_i) + \epsilon_i \qquad \text{(eqn. 1)}$$

where

Z_i	is the value of the dependent variable at point i,
X_i , Y_i	are the first 2 coordinates at that point,
$f(X_i, Y_i)$	is the trend function, and
ϵ_i	is the residual (random) component.

The general form of the polynomial function sought is:

$$Z = \sum_{i=0}^{P} b_{rs} X^r Y^s + \epsilon \qquad \text{(eqn. 2)}$$

where

Z	is the value at point X, Y,
p	is the order of the polynom,
b_{rs}	are the polynomial coefficients;
r, s	are the exponents that behave such, that r=p in the first term and decreases from term to term about 1 down to 0 (in the last term), s=0 in the first term and increases from term to term about 1 up to p (in the last term).
ϵ	is the residual component.

There are $\frac{(p+1)(p+2)}{2}$ coefficients.

The first few polynomial functions are:

b_0

flat function; b_0 is the overall average of the dependent variable (being derived from the two dimensional case this is the intercept on the dependent variables' axis at the location 0/0).

$b_0 + b_1 X + b_2 Y$

linear function (1st order); b_1 is the linear slope along the X-axis, b_2 is the linear slope along the Y-axis.

$b_0 + b_1 X + b_2 Y + b_3 X^2 + b_4 XY + b_5 Y^2$

quadratic function (2nd order); b_3 and b_5 are the quadratic slopes along the X- and Y-axis respectively, b_4 is the combinatorial slope of both axes.

To clarify the underlying principle of how the coefficients and how the powers of X and Y are set, the cubic function (3rd order) is presented here:

$$b_0 + b_1 X + b_2 Y + b_3 X^2 + b_4 XY + b_5 Y^2 +$$

$$+ b_6 X^3 + b_7 X^2 Y + b_8 XY^2 + b_9 Y^3$$

To solve the system the polynomial coefficients are conceived as regression coefficients (Unwin 1975). Their calculation is performed using the general least squares method's formula:

$$\sum_{i=1}^{n} \epsilon_1^2 = \sum_{i=1}^{n} (Z_{i\,(Obs)} - Z_{i\,(Calc)})^2 \rightarrow \min \qquad \text{(eqn. 3)}$$

where

n	is the number of observations,
$Z_{i\,(Obs)}$	is the observed value at point i,
$Z_{i\,(Calc)}$	is the value of the polynomial function at i.

This standard method to find the actual values of the coefficients (namely multiple regression) can be consulted in textbooks such as Burden and Faires (1985), Upton and Fingleton (1985), ter Braak and Looman (1987, p. 53) or specialised papers by Gittins (1968) and Mucina et al. (1988).

Advantages

1) As in regression analysis the polynom of the surface can be used to predict or interpolate Z values

of unsampled sites.

As the surface usually will not go through the observed points (except in the case when the order of the polynom equals the number of observations, but see the cons below!) it will not be an exact interpolator. The least squares criterion can be used to judge the goodness of fit of the surface to the data. This is simply the fraction of variance accounted for by the explanatory variable, also called the adjusted coefficient of determination:

$$R^2_{adj} = 1 - \left(\frac{\text{residual variance}}{\text{total variance}}\right) \qquad \text{(eqn. 4)}$$

This coefficient takes the number of fitted parameters in to account.

In comparison, the unadjusted coefficient of determination

$$R^2 = 1 - \left(\frac{\text{residual sum of squares}}{\text{total sum of squares}}\right) \qquad \text{(eqn. 5)}$$

does not.

When a large number of parameters is fitted, R^2 may yield a value close to 1, even when the expected response does not depend on the explanatory variable (ter Braak and Looman 1987, p. 33).

The multiple correlation coefficient (=product-moment correlation between observed and fitted values) is the square root of the coefficient of determination (ter Braak and Looman l.c.; see also Mucina et al. 1988).

Finally one can use the variance ratio (F) of the mean squares of the regression and the residuals

$$F = \frac{\text{mean square regression}}{\text{mean square residual}} \qquad \text{(eqn. 6)}$$

An F close to 1 says that the regression is of no assistance in explaining the Z variation. The more that the value exceeds 1, the better the surface (Upton and Fingleton 1985, p. 272).

In addition, F assumes that the residuals are independently and normally distributed (see the problems below).

2) Overall TSA is good for finding basic trends or clines in the observed area (especially when the distribution of residuals is random it can be explained as noise with almost complete certainty). Analogous to sampling at different scales the order of the polynom can be used as a filter. Low order polynomials will yield more overall trends (larger patterns), high order polynomials will give a more fine-grained picture.

Problems

1) Especially when only a few data points are provided, extreme values can distort the surface.

2) The method of minimizing the least squares requires that the errors ϵ are normally distributed and not the observed Z values.

There is no general rule of thumb to say that, if the Z values are normally distributed, the errors will have such an distribution too (and vice versa). It can be nothing else but a guess that there is a connection between the distribution of the actual values and the errors.

In doing so, one should first check the distribution of the observed values. When there is an apparently skewed distribution one should first transform the original data (e.g. log-transformation of positively skewed data; see Burrough 1987). The transformed data should then be used for multiple regression and the errors checked for normality. When they are not normally distributed one can say nothing about the distinction between the long-range and short-range variation.

Ripley (1981) recommends that all distances therefore should be rescaled in the range between −1 and +1.

3) When using an F test one assumes that the residuals are independent from each other and normally distributed. This is in most cases unrealistic, because the deviations from the surface are mostly correlated over short distances (Burrough 1987, p. 218).

Imagine a low-order polynomial surface that does not fit some area of the data points very well (e.g. and edge of the observation area); the residuals for those points (whose position is close to each other and are therefore correlated by distance) will all be high.

4) By increasing the number of parameters (=coefficients) up to the number of data points the surface will be an exact fit. The inflections increase up to 1 less than the order of the surface and the determination coefficient R^2 and the distinction between long-range and short-range variation again is meaningless. In fact the law of simplicity should rule here (as ever) to find the compromise between as few as possible parameters and minimum sum of squares.

5) The edges of the sampled area (and not only there) are locations where the surface will "behave" markedly different than the inner parts of the sample plot. This is because there are no further data points that can hold the surface "in bounds". The result is that the residuals in those areas will mostly be greater than in other parts (Ripley 1981).

This affects the validity of any F test or R^2 test, as mentioned above. Cliff and Ord (1981) therefore test residuals for autocorrelation.

Upton and Fingleton (1985, p. 324) suggest solving the problem by relocation of the boundaries. The best way to do so is to sample and calculate TSA in a bigger area than you finally interpret (and show).

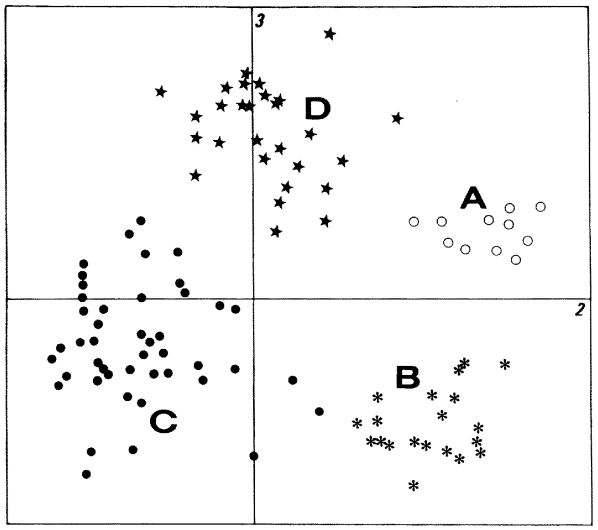


Fig. 1. Ordination plane of axes 2 and 3 of a PCA (non-transformed) of the Vorarlberg woodland data. For interpretation of the clusters A through D see the text.

Local Trends

Burrough (1987, p. 222) gives an alternative to or refinement of the method mentioned above. Instead of fitting a global trend surface one can divide the whole study area into smaller parts and calculate local trends for each of the parts separately. The big advantage is that major outliers affect only their own localities and not the whole plane. Burrough gives an example using small squares (= windows), but one could also try to do this using triangles or hexagons. This should be done with more than one window size and more than one order of local trend.

Burrough does not say anything about overlapping windows, but some smoothing could possibly be made using overlapping windows and calculating for each interpolated point the average value from the number of "window-hits". In that case it could probably be possible to use circles, to avoid "squared", "triangular" etc. bias when using quadratic, rectangular, triangular etc. windows, especially for big ones.

Canonical Trend Surface Analysis (CTS)

This method is based upon the canonical correlations between two sets of orthogonal axes, one carrying the geographic information, the other derived from a set of multivariate observations. To fit a surface on multivariate data this method should be chosen.

For a more detailed description see Wartenberg (1985).

Moving weighted averages

To avoid some of the problems when using trend surfaces to interpolate the values of the attribute at non-sampled sites it is better to use another underlying principle:

Autocovariation, which implies that observations located close together tend to be more alike than observations spaced further apart. To do so one solution is that observed Z values close to any point we want to interpolate have greater influence than values observed farther apart. What we need is a weighting function (w_i) for the observed Z values based on a distance measure, that returns a value of 0 when the distance between the observed point and the estimated point $\rightarrow^{\infty}$.

The weights used are:

1) negative exponential functions as

$$\exp(-\alpha d) \text{ or } \exp(-\alpha d^2)$$

2) reciprocal functions as

$$\frac{1}{d} \text{ or } \frac{1}{d^2}$$

where

d is the distance between the 2 points and
α is a constant >0.

The weighting moving average ($\hat{Z}$; the hat (^) over Z means the estimate of Z built from our sampled data) is computed using the following formula:

$$\hat{Z}(x_i) = \frac{\sum_{i=1}^{n} w_i Z(x_i)}{\sum_{i=1}^{n} w_i} \qquad \text{(eq. 7)}$$

where

x_j is the point at which we interpolate,
x_i is an observed point with i running from 1 to
n the number of observed points.

Burrough (1987, p. 237) mentions as the most common form of welghting function the inverse squared distance weighting that results in the following formula for the moving average:

$$\hat{Z}(x_j) = \frac{\sum_{i=1}^{n} \frac{Z(x_i)}{d^2_{(i,j)}}}{\sum_{i=1}^{n} \frac{1}{d^2_{(i,j)}}} \qquad \text{(eqn. 8)}$$

The "moving" in the name of the interpolation method comes from the way the newly obtained Z values are calculated. Although it is possible to calculate

for every interpolated value the average using all observed points, it is more common (and less time consuming) to move a window of such a size over the whole region that there are 8-12 observed data points inside the frame. The size, shape, and orientation may have a big influence on the results. This leads towards a time-consuming test using different values for each of the window-parameters and comparing the results. To avoid many of those problems one should consider the next method.

Kriging

Developed by D.G. Krige (1966), a South African mining engineer, and the French geomathematician G. Matheron (1971) the underlying principle is that the weights (w_i) are estimated using the semivariogram (Matheron 1962) of the data (explained below). To distinguish the weights derived from the semivariogram from the weights discussed above, they will here be denoted as λ_i.

But before we step deep into the details of kriging we have to clarify the basic underlying principles (Delhomme 1978).

1) What we want to track is a so called regionalized variable (ReV), that is a variable showing certain structure in space and/or time. A ReV can be expressed simply as a function z (x), where x is a point in space.

It is convenient to deal with ReV's by using the probabilistic theory of random functions. Therefore a ReV z (x) can be interpreted as a realization of a random function, denoted by Z (x) (the capitalization indicates the random function). As it is impossible to obtain the moments of a random function (= expectation (E) of the mean (μ), variance (var) or probability density function (pdf)) from only one realization (= sampled point), and as the observed values have some statistical interference upon each other (and the interpolations as well), there has to be a minimal amount of observations provided (the decision as to how many of them and how dense will be part of the result, as we will see later).

2) As x is a point with the coordinates (ξ, η), any h = x—x' will consequently be a vector.

The second hypothesis we have to rely on (when using random functions) is the principle of stationarity: a random function is said to be stationary if its joint pdf at k arbitrary points is invariant under simultaneous translation of all these points.

In most cases this will be only a weak stationarity of order 2 (= only the expectations (E) of μ and var are invariant under transition).

The expectation (E) of a value at any x is then a constant, the mean (μ) (= 1st moment) is independent of x

$$E\ [Z\ (x)] = m\ (x) = \mu \qquad \text{(eqn. 9)}$$

and the covariance (cov) between two points x and x' does not depend on any of those two separately, but only on the vector (h):

$$E\ [Z\ (x)-\mu]\ [Z\ (x')-\mu] = \text{cov}\ (x-x') \qquad \text{(eqn. 10).}$$

Furthermore, the variance (var) (being nothing else but a special form of the covariance) is independent of x

$$E\ [(Z\ (x)-\mu)^2] = \text{var}\ [Z\ (x)] \qquad \text{(eqn. 11)}$$

(the last two equations show the expectations of the 2nd moment).

The derivation of the kriging equations requires weak stationarity not for Z (x) itself, but only for its increments.

Thus we need only the hypothesis that, for any vector h, the increment Z(x+h) — Z(x) has zero expectation and a variance independent of x. We can therefore write

$$E\ [Z\ (x+h)-Z\ (x)] = 0 \qquad \text{(eqn. 12)}$$

$$\text{var}\ [Z\ (x+h)-Z\ (x)] = 2\ \gamma\ (h) \qquad \text{(eqn. 13)}$$

This hypothesis is weaker than stationarity of order 2 and is called the intrinsic hypothesis. Any random function that satisfies this hypothesis is called an intrinsic random function.

The function γ (h) is called the semivariogram (often referred to - as here - simply as the variogram).

We can write the formula of the variogram in two ways:

$$\gamma\ (h) = \frac{1}{2}\ \text{var}\ [Z\ (x+h)-Z\ (x)] \qquad \text{(eqn. 14)}$$

$$\gamma\ (h) = \frac{1}{2}\ E\ [(Z\ (x+h)-Z\ (x))^2] \qquad \text{(eqn. 15)}$$

and therefore the 2 γ (h) of equation 13 is the mean squared difference for two points separated by a distance (or lag) h.

To estimate the variogram we have to calculate it for any distance (or distance class) in our data using

$$\gamma\ (h) = \frac{1}{2N\ (h)} \sum_{i=1}^{N\ (h)} [z\ (x_i+h)-z\ (x_i)]^2 \qquad \text{(eqn. 16)}$$

where

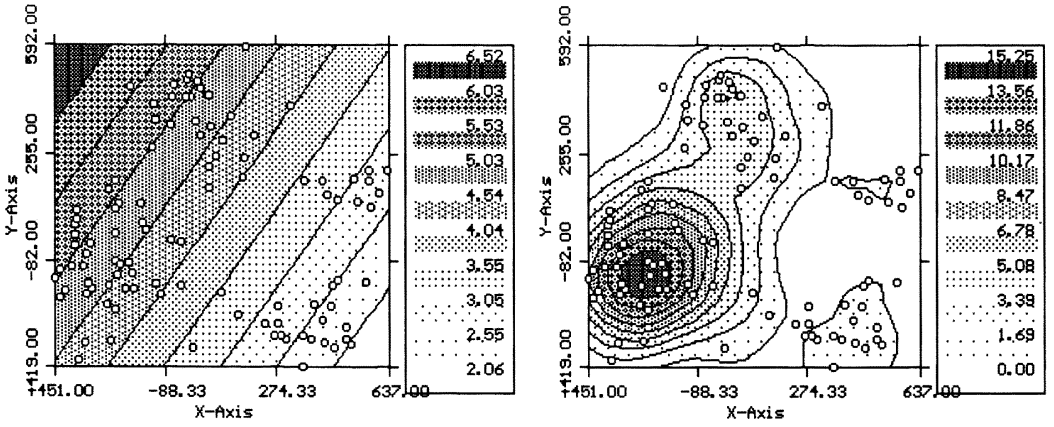
$z\ (x_i)$ is the observed value at point x_i .

N (h) is the number of distances we can obtain using our data.

In a one-dimensional sample the lag h is only a scalar. When the independent coordinates of x are two-

Fig. 2. TSA on pH values for soil as measured in 105 sampling stations (indicated by small circles inside the maps) of the Vorarlberg deciduous and mixed woods. The left-hand subfigures are the trend surface maps; those on the right are the maps of residuals (the weighting function was 200,00). The designations (a) through (c) correspond to trend degrees. For goodness-of-fit see Tab. 1.

A

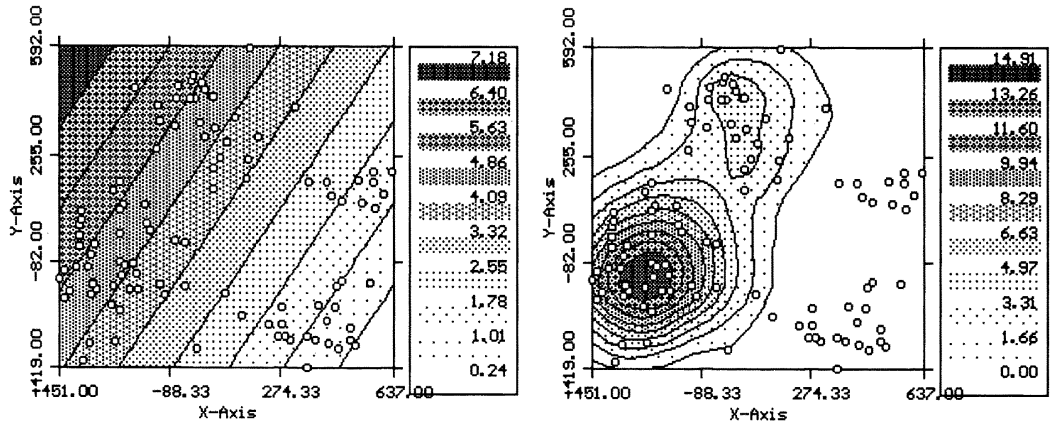


The TRENDSurface of Order 1

The RESIDUALSurface
Ratio: 200.0000

R^2: 0.5527

B

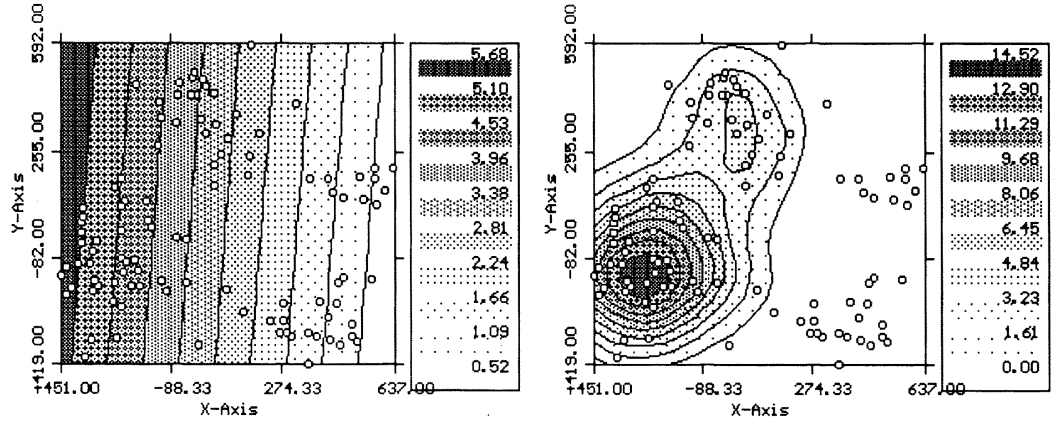


The TRENDSurface of Order 3

The RESIDUALSurface
Ratio: 200.0000

R^2: 0.6202

C



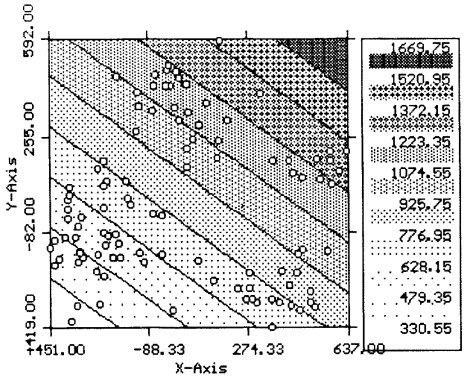
The TRENDSurface of Order 5

The RESIDUALSurface
Ratio: 200.0000

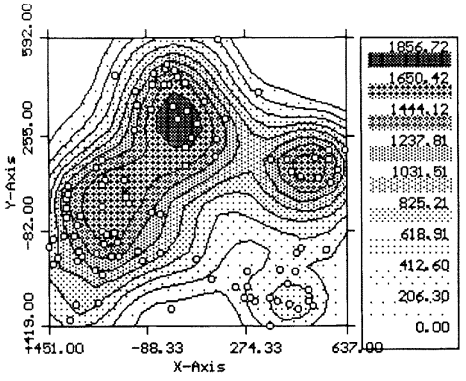
R^2: 0.6760

Fig. 3. TSA on altitude of the Vorarlberg woodland data. For further details see Fig. 2 and for goodness-of-fit consult Tab. 1.

A

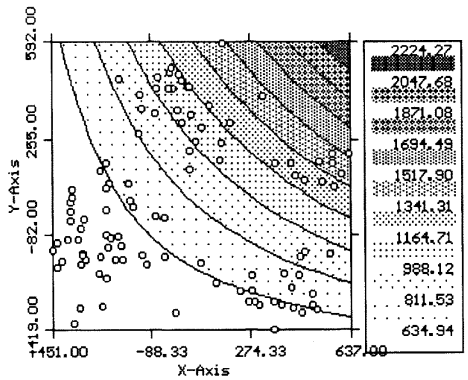


The TRENDSurface of Order 1

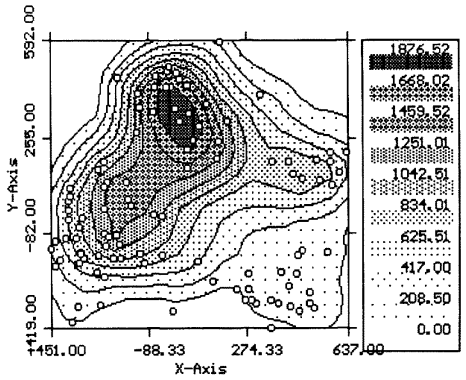


The RESIDUALSurface
Ratio: 0.9000

B

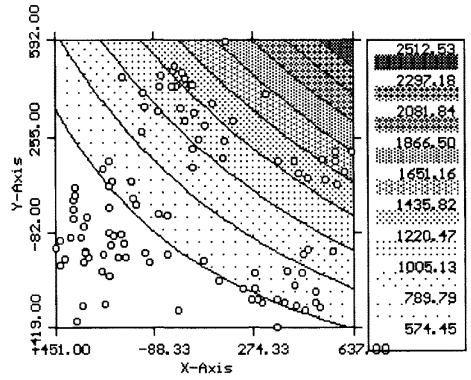


The TRENDSurface of Order 2

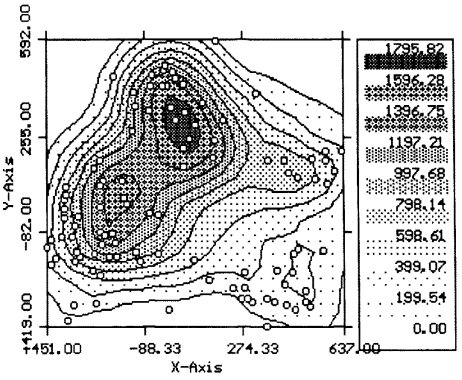


The RESIDUALSurface
Ratio: 0.9000

C

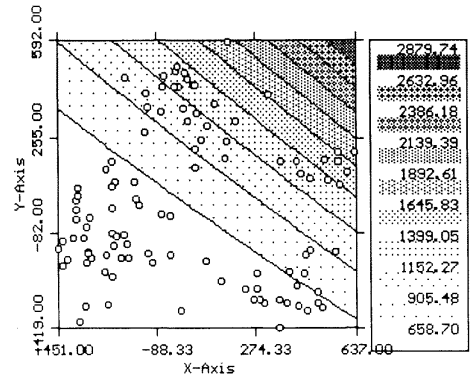


The TRENDSurface of Order 3

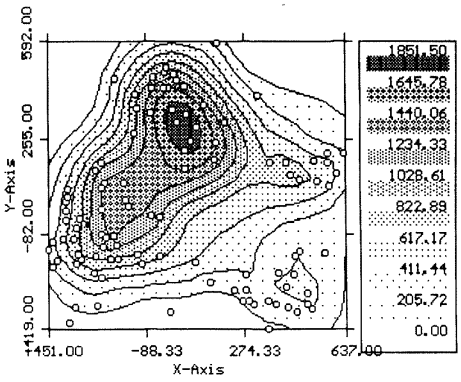


The RESIDUALSurface
Ratio: 0.9000

D



The TRENDSurface of Order 4



The RESIDUALSurface
Ratio: 0.9000

dimensional (as in our consideration) the variogram is then a function of its two components (in polar coordinates now):

1) the modulus |h| (=distance or lag), and
2) the polar angle θ.

Practically we do this such, that we compute variograms along two, three (or probably more) directions, divide the existing lags into distance classes and, to provide sufficient variances to measure, do this rather in a pie segment around the direction than merely along the direction vector (and obtain therefore direction classes).

A comparison of the variograms calculated for different directions will show in most cases the anisotropies of the observed phenomenon (compared with the isotropic case, where γ (h) is a function of |h| only).

Furthermore we can distinguish 4 classical types of behavior at the origin of the variogram (Delhomme 1978, Burrough 1987, Legendre and Fortin 1989 give further explanations about the form of a semivariogram and its meaning):

1) The parabolic shape that characterizes an extremely regular variable (either exponential, gaussian or spherical) is differentiable by mean square convergence.

2) Linear shape typifies a variable less regular than in the above case; not differentiable.

3) Discontinuity at the origin (nugget effect C) means that the variable is very irregular (not continuous in mean square convergence). Two distinct points, even when they are very close to each other, show a difference with a variance at least equal to the nugget effect. What has happened is that there is an internal discontinuity on a scale much smaller than the spacing of the sampled data points; alternatively, it may be due to measurement errors.

4) Flat variogram or pure white noise show the purely random case. Z(x) and Z(x+h) are by no means correlated.

On the other hand we can distinguish different behaviors at large distances (or at infinity distance):

1) The variogram increases indefinitely (as in the linear case), or

2) it stabilizes around some limiting value, called the sill. The distance at which the sill is reached is called the range.

The range shows the extent of the zone of influence of an experimental point: at distances beyond the range correlations between points are nil (as the reader can see here lies the solution for the window size problem mentioned above while discussing moving weighted averages).

How we try to fit a theoretical model upon our estimated variogram. We will use linear combinations of Z (x) to derive the kriging equations. A model for γ (h) must be such that for every set n of points x_i and for n arbitrary coefficients λ_i , that verify

$$\sum_{i=1}^{n} \lambda_i = 0 \qquad \text{(eqn. 17)}$$

(so that we get a linear combination of increments, which are the only quantities we can rely on when working under the principles of the intrinsic hypothesis) is always obtained a positive result when calculating

$$\text{var} \left(\sum_{i=1}^{n} \lambda_i Z(x_i) \right) = - \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j \gamma (x_i - x_j) \qquad \text{(eqn. 18)}$$

Any nugget effect (C) results in the addition of a constant C to the model of γ (h) so that we have:

$$\gamma (h) = C (1-\delta) + \gamma_0 (h) \qquad \text{(eqn. 19)}$$

where

δ is a so called Dirac measure, and
γ_0 (h) is the variogram without the nugget effect.

We now extend the intrinsic hypothesis to make it possible to deal with variables for which that hypothesis of constant mean cannot be reasonably made.

We suppose that the expectation of the mean (see equation 9 above) varies only slowly relative to the working scale. This function m (x), called the drift, will be assumed to be regular enough to be expressed locally as

$$m (x) = \sum_{L=1}^{k} a_L f^L (x) \qquad \text{(eqn. 20)}$$

where

a_L are the unknown coefficients,
f^L are the given basic functions, and
k is the total number of both of them.

In common practice these functions (f^L) are monomials.

For any point x that has the coordinates (ζ, η) we can therefore write

$$m (\zeta, \eta) = a_1 + a_2 \zeta + a_3 \eta \qquad \text{(eqn. 21)}$$

for the linear case (compare this with the polynomial functions mentioned in the TSA chapter above and you see that the drive behaves exactly like the surface computed there).

The second part of the intrinsic hypothesis (equation 13) will be maintained. All that remains to be resolved

is the inference of the variogram when equation 12 is not fulfilled.

As we know the values of the variable z (x) at n observation points x_1 , x_2 , ..., x_n , we have to estimate quantities z_0 which are:

1) $z_0 = z(x_0)$ (eqn. 22)
is the value of z at point $x = x_0$

2) $z_0 = \frac{1}{s} \int_{s_0} z(x)\, dx$ (eqn. 23)

is the average value of z (x) over the mesh s_0 of area s, centered at point $x = x_0$.

3) $z_0 = \frac{1}{S} \int_{S} z(x)\, ds$ (eqn. 24)

is the average value of z (x) over the domain S.

We derive the calculations here only for the third case (the first case can be seen as a border line case where the domain is reduced to a point, the second case is a more special form of the third). We estimate z_0 using a weighted average

$$\hat{z}_0 = \sum_{i=1}^{n} \lambda_i z(x_i) \qquad \text{(eqn. 25)}$$

so that the linear combinations of λ_i give the best possible estimation. Recall that the probabilistic interpretation of the regionalized variable z (x), $\hat{z}_0$ is nothing else but a realization of a random function:

$$\hat{Z}_0 = \sum_{i=1}^{n} \lambda_i Z(x_i) \qquad \text{(eqn. 26)}$$

(Remember: capitalized letters refer to random functions, small letters to observed values).

The weights λ_i must be such that the estimator $\hat{Z}_0$ is unbiased (and gives therefore no systematic under- or over-estimation) and optimal in the sense of mean square error. This can be expressed as:

$$E[\hat{Z}_0 - Z_0] = 0 \qquad \text{(eqn. 27)}$$

$$\text{var}[\hat{Z}_0 - Z_0] \rightarrow \min \qquad \text{(eqn. 28).}$$

With bias involved equation 27 becomes

$$\sum_{i=1}^{n} \lambda_i \sum_{L=1}^{k} a_L f_i^L - \sum_{L=1}^{k} a_L f_s^L = 0 \qquad \text{(eqn. 29)}$$

to handle the drift, where

$f_i^L = f^L(x_i)$. is a monomial as in equation 21

and

$f_s^L = \frac{1}{S} \int_s f^L(x)\, dx$ as in equation 24.

Equation 29 must be fulfilled, whatever the unknown values of a_L are, therefore we have to change it into

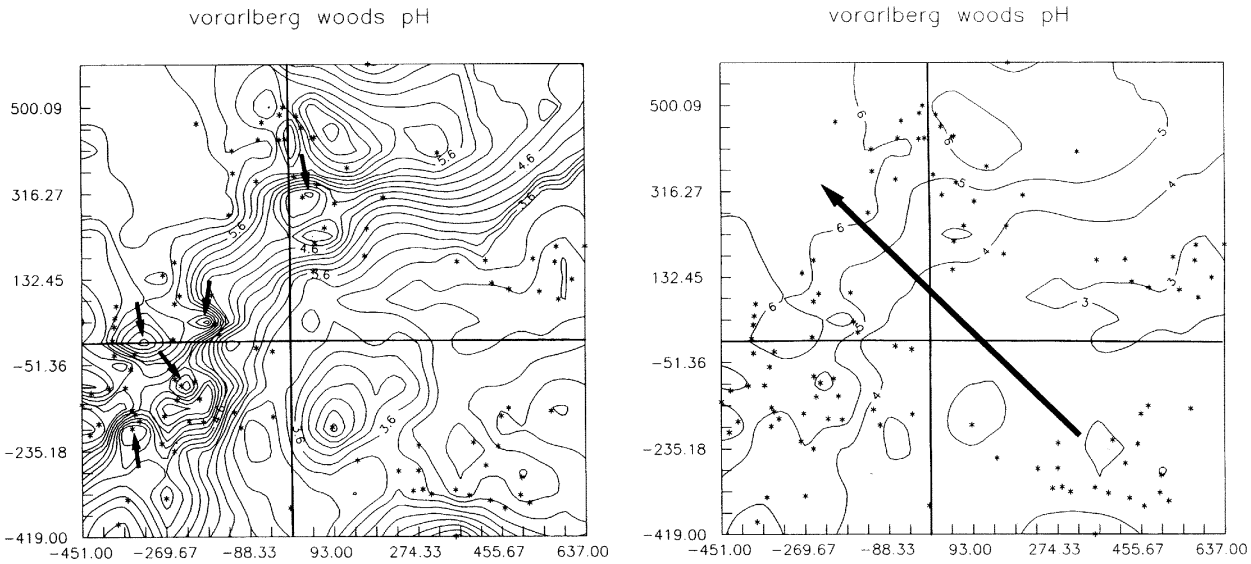


Fig. 4. A kriging analysis of the soil pH data (Vorarlberg woodlands) on the same ordination plane as in Fig. 1. Legend: (a) - high-density interpolation; the arrows point towards regions with extremely high original values of pH; (b) a thinned kriging map (the long arrow indicates the direction of generally increasing values of pH).

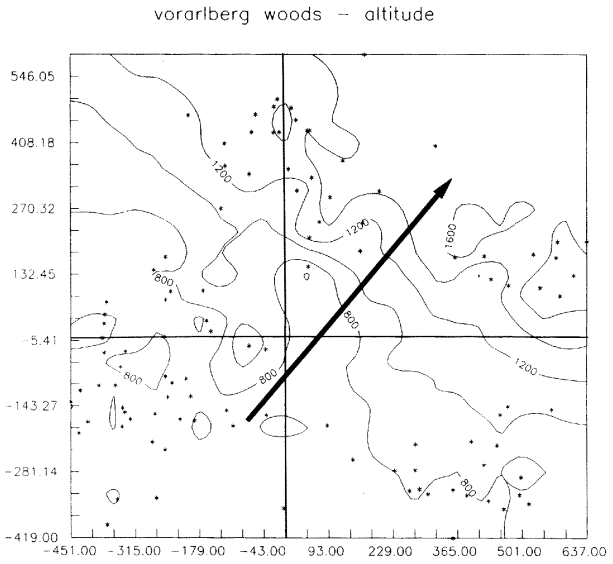


Fig. 5. A thinned kriging map of altitude (Vorarlberg woodland data). The long arrow indicates the direction of the increasing altitude.

$$\sum_{i=1}^{n} \lambda_i f_i^L - f_S^L = 0; \qquad L = 1, \ldots, k \qquad \text{(eqn. 30)}$$

now we can start to solve the whole system of equations from the bottom up, taking $f^1 \equiv 1$ we get as first constraint:

$$\sum_{i=1}^{n} \lambda_i - 1 = 0 \qquad \text{(eqn. 31)}$$

Using equation 31 we can write the error $(\hat{Z}_0 - Z_0)$ as a combination of the increments of the random function Z:

$$(\hat{Z}_0 - Z_0) = \sum_{i=1}^{n} \lambda_i Z(x_i) - Z_0 = \sum_{i=1}^{n} \lambda_i (Z(x_i) - Z_0) \qquad \text{(eqn. 32)}$$

Thus, the variance of this error may be expressed as

$$\text{var}\,[\hat{Z}_0 - Z_0] = -\sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \lambda_j \gamma_{ij} + 2\sum_{i=1}^{n} \lambda_i \gamma_{iS} - \gamma_{SS} \qquad \text{(eqn. 33)}$$

where

$\gamma_{ij} = \gamma(x_i - x_j)$ is the value of the variogram between two observation points x_i and x_j.

$\gamma_{iS} = \frac{1}{S}\int_S \gamma(x_i - x)\,dx$ is the mean value of the variogram between the point x_i and a point describing the domain S.

$\gamma_{SS} = \frac{1}{S^2}\int_S \int_S \gamma(x - x')\,dx\,dx'$ is the mean value of the variogram between two points independently sweeping over the whole domain S.

What is left is to minimize the quadratic form in equation 33 (under the k constraints of equation 30) to obtain the following linear (kriging) system:

$$\sum_{j=1}^{n} \lambda_j \gamma_{ij} + \sum_{L=1}^{k} \mu_L f_i^L = \gamma_{iS} \qquad i = 1, \ldots, n \qquad \text{(eqn. 34)}$$

$$\sum_{j=1}^{n} \lambda_j f_j^L = f_S^L \qquad L = 1, \ldots, k$$

To get the case where the domain is reduced to a point, we have only to substitute:

f_S^L by $f_0^L = f^L(x_0)$

γ_{iS} by $\gamma_{i0} = \gamma(x_i - x_0)$

γ_{SS} by $\gamma_{00} = \gamma(x_0 - x_0) = 0$

The solution of this system of n+k equations and n+k unknowns gives the n weights λ_i and the k Lagrange multipliers μ_L.

The weights λ_i are subsequently introduced into equation 25 to give the kriging estimate (=kriged value). As a by product, the estimation variance (=kriging variance) is obtained by substituting

$$\sum_{i=1}^{n} \lambda_i \gamma_{ij} = \gamma_{iS} - \sum_{L=1}^{k} \mu_L f_S^L \qquad \text{(eqn. 35)}$$

$$\sigma_k^2 = \text{var}\,[\hat{Z}_0 - Z_0] = \sum_{i=1}^{n} \lambda_i \gamma_{iS} + \sum_{L=1}^{k} \mu_L f_S^L - \gamma_{SS} \qquad \text{(eqn. 36)}$$

Advantages

1) The result of the kriging process is a map with interpolated values that coincide with the sampled data points. The weights obtained among neighboring data points get split, so the local density of data points has no influence on the interpolated results. The obtained

Tab. 1. The goodness-of-fit (the variation accounted for by a particular trend degree) for analysed data sets.

Trend degree	1	2	3	4	5
woodland data pH	0.55	0.60	0.62	0.65	0.68
woodland data altitude	0.69	0.76	0.77	0.78	0.78
dune meadows A1 horizon	0.37	0.40	0.51	0.78	—
dune meadows soil moisture	0.77	0.79	0.89	0.93	—

(a-indicates a A^2 of 1.0)

interpolations are in general better than ones obtained using moving weighted averages, because we do not have to think about what window size or weight to choose.

Furthermore, we obtain a map of the standard deviations of the estimations (or from the estimation variance) to provide ideas on where to intensify the sampling (Legendre and Fortin 1989).

2) The weights are tailored to the variability of the phenomenon of interest. A regular variable results in higher weights to the closest data points, an irregular variable results in dampening of the weights of the closest observed points.

3) When a calculated point x_0 coincides with an observed point x_i the solution of the system is $\lambda_i = 1$, $\lambda_j = 0$ for $j \neq i$.

This results in $\hat{Z} = Z(x_i)$ and $\sigma_k^2 = 0$. Thus kriging is an exact interpolator.

4) σ_k^2 is based upon the relationship between the observed data points and the points to be estimated only, therefore it does not depend on the real values of $z(x_i)$.

Problems

1) The method is not so well known (as TSA for example), therefore applications are not so widespread. Some packages (e.g. SURFER V3.0 from Golden Software Inc., 1987) provide the interpolation map, but no map of the deviations. Furthermore they do not say anything about how the weights (w_i) were obtained.

While the solutions for TSA are well known (and therefore it was relatively simple to write an application for it, see below), the exact formulations for kriging are hard to find (and indeed to derive) and so are the applications. Legendre and Fortin (1989) give a list of the packages available.

2) For the estimation of point values (as for average values over the whole lattice), there are as many kriging systems as there are points to be estimated. When all available data points are taken into account every time, only the right-hand side of the equations changes and the matrix of the kriging system needs to be inverted only once (we call that neighborhood unique).

A disadvantage of such an elegant solution is the limitation coming from the number of observations. Too many observations result in an unacceptable computing time (at least on small systems). To get around that problem one may compute the kriging equations for each point by using only 10-20 points of its immediate neighborhood (but that leads again into the problem of estimating the window size or number of points to involve).

3) There may be an advantage when using the values of the next few neighboring points only (the so called block kriging (see Burrough 1987) - versus the simple or point kriging, whose equation system has been solved above). Any nugget effect (C), that may be involved when using the whole sampling area, may be introduced only from some subregions of the area. Therefore such partitioning of the area will give better results for most blocks, leaving only the few with the chaotic component apart (but see the next point below).

4) Whenever a nugget effect (C) arises, the chaotic components of the phenomenon cannot be predicted (as they come probably from a smaller scale than our observation scale), as it is not spatially correlated. At any observed point, kriging gives the exact value (as measured, see above), but outside, where points are estimated, this chaotic element is filtered out (which may be exactly what we want).

Linear interpolation

Of all methods mentioned here this is the most simple and most crude one. The original data points are

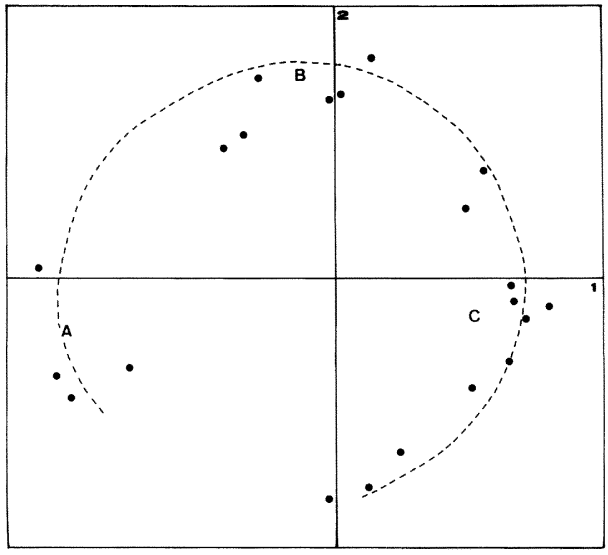
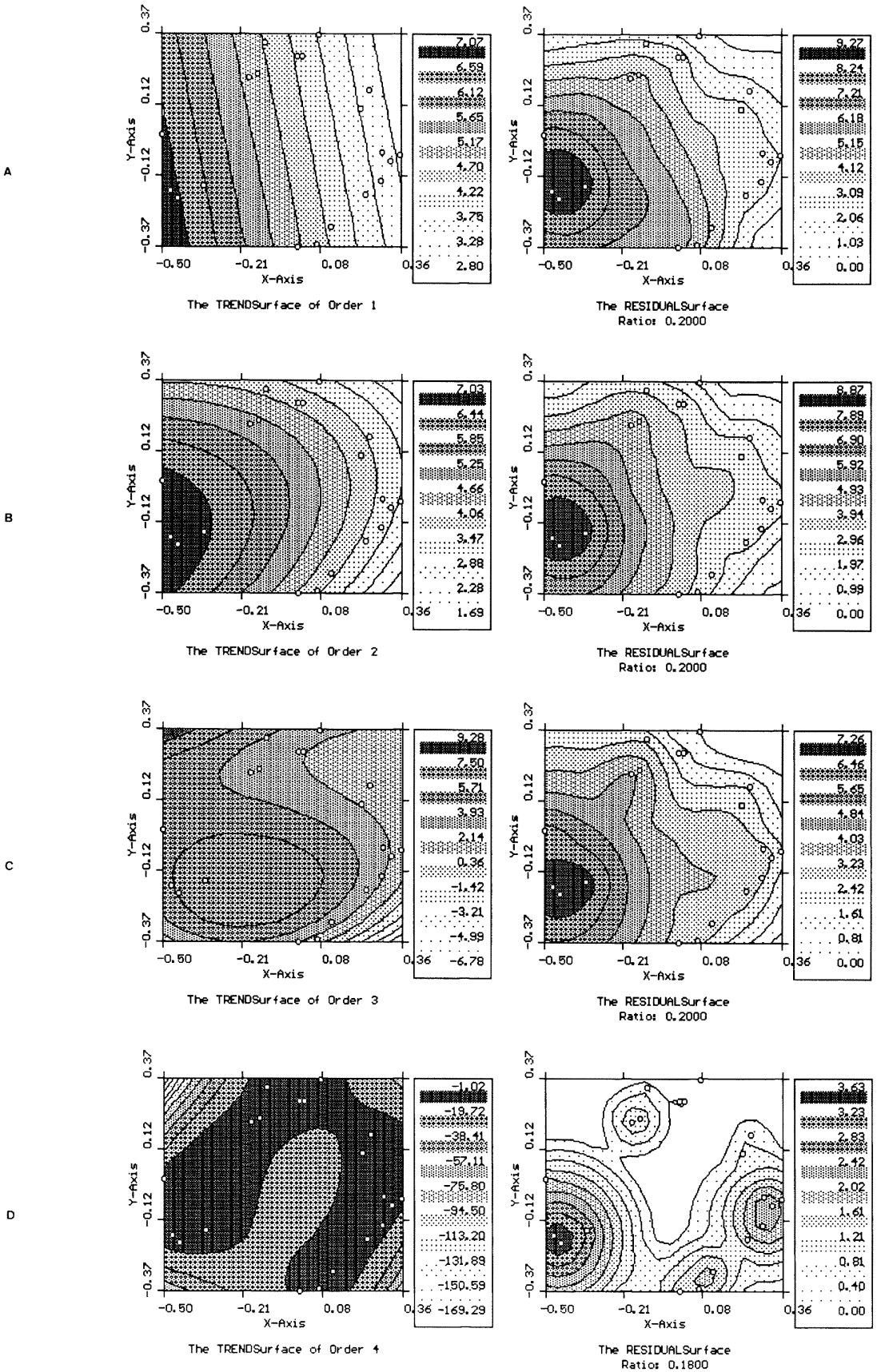


Fig. 6. Ordination plane of axes 1 and 2 of a Principal Coordinate Analysis of the floristic data on dune meadows from the Island of Terschelling (the Netherlands). The dotted line follows the coenocline, the signs A, B, and C indicate the position of loosely defined relevé clusters (see text for interpretation). The full circles lie in the ordinated positions of the 22 relevés.

Fig. 7. TSA on data on depth of A1 soil horizon on the PCOR ordination plane of the dune meadows. The weighting function in the residual maps was 0,2. The subfigures (a) through (d) stay for trend degrees 1 through 4, respectively.



connected using triangles (for more references about Delaunay triangulation see Ripley 1981, Watson 1982, Upton and Fingleton 1985), or, when the data has been sampled in a regular grid, one can connect the points using squares and eventually the diagonals as well.

Using a rectangular grid creates a problem when interpolating along the diagonals between the four corner points of any grid cell. The Z value at the point where the diagonals cross is the same for both diagonals only when the surface spanned between the 4 points is totally plain (=of 1st order). A high order surface will produce different Z-coordinates for each diagonal. Every reader can check this by bending up one corner of a sheet of paper and comparing the linear diagonals between the diagonally opposite corner points.

To overcome the problem there are three solutions:

1) interpolating only along the X or Y axis and not along the diagonals,

2) interpolating only along one diagonal direction (each small grid cell will be divided into two rectangular triangles) and ignoring the other diagonal,

3) interpolating along both diagonals and taking the mean of the 2 Z values obtained where the lines cross as the average Z value.

In most cases the resulting surface built from the original polygons will be informative enough (understood as a tool to view the overall structure of the data), without need to interpolate to obtain further points. Then the only interpolation that takes place is the construction of the polygons on a graphics-device, namely a computer monitor. In fact most surfaces constructed and visualized by computers are composed out of such squares or triangles. Only in the cases when there is need of Z values for subregions that are only sparse or not at all observed should one interpolate.

Advantages

The prime use for this technique is to obtain a rough and dirty overview of the behavior of the interesting variable in the observation plane. Extreme values can be easily detected (which gives us for instance an opportunity to handle them separately - or remove them - before feeding the data into further calculation processes).

Problems

Any interpolation done in a linear way may be justified only in the simplest cases (in means of expected surface), or when the sampled points are so dense that any more sophisticated calculation would not result in a significantly better result (this is a good justification of the law of simplicity).

Applications of trend-surface analysis and kriging

In order to illustrate the possibilities (some merits and flaws) of the trend-surface analysis and kriging we present two examples of ordination diagrams differing in the dispersion pattern of points on the ordination plane (without and with a horseshoe).

The handling of metric and ordinal interpolated data is also illustrated.

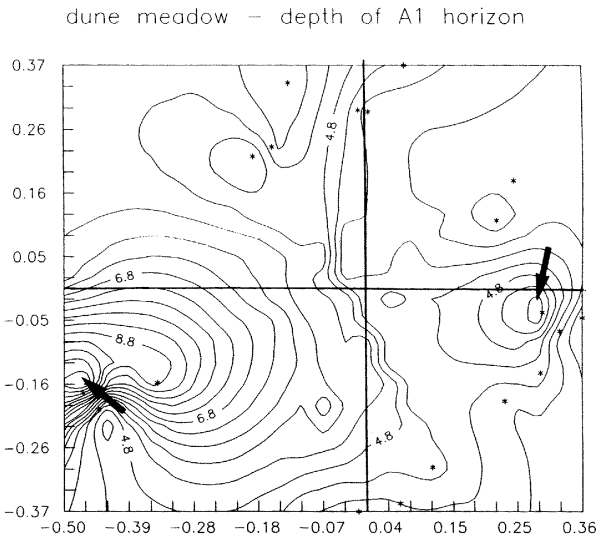


Fig. 8. A kriging map of the soil data (depth of A1 horizon) in the dune meadows. The arrows indicate the extremities in the fitted values. For details consult Fig. 2 and Tab. 1.

Example 1

Data

These data were collected in deciduous and mixed submontane to subalpine woods of Vorarlberg (western Austria). The scale of Braun-Blanquet (1964) was used in making phytosociological relevés (105 in total). The original abundance/dominance scale data were coded according to van der Maarel (1979). For each site data on topography and soil properties were collected (Mucina and Grabherr in prep.). In the present interpretations only pH of soil and altitude were considered.

Ordination method

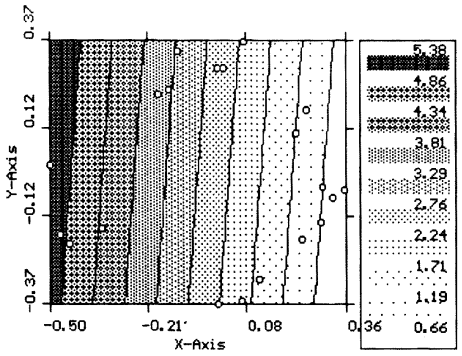
The transformed species-by-relevé matrix was ordinated by non-centered principal component analysis (PCA); the programe package CANOCO (ter Braak 1987) was used for computations. As the first axis of this variant of PCA is monotonic, only the diagram of axes 2 and 3 was considered.

Interpretation

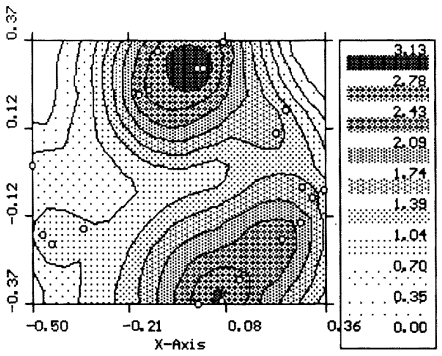
Each relevé on the ordination diagram (Fig. 1) was assigned to a class (syntaxon) according to a previous classification of the Vorarlberg woods by Grabherr and

Fig. 9. TSA for soil-moisture data (ordinal scale) for the dune meadow PCOR ordination. For details consult Fig. 2 and Tab. 1.

A

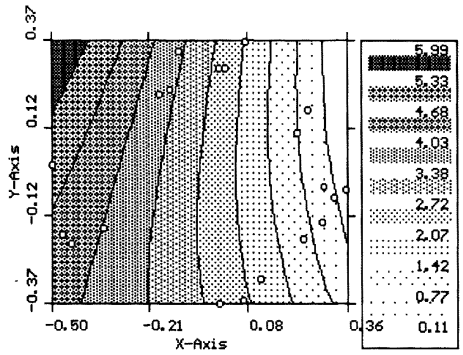


The TRENDSurface of Order 1

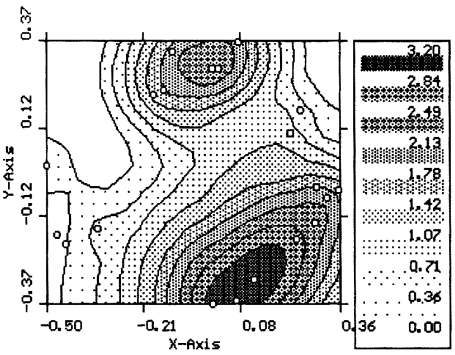


The RESIDUALSurface
Ratio: 0.3000

B

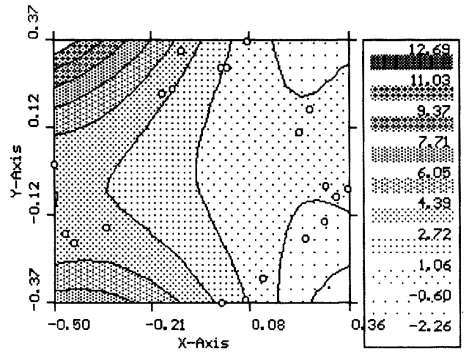


The TRENDSurface of Order 2

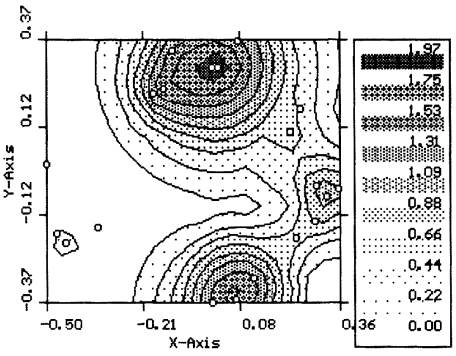


The RESIDUALSurface
Ratio: 0.3000

C

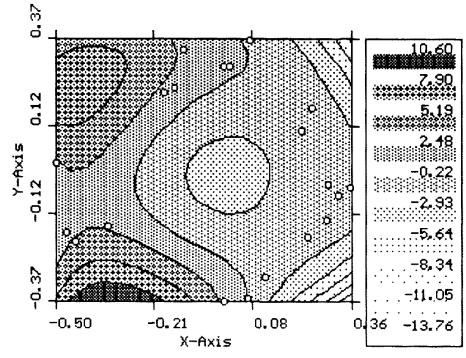


The TRENDSurface of Order 3

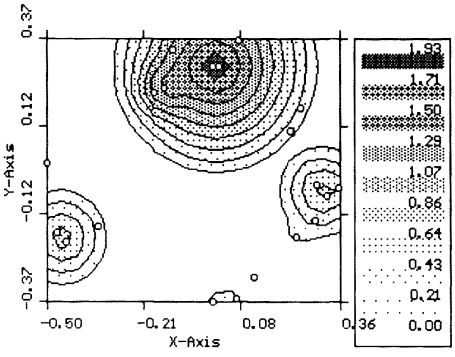


The RESIDUALSurface
Ratio: 0.3000

D



The TRENDSurface of Order 4



The RESIDUALSurface
Ratio: 0.3000

Mucina (1989). Basically 4 well-separable clusters (A through D) were recognized. These can be identified with subalpine needle woods (Homogyno-Piceetum: cluster A), plateau fir woods (Bazzanio-Abietetum: cluster B), mixed montane beech-spruce woods (Abieti-Fagetum sensu lato: cluster C) and tall-herb spruce woods (Adenostylo-Piceetum sensu lato: cluster D).

Each cluster occupies a separate quarter on the ordination diagram. TSA on the superimposed values of pH in locations of relevés on the ordination diagram shows already by trend 1 (Fig. 2a) a very good fit (55%) with a clear trend from clusters A and B towards clusters C and D. The former two clusters represent woods on acid soils. The 3rd trend-degree (Fig. 2b) shows even better fit (62%), but the direction and shape of the fitted surfaces remained the same. The direction of trend surface 5 (Fig. 2c) did slightly change, but the overall fit increase was negligible. The residuals (Fig. 2) show high values for all fitted trends especially within cluster C and less pronouncedly in cluster D. Both clusters, unlike A and B, contain associations occurring both on calcium rich and calcium poor soils.

The trend analysis of altitude as the superimposed external variable shows an analogous pattern (Fig. 3): already the first trend degree explains much information (almost 69%), the direction of the trend surfaces is identical for all analysed degrees (1 though 5), and the highest residuals are encountered in locations in clusters C and D - again being rather heterogeneous in terms of altitudinal position of relevés.

In summary, a TSA analysis of this type of ordination patterns was very effective which can be judged from (a) the constant direction of the trend surfaces regardless of the trend degree, i.e. except for coefficient b_0 the remaining coefficients approach zero, and from (b) negligible increase of goodness-of-fit along the gradient of the trend degrees (Tab. 1).

A kriging analysis of both soil reaction (Fig. 4) and altitude (Fig. 5) data reveals the same patterns as the TSA when a proper "resolution grain" is selected. Too detailed isolining (Fig. 4a and b) makes the reading of the map more complicated. Inspection of Fig. 4a, specifically the upper right (position of cluster D in Fig. 1) and lower left (position of cluster C in Fig. 1) quarters of the diagram, reveals the presence of local data anomalies which have already been identified as sources of residuals by the TSA.

Example 2

Data

This data set contains 20 phytosociological relevés from dune meadows on Terschelling Island (the Netherlands). It was used as a reference data set in testing various multivariate methods by Jongman et al. (1987, p. xvii-ixx). The relevé scale and its transformation of data was carried out as described in Example 1.

Of several environmental data the thickness of the A1 horizon (measured on a metric scale) and site-moisture value (estimated on a 5-degree ordinal scale) were used as external variables.

Ordination method

Principal coordinate analysis (PCOR; see eg., Pielou 1984 for details) was performed on the data set using Similarity Ratio (eg. Hajdu 1982) as the resemblance measure. The computations were accomplished with the programe PRINCOOR from package SYNTAX III (Podani 1988).

Interpretation

The ordination plane of axes 1 and 2 (Fig. 6) shows a clear horseshoe (see Wartenberg et al. 1987 and Peet et al. 1988 for discussions) composed of 3 clusters. The cluster A occupies the negative pole of axis 1. It is characterized by wetland species such as Ranunculus flammula and Eleocharis palustris. Cluster B is outlying on the positive pole of axis 2 and comprises relevés with Alopecurus geniculatus and Sagina procumbens. The biggest cluster C lies on the opposite pole to cluster A and includes species such as Plantago lanceolata, Anthoxanthum odoratum and Achillea millefolium. The sequence of these clusters makes up a coenocline ranging from wetland (A) to mesophilous grasslands (C).

Unlike Example 1, the 1st-degree trend of the A1-horizon depth (Fig. 7a) accounts for only approx. 37% of variation. The trends of degrees 2 and 3 (Figs. 7b and c, respectively) show considerable linear increase of explained variation which then rises dramatically by computing the 4th-degree trend. In the latter image (Fig. 7d) the optically identified horseshoe structure was identified. The highest residuals are at positions of relevés 11 and 14 (situated on opposite poles of axis 1; Fig. 7d). The values of A1 horizon depth were, indee, extraordinarily high. Generally there is an trend of increasing depth of A1 horizon from wetland towards mesic grasslands.

The image of kriging analysis (Fig. 8) is less informative. This isoline map mirrors well the spatial pattern of extrapolated values as well as the position of the relevés with an extraordinarily deep upper soil horizon, but the overall trend is rather unclear.

Ordinal data (if handled as metric) can also be used in the trend-surface analysis, inverse-weighting interpolation or kriging. However, the interpolated values between the integer values must not be interpreted. The interpolation analyses can serve only to detect boundaries separating regions with data points scoring in the ordinal values. Therefore, for the purpose

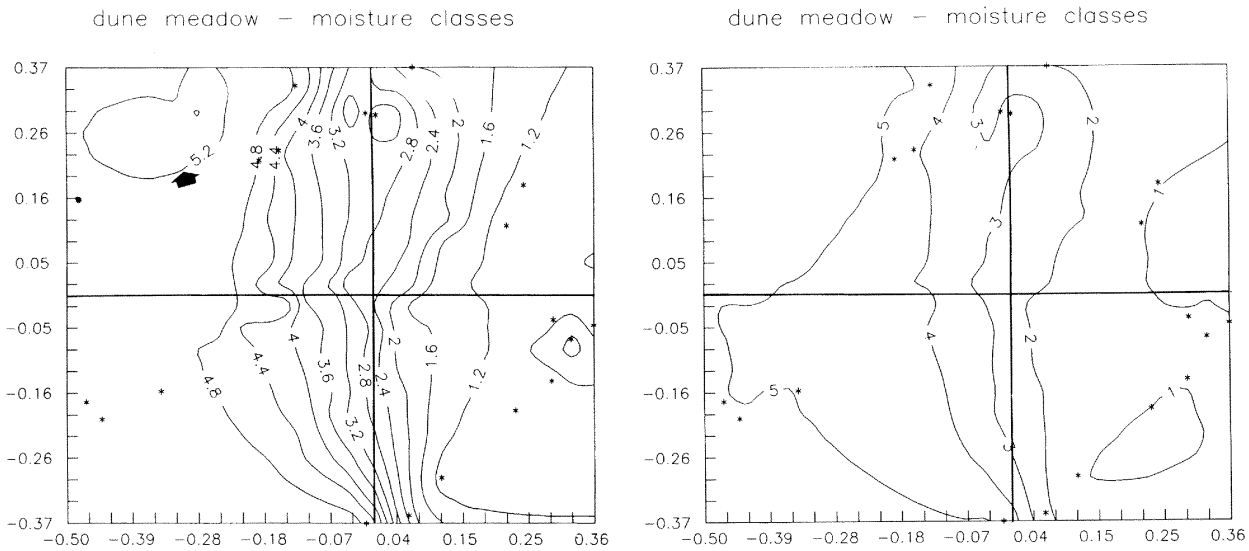


Fig. 10. Kriging maps for the fitted soil-moisture data on the ordination of the dune meadow data. (a) - high-density line map showing the artefact values (small arrow). (b) thinned map adjusted to integer values.

of illustration, the limits or isolines to show trends should be given in the units of the ordinal scale.

Soil-moisture value is such an ordinal variable in the dune meadows data set. The 1st degree trend (Fig. 9a) accounted for a large portion of information (app. 77%) and revealed a clear linear trend correlating negatively with axis 1 of the ordination (Fig. 6). The 2nd-degree trend (Fig. 9b) did not show much change, but on the level of the subsequent trend degrees the horseshoe effect showed an impact on the trend pattern (Fig. 9c and d).

Although the goodness-of-fit increased to 93% (5th degree), the interpretability of the detected pattern did not improve.

Kriging analysis of the soil-moisture data was performed with various densities of interpolated isolines (Fig. 10). The latter figure shows clearly the importance of appreciation of the ordinal character of the data. When the isolines are not settled to correspond to actual values of the ordinal data, "overprecision" phenomena emerge: a part of the interpolated region (Fig. 10a) shows interpolated values above 5.2 which is an artefact as the original values did not exceed 5. Fig. 10b shows a more realistic image of the interpolated pattern as the isolines were settled equivalent to the ordinal scale.

Appendix

As we has no rigorous, user-friendly software for TSA available, the first author of this paper programmed two applications for that use.

The first, named TSA, computes trend surfaces up to the 10th order using the input of up to 238 data-points (an expansion for more data-points is planned).

The input data is read in from a free-format file.

The outputs are a complete report file (with diagrams in text-mode), and in addition - for each order computed - an input-file for the second program, TSAGRAPH, the viewing-program. It runs completely under a graphical-user-interface (mouse-interface, menu-bar, windows, dialog-boxes). The screen is divided into 2 windows, the left one displays a map of the trend surface, the right one a map of the residuals. Free choice of the ratio of influence in the residual-map allows one to spread out (and therefore to view in the form of a map) the goodness of fit for every observed point.

Almost every parameter (size, location and position of both maps, legends, markers and lines for the axes, and surface-levels) is computed while loading a new data-set, but can be changed by the user at any stage of the analysis.

A printout can be made using a Epson FX (compatible) printer (as we did for the graphs in this paper).

Both programs are at the moment available only for the IBM-PC (or compatible) under DOS 2.1 (or higher), with a coprocessor installed. A harddisk is recommended. For TSAGAAPH a Hercules-, EGA- or VGA-screen and - card and a Microsoft (compatible) mouse is necessary. Programming has been done using Turbo Pascal V5.5 (some routines in Turbo Assembler V1.0).

Parting to other environment (Apple Macintosh) is partly under way.

As it is the intention of the first author to produce good, user-friendly software that anyone can afford, these two programs (and other software) not mentioned here) are distributed as SHAREWARE.

Any reader interested in obtaining further details should contact him at the address mentioned above.

Acknowledgements. The authors thank Georg Grabherr for permission to use his Vorarlberg woodland data.

REFERENCES

AUSTIN, M.P. 1977. Use of ordination and other multivariate descriptive methods to study succession. Vegetatio 35: 165-175.

BIRKS, H.J.B. 1987. Recent methodological developments in quantitative descriptive biogeography. Ann. Zool. Fennici, Helsinki, 24: 165-178.

BRAUN-BLANQUET, J. 1964. Pflanzensoziologie. Grundzüge der Vegetationskunde. 3. Aufl. Springer-Verlag., Wien.

BURDEN, R.L. and J.D. FAIRES. 1985. Numerical analysis. 3rd ed. Prindle, Weber and Schmidt, Boston.

BURROUGH, P.A. 1987. Spatial aspects of ecological data. In: Jongman, R.H.G. ter Braak, C.J.F. and van Tongeren, O.F.R. (eds.), Data analysis in community and landscape ecology. pp. 213-251. Pudoc, Wageningen.

CLIFF, A.D. and J.K. ORD. 1981. Spatial processes: models and applications. Pion, London.

DARGIE, T.C.D. 1984. On the integrated interpretation of indirect site ordinations: a case study using semi-arid vegetation in southeastern Spain. Vegetatio 55: 37-55.

DELHOMME, J.P. 1978. Kriging in the hydrosciencs. Adv. Water Res. 1: 251-266.

DIGBY, P.G.N. and R.A. KEMPTON. 1987. Multivariate analysis of ecological communities. Chapman and Hall, London.

ELLENBERG, H. 1979. Zeigerwerte der Gefässpflanzen Mitteleuropas. Verlag Erich Goltze KG, Göttingen.

FEOLI, E. 1983. Predictive use of classification and ordination methods in plant community ecology. A summary with examples. In: Ferrari, C. Gentile, S. Pignatti, S. and Poli Marchese, E. (eds.), Le comunità vegetali come indicatori ambientali, pp. 83-108. Regione Emilia-Romana e Società Italiana di Fitosociologia, Bologna.

FEOLI-CHIAPELLA, L. and E. FEOLI. 1977. A numerical phytosociological study of the summits of the Majella massive (Italy). Vegetatio 34: 21-29.

FORTIN, M.J., P. DRAPEAU and P. LEGENDRE. 1989. Spatial autocorrelation and sampling design in plant ecology. Vegetatio 83: 209-222.

GITTINS, R. 1968. Trends-surface analysis of ecological data. J. Ecol. 56: 845-869.

GRABHERR, G. and L. MUCINA. 1989. Übersicht der Wälder und Waldstandorte in Vorarlberg. Lebensraum Vorarlberg, Bregenz, 3: 9-46.

HAJDU, L.J. 1981. Graphical comparison of resemblance measures in phytosociology. Vegetatio 48: 47-59.

HILL, M.O. 1974. Correspondence analysis: a neglected multivariate method. Appl. Stat. 23: 340-353.

JONGMAN, R.H.G. ter BRAAK, C.J.F. and O.F.R. VAN TONGEREN. 1987. Data analysis in community and landscape ecology. Pudoc, Wageningen.

KRIGE, D.G. 1966. Two dimensional weighted moving average trend surfaces for are evaluation. J.S. Afr. Inst. Min. Metall. 66: 13-38.

LEGENDRE, L. and P. LEGENDRE. 1983. Numerical ecology. Elsevier, Amsterdam.

LEGENDRE, P. 1990. Quantitative methods and biogeographic analysis. In: Garbary, D. and South G.R. (eds.), Evolutionary biogeography of the marine algae of the North Atlantic. pp. 1-25. Springer-Verlag, Berlin.

LEGENDRE, P. and M.J. FORTIN. 1989. Spatial pattern and ecological analysis. Vegetatio 80: 107-138.

LEGENDRE, P. and L. LEGENDRE, (eds.). 1987. Developments in numerical ecology. Springer Verlag, Berlin.

LUDWIG, J.A. and J.F. REYNOLDS. 1988. Statistical ecology. A primer on methods and computing. J. Wiley and Sons, New York.

MATHERON, G. 1962. Traité de géostatistique appliquée. Tome 1. Éditions Technip, Paris.

MATHERON, G. 1971. The theory of regionalized variables and its applications. Cah. Centre Morphol. Mathem. de Fontainebleu.

MILES, J., W. SCHMIDT and E. VAN DER MAAREL (eds.). 1988. Temporal and spatial patterns of vegetation dynamics. Kluwer, Dordrecht.

MUCINA, L. 1982. Numerical classification and ordination of ruderal plant communities (Sisymbrietalia, Onopordetalia) in the western part of Slovakia. Vegetatio 48: 267-275.

MUCINA, L. 1989. Syntaxonomy of the Onopordum acanthium communities in temperate and continental Europe. Vegetatio 81: 107-115.

MUCINA, L. and M.B. DALE (eds.). 1989. Numerical syntaxonomy. Kluwer, Dordrecht.

MUCINA, L. and S. POLÁČIK. 1982. Principal components analysis and trend surface analysis of a small-scale pattern in a transition mire. Vegetatio 48: 165-173.

MUCINA, L. and E. VAN DER MAAREL. 1989. Twenty years of numerical syntaxonomy. Vegetatio 81: 1-15.

MUCINA, L., V. ČÍK and P. SLAVKOVSKÝ. 1988. Trend surface analysis and splines for pattern determination in plant communities. Coenoses 3: 89-105.

MUCINA, L. and G. GRABHERR. (in prep.). Ecology of montane and subalpine woods in Vorarlberg western Austria.

NICHOLS, S. 1977. On the interpretation of principal components analysis in ecological contexts. Vegetatio 34: 191-197.

ORLÓCI, L. 1978. Multivariate analysis in vegetation research. 2nd ed. Dr. W. Junk, The Hague.

PEET, A.K., R.G. KNOX, J.S. CASE and R.B. ALLEN. 1988. Putting things in order: the advantages of detrended correspondence analysis. Am. Nat. 131: 924-934.

PERSSON, S. 1981. Ecological indicator values as an aid in the interpretation of diagrams. J. Ecol. 69: 71-84.

PICKETT, S.T.A. and J. KOLASA. 1989. Structure of theory in vegetation science. Vegetatio 83: 7-15.

PIELOU, E.C. 1984. The interpretation of ecological data. A primer on classification and ordination. J. Wiley and Sons, New York.

PODANI, J. 1988. SYN-TAX III. User's manual. Abstr. Bot., Budapest, 12: 1-183.

RIPLEY, B.D. 1981. Spatial statistics. J. Wiley and Sons, New York.

TER BRAAK, C.J.F. 1987. CANOCO - a FORTRAN program for canonical community ordination by (partial) (detrended) (canonical) correspondence analysis, principal components analysis and redundancy analysis (version 2.1). TNO Institute of Applied Computer Science, Wageningen.

TER BRAAK, C.J.F. and C.W.N. LOOMAN. 1987. Regression. In: Jongman, R.H.G. ter Braak C.J.F. and van Tongeren O.F.R. (eds.), Data analysis in community and landscape ecology. pp. 29-77. Pudoc, Wageningen.

UNWIN, D. 1975. An introduction to trend surface analysis. Concepts Tech. Mod. Georg. 5: 1-40.

UPTON, G.J. and B. FINGLETON. 1985. Spatial data analysis by example. Volume 1. Point pattern and quantitative data. J. Wiley and Sons, Chichester.

VAN DER MAAREL, E. 1969. On the use of ordination models in phytosociology. Vegetatio 19: 21-46.

VAN DER MAAREL, E. 1979. Transformation of cover-abundance values in phytosociology and its effects on community similarity. Vegetatio 39: 97-114.

VAN DER MAAREL, E. 1980. On the interpretability of ordination diagrams. Vegetatio 42: 43-45.

VAN DER MAAREL, E. 1989. Theoretical vegetation science on the way. Vegetatio 83: 1-6.

VAN DER MAAREL, E, I. ESPEJEL and P. MORENO-CASASOLA. 1987. Two-step vegetation analysis based on very large data sets. Vegetatio 68: 139-143.

VAN GROENEWOUD, H. 1965. Ordination and classification of Swiss and Canadian coniferous forests by various biometric and other methods. Ber. Geobot. Inst. ETH, Stiftung Rübel 36: 28-102.

WARTENBERG, D. 1985. Canonical trend surface analysis: a method for describing geographic patterns. Syst. Zool. 34: 259-279.

WARTENBERG, D., S. FERSON and F.J. ROHLF. 1987. Putting things in order: a critique of detrended correspondence analysis. Am. Nat. 129: 434-448.

WATSON, D.F. 1982. ACORD: Automatic contouring of raw data. Comp. Geosci. 8: 97-101.

Manuscript received: October 1991