All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices

Daniel Borcard *, Pierre Legendre

Département de sciences biologiques, Université de Montréal, C.P. 6128, succ. Centre-ville, Montreal Quebec, Canada H3C 3J7

Abstract

Spatial heterogeneity of ecological structures originates either from the physical forcing of environmental variables or from community processes. In both cases, spatial structuring plays a functional role in ecosystems. Ecological models should explicitly take into account the spatial structure of ecosystems. In previous work, we used a polynomial function of the geographic coordinates of the sampling sites to model broad-scale spatial variation in a canonical (regression-type) modelling context. In this paper, we propose a method for detecting and quantifying spatial patterns over a wide range of scales. This is obtained by eigenvalue decomposition of a truncated matrix of geographic distances among the sampling sites. The eigenvectors corresponding to positive eigenvalues are used as spatial descriptors in regression or canonical analysis. This method can be applied to any set of sites providing a good coverage of the geographic sampling area. This paper investigates the behaviour of the method using numerical simulations and an artificial pseudo-ecological data set of known properties. © 2002 Elsevier Science B.V. All rights reserved.

Keywords: Geographic distances among sampling sites; Principal coordinate analysis; Variation partitioning; Scale; Spatial analysis

1. Introduction

In ecological theory, a major paradigm states the importance of spatial structure, not only as a potential nuisance for sampling or statistical testing, but also as a functional necessity, to be studied for its own sake and included into ecological modelling (Legendre and Fortin, 1989; Legendre, 1993; Legendre and Legendre, 1998). In the framework of multivariate data analysis, several methods have been proposed to include space as

an explicit predictor. Legendre and Troussellier (1988) used a matrix of Euclidean (geographic) distances among their sampling sites in a series of Mantel and partial Mantel tests. Legendre (1990) proposed using geographic coordinates directly as explanatory variables in constrained ordination techniques (redundancy analysis, RDA, and canonical correspondence analysis, CCA), by placing the terms of a cubic trend-surface equation into the explanatory (i.e. constraining) data matrix. This approach, called multivariate trend-surface analysis, was later integrated into a method of variation partitioning, where ecological variation was decomposed into four fractions (pure environmental, pure spatial, explained both

---

* Corresponding author. Fax: + 1-514-343-2293.
  E-mail address: borcardd@magellan.umontreal.ca (D. Borcard).

by space and environment, and unexplained) using partial constrained ordination (Borcard et al., 1992; Borcard and Legendre, 1994; Méot et al., 1998). This technique, which was summarised by Legendre and Legendre (1998), Section 13.5, has proved very successful and is now widely applied in various fields of ecology (see references in Legendre and Legendre (1998), p. 775).

The coarseness of trend-surface analysis presents a problem, however. This method is devised to model broad-scale spatial structures with simple shapes like planes, saddles, or parabolas representing bumps or troughs. Finer structures cannot be adequately modelled by this method: too many parameters would be required to do so.

In recent years, researchers have increased their awareness of the fact that ecological processes occur at defined scales, and that their perception depends upon a proper matching of the sampling strategy to the size, grain and extent of the study, and the statistical tools used to analyse the data. This has generated the need for analytical techniques devised to reveal the spatial structures of a data set at any scale that can be perceived by the sampling design. In this paper, we propose a method for detecting and quantifying spatial patterns over a wide range of scales. This method can be applied to any set of sites providing a good coverage of the geographic sampling area. This method will first be presented in the unidimensional context, where it has the further advantage of being usable even for short ($n > 25$) data series.

2. The method

The analysis begins by coding the spatial information in a form allowing us to recover various structures over the whole range of scales encompassed by the sampling design. This technique will work on data sampled along linear transects as well as on geographic surfaces or in three-dimensional space. This paper will focus on the unidimensional case, demonstrating the efficiency of the method by way of simulations of simple and complex data.

In the framework of linear modelling, the most straightforward technique for modelling spatial structures is polynomial regression (trend-surface analysis in the bidimensional case), where the spatial variables are used to generate a polynomial function of the X (or X and Y, or X, Y and Z) coordinates of the sampling units (Legendre, 1990; Borcard et al., 1992; Borcard and Legendre, 1994). For a linear transect, using the X coordinates of the sampling units as an explanatory variable allows one to model a linear trend that may be present in the data. Adding a second-order (X^2) monomial term allows the model to be bent once in the form of a parabola. Each higher-order term generates one more bend, and hence increases the fit of the model to finer-scale spatial structures. One major problem with this approach is that the individual terms are highly correlated, thereby preventing the modelling of independent structures at different scales. Furthermore, especially in the bidimensional case, the number of terms of the polynomial function grows very quickly, making the third order (with nine terms) the highest one to be usable practically, despite its coarseness in terms of spatial resolution. Polynomials can be turned into orthogonal polynomials, either by using a Gram–Schmidt orthogonalization procedure, or by carrying out a principal component analysis (PCA) on the matrix of monomials. A new difficulty arises: each new orthogonal variable is a linear combination of several (in the case of the Gram–Schmidt orthogonalization) or all (in the case of PCA) the original variables; it does not represent a single scale any longer. To solve these problems, our new approach has a different starting point involving the close neighbourhood relationships among the sampling sites.

2.1. Modified matrix of Euclidean distances

Fig. 1 displays the steps of a complete spatial analysis using the new method based on principal coordinates of neighbour matrices (PCNM).

First, we construct a matrix of Euclidean distances among the sites. Then, we define a threshold under which the Euclidean distances are kept as measured, and above which all distances are considered to be 'large', the corresponding numbers being replaced by an arbitrarily large value. For a reason explained later, this large

value can be set equal to four times the threshold value. For instance, in the case of a linear transect made of sampling points regularly spaced 1 m apart, we could set the threshold at 1 m to retain only the closest neighbours, and replace all other distances in the matrix by $1.0 \text{ m} \times 4 = 4.0 \text{ m}$.

Contrary to the polynomial regression/trend-surface analysis approach, our idea is to reconstruct the spatial patterns by starting from the fine-scale relationships instead of the broad-scale trends.

Fig. 1. Principle of the new method of spatial analysis based on principal coordinate analysis of a truncated matrix of Euclidean (geographic) distances among sampling sites (PCNM).

2.2. Principal coordinate analysis

The second step is to compute the principal coordinates of the modified distance matrix. This is necessary because we need our spatial information to be represented in a form compatible with applications of multiple regression or canonical ordination (RDA or CCA) i.e. as an object-by-variable matrix. We obtain one or several null, and several negative eigenvalues. Principal coordinate analysis of the truncated distance matrix makes it impossible to represent the distance matrix entirely in a space of Euclidean or complex coordinates. The negative eigenvalues cannot be used as such because the corresponding axes are complex (i.e. the coordinates of the sites along these axes are complex numbers). In any case, the positive eigenvalues represent the Euclidean components of the neighbourhood relationships of our truncated matrix; these are the components that are of interest to us. Of course, one could correct for the negative eigenvalues using one of the methods described in Gower and Legendre (1986) or Legendre and Legendre (1998). These methods consist in adding a constant either to the original non-diagonal distances, or to the squared non-diagonal distances in the matrix. By doing so, one increases all distances and changes the reconstructed spatial arrangement of the sampling sites. We have empirically compared the results obtained using either the original real-number axes corresponding to positive eigenvalues, or all axes after correction for negative eigenvalues. Results indicate that a good reconstruction of the spatial structures is obtained by using the former method, i.e. only the axes corresponding to positive eigenvalues, without correcting the axes having negative eigenvalues.

The principal coordinates derived from these positive eigenvalues can now be used as explanatory variables in multiple regression, RDA, or CCA, depending on the context.

To investigate the process of truncation of the distance matrix, explained above, we built a series of truncated distance matrices with 'large distance' values made of a series of factors ranging from 2 to 8. We observed that beyond a factor of four times the threshold for the 'large' distances,

the principal coordinates remain the same to within a multiplicative constant. In other words, the first principal coordinate obtained with a factor of 4 had a correlation of 1.0 with the first principal coordinate obtained with a factor of 5, or any other value; the same was true for the whole series of principal coordinates. Consequently, multiple regressions using principal coordinates obtained with a multiplicative constant of 4 and above will yield the same R^2 and the same P-values as with any other multiplicative constant larger than 4. Thus we decided to apply the factor 4 in all subsequent steps of our investigation.

When computed from a distance matrix corresponding to n equidistant objects arranged as a straight line, as in Fig. 1, truncated with a threshold of one unit (MAX = 1 i.e. only the immediate neighbours are retained), the principal coordinates correspond to a series of sinusoids with decreasing periods (Fig. 2); the largest period is $n + 1$, and the smallest one is equal to or slightly larger than 3. The number of principal coordinates is a round integer corresponding to two-thirds of the number of objects. If the truncation threshold is larger than 1, fewer principal coordinates are obtained, and several among the last (finer) ones are distorted, showing aliasing of structures having periods too short to be represented adequately by the discrete site coordinates. This behaviour will later be shown to have important consequences on the performance of the method. Thus, our method presents a superficial resemblance to Fourier analysis and harmonic regression, but it will be shown to be more general since it can model a wider range of signals, and can be used with irregularly spaced data.

3. Numerical simulations

We conducted extensive simulations to explore the behaviour of the method regarding type I error and power to detect various kinds of signals (Fig. 3); power was measured as the rate of rejection of the null hypothesis, at significance level 0.05, when an effect was present in the data. The simulation setup consisted of a straight line with 100 equidistant points representing the sampling
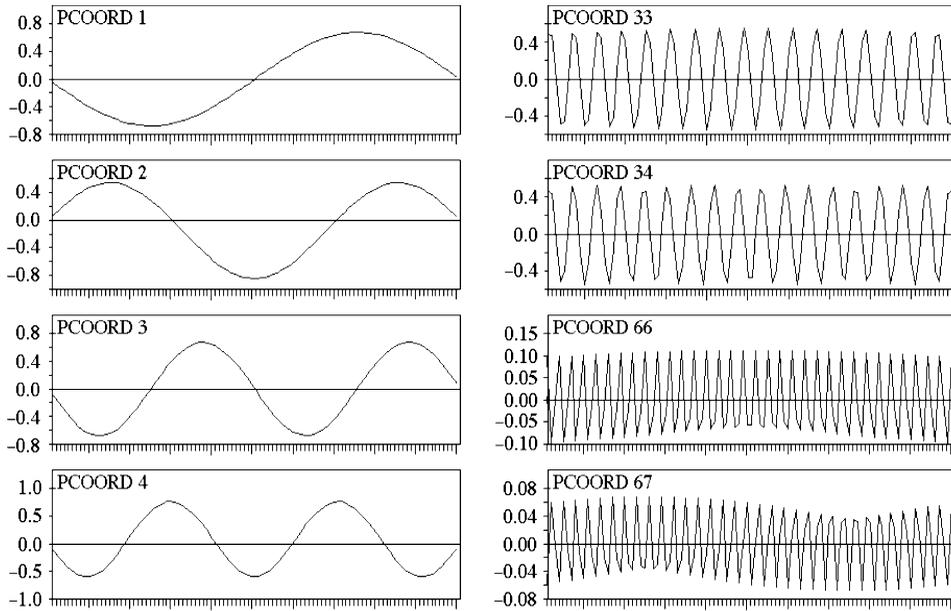
Fig. 2. Eight of the 67 principal coordinates obtained by principal coordinate analysis of a matrix of Euclidean distances among 100 objects, truncated after the first neighbours (MAX = 1).
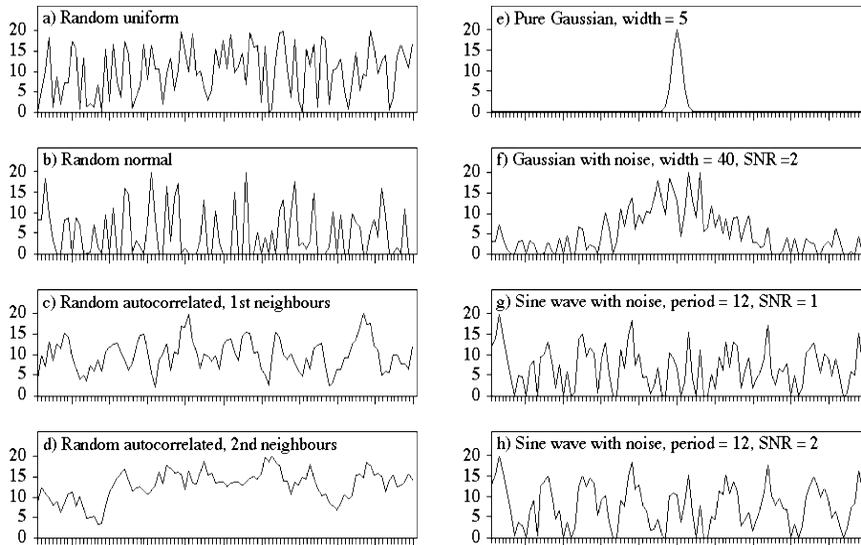
Fig. 3. Eight examples of dependent variables used in the simulation study.

sites. We experimented with the method using multiple regression (i.e. with a single dependent variable at a time), for various types of dependent variables. Every simulation run, described below,

consisted of 5000 independent analyses. The global statistical significance of each analysis was tested by a permutation test involving 999 permutations. Notwithstanding the fact that one can

never simulate all possible or relevant ecological situations, the following results are presented as support for the sensitivity of the method described in this paper.

For statistical testing, we used the method of permutation of residuals under a full model (ter Braak, 1990, 1992; Anderson and Legendre, 1999); in this method, the permutable units are the residuals of the multiple regression. As a test statistic for the global test, we used the R^2 of the multiple regression; within any given permutation test, the values of R^2 and F are monotonic to each other and, thus, represent equivalent statistics for permutation testing. In applications on complex data involving tests of individual regression coeffi-

cients, we also tested by permutation the t-statistic associated with each regression coefficient.

3.1. Random variables and type I error

The first series of simulations focussed on the type I error of the method. The dependent variables were random numbers drawn from four different distributions: uniform (Fig. 3a), normal (Fig. 3b), exponential, and (as an extreme case) exponential cubed, following Manly (1997) and Anderson and Legendre (1999). Fig. 4 shows the results of these runs, carried out using two different truncation thresholds (value MAX) of the spatial matrix. Several independent series of 5000
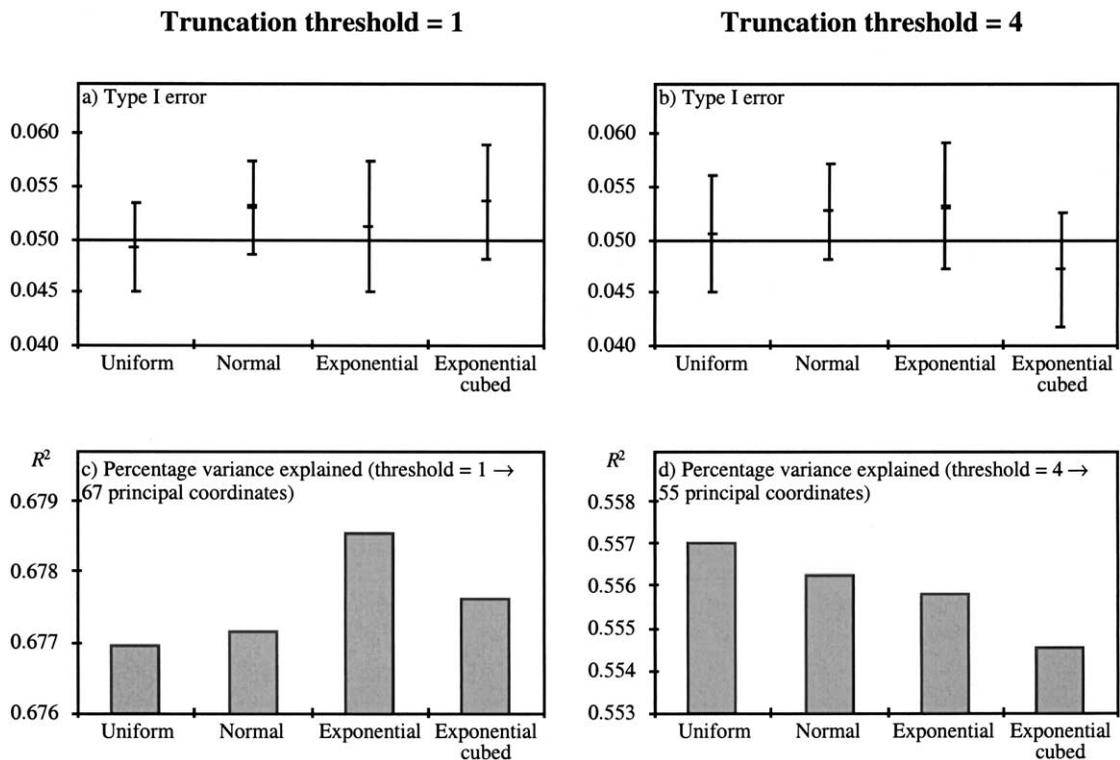
Fig. 4. Type I error of the method (a, b) and percentage of variance explained (R^2: c, d) on series of 100 data points randomly drawn from four different distributions, with spatial matrices with two different resolutions set by the truncation threshold (value MAX) of the Euclidean distance matrix. Each run consisted of 5000 independent simulations. The error bars in (a) and (b) represent 95% confidence intervals.

simulations showed that the frequency distribution of the *P*-values was approximately flat (results not illustrated), yielding an appropriate number of type I error cases (Fig. 4a and b). Thus, the method showed good performance on this crucial aspect. Note also the percentage of variance 'explained' (non-significantly): in linear regression, when the dependent variable is completely random, as it is the case in these simulations, the expected value of R^2 is equal to the ratio between the number of explanatory variables and the number of objects minus 1. The results presented in Fig. 4c and d are right on the spot.

3.2. Power to detect autocorrelation in random response variables

In a second series of simulations, we added autocorrelation to the random response variables (Fig. 3c and d). After generating a series of random numbers drawn from a standard normal distribution, we computed moving averages on the series, with window widths varying from 3 (i.e. one random value and its first neighbours on either side) to 9 (one value and its four neighbours on either side). The number of neighbours (on one side of a point) that are included in the moving average are used as a measure of the range of the autocorrelation: the window width of 3 has a range of 1 while the window width of 9 has a range of 4. The results, displayed in Fig. 5, show that the method detected spatial structures generated by autocorrelation almost faultlessly (power = 1), provided that the truncation value of the spatial matrix was smaller than the window width of the autocorrelation. This is an important result which can be related to the number and shape of the principal coordinates described before. Retaining more distant neighbours produced fewer principal coordinates, the ones representing the finest structures being lost. Besides, the last part of the series of principal coordinates, which should have contained the sines with the shortest periods, were distorted if the truncation of the matrix of Euclidean distances retained more than the first neighbours. This in turn decreased the ability of the set of principal coordinates to detect fine spatial structures in the dependent variables.

As will be shown below, this property of the method also appears when one looks for other types of spatial structures in the dependent variables.

3.3. Power to detect Gaussian curves

The third series of simulations was devoted to the detection of a single, Gaussian-shaped bump, because species frequencies often have unimodal (Gaussian-like) distributions along environment gradients (Austin, 1976). A Gaussian function (i.e. a normal density function) was computed, with a given maximum height and width; the width was defined as two standard deviations on either side of the mean, measured in sampling intervals. Within each simulation run, the width and maximum height were fixed, while the position of the mean of the curve along the transect varied at random.

A first series of runs involved only a thin Gaussian curve (5 sampling intervals wide, Fig. 3e), which was submitted to spatial matrices of increasing coarseness (Fig. 6a). These simulations showed that the method had no problem in detecting the signal, provided that the threshold of the spatial matrix was lower than or equal to the width of the Gaussian curve. The percentage of variance explained varied, though, increasing when the spatial matrix allowed finer resolution i.e. when the truncation value tended to 1 (Fig. 6b).

In another series of runs, normal noise was added to the Gaussian curve (Fig. 3f) with a signal-to-noise ratio (SNR) arbitrarily defined as the maximum height of the Gaussian curve divided by twice the standard deviation of the noise. With this definition, a SNR equal to 1 implies a normal noise component where 68% of the random values fall within a range equal to the maximum height chosen for the Gaussian curve. The data generation algorithm was the following:

1. Generate a Gaussian curve with a known maximum height MH (for instance MH = 20).
2. Draw random values from a normal distribution with mean = 0 and standard deviation = 1.
3. Select a SNR (for instance SNR = 2).
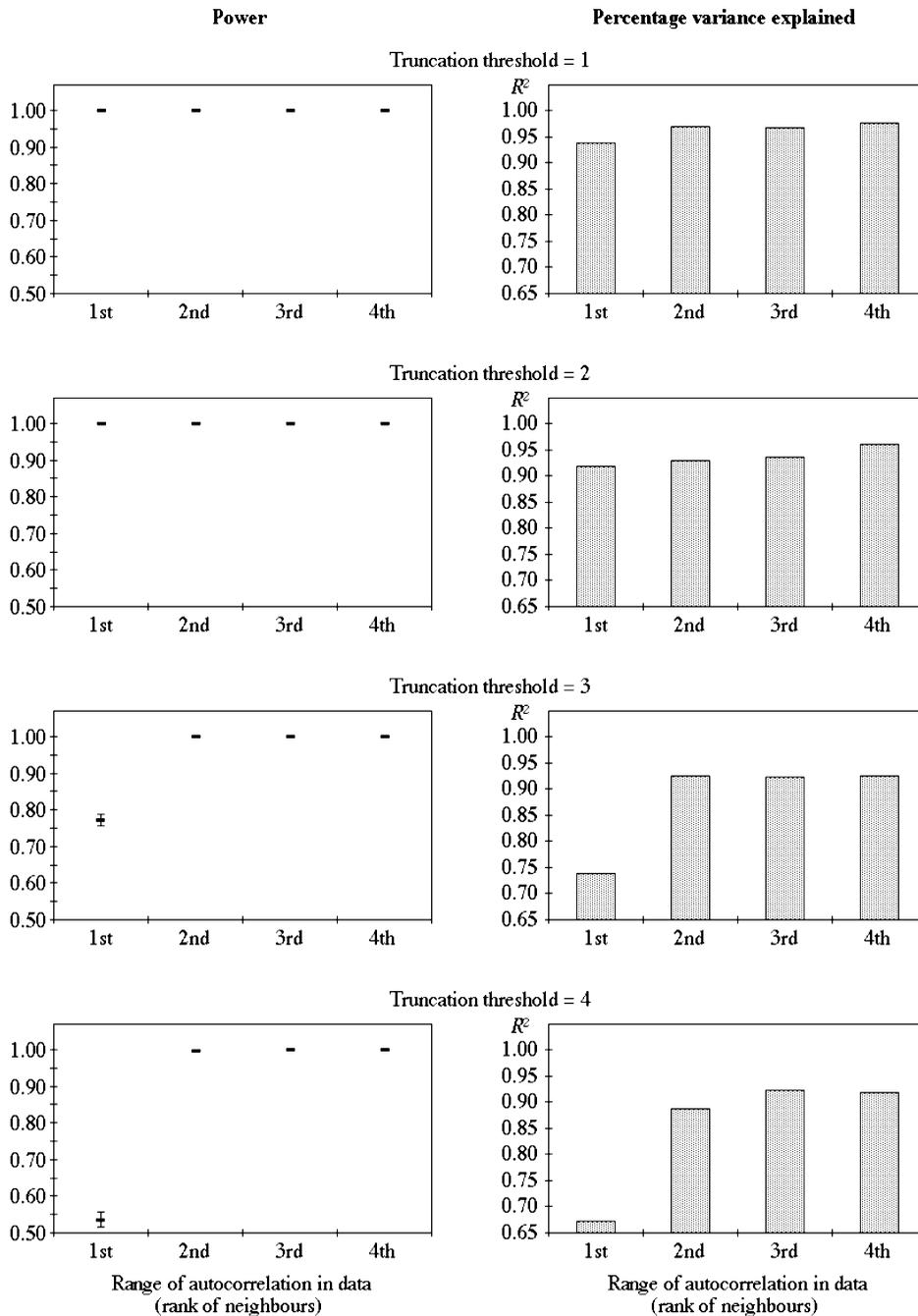
Fig. 5. Power to detect random autocorrelated variables (left) and percentage of variance explained (R^2: right) for spatial matrices with various resolutions, set by the truncation threshold of the Euclidean distance matrix, and various ranges of autocorrelation (abscissa) in the dependent variable. Each run consisted of 5000 independent simulations. The 95% confidence intervals are omitted if they are narrower than the symbols.
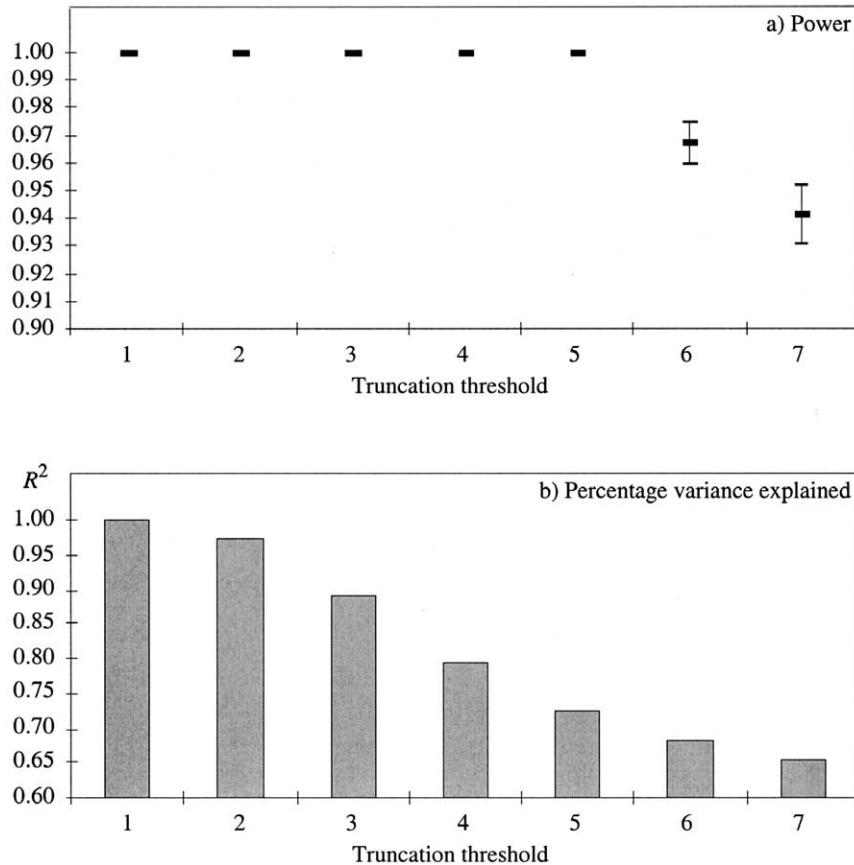
Fig. 6. Power to detect Gaussian-shaped bumps (a) and percentage of variance explained (R^2: b) for spatial matrices with various resolutions set by the truncation threshold of the Euclidean distance matrix (abscissa). Each run consisted of 5000 independent simulations. The 95% confidence intervals are omitted if they are narrower than the symbols. In the absence of noise in the data, the variation in the results (shown by the error bars) are due to some dependence of the power to the position of the Gaussian bump in the data series.

4. Adjust the standard deviation of the noise by multiplying its values by MH/(SNR × 2); this works because multiplying the values of a random normal series by a given number multiplies its standard deviation by that number. In our example: noise SD = 20/(2 × 2) = 5, so multiply all values obtained in (2) by 5.

5. Add signal (obtained in (1)) and noise (obtained in (4)).

6. Set all negative values of the data obtained in (5) to zero. Because they simulate species distributions, these values cannot be negative.

7. Rescale the data so obtained to the pre-selected maximum height.

Fig. 7 shows that the capacity of detection of the method depended on both the SNR and the width of the Gaussian curve. With SNR = 1 (Fig. 7a–d), power was 0.74 in the best case. This occurred with the broadest curve (40 sampling intervals) and, interestingly, the coarsest spatial matrix (truncation threshold = 4: Fig. 7d). This feature, better power obtained when using a coarser spatial matrix, held true for all but the narrowest Gaussian curve. The coarseness of the spatial matrix seemed to allow it to 'see' through the noise better than finer ones. As expected, the broader the Gaussian curve, the easier its detection is by our method.

With less noise in the data (SNR = 2, Fig. 7e–h), power increased dramatically, varying between 0.93 and 1.00 for all but the narrowest

Gaussian curve. Here again, power was better when the spatial matrix was coarser. Observe, however, that again the percentage of variance
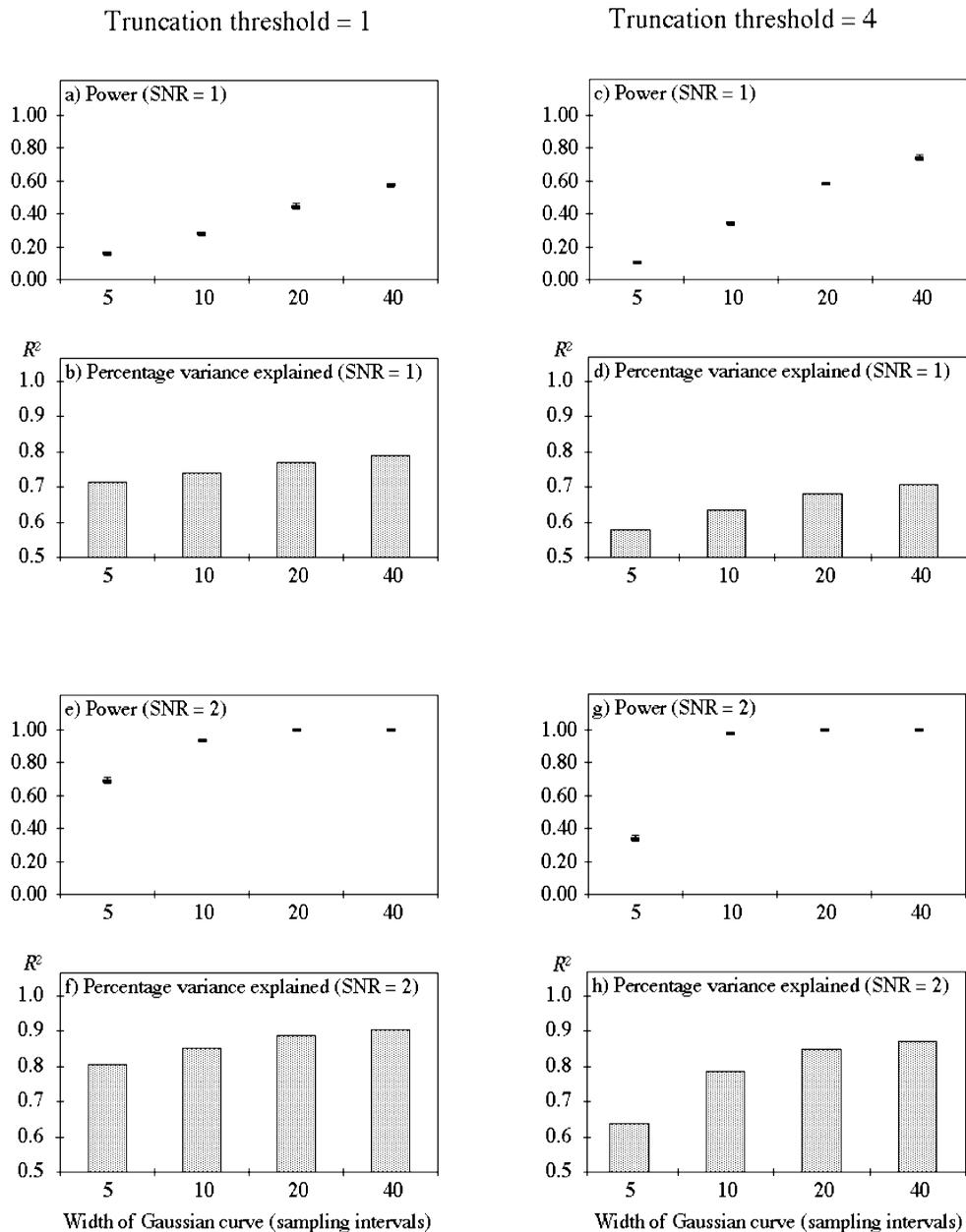
Fig. 7. Power to detect Gaussian-shaped bumps (a, c, e, g) and percentage of variance explained (R^2: b, d, f, h) for spatial matrices with two resolutions (left, right) set by the truncation threshold of the Euclidean distance matrix) and noise added to the data. SNR = signal-to-noise ratio; see definition in text. Each run consisted of 5000 independent simulations. The 95% confidence intervals are omitted if they are narrower than the symbols.

explained was higher when the spatial matrix was finer (truncation threshold = 1); this is a logical result considering that finer explanatory variables can model finer structures in the data.

3.4. Power to detect sine curves

A fourth series of simulations was devoted to sinusoids, with periods varying from 5 to 40 inter-site distances, and SNR having values of 1 (Fig. 3g) and 2 (Fig. 3h). The results, shown in Fig. 8, bear some resemblance with those obtained for Gaussian curves: power depended on the amount of noise in the data (compare Fig. 8a–e and Fig. 8c–g), and more variance was explained when the spatial matrices were finer (truncation threshold = 1; compare Fig. 8b–d and Fig. 8f–h). Interestingly and predictably, the finest sinusoids were never detected when the spatial matrix was too coarse (Fig. 8c–g, periods of 5). In that case, the period of the finest spatial principal coordinate was larger than that of the dependent variable. Another noteworthy feature is that, contrary to the Gaussian case, the amount of variance explained did not vary with the period of the sinusoid for given combinations of SNR and spatial matrix resolution, provided that the latter was fine enough to detect the signal. When this was not the case, no combination of the available variables could adequately model the dependent variable.

3.5. Power to detect gradients

Our last series of simulations focussed on linear gradients with SNR values of 1 and 2. Strikingly, the method never failed to detect the pattern, and always explained practically all the variance; the proportion of the dependent variable's variance explained was between 0.998 and 0.999.

4. Test on complex data

This section is devoted to the illustration of the use of our method with actual data sets. Our example involves artificial data constructed by combining various kinds of signals usually present in real data, plus two types of noise. This provides a pattern that has the double advantage of being realistic and controlled, thereby permitting a precise assessment of the potential of the method to recover the structured part of the signal and to dissect it into its primary components. Other papers will be devoted to the application of the method to real ecological data sets.

4.1. The data

We constructed the data by adding the following components together (Fig. 9) into a transect consisting of 100 equidistant observations:
1. a linear trend (Fig. 9a);
2. a single normal patch in the centre of the transect (Fig. 9b);
3. four waves (i.e. a sine wave with a period of 25 sampling units) (Fig. 9c);
4. 17 waves (i.e. a sine wave with a period of ≈ 5.9 sampling units) (Fig. 9d);
5. a random autocorrelated variable, with auto-correlation determined by a spherical vari-ogram with nugget value = 0 and range = 5 (Fig. 9e);
6. a noise component drawn from a random normal distribution with mean = 0 and vari-ance = 4 (Fig. 9f).

Fig. 9 shows the partial contributions of the six components to the variance of the final artificial response variable. The random noise (Fig. 9f) contributed for more than half of the total variance. Thus, the spatially structured components of the compound signal (Fig. 9a–e) were well hidden in the noise, as it is often the case with real ecological data.

4.2. Analytical procedure

4.2.1. Detrending the dependent variable
As it is the case in most techniques of spatial analysis, the first step is to detrend the data. We recommend doing it whenever a significant linear trend is detected even though our method is able to model linear trends, because this preliminary step allows a separate modelling of the trend
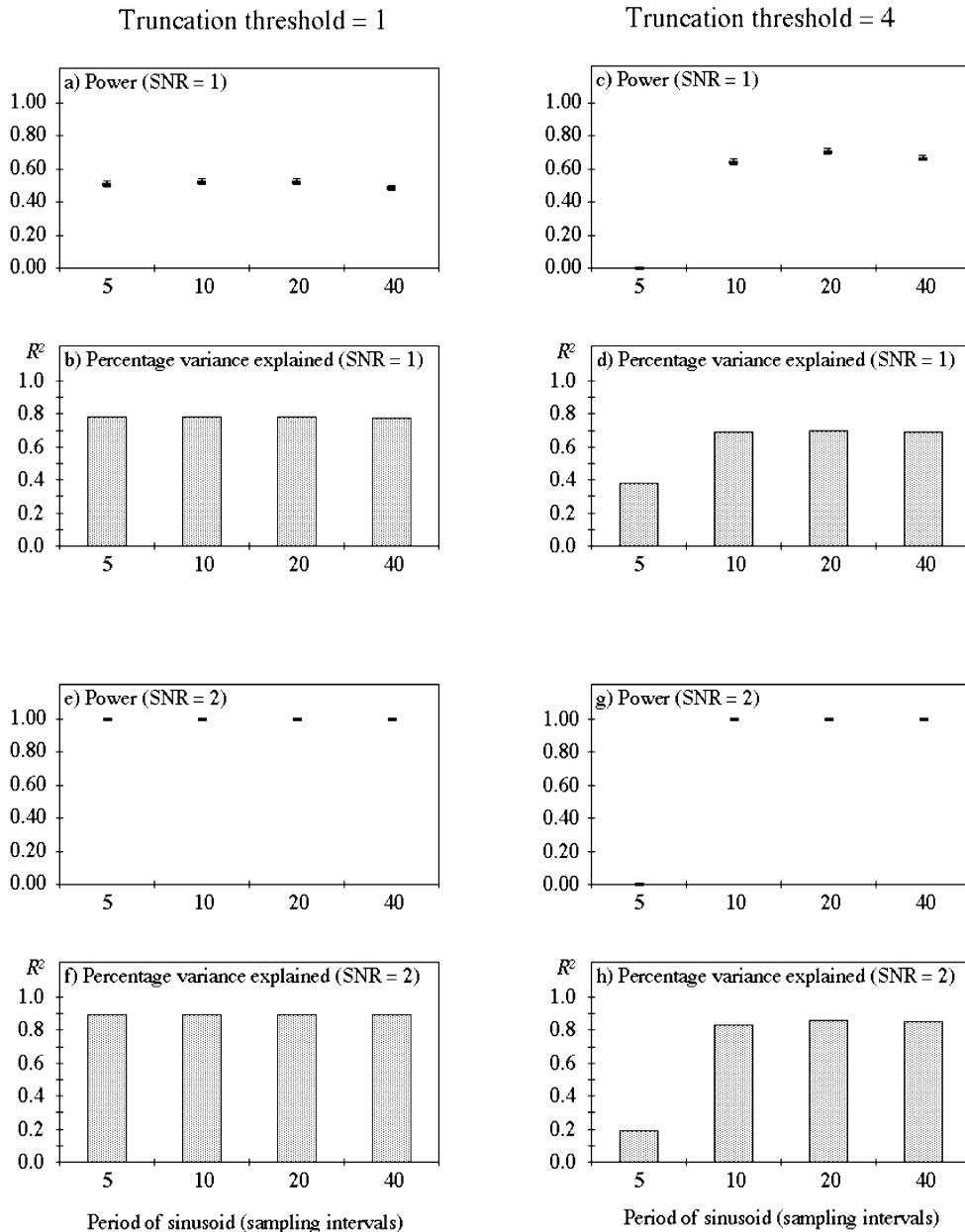
Fig. 8. Power to detect sine curves (a, c, e, g) of different periods (abscissa measured in sampling intervals) and percentage of variance explained (R^2: b, d, f, h) for spatial matrices with two resolutions set by the truncation threshold of the Euclidean distance matrix (left, right) and noise added to the data. Each run consisted of 5000 independent simulations. The 95% confidence intervals are omitted if they are narrower than the symbols.

while retaining all the potential of the principal coordinates to model more complex features. While our example is unidimensional and thus requires only the fitting of a straight line by simple linear regression, bidimensional data sets must be detrended by fitting a plane.

In our example, the trend was significant ($P =$ 0.001 after 999 permutations), with an R^2 of 0.172, which is a bit higher than the 12.1% that we had built into the data (Fig. 9a). This preliminary regression has thus modelled 5.1% of additional trend present in the combination of the other components, particularly the random normal component which is a finite sample of a distribution with mean $= 0$, and can thus show a local trend.

After this step, one works on the detrended data. In our case, the single central bump that we had built into the data accounted for 8% of the variance of the detrended data, the 4 waves for 17%, the 17 waves for 11% and the random autocorrelated signal for 11%. The spatially structured components thus contributed 47% of the total variance of the detrended data, and the pure random component 53%. These values are the contributions of the variables to the variation of the detrended data, in the sense of Scherrer (1984). They were obtained as follows. (1) Regress
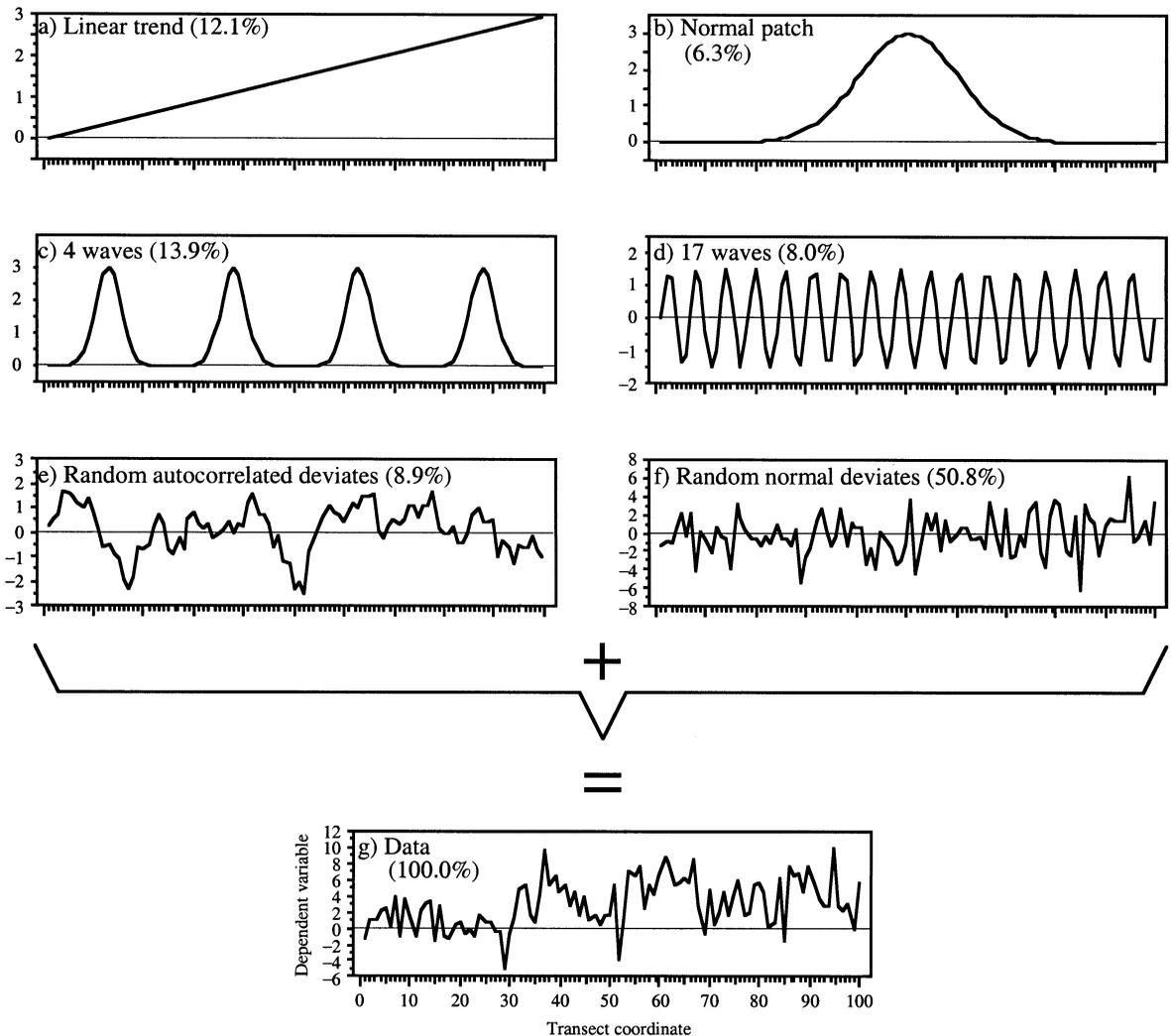
Fig. 9. Construction of the artificial pseudo-ecological data set of known properties. The six components added together are shown, with their contributions to the variance of the final signal.

the detrended data on the central bump, 4 waves, 17 waves, random autocorrelated and random signals originally built into the data. (2) Compute the correlation of each variable with the detrended data. (3) Compute each contribution as the product of the standardised (partial) regression coefficient with the correlation coefficient. These contributions will be compared later with the structures revealed by the PCNM method.

4.2.2. Building the matrix of spatial variables

This step involves the procedure described in Section 2: build a matrix of Euclidean distances among objects, truncate it to the first neighbours, replace the removed values by the highest value retained multiplied by 4, and compute the principal coordinates of the resulting matrix. Our example being built upon the same linear transect of 100 objects as the one used in the simulations, the results were the same: we obtained 67 principal coordinates representing a series of sine waves of decreasing periods, starting from a period of 101, and ending with a period slightly larger than 3. These are the spatial variables that will be used in the next steps.

4.2.3. Running the spatial analysis

Since our example involves a single dependent variable, the spatial analysis consists in a multiple linear regression of the detrended dependent variable onto the 67 spatial variables built in step 2. The main question at this step is to decide what kind of model is appropriate: a global one, retaining all the spatial variables and yielding an R^2 as high as possible, or a more parsimonious model based on the most significant spatial variables? The answer may depend on the problem, but in our opinion the general procedure should include some sort of thinning of the model. Remember that the number of parameters of the global model is equal to about 67% of the number of objects, a situation which may often lead to an overstated value of R^2 by chance alone. The solution that we propose, and have applied to this example, consists in testing the significance of all the (partial) regression coefficients and retaining only the principal coordinates that are significant at a predetermined (one-tailed) probability value.

All tests can be done using a single series of permutations if the permutable units are the residuals of a full model (Anderson and Legendre, 1999; Legendre and Legendre 1998), which is the case here. The explanatory variables being orthogonal, no recomputation of the coefficients of the 'minimum' model are necessary. Note, however, that a new series of statistical tests based upon the minimum model would give different results, since the denominator (residual mean square) of the F statistic would have changed.

The analysis of our detrended artificial data yielded a complete model explaining 75.3% of the variance when using the 67 explanatory variables. Reducing the model as described above allowed us to retain 8 spatial variables at $P = 0.05$, explaining together 43.3% of the variance. This value compares well with the 47% of the variance representing the contributions of the single bump, the two variables with 4 and 17 waves, and the random autocorrelated component of the detrended data. The spatial variables retained were principal coordinates no. 2, 6, 8, 14, 28, 33, 35 and 41.

4.2.4. Dissecting the spatial model

One major advantage of our method is that the components of the spatial model obtained are orthogonal, and can thus be either examined separately or combined at will into independent submodels that can be interpreted with the help of external information. When such knowledge is not available, the submodels may help generate hypotheses about the underlying processes that have generated the structures.

It often happens that the significant variables are grouped in series of roughly similar periods. In our example, for instance, there is a clear gap between the first four significant variables and the last four. Thus, a first step may be to draw two submodels, one involving variables 2, 6, 8 and 14 (added together, using their regression coefficients as weights) and the other involving variables 28, 33, 35 and 41. The results are shown in Fig. 10a–d, respectively. The 'broad-scale' submodel (Fig. 10a) shows 4 major bumps, the two central ones being much higher than the two lateral ones. This may indicate that two mechanisms are actu-
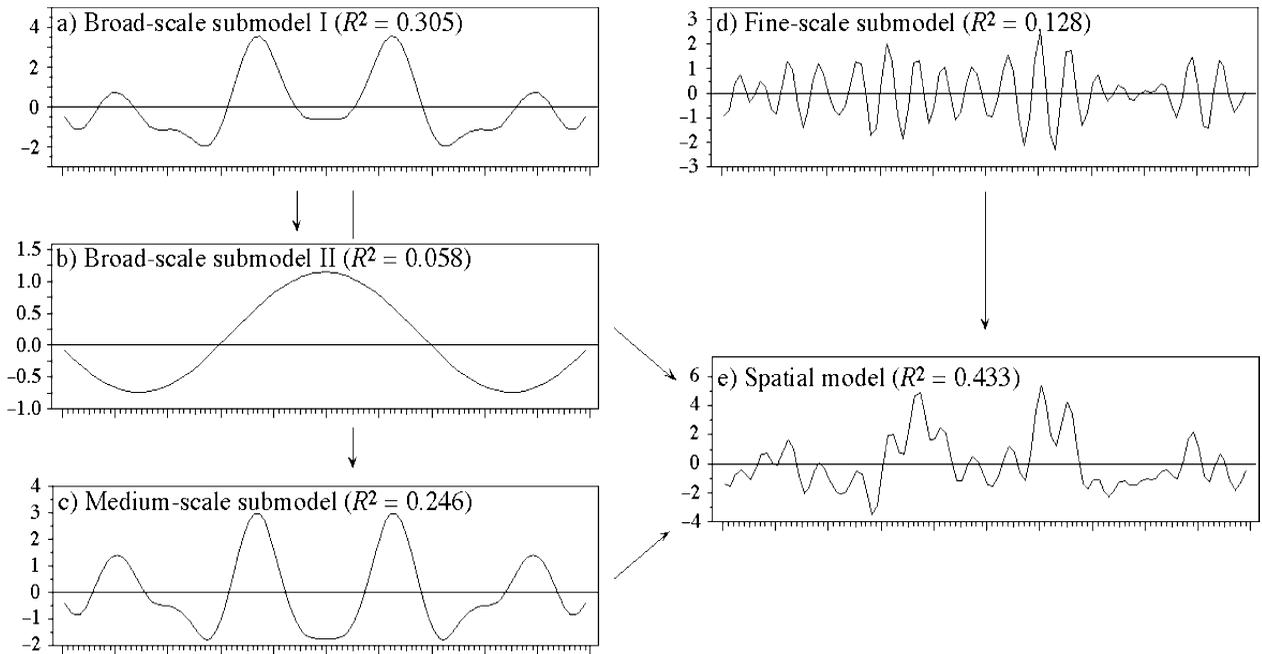
Fig. 10. Minimum spatial model and its additive submodels obtained by the new method of spatial analysis.

ally confounded, one producing the four bumps and another process elevating the two central ones. Subdividing this submodel further by separating variable 2 from variables 6, 8 and 14 allowed indeed to distinguish a central bump (Fig. 10b) and 4 waves (Fig. 10c). The fine-scale submodel (Fig. 10d) shows 17 waves, with hints of a 4-bump pattern. The spatial model made of the 8 variables is shown in Fig. 10e.

The method has successfully revealed the four deterministic components that we built into the data: trend, single central bump, 4 waves and 17 waves, despite the large amount of noise added. The amount of variance explained by the model suggests that most of the spatially-structured information present in the random autocorrelated component of the data is also contained in the model (in accordance with our simulation results), but that it could not be separated from the periodic signals because it was 'diluted' over several scales. The successful extraction of the structured information can be further illustrated by comparing the model of the detrended data obtained above (Fig. 11b) to the sum of the four compo-

nents 'central bump', '4 waves', '17 waves' and 'random autocorrelated' (Fig. 11a), and by comparing the residuals of the spatial model (Fig. 11d) to the real noise built into the data, i.e. the uncorrelated random variate (Fig. 11c).

5. Discussion

For most applications to ecology, the method presented above will be one element integrated into an analytical procedure involving not only the dependent and spatial variables, but also some environmental data. Procedures to interpret the spatial structures revealed by the analysis described above will depend on the context. When one has several environmental variables available for analysis, one possible way is to run a separate, complete spatial analysis for each environmental variable, and subsequently look for structures emerging at similar scales in the dependent and one or several of the environmental variables. A possible shortcut is to analyse the environmental data together with a set of spatial variables re-

stricted to those significant for the dependent variable. This would allow a quicker assessment of the similarity of significant scales.

Another approach is to use the Borcard et al. (1992), Borcard and Legendre (1994) method of variation partitioning, where the sets of environmental and spatial variables are used alone, in combination, and as partial explanatory variables, to assess the amounts of variation explained together by environment and space, or by one component alone while controlling for the other. The efficiency of the variation partitioning method will be greatly improved by replacing the traditional polynomial function by the spatial variables resulting from our new PCNM method. For exam-

ple, in the case of the Oribatid mite data presented in Borcard et al. (1992), fraction [a] (nonspatial environmental component) accounted for 13.1% of the total variation with the polynomial approach, and 12.1% with the PCNM method. Fraction [b] (explanation shared by the environmental and spatial variables) changed from 31.0 to 32.6%, and the 'pure' spatial component increased from 12.2 to 23.8%. The overall percentage of explained variation increased from 56.9 to 68.5%. Thus, in this case, the improvement consisted of the increase of the pure spatial fraction: more spatial structures not explained by the environmental variables have been detected. In other cases, however, the change may also affect the
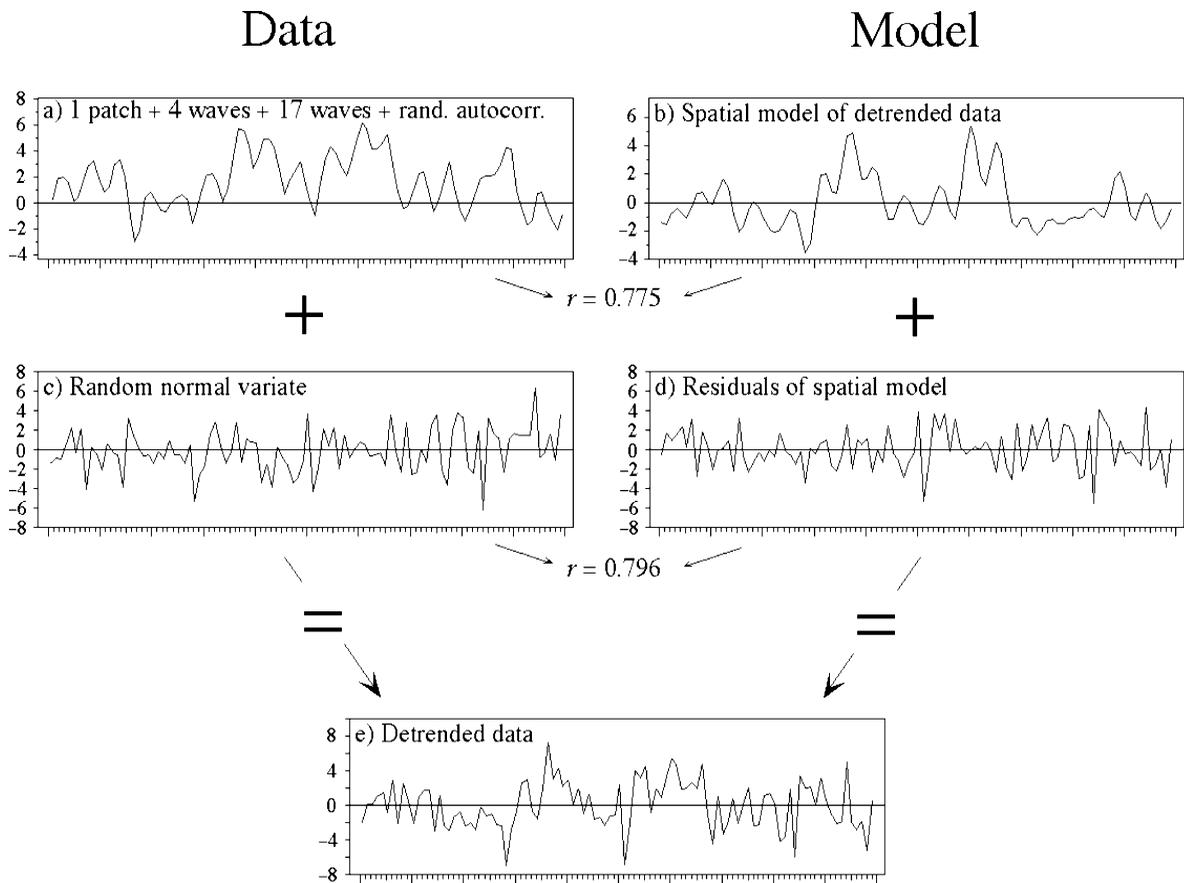
Fig. 11. Comparison of the structured (a) and random (c) components of the data on the one hand, and the spatial model (b) and its residuals (d) on the other hand, and correlations between the homologous components. (c) The left-hand (a + c) and right-hand (b + d) sides of the figure sum to the original detrended data.

other components. For instance, the detection of fine spatial structures also explained by the environmental variables would increase the share of component [b] at the expense of fraction [a].

These considerations open the door to applications of the PCNM method to multivariate data. In this case the multiple regression used above is replaced by a method of constrained ordination suitable to the data: RDA (Rao, 1964), CCA (ter Braak, 1986), or redundancy analysis on transformed species data (Legendre and Gallagher, 2001). For these applications, however, no existing program provides the appropriate statistical tests on individual regression or canonical coefficients to allow the selection of a proper subset of spatial variables. One would have to rely upon more traditional, non-permutational tests for approximate results, or on stepwise procedures. The t-values of regression coefficients provided by the program Canoco (ter Braak and Smilauer, 1998) can also be used for assessment of the most important spatial variables, although they are not accompanied by permutational probabilities. In the future, programs of canonical analysis should include permutational tests of significance of individual regression coefficients.

Another possible extension concerns data sampled across a surface i.e. bidimensional spatial data. Preliminary attempts in this direction show that our method still provides periodic spatial variables if the data are regular, but that the spatial resolution is about half that obtained in the unidimensional case; the bidimensional model includes the same number of principal coordinates as the unidimensional. Besides, the principal coordinates do not show the simple scale-to-variance relationship which allows them to appear readily in decreasing order of periods in the unidimensional case.

Finally, a word must be said about irregularly sampled data. The effect of a missing data point in a regular series is to disrupt the sine waves provided by the principal coordinate analysis. This disruption acts on the amplitude, phase and period of the sines, thereby affecting the interpretation of the spatial variables in terms of scales. Truly irregular sampling patterns result in totally irregular principal coordinates. Note that these are still suitable spatial descriptors, but their interpretation is complicated by the fact that each one of them often bears structures at several scales.

In cases where a regular sampling series suffers from one or a few missing observations, there is a simple way of overcoming the problem. It consists in filling the voids i.e. adding points where they are missing in the file of spatial coordinates; nothing is added or interpolated in the dependent variable. The filled-up series is then submitted to the analysis yielding the principal coordinates, and then the supplementary objects are removed from the matrix of principal coordinates before it is used as a set of spatial explanatory variables. This trick has a cost, however: removal of the supplementary objects introduces some correlation among the spatial variables, which were previously uncorrelated. As long as the number of supplementary objects remains low in comparison to the number of observed objects, these correlations are low. But an exaggerated use of supplementary objects may introduce unwanted amounts of correlation among the spatial variables, thereby compromising one major feature of the method, i.e. the independence of the spatial variables, which is required for their combination into orthogonal submodels.

This paper raises a number of mathematical questions; for instance, the relationship between our method of decomposition of the spatial relationships among sites and the one proposed by Méot et al. (1993), Fourier analysis, and the decomposition of Toeplitz matrices. We hope that the paper will attract the interest of mathematicians who can help us understand these properties and develop methods of spatial modelling further.

A FORTRAN program (SPACEMAKER: source code, compiled versions for Macintosh and DOS, and program documentation) to carry out the principal coordinate decomposition of spatial locations described in this paper is available on the WWWeb site http://www.fas.umontreal.ca/biol/legendre/ or via anonymous ftp at ftp://ftp.umontreal.ca/pub/casgrain/labo/SPACEMAKER/.

Acknowledgements

References

Anderson, M.J., Legendre, P., 1999. An empirical comparison of permutation methods for tests of partial regression coefficients in a linear model. J. Statist. Comput. Simul. 62, 271–303.

Austin, M.P., 1976. On nonlinear species response models in ordination. Vegetatio 33, 33–41.

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

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

Gower, J.C., Legendre, P., 1986. Metric and Euclidean properties of dissimilarity coefficients. J. Classif. 3, 5–48.

Legendre, P., 1990. Quantitative methods and biogeographic analysis. In: Garbary, D.J., South, R.G. (Eds.), Evolutionary biogeography of the marine algae of the North Atlantic. NATO ASI Series, vol. G 22. Springer-Verlag, Berlin, pp. 9–34.

Legendre, P., 1993. Spatial autocorrelation: trouble or new paradigm? Ecology 74, 1659–1673.

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

Legendre, P., Gallagher, E.D., 2001. Ecologically meaningful transformations for ordination of species data. Oecologia, 129, 271–280.

Legendre, P., Legendre, L., 1998. Numerical ecology, second English ed. Elsevier Science BV, Amsterdam, p. 853.

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

Manly, B.J.F., 1997. Randomization, bootstrap and Monte Carlo methods in biology, second ed. Chapman and Hall, London, p. 399.

Méot, A., Chessel, D., Sabatier, R., 1993. Opérateurs de voisinage et analyse des données spatio-temporelles. In: Lebreton, J.D., Asselain, B. (Eds.), Biométrie et environnement. Masson, Paris, pp. 46–71.

Méot, A., Legendre, P., Borcard, D., 1998. Partialling out the spatial component of ecological variation: questions and propositions in the linear modeling framework. Environ. Ecol. Stat. 5, 1–26.

Rao, C.R., 1964. The use and interpretation of principal component analysis in applied research. Sankhyaá A 26, 329–358.

Scherrer, B., 1984. Biostatistique. Gaëtan Morin, Chicoutimi, pp. 850.

ter Braak, C.J.F., 1986. Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. Ecology 67, 1176–1179.

ter Braak, C.J.F., 1990. Update notes: CANOCO version 3.10. Agricultural Mathematics Group, Wageningen.

ter Braak, C.J.F., 1992. Permutation versus bootstrap significance tests in multiple regression and ANOVA. In: Jöckel, K.-H., Rothe, G., Sendler, W. (Eds.), Bootstrapping and related techniques. Springer, Berlin, pp. 79–86.

ter Braak, C.J.F. and Smilauer, P., 1998. CANOCO reference manual and user's guide to Canoco for Windows—Software for canonical community ordination (version 4). Microcomputer Power, Ithaca, New York, pp. 351.