6.3 Modeling spatial structures

6.3.1 Introduction: the 3 components of spatial structure

For a good understanding of the nature of spatial variation, it is useful to decompose it into three independent components (Figure 43):

1. A major structural component: the overall **mean** of the variable(s) across the whole sampling area. This mean may vary in a continuous way on one or several axes across the area. In this case it is said to show a **trend**. The presence of a trend in ecological data is generally interpreted as the action of a factor at a scale larger than the study area.

2. A variation component that is spatially autocorrelated, but at a finer scale than the trend, called a **regional** scale. This variation can often be interpreted as the result of either biotic processes or the action of environmental forcing on the studied variables.

3. A **random, uncorrelated variation**, arising from observation or analytical error, or from variation that may be structured (correlated) at a scale too fine to be resolved by the sampling design.

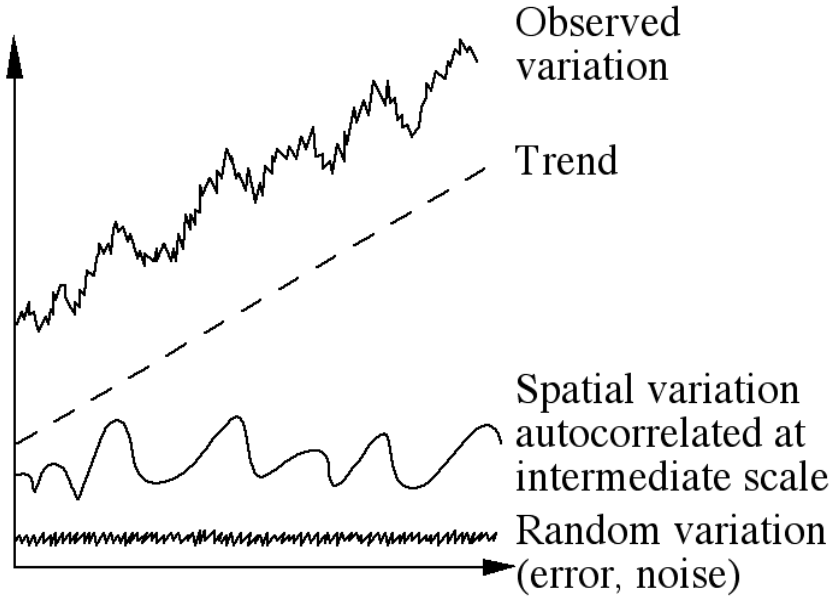

Figure 43 - The three components of spatial variation

A spatial structure may appear in a variable *y* because the process that has produced the values of *y* is spatial and has generated **autocorrelation** in the data; or it may be caused by **dependence** of *y* upon one or several causal variables *x* which are spatially structured; or both.

The following sections present two techniques to model spatial structures in the univariate or multivatiate context. The first technique, *trend surface analysis*, is a crude method mainly adequate to model simple gradients, or remove them from the data (when detrending is necessary). The second method, *PCNM analysis*, has been developed to model spatial structures at all scales resolved by a given sampling design.

6.3.2 Trend surface analysis

This technique is a **particular case of multiple regression**, where the explanatory variables are geographical (x-y) coordinates, sometimes completed by higher order polynomials. When applying this method, one generally supposes that the spatial structure of the observed variable is a result of one or two generating processes that spread over the whole studied area, and that the resulting broad-scale structure of the dependent variable can be modelled by means of a polynomial of the spatial coordinates of the samples. A simple example follows:

Imagine a soil arthropod, the density of which (let us call it *z*) increases from 0 (near a stream) to 100 individuals per square meter (in a nearby meadow). If this density variation is linear, a simple linear regression, with the distance to the stream (*x*) acting as explanatory variable, is enough to model the arthropod density in the whole meadow (Figure 44):

$$\hat{z} = b_0 + b_1 x$$

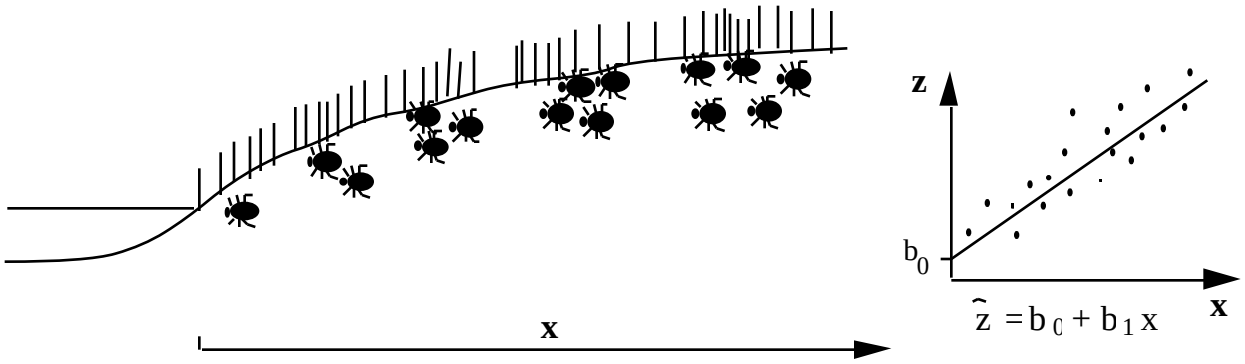

Figure 44 - Density of an arthropod species along a gradient and linear model.

Now, if the stream (with its neighbouring meadows) extends from higher mountains to sea level, perhaps the arthropod density varies also with the altitude (*y*). A second explanatory variable is necessary, i.e. the altitude, or possibly the distance to the source along the stream. If the density variation with respect to the altitude is also linear, one gets a first order multiple regression equation of the form:

$$\hat{z} = b_0 + b_1 x + b_2 y$$

The result is thus a *regression plane* fitted through the *z* data (densities) by means of the x-y coordinates of the arthropod sampling points (Figure 45).

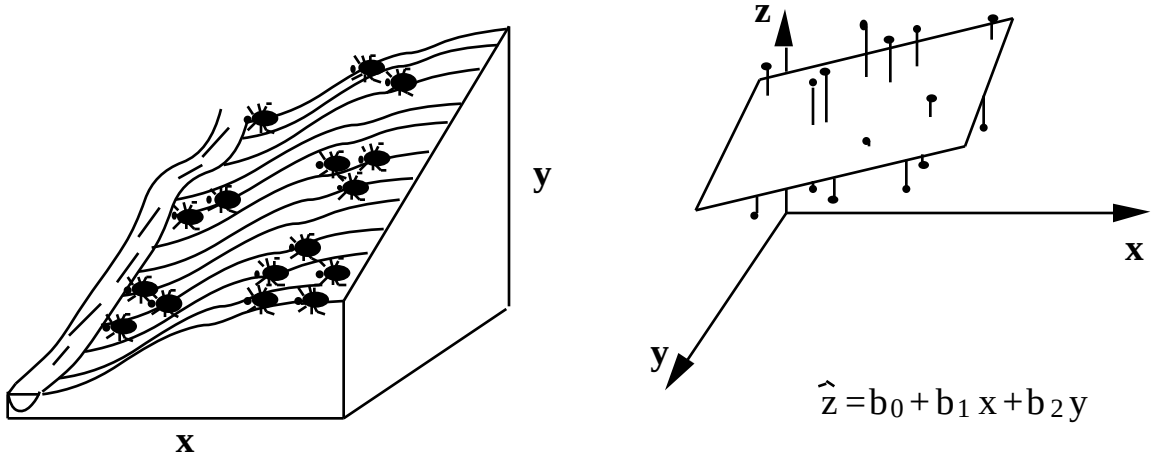

Figure 45 - Density of an arthropod species along a double gradient and linear model.

If a plane does not explain enough variation, one can try to fit higher order polynomials, by adding second, third ...order x-y terms and their products. The following equation is a cubic trend surface equation:

$$\hat{z} = b_0 + b_1 x + b_2 y + b_3 x^2 + b_4 xy + b_5 y^2 + b_6 x^3 + b_7 x^2 y + b_8 xy^2 + b_9 y^3$$

It is easy to visualize the outcome of the addition of one order to a trend surface model by remembering that each addition of an order allows one more fold to the surface (Figure 46):

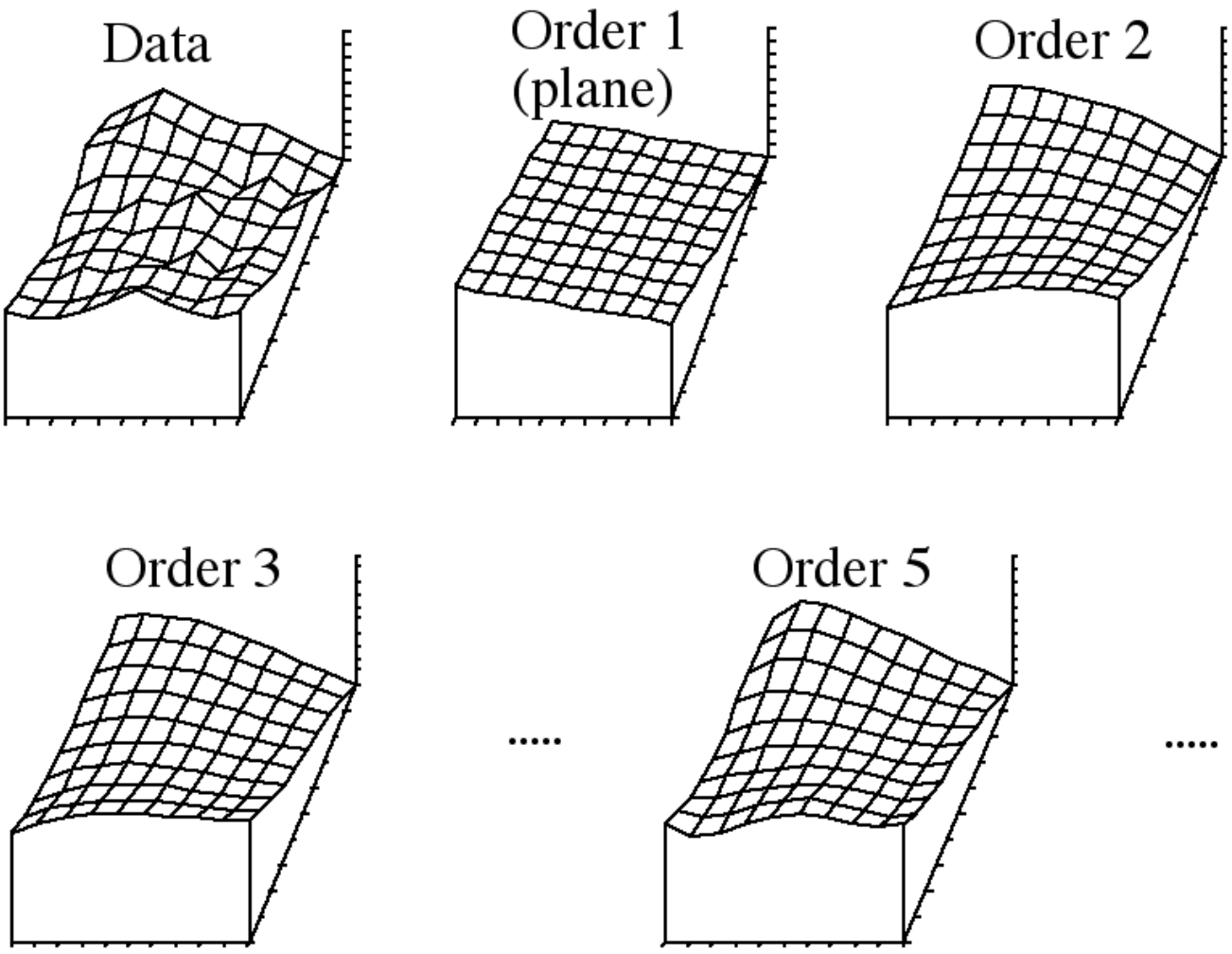

Figure 46 - Example of trend surface analysis, equations of order 1, 2, 3 and 5.

Trend surface analysis can model relatively simple structures with a reasonable amount of “hills” and “ holes” resulting of one or two long-range trends (hence the name) across the sampling area. But this method, although easy to compute, suffers from several conceptual and practical problems, and should be used with great care. Here are some of these problems:

Conceptual problem:

- fitting a trend surface is useful only when the trend has an underlying physical or biological explanation, or if it can help generating biological hypotheses; interpretation of individual terms is often difficult;

Practical problems:

- when data points are few, extreme values can seriously distort the surface;

- the surfaces are extremely susceptible to edge effects. Higher-order polynomials can turn abruptly near area edges, leading to unrealistic values;

- trend surfaces are inexact interpolators. Because they are long-range models, extreme values of distant data points can exert an unduly large influence, resulting in poor local estimates of the studied variable.

Detrending

Despite its problems, trend surface analysis is very useful in one specific case. It has been said in Section 6.2.5 that for testing, the condition of second-order stationarity or, at least, the intrinsic assumption must be satisfied. Removing a trend from the data at least makes the mean constant over the sampling area (although it does not address any problem of heterogeneity of variance). Furthermore, most methods of spatial analysis are devised to model the intermediate-scale component of spatial variation and are therefore much more powerful on detrended data. For these reasons, trend surface analysis is often

used to detrend data: one fits a plane on the data and, if the trend is significant, one proceeds to analyze the finer scale structure on the residuals of this regression (this is equivalent to subtract the fitted values from the raw data and to work with what remains).

6.3.3 Principal Coordinates of Neighbour Matrices (PCNM)

As said above, the coarseness of trend-surface analysis presents a problem: fine structures cannot be adequately modelled by this method. Too many parameters would be required to do so, 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 spatial scale or direction any longer.

In recent years, researchers have become more aware of the fact that ecological processes occur at defined scales, and that their perception and modeling depends upon a proper matching of the sampling strategy to the size, grain and extent of the study, and on the statistical tools used to analyze 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. This is why Borcard & Legendre (2002)[1] and Borcard *et al.* (2004)[2] have proposed a method for detecting and quantifying spatial patterns over

[1] Borcard, D. & Legendre, P. 2002. All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecological Modelling* 153: 51-68.

[2] Borcard, D., P. Legendre, Avois-Jacquet, C. & Tuomisto, H. 2004. Dissecting the spatial structures of ecologial data at all scales. *Ecology* 85(7): 1826-1832.

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 be presented below in the unidimensional context, where it has the further advantage of being usable even for short ($n > 25$) data series. Most of the text below is adapted from Borcard & Legendre (2002).

The analysis begins by coding the spatial information in a form allowing to recover various structures over the whole range of scales encompassed by the sampling design. This technique works on data sampled along linear transects as well as on geographic surfaces. The demonstration below is made on a univariate, unidimensional case for the sake of clarity. Figure 47 displays the steps of a complete spatial analysis using principal coordinates of neighbour matrices (PCNM).

A. Modified (truncated) matrix of Euclidean distances

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. This "large" value has been empirically set equal to four times the threshold value. Beyond this value, the principal coordinates remain the same to within a multiplicative constant.

For instance, in the case of a linear transect made of sampling points regularly spaced 1 metre apart, we could set the threshold at 1 metre to retain only the closest neighbours, and replace all other distances in the matrix by 1.0 m × 4 = 4.0 m.

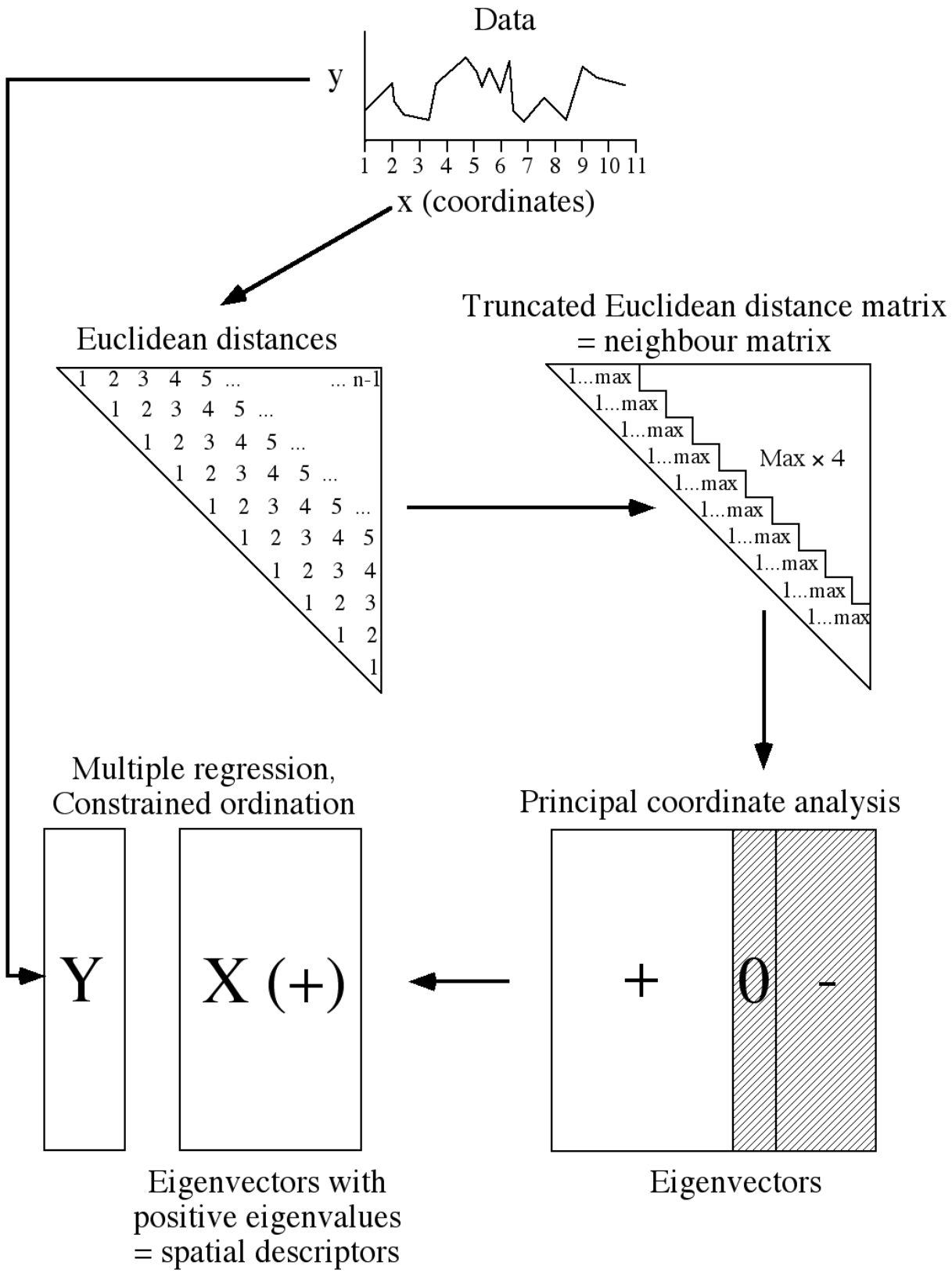

Figure 47 - The computational steps of a PCNM analysis.

B. Principal coordinate analysis of the truncated distance matrix

The second step is to compute the principal coordinates (PCoA) 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 (preferably redundancy analysis, RDA), i.e., as an object-by-variable matrix. We obtain several positive, one or several null, and several negative eigenvalues. Principal coordinate analysis (PCoA) of the truncated distance matrix makes it impossible to represent the distance matrix entirely in a space of Euclidean coordinates because the truncated distance matrix is not Euclidean. When the PCoA is computed in the usual manner, 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). A modified form of the analysis allows them to be computed, but it will not be detailed here.

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

When computed from a distance matrix corresponding to *n* equidistant objects arranged as a straight line, as in Figure 47, 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 sine waves with decreasing periods (Figure 48); 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, a behaviour that alters the performance of the method.

Thus, the PCNM method presents a superficial resemblance to Fourier analysis and harmonic regression, but it is more general since it can model a wider range of signals, and can be used with irregularly spaced data.

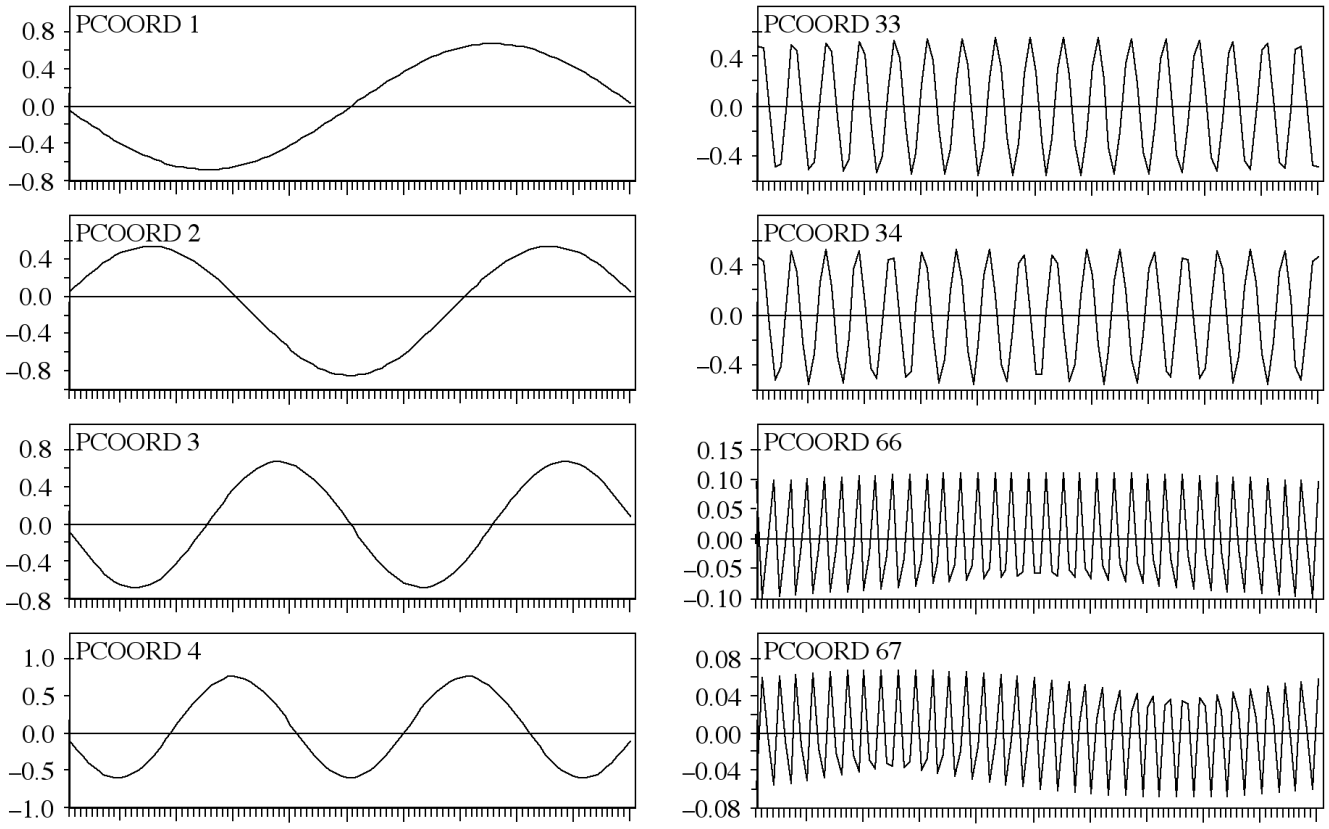

Figure 48 - 8 of the 67 principal coordinates obtained by principal coordinate analysis of a matrix of Euclidean distances among 100 objects, truncated after the first neighbours.

Borcard & Legendre (2002) have shown by simulations that PCNM analysis has a correct type I error and is powerful to detect various types of spatial structures: gradients, single bumps, sine waves, as well as random but spatially autocorrelated signals.

When used on structurally complex data, PCNM analysis also succeeds in recovering spatial structures at various scales. This can be achieved by building subsets of PCNM variables, thereby constructing

additive submodels that can be interpreted *a posteriori* by means of environmental variables or used to build hypotheses about the processes that have generated the structures. Real-world applications are presented by Borcard *et al.* (2004) and, for instance, Brind'Amour *et al.* (2005)[3].

C. Example on artificial data

Borcard *et al.* (2002) present an example involving 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.

Construction of the artificial data - The data were constructed by adding the following components together (Figure 49) into a transect consisting of 100 equidistant observations:

1) a linear trend (Fig. 49a);

2) a single normal patch in the centre of the transect (Fig. 49b);

3) 4 waves (= a sine wave with a period of 25 units) (Fig. 49c);

4) 17 waves (i.e., a sine wave with a period of approximately 5.9 sampling units) (Fig. 49d);

5) a random autocorrelated variable, with autocorrelation determined by a spherical variogram with nugget value = 0 and range = 5 (Fig. 49e);

6) a noise component drawn from a random normal distribution with mean = 0 and variance = 4 (Fig. 49f).

[3] Brind'Amour, A., D. Boisclair, P. Legendre and D. Borcard. 2005. Multiscale spatial distribution of a littoral fish community in relation to environmental variables. Limnology and Oceanography 50: 465-479.

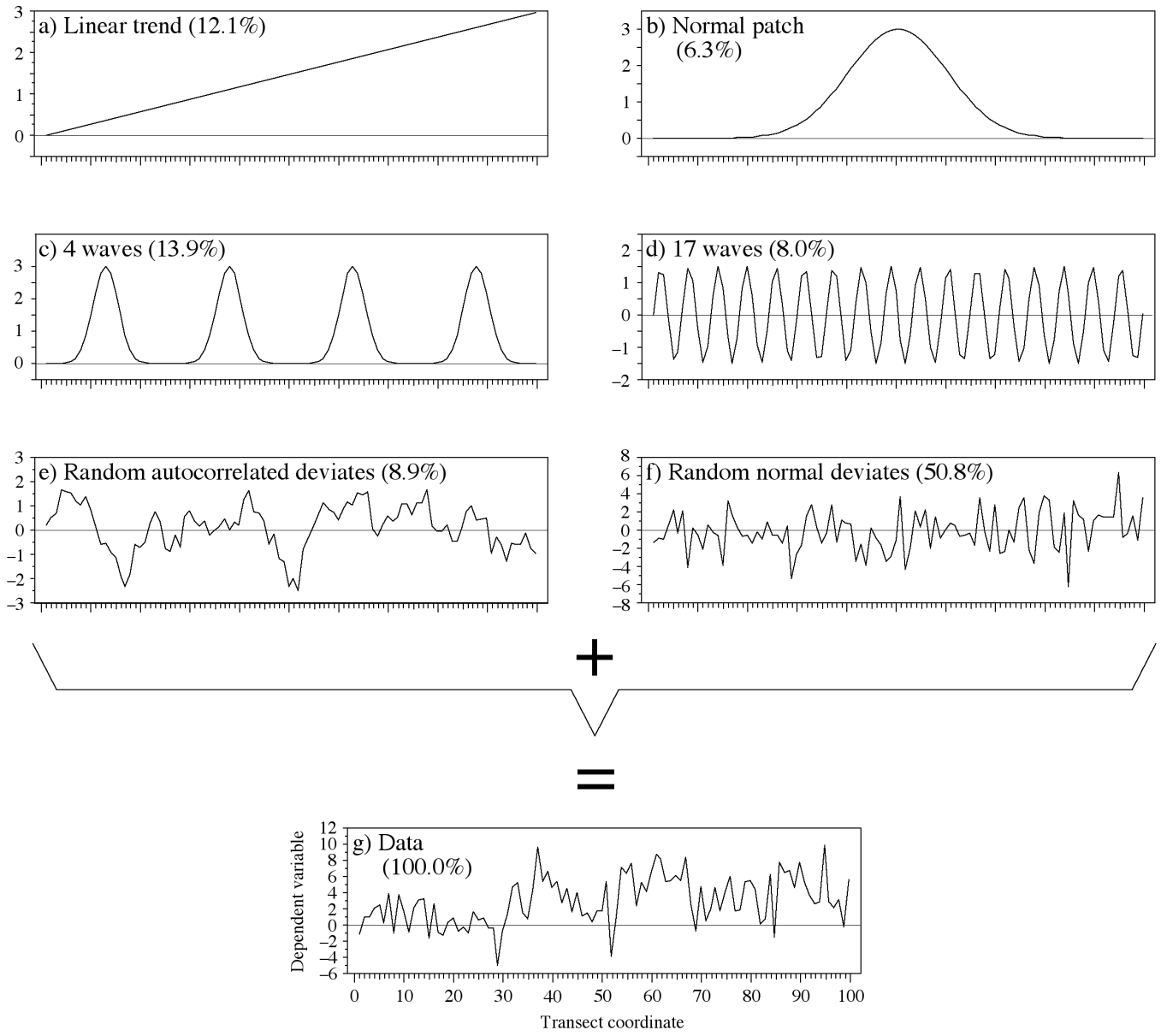

Figure 49 - Construction of the artificial pseudo-ecological data set of known properties. The six components added together are shown, with their contributions to the variance of the final signal.

In the final artificial response variable, the random noise (Fig. 49f) contributed for more than half of the total variance. Thus, the spatially-structured components of the compound signal (Fig. 49a to 49e) were well hidden in the noise, as it is often the case with real ecological data.

Data analysis - The spatial analysis consists in the following steps:

(1) Detrending of the dependent variable (done here because a strong and significant trend was present).

(2) Since this example involves a single dependent variable, multiple linear regression of the detrended dependent variable onto the 67 spatial variables built as explained before.

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 most applications so far included 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 (this can be corrected by the use of an adjusted R^2, however). At the state of our knowledge at that time (early 2000s), a convenient solution consisted 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 & Legendre, 1999[4]; 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. Note also that this procedure is different from that advocated nowadays and explained in chapter 4b (forward selection with a double stopping criterion, i.e., the usual α-level and the R^2_{adj} of the complete model).

[4] Anderson, M.J. and 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.

Analytical results - The analysis of the (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 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 PCNM variables retained were principal coordinates no. 2, 6, 8, 14, 28, 33, 35 and 41.

Additive submodels - It often happens that the significant variables are grouped in series of roughly similar periods. In these data, for instance, there is a clear gap between the first four significant PCNM 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 in the minimum model as weights) and the other involving variables 28, 33, 35 and 41. The results are shown in Figures 50a and 50d respectively. The "broad-scale" submodel (Fig. 50a) shows four major bumps, the two central ones being much higher than the two lateral ones. This may indicate that two mechanisms are actually 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. 50b) and 4 waves (Fig. 50c). The fine-scale submodel (Fig. 50d) shows 17 waves, with hints of a 4-bump pattern. The spatial model made of the 8 variables is shown in Figure 50e.

The method has successfully revealed the four deterministic components that were 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 the simulation results), but that it could not be separated from the periodic signals because it was “diluted” over several scales.

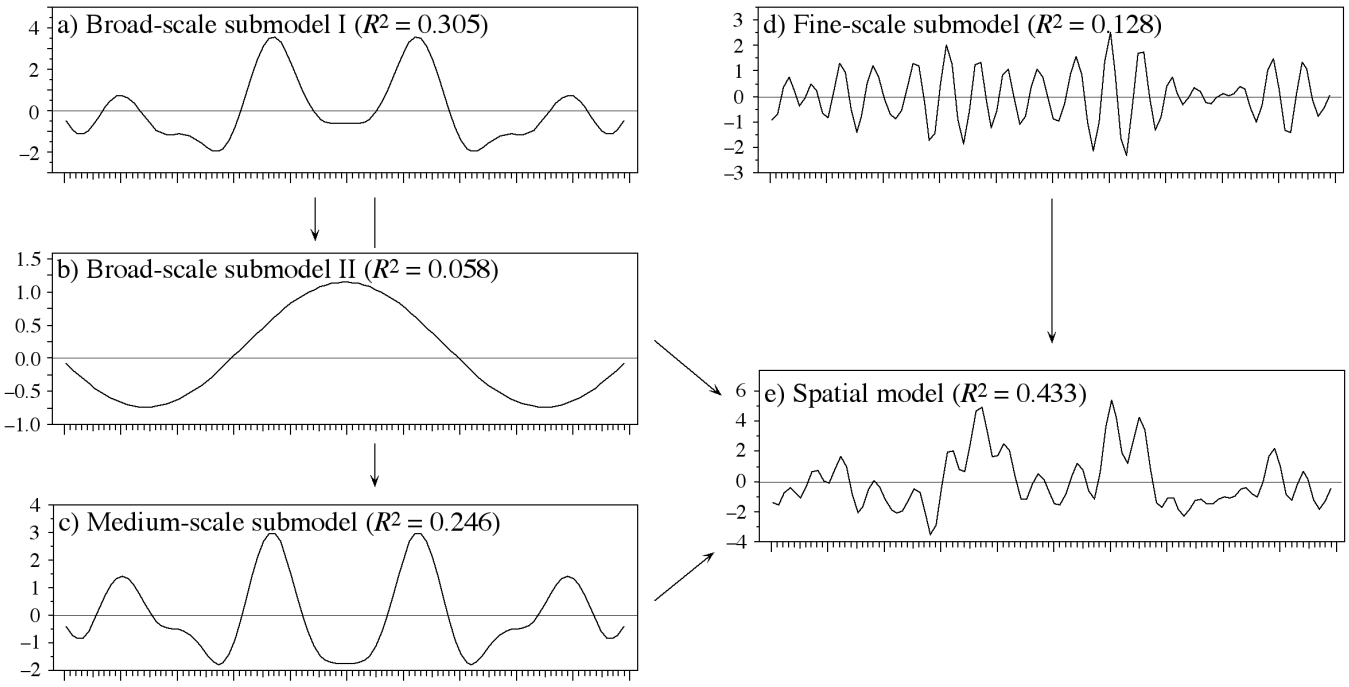

Figure 50 - Minimum spatial model and its additive submodels obtained by PCNM analysis on the (detrended) artificial data shown in Figure 49.

The successful extraction of the structured information can be further illustrated by comparing (Figure 51):

- the model of the detrended data obtained above (reproduced in Fig. 51b) to the sum of the four components “central bump”, “4 waves”, “17 waves” and “random autocorrelated” (Fig. 51a), and

- the residuals of the spatial model (Fig. 51d) to the real noise built into the data, i.e., the uncorrelated random variate (Fig. 51c).

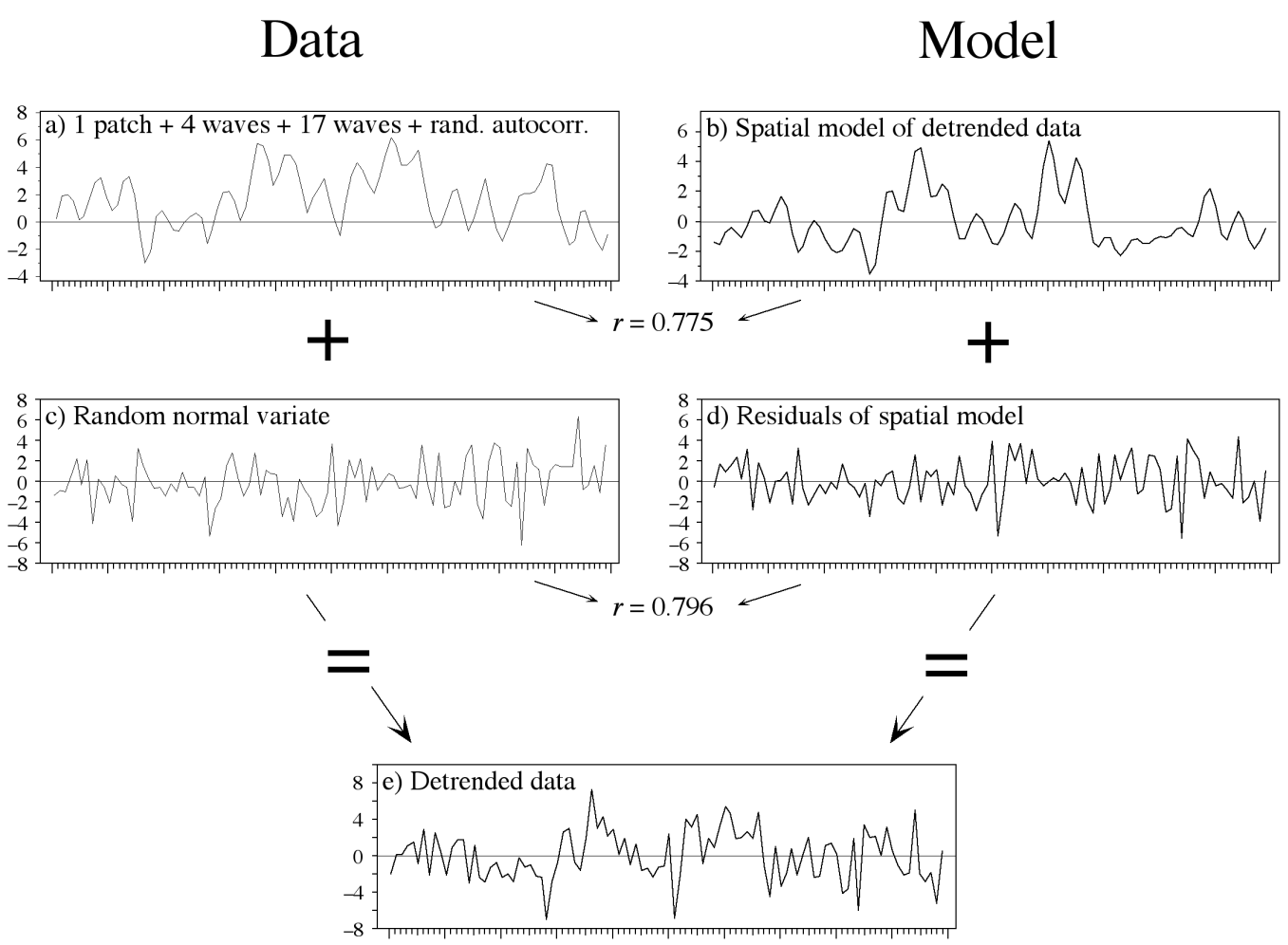

Figure 51 - Comparison of the structured (a) and random (c) components of the data on the one hand, and the spatial model (b) and its residuals (d) on the other hand, and correlations between the homologous components.

Ecological interpretation of a PCNM analysis

In the univariate case (as above), the simplest way of interpreting the results of a PCNM analysis is to regress the fitted values of the PCNM model on the environmental variables available in the study. This ensures that only the spatialized fraction of variation of the response variable is interpreted, but it bears the inconvenient that all spatial scales are confounded in the model. To unravel the scales where ecological processes take place, it is generally more fruitful to

decompose the PCNM model into submodels as above, and regress the fitted values of these submodels separately on the environmental variables. Each submodel is likely to be explainable by a different subset of environmental variables and, since the submodels are orthogonal to one another, the results will reveal scale-dependent processes that act independently on the response variable. Examples can be found in Borcard *et al.* (2004).

Setup and interpretation of a PCNM analysis in the multivariate case

If the research involves a matrix of response variables **Y** (e.g. a matrix of species abundances), the PCNM analysis is run by canonical ordination instead of multiple regression. A subset of significant PCNM base functions can still be selected (for instance by forward selection with double stopping criterion). If RDA is used, one obtains an R^2 (called in this case a bimultivariate redundancy statistic) that can be adjusted for the number of objects and explanatory variables in the same way as an ordinary R^2 can be. After this selection, several paths can be followed to further interpret the results:

Path 1: the RDA is run normally, and the fitted site scores of the most important canonical axes are regressed on the environmental variables as above. This path produces one orthogonal model of spatially structured data for each canonical axis, but since all PCNM base functions (=PCoA axes=PCNM variables) are involved in each axis, the spatial scales are confounded.

Path 2: this path consists in grouping the significant PCNM variables into scales (as in the artificial example above), and running a separate RDA for each group of PCNM variables. Each RDA will yield a series of canonical axes that will be spatially structured at a scale defined by the subset of PCNM variables used in the analysis. The most important axes of each RDA can be explained by regressing them on the environmental variables.

Path 3: a more complex, but potentially very powerful approach, is to combine PCNM analysis with variation partitioning. For instance, one could proceed as follows:

- forward-select the significant PCNM variables;

- group the significant PCNM variables into *k* subgroups of different spatial scales (for instance *k* = 3);

- forward-select the environmental variables;

- run a variation partitioning using the *k* subgroups of PCNM variables as well as the significant environmental variables (taken as one separate group of explanatory variables).

This path will yield a detailed assessment of the amount of spatial and nonspatial variation explained by or without the environmental variables at all scales.

Further remarks and summary notes on PCNM base functions

PCNM variables represent a spectral decomposition of the spatial relationships among the study sites.

They can be computed for regular or irregular sets of points in space or time.

PCNM variables are orthogonal. If the sampling design is regular, they look like sine waves. This is a property of the eigen-decomposition of the centered form of a distance matrix (Laplacian). If the sampling design is irregular, the PCNM variables have irregular shapes as well, but they can still be roughly ordered from broad-scale to fine-scale.

The grouping of PCNM variables into submodels of various scales implies arbitrary decisions about the building of the groups.

PCNM variables can also be computed for circular sampling

designs. An example can be found in Brind'Amour et al. (2005).

PCNM analysis can be used for temporal analysis, as well as spatio-temporal analysis. Research is presently underway to allow the analysis of spatio-temporal designs without spatial replication while still testing the interaction[5].

The concept of PCNM has been recently generalized to that of Moran's Eigenvector Maps (MEM); several ways of computing such vectors are now available (Dray *et al.*, 2006)[6]. Furthermore, another offspring of PCNM, a method called *asymmetric eigenvector maps* (AEM) has recently been developed to model directional processes (as can be found for instance in streams)[7].

[5] Pierre Legendre, Miquel De Caceres and Daniel Borcard (*submitted to Ecology*). Community surveys through space and time to assess environmental changes: testing the space-time interaction in the absence of replication.

[6] Dray, S., P. Legendre and P. Peres-Neto. Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *Ecological Modelling* **196**: 483-493.

[7] Blanchet, F. G., Pierre Legendre and Daniel Borcard (*accepted pending minor corrections*): Modelling directional spatial processes in ecological data (*Ecological Modelling*).