

DISSECTING THE SPATIAL STRUCTURE OF ECOLOGICAL DATA AT MULTIPLE SCALES

DANIEL BORCARD,^{1,4} PIERRE LEGENDRE,¹ CAROL AVOIS-JACQUET,^{1,2} AND HANNA TUOMISTO³

¹Département de sciences biologiques, Université de Montréal, C.P. 6128, succursale Centre-ville, Montréal, Québec, Canada H3C 3J7

²Station d'Hydrobiologie Lacustre, INRA, BP 511, F-74203 Thonon-les-Bains Cedex, France

³Department of Biology, University of Turku, FIN-20014 Turku, Finland

Abstract. Spatial structures may not only result from ecological interactions, they may also play an essential functional role in organizing the interactions. Modeling spatial patterns at multiple spatial and temporal scales is thus a crucial step to understand the functioning of ecological communities. PCNM (principal coordinates of neighbor matrices) analysis achieves a spectral decomposition of the spatial relationships among the sampling sites, creating variables that correspond to all the spatial scales that can be perceived in a given data set. The analysis then finds the scales to which a data table of interest responds. The significant PCNM variables can be directly interpreted in terms of spatial scales, or included in a procedure of variation decomposition with respect to spatial and environmental components. This paper presents four applications of PCNM analysis to ecological data representing combinations of: transect or surface data, regular or irregular sampling schemes, univariate or multivariate data. The data sets include Amazonian ferns, tropical marine zooplankton, chlorophyll in a marine lagoon, and oribatid mites in a peat bog. In each case, new ecological knowledge was obtained through PCNM analysis.

Key words: chlorophyll a; oribatid mites; principal coordinates of neighbor matrices (PCNM); sampling design; scale; spatial analysis; tropical ferns; tropical zooplankton; variation partitioning.

INTRODUCTION

The importance of spatial ecological structures is now widely recognized in ecological theory (Legendre and Fortin 1989, Legendre 1993, Peterson and Parker 1998). The interactions between living communities and their physical environment, and among the organisms themselves, occur at definite spatial and temporal scales, and give rise to spatial patterns that need to be assessed to untangle the processes structuring these communities. This assessment is not trivial when one's objective is to include in the model all the scales perceived in a given data set. Among the methods that have been proposed to include space as an explicit predictor in ecological modeling, Legendre and Troussellier (1988) built a matrix of Euclidean distances to be used in a series of Mantel and partial Mantel tests, and Legendre (1990) proposed to use the geographic coordinates directly as explanatory variables in constrained ordination techniques, augmented by all terms of a cubic trend-surface equation. The latter approach was integrated into a method of variation partitioning, where ecological variation was decomposed into four fractions using partial constrained regression or ordination methods (Borcard et al. 1992, Borcard and Legendre 1994, Legendre and Borcard 2004). This technique has proved very successful and is now widely

applied in ecology; see references in Legendre and Legendre (1998:775).

Trend-surface analysis only allows the broad-scale spatial variation (at the scale of the extent of the sampling campaign) to be modeled. Therefore, Borcard and Legendre (2002) proposed a new approach devised to identify spatial patterns across the whole range of scales perceptible with a given data set. This method (Fig. A1 in Appendix A) is based on the computation of the principal coordinates of a matrix of geographic neighbors among the sampling sites (PCNM, acronym for principal coordinates of neighbor matrices). The present paper illustrates the application of PCNM analysis to real ecological data observed using various spatial designs: linear (transect) and two-dimensional (surface), regular or irregular sampling schemes, and presents various ways to obtain new ecological knowledge from the results.

THE METHOD OF PCNM ANALYSIS

A detailed account of PCNM analysis is found in Borcard and Legendre (2002). The method creates a set of spatial explanatory variables that have structure at all the scales encompassed by the data matrix (akin to the series of sines and cosines used in Fourier analysis), and determines to which of these variables the response data (univariate or multivariate) are statistically responding. The steps to create the PCNM variables are summarized in Table A1 of Appendix A. Note that using this method does not imply that one expects

Manuscript received 4 August 2003; revised and accepted 9 January 2004; final version received 24 February 2004. Corresponding Editor: G. M. Henebry.

⁴ E-mail: Daniel.Borcard@umontreal.ca

to find periodic spatial structures: the spectral variables are used mainly to model nonlinear structures through a flexible combination of sine waves.

When the sampling design is a linear transect with equidistant points, the spatial variables are a series of sine waves with periods decreasing from $n + 1$ (where n is the number of points in the transect) to ~ 3 . The number of variables obtained in this case is $2n/3$ rounded to the upper integer value. Under other sampling designs, the number and shape of the PCNM variables may differ, but they always correspond to the number of positive eigenvalues of the principal coordinate analysis.

The response data should be checked for linear trends prior to the application of PCNM analysis. A trend indicates the presence of a spatial structure at broader scale than the sampling extent. Although PCNM analysis is capable of recovering linear trends, it is pointless to use a combination of sine waves to model gradients. This may also obscure other structures that could have been recovered from the data; half the available PCNM variables would be needed to model the trend, so that their role in modeling finer structures could go unnoticed.

ILLUSTRATION USING REAL ECOLOGICAL DATA

The various data sets presented here deal with different kinds of data and sampling strategies, and we propose several ways of interpreting the results, which may suit different research purposes.

Regular one-dimensional sampling (transect): a fern species in upper Amazonia

The data were gathered during a research program on vegetation ecology of Amazonian tropical rain forests. PCNM (principal coordinates of neighbor matrices) analysis was applied to two linear transects located in Northeastern Peru, in the regions of Nauta ($4^{\circ}27' S$, $73^{\circ}35' W$) and Huanta ($3^{\circ}17' S$, $71^{\circ}51' W$). Previous analyses on these transects can be found in, e.g., Tuomisto et al. (1995) and Tuomisto and Poulsen (2000). Each transect consists of 260 adjacent, square (5×5 m) subplots where the abundances (number of individuals) of all species of ferns were recorded, as well as the following environmental variables: elevation above the lowest point in the transect (in meters), thickness of organic soil horizon (in centimeters; a single measurement at the center of the subplot), degree of drainage (six ordered classes), canopy height (in meters), number of tree stems in five ordered classes of diameter at breast height (dbh), number of lianas in three diameter classes, tree canopy coverage (percentage), shrub coverage (percentage) and (at Nauta only) herb coverage (percentage). We analyzed the spatial distribution of the most abundant species, *Adiantum tomentosum* Klotzsch, to assess the spatial scales at which the abundance of this species is structured, and relate these scales to the environmental variables. We could have analyzed the whole fern com-

munity if the study had been aimed at answering community-level questions, but the results would have been less precise for any one species. The Nauta results are presented in detail here; the Huanta results are found in Appendix B.

Step 1: PCNM analysis of A. tomentosum at Nauta.—The *A. tomentosum* abundances were first square-root transformed to make them more symmetrical (if not strictly normal). They presented a significant linear spatial trend ($R^2 = 0.102$, $P = 0.001$, 999 permutations). Therefore, we detrended the variable before PCNM analysis. PCNM generation over 260 quadrats produced 176 PCNM variables. Among them, 50 were significant, yielding a parsimonious model with $R^2 = 0.815$. The spatial model was arbitrarily partitioned into four additive submodels corresponding to the following scales: very broad (the 10 first PCNM variables, which were all significant; $R^2 = 0.333$), broad (8 significant PCNM from the group 11 to 20; $R^2 = 0.239$), medium (12 significant PCNM from 21 to 50; $R^2 = 0.126$) and fine (20 significant PCNM from 51 to 176; $R^2 = 0.117$). To identify the significant periods in the submodels, we computed a contingency periodogram (a method for periodic analysis of short data sets; Legendre et al. 1981) on each of them. Significant periods of 50 and between 71 and 88 subunits, i.e., of 250 m and between 355 and 440 m, were found in the very-broad-scale submodel; 36 subunits (180 m) in the broad-scale submodel; 18 subunits (90 m) at medium scale; and 10 and 13 subunits (50 and 65 m) at fine scale. The data, model, and submodels are displayed in Fig. 1.

Step 2: regression of fern data and PCNM submodels on environmental variables.—To identify the environmental variables related to *A. tomentosum* abundance at all scales, the (square-root-transformed and detrended) fern data as well as the four spatial submodels obtained above were submitted to multiple regressions with backward elimination of explanatory variables. Linearity of the relationship between the dependent and each (untransformed) explanatory variable was checked visually. The significant variables (with probabilities) are given in Table B1 of Appendix B. Most environmental variables explaining one or more spatial submodels are significant in a classical, non-spatial regression involving the fern counts. Most prominent is elevation, significant at very broad and broad scales. This variable acts as a proxy for unmeasured edaphic factors that vary at these scales, for instance soil differences related to catena formation. Of the other soil properties, thickness of organic soil horizon is significant in the models at very broad scale, and drainage at medium scale. The influence of vegetation structure is represented by several variables. Canopy height is the most important at very broad scale. The number of trees with dbh 3–7.5 cm has some influence at broad scale. The canopy, shrub, and herbaceous coverage variables each explain one submodel significantly, without

appearing significant in the regression of fern counts on the original environmental variables. Their effect is thus better detected by a filtered variable focusing on the relevant spatial scale. No environmental influence could be related to the fine-scale pattern; see *Discussion*, below.

At Huanta (Table B2 and Fig. B1 in Appendix B), the soil-related variables are the same and act at the same scales as at Nauta, although the spatial component in fern abundance was much weaker than at Nauta. None of the vegetation coverage variables is significant, and canopy height acts at a finer scale. Several variables of tree and liana counts are significant at very broad scale, which is probably related to the peculiar shape of this submodel, showing essentially one major bump and trough in the right-hand portion of the transect.

*Irregular one-dimensional sampling (transect):
zooplankton in Guadeloupe*

PCNM analysis was used to describe the multiscale spatial variation of zooplankton biomass inhabiting a coastal reef lagoon (Guadeloupe, 16°18' N, 61°34' W) and test hypotheses about the biological and physical processes causing zooplankton patchiness (Avois-Jacquet 2002). The zooplankton (two size classes: 190–600 μm and $>600 \mu\text{m}$) and environmental variables were sampled at 51 irregularly spaced sites (sampling interval: 15–637 m), arranged along a straight-line cross-shelf transect 8.4 km in length. A significant spatial trend was found in the two zooplankton size classes and extracted by regression. PCNM analysis was carried out on the detrended data.

Production of the PCNM variables required a special treatment. The site spacing along the transect was irregular. In order to allow the PCNM analysis to detect features smaller than the widest gap (637 m) between neighboring sites, six supplementary objects were added into the data matrix containing the site coordinates. This filled the voids, leaving a widest gap of 227 m, which provided the truncation distance of the matrix. After principal-coordinate analysis, the supplementary objects were removed from the matrix of PCNM variables. This operation did not affect the dependent variables: it simply allowed us to obtain finer-scale PCNM variables. This trick comes with a price. Since the supplementary objects were removed following the computation of the principal coordinates, there is a loss of orthogonality among the coordinates. Actually, in this case, the largest correlation (between PCNM variables 10 and 26) was 0.295, but most other values were below 0.10. For irregular sampling designs, one has to find a trade-off between fine resolution and orthogonality of the spatial variables, two key properties of PCNM analysis (Borcard and Legendre 2002).

Due to the irregularity of the sampling, the 28 PCNM variables do not portray regular sine waves along the transect, but they can nevertheless be sorted into scales of increasing fineness by examining their shapes or by

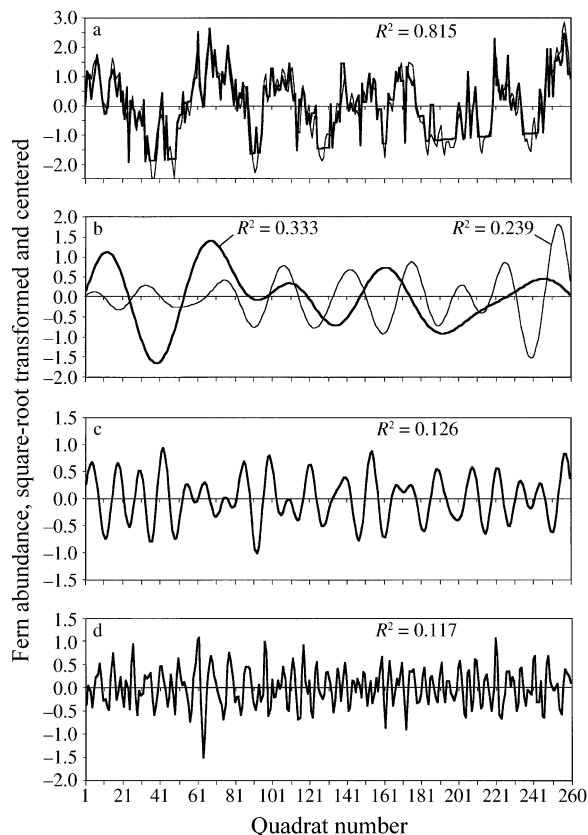


FIG. 1. Square-root-transformed *Adiantum tomentosum* abundances and spatial PCNM (principal coordinates of neighbor matrices) model and submodels (ordinate) along the Nauta transect (abscissa; 260 quadrats): (a) *A. tomentosum* (thick line) and parsimonious PCNM model; (b) very-broad-scale (thick line) and broad-scale models; (c) medium-scale model; (d) fine-scale model. The R^2 values are calculated with respect to the total variation in the transformed fern count data; they are thus additive. Fern-abundance data before transformation were measured as number of ferns per 25-m² quadrat.

computing variograms or correlograms. Examples are shown in Fig. C1 of Appendix C). In regular one-dimensional designs, the order of the PCNM axes always corresponds to the size of the patterns they describe. As usual, only a subset of the PCNM variables significantly explained the variance of the two size classes of zooplankton biomass (Table C1 in Appendix C); they were arbitrarily subdivided into broad-scale (1–5 km: PCNM variables 1 to 9), and medium-scale influences (400–1000 m: PCNM variables 10 and above).

PCNM analysis yielded an $R^2 = 0.579$ with 10 significant PCNM variables for the small zooplankton (Fig. C2 in Appendix C) and 0.339 with 4 PCNM for the large zooplankton (Fig. C3 in Appendix C). The significant PCNM variables for the two fractions are given in Table C1 of Appendix C. PCNM variables 6 and 9 were significant for both fractions of zooplankton, but the small organisms showed significant spatial variation at more scales than the large ones. The large

and small zooplankton differed strongly in the intensity and spectral composition of their patchiness.

At the scale of the whole lagoon, the biomass pattern consisted in a decreasing gradient from the coast to the barrier reef. The broad- and medium-scale patterns revealed by PCNM analysis (Figs. C2 and C3 in Appendix C) formed waves with periods ranging from 800 m to 2 km. From 0 to 4 km from the shore, large and small zooplankters exhibited low biomass near the coast and high biomass in the middle of the lagoon (Fig. C2c and C3c in Appendix C). At finer scales (Fig. C2d and C3d in Appendix C), they showed patchiness across the whole transect.

To identify the environmental processes that contribute to the generation and maintenance of multiscale spatial variation in the zooplankton, multiple regression with forward selection was used to select the environmental variables that significantly explained each spatial pattern of biomass for each zooplankton size class (i.e., 190–600 μm and $>600 \mu\text{m}$). The undetrended environmental data were used first to explain the trend identified in the zooplankton data (i.e., the decreasing biomass gradient from the coast to the barrier reef). Second, the detrended environmental data were used to explain the PCNM models of the detrended zooplankton data. Detailed results are presented in Avois-Jacquet (2002); a short account is given here.

At the scale of the trend, which corresponds to processes larger than the extent of the survey, the decreasing gradient of zooplankton biomass was mainly explained by decreasing phytoplankton biomass. Salinity increased from the coast to the barrier reef, whereas wind speed and dissolved oxygen decreased.

The response of zooplankton at the scale represented by PCNM 6 was mostly linked to the presence of cays and seagrass beds, followed by swell height and dissolved oxygen. The cays had higher zooplankton biomass, greater swell height, and lower values of dissolved oxygen than the seagrass beds. At the scale of PCNM 9, deep muddy bottoms and seagrass beds were negatively linked to zooplankton biomass. Turbidity and swell height were positively linked to zooplankton biomass and showed high values on the cays. The finer-scale patterns of small (PCNM 12) and large (PCNM 15) zooplankton were mainly explained by the habitat variables. Wind speed explained PCNM 12 significantly, and large zooplankton showed a negative relationship with phytoplankton biomass at the scale of PCNM 15; this relationship was positive at the scale of the whole lagoon. The results support the hypothesis that the structure of zooplankton biomass and of its generating processes (types and effects) vary across spatial scales (Mackas et al. 1985), as well as the hypothesis that zooplankton spatial structuring is size specific (Piontkovski and William 1995). But globally, at the trend- to broad-scale, the spatial structure of zooplankton biomass was mostly explained by hydrody-

namic and biological variables, while across broad to medium scales, the types of habitat dominated.

*Regular two-dimensional sampling grid:
chlorophyll in a brackish lagoon*

First analyzed using Mantel tests by Legendre and Troussellier (1988), this is one of the data sets used to demonstrate the method of variation partitioning (Borcard et al. 1992). Fig. D1 in Appendix D shows the results of a PCNM analysis of a single variable, chlorophyll *a*, over a geographic surface, the Thau marine lagoon (Hérault, southern France: from 43°20' to 43°28' N, and from 3°32' to 3°42' E; 19 km in length). Chlorophyll *a* (Chl *a*) is an often-used indicator of the abundance of phytoplankton in aquatic ecosystems. A regular sampling grid, with 1-km mesh size, was used to sample 63 sites in the Thau lagoon on 17 June 1986 (Amanieu et al. 1989). Principal-coordinate decomposition of the truncated distance matrix (truncation distance: 1 km) produced 45 PCNM variables; four of them are shown in Fig. D1 (e–h). Chl *a* (Fig. D1a) was regressed on the 45 PCNM variables. Backward elimination was applied, resulting in 12 significant PCNM variables explaining together 78% of the spatial variance of the Chl *a* data. The 12 PCNM variables were divided in three groups corresponding to the broad (PCNM 1, 3, 5, and 8), intermediate (PCNM 13, 14, 17, 19, and 20), and fine scales (PCNM 24, 28, and 36). The values of Chl *a* adjusted to these submodels were computed and plotted in Fig. D1b–d. The broad-scale model describes the main hydrological gradient across the lagoon, controlled by inflow of marine water, mainly through the harbor of Sète near the eastern end of the lagoon, and inflow of freshwater from Canal du Midi at the western end of the lagoon and other rivers of the drainage basin. The intermediate-scale model predicts high values near the largest towns around the lagoon, except Mèze (number 5 in Fig. D1c), which has a highly efficient sewage treatment plan (Troussellier et al. 1986). The fine-scale model, which is also statistically significant, predicts high values of Chl *a* in specific areas in the lagoon, but we could find no ecological interpretation of the spatial variation predicted by this model; see *Discussion*, below.

*Irregular two-dimensional sampling:
oribatid mites in a Sphagnum carpet*

One of the data sets used by Borcard et al. (1992) to illustrate the variation partitioning method is here reexamined using PCNM analysis. Seventy soil and *Sphagnum* cores (5 cm in diameter, 7 cm deep) were collected from a small, 10 \times 2.6 m area in the partially floating vegetation mat extending from the forest border into a Laurentian peat-bog lake (Lac Geai, 46° N, 74° W, Station de biologie des Laurentides, Québec, Canada). Details about the site and the field methods are found in Borcard and Legendre (1994). The sampling design formed an irregular pattern.

The species matrix contained 70 objects (cores) and 35 species; the environmental matrix contained the following variables: type of substratum (seven unordered classes), microtopography (blanket or hummock), coverage density of the shrub cover (three semi-quantitative classes), water content in grams per liter of fresh uncompressed matter (quantitative), and bulk density in grams per liter of dry uncompressed matter.

In the context of variance partitioning, Borcard and Legendre (2002) showed how the substitution of the third-order polynomial by a set of PCNM variables improved the modeling of fraction [c] (strictly spatial) without much affecting fractions [a] and [b]. Here we show how the modeling of medium and fine-scale patterns can be refined, compared to the results obtained by means of a third-order polynomial (Borcard and Legendre 1994).

A preliminary redundancy analysis (RDA; Rao 1964) involving only the *X* and *Y* geographic coordinates as explanatory variables, and thus looking for a linear gradient, significantly explained 28.4% of the variance of the data. Prior to that RDA, the species data had been Hellinger-transformed following Legendre and Gallagher (2001). An a posteriori multiple regression of the first (and only significant) RDA axis on the set of environmental variables showed that it can be explained by a combination of water content and substratum density (Fig. E1 in Appendix E).

For PCNM analysis, starting from the *X* and *Y* geographic coordinates, a matrix of PCNM variables was created with a truncation distance of 1.012 m. This critical distance was obtained by running a single linkage clustering on a matrix of Euclidean distances among the sites, and computing the chain of primary connections. The largest of these connections is the shortest distance to retain in the truncation process of the PCNM method to maintain the graph of all the objects connected. This procedure yielded 43 PCNM variables; they did not display the regular, sine-shaped patterns that are obtained for regular sampling lines or grids. The progression from broad to fine scale is preserved, however: the first PCNM variables represent coarse patterns and the last ones finer-scale patterns (Fig. E2 in Appendix E).

Since the preliminary RDA above had demonstrated the presence of a linear gradient in the data, the analysis presented here was a partial RDA with forward selection among the 43 PCNM variables, with the *X* and *Y* coordinates as covariables to control for the gradient, while modeling the spatial structure using PCNM variables.

Forward selection retained 12 PCNM variables. The model explained 45.1% of the variance of the detrended data (i.e., after controlling for the effect of the covariables), or 32.3% of the undetrended data. To interpret the three significant canonical axes, we regressed each of them on our set of environmental variables, with backward elimination of explanatory variables. Qual-

itative variables had been recoded using dummy variables.

The first axis, explaining 22.6% of the detrended data, is mainly a combination of PCNM 4, 11, 3, and 1 (in decreasing order of importance in the model). It represents a broad-scale alternation of troughs and bumps, the latter corresponding mainly to the zones with few or no shrubs (shrubs are the most significant environmental variables; multiple regression $R^2 = 0.481$). Shrubs are distributed in coarse patches across the sampling area, which translates into a spatial model that features mainly broad-scale PCNM variables (Fig. E3 in Appendix E).

The second canonical axis, which explains 8.4% of the detrended data, is mainly a combination of PCNM 1, 3, and 5 and thus it features patterns at the same broad scale as the first one; by definition of canonical axes, they are linearly independent of the first axis. Axis 2 is explained by the zones with few shrubs, but excluding (negative correlation) those with hummocks (Fig. E4 in Appendix E). No other available environmental variable was significant to explain these patterns, and the regression had an R^2 of only 0.114, which suggests that other, unidentified mechanisms are responsible for at least part of that spatial structure.

The third canonical axis, which explains 4.5% of the detrended data (not shown), is mainly a combination of PCNM 37, 6, and 16. The pattern represents patches with low water content and no shrubs ($R^2 = 0.337$).

When compared to the results obtained by means of a third-order polynomial (Borcard and Legendre 1994), these new analyses confirm that most of the spatial patterns that are linked to known environmental variables occur at broad to medium scales. But there remains some spatial variation left to model. A partial RDA of the species data by the PCNM variables, controlling for the effects of the environmental variables and the *X*-*Y* trend, yielded two significant axes (Fig. E5 in Appendix E). As explained by Borcard and Legendre (1994), such patterns may have been produced by several unmeasured abiotic or biotic mechanisms. Here, for instance, the distribution of some sources of food like fungal hyphae are good candidates. Alternatively, the patterns may also be the result of biotic processes generating spatial autocorrelation in the response variables. Testing these hypotheses could be the subject of a follow-up study.

DISCUSSION

PCNM variables, which have the familiar shape of sine waves in the case of regular sampling, are obtained because an eigenvalue decomposition of the Laplacian (centered similarity matrix) representing a neighborhood graph produces a spectral decomposition of space (Brillinger 1981, Chung 1997). PCNM analysis can model spatial structures at all the spatial scales that can be perceived by the data set. This means that it cannot model structures larger than the extent of the sampling

design, nor features smaller than the distance used to truncate the distance matrix. In time-series analysis, the “observation window” is bounded by similar constraints (Legendre and Legendre 1998: Chapter 12).

In each example presented in this paper, new insights were obtained by PCNM analysis, compared to classical regression or canonical analysis against environmental explanatory variables, because the species–environment relationships could be related to one or several spatial scales. This often represents crucial information to interpret the results of a study, because the scale of observation strongly influences the perception of relationships among variables. For instance, net primary production is mainly driven by temperature and moisture at continental scale, but at regional scale, where the climate is relatively homogeneous, aspect and soil are the main factors (O’Neill and King 1998). At finer scales yet, plant communities may be structured by biotic processes such as competition, and the resulting patterns could be modeled by fine-scale PCNM variables.

As in any other statistical model, there is no warranty that the environmental variables that are related to the species PCNM spatial patterns have a direct causal link to the species themselves. Relationships may be found because of common links to other factors. In all regression-based analyses, causality resides in the hypotheses. Finding patterns that are consistent with the hypotheses reinforces them but does not formally prove them.

In the framework of variation partitioning, a significant fraction [c] (i.e., the fraction of variation explained by the spatial variables only) suggests that some natural process (see below) has generated the identified spatial structure, even though no explanatory variable is presently available to explain it. Mapping fraction [c] helps ecologists formulate new hypotheses about these processes. Besides the environmental-control model, ecologists may call upon other classes of processes to explain this fraction of variation: population or community dynamics, other biotic processes like predator–prey interactions, and historical dynamics; explanations are given in Borcard and Legendre (1994: Table 3) and in Legendre and Legendre (1998: Table 13.3).

Fine-scale patterns identified by PCNM analysis are often explained by none of the available environmental variables. This was the case in all the examples in this paper. Fine-scale spatial patterns may depict spatial autocorrelation, which is generated by dynamic processes within the dependent variables themselves (e.g., biotic interactions among individuals). If that is the case and the analysis could be repeated at several identical sites, similar patterns would be identified at fine scale, but the autocorrelated patches would be in different locations. An example is shown in Fig. 1 of Legendre et al. (2002). Of course, one cannot exclude

the possible influence of unmeasured, fine-scale-structured abiotic variables.

PCNM analysis is a powerful way to model spatial structures, revealing their appearances. Interpretation of the results in terms of ecological processes requires, however, a deep ecological understanding of the systems under study, since PCNM analysis does not model processes directly. This is also the case for all other methods of spatial analysis.

The data sets used in the four examples presented in this paper are found in the Supplement.

ACKNOWLEDGMENTS

The work reported in this paper was supported by NSERC grant number OGP0007738 to P. Legendre.

LITERATURE CITED

- Amanieu, M., P. Legendre, M. Troussellier, and G.-F. Frisoni. 1989. Le programme Écothau: théorie écologique et base de la modélisation. *Oceanologica Acta* **12**:189–199.
- Avois-Jacquet, C. 2002. Variabilité spatiale multiéchelle du zooplancton dans un lagon récifal côtier. Dissertation. Université Paris 6, Paris, France.
- Borcard, D., and P. Legendre. 1994. Environmental control and spatial structure in ecological communities: an example using oribatid mites (Acari, Oribatei). *Environmental and Ecological Statistics* **1**:37–61.
- Borcard, D., and P. Legendre. 2002. All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecological Modelling* **153**:51–68.
- Borcard, D., P. Legendre, and P. Drapeau. 1992. Partialling out the spatial component of ecological variation. *Ecology* **73**:1045–1055.
- Brillinger, D. R. 1981. Time series. Data analysis and theory. Expanded edition. Holden-Day, San Francisco, California, USA.
- Chung, F. R. K. 1997. Spectral graph theory. CBMS [Conference Board of the Mathematical Sciences] Regional conference series in mathematics, 0160-7642, number 92. American Mathematical Society, Providence, Rhode Island, USA.
- Legendre, P. 1990. Quantitative methods and biogeographic analysis. Pages 9–34 in D. J. Garbary and R. G. South, editors. *Evolutionary biogeography of the marine algae of the North Atlantic*. NATO ASI series, volume G 22. Springer Verlag, Berlin, Germany.
- Legendre, P. 1993. Spatial autocorrelation: trouble or new paradigm? *Ecology* **74**:1659–1673.
- Legendre, P., and D. Borcard. 2004. Quelles sont les échelles spatiales importantes dans un écosystème? Chapter 19 in J.-J. Droesbeke, M. Lejeune, and G. Saporta, editors. *Analyse statistique de données spatiales*. Éditions TECHNIP, Paris, France.
- Legendre, P., M. R. T. Dale, M.-J. Fortin, J. Gurevitch, M. Hohn, and D. Myers. 2002. The consequences of spatial structure for the design and analysis of ecological field surveys. *Ecography* **25**:601–615.
- Legendre, P., and M.-J. Fortin. 1989. Spatial pattern and ecological analysis. *Vegetatio* **80**:107–138.
- Legendre, L., M. Fréchette, and P. Legendre. 1981. The contingency periodogram: a method of identifying rhythms in series of nonmetric ecological data. *Journal of Ecology* **69**:965–979.
- Legendre, P., and E. D. Gallagher. 2001. Ecologically meaningful transformations for ordination of species data. *Oecologia* **129**:271–280.
- Legendre, P., and L. Legendre. 1998. Numerical ecology. Second English edition. Elsevier Science BV, Amsterdam, The Netherlands.

- Legendre, P., and M. Troussellier. 1988. Aquatic heterotrophic bacteria: modeling in the presence of spatial autocorrelation. *Limnology and Oceanography* **33**:1055–1067.
- Mackas, D. L., K. L. Denman, and M. R. Abbott. 1985. Plankton patchiness: biology in the physical vernacular. *Bulletin of Marine Science* **37**:652–674.
- O'Neill, R. V., and A. W. King. 1998. *Homage to St. Michael; or, Why are there so many books on scale?* Pages 3–15 in D. L. Peterson and V. T. Parker, editors. *Ecological scale—theory and applications*. Columbia University Press, New York, New York, USA.
- Peterson, D. L., and V. T. Parker, editors. 1998. *Ecological scale—theory and applications*. Columbia University Press, New York, New York, USA.
- Piontkovski, S. A., and R. Williams. 1995. Multiscale variability of tropical ocean zooplankton biomass. *ICES Journal of Marine Science* **52**:643–656.
- Rao, C. R. 1964. The use and interpretation of principal component analysis in applied research. *Sankhyā A* **26**:329–358.
- Troussellier, M., P. Legendre, and B. Baleux. 1986. Modeling of the evolution of bacterial densities in an eutrophic ecosystem (sewage lagoons). *Microbial Ecology* **12**:355–379.
- Tuomisto, H., and A. D. Poulsen. 2000. Pteridophyte diversity and species composition in four Amazonian rain forests. *Journal of Vegetation Science* **11**:383–396.
- Tuomisto, H., K. Ruokolainen, R. Kalliola, A. Linna, W. Danjoy, and Z. Rodriguez. 1995. Dissecting Amazonian biodiversity. *Science* **269**:63–66.

APPENDIX A

A table presenting the steps used to create a set of principal coordinates of neighbor matrices (PCNM) variables and a figure showing a schematic description of PCNM analysis is available in ESA's Electronic Data Archive: *Ecological Archives* E085-050-A1.

APPENDIX B

A figure and two tables showing results of the analysis of the Amazonian fern data (Peru) are available in ESA's Electronic Data Archive: *Ecological Archives* E085-050-A2.

APPENDIX C

Three figures and a table showing results of the analysis of the zooplankton data from Guadeloupe are available in ESA's Electronic Data Archive: *Ecological Archives* E085-050-A3.

APPENDIX D

A figure showing results of the analysis of the chlorophyll *a* from the Thau (France) lagoon is available in ESA's Electronic Data Archive: *Ecological Archives* E085-050-A4.

APPENDIX E

Five figures showing results of the analysis of the oribatid mites from Saint-Hippolyte, Quebec, are available in ESA's Electronic Data Archive: *Ecological Archives* E085-050-A5.

SUPPLEMENT

Data used in the four applications of principal coordinates of neighbor matrices (PCNM) analysis—abundance of ferns, transect coordinates, and environmental variables (Peru); biomass of zooplankton, transect coordinates, and environmental variables (Guadeloupe); chlorophyll *a* and spatial coordinates (Thau, France); and oribatid mite species counts, spatial coordinates, and environmental variables (Saint-Hippolyte, Quebec)—are available in ESA's Electronic Data Archives: *Ecological Archives* E085-050-S1.

Appendices to:

Borcard, D., P. Legendre, C. Avois-Jacquet and H. Tuomisto. 2004. Dissecting the spatial structure of ecological data at multiple scales. *Ecology* 85: 1826-1832.

APPENDIX A

Ecological Archives E085-050-A1

A TABLE PRESENTING THE STEPS USED TO CREATE A SET OF PRINCIPAL COORDINATES OF NEIGHBOR MATRICES (PCNM) VARIABLES (TABLE A1) AND A FIGURE SHOWING A SCHEMATIC DESCRIPTION OF PCNM ANALYSIS (FIG. A1).

TABLE A1. Summary of the steps used to create a set of PCNM variables. See also Fig. A1.

-
1. From the geographic coordinates of the sites, compute a matrix of Euclidean distances.
 2. Truncate the matrix obtained in step 1 at or just above the distance between nearest neighbours. If the first neighbours are not all equally distant, truncate just above the smallest distance that keeps all sites connected into a single network (connected graph). This can be obtained by computing a single linkage clustering of the matrix of Euclidean distances, and retaining the largest distance found in the chain of primary connexions.
 3. Fill the truncated portion of the distance matrix with an arbitrarily large distance value. A value equal to four times the largest distance retained in the truncated matrix is sufficient and appropriate.
 4. Compute the principal coordinates of the matrix obtained in step 3.
 5. Retain only the principal coordinates corresponding to positive eigenvalues. These are the spatial variables that can now be used as explanators in multiple regression or canonical ordination analysis.
 6. The number of PCNM spatial variables created by the principal coordinate analysis being high (about two thirds the number of objects), reduce the number of explanatory variables using an appropriate selection method, retaining only those that significantly contribute to the explanation of the response data. Possible ways of thinning the model to make it parsimonious include tests of significance of individual multiple regression coefficients (preferably by permutation), retaining the significant PCNM variables only, as well as backward, forward, or stepwise selection in regression, or forward selection of the explanatory variables in canonical analysis.
-

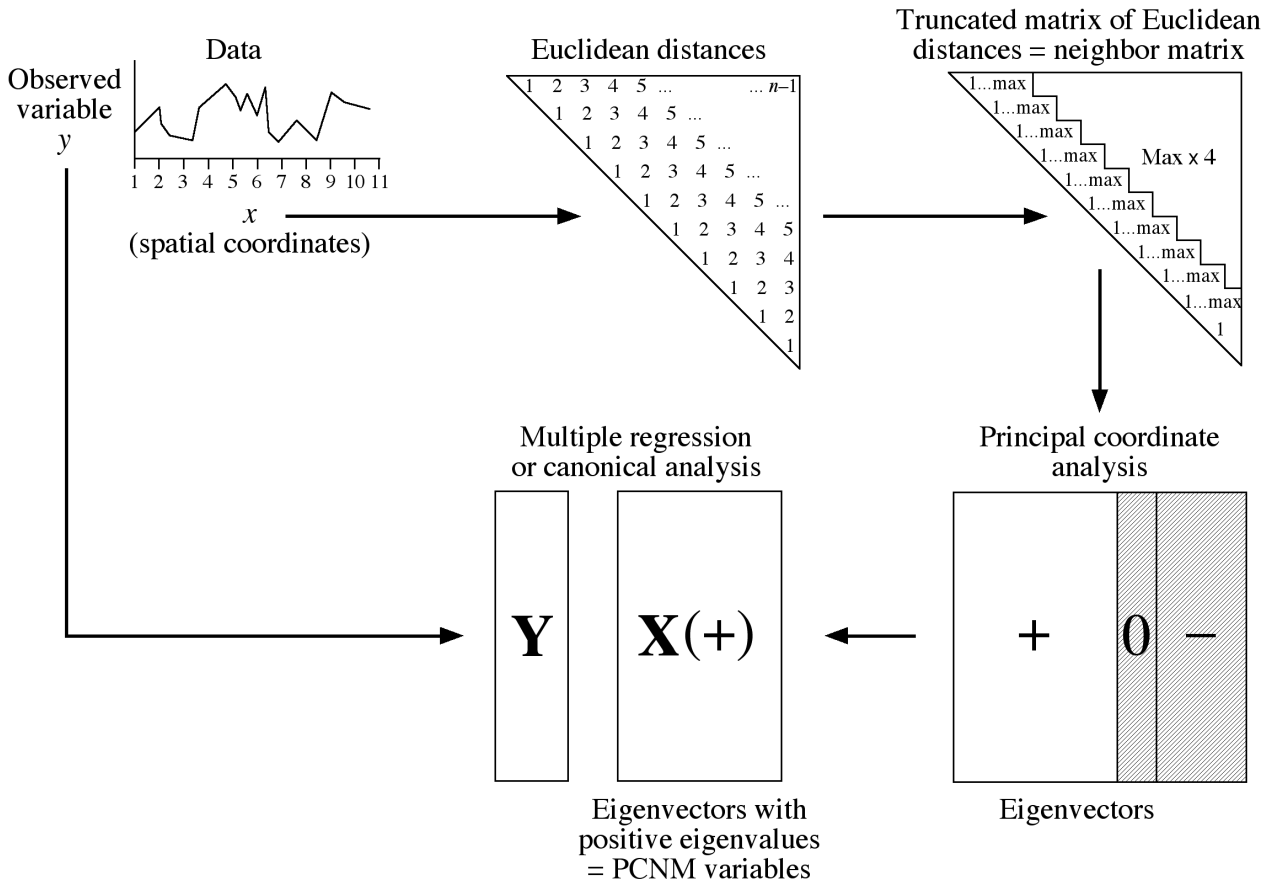


FIG. A1. Schematic description of PCNM analysis. The descriptors of spatial relationships (PCNM variables) are obtained by principal coordinate analysis of a truncated matrix of Euclidean (geographic) distances among the sampling sites. For regular sampling designs, truncation at the distance of first neighbors is recommended.

APPENDIX B

Ecological Archives E085-050-A2

TWO TABLES (TABLES B1 AND B2) AND A FIGURE (FIG. B1) SHOWING RESULTS OF THE ANALYSIS OF THE AMAZONIAN FERN DATA (PERU).

TABLE B1. R^2 and probabilities related to the spatial analysis of the fern *Adiantum tomentosum* at Nauta (Peru). First line: R^2 of each spatial submodel. Second line: R^2 of the regression of the submodel (fitted values) on a subset of backward-selected environmental variables. Third line: product of the two previous lines, i.e. variation of the fern data explained by the environmental variables at the scale considered. Other lines: p-values of the regression coefficients of the environmental variables in the model considered.

	A. <i>tomentosum</i>	Very broad	Broad	Medium	Fine
R^2 of PCNM submodel on <i>A. tomentosum</i>		0.333	0.239	0.126	0.117
R^2 of envir. on submodel		0.347	0.334	0.157	
R^2 of envir. on <i>A. tomentosum</i>	0.436	0.116	0.080	0.020	
Elevation (m)	< 0.0001	< 0.0001	< 0.0001		
Thickness of soil organic horizon (cm)	0.0004	< 0.0001			
Drainage	0.0393			0.0007	
Canopy height (m)	0.0011	< 0.0001			
Canopy coverage (%)				0.0024	
Shrub coverage (%)	0.0621	0.0341		0.0001	
Herb coverage (%)		0.0002			
Trees 3 - 7.5 cm DBH	0.0083		0.0009		
Lianas 3 - 7.5 cm diameter				0.0413	
Lianas 8 - 15 cm diameter				0.0183	

TABLE B2. R^2 and probabilities related to the spatial analysis of the fern *Adiantum tomentosum* at Huanta (Peru). First line: R^2 of each spatial submodel. Second line: R^2 of the regression of the submodel (fitted values) on a subset of backward-selected environmental variables. Third line: product of the two previous lines, i.e. variation of the fern data explained by the environmental variables at the scale considered. Other lines: p-values of the regression coefficients of the environmental variables in the model considered.

	<i>A. tomentosum</i>	Very broad	Broad	Medium	Fine
R^2 submodel on <i>A. tomentosum</i>		0.324	0.047	0.17	0.152
R^2 envir. on submodel		0.408	0.099	0.072	0.039
R^2 envir. on <i>A. tomentosum</i>	0.227	0.132	0.005	0.012	0.006
Elevation (trend)		0.0107			
Elevation (detrended)	< 0.0001	< 0.0001	< 0.0001		
Thickness of soil organic horizon (cm)		0.0087			0.0420
Drainage	0.0047			0.0002	0.0188
Canopy height (m)	0.0034			0.0006	
Trees 3 - 7.5 cm DBH		0.0007			
Trees 31.5 - 62.5 cm DBH		0.0215			
Trees > 63 cm DBH		0.0384			
Lianas 8 - 15 cm diameter		0.0017	0.001		

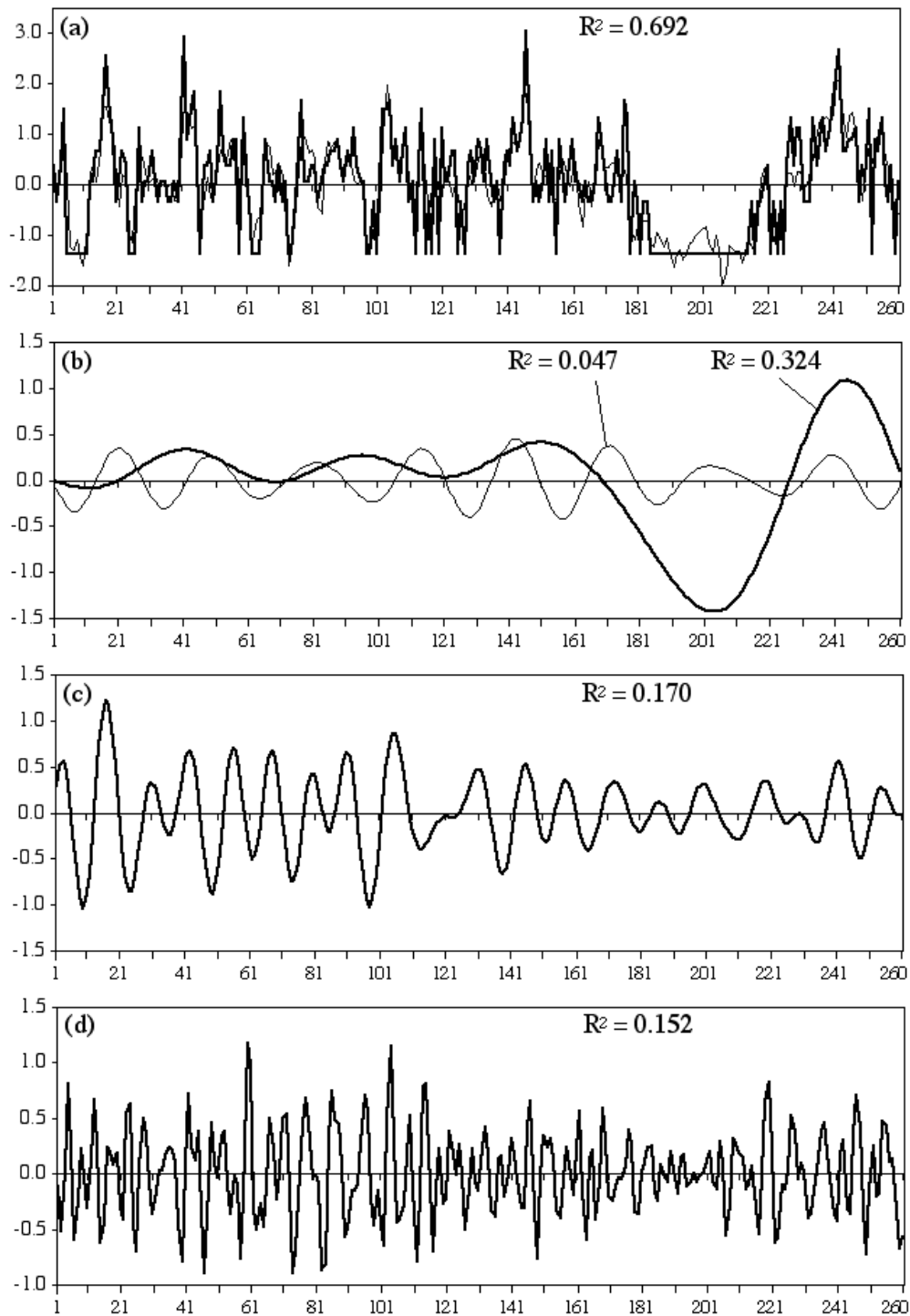


FIG. B1. Square-root-transformed *Adiantum tomentosum* abundances and spatial PCNM model and submodels (ordinate) along the Huanta transect (abscissa, 260 quadrats). (a) *A. tomentosum* (thick line) and parsimonious PCNM model. (b) very broad (thick line) and broad scale models. (c) medium scale model. (d) Fine scale model. The R^2 are calculated with respect to the total variation in the transformed fern count data; they are thus additive.

APPENDIX C

Ecological Archives E085-050-A3

A TABLE (TABLE C1) AND THREE FIGURES (FIG. C1, C2 AND C3) SHOWING RESULTS OF THE ANALYSIS OF THE ZOOPLANKTON DATA FROM GUADELOUPE.

TABLE C1. Variance of the two size classes of zooplankton and its partitioning among various fractions. Percentages are given with respect to the total variance of the data and to the variance of the detrended data. The PCNM variables are not strictly orthogonal in this case, so that the amounts of variance explained by each PCNM do not add exactly to the total obtained by a multiple regression. PCNM up to 9 are arbitrarily considered as broad scale, the other ones as medium scale. Detr. = detrended.

	Variance					
	190–600 μm	% (total)	% (detr.)	> 600 μm	% (total)	% (detr.)
Total	0.4178			0.4641		
Trend model	0.2097	50.2		0.0492	10.6	
Detrended	0.2081			0.4149		
PCNM model	0.1206	28.9	57.9	0.1407	30.3	33.9
PCNM 4				0.0463	10.0	11.2
PCNM 6	0.0277	6.6	13.3	0.0552	11.9	13.3
PCNM 7	0.0046	1.1	2.2			
PCNM 8	0.0091	2.2	4.4			
PCNM 9	0.0038	0.9	1.8	0.0286	6.2	6.9
PCNM 12	0.0081	1.9	3.9			
PCNM 14	0.0075	1.8	3.6			
PCNM 15				0.0145	3.1	3.5
PCNM 16	0.0178	4.3	8.5			
PCNM 18	0.0090	2.2	4.3			
PCNM 23	0.0373	8.9	17.9			
PCNM 24	0.0086	2.1	4.1			

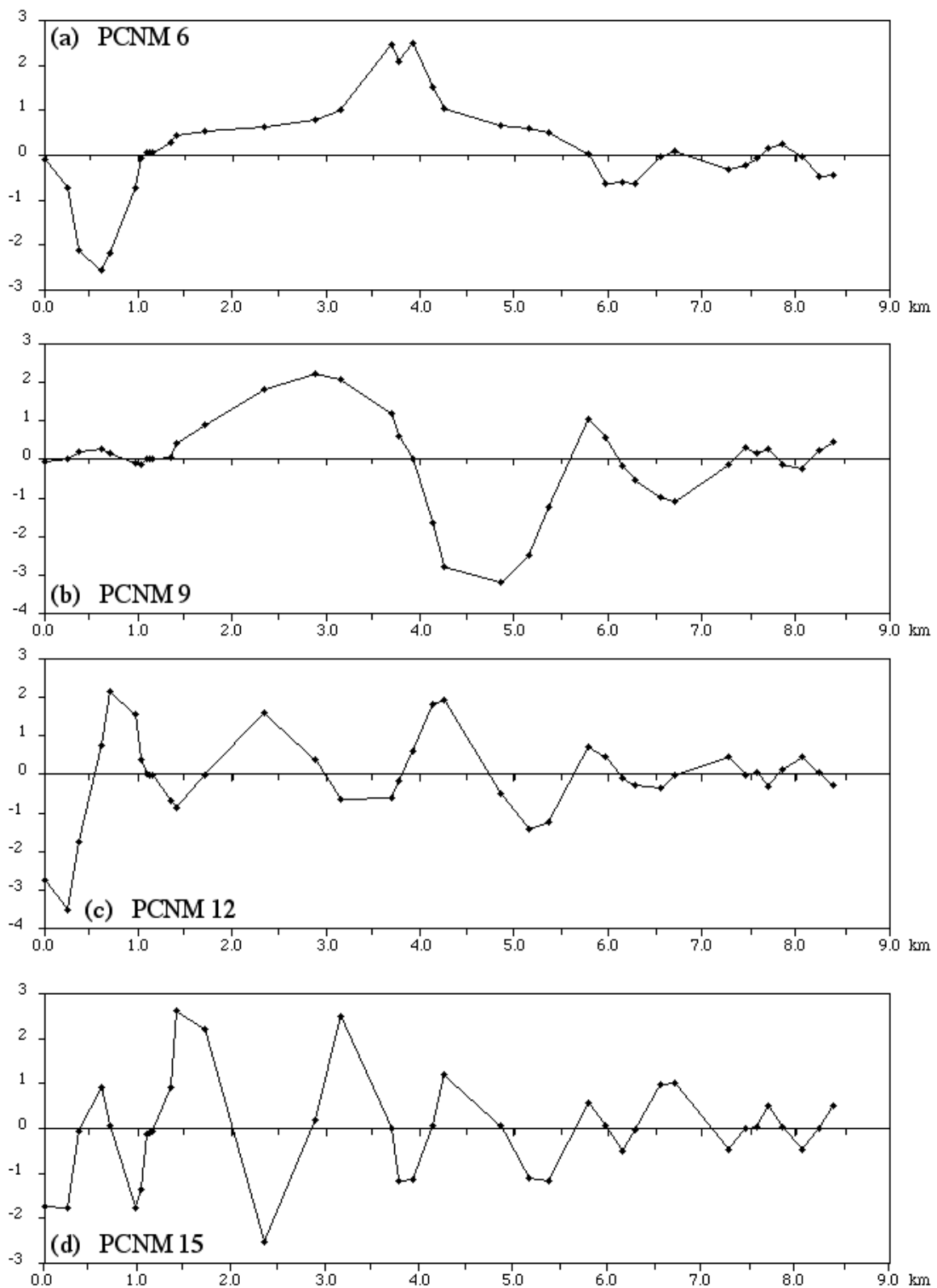


FIG. C1. Four examples of PCNM variables characterizing the Guadeloupe transect. Abscissa: distance from the coast.

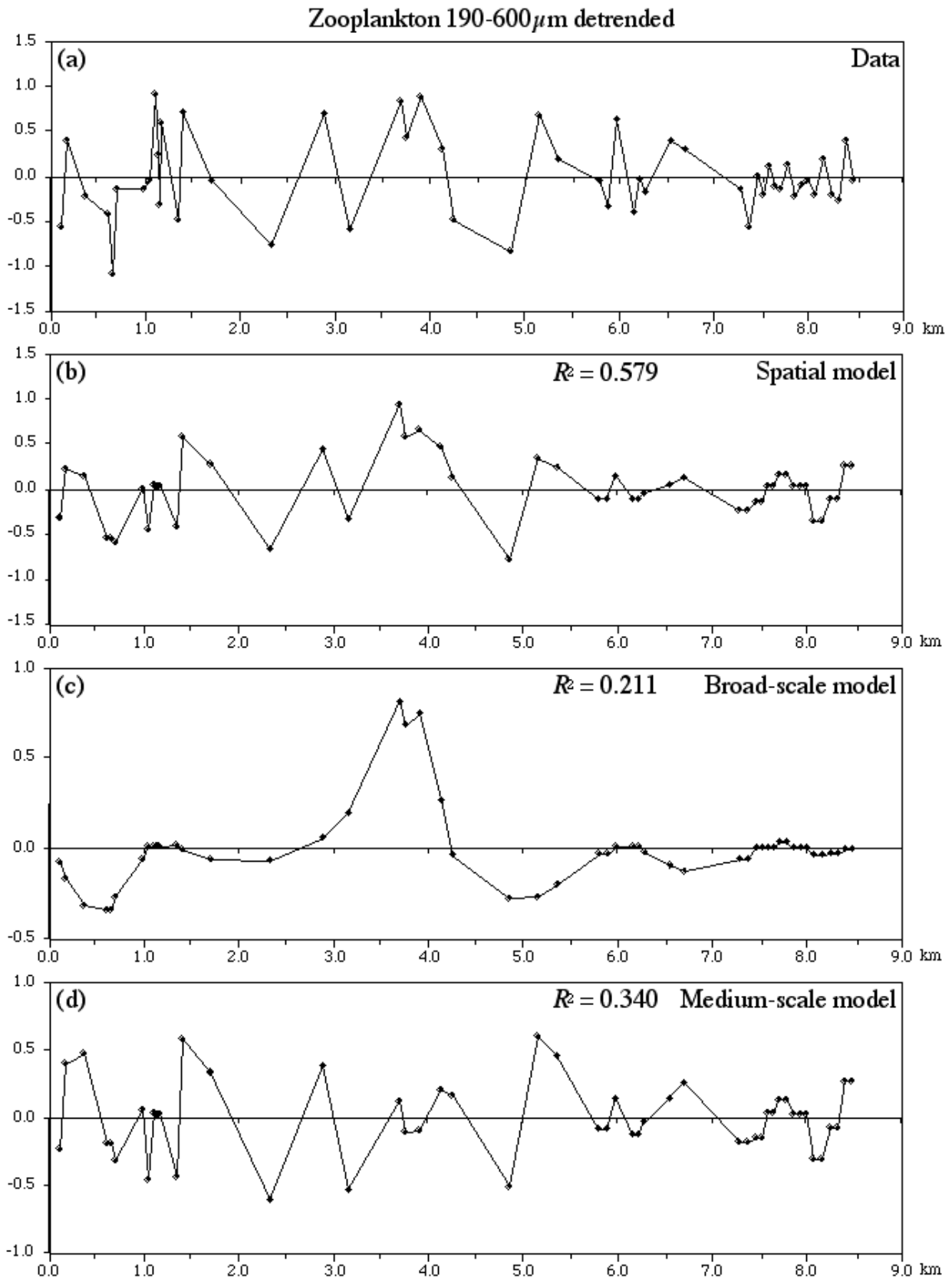


FIG. C2. Small zooplankton biomass data (detrended) and three spatial submodels. Abscissa: distance from the coast.

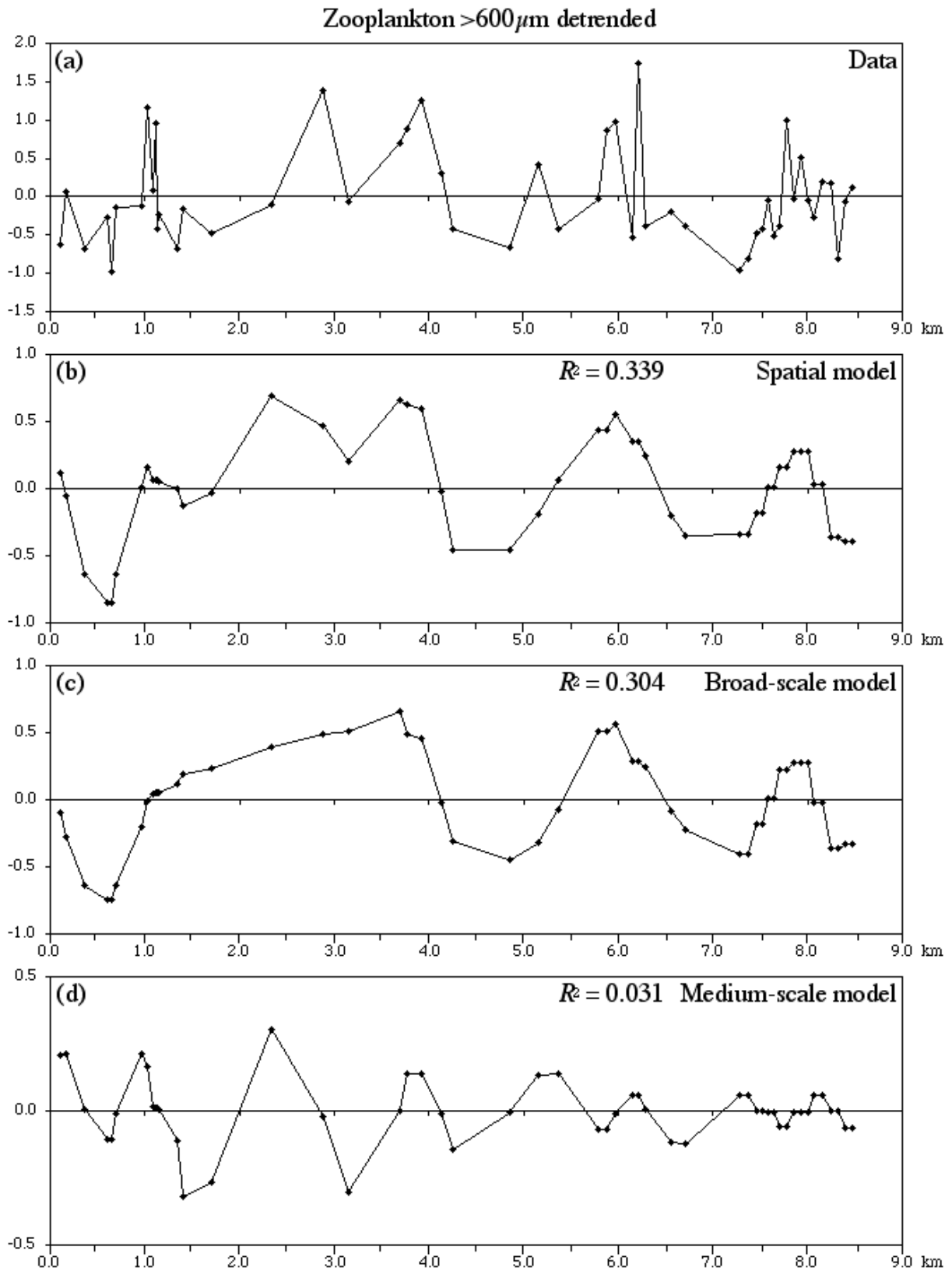


FIG. C3. Large zooplankton biomass data (detrended) and three spatial submodels. Abscissa: distance from the coast.

APPENDIX D

Ecological Archives E085-050-A4

A FIGURE (FIG. D1) SHOWING RESULTS OF THE ANALYSIS OF THE CHLOROPHYLL *a* DATA FROM THE THAU (FRANCE) LAGOON.

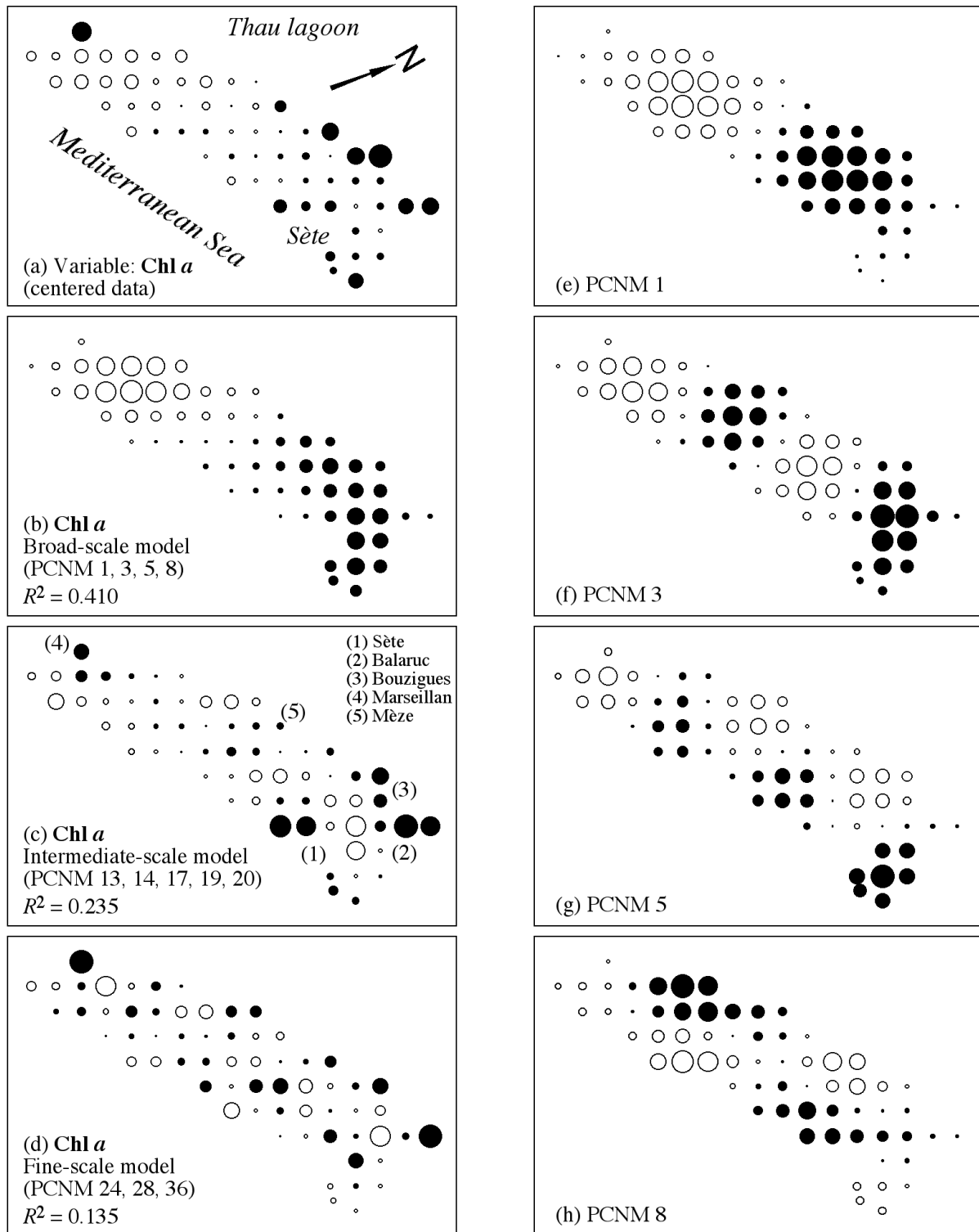


FIG. D1. PCNM analysis of Chl *a* at 63 sampling sites in the Thau marine lagoon. (a) Map of the values of Chl *a* (centered on 0). (b) Broad-scale spatial model of Chl *a*. (c) Intermediate-scale model. (d) Fine-scale model. (b-d) The values forecasted by each model were centered on 0 before mapping. (e-h) Maps of PCNM variables 1, 3, 5, and 8 which, together, form the broad-scale model represented in panel b. Filled bubbles: positive values; empty bubbles: negative values.

APPENDIX E

Ecological Archives E085-050-A5

FIVE FIGURES (FIG. E1 TO E5) SHOWING RESULTS OF THE ANALYSIS OF THE ORIBATID MITE DATA FROM ST-HIPPOLYTE (QUÉBEC, CANADA).

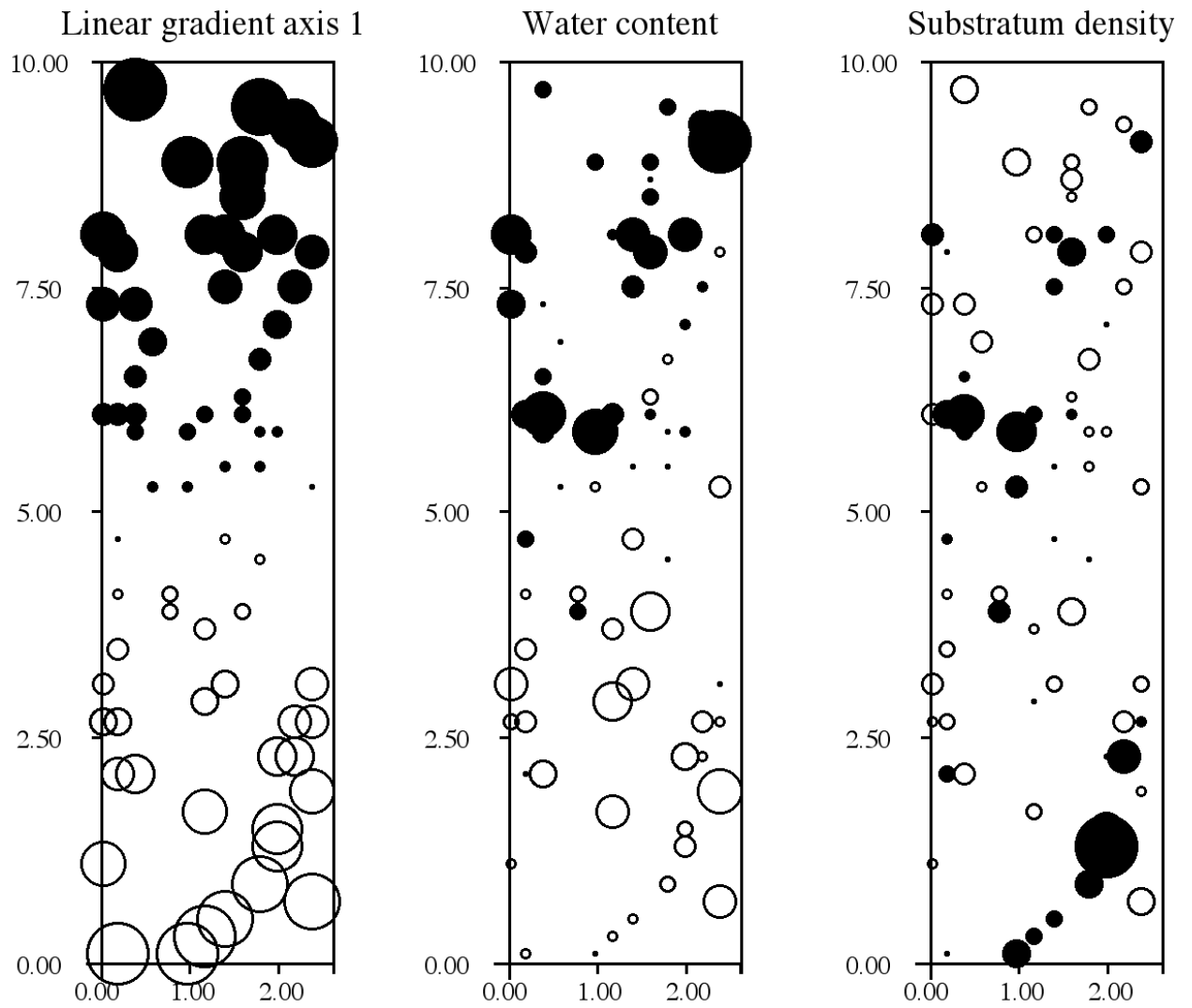


FIG. E1. First canonical axis of an RDA of the Hellinger-transformed Oribatid mite species data explained by the X and Y coordinates of the cores, and two environmental variables explaining this axis significantly.

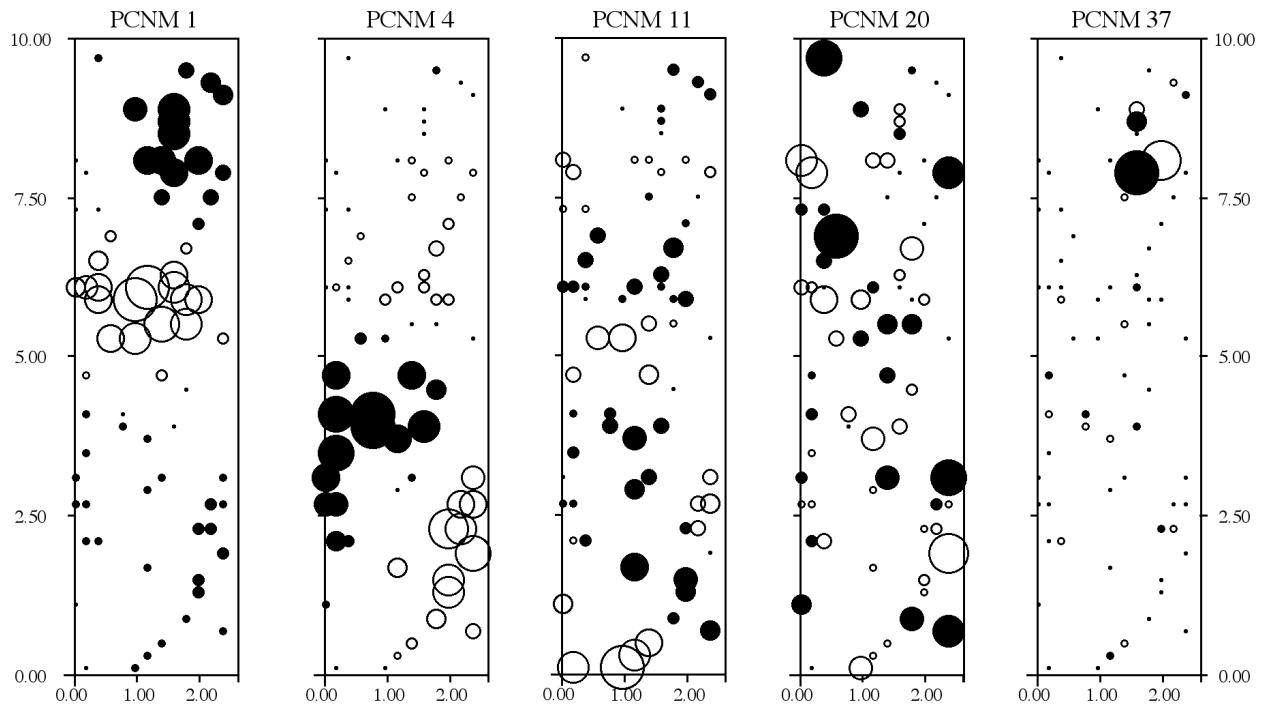


FIG. E2. PCNM 1, 4, 11, 20 and 37 obtained from the irregular sampling design used in the St-Hippolyte Sphagnum moss mat.

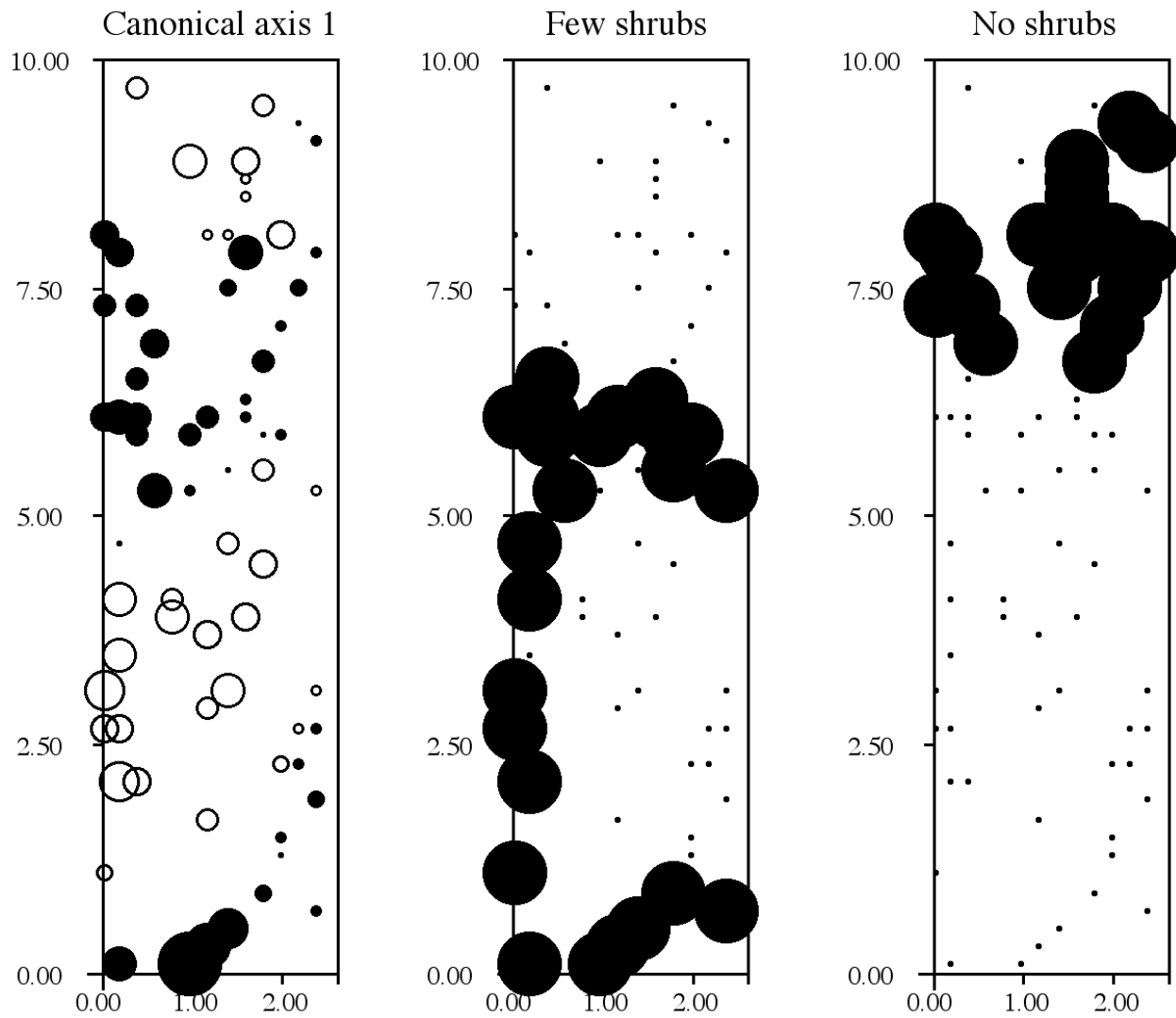


FIG. E3. First canonical axis of a partial RDA of the Hellinger-transformed Oribatid mite species data explained by 12 PCNM variables, with the X and Y coordinates of the cores as covariables, and two environmental variables explaining this axis significantly.

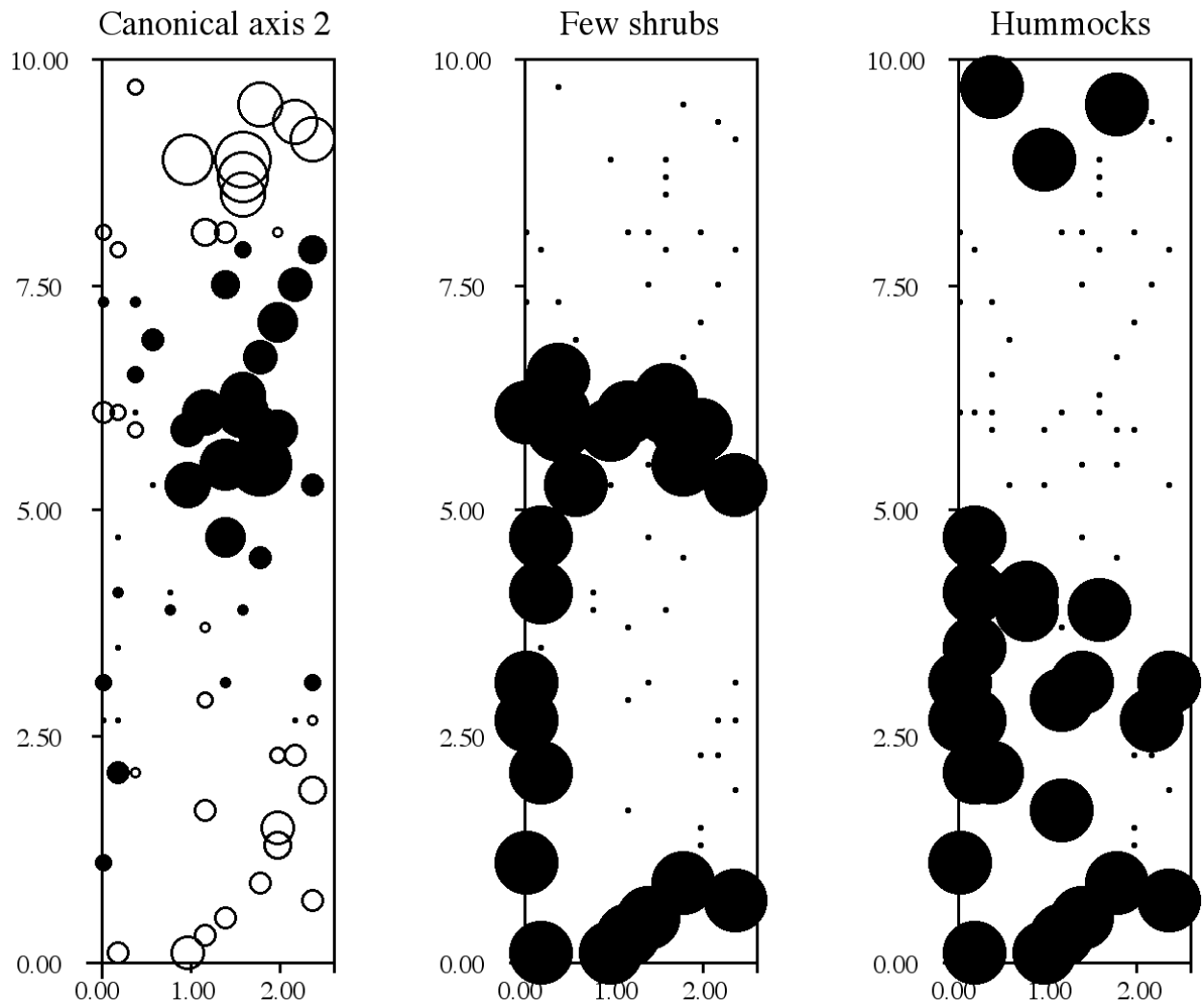


FIG. E4. Second canonical axis of a partial RDA of the Hellinger-transformed Oribatid mite species data explained by 12 PCNM variables, with the X and Y coordinates of the cores as covariables, and two environmental variables explaining this axis significantly.

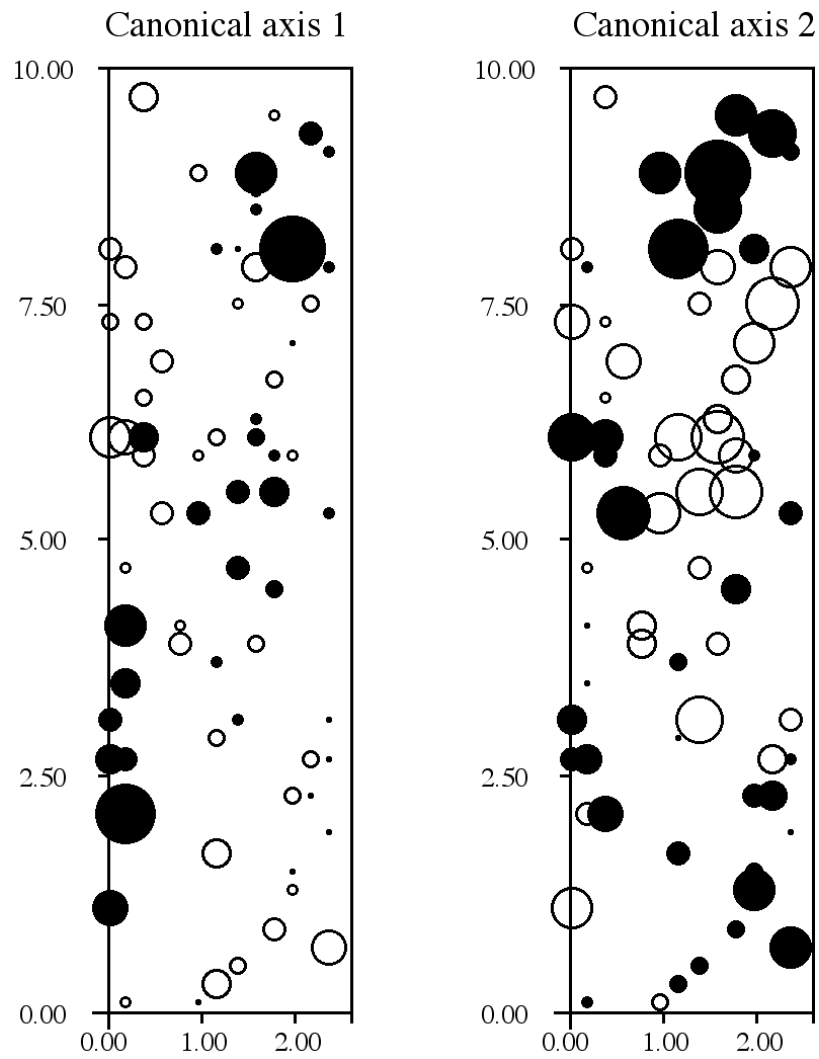


FIG. E5. First and second canonical axis of a partial RDA of the Hellinger-transformed Oribatid mite species data explained by 12 PCNM variables, with the 14 environmental variables as well as the X and Y coordinates of the cores as covariables.