



Partialling out the Spatial Component of Ecological Variation

Daniel Borcard; Pierre Legendre; Pierre Drapeau

Ecology, Volume 73, Issue 3 (Jun., 1992), 1045-1055.

Stable URL:

<http://links.jstor.org/sici?sici=0012-9658%28199206%2973%3A3%3C1045%3APOTSCO%3E2.0.CO%3B2-O>

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

Ecology is published by The Ecological Society of America. Please contact the publisher for further permissions regarding the use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/esa.html>.

Ecology

©1992 The Ecological Society of America

JSTOR and the JSTOR logo are trademarks of JSTOR, and are Registered in the U.S. Patent and Trademark Office. For more information on JSTOR contact jstor-info@umich.edu.

©2003 JSTOR

PARTIALLING OUT THE SPATIAL COMPONENT OF ECOLOGICAL VARIATION¹

DANIEL BORCARD

Institut de Zoologie, Université de Neuchâtel, Chantemerle 22, CH-2007 Neuchâtel, Switzerland

PIERRE LEGENDRE AND PIERRE DRAPEAU

*Département de Sciences biologiques, Université de Montréal, C.P. 6128, succursale A,
Montréal, Québec, Canada H3C 3J7*

Abstract. A method is proposed to partition the variation of species abundance data into independent components: pure spatial, pure environmental, spatial component of environmental influence, and undetermined. The new method uses pre-existing techniques and computer programs of canonical ordination. The intrinsic spatial component of community structure is partialled out of the species–environment relationship in order to see if the environmental control model still holds. The method is illustrated using oribatid mites in a peat blanket, forest vegetation data, and aquatic heterotrophic bacteria. In this latter example, the new method is shown to be complementary to another approach based on partial Mantel tests.

Key words: *canonical correspondence analysis; community structure; constrained ordination; environmental control model; forest vegetation; heterotrophic bacteria; modeling ecological relationships; oribatid mites; partial canonical ordination; redundancy analysis; spatial analysis; variation partitioning.*

INTRODUCTION

The spatial heterogeneity that may be observed in plant or animal communities generally has a multiplicity of origins. Two now classical models are found in the literature (May 1984): the environmental control model, where environmental variables are deemed responsible for the observed variations in the presence or abundance of the organisms (Whittaker 1956, Bray and Curtis 1957); and the biotic control model, where the links among organisms, horizontal (competition: Connell 1983, Schoener 1983) or vertical (predation, etc.: Mech 1981, Langeland 1982, Reinertsen et al. 1986), are considered to be the primary factors structuring communities (Southwood 1987). These models have often been viewed as competing or mutually exclusive hypotheses (see also May 1984). Quinn and Dunham (1983) suggested that it may be misleading to model the observed variations in patterns and processes of natural communities in terms of single causes. They propose a different approach to the understanding of biological phenomena by viewing alternative causes in the non-mutually-exclusive sense, that is, in terms of the relative contribution of each alternative.

The spatial structuring of natural communities clearly poses the problem of the relative contribution of different factors whose interaction often results in an overlaid effect in space. These factors include historical events such as fires in terrestrial ecosystems, disturbances of various kinds, and so on; contagious processes related to the mode of reproduction; circulation in fluid ecosystems; or complex behavioral interactions

among members of the community itself, such as predation and competition (Finegan 1984, Boerner 1985, Hughes and Fahey 1988, Legendre and Troussellier 1988, Legendre and Fortin 1989, Hudon et al. 1990, Leduc et al. 1992). In the present paper, we will call “environmental variables” all the independent variables of the model, biotic or abiotic, that are hypothesized to determine the variation of the dependent taxonomic group (species or assemblage) under study. In what follows the measures of variation are sums of squares, expressed as portions of the total sum of squares, as in nonadjusted coefficients of determination.

During the past decade, ecologists have become interested in measuring the spatial structure of biotic or abiotic environmental variables, either a priori to statistical testing, or to describe the shape of their spatial pattern. When studying the causes of spatial variation, the spatial structure of ecological data may act as a synthetic variable for the underlying processes that have generated it. Methods have been proposed recently, based on Mantel (1967) and partial Mantel testing (Smouse et al. 1986), for integrating space as a predictive variable into ecological models (Legendre and Troussellier 1988, Legendre and Fortin 1989, Leduc et al. 1992). However, the Mantel tests used in these papers only seek a statistically significant relationship between two or more similarity or distance matrices, and cannot be used to measure the fraction of the variation in the species matrix explained either by the environmental variables alone or the spatial structure of the species data alone, or to give any clue as to how much variation can be related simultaneously to both sets of explanatory variables.

¹ Manuscript received 23 July 1990; revised 21 February 1991; accepted 25 February 1991.

In this paper, we propose a quantitative approach to this problem, based on constrained and partial canonical ordination techniques: canonical correspondence analysis (ter Braak 1986, 1988a) and redundancy analysis (van den Wollenberg 1977).

THEORY

Constrained and partial ordination

With the increasing use of computers in the 1970s, multivariate methods like clustering and ordination have become very popular to help interpret ecological data. Among ordinations, two closely related eigenvector methods are broadly used: principal component analysis (PCA), which preserves the Euclidean distance among points, is used when one expects the responses of the species to the conditions of their habitat to be linear (as in short segments of ecological gradients; see Jongman et al. 1987); and correspondence analysis (CA) (Benzecri et al. 1973, among others), also called reciprocal averaging (Hill 1973), which preserves the so-called chi-square distance among points, and is appropriate when the responses of the dependent variables are expected to be unimodal along environmental gradients. Both methods, like others of the same kind, are intended to display the main trends of variation of a multidimensional data set in a reduced space of a few, linearly independent dimensions. Interpretation of the emerging features is done a posteriori through correlation of the main axes with another matrix of explanatory variables. This two-step analysis is called *indirect gradient analysis* by ter Braak (1986). For more detailed discussions of these methods and their ecological applications, we refer to the textbooks of Benzecri et al. (1973), Gauch (1982), Jongman et al. (1987), Legendre and Legendre (1983), Orłóci (1978), and Pielou (1984), or to the excellent review paper of Gower (1987).

In recent years, some authors have emphasized the interest of the direct counterpart of these procedures for ecological investigation, namely *direct gradient analysis*. In this case, instead of correlating the explanatory variables a posteriori with the main ordination axes of the analyzed data set, one constrains the ordination axes themselves to be linear combinations of the supplied set of explanatory variables. The methods of *constrained ordination* corresponding to PCA and CA are *redundancy analysis* (RDA) (van den Wollenberg 1977) in the linear context, and *canonical correspondence analysis* (CCA) (ter Braak 1986) in the unimodal context. A further and very interesting development of these procedures by ter Braak (1988a) allows *partial constrained ordinations*, where the above computations are made after removing, by multiple linear regression, the effects of known or undesirable variables, called covariables. Integration of these techniques into the tool box of ecological analyses, by means of the CANOCO program of ter Braak (1988b), pro-

vides a major improvement towards detailed causal interpretation of natural phenomena.

The method proposed in the present paper to partition the variation of species assemblages is based upon a few general measurements of the variation of the analyzed data sets, involving the eigenvalues of different constrained and partial analyses, that can be obtained from existing computer programs. The new method will be tested using data pertaining to different kingdoms and ecosystems: oribatid mites living in peat moss at the interface between a terrestrial and an aquatic ecosystem, tree assemblages in a forest community, and heterotrophic bacteria in a brackish lagoon.

"Environmental" and "spatial" variation

In the process of explaining the variation of a species abundance matrix, one usually relates the data to a set of environmental variables consisting, for instance, of physico-chemical, climatological, or geomorphological descriptors, which are often considered to be the most important determinants of plant or animal species assemblage patterns. Apart from the detailed analytical results of the constrained ordination, the use of RDA or CCA allows one to measure the amount of variation (sum of canonical eigenvalues) in the species data that can be explained by the set of environmental variables (see examples below). In situations where environmental gradients determine most of the variation in the living community, this amount of explained variation is fairly high. Legendre and Fortin (1989) have discussed the major role of spatial heterogeneity in ecological theories, as well as its importance in the choice of statistical tests; it follows that it is always important to get a measure of the amount of spatial structuring in species abundance data. Thus, a further step in the analysis should be to relate the species data set to the spatial coordinates of the samples, as suggested by ter Braak (1987a) and demonstrated by Legendre (1990), who presents a complete example. As above, one gets a measure of the amount of variation in the species data that can be explained by the supplied "spatial matrix."

Partitioning the variation

By making two canonical ordinations, each of them constrained by one of the sets of explanatory variables, one gets a measure of the total importance for the species data of (1) the effects of the environmental conditions and (2) the spatial structure. But, as suggested in the Introduction, some species and environmental variables may share a common spatial structuring. This may be due to the effect of spatially structured environmental descriptors on the dependent biotic variable(s), or to some spurious effect of an extraneous variable, not included in the model, that causes a common spatial structure to show up in both the independent and the dependent variables of the model (see *Discussion*). Thus, in the above analyses, the

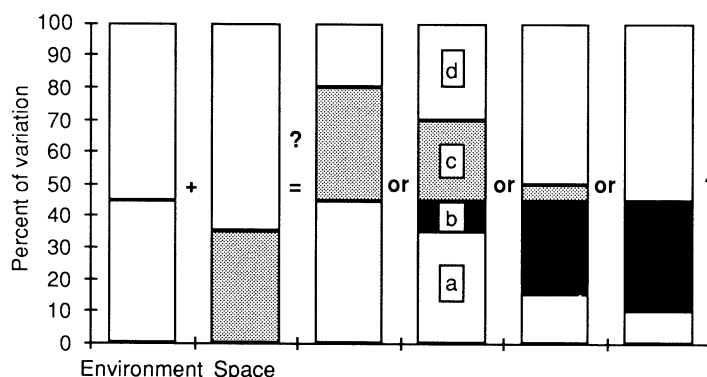


FIG. 1. Percentages of variation of a species data matrix explained by environment and by space (two left-hand columns). Separate analyses do not allow discrimination between the four situations represented at the right of the equals sign. a = environment, b = environment + space, c = space, d = undetermined.

amount of variation in the species data that is due to this common spatial structuring has been extracted by both the environmental and the spatial sets of explanatory variables. In terms of amount of explained variation, these analyses are therefore partially redundant (Fig. 1). To get a more complete picture of the situation, one should be able to partition the total variation of the species data as follows (Fig. 2):

a) The nonspatial environmental variation in the species data, which is the fraction of the species variation that can be explained by the environmental descriptors independently of any spatial structure;

b) The spatial structuring in the species data that is shared by the environmental data. This common variation is partly a consequence of the relations of the species with spatially structured environmental conditions, but a certain amount of it could be noncausal, i.e., due to separate relations of both sets of variables with some external space-structuring processes (unidentified at this stage);

c) The spatial patterns in the species data that are not shared by the environmental data. In general terms, these patterns may reflect some contagious biological process like growth, reproduction, or predation, without environmental component (or, more precisely, without relation with the variables that were actually included in the analysis);

d) The fraction of the species variation explained neither by spatial coordinates nor by environmental data.

In the univariate, nonspatial case, this type of decomposition has been described by Whittaker (1984).

Knowing the relative weight of these items can be of decisive importance when applying causal hypotheses to one's data in the framework of some precise ecological theory. We propose to achieve this partitioning by partial canonical ordination (RDA or CCA) of the species and spatial matrices while controlling for the effect of the environmental descriptors, which provides the value of item (c) in Figs. 1 and 2, and by a

partial canonical ordination of the species and environmental data sets, controlling for space, which gives the value of item (a).

The fractions of variation explained in the different analyses are actually estimated as follows:

In a linear context: Assume that the variation in the species data equals 1 (this is the case in the RDA output of the CANOCO program; this can always be achieved by multiplying all species data by an appropriate constant). The fraction of variation accounted for by the canonical (constrained) axes is then simply given by the sum of their eigenvalues. In partial RDA, the effect of the covariables is removed before performing the canonical analysis. Thus, the sum of all eigenvalues drops below the value obtained in the previous analysis (RDA), which had been scaled to equal 1, and gives a measure of the remaining variation of the species data. After the analysis, the fraction of the variation accounted for by the explanatory variables is obtained by summing the canonical eigenvalues.

In a unimodal context: When CCA and partial CCA are in use, the canonical eigenvalues are still measures of the amount of variation accounted for by the explanatory variables. To transform these measures into percentages of the total variation of the species data, one has to divide the individual eigenvalues (in a detailed analysis) or the sum of all *canonical* eigenvalues (in our problem) by the total inertia, or trace, or sum of all eigenvalues of a correspondence analysis of the species matrix.

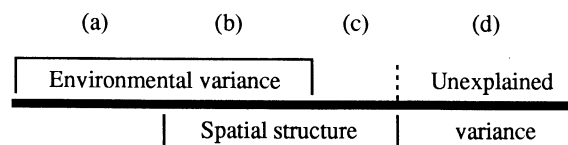


FIG. 2. Variation partitioning of a species data table, showing that fraction (b) is the intersection of the environmental and spatial components of the species variation.

As a support to our discussion, we propose three real examples covering different ecological situations. The whole process is based on computations made with the CANOCO program of ter Braak, release 3.11, 1990 (ter Braak 1988b).

TESTS ON REAL DATA

Test 1: Oribatid mites (Acari, Oribatei) of the Sphagnum mosses of Lac Geai, Station de Biologie de l'Université de Montréal, Saint-Hippolyte, Québec

In June 1989, 70 cores of *Sphagnum* moss (5 cm in diameter and 7 cm deep) were extracted from a 10 × 2.5 m sampling area in a floating moss and peat blanket extending from the forest border into the lake; the blanket is 10 m broad on the average. Sampling was stratified by types of substrate, with the number of cores in each subsample proportional to the surface area of its stratum.

One of the aims of the sampling was to study the spatial distribution of the oribatid community in the mosses, and to verify whether this structure could be related to that of environmental variables. Within the sampling area, mosses have not been disturbed during recent years, so we considered their spatial patterns to be natural.

The environmental variables were: substratum (seven unordered qualitative classes: four species of *Sphagnum*, ligneous litter, bare peat, interface between *Sphagnum* species), morphology of the substratum (two qualitative classes: blanket or hummock), coverage density of the shrub cover (three semi-quantitative classes), water content in percent (quantitative), density of the substratum in grams per litre of dry uncompressed matter (quantitative). Thus, apart from the last two, the descriptors of the life conditions of the oribatid community are not abiotic sensu stricto, but rather refer to some structural aspects of other living organisms, which are likely to show spatial patterns in response to their own environmental constraints.

The matrix of two-dimensional geographical coordinates, x and y , has been completed as suggested by Legendre (1990), by adding all terms for a cubic trend surface regression of the form

$$\hat{z} = b_1x + b_2y + b_3x^2 + b_4xy + b_5y^2 + b_6x^3 + b_7x^2y + b_8xy^2 + b_9y^3. \quad (1)$$

This ensures not only the linear gradient patterns in the species data to be extracted, but also more complex features like patches or gaps, which require the quadratic and cubic terms of the coordinates and their interactions to be correctly described. There is no need for an intercept term (b_0) since the species data are centered on the origin in the correspondence analysis solution. In order to avoid artificial increase of the explained variation by mere chance, the nine terms of the equation have been submitted to the CANOCO

procedure of "forward selection of explanatory variables," a multivariate extension of the stepwise regression method. The following terms of the equation were retained in this example:

$$\hat{z} = b_1x + b_2y + b_4xy + b_5y^2 + b_9y^3. \quad (2)$$

Results and discussion

The detailed results of the study will be reported in other papers. We will focus here only on the application of the new method of variation partitioning.

From the 9850 individuals and 49 taxa captured, 14 species, involving only 50 individuals, were eliminated; their very poor representation introduced a large number of zeroes in the data matrix. The 35 remaining species (9800 individuals) were passed through the steps of analysis described above. The species abundances, showing a large range of variation (from 0 to 723) were first transformed by taking Napierian logarithms [$y' = \ln(y + 1)$], following Berthet and Gérard (1965) and ter Braak (1987b). Classes of the qualitative environmental variables were transformed into dummy binary variables, as recommended by ter Braak (1987a). The same holds for the semi-quantitative variable "shrub cover"; it could have been treated as a quantitative variable, but there was not enough precision in the cover estimations to do so.

As living organisms often show unimodal responses to the environmental gradients of their biotopes, the analyses were made using CCA. The four analyses gave the following global results:

1) CCA of the species matrix, constrained by the environmental matrix: sum of all canonical eigenvalues = 0.521.

2) CCA of the species matrix, constrained by the extended matrix of geographical coordinates: sum of all canonical eigenvalues = 0.503.

3) like (1), after removing the effect of the geographical matrix: sum of all canonical eigenvalues = 0.160.

4) like (2), after removing the effect of the environmental variables: sum of all canonical eigenvalues = 0.142.

The sum of all eigenvalues in a correspondence analysis of the species matrix is 1.164. Thus the percentage of the total variation of the species matrix accounted for by each step of the analysis is obtained as follows:

$$\begin{aligned} \text{step (1): } & 0.521 \cdot 100 / 1.164 = 44.8\% \\ \text{step (2): } & 0.503 \cdot 100 / 1.164 = 43.2\% \\ \text{step (3): } & 0.160 \cdot 100 / 1.164 = 13.7\% \\ \text{step (4): } & 0.142 \cdot 100 / 1.164 = 12.2\% \end{aligned}$$

The overall amount of explained variation (in percentage of the total variation of the species matrix) is obtained either by summing the results of steps (1) and (4), or those of (2) and (3): $44.8 + 12.2 \approx 43.2 + 13.7 \approx 57.0\%$. The slight difference is due to successive approximations during the calculations.

The whole variation of the species matrix can be

partitioned as follows (Fig. 3): (a) nonspatial environmental variation [step (3)]: 13.7%; (b) spatially structured environmental variation [step (1) – step (3), or step (2) – step (4)]: 31.0%; (c) spatial species variation that is not shared by the environmental variables [step (4)]: 12.2%; (d) unexplained variation and stochastic fluctuations: $100 - 57.0 = 43.0\%$. Note that in theory, value (c) can be negative (Whittaker 1984). In ecological practice, however, this is unlikely to occur.

Fig. 3 illustrates the relative importance of the various processes that control the variation of the oribatid mites.

The environmental variables explain $\approx 45\%$ of the variation in the species matrix (step 1). Roughly two-thirds of this amount (part b in Fig. 3) can also be predicted by the supplied function of the geographical coordinates of the samples. This means that the species and the environmental data have a fairly similar spatial structuring, resulting partly from the same response to some common underlying causes (as, for instance, the humidity gradient, that acts on the oribatid community as well as on the distribution of the *Sphagnum* mosses), and partly from the direct response of the oribatid community to some spatially structured features of their substratum. Part a involves local effects of the environmental variables on the oribatid community, with no spatial pattern that could be assessed by means of our cubic trend surface function of geographic coordinates.

About one-fourth of the explained variation (part c) is assessable by the spatial matrix, but cannot be related to the measured environmental variables. Thus the spatial matrix acts partly as a synthetic descriptor of unmeasured underlying processes: external causes or biotic factors like, for instance, social aggregation. In this specific case, the small amount of total variation accounted for by space alone (12.2%) shows that no fundamental spatial-structuring process has been missed. The structure assessed as part c is linked to processes that have probably also generated (as local effects) a part of the totally unexplained variation of the species matrix (part d).

The amount of unexplained variation (d) is fairly high (43.0%), even assuming that part of it is due to nondeterministic fluctuations. Although the underlying processes cannot be identified from the available data, the analyses give some information about them: they are (at least partly) independent of the measured environmental variables (which do not pretend to be exhaustive), and their action on the oribatid community structure cannot be totally predicted by the supplied function of the spatial coordinates. In other words, a fair amount of variation is due to local effects of unmeasured (biotic or abiotic) controlling variables, or to spatial structures that have been missed because they require more complex functions to be described. For instance, the distribution of some classes of nutrients such as ligneous matter, fungi, algae, or pollen, is very

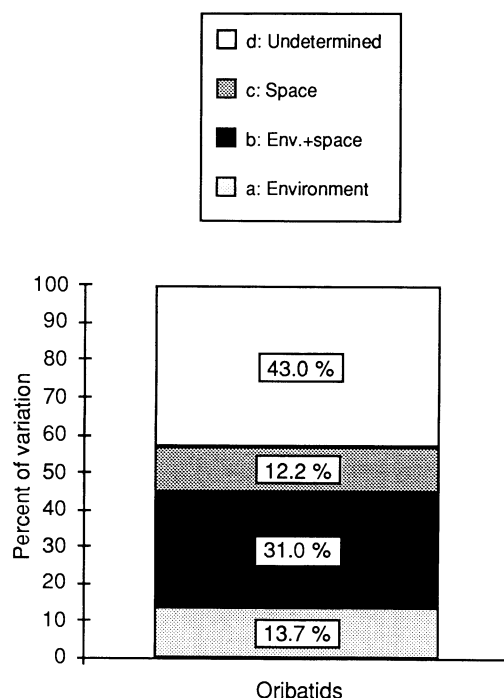


FIG. 3. Variation partitioning of the oribatid mites data matrix.

likely to determine some spatial structures in the mite community, but probably at a very local scale (within a few centimetres).

Test 2: Forest vegetation of the Municipalité régionale de Comté du Haut-Saint-Laurent, Québec

These data, collected during an ecological study of a terrestrial ecosystem in southwestern Québec (Bouchard et al. 1985, Bergeron et al. 1988, Brisson et al. 1988, Leduc et al. 1992), have already been used as test data for spatial analysis by Legendre and Fortin (1989). In an area of $\approx 0.5 \text{ km}^2$, 200 vegetation quadrats, each $10 \times 20 \text{ m}$ in size, were surveyed by means of a systematic sampling design. The available data comprise the abundances of trees (at the species level), six geomorphological variables, and the geographical locations of the samples.

Among the 28 species of trees that were present in the area, the 12 most abundant were retained for this example, and their abundances transformed into Napierian logarithms, as in the previous example.

The six environmental descriptors are: quality of drainage (7 semi-quantitative classes), stoniness of the soil (7 semi-quantitative classes), topography (11 unordered qualitative classes), directional exposure (the 8 sectors of the compass card, plus class 9 = flat land), texture of the horizon 1 (the upper mineral layer) of the soil (8 unordered qualitative classes), and geomorphology (6 unordered qualitative classes). These data

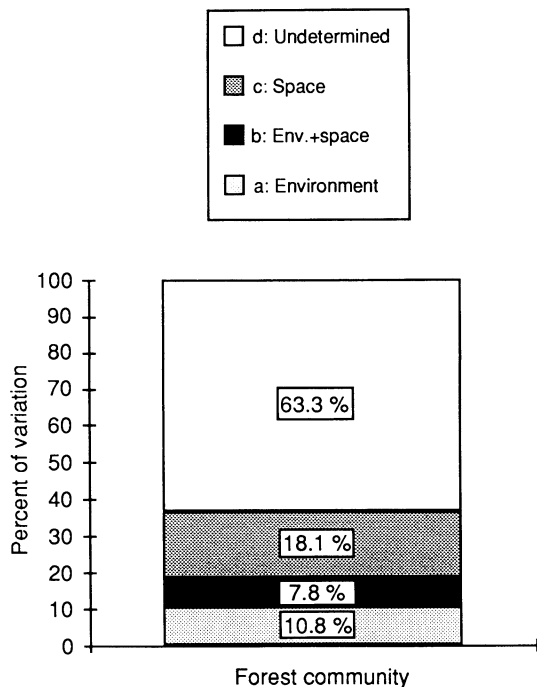


FIG. 4. Variation partitioning of the forest community data matrix.

were used to compute an Estabrook-Rogers similarity index among quadrats (Estabrook and Rogers 1966, Legendre and Legendre 1983). The "environmental matrix" that was finally used in this example was built by taking the 6 principal coordinates of the association matrix based upon this index that were associated with positive eigenvalues. This procedure preserved the whole variation of the environmental data; the new quantitative eigenvariables (principal axes) provide a more compact matrix than the original one, where every qualitative variable would have had to be transformed into a set of dummy binary variables before being used in this study.

The matrix of x - y geographical coordinates of the quadrats was completed in the same way as in the previous example, by adding the terms of a cubic trend surface regression (Eq. 1 above). After applying the same forward selection procedure as above, the following terms of the equation were retained:

$$\hat{z} = b_1x + b_2y + b_3x^2 + b_5y^2 + b_6x^3 + b_7x^2y + b_8xy^2. \quad (3)$$

Results and discussion

The four steps of analysis of our procedure were applied to these data, using CCA, and gave the following results:

1) CCA of the species matrix, constrained by the environmental matrix: sum of all canonical eigenvalues = 0.268.

2) CCA of the species matrix, constrained by the extended matrix of geographical coordinates: sum of all canonical eigenvalues = 0.373.

3) like (1), after removing the effect of the geographical matrix: sum of all canonical eigenvalues = 0.156.

4) like (2), after removing the effect of the environmental variables: sum of all canonical eigenvalues = 0.261.

The sum of all eigenvalues of a correspondence analysis of the species matrix equals 1.443. Thus the percentage of the total variation of the species matrix accounted for by each step of analysis is obtained as follows:

$$\text{step (1): } 0.268 \cdot 100 / 1.443 = 18.6\%$$

$$\text{step (2): } 0.373 \cdot 100 / 1.443 = 25.9\%$$

$$\text{step (3): } 0.156 \cdot 100 / 1.443 = 10.8\%$$

$$\text{step (4): } 0.261 \cdot 100 / 1.443 = 18.1\%.$$

Overall amount of explained variation:

$$18.6 + 18.1 = 25.9 + 10.8 = 36.7\%.$$

The total variation of the species data set can thus be partitioned as follows (Fig. 4): (a) non-spatial environmental variation: 10.8%; (b) spatially structured environmental variation: 7.8%; (c) spatial species variation that is not shared by the environmental variables: 18.1%; (d) unexplained variation and stochastic fluctuations: 63.3%.

The proportions of the three parts of explained variation are different from those of the first example. Here, a good half of the variation explained by the environmental variables is due to local effects, while over two-thirds of the variation explained by the spatial matrix is independent of the environmental descriptors. Leduc et al. (1992) have shown by means of Mantel tests that the spatial variation of the trees, when viewed within the framework of disturbance dynamics, can be indicative of different processes. They propose hypotheses about the mechanisms involved (regenerative strategies, competition for space). Here, when we consider the partitioning of the variation observed in this data set, we see that the major causes of the spatial patterns in the species data are not to be found in the environmental descriptors (compare fractions b and c).

One striking fact in this example is the large amount of unexplained variation: more than half of the total variation of the species matrix remains unexplained. Whether this is due to some overlooked factors or to a large amount of stochastic variation remains unclear. In any case, the available environmental variables globally play a significant role ($P = .001$), as can be demonstrated by a permutation test on the trace statistic of the analysis provided by CANOCO.

Test 3: Factors influencing the growth of two groups of bacteria in the Thau brackish lagoon, southern France

As a part of a multidisciplinary research program on the Thau lagoon located near Sète on the Mediterranean shore of France, two groups of aquatic bacteria have been sampled, and two environmental variables measured at 63 sampling sites in June 1986 (Amanieu et al. 1989). In our example, the "species matrix" contains two columns. The first one gives the number of colony-forming units of aerobic heterotrophs growing on bioMérieux nutrient agar (low NaCl concentration), called BNA hereafter, which are presumably of continental origin; the second column is the number of colonies growing on marine agar (3.4% salinity = 34 g NaCl per litre of agar), called MA, which are expected to be mostly of marine origin. The environmental descriptors are the concentrations of chlorophyll *a* ("CHL *a*," in micrograms per litre) and the particulate organic carbon ("POC," in milligrams per litre). The locations of the samples are known by their coordinates, *x* and *y*; these will be used as spatial descriptors to ensure full comparability with the results of Legendre and Troussellier (1988). These authors used a simple matrix of Euclidean distances to test for a spatial structure in their data. In the case of the BNA bacteria, this arbitrary choice results in practically no change in the results, since an application of the same forward selection procedure as in the other examples leads to the selection of *x* only, *y* having no significant explanation power. Insofar as the MA bacteria are concerned, one could get a slightly greater explanation of the variation by retaining the *x*² term, but even this larger fraction does not remain significant when the effect of the environmental factors is removed.

The scientific question to be solved is exposed in detail by Legendre and Troussellier (1988); one of the purposes of this study is to show that "some assumed relationships between bacteria and environmental variables can be spurious, implying a common spatial gradient, while others are real." To elucidate which relations are real and which are spurious, these authors proposed a way of modeling implying a series of partial Mantel tests, and they produced a synthetic model of causal relationships for each category of bacteria. These models are reproduced in Fig. 5.

Here we propose another way of looking at this problem, in showing how much of the variation in each bacterial variable can be explained by the environmental variables taken globally, when controlling, or not, for the effect of space.

In this example, dispersion diagrams between abundances of bacteria and the environmental factors show clearly that the statistical relationships are linear within the observation range. Hence a linear method of canonical ordination is appropriate, namely redundancy analysis (RDA), instead of the unimodal CCA method. Since we will analyze each bacterial variable separately, this example is nothing but an application of CANOCO for performing multiple and partial linear regression. The problem would become really multivariate if we

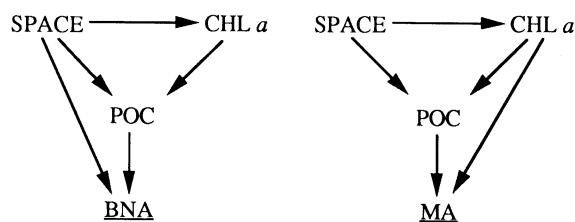


FIG. 5. Synthetic models of causal relations between space, CHL *a*, POC, and the BNA (left) and MA (right) heterotrophic bacteria based on partial Mantel tests. For explanations of the symbols, see *Tests on real data: Test 3* Significant links are indicated by arrows. After Legendre and Troussellier (1988).

had to explain instead the distribution of a whole community of aquatic bacteria. But considering this simple case in the context of multivariate quantitative modeling provides a good insight into the principles of the method. Notice that this approach can be generalized to situations involving unimodal relations between the data to be explained and the set of predictors, by means of CCA, which is not possible with multiple linear regression.

Aerobic heterotrophs growing on bioMérieux nutrient agar (BNA).—The four analyses give the following global results (*P*-values determined by Monte Carlo permutation tests):

- 1) RDA of the "BNA" submatrix, constrained by the environmental matrix: sum of all canonical eigenvalues = 0.451 (*P* = .001)
- 2) RDA of the "BNA" submatrix, constrained by the matrix of geographical coordinates: sum of all canonical eigenvalues = 0.595 (*P* = .001)
- 3) like (1), after removing the effect of the geographical matrix: sum of all canonical eigenvalues = 0.053 (*P* = .018)
- 4) like (2), after removing the effect of the environmental variables: sum of all canonical eigenvalues = 0.197 (*P* = .001).

We note that the Monte Carlo permutation tests on the trace statistics of the RDAs (1), (2), and (4) are significant at a Bonferroni-corrected level of 0.05/4 = 0.0125.

In the CANOCO version of RDA, these sums of eigenvalues can directly be read as fractions of explained variation. Thus, the total explained variation equals:

$$45.1 + 19.7 = 59.5 + 5.3 = 64.8\%$$

The covariation between the environmental variables and space is:

$$45.1 - 5.3 = 59.5 - 19.7 = 39.8\% \text{ of the total variation of the "BNA" variable.}$$

The tests of statistical significance of the sums of canonical eigenvalues, reported above, have been made

to provide a comparison with the results of Legendre and Troussellier (1988). The main feature of these results is the very small amount of common variation (0.053) between BNA and the environmental variables when holding space constant, as pointed out by these authors who discussed the same data in terms of partial Mantel correlations. BNA and the environmental variables are obviously driven by a common spatial gradient, a situation that generates a spurious, noncausal correlation between them (Legendre and Troussellier 1988).

Aerobic heterotrophs growing on marine nutrient agar (MA). — Results of the four analyses:

1) RDA of the “MA” submatrix, constrained by the environmental matrix: sum of all canonical eigenvalues = 0.311 ($P = .001$)

2) RDA of the “MA” submatrix, constrained by the matrix of geographical coordinates: sum of all canonical eigenvalues = 0.126 ($P = .015$)

3) like (1), after removing the effect of the geographical matrix: sum of all canonical eigenvalues = 0.188 ($P = .002$)

4) like (2), after removing the effects of the environmental variables: sum of all canonical eigenvalues = 0.003 ($P = .903$)

The Monte-Carlo permutation tests on the trace statistics of the RDAs (1) and (3) are significant at a Bonferroni-corrected level of $0.05/4 = 0.0125$.

Total explained variation:

$$31.1 + 0.3 = 12.6 + 18.8 = 31.4\%$$

Covariation between the environmental variables and space:

$$31.1 - 18.8 = 12.6 - 0.3 = 12.3\%$$

of the total variation of the “MA” submatrix.

Contrary to the “BNA” group, these bacteria show a significant relation to the measured environmental variables, with almost no residual spatial structure (0.3% of total variation) left after removing the effect of these descriptors.

These results, which agree entirely with those of Legendre and Troussellier, are represented in two different ways in Fig. 6. First we show the same diagrams of the partitioned variation as in the previous examples. Second, we propose, for each type of bacteria, a synthetic model based upon those of Legendre and Troussellier, but with an additional feature, the part of “unknown factors and stochastic variations”; the thickness of the arrows indicates our estimation of the proportion of the total variation of the “species” matrix explained by each set of variables. Such an illustrative model provides a quantitative insight into the relations between all the measured variables, as well as a visual estimate of the residual (unexplained) variation of the dependent variable (or, in a multivariate

case, the species data set). For instance, the “MA” model shows that, despite the fact that the relations between the bacteria and the environmental variables are statistically significant, as shown by Legendre and Troussellier, there must be some other important yet unmeasured factors influencing the bacteria. One can also see that these unstudied factors have no covariation with the supplied set of spatial coordinates.

Comparison between Figs. 5 and 6 shows the complementarity of the two approaches: the Mantel tests provide an assessment of the significant interactions and a construction of the general model, while the partial canonical ordination analyses, which also provide tests of significance, allow us to quantify the major sources of variation in the set of dependent variables.

GENERAL DISCUSSION

What we propose here is a new method for partitioning the variation of species assemblages, that allows one to measure the relative contribution of sets of explanatory variables by using eigenvalues of constrained and partial ordinations. Our method is conceptually linked to the idea that ecological phenomena are explained by non-mutually-exclusive abiotic and biotic processes that overlap in space and time (Quinn and Dunham 1983). Operationally, it is based on the use of pre-existing methods of canonical ordination (RDA: van den Wollenberg 1977; CCA: ter Braak 1986) and computer programs; ter Braak's CANOCO was used in our examples. An interesting feature of our method is that it allows one to partial out the spatial component of community data sets. Considering the increasing importance of space in ecological theory (Wiens 1989), and being aware that most ecological data are spatially autocorrelated, it becomes necessary to develop quantitative methods to help us to understand and model such situations in a way that allows a clear assessment of: (1) what can be directly interpreted in terms of interactions between the species (or, more generally, the dependent set of variables) and their environment; (2) which part of the variation can be predicted by means of a spatial matrix (but not necessarily interpreted in terms of underlying causes); (3) how much variation is not explainable or predictable at all by means of the known variables.

Let us comment briefly on these points. (1) As has been shown by Legendre and Troussellier (1988), the presence of a spatial structure shared by the species and the environmental data sets leads to an overestimation of the interactions between the species and the measured environmental conditions. Sometimes these interactions don't even exist (as in the case of the BNA bacteria), sometimes their importance is simply less than could be estimated by means of simple among-set correlations. Our method allows an accurate quantification of this fraction of variation, that could subsequently undergo a detailed analysis by classical modeling, by use of the supplied canonical coefficients

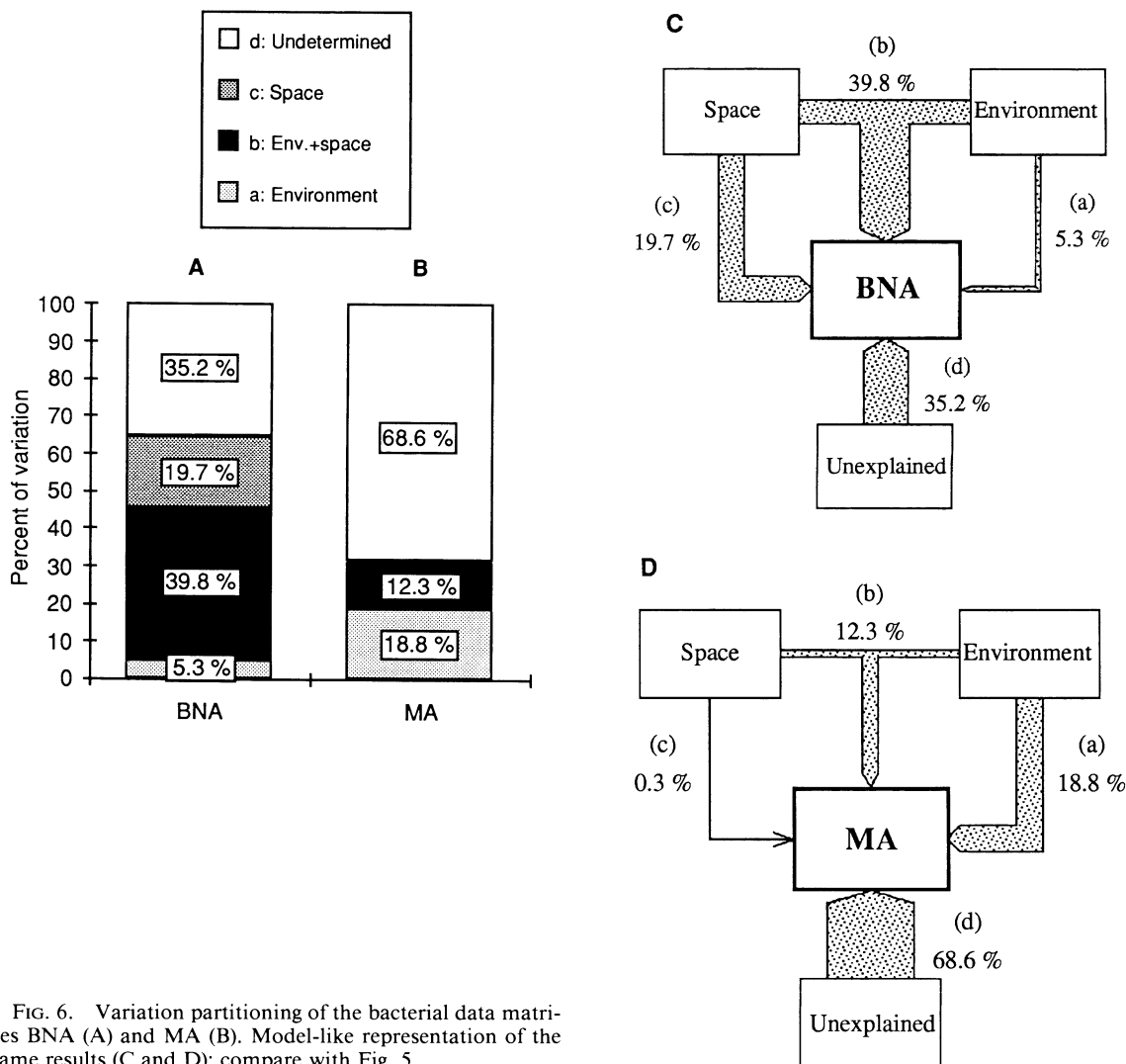


FIG. 6. Variation partitioning of the bacterial data matrices BNA (A) and MA (B). Model-like representation of the same results (C and D): compare with Fig. 5.

(or the regression coefficients if there is a single variable to explain). (2) The variation explained by environment *and* space is also quantified, but the reasons of this sharing cannot be fully deduced from our procedure. Here, a series of models whose predictions are tested by Mantel tests, as proposed by Legendre and Troussellier (1988), could clarify the relations: common underlying cause, or factors influencing the environmental variables, which in turn influence the species? (3) The amount of "strictly spatial" variation can be of particular importance in ecological investigations when one seeks to bring some hardly measurable influences to light, as for instance disturbance events against environmental factors (as in Leduc et al. 1992). Note that careful attention must first be given to the search for plausible environmental factors that could explain some of that remaining spatial variation. A high amount of strictly spatial variation, as in our second example, can then be interpreted in terms of some

alternative hypothesis, once the environmental control hypothesis has been discarded.

One should, however, be particularly cautious when trying to interpret this "strictly spatial" fraction of variation. Although the integration of space as a "full-right" explanatory variable into models can be very useful, as we have shown, it should always be framed within an ecological theory: disturbance dynamics, predation, competition, and so on. In this context, our method could also be considered as a way of confirming the predictions of such theories when spatial structures can be shown to be independent of the environmental descriptors, both for observational and for experimental studies.

The orthogonal partition of the variation that we obtain limits the variation decomposition to blocks of explanatory variables. Because of that, one has to look at the number of independent variables per block when comparing the amounts of explained variation. Indeed,

as our method does not provide a distinction between true structural and stochastic variation, each explanatory variable is likely to increase, by chance alone, the amount of explained variation. In the case of the spatial explanatory set, we have proposed here to overcome at least part of this problem by applying a reasoned choice of spatial variables, in the manner of multivariate stepwise selection. Other stopping criteria can be imagined, as, for instance, adding terms of the polynomial trend-surface equation until the Mantel correlogram of the residual surface becomes statistically nonsignificant. In general, one should try to use comparable numbers of carefully chosen explanatory variables in each block, if possible. In our example 1, there were 11 environmental and 5 spatial variables; in example 2 there were 6 environmental and 7 spatial variables; in example 3 there were 2 environmental and 2 spatial variables. Stated in a regression modeling context, our partition of the variation only aims at giving to both the environmental and spatial components as much chance as possible to express themselves in the explanation of the dependent variables' data table.

Some common features of the results presented in our three examples may be surprising. The amount of unexplained variation, for instance, is always fairly high. In the state of the art, however, one cannot discriminate between the "potentially explainable" variation and the "real" stochasticity in that unexplained variation. It may not be feasible to measure all the environmental variables (in the broad sense: biological interactions and external environmental factors) that are relevant in an ecological study. On the other hand, by using enough terms in the "space" matrix, one makes sure that one explains as much of the spatial variation of the species data as possible; so if the unexplained variation contains deterministic components, these effects will be local and not spatially organized over the sampling area.

Given these constraints, the amounts of variation involved in the main explained trends of the example data sets may seem proportionally low, but the underlying causes found to be significant can nevertheless be considered as important in the structuring of these communities.

There are still investigations to be made of the robustness of our method. For instance, we have to verify whether random perturbations of the data, or the addition of random variables, lead to important shifts of the results; preliminary attempts seemed to generate no great perturbations, however, at least with reasonably large data sets. These issues will be discussed in a separate paper.

Our new method is presented as complementary to the technique of partial Mantel modeling of Legendre and Troussellier (1988), that allows only simple models of relationships to be tested; the method proposed in this paper quantifies more precisely the partitioning of

the variation between the spatial and environmental components. When Mantel modeling results are available, as for instance with the bacterial data, one sees that there is a high degree of concordance between the results of the two methods. The next step will be to extend these methods to allow modeling the relationships among more than three data sets.

ACKNOWLEDGMENTS

It is a pleasure to acknowledge the interesting suggestions made by Cajo C.J.F. ter Braak and by another anonymous reviewer. This research was carried out during tenure of a Postdoctoral Fellowship of the Swiss National Foundation for Scientific Research by D. Borcard, and of a Killam Research Fellowship of the Canada Council by P. Legendre. It was also supported by NSERC grant No. A7738 to P. Legendre. This is contribution No. 371 of the Groupe d'Écologie des Eaux douces, Université de Montréal.

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.
- Benzecri, J. P., et al. 1973. L'analyse des données. Two volumes. Dunod, Paris, France.
- Bergeron, Y., A. Bouchard, and A. Leduc. 1988. Les successions secondaires dans les forêts du Haut-Saint-Laurent, Québec. *Naturaliste Canadien* 115:19-38.
- Berthet, P., and G. Gérard. 1965. A statistical study of microdistribution of Oribatei (Acari). Part I: the distribution pattern. *Oikos* 16:214-227.
- Boerner, R.E.J. 1985. Alternate pathways of succession on the Lake Erie islands. *Vegetatio* 63:35-44.
- Bouchard, A., Y. Bergeron, C. Camiré, P. Gangloff, and M. Gariépy. 1985. Proposition d'une méthodologie d'inventaire écologique et de cartographie écologique: le cas de la MRC du Haut-Saint-Laurent. *Cahiers de Géographie du Québec* 29:79-95.
- Bray, R. J., and J. T. Curtis. 1957. An ordination of the upland forest communities of southern Wisconsin. *Ecological Monographs* 27:325-349.
- Brisson, J., Y. Bergeron, and A. Bouchard. 1988. Les successions secondaires sur sites mésiques dans le Haut-Saint-Laurent, Québec, Canada. *Canadian Journal of Botany* 66:1192-1203.
- Connell, J. H. 1983. On the prevalence and relative importance of interspecific competition: evidence from experiments. *American Naturalist* 122:661-697.
- Estabrook, G. F., and D. J. Rogers. 1966. A general method of taxonomic description for a computed similarity measure. *BioScience* 16:789-793.
- Finegan, B. 1984. Forest succession. *Nature* 312:109-114.
- Gauch, H. G., Jr. 1982. Multivariate analysis in community ecology. Cambridge University Press, New York, New York, USA.
- Gower, J. C. 1987. Introduction to ordination techniques. Pages 3-64 in P. Legendre and L. Legendre, editors. *Developments in numerical ecology*. NATO ASI Series, Volume G 14. Springer-Verlag, Berlin, Germany.
- Hill, M. O. 1973. Reciprocal averaging: an eigenvector method of ordination. *Journal of Ecology* 61:237-249.
- Hudon, C., P. Legendre, A. Lavoie, J.-M. Dubois, et G. Vigeant. 1990. Effets du climat et de l'hydrologie sur le recrutement du homard américain (*Homarus americanus*) dans le nord du golfe du Saint-Laurent. Pages 161-177 in J.-C. Therriault, editor. *Le golfe du Saint-Laurent: petit océan ou grand estuaire?* Canadian Special Publications in Fisheries and Aquatic Sciences 113.
- Hughes, J. W., and T. J. Fahey. 1988. Seed dispersal and

- colonization in a disturbed northern hardwood forest. *Bulletin of the Torrey Botanical Club* **115**:89–99.
- Jongman, R. H. G., C. J. F. ter Braak, and O. F. R. van Tongeren, editors. 1987. *Data analysis in community and landscape ecology*. PUDOC, Wageningen, The Netherlands.
- Langeland, A. 1982. Interactions between zooplankton and fish in a fertilized lake. *Holarctic Ecology* **5**:273–310.
- Leduc, A., P. Drapeau, Y. Bergeron, and P. Legendre. 1992. Importance of spatial components in the analysis of forest cover variations. *Journal of Vegetation Science*, *in press*.
- Legendre, L., and P. Legendre. 1983. *Numerical ecology. Developments in environmental modelling*, 3. Elsevier, Amsterdam, The Netherlands.
- Legendre, P. 1990. Quantitative methods and biogeographic analysis. Pages 9–34 in D. J. Garbary and R. R. South, editors. *Evolutionary biogeography of the marine algae of the North Atlantic*. NATO ASI Series, Volume G 22. Springer-Verlag, Berlin, Germany.
- Legendre, P., and M.-J. Fortin. 1989. Spatial pattern and ecological analysis. *Vegetatio* **80**:107–138.
- Legendre, P., and M. Troussellier. 1988. Aquatic heterotrophic bacteria: modeling in the presence of spatial autocorrelation. *Limnology and Oceanography* **33**:1055–1067.
- Mantel, N. 1967. The detection of disease clustering and a generalized regression approach. *Cancer Research* **27**:209–220.
- May, R. M. 1984. An overview: real and apparent patterns in community structure. Pages 3–16 in D. R. Strong, D. Simberloff, L. G. Abele, and A. B. Thistle, editors. *Ecological communities: conceptual issues and the evidence*. Princeton University Press, Princeton, New Jersey, USA.
- Mech, L. D. 1981. *The wolf. The ecology and behavior of an endangered species*. University of Minnesota Press, Minneapolis, Minnesota, USA.
- Orlói, L. 1978. *Multivariate analysis in vegetation research*. Dr. W. Junk B.V., The Hague, The Netherlands.
- Pielou, E. C. 1984. *The interpretation of ecological data. A primer on classification and ordination*. Wiley, New York, New York, USA.
- Quinn, J. F., and A. E. Dunham. 1983. On hypothesis testing in ecology and evolution. *American Naturalist* **122**:602–617.
- Reinertsen, H., A. Jensen, A. Langeland, and Y. Olsen. 1986. Algal competition for phosphorus: the influence of zooplankton and fish. *Canadian Journal of Fisheries and Aquatic Sciences* **43**:1135–1141.
- Schoener, T. W. 1983. Field experiments on interspecific competition. *American Naturalist* **122**:240–285.
- Smouse, P. E., J. C. Long, and R. R. Sokal. 1986. Multiple regression and correlation extensions of the Mantel test of matrix correspondence. *Systematic Zoology* **35**:627–632.
- Southwood, T. R. E. 1987. The concept and nature of the community. Pages 3–27 in J. H. R. Gee and P. S. Giller, editors. *Organization of communities past and present*. Blackwell Scientific, Oxford, England.
- ter Braak, C. J. F. 1986. Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. *Ecology* **67**:1167–1179.
- . 1987a. The analysis of vegetation–environment relationships by canonical correspondence analysis. *Vegetatio* **69**:69–77.
- . 1987b. Ordination. Pages 91–173 in R. H. G. Jongman, C. J. F. ter Braak, and O. F. R. van Tongeren, editors. *Data analysis in community and landscape ecology*. PUDOC, Wageningen, The Netherlands.
- . 1988a. Partial canonical correspondence analysis. Pages 551–558 in H. H. Block, editor. *Classification and related methods of data analysis*. North Holland Press, Amsterdam, The Netherlands.
- . 1988b. CANOCO—an extension of DECORANA to analyze species–environment relationships. *Vegetatio* **75**:159–160.
- van den Wollenberg, A. L. 1977. Redundancy analysis. An alternative for canonical correlation analysis. *Psychometrika* **42**:207–219.
- Whittaker, J. 1984. Model interpretation from the additive elements of the likelihood function. *Applied Statistics* **33**:52–64.
- Whittaker, R. H. 1956. *Vegetation of the Great Smoky Mountains*. *Ecological Monographs* **26**:1–80.
- Wiens, J. A. 1989. *The ecology of bird communities*. Volume 2: Processes and variations. Cambridge University Press, Cambridge, England.