Environmental control and spatial structure in ecological communities: an example using oribatid mites (Acari, Oribatei)

DANIEL BORCARD* and PIERRE LEGENDRE[‡]

*Institut de Zoologie, Université de Neuchâtel, Chantemerle 22, CH-2000 Neuchâtel, Switzerland
[†]Departemente de Sciences Biologiques, University de Montreal, CP 6128, Succursale A, Montreal, Quebec H3C 3J7, Canada

We have recently proposed to use partial canonical ordinations to partition the variation of species abundance data into four additive components: environmental at a local scale, the spatial component of the environmental influence, pure spatial, and an undetermined fraction. By means of an example, we show how to use the information contained in these fractions to provide better insight into the data. In particular, the interpretation is assisted by separately mapping the various canonical axes and relating them to possible generating processes. We derive a general framework for the causal interpretation of the various fractions of this partition, which includes the environmental and the biotic control models, as well as historical dynamics.

Keywords: biotic control, canonical correspondence analysis, cryptostigmatic mites, environmental control, historical dynamics, mapping, modeling ecological relationships, spatial patterns, variation partitioning

1. Introduction

In the ecological literature, the causes of community structuring are generally related to two

classical models (May, 1984): the environmental control model (Whittaker, 1956; Bray and Curtis, 1957; Hutchinson, 1957) and the biotic control model (Lindeman, 1942; Southwood, 1987). Often considered as mutually exclusive (May, 1984), these two classes of influence factors could in fact be responsible for complementary aspects of community structure (Quinn and Dunham, 1983), together with other factors like historical events, inputs of auxiliary energy, or disturbances of various kinds.

Due to the diverse (and often partly unknown) nature of these structuring causes, assessment of their relative contribution is quite difficult. While it is often possible to obtain measures of the main abiotic site descriptors, biotic interactions like predation, competition and so on, or past (historical) events like fires or human alterations, cannot be directly measured. This is one reason for the increasing interest, shown by ecologists, in quantifying the spatial structure of biotic or abiotic variables: the spatial structure can be considered as evidence for the various processes that have generated it, and can therefore be used as a synthetic, indirect descriptor of these processes. Models integrating space as a predictive variable, based on Mantel (1967) and partial Mantel testing (Smouse *et al.*, 1986), have been used for this purpose by Legendre and Troussellier (1988), Legendre and Fortin (1989), and Leduc *et al.* (1992). Borcard *et al.* (1992) describe a method of variation partitioning, based on canonical ordinations, that allows quantification of the main classes of influence (environment and 'space' alone, and combination) acting on a species assemblage.

In the present paper, we want to show that this modeling method, combined with contouring of the various fractions of the variation, is a useful tool in ecological and environmental science. It allows testing the hypotheses that motivated the study and also generating new hypotheses to account for unexplained but spatially structured components of the variation. This extension of classical linear modeling is part of a general causal framework, presented in the Discussion and Conclusions section, incorporating the various types of models and effects that may be associated with the various fractions of variation.

The method was applied to an Oribatid mite community data set sampled through space. Environmental factors such as humidity, soil type and vegetation have been recognized as significant determinants of Oribatid community structure (West, 1984; Wauthy *et al.*, 1989; Borcard, 1991). We will discuss how much of our mite community spatial variation can be explained by these factors. Food resources such as microorganisms also seem to have great importance (Cancela da Fonseca and Poinsot-Balaguer, 1983), but their abundance and availability for the mites is difficult to measure directly. Among the biotic factors, predation seems to be a major structuring process, as suggested by Lebrun *et al.* (1991). Other biotic interactions such as competition have never been clearly shown to act significantly, so that some authors suggest that soil mite communities could often be in a state of non-equilibrium in the sense of Wiens (1984) (Wauthy *et al.*, 1989). Other authors propose alternatives to the classical but somewhat rigid competitive exclusion principle, in the form of time-dependent exploitation of alternate resources (Ponge, quoted by Vannier, 1985). Another class of factors has been advocated as a structuring influence on soil communities, namely human alterations; for instance, trampling has a profound effect on the numerical balance of communities, and (somewhat less) on species composition (Garay and Nataf, 1982; Borcard, 1988).

No complete and standard analytical procedure is available to discriminate between these various sources of influence. We propose to scrutinize independent fractions of the variation of a soil mite community after mapping of these fractions. The method of Borcard *et al.* (1992), referred to above, achieves a partition of the spatial variation of species data in two mappable components: (i) the local (fraction *a*) and spatially structured (fraction *b*) covariation between

the mite community and the measured environmental features, and (ii) the community spatial structure that is independent of these environmental variables (fraction c).

2. Sampling site and methods

The study site is located on the southern shore of a small Laurentide lake, Lake Geai, located on the territory of the Station de Biologie des Laurentides (74°W, 46°N). According to Harper and Cloutier (1985), the main characteristics of the lake are as follows: small (0.92 ha) dystrophic, dimictic lake, with acidic water (mean pH of 6.0, after Plinski and Magnin, 1979). The water is heavily coloured, with a high content of organic matter (mean value of total organic C: 10 mg L^{-1}), but weakly mineralized (mean conductivity: 26 mS cm^{-1}). The surrounding forest of *Betula papyrifera*, which developed after a fire around 1920, harbors some *Acer saccharum* and, in the immediate vicinity of the lake, coniferous species such as *Picea mariana*, *Abies balsamea* and *Thuja occidentalis*.

In June 1989, we delineated a 10 × 2.6 m transect on the partially floating vegetation mat surrounding the lake and extending from the forest border to the free water, and mapped the main environmental features of the site (Figure 1). In order to get an accurate representation of all the elements of the vegetation mosaic, the following sampling strategy was used: a grid with 20 × 20 cm mesh was defined over the area; stratified sampling was performed, in which the seven substratum classes acted as strata (see next paragraph and Figure 2; notice that the interface between *Sphagnum* patches is one of those strata). Within each stratum, we applied a systematic sampling design, with sampling effort proportional to the square root of the surface area of the stratum estimated by means of the grid. This procedure allowed the sampling to be spread over most of the delimited area and to be weighted in proportion to site heterogeneity. A total of 70 cores, each 5 cm in diameter and 7 cm in depth, were obtained by means of a sampler described by Borcard (1986); extraction of the mite fauna was performed with a multiple extractor of the type described in the same paper.

Besides the counts of adult Oribatid mites, we noted the spatial location of the samples, as well as the following variables: substratum (seven nominal classes: *Sphagnum magellanicum* (with a majority of *Sphagnum rubellum*), *Sphagnum rubellum*, *Sphagnum nemoreum* (with a minority of *Sphagnum angustifolium*), *Sphagnum rubellum* + *Sphagnum magellanicum* (equal parts), ligneous litter, bare peat, interface between *Sphagnum* species), microtopography of the substratum (2 qualitative classes: blanket (flat) or hummock (raised)), coverage density of the shrub cover (3 semi-quantitative classes), water content in percent (quantitative), density of the substratum in g L^{-1} of dry uncompressed matter (quantitative). The four groups of *Sphagnum* mosses are called 'group 1' to 'group 4' in the Figures and Tables. The substratum type variable acts as a global descriptor for features such as substratum-dependent types and quantity of food for the mites (fungi, algae, bacteria), as well as microstructural characteristics that would be too laborious, or even impossible, to measure directly, such as the arrangement of stems and leaves. Microtopography of the substratum refers to the fact that some *Sphagnum* species (mostly *S. magellanicum* in our case), instead of growing as a horizontal carpet, sometimes build more or less hemispherical hummocks. These formations often appear in sites where the water table is low; since closed and dense formations prevent evaporation (as in alpine plant cushions), these hummocks are considered to be reactions of the sphagna against desiccation (Matthey, 1965). Other types of hummocks appear at the 'waterfront' of *Sphagnum* carpets, where the mosses grow their way into the water. Although the hummocks recreate the same humidity conditions as those prevailing in the normally wet flat blankets, we were interested in whether the mite community reacted to
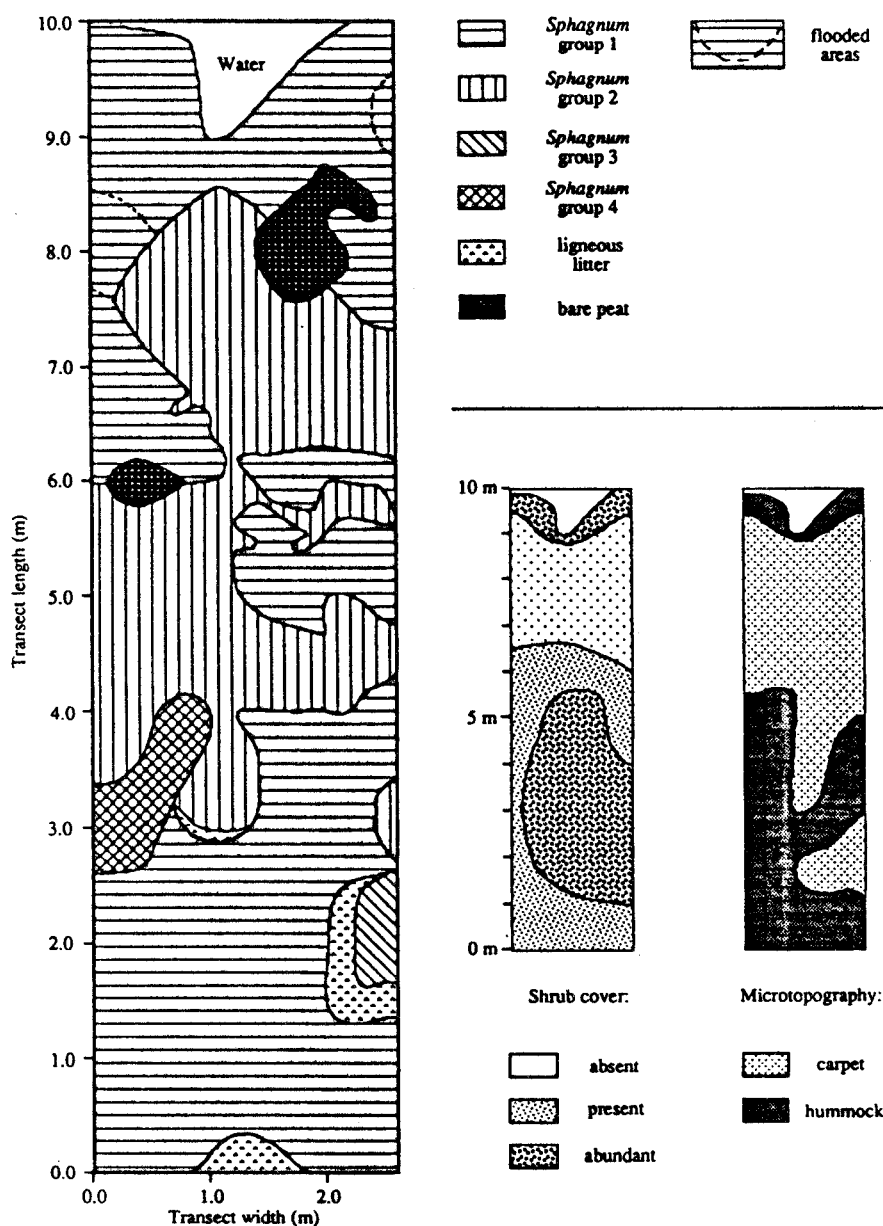
Figure 1. General map of the sampling area, showing the distribution of variables substratum, shrub coverage and microtopography.

topographic modifications of their habitat. Shrub cover density was coarsely recorded (three semi-quantitative classes: absent, present, abundant) to check whether the larger amount of ligneous matter and the shade provided by these shrubs could explain the possible presence of forest species in the mite community. The *Sphagnum* variables and the shrub cover are the only ones that unambiguously (though not exclusively) characterize the trophic environment of the mites; the other variables describe, to various extents, their physical environment.
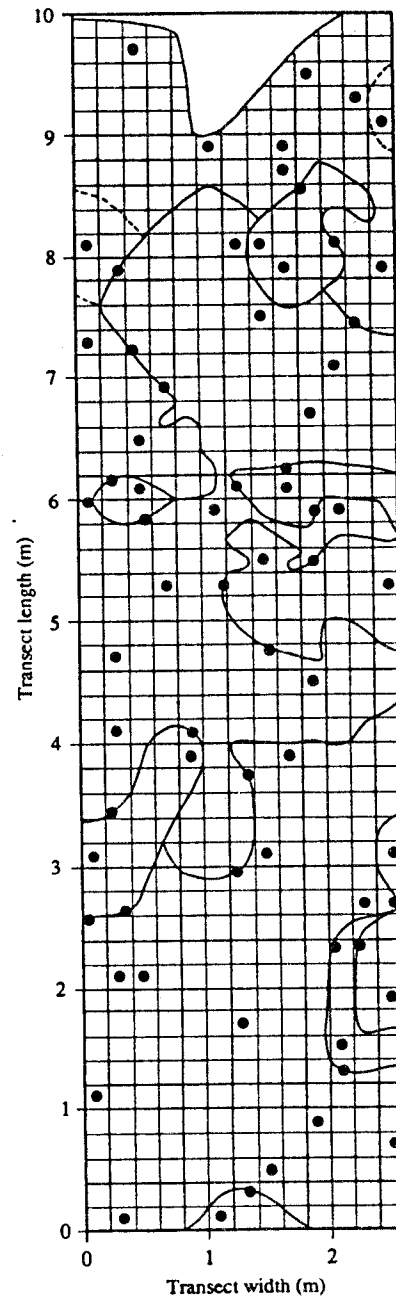
Figure 2. Map of the sampling area, showing the 20 × 20 cm sampling grid and the location of the 70 samples.

3. Mathematical methods

We used canonical and partial canonical correspondence analysis (ter Braak, 1986, 1988a) for gradient analysis and variation partitioning, as proposed by Borcard *et al.* (1992), and inverse squared distance interpolation for gridding.

Canonical correspondence analysis (CCA) is a combination of 'classical' multivariate ordination and multiple regression analysis; the species ordination axes are constrained to be linear combinations of the environmental variables, provided in a second matrix. This ensures that the canonical ordination axes are optimally related to the supplied environmental descriptors. The method provides two sets of sample scores: the sample scores that are weighted averages of the species scores (as in classical correspondence analysis), and the sample scores that are linear combinations of environmental variables (i.e., the scores that are predicted by the multiple regression model). We call these two categories the 'species-derived scores' and the 'model scores', respectively. The partial version of CCA allows removing, by multiple linear regression, the effect of known or undesirable variables before computing the constrained ordination. All these procedures are available in release 3.11 of the CANOCO^{TM} program (ter Braak, 1988b, 1990). Further explanations can be found in the author's papers (ter Braak, 1986, 1987a, 1987b, 1988a, 1988b). For examples of applications, see for instance Birks *et al.* (1990), Stevenson *et al.* (1989), Siepel (1989) (calibration of indicator species), and Borcard *et al.* (1992) (community structure analysis).

Inverse squared distance weighting is a classical interpolation method for gridding data; it uses a reasonable compromise between the hypotheses of a very large and that of a very weak influence of remote neighbours; according to Isaaks and Srivastava (1989, p. 259), it represents the most common choice. It uses the following equation:

$$z(\text{node}) = \frac{\Sigma_i\left(\dfrac{z(\text{point}_i)}{d^2(\text{node, point}_i)}\right)}{\Sigma_i\left(\dfrac{1}{d^2(\text{node, point}_i)}\right)} \tag{1}$$

Maps were produced with the MacGridzo^{TM} program; the grid mesh width was set to 0.2 m.

The CANOCO program can be obtained from Microcomputer Power, 111 Clover Lane, Ithaca, NY 14850, USA, or Scientia Publishing, pf. 658, H-1365 Budapest, Hungary. MacGridzo is a product of RockWare, Inc., 4251 Kipling St., Suite 595, Wheat Ridge, Colorado 80033, USA.

4. General results

We captured a total of 9850 adult Oribatid mites, which represents a mean density of about 72 000 individuals per square metre. Forty-nine taxa were recognized as species, though many were not given a species name, owing to the incomplete stage of systematic knowledge of the North American Oribatid fauna. Fourteen taxa, involving only 50 individuals, were subsequently eliminated, because their very poor representation introduced a large number of zeros in the data matrix, a situation that could be troublesome in the data analysis, without providing relevant information about the main trends of the community. Thus, the Oribatid data matrix to be analyzed has a 70 (samples) × 35 (species) dimension.

Table 1. Correlations between the four first species axes and the environmental variables of the CCA 'mites' by 'substratum variables'.

	Axis 1	Axis 2	Axis 3	Axis 4
Sphagnum group 1	−0.135	0.183	−0.502	−0.186
Sphagnum group 2	0.346	−0.123	0.214	−0.015
Sphagnum group 3	−0.200	−0.038	0.157	0.496
Sphagnum group 4	−0.082	−0.056	−0.164	0.060
Litter	−0.233	0.040	0.479	−0.326
Bare peat	0.106	0.520	0.143	0.130
Interfaces	0.053	−0.178	0.148	0.123

5. Contribution of the substratum type to the mite community variation

What follows is an example to show how maps of the results of a partial analysis can be exploited to generate new ecological hypotheses. We have at our disposal a data base with all the variables presented above. A normal descriptive run of analyses would consist in a first canonical correspondence analysis using all environmental variables, followed by a variation partitioning after Borcard *et al.* (1992) to check whether there remains an important fraction of unexplained spatial variation. In other cases, one may wish to test more specific hypotheses, involving only a subset of the explanatory variables. A subsequent examination of the remaining variation could then suggest the addition of other variables to increase the explanatory power of the model.

Among the variables at our disposal, there are seven classes of substratum. Most of these classes define *Sphagnum* species or decaying organic matter. So, this variable acts as a synthetic descriptor both for the physical environment of the mites (morphology, available space) and, at least partly, the food available. Oribatid mites mostly feed on mycelia, algae, bacteria or decaying organic matter, all these varying with the substratum. Could this simple substratum variable be sufficient to explain an important part of the mite community variation?

To test this hypothesis we performed a CCA of the mite data matrix, constrained by the seven substratum classes. The result is poor: the substratum variables explain altogether 13.7% of the mite community variation. The species-environment relationship is globally significant ($p = 0.009$ after 999 permutations), but the individual canonical axes are not. Correlations between the four main canonical axes and the seven substratum variables are shown in Table 1.

Despite the weak correlation with the environmental variables, we present a map of the first ordination axis (species-derived scores, $p = 0.107$) in Figure 3a. This map indicates that this axis is essentially made of a longitudinal trend. The second one displays striking local peaks (Figure 3b), that are easy to associate with spots of bare peat and flooded zones where the *Sphagnum* carpet is very thin.

6. Variation partitioning

At this step it is useful to introduce a matrix of spatial descriptors, in order to partition the mite data variation as proposed by Borcard *et al.* (1992). The matrix is made of the x and y coordinates of the samples, and some terms of a third-order trend surface regression equation; these terms have
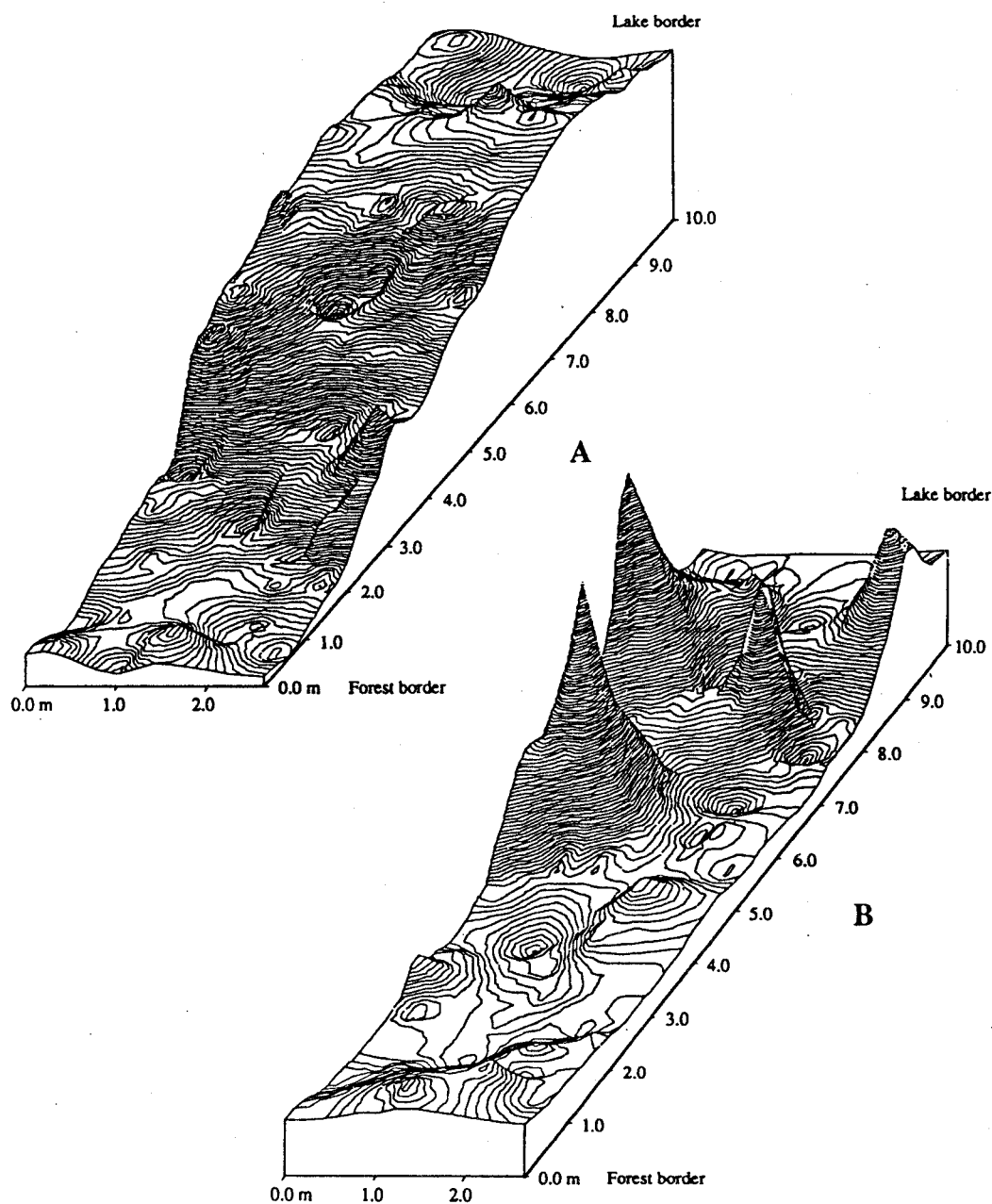
Figure 3. (a) Map of axis 1 of the 'mites' by 'substratum variables' CCA. (b) Map of axis 2 of the 'mites' by 'substratum variables' CCA. As in the following figures, no vertical scale is provided, since the maps have been rescaled from different units to help visual comparison.

been selected by means of a forward selection procedure from among the nine terms of a full third-order polynomial equation of the x and y coordinates. The partitioning will yield four fractions of the mite data variation (Figure 4a): fraction a (local species variation, explained by the environmental descriptors independently of any spatial structure), b (spatial structure in the species data
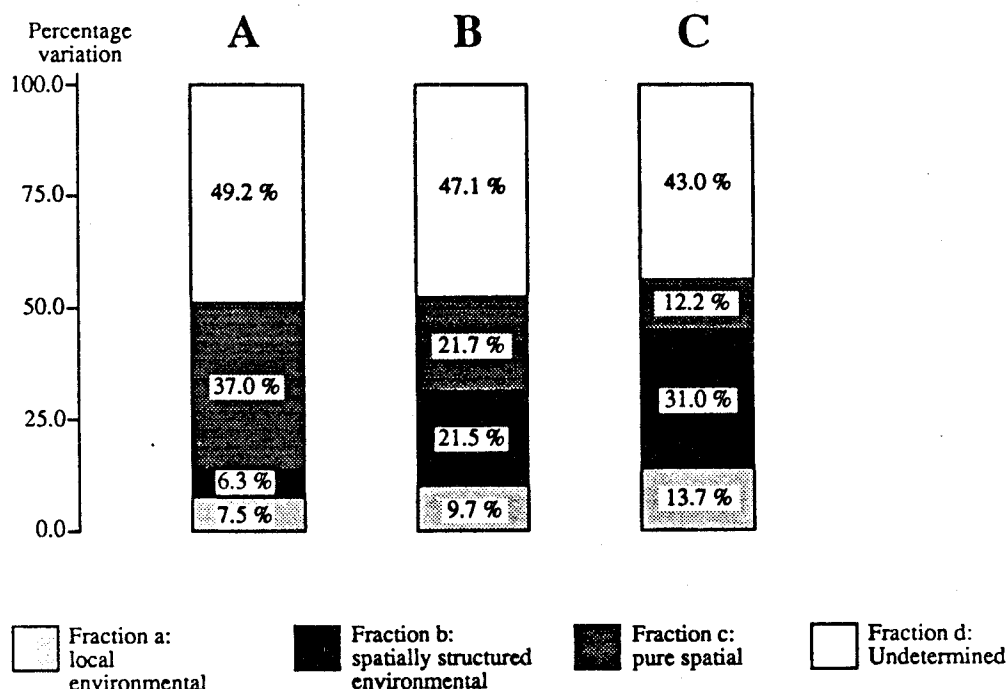
Figure 4. Variation partitioning of the Oribatid mite data table. A Environmental matrix made of the substratum variables. B Like A, with humidity added in the environmental matrix. C Complete environmental matrix, as in Borcard *et al.* (1992).

that is shared by the environmental descriptors), *c* (spatial patterns in the species data that are not shared by the environmental variables), and *d* (unexplained variation).

The result of the variation partitioning shows that 37.0% of the mite data variation can be explained significantly by the spatial matrix (fraction *c*, $p = 0.001$), after the effect of the substratum variables has been removed. The *a* and *b* fractions are comparatively quite small, reflecting the minor contribution of the substratum variables to the explanation of the species variation. A map of the first axis of fraction *a* (Figure 5) is almost identical to that of Figure 3b, confirming the local character of the variation due to the spots of bare peat. Also to be mentioned is that even complex data manipulations like those performed by CANOCO in these two CCA and partial CCA runs do not alter the data structure in any critical manner. A map of the first strictly spatial axis (fraction *c*, Figure 6) shows that there remains a strong, highly significant ($p = 0.001$), longitudinal gradient in the data.

7. Contribution of humidity

Table 2a shows how the environmental variables that have not yet been included in the analysis correlate with the first four strictly spatial canonical axes (fraction *c*). The variable most strongly correlated is humidity ($r = 0.696$), which thus becomes a good candidate for inclusion in our model. A new variation partitioning analysis including this variable produces results (Figure 4B)

Lake border

10.0

9.0

8.0

7.0

6.0

5.0

4.0

3.0

2.0

1.0

0.0 m   Forest border

0.0 m   1.0   2.0
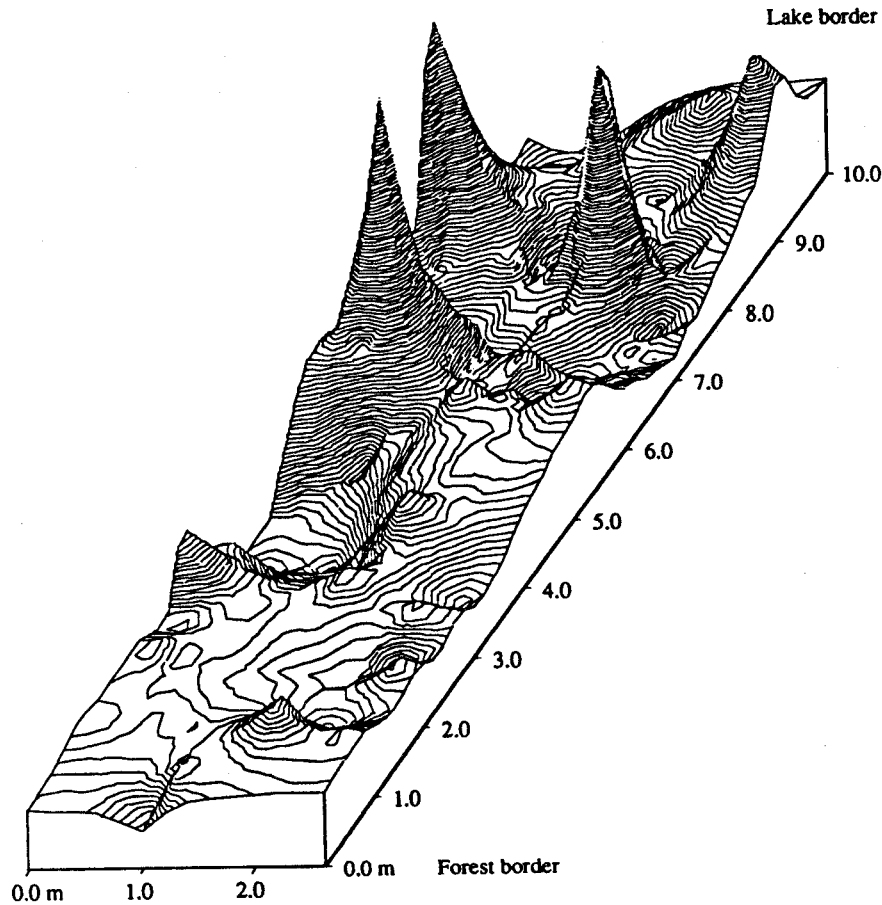
Figure 5. Map of axis 1 of the 'mites' by 'substratum variables' partial CCA, after removing the effect of space. Compare with Figure 3b.

that look quite different from the previous one (Figure 4A). The proportion of explained variation is now 31.2%, indicating that humidity has added 17.4% to the environmental explanation. Most strikingly, 15.2% of the variation has moved from fraction c (pure spatial) to b (spatially structured environmental variation). The first axis is now mainly a 'humidity axis' and is highly significant ($p = 0.001$). Fraction d (unexplained variation) has decreased only 2.1% to the profit of fraction a (local environmental influences).

8. Adding more variables

There is still 21.7% of the spatial variation which remains unexplained. A map of the first spatial axis (Figure 7a) displays a shape that is not a simple gradient any more. Correlating it with the remaining environmental variables (Table 2b) shows that the next good candidate for inclusion in the model would be substratum density ($r = -0.506$ with spatial axis 1), which was not correlated

Table 2. Correlations between the four first spatial axes of the partial CCA 'mites' by 'space' and the environmental variables not yet included in the model, (a) after removing the effect of the substratum variables, (b) after removing the effect of the substratum variables and humidity.

(a)	Axis 1	Axis 2	Axis 3	Axis 4
No shrubs	0.632	0.233	-0.062	-0.074
Few shrubs	-0.124	-0.201	0.404	0.309
Many shrubs	-0.462	-0.004	-0.350	-0.243
Hummock	-0.535	0.139	-0.262	0.194
Substratum density	-0.101	0.220	0.196	0.066
Humidity	0.696	0.297	0.058	0.088

(b)	Axis 1	Axis 2	Axis 3	Axis 4
No shrubs	0.334	0.320	-0.048	0.025
Few shrubs	-0.110	-0.331	0.356	-0.255
Many shrubs	-0.199	0.037	-0.315	0.234
Hummock	-0.359	0.095	-0.263	-0.208
Substratum density	-0.506	-0.082	0.141	0.084

with the spatial axes at all in the previous run (Table 2a). A map of the density itself (Figure 7b) indeed resembles that of the first spatial axis in many aspects.

Inclusion of all the remaining variables into the analysis gives the partitioning reported by Borcard *et al.* (1992) and reproduced here (Figure 4C); the gain in explained variation from Figure 4B to 4C is only 4.1%, and this goes into fraction *a*; fraction *c* drops to 12.2%, adding 9.5% to fraction *b*.

9. Non-environmental community spatial structure: a clue to hidden processes?

In Figure 4C, 12.2% of the variation in the mite data matrix can still be explained using only the spatial data matrix, independently of any reference to the environmental variables. The first three ordination axes of fraction *c*, explaining together 11.8% of the mite community matrix variation (i.e., most of the total of 12.2% for the spatial, non-environmental effects), are significantly related to the spatial descriptors: $p = 0.001$, 0.001 and 0.014 respectively.

Is there some explanation for these results? On a general level, it is not surprising that the available explanatory variables do not explain all the spatial structure of a living community data set. First, some significant, spatially structured environmental variables which are important for the mite species assemblage (physical or chemical characteristics, food resources, etc.) may have been overlooked in the research project. Second, the community under study may have undergone in the past various kinds of disturbances – resulting from human activity, for instance – that are still reflected on its present spatial structure. Third, processes related to the population dynamics of the species involved, such as social behaviour or predator–prey interactions, are very likely to generate measurable spatial structures (biotic control model).

Lake border

Figure 6. Map of axis 1 of the 'mites' by 'space' partial CCA, after removing the effect of the substratum variables.

10. Discussion and conclusions

The simple case of environmental modeling presented in this paper shows how a set of spatial variables can be used to estimate the proportion of spatial variation remaining in a data set after the first hypothesis has been tested, and, with good knowledge of the organisms under study, how to generate new hypotheses regarding the causes of these remaining structures.

This method reflects the way ecologists actually work with data, testing simple models first, and then refining them using more data. In studies based upon previously assembled data bases, it is easy, as in the example above, to return to the files for additional data corresponding to the new hypotheses that are generated by examination of the maps. In hypothesis-driven studies, contemplation of the maps of the unexplained spatial fraction may lead to new hypotheses that may be included in the design of new field programs.

We mention two real-case studies. The first one concerns the structure displayed by 28 tree species in a 50 ha plot in southwestern Québec. This data set had been analyzed by Borcard *et al.* (1992); the variation was partitioned as described above, and it was found that fraction *d*
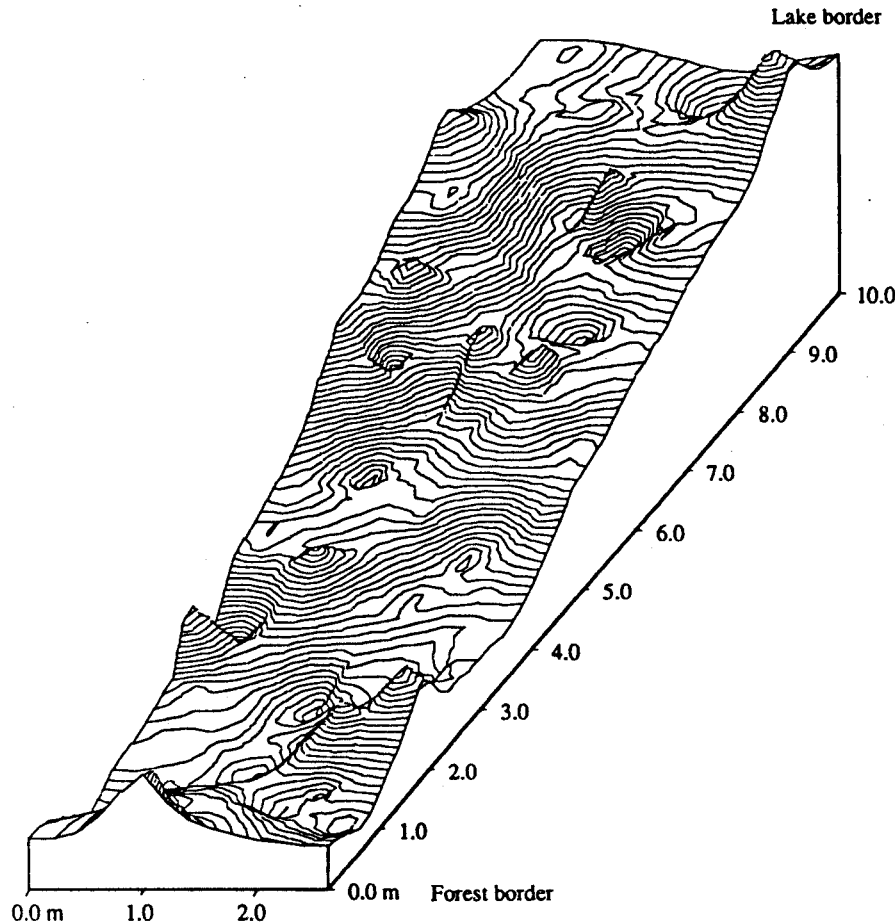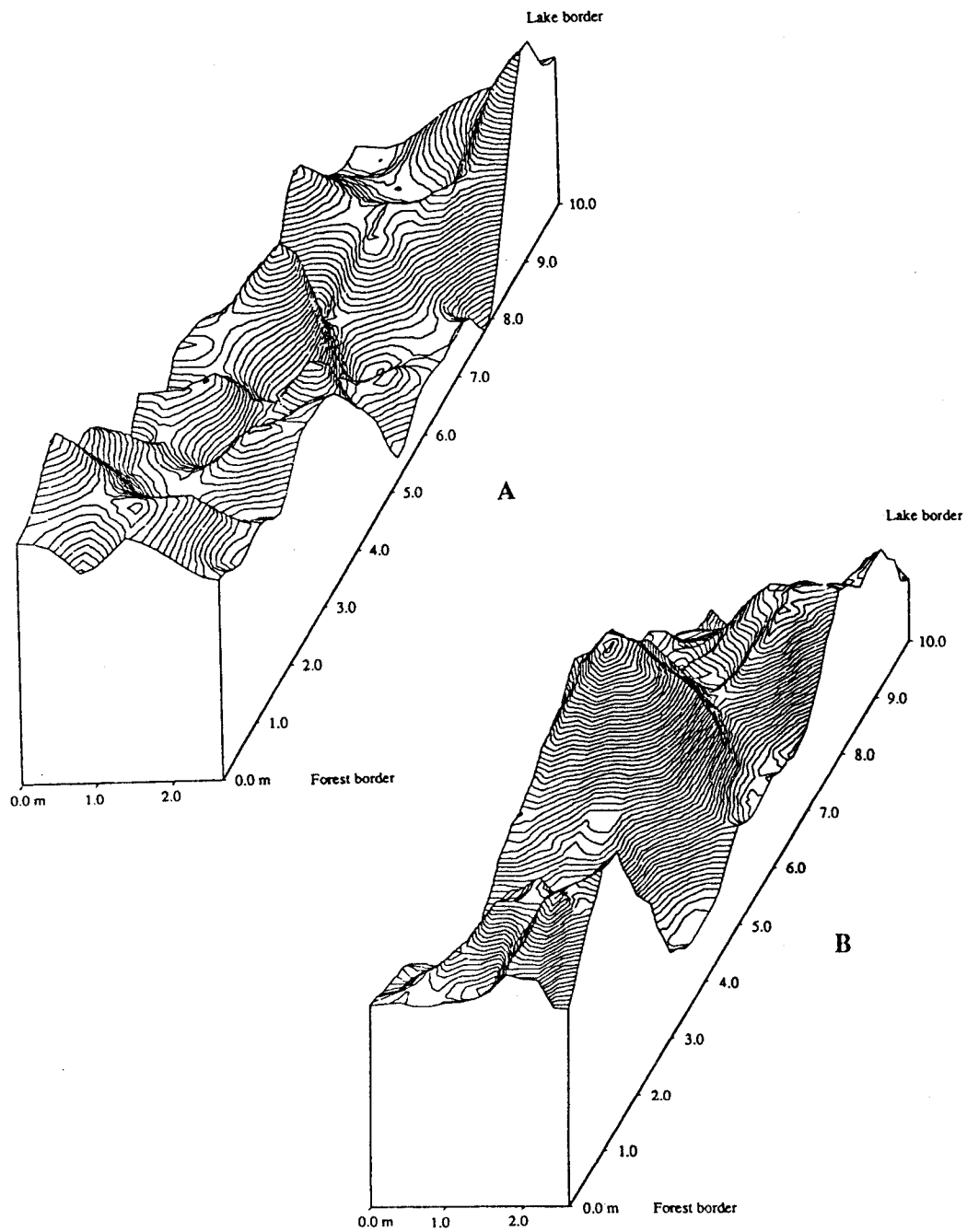
Figure 7. (a) Map of axis 1 of the 'mites' by 'space' partial CCA, after removing the effect of the substratum variables and humidity. (b) Map of variable 'substratum density'.

Table 3. Theoretical causal relationships between environmental variables (representing processes) and community structure. Fractions (a), (b), (c) and (d) of the community data variation refer to Figure 4.

Fraction	Causal factor	Process	Effect
(a)	Environmental factor	ECM	– Community structure
(a)*	Non-spatially structured factor not included in the analysis	ECM	– Env. variable in the analysis – Non-spatial community var.
	Historical events without spatial structure at the study scale	HD	– Env. variable in the analysis – Non-spatial community var.
(b)	Env. factor with spatial structure	ECM	– Community spatial structure
(b)*	Spatially structured env. factor not included in the analysis	ECM	– Env. variable in the analysis – Community spatial structure
	Spatially structured historical events	HD	– Env. variable in the analysis – Community spatial structure
(c)*	Spatially structured factors not included in the analysis	ECM	– Community spatial structure
	Spatially structured historical events	HD	– Community spatial structure
	Predation, competition, etc.	BCM	– Community spatial structure
(d)*	Factor not included in the analysis, not spatially structured (at study scale)	ECM	– Non-explained community variation
	Biotic control factors not spatially structured (at study scale)	BCM	– Non-explained community variation
	Random variation, sampling error, etc.	Noise	– Non-explained community variation

ECM = Environmental control model. BCM = Biotic control model. HD = Historical dynamics. Asterisks (*) indicate factors not explicitly spelled out in the model.

(unexplained variation) contained 63.3% of the variation. This led to another piece of research (Leduc *et al.*, 1992, using partial Mantel testing instead of partial canonical analysis), where additional community dynamics hypotheses were tested, dealing especially with the relationship between trees of the various species, their saplings, and the environmental variables; it was found, in particular, that species associated with hydric or mesic conditions respond differently in the models. In another study, Legendre and Troussellier (1993) investigated models to explain the spatial variations of abundance of phytoplankton and of two categories of heterotrophic bacteria in a brackish lagoon. Classical regression modeling, using only environmental descriptors as independent variables, left 42% to 100% of the spatial variation unexplained; mapping (Legendre *et al.*, 1989) as well as Borcard *et al.* (1992) variation partitioning of these variables, gave the clue that in many cases, spatial structuring played an important role; using a water circulation model among the predictive variables helped reduce considerably the amount of unexplained variation in most models.

As in any causal analysis, causality resides in our hypotheses, and the objective is to verify how much of the variation can be explained by a consistent body of hypotheses (i.e., without incompatibilities). One has to be aware, however, of the problem that may be created by causal factors

that are not explicitly spelled out in the model. Not only can they contribute to the unexplained variation; they can also produce false correlations, if they are causally anterior to both the variables to be explained and some of the variables hypothesized as explanatory in the model. In community analysis, furthermore, we are faced with a multiplicity of potential causal agents for the observed structures, acting at different spatial and temporal scales. Three main sources of causal variation were mentioned in the Introduction. These relate to the environmental control model (ECM), the biotic control model (BCM), and to other factors such as historical events (anthropogenic: agriculture, logging, constructions of various sizes, etc.; or natural: isolation by geographic barriers, etc.) or disturbances of various kinds (storms, forest fires, volcanic eruptions, landfalls, etc.), that can collectively be referred to as historical dynamics (HD). Table 3 lists the various cause-to-effect relations that can be contained in each of fractions *a* to *d*. First, the effect 'in the model' is listed for each fraction, followed by other possible causes and effects not explicitly spelled out in the model (*). As an illustration of the use of this table, we propose some examples of hidden causes for the Oribatid mite example (the examples are presented in the same order as the general mechanisms in Table 3).

(a)* Local, unmeasured variations in substratum chemistry are likely to affect both the mites and the vegetation in which they live, a case that would lead to covariation between them. Imagine that, one month before the sampling, a flock of birds has spent a few hours on the sampling site. Their randomly distributed faeces could have altered the vegetation and thus affected soil animals.

(b)* Suppose that some years ago a tree fell across our sampling site and rotted away. By now it would have disappeared under the sphagna, but these mosses would have been affected by it, and the mite community would certainly reflect the presence of wood residues, for instance by an increase in the density of the Phthiracarids. Consequently a part of the vegetation-mite covariation would be caused by this past event.

(c)* Let us take again our example of the fallen tree. Had we not noted the vegetation patterns, the effect of the wood on the mite community would not have covaried with an explicit environmental factor, and consequently would have been displayed in fraction *c* of the spatial variation of unknown origin.

Partitioning the variation of the community structure into its components and mapping them has allowed, not only finding interesting correlations that support preconceived models, but quantifying and mapping where the hypotheses fall short of the reality. Ecologists can undoubtedly use such insights to go back to the field with better models in mind.

References

Birks, H.J.B., Juggins, S. and Line, J.M. (1990). Lake surface-water chemistry reconstructions from palaeolimnological data. In: *The Surface Waters Acidification Programme* (B.J. Mason, ed.) Cambridge University Press, pp. 301–13.

Borcard, D. (1986). Une sonde et un extracteur destinés à la récolte d'Acariens (Acari) dans les sphaignes (*Sphagnum* spp.). *Bull. Soc. Entomol. Suisse*, 59, 283–8.

Borcard, D. (1988). Les Acariens Oribates des sphaignes de quelques tourbières du Haut-Jura suisse. PhD Thesis, Université de Neuchâtel, Switzerland.

Borcard, D. (1991). Les Oribates des tourbières du Jura suisse (Acari, Oribatei): écologie. II. Les relations Oribates – environnement à la lumière du test de Mantel. *Rev. d'Écol. Biol. Sol*, 28, 323–39.

Borcard, D., Legendre, P. and Drapeau, P. (1992). Partialling out the spatial component of ecological variation. *Ecology*, 73, 1045–55.

Bray, R.J. and Curtis, J.T. (1957). An ordination of the upland forest communities of southern Wisconsin. *Ecol. Monographs*, **27**, 325–49.

Cancela da Fonseca, J.P. and Poinsot-Balaguer, N. (1983). Les régimes alimentaires des Microarthropodes du sol en relation avec la décomposition de la matière organique. *Bull. Soc. Zool. France*, **108**, 371–88.

Garay, I. and Nataf, L. (1982). Microarthropods as indicators of human trampling in sub-urban forests. In: *Urban Ecology* (R. Bornkamm, J.A. Lee and M.R.D. Seaward, eds) Blackwell, Oxford. pp. 201–207.

Harper, P.-P. and Cloutier, L. (1985). Composition et phénologie de communautés d'insectes du lac Geai, lac dystrophe des Laurentides (Québec). *Natur. Canadien (Revue d'Écol. Systém.)*, **112**, 405–15.

Hutchinson, G.E. (1957). Concluding remarks. *Cold Spring Harbor Symp. Quant. Biol.*, **22**, 415–27.

Isaaks, E.H. and Srivastava, R.M. (1989). *Applied Geostatistics*. Oxford University Press, New York.

Lebrun, P., van Impe, G., de Saint Georges-Gridelet, D., Wauthy G. and André, H.M. (1991). The life strategies of mites. In: *The Acari. Reproduction, Development and Life-history Strategies* (R. Schuster and P.W. Murphy, eds) Chapman & Hall, London.

Leduc, A., Drapeau, P., Bergeron, Y. and Legendre, P. (1992). Study of spatial components of forest cover using partial Mantel tests and path analysis. *J. Veg. Sci.*, **3**, 69–78.

Legendre, P. and Fortin, M.-J. (1989). Spatial pattern and ecological analysis. *Vegetatio*, **80**, 107–38.

Legendre, P. and Troussellier, M. (1988). Aquatic heterotrophic bacteria: modeling in the presence of spatial autocorrelation. *Limnol. Oceanogr.*, **33**, 1055–67.

Legendre, P., Troussellier, M., Jarry, V. and Fortin, M.-J. (1989). Design for simultaneous sampling of ecological variables: from concepts to numerical solutions. *Oikos*, **55**, 30–42.

Legendre, P. and Troussellier, M. (1993). Origin of spatial structures in aquatic bacterial communities: From hypotheses to numerical solutions. In: *Trends in Microbial Ecology* (R. Guerrero and C. Pedrós-Alió, eds) Proceedings of the Sixth International Symposium on Microbial Ecology (ISME-6), Barcelona, 6–11 September 1992. pp. 353–8.

Lindeman, R.L. (1942). The trophic-dynamic aspect of ecology. *Ecology*, **23**, 399–418.

Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Res.*, **27**, 209–20.

Matthey, W. (1965). Signification de la butte à sphaignes dans l'évolution de la haute tourbière. *Comptes Rendus Soc. Biogéog.*, **371–2**, 119–30.

May, R.M. (1984). An overview: real and apparent patterns in community structure. In: *Ecological Communities: Conceptual Issues and the Evidence* (D.R. Strong Jr, D. Simberloff, L.G. Abele and A.B. Thistle, eds) Princeton University Press. pp. 3–16.

Plinski, M. and Magnin, E. (1979). Analyse écologique du phytoplancton de trois lacs des Laurentides (Québec, Canada). *Canadian J. Bot.*, **57**, 2791–9.

Quinn, J.F. and Dunham, A.E. (1983). On hypothesis testing in ecology and evolution. *Am. Naturalist*, **122**, 602–17.

Siepel, H. (1989). Objective selection of indicator species for nature management. Comptes–rendus du Symposium 'Invertébrés de Belgique', pp. 415–18.

Smouse, P.E., Long, J.C. and Sokal, R.R. (1986). Multiple regression and correlation extensions of the Mantel test of matrix correspondence. *System. Zool.* **35**, 627–32.

Southwood, T.R.E. (1987). The concept and nature of the community. In: *Organization of Communities: Past and Present* (J.H.R. Gee and P.S. Giller, eds) Blackwell Scientific Publications, Oxford. pp. 3–27.

Stevenson, A.C., Birks, H.J.B., Flower, R.J. and Battarbee, R.W. (1989). Diatom-based pH reconstruction of lake acidification using canonical correspondence analysis. *Ambio*, **18**, 228–33.

ter Braak, C.J.F. (1986). Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. *Ecology*, **67**, 1167–79.

ter Braak, C.J.F. (1987a). The analysis of vegetation-environment relationships by canonical correspondence analysis. *Vegetatio*, **69**, 69–77.

ter Braak, C.J.F. (1987b). Ordination. In: *Data Analysis in Community and Landscape Ecology* (R.H.G. Jongman, C.J.F. ter Braak and O.F.R. van Tongeren, eds) PUDOC, Wageningen. pp. 91–173.

ter Braak, C.J.F. (1988a). Partial canonical correspondence analysis. In: *Classification and Related Methods of Data Analysis* (H.H. Block, ed.) North Holland Press, Amsterdam. pp. 551–8.

ter Braak, C.J.F. (1988b). CANOCO – an extension of DECORANA to analyze species-environment relationships. *Vegetatio*, **75**, 159–60.

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

Vannier, G. (1985). Mode d'exploitation et de partage des ressources alimentaires dans le système saphrophage par les microarthropodes du sol. *Bull. d'Écol.*, **16**, 19–34.

Wauthy, G., Noti, M.-I. and Dufrêne, M. (1989). Geographic ecology of soil oribatid mites in deciduous forests. *Pedobiologia*, **33**, 399–416.

West, C.C. (1984). Micro-arthropod and plant species associations in two subantarctic terrestrial communities. *Oikos*, **42**, 66–73.

Wiens, J.A. (1984). On understanding a non-equilibrium world: Myth and reality in community patterns and processes. In: *Ecological Communities: Conceptual Issues and the Evidence* (D.R. Strong Jr, D. Simberloff, L.G. Abele and A.B. Thistle, eds) Princeton University Press. pp. 439–57.

Whittaker, R.H. (1956). Vegetation of the Great Smoky Mountains. *Ecol. Monographs*, **26**, 1–80.

Discussion

DONALD E. MYERS

Department of Mathematics, University of Arizona

1. Overview

Spatial structure, spatial correlation, spatial dependence, spatial heterogeneity and spatial variability are all terms that are used to denote some form of known or presumed relationship between a variate of interest and geographical positioning. That is, the data is spatially located in one of several possible senses. In some instances the variable represents a value for a region and in others it may represent a measure at a point or small area/volume. If the variable of interest is additive then these two perspectives can be merged. It will make some difference whether the variable(s) are categorical, nominative or continuous. In most cases the objective is to quantify the spatial structure or spatial variability. Spatial structure is also related to interpolation methods. Because both of these aspects are central to this paper it is useful to consider other methods for quantifying spatial structure and alternative interpolation methods.

There are at least two ways to characterize spatial structure, one of which is to explicitly incorporate spatial coordinates in the quantification of the spatial correlation. The authors have used a second approach, implicit quantification by a multivariate analysis technique, canonical correspondence analysis (CCA). While CCA does not specifically relate geographical position or distance to correlations it has the potential for identifying factors contributing to the spatial variability. This is especially useful when those factors are not directly observable or measurable. Similar or related techniques include principal components analysis, factor analysis and correspondence analysis. Ordinary, i.e., non-canonical, correspondence analysis was first introduced for categorical variables but has been extended and applied to continuous environmental data by a number of authors. These are all data driven techniques in that there is no model assumed a priori in order to perform the analysis. One disadvantage is that the method does not incorporate an adjustment for the support of the sample; in this particular application the support would correspond to the shape and volume of the cores. In general, variability will decrease as the

support is increased especially in relationship to the geographical extent of the area sampled. Another disadvantage is that the multivariate characterization of spatial structure does not incorporate directional dependence.

2. Variograms

Alternatively, spatial correlation can be quantified by the use of a (spatial) autocorrelation function. The variogram, first introduced in the context of ore reserve estimation problems in mining, has certain advantages over the autocorrelation function but is essentially comparable. The variogram explicitly relates spatial correlation of a variable to itself in terms of the separation vector between sample locations; it is also adaptable to changes in sample support. A further advantage is that the variogram can then be used in the spatial interpolation stage. Estimation and modeling of variograms is described in a number of standard references on geostatistics. The variogram also can incorporate directional dependence.

3. Inverse distance weighting

Contour maps, using irregularly spaced data, are produced in two stages. First the data is interpolated to a regular grid and then the contouring algorithm produces the contour lines from the gridded data. There are a variety of techniques that can be used to interpolate from irregularly spaced data (locations) to a grid. Inverse distance weighting, IWD, was used by the authors. While it is common to use the square of the distance in determining the weights, there is no intrinsic reason for doing so. As was shown in Kane *et al.* (1982) the results may or may not be sensitive to the choice of the exponent. IWD has a number of disadvantages. First of all it only incorporates the spatial correlation between individual sample locations and the location to be interpolated. Secondly IWD only incorporates distance and not direction in assigning weights to sample data. It does not incorporate non-punctual sample support and does not provide for direct estimation of spatial averages. Finally it is not an exact interpolator. It does have the advantage of not requiring any model assumptions. While the geostatistical interpolator (kriging) does require some model assumptions it does not have the disadvantages noted for IWD. For an overview of geostatistical methods and available public domain software see Myers (1991).

4. Interpolation and contouring

Interpolation methods nearly always result in smoothing the data and hence reduce the variability. In appraising the goodness of a contour map one may be tempted to assume smoothness is a desirable characteristic but this may also be misleading. Some measures of the interpolation errors would provide a better gauge of the goodness of a contour map. In particular when the data locations are irregularly spaced the contour map will not be equally reliable in all parts. Unfortunately IWD does not incorporate any measure of the interpolation errors and hence does not provide a measure of goodness for the resulting contour maps. When contour maps are produced from irregularly spaced data the reliability of any portion of a map is directly related to the density of the data locations used in producing the map as well as the interpolation method. Geostatistical techniques have the advantage of providing a measure of the variability of the interpolation errors; other well-known techniques such as thin-plate splines are incorporated as special cases of the

geostatistical interpolators. Geostatistical techniques have been extended to multivariate data and incorporate intervariable as well as spatial correlations.

Additional references

Kane, V., Butz, T., Begovich, C. and Myers, D.E. (1982). Interpretation of Regional Geochemistry. *Comp. Geosci.*, **8**(2), 117–36.

Myers, D.E. (1991). Interpolation and Estimation with Spatially Located Data. *Chemomet. Intell. Lab. Syst.*, **11**, 209–28.

Myers, D.E. (1992). Correspondence Analysis applied to Environmental data sets. In: *Multivariate Environmental Statistics* (G.P. Patil and C.R. Rao, eds) North Holland/Elsevier Science Publishers, Amsterdam. (To appear).

Discussion

H.J.B. BIRKS

Botanical Institute, University of Bergen

As an ecologist, I was particularly pleased to read Borcard and Legendre (1994) (hereafter B&L) for several reasons. First, it provides a novel and exciting way of identifying and exploring, for the purposes of hypothesis-generation, possible causes of ecological variation operative at a variety of spatial scales. Second, it brings some ecological and geographical life into the display of ordination results, so frequently presented in ecological papers as rather lifeless two-dimensional scatter plots of site scores on axes 1 and 2, etc. Third, it shows that ordination and contour mapping, commonly regarded by many ecologists as totally separate activities (but see, for example, MacDonald and Waters, 1988) can be usefully combined to provide new, effective, and revealing displays of ecological patterns. Fourth, B&L is a fine example of empirical environmental modelling in community ecology. It starts with a simple ecological species-substrate model and progressively improves on this by incorporating further explanatory variables to generate new and increasingly more precise and realistic ecological hypotheses. It is thus a clear example of 'the method of successive approximation' in statistical ecology.

The idea of using constrained ordination techniques such as canonical correspondence analysis or redundancy analysis to detect spatial gradients in ecological data was first presented by ter Braak (1987) and developed by Legendre (1990) and Hill (1991) with the use of quadratic or cubic terms of the basic geographical coordinates. Borcard *et al.* (1992) exploited this type of analysis, along with ter Braak's (1988) partial constrained ordination techniques, to decompose the variance in species-abundance data into four independent, additive components – a purely environmental component (fraction *a* in Figure 4 of B&L), a spatially covarying environmental component (fraction *b*), a purely spatial component (fraction *c*), and an unexplained component (fraction *d*). This general variance-partitioning procedure for multivariate ecological 'response variable-type' data has wide potential applicability in statistical ecology (Legendre, 1993). For example, it has now been used to decompose the variance of plant abundances over a 10-year-period into spatial and temporal components (ter Braak and Wiertz, 1993), the variation in late-glacial-pollen stratigraphical data from several sequences into within-sequence temporal and between-sequence

spatial components (Ammann *et al.*, 1993), the variances in different biostratigraphical data-sets into long-term and short-term temporal fractions (Lotter and Birks, 1994), and the total variation in plant-distributional data into modern ecological and historical components (Birks, 1994).

In many fine-scale ecological investigations, such as the one presented by B&L, the purely environmental fraction *a* may largely reflect local, site-scale ecological patterns determined by local environmental factors. In contrast, the spatially covarying environment component *b* may relate more to regional-scale ecological patterns influenced by environmental variables that covary geographically when space is modelled as a linear, quadratic, or cubic trend-surface. It is less clear what fraction *c*, the purely spatial fraction, might reflect ecologically.

It is not uncommon, as here, to find that a linear combination of geographical coordinates 'explains' the biological data remarkably well, without considering any specific direct or indirect ecological variables (see also Hill, 1991; Birks, 1994). This suggests that species may share a common spatial structuring with important explanatory environmental variables. This highlights a major problem in the analysis of spatially-arranged ecological data (note that nearly all ecological data are spatially arranged!), namely that spatial geographical and regional environmental variables are often so highly correlated that they may be inseparable statistically (e.g. Legendre and Fortin, 1989; Legendre and Troussellier, 1988). What is surprising in B&L's results is that as much as 12.2% of the variance (fraction *c*) in their mite data is 'explained' by space independent of ecological factors such as moisture, micro-topography, shade, or substrate type. By comparison, local environmental variables independent of any spatial structure (fraction *a*) only capture 13.7% whereas the spatially covarying regional environmental factors (fraction *b*) explain 31% of the biological variance. A future task for community ecologists is thus to generate (and subsequently test, if at all possible) hypotheses about the likely processes influencing the 12.2% pure spatial and, of course, the 43% unexplained fractions. B&L present in Table 3 possible rather generalised causal factors that could be invoked to explain the four variance components. I found their possible causes for the 12% pure spatial and 43% unexplained variance components under their general historical dynamics and environmental control and biotic models vague and rather unconvincing, but that may simply reflect my total ignorance of oribatid mites and their ecology! As a palaeo- and historical ecologist I was, however, comforted to see that historical dynamics are included in their long list of possibilities. It is a constant source of surprise to me how many ecologists persist in believing that 'history is bunk!'

The challenge is how to refine the interpretations of components (c) and (d). For this, further exploration of B&L's data is needed. I feel that additional contoured maps for the different components (*a*, *b*, *c*) shown in Figures 4a, 4b, and 4c are needed to help explore the form of the spatial patterns more fully. B&L and Legendre (1993) suggest that fraction *c* might reflect spatial population- or community-based processes, unmeasured abiotic environmental factors, or some form of habitat disturbance in the past that is reflected today by the spatial structure of fraction *c*. I would like to see the spatial patterns of this fraction before evaluating these hypotheses.

I would also be fascinated to see contoured maps of the squared residual distance and the percentage fit for individual sites on the first few ordination axes, one of the ordination diagnostics optionally provided by ter Braak's (1990) CANOCO 3.1x program. Perhaps not all samples are equally well modelled in the (partial) constrained ordinations. It would certainly be important to know and hopefully the readers' mind could be reassured about the robustness of B&L's results.

Turning to fraction *d*, the unexplained portion, why not map it too. If Borcard and Legendre did a partial correspondence analysis ordination (partial the effects of the various environmental and spatial variables as co-variables prior to an ordinary unconstrained ordination – see Bakker *et al.* (1990) for an example of such a partial ordination), it would then be possible to map fraction *d* as the site ordination scores, based on the residual variation in the mite data that is unexplained by

the available spatial and environmental variables. Such maps might aid in suggesting ecological hypotheses about the causes of the largest single component of the variance in B&L's oribatid mite data, namely the unexplained 43% fraction. Are there any coherent spatial patterns in this fraction or are there seemingly random and/or spatially incoherent variations from site to site? Mapping the spatial arrangement of this fraction might thus help to suggest whether it primarily represents unmeasured but localised factors such as abiotic variables (e.g. silt content, nitrogen levels, redox potential of the peat, etc.) and/or unmeasured (and unmeasurable!) biotic processes (e.g. behavioural responses, predation, etc.) whose spatial structure cannot be modelled by a cubic trend-surface, or whether it is largely stochastic variation.

Additional references

Ammann, B., Birks, H.J.B., Drescher-Schneider, R., Juggins, S., Lang, G. and Lotter, A.F. (1994). Patterns of variation in late-glacial pollen stratigraphy along a north-west-south-east transect through Switzerland – a numerical analysis. *Quaternary Sci. Rev.* (Submitted).

Bakker, C., Herman, P.M.J. and Vink, J. (1990). Changes in seasonal succession of phytoplankton induced by the storm-surge barrier in the Oosterschelde (S.W. Netherlands). *J. Plankton Res.*, **12**, 947–72.

Birks, H.J.B. (1994). Is the hypothesis of survival on glacial nunataks necessary to explain the present-day distributions of Norwegian mountain plants? *Phytocoenol.* (Submitted).

Hill, M.O. (1991) Patterns of species distribution in Britain elucidated by canonical correspondence analysis. *J. Biogeography*, **18**, 247–55.

Legendre, P. (1990). Quantitative methods and biogeographic analysis. In: *Evolutionary Biogeography of the Marine Algae of the North Atlantic* (D.J. Garbary and R.R. South, eds) Springer-Verlag, Berlin. pp. 9–34.

Legendre, P. (1993). Spatial autocorrelation: Trouble or new paradigm? *Ecology*, **74**, 1659–73.

Lotter, A.F. and Birks, H.J.B. (1994). The impact of the Laacher See Tephra on terrestrial and aquatic ecosystems in the Black Forest (Southern Germany). *J. Quaternary Sci.* (Submitted).

MacDonald, G.M. and Waters, N.M. (1988). The use of most predictable surfaces for the classification and mapping of taxon assemblages. *Vegetatio*, **74**, 125–35.

ter Braak, C.J.F. (1990). CANOCO – a FORTRAN program for CANOnical Community Ordination by [partial] [detrended] [canonical] correspondence analysis, principal components analysis, and redundancy analysis, version 3.10. Microcomputer Power, Ithaca, New York.

ter Braak, C.J.F. and Wiertz, J. (1994). On the statistical analysis of vegetation change: changes in a Junco-Molinion grassland after ten years with water withdrawal and soil acidification. *J. Veg. Sci.* (Submitted).

Rejoinder

PIERRE LEGENDRE and DANIEL BORCARD

Since we wrote the original paper (Borcard *et al.*, 1992) describing the method for partitioning the variance among spatial and environmental components, as well as the one published in the present issue of the *Journal of Environmental Statistics* emphasizing various aspects of the interpretation of the components and the need to look at maps to carry out this task in an intelligent way, the method has been tried by us and by others on a variety of data sets and ecological situations;

several new, exciting applications in papers that have just been submitted for publication are brought to our attention for the first time in Professor Birks' Discussion (above). Some of these results were reported during the annual symposium of G.R.I.L. (*Groupe de recherche interuniversitaire en limnologie et en environnement aquatique*) held on February 25 and 26, 1993, at the *Station de biologie des Laurentides* of Université de Montréal. This meeting was the opportunity to raise (Rasmussen, 1993) and discuss problems that may arise in the use of the method and to look for solutions; some of these discussions cross-check with the points raised in Professor Myers' and Birks' Discussions.

It is important to understand that the spatial polynomial, that we proposed as a method to explicitly incorporate the spatial structure into our variance partitioning model, can capture only the large-scale structures of the dependent variable under study if the polynomial of the geographic co-ordinates (x and y) remains of low order, as it should. Indeed, this is but a multivariate application of the old method of trend surface analysis first advocated by Student (1914)! When we first thought of using this polynomial in the context of partial regression or partial canonical variance partitioning, we were happy to have found at least one way of explicitly incorporating spatial structures into ecological models. We never pretended that this method was appropriate for all situations, being conscious of the fact that small-scale structures are not captured by this polynomial. The method, however, seemed relevant to a whole category of problems that are of concern to population and community ecologists, as shown by the examples that we analyzed in the two above-mentioned papers.

The first analyses of a new data set should involve 'structure functions' and maps of various kinds; Legendre and Fortin (1989) have summarized these approaches. Structure functions, advocated by Professor Myers in his Discussion, include variograms and correlograms, univariate or multivariate; this depends on whether one is looking at the spatial behavior of a single dependent variable at a time, a single summary variable (principal axis, diversity measure, and so on), or at the structure of the whole dependent multivariate data table of interest through multivariate variograms (Bourgault and Marcotte, 1991) or multivariate Mantel correlograms (Sokal, 1986; Legendre and Fortin, 1989). Structure functions allow to decompose spatial variability as a function of scale (distance classes) and to get an understanding of the scales at which important variability occurs in the data set. In the same way, univariate maps of single variables may be obtained by one of the contour mapping methods discussed by Professor Myers; maps of multivariate data may be obtained either through constrained clustering (see Legendre and Fortin, 1989, for a review; also Bourgault *et al.*, 1992), by co-kriging (Matheron, 1970; Myers, 1982, 1983) or vector kriging (Young, 1987), or by separately mapping multivariate ordination axes. If these preliminary studies show that variability is mostly large-scale, or if the hypothesis of interest focuses on large scales (given the size of the study area), then the spatial polynomial may be useful to express it in the model. (In other cases, the large-scale fraction of variability may be trivial and the worker may have to extract it from the data in order to pursue the analysis on the residuals of the trend-surface equation.) If it is found that small-scale autocorrelation is the fraction of interest which ought to be included in the model, other ways should be found to take it into account in the analysis (below).

Geometric anisotropy can easily be identified by computing directional variograms or correlograms. Notice however that it does not affect the fraction of variability extracted by the spatial polynomial. Proper identification of the various forms of anisotropy is important in contour mapping by the geostatistical method of kriging, as well as for the proper understanding of the spatial process having generated the observed structures.

The terms of the spatial polynomials originally proposed by Legendre (1990) are not independent of one another. If the interpretation of the regression or canonical coefficients relating these

terms to the community structure is of special interest, orthogonal polynomials should be used instead of the classical polynomials. Orthogonal polynomials explain exactly the same amount of variability as the classical polynomials from which they are built.

Small-scale spatial structures could be incorporated into models by various methods whose development is now considered in P.L.'s laboratory; large-scale trends should have been extracted from the data first. The first such method (Niyonsenga, 1993) consists of empirically estimating the autocorrelation function and incorporating it as an explicit among-point (among-locality, $p \times p$) autocorrelation matrix in the computation of the regression or canonical analysis describing the relation between the dependent and independent variables (Ripley, 1981; Haining, 1990).

In the second method that we have experimented with, the range of distances among the p points is divided in distance classes. For each class, a locality-by-locality ($p \times p$) weight matrix NM is written, containing values $(1/p_i)$ for pairs of localities that are located within the given distance class, and 0's elsewhere; if p_i is the number of first neighbours (or second neighbours, etc.) of object i, then $(1/p_i)$ is the weight given to each of its neighbours under the assumption of equal contribution. Each such matrix can be called a Neighbourhood Matrix of order 1 (NM_1 for first-distance neighbours), order 2 (NM_2 for second-distance neighbours), etc.; the large-distance neighbour matrices, corresponding to distances that exceed the range of the variogram or correlogram, can be discarded. Postmultiplying each such matrix by the vector $y(p \times 1)$ or matrix $Y(p \times n)$ of observed values for the n dependent variable(s) produces a vector $y'(p \times 1)$ or matrix $Y'(p \times n)$ giving the contribution, to each location, of the values of the dependent variable(s) from neighbouring sites, according to the autocorrelation model:

$$NM_i \cdot y = y_i'$$ (1)

or

$$NM_i \cdot Y = Y_i'$$ (2)

These vectors or matrices can now be used in a Borcard *et al.* (1992) partitioning as explicit expressions of the effect of small-scale autocorrelation. Preliminary experimentation shows that, as expected, the fraction of variability extracted using this method only partially overlaps the spatial component extracted using the spatial polynomial and is thus complementary to it.

The need to map the various fractions of variation is emphasized in both Birks' and Myers' Discussions. Contour mapping is easily done for the spatially-structured fractions b and c. The interpretation of fraction b is critical since it may involve two opposite processes, as explained in Table 3 of the main paper: it may correspond either to a structuring of the dependent variable which is due to the fact that the measured causal environmental variables are themselves spatially structured, or to a structuring by other spatially-structured variables or processes, not included in the analysis, which are causal to both the dependent and independent environmental variables. Consider for example a river with a pollution gradient and a gradient in fish assemblages; sampling has been carried out at various locations along the river. One is likely to find that a large fraction of the community-pollution relationship will fall in fraction b. Is this a response of the fish to the spatially-structured pollution gradient, or would we not find the same fish spatial structure even without the pollution effect (in which case the structuring factor is the river continuum itself)? Professor Birks is right, of course, when he emphasizes the need for a thorough analysis of the maps of the various components of fraction c. These maps are the first opportunity for ecologists to examine a spatially structured, yet unexplained component of variation, and they may be tremendously helpful in generating new ideas and hypotheses about the large-scale phenomena that may have generated these structures. One should not be surprised, however, to find significant spatial variation remaining in the unexplained fraction c, given the limits of all field ecological investigations.

Mapping the non-spatially-structured fractions (*a*: non-spatially–structured environmental variation) and (*d*: residual), is not an easy task: how can one map something which is not spatially structured? We have argued quite a bit about that among ourselves over the past few months, and with colleagues who were analyzing data using our method. We have come to the conclusion that there is a need for such maps, but that these maps should not give a false impression of spatial continuity. They could be drawn using either three-dimensional pin charts representing the values measured at the sampling locations only, or influence (Thiessen, or Voronoï) polygon charts in which all points closer to an actual sampling locality receive the same value as that locality, thus creating surfaces with sharp discontinuities. But Professor Birks is right here: maps may help identify important but unmeasured locally-operating environmental or historical factors that are responsible for this residual variation.

In this short discussion, we have shown where the spatial polynomial stands in a general process aimed at explicitly expressing spatial relationships in our ecological models. A preliminary study of the data set is necessary to assess which part of the variability is small- or large-scale, and whether the large-scale component is of interest to the ecological hypotheses under study, or whether it is trivial and should be extracted from the data in order to allow modeling the small-scale effects. Directions of research have been proposed in order to incorporate small-scale spatial structures into models. The next step needed to turn this approach into a flexible statistical tool for environmental modeling is to break the three-matrix ceiling (one dependent and two independent matrices describing environmental and spatial components) and learn how to combine the information from several independent variable matrices in such a model; typically these matrices may represent several types of environmental components (biotic and abiotic, for instance) as well as small- and large-scale spatial effects. At the present moment this requires clumsy combinations of all matrices of interest into various triplets, with possible loss of precision during the partitioning, since the method available is limited to the study of three matrices only at a time.

Additional references

Bourgault, G. and Marcotte, D. (1991). Multivariable variogram and its application to the linear model of coregionalization. *Math. Geol.*, **23**, 899–928.

Bourgault, G., Marcotte, D. and Legendre, P. (1992). The multivariate (co)variogram as a spatial weighting function in classification methods. *Math. Geol.*, **24**, 463–78.

Haining, R. (1990). *Spatial Data Analysis in the Social and Environmental Sciences.* Cambridge University Press, Cambridge.

Legendre, P. (1990). Quantitative methods and biogeographic analysis. In: *Evolutionary Biogeography of the Marine Algae of the North Atlantic* (D.J. Garbary and G.R. South, eds) NATO ASI Series, Vol. G22, Springer-Verlag, Berlin. pp. 9–34.

Matheron, G. (1970). La théorie des variables régionalisées, et ses applications. *Les Cahiers du Centre de Morphologie Mathématique de Fontainebleau*, Fascicule 5.

Myers, D.E. (1982). Matrix formulation of co-kriging. *Math. Geol.*, **14**, 249–57.

Myers, D.E. (1983). Estimation of linear combination and co-kriging. *Math. Geol.*, **15**, 633–7.

Niyonsenga, T. (1993). Modélisation spatiale: Faudrait-il tenir compte de l'autocorrélation spatiale dans les analyses canoniques? Paper presented at the 3rd Annual Symposium of G.R.I.L., Station de biologie des Laurentides, Université de Montréal, February 25, 1993.

Rasmussen, J. (1993). Space: The final frontier? Paper presented at the 3rd Annual Symposium of G.R.I.L., Station de biologie des Laurentides, Université de Montréal, February 25, 1993.

Ripley, D.B. (1981). *Spatial Statistics.* Wiley, New York.

Sokal, R.R. (1986). Spatial data analysis and historical processes. In: *Data Analysis and Informatics, IV*, (E. Diday *et al.*, eds.) Proc. Fourth Int. Symp. Data Anal. Informatics, Versailles, France, 1985, North-Holland, Amsterdam. pp. 29–43.

Student [Gosset, W.S.] (1914). The elimination of spurious correlation due to position in time or space. *Biometrika*, **10**, 179–80.

Young, D.S. (1987). Random vectors and spatial analysis by geostatistics for geotechnical applications. *Math. Geol.*, **19**, 467–79.