

To be published in *Ecology*

Preprint version: January 6, 2005

A generalized, variogram-based framework for multi-scale ordination

Pierre COUTERON¹ and Sébastien OLLIER²

1 Ecole Nationale du Génie Rural des Eaux et des Forêts / UMR botAnique et bioinforMatique de l'Architecture des Plantes (AMAP), Boulevard de la Lironde, TA40/PS2, 34398 Montpellier Cedex 05, France, and French Institute of Pondicherry, 11 Saint Louis Street, Pondicherry 605001, India

2 UMR CNRS 5558, Laboratoire de Biométrie et Biologie Evolutive, Université Claude Bernard Lyon 1, 69 622 Villeurbanne Cedex, France.

Present address for direct correspondence

Pierre COUTERON
French Institute of Pondicherry, 11 Saint Louis Street, Pondicherry 605001, India
Tel: +91 413 2334 168 , Fax: +91 413 2339 534
Pierre.couteron@ifpindia.org

Abstract

Multi-scale ordination (MSO) deals with potential scale-dependence in species assemblages, by studying how results from multivariate ordination may be different on different spatial scales. MSO methods were initially based on two-term local covariances between species, and therefore required sampling designs composed of adjacent quadrats. A variogram-based MSO, applicable to very diverse sampling designs has recently been introduced by H.H. Wagner (2003, 2004). This refers to Principal Component Analysis, Correspondence Analysis and derived "two-table" (also called "direct") ordination methods, i.e., Redundancy Analysis and Canonical Correspondence Analysis.

In this paper we put forward an enlarged framework for variogram-based MSO which relies on a generalized definition of inter-species covariance and on matrix expression of spatial contiguity between sampling units. This enables us to provide distance-explicit decompositions of variances and covariances (in their generalized meaning), that are consistent with many ordination methods in both their single- and two-table versions. A spatially explicit apportioning of diversity indices is proposed for some particular definitions of variance. Referring to two-table ordination methods allowed the multi-scale study of residual spatial patterns after factoring out of available environmental variables. Some aspects of the approach are briefly illustrated with vegetation data from a neotropical rainforest in French Guiana.

Key words: diversity apportioning; species assemblages; multi-scale ordination; multivariate geostatistics; canonical correspondence analysis; spatial contiguity; tropical rainforest; variogram.

Introduction

Determining to what extent multi-species patterns of association may be different on different spatial scales is obviously a central issue in ecology (Levin 1992). The concern to integrate space in numerical studies of inter-species association has led to the development of a method of multi-scale ordination (MSO; Ver Hoef and Glenn-Lewin 1989). This requires data from continuous sampling designs (e.g. belt transects) since it is based on the computation of two-term local covariances between species (Greig-Smith 1983). In two recent papers, H.H. Wagner (2003, 2004) proposed a new method for MSO based on the variogram (Wackernagel 1998) thereby allowing the use of data collected by means of very diverse sampling designs. This insightful approach was proposed for two usual ordination methods, i.e., Principal Component Analysis (PCA) and Correspondence Analysis (CA). Extension to "direct" ordination methods (Legendre and Legendre 1998), using two data tables, such as Redundancy Analysis (RDA, relating to PCA) and Canonical Correspondence Analysis (CCA, relating to CA) was also proposed by Wagner (2004).

This most recent contribution is a considerable step forward since CA is generally preferred to PCA for the study of inter-species associations. CA is a very popular and powerful method that positions species and sites along common ordination axes, by applying the same centering and weighting options to rows and columns of the site by species table. However, in spite of several attractive properties, there is no reason to consider CA as being automatically the most appropriate ordination method whatever the characteristics of the data and the aim of the study (Gimaret-Carpentier et al. 1998). Several alternatives to CA, with distinct properties, can be defined by changing weighting options for either sites or species. For instance, Pélissier et al. (2003) demonstrated that changing species weighting, i.e. placing varying degrees of emphasis on scarce species, could be used to define three ordination methods (including CA) and that each was consistent with one classical diversity index (richness, Shannon's, Simpson's). On the other hand, Dolédec et al. (2000) used a uniform weighting of sites to derive an alternative to CCA with interesting properties for the separation of species niches. Hence, it would be preferable for ecologists to become aware of the potential adaptability of both single- and two-table methods of ordination to their specific aims and to the characteristics of their data. Such adaptability should also encompass the emerging field of spatially explicit ordinations.

The paper presented here was triggered by the pioneering work conducted by Wagner (2003, 2004) but aims to define a broader framework for variogram-based multi-scale ordinations. We intend to demonstrate that it is possible to partition by distance the results of very diverse ordination methods, as defined by re-scaling and weighting options for the rows and columns of the data tables. To do so, we will introduce a generalized definition of covariance between species which encompasses several ordination methods while being amenable to scale-explicit decompositions. Consistency with the additive decomposition of common diversity indices (Pélissier et al. 2003; Couteron and Pélissier 2004) will be highlighted, while referring to methods of two-table "direct" ordination will allow the explicit analysis of residual spatial patterns after the factoring out of some environmental variables. This aspect will be emphasized in a brief illustration based on vegetation data from a neotropical rainforest. In this report, we have chosen to keep mathematical developments to a minimum while providing a complete treatment in matrix form in an appendix published in the Electronic Ecological Archives. Computer programs are freely available from the first author (Matlab®

version) or on <http://pbil.univ-lyon1.fr/CRAN/> (R [Ihaka and Gentleman 1996] version integrated in the ade4 package).

A generalized definition of covariance

Data tables containing counts of individual organisms by sampling sites (say "quadrats") and taxa (usually species) are both a central and general feature of ecological studies. Let us consider such a table, based on N sampled individuals, for which f_{ai} is the total number of individuals belonging to species i ($1 < i < S$) that were counted in quadrat a ($1 < a < Q$). Let p_{ai} be the corresponding relative frequency ($p_{ai} = f_{ai}/N$) while p_{a+} and p_{+i} are the relative frequencies for quadrat a and species i , respectively. We have introduced our topic with explicit reference to counted individuals, though the above parameters remain meaningful as long as f_{ai} is a non-negative value (biomass measurements, semi-quantitative indices of abundance, presence/absence, ...) expressing the abundance of species i in quadrat a .

Ordination methods such as correspondence analysis CA and various versions of PCA (ter Braak 1983) are the usual tools employed to analyze quadrats by species tables. Central to all these methods is the application of singular values decomposition (svd), also called eigenanalysis, to a square S by S matrix, which is the usual variance-covariance matrix, \mathbf{C} , for the species-centered (non-standardized) PCA and which is another matrix \mathbf{Q}^2 in the case of CA (see Legendre and Legendre 1998:453 and Wagner 2004 for details). In \mathbf{Q}^2 , terms on the diagonal are homologous to variances and are proportional to the portions of the total chi-square of the data table (Legendre and Legendre 1998:452) that are attached to each of the S species. Off-diagonal terms are homologous to the usual pairwise covariances and measure to what extent two arbitrary species may conjointly depart from expected abundance values.

We can see matrices \mathbf{C} and \mathbf{Q}^2 as nothing more than special cases of a square matrix $\mathbf{G}_T = [g_{ij}]_{1 \leq i \leq S, 1 \leq j \leq S}$ based on an appropriate generalization of the notions of species variance (diagonal values) and covariance (off-diagonal values). This generalized measure of covariance is, for two arbitrary species i and j :

$$g_{ij} = \frac{1}{2} \sum_{a=1}^Q \sum_{b=1}^Q (x_{ai} - x_{bi}) \sqrt{w_i w_j} (x_{aj} - x_{bj}) \delta_a \delta_b \quad (1)$$

variance being a special case with $i=j$.

Here, w_i weighs the influence of species i , while δ_a and δ_b are the weights given to quadrats a and b , respectively; x_{ai} denotes any measure of abundance of species i in quadrat a that can be derived from the initial value f_{ai} via re-scaling options (Table 1).

The choice of weighting options is a central yet often eluded question when using multivariate techniques, since weighting along with re-scaling and centering defines the nature of the distance between quadrats and, for some methods, also between species. Moreover, the choice of weighting options relates to very practical questions concerning, for instance, the influence that it seems meaningful to confer to a particular species in the definition of a multi-specific assemblage, or to a given quadrat in the investigation of an ecological gradient. Addressing such questions means that the biogeographic context must be taken into account (e.g., are there many scarce species? How abundant are the most frequent species?) along with the sampling design (does it give a fair estimate of species abundance in a region?) and, for two-table methods, the nature of the ecological gradients under study (are there strong limiting factors or threshold effects?). More detailed discussions on the consequences of weighting can

be found in Dolédec et al. (2000, regarding quadrats in direct gradient analysis) and in Pélissier et al. (2003, regarding species).

Combining re-scaling and weighting options opens up a wide selection of ordination methods and associated properties. Some examples, based on published methods, are presented in Table 1, but other possibilities are obviously imaginable. The presentation of classical ordination methods in terms of weighting of rows and columns was introduced by Escoufier (1987) and was used by Sabatier et al. (1989) and Dolédec et al. (2000) for several single- and two-table methods (including CA, CCA and classical versions of PCA and RDA). Pélissier et al. (2003) used this presentation to compare the properties of the three methods corresponding to option IV in Table 1, and to investigate their relationship with diversity measures. All these authors based their presentation of the methods on a species-centered version of the data table containing differences between individual observations, x_{ai} , and the δ_a -weighted mean value, \bar{x}_i , found for each species. Alternatively, in Eq. 1, we use all pairwise differences between observations to compute variance and covariance. The equivalence of the two approaches is explained in the appendix (see Eq. A.5 to Eq. A.10).

Generalized spatial covariance

From Eq. 1, the contribution made by a given pair (a, b) of quadrats to the covariance between two species can be expressed as:

$$g_{ij}(a, b) = \frac{1}{2} (x_{ai} - x_{bi}) \sqrt{w_i w_j} (x_{aj} - x_{bj}) \delta_a \delta_b \quad (2)$$

This translates easily into a generalized version of cross-variograms ($i \neq j$) and variograms ($i = j$), namely:

$$\gamma_{Gij}(h) = \frac{1}{K(h)} \sum_{a, b | h_{ab} \approx h} g_{ij}(a, b) \quad (3)$$

where h is the central value of a given distance class, and where $K(h)$ is a scaling coefficient,

$$\text{such as: } K(h) = \sum_{a, b | h_{ab} \approx h} \delta_a \delta_b \quad (4)$$

Considering all species together leads to a generalized variogram of species composition ("generalized" since potentially relating to several ordination methods and distance metrics):

$$\mathcal{Y}_G(h) = \sum_i \mathcal{Y}_{Gii}(h) \quad (5)$$

Eq. 4 is a crucial point since $K(h)$ standardizes $\mathcal{Y}_G(h)$ in such a manner to equate its expected value (sill) with the total variance of the ordination method defined by weighting options. This is completely different from computing the usual experimental variogram from ordination scores, except for the special case of uniform quadrat weights (as in Wagner 2003) where $K(h)$ is proportional to the number, n_h , of pairs of quadrats relating to distance class h . Conversely, if quadrat weights are not uniform, scaling by $K(h)$ is the only manner to ensure that, whatever the distance class, the expected value of $\mathcal{Y}_G(h)$ is the total variance ("inertia") attached to matrix \mathbf{G}_T and computed from the sum of its diagonal elements. Note that such a property is not guaranteed by the manner in which Wagner (2004) defined her version (denoted as $\mathcal{Y}_Q(h)$) of the CA-related variogram since the corresponding scaling remains proportional to n_h despite the fact that quadrat weights are not uniform. The scaling by $K(h)$ is of particular interest if weightings of both species and quadrats are chosen as to relate to a

diversity measurement (option IV in Table 1). In this case, the trace of \mathbf{G}_T is the diversity among quadrats (Couteron and Pélissier 2004), which means that $\mathcal{Y}_G(h)$ measures the average beta-diversity between pairs of quadrats corresponding to distance class h . Equivalently,

$$VAR(a,b) = \sum_i g_i(a,b) \quad (6)$$

quantifies the contribution made by a given couple (a,b) of quadrats to beta-diversity. Some classical dissimilarity indices, such as Jaccard's or Sorensen's (Legendre and Legendre 1998:256) are often used to quantify beta-diversity, though these have no direct connection with either geostatistical tools or ordination methods. Conversely, Eqs.2 and 6 provide a family of dissimilarity indices some of which relate directly to both.

Variograms and cross-variograms of ordination axes

Regardless of the reference ordination method chosen, a generalized variance-covariance matrix, \mathbf{G}_h , is computed for each distance class h . To ensure efficient computations by any matrix-oriented programming language, as we did with Matlab® and R (Ihaka and Gentleman 1996), we introduced a matrix formulation of the method. It is based of a contiguity relationship (Thioulouse et al. 1995) consistent with the variogram, which considers two quadrats as "neighbors at scale h " if the distance between them is within the bounds of the class centered around h (see Appendix). Assuming that distance classes include all pairs of quadrats while being mutually exclusive, we demonstrated (see Appendix, Eq. A.13) that the matrices \mathbf{G}_h sum to \mathbf{G}_T , whatever the initial choice of the reference ordination method by weighting options.

The eigenvectors and eigenvalues originating from the singular values decomposition (svd) of \mathbf{G}_T can be partitioned with respect to distance classes (Appendix), as a generalization of the fundamental principle introduced by Ver Hoef and Glenn-Lewin (1989). But the complete variance-covariance matrix, \mathbf{F}_h , between eigenvectors can also be obtained (Appendix, Eq. A.13). Considering off-diagonal elements of \mathbf{F}_h , namely covariances at scale h between eigenvectors, is a new perspective in MSO which can be used to investigate the potential existence of a scale-dependent covariance between distinct ordination axes. This question, though ignored by most papers devoted to MSO, is closely related to the initial concern of Noy-Meir and Anderson (1971), namely that ordination results may substantially vary with spatial scales. This would mean, for example, that species displaying the most prominent variations of abundance may not be the same depending on the average distance between the quadrats, or that distinct species assemblages may be found for different distance classes. How can we test whether this is the case or not? One way would be to carry out an ordination for each of the \mathbf{G}_h matrices and compare the results, but this is likely to be cumbersome while objective criteria for the comparison are not straightforward to define. We propose a more efficient approach by constructing cross-variograms of eigenvectors from the off-diagonal values of \mathbf{F}_h matrices after appropriate scaling by $K(h)$. All these cross-variograms have an expectation of zero, since the eigenvectors of \mathbf{G}_T are globally uncorrelated, but some may have significant departures from this expectation on particular scales. (Of course, only the cross-variograms for the most prominent eigenvectors are to be analyzed.) If this is the case, it is possible to know at which scales it may be worthwhile carrying out specific ordination analyses via the svd of the corresponding \mathbf{G}_h matrices.

Taking environmental heterogeneity into account

If the species by quadrats table is accompanied by environmental variables assessed at the quadrat scale, it is advantageous to factor out the influence of such variables prior to analyzing the residual spatial patterns of species composition. Technically, this verifies whether some basic assumptions, such as "intrinsic" stationarity (used to interpret the empirical variogram), or independence of residuals (assumed to fit a linear model of species-environment relationship) are met by the data (see Wagner 2004 for an extensive discussion). In terms of ecological interpretation, it is judicious to see residual spatial patterns of community composition as predominantly shaped by biotic processes, such as species dissemination or species interactions (Wagner 2004) and as potentially informative on the scale at which such processes may operate.

Any two-table "direct" ordination starts from the decomposition of the quadrats by species table, \mathbf{X} , into an approximated table \mathbf{A} , modelled from environmental variables by a weighted linear regression and a residual table \mathbf{R} . Such a decomposition may be carried out in a manner consistent with a two-table version of any of the ordination methods mentioned in Table 1 (Sabatier et al. 1989; Pélissier et al. 2003) when the linear regression uses the quadrat weights defining the ordination. (For instance, defining table \mathbf{X} along with species and quadrat weights so as to make them consistent with CA means that an ordination on \mathbf{A} would be a CCA.) To study residual spatial patterns, it is possible to break down \mathbf{G}_R , i.e., the variance-covariance matrix computed from \mathbf{R} , into additive variance-covariance matrices, \mathbf{G}_{R_h} , each corresponding to a certain distance class, and on which a variogram-based multi-scale analysis can be based (see Appendix).

Brief illustration based on tropical rain forest data

We considered 7,189 trees (diameter at breast height above 10 cm) belonging to 59 species sampled in a lowland tropical rain forest of ca. 10,000 ha in French Guiana. The sampling design was based on 411 rectangular quadrats of 0.3 ha each, located at the nodes of a 400 m by 500 m grid. Environmental information at the quadrat scale was expressed by a synthetic nominal variable (12 categories) primarily based on topography and soil water regime (see Coutron et al. 2003 for details). Performing CA on the quadrats by species table showed two main floristic gradients corresponding to the second and third axes (CA2 and CA3). The first axis (CA1) resulted from the spurious occurrence of a scarce species (17 trees) in a particular quadrat and this illustrates a well-known drawback of CA. Results of the non-symmetric correspondence analysis (NSCA) were free from this problem, while the two main axes, NSCA1 and NSCA2, correlated strongly with CA3 and CA2 ($r = 0.76$ and $r = 0.84$, respectively) despite being defined from distinct species. For this data set, shifting emphasis from scarce to abundant species changed the hierarchy between the ordination axes, but the detection of two main floristic gradients proved robust with respect to species weighting. To go beyond these results established by a previous study (Coutron et al. 2003) we explicitly considered inter-quadrats distances by applying the generalized variogram-based MSO with CA and NSCA as reference ordination methods. First, we partitioned the total variance attached to each ordination axis (eigenvalue) among distance classes. Since diversity-related ordinations were used (Pélissier et al. 2003), it was the total among-quadrats diversity (*sensu* the species richness for CA or the Simpson-Gini index for NSCA) which was successively broken down with respect to main floristic gradients (eigenvalues) and distance classes.

The floristic gradient defined by CA2 and NSCA2 failed to show any obvious spatial pattern since variograms were found to waver between confidence envelopes, and this regardless of the reference ordination (Fig. 1-a and b). The study of residual patterns, after factoring out the 12 environmental categories (unconstrained ordinations on the residual variance-covariance matrix, \mathbf{R}) showed significant departures of the CA-based variogram for distances under 2 km. Such a change in the variogram stemming from the partialling out of the environmental variable typifies the complex interaction which can be expected between the environmental heterogeneity and spatial patterns of species assemblages. It also illustrates the advantage of studying such an interaction within a unified theoretical framework of multi-scale ordination since it enabled us to express in the same unit all kinds of results derived from a particular ordination method. This renders variograms of both initial and residual patterns directly comparable, i.e. a desirable property that could not have been achieved by the computation of classical variograms from ordination scores. The other floristic gradient (defined by CA3 and NSCA1) showed a strong spatial pattern that pointed toward non-stationarity (see Wagner 2004 for a detailed definition) since both initial variograms (Fig. 1-c and 1-d) continued to rise up to 8 km without reaching a sill. The variograms of the homologous axes provided by CA and NSCA after factoring out the qualitative environmental categories appeared to be very similar. This indicated that the observed spatial patterns relating to this floristic gradient were not determined by the spatial distribution of the environmental categories.

By separately analyzing spatial patterns of distinct ordination axes we have implicitly hypothesized the absence of any scale-dependent relationship between the ordination axes or, equivalently, the stability across scales of inter-species covariances (“intrinsic” covariances *sensu* Wackernagel, 1998). Such a hypothesis can be easily addressed by computing the cross-variograms between the ordination axes (from off-diagonal elements of matrices \mathbf{F}_h , Eq. A.18 in the Appendix). Only cross-variograms computed from the residual variance-covariance matrix (\mathbf{R}) are shown (Fig. 1-e and -f) since homologous cross-variograms from the initial data table were very similar. No scale dependence was observed between the two ordination axes given by NSCA (Fig. 1-f) and covariances between abundant species thus appeared to be stable across scales. This was not the case when the emphasis was placed on scarcer species by the use of CA since most values of the corresponding cross-variogram were outside the confidence envelopes (Fig. 1-e). Indeed, by diagonalizing the pooled variance-covariance matrices for distances under 4 km vs. distances above 4 km, we obtained two clearly distinct sets of species with high loadings on the ordination axes (results not shown). This result exemplified how scale-dependence may be detected by analyzing cross-variograms between ordination axes, while also illustrating the influence that species weighting may have: CA results proved scale-dependent though NSCA results did not.

Concluding remarks

In the above illustration, we deliberately restricted ourselves to some particular analyses that can be obtained from the generalized variogram-based MSO, but other kinds of analyses are clearly possible. For instance, it may be of interest to compare the spatial patterns of all individual species and identify scales on which some patterns may differ from others. This can be done by analysing, for instance by PCA, the table containing the generalized variograms of all species (prior standardization by variances of individual species as to have all sills equal to one is likely to be preferable). It may also be of interest to study the manner in which the distribution of the eigenvalues changes with scale, by analyzing the table

containing the generalized variogram of the eigenvalues. Any type of MSO can address these questions, but our unifying approach also provides a choice between ordination methods while offering links with diversity measurement and apportioning. As a consequence, future users may be able to select the particular ordination method (either direct or indirect) that best suits their data and aims.

Furthermore, the presentation of the method in matrix-form, with the use of contiguity matrices (Appendix), not only allows for efficient programming, but also opens up interesting methodological perspectives. Indeed, two-term local variances and covariances (TTLV/TTLC) on which the "classical" MSO relies (Ver Hoef and Glenn-Lewin 1989) have been also formulated using contiguity matrices (Di Bella and Jona-Lasinio 1996, Ollier et al. 2003). Such a formulation is obviously a sound basis, not only for a further generalization of the TTLV/TTLC-based MSO, as performed by us with the variogram-based MSO, but also for a thorough investigation of the respective properties of the two approaches.

Acknowledgements

We are indebted to D. Chessel (University of Lyon-1) for important insights into several topics mentioned in this paper, to R. Pélissier (IRD/UMR AMAP) for his valuable comments on preliminary drafts, and to H. H. Wagner for pertinent comments on the initial version.

Literature cited

- Couteron, P., and R. Pélissier. 2004. Additive apportioning of species diversity: towards more sophisticated models and analyses. *Oikos*, **107**: 215-221..
- Couteron, P., R. Pélissier, D. Mapaga, J.-F. Molino, and L. Teillier. 2003. Drawing ecological insights from a management-oriented forest inventory in French Guiana. *Forest Ecology and Management* **172**:89-108.
- Di Bella, G., and G. Jona-Lasinio. 1996. Including spatial contiguity information in the analysis of multispecific patterns. *Environmental and Ecological Statistics* **3**: 269-280.
- Dolédec, S., D. Chessel, and C. Gimaret-Carpentier. 2000. Niche separation in community analysis: a new method. *Ecology* **81**:2914–2927.
- Escoufier, Y. 1987. The duality diagram: a means of better practical applications. Pages 139-156 *in* P. Legendre and L. Legendre, editors. *Development in numerical ecology*. Springer-Verlag, Berlin, Germany.
- Gimaret-Carpentier, C., D. Chessel, and J.-P. Pascal. 1998. Non-symmetric correspondence analysis: an alternative for species occurrences data. *Plant Ecology* **138**:97–112.
- Greig-Smith, P. 1983. *Quantitative Plant Ecology*. 3rd ed. Blackwell Science Publishing, Oxford, UK.
- Ihaka, R., and R. Gentleman. 1996. R: a language for data analysis and graphics. *Journal of Computational and Graphical Statistics* **5**:299-314.
- Legendre, P., and L. Legendre. 1998. *Numerical ecology*. Elsevier, Amsterdam, The Netherlands.
- Levin, S. 1992. The problem of pattern and scale in ecology. *Ecology* **73**: 1943-1967.

- Noy-Meir, I., and D. Anderson. 1971. Multiple pattern analysis or multiscale ordination: towards a vegetation hologram. Pages 207-232 in G. P. Patil, E.C. Pielou, and E.W. Water, editors. *Statistical ecology: populations, ecosystems, and systems analysis*. Pennsylvania State University Press, University Park, Pennsylvania, USA.
- Ollier, S., D. Chessel, P. Couteron, R. Pélissier, and J. Thioulouse. 2003. Comparing and classifying one-dimensional spatial patterns: an application to laser altimeter profiles. *Remote Sensing of Environment* **85**: 453-462.
- Pélissier, R., P. Couteron, S. Dray, and D. Sabatier. 2003. Consistency between ordination techniques and diversity measurements: two strategies for species occurrence data. *Ecology*, **84**:242-251.
- Sabatier, R., J.-D. Lebreton, and D. Chessel. 1989. Principal component analysis with instrumental variables as a tool for modelling composition data. Pages 341–352 in R. Coppi and S. Bolasco, editors. *Multiway data analysis*, Elsevier Science Publishers, The Netherlands.
- ter Braak, C. J. F. 1983. Principal components biplots and alpha and beta diversity. *Ecology* **64**: 454-462.
- Thioulouse, J., D. Chessel, and S. Champély. 1995. Multivariate analysis of spatial patterns: a unified approach to local and global structures. *Environmental and Ecological Statistics* **2**: 1-14.
- Ver Hoef, J. M., and D. C. Glenn-Lewin. 1989. Multiscale ordination: a method for detecting pattern at several scales. *Vegetatio* **82**: 59-67.
- Wackernagel, H. 1998. *Multivariate geostatistics*. Second completely revised edition. Springer, Berlin, Germany.
- Wagner, H. H. 2003. Spatial covariance in plant communities: integrating ordination, geostatistics, and variance testing. *Ecology* **84**: 1045-1057.
- Wagner, H. H. 2004. Direct multi-scale ordination with canonical correspondence analysis. *Ecology* **85**: 342-351.

Table 1. Definition of some ordination methods from re-scaling and weighting options

D is the total diversity computed from the species frequency distribution

($D = \sum_i w_i [f_{i+} (1 - f_{i+})]$), while I_N is the total "inertia" corresponding to the ordination of the

species by quadrats table (trace of the generalized variance-covariance matrix, G_T). See text for other denotation.

Re-scaling options	Weighting options for quadrats, δ_a	Weighting options for species, w_i	Corresponding ordination method	Link with diversity indices
(I) $x_{ai} = p_{ai}$	$\delta_a = 1/Q$	$w_i = 1$	Species-centered PCA*	
(II) $x_{ai} = \frac{p_{ai}}{\sqrt{V_i}}$ **	$\delta_a = 1/Q$	$w_i = 1$	PCA on the species correlation matrix*	$I_N = S$ (richness)
(III) $x_{ai} = p_{ai}/p_{a+}$	$\delta_a = 1/Q$	$w_i = 1$	Species-centered PCA on proportions (ter Braak 1983)	$I_N = \text{Simpson-Gini}$
(IV) $x_{ai} = p_{ai}/p_{a+}$	$\delta_a = p_{a+}$	$w_i = 1/p_{+i}$	Correspondence analysis (CA)*	$D = S-1$ (richness - 1)
		$w_i = \log(1/p_{+i}) / (1-p_{+i})$	(Pélissier et al. 2003)	$D = \text{Shannon}$
		$w_i = 1$	Non-symmetric correspondence analysis (NSCA) (Gimaret-Carpentier et al. 1998)	$D = \text{SimpsonGini}$

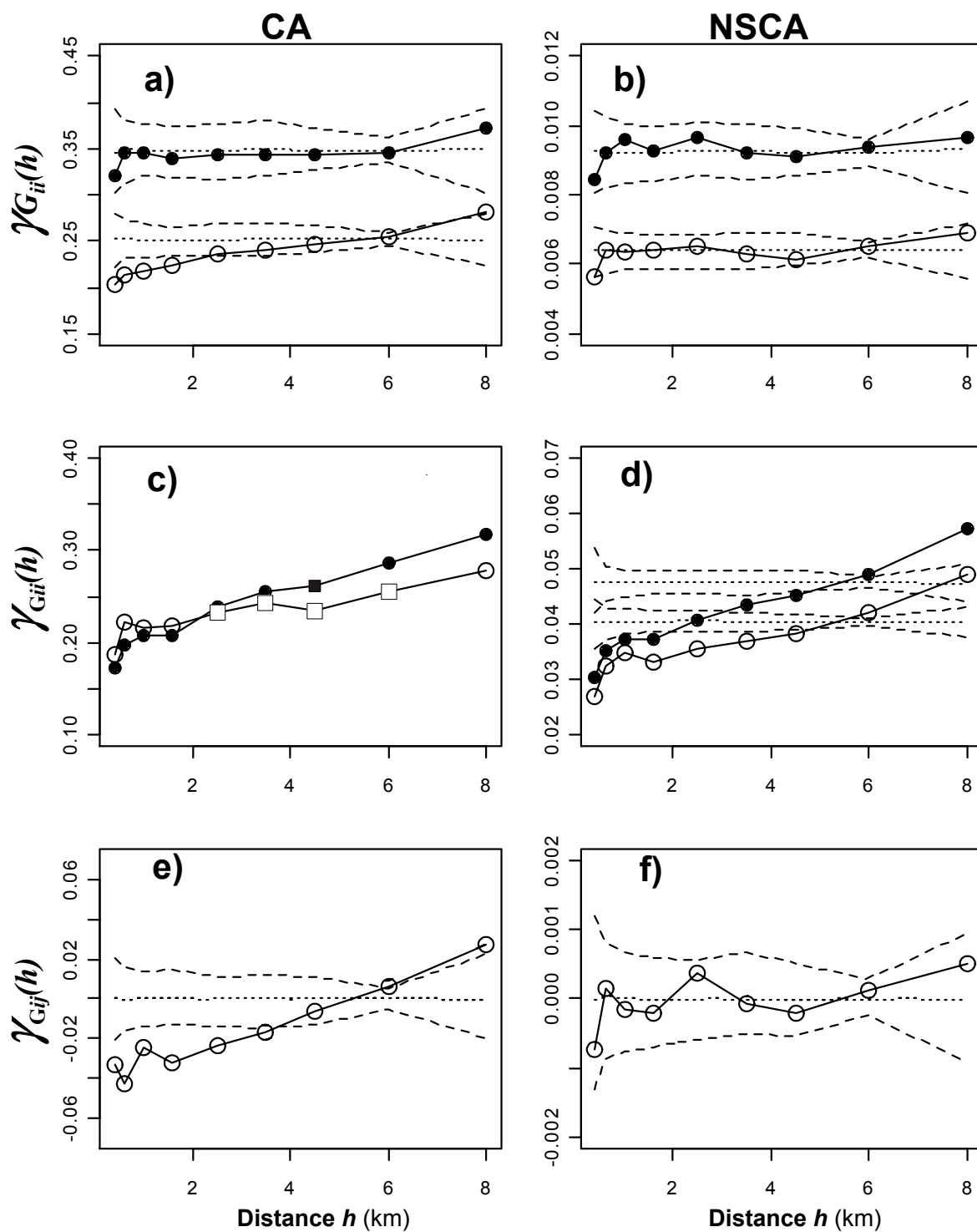
* (Legendre and Legendre 1998); ** $V_i = \frac{1}{Q} \sum_a (p_{ai} - p_{+i})^2$

Legend for Figure 1

Spatial patterns shown by the main ordination axes provided by the application of correspondence analysis (CA) and non-symmetric correspondence analysis (NSCA) to vegetation data from the COUNAMI Forest Reserve in French Guiana (7,189 trees of 59 species sampled in 411 quadrats of 0.3 ha each). a) Generalized variogram for axis CA2 of the initial table (filled circles) and for the homologous axis from the residual table, after factoring out of 12 environmental categories (open circles). The dashed lines denote the 95% bilateral envelopes computed from 300 re-allocations of the specific composition to geographical locations (complete randomization for variograms from the initial data table and randomization within environmental categories for variograms from the residual table). The dotted line denotes the mean values for randomizations. b) Same as a) but for axis NSCA2. c) Same as a) but for CA3 (confidence envelopes are omitted for legibility, values within envelopes are marked by a square). d) Same as a) but for NSCA1. e) Generalized cross-variograms between the two main CA axes of the residual table. f) same as e) for NSCA.

1

Figure 1



Ecological Archives E086-042-A1

Pierre Couteron, and Sébastien Ollier. 2005. A generalized, variogram-based framework for multi-scale ordination. *Ecology* 86:828-834.

Appendix A. Matrix-algebraic presentation of the concepts and computations..

General denotation

Let \mathbf{X} be a table expressing a measure of the abundance x_{ai} of S species (columns) within Q quadrats (rows). \mathbf{x}_i and \mathbf{x}_j are two columns of table \mathbf{X} , relating to species i and j , respectively. Let \mathbf{D} be a matrix containing quadrat weights (δ_a , $\sum_a \delta_a = 1$) on its main diagonal and zeros for all off-diagonal values, and let \mathbf{W} be a S by S matrix containing the square root of species weights ($\sqrt{w_i}$) on its main diagonal and zeros outside. (In the main paper, Table 1 gives some options for abundance re-scaling and for quadrat and species weighting.)

Contiguity relationships

Let \mathbf{L}_h be a Q by Q matrix expressing a contiguity relationship (*sensu* Lebart 1969) between the quadrats. For our variogram-based approach, we consider quadrats a and b as neighbors if the distance between the two is within the bounds of the distance class centered around h :

$$\mathbf{L}_h(a,b)=1 \text{ if } h_{a,b} \approx h \text{ and } \mathbf{L}_h(a,b)=0 \text{ otherwise.} \tag{A.1}$$

To introduce quadrat weights into the analysis, we define the matrix \mathbf{M}_h and the vector \mathbf{E}_h such that:

$$\mathbf{M}_h = \mathbf{D}\mathbf{L}_h\mathbf{D} \text{ and } \mathbf{E}_h = \mathbf{M}_h\mathbf{1}_Q \tag{A.2}$$

where $\mathbf{1}_Q$ is the vector containing Q values equal to 1.

\mathbf{M}_h contains, for each pair (a,b) of neighboring quadrats at "scale" h , the product $\delta_a \delta_b$ of their weights. \mathbf{E}_h features, for each quadrat a , the sum of the weights of its neighbors multiplied by δ_a . Let \mathbf{N}_h be the Q by Q matrix with \mathbf{E}_h on its main diagonal and zeros elsewhere.

We shall assume that distance classes include all pairs of quadrats while being mutually exclusive. In such a case, the two following matrices:

$$\mathbf{M}_T = \sum_h \mathbf{M}_h \quad \text{and} \quad \mathbf{N}_T = \sum_h \mathbf{N}_h \quad (\text{A.2b})$$

are such that \mathbf{M}_T is a Q by Q matrix that containing zeros on the diagonal while all values off the diagonal are equal to $\delta_a \delta_b$; \mathbf{N}_T is a Q by Q matrix that containing $(1-\delta_a)\delta_a$ values on the diagonal and zeros elsewhere. With \mathbf{M}_T and \mathbf{N}_T it is as if each quadrat has all other quadrats as neighbors. Denoting \mathbf{I}_Q the Q by Q diagonal identity matrix, we can also write:

$$\mathbf{N}_T = \mathbf{D}(\mathbf{I}_Q - \mathbf{D}) \quad \text{and} \quad \mathbf{M}_T = \mathbf{D}(\mathbf{1}_Q \mathbf{1}_Q^t - \mathbf{I}_Q) \mathbf{D} \quad (\text{A.3})$$

(where the exponent 't' is the matrix transpose). Thus:

$$\mathbf{N}_T - \mathbf{M}_T = \mathbf{D} - \mathbf{D} \mathbf{1}_Q \mathbf{1}_Q^t \mathbf{D} \quad (\text{A.4})$$

Equivalent expressions of the generalized variance-covariance matrix

Let \mathbf{G}_T be the generalized variance-covariance matrix, irrespective of distance classes, that can be directly computed from table \mathbf{X} using weighting options for rows and columns defined by matrices \mathbf{D} and \mathbf{W} , respectively. \mathbf{G}_T contains, for each species couple (i,j) , the generalized covariances, g_{ij} as defined by Eq. 1 and Eq. 2 in the main paper:

$$g_{ij} = \sum_{a,b} g_{ij}(a,b) \quad (\text{A.5})$$

Usual algebraic manipulations allow us to re-write Eq. 1 and Eq. A.5 as:

$$g_{ij} = \sqrt{w_i w_j} \left(\sum_a^Q \delta_a x_{ai} x_{aj} - \bar{x}_i \bar{x}_j \right) \quad (\text{A.6})$$

where \bar{x}_i and \bar{x}_j are the \mathbf{D} -weighted means of \mathbf{x}_i and \mathbf{x}_j , respectively.

$$(\bar{x}_i = \sum_a^Q \delta_a x_{ia} \quad \text{or, equivalently,} \quad \bar{x}_i = \mathbf{X}_i^t \mathbf{D} \mathbf{1}_Q)$$

The matrix expression of g_{ij} is thus:

$$g_{ij} = \mathbf{W}(\mathbf{x}_i^t \mathbf{D} \mathbf{x}_j - \mathbf{x}_i^t \mathbf{D} \mathbf{1}_Q \mathbf{x}_j^t \mathbf{D} \mathbf{1}_Q) \mathbf{W} \quad (\text{A.7})$$

which generalizes into:

$$\mathbf{G}_T = \mathbf{W}(\mathbf{X}^t \mathbf{D} \mathbf{X} - \bar{\mathbf{X}}^t \mathbf{D} \bar{\mathbf{X}}) \mathbf{W} \quad (\text{A.8})$$

$$\text{where } \bar{\mathbf{X}} = \mathbf{1}_Q \mathbf{1}_Q^t \mathbf{D} \mathbf{X} \quad \text{and where } \bar{\mathbf{X}}^t \mathbf{D} \bar{\mathbf{X}} = [\bar{x}_i \bar{x}_j] \quad (\text{A.9})$$

Note that we may also write:

$$\mathbf{G}_T = \mathbf{W}(\mathbf{X} - \bar{\mathbf{X}})^t \mathbf{D} (\mathbf{X} - \bar{\mathbf{X}}) \mathbf{W} \quad (\text{A.10})$$

On the other hand, it is important to note that \mathbf{G}_T can be directly computed as:

$$\mathbf{G}_T = \mathbf{W}\mathbf{X}'(\mathbf{N}_T - \mathbf{M}_T)\mathbf{X}\mathbf{W} \quad (\text{A.11})$$

Proof of Eq. A.11:

$$\mathbf{X}'(\mathbf{N}_T - \mathbf{M}_T)\mathbf{X} = \mathbf{X}'(\mathbf{D} - \mathbf{D}\mathbf{1}_Q\mathbf{1}_Q'\mathbf{D})\mathbf{X} \quad (\text{using Eq. A.4})$$

$$\mathbf{X}'(\mathbf{D} - \mathbf{D}\mathbf{1}_Q\mathbf{1}_Q'\mathbf{D})\mathbf{X} = \mathbf{X}'\mathbf{D}\mathbf{X} - \mathbf{X}'\mathbf{D}\bar{\mathbf{X}} \quad (\text{using Eq. A.9})$$

Noting that $\mathbf{X}'\mathbf{D}\bar{\mathbf{X}} = \bar{\mathbf{X}}'\mathbf{D}\bar{\mathbf{X}}$, allows us to write:

$$\mathbf{X}'(\mathbf{N}_T - \mathbf{M}_T)\mathbf{X} = \mathbf{X}'\mathbf{D}\mathbf{X} - \bar{\mathbf{X}}'\mathbf{D}\bar{\mathbf{X}} \quad (\text{A.12})$$

Partition of the generalized variance-covariance matrix among distance classes

The very definition of matrices \mathbf{N}_T and \mathbf{M}_T (Eq. A.2b), along with Eq. A.11, enables partition of \mathbf{G}_T into strictly additive components, \mathbf{G}_h , that relate each to a distance class:

$$\mathbf{G}_T = \sum_h \mathbf{G}_h = \sum_h \mathbf{W}\mathbf{X}'(\mathbf{N}_h - \mathbf{M}_h)\mathbf{X}\mathbf{W} \quad (\text{A.13})$$

\mathbf{G}_h is the generalized variance-covariance matrix defined for the distance class h by the neighboring relationship expressed by the matrices \mathbf{N}_h and \mathbf{M}_h . \mathbf{G}_h translates easily into generalization of Wagner's variogram matrix (2003) by a division of all its values by

$$K(h) = \sum_{a,b|h_{ab} \approx h} \delta_a \delta_b \quad \text{or} \quad K(h) = \mathbf{1}_Q' \mathbf{M}_h \mathbf{1}_Q \quad (\text{A.14})$$

Equations A.2, A.13 and A.14 are used for easy programming of the method as well as efficient computations via any matrix-oriented programming environment, as we did with Matlab® and R (Ihaka and Gentleman 1996): see the freely available library "msov" on <http://pbil.univ-lyon1.fr/CRAN/>.)

For a particular species couple i and j we obtain:

$$g_{ij}(h) = \mathbf{W}\mathbf{x}_i'(\mathbf{N}_h - \mathbf{M}_h)\mathbf{x}_j\mathbf{W} \quad (\text{A.15})$$

Dividing by the scaling factor $K(h)$ gives the value at "scale" h of the generalized version of either cross-variogram ($i \neq j$) or variogram ($i = j$):

$$\gamma_{G_{ij}}(h) = \frac{1}{K(h)} g_{ij}(h) \quad (\text{A.16})$$

Multi-scale ordination

All the ordination methods mentioned in Table 1 of the main paper are based on the singular values decomposition (svd) of the appropriate version of \mathbf{G}_T to compute eigenvectors, \mathbf{u}_f , and associated eigenvalues, λ_f . Let \mathbf{U}_f be the matrix having all the eigenvectors \mathbf{u}_f as columns and let $\mathbf{\Lambda}$ be the diagonal matrix having the eigenvalues λ_f on its diagonal. Both eigenvectors and eigenvalues of \mathbf{G}_T can be partitioned by distance classes:

$$\mathbf{F}_h = \mathbf{U}_f' \mathbf{G}_h \mathbf{U}_f \text{ and } \lambda_f(h) = \mathbf{u}_f' \mathbf{G}_h \mathbf{u}_f \quad (\text{A.17})$$

\mathbf{F}_h is the variance-covariance matrix of the eigenvectors at scale h . Scale-dependent variogram/cross-variogram matrices of the eigenvectors are deduced by the appropriate scaling (Eq. A.16). Note also that:

$$\sum_h \mathbf{F}_h = \mathbf{U}_f' \left(\sum_h \mathbf{G}_h \right) \mathbf{U}_f = \mathbf{U}_f' \mathbf{G}_T \mathbf{U}_f = \mathbf{\Lambda} \quad (\text{A.18})$$

Taking environmental heterogeneity into account

Let us now suppose that a table, \mathbf{Z} , containing assessments of P environmental variables for the Q quadrats, is available in addition to table of species composition. It is well established that the centered by columns table, \mathbf{X}_c , may be partitioned into an approximated table,

$$\mathbf{A}_c = \mathbf{Z}(\mathbf{Z}'\mathbf{D}\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{D}\mathbf{X}_c \quad (\text{A.19})$$

and a residual table, $\mathbf{R}_c = \mathbf{X}_c - \mathbf{A}_c$ (Sabatier et al. 1989).

In the same manner, it may also have a direct decomposition of the initial table \mathbf{X} :

$$\mathbf{A} = \mathbf{Z}(\mathbf{Z}'(\mathbf{N}_T - \mathbf{M}_T)\mathbf{Z})^{-1}\mathbf{Z}'(\mathbf{N}_T - \mathbf{M}_T)\mathbf{X} \text{ and } \mathbf{R} = \mathbf{X} - \mathbf{A} \quad (\text{A.20})$$

After factoring out the environmental variables, residual spatial patterns may be studied by the multi-scale analysis of spatial covariances derived from table \mathbf{R} or \mathbf{R}_c . The total residual variance-covariance matrix, \mathbf{G}_{RT} , is computed as:

$$\mathbf{G}_{RT} = \mathbf{W}\mathbf{R}'(\mathbf{N}_T - \mathbf{M}_T)\mathbf{R}\mathbf{W} = \mathbf{W}\mathbf{R}_c'(\mathbf{N}_T - \mathbf{M}_T)\mathbf{R}_c\mathbf{W} \quad (\text{A.21})$$

and is broken down with respect to distance classes:

$$\mathbf{G}_{R_h} = \mathbf{W}\mathbf{R}'(\mathbf{N}_h - \mathbf{M}_h)\mathbf{R}\mathbf{W} = \mathbf{W}\mathbf{R}_c'(\mathbf{N}_h - \mathbf{M}_h)\mathbf{R}_c\mathbf{W} \quad (\text{A.22})$$

The additive partitioning of \mathbf{G}_{RT} with respect to distance classes thus enables an investigation of the residual spatial patterns by a multi-scale ordination scheme analogous to that defined by Eq. A.17 and Eq. A.18.

Literature cited in this appendix

Ihaka, R., and R. Gentleman. 1996. R: a language for data analysis and graphics. *Journal of Computational and Graphical Statistics* **5**:299-314.

Lebart, L. 1969. Analyse statistique de la contiguïté. *Publications de l'Institut de Statistiques de l'Université de Paris* **28**: 81-112.

Sabatier, R., J.-D. Lebreton, and D. Chessel. 1989. Principal component analysis with instrumental variables as a tool for modelling composition data. Pages 341–352 *in* R. Coppi and S. Bolasco, Editors. *Multiway data analysis*, Elsevier Science Publishers, Amsterdam, The Netherlands.

Wagner, H.H. 2003. Spatial covariance in plant communities: integrating ordination, geostatistics, and variance testing. *Ecology* **84**: 1045-1057.