A GENERALIZED, VARIOGRAM-BASED FRAMEWORK FOR MULTI-SCALE ORDINATION

PIERRE COUTERON1,3 AND SÉBASTIEN OLLIER2

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

Abstract. Multiscale ordination (MSO) deals with potential scale dependence in species assemblages by studying how results from multivariate ordination may be different at 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, recently introduced by H. H. Wagner, is applicable to very diverse sampling designs and for use with principal-components 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 that 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 available environmental variables. Some aspects of the approach are briefly illustrated with vegetation data from a Neotropical rain forest in French Guiana.

Key words: canonical correspondence analysis; diversity apportioning; multi-scale ordination; multivariate geostatistics; spatial contiguity; species assemblages; 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, 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- table ordination methods, i.e., principal-components 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.

Our work 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 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 introduce a generalized definition of co-variation between species that encompasses several ordination methods while being amenable to scale-explicit decompositions. We also highlight the link with the additive decomposition of common diversity indices (Peñissier et al. 2003, Couteron and Peñissier 2004) and refer to methods of two-table “direct” ordination to carry out the explicit analysis of residual spatial patterns after factoring out some environmental variables. This aspect is then 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.

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 nonnegative 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 Principal-components analysis (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, \( C \), for the species-centered (non-standardized) PCA and which is another matrix \( Q^2 \) in the case of CA (see Legendre and Legendre [1998:453] and Wagner [2004] for details). In \( Q^2 \), 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 jointly depart from expected abundance values.

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

\[
g_{ij} = \frac{1}{2} \sum_{a=1}^{Q} \sum_{t=1}^{S} (x_{ai} - x_{bj}) \sqrt{w_i w_j (x_{ai} - x_{bj}) \delta_i \delta_j} \tag{1}
\]

with variance being a special case where \( i = j \).

Here, \( w_i \) weights the influence of species \( i \), while \( \delta_i \) and \( \delta_j \) 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-overlooked 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 Péllissier 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),

\footnote{http://pbil.univ-lyon1.fr/CRAN/}
and was used by Sabatier et al. (1989) and Dolédec et al. (2000) for several single- and two-table methods (including CA, CCA [canonical correspondence analysis] and classical versions of PCA and RDA [redundancy analysis]). 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 \( \delta_s \)-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

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

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

\[
\gamma_{C_s}(h) = \frac{1}{K(h)} \sum_{a,b} g_{ij}(a, b)
\]

where \( h \) is the central value of a given distance class, and where \( K(h) \) is a scaling coefficient, such as the following:

\[
K(h) = \sum_{a,b|a=b} \delta_a \delta_b.
\]

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

\[
\gamma_C(h) = \sum_{i} \gamma_{C_i}(h).
\]

Eq. 4 is a crucial point since \( K(h) \) standardizes \( \gamma_C(h) \) in such a manner as 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, \( n_p \), 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 \( \gamma_C(h) \) is the total variance ("inertia") attached to matrix \( G_s \) 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 \( \gamma_C(h) \) of the CA-related variogram since the corresponding scaling remains proportional to \( n_i \) 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 so as to relate to a diversity measurement (option IV in Table 1). In this case, the trace of $G_T$ is the diversity among quadrats (Couteron and Pellissier 2004), which means that $\gamma_d(h)$ measures the average beta diversity between pairs of quadrats corresponding to distance class $h$. Equivalently,

$$\text{VAR}(a, b) = \sum g_{a}(a, b)$$

(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, $G_h$, 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 on 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 $G_h$ sum to $G_T$, whatever the initial choice of the reference ordination method by weighting options.

The eigenvectors and eigenvalues originating from the singular values decomposition (SVD) of $G_T$ 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, $F_h$, between eigenvectors can also be obtained (Appendix, Eq. A.17). Considering off-diagonal elements of $F_h$, namely covariances at scale $h$ between eigenvectors, is a new perspective in multi-scale ordination (MSO) that 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 $G_h$ 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 $F_h$ matrices after appropriate scaling by $K(h)$. All these cross-variograms have an expectation of zero, since the eigenvectors of $G_T$ 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 $G_h$ 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, $X$, into an approximated table $A$, modeled from environmental variables by a weighted linear regression and a residual table $R$. 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, Pellissier et al. 2003) when the linear regression uses the quadrat weights defining the ordination. (For instance, defining table $X$ along with species and quadrat weights so as to make them consistent with CA means that an ordination on $A$ would be a CCA.) To study residual spatial patterns, it is possible to break down $G_R$, i.e., the variance-covariance matrix computed from $R$, into additive variance-covariance matrices, $G_{R_a}$ 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 7189 trees (diameter at breast height above 10 cm) belonging to 59 species sampled in a
lowland tropical rain forest of ~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 × 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 Couteron et al. [2003] for details). Performing correspondence analysis (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 nonsymmetric 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 (Couteron et al. 2003) we explicitly considered inter-quadrat distances by applying the generalized variogram-based multi-scale ordination (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 (Péllissier et al. 2003), it was the total among-quadrats diversity (sensu the species richness for CA or the Simpson-Gini index for NSCA) that 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. 1a and b). The study of residual patterns, after factoring out the 12 environmental categories (unconstrained ordinations on the residual variance–covariance matrix, 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 that 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 MSO 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—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. 1c and 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 F_r Eq. A.18 in the Appendix). Only cross-variograms computed from the residual variance–covariance matrix (R) are shown (Fig. 1e 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. 1f) 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. 1e). 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 multiscale ordination (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 at which some patterns may differ from others. This can be done by analyzing, for instance by PCA, the table containing the generalized variograms of all species (prior standardization by variances of individual species so as to have all sills equal to 1 is likely to be preferable [a “sill” is the expected variance, i.e., the value at which a variogram is expected to level off]). 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
FIG. 1. Spatial patterns shown by the main ordination axes provided by the application of correspondence analysis (CA) and nonsymmetric correspondence analysis (NSCA) to vegetation data from the Coumani Forest Reserve in French Guiana (7189 trees of 59 species sampled in 411 quadrats of 0.3 ha each). (a) Generalized variogram $\gamma_{G}(h)$, for axis CA2 of the initial table (solid circles) and for the homologous axis from the residual table, after factoring out of 12 environmental categories (open circles). Dashed lines denote the 95% bilateral envelopes computed from 300 reallocations 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). Each dotted line denotes the mean values for randomizations. (b) Same as panel (a), but for axis NSCA2. (c) Same as panel (a), but for CA3 (confidence envelopes are omitted for legibility, values within envelopes are marked by a square). (d) Same as panel (a), but for NSCA1. (e) Generalized cross-variograms, $\gamma_{xy}$, between the two main CA axes of the residual table. (f) Same as panel (e), but for NSCA.

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.
ACKNOWLEDGMENTS

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


APPENDIX

A matrix-algebraic presentation of the concepts and computations is available in ESA’s Electronic Data Archive: Ecological Archives E086-042-A1.