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 **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 **X**, relating to species i and j, respectively. Let **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 **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:

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

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

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

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:

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

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 $(1\text{-}\delta_a)\delta_a$ 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

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

(where the exponent ' ^{t} ' is the matrix transpose). Thus

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

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:

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

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

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

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

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

The matrix expression of g_{ij} is thus

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

which generalizes into

$$\mathbf{G}_\mathbf{T} = \mathbf{W}(\mathbf{X}^t \mathbf{D} \mathbf{X} - \overline{\mathbf{X}}^t \mathbf{D} \overline{\mathbf{X}})\mathbf{W} \tag{A.8}$$

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

Note that we may also write:

$$\mathbf{G}_\mathbf{T} = \mathbf{W}(\mathbf{X} - \overline{\mathbf{X}})^t \mathbf{D}(\mathbf{X} - \overline{\mathbf{X}})\mathbf{W} \tag{A.10}$$

On the other hand**,** it is important to note that $\mathbf{G_T}$ can be directly computed as

$$\mathbf{G}_\mathbf{T} = \mathbf{W}\mathbf{X}^t (\mathbf{N}_\mathbf{T} - \mathbf{M}_\mathbf{T})\mathbf{X}\mathbf{W} \tag{A.11}$$

Proof of Eq. A.11:

$$\mathbf{X}^t (\mathbf{N}_\mathbf{T} - \mathbf{M}_\mathbf{T})\mathbf{X} = \mathbf{X}^t (\mathbf{D} - \mathbf{D}\mathbf{1}_Q \mathbf{1}_Q{}^t \mathbf{D})\mathbf{X} \quad \text{(using Eq. A.4)}$$

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

Noting that $\mathbf{X}^t \mathbf{D} \overline{\mathbf{X}} = \overline{\mathbf{X}}^t \mathbf{D} \overline{\mathbf{X}}$, allows us to write

$$\mathbf{X}^t (\mathbf{N}_\mathbf{T} - \mathbf{M}_\mathbf{T})\mathbf{X} = \mathbf{X}^t \mathbf{D} \mathbf{X} - \overline{\mathbf{X}}^t \mathbf{D} \overline{\mathbf{X}} \tag{A12}$$

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

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

$\mathbf{G}_h$ 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

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

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

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

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 ($i=j$)

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

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 Λ 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

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

$\mathbf{F}_h$ 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

$$\sum_h \mathbf{F}_h = \mathbf{U}_f{}^t \left(\sum_h \mathbf{G}_h \right) \mathbf{U}_f = \mathbf{U}_f{}^t \mathbf{G}_\mathrm{T} \mathbf{U}_f = \mathbf{\Lambda} \qquad \text{(A.18)}$$

Taking environmental heterogeneity into account

Let us now suppose that a table, **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, $\mathbf{X}_\mathrm{C}$, may be partitioned into an approximated table,

$$\mathbf{A}_\mathrm{C} = \mathbf{Z}(\mathbf{Z}^t\mathbf{D}\mathbf{Z})^{-1}\mathbf{Z}^t\mathbf{D}\mathbf{X}_\mathrm{C} \qquad \text{(A.19)}$$

and a residual table, $\mathbf{R}_\mathrm{C} = \mathbf{X}_\mathrm{C} - \mathbf{Ac}$ (Sabatier et al. 1989).

In the same manner, it may also have a direct decomposition of the initial table **X**:

$$\mathbf{A} = \mathbf{Z}(\mathbf{Z}^t(\mathbf{N}_\mathrm{T} - \mathbf{M}_\mathrm{T})\mathbf{Z})^{-1}\mathbf{Z}^t(\mathbf{N}_\mathrm{T} - \mathbf{M}_\mathrm{T})\mathbf{X} \text{ and } \mathbf{R} = \mathbf{X} - \mathbf{A} \qquad \text{(A.20)}$$

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

$$\mathbf{G}_\mathrm{RT} = \mathbf{W}\mathbf{R}^t(\mathbf{N}_\mathrm{T} - \mathbf{M}_\mathrm{T})\mathbf{R}\mathbf{W} = \mathbf{W}\mathbf{R}_\mathrm{C}{}^t(\mathbf{N}_\mathrm{T} - \mathbf{M}_\mathrm{T})\mathbf{R}_\mathrm{C}\mathbf{W} \qquad \text{(A.21)}$$

and is broken down with respect to distance classes

$$\mathbf{G}_{\mathrm{R}h} = \mathbf{W}\mathbf{R}^t(\mathbf{N}_h - \mathbf{M}_h)\mathbf{R}\mathbf{W} = \mathbf{W}\mathbf{R}_\mathrm{C}{}^t(\mathbf{N}_h - \mathbf{M}_h)\mathbf{R}_\mathrm{C}\mathbf{W} \qquad \text{(A.22)}$$

The additive partitioning of $\mathbf{G}_\mathbf{RT}$ 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.