Journal of Vegetation Science 19: 45-56, 2008
DOI : 10.3170/8-18312

Spatial ordination of vegetation data using a generalization of Wartenberg's multivariate spatial correlation

Dray, Stéphane[1*]; Saïd, Sonia[2] & Débias, François[1]

[1]*Université de Lyon, université Lyon 1, CNRS, UMR 5558, Laboratoire de Biométrie et Biologie Evolutive, 43 boulevard du 11 novembre 1918, Villeurbanne FR-69622, France;*
[2]*Office National de la Chasse et de la Faune Sauvage, Centre National d'Etudes et de Recherches Appliquées Cervidés-Sanglier, 85bis avenue de Wagram, FR-75017, Paris, France;*
[*]*Corresponding author: Fax +33 472431388; E-mail dray@biomserv.univ-lyon1.fr*

Abstract

Question: Are there spatial structures in the composition of plant communities?

Methods: Identification and measurement of spatial structures is a topic of great interest in plant ecology. Univariate measurements of spatial autocorrelation such as Moran's *I* and Geary's *c* are widely used, but extensions to the multivariate case (i.e. multi-species) are rare. Here, we propose a multivariate spatial analysis based on Moran's *I* (MULTISPATI) by introducing a row-sum standardized spatial weight matrix in the statistical triplet notation. This analysis, which is a generalization of Wartenberg's approach to multivariate spatial correlation, would imply a compromise between the relations among many variables (multivariate analysis) and their spatial structure (autocorrelation). MULTISPATI approach is very flexible and can handle various kinds of data (quantitative and/or qualitative data, contingency tables).

A study is presented to illustrate the method using a spatial version of Correspondence Analysis.

Location: Territoire d'Etude et d'Expérimentation de Trois-Fontaines (eastern France).

Results: Ordination of vegetation plots by this spatial analysis is quite robust with reference to rare species and highlights spatial patterns related to soil properties.

Keywords: Correspondence Analysis; Moran's *I*; Multivariate analysis; Spatial autocorrelation; Spatially Constrained Ordination.

Abbreviations: CCA = Canonical Correspondence Analysis; CA = Correspondence Analysis; MCA = Multiple Correspondence Analysis; MULTISPATI = Multivariate spatial analysis based on Moran's I; TF = Territoire d'Etude et d'Expérimentation de Trois-Fontaines.

Nomenclature: Tutin et al. (2001).

Introduction

Multivariate methods have often been used to summarize ecological datasets. For instance, species-by-site tables can be analysed by Principal Component Analysis (PCA, Hotelling 1933) or Correspondence Analysis (CA, Greenacre 1984). CA is often preferred by plant ecologists because it considers relative composition (when PCA is based on abundance) and it is related to unimodal response models (Whittaker 1967; ter Braak 1985). However, the use of CA is sometimes problematic because the χ^2 metric (used by CA) tends to overemphasize the importance of rare species. Various methods can be used to analyse environmental variables: PCA for quantitative data, Multiple Correspondence Analysis (MCA, Tenenhaus & Young 1985) for qualitative data, while alternatives are available for mixtures of quantitative and qualitative data (PCAMIX, Kiers 1994).

Ecological datasets are often geo-referenced; one major question concerns the identification and explanation of the spatial variability of ecological structures (Cormack & Ord 1979). Standard multivariate techniques are often used on geo-referenced datasets and often successfully. The usual approach consists in performing multivariate analysis to identify the main ecological processes and then interpreting the spatial components of structures observed on the first few axes. The second step can be achieved by mapping the scores in geographical space (e.g. Goodall 1954; Kadmon & Danin 1997; Dray et al. 2003c) or by using geostatistical tools such as spatial autocorrelation indices (Selmi et al. 2003). However, standard multivariate analyses do not directly take into account spatial relations in their computation and are not specifically designed to identify spatial structures.

The identification and measurement of the spatial component of a single variable has been a major issue in applied geography. Global indices such as Moran's *I* and Geary's *c* (Moran 1948; Geary 1954; Cliff & Ord 1973) and their local extensions (Anselin 1995) have been widely used to measure the spatial dependence and its local variations for one quantitative variable.

Indices for nominal data are also well known (Krishna Iyer 1949). However, all these approaches focus on the spatial structure of a single variable and cannot be used in a multivariate context.

Ecological data are often multivariate and spatialized and their analysis is closely related to the development of spatial multivariate techniques, i.e. methods of multivariate analysis aiming at the identification of spatial structure (e.g. spatial patches, regional trends). These methods, which require the inclusion of the spatial dependence between observations in multivariate analysis, are relatively undeveloped although this problem is encountered across a wide range of fields. The first interesting attempt that aimed at depicting basic multivariate spatial patterns was due to Wartenberg (1985). Lee (2001) showed that Wartenberg's approach had major drawbacks and proposed a bivariate spatial association measure which can be easily used for spatial multivariate analysis. Other methods (see review in Bailey & Krzanowski 2000) have been developed in various fields such as spatial imagery (Switzer & Green 1984) or geosciences (Grunsky & Agterberg 1991). All these approaches (including Wartenberg's method) include the diagonalization of a spatial covariance or correlation matrix to identify multivariate spatial association and are restricted to the case of quantitative (normalized) variables.

In this paper, we propose a new method of spatial multivariate analysis. Contrary to the above-cited methods that deal only with (normalized) quantitative data, our approach is very general. It introduces a spatial constraint in classical multivariate methods and allows users to perform spatial CA, for example. It can be seen as a generalization of Wartenberg's approach taking into account the pitfalls pointed out by Lee (2001). We first introduce some simple elements of spatial analysis and we then present the general framework of multivariate analysis using the statistical triplet notation. Next, the principles of a new spatial multivariate analysis are given and the method is illustrated on a real data set.

Measurements of spatial association

Geary's c and Moran's I

Let us consider x, a vector composed by the measurements of a variable for n spatial units, i.e. $\mathbf{x}^t = [x_1,\ldots x_n]$. Moran's I is given by:

$$I(\mathbf{x}) = \frac{n\sum_{(2)} c_{ij}(x_i-\bar{x})(x_j-\bar{x})}{\sum_{(2)} c_{ij}\sum_{i=1}^{n}(x_i-\bar{x})^2} \qquad (1)$$

and the Geary's c is :

$$c(\mathbf{x}) = \frac{(n-1)\sum_{(2)} c_{ij}\left(x_i-x_j\right)^2}{2\sum_{(2)} c_{ij}\sum_{i=1}^{n}(x_i-\bar{x})^2} \qquad (2)$$

where $\sum_{(2)} = \sum_{i=1}^{n}\sum_{j=1}^{n}$ with $i \neq j$

and $\mathbf{C} = [c_{ij}]$ is a spatial connectivity matrix.

Geary's c is always positive while Moran's I can be positive or negative. If we consider the vector $\mathbf{z}^t = [z_i] = [x_i - \bar{x}]$ of centered values, Moran's I is equal to:

$$I(\mathbf{x}) = \frac{n\sum_{(2)} c_{ij} z_i z_j}{\sum_{(2)} c_{ij}\sum_{i=1}^{n} z_i^{\,2}} \qquad (3)$$

The spatial weighting matrix

The matrix $\mathbf{C} = [c_{ij}]$ is a weighting matrix (Bavaud 1998; Tiefelsdorf et al. 1999) which indicates the strength of the potential interaction between spatial units. Cliff & Ord (1973, p. 12), specified that "the use of a generalised weighting matrix [...] allows the investigator to choose a set of weights which he deems appropriate from prior considerations. This allows great flexibility". The binary connectivity version of $\mathbf{C}$ ($\mathbf{B}$) whose elements equal 1 for contiguous spatial units and 0 otherwise is often used. Econometricians prefer to use the row-sum standardized version of $\mathbf{C}\left(\mathbf{W} = \left[c_{ij} / \sum_{j=1}^{n} c_{ij}\right]\right)$ which allows easier interpretation of autoregressive models (Ord 1975). A double standardized spatial weighting matrix is also often used with sum of all elements equal to $1\left(\mathbf{F} = \left[c_{ij} / \sum_{(2)} c_{ij}\right]\right)$ or to n ($n\mathbf{F}$). De Jong et al. (1984) provided exact lower and upper bounds for c and I for a given connection matrix. These extremes are given by the smallest and largest eigenvalues of **NBN** and $\mathbf{N}\left(\mathbf{W}+\mathbf{W}^t\right)\mathbf{N}/2$ (where $\mathbf{N} = \left(\mathbf{I} - (1/n)\mathbf{1}\mathbf{1}^t\right)$ is a centering operator, $\mathbf{1}^t = [1,\ldots,1]$ a (1 by n) row vector and $\mathbf{I}$ the identity matrix), for the $\mathbf{B}$ and $\mathbf{W}$ weighting options respectively, while the eigenvectors of these matrices (Griffith 1996, 2000a) can be used for spatial filtering purposes (Griffith 2000b; Getis & Griffith 2002). This diagonalization is closely related to the PCNM approach (Borcard & Legendre 2002) as demonstrated by Dray et al. (2006).

The choice of the spatial weighting matrix is the most critical step in computing a measure of spatial associa-

tion because it can influence the significance of the test (Tiefelsdorf et al. 1999). Moreover, it defines also the limits of autocorrelation measures.

When **W** is applied, Lee (2001) proposes a nice decomposition of Moran's *I* into two parts using the concept of spatial lag (Anselin 1996). The lag vector is composed of the averages of neighbours weighted by the spatial connection matrix and is computed by:

$$\tilde{\mathbf{x}} = \mathbf{W}\mathbf{x} \text{ (i.e. } \tilde{x}_i = \sum_{j=1}^{n} w_{ij} x_j) \quad (4)$$

Anselin (1996) proposed to study spatial autocorrelation with a Moran scatterplot by plotting the original variable (**x**) against the spatial lag of the variable ($\tilde{\mathbf{x}}$). The use of the weighting matrix **W** reduces (3) to:

$$I(\mathbf{x}) = \frac{\sum_{(2)} c_{ij} z_i z_j}{\sum_{i=1}^{n} z_i^2} = \frac{\sum_{i=1}^{n} z_i \tilde{z}_i}{\sum_{i=1}^{n} z_i^2} = \frac{\mathbf{z}^t \tilde{\mathbf{z}}}{\mathbf{z}^t \mathbf{z}} \quad (5)$$

where $\tilde{\mathbf{z}} = \mathbf{W}\mathbf{z}$. Row-sum standardization implies that Moran's *I* is reduced to a ratio of quadratic forms which can be easily interpreted and it provides a sort of smoothing operator (lag vector). In the case of a regular lattice, the number of neighbours tends to be constant and the use of **B** (or **F**) weights can be justified. One can think of a 'spatial lag' as $\sum_{j=1}^{n} c_{ij} z_j$ (as in eq. 3) as a sum of the neighbouring values. If the number of neighbours is not constant, these values are a function of the number of neighbours. This makes sense in some contexts, but there are many where it doesn't. The row-standardization avoids this problem and creates a variable that is simply an average of the neighbours and thus will be comparable to the value of the original observations. Moreover, it facilitates comparisons between spatial parameters in the case of spatial autoregressive model. The parameter space (the range of allowed values) is determined by the values in the weight matrix. For a row-standardized weights matrix, the maximum is always 1. For an unstandardized weights matrix, the maximum depends on the values in the matrix: high weights yield small parameter values and vice versa (L. Anselin pers. comm.).

All these considerations justify the preference for selecting the **W** weighting option to compute Moran's *I*. Another argument arises from the work of Lee (2001) who rewrote Moran's *I* to develop a bivariate spatial association measure:

$$I(\mathbf{x}) = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(\tilde{x}_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} = \sqrt{\frac{\sum_{i=1}^{n} (\tilde{x}_i - \bar{x})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}} \cdot \underbrace{\sqrt{\frac{\sum_{i=1}^{n} (\tilde{x}_i - \bar{\tilde{x}}_i)^2}{\sum_{i=1}^{n} (\tilde{x}_i - \bar{x})^2}}}_{\equiv 1} \cdot r_{\mathbf{x},\tilde{\mathbf{x}}} \cong \sqrt{SSS_{\mathbf{x}}} \cdot r_{\mathbf{x},\tilde{\mathbf{x}}} \quad (6)$$

As the second part of the middle element is approximately 1, Moran's *I* can be regarded as the product of a spatial smoothing scalar (SSS) by the Pearson correlation between the variable and its spatial lag. Then, Lee (2001, p. 376) suggested that a bivariate spatial association measure should include a "point-to-point association between two variables, which requires the inclusion of a certain form of Pearson's correlation between the two variables" and "should reflect the degrees of spatial autocorrelation for both variables under investigation. In other words, it should respond to the collective effect of the SSSs of the variables". Unfortunately, his approach considers the correlation between the spatial lag vectors ($r_{\tilde{\mathbf{x}},\tilde{\mathbf{y}}}$) but not between the original variables ($r_{x,y}$), and he found that these two quantities ($r_{\tilde{\mathbf{x}},\tilde{\mathbf{y}}}$ and $r_{x,y}$) could have different signs.

A new spatial multivariate analysis: MULTISPATI analysis

Standard multivariate analysis is a natural tool to summarize large data sets. Various methods are available to take into account the different characteristics of the data (quantitative or qualitative variables, contingency tables…). For a review in ecology, the reader should consult Dray et al. (2003a). The notion of statistical triplet (Cailliez & Pagès 1976; Escoufier 1987) provides a theoretical framework and an efficient way to define multivariate analyses. Applying a multivariate method (e.g., PCA, CA, …) to a data table **X** corresponds to the analysis of a statistical triplet (**X**, **Q**, **D**) (see App. 1 for more details).

We present a new approach, Multivariate spatial analysis based on Moran's *I* (MULTISPATI). The method originates in a course in French (Chessel et al. 2004a) and introduces the row-sum standardized weight matrix **W** in the analysis of a statistical triplet (**X**, **Q**, **D**). It is possible to extend the concept of lag vector to construct a lag matrix $\tilde{\mathbf{X}} = \mathbf{W}\mathbf{X}$. The two tables $\tilde{\mathbf{X}} = \mathbf{W}\mathbf{X}$ and **X** are fully matched, i.e. it contains the measurements of the same variables for the same sites. The principle of MULTISPATI consists of the analysis of this pair of tables by the coinertia analysis (Dolédec & Chessel 1994; Dray et al. 2003a) of a pair of fully matched tables (Torre & Chessel 1995; Dray et al. 2003b). MULTISPATI seeks for $\mathbf{u}_1$ (with $\|\mathbf{u}_1\|_{\mathbf{Q}}^2 = \mathbf{u}_1^t \mathbf{Q} \mathbf{u}_1 = 1$) maximizing the quantity (see App. 1 for more details):

$$Q(\mathbf{u}_1) = \mathbf{a}_1{}^t \mathbf{D} \tilde{\mathbf{a}}_1 \quad (7)$$

This analysis maximizes the scalar product between a linear combination of original variables ($\mathbf{a}_1 = \mathbf{X}\mathbf{Q}\mathbf{u}_1$) and

a linear combination of lagged variables ($\tilde{\mathbf{a}}_1 = \mathbf{WXQu}_1$). Eq. (7) can be rewritten as:

$$Q(\mathbf{u}_1) = I_{\mathbf{D}}(\mathbf{a}_1) \|\mathbf{a}_1\|^2_{\mathbf{D}} \quad (8)$$

This formulation shows that MULTISPATI finds coefficients ($\mathbf{u}_1$) to obtain a linear combination of variables ($\mathbf{a}_1 = \mathbf{XQu}_1$) which maximizes a compromise between the classical multivariate analysis ($\|\mathbf{a}_1\|^2_{\mathbf{D}}$) and a generalized version of Moran's I ($I_{\mathbf{D}}(\mathbf{a}_1)$). The only difference between the generalized $I_{\mathbf{D}}$ and the classical Moran's I (eq. 3) is that the first one used a general matrix of weights $\mathbf{D}$ while the second considers only the usual case where $\mathbf{D} = \frac{1}{n}\mathbf{I}$.

In practice, it is preferable to diagonalize the $\mathbf{Q}$-symmetric matrix $\mathbf{H} = (1/2)(\mathbf{X}^t (\mathbf{W}^t \mathbf{D} + \mathbf{DW}) \mathbf{XQ})$ instead of $\mathbf{X}^t\mathbf{DWXQ}$ which is not symmetric. The maxima of eq. 8 is equal and given by the first eigenvalue (λ_1) of $\mathbf{H}$.

In the case of the normalised PCA, MULTISPATI is equivalent to Wartenberg's approach using a row-sum weighting scheme (more details are given in the Discussion section and in App. 1).

In order to test the statistical significance of the spatial structure of the table $\mathbf{X}$, a permutation procedure can be used. The statistic used is equal to *trace*($\mathbf{X}^t$ $\mathbf{DWXQ}$). The *p*-value is computed by comparing the observed value to those obtained by permutation of the rows of the table $\mathbf{X}$.

The MULTISPATI approach has been implemented in the R software as a function of the ade4 package (Chessel et al. 2004b). The data set analysed in this paper is also available in the package under the name 'vegtf'.

Application

Study area and data collection

Data were collected in the Territoire d'Etude et d'Expérimentation de Trois-Fontaines (TF), a 1360-ha enclosed forest. TF is situated in north-eastern France (48°43' N, 4°56' W) and has a continental climate, characterized by cold winters and hot summers. The forest overstorey is dominated by *Quercus* sp., *Fagus sylvatica* and *Carpinus betulus*. The data set contains information about vegetation accessible to roe deer (height < 1.20 m, Duncan et al. 1998). This data set was collected at the scale of 1-m² sample plots and has been geo-referenced and introduced in a Geographic Information System. This sample technique is part of a population dynamics study aiming to understand relationships between roe deer population and their available food. On each plot, the abundance-dominance of all vascular plant species was recorded using a 7-point scale (Braun-Blanquet 1932): absence; rare and cover < 5 %; abundant and cover < 5 %; 5 < cover < 25 %; 25 < cover < 50 %; 50 < cover < 75 %; 75 < cover < 100 %. For the analysis, these abundance values were coded from 0 to 7. In total, 337 plots systematically distributed on a grid (grid size = 333 m) were sampled and a neighbourhood graph was constructed using the queen definition (rectangular and diagonal connections). In total, 116 species were recorded. Only species which occur in at least four plots were kept for the analysis.

We applied CA and MULTISPATI-CA to the vegetation data set (337 plots, 80 species). One property of CA (scaling type 1) is searching for a score of species of unit weighted variance (weights are relative frequency of species), samples are plotted at the weighted centroids of the species using the same weights and CA maximizes the weighted variance of plots (weights are relative frequency of plots). Interpretation of MULTISPATI-CA is exactly the same except that the quantity maximized is the product of the weighted variance of plots (criteria maximized in CA) and the generalized version of Moran's I (using weights equal to the relative frequency of plots).

Results

A bar plot of the eigenvalues of CA (Fig 1a) showed a continuous decrease (4.02%, 3.72%, 3.38%, 3.24% and 3.07%) of the variability explained by the first five axes. It is then difficult to choose the number of axes to keep and this trend can be seen as a complete absence of ecological structure in the data. Subsequently, we tried to interpret the results using an ordination diagram of species (Fig. 1b) and mapping of scores of plots (Fig. 1c, d) for the first two axes of the analysis. CA is very sensitive to rare species which are abundant in poor plots (e.g., see Table 1 in Dray et al. 2003a). This is illustrated by the position of the most discriminated species: *Athyrium filix-femina*. This species occurs in 12 plots and 55.8 % of its occurrences are in plots where richness is less than 4 species. Species characterising (Fig. 1b) the negative side of the first axis of CA are essentially heliophilic species with low cover (*Dactylis glomerata*, *Plantago* sp., *Trifolium* sp.).

The positive side corresponds to sciaphilic species with high cover (*Fagus sylvatica*, *Ilex aquifolium*, *Fraxinus excelsior*). Axis 2 of CA opposes the pole of neutrophilic-calcareous species (*Adoxa moschatellina*, *Fraxinus excelsior*) to acidicline-acidiphilic species (*Athyrium filix-femina*, *Luzula pilosa*, *Convallaria majalis*) which are mostly located in the southern part of TF. Results of CA depict local structures and do not reveal

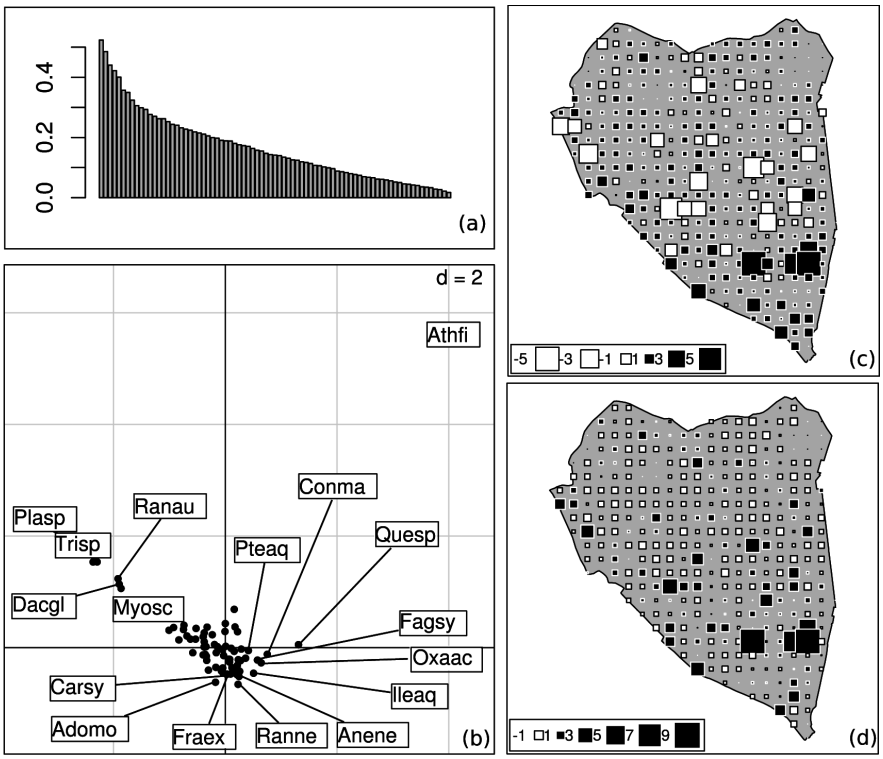

Fig. 1. Results of Correspondence analysis (CA). Eigenvalues (**a**), scores of species (**b**) and mapping of scores of plots on the first (**c**) and second (**d**) axis. The values of *d* indicates the size of squares of the grid. Species names: *Adoxa moschatellina* (Adomo), *Anemone nemorosa* (Anene), *Athyrium filix-femina* (Athfi), *Carex sylvatica* (Carsy), *Convallaria majalis* (Conma), *Dactylis glomerata* (Dacgl), *Fagus sylvatica* (Fagsy), *Fraxinus excelsior* (Fraex), *Ilex aquifolium* (Ileaq), *Myosotis scorpioides* (Myosc), *Oxalis acetosella* (Oxaac), *Plantago* sp. (Plasp), *Pteridium aquilinum* (Pteaq), *Quercus* sp. (Quesp), *Ranunculus auricomus* (Ranau), *Ranunculus nemorosus* (Ranne), *Trifolium* sp. (Trisp).

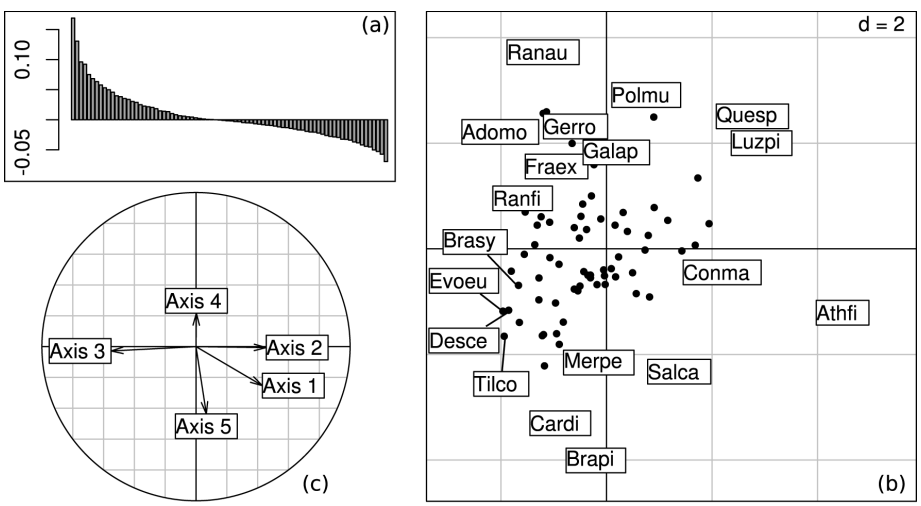

Fig. 2. Results of spatial correspondence analysis (MULTISPATI-CA). Eigenvalues (**a**), scores of species (**b**) and projections of the first five principal axes of CA onto the first two principal axes of MULTISPATI-CA (**c**). As principal axes are normalised, this figure represents correlations. The value of *d* indicates the size of squares of the grid. Species names: *Adoxa moschatellina* (Adomo), *Athyrium filix-femina* (Athfi), *Brachypodium pinnatum* (Brapi), *Brachypodium sylvaticum* (Brasy), (Cardi), *Convallaria majalis* (Conma), *Deschampsia cespitosa* (Desce), *Evonymus europaeus* (Evoeu), *Fraxinus excelsior* (Fraex), *Galium aparine* (Galap), *Geranium robertianum* (Gerro), *Luzula pilosa* (Luzpi), *Mercurialis perennis* (Merpe), *Polygonatum multiflorum* (Polmu), *Quercus* sp. (Quesp), *Ranunculus auricomus* (Ranau), *Ranunculus ficaria* (Ranfi), *Salix caprea* (Salca), *Tilia cordata* (Tilco).

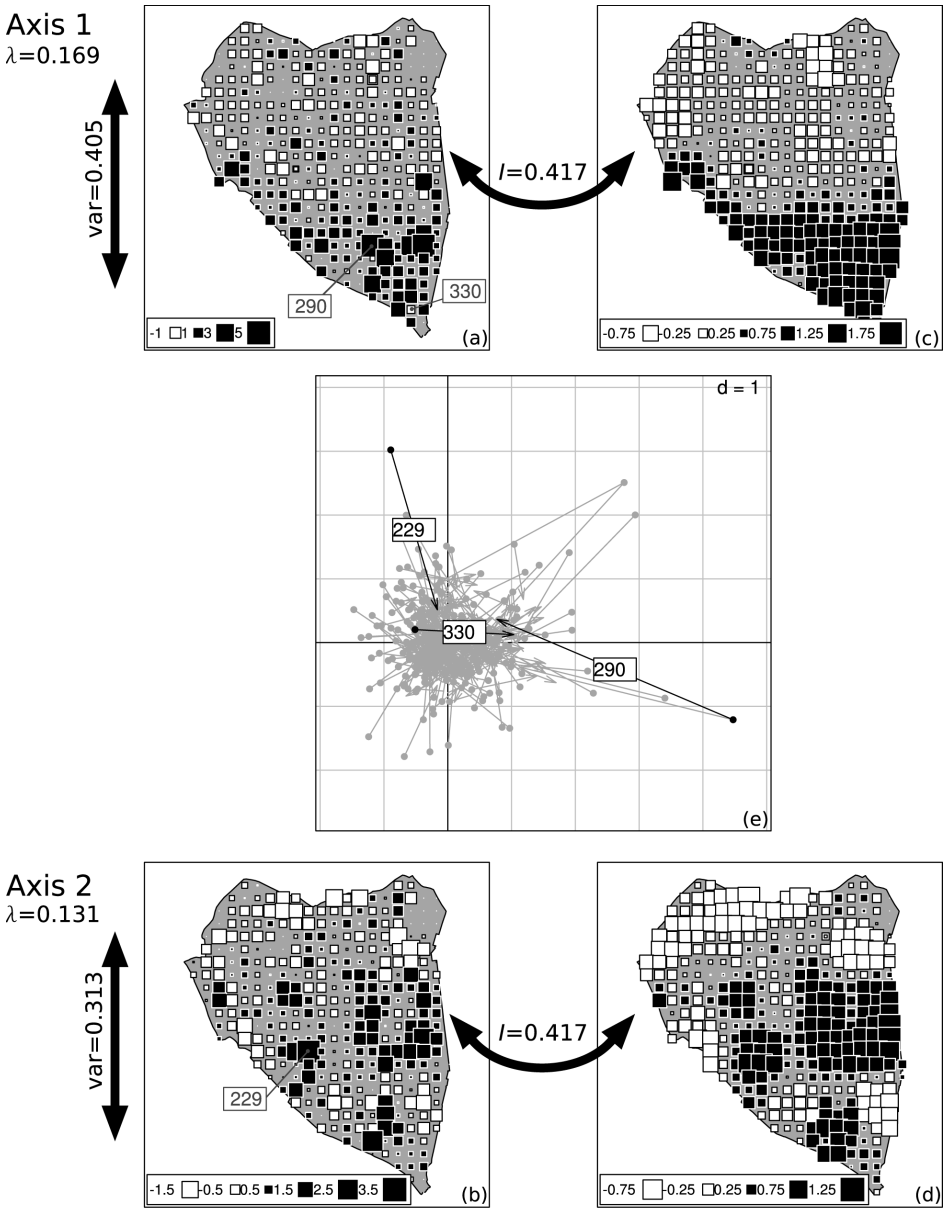

Fig. 3. Results of Spatial Correspondence Analysis (MULTISPATI-CA). Mapping of scores of plots on the first (**a**) and second (**b**) axis and of lagged score (averages of neighbours weighted by the spatial connection matrix) for the first (**c**) and second (**d**) axis. Representation of scores and lagged scores (**e**) of plots (for each site, the arrow links the score to the lagged score). Only plots discussed in the text are indicated by their labels. The value of *d* indicates the size of squares of the grid.

any clear spatial pattern (Fig. 1c, d). Maps of plots scores are then poorly spatially structured (Moran's *I* is equal to 0.086 and 0.080 for axis 1 and 2 respectively).

Results obtained by MULTISPATI-CA are easier to interpret (Figs. 2, 3). The bar plot of eigenvalues (Fig. 2a) suggests two main structures associated to the first two positive eigenvalues (corresponding to positive spatial autocorrelation). Ordination of species (Fig. 2b) shows that MULTISPATI-CA is much less sensitive to rare species than CA. The positive side of the first axis corresponds to acidicline species (e.g. *Athyrium filix-femina*, *Luzula pilosa*, *Convallaria majalis*) which are mainly located in the southern part of TF (Fig. 3a). Species on the negative side of the first axis are essentially neutrophilous and calcareous species, with low cover and often demanding in terms of soil quality (*Adoxa*

moschatellina, *Tilia cordata*, *Evonymus europaeus*, *Deschampsia cespitosa*, *Brachypodium sylvaticum*). Scores of plots are positively autocorrelated (Moran's I = 0.417) and their mapping reveals a north-south structure in TF. Second axis of MULTISPATI-CA opposes the pole of the calcareous species with *Brachypodium pinnatum*, *Mercurialis perennis*, *Carex digitata* (negative side) to nitrophilic species with *Geranium robertianum*, *Ranunculus auricomus*, *Fraxinus excelsior*, *Adoxa moschatellina* and *Galium aparine* (positive side). Map of plots scores (Fig. 3b) reveals a spatial structure (Moran's I = 0.417). Spatial structures detected by MUTLISPATI-CA appear as mixture of structures obtained by CA (Fig 2c). The first axis of MUTLISPATI-CA is negatively correlated with Axis 1 (– 0.431) and Axis 2 (– 0.456) of CA and positively correlated with Axis 3 (0.551). Second axis of MULTISPATI-CA is correlated with Axis 1 (– 0.253), Axis 4 (0.217) and Axis 5 (– 0.437) of CA. MULTISPATI-CA maximizes the product between the variance and the spatial autocorrelation of plots scores while CA maximizes only the variance. The loss of variance (due to the maximization of the product) is quite small: 0.405 versus 0.524 for axis 1 and 0.313 versus 0.482 for axis 2. On the other hand, the gain of spatial autocorrelation (Moran's I) is important: 0.417 versus 0.086 for axis 1 and 0.417 versus 0.080 for axis 2.

Spatial autocorrelation can be seen as the link between one variable and the lagged vector (eq. 6). Hence, the spatial part of MULTISPATI-CA can be analysed through the link between scores and lagged scores (Fig. 3). Each plot can be represented on the factorial map by an arrow (Fig. 3e, the bottom corresponds to its score, the head corresponds to its lagged score). A short arrow reveals a local spatial similarity (between one plot and its neighbours) while a long arrow reveals a spatial discrepancy. This viewpoint can be interpreted as a local index of spatial association (Anselin 1995). For instance, plot 330 has a quite long right horizontal arrow because its neighbours contain at least one acidicline species (*Athyrium filix-femina*, *Luzula pilosa*, *Convallaria majalis*, *Quercus* sp.) while it does not contain any one. On the other hand, plot 290 which has a long left arrow contains only *Athyrium filix-femina* while there is only one occurrence of acidicline species (*Luzula pilosa*) in one of its eight neighbours. For the second axis, plot 229 corresponds to a long vertical arrow because it contains only nitrophilic species (*Geranium robertianum*, *Ranunculus auricomus*) and its neighbours do not contain these species. Lastly, note that lagged scores (Fig. 3c, d) could be used to perform a spatial classification of the main vegetation types, an objective of prime interest in wildlife management (e.g. Pettorelli et al. 2005).

Discussion

One major objective of multivariate analysis is to describe the main ecological structures while the MULTISPATI approach aims to describe only spatial ecological structures. Hence, the use of MULTISPATI improves the description of spatial patterns but non-spatial information is discarded. This point of view can be related to the comparison between ordination (e.g. CA) and constrained ordination methods (e.g. Canonical Correspondence Analysis, CCA). In CCA, we decide to lose the optimality of CA in order to study species responses to environmental variables of interest. In MULTISPATI-CA, we decide to lose the optimality of CA in order to study spatial patterns of ecological structures. This choice is clearly related to the objective of the study.

The MULTISPATI approach is based on the introduction of the spatial weighting matrix in the statistical triplet notation. Hence, one can perform MULTISPATI-PCA for quantitative data, MULTISPATI-CA for contingency tables, MULTISPATI-MCA for qualitative data. As shown before (Eq. 1), Moran's I is negative when negative autocorrelation appears and our approach, which is based on this index, will produce negative eigenvalues. In the case of a high negative eigenvalue, it is important to inspect the associated eigenvectors which can depict local structures of interest (negative autocorrelation). On the contrary, Geary's c is always positive (Eq. 2) and that is probably why the first attempts to spatial multivariate analysis are based on this index. Following the initial work of Lebart (1969), many methods have been mainly developed by the French school of statisticians (Le Foll 1982; Benali & Escofier 1990; Chessel & Sabatier 1993; Méot et al. 1993) and by Italians in the context of multiscale analysis (Di Bella & Jona-Lasinio 1996). Although these methods have the advantage to produce only positive eigenvalues, they have a major drawback in their objectives: they maximize the local variance (i.e. difference between neighbours) while often users want to minimize this quantity and maximize the spatial correlation.

Wartenberg (1985) was the first to develop a multivariate analysis based on Moran's I. In his work, he proposed to diagonalize $\mathbf{M} = \mathbf{X}^t\mathbf{F}\mathbf{X}$, where $\mathbf{X}$ contains normed and centered variables, and presented examples where the connectivity matrix was based on inter-points distances and always symmetric. Lee (2001) criticized Wartenberg's work and proved that this approach is not valuable with row-sum standardized weights. He shows that in this case, the spatial bivariate association measure is not correct because it is asymmetric. The derivation of Lee (2001) is correct but rather naïve because this formulation yields an asymmetric matrix $\mathbf{M}$. Finding eigenvalues of such a matrix is difficult because they can be complex. The correct derivation must use $(\mathbf{W} + \mathbf{W}^t)$

instead of **W** (de Jong et al. 1984). In this case, Wartenberg's (1985) approach is exactly the MULTISPATI-(normalised) PCA. The third part of the appendix shows that in this case, the spatial bivariate association measure is symmetric and satisfies Lee's (2001) conditions cited above. Moreover, it is interesting to note that this measure considers the correlations between one original variable and another one lagged variable ($r_{\tilde{\mathbf{y}}_j \mathbf{y}_k}$, $r_{\mathbf{y}_j \tilde{\mathbf{y}}_k}$) and thus is intimately linked to the multivariate extension of a Moran scatterplot which plots $\tilde{\mathbf{y}}_j$ versus $\mathbf{y}_k$ or $\tilde{\mathbf{y}}_k$ versus $\mathbf{y}_j$ (Anselin et al. 2002).

The MULTISPATI approach maximizes the compromise between the original analysis and the autocorrelation (Eq. 8). Note that an additional constraint can be intoduced so that the MULTISAPTI analysis will maximize only the spatial part (i.e. the first part of the product in Eq. 8). One alternative for this purpose is to use the principal components of **X** instead of the table **X**. This corresponds to an implicit use of the Mahalanobis metrics (i.e. inverse of the variance-covariance matrix) and therefore to a previous decorrelation of the original variables. However, this step requires many observations compared to the number of variables in order to avoid problems of numerical instability. This special case of MULTISPATI on quantitative data is equivalent to the spatial factor analysis and their extensions (Switzer & Green 1984; Green et al. 1988; Grunsky & Agterberg 1991; Bailey & Krzanowski 2000).

The two points of view (Geary's *c* and Moran's *I*) have been reconciled by Thioulouse et al. (1995) who use the spatial weights from **B** to normalize the data. Although this approach is very elegant in a mathematical point of view, it requires that the mean and the variance of the original variables are computed taking into account the spatial connectivity. As summing the elements of **B** by row will not lead usually to uniform weights, the mean and the variance of the original variables are not invariant when permuting the rows of **X**. This is problematic because in the case of the null hypothesis (no spatial structure) of inferential tests, it would be required that all spatial units have the same weight in the computation of the mean and the variance. This problem does not appear in MULTISPATI test but further study is required in order to examine the inferential properties of the testing procedure.

Various methods have been proposed in ecology to take into account space in multivariate analysis. The most traditional approach considers two steps. Firstly, a table representing explicitly some spatial structures is constructed (this table is often called 'space' in scientific papers). Then, this table is used as a predictor or covariable in a (partial) canonical ordination method such as Redundancy Analysis (RDA, Rao 1964). Methods of variation partitioning (Peres-Neto et al. 2006) can then be used to evaluate the part of the variation in species composition which is due to space, environment or both. In the literature, different tools have been proposed to create the space table. Borcard et al. (1992) used a polynomial of degree 3 while Borcard & Legendre (2002) developed the PCNM approach. Dray et al. (2006) demonstrated that the PCNM method is a particular case of the more general framework of Moran's eigenvector maps (MEM). MEM are the eigenvectors of a double-centered spatial weighting matrix. MEM provides a set of orthogonal predictors which maximize the spatial autocorrelation (i.e. Moran's *I*). In a theoretical point of view, MEM and MULTISPATI are quite close: they use a spatial weighting matrix and measure spatial autocorrelation by Moran's *I*. However, when MULTISPATI seeks for linear combinations of variables that maximize the product of the autocorrelation by the variance, RDA with MEM maximizes the variance explained by the spatial descriptors (which maximize spatial autocorrelation). This theoretical difference implies some practical considerations. For a table with *n* rows (sites), the number of MEM, which are used as descriptors or covariables in RDA, can be equal to $(n - 1)$. In this case, the regression step of RDA can be problematic because the species table would be completely predicted by the high number of spatial descriptors. It is thus necessary to reduce the number of MEM before the regression. Classical forward selection is too liberal in this context (many orthogonal predictors) and tends to select too much regressors (see Dray et al. 2006 for more details). MULTISPATI maximizes directly the spatial autocorrelation and so avoids all the problems related to the regression step of RDA. Hence, MULTISPATI could be preferred if one want to study the spatial structures of one data set. If one want also to include other descriptors (e.g. environmental variables) than space and to perform variation partitioning, MULTISPATI can not be used as it works with autocorrelations and not variances. If we suppose that a subset of MEM has been properly selected, RDA with MEM has to be preferred for this objective. In this context, methodological developments are required to extend the MULTISPATI approach to the case of methods for relating two data tables such as Redundancy Analysis or co-inertia analysis (Dolédec & Chessel 1994; Dray et al. 2003a).

Acknowledgements. This work was made possible by financial support from the 'Office National de la Chasse et de la Faune Sauvage' and co-managed by the 'Office National des Forêts'(ONF) and the 'Office National de la Chasse et de la Faune Sauvage' (ONCFS). We thank Daniel Chessel, Jean-Luc Dupouey, Jean-Michel Gaillard, François Klein and Sebastien Ollier for valuable discussions and review of earlier versions of this manuscript. We are grateful to Olivier Widmer and Olive Saïd for field assistance. We also thank Jari Oksanen and two anonymous reviewers.

References

Anselin, L. 1995. Local indicators of spatial association. *Geograpical Analysis* 27: 93-115.

Anselin, L. 1996. The Moran scatterplot as an ESDA tool to assess local instability in spatial association. In: Fischer, M.M., Scholten, H.J. & Unwin, D. (eds.) *Spatial analytical perspectives on GIS*, pp. 111-125. Taylor and Francis, London, UK.

Anselin, L., Syabri, I. & Smirnov, O. 2002. Visualizing multivariate spatial correlation with dynamically linked windows. In: Anselin, L. & Rey, S. (eds.) *New tools for spatial data analysis: Proceedings of a Workshop*, CSISS, Santa-Barbara, CA, US.

Bailey, T.C. & Krzanowski, W.J. 2000. Extensions to spatial factor methods with an illustration in geochemistry. *Mathematical Geology* 32: 657-682.

Borcard, D., Legendre, P. & Drapeau, P. 1992. Partialling out the spatial component of ecological variation. *Ecology* 73: 1045-1055.

Borcard, D. & Legendre, P. 2002. All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecological Modelling* 153: 51-68.

Bavaud, F. 1998. Models for spatial weights: A systematic look. *Geograpical Analysis* 30: 153-171.

Benali, H. & Escofier, B. 1990. Analyse factorielle lissée et analyse factorielle des différences locales. *Revue de Statistique Appliquée* 38: 55-76.

Braun-Blanquet, J. 1932. *Plant sociology, the study of plant communities*. McGraw Hill, New-York, NY, US.

Cailliez, F. & Pagès, J.P. 1976. *Introduction à l'analyse des données*. SMASH, Paris, FR.

Chessel, D. & Sabatier, R. 1993. Couplage de triplets statistiques et graphes de voisinage. In: Asselain, B., Boniface, M., Duby, C., Lopez, C., Masson, J.P. & Tranchefort, J. (eds.) *Biométrie et données spatio-temporelles*, pp. 28-37. Société Française de Biométrie, ENSAR, Rennes, FR.

Chessel, D., Ollier, S. & Dray, S. 2004a. *Ordination sous contraintes spatiales*. http://pbil.univ-lyon1.fr/R/stage/stage8.pdf

Chessel, D., Dufour, A.B. & Thioulouse, J. 2004b. The ade4 package-I- One-table methods. *R News* 4: 5-10.

Cliff, A.D. & Ord, J.K. 1973. *Spatial autocorrelation*. Pion, London, UK.

Cormack, R.M. & Ord, J.K. 1979. *Spatial and temporal analysis in ecology*. International Co-operative Publishing House, Fairland.

de Jong, P., Sprenger, C. & van Veen, F. 1984. On extreme values of Moran's *I* and Geary's *c*. *Geograpical Analysis* 16: 17-24.

Di Bella, G. & Jona-Lasinio, G. 1996. Including spatial contiguity information in the analysis of multispecific patterns. *Environmental and Ecological Statistics* 3: 269-280.

Dolédec, S. & Chessel, D. 1994. Co-inertia analysis: an alternative method for studying species-environment relationships. *Freshwater Biology* 31: 277-294.

Dray, S., Chessel, D. & Thioulouse, J. 2003a. Co-inertia analysis and the linking of ecological data tables. *Ecology* 84: 3078-3089.

Dray, S., Chessel, D. & Thioulouse, J. 2003b. Procrustean co-inertia analysis for the linking of multivariate data sets. *Ecoscience* 10: 110-119.

Dray, S., Pettorelli, N. & Chessel, D. 2003c. Multivariate analysis of incomplete mapped data. *Transactions in GIS* 7: 411-422.

Dray, S., Legendre, P. & Peres-Neto, P.R. 2006. Spatial modeling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *Ecological Modelling* 196: 483-493.

Duncan, P., Tixier, H., Hoffman, R.R. & Lechner-Doll, M. (1998). Feeding strategies and the physiology of digestion in roe deer. In: Andersen, R., Duncan, P. & Linnell, J.D.C. (eds.) *The European roe deer: the biology of success*, pp. 91-116. Scandinavian University Press, Oslo, NO.

Escoufier, Y. 1987. The duality diagram: a means of better practical applications. In: Legendre, P. & Legendre, L. (eds.) *Developments in numerical ecology*, pp. 139-156. Springer Verlag, Berlin, DE.

Geary, R.C. 1954. The contiguity ratio and statistical mapping. *The Incorporated Statistician* 5: 115-145.

Getis, A. & Griffith, D.A. 2002. Comparative spatial filtering in regression analysis. *Geograpical Analysis* 34: 130-140.

Goodall, D.W. 1954. Objective methods for the classification of vegetation III. An essay on the use of factor analysis. *Australian Journal of Botany* 2: 304-324.

Green, A.A., Berman, M., Switzer, P. & Graig, M.D. 1988. A transformation for ordering multispectral data in terms of image quality with implications for noise removal. IEEE *Transactions in Geoscience Remote Sensing* 26: 65-74.

Greenacre, M.J. 1984. *Theory and applications of Correspondence Analysis*. Academic Press, London, UK.

Griffith, D.A. 1996. Spatial autocorrelation and eigenfunctions of the geographic weights matrix accompanying georeferenced data. *The Canadian Geographer* 40: 351-367.

Griffith, D.A. 2000a. Eigenfunction properties and approximations of selected incidence matrices employed in spatial analyses. *Linear Algebra Applications* 321: 95-112.

Griffith, D.A. 2000b. A linear regression solution to the spatial autocorrelation problem. *Journal of Geographical Systems* 2: 141-156.

Grunsky, E.C. & Agterberg, F.P. 1991. SPFAC: a FORTRAN-77 program for spatial factor analysis of multivariate data. *Computational Geosciences* 17: 133-160.

Hotelling, H. 1933. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology* 24: 417-441.

Kadmon, R. & Danin, A. 1997. Floristic variation in Israel: a GIS analysis. *Flora* 192: 341-345.

Kiers, H.A.L. 1994. Simple structure in component analysis techniques for mixtures of qualitative and quantitative variables. *Psychometrika* 56: 197-212.

Krishna Iyer, P.V. 1949. The first and second moments of some probability distributions arising from points on a lattice and their application. *Biometrika* 36: 135-141.

Le Foll, Y. 1982. Pondération des distances en analyse factorielle. *Statistique et Analyse des Données* 7: 13-31.

Lebart, L. 1969. Analyse statistique de la contiguïté. Publications de l'lnstitut Statistique Universitaire de Paris 28: 81-112.

Lee, S.-I. 2001. Developing a bivariate spatial association measure: An integration of Pearson's *r* and Moran's *I*. *Journal of Geographical Systems* 3: 369-385.

Méot, A., Chessel, D. & Sabatier, R. 1993. Opérateurs de voisinage et analyse des données spatio-temporelles. In: Lebreton, J.D. & Asselain, B. (eds.) *Biométrie et environnement*, pp. 45-72. Masson, Paris, FR.

Moran, P.A.P. 1948. The interpretation of statistical maps. *Journal of the Royal Statistical Society Ser. B* 10: 243-251.

Ord, J.K. 1975. Estimation methods for models of spatial interaction. *Journal of the American Statistical Association* 70: 120-126.

Peres-Neto, P.R., Legendre, P., Dray, S. & Borcard, D. 2006. Variation partitioning of species data matrices: estimation and comparison of fractions. *Ecology* 87: 2614-2625.

Pettorelli, N., Dray, S. & Maillard, D. 2005. Coupling principal component analysis and GIS to map deer habitats. *Wildlife Biology* 11: 363-370.

Rao, C.R. 1964. The use and interpretation of principal component analysis in applied research. *Sankhya A* 26: 329-359.

Selmi, S., Boulinier, T. & Faivre, B. 2003. Distribution and abundance patterns of a newly colonizing species in Tunisian oases: the Common Blackbird *Turdus merula*. *Ibis* 145: 681-688.

Switzer, P. & Green, A.A. 1984. Min/max autocorrelation factors for multivariate spatial imagery. In: pp. 23. Technical Report 6, Stanford University, Stanford, CA, US.

Tenenhaus, M. & Young, F.W. 1985. An analysis and synthesis of multiple correspondence analysis, optimal scaling, dual scaling, homogeneity analysis and other methods for quantifying categorical multivariate data. *Psychometrika* 50: 91-119.

ter Braak, C.J.F. 1985. Correspondence analysis of incidence and abundance data: properties in terms of a unimodal reponse model. *Biometrics* 41: 859-873.

ter Braak, C.J.F. & Gremmen, J.M. 1987. Ecological amplitudes of plant species and the internal consistency of Ellenberg's indicator values for moisture. *Vegetatio* 69: 79-87.

Thioulouse, J., Chessel, D. & Champely, S. 1995. Multivariate analysis of spatial patterns: a unified approach to local and global structures. *Environmental and Ecological Statistics* 2: 1-14.

Tiefelsdorf, M., Griffith, D.A. & Boots, B. 1999. A variance-stabilizing coding scheme for spatial link matrices. *Environmental Planning* A 31: 165-180.

Torre, F. & Chessel, D. 1995. Co-structure de deux tableaux totalement appariés. *Revue de Statistique Appliquée* 43: 109-121.

Tutin, T.G., Heywood, V.H., Burges, N.A., Valentine, D.H., Walters, S.M. & Webb, D.A. 2001. *Flora Europaea, Vols. 1-5*. Cambridge University Press, Cambridge, UK.

Wartenberg, D. 1985. Multivariate spatial correlation: a method for exploratory geographical analysis. *Geographical Analysis* 17: 263-283.

Whittaker, R.H. 1967. Gradient analysis of vegetation. *Biological Review* 42: 207-264.

Received 29 November 2006;
Accepted 11 May 2007;
Co-ordinating Editor: J. Oksanen.

App. 1. Mathematics

Classical multivariate analysis:

Each multivariate method (e.g., PCA, CA, …) corresponds to a statistical triplet (**X**, **Q**, **D**) where **X** is a $(n \times p)$ matrix derived from any data table, **D** be a scalar product of $\mathbb{R}^n$ (*n* by *n* symmetric matrix) and **Q** be scalar product of $\mathbb{R}^p$ (*p* by *p* symmetric matrix). The analysis of a triplet consists of finding a vector $\mathbf{u}_1$ (first principal axis) so that the quadratic form:

$$Q(\mathbf{u}_1) = \left\|\mathbf{XQu}_1\right\|^2_{\mathbf{D}} = \mathbf{u}_1{}^t\mathbf{QX}^t\mathbf{DXQu}_1 \tag{A.1}$$

is maximized under the constraints that $\left\|\mathbf{u}_1\right\|^2_{\mathbf{Q}} = \mathbf{u}_1{}^t\mathbf{Qu}_1 = 1$.

If *r* is the rank of the matrix **X**, then the second and further principal axes ($\mathbf{u}_2$, $\mathbf{u}_3$, …, $\mathbf{u}_r$) maximize the same quantity, but are subjected to extra constraints of orthogonality, i.e. for all $s \neq t$ $\left(\mathbf{u}_s \mid \mathbf{u}_t\right)_{\mathbf{Q}} = 0$.

In practice, the solution vectors $\mathbf{u}_j$ ($1 \le j \le r$) are obtained as the right-hand eigenvectors of $\mathbf{X}^t\mathbf{DXQ}$, and the maximum of $Q(\mathbf{u}_j)$ is equal and given by the *j*-th eigenvalue λ_j.

The general framework of MULTISPATI:

MULTISPATI corresponds to the diagonalization of the statistical triplet $(\mathbf{X}, \mathbf{Q}, 1/2(\mathbf{W}^t\mathbf{D} + \mathbf{DW}))$. It seeks for a vector $\mathbf{u}_1$ (with $\left\|\mathbf{u}_1\right\|^2_{\mathbf{Q}} = 1$) maximizing the quantity:

$$\begin{aligned} Q(\mathbf{u}_1) &= \mathbf{u}_1{}^t\mathbf{Q}^t\mathbf{X}^t(1/2(\mathbf{W}^t\mathbf{D} + \mathbf{DW}))\mathbf{XQu}_1 \\ &= 1/2(\mathbf{u}_1{}^t\mathbf{Q}^t\mathbf{X}^t\mathbf{W}^t\mathbf{DXQu}_1 + \mathbf{u}_1{}^t\mathbf{Q}^t\mathbf{X}^t\mathbf{DWXQu}_1) \\ &= 1/2\left(\mathbf{XQu}_1 \mid \mathbf{WXQu}_1\right)_{\mathbf{D}} + 1/2\left(\mathbf{WXQu}_1 \mid \mathbf{XQu}_1\right)_{\mathbf{D}} \\ &= \mathbf{u}_1{}^t\mathbf{Q}^t\mathbf{X}^t\mathbf{DWXQu}_1 = \mathbf{a}_1{}^t\mathbf{DWa}_1 = \mathbf{a}_1{}^t\mathbf{D\tilde{a}}_1 \end{aligned} \tag{A.2}$$

Equation (A.2) can be rewritten as:

$$\begin{aligned} Q(\mathbf{u}_1) &= \frac{\mathbf{u}_1{}^t\mathbf{Q}^t\mathbf{X}^t\mathbf{DWXQu}_1}{\mathbf{u}_1{}^t\mathbf{Q}^t\mathbf{X}^t\mathbf{DXQu}_1}\mathbf{u}_1{}^t\mathbf{Q}^t\mathbf{X}^t\mathbf{DXQu}_1 \\ &= I_{\mathbf{D}}\left(\mathbf{XQu}_1\right)\left\|\mathbf{XQu}_1\right\|^2_{\mathbf{D}} = I_{\mathbf{D}}\left(\mathbf{a}_1\right)\left\|\mathbf{a}_1\right\|^2_{\mathbf{D}} \end{aligned} \tag{A.3}$$

MULTISPATI-(normalised)-PCA:

In the case of the normalised PCA of an original table **Y** (i.e., $\mathbf{X} = [x_{ij}] = \left[(y_{ij} - \overline{y}_j) / \sigma_j \right]$), the elements in matrix **H** can be written as:

$$
\begin{aligned}
\mathbf{H}_{jk} &= (1/2)(\mathbf{x}_j{}^t(\mathbf{W}^t\mathbf{D} + \mathbf{D}\mathbf{W})\mathbf{x}_k\mathbf{Q}) \\
&= \frac{1}{2n}\sum_{i=1}^{n} \frac{\tilde{y}_{ij} - \overline{y}_j}{\sigma_j}\frac{y_{ik} - \overline{y}_k}{\sigma_k} + \frac{1}{2n}\sum_{i=1}^{n} \frac{y_{ij} - \overline{y}_j}{\sigma_j}\frac{\tilde{y}_{ik} - \overline{y}_k}{\sigma_k} \\
&= \frac{1}{2}\sum_{i=1}^{n} \frac{\tilde{y}_{ij} - \overline{y}_j}{\sqrt{\sum_{i=1}^{n}(y_{ij} - \overline{y}_j)^2}}\frac{y_{ik} - \overline{y}_k}{\sqrt{\sum_{i=1}^{n}(y_{ik} - \overline{y}_k)^2}} \\
&+ \frac{1}{2}\sum_{i=1}^{n} \frac{y_{ij} - \overline{y}_j}{\sqrt{\sum_{i=1}^{n}(y_{ij} - \overline{y}_j)^2}}\frac{\tilde{y}_{ik} - \overline{y}_k}{\sqrt{\sum_{i=1}^{n}(y_{ik} - \overline{y}_k)^2}} \\
&\cong \frac{1}{2}\sqrt{SSS_{\mathbf{y}_j}}\, r_{\tilde{\mathbf{y}}_j\mathbf{y}_k} + \frac{1}{2}\sqrt{SSS_{\mathbf{y}_k}}\, r_{\tilde{\mathbf{y}}_k\mathbf{y}_j}
\end{aligned}
\qquad \text{(A.4)}
$$

and using (6) and (A.4), it is easy to show that $\mathbf{H}_{kk} = I(\mathbf{y}_k)$. In this case, MULTISPATI is equivalent to Wartenberg's approach using a row-sum weighting scheme.