



**HAL**  
open science

## Co-inertia analysis and the linking of ecological tables

Stéphane Dray, Daniel Chessel, Jean Thioulouse

► **To cite this version:**

Stéphane Dray, Daniel Chessel, Jean Thioulouse. Co-inertia analysis and the linking of ecological tables. *Ecology*, 2003, 84 (11), pp.3078-3089. <10.1890/03-0178>. <hal-00427392>

**HAL Id: hal-00427392**

**<https://hal.science/hal-00427392v1>**

Submitted on 26 Aug 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY-NC 4.0 - Attribution - Non-commercial use - International License

## CO-INERTIA ANALYSIS AND THE LINKING OF ECOLOGICAL DATA TABLES

STÉPHANE DRAY,<sup>1</sup> DANIEL CHESSEL, AND JEAN THIOULOUSE

UMR CNRS 5558, Laboratoire de Biométrie et Biologie Evolutive, Université Claude Bernard Lyon 1,  
69622 Villeurbanne Cedex, France

**Abstract.** Ecological studies often require studying the common structure of a pair of data tables. Co-inertia analysis is a multivariate method for coupling two tables. It is often neglected by ecologists who prefer the widely used methods of redundancy analysis and canonical correspondence analysis. We present the co-inertia criterion for measuring the adequacy between two data sets. Co-inertia analysis is based on this criterion as are canonical correspondence analysis or canonical correlation analysis, but the latter two have additional constraints. Co-inertia analysis is very flexible and allows many possibilities for coupling. Co-inertia analysis is suitable for quantitative and/or qualitative or fuzzy environmental variables. Moreover, various weighting of sites and various transformations and/or centering of species data are available for this method. Hence, more ecological considerations can be taken into account in the statistical procedures. Moreover, the principle of this method is very general and can be easily extended to the case of distance matrices or to the case of more than two tables. Simulated ecological data are used to compare the co-inertia approach with other available methods.

**Key words:** canonical correlation analysis; canonical correspondence analysis; co-inertia analysis; multi-table analysis; ordination; redundancy analysis; statistical analysis.

### INTRODUCTION

Problems in applied or theoretical ecology often deal with the study of a pair of numerical data tables. Some relate species traits to species composition (Ojeda et al. 1998) or habitat utilization to species traits (Willby et al. 2000), while others link experimental conditions (e.g., geographic locations or date) to species composition (Dolédec and Chessel 1987). One of the major tasks of ecological studies is to analyze the response of community composition to environmental conditions and this often requires the use of multivariate analyses. Gauch (1982:1) summarizes the reasons for this choice in a clear and concise way:

*Community ecology concerns assemblages of plants and animals living together and the environmental and historical factors with which they interact. . . . Community data are multivariate because each sample site is described by the abundances of a number of species, because numerous environmental factors affect communities, and so on. . . . The application of multivariate analysis to community ecology is natural, routine and fruitful.*

Ordination methods allow detecting the underlying data structure, but “a major purpose is interpretation of community relationships to environment, and not simply the representation of numerical relationships among samples or species in a hyperspace with a lim-

ited number of axes” (Gauch and Wentworth 1976:17). Among the available methods for this task are: canonical correlation analysis (CANCOR, Hotelling 1936); principal component analysis with instrumental variables (Rao 1964), also named “redundancy analysis” (RDA); co-inertia analysis (COIA, Dolédec and Chessel 1994); and canonical correspondence analysis (CCA, ter Braak 1986).

Various kinds of data such as numbers of individuals, presence absence data, abundance indexes, biomass, etc., can fill the species table. Environmental conditions are recorded at each site by way of quantitative, qualitative, or fuzzy variables (Chevenet et al. 1994). In addition, ecologists have learned to use various transformations of environmental data as well as species data (Noy-Meir 1973, Noy-Meir et al. 1975) before performing multivariate analysis.

The diversity of biological questions has led to the study of various living creatures, implying a diversity of data types, a diversity of numerical conditions, and a diversity of statistical approaches. In a given situation, characterized by the properties of collected data and the objectives of the study, the choice of the “good” statistical method can be very difficult for ecologists. Unfortunately, this choice is often guided by practical considerations such as the possibilities proposed by a statistical software package, and the theoretical considerations concerning the characteristics of the data and the objectives of the study are neglected. Co-inertia analysis, which has been implemented only in the ADE-4 software (Thioulose et al. 1997), is much less used than CCA (Birks et al. 1996). The var-

Manuscript received and accepted 17 March 2003 (originally submitted 19 July 2002). Corresponding Editor: B. Shipley.

<sup>1</sup> E-mail: dray@biomserv.univ-lyon1.fr

ious statistical methods available for ecological studies do not provide optimal results for all ecological situations. Now, almost 10 years after Palmer's (1993) paper in *Ecology*, the controversy about the choice of methods for analysis and detrending (Wartenberg et al. 1987, Peet et al. 1988, Jackson and Somers 1991) has nearly settled down. Indeed, RDA and CCA have become the most widely used methods. The success of CCA (more than 800 references vs. 80 references for COIA [based on Science Citation Index, through April 2002]) is probably due to the availability of this method in statistical packages and to the previous success of correspondence analysis (Hill 1974). But this success should not hide the fact that CCA is only suited to gradient analysis (Palmer 1993) and is not always appropriate for the coupling of two tables. Indeed, CCA is very stringent and requires that the species table is analyzed by correspondence analysis (CA) and that the sites are weighted by their richness. These considerations are not suitable for all situations.

In this paper, we present the principles of COIA, showing the numerous possibilities available for coupling two tables. The co-inertia criterion, which has been often neglected, is presented and appears as a central concept when analyzing a pair of tables. Moreover, we emphasize the generality of the principle of COIA that can be extended to the case of linking more than two tables.

A GLOBAL MEASURE OF CO-STRUCTURE

Different statistics such as Pearson correlation coefficient or covariance can be used to measure the relation between two variables. The purpose of this section is to define a statistic that measures the relation between two (or more) sets of variables. Let  $\tilde{\mathbf{X}}$  be a table containing the values of  $p$  environmental variables (columns) measured at  $n$  sites (rows). Each site can be represented as a point in an ecological hyperspace with  $p$  dimensions where each axis represents an environmental variable. If  $\mathbf{D}$  is the diagonal matrix ( $n \times n$ ) of site weights ( $\mathbf{D} = \text{diag}(w_1, \dots, w_n)$ ) and if  $\mathbf{Q}$  ( $p \times p$ ) is a metric of this hyperspace, then the inertia of the "cloud of sites" around the reference point  $o$  is simply

$$I_0 = \sum_{i=1}^n w_i \|\tilde{\mathbf{X}}_i - o\|_{\mathbf{Q}}^2 = \sum_{i=1}^n w_i \|\mathbf{X}_i\|_{\mathbf{Q}}^2 = \text{trace}(\mathbf{XQX}^T\mathbf{D}). \tag{1}$$

This total inertia is a global measure of the variability of the data. It is the weighted sum of square distances measured with  $\mathbf{Q}$ , between the points of  $\tilde{\mathbf{X}}$  ( $n$  sites) and the reference point  $o$ . If  $\mathbf{Q}$  is the Euclidean metric and  $\mathbf{D}$  the diagonal matrix of uniform weights ( $w_i = 1/n$ ) and if  $o$  is the centroid of the cloud ( $o = [\bar{x}_1, \dots, \bar{x}_p]$  and  $\mathbf{X}$  is the centered-by-species table), the inertia is simply a sum of variances.

The sites  $\mathbf{X}_i$  can be projected on a  $\mathbf{Q}$ -normed vector  $\mathbf{u}$  and the projected inertia is expressed by

$$I(\mathbf{u}) = \mathbf{u}^T\mathbf{QX}^T\mathbf{DXQu}.$$

The total inertia can be easily decomposed on a set of  $p$  orthogonal  $\mathbf{Q}$ -normed vectors  $\mathbf{u}_k$ :

$$I_0 = \sum_{k=1}^p I(\mathbf{u}_k) = \sum_{k=1}^p \mathbf{u}_k^T\mathbf{QX}^T\mathbf{DXQu}_k = \sum_{k=1}^p \|\mathbf{XQu}_k\|_{\mathbf{D}}^2.$$

Let  $\tilde{\mathbf{Y}}$  be the species table, derived from an original table  $\tilde{\mathbf{Y}}$ , with  $n$  rows (sites) and  $q$  columns (species). In the species hyperspace, each site  $\mathbf{Y}_i$  is represented by a point. If each site has the same weight ( $\mathbf{D} = \text{diag}(w_1, \dots, w_n)$ ) and if metric  $\mathbf{R}$  is used in the species hyperspace, then the inertia is

$$J_0 = \text{trace}(\mathbf{YRY}^T\mathbf{D})$$

and can be decomposed as above on a set of vectors  $\mathbf{v}_k$ .

It is not much more difficult to study the common geometry of the two clouds. Co-inertia is a global measure of the co-structure of sites in the environmental and species hyperspaces: it is high when the two structures vary simultaneously (and also when they vary inversely), and low when they vary independently, or when they do not vary. It is defined by

$$\text{Co-}I = \sum_{k=1}^p \sum_{j=1}^q (\mathbf{u}_k^T\mathbf{QX}^T\mathbf{DYRv}_j)^2 = \sum_{k=1}^p \sum_{j=1}^q ([\mathbf{X}^k]^T\mathbf{DY}^j)^2 = \text{trace}(\mathbf{XQX}^T\mathbf{DYRY}^T\mathbf{D}).$$

If the clouds are centered, then inertia is a sum of variances and co-inertia is a sum of square covariances.

PRINCIPLES OF CO-INERTIA ANALYSIS

The co-inertia criterion measures the concordance between two data sets, and a multivariate method based on this statistic has been developed. Co-inertia analysis (Chessel and Mercier 1993, Dolédec and Chessel 1994) is a symmetric coupling method that provides a decomposition of the co-inertia criterion on a set of orthogonal vectors. It is defined by the analysis of statistical triplet  $(\mathbf{Y}^T\mathbf{DX}, \mathbf{Q}, \mathbf{R})$ . Different types of data lead to different transformations (centering, normalization, ...) of  $\mathbf{X}$  and  $\mathbf{Y}$  and to different metrics (weights)  $\mathbf{Q}$  and  $\mathbf{R}$ . Co-inertia analysis aims to find a vector  $\mathbf{v}_1$  in the species space and a vector  $\mathbf{u}_1$  in the environmental space with maximal co-inertia. If  $\mathbf{X}$  and  $\mathbf{Y}$  are centered, then co-inertia analysis (COIA) maximizes the square covariance between the projection of  $\mathbf{X}$  on  $\mathbf{u}_1$  and the projection of  $\mathbf{Y}$  on  $\mathbf{v}_1$ :

$$\text{cov}^2(\mathbf{XQu}_1, \mathbf{YRv}_1) = \text{corr}^2(\mathbf{XQu}_1, \mathbf{YRv}_1) \times \text{var}(\mathbf{XQu}_1) \times \text{var}(\mathbf{YRv}_1)$$

i.e., (a) = (b)  $\times$  (c)  $\times$  (d).

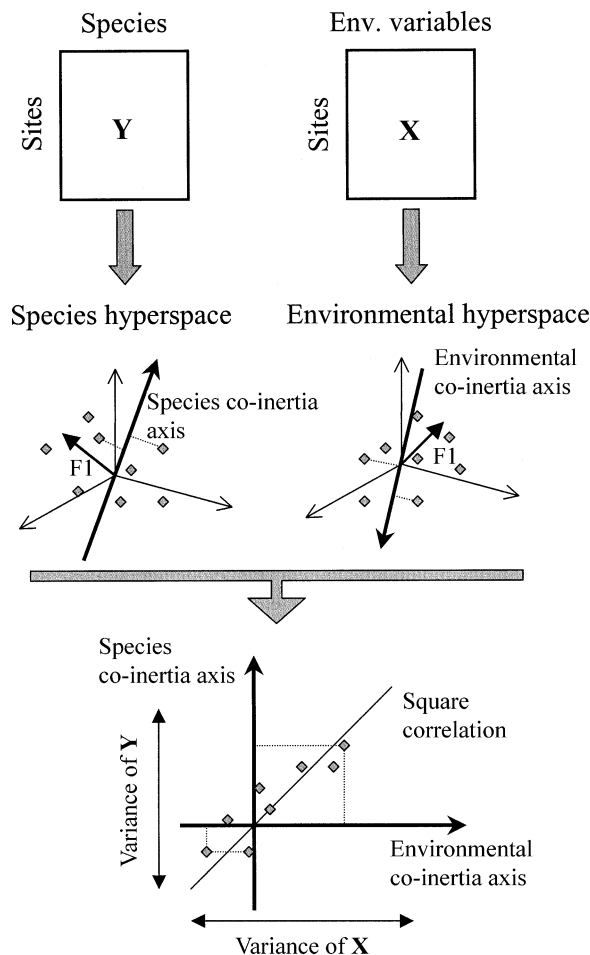


FIG. 1. Principles of co-inertia analysis (COIA). The two ecological data tables  $\mathbf{X}$  and  $\mathbf{Y}$  produce two representations of the sites in two hyperspaces. Separate analyses find axes maximizing inertia in each hyperspace (F1 [first factorial axis]). COIA aims to find a couple of co-inertia axes on which the sites are projected. COIA maximizes the square covariance between the projections of the sites on the co-inertia axes.

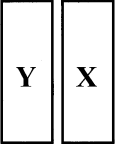
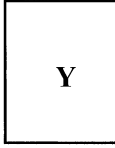

This square covariance (a) can be easily decomposed, showing that COIA finds a compromise between the correlation (b), the variance of sites in the species viewpoint (d) and the variance of sites in the environmental viewpoint (c) (Fig. 1). The second and further pairs of vectors ( $\mathbf{u}_2, \mathbf{v}_2, \dots$ ) maximize the same quantity but are subject to extra constraints of orthogonality. CANCOR (canonical correlation analysis), CCA (canonical correspondence analysis), and RDA (redundancy analysis) also maximize the square covariance but with additional constraints (Fig. 2) influencing the robustness of the analysis relative to the number of variables. CANCOR is defined by the use of two Mahalanobis metrics  $(\mathbf{X}^T \mathbf{D} \mathbf{X})^{-1}$  and  $(\mathbf{Y}^T \mathbf{D} \mathbf{Y})^{-1}$ ; CCA and RDA have only one Mahalanobis metric  $(\mathbf{X}^T \mathbf{D} \mathbf{X})^{-1}$ . The Mahalanobis metric takes into account the correlation in the data since it is calculated using the inverse

of the variance-covariance matrix (De Maesschalck et al. 2000). The use of this metric adds constraints in the analysis and its calculation implies precautions concerning the dimensions of the tables when using CCA, RDA, or CANCOR. The additional constraints are that when COIA maximizes the square covariance, CANCOR maximizes the square correlation and CCA or RDA maximizes the product of the square correlation by the species variance, which is simply the variance explained by the instrumental variables (sensu Rao 1964) of  $\mathbf{X}$ . That is why CCA and RDA permit variance partitioning (e.g., Méot et al. 1998) by using partial ordination (Rao 1964, ter Braak 1988) to measure the variance explained by different sets of instrumental variables.

Co-inertia analysis is a general coupling method that maximizes the co-inertia between the variables of two tables. Separate tables  $\mathbf{X}$  and  $\mathbf{Y}$  can be analyzed by various methods, leading to different coupling methods. Hence, Reynaud and Thioulouse (2000) perform a correspondence analysis (CA) on  $\mathbf{Y}$  and a multiple correspondence analysis (MCA, Tenenhaus and Young 1985) on  $\mathbf{X}$  and link these two analyses through COIA (CA-MCA COIA). Various couplings have been performed in ecological studies, such as PCA-PCA COIA (Cadet and Thioulouse 1998; PCA = principle component analysis) or CA-PCA COIA (Lods-Crozet et al. 2001) but many other possibilities have not yet been used. The CA-PCA or CA-MCA COIA is very similar to CCA and the two approaches aim to find a site score that is a linear combination of environmental variables maximizing the variance of species centroid (i.e., separation of species niches). The only difference is that CCA has an additional constraint (the total variance must be equal to 1), and must be avoided in the case of numerous environmental variables.

COIA is very general and some existing methods appear as special cases of it. Inter-battery analysis (Tucker 1958) is mathematically equivalent to a PCA-PCA COIA. The MCA-MCA COIA is equivalent to the correspondence analysis of the Burt's table crossing two qualitative tables (Cazes 1980). When table  $\mathbf{X}$  contains qualitative variables and  $\mathbf{Y}$  contains species numbers, it is usual to cross tables  $\mathbf{X}$  and  $\mathbf{Y}$  to obtain a matrix containing the distribution of species among the environmental variables categories. A simple CA of this new table allows one to ordinate the species and the environmental classes (analysis of ecological profiles, Romane 1972, Montana and Greig-Smith 1990). Binary discriminant analysis (Strahler 1978), which has been used in ecology (Del Moral 1982, Huang and Del Moral 1988), is mathematically equivalent to Romane's CA. Although this approach allows plotting of species and environmental classes (Ben-Shahar 1987, Ben-Shahar and Skinner 1988), no information about ordination of sites is available. It is easy to demonstrate that CA of the table of ecological profiles is a CA-MCA COIA (Mercier et al. 1992).

FIG. 2. Criteria maximized in canonical correlation analysis, co-inertia analysis, and analysis with respect to instrumental variables. Canonical correlation analysis (CCA) requires that there be few species and few environmental variables compared to site number; analysis with respect to instrumental variables (e.g., CCA or RDA [redundancy analysis]) requires only that there be few environmental variables, while there is no constraint in the co-inertia approach.

Maximization of $\text{cov}^2(\mathbf{XQu}_i, \mathbf{YRv}_i) = \text{corr}^2(\mathbf{XQu}_i, \mathbf{YRv}_i) \times \text{var}(\mathbf{XQu}_i) \times \text{var}(\mathbf{YRv}_i)$		
with the constraints		
$\text{var}(\mathbf{XQu}_i) = 1$ $\text{var}(\mathbf{YRv}_i) = 1$	$\text{var}(\mathbf{XQu}_i) = 1$	No constraint
Canonical correlation analysis  	Analysis with respect to instrumental variables (CCA, RDA)  	Co-inertia analysis  

THE FLEXIBILITY OF CO-INERTIA ANALYSIS (COIA)

As seen before, co-inertia analysis allows many couplings of ecological data. The two tables can be analyzed by various analyses (e.g., CA [correspondence analysis], centered PCA [principle component analysis], normed PCA, etc.) with the only constraint that the sites are weighted in the same way for the two separate analyses. Hence, COIA can analyze quantitative, qualitative, or even fuzzy environmental variables. For quantitative variables, the environmental table can be analyzed by at least 10 different PCA options. In the case of qualitative variables, MCA (multiple correspondence analysis) is applied on the environmental table. Fuzzy correspondence analysis (FCA, Chevenet et al. 1994) is suitable for fuzzy variables. Hill and Smith analysis (Hill and Smith 1976, Kiers 1994) allows one to analyze qualitative and quantitative environmental variables simultaneously. All types of environmental variables can be incorporated in COIA and this flexibility is also available for the species data.

The choice of the analysis for species data is decisive for the coupling because different analyses imply different ecological considerations. The first element to take into account is the shape of the species response curve in relation to environmental variables. In the case of unimodal response, the niche centroid (i.e., the average of environmental variables per species) is a good summary of species distribution (ter Braak and Looman 1986). Computations of niche centroids are based on the weighted-average principle and are computed in table  $\mathbf{Y}^t\mathbf{D}\mathbf{X}$  when  $\mathbf{Y}$  has been transformed by  $y_{ij}/y_{+j} = p_{ij}$ . This transformation is introduced when  $\mathbf{Y}$  is analyzed by a CA, by a PCA on species profiles, or by a nonsymmetric correspondence analysis (NSCA, Lauro and D'ambra 1984, Gimaret-Carpentier et al. 1998a, Kroonenberg and Lombardo 1999). The square root of

the transformation into species profiles is also the basis of the Hellinger transformation (Legendre and Gallagher 2001).

The use of CA in the instrumental variables approach leads to CCA (canonical correspondence analysis), and the use of PCA on species profiles in the co-inertia approach leads to OMI analysis (outlying mean index analysis, Dolédec et al. 2000). But other coupling possibilities exist, such as a CA of  $\mathbf{Y}$  in the co-inertia approach. When species response is assumed to be linear, the species–environment relation is well summarized by correlation coefficients. A simple PCA on  $\mathbf{Y}$  can then be applied, and leads to RDA (redundancy analysis) in the instrumental-variables approach, and to inter-battery analysis (Tucker 1958) in the co-inertia approach. In the case of a short gradient, methods based on correlation coefficients or on weighted averaging produce similar results. There are many zeroes (i.e., species absences) before *or* after the species optimum and so linear or unimodal models can be fitted adequately to species responses. If the studied gradient is longer, there are many zeroes before *and* after the species optimum. In this case, a correlation coefficient cannot summarize the species response, and a linear model is also inadequate.

Transformations of  $\mathbf{Y}$  by  $y_{ij}/y_{+j} = p_{ij}$  (species profiles) or by  $y_{ij}/y_{i+} = p_{ji}$  (sites profiles) clearly has an influence on the orientation of the study. In the table of species profiles, species are considered as frequency distributions among sites. As abundance values are divided by species totals, the differences in global abundance between species are not taken into account. The analysis of this table is focused on the relative distribution of species over the sites and aims to compare the ecological preferences of species. This implies that the species is the unit of interest and that the study consists in species niche separation. Transformation of

**Y** into a table of sites profiles implies that species are considered as variables and sites as frequency distributions among species. Abundances values are divided by sites totals and so differences in species richness of sites are removed. In this case, the site is the unit of interest and the study is focused on the species composition of the site (ter Braak 1983). This approach is able to detect changes in the relative species composition of sites, which can be linked to changes in the environment and is suitable for biomonitoring purposes. If data are not modified then the unit of interest is *the occurrence* of a species in a site. The importance of a site is proportional to its richness, and the importance of a species is proportional to its abundance. Double profile ( $y_{ij}/y_{+j}y_{i+}$ ) is a compromise between sites-orientated studies and species-orientated studies that removes the variations of richness among sites and the variations of abundances among species. This option is used in CA, so it must be emphasized that CCA removes the abundance effect, and the information given by species absences is not considered. When the aim of the study is, for example, to analyze limiting factors or pollution effects, species absence is a piece of information and CCA should not be used. Species profiles or unmodified data allow one to take into account species absence and thus to reveal a limiting factor. This option is used in classical PCA (linear model) and in species profile PCA (unimodal model) and leads to OMI analysis (Dolédéc et al. 2000) in the co-inertia approach.

Various centerings can also be applied to table **Y**. Mathematically and geometrically, and also ecologically, centering involves a point of reference for the study. In the case of noncentered data, the point of reference is the all-zero record—an empty site or a species that is always absent. Information is all that deviates from this absolute zero and study is focused on absolute variations and not on the deviations from a simple model.

Centering by species implies that the reference point is a hypothetical site where the species composition is simply the mean species composition computed for all sites. Information is given by a site when it deviates from this hypothetical site and a species is taken into account if it departs from a uniform distribution over all sites. The study is focused on composition differences between sites and species and not on absolute composition.

Centering by site implies that the reference point is an average species for which the abundance in a site is a constant proportion of the species total abundance in this site. Information is given by a species only if its distribution differs from the distribution of the total abundance. A site is informative if its composition deviates from equal proportions of all species.

All these considerations have to be taken into account and the choice must and can only be made by ecologists.

The last element to take into account concerns weighting of species and sites. Species and sites can be uniformly weighted but one may wish to weight species proportionally to their global abundances or sites proportionally to their richness. Weighting is related to the ability of the sampling protocol to be representative of the community. Sampling selectivity is a reason for nonrepresentativeness and is due to the fact that many species are rare in the sample not because they are rare in the studied area, but because the collecting method is not efficient for capturing them (Bayley and Peterson 2001). Hence, information given by an abundant species is more reliable than that given by a rare species and more weight must be assigned to abundant species. Unequal sampling effort also introduces biases in the analysis (Cao et al. 2001, 2002) because sampling size is closely related to species richness. Species–area relationships (e.g., Soberon and Llorente 1993) have been often studied and are a consequence of two independent phenomena. The total number of individuals increases with area, leading to an increased probability of encountering more species with larger areas (Palmer and White 1994). Hence, estimation of species richness is more reliable in a rich site that has been more sampled than in a poor site, and that is why sites must be weighted by their richness. Unequal sampling effort can also affect species differently; with a low sampling effort it is more difficult to detect the presence of a rare species than that of an abundant species. In that case, sites must be weighted by their richness and species by their global abundance.

These different weighting options influence the total inertia of the statistical triplet (**Y**, **R**, **D**). If we consider the case of two diagonal matrices ( $\mathbf{D} = \text{diag}(w_1, \dots, w_n)$ ) and  $\mathbf{R} = \text{diag}(\rho_1, \dots, \rho_q)$ , the inertia is simply

$$J_0 = \sum_{i=1}^n \sum_{j=1}^q w_i \rho_j y_{ij}^2.$$

This equation shows that the total inertia can be partitioned into each cell of the table **Y**. According to the orientation of the study and to the weighting options, it is evident that part of inertia due to each cell of **Y** will be more or less influenced by different parameters (Table 1). Table 1 shows clearly that CA is a compromise between the species-profiles NSCA and sites-profiles NSCA (Gimaret-Carpentier et al. 1998b). Study of species niche differentiation by CA gives more weight to low-species-richness-sites than species-profiles NSCA. Study of composition of sites by CA gives more weight to rare species than sites-profiles NSCA. The numerous options presented in Table 1 are good starting point to develop new methods to analyze species–environment relationships. For example, a coupling method based on NSCA of the species table would be more efficient with rare species or poor sites than classical CCA based on CA (Couteron et al. 2003).

TABLE 1. Total inertia for various situations of weighting and profiles.

Weighting†	No profile	Sites profiles	Species profiles	Double profiles
	$P_{ij}$	$\frac{P_{ij}}{P_{i+}} = p_{ji}$	$\frac{P_{ij}}{P_{+j}} = p_{ij}$	$\frac{P_{ij}}{P_{i+}P_{+j}}$
Uniform weighting ( $\rho_j = 1, w_i = 1$ )	$\sum_{i=1}^n \sum_{j=1}^q P_{ij}^2$ (PCA)	$\sum_{i=1}^n \sum_{j=1}^q \frac{P_{ij}^2}{P_{i+}^2}$ (sites-profiles PCA)	$\sum_{i=1}^n \sum_{j=1}^q \frac{P_{ij}^2}{P_{+j}^2}$ (species-profiles PCA)	$\sum_{i=1}^n \sum_{j=1}^q \frac{P_{ij}^2}{P_{i+}^2 P_{+j}^2}$
Sites weighting ( $\rho_j = 1, w_i = p_{i+}$ )	$\sum_{i=1}^n \sum_{j=1}^q P_{ij}^2 P_{i+}$	$\sum_{i=1}^n \sum_{j=1}^q \frac{P_{ij}^2}{P_{i+}}$ (sites-profiles NSCA)	$\sum_{i=1}^n \sum_{j=1}^q \frac{P_{ij}^2 P_{i+}}{P_{+j}^2}$	$\sum_{i=1}^n \sum_{j=1}^q \frac{P_{ij}^2}{P_{i+} P_{+j}^2}$
Species weighting ( $\rho_j = p_{+j}, w_i = 1$ )	$\sum_{i=1}^n \sum_{j=1}^q P_{ij}^2 P_{+j}$	$\sum_{i=1}^n \sum_{j=1}^q \frac{P_{ij}^2 P_{+j}}{P_{i+}^2}$	$\sum_{i=1}^n \sum_{j=1}^q \frac{P_{ij}^2}{P_{+j}}$ (species-profiles NSCA)	$\sum_{i=1}^n \sum_{j=1}^q \frac{P_{ij}^2}{P_{i+}^2 P_{+j}}$
Double weighting ( $\rho_j = p_{+j}, w_i = p_{i+}$ )	$\sum_{i=1}^n \sum_{j=1}^q P_{ij}^2 P_{i+} P_{+j}$	$\sum_{i=1}^n \sum_{j=1}^q \frac{P_{ij}^2 P_{+j}}{P_{i+}}$	$\sum_{i=1}^n \sum_{j=1}^q \frac{P_{ij}^2 P_{i+}}{P_{+j}}$	$\sum_{i=1}^n \sum_{j=1}^q \frac{P_{ij}^2}{P_{i+} P_{+j}}$ (CA)

Notes: In order to simplify the equations, we consider uniform weights equal to 1 (in the place of  $1/n$  and  $1/p$ ), and the table of abundance values has been divided by its total ( $p_{ij} = y_{ij}/y_{++}$ );  $p_{i+} = y_{i+}/y_{++}$ ,  $p_{+j} = y_{+j}/y_{++}$ . Information in parentheses is the name of the statistical method based on the profile/weighting option directly above it; if there is no information in parentheses it means that no existing method is based on this profile/weighting option. CA = correspondence analysis, PCA = principle-component analysis, and NSCA = nonsymmetric correspondence analysis.

† For weightings,  $\rho_j$  is the weight associated to species  $j$ , and  $w_i$  is the weight associated to site  $i$ .

EXTENSIONS OF CO-INERTIA ANALYSIS

The concept of co-inertia appears as a central part in two-tables coupling methods, and is general enough to be extended to other cases (Esposito Vinzi 2001). If the use of Euclidean or chi-square distances is not satisfactory, ecological distances computed on species data may be preferred (Legendre and Anderson 1999, Legendre and Gallagher 2001, McArdle and Anderson 2001). This strategy is also available in COIA (Fig. 3b) and distance matrices can be analyzed to improve the ecological sense of the study (Dray et al. 2003). Moreover, the possibility to extend the co-inertia criterion to the case of more than two tables provides new efficient tools for the ecologist (Fig. 3). For example, RLQ analysis (Fig. 3c; Dolédec et al. 1996) is simply the extension of co-inertia analysis to analyze the relationships between species traits and environmental variables through a species-by-sites table (Legendre et al. 1997) (the RLQ method uses linear combination of the variables in table **R** [external information about rows] and of the variables in table **Q** [external information about columns] of maximal covariance weighted by data contained in table **L** [link table]. Dray et al. (2002) are adapting RLQ analysis to study the relationships between two data sets arising from different sampling schemes with different sampling locations. If the sites are partitioned, within-class or between-class co-inertia analysis (Fig. 3f; Franquet and Chessel 1994, Franquet et al. 1995) allows one to analyze the species–environment relationships while taking into account this partition. Co-inertia has also been extended to the

case of coupling  $k$  tables ( $k > 2$ ), under the name of “multiple co-inertia analysis” (Fig. 3d; Chessel and Hanafi 1996) to study spatio-temporal variations of species composition. Another extension, concordance analysis (Fig. 3e; Lafosse and Hanafi 1997) is suitable for coupling  $k$  tables with a reference table and can be used to study the relationships between the environment and various groups of species. Lastly, co-inertia is useful in the case of the analysis of a series of  $k$  pairs of tables (STATICO, Fig. 3g; Simier et al. 1999, Thioulouse et al. 2004). The STATICO method is particularly well adapted to the study of the modifications of species–environment relationships during several sampling years.

APPLICATION

In this section we provide an ecological example to illustrate the co-inertia approach. We use the program COMPAS (Minchin 1987) to create species responses to a two-dimensional environmental space. This program has been used in many papers to compare the results of different ordination methods (e.g., Austin et al. 1995, McCune 1997). The sampling pattern was a  $10 \times 10$  grid over the entire ecological space. Species responses were generated for 40 species without systematic trend and noise. Species responses are symmetrical and parameters have been chosen as follows: lograndom distribution of modal abundances, normal distribution of ranges on the two gradients, and uniform random distribution of modal coordinates on the gradients. The average richness was 21 species per sam-

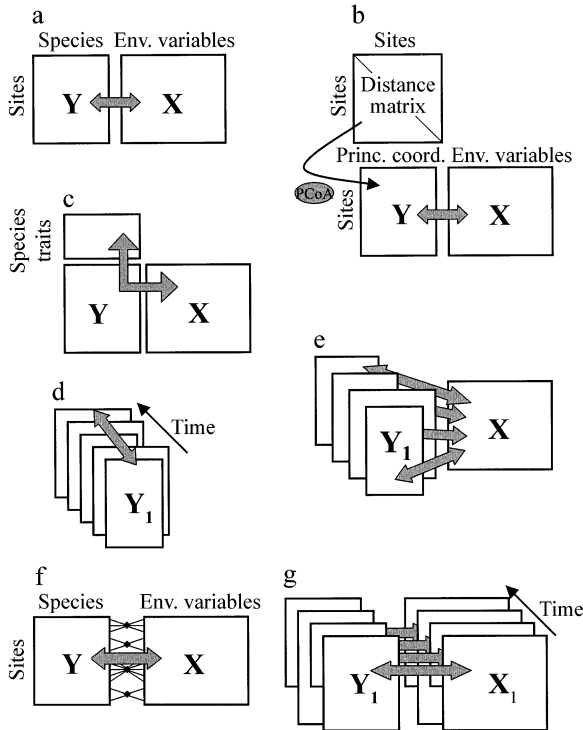


FIG. 3. Some extensions based on the co-inertia criterion. **Y** is a sites-by-species table, and **X** is a sites-by-environmental-variables table. (a) Classical co-inertia analysis. (b) Coupling an environmental table and principal coordinates obtained by the principal coordinate analysis (PCoA) of a distance matrix. (c) RLQ analysis to link environmental variables with species traits through **Y**. (d) Multiple co-inertia analysis to link  $k$  tables measuring  $k$  sets of variables for the same sites. (e) Concordance analysis to link  $k$  tables to a reference table. (f) Between-class or within-class co-inertia analysis to study the species-environment relationships according to groups of sites. (g) STATICO to link  $k$  couples of tables.

pling unit (range 6–32 species) and the  $\beta$  diversity for gradient 1 was 2.42 standardized half changes (1.02 R units) and for gradient 2, 1.71 standardized half changes (0.99 R units) (see Minchin [1987] for R units). We performed two-dimensional ordinations of the simulated data (100 sites and 40 species). Environmental variables was orthogonal and consist of coordinates ( $x$ ,  $y$ ) of the nodes of the grid in the ecological space.

Procrustean analysis (Digby and Kempton 1987) is used to quantitatively compare the configuration of sites given by the two environmental variables to the configuration of sites obtained by the first two environmental-constrained scores of ordination. We measure the fit between original data and ordination results by  $m$ , the sum of singular values of procrustes analysis (Jackson 1995, Dray et al. 2003). The  $m$  statistic increases when the fit between the two configurations is better.

We first perform CCA (canonical correspondence analysis) and CA-PCA COIA (correspondence analy-

sis-principle-component analysis) on simulated data sets. The two methods produce similar results ( $m_{\text{COI}} = m_{\text{CCA}} = 0.98$ ) and environmental scores of sites produced by the two methods are very similar to the original data (Fig. 4a and b). This confirms that results obtained by CCA and CA-PCA COIA are very similar in most cases.

We next create random environmental variables from a random number generator with a uniform distribution between 0 and 1. We apply CCA and CA-PCA COIA with 22 environmental variables (the two original variables and 20 random environmental variables). Although the two methods are affected by this noise (Fig. 4c and d), it seems that results are better for CCA than for COIA ( $m_{\text{COI}} = 0.93$ ,  $m_{\text{CCA}} = 0.97$ ). This trend is confirmed by the decrease of the  $m$  statistic when we add from 1 to 40 random environmental variables (Fig.

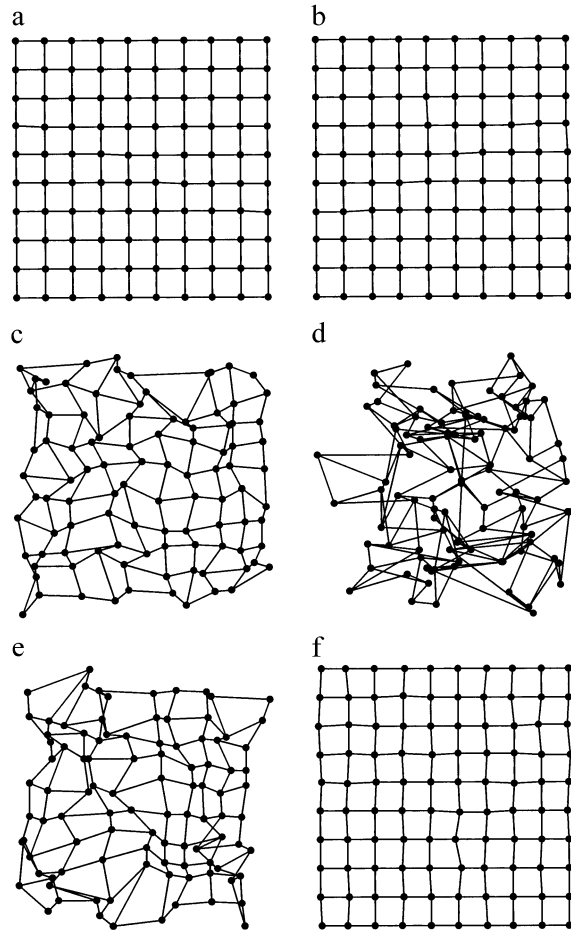


FIG. 4. Environmental-constrained sites scores of canonical correspondence analysis (CCA) (a, c, and e) and co-inertia analysis COIA (b, d, and f) for various data sets. In the first data set analyzed (a and b) the two environmental variables consist of the coordinates of samples in the ecological space. We add 20 random variables to the two original ones for the second data set (c and d). For the third data set, 20 correlated variables are added to the two original ones (e and f).

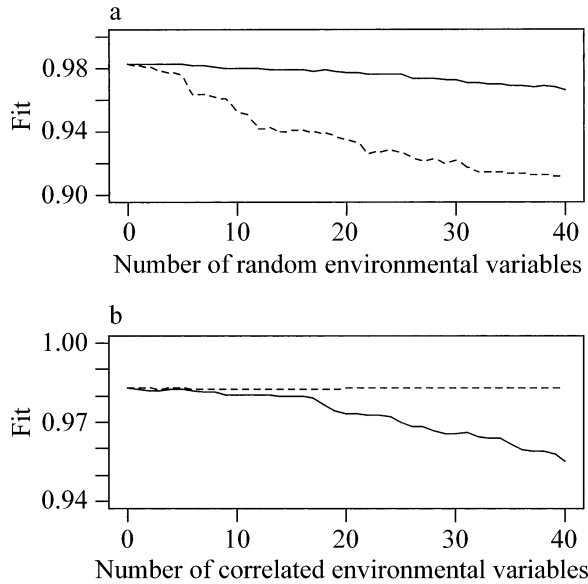


FIG. 5. Influence of the number (a) random and (b) correlated variables on the results of CCA (canonical correspondence analysis) and COIA (co-inertia analysis). We measure the fit between original environmental variables and the environmental-constrained sites scores for the first two axes by the  $m$  statistic (sum of singular values of Procrustean analysis). The dashed lines correspond to COIA, and the solid lines to CCA.

5a) and is not very surprising. CCA maximizes the proportion of variance in the species matrix that is explained by the environmental matrix while COIA maximizes the square covariance between the species score and the environmental score. This difference is due to the Mahalanobis metric of CCA, which is linked to the underlying regression step. Adding random environmental variables implies that the multiple regression of CCA explains the species matrix better. If the number of variables is equal to the number of sites minus 1, the results of CCA are the same as those given by CA of the species data because the species matrix is fully explained by the environmental matrix. As the simulated data are very structured, CA produces good results ( $m_{CA} = 0.86$ ) and so the decrease of  $m_{CCA}$  is low (Fig. 5a). Concerning the results of CA-PCA COIA, adding random environmental variables increases the total inertia of the environmental table while the structured inertia (due to the two original environmental variables) is constant. Therefore, the structured-inertia part decreases and the principle axes of PCA maximizing the projected variance of the sites are mostly defined by the random variables. This has an influence on COIA, which finds a compromise among the variance of sites in the species viewpoint, the variance of sites in the environmental viewpoint, and the square correlation. The random variables modify the ordination of sites in the environmental hyperspace and COIA finds an arrangement of sites in this hyperspace that is

less correlated to the original one but with high “random” variance and so high covariance.

We then create variables very correlated to the first original environmental variables by adding a noise (normal distribution with mean = 0 and varying standard deviation) to this variable. We apply CCA and CA-PCA COIA with 22 environmental variables (the two original variables and 20 correlated environmental variables). In this case, COIA is not affected by the new correlated variables (Fig. 4f,  $m_{COI} = 0.98$ ) whereas CCA results are affected (Fig. 4e,  $m_{CCA} = 0.97$ ). This trend is confirmed when we add from 1 to 40 correlated environmental variables (Fig. 5b). In that case, the structured-inertia part of the environmental table does not change and so the results of COIA are very stable. For CCA, the same problem appears: the number of variables increases, the species table is better explained, and results lead to those of CA. However, the problem of collinearity, which is well known in the case of multivariate linear regression (e.g., Legendre and Legendre 1998:518–521), is also present in CCA. The collinearity among environmental variables implies instability for coefficients of environmental variables used to construct axes and the ecological interpretation of biplot is then very dubious (Table 2). In this case, it is very important to observe the consistency (same sign and similarity of values) between coefficient of environmental variables and correlation of these variables with axes (ter Braak 1990, ter Braak and Looman

TABLE 2. Consistency between coefficient (Coef.) and correlation (Cor.) of environmental variables with the first environmental-constrained axis of CCA and CA-PCA COIA.

Variables	CA-PCA COIA		CCA	
	Coef.	Cor.	Coef.	Cor.
Env1	0.22	1	-1.3	0.99
Env2	0	0	0	0
Cor1	0.22	1	-0.5	0.98
Cor2	0.22	1	1.37	0.99
Cor3	0.22	1	-0.55	0.98
Cor4	0.22	1	-0.22	0.98
Cor5	0.22	1	-0.04	0.98
Cor6	0.22	1	0.9	0.99
Cor7	0.22	1	0.87	0.98
Cor8	0.22	1	0.92	0.99
Cor9	0.22	1	-0.36	0.98
Cor10	0.22	1	0.3	0.99
Cor11	0.22	1	0.36	0.98
Cor12	0.22	1	0.02	0.98
Cor13	0.22	1	-0.37	0.98
Cor14	0.22	1	-0.55	0.98
Cor15	0.22	1	-0.45	0.98
Cor16	0.22	1	-0.33	0.99
Cor17	0.22	1	-0.13	0.98
Cor18	0.22	1	2.34	0.99
Cor19	0.22	1	-1.64	0.98
Cor20	0.22	1	0.19	0.98

Notes: CCA = canonical correspondence analysis; CA = correspondence analysis, PCA = principle-component analysis, and COIA = co-inertia analysis. The first two original variables are named Env1 and Env2, and the 20 variables correlated to Env1 are named Cor1, . . . , Cor20.

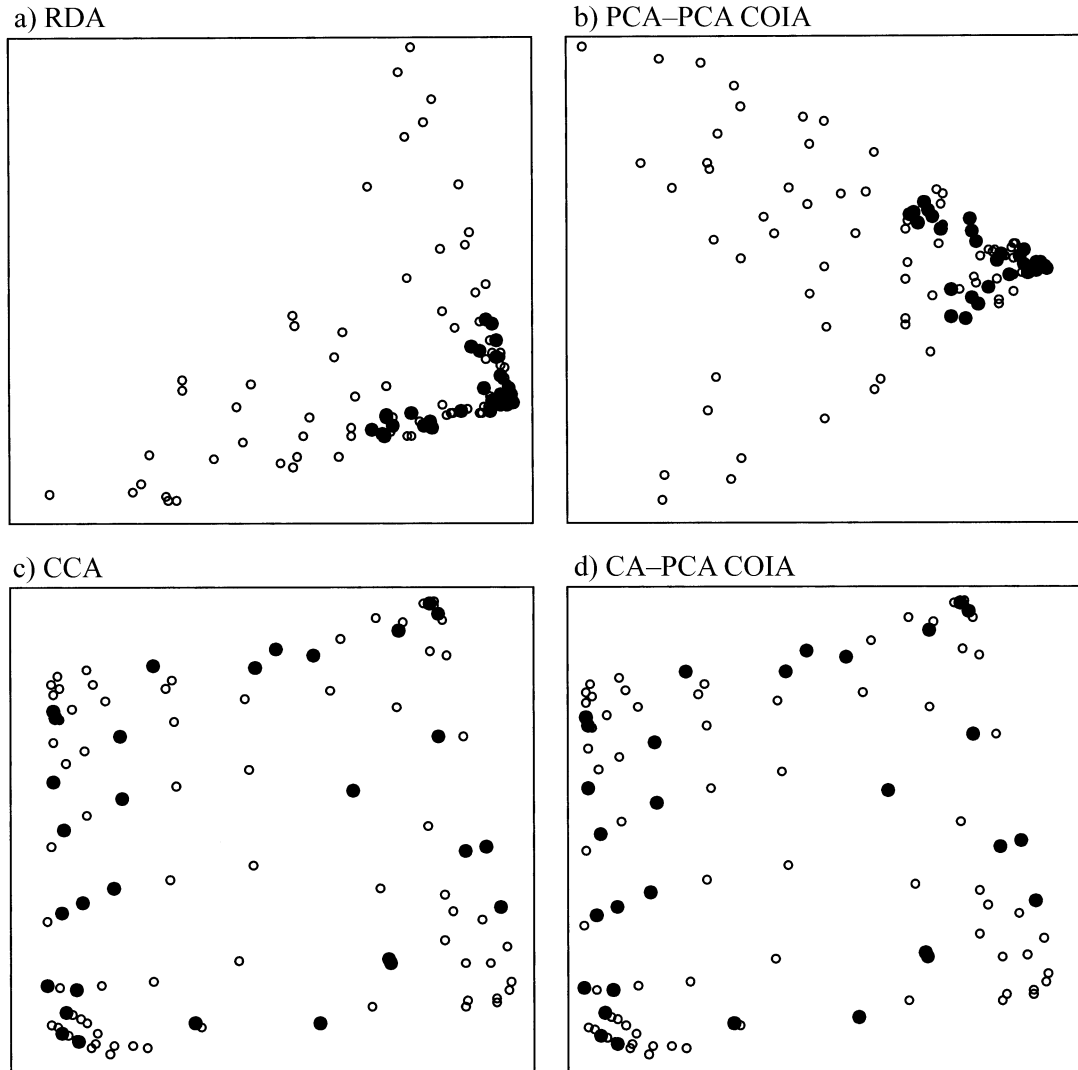


FIG. 6. Ability of the four methods to detect a pollution effect (decrease of abundances of all species in 30 sites). Species-constrained sites scores are presented (without numeric scales, since it is the shape of the cloud of points that is of interest); polluted sites are represented by solid black circles.

1994). COIA is not affected by collinearity and the consistency between correlation and coefficient is very good (Table 2), but this is not case for CCA, which produces results difficult to interpret. Indeed, there are many inconsistencies concerning coefficients values of variables that are very correlated to the first axis (e.g., coefficient for Env1 is  $-1.3$  while coefficient for Cor18 is  $2.34$ , Table 2).

Lastly, we simulate the effect of pollution by multiplying by  $0.3$  the abundance values of all species in 30 sites randomly selected. We apply CCA, RDA (redundancy analysis), CA-PCA COIA and PCA-PCA COIA on this data to test if these methods can detect this decrease of global abundance due to pollution. We plot species-constrained sites scores for the four methods (Fig. 6). In this situation, CCA and CA-PCA COIA, which are based on species composition, are

inefficient. RDA and PCA-PCA COIA, which focus on absolute composition, are able to identify the effects of pollution.

#### DISCUSSION

Analyses with respect to instrumental variables (e.g., CCA [canonical correspondence analysis], RDA [redundancy analysis], etc.) require a small number of environmental variables, and CANCOR (canonical correlation analysis) requires a small number of species and environmental variables compared to the number of sites (Fig. 2). This comes from the two simultaneous multivariate regressions in CANCOR, and from the multivariate regression in analyses with respect to instrumental variables. In contrast, co-inertia analysis (COIA) is linked to partial least-squares regression, a robust alternative to classical regression (Tenenhaus

1998). In the case of qualitative environmental variables, the dimension of the environmental space increases quickly with the number of variables, and co-inertia analysis is often the only alternative. If there are few uncorrelated (i.e., orthogonal) variables, the multivariate-regression step is simply an extension of univariate regression (Legendre and Legendre 1998) and CCA or RDA are very efficient. If there are many variables, results of CCA tend toward those of CA (correspondence analysis) and the results of RDA toward those of PCA (principle component analysis). If the variables are correlated, CCA and RDA become unstable. For these cases, COIA is a good alternative. Another alternative could be to (1) perform PCA on environmental variables to obtain uncorrelated variables and (2) apply CCA between the species matrix and principal axes of the previous PCA. This second alternative provides exactly the same results as COIA of the two original tables except that, unlike COIA, it does not allow one to plot directly the original environmental variables, which have been replaced by their principal axes. In the same way, CANCECOR between the principal axes of two tables is equivalent to COIA of the original tables.

The previous considerations concerning the number of variables, the species response model, the units of interest, the point of reference of the study, and the significance of species absence must be taken into account in order to choose the "good" statistical method. The theoretical framework induced by co-inertia will probably introduce new coupling methods. Adding new methods will not make the choice for ecologists easier but we hope that the previous considerations will give practical elements to guide this choice. Experience shows that in many cases different methods will give similar results, but that in particular situations the results of a study can greatly depend on the choice of the multivariate method. This can be illustrated by the detection of pollution by RDA or PCA-PCA COIA while CCA or CA-PCA COIA fail (Fig. 6).

Moreover, the generalization of the co-inertia criterion for linking more than two tables provides new efficient tools to study species-environment relationships. The introduction of species traits and the possibility to study simultaneously the spatial and temporal variations will probably greatly improve the quality of results in ecological studies. Even if some answers to these questions are given by classical coupling, the use of more sophisticated methods provides additional results of great interest to ecologists.

All these methods are available in the free ADE-4 software (Thioulouse et al. 1997), which is also distributed under the form of an R package (Ihaka and Gentleman 1996).

#### ACKNOWLEDGMENTS

We wish to thank P. Legendre and Mike Palmer, whose suggestions and comments have allowed us to improve the

first version of this text. We also wish to thank Peter Minchin for his very useful COMPAS program.

#### LITERATURE CITED

- Austin, M. P., J. A. Meyers, L. Belbin, and M. D. Doherty. 1995. Simulated data case study, Sub-project 5, Modelling of landscape patterns and processes using biological data. Available by loan from Division of Wildlife and Ecology, Commonwealth Scientific and Industrial Research Organisation, Canberra, Australian Capital Territory, Australia.
- Bayley, P. B., and J. T. Peterson. 2001. An approach to estimate probability of presence and richness of fish species. *Transactions of the American Fisheries Society* **130**:620–633.
- Ben-Shahar, R. 1987. Grasses and habitat relationships on a sour bushveld nature reserve. *Vegetatio* **72**:45–49.
- Ben-Shahar, R., and J.-D. Skinner. 1988. Habitat preferences of african ungulates derived by uni- and multivariate analyses. *Ecology* **69**:1479–1485.
- Birks, H. J. B., S. M. Peglar, and H. A. Austin. 1996. An annotated bibliography of canonical correspondence analysis and related constrained ordination methods 1986–1993. *Abstracta Botanica* **20**:17–36.
- Cadet, P., and J. Thioulouse. 1998. Identification of soil factors that relate to plant parasitic nematode communities on tomato and yam in the French West Indies. *Applied Soil Ecology* **8**:35–49.
- Cao, Y., D. P. Larsen, R. M. Hughes, P. L. Angermeier, and T. M. Patton. 2002. Sampling effort affects multivariate comparisons of stream assemblages. *Journal of the North American Benthological Society* **21**:701–714.
- Cao, Y., D. P. Larsen, and R. S. J. Thorne. 2001. Rare species in multivariate analysis for bioassessment: some considerations. *Journal of the North American Benthological Society* **20**:144–153.
- Cazes, P. 1980. L'analyse de certains tableaux rectangulaires décomposé en blocs: généralisation des propriétés rencontrées dans l'étude des correspondances multiples. *Les cahiers de l'analyse des données* **5**:145–161.
- Chessel, D., and M. Hanafi. 1996. Analyse de la co-inertie de K nuages de points. *Revue de Statistique Appliquée* **44**:35–60.
- Chessel, D., and P. Mercier. 1993. Couplage de triplets statistiques et liaisons espèces—environnement. Pages 15–43 in J. D. Lebreton and B. Asselain, editors. *Biométrie et Environnement*. Masson, Paris, France.
- Chevenet, F., S. Dolédec, and D. Chessel. 1994. A fuzzy coding approach for the analysis of long-term ecological data. *Freshwater Biology* **31**:295–309.
- 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.
- Del Moral, R. 1982. Control of vegetation on contrasting substrates: herb patterns on serpentinite and sandstone. *American Journal of Botany* **69**:227–238.
- De Maesschalck, R., D. Jouan-Rimbaud, and D. L. Massart. 2000. The Mahalanobis distance. *Chemometrics and Intelligent Laboratory Systems* **50**:1–18.
- Digby, P. G. N., and R. A. Kempton. 1987. *Multivariate analysis of ecological communities*. Chapman and Hall, London, UK.
- Dolédec, S., and D. Chessel. 1987. Rythmes saisonniers et composantes stationnelles en milieu aquatique. I. Description d'un plan d'observations complet par projection de variables. *Acta Oecologica—Oecologia Generalis* **8**:403–426.
- Dolédec, S., and D. Chessel. 1994. Co-inertia analysis: an alternative method for studying species-environment relationships. *Freshwater Biology* **31**:277–294.

- Dolédéc, S., D. Chessel, and C. Gimaret-Carpentier. 2000. Niche separation in community analysis: a new method. *Ecology* **81**:2914–2927.
- Dolédéc, S., D. Chessel, C. J. F. ter Braak, and S. Champely. 1996. Matching species traits to environmental variables: a new three-table ordination method. *Environmental and Ecological Statistics* **3**:143–166.
- Dray, S., D. Chessel, and J. Thioulouse. 2003. Procrustean co-inertia analysis for the linking of multivariate data sets. *Ecoscience* **10**:110–119.
- Dray, S., N. Pettorelli, and D. Chessel. 2002. Matching data sets from two different spatial samplings. *Journal of Vegetation Science* **13**:867–874.
- Esposito Vinzi, V. 2001. Explanatory methods for comparative analyses. *Chemometrics and Intelligent Laboratory Systems* **58**:275–286.
- Franquet, E., and D. Chessel. 1994. Approche statistique des composantes spatiales et temporelles de la relation faune-milieu. *Comptes Rendus de l'Académie des Sciences Serie III—Sciences de la Vie* **317**:202–206.
- Franquet, E., S. Dolédéc, and D. Chessel. 1995. Using multivariate analyses for separating spatial and temporal effects within species–environment relationships. *Hydrobiologia* **300/301**:425–431.
- Gauch, H. G. 1982. *Multivariate analysis in community ecology*. Cambridge University Press, Cambridge, UK.
- Gauch, H. G., and T. R. Wentworth. 1976. Canonical correlation analysis as an ordination technique. *Vegetatio* **33**:17–22.
- Gimaret-Carpentier, C., D. Chessel, and J.-P. Pascal. 1998a. Non-symmetric correspondence analysis: an alternative for species occurrences data. *Plant Ecology* **138**:97–112.
- Gimaret-Carpentier, C., D. Chessel, J.-P. Pascal, and B. R. Ramesh. 1998b. Advantages of non-symmetric correspondence analysis in identifying multispecific spatial patterns in the rain forest of the western Ghats. Pages 397–411 in Y. Laumier, B. King, C. Legg, and K. Rennolls, editors. *Data management and modelling using remote sensing and GIS for tropical forest land inventory*. Rodeo International Publishers, Jakarta, Indonesia.
- Hill, M. O. 1974. Correspondence analysis: a neglected multivariate method. *Applied Statistics—Journal of the Royal Statistical Society Series C* **23**:340–354.
- Hill, M. O., and A. J. E. Smith. 1976. Principal component analysis of taxonomic data with multi-state discrete characters. *Taxon* **25**:249–255.
- Hotelling, H. 1936. Relations between two sets of variates. *Biometrika* **28**:321–377.
- Huang, C. L., and R. Del Moral. 1988. Plant–environment relationships on the Montlake wildlife area, Seattle Washington, USA. *Vegetatio* **75**:103–113.
- Ihaka, R., and R. Gentleman. 1996. R: a language for data analysis and graphics. *Journal of Computational and Graphical Statistics* **5**:299–314.
- Jackson, D. A. 1995. PROTEST: a PROcrustean randomization TEST of community environment concordance. *Ecoscience* **2**:297–303.
- Jackson, D. A., and K. M. Somers. 1991. Putting things in order: the ups and downs of detrended correspondence analysis. *American Naturalist* **137**:704–712.
- Kiers, H. A. L. 1994. Simple structure in component analysis techniques for mixtures of qualitative and quantitative variables. *Psychometrika* **56**:197–212.
- Kroonenberg, P. M., and R. Lombardo. 1999. Nonsymmetric correspondence analysis: a tool for analysing contingency tables with a dependence structure. *Multivariate Behavioral Research* **34**:367–396.
- Lafosse, R., and M. Hanafi. 1997. Concordance d'un tableau avec K tableaux: définition de K+1 uples synthétiques. *Revue de Statistique Appliquée* **45**:111–126.
- Lauro, N., and L. D'ambra. 1984. L'analyse non symétrique des correspondances. Pages 433–446 in E. Diday, M. Jambu, L. Lebart, J. P. Pagès, and R. Tomassone, editors. *Data analysis and informatics III*. Elsevier Science, Amsterdam, The Netherlands.
- Legendre, P., and M. J. Anderson. 1999. Distance-based redundancy analysis: testing multispecies responses in multifactorial ecological experiments. *Ecological Monographs* **69**:1–24.
- Legendre, P., and E. D. Gallagher. 2001. Ecologically meaningful transformations for ordination of species data. *Oecologia* **129**:271–280.
- Legendre, P., R. Galzin, and M. L. Harmelin-Vivien. 1997. Relating behavior to habitat: solutions to the fourth-corner problem. *Ecology* **78**:547–562.
- Legendre, P., and L. Legendre. 1998. *Numerical ecology*. Second English edition. Elsevier Science, Amsterdam, The Netherlands.
- Lods-Crozet, B., E. Castella, D. Cambin, C. Ilg, S. Knispel, and H. Mayor-Simeant. 2001. Macroinvertebrate community structure in relation to environmental variables in a Swiss glacial stream. *Freshwater Biology* **46**:1641–1661.
- McArdle, B. H., and M. J. Anderson. 2001. Fitting multivariate models to community data: a comment on distance-based redundancy analysis. *Ecology* **82**:290–297.
- McCune, B. 1997. Influence of noisy environmental data on canonical correspondence analysis. *Ecology* **78**:2617–2623.
- Méot, A., P. Legendre, and D. Borcard. 1998. Partialling out the spatial component of ecological variation: questions and propositions in the linear modelling framework. *Environmental and Ecological Statistics* **5**:1–27.
- Mercier, P., D. Chessel, and S. Dolédéc. 1992. Complete correspondence analysis of an ecological profile data table: a central ordination. *Acta Oecologica—International Journal of Ecology* **13**:25–44.
- Minchin, P. R. 1987. Simulation of multidimensional community patterns: towards a comprehensive model. *Vegetatio* **71**:145–156.
- Montana, C., and P. Greig-Smith. 1990. Correspondence analysis of species by environmental variable matrices. *Journal of Vegetation Science* **1**:453–460.
- Noy-Meir, I. 1973. Data transformation in ecological ordination. I. Some advantages of non-centring. *Journal of Ecology* **61**:329–341.
- Noy-Meir, I., D. Walker, and W. T. Williams. 1975. Data transformation in ecological ordination. II. On the meaning of data standardization. *Journal of Ecology* **63**:779–800.
- Ojeda, F., J. Arroyo, and T. Marañón. 1998. The phytogeography of European and Mediterranean heath species (Ericoideae, Ericaceae): a quantitative analysis. *Journal of Biogeography* **25**:165–178.
- Palmer, M. W. 1993. Putting things in even better order: the advantages of canonical correspondence analysis. *Ecology* **74**:2215–2230.
- Palmer, M. W., and P. S. White. 1994. Scale dependence and the species–area relationship. *American Naturalist* **144**:717–740.
- Peet, R. K., R. G. Knox, J. S. Case, and R. B. Allen. 1988. Putting things in order: the advantages of detrended correspondence analysis. *American Naturalist* **131**:924–934.
- Rao, C. R. 1964. The use and interpretation of principal component analysis in applied research. *Sankhya A* **26**:329–359.
- Reynaud, P. A., and J. Thioulouse. 2000. Identification of birds as biological markers along a neotropical urban–rural gradient (Cayenne, French-Guiana), using co-inertia analysis. *Journal of Environmental Management* **59**:121–140.
- Romane, F. 1972. Utilisation de l'analyse multivariée en phytocologie. *Investigacion Pesquera* **36**:131–139.

- Simier, M., L. Blanc, F. Pellegrin, and D. Nandris. 1999. Approche simultanée de K couples de tableaux: application à l'étude des relations pathologie végétale-environnement. *Revue de Statistique Appliquée* **47**:31–46.
- Soberon, J. M., and J. B. Llorente. 1993. The use of species accumulation functions for the prediction of species richness. *Conservation Biology* **7**:480–488.
- Strahler, A. H. 1978. Binary discriminant analysis: a new method for investigating species–environment relationships. *Ecology* **59**:108–116.
- Tenenhaus, M. 1998. La régression PLS. Théorie et pratique. Editions Technip, Paris, France.
- Tenenhaus, M., and F. W. Young. 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. 1983. Principal components biplots and alpha and beta diversity. *Ecology* **64**:454–462.
- ter Braak, C. J. F. 1986. Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. *Ecology* **67**:1167–1179.
- ter Braak, C. J. F. 1988. Partial canonical correspondence analysis. Pages 551–558 in H. H. Bock, editor. Classification and related methods of data analysis. Elsevier Science, Amsterdam, The Netherlands.
- ter Braak, C. J. F. 1990. Interpreting canonical correlation analysis through biplots of structure correlations and weights. *Psychometrika* **55**:519–531.
- ter Braak, C. J. F., and C. W. N. Looman. 1986. Weighted averaging, logistic regression and the Gaussian response model. *Vegetatio* **65**:3–11.
- ter Braak, C. J. F., and C. W. N. Looman. 1994. Biplots in reduced-rank regression. *Biometrical Journal* **36**:983–1003.
- Thioulouse, J., D. Chessel, S. Dolédec, and J. M. Olivier. 1997. ADE-4: a multivariate analysis and graphical display software. *Statistics and Computing* **7**:75–83.
- Thioulouse, J., M. Simier, and D. Chessel. 2004. Simultaneous analysis of a sequence of paired ecological tables. *Ecology* **85**, in press.
- Tucker, L. R. 1958. An inter-battery method of factor analysis. *Psychometrika* **23**:111–136.
- Wartenberg, D., S. Ferson, and F. J. Ohlf. 1987. Putting things in order: a critique of detrended correspondence analysis. *American Naturalist* **129**:434–448.
- Willby, N. G., V. J. Abernethy, and B. O. L. Demars. 2000. Attribute-based classification of European hydrophytes and its relationship to habitat utilization. *Freshwater Biology* **43**:43–74.