*Freshwater Biology* (1994) 31, 277–294

# Co-inertia analysis: an alternative method for studying species–environment relationships

SYLVAIN DOLÉDEC* AND DANIEL CHESSEL
*URA CNRS 1451 'Ecologie des Eaux Douces et des Grands Fleuves', Université Lyon I, 69622 Villeurbanne Cedex, France*

**Author to whom correspondence should be sent*

## SUMMARY

1. Methods used for the study of species–environment relationships can be grouped into: (i) simple indirect and direct gradient analysis and multivariate direct gradient analysis (e.g. canonical correspondence analysis), all of which search for non-symmetric patterns between environmental data sets and species data sets; and (ii) analysis of juxtaposed tables, canonical correlation analysis, and intertable ordination, which examine species–environment relationships by considering each data set equally. Different analytical techniques are appropriate for fulfilling different objectives.
2. We propose a method, co-inertia analysis, that can synthesize various approaches encountered in the ecological literature. Co-inertia analysis is based on the mathematically coherent Euclidean model and can be universally reproduced (i.e. independently of software) because of its numerical stability. The method performs simultaneous analysis of two tables. The optimizing criterion in co-inertia analysis is that the resulting sample scores (environmental scores and faunistic scores) are the most covariant. Such analysis is particularly suitable for the simultaneous detection of faunistic and environmental features in studies of ecosystem structure.
3. The method was demonstrated using faunistic and environmental data from Friday (*Freshwater Biology* 18, 87–104, 1987). In this example, non-symmetric analyses is inappropriate because of the large number of variables (species and environmental variables) compared with the small number of samples.
4. Co-inertia analysis is an extension of the analysis of cross tables previously attempted by others. It serves as a general method to relate any kinds of data set, using any kinds of standard analysis (e.g. principal components analysis, correspondence analysis, multiple correspondence analysis) or between-class and within-class analyses.

## Introduction

A fundamental property of biological systems is their ability to evolve, which is dependent on the system structure as well as on the relationships between the species and their environment (e.g. Prodon, 1988). As underscored by Wiens (1986, p. 154), 'the challenge of community ecology is to discover the patterns of natural assemblages of organisms and to explain them in terms of controlling processes'. In particular, Townsend & Hildrew (1994) indicate that a central purpose of ecology is to understand the relationships between the bewildering diversity of organisms and environments. The aim proposed by Townsend & Hildrew essentially involves the study of the link between species traits and environmental variability or, at least, the link between species traits and species utilization of environmental units that have a particular level of spatial and temporal variability.

The study of the relationships between a fauna (or a flora) and its environment usually leads to two sets: (i) a faunistic array that contains the abundances or occurrences of a number of taxa in a set of sites; and (ii) an environmental array that includes quantitative or categorical measurements from the same sites. Two main objectives are usually involved in the statistical study of the above two data sets: (i) an inference of the faunistic variation is drawn from the environment variation and, conversely, a prediction of environmental parameters is made using the taxa list; and (ii) the covariation between a sample ordination made via the faunistic data set and a sample ordination made via the environment data set is examined. The great diversity of organisms living under various environmental conditions expectedly generates a great diversity in data. Consequently, the above objectives have been examined extensively and many statistical solutions have been proposed. The objective of this paper is to review the various statistical approaches to the analysis of species–environment relationships, and to describe an alternative method that we call co-inertia analysis. Co-inertia analysis is used in other papers of this issue to check for a co-structure (i.e. a relationship) between species traits and habitat utilization in the Upper Rhône River.

## A review of linear ordination methods for studying species–environment relationships

### *Indirect and direct gradient analysis*

A classical method to interpret floro-faunistic structures is the so-called indirect gradient analysis (Whittaker, 1967), or continuum analysis (Anderson, 1963), or vegetational ordination (Austin, 1968; Fig. 1a). It consists of ordering samples along one or several axes using synthetic scores obtained from linear or non-linear methods (reviewed in Dale, 1975). For example, Rabeni & Gibbs (1980) used the Bray & Curtis method to order macroinvertebrate populations, and then plotted the values of environmental variables to interpret the faunistic ordination. Among linear methods, the most commonly used are principal component analysis (e.g. Goodall, 1954; with an early application in animal ecology by Brian, 1964) generally centred by taxa, correspondence analysis (e.g. Roux & Roux, 1967; Hill, 1974), and the more recent detrended version of correspondence analysis (Hill & Gauch, 1980). The sample scores are then interpreted in relation with environmental parameters. Some authors have correlated the sample scores with each environmental variable (e.g. Vincent, 1981; Gibson & Hurlbert, 1987). Moreover, Barkham & Norris (1970) calculated correlations of the scores resulting from the analysis of a floristic table with the scores resulting from the analysis of soil data. Finally, multiple regression has been used to relate a linear combination of environmental parameters and scores resulting from the analysis of the faunistic table (e.g. Prodon & Lebreton, 1981, in bird ecology; Chang & Gauch, 1986, in plant ecology; Storey *et al.*, 1990, in stream ecology).

Conversely, direct gradient analysis (Whittaker, 1967) or environmental ordination (Austin, 1968) consists of ordering samples along simple (one variable) or complex (multivariate) environmental gradients (Fig. 1b). Species are then plotted on the environmental axes to study their response to environmental gradients. Species may be used as supplementary individuals in the analysis of environmental parameters (e.g. Lapchin & Roux, 1977).

In a summary, Orloci (1988, p. 174) underscored that among 'the many numerical methods available for seeking vegetation patterns in compositional terms, ordination stands out as a most important methodology' because it detects 'patterns as arrangements of units on axes'.

### *Multivariate direct gradient analysis*

In direct and indirect gradient analysis, the problem is to find the best synthetic scores, i.e. the scores that provide the optimal interpretation of species–environment relationships. However, Ter Braak (1986, p. 1167) emphasized the difficulty 'to detect by indirect gradient analysis the effects on community composition of a subset of environmental variables in which one is particulary interested' and the need of a multivariate direct gradient analysis to overcome such difficulties. The author proposed a method named canonical correspondence analysis (Fig. 1c) 'because it is a correspondence analysis technique in which the axes are chosen in the light of the environmental variables' (Ter Braak, 1986, p. 1168). This technique calculates a derived environmental variable that is a linear combination of the original environmental variables, which is optimal in the sense that it minimizes the size of species niches, i.e. it minimizes the ratio of within-species variance to total variance. As stated by Hill (1988, p. 139) 'using canonical correspondence analysis is similar to making an ordination and then looking for an estimate of the sample scores based on a multiple regression of the

Dolédec, S., and D. Chessel. 1994. Co-inertia analysis: an alternative method for studying species-environment relationships. Freshwater Biology 31:277-294.
http://pbil.univ-lyon1.fr/R/articles/arti084.pdf

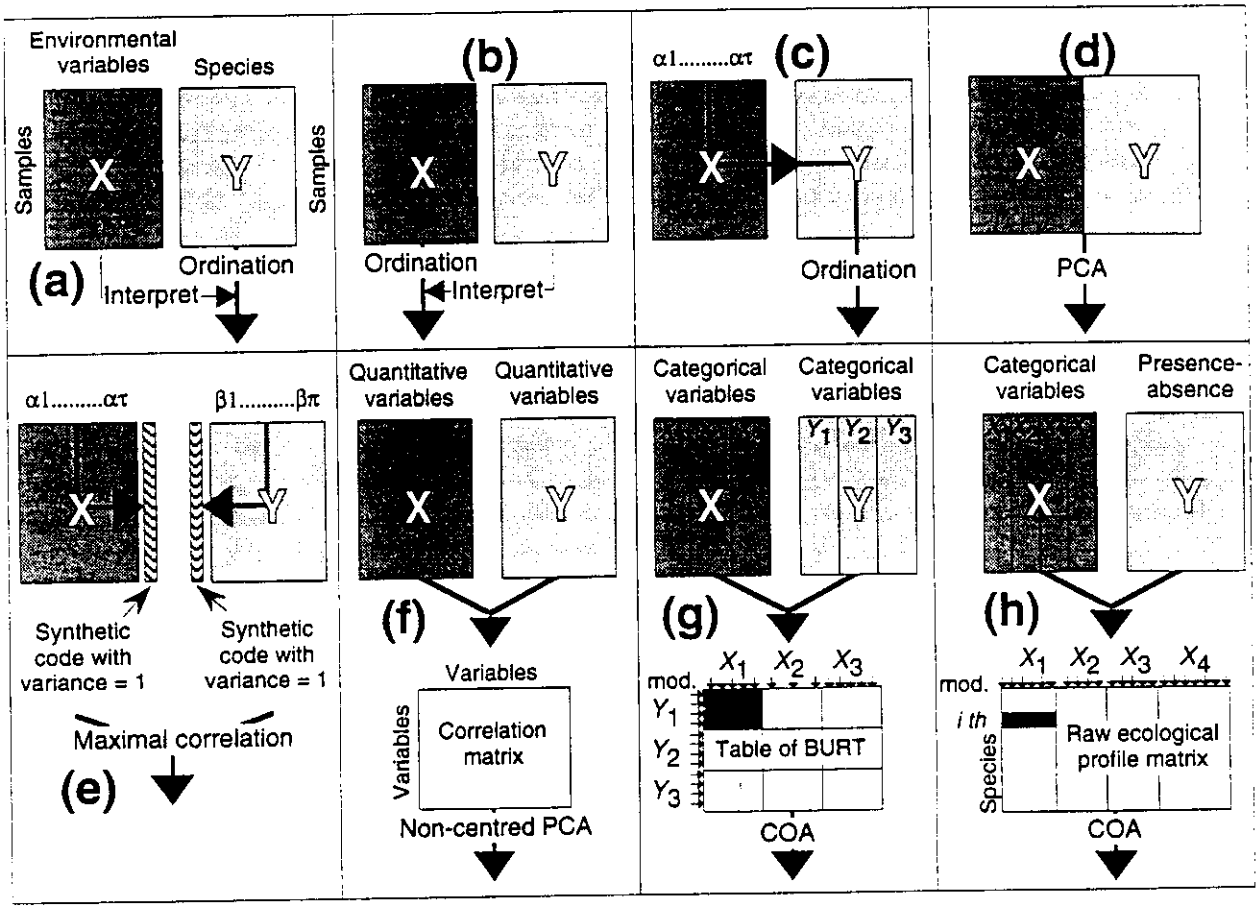


**Fig. 1** Various methods used by community ecologists to study species–environment relationships (see text for citations and explanations). In the following representations, X is the environmental data set and Y is the floristic or the faunistic data set. (a) Indirect gradient analysis. (b) Direct gradient analysis. (c) Multivariate direct gradient analysis (principal component analysis and correspondence analysis with respect to instrumental variables, canonical correspondence analysis). (d) Juxtaposition of the environmental and faunistic data sets; the resulting array is processed by a principal component analysis (abbreviated PCA). (e) Canonical correlation analysis. (f) Intertable analysis using a correlation matrix then processed by a non-centred principal component analysis (abbreviated PCA). (g) Intertable analysis using a table of Burt then processed by a correspondence analysis (abbreviated COA). (h) Intertable analysis using a raw ecological profile table then processed by a correspondence analysis (abbreviated COA).

environmental variables'. Hence, Lebreton *et al.* (1988) proposed the name 'correspondence analysis with respect to instrumental variables', following the demonstration by Chessel, Lebreton & Yoccoz (1987) that canonical correspondence analysis consisted of a projection of the species array on to the subspace generated by environmental variables. The use of canonical correspondence analysis is illustrated by Ter Braak (1987), Johnson & Wiederholm (1989), and many others.

Sabatier & Van Campo (1984) introduced principal component analysis with respect to instrumental variables to study climate changes in Greece over 18 000 years. Sabatier (1987) and Sabatier, Lebreton & Chessel (1989) then related canonical correspondence analysis and principal component analysis with respect to instrumental variables, and Lebreton *et al.* (1991) showed how the choice of weight matrices for each data set creates particular cases of principal component analysis with respect to instrumental variables.

The synthesis by Ludwig & Reynolds (1988) ends with a perspective of what canonical correspondence analysis offers in ecological research, and it clearly appears that multivariate methods with respect to instrumental variables are the best adapted to non-

Dolédec, S., and D. Chessel. 1994. Co-inertia analysis: an alternative method for studying species-environment relationships. Freshwater Biology 31:277-294.
http://pbil.univ-lyon1.fr/R/articles/arti084.pdf

symmetric studies (prediction of environment or response of communities). Finally, Birks & Austin (1992, p. 1) in their review indicated that the development of canonical correspondence analysis has 'revolutionised quantitative community ecology and related subjects such as limnology' by incorporating 'regression and ordination into a single extremely powerful method for multivariate direct gradient analysis called canonical or constrained ordination'.

Hence, in the above methods, the relationships between the two tables are non-symmetrical because one set of variables is considered as a predicting set whereas the second set of variables is considered as a response. By contrast, as explained by Gittins (1985), the methods reviewed below are symmetric since they involve an equal footing of the two sets of variables.

### *Analysis of juxtaposed tables*

A method proposed by Dagnélie (1965) is to mix all the species and environmental data together into a compound analysis (Fig. 1d). This was done by Depiereux, Feytmans & Micha (1983) to study the relationships between macroinvertebrate distributions and water quality measurements. Hill (1988, p. 140) emphasized that 'this approach is often logically unsound' because it depends on the weight given to variables (species and environmental parameters); for example, 'by including many environmental variables, the analysis can become effectively an environmental ordination with passive species variables'. Conversely, he notes that 'if many species are present at each site and only a few environmental variables are measured, then the environmental variables may be effectively passive'. Austin (1968, p. 740) also reported that such an operation 'may result in principal components which, though accounting for the major variation, do not necessarily include high loadings for any of the vegetational variables'.

### *Canonical correlation analysis*

Canonical correlation analysis (Fig. 1e) also simultaneously analyses faunal and environmental data sets (see review in Gittins, 1985). The two data sets are made using quantitative variables measured at the same sites. The analysis determines the linear combination of variables (called canonical variates) from each of the data sets such that the correlation between the canonical variates from each set is maximized. Canonical correlation analysis has been performed, for example, in stream ecology by Corkum & Ciborowski (1988) to examine relationships between lotic invertebrate assemblages and environmental variables. This method requires similar numbers for environmental variables and species. Several criticisms to this method have been made regarding its relevance for ecological data. For example Austin (1968, p. 754), studying vegetation samples, indicated that 'canonical correlation does not appear to provide a very satisfactory technique; the mathematical model on which it is based, with its requirements for orthogonal correlations between vegetation and environment and complete linearity, appears to be too stringent'.

Multiple discriminant analysis, also known as canonical variate analysis, is a particular case of canonical correlation analysis. Usually, floristic or faunistic samples are classified by groups and discriminant analysis is then applied to relate the site grouping to environmental data (e.g. Cassie, 1972, in marine ecology; Bonin & Roux, 1978, in plant ecology; Wright *et al.*, 1984, in stream ecology).

### *Intertable analyses: towards co-inertia analysis*

Three approaches have previously been used for analyses of cross tables (intertable ordination) using different types of data:

1 Each data set contains quantitative measurements and the cross table is a rectangular table of correlation coefficients (Fig. 1f). The resulting table is processed by principal component analysis. Such an analysis was described by Tucker (1958) under the name of inter-battery analysis to study relationships between the results obtained by the same individuals for two batteries of psychometric tests. Inter-battery analysis is also known as the first step of a partial least square regression. Hoskuldsson (1988) reported that partial least square regression methods can be expected to be useful for modelling purposes when many variables are measured in few samples. This situation is not rare in ecology, because often the cost of an extra individual sample is high especially when many variables are measured.

2 Each data set contains categorical variables and the cross table contains the number of individuals for

Dolédec, S., and D. Chessel. 1994. Co-inertia analysis: an alternative method for studying species-environment relationships. Freshwater Biology 31:277-294.
http://pbil.univ-lyon1.fr/R/articles/arti084.pdf

all couples of categories (= modalities = mod. in Fig.1g). This cross table of Burt (Cazes, 1980) is then processed by a correspondence analysis. Such an analysis is known as canonical analysis on categorical variables.

3 One data set contains ecological variables by category and the other data set contains the presence or absence of species. The cross table is a contingency table made of raw ecological profiles (Fig. 1h), i.e. it contains the frequencies of each species for each modality (= mod. in Fig. 1h) of the ecological variables (see also Feoli & Orloci, 1985); the resulting table is then processed by a correspondence analysis (Romane, 1972). The work of Romane was recently rediscovered by Montana & Greig-Smith (1990) as a convenient alternative to canonical correspondence analysis.

## Co-inertia analysis

Developments in statistical analysis (Mercier, Chessel & Dolédec, 1992; Chessel & Mercier, 1993) have focused on intertable ordination. The solution that we propose is called co-inertia analysis. Such a method comes as an extension to the approach by Tucker (1985). It works on a covariance matrix (species × environment) instead of a correlation matrix. Hence, the essential mathematical property of co-inertia analysis is a generalization of Tucker's inter-battery analysis to any type of analysis. As a consequence, it enables various standard analyses such as correspondence analysis and principal component analysis to be connected after any transformation of the raw data sets (e.g. centred, standardized, double centred, row centred, ordinally coded variables). As indicated above (Hoskuldsson, 1988), this is the only way to search for species–environment relationships when many variables (i.e. many species and several environmental variables) are sampled in few sites [i.e. the number of environmental and faunistic (e.g. number of taxa) variables is higher than that of samples; see Friday's example below]. In such a case, canonical correspondence analysis is reduced to correspondence analysis of the faunistic table because a number of random linear combinations of the environmental parameters may predict the faunistic structure.

### *Principle*

Let X be the environmental table and Y the taxa table (Fig. 2). Separate analysis of each data table brings out a principal axis (noted F1 in Fig. 2), which is the vector direction maximizing the projected variability (or inertia) in each array independently. The sampling units may be ordinated along the resulting axes as in standard analyses (e.g. principal component analysis or correspondence analysis). Moreover, it may be conceivable to isolate a new axis in one multidimensional space (noted as 'environment axis' in Fig. 2), and a new axis in the other multidimensional space (noted as 'fauna axis' in Fig. 2), so that the covariance between the two new sets of projected scores is maximal. This maximal covariance means a maximal correlation and simultaneously maximal standard deviations of both new environment and faunistic scores (see Appendix 1 for the mathematical model).

### *Comparison with other methods*

The popular detrended version of correspondence analysis (Hill & Gauch, 1980) is a pragmatic procedure for removing the arch (or horseshoe) effect generally associated with correspondence analysis of faunistic data. Co-inertia analysis that uses an environmental and a faunistic table will also remove such an artefact because the arch structure of faunistic data generally has no equivalent in the structure of environmental data (quantitative measurements); thus, the arch effect is removed in the co-structure of faunistic and environmental data. Here, the advantage of co-inertia analysis is that this method is mathematically defined (cf. Appendix 1), and may be reproduced independently of the software support.

Multidimensional scaling methods are also of potential use for species–environment relationships (e.g. Faith & Norris, 1989). The objective of these methods is to use, for example, Euclidean distance to approximate given dissimilarities arising from a set of points. The 'Stress' (and 'Sstress') function is then used as the most important criterion to be minimized while estimating the set of points. Many algorithms may be used to minimize the 'Stress' function (Gower, 1984).

According to Gower (1987), the ordination methods based on eigenstructure (e.g. our co-inertia analysis) are associated with general, well-understood math-

Dolédec, S., and D. Chessel. 1994. Co-inertia analysis: an alternative method for studying species-environment relationships. Freshwater Biology 31:277-294.
http://pbil.univ-lyon1.fr/R/articles/arti084.pdf

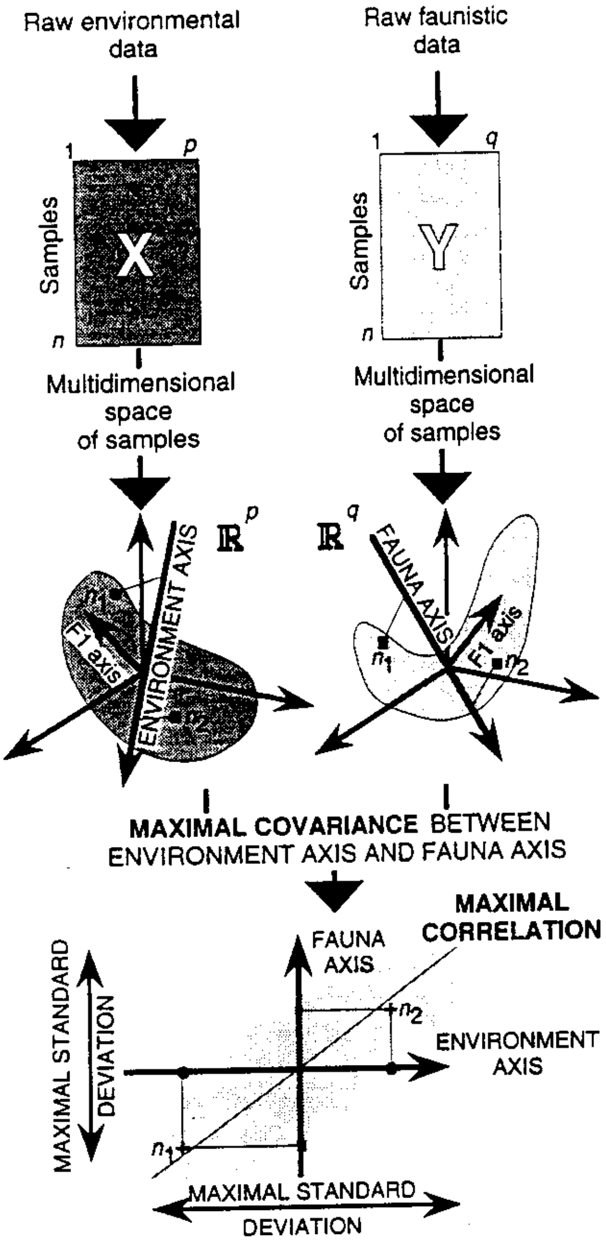


**Fig. 2** Co-inertia analysis as an extension of the initial approach of Tucker (1958), which was performed in connecting two standardized principal component analyses (the upper arrows symbolize the initial transformation of raw data, i.e. centring, double centring, etc.). Let X be the environmental array. Let Y be the faunistic array. Samples $n_1$ and $n_2$ are two given stands that are to be projected on both the environment axis and the fauna axis of the co-inertia analysis. These projections define new scores of stands that are the most covariant, i.e. these new scores have a maximal correlation and the standard deviations of the new scores are maximal.

ematical properties, whereas methods 'based on Stress and Sstress are much less well-understood mathematically' (Gower, 1987, p. 31). Moreover, 'iterative computer algorithms are continually improving but the mathematical fact that solutions are not nested and the lack of information on the occurrence of local optima are a problem'. This forces the user of multidimensional scaling methods to repeat calculations with various starting configurations to avoid suboptimal solutions. For example, Belbin (1991) found it reasonable to compute based on ten random starting configurations. In contrast, linear ordination methods represent a mathematically coherent (Euclidean model) and diversified 'toolbox', and can be universally reproduced because of their numerical stability. As a consequence, linear ordination methods are independent of the software support. Another difference is that linear ordination methods are based on the diagonalization (see Appendix 1) of a symmetric matrix whereas in multidimensional scaling methods there is a selection of the space (sites or species) in which the set of points is projected.

Co-inertia analysis is a two-table ordination method, as is the canonical correspondence analysis of Ter Braak (1986, 1987) or canonical correlation analysis (Gittins, 1985). The differences among these methods are the conditions that limit the demonstration of any pertinent structure. In canonical correlation analysis the number of species and the number of environmental variables must be much lower than the number of samples. In canonical correspondence analysis, a small number of environmental variables is required to predict the faunistic structure. Co-inertia analysis enables the connection of tables having similar (even low) as well as different numbers of environmental variables, species, and/or samples.

In synthesizing long-term ecological research, one is often confronted with a large number of species and a large number of modalities characterizing the environment. Another scientific field also confronted with such a problem is chemistry. For example, the most successful method used to compare the structure of molecules and their biological activity (QSAR) is the partial least square regression (Hoskuldsson, 1988), and co-inertia analysis is the first step of partial least square regression (cf. Lindgren, Geladi & Wold, 1993). The main reason for the high popularity of partial least square regression in chemistry is the numerical stability of the results, which provides its robustness.

Dolédec, S., and D. Chessel. 1994. Co-inertia analysis: an alternative method for studying species-environment relationships. Freshwater Biology 31:277-294.
http://pbil.univ-lyon1.fr/R/articles/arti084.pdf

Thus, in comparison with the above discussed methods (e.g. canonical correspondence analysis, canonical correlation analysis), co-inertia analysis is the simplest and most robust approach for matching two tables.

### Co-inertia analysis of Friday's (1987) data

To illustrate the technique of co-inertia analysis, data were selected from Friday's (1987) study of the diversity of macroinvertebrate and macrophyte communities in ponds. After reviewing the wide range of factors (from the biogeographic scale to the habitat scale) that could explain inter-pond variation, Friday (1987, p. 87) attempted to establish the causes of 'interpond variation in the number and identification of macroinvertebrate species'. From the large number of water bodies created by open-cast extraction of ball-clay in the Isle of Purbeck, Dorset (U.K.), the author selected sixteen ponds (Fig. 3a); chemical measurements were made and macroinvertebrate fauna was collected to investigate if chemistry, and especially pH, acts as a limiting factor. In her work, Friday (1987) used simple and multiple regressions to determine which chemical or morphometric parameters were related to the distribution and abundance of macroinvertebrate species.

From that study, we have used ninety-one macroinvertebrate taxa and eleven environmental variables: pond area (ha), vegetated area (ha), maximum depth (m), pH, conductivity ($\mu S\,cm^{-2}$), 5-days biological oxygen demand (BOD; mg $O_2 l^{-1}$), total hardness ($mg\,l^{-1}$), alkalinity ($mg\,l^{-1}$), calcium ($mg\,l^{-1}$), orthophosphate ($mg\,l^{-1}$), nitrate ($mg\,l^{-1}$), and turbidity (Formazin units). Clearly, the large number of taxa and the large number of variables, relative to the low number of samples ($n = 16$), calls for co-inertia analysis.

A first step consisted of conducting separate analyses to interpret both the environmental structure and the faunistic structure. The second step consisted of comparing the resulting ordinations to that obtained using co-inertia analysis. We applied the logic of principal component analysis because the environmental factor related to the distribution of species was supposed to be limiting (cf. above review and Fig. 1f).

All calculations and graphics were made using ADE (Chessel & Dolédec, 1992).

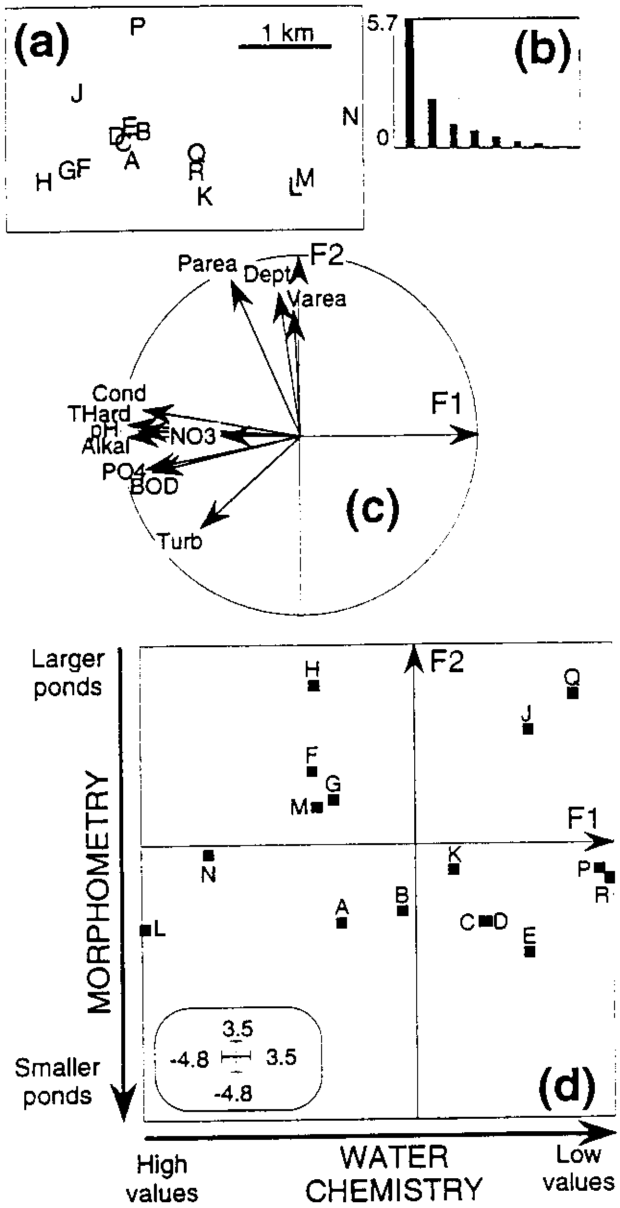


**Fig. 3** Results of the standardized principal component analysis of eleven environmental measurements made by Friday (1987) in sixteen ponds. (a) Geographic map of the sampling area (letters indicate position of ponds). (b) Histogram of eigenvalues. (c) Correlation circle (Varea = vegetated area; Dept = depth; Parea = pond area; Cond = conductivity; THard = total hardness; pH = pH; NO3 = nitrate; Alkal = alkalinity; PO4 = phosphate; BOD = 5-day biological oxygen demand; Turb = turbidity). (d) F1 × F2 factorial plane of ponds.

#### *Use of separate analyses*

The environmental data were processed by standardized principal component analysis. According to the eigenvalues (Fig. 3b), the two first axes are

Dolédec, S., and D. Chessel. 1994. Co-inertia analysis: an alternative method for studying species-environment relationships. Freshwater Biology 31:277-294.
http://pbil.univ-lyon1.fr/R/articles/arti084.pdf

sufficient for demonstrating the structure in the data. The correlation circle (Fig. 3c) demonstrates a clear influence of chemical variables, that have high correlations with axis F1 (conductivity, total hardness, pH, alkalinity, phosphate concentration, BOD). The morphometric features (pond area, vegetated area and depth) are taken into consideration by axis F2. Consequently, the factorial plane of the two first axes (Fig. 3d) arranges ponds according to water chemistry and morphometry.

The macroinvertebrate communities were then subjected to a centred principal component analysis. This put the emphasis on variations in invertebrate abundance and species richness because these variations were the major characteristics studied via multiple regression by Friday (1987). According to the eigenvalues obtained (Fig. 4a), three axes demonstrate the structure of the faunistic data. The positions of the squares on the factorial plane F1 × F2 of the taxa (Fig. 4b) indicate that most species are on the positive side of the F1 axis, in particular for Ephemeroptera, Malacostracea, and Mollusca, but also for Trichoptera, Oligochaeta, and Diptera. Consequently, the sixteen ponds are differentiated on this plane (Fig. 4c) according to species richness. Lower values of species richness are encountered for ponds R, P, E, J, and Q, and higher values are found for ponds N and H. Moreover, ponds L and F have

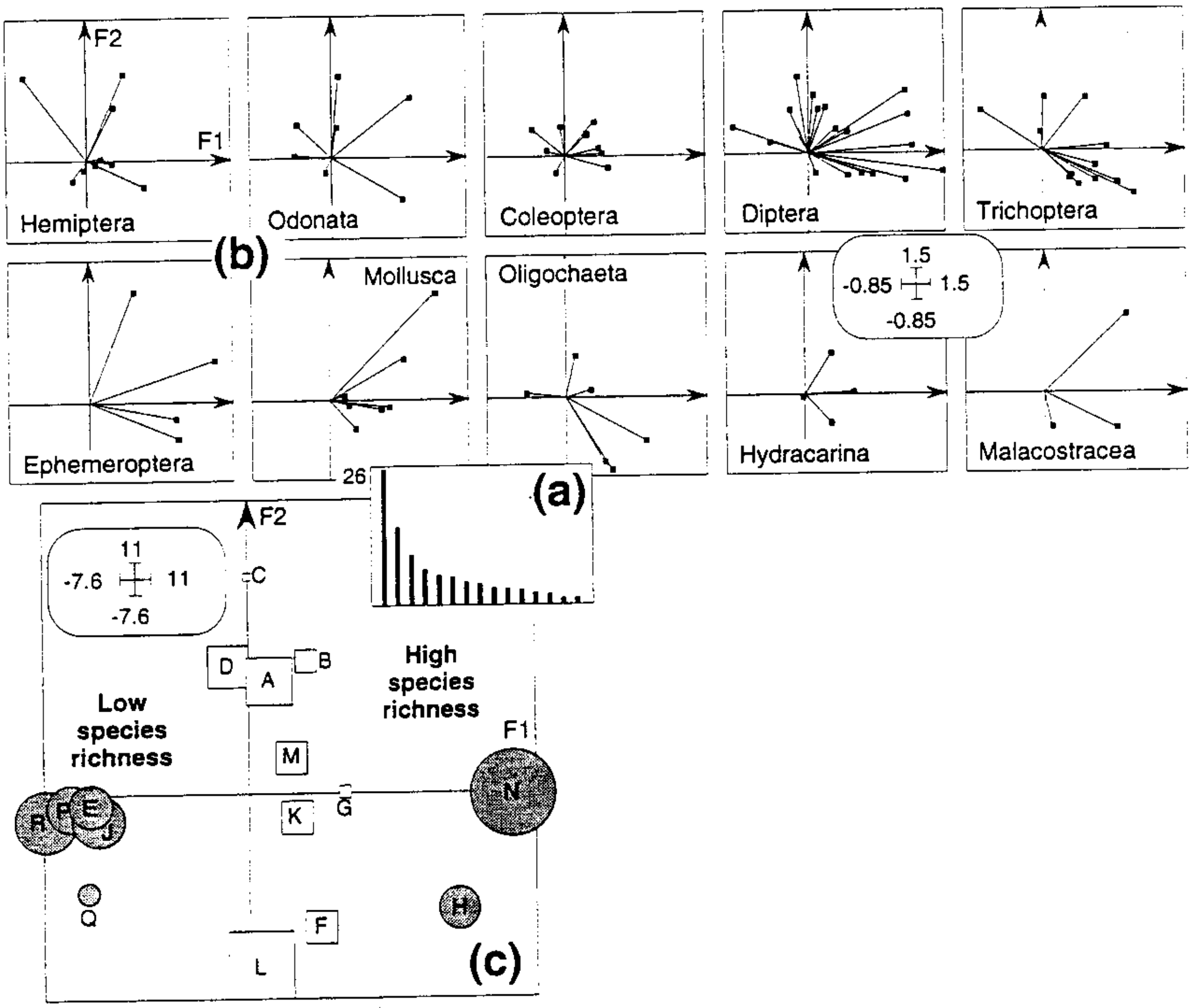


Fig. 4 Results of the centred principal components analysis processed on the abundance table of Friday (1987) (ninety-one taxa collected in sixteen ponds). (a) Histogram of eigenvalues. (b) F1 × F2 factorial plane of taxa (taxa are separated by taxonomic groups for readability). (c) F1 × F2 factorial plane of ponds (cf. Fig. 3a). The pond coordinates on the F3 axis are plotted on the F1 × F2 factorial plane and are proportional to the area of squares (negative values) or circles (positive values).

taxa different from those of ponds A, B, C, and D. Positive values on factorial axis F3 group ponds without *Tubifex tubifex* (R, P, E, J, Q, N, and H; Fig. 4c).

### *Use of co-inertia analysis*

Co-inertia analysis was then processed using the above data. The co-structure described by co-inertia axis F1 is not far from the structures of each data set described by axes F1 in each separate analysis (Fig. 5) because the values of projected variances on axis F1 of the co-inertia analysis were close to the values of projected variances on axes F1 of the standard analyses (Table 1). Hence, the co-inertia analysis was able to demonstrate a co-structure between the environment data data set and the faunistic data set. Such a co-structure is determined by the maximization of the covariance between the two new sets of projected coordinates. This means a simultaneous maximization of the square-rooted projected inertia (which defines the structure of each table separately) and of the correlation between the two new sets of projected coordinates.

To check the significance of the resulting correlation (noted *R* value in Table 1) between the two sets of coordinates resulting from the co-inertia analysis, we carried out 100 co-inertia analyses of the environmental and the faunistic data sets after random permutation of their rows. $R^2$ values were calculated and the frequency distributions of the 100 random $R^2$ values for co-inertia axes F1 and F2 were elaborated. The test for the observed $R^2$ value of co-inertia axis F1 was significant (Fig. 6a) because the observed value was in a class containing only two random values among the 100 possible (i.e. $P < 0.05$). By contrast, almost 10% of the random values are higher than the observed $R^2$ value of co-inertia axis F2 (Fig. 6b). This leads us to keep only one co-structure variable between the environmental data set and the faunistic data set.

**Table 1** Comparison of inertia resulting from the co-inertia analysis with inertia resulting from the separate analyses of each data set. Two co-inertia axes (labelled F1 and F2) are selected. VarE = inertia of the environment table projected on to co-inertia axes; VarF = inertia of the faunistic table projected on to co-inertia axes; InerE = maximal projected inertia of the environment table (first and second eigenvalue of the separate analysis); InerF = maximal projected inertia of the faunistic table (first and second eigenvalue of the separate analysis); Covar = covariance of the two sets of coordinates projected on to co-inertia axes; *R* value = correlation between the two new sets of coordinates resulting from the co-inertia analysis

| Axis | VarE | VarF | InerE | InerF | Covar | *R* value |
|---|---|---|---|---|---|---|
| F1 | 5.23 | 23.66 | 5.34 | 25.26 | 10.27 | 0.925 |
| F2 | 0.96 | 12.35 | 1.54 | 14.66 | 2.95 | 0.855 |

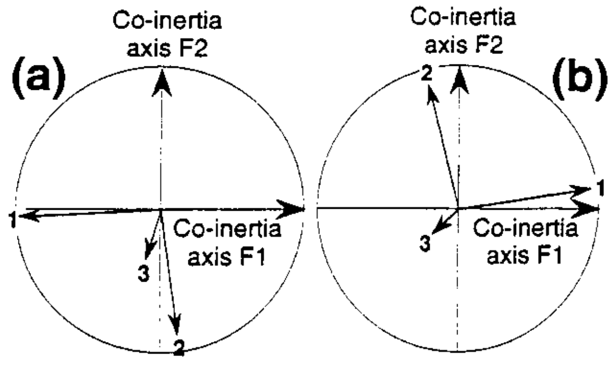


**Fig. 5** Relation between separate analyses and co-inertia analysis. Each smaller arrow represents an axis of the standard analysis. Numbers stand for the axis numbers of the standard analysis. (a) Components of the standardized principal component analysis of the environmental data set projected on to the co-inertia axes. (b) Components of the centred principal component analysis of the faunistic data set projected on to the co-inertia axes.

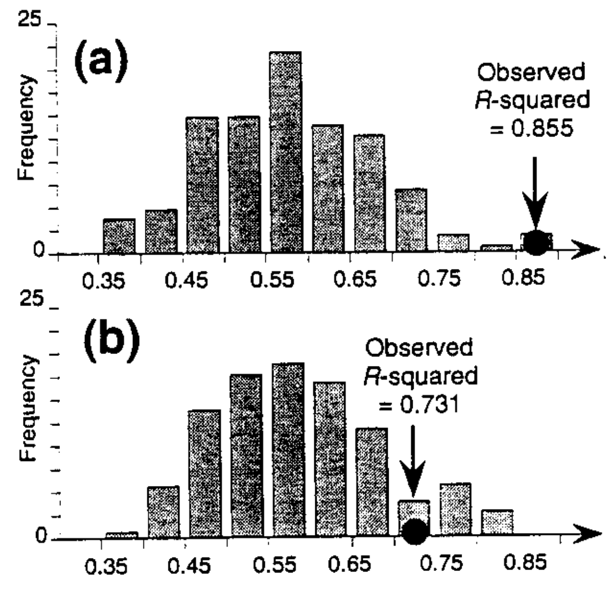


**Fig. 6** Frequency distribution of the first (a) and second (b) axis $R^2$ values for 100 random co-inertia analyses and the observed values for Friday's data (cf. Table 1).

Dolédec, S., and D. Chessel. 1994. Co-inertia analysis: an alternative method for studying species-environment relationships. Freshwater Biology 31:277-294.
http://pbil.univ-lyon1.fr/R/articles/arti084.pdf

The eigenvalue diagram of the co-inertia analysis emphasizes the importance of the first axis (Fig. 7a). However, we have retained two axes for making a graphical representation for separate comparisons with the structure of each data set. Considering only the first co-inertia axis, it is clear that water chemistry of ponds, and especially pH and alkalinity (Fig. 7b), is the most important feature correlated with the distribution of invertebrate fauna and extracted by co-inertia analysis. According to the environmental parameter scores, an environmental ordination of the ponds is achieved that enables four groups (contrasted grey areas in Fig. 7c) to be distinguished from the lower pH (pond R and P) to the higher pH (pond N and L) values. Symmetrically, an ordination diagram of ponds (Fig. 7d) may be achieved using species richness. Furthermore, it is demonstrated that the covariation with pH is different from one taxonomic group to another (Fig. 7e). Hence, Ephemeroptera, Hydracarina, Malacostraca, and Mollusca taxa seem to be the most sensitive to low pH whereas Hemiptera, Coleoptera, Odonata, and Diptera seem to be more or less sensitive. Among Oligochaeta and Trichoptera, only one taxon lives in low pH ponds. Such a decrease in species richness related to low pH was already observed in other freshwater ecosystems (e.g. Raddum & Fjelleim, 1984, in lakes; Weatherley & Ormerod, 1987, in streams). For example, studying Norwegian lakes, Raddum & Fjellheim (1984) indicated that low pH (<5) is associated with an elimin-

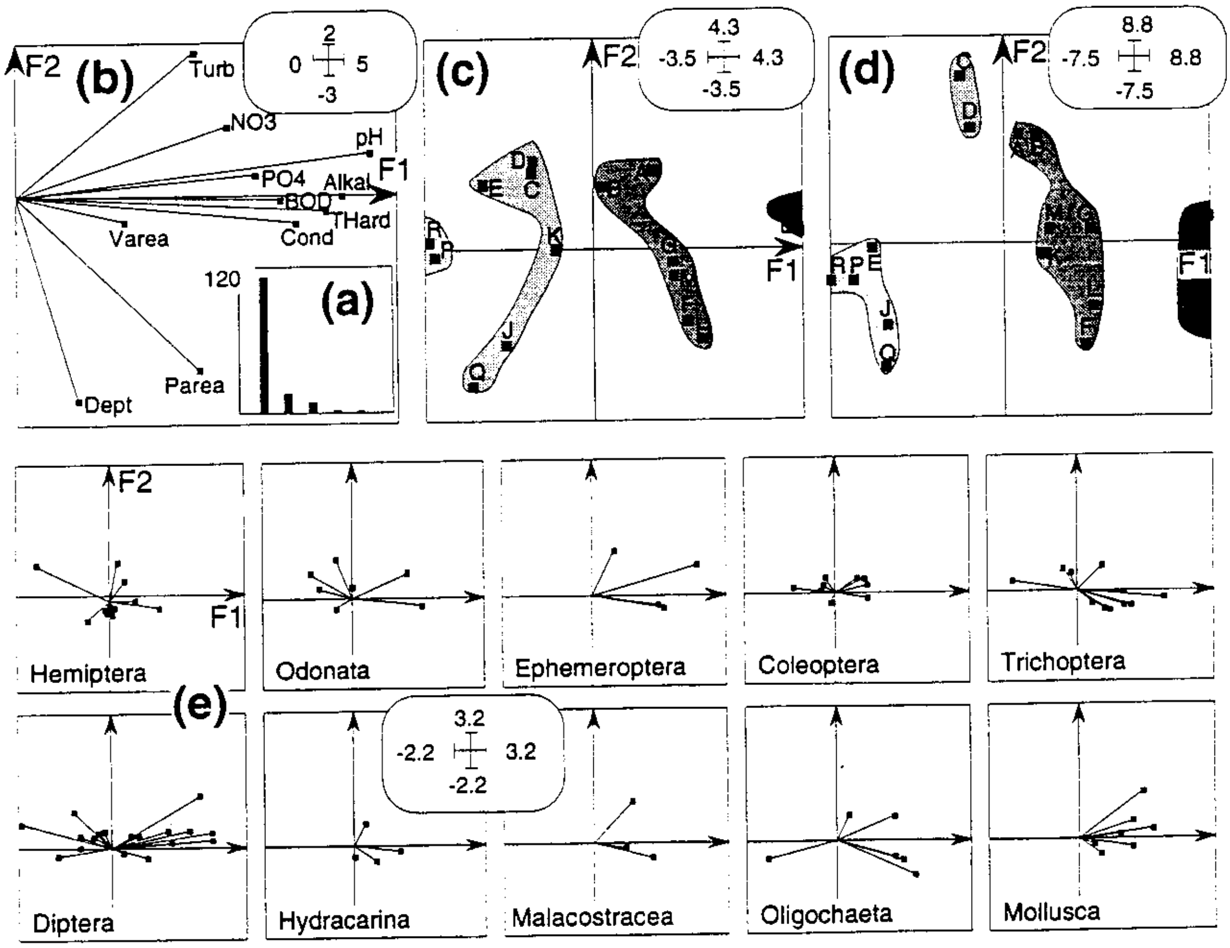


Fig. 7 Results of the co-inertia analysis performed on the eleven environmental variables and the ninety-one taxa. (a) Histogram of eigenvalues. (b) Position of the environmental variables on the F1 × F2 co-inertia plane (cf. Fig. 3c for variable labels). (c) Position of ponds (cf. Fig. 3a) on the F1 × F2 co-inertia plane, using environmental variable co-inertia weights. (d) Position of ponds on the F1 × F2 co-inertia plane, using taxa co-inertia weights. (e) Position of taxa on the F1 × F2 co-inertia plane (taxa are separated by taxonomic groups for readability).

Dolédec, S., and D. Chessel. 1994. Co-inertia analysis: an alternative method for studying species-environment relationships. Freshwater Biology 31:277-294.
http://pbil.univ-lyon1.fr/R/articles/arti084.pdf

ation of snails and leeches, and with a large decrease in species richness for some groups of insects (e.g. Plecoptera, Trichoptera, and Ephemeroptera).

As shown in Table 1, the correlation between the new environmental ordination of ponds and the new faunistic ordination of ponds was high for axis F1 ($R = 0.925$). Hence, ponds were positioned according to their environmental and faunistic scores (Fig. 8a), which illustrates a good fit between the two new ordination sets. Another way to compare the two ordinations from co-inertia analysis is to plot the standardized environmental and faunistic scores together on a factorial map and to link the two positions of a given pond by an arrow (Fig. 8b). Hence, it is possible to discuss the correlation between the list of collected species and the environmental features of ponds. For example, ponds A and B simultaneously have fairly different values of environmental variables (the letters A and B are separated in Fig. 8b) and close faunistic contents (the arrows point to the same area in Fig. 8b).

Finally, some taxa were positioned according to the new environmental ordination (Fig. 9a) to visualize their response to pH and alkalinity increase (Fig. 9b). In this presentation, taxa were arranged according to their co-inertia scores on axis F1, i.e.

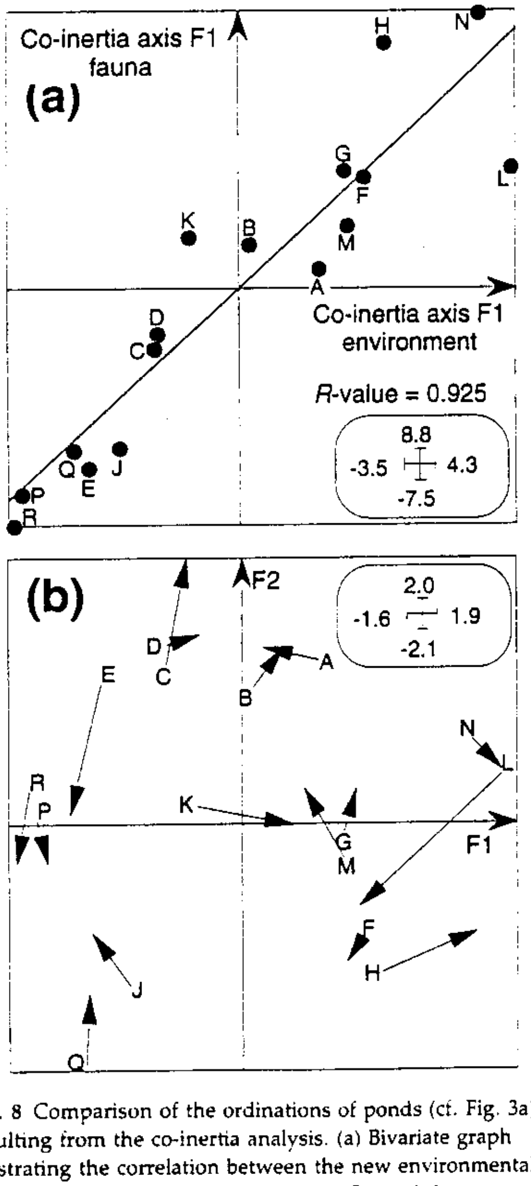


**Fig. 8** Comparison of the ordinations of ponds (cf. Fig. 3a) resulting from the co-inertia analysis. (a) Bivariate graph illustrating the correlation between the new environmental ordination of ponds (environment axis F1) and the new faunistic ordination of ponds (fauna axis F1). (b) Positions of ponds on the F1 × F2 factorial plane using standardized coordinates. Capital letters indicate the positions of ponds resulting from the new environmental ordination. For each pond, the arrow links the latter position to the position resulting from the new faunistic ordination.

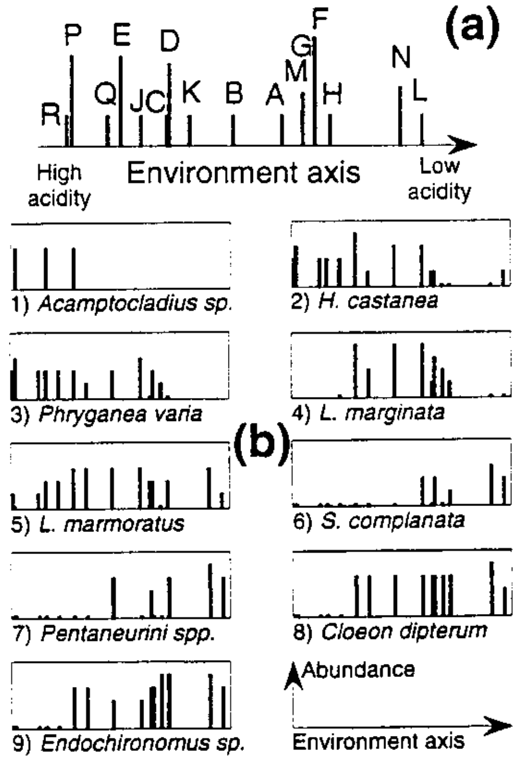


**Fig. 9** (a) New environmental ordination of ponds (cf. Fig. 3a) by a gradient of acidity on co-inertia axis F1. (b) Abundance of selected taxa in the ponds arranged according to their co-inertia scores on axis F1 along with the gradient of acidity. The first three (1–3) taxa show higher abundance in acid ponds, whereas the last four (6–9) taxa demonstrate a higher abundance in low-acidity ponds. The two intermediate taxa (*Leptophlebia marginata, Limnephilus marmoratus*) show higher abundance at intermediate levels of acidity.

Dolédec, S., and D. Chessel. 1994. Co-inertia analysis: an alternative method for studying species-environment relationships. Freshwater Biology 31:277-294.
http://pbil.univ-lyon1.fr/R/articles/arti084.pdf

to their relationships with acidity. For example, *Acamptocladius* sp., *Hesperocorixa castanea*, and *Phryganea varia* are found in high-acidity ponds whereas *Segmentina complanata*, *Endochironomus* sp., *Cloeon dipterum*, and *Pentaneurini* spp. seem to be distributed throughout low-acidity ponds. *Leptophlebia marginata* and *Limnephilus marmoratus* are prominent at an intermediate level of acidity.

## Discussion

As underscored by Mercier (1991) and Mercier *et al.* (1992), the widely used canonical correspondence analysis provides an efficient way of conducting direct gradient analysis. By contrast, co-inertia analysis is more suitable to symmetric analysis and avoids the multicolinearity problem associated with canonical correspondence analysis when the number of environmental variables is close to the number of sampling sites. The major advantage of co-inertia analysis is its ability to maximize the covariance of projected scores. Furthermore, it enables linkage of every standard analysis such as principal component analysis, correspondence analysis, and multiple correspondence analysis processed on the environmental and/or the faunistic data sets. Hence, Tucker's (1958; Fig. 1f) inter-battery analysis was actually the co-inertia analysis of two standardized principal component analyses; Cazes' (1980; Fig. 1g) canonical analysis on categorical variables was actually the co-inertia analysis between two multiple correspondence analyses; and Romane's (1972; Fig. 1h) approach was actually the co-inertia analysis between a correspondence analysis and a multiple correspondence analysis. Thus, co-inertia analysis enables the user to explore various combinations of various types of data.

Ecologists are now confronted with many possible ways of analysing data. While the multiple algorithms associated with multidimensional scaling methods are not well understood mathematically (Gower, 1987), linear ordination methods such as co-inertia analysis have a strong mathematical coherence supported by the Euclidean model, which results in numerical stability and thus robustness. As suggested by Kenkel & Orloci (1986, p. 919), 'none of the currently available ordination strategies is appropriate under all circumstances, and the future research in ordination methodology should emphasize a statistical rather than empirical approach'. In this context, co-inertia analysis serves as a unifying method that extends the usual inertia analyses of one table to the simultaneous analysis of two tables.

Perspectives for freshwater and general ecology can be drawn from such results, and we present briefly five examples that may be developed further in the context of synthesizing large data sets (Fig. 10):
1 Interest may be focused on the typological value of faunistic groups (Fig. 10a) because the costs of complete species surveys are very high as community

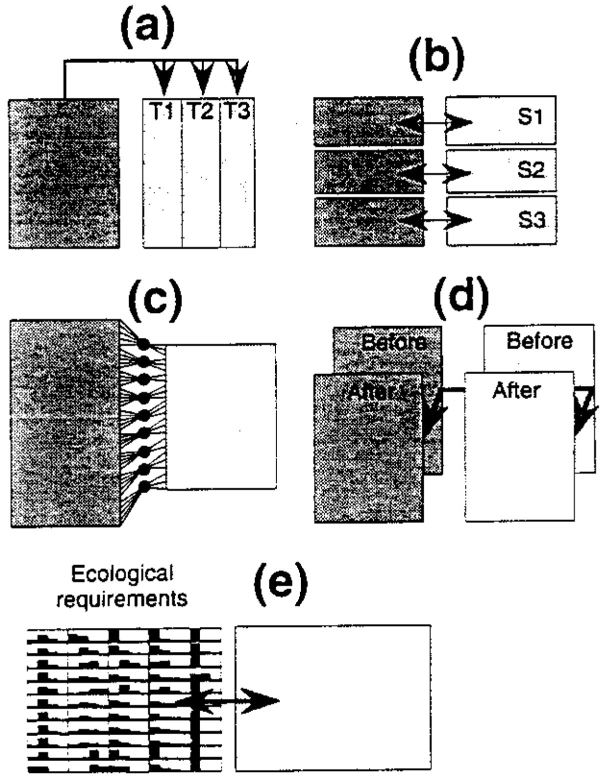


**Fig. 10** Some examples of ecological problems for which co-inertia analysis is an appropriate tool (the dark grey areas indicate the environmental data sets and the light grey areas indicate the faunistic data sets). (a) Study of the typological value of different taxonomic groups (noted T1, T2, T3). (b) Study of the temporal stability of species–environment relationships (S1, S2, S3 are three situations, e.g. sampling dates in which sampling sites are investigated for the same environmental variables and the same faunistic assemblage). (c) Study of the between-class species–environment relationships in the case of a different sampling effort between the environmental and the faunistic data sets such as in long-term biomonitoring. (d) Study of the impact of a human-made disturbance of the species–environment relationships. (e) Connecting ecological requirements accumulated from the literature to a faunistic table.

Dolédec, S., and D. Chessel. 1994. Co-inertia analysis: an alternative method for studying species-environment relationships. Freshwater Biology 31:277-294.
http://pbil.univ-lyon1.fr/R/articles/arti084.pdf

ecologists typically are confronted with many taxonomic groups. Thus, it may be useful to know which taxonomic group is the most relevant for examining a given ecological question. Therefore, this topic aims to relate an environmental array with a faunistic array that is divided into taxonomic (or functional) groups. For example, this topic is underlying the work of Friday (1987) and Richardot-Coulet, Chessel & Bournaud (1986).

2 The second example concerns the intensity or the dynamics of the species–environment relationships (Fig. 10b), which would be particularly useful in freshwater research for studying disturbance and recovery of the species–environment relationships as, for example, affected by flow variations. It has been shown in previous papers (Dolédec & Chessel, 1987, 1989) that within-class analyses could be performed either on faunistic tables or on environmental tables for the study of three-way tables (variable × site × time). Thus, it may be possible to connect these resulting within-class analyses by means of a co-inertia analysis; as a result, the stability of a system could be examined in terms of the variability of the intensity of the link between species and environment.

3 The third example is an extension of the second one (Fig. 10c) and is concerned with long-term surveys. Because of the difference in costs between macroinvertebrate sampling (usually replicated) and environmental parameter measurements (usually not replicated and therefore cheaper than the former), few faunistic samples may be made at one sampling station but far more environmental measurements could be taken. Thus, a comparison of the information given by environmental parameters with the information given by examining macroinvertebrates could indicate the most cost-effective approach to use. The method of doing this would involve studying the between-sample species–environment relationships.

4 The fourth example (Fig. 10d) in freshwater research involves the study of the link between the modification of physical habitats and the modification of species assemblages before and after the incorporation of a management measure (e.g. impoundment, effluent, water diversion). The simultaneous ordination of a table containing the difference in species abundance and a table containing variation of environmental characteristics before and after the implementation of a management plan could be useful in assessing the impact of that management measure on the species–environment relationship.

5 A final example concerns the work of Bournaud, Richoux & Usseglio-Polatera (1992). Their study integrates the ecological requirements of aquatic Coleoptera (Fig. 10e) in the exploratory analysis of a faunistic table via co-inertia analysis, and the authors illustrate the difficulties in gathering, summarizing, and expressing large amounts of ecological information from published or unpublished reports, and even incorporation of the 'experience' from experts in this or that taxonomic group. Bournaud *et al.* (1992, p. 165) propose to consider 'how a species is linked with different qualitative situations (or modalities) that may occur for an ecological or biological variable'. Bournaud *et al.* (1992, p. 166) evaluate such ecological information 'by four levels: 0, the species is never found in the situation (or category); 1, the species has weak links with the situation; 2, the species is fairly strongly related to it; and 3, the species is strongly related to it'. This example is further developed in the subsequent papers of this issue. Furthermore, the authors of papers in this issue have expanded the notion of ecological requirements to include information of species traits based on the available biological information.

## Acknowledgments

Many thanks to Vincent Resh (Department of Entomological Sciences, University California, Berkeley, U.S.A.), Bernhard Statzner (URA CNRS 1451 'Ecologie des Eaux Douces et des Grands Fleuves', Université Lyon I, Villeurbanne, France), Richard Norris (Water Research Centre, University of Canberra, Australia) and Jean-Dominique Lebreton (CEPE CNRS, Montpellier, France), and two anonymous referees for stimulating discussions and helpful comments made on the manuscript.

## Software

Software to perform co-inertia analysis is incorporated in ADE version ≥3.3 (Chessel & Dolédec, 1992). ADE is available free for research and teaching uses, on request to the senior author of this paper. It is also available by anonymous FTP server on biom3.univ-lyon1.fr (134.214.100.42; directory/pub/ADE).

Dolédec, S., and D. Chessel. 1994. Co-inertia analysis: an alternative method for studying species-environment relationships. Freshwater Biology 31:277-294.
http://pbil.univ-lyon1.fr/R/articles/arti084.pdf

## References

Anderson D.J. (1963) The structure of some upland plant communities in Caernarvonshire. III. The continuum analysis. *Journal of Ecology*, **51**, 403–414.

Austin M.P. (1968) An ordination study of a chalk grassland community. *Journal of Ecology*, **56**, 739–757.

Barkham J.P. & Norris J.M. (1970) Multivariate procedures in an investigation of vegetation and soil relations of two beach woodlands, Cotswold Hills, England. *Ecology*, **51**, 630–639.

Belbin L. (1991) Semi-strong hybrid scaling, a new ordination algorithm. *Journal of Vegetation Science*, **2**, 491–496.

Birks H.J.B. & Austin H.A. (1992). *An Annotated Bibliography of Canonical Correspondence Analysis and Related Constrained Ordination Methods 1986–1991*. University of Bergen, Norway, 29 pp.

Bonin G. & Roux M. (1978) Utilisation de l'analyse factorielle des correspondances dans l'étude phyto-écologique de quelques pelouses de l'Apennin lucano-calabrais. *Acta Œcologica*, **13**, 121–138.

Bournaud M., Richoux P. & Usseglio-Polatera P. (1992) An approach to the synthesis of qualitative ecological information issued from aquatic Coleoptera communities. *Regulated Rivers*, **7**, 165–180.

Brian M.V. (1964) Ant distribution in a southern English heath. *Journal of Animal Ecology*, **33**, 451–461.

Cailliez F. & Pagès J.P. (1976) *Introduction à l'analyse des données*. SMASH. Paris, 616 pp.

Cassie R.M. (1972) Fauna and sediments of an intertidal mud-flat: and alternative multivariate analysis. *Journal of Experimental Marine Ecology*, **9**, 55–64.

Cazes P. (1980) L'analyse de certains tableaux rectangulaires décomposé en blocs: généralisation des propriétés rencontrées dans l'étude des correspondances multiples. I. Définitions et applications à l'analyse canonique des variables qualitatives. *Cahiers de l'Analyse des Données*, **5**, 145–161.

Chang D.M.S. & Gauch H.G. (1986) Multivariate analysis of plant communities and environmental factors in NGARI. *Ecology*, **67**, 1568–1575.

Chessel D. & Dolédec S. (1992) *ADE Software. Multivariate Analyses and Graphical Display for Environmental Data (version 3.4). Vol. 1. User's Manual* (in English), *Vol. 2. Examples* (in French), *Vol. 3. Graphics Documentation* (in French). Université Lyon I, 523 pp.

Chessel D. & Mercier P. (1993) Couplage de triplets statistiques et liaisons espèces-environnement. *Biométrie et Environment* (Eds J.D. Lebreton and B. Asselain), pp. 15–44. Masson, Paris.

Chessel D., Lebreton J.D. & Yoccoz N. (1987) Propriétés de l'analyse canonique des correspondances. Une utilisation en hydrobiologie. *Revue de Statistique Appliquée*, **35**, 55–72.

Corkum L.D. & Ciborowski J.H. (1988) Use of alternative classifications in studying broad-scale distributional patterns of lotic invertebrates. *Journal of the North American Benthological Society*, **7**, 167–179.

Dagnélie P. (1965) L'étude des communautés végétales par l'analyse statistique des liaisons entre les espèces et les variables écologiques: principes fondamentaux, un exemple. *Biometrics*, **21**, 345–61.

Dale M.B. (1875) On objectives of ordination methods. *Vegetatio*, **30**, 15–32.

Depiereux E., Feytmans E. & Micha J.C. (1983) Utilisation critique de l'analyse en composantes principales et du cluster analysis pour la description d'échantillons d'invertébrés benthiques en eau douce. *Oikos*, **40**, 81–94.

Dolédec S. & Chessel D. (1987) Rythmes saisonniers et composantes stationnelles en milieu aquatique. I. Description d'un plan d'observation complet par projection de variables. *Acta Œcologica*, **8**, 403–426.

Dolédec S. & Chessel D. (1989) Rythmes saisonniers et composantes stationnelles en milieu aquatique. II. Prise en compte et élimination d'effets dans un tableau faunistique. *Acta Œcologica*, **10**, 207–232.

Dolédec S. & Chessel D. (1991) Recent developments in linear ordination methods for environmental sciences. *Advances in Ecology*, **1**, 133–155.

Escoufier Y. (1987) The duality diagram: a means of better practical applications. *Development in Numerical Ecology* (Eds P. Legendre and L. Legendre), pp. 139–156. Springer-Verlag, Berlin.

Faith D.P. & Norris R.H. (1989) Correlation of environmental variables with patterns of distribution and abundance of common and rare freshwater macroinvertebrates. *Biological Conservation*, **50**, 77–98.

Feoli E. & Orloci L. (1985) Species dispersion profiles of anthropogenic grasslands in the Italian Eastern Pre-Alps. *Vegetatio*, **60**, 113–118.

Friday L. (1987) The diversity of macroinvertebrate and macrophyte communities in ponds. *Freshwater Biology*, **18**, 87–104.

Gibson D.J. & Hulbert L.C. (1987) Effects of fire, topography and year to year climatic variation on species composition in tallgrass prairie. *Vegetatio*, **72**, 175–185.

Gittins R. (1985) *Canonical Analysis, a Review with Applications in Ecology*. Springer-Verlag, Berlin, 350 pp.

Goodall D.W. (1954) Objective methods for the classification of vegetation III. An essay in the use of factor analysis. *Australian Journal of Botany*, **2**, 304–324.

Gower J.C. (1984) Multivariate analysis: ordination multidimensional scaling and allied topics. *Handbook of Applicable Mathematics. Statistics* (Ed E.H. Lloyd), pp. 721–781. Wiley, Chichester.

Gower J.C. (1987) Introduction to ordination techniques. *Development in Numerical Ecology* (Eds P. Legendre and L. Legendre), pp. 3–64. Springer-Verlag, Berlin.

Hill M.O. (1974) Correspondence analysis: A neglected multivariate method. *Journal of the Royal Statistical Society (London)*, **23**, 340–354.

Hill M.O. (1988) How effective is ordination as means of relating vegetation to ecological factors? *Bulletin de la Société Royale de Botanique de Belgique*, **121**, 134–141.

Hill M.O. & Gauch H.C. (1980) Detrended correspondence analysis: an improved ordination technique. *Vegetatio*, **42**, 47–58.

Hoskuldsson A. (1988) Partial least square regression methods. *Journal of Chemometrics*, **2**, 211–228.

Johnson R.K. & Wiederholm T. (1989) Classification and ordination of profundal macroinvertebrate communities in nutrient poor, oligo-mesohumic lakes in relation to environmental data. *Freshwater Biology*, **21**, 375–386.

Kenkel N.C. & Orloci L. (1986) Applying metric and nonmetric multidimensional scaling to ecological studies: some new results. *Ecology*, **67**, 919–928.

Lapchin L. & Roux C. (1977) Utilisation de l'analyse des correspondances pour l'étude de la répartition des invertébrés benthiques dans deux ruisseaux à Salmonidés de Bretagne. *Annales d'Hydrobiologie*, **8**, 333–354.

Lebreton J.D., Chessel D., Prodon R. & Yoccoz N. (1988) L'analyse des relations espèces–milieu par l'analyse canonique des correspondances. I. Variables de milieu quantitatives. *Acta Œcologica*, **9**, 53–67.

Lebreton J.D., Sabatier R., Banco G. & Bacou A.M. (1991) Principal component and correspondence analyses with respect to instrumental variables: an overview of their role in studies of structure–activity and species–environment relationships. *Applied Multivariate Analysis in SAR and Environmental Studies* (Eds J. Devillers and W. Karcher), pp. 85–114. Kluwer Academic, Dordrecht.

Lindgren F., Geladi P. & Wold S. (1993) The kernel algorithm for PLS. *Journal of Chemometrics*, **7**, 45–59.

Ludwig J.A. & Reynolds J.F. (1988) *Statistical Ecology*. Wiley, New York, 337 pp.

Mercier P. (1991) *Etude des relations espèces-environnement et analyse de la co-structure d'un couple de tableaux*. Thèse, Université Lyon I, 168 pp.

Mercier P., Chessel D. & Dolédec S. (1992) Complete analysis of an ecological profile data table: a central ordination method. *Acta Œcologica*, **13**, 25–44.

Montana C. & Greig-Smith P. (1990) Correspondence analysis of species by environmental variable matrices. *Journal of Vegetation Sciences*, **1**, 453–460.

Orloci L. (1988) Detecting vegetation pattern. *Atlas of Science: Plants & Animals*, **1**, 173–177.

Prodon R. (1988) *Dynamique des systèmes avifaune-végétation après déprise rurale et incendies dans les pyrénées méditerranéennes siliceuses*. Thèse, Université Paris VI. 333 pp.

Prodon R. & Lebreton J.D. (1981) Breeding avifauna of a Mediterranean succession: the holm oak and cork oak series in the eastern Pyrénées. 1. Analysis and modelling of the structure gradient. *Oikos*, **37**, 21–38.

Rabeni C.F. & Gibbs E. (1980) Ordination of deep river invertebrate communities in relation to environmental variables. *Hydrobiologia*, **74**, 67–76.

Raddum G.G. & Fjellheim A. (1984) Acidification and early warning organisms in freshwater in western Norway. *Verhandlungen der Internationalen Vereinigung für theoretische und angewandte Limnologie*, **22**, 1973–1980.

Richardot-Coulet M., Chessel D. & Bournaud M. (1986) Typological value of the benthos of old beds of a large river. Methodological approach. *Archiv für Hydrobiologie*, **107**, 363–383.

Romane F. (1972) Un exemple d'utilisation de l'analyse factorielle des correspondances en écologie végétale. *Basic Problems and Methods in Phytosociology* (Eds E. van der Maarel and R. Tüxen), pp. 151–167. Junk, The Hague.

Roux G. & Roux M. (1967) A propos de quelques méthodes de classification en phytosociologie. *Revue be Statistique Appliquée*, **15**, 59–72.

Sabatier R. (1987) Analyse factorielle de données structurées et métriques. *Statistique et Analyse des Données*, **12**, 75–96.

Sabatier R. & Van Campo M. (1984) L'analyse en composantes principales de variables instrumentales appliquées à l'estimation des paléoclimats de la Grèce, il y a 18 000 ans. *Bulletin de la Société Botanique Française*, **131**, 85–96.

Sabatier R., Lebreton J.D. & Chessel D. (1989) Principal component analysis with instrumental variables as a tool for modelling composition data. *Multiway Data Analysis* (Eds R. Coppi and S. Bolasco), pp. 341–352. Elsevier, Amsterdam.

Storey A.W., Bunn S.E., Davies P.M. & Edward D.H. (1990) Classification of the macroinvertebrate fauna of two river systems in southwestern Australia in relation to physical and chemical parameters. *Regulated Rivers*, **5**, 217–232.

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. (1987) The analysis of vegetation–environment relationships by canonical correspon-

Dolédec, S., and D. Chessel. 1994. Co-inertia analysis: an alternative method for studying species-environment relationships. Freshwater Biology 31:277-294.
http://pbil.univ-lyon1.fr/R/articles/arti084.pdf

dence analysis. *Vegetatio*, **69**, 69–77.

Townsend C.R. & Hildrew A.G. (1994) Species traits in relation to a habitat templet for river systems. *Freshwater Biology*, **31**, 265–275.

Tucker L.R. (1958) An inter-battery method of factor analysis. *Psychometrika*, **23**, 111–136.

Vincent B. (1981) Profondeur, vase et courant, facteurs de micro-répartition transversale du benthos dans l'estuaire d'eau douce du Saint-Laurent (Québec). *Canadian Journal of Zoology*, **59**, 2297–2305.

Weatherley N.S. & Ormerod S.J. (1987) The impact of acidification on macroinvertebrate assemblages in Welsh streams: towards an empirical model. *Environmental Pollution*, **46**, 223–240.

Whittaker R.H. (1967) Gradient analysis of vegetation. *Biological Reviews of the Cambridge Philosophical Society*, **42**, 207–264.

Wiens J.A. (1986) Spatial scale and temporal variation in studies of shrubsteppe birds. *Community Ecology* (Eds J. Diamond and T.J. Case), pp. 154–172. Harper and Row, New York.

Wright J.F., Moss D., Armitage P.D. & Furse M.T. (1984) A preliminary classification of running-water sites in Great Britain based on macro-invertebrate species and the prediction of community type using environmental data. *Freshwater Biology*, **14**, 221–256.

(*Manuscript accepted October 1993*)

## Appendix 1 The mathematical model of co-inertia analysis

The mathematical model of co-inertia analysis may be examined using the duality diagram synthesized in Cailliez & Pagès (1976) and introduced in statistical ecology by Escoufier (1987). Using such notations, the term co-inertia may be justified as follows.

Let $(\mathbf{X}, \mathbf{D}_p, \mathbf{D}_n)$ and $(\mathbf{Y}, \mathbf{D}_q, \mathbf{D}_n)$ be two statistical triplets, table **X** is the environmental data set (after an initial transformation); $\mathbf{D}_p$ contains the weight associated with the columns of table X; $\mathbf{D}_n$ contains the weight associated with the rows of table **X**. Table **Y** is the faunistic data set (after an initial transformation); $\mathbf{D}_q$ contains the weight associated with the columns of table Y; $\mathbf{D}_n$ contains the weight associated with the columns of table Y (see Dolédec & Chessel, 1991).

The first statistical triplet $(\mathbf{X}, \mathbf{D}_p, \mathbf{D}_n)$ defines an inertia analysis of **n** points in a multidimensional space noted $\mathbb{R}^p$ and of $p$ points in a multidimensional space noted $\mathbb{R}^n$. After diagonalization, $r$ principal axes preserved and matrices $\mathbf{R}_r$, $\mathbf{C}_r$, $\mathbf{N}_r$ are generated. $\mathbf{R}_r$ contains the scores of the $n$ rows on the $r$ axes. $\mathbf{C}_r$ contains the scores of the $p$ rows on the $r$ axes. $\mathbf{N}_r$ contains the eigenvalues $(\nu_1 \ldots \nu_r)$.

The second statistical triplet $(\mathbf{Y}, \mathbf{D}_q, \mathbf{D}_n)$ defines an inertia analysis or $n$ points in a multidimensional space noted $\mathbb{R}^q$ and of $q$ points in a multidimensional space noted $\mathbb{R}^n$. After diagonalization, $s$ principal axes are preserved and matrices $\mathbf{R}_s$, $\mathbf{C}_s$, $\mathbf{M}_s$ are generated. $\mathbf{R}_s$ contains the scores of the $n$ rows on the $s$ axes. $\mathbf{C}_s$ contains the scores of the $q$ rows on the $s$ axes. $\mathbf{M}_s$ contains the eigenvalues $(\mu_1 \ldots \mu_s)$.

Let **u** and **v** be a pair of vectors. The former is normalized by matrix $\mathbf{D}_p$ in the multidimensional space $\mathbb{R}^p$ and the latter is normalized by matrix $\mathbf{D}_q$ in the multidimensional space $\mathbb{R}^q$. The projection of the multidimensional space associated with table **X** on to vector **u** generates $n$ coordinates in a column matrix:

$$\xi = \mathbf{X}\mathbf{D}_p\mathbf{u} \tag{1}$$

The projection of the multidimensional space associated with table **Y** on to vector **v** generates $n$ coordinates in a column matrix:

$$\psi = \mathbf{Y}\mathbf{D}_q\mathbf{v} \tag{2}$$

Co-inertia associated with the pair of vectors **u** and **v** is equal to:

$$H(\mathbf{u}, \mathbf{v}) = \xi^t \mathbf{D}_n \psi \text{ (t for transposed)} \tag{3}$$

If the initial tables (**X** and **Y**) are centred, then the co-inertia is the covariance between the two new scores:

$$Cov(\xi, \psi) = (Iner_1(\mathbf{u}))^{\frac{1}{2}}(Iner_2(\mathbf{v}))^{\frac{1}{2}}Corr(\xi, \psi) \tag{4}$$

with $Iner_1$ (**u**) as the projected inertia on to vector **u**, i.e. the variance of the new scores on **u**, $Iner_2$ (**v**) as the projected inertia on to vector **v**, i.e. the variance of the new scores on **v**, and $Corr(\xi, \psi)$ as the correlation between the two coordinate systems. Note that the square of the latter entity $Corr(\xi, \psi)$ is maximized via canonical correlation analysis. In contrast, a co-inertia axis associated with a pair of vectors **u** and **v** maximizes $Cov(\xi, \psi)$.

To obtain co-inertia axes, one diagonalizes the matrix **W**:

Dolédec, S., and D. Chessel. 1994. Co-inertia analysis: an alternative method for studying species-environment relationships. Freshwater Biology 31:277-294.
http://pbil.univ-lyon1.fr/R/articles/arti084.pdf

$$W = D_p^{\frac{1}{2}} X^t D_n Y D_q Y^t D_n X D_p^{\frac{1}{2}} \quad (5)$$

(for a proof see Tucker, 1958, for a matching between two principal component analyses, and Chessel & Mercier, 1993, for the general case). Let $\mathbf{U}_z$ be the matrix that contains the first $z$ normalized eigenvectors of $\mathbf{W}$ and $\Lambda_z$ be the matrix that contains the first $z$ corresponding eigenvalues (noted $\lambda_k$, $1 \le k \le z$). The first $z$ co-inertia axes (associated norm $\mathbf{D}_p$) in $\mathbb{R}^p$ are obtained as:

$$A_z = D_p^{-\frac{1}{2}} U_z \quad (6)$$

The first $z$ co-inertia axies (associated norm $\mathbf{D}_q$) in $\mathbb{R}^q$ are obtained as:

$$B_z = Y^t D_n X D_p^{\frac{1}{2}} U_z \Lambda_z^{-\frac{1}{2}} \quad (7)$$

$\mathbf{Y}^t$ the transposed matrix of $\mathbf{Y}$. $\mathbf{A}_z$ and $\mathbf{B}_z$ are called optimal co-inertia weights of the variables, respectively, in table X and table Y. The co-inertia scores of table X rows are obtained as:

$$X_z^* = X D_p A_z \quad (8)$$

The co-inertia scores of table Y rows are obtained as:

$$Y_z^* = Y D_q B_z \quad (9)$$

Furthermore, one may compare the projected variability resulting from the separate analyses and that from co-inertia analysis by calculating the scores of the initial inertia axes projected on to the co-inertia axes. Let $\mathbf{C}_r^*$ and $\mathbf{C}_s^*$ be the resulting scores with:

$$C_r^* = N_r^{-\frac{1}{2}} C_r^t D_p A_z \quad (10)$$

and

$$C_s^* = M_s^{-\frac{1}{2}} C_s^t D_q B_z \quad (11)$$

We call the diagonal elements of matrix $(\mathbf{X}_z^*)^t \mathbf{D}_n \mathbf{X}_z^*$ and matrix $(\mathbf{Y}_z^*)^t \mathbf{D}_n \mathbf{Y}_z^*$ pseudo-eigenvalues. Let $\nu_k^*$ be the $k$th pseudo-eigenvalue of table X and $\mu_k^*$ be the $k$th pseudo-eigenvalue of table Y. Such values are useful in situating the value of co-inertia axes in comparison with inertia axes. Finally, the quantity:

$$\rho_k^{*2} = \frac{\lambda_k}{\nu_k^* \mu_k^*} \quad (12)$$

is an expression of the correlation between the two new sets of coordinates.

Dolédec, S., and D. Chessel. 1994. Co-inertia analysis: an alternative method for studying species-environment relationships. Freshwater Biology 31:277-294.
http://pbil.univ-lyon1.fr/R/articles/arti084.pdf