

SIMULTANEOUS ANALYSIS OF A SEQUENCE OF PAIRED ECOLOGICAL TABLES

JEAN THIOULOUSE,^{1,3} MONIQUE SIMIER,² AND DANIEL CHESSEL¹

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

²Centre de Recherche Halieutique IFREMER/IRD, Avenue Jean Monnet, B.P. 171, 34203 Sète Cedex, France

Abstract. In this paper, we present a statistical method called STATICO that can be used to analyze series of pairs of ecological tables. The objective of this method is to find the stable part in the dynamics of the relationships between the species and their environment. The treated data are a sequence of pairs of ecological tables. Each pair is made of a species abundance table (species in columns) and an environmental variables table (variables in columns). The sampling sites (in rows) must be the same for the two tables of one pair, but they may be different among the pairs. The environmental variables must be the same in all the environmental tables, and the list of species must be the same in all the species tables too, although some species may be absent from some tables (the corresponding columns will contain all zeros). From a statistical point of view, STATICO is a multitable analysis (partial triadic analysis) performed on the series of cross-tables resulting from the co-inertia analysis of each pair of tables. A small ecological example data set is analyzed and the results are discussed to show how this method can be used to extract the stable part of species–environment relationships. All computations and graphical displays can be performed with free software available on Internet.

Key words: ADE-4; canonical correspondence analysis; co-inertia analysis; compromise; inter-structure; partial triadic analysis; redundancy analysis; series of paired ecological tables; STATICO; STATIS; three-way table.

INTRODUCTION

The analysis of two data tables by multivariate analysis methods leads to two ordinations of the sampling sites. When the two tables make up an ecological pair (one table containing species composition and the other one environmental conditions for the same sampling sites), the relationships between these two ordinations are interesting to explore. In this case, the structure common to these two ordinations (called the co-structure) can be analyzed by canonical correspondence analysis (CCA, ter Braak 1986), redundancy analysis (RDA, Wollenberg 1977), or co-inertia analysis (Dolédec and Chessel 1994).

When the pair of ecological tables is repeated, for example in time or in space, one can try to examine the temporal or spatial stability of the relationships between the two tables. Within-class and between-class co-inertia analysis (Franquet and Chessel 1994, Franquet et al. 1995) were proposed with this aim, considering each repetition of the pair as a class. More recently, Simier et al. (1999) have proposed the STATICO (STATIS and CO-inertia) method, which is an application of the STATIS (Structuration des TABLEaux à Trois Indices de la Statistique) method (Lavit et al. 1994) to co-inertia operators. In fact, STATICO is a partial triadic analysis (PTA, Thioulouse and Chessel

1987) of the series of cross-tables coming from the co-inertia analyses performed on each pair of tables. This method combines the objectives of STATIS (finding the stable part of the structure of a series of tables) and the objectives of co-inertia analysis (finding the common structure of two data tables). The advantage of STATICO over within-class and between-class co-inertia analysis and other ad hoc methods is that it provides a complete and consistent analysis framework, with a much stronger mathematical background. It offers a detailed pathway through the numerous steps of analyzing a sequence of paired ecological tables, with a rigorous and reproducible approach, and many graphical aids.

Starting from an ecological standpoint, we describe the advantages of the method and compare it to other ad hoc approaches. We discuss the advantages of using co-inertia analysis in this context over CCA or RDA. The mathematical description of co-inertia analysis, PTA, and STATICO are presented in several appendices using simple matrix notations. We present an application of the method using a simple ecological data set, with a detailed appendix enabling users to redo all the computations step by step.

METHODOLOGICAL BASES

Co-inertia analysis

Co-inertia analysis belongs to the family of two-table coupling methods, which include canonical correspondence analysis (CCA), redundancy analysis (RDA), and

Manuscript received 30 September 2002; revised 25 March 2003; accepted 26 March 2003. Corresponding Editor: N. Kenkel.

³ E-mail: jthioulouse@biomserv.univ-lyon1.fr

canonical analysis (Gittins 1985). However, co-inertia analysis differs in two ways from CCA and RDA, and also from canonical analysis. The principle of co-inertia analysis is very simple: it searches for axes that maximize the covariance between the row coordinates of the two tables. This principle is similar to that of canonical analysis, except that the covariance is maximized in co-inertia analysis, instead of the correlation as in canonical analysis. An important step in co-inertia analysis is the computation of a cross-table. For example, in the case of a PCA–PCA co-inertia analysis (i.e., a co-inertia analysis coupling two principal components analyses, which is exactly the inter-battery analysis of Tucker [1958]), this cross-table contains the cross-covariances between the variables of the two tables. Appendix A gives a formal description of co-inertia analysis in matrix form.

Two-table coupling methods can be separated into symmetric and asymmetric methods, according to the way the two tables are treated. CCA and RDA are often seen as asymmetric methods, because the focus of the analysis is on the species table, considered as representing a response to environmental forcings. Indeed, in CCA and RDA there is a regression step ensuring that sampling scores are linear combinations of environmental variables. From this point of view, canonical analysis and co-inertia analysis are symmetric, because the two tables have symmetric roles in the analysis. They are different however, because in co-inertia analysis there is no regression step, while canonical analysis can be seen as a double reciprocal multiple regression (Lebart et al. 1984: Chapter III).

These methodological differences between symmetric and asymmetric methods have important consequences when considering the properties of the methods: the regression step implies that in CCA, RDA, and canonical analysis, the number of samples must be high compared to the number of environmental variables. Co-inertia analysis comprises no regression step and does not have this constraint. Dray et al. (2003) show the importance of this problem when comparing CCA and co-inertia analysis results, and the potential drawbacks of the regression step.

But there is another dissymmetry that has more biological meaning, and one where co-inertia analysis has still more advantages. It is the dissymmetry in the biological meaning of the two tables, and in the way that this meaning is taken into account by the analysis. In this sense, RDA is a symmetric method, while CCA is an asymmetric method: in RDA, both tables are analyzed by a PCA, while in CCA, one table is analyzed by a PCA and the other by a correspondence analysis (CA). In the same way, a PCA–PCA co-inertia analysis is a symmetric method, while a PCA–CA co-inertia analysis is an asymmetric method. In RDA and in a PCA–PCA co-inertia analysis, the scalar product between two columns x and y (one in each table) is equal to the covariance between the two columns: $\langle x|y \rangle_{D_n}$

$= 1/n_j \sum_i x_i y_i = \text{cov}(x, y)$, n being the number of samples. This is why RDA is considered as a method that can only be used when the species response curve along the ecological gradient is linear: the scalar product gives directly the slope of the regression line. In CCA and in a PCA–CA co-inertia analysis, this scalar product is equal to the mean (m_j) of the species distribution along the ecological gradient. Indeed, in this case, the x_i are the species abundance in the samples: $\langle x|y \rangle_{D_n} = 1/n_j \sum_i x_i y_i = m_j$ (the n_j are the species totals). This is why CCA can be used when the species response curve is unimodal: a unimodal response curve is more appropriately approximated by its mean than a linear trend.

The same reasoning can be used in co-inertia analysis, but instead of being limited to these two alternatives (CCA and unimodal species response curves vs. RDA and linear species response curves), many other methods can be explored. The two simplest ones, corresponding to RDA and CCA, are the PCA–PCA co-inertia analysis, where the two tables are analyzed by PCA, and the PCA–CA co-inertia analysis, where the environmental variables table is analyzed with a PCA and the species table is analyzed with a correspondence analysis. The case of the MCA–CA co-inertia analysis is also particular: it links a multiple correspondence analysis (MCA, Tenenhaus and Young 1985) and a correspondence analysis, and it is exactly the same method as the analysis of ecological profiles (Romane 1972, Montaña and Greig-Smith 1990, Mercier et al. 1992, Reynaud and Thioulouse 2000).

Co-inertia analysis can also be used with many other analyses. Dray et al. (2003) give several examples and a complete presentation of these possibilities. The number of possibilities allowed for each of the two tables is high, and the number of possible co-inertia analyses is two times higher (since two tables are involved). Only a small number of these possibilities have so far been explored. All of them have different mathematical properties that need to be properly examined, although probably not all of them are equally interesting. But this flexibility will allow ecologists to outgrow the opposition between unimodal vs. linear species response curve models, and provide them with statistical analysis methods much more adapted to their data than only two standard procedures.

Partial triadic analysis

Partial triadic analysis (PTA) was derived from triadic analysis (Tucker 1966) and introduced in ecology by Thioulouse and Chessel (1987). The name “partial” triadic analysis comes in fact from Kroonenberg (1989). Several examples of application can be found in ecology (Dolédec 1988, Centofanti et al. 1989, Aliaume et al. 1993, Blanc and Beaudou 1998, Blanc et al. 1998). The aim of this method is to analyze a three-way table (i.e., a data cube; for example, species \times sites \times time), seen as a sequence of two-way tables.

It is therefore different from Kroonenberg's three-mode PCA (Kroonenberg 1983), and from Tucker's triadic analysis. Partial triadic analysis belongs to the family of the STATIS methods, as described in the book by Lavit (1988), but in PTA all the tables must have the same rows and the same columns, while in the STATIS method, the rows or the columns (but not both) may be different among the tables. These methods are based on the concepts of vector variance, vector covariance, and vector correlation (see Escoufier [1973] for a mathematical presentation of these notions). The vector variance of a table is simply the variance of the vector obtained by rewriting the columns of the centered table one below the other instead of side by side. The vector covariance and vector correlation between two tables is simply the covariance and the correlation of the two vectors obtained in the same way. They are not to be mistaken for the matrix of cross-covariances between the variables of two tables, which is used for example in co-inertia analysis.

Partial triadic analysis, like all the methods of the STATIS family, comprises three steps: the interstructure, the compromise, and the trajectories (Lavit et al. 1994). These three steps are explained below, and Appendix B gives a formal description of PTA in matrix form.

First step: the interstructure. This step gives the "importance" of each table: a matrix of scalar products between the tables is computed, and the components of the first eigenvector of this matrix are used as weights of the tables to compute the compromise. An important question here is the choice of the components of this first eigenvector to weight the tables. It is justified by the fact that the linear combination obtained with this weighting is mathematically optimal. The inertia (sum of all the variances) of the compromise is maximal (with the constraint that the sum of the squares of the weights must be equal to one) when using the components of the first eigenvector. Any other weighting (uniform weights, for example) would produce a compromise having a smaller inertia, and would thus be less representative of the series of tables. Note that this mathematical property is not particular to PTA: it is used in all the methods of the STATIS family. One particularity of PTA is that the components of the first eigenvector are not always strictly positive, contrary to the other STATIS methods. This may be a problem, as a negative component would mean that the corresponding table has a meaning opposed to the others, and that the overall structure of the series of tables is weak. In such a case, PTA should be avoided, but could be replaced by a classical STATIS method (STATIS on operators). Another solution would be to remove the table having the negative component.

The first two axes can also be used to display the series of tables on a factor map, as in a simple PCA. This factor map summarizes the global structure and

the relationships between the tables. This is not, however, the main goal of this step.

Second step: the compromise. This is the main step of the analysis. The compromise table is computed as the weighted mean of all the tables of the series, using the components of the first eigenvector of the interstructure as weights. This table is called the compromise, and it has the same dimensions and the same structure and meaning as the tables of the series. It is analyzed by a PCA, giving a picture of the structures common to all the tables.

Third step: the trajectories. The rows and columns of all the tables of the series are projected on the factor map of the PCA of the compromise as additional elements. The points can then be linked by lines to underline their trajectories. Each table can also be analyzed independently (for example, by a separate PCA of each table) and the axes and components of these PCAs can be projected on the factor maps (of the rows and of the columns, respectively) of the PCA of the compromise. This step summarizes the variability of the series of tables around the common structure defined by the compromise.

THE STATICO METHOD

STATICO is a three-step method: (1) each table is first analyzed by a basic analysis (i.e., a one-table method), for example PCA, CA, or MCA; (2) each pair is then linked by a co-inertia analysis, producing a cross-table; (3) PTA is used to analyze the series of cross-tables. Fig. 1 shows a summarized flow chart of the STATICO method.

The series of tables is made up of K pairs. The k th pair consists of one table of species composition and one table of environmental conditions. All the species composition tables must have the same species (in columns), and all the environmental conditions tables must have the same environmental variables (in columns). However, not all species need be present in all tables. The sampling sites (in rows) must be the same for the two tables of one pair, but they may vary among the pairs, for example for different geographic regions. Each pair is then linked by a co-inertia analysis, leading to a cross-table.

The series of cross-tables obtained in this way is then analyzed by a PTA, concluding the STATICO analysis. The compromise of this PTA is a cross-table equal to a weighted mean of all the cross-tables. The analysis of this compromise gives an ordination of the species and an ordination of the environmental variables representing the stable part of the species-environment relationships across the series. The species and environmental variables of each pair of tables can then be projected as additional elements on the PCA of this compromise, showing how the common structure is taken into account by each pair. Moreover, the rows of all the tables can also be projected on the compromise,

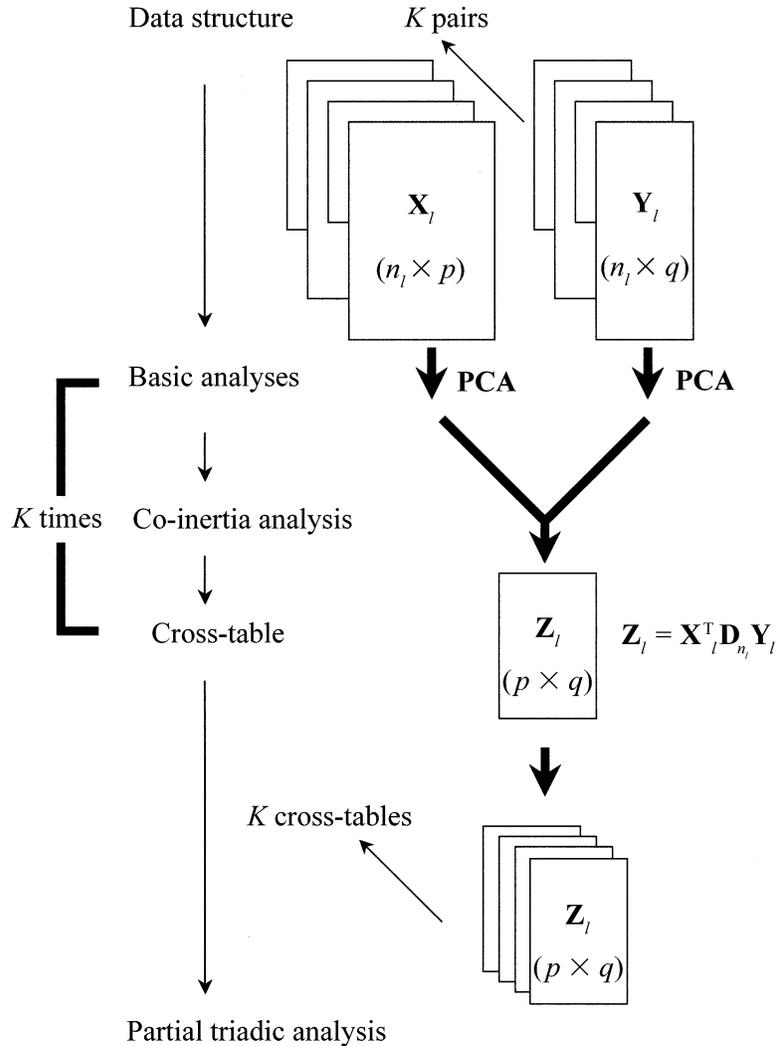


FIG. 1. STATICO flow chart. The data structure is a sequence of pairs of ecological tables. Basic analyses (here PCA) are performed on each table. Co-inertia analyses allow linkage of the pairs of PCA, producing a sequence of cross-tables. PTA is finally used to analyze this sequence.

showing the ordinations of the sites and their relationships across the series of tables.

APPLICATION EXAMPLE

The example data set was extracted from Pegaz-Maucet (1980), and was also used by Blanc et al. (1998). The species data consist of 13 species in six sites, measured four times, and the environmental data consist of 10 variables measured in the same sites and at the same dates. So in this case, the sites are the same at all dates, but this is not a constraint of the STATICO method: only the species and the environmental variables must be the same at all dates.

The species data are the abundance of 13 Ephemeroptera taxa and the environmental data are the values of 10 physicochemical variables measured four times a year (spring, summer, autumn, and winter) in

samples coming from six sites ordered along a small stream in eastern France, the Méaudret (Fig. 2). The series is therefore made of four pairs of tables. Each pair is made of one environmental variables table with six rows and 10 columns (Table 1) and one species composition table with six rows and 13 columns (Table 2).

Autrans, France, is a mountain resort, with tourists coming both in summer and winter. However, the stream flow is much lower in summer and autumn, and pollutants are therefore much more concentrated. Site 1 is upstream from the village, so it is not polluted. Site 2 is just below and receives the village sewage, so the pollution is maximum in this site. Recovery (the process of natural decontamination of the stream) takes place along Sites 3, 4, and 5, and Site 6 is a control upstream on the Bourne River, into which the Méaudret

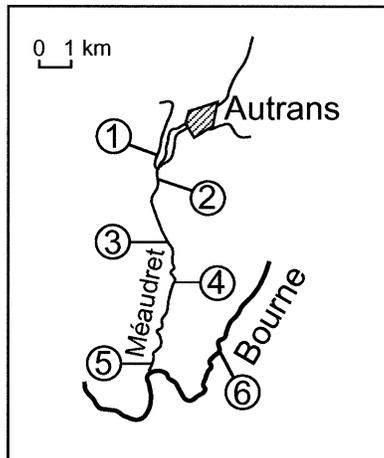


Fig. 2. Geographical map showing the six sampling sites. Autrans, France, is a mountain resort (45°10' N, 5°33' E; elevation = 1250 m). Site 1 is situated just upstream from the Autrans sewer system output, while Site 2 is just below. Sites 3, 4, and 5 are situated downstream along the Méaudret, and Site 6 is upstream on the Bourne, into which the Méaudret flows.

flows. The details of sampling methods can be found in Pegaz-Maucet (1980).

The aim of the study is to analyze the stability of the relationships between the physicochemical composition of water samples, indicating the pollution by the Autrans village, and the species composition found at the corresponding sites. Only the results of the interstructure, compromise, and trajectories steps of the STATICO method will be given here. The separate analyses of each table (normed PCA of the environmental variables and centered PCA of the species data) and the PCA-PCA co-inertia analysis of each pair of ecological tables is not presented here, but a more complete presentation is given in Appendix C.

Interstructure

Fig. 3 shows the factor map of the interstructure. The third sampling date (autumn), with the longest arrow, is clearly more important than the three others, meaning that the compromise will be more influenced by this date. Spring and winter have short arrows, which means that the corresponding tables are less structured and that their importance in the compromise will be lower.

Compromise

Fig. 4 shows the factor maps of the PCA of the compromise, for the physicochemical variables (Fig.

TABLE 1. The 10 physicochemical variables measured four times (spring, summer, autumn, and winter) at six sampling sites along the stream.

Season and site	Temp. (°C)	Flow (m ³ /s)	pH	Conduct. (µmhos/cm)	Oxygen (mg/L)	BDO5 (mg/L)	Oxydab. (mg/L)	Ammo. (mg/L)	Nitrates (mg/L)	Phosph. (mg/L)
Spring										
Site 1	10	41	8.5	295	110	2.3	1.4	0.12	3.4	0.11
Site 2	11	158	8.3	315	13	7.6	3.3	2.85	2.7	1.5
Site 3	11	198	8.5	290	113	3.3	1.5	0.4	4	0.1
Site 4	12	280	8.6	290	126	3.5	1.5	0.45	4	0.73
Site 5	13	322	8.5	285	117	3.6	1.6	0.48	4.6	0.84
Site 6	11	303	8.5	245	100	1.7	0.9	0.05	2.7	0.16
Summer										
Site 1	13	62	8.3	325	95	2.3	1.8	0.11	3	0.13
Site 2	13	80	7.6	380	20	21	5.7	9.8	0.8	3.65
Site 3	15	100	7.8	385	46	15	2.5	7.9	7.7	4.5
Site 4	16	140	8	360	76	12	2.6	4.9	8.4	3.45
Site 5	15	160	8.4	345	91	1.7	1.9	0.22	10	1.74
Site 6	13	310	8.2	285	82	8.5	1.6	0.59	3.7	0.6
Autumn										
Site 1	1	25	8.4	315	91	1.6	0.5	0.07	6.4	0.03
Site 2	3	63	8	425	38	36	8	12.5	2.2	6.5
Site 3	2	79	8.1	350	84	7.1	1.9	2.7	13.2	3.7
Site 4	3	85	8.3	330	106	2	1.4	0.42	12	1.6
Site 5	2	72	8.6	305	91	1.6	0.9	0.1	9.5	1.25
Site 6	4	181	8.6	270	105	2.8	0.5	0.1	3.66	0.43
Winter										
Site 1	3	118	8	325	100	1.6	1.2	0.17	1.8	0.19
Site 2	3	252	8.3	360	100	9.5	2.9	2.52	4.6	1.6
Site 3	3	315	8.3	370	100	8.7	2.8	2.8	4.8	2.85
Site 4	3	498	8.3	330	100	4.8	1.6	1.04	4.4	0.82
Site 5	2	390	8.2	330	100	1.7	1.2	0.56	5	0.6
Site 6	3	480	8.2	290	100	1.3	0.8	0.04	2.2	0.13

Note: Table heading abbreviations: Temp. = temperature, Conduct. = conductivity, BDO5 = biological demand for oxygen, Oxydab. = oxydability, Ammo. = ammonium, Phosph. = phosphates.

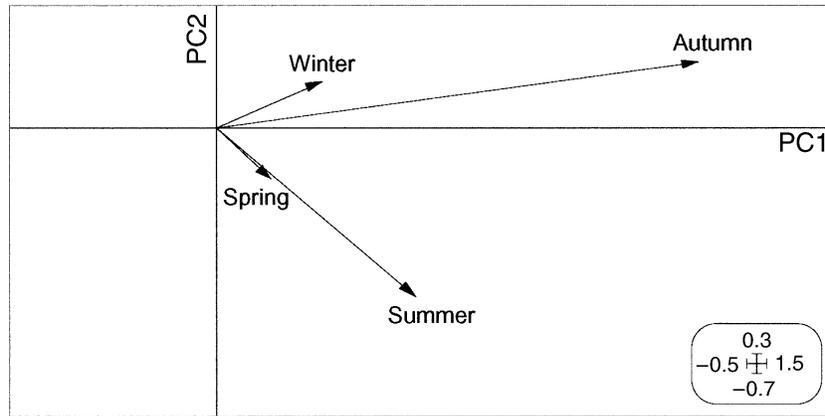


FIG. 3. Interstructure factor map of the STATICO analysis on the Méaudret data. This map shows the importance of each sampling date in the compromise. PC1 = first principal component; PC2 = second principal component. The scale of the graph is given in the rounded box.

4A) and for the Ephemeroptera species (Fig. 4B). The factor map of the physicochemical variables (Fig. 4A) opposes on the first axis the indicators of pollution (on the right: conductivity, ammonium, phosphorus, oxidability, and biological demand for oxygen [BDO5]) to the oxygen concentration and the pH (on the left). This first axis is therefore a pollution gradient, with

the highest pollution on the right. On the second axis, nitrates are isolated in the upper part of the graphic, but temperature and flow are in the same direction. Nitrates are an intermediate step in the ammonium–nitrates–mineral nitrogen sequence, and they are characteristic of the recovery process of the stream. Temperature and flow are also increasing downstream. The

TABLE 2. Abundance of 13 Ephemeroptera species in six sites along the stream at four times.

Season and site	Eda	Bsp	Brh	Bni	Bpu	Cen	Ecd	Rhi	Hla	Hab	Par	Cae	Eig
Spring													
Site 1	4	7	10	9	0	0	0	5	9	0	4	0	0
Site 2	0	0	8	0	0	0	0	0	4	0	0	0	0
Site 3	0	5	5	0	0	0	0	2	5	0	0	0	0
Site 4	0	3	6	0	0	0	0	3	6	0	0	0	0
Site 5	0	5	6	0	0	0	5	0	4	0	0	0	4
Site 6	2	7	9	4	0	0	0	5	6	0	2	0	10
Summer													
Site 1	6	7	10	0	10	0	0	2	7	0	0	0	2
Site 2	0	0	9	0	0	0	0	0	0	0	0	0	0
Site 3	0	6	8	0	0	2	0	0	0	0	0	0	0
Site 4	0	7	11	0	0	2	0	0	2	0	0	5	5
Site 5	0	6	9	2	3	0	4	0	0	0	0	2	7
Site 6	0	8	9	3	4	0	2	4	3	0	0	0	8
Autumn													
Site 1	4	5	8	0	9	6	0	5	9	0	7	0	0
Site 2	0	0	1	0	0	0	0	0	0	0	0	0	0
Site 3	0	9	10	0	0	0	0	0	4	0	3	0	0
Site 4	0	10	13	0	0	3	0	5	5	1	4	2	4
Site 5	2	10	12	0	4	0	8	4	4	2	5	1	6
Site 6	0	12	13	0	8	3	6	9	7	0	0	0	5
Winter													
Site 1	3	6	7	0	6	7	0	4	8	0	4	0	0
Site 2	0	3	6	0	0	5	0	4	3	0	1	0	0
Site 3	0	0	3	0	0	1	0	1	0	0	0	0	0
Site 4	0	6	10	0	0	5	1	3	5	0	2	0	0
Site 5	1	9	11	0	3	6	8	3	5	2	5	0	0
Site 6	0	9	12	0	7	4	5	8	6	0	1	0	0

Note: The species codes are as follows: Eda = *Ephemera danica*, Bsp = *Baetis* sp., Brh = *Baetis rhodani*, Bni = *Baetis niger*, Bpu = *Baetis pumilus*, Cen = *Centroptilum*, Ecd = *Ecdyonurus*, Rhi = *Rhithrogena*, Hla = *Habrophlebia lauta*, Hab = *Habroloetoides modesta*, Par = *Paraletophlebia*, Cae = *Caenis*, Eig = *Ephemerella ignita*.

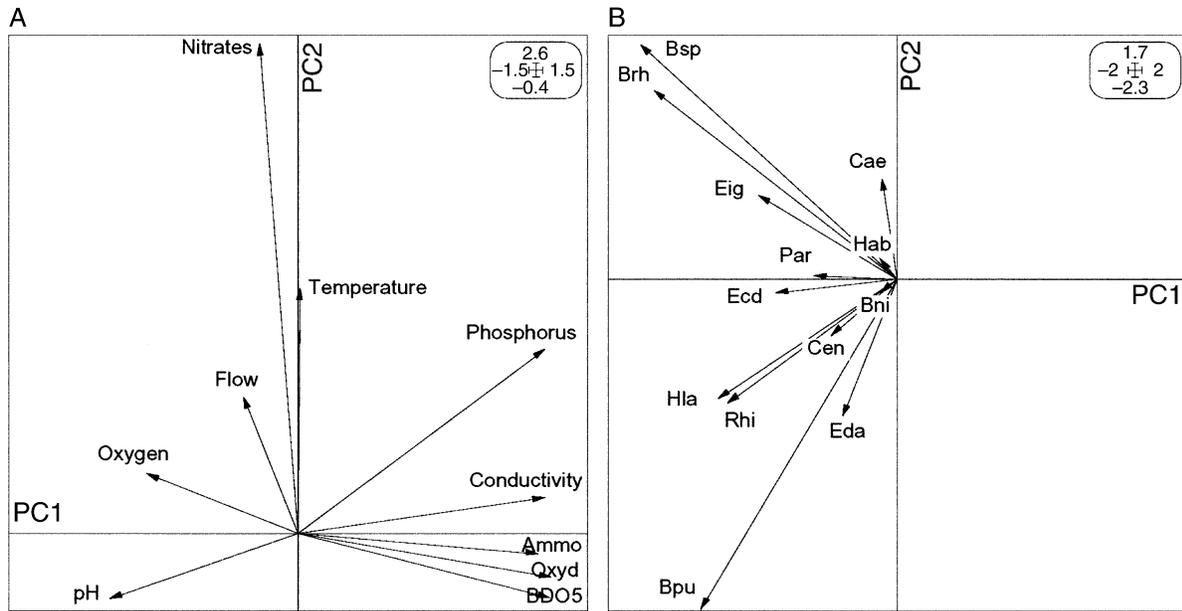


FIG. 4. Compromise factor maps: (A) physicochemical variables; (B) Ephemeroptera species. This map shows the stable part of the species–environment relationships. See Table 2 for species names. PC1 = first principal component; PC2 = second principal component. The scale of the graph is given in the rounded box.

second axis is therefore an upstream–downstream axis, correlated with water quality recovery.

Fig. 4B shows a size effect, with all Ephemeroptera species on the left. This is a characteristic effect obtained when using PCA on species data, and it allows the effect of the pollution on species abundance to be clearly seen. Species with long arrows are more abundant, and species with short arrows are less abundant. Species in the upper left part of the graphic (Bsp and Brh) are characteristic of places where recovery has taken place, while species in the lower left part of the graphic (Bpu) are characteristic of non-polluted sites.

So the stable part of the species–environment dynamics extracted by this compromise is as follows: (1) on the first factor, the pollution gradient, linked to the overall toxicity of pollution on Ephemeroptera species, and (2) on the second factor, the upstream–downstream recovery process, linked to the opposition between species characteristic of non-polluted sites (Bpu in Sites 1 and 6) and species present in intermediate sites (Bsp and Brh in all sites except Site 2).

The next step of the STATICO analysis is a detailed description of each date, obtained by projecting the rows and columns of the initial tables in the compromise analysis.

Trajectories

Fig. 5 shows the projection of the six sites at the four dates on the factor map of the PCA of the compromise. The graphic has been split according to seasons. Each site is represented by two points: one is the projection of the row of the physicochemical table (ar-

row tip), and the other is the projection of the row of the species table (circle). The structure is the strongest in autumn, when the pollution is maximum. The species composition and the water physical chemistry are directly linked to the pollution–recovery gradients that were evidenced in the compromise analysis. Site 1 is unharmed because it is upstream from Autrans. Site 2 is just below and receives the sewage, so the pollution is maximum. Recovery occurs along Sites 3, 4, and 5, and pollution is absent from Site 6, which is a control on another stream. In spring the pollution has not yet begun, and the structures are weak. In summer, the pollution has not achieved its maximum, because the recovery process limits the pollution. In autumn, the structure is very strong because the stream flow is the lowest, natural recovery capacities of the stream have been exceeded, and the concentration of pollutants is maximum. In winter, the pollution has decreased because summer tourists have left, and the stream flow is much higher. This graphic very clearly shows the dynamics of species–environment relationships. The species follow the environmental modifications, with a time lag in winter and summer.

It is also possible to project the columns of the tables (i.e., the physicochemical variables and the Ephemeroptera species at each date on the compromise PCA factor map). Fig. 6 shows the projection of the ten physicochemical variables, and Fig. 7 shows the projection of the 13 Ephemeroptera species. Fig. 6 clearly confirms that the physicochemical structures are the strongest in autumn, and that the pollution–recovery process reaches a maximum during this season. On the

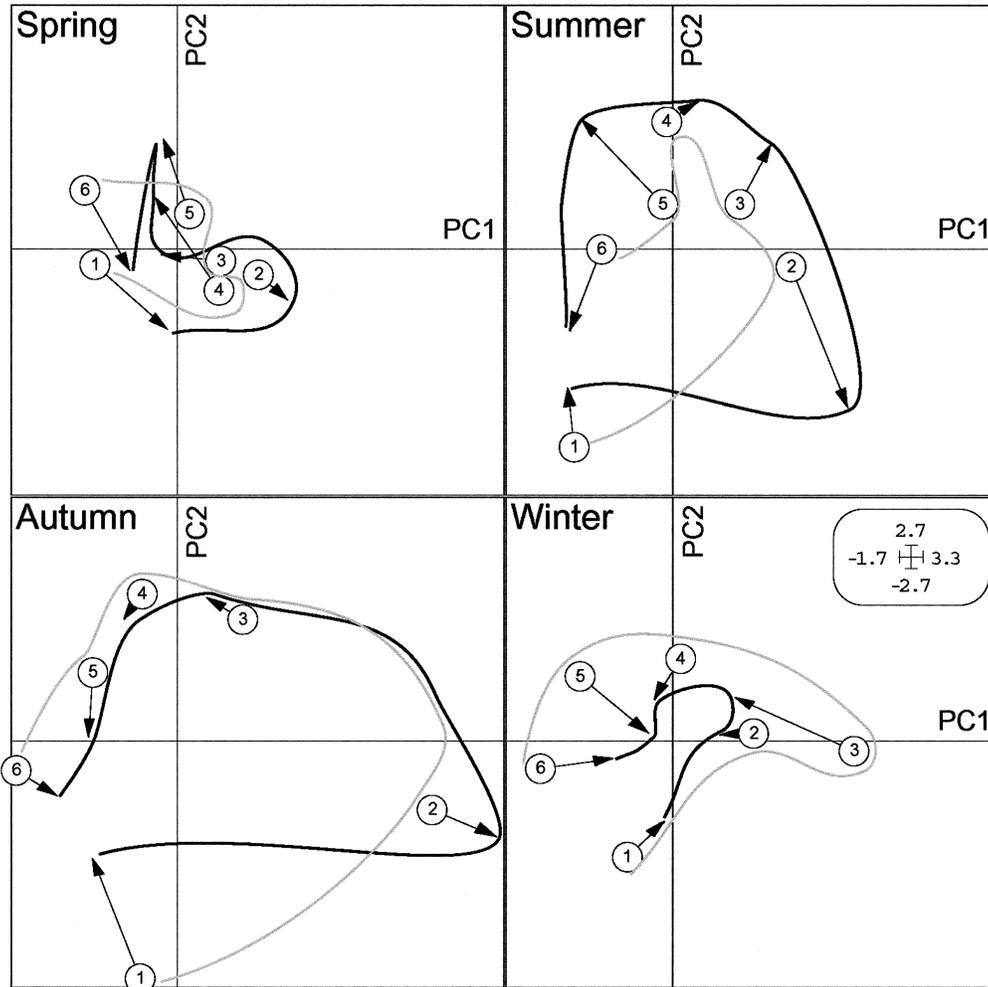


FIG. 5. Trajectories factor map: projection of the coordinates of the six sampling sites on the PCA factor map of the compromise. The black lines correspond to the trajectories of the rows of the physicochemical variables table (arrow tips), and the grey lines correspond to the trajectories of the rows of the Ephemeroptera species table (circles). PC1 = first principal component; PC2 = second principal component. The scale of the graph is given in the rounded box.

contrary, the physicochemical structures are very low in spring and winter. Fig. 7 confirms the modifications of the species distributions during the year. For example, Brh (*Baetis rhodani*) and Bsp (*Baetis* sp.) are abundant in autumn, but they are found only downstream to Site 2, where pollution has already decreased. On the contrary, Bpu (*Baetis pumilus*) is mostly present upstream to Site 2 (in Site 1), or in Site 6 where pollution is absent. These three species have long arrows pointing to the left because they are abundant in all sites, except in Site 2 due to pollution.

DISCUSSION

This small example shows that the STATICO method is an efficient tool to analyze sequences of paired ecological tables. Its flexibility allows it to be adapted to very different experimental conditions. This flexibility comes in part from the flexibility of co-inertia analysis,

which can be used with any kind of pairs of ecological tables. Many other forms of STATICO could be developed, based for example on other types of co-inertia analyses. It is important to note that all the co-inertia analyses performed on the pairs of tables need not to be the same: one can mix PCA-PCA, PCA-CA, MCA-CA, or any other form of co-inertia analyses, according to the aims of the study. The only constraints are that all the cross-tables must have the same rows and the same columns (i.e., that the species and environmental variables must be the same in all the pairs of tables). These constraints come from partial triadic analysis (PTA), and one of them (but not both) can be removed by using STATIS instead of PTA. In this case, the interpretation of the removed constraint is lost. If environmental variables vary among tables, the analysis of the compromise will not describe them, because the compromise will be a square matrix of scalar products

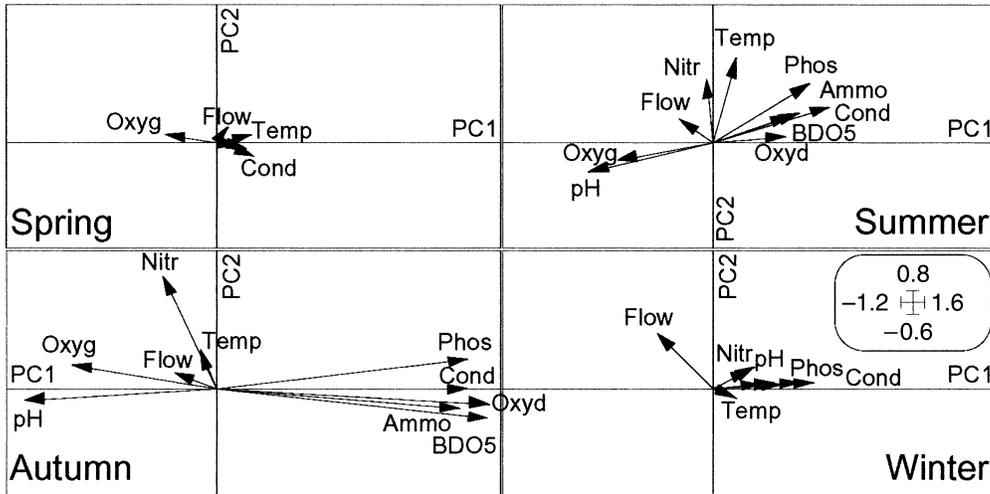


FIG. 6. Trajectories factor map: projection of coordinates of the 10 physicochemical variables on the PCA factor map of the compromise. The four seasons have been plotted on four separate graphics. In spring and winter, some of the labels of the variables near the origin were superimposed and could not be read. Since these variables have no importance in this graphic, they have been removed. PC1 = first principal component; PC2 = second principal component. The scale of the graph is given in the rounded box.

among species. Conversely, if species vary among tables, the analysis of the compromise will not describe them, because the compromise will be a square matrix of scalar products among environmental variables. Keeping both constraints allows PTA to compute a compromise that is a species \times environmental variables cross-table, leading to the simultaneous description of species and environmental variables.

STATICO could also be based on a CCA and/or a RDA on each pair of tables. However, methodological problems are raised by the regression step involved in these methods: if the number of samples is low, CCA

tends toward a simple CA, and the role of environmental variables may be underestimated. See Dray et al. (2003) for a study of this problem, and comparisons between CCA and co-inertia analysis based on simulated data. This problem can be avoided using CCA-PLS (partial least squares CCA; ter Braak and Verdonschot 1995), but with a substantial increase of complexity. Moreover, using these methods is an implicit choice of the unimodal vs. linear species response curve model that may not be justified by the aims of the study, or not be supported by the data. Using MCA, NSCA (non-symmetric correspondence analysis, Gi-

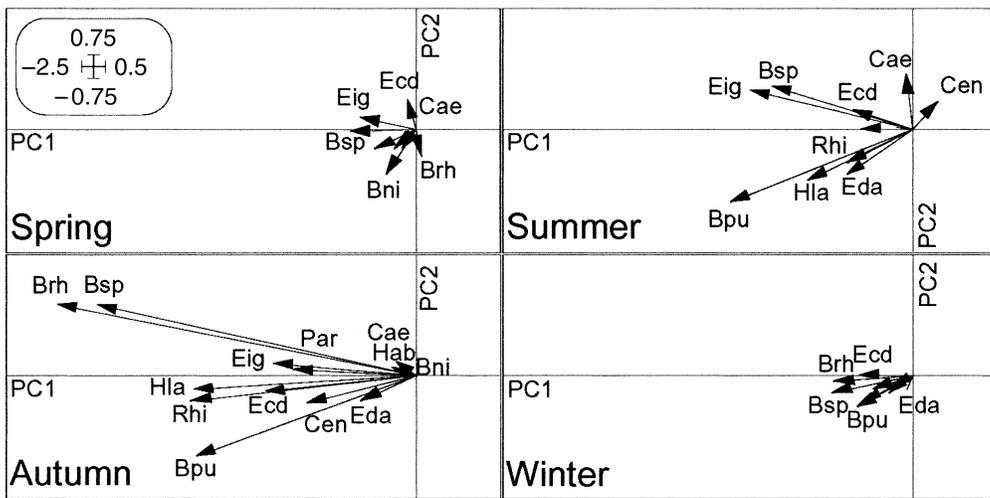


FIG. 7. Trajectories factor map: projection of the coordinates of the 13 Ephemeroptera species on the PCA factor map of the compromise. The four seasons have been plotted on four separate graphics. As in Fig. 6, some of the species codes near the origin were removed to improve the presentation. PC1 = first principal component; PC2 = second principal component. The scale of the graph is given in the rounded box.

maret-Carpentier et al. 1998), FCA (fuzzy correspondence analysis, Chevenet et al. 1994), PCA variants, or other basic analyses linked by a co-inertia analysis allows researchers to take into account more biological information, to use methods more adapted to the data, and to produce more accurate statistical results.

The comparison between STATICO and ad hoc approaches can be based on two types of considerations: the mathematical properties of the methods, and the biological aims of the study. From a mathematical point of view, STATICO has big advantages. Two examples of ad hoc methods are: (1) performing a CCA or a RDA on the joint tables (i.e., the two tables obtained by pasting the tables of a sequence one below the other), and (2) performing a CCA or a RDA on the average tables and then projecting each initial table in this analysis. The advantage of STATICO is the mathematical optimality theorems on which it is based: no other compromise can be better than the PTA compromise. The ad hoc methods will give results similar to STATICO when the structure of the sequence is strong (i.e., when the stable part is important) and all the tables have almost the same contribution to this structure. But when the structure is less strong, and when some tables have minor or different contributions to this structure, the STATICO compromise will be better, because instead of a simple average, it uses a non-uniform weighting (components of the first eigenvector of the interstructure). In this case, STATICO will downweight the tables with minor contribution or different structure, and the resulting compromise will be nearer to the other tables.

From a more general point of view, using methods precisely adapted to the biological aims of the study instead of using ad hoc methods is preferable. The search for a well adapted method, instead of the immediate use of an easy ad hoc method, will prompt further thinking about the aims of the study, and will allow a better interpretation of the results.

Many graphical displays can be drawn to help interpret the numerical outputs of STATICO, and not all of them have been shown in this paper. We have chosen a very small example, and many of our conclusions could have been drawn from simpler ad hoc methods, like the PCA of joint tables. However, on large data sets with more complex structures, STATICO will allow a much more detailed analysis and will make the study of the stability of species–environment relationships easier.

All the computations and graphical displays were performed using the free ADE-4 multivariate analysis and graphical display software (Thioulouse et al. 1997; available online).⁴

ACKNOWLEDGMENTS

We are very grateful to N. Kenkel, P. Legendre, and an anonymous referee for their help in improving the first version of this paper.

⁴ URL: (<http://pbil.univ-lyon1.fr/ADE-4>)

LITERATURE CITED

- Aliaume, C., C. Monteiro, M. Louis, T. Lam Hoai, and G. Lasserre. 1993. Organisation spatio-temporelle des peuplements ichthyologiques dans deux lagunes côtières: au Portugal et en Guadeloupe. *Oceanologica Acta* **16**:291–301.
- Blanc, L., and D. Beaudou. 1998. Stabilité temporelle des structures spatiales des peuplements piscicoles des régions Languedoc-Roussillon et Provence-Alpes-Côte d'Azur. *Bulletin Français de la Pêche et de la Pisciculture* **348**:23–45.
- Blanc, L., D. Chessel, and S. Dolédec. 1998. Etude de la stabilité temporelle des structures spatiales par analyse d'une série de tableaux faunistiques totalement appariés. *Bulletin Français de la Pêche et de la Pisciculture* **348**:1–21.
- Centofanti, M., D. Chessel, and S. Dolédec. 1989. Stabilité d'une structure spatiale et compromis d'une analyse statistique multi-tableaux: application à la physico-chimie d'un lac réservoir. *Revue des Sciences de l'Eau* **2**:71–93.
- 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.
- Dolédec, S. 1988. Les analyses multi-tableaux en écologie factorielle. II Stratification longitudinale de l'Ardèche à partir des descripteurs physico-chimiques. *Acta Œcologica, Ecologia Generalis* **9**:119–135.
- Dolédec, S., and D. Chessel. 1994. Co-inertia analysis: an alternative method for studying species–environment relationships. *Freshwater Biology* **31**:277–294.
- Dray, S., D. Chessel, and J. Thioulouse. 2003. Co-inertia analysis and the linking of ecological data tables. *Ecology* **84**:3078–3089.
- Escoufier, Y. 1973. Le traitement des variables vectorielles. *Biometrics* **29**:750–760.
- Franquet, E., and D. Chessel. 1994. Approche statistique des composantes spatiales et temporelles de la relation faunemilieu. *C.R. Acad. Sci. Paris, Sciences de la vie* **317**:202–206.
- Franquet, E., S. Dolédec, and D. Chessel. 1995. Using multivariate analyses for separating spatial and temporal effects within species–environment relationships. *Hydrobiologia* **300–301**:425–431.
- Gimaret-Carpentier, C., D. Chessel, and J. P. Pascal. 1998. Non-symmetric correspondence analysis: an alternative for community analysis with species occurrences data. *Plant Ecology* **138**:97–112.
- Gittins, R. 1985. Canonical analysis, a review with applications in ecology. Springer-Verlag, Berlin, Germany.
- Kroonenberg, P. M. 1983. Three-mode principal component analysis. DSWO Press, Leiden, Germany.
- Kroonenberg, P. M. 1989. The analysis of multiple tables in factorial ecology. III. Three-mode principal component analysis: “analyse triadique complète.” *Acta Œcologica, Ecologia Generalis* **10**:245–256.
- Lavit, C. 1988. Analyse conjointe de tableaux quantitatifs. Masson, Paris, France.
- Lavit, C., Y. Escoufier, R. Sabatier, and P. Traissac. 1994. The ACT (Statis method). *Computational Statistics and Data Analysis* **18**:97–119.
- Lebart, L., L. Morineau, and K. M. Warwick. 1984. Multivariate descriptive analysis: correspondence and related techniques for large matrices. John Wiley and Sons, New York, New York, USA.
- Mercier, P., D. Chessel, and S. Dolédec. 1992. Complete correspondence analysis of an ecological profile data table: a central ordination method. *Acta Œcologica* **13**:25–44.
- Montaña, C., and P. Greig-Smith. 1990. Correspondence analysis of species by environmental variable matrices. *Journal of Vegetation Science* **1**:453–460.

- Pegaz-Maucet, D. 1980. Impact d'une perturbation d'origine organique sur la dérive des macro-invertébrés benthiques d'un cours d'eau. Comparaison avec le benthos. Thèse de 3^e cycle, Université Lyon 1, France.
- 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 Phytécologie. *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.
- 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. 1986. Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. *Ecology* **67**:1167–1179.
- Ter Braak, C. J. F., and P. F. M. Verdonschot. 1995. Canonical correspondence analysis and related multivariate methods in aquatic ecology. *Aquatic Sciences* **57**:255–289.
- Thioulouse, J., and D. Chessel. 1987. Les analyses multivariées en écologie factorielle. I De la typologie d'état à la typologie de fonctionnement par l'analyse triadique. *Acta Oecologica, Oecologia Generalis* **8**:463–480.
- 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.
- Tucker, L. R. 1958. An inter-battery method of factor analysis. *Psychometrika* **23**:111–136.
- Tucker, L. R. 1966. Some mathematical notes on three-mode factor analysis. *Psychometrika* **31**:279–311.
- Wollenberg, A. L. 1977. Redundancy analysis, an alternative for canonical analysis. *Psychometrika* **42**:207–219.

APPENDIX A

CO-INERTIA ANALYSIS RATIONALE

Co-inertia analysis is a two-tables analysis. It searches for axes that maximize the covariance between the coordinates of the rows of the two tables. See Dray et al. (2003) and Dolédec and Chessel (1994) for a more complete description of co-inertia analysis.

Let \mathbf{X} ($n \times p$) and \mathbf{Y} ($n \times q$) be the two tables: in an ecological situation, \mathbf{X} may be, for example, an environmental variables table, and \mathbf{Y} a species abundance table. \mathbf{X}^T is the transpose of \mathbf{X} . Let \mathbf{D}_p ($p \times p$) and \mathbf{D}_q ($q \times q$) be the diagonal matrices of column weights. These weights are often equal to one (as in a PCA, for example). An exception is CA, where they are equal to n_j/n , with $n_j = \sum_i n_{ij}$, $n = \sum_j n_j$, n_{ij} being the abundance of species j in sample i . Other examples of analyses where non-uniform column weights are used are FCA (fuzzy correspondence analysis, Chevenet et al. 1994), and NSCA (non-symmetric correspondence analysis, Gimaret-Carpentier et al. 1998).

Let \mathbf{D}_n ($n \times n$) be the diagonal matrix of row weights. These weights are also very often uniform, all equal to $1/n$ in PCA, for example. In CA, they are equal to n_i/n , with $n_i = \sum_j n_{ij}$, $n = \sum_i n_i$. The row weights of the two tables must be equal to do the co-inertia analysis. This means that, for example in a PCA-CA co-inertia analysis, the initial PCA of environmental variables must be computed with the row weights of the species table CA (instead of $1/n$).

The PCA of \mathbf{X} is the eigenanalysis of matrix $\mathbf{C} = \mathbf{X}^T \mathbf{D}_n \mathbf{X} \mathbf{D}_p = 1/n \mathbf{X}^T \mathbf{X}$. If the variables of \mathbf{X} are centered, \mathbf{C} is the covariance matrix, and if they are normed, it is the correlation matrix.

The co-inertia analysis of \mathbf{X} and \mathbf{Y} is the eigenanalysis of matrix \mathbf{W} :

$$\mathbf{W} = \mathbf{X}^T \mathbf{D}_n \mathbf{Y} \mathbf{D}_q \mathbf{Y}^T \mathbf{D}_n \mathbf{X} \mathbf{D}_p.$$

The PCA-PCA co-inertia analysis of \mathbf{X} and \mathbf{Y} (Tucker's inter-battery analysis) is the eigenanalysis of matrix $\mathbf{W}_{\text{PCA-PCA}}$:

$$\mathbf{W}_{\text{PCA-PCA}} = 1/n^2 \mathbf{X}^T \mathbf{Y} \mathbf{Y}^T \mathbf{X}.$$

If the variables of \mathbf{X} and \mathbf{Y} are normed, this is a matrix ($p \times p$) of sums of products of cross-correlations between \mathbf{X} and \mathbf{Y} : $\sum_i \text{cor}(x_i, y_j) \text{cor}(x_i, y_k)$, which is high when both correlations are high. For example, for two environmental variables x_1 and x_2 , this value will be high if the correlations between these variables and the species are high for the same species (i.e., two environmental variables will be linked if they have strong (linear) correlations with the same species).

In the case of a PCA-CA co-inertia analysis, the row weights are not uniform, and matrix \mathbf{W} cannot be simplified. Since the row weights (\mathbf{D}_n) must be the same for \mathbf{X} and \mathbf{Y} , the PCA of \mathbf{X} is done using the row weights of the CA of \mathbf{Y} . For two environmental variables x_1 and x_2 , the values in matrix \mathbf{W} will be high if the position of the optimum of the species response curve is similar on these two variables.

Many other forms of co-inertia analysis exist, with different mathematical properties. They can be obtained by coupling other basic analyses. A recent example is the OMI (outlying mean index) analysis (Dolédec et al. 2000), which uses a PCA on species profiles.

APPENDIX B

PARTIAL TRIADIC ANALYSIS

Partial triadic analysis is an analysis of a sequence of tables. Let \mathbf{X}_k be the k^{th} table of the series. All the tables must have the same number of rows (n) and columns (p). Let \mathbf{D}_n and \mathbf{D}_p be the diagonal matrices of row and column weights.

INTERSTRUCTURE

The interstructure consists of computing a matrix of scalar products between tables (i.e., the vector covariance matrix: $\text{Covv}(\mathbf{X}_k, \mathbf{X}_l) = \text{Trace}(\mathbf{X}_k^T \mathbf{D}_n \mathbf{X}_l \mathbf{D}_p)$). The eigenanalysis of this matrix gives a first eigenvector, in which components α_k are used as weights to compute the compromise. Alternatively, a matrix of vector correlations can be used to rescale

the importance of the tables: $\mathbf{Rv}(\mathbf{X}_k, \mathbf{X}_l) = \text{Covv}(\mathbf{X}_k, \mathbf{X}_l) / \sqrt{\text{Vav}(\mathbf{X}_k) \text{Vav}(\mathbf{X}_l)}$. $\text{Vav}(\mathbf{X}_k)$ is the vector variance of table \mathbf{X}_k : $\text{Vav}(\mathbf{X}_k) = \text{Trace}(\mathbf{X}_k^T \mathbf{D}_n \mathbf{X}_k \mathbf{D}_p)$. It is simply the variance of the vector obtained by putting all the columns of table \mathbf{X}_k one below the other. The vector correlation matrix and the vector covariance matrix are linked by the same relationships as the normal correlation and covariance matrices.

COMPROMISE

The compromise \mathbf{X}_c is a linear combination of the initial tables, weighted by the components of the first eigenvector of the interstructure: $\mathbf{X}_c = \sum_k \alpha_k \mathbf{X}_k$. The analysis of this compromise table (for example by a simple PCA) gives factor

maps that can be used to interpret the structures of this compromise.

TRAJECTORIES

The rows and columns of the initial tables are projected in the PCA of the compromise (i.e., the rows are projected on the principal axes, and the columns are projected on the prin-

cipal components). This step is performed like in a simple PCA: let \mathbf{U} be the matrix of the eigenvectors of the analysis of the compromise. The coordinates of the rows of table \mathbf{X}_k are $\mathbf{R}_k = \mathbf{X}_k \mathbf{D}_p \mathbf{U}$, and the coordinates of its columns are $\mathbf{C}_k = \mathbf{X}_k^T \mathbf{D}_n \mathbf{X}_c \mathbf{D}_p \mathbf{U} \Lambda^{-1/2}$, $\Lambda^{-1/2}$ being the diagonal matrix of the inverses of the square root of the eigenvalues of the analysis of the compromise.

APPENDIX C

STATICO computations are available in ESA's Electronic Data Archive: *Ecological Archives* E085-004-A1.