



## **SIMULTANEOUS ANALYSIS OF SEVERAL PAIRS OF ECOLOGICAL TABLES: THE CO-ATPS METHOD**

**RODNELLIN ONESIME MALOUATA<sup>1</sup>, GABRIEL KISSITA<sup>2</sup>, ROGER  
ARMAND MAKANY<sup>3</sup> and LÉONARD NIÉRÉ<sup>2</sup>**

<sup>1</sup>Faculty of Science and Technology  
University Marien Ngouabi  
Brazzaville-Congo, BP 69

<sup>2</sup>Higher Institute of Management  
University Marien Ngouabi  
Brazzaville-Congo, BP 69

<sup>3</sup>Faculty of Economics  
University Marien Ngouabi  
Brazzaville-Congo, BP 69

### **Abstract**

The problem of link analysis between two vertical multi-tables has been solved for the first time in STATICO method. But the STATICO method has the disadvantage of providing in some cases uninterpretable compromises because of default of Partial Triadic Analysis that gives weighting coefficients that are not all of the same sign in some cases. In this article, we propose two methods called CO-ATPS and ATPSCO. The first solves the same question about the relationship between two vertical multi-tables or two cubes and the second only in the case of two cubes. The CO-ATPS method is applicable even if the weighting coefficients are not all the same sign. To show the interest of our approaches, we use a set of environmental data already processed by some authors.

Received December 27, 2013

2010 Mathematics Subject Classification: 62H20.

Keywords and phrases: partial triadic analysis, STATIS, co-inertia analysis, STATICO, CO-ATPS.

© 2013 Pioneer Scientific Publisher

## 1. Introduction

In ecology, the study of the relationship between living beings and their environment is a central problem. He is interested in the relationship between the two basic compartments of the ecosystem, biocenosis (living beings) and biotope (environment).

This study considers the ecosystem as a functional entity where the physical and biological processes are constantly interacting, the importance of what is widely emphasized in the ecological literature (see Crocker ([7], Margalef [18]). To investigate this relationship, the sampling campaign usually leads to the acquisition of a set of data collected on a set of stations (sites, statements) and is presented in the form of two partitioned tables or not partitioned tables called floro-faunistical tables (measures of taxa) and mesological table (environment, region).

Factor analysis of these two data tables on the same individuals allow to get two distinct typologies of these individuals. This is the Co-inertia analysis proposed by Chessel and Mercier [5] and Dolédec and Chessel [8] which made possible the coupling of two tables and obtaining these two typologies. This notion of coupling is based on the duality diagram proposed by Escoufier [10]. The Co-inertia Analysis covers the Inter-Battery Analysis of Tucker [26], the Canonical Analysis of Hotelling [13] and the Canonical Analysis of qualitative variables Cazes [4]. Its purpose is to seek Co-inertia axes maximizing the covariance between the coordinates of the projections of the rows of each table. Or, if we have two tables of data on the same individuals which are respectively measured on two quantitative variables, the Co-inertia Analysis between these two tables, is equivalent to determine a number of Co-inertia axes, each associated with the first group of variables and the other in the second group of variables maximizing the squared covariance between synthetic components of the group 1 and group 2 subject to the constraints of standard unit on the Co-inertia axes associated with each component of each table. The Co-inertia axes are obtained by diagonalization of a matrix, these axes can also be obtained by the Principal Component Analysis (PCA) of the cross table or Co-inertia table between the two groups of variables.

The Co-inertia Analysis was extended to several tables in Multiple Co-inertia Analysis (MCOIA) by Chessel and Hanafi [6] and Hanafi and Kiers [12]. Recently, some methods based on two multi-tables with the same number of individuals have been proposed by Vivien and Sabatier (DO-ACT [28]), by Vivien (ACIMOG [27]) and Kissita et al. (CONCORG and CONCORG M [14]).

But when the partition is made at the individuals level, the variables of the tables being the same, we find our-selves in the context of the method *STATICO* (combination of *STATIS* method of L'Hermier des Plantes [17] and Co-inertia Analysis of Chessel and Mercier [5]) proposed by Simier et al. [24]). The two partitioned tables become respectively multi-tables associated with group 1 and group 2. The *STATICO* method proposed by Simier et al. [24]) is a *STATIS* on original tables, or a Partial Triadic Analysis (PTA) of cross-covariance tables between subtables of each pair of variables. Recently, in the research of the link between two cubes, Thioulouse [25] proposed the *COSTATIS* method allowing to one hand to determine from both Partial Triadic Analysis two compromises of each cube and secondly to make Co-inertia Analysis of these compromises. But *STATICO* and *COSTATIS* methods bear the default of the Partial Triadic Analysis (PTA) providing weighting coefficients in some cases which are not all of the same sign, which leads the user to adopt another method. As alternatives to the Partial Triadic Analysis (PTA), leading to produce weighting coefficients having of the same sign, for example, we can be made two recent approaches *sPTA* (successive Partial Triadic Analysis) and *SPTA* (Simultaneous Partial Triadic Analysis), which were respectively designed by Mizere et al. [19] and Nguessolta et al. [20]. In the search of the relationship between two multi-tables, other methods have been proposed by Kissita et al. [15].

It is discussed in this article to propose two methods: one in the same context as that of *STATICO* and the other in the same context as that of *COSTATIS*. These new methods are called *CO-ATPS* and *ATPSCO*. They are based on criteria, that is to say, we proceed respectively like *STATICO* and *COSTATIS*, instead of the Partial Triadic Analysis (PTA) in *STATICO* and *COSTATIS*, we substitute respectively by the Simultaneous Partial Triadic Analysis (PTA) in *CO-ATPS* and in *ATPSCO*. This way is interesting for the *CO-ATPS* in so far as we can make the Principal Component Analysis (PCA) of compromise table if the weighting coefficients are of the same sign and on the other hand another Simultaneous Partial Triadic Analysis (*SPTA*) in the space of variables if the weighting coefficients are not of the same sign. It is also possible to use the successive Partial Triadic Analysis (*sPTA*) instead of Simultaneous Partial Triadic Analysis (*SPTA*).

The paper is organized as follows:

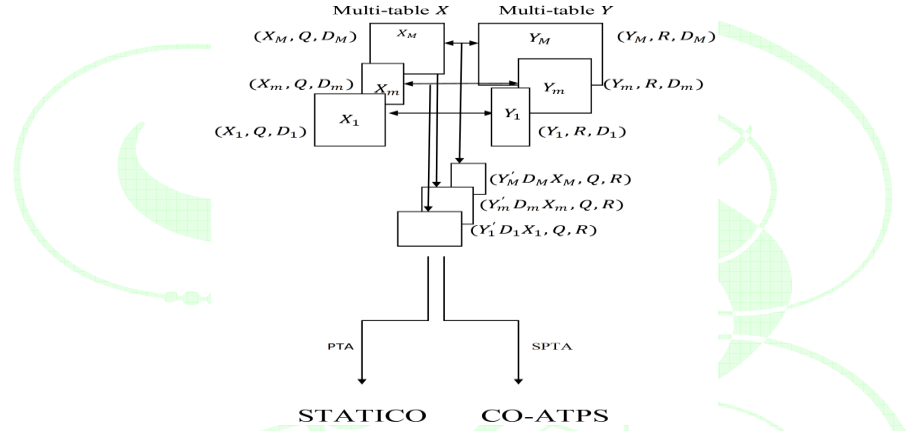
In Section 2, before proposing *CO-ATPS* and *ATPSCO*, we recall the used material and the methodological basis of these approaches. In Section 3, a discussion and conclusion are proposed.

## 2. Materials and Methods

In this section, we will first define the data and then we will process them by CO-ATPS and ATPSCO methods.

### 2.1. Materials

#### 2.1.1. Context of the method



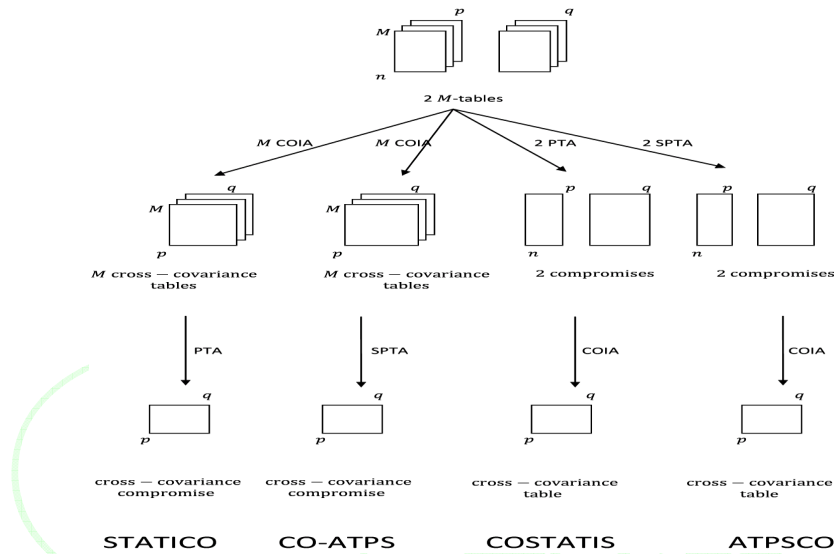
**Figure 1.** Comparison of two systems in the case of a non-comprehensive sampling of all individuals at all dates of measurement.

The context of the method is the method STATICO. Being given  $M$  couples of triplets of tables  $\{(X_m, Q, D_m), (Y_m, R, D_m)\}$  for  $m = 1, \dots, M$ ,  $X_m$  is a table of dimensions  $(n_m, p)$  whose rows are observations of  $n_m$  individuals and columns are measures of  $p$  variables,  $Y_m$  is a table of dimensions  $(n_m, q)$  with the same  $n_m$  individuals as table  $X_m$ , with  $q$  the number of variables, where  $n = \sum_{m=1}^m n_m$ .

We denote by  $D_m$  diagonal metric weight of  $n_m$  individuals defined in  $\mathbb{R}^{n_m}$  for all  $m = 1, \dots, M$ ,  $Q$  and  $R$  are respectively metrics defined in  $\mathbb{R}^p$  and  $\mathbb{R}^q$ .

One study to another the list of variables is constant while the list of individuals may vary (Figure 1).

In the case of a complete sample of all individuals in all studies, we obtain a couple of datacube (Figure 2).



**Figure 2.** Comparison of four systems. STATICO is a Partial Triadic Analysis on series of couples of product of tables obtained by coupling two tables a couple at each date. CO-ATPS is a Simultaneous Partial Triadic Analysis on series of couples of tables product obtained by coupling two tables a couple at each date. COSTATIS two Partial Triadic Analyses are used to calculate the compromises of two  $M$  tables. ATPSCO two Simultaneous Partial Triadic Analyses are used to calculate the compromises of two  $M$  tables.

### 2.1.2. Example data

The data are analyzed in this example have also been treated by Pegaz-Maucet [21], by Hanafi [11], Blanc et al. [1], Thioulouse et al. [23], Thioulouse [25] and recently by Kissita et al. [15]. These environmental data were measured on Méaudret river of France. Six (6) stations distributed in Méaudret river were each visited once per season (1-Spring, 2-Summer, 3-Autumn, 4-Winter). Each time were measured at each station 10 physico-chemical parameters of the river (Temp = Temperature, Flow, pH, Cond = conductivity, oxyg = oxygen, BOD5 = Biological demand for oxygen, Oxyd = oxydability, Ammo = ammonium, Nitra = Nitrates, Phosp = phosphates) and identified 13 present species (ephemeroptera) (Eda = Ephemera danica, Bsp = Baetis sp, Brh = Baetis Rhodani, Bni = Baetis niger, Bpu = Baetis pumilus, Cen = Centropilum, Ecd = Ecdyonurus, Rhi = Rhithrogena Hla = Habrophlebialauta, Hab = Habroletoides modesta, Par = Paraletophlebia, Caе = Caenis, Eig = Ephemera ignita).

We obtain a total of 24 statements (6 stations  $\times$  4 seasons). The faunistical table  $X$  of dimension (24, 13) is made with four tables of  $X_m$  of dimension (6, 13). Similarly, mesological table  $Y$  of dimension (24, 10) consists four tables  $Y_m$  of dimension (6, 10). The two tables  $X_m$  (species) and  $Y_m$  (environment) correspond to the same season  $m$  ( $m = 1, 2, 3, 4$ ). The objective of this study is to describe the evolution of species-environment relationship from one season to another. This is to describe the evolution of the species-environment relationship of a season to another. We do not make the interpretation a priority, even if the elements in this direction are discussed.

The purpose of this example is to analyze these data by the methods CO-ATPS and ATPSCO.

## 2.2. Methods

We will first give some methodological bases allowing to propose CO-ATPS and ATPSCO methods.

### 2.2.1. Partial triadic analysis (PTA)

Partial Triadic Analysis (PTA) is a method of data analysis that allows to compare several tables having in common individuals (rows) and variables (columns). It was introduced by Thioulouse and Chessel [22] in Ecology and taken again by Blanc [2].

Thus, we consider the triplet  $(X_m, Q, D)$  totally matched in rows and columns. Tables  $X_m$  are centered and possibly reduced STATIS on tables is also called Partial Triadic Analysis (PTA). The difference with STATIS and dual STATIS (L'Hermier des Plantes [17]) is that it works directly on the original tables  $X_m$  without using operators  $W_m D$ . Partial Triadic Analysis (PTA) belongs to the family of STATIS methods. This method is based on the concepts defined by Escoufier [9], vector variance of a table  $Van(X_m)$ , vector covariance between tables  $Covv(X_m, X_k)$  and vector correlation between tables  $Rv(X_m, X_k)$ .

The analysis takes place in three stages, corresponding to the steps defined by Lavit et al. [16]: interstructure, compromise and intrastructure. The interstructure allow to compare tables between them, it consist to diagonalize the matrix  $S$  of size  $M \times M$  whose the general term  $S_{mk}$  is given by  $tr(X'_m D X_k Q)$ .

The first  $a_1$  own standard unit vector is then used to construct the compromise table  $X_c$ . The coefficients  $\alpha_{m,1}$ , which are the components of the vector  $a_1$  are not all of the same sign in some cases, contrary to STATIS method. This raises problems of interpretation and leads the user to adopt a different strategy.

The step of the compromise of PTA is based on finding a compromise matrix

$$X_c^{(1)} = \sum_{m=1}^M \alpha_{m,1} X_m$$

which expresses the common structure with  $M$  tables. The coefficients  $\alpha_{m,1}$  are the components of the vector  $a_1$  solution of the maximizing  $\|X_c\|_{HS}^2 = tr(X_c' Q D X_c)$  subject to the constraint of unit norm of the vector  $a$  ( $a'a = 1$ ).

The intrastucture provides principal axes of the compromise  $X_c^{(1)}$  on which project the rows, and the principal components of which project the columns. It is also used to trace the trajectories that allow to project each individual of each table as supplementary individual on the principal axes of the compromise, that is to say, to calculate the components  $\psi_m^{(1)} = X_m Q \omega_1$ , ( $m = 1, \dots, M$ ) with  $\omega_1$  the eigenvector  $Q$ -normed order 1 of the matrix  $V_{X_c^{(1)}} Q = X_c^{(1)'} D X_c^{(1)} Q$ . It is the same for the columns of the table  $X_m$  that are also considered supplementary variables. The coordinates of the supplementary variables are housed in the vector  $\phi_m^{(1)} = X_m' D v_1$  with  $v_1$  the eigenvector  $D$ -normed operator  $W_{X_c^{(1)}} D$ .

Partial Triadic Analysis does not guarantee mathematically that the coefficients  $\alpha_m$  associated with each table are all of the same sign. If this is not the case, it should be wary seriously: the existence of a compromise is not established.

This led initially to the sPTA proposed Mizere et al. [19].

### 2.2.2. Simultaneous partial triadic analysis (SPTA)

When Partial Triadic Analysis is replaced by the simultaneous Partial Triadic Analysis (SPTA) which the methodological bases are recalled in the following manner: one searches the links between subtables  $X_m$  ( $m = 1, \dots, M$ ) that are fully

matched to the same  $n$  individuals and to the same  $p$  variable. Partial Triadic Analysis (SPTA) of Nguessolta et al. [20] is a simultaneous method in determining the solution, unlike Successive Partial Triadic Analysis (sPTA) of Mizere et al. [19]. These methods aim to build compromises like the Partial Triadic Analysis (PTA) of Thioulouse and Chessel [22] that express the common structure with  $M$  tables  $X_m$ .

More specifically, one seeks the compromise

$$X_c = \sum_{m=1}^M \alpha_m X_m$$

such that the function

$$\begin{aligned} f(U, a) &= tr \left[ \left( \sum_{m=1}^M \alpha_m X_m Q U \right)' D \left( \sum_{m=1}^M \alpha_m X_m U \right) \right] \\ &= tr \left[ U' Q \left( \sum_{m=1}^M \alpha_m X_m \right)' D \left( \sum_{m=1}^M \alpha_m X_m \right) Q U \right] \\ &= tr(U' Q X_c' D X_c Q U) \end{aligned}$$

subject to the constraints  $U' Q U = I_r$  and  $a' a = 1$ , where  $a = [\alpha_1, \dots, \alpha_M]'$  is the vector containing the weights of tables  $X_m$  and  $X_c = \sum_{m=1}^M \alpha_m X_m$  the compromise of tables  $X_m$ .

Solution of this problem is given by an algorithm that converges to a local solution in which the summary is given by the following points:

- (1) We randomly select the matrice  $U_0$  such as  $U_0' Q U_0 = I_r$  and the normalized vector  $a_0$  such as  $a_0' a_0 = 1$ , and  $\varepsilon$  (e. g., 0.00001)
- (2) Compute:  $E_{U_0 a_0}$
- (3) Perform singular value decomposition of  $E_{U_0 a_0} U_0 = L \Delta G'$
- (4) Set for Update  $U_0$  the matrix  $Z = L G'$



(5) Compute the matrix:

$$S_Z = [\text{vec}(D^{1/2}X_1QZ) | \text{vec}(D^{1/2}X_2QZ) | \dots | \text{vec}(D^{1/2}X_MQZ) |].$$

(6) Compute:  $R_Z = S_Z'S_Z$ .

(7) Determine the eigenvalues and eigenvectors of  $R_Z$ .

(8) Compute for updates of  $a_0$  the first eigenvector  $a^*$  of  $R_Z$  associated with the largest eigenvalue.

(9) If  $f(Z, a^*) - f(U_0, a_0) \geq \epsilon$ , then  $U_0 = Z$  and  $a_0 = a^*$  go to 2, otherwise the algorithm has converged.

(10) One computes projected inertia  $\rho_{m,k} = \text{var}(X_mQz_k)$ , where  $z_k$  ( $k = 1, \dots, r$ ) is the  $k$ th column vector of the matrix  $Z$ .

(11) One computes the compromise

$$X_c^* = \sum_{m=1}^M \alpha_m^* X_m,$$

where  $a^* = [\alpha_1^*, \dots, \alpha_M^*]'$  is the vector containing the weights that leads to the convergence of the algorithm.

If the weighting coefficients  $\alpha_m^*$  of tables  $X_m$  are all of the same sign, which is the advantage in the construction of the compromise, we make the PCA of compromise. On the other hand, if are not of the same sign, the compromise are a linear combination of matrices  $X_m$  and we make another SPTA in the space of variables by maximizing

$$f(V, b) = \text{tr}(V'DX_cQX_c'DV) \quad (1)$$

subject to the constraints  $V'DV = I_r$  and  $b'b = 1$ , where  $b = [\beta_1, \dots, \beta_M]'$  is the vector containing weights of tables  $X_m$  and  $X_c = \sum_{m=1}^M \alpha_m X_m$  is a linear combination of matrices  $X_m$ .

At the optimum  $a = b$ .

In this way, the found sets  $U$  and  $V$  are used to represent the individuals and the variables.

Proof of this result that we recall here is found in the paper of Nguesso et al. [20]. The diagonalization of the matrix  $R_Z$  leads to the stationary equation:

$$R_Z a^* = \rho a^*.$$

or  $R_Z = S'_Z S_Z$  where

$$S_Z = [\text{vec}(D^{1/2} X_1 QZ) | \text{vec}(D^{1/2} X_2 QZ) | \dots | \text{vec}(D^{1/2} X_M QZ) |]$$

is the summary matrix associated with cube  $X$ .

If we set by  $s_m(Z) = \text{vec}(D^{1/2} X_m QZ)$  and

$$s_c(Z) = \sum_{h=1}^M \alpha_h^* s_h(Z) = \text{vec} \left( D^{1/2} \sum_{h=1}^M \alpha_h^* X_h QZ \right) = \text{vec}(D^{1/2} X_c QZ),$$

we get for  $m = 1, \dots, M$

$$\begin{aligned} \alpha_m^* &= \frac{1}{\rho} s'_m(z) \sum_{h=1}^M \alpha_h^* s_h(Z) \\ &= \frac{1}{\rho} s'_m(Z) s_c(Z) \\ &= \frac{1}{\rho} (\text{vec}(D^{1/2} X_m QZ)' \text{vec}(D^{1/2} X_c QZ)) \\ &= \frac{1}{\rho} \text{tr}(Z' Q X'_m D X_c QZ). \end{aligned}$$

If we set by  $P_{s_c(Z)} = \frac{1}{\|s_c(Z)\|^2} s_c(Z) s'_c(Z)$  the orthogonal projection on the

subspace of  $\mathbb{R}^{nr}$  spanned by the vector  $s_c(Z)$  with the usual metric, we obtain

$$(\alpha_m^*)^2 \rho^2 = s'_m(Z) s_c(Z) s'_c(Z) s_m(Z) = \|s_c(Z)\|^2 \|P_{s_c(Z)} s_m(Z)\|^2$$

As a result for all  $m$

$$|\alpha_m^*| = \|s_c(Z)\| \frac{\|P_{s_c(Z)} s_m(Z)\|}{\rho} = \frac{|tr(Z' Q X'_m D X_c Q Z)|}{\rho}$$

$\rho = a^{*'} R_Z a^* = \|S_Z a^*\|^2$  is positive quantity. We get the following result for all  $m$ :

$$|tr(Z' Q X'_m D X_c Q Z)| = \|s_c(Z)\| \|P_{s_c(Z)} s_m(Z)\|.$$

Consequently the sign of  $\alpha_m^*$  depends on the sign of the quantity  $tr(Z' Q X'_m D X_c Q Z)$  for all  $m$  because  $\rho$  is positive.

### 2.2.3. STATICO method

We consider two multi-tables  $X = [X'_1 | X'_2 | \dots | X'_M]'$  and  $Y = [Y'_1 | Y'_2 | \dots | Y'_M]'$  associated respectively with group 1 and group 2. The STATICO method is a combination of Co-inertia and STATIS. It allow to analyze simultaneously a series of  $M$  couples of tables  $(X_m, Y_m)$ .

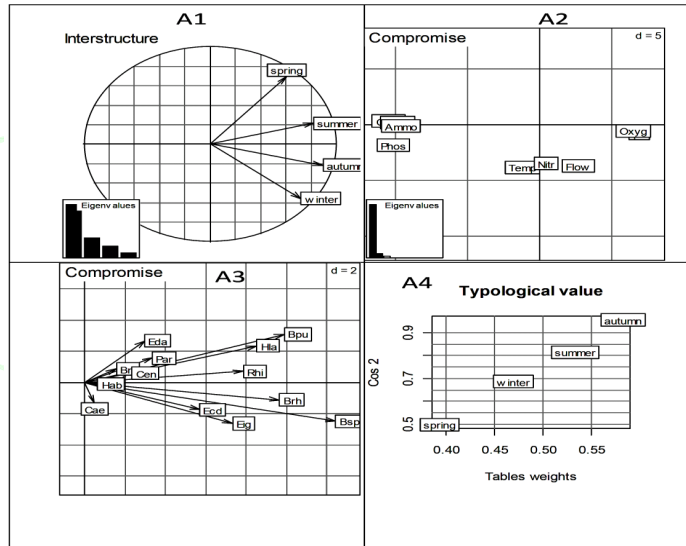
Let us denote by  $Z_m = Y'_m D_m X_m$  matrices of Co-inertia between tables  $Y_m$  and  $X_m$  for  $m = 1, 2, \dots, M$ .

The STATICO method of Simier et al. [24] is a STATIS on original tables or a Triadic Analysis Partial (PTA) of tables  $Z_m = Y'_m D_m X_m$ . More concretely, we search the compromise  $Z_c = \sum_{m=1}^M \alpha_m Z_m$  such as the inertia  $\|Z_c\|_{HS}^2 = tr(Z'_c D Z_c Q)$  is maximal subject to the constraint  $a'a = \sum_{m=1}^M \alpha_m^2 = 1$ .

The solution of this problem is given by diagonalization of the matrix  $S$  of dimension  $M \times M$  whose the general term  $S_{mk}$  is given by  $tr(Z'_m D Z_k Q)$ .

The step of the compromise of PTA is based on finding a compromise matrix  $Z_c^{(1)} = \sum_{m=1}^M \alpha_{m,1} Z_m^{(1)}$  which expresses the common structure with  $M$  tables. The coefficients  $\alpha_{m,1}$  are the components of the vector  $a_1$  solution of the maximizing  $\|Z_c\|_{HS}^2 = tr(Z'_c D Z_c Q)$  subject to the constraint of unit norm of the vector  $a$  ( $a'a = 1$ ). They are not all of the same sign some cases, which poses problems in STATICO method.

**Illustration of the results by the STATICO analysis.** The results of the data used in the STATICO method are available in the R software package especially in ade4. The ade4 package can be downloaded for free in the website: <http://biomserv.univ-lyon1.fr/ade-4.html>. We print the results available and we analyze them briefly for discussion purposes that will be presented in the next section.



**Figure 3.** This Figure contains four sub Figures. (A1) Position of the seasons on the two axes of the factorial map of the compromise, (A2) Representation of environmental variables on the factorial map of the compromise. (A3) Representation of species on the factorial map of the compromise. (A4) Typological values of the four tables.

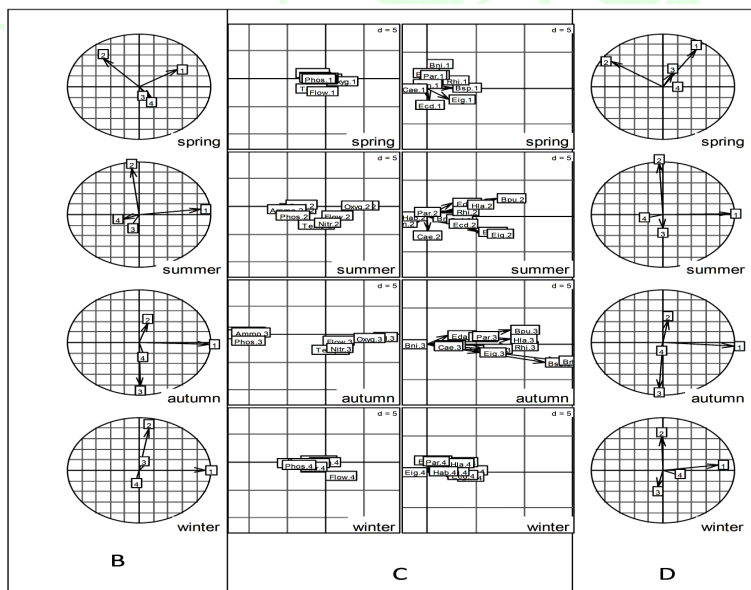
Figure 3 (A1) represents the factorial map of interstructure. It shows the contribution of each season in the formation of the compromise. Although the difference in contribution between the Autumn and Summer is only 0.048, Autumn season present as season which participate more in the formation of the compromise while Spring and Winter participate less with compromise and their tables are less structured.

Figure 3 (A2 and A3) shows the factorial map of the compromise. The results derived from Principal Component Analysis (PCA) of a share of physicochemical variables and other ephemeroptera species. These two typologies associated define the stability of the average species-environment relationship.

In Figure 3 (A2), Nitrates, the flow and temperature determine the axis 2 because one gets a specific distribution of faunistical composition that combines with environmental factors.

In Figure 3 (A3), we observe a size effect characterized by the species in combination with environmental factors. Watercourse pollution is represented by axis 1, the river restoration and the faunistical composition of the upstream different from downstream are described by the axis 2. So, each relationship “pollution-restoration” correspond a spatio-temporal stability of type upstream-downstream.

Figure 3 (A3) shows the non-pollution of stations characterized by the Bpu species are noticed in S1 and S6 stations on the one hand and the presence of species and Bsp Brh in all stations except in station S2 on the other hand. It also shows the faunistical composition in upstream of species (Bsp, Brh, Eig) and in downstream of species (Bpu, Hla, Eda).



**Figure 4 (B and D).** Representation of the stations per seasons on the factorial map of the PCA of the compromise. (C) Representation of the species and physicochemical variables seasons on the factorial map of the compromise.

Figure 4 (B and D) shows the results of the Principal Component Analysis (PCA) of the compromise on which one projects the six stations of the four seasons on the factorial map given by the PCA. The rows of table of the physicochemical

variables and of table of the ephemeroptera species are projected to represent the two points of each station. Species composition and the physical properties of water are directly related to the gradients of restoration of pollution.

Figure 4 (C) shows the projections of the 10 physicochemical variables and 13 species on the factorial map of the PCA of the compromise. We can find for all seasons more or less in the station S6 the species “Bsp” and “Brh”.

The axis 1 opposes on the one hand the station S6 and on the other hand the station S2 for all seasons. In Spring, station S5 is characterized by high temperatures and flow. In autumn, the station S2 is characterized by Phosphates and the Ammonium. Near the center spaces, there are rare species (“Hab”, “Bni”) in Autumn and (“Cae”, “Brh”) in Spring. The Axis 2 opposes generally on the one hand station S1 and S4 and S6 stations on the other hand.

In view of the positions of the environmental variables, we notice a some difference where the “Ammo” and “Nitra” exchanged positions. In Summer, the presence of oxygen is very much important in all seasons except in winter where the flow in the station S5 and the conductivity of the water in the station S2 are very high. In autumn, except of temperature and flow that are low, all physicochemical variables are abundant at all stations except the station S1.

#### 2.2.4. COSTATIS method

The COSTATIS method is proposed by Thioulouse [25]. It is based on two Partial Triadic Analyses and a Co-inertia Analysis of the compromises. The method involves two steps. The first step is a Partial Triadic analysis on each  $M$  tables. We then find two compromises  $X_c = \sum_m^M \alpha_m X_m$  of dimension  $(n \times p)$  for the first Partial Triadic Analysis and  $Y_c = \sum_m^M \beta_m Y_m$  of dimension  $(n \times q)$  for the second Partial Triadic Analysis.

$$n = \sum_m^M n_m.$$

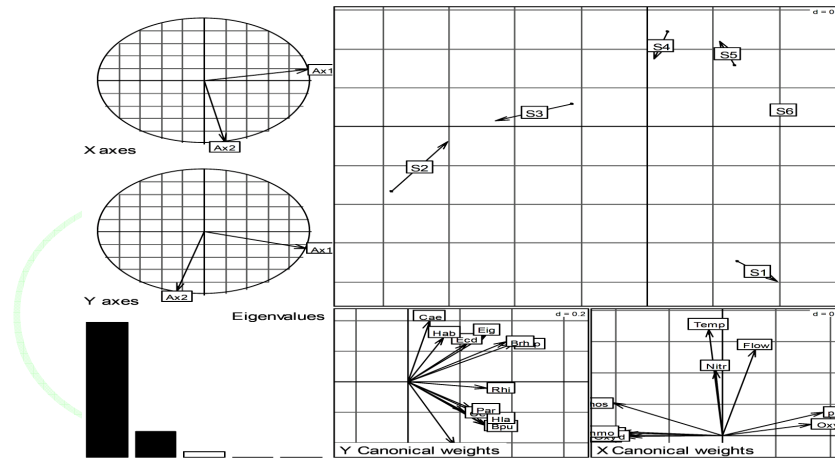
The second step is a Co-inertia analysis to maximize

$$\text{trace}(X_c Q X_c' D Y_c R Y_c' D)$$

subject to the constraints  $\sum_m^M \alpha_m^2 = 1$  and  $\sum_m^M \beta_m^2 = 1$ .

We recall that when the coefficients  $\alpha_m$  and  $\beta_m$  are not the same sign, the method no longer applies.

### Illustration of the results by the COSTATIS analysis.



**Figure 5.** Representation of the species and environmental variables per season on the first two of Co-inertia axes. Projection of the coordinates of the 10 physicochemical variables and 13 species on the first two Co-inertia axes.

La Figure 5 shows the results of physicochemical variables, ephemeroptera species and the position of the stations on the first two Co-inertia axes of the compromises. These axes are obtained by maximizing the covariance between the coordinates of the compromise variables of environment and the compromise species.

As in STATICO, axis 1 is an axis of pollution gradient and axis 2 the opposition upstream-downstream.

Figure 5 also shows the projection of the six stations in four seasons on the first two axes of Co-inertia compromise. In Winter, except the station S3 where pollution is low, the station S2 has a high pollution in all the seasons. This pollution reduces upstream throughout of the stations S3, S4 and S5 and is lower in the station S6. It also shows the opposition upstream-downstream of the ephemeroptera species (see Thioulouse [25]).

On one side, upstream, we have the species (“Bsp”, “Brh”, “Eig”) more abundant in stations S5 and S6 and on the other downstream side, we have the species “Eda” in very high quantities.

**2.2.5. ATPSCO method**

ATPSCO is an alternative method to COSTATIS. It is based on two simultaneous Partial Triadic Analyses and a Co-inertia Analysis of the compromises. The method involves two steps.

The first step is a simultaneous Partial Triadic analysis on each  $M$  tables. We then find two compromises  $X_c = \sum_m^M \alpha_m X_m$  of dimension  $(n \times p)$  for the first simultaneous Partial Triadic Analysis and  $Y_c = \sum_m^M \beta_m Y_m$  of dimension  $(n \times q)$  for the second simultaneous Partial Triadic Analysis.

$$n = \sum_m^M n_m.$$

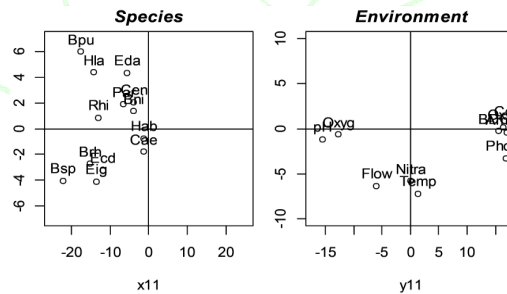
The second step is a Co-inertia analysis to maximize

$$trace(X_c Q X_c' D Y_c R Y_c' D)$$

subject to the constraints  $\sum_m^M \alpha_m^2 = 1$  and  $\sum_m^M \beta_m^2 = 1$ .

We recall that when the coefficients  $\alpha_m$  and  $\beta_m$  are not the same sign, the COSTATIS method no longer applies, it must then use the ATPSCO method which gives the coefficients of the same sign.

**Illustration of the results by the ATPSCO analysis.**



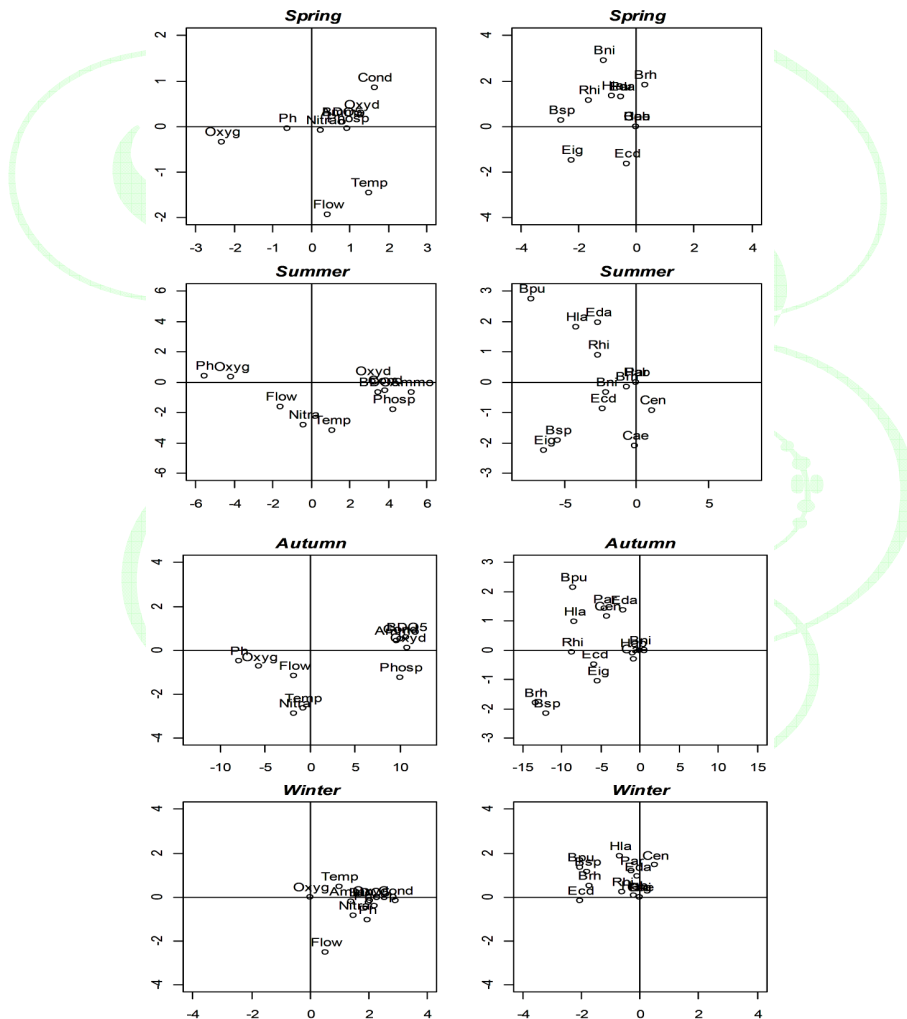
**Figure 6.** Representation of the species and environmental variables on the Co-inertia axes of the Co-inertia analysis of two compromises.

There is a centering per season for the species and environmental variables. This centering is followed the overall normalization (Bouroche [3]) of the environmental variables. This normalization of Bouroche has a more ecological sense since it will



keep the inter-season variability. It varies from one season to another and from one variable to another.

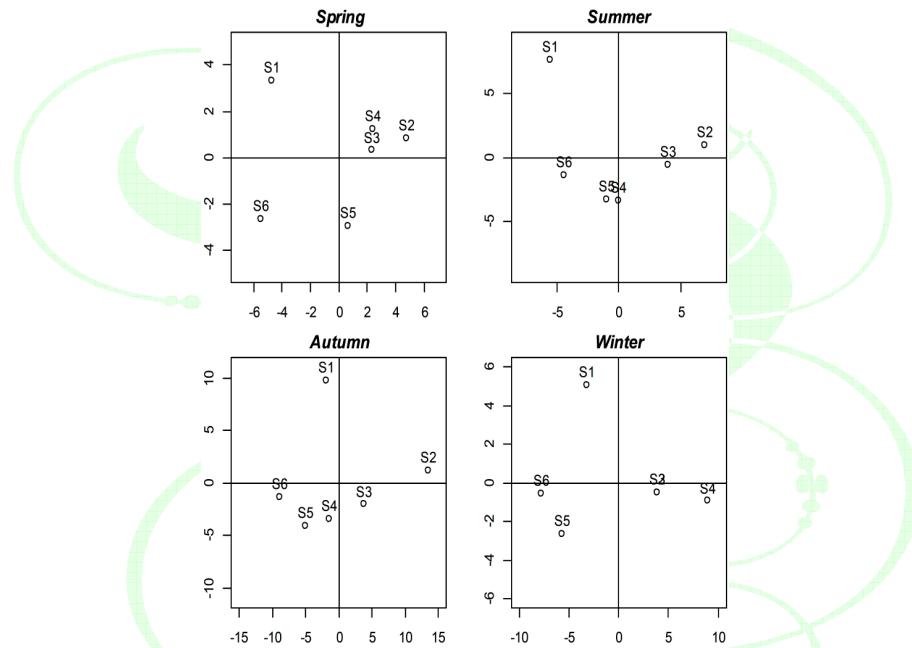
For reasons of clarity, the graphics are not superimposed. The same structure we have found for STATICO and COSTATIS is observed. Figure 6 shows two graphs: one for ephemeroptera species and another for environmental variables. It shows that the environmental variables have a structure along the axes and faunistical variables, a structure along the bisectrices.



**Figure 7.** Projection of the coordinates of the 13 species and 10 environmental variables on the Co-inertia axes of the Co-inertia analysis of two compromises.

As in COSTATIS and STATICO, axis 1 corresponds to the pollution of river and axis 2 describes the restoration of the river. But the composition of the upstream is different from downstream.

Figure 7 shows the projection on the axes of Co-inertia given by the Co-inertia analysis of two compromises of all species and environmental variables per season. It shows the same opposition upstream-downstream characteristic of the ephemeroptera species.



**Figure 8.** Projection of the coordinates of the six stations on the Co-inertia axes given by the Co-inertia analysis of two compromises.

When looking at easily in Figure 8 which shows the projection of all lines (stations) on the axes of Co-inertia and Figure 7, we observe the abundance of species in almost all stations except stations S2 and S3 that have a low number of ephemeroptera species. Thus, the station S1 contains a high number of species “Eda” and the stations S5 and S6 have a high number of species “Bsp”, “Brh” and “Eig”.

This example illustrates perfectly that ATPSCO present the same structure as the other methods. It presents the axis 1 as characteristic of the pollution of rivers impacting on species abundance and the axis 2 as opposition in upstream and downstream.

### 2.2.6. CO-ATPS method

The context of the CO-ATPS method is identical to that of STATICO. Thus, we consider two multi-tables  $X = [X'_1 | X'_2 | \dots | X'_M]'$  and  $Y = [Y'_1 | Y'_2 | \dots | Y'_M]'$  associated respectively with the Group 1 and Group 2 as defined under Section 2.4 let us recall it.

The CO-ATPS method is a simultaneous Partial Triadic Analysis (SPTA) of the tables  $Z_m = Y'_m D_m X_m$ . More concretely, one seeks the compromise  $Z_c = \sum_{m=1}^M \alpha_m Z_m$  such as the function  $f$  defined below is maximum

$$f(U, a) = \text{tr}(U' Q Z'_c D Z_c Q U)$$

subject to the constraints  $U' Q U = I_r$  and  $a' a = 1$ . The matrix  $U$  and the vector  $a$  are determined by the algorithm of the SPTA.

If the weighting coefficients  $\alpha_m^*$  of tables  $X_m$  are all of the same sign, which is the advantage in the construction of the compromise, we make the PCA of compromise. On the other hand, if are not of the same sign, the compromise are a linear combination of matrices  $X_m$  and we make another SPTA in the space of variables by maximizing

$$f(V, b) = \text{tr}(V' D X_c Q X'_c D V) \quad (2)$$

subject to the constraints  $V' D V = I_r$  and  $b' b = 1$  where  $b = [\beta_1, \dots, \beta_M]'$  is the vector containing weights of tables  $X_m$  and  $X_c = \sum_{m=1}^M \alpha_m X_m$  is a linear combination of matrices  $X_m$ .

At the optimum  $a = b$ .

In this way, the found sets  $U$  and  $V$  are used to represent the individuals and the variables. The CO-ATPS method operates on the tables  $Z_m = Y'_m D_m X_m$  in three steps like the STATICO method.

**(a) Definition of the average.** From the tables  $X_m$  and  $Y_m$  for  $m = 1, \dots, M$ , we build cross tables Co-inertia  $Z_m = Y'_m D_m X_m$  for  $m = 1, \dots, M$  of dimension

$q \times p$  and whose associated statistical triplet is  $(Z_m = Y'_m D_m X_m, Q, R)$ . From the triplet  $(Z_m = Y'_m D_m X_m, Q, R)$ , one seeks the triplet  $\left( Z_c = \sum_{m=1}^M \alpha_m Z_m, Q, R \right)$  such as the function  $f$  defined below is maximum subject to the constraints  $U'QU = I_r$  and  $a'a = 1$ . This maximizing implies the execution of an iterative algorithm and converge. The coordinates  $a$  are numbers  $\alpha_m$  and are all of the same sign, they are considered all positive.

This algorithm involves the determination of the matrix  $U$ .

**(b) Analysis of the average.** The analysis of the average can be examined with care the average relationship between the two groups of variables.

The average  $Z_c = \sum_{m=1}^M \alpha_m Z_m$  is the matrix of Co-inertia  $Y'_c D X_c$  of two compromise tables  $X_c$  and  $Y_c$ . Find a compromise co-inertia, it is to find numbers  $\alpha_m$  such as triplet  $\left( Z_c = \sum_{m=1}^M \alpha_m Z_m, Q, R \right)$  present a maximum inertia. This analyze provides two sets of axes, one in  $\mathbb{R}^p$   $Q$ -orthonormal which we note by  $F$ , another in  $\mathbb{R}^q$   $R$ -orthonormal which we note by  $H$  which are often Co-inertia axes. The table columns  $Z_c$  are columns of multi-table  $X$  and its rows are the columns of multi-table  $Y$ .

Each row (individual) of the Co-inertia table  $Z_m$  is a point of  $\mathbb{R}^p$  that can be projected on the set of Co-inertia axes  $F$ . Similarly, each column (variable) of the  $Z_m$  is a point of  $\mathbb{R}^q$  that can be projected on the set of Co-inertia axes  $H$ . Thus, we obtain average co-structure between two multi-tables  $X$  and  $Y$ .

The concordance of the two projected clouds associated with each of all couples formed a triplet of the first group of variables and a triplet of the second group of variables at the same stage is obtained by normalization and superposition of the two projected clouds, for measure the part of the realization of the average relationship by each of all the couples of triplets. The stability of the relationship between tables  $X_m$  and  $Y_m$  can be evaluated using the weighting coefficients  $\alpha_m$  that allow to build the compromise. But we can also study the stability of the relationship between couple  $(X_m, Y_m)$  calculate the correlation coefficients between partial components

$c_{X_{m,s}} = X_m Q f_s$  and  $c_{Y_{m,s}} = Y_m R h_s$ , where  $f_s$  and  $h_s$  are respectively the columns of the matrices  $F$  and  $H$  are obtained by analyzing the mean triplet  $(Z_c, Q, R)$ .

**(c) Stability of the average.** This is to study the stability around of the compromise. The Co-inertia axes determined during the analysis of the average are the result of the study of the temporal variability of the average common structure described by the compromise across all the cross tables  $Z_m$ .

The compromise is a mean of the structures, it should be represented in the infrastructure (the reproducibility by study of this structure) expression elements of the compromise, therefore clouds of points, clouds of couples of points. To do this, the Simultaneous Partial Triadic Analysis associated with the CO-ATPS method uses the technique of supplementary elements. We can project on the axes and the principal component of the compromise the rows, the columns, the axes and the principal component of the separated analyses of each table.

**Illustration of the results by the CO-ATPS analysis.** Table compromise  $Z_c$  to 40 rows and 13 columns is formed of the superposition of 4 cross table of 10 rows-variables and 13 columns-species.

The compromise analysis provides two factorial maps: that of the physicochemical variables and that of the ephemeroptera variables (species). This figure shows the stability of the species-environment relationship.

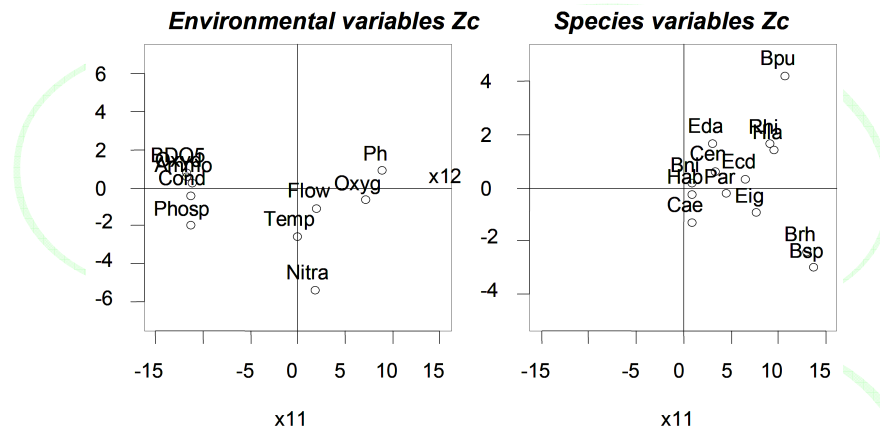
The first factorial map of physicochemical variables on the first axis opposes environmental factors. At the left, we find: Conductivity, Ammonium, Phosphates, the Oxydability and biological demand for oxygen and pH.

**Table 1.** This table contains the weighting coefficients of the matrices  $Z_m = Y'_m D_m X_m$

Methods	Seasons			
	Spring	Summer	Autumn	Winter
CO-ATPS ( $\alpha_m$ )	0.106	0.379	0.909	0.130
STATICO ( $\alpha_m$ )	0.393	0.533	0.582	0.469

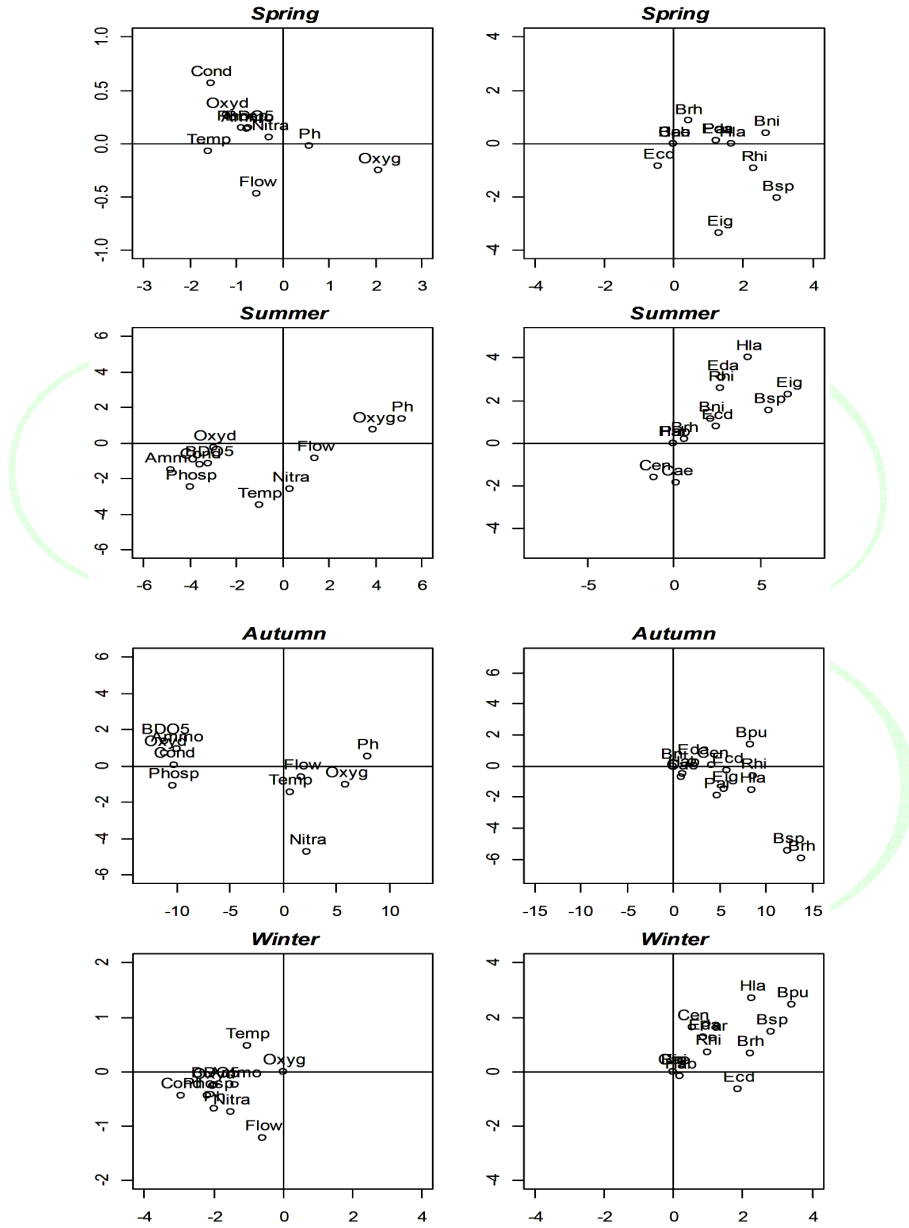
**Table 2.** Squared cosine between  $Z_m = Y'_m D_m X_m$  and  $Z_c$ 

	Spring	Summer	Autumn	Winter
$\cos^2$	0.183	0.578	0.975	0.276



**Figure 9.** Representation of the compromise variables of the environment and faunistical compromise variables on the first two axes of the PCA of the compromise. Projection of the coordinates of 10 physicochemical variables and 13 ephemeroptera species on the factorial map of the PCA of the compromise.

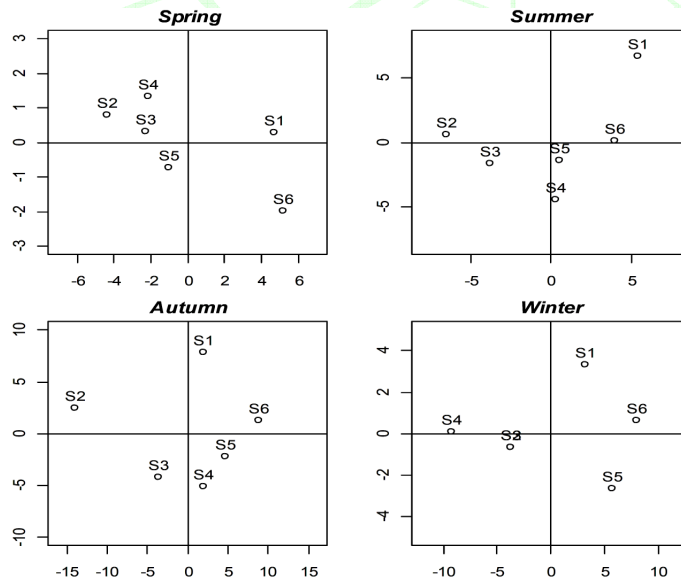
So, the first axis characterizes the pollution of the environment. On the second axis the Nitrates variable is good represented in so far as it has a better position relative to the temperature and flow which are almost in the same direction. This proves that in Autumn, the importance of nitrates is greatest in station S4. The second axis characterizes the opposition upstream-downstream of the species-environment relationship. We find the same structure as observed by STATICO, COSTATIS and ATPSCO. The second factorial map provides an image of the ephemeroptera species. We recognize above on the map 1-2 a cloud of points-species characteristic of the size effect and cloud of point variables characteristic elements pollution-restoration of the river. Pollution measured by highly redundant descriptors (Ammonium, oxygen, BOD5, Oxydability) and its link to the mineral load and organic load (Conductivity, Phosphates) generates high inertia.



**Figure 10.** Representation of the species and environmental variables per seasons on the first two axes of the PCA of the compromise. Projection of the coordinates of 10 physicochemical variables and 13 ephemeroptera species on the first two axes of the PCA of the compromise.

Table 1 illustrates that the CO-ATPS method shows good that the Autumn contributes more to the formation of compromise with differential contribution on the Summer 0.430 stronger than STATICO that is only 0.487. Seasons Spring (0.106) and Winter (0.130) does not contribute much to the definition of compromise. Simultaneously, the environmental variability (mesological) and faunistical diversity is low in winter and spring.

Table 2 contains the squared cosine between tables. It is clear from this table that the Fall and Summer have raised squared cosines. CO-ATPS discriminate better the environmental variables giving on the axe 2 by a broad advance to the Nitrates on temperature whereas in STATICO and in COSTATIS these variables occupy almost the same position. In autumn, Bsp, Eig and Brh species are more abundant in the stations S4 and S5 while Bpu species are more numerous in the stations S1 and S6. In summer, the pollution is stronger and it is characterized by physicochemical descriptors (pH, oxygen) in the station S6 and (Ammo, Phosp, BOD5) in the station S3; restoration is low. After pollution, Méaudret river restores, but its flow and its temperature increase, and the faunistical content downstream is distinct from the faunistical content upstream. Hence a structure along the axes for environmental variables and a structure along the bisectrices for the faunistical variables.



**Figure 11.** Representation of the stations son the first axes of the PCA of the compromise.



### 3. Discussion

All these methods have advantages and some disadvantages which can be summarized as follows:

**STATICO:** A big advantage of this method is to summarize optimally all the information contained in the original tables. It allows to see a joint structure and characterizes the stability of relations between mesological and faunistical surveys. As with any strategy of the STATIS type, this method calculates an average (compromise), analyzes the average (compromise analysis) and uses the results to examine the components of this average (intrastructure).

The major disadvantage of this method is that it is not applicable in case the weights are not the same sign.

**COSTATIS:** As its name suggests (Co-inertia and STATIS), this method is the Co-inertia analysis of compromise. It first summarizes optimally the information contained in both  $M$ -tables: mesological table and faunistical table thus giving two compromises via two Partial Triadic Analyses, which are then analyzed by the Co-inertia to search for relations between two stable structures.

The major disadvantage of this method is that it is not applicable in case the weights are not the same sign. Furthermore, the method works only in the case of two cubes.

**ATPSCO:** The theoretical operation of the COSTATIS method is identical to that of the method ATPSCCO. It first summarizes optimally the information contained in both  $M$ -tables: mesological table and faunistical table thus giving two compromises via two Simultaneous Partial Triadic Analyses, which are then analyzed by the Co-inertia to search for relations between two stable structures.

Furthermore, the method only works in the case of two cubes.

**CO-ATPS:** The theoretical operation of the CO-ATPS method is identical to that of STATICO method. It comes in three stages: the definition of the average, the average analysis, stability analysis of the average. To determine the weighting coefficients, STATICO proceeds by the diagonalization of the matrix of scalar products between crossed triplets  $(Z_m = Y'_m D_m X_m, Q, R)$  for  $m = 1, \dots, M$ , but the CO-ATPS proceeds by an algorithm to find the weights. But the CO-ATPS method is more interesting than the STATICO method since STATICO poses

problems because if the weighting coefficients that allow the construction of the compromise are not all of the same sign in some cases, the STATICO method does not apply, on the other hand, the CO-ATPS method is always applicable. For each table and for each solution, the quantity is computed

$$Cmc = \cos^2(Z_m, Z_c) = \frac{(tr^2(H'RZ'_mQZ_cRH))}{tr(H'QZ'_mQZ_mRH)tr(H'RZ'_cQZ_cRH)}$$

This quantity is an indicator of the quality of the representation of each table  $Z_m$  in the compromise  $Z_c$ .

#### 4. Conclusion

The CO-ATPS operate as STATICO. It determines a compromise which is an average when the weighting coefficients are all of the same sign and a linear combination of matrices when they are not them. In both cases the CO-ATPS determine two sets of representation in which one represents the variables of two groups. These sets of axes allow to visualize the individuals. The CO-ATPS is interesting than the ATPSCO in so far as the ATPSCO is applied that in the case of two cubes.

#### References

- [1] L. Blanc, D. Chessel and S. Dolédec, Étude de la stabilité temporelle des structures spatiales par les analyses d'une série de tableaux de relevés fauniques totalement appariés, *Bulletin Français de la Pêche et de la Pisciculture* (1998), 1-21.
- [2] L. Blanc, Données spatio-temporelles en écologie et analyses multi-tableaux : examen d'une relation, Thèse de Doctorat unique, Université de Lyon I, France, 2000.
- [3] J. D. Bouroche, Analyse des données ternaires: La double analyse en composantes principales, Thèse de 3-ème cycle, Université de Paris VI. France, 1975.
- [4] P. Cazes, L'analyse de certains tableaux rectangulaires décomposé en blocs: généralisation des propriétés rencontrées dans l'étude des correspondances multiples, *Les cahiers de l'analyse des données* 5 (1980), 145-161 and 387-406.
- [5] D. Chessel and P. Mercier, Couplage de triplets statistiques et liaisons espèces-environnement, *Biométrie et Environnement*, J. D. Lebreton and B. Asselain, eds., Masson, Paris, 1993, pp. 15-44.

- [6] D. Chessel and M. Hanafi, Analyse de la Co-inertie de  $K$  nuages de points, *Revue de Statistiques Appliquées* 44 (1996), 35-60.
- [7] R. L. Crocker, Soil genesis and the pedogenic factors, *Quart. Rev. Biol.* 27 (1952), 136-168.
- [8] S. Dolédec and D. Chessel, Co-inertia analysis: an alternative method for studying species-environment relationships, *Freshwater Biology* 31 (1994), 277-294.
- [9] Y. Escoufier, Le traitement des variables vectorielles, *Biometrics* 29 (1973), 750-760.
- [10] Y. Escoufier, The duality diagram: a means of better practical, *Developpementin numerical ecology*, P. Legendre and L. Legendre, eds., NATO Advanced Institute, Serie G, Springer Verlags, Berlin, 1987, pp. 139-156.
- [11] M. Hanafi, Structure de l'ensemble des analyses multivariées des tableaux de données à trois entrées: Eléments théoriques et appliqués, Thèse de Doctorat unique, Université Claude Bernard de Lyon 1, France, 1997.
- [12] M. Hanafi and H. A. L. Kiers, Analysis of  $K$  sets of data, with differential emphasis on agreement between and within sets, *Computational Statistics and Data Analysis* 51 (2006), 1491-1508.
- [13] H. Hotelling, Relations between two sets of variates, *Biometrika* 28 (1936), 321-377.
- [14] G. Kissita, P. Cazes, M. Hanafi and R. Lafosse, Deux méthodes d'analyse factorielle du lien entre deux tableaux de variables partitionnés, *Revue de Statistique Appliquée* LII(3) (2004), 73-92.
- [15] G. Kissita, L. Niéré, R. A. Makany and D. Mizère, Canonical correlation analysis between two vertical multi-block tables and applications to qualitative variables, *Advances and Applications in Statistics* 34(1) (2013), 11-38.
- [16] Ch. Lavit, Y. Escoufier, R. Sabatier and P. Traissac, The ACT (STATIS method), *Computational Statistics and Data Analysis* 18 (1994), 97-119.
- [17] H. L'Hermier Des Plantes, Structuration de tableaux à trois indices de la statistique. Théorie et applications d'une méthode d'analyse conjointe, Thèse de 3<sup>ème</sup> cycle, Université Montpellier II (1976), 98 p.
- [18] R. Margaref, *Perspectives in ecological theory*, Univ. Chicago Press, Chicago, 1968.
- [19] D. Mizère, E. Nguessolta, G. Kissita and R. A. Makany, The successive partial triadic, analysis, *Far East Journal of Mathematical Sciences (FJMS)* 74(2) (2013), 341-360.
- [20] E. Nguessolta, D. Mizère, G. Kissita and R. A. Makany, L'Analyse Triadique Partielle Simultanée, Submit to the *Revue de Statistique Appliquée*, 2013.
- [21] D. Pegaz-Maucet, 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<sup>ème</sup> cycle, Université Lyon I, France, 1980.

- [22] J. Thioulouse and D. Chessel, Les analyses multitableaux en ecologie factorielle, i: De la typologie d'état à la typologie de fonctionnement par l'analyse triadique, *Acta Oecologica, Oecologia Generalis* 8 (1987), 463-480.
- [23] J. Thioulouse, M. Simier and D. Chessel, Simultaneous analysis of sequence of paired ecological tables, *Ecology* 85(1) (2004), 272-283.
- [24] M. Simier, L. Blanc, F. Pellegrin and D. Nandris, Approche simultanee de  $k$  couples de tableaux: Application à l'étude des relations pathologie végétale-environnement, *Rev. Statist. Appl.* 47 (1999), 31-46.
- [25] J. Thioulouse, Simultaneous analysis of sequence of paired ecological tables: a comparison of several methods, *The Annals of Applied Statistics* 5(4) (2011), 2300-2325.
- [26] L. R. Tucker, An inter-battery method of factor analysis, *Psychometrika* 23 (1958), 111-136.
- [27] M. Vivien, Approches PLS linéaires et non linéaires pour la modélisation de multi-tableaux, *Théories et Applications*, Thèse, Université de Montpellier I, 2002.
- [28] M. Vivien and R. Sabatier, A generalization of STATIS-ACT strategy: DO-ACT for two multiblocks tables, *Computational Statistics and Data Analysis* 46 (2004), 155-171.