

# Using Partial Triadic Analysis for Depicting the Temporal Evolution of Spatial Structures: Assessing Phytoplankton Structure and Succession in a Water Reservoir

**Frédéric Bertrand**

*Institut de Recherche Mathématique Avancée, Université de Strasbourg & CNRS, France*

**Myriam Maumy**

*Institut de Recherche Mathématique Avancée, Université de Strasbourg & CNRS, France*

*Partial triadic analysis is a multiway analysis method that is a well suited statistical tool to get a clear representation of a chronological series of matrices, one for each sample date. It allows the simultaneous principal component analyses of several matrices and permits one to find a spatial structure common to every matrix and to study its temporal stability. Partial triadic analysis begins by searching for an average table called compromise. The compromise table is then analyzed and its reproducibility by each initial table is finally investigated.*

*A partial triadic analysis was applied to a phytoplankton dataset that was collected in 2006 at six stations in the Marne Reservoir, located in France in the Seine catchment area. The spatial and temporal organizations of the assemblages of these different species were derived and hence the existence of some changes in water quality could be assessed since micro-organisms, especially phytoplankton species, may be considered as potential indicators of local and more global changes in aquatic ecosystems and may thus constitute an excellent biomarker of water quality. This example demonstrates the power of partial triadic analysis for depicting the temporal evolution of spatial structures.*

*The exposition is accessible to readers with an intermediate to advanced knowledge of statistics. Some prior exposure to principal component analysis is required for reading the article which can be viewed as a sequel to Bertrand et al. (2007). A basic knowledge of R is helpful.*

**Keywords:** *Multiway analysis, STATIS-ACT method, partial triadic analysis, principal component analyses, duality diagrams.*

## 1. Theory

### 1.1. Introduction

In this case study we investigate the potential of partial triadic analysis (PTA), a special kind of multivariate analysis. Partial triadic analysis is meant to perform a statistical analysis of experiments when the same variables are measured on the same individuals at different points in time. From a mathematical point of view, the aim of a PTA is to analyze a three-way table, also called a data cube i.e. a matrix with three subscripts  $X_{i,j,k}$  which is seen as a sequence of two-way tables – matrices with two subscripts  $X_{i,j}$ .

Partial triadic analysis belongs to the family of STATIS methods (Lavit 1988). It involves three essential steps: the interstructure, the compromise, and the intrastructure analysis.

Partial triadic analysis is one of the simplest multiway methods (Escoufier 1980). It was derived from triadic analysis (Tucker, 1966), described in Jaffrenou (1978) and introduced by Thioulouse and Chessel (1987) in ecology as “triadic analysis”. This method was renamed “partial triadic analysis” by Kroonenberg (1989) to highlight the difference with the original triadic analysis. It is also known as “Pre-STATIS” or “STATIS on the X” in Leibovici (1993) or “PCA-SUP” – PCA of a derived two-way supermatrix – in Kiers (1991). PTA has been used in several ecological studies (Dolédéc and Chessel 1991, Blanc *et al.* 1998, Blanc and Beaudou 1998, Gaertner 2000, Rossi 2003, Jiménez *et al.* 2006, Ernoult *et al.* 2006, Carassou and Ponton 2007, Pavoine *et al.* 2007 and Rolland *et al.* 2009).

More precisely, PTA searches for structures that are stable along the sequence of two-way tables. Its first step is called the interstructure. It corresponds to a global representation and gives the “importance” of each table. During this step, a matrix of scalar products between tables is computed.

The second step, the compromise computation and analysis, is the main step of the method. 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 has the same dimensions and the same structure and meaning as the tables of the series. The compromise table exhibits the best summary properties of the initial tables. It is analyzed by a PCA, providing a picture of the structures common to all the tables and a simultaneous representation of individuals and variables.

The third step of a partial triadic analysis is the analysis of the intrastructure. The rows and columns of all the tables of the sequence are projected on the factor map of the PCA of the compromise as additional elements. This step summarizes the variability of the series of tables around the common structure defined by the compromise.

### 1.2. Duality diagrams

Partial triadic analysis can be included in a broader setting using duality diagrams (Escoufier 1987, Holmes 2008). The terms duality diagram or statistical triplet are often used interchangeably. Such a “triplet”  $(X, D_p, D_n)$  can be used to define a multivariate data analysis from a geometrical point of view.  $X$  is the  $n \times p$  table to analyze,  $D_p$  the most often diagonal  $p \times p$  symmetric positive definite matrix of column weights and  $D_n$  is an  $n \times n$  matrix of weights on the “observations”, which is most often diagonal too.

For instance, a principal component analysis (PCA) of a matrix  $Z$  amounts to the statistical triple  $(X, I_p, 1/nI_n)$  where  $X_{ij}$  is given by  $Z_{ij} - (\mu_{zj})$  with  $(\mu_{zj})$  the mean of the  $j^{\text{th}}$  variable. Given a duality diagram and from a theoretical point of view, one proceeds to the eigendecomposition of the matrix  $X'D_nXD_p$ , if  $n > p$ , or of the matrix  $XD_pX'D_n$  if  $n < p$ , where  $X'$  is the transpose of the matrix  $X$ .

Table 1. Steps of a duality diagram analysis

Relation used	Result
$X'D_nXD_p = \lambda_s u_s$	Eigendecomposition ( $\lambda_s$ $s^{\text{th}}$ eigenvalue and $u_s$ $s^{\text{th}}$ eigenvector).
$G_s = \lambda_s^{1/2} u_s$	Coordinates of the active columns (X).
$F_s = XD_p u_s$ $F_{\text{sup},s} = X'_{\text{sup}} D_n u_s$	Coordinates of the active (X) or supplementary ( $X_{\text{sup}}$ ) rows.
$G_{\text{sup},s} = \lambda_s^{1/2} X'_{\text{sup}} D_n F_s$	Coordinates of the supplementary columns ( $X_{\text{sup}}$ ).

Yet since the numerical algorithms for computing the eigendecomposition of a real matrix are far more efficient if the matrix is symmetric we actually proceed to the eigendecomposition of the symmetric matrix  $A = D_p^{1/2} X' D_n X D_p^{1/2}$ , if  $n > p$ , or of the symmetric matrix  $B = D_n^{1/2} X D_p X' D_n^{1/2}$  if  $n < p$ . The eigenvalues of  $A$  are the same as those of  $X' D_n X D_p$  and the eigenvalues of  $B$  are the same as those of  $XD_p X' D_n$ . The eigenvectors of  $A$  multiplied by  $D_p^{1/2}$  are those of  $X' D_n X D_p$  and the eigenvectors of  $B$  multiplied by  $D_n^{1/2}$  are those of  $XD_p X' D_n$ . In the case of a PCA, this leads to finding an eigendecomposition of  $A = (1/n) X' X$  or  $B = (1/n) X X'$  and thus, since the matrix  $X$  is centered, to find an eigendecomposition of the covariance matrix of the variables  $Z_1, \dots, Z_p$  which is what PCA is.

We will use the formulas summed up in Table several times in what follows.

### 1.3. The STATIS-ACT method

The STATIS-ACT method is an exploratory technique of multivariate data analysis which is a generalization of principal component analysis. Its goal is to analyze several sets of variables collected on the same set of observations. It is attributed to Escoufier (1980) and L'Hermier des Plantes (1976, see also Lavit *et al.*, 1994), a related approach is known, in the English speaking community, as procrustes matching by congruence coefficients (Korth and Tucker, 1976). For a quick yet thoroughly detailed introduction to the STATIS method and some of its computational details we point the reader to Abdi and Valentin (2007).

The STATIS-ACT method is based on linear algebra and especially Euclidean vector spaces (STATIS stands for Structuration des Tableaux A Trois Indices de la Statistique, ACT stands for Analyse Conjointe de Tableaux). It has been devised for multiway data situations on the basic idea of computing Euclidean distances between configurations of points (Escoufier, 1973).

Any STATIS analysis is made of three successive steps.

1. **Interstructure:** compare and analyze the relationship between the different data sets.
2. **Compromise:** combine them into a common structure called a compromise which is then analyzed via PCA to reveal the common structure between the observations, see Lazraq (2008) for an inferential approach to validating the compromise.
3. **Trajectories:** project each of the original data sets onto the compromise to analyze communalities and discrepancies.

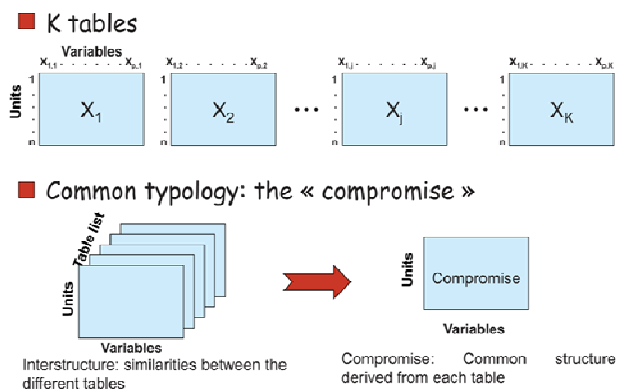
STATIS is used in very different areas such as sensory evaluation and food science (Qannari *et al.* 1995, Schlich 1996, Martin *et al.* 2000, Meyners *et al.* 2000, Meyners 2002, Chaya *et al.* 2004, Perez-Hugalde *et al.* 2004, Perrin *et al.* 2008), molecular imaging (Coquet *et al.* 1996) brain imaging (Kherif *et al.* 2003), ecology (Dolédéc 1988, Aliaume *et al.* 1993, Simier *et al.* 1999, Cadet *et al.* 2005), coastal marine ecology (Gaertner *et al.* 1998, Licandro and Ibanez 2000, Lekve *et al.* 2002, Gailhard *et al.* 2003, Muiño *et al.* 2003, Simier *et al.* 2004, Lobry *et al.* 2006), limnology (Centofanti *et al.* 1989, Anneville *et al.* 2002), chemometrics (Stanimirova *et al.*, 2004) and marketing (Perez Aparicio *et al.* 2007).

The main idea of the ACT technique is to compare configurations of the same observations obtained in different circumstances. As a consequence, a measure of similarity between two configurations has to be introduced. This is equivalent to defining a distance between the corresponding scalar product matrices. We now follow the article of Lavit *et al.* (1994) to give more insight into the core of the STATIS-ACT method.

### 1.4. Partial triadic analysis

From a statistical point of view, partial triadic analysis is a multivariate method that analyzes matrices in a three dimensional data array. It is based on the logic of Principal Component Analysis (PCA). It is designed to study simultaneously several sub-matrices of quantitative data and to detect within the structure any pattern common to these different sub-matrices in order to extract a multivariate structure that is expressed through the different dates, see Figure 1.

PTA is a STATIS-like method; the only noticeable difference between PTA and STATIS is that, since the individuals and the variables are identical from a table to another, the steps of the STATIS analysis are applied directly to the observed tables  $X_k$  instead of the scalar products  $W_k = X_k'X_k$  derived from the observed tables. It is the reason why PTA is also called STATIS on the "X".



**Figure 1.** PTA: a method from the STATIS family  
 Note. Interstructure: similarities between the different tables.  
 Compromise: common structure derived from each table.

#### 1.4.1. Interstructure analysis

The derivation of the interstructure is based on the concepts of vector variance  $Var_V$ , vector covariance  $Cov_V$ , and vector correlation  $R_V$  (Escoufier, 1973). Let  $(S, D_p, D_n)$  and  $(T, D_p, D_n)$  be two duality diagrams. The vector covariance  $Cov_V(S, T)$  between the two duality diagrams  $(S, D_p, D_n)$  and  $(T, D_p, D_n)$  is defined by:

$$\text{Cov}_V(S, T) = \text{trace}(S' D_n T D_p).$$

The vector variance  $\text{Var}_V(S)$  of the duality diagram  $(S, D_p, D_n)$  is given by:

$$\text{Var}_V(S) = \text{Cov}_V(S, S) = \text{trace}(S' D_n S D_p).$$

The distance  $d^2(S, T) = \text{Tr}[(D_n^{1/2} S D_p^{1/2} - D_n^{1/2} T D_p^{1/2})'(D_n^{1/2} S D_p^{1/2} - D_n^{1/2} T D_p^{1/2})]$  deduced from the scalar product  $\langle S | T \rangle = \text{Tr}(D_p^{1/2} S' D_n^{1/2} D_n^{1/2} T D_p^{1/2}) = \text{Tr}(S' D_n T D_p) = \text{Cov}_V(S, T)$  can be used to compare the  $X_i$ 's. The square of the norm of a duality diagram is thus equal to  $\text{Cov}_V(S, S) = \text{Var}_V(S)$ .

Another question that naturally arises is: Should we compare the  $X_k$ 's or the normed  $X_k$ 's?

A large distance  $d^2(X_k, X_{k'})$  points out a strong difference between  $X_k$  and  $X_{k'}$ . Is it a difference in shape or a difference in size? To eliminate the second effect, we deal with the normed duality diagram deduced from  $(X_k, D_p, D_n)$ :

$$\frac{X_k}{\|X_k\|} = \frac{X_k}{\sqrt{\text{Var}_V(X_k)}} = \frac{X_k}{\sqrt{\text{tr}(X_k' D_n X_k D_p)}}.$$

As a consequence, we compute the normed scalar products, the cosine, between the duality diagrams  $(X_k, D_p, D_n)$  and  $(X_{k'}, D_p, D_n)$ , also known as the  $R_V$  correlation coefficient:

$$\begin{aligned} \left\langle \frac{X_k}{\|X_k\|} \middle| \frac{X_{k'}}{\|X_{k'}\|} \right\rangle &= \frac{\langle X_k | X_{k'} \rangle}{\sqrt{\text{Var}_V(X_k)} \sqrt{\text{Var}_V(X_{k'})}} \\ &= \frac{\text{Cov}_V(X_k, X_{k'})}{\sqrt{\text{Var}_V(X_k)} \sqrt{\text{Var}_V(X_{k'})}} \\ &= R_V(X_k, X_{k'}) \end{aligned}$$

Let  $C$  be the cosine matrix, also called the interstructure matrix, be a real square matrix of order  $K$  whose generic term  $c_{kk'}$  is the cosine between the  $X_k$  and  $X_{k'}$ :

$$C = ((R_V[X_k, X_{k'}]))_{1 \leq k, k' \leq K}.$$

The computation of the  $R_V$  coefficient matrix between the tables allows the comparison of the tables and the representation of the proximity between tables (Robert and Escoufier, 1976).

In the case of a PTA where we deal with standardized variables and always use the norm  $D_n = I_n$ ,  $D_p = 1/p I_p$ , the vector covariance between the tables  $X_j$  and  $X_k$  is the sum of the correlations between identical couples of variables:

$$\text{Cov}_V(X_j, X_k) = \text{tr}(X_j' X_k) / p = \sum_{i=1}^p \text{cor}(X_j^i, X_k^i)$$

The vector variance of table  $X_j$  is hence equal to:

$$\text{Var}_V(X_j) = \text{tr}(X_j' X_j) / p = \sum_{i=1}^p \text{cor}(X_j^i, X_j^i) = p$$

Consequently, the  $R_V$  coefficient is the mean of the correlations between identical couples of variables:

$$\begin{aligned} R_V(X_k, X_{k'}) &= \frac{\text{Cov}_V(X_k, X_{k'})}{\sqrt{\text{Var}_V(X_k)} \sqrt{\text{Var}_V(X_{k'})}} \\ &= \frac{\sum_{i=1}^p \text{cor}(X_k^i, X_{k'}^i)}{p} \end{aligned}$$

To plot the  $K$  stages in a  $h$ -dimensional space, we use the least squares approximation  $C_h$  of  $C$ , equal to the  $h$  first elements of the spectral decomposition  $\sigma_1 p_1 p_1' + \dots + \sigma_r p_r p_r' = P \Sigma P'$  of  $C$ , with  $P'P = I$ . Stage  $k$  is plotted as a point  $M_k$  whose coordinates are the  $h$  first elements of the  $k^{\text{th}}$  row of  $P \Sigma^{1/2}$ . The points  $M_1, \dots, M_k$  satisfy  $\langle OM_k | OM_{k'} \rangle = (C_h)_{kk'}$ . The loss function is:

$$\|C - C_h\|^2 = \sum_{k, k'} [(C - C_h)_{kk'}]^2 = \sum_{l=h+1}^r \sigma_l^2.$$

On such a graph, scalar products are not easily readable, except for the norm  $\|X_k\|$  approximated by the length of the vector  $OM_k$  and for the scalar product between normed  $X_k$  and normed  $X_{k'}$ , approximated by the cosine of  $(OM_k, OM_{k'})$ . The projected distance induced by the least squares approximation is systematically lower than  $d^2(X_k, X_{k'})$ .

#### 1.4.2. Compromise construction

At this point, in the STATIS method, a positive semi-definite assumption is necessary to derive the key Property 1 whereas this restrictive assumption was not necessary in the interstructure derivations.

**Property 1 (Lavit et al. 1994).** Let  $W_1 = \sigma_1 p_1 p_1'$  be the first element of the spectral decomposition of the interstructure matrix  $W$ . Components of  $p$  can be chosen positive.

This property is no longer true for the PTA since we work on the  $X$ 's tables instead of dealing with the  $W$ 's. As a consequence, if all the  $X$ s tables are not positively correlated with the first component,  $C_1 = \sigma_1 p_1 p_1'$ , or if any two of the  $X$ s tables are negatively correlated, then we will be unable to use the PTA and will have to use the STATIS method instead. We now assume that the components of  $p$  can be chosen positive.

Moreover if all the components of the  $C$  matrix are positive, then, on the plot of the  $K$  stages, cosines are

positive and thus all the points  $M_k$  are situated inside a convex cone.

**Definition 1 (Lavit et al. 1994).** The  $n \times p$  compromise matrix  $X^+$  is defined as a weighted sum  $\sum_k \alpha_k X_k$ . The coefficient  $\alpha_k$  is the coordinate of stage  $k$  in the one-dimensional plot deduced from the first element  $C_1$  of the spectral decomposition of  $C$ .

The  $C$  matrix is not centered, thus the first eigenvector of  $C$  highlights what is common to the different tables. Tables with larger values on the first eigenvector will have a larger weight.

**Property 3 (adapted from Lavit et al. 1994).** The compromise matrix  $X^+ = \sum_k \alpha_k X_k$  is the linear combination of the  $X_k$ 's which is the most closely related to the  $X_k$ 's. In other words,  $X^+$  maximizes:

$$f(\alpha_1, \dots, \alpha_K) = \sum_l \left\langle \sum_k \alpha_k X_k \middle| X_l \right\rangle^2 / \sum_k \alpha_k^2.$$

**Proof (adapted from Lavit et al. 1994).** We develop the numerator in:

$$\begin{aligned} f(\alpha_1, \dots, \alpha_K) &= \sum_l \sum_{k,k'} \alpha_k \alpha_{k'} \langle X_k | X_l \rangle \langle X_{k'} | X_l \rangle \\ &= \sum_{k,k'} \alpha_k \alpha_{k'} \sum_l \langle X_k | X_l \rangle \langle X_l | X_{k'} \rangle \\ &= \sum_{k,k'} \alpha_k \alpha_{k'} \sum_l C_{k'l} C_{lk} = \sum_{k,k'} \alpha_k \alpha_{k'} (C^2)_{k'k}. \end{aligned}$$

In the spectral decomposition of  $C$ ,  $p_1$  is the eigenvector of  $CC' = C^2$  associated to the largest eigenvalue. Thus the quotient can be written as the Rayleigh quotient  $x'CC'x / \|x\|^2$  with  $x = (\alpha_1, \dots, \alpha_K)'$ , which is maximum for  $x = p_1$ . Note that this quotient is also maximum for  $p_1 / \|p_1\|^2$ . As a consequence we can add the constraint that  $\sum_k \alpha_k^2 = 1$  in the definition of the compromise matrix  $X^+$ .

In the case of a PTA for which we deal with standardized tables we have:

$$\begin{aligned} f(\alpha_1, \dots, \alpha_K) &= \sum_{k,k'} \alpha_k \alpha_{k'} \text{tr}(X_k' D_n X_{k'} D_p) \\ &= (\alpha_1, \dots, \alpha_K)' C (\alpha_1, \dots, \alpha_K) \end{aligned}$$

which is maximum, under the constraint  $\sum_k \alpha_k^2 = 1$ , for  $(\alpha_1, \dots, \alpha_K)' = p_1$ .  $\square$

What does the compromise mean? If  $X_k$  and  $X_{k'}$  correspond to similar tables in shape and size, the angle  $(OM_k, OM_{k'})$  is small and the lengths  $\|OM_k\|$  and  $\|OM_{k'}\|$  are nearly the same. This case leads to identical values for  $\alpha_k$  and  $\alpha_{k'}$ . On the other hand, a large difference between

the two tables induces either a large angle  $(OM_k, OM_{k'})$  or unequal lengths for  $OM_k$  and  $OM_{k'}$  and hence different values for  $\alpha_k$  and  $\alpha_{k'}$  in both cases. Consequently  $C$  gives relatively less weight to outliers, and leads to a compromise configuration which reflects the inter-element distances as they are seen by the majority.

We now provide a graphical representation of the compromise. To plot the compromise  $X^+$  in a  $h$ -dimensional space, we use the least squares approximation  $X^+_h$  of  $X^+$ . We compute a non-centered non-scaled PCA of the transpose of the compromise matrix  $X^+$  and retain the  $h$  first elements of the eigendecomposition  $(1/p)X^+X^{+'} = P\Delta P^{-1}$  with  $PP^{-1} = I$ . The duality diagram to be used is  $(X^+, D_n = I_n, D_p = 1/pI_p)$ . Both column element  $j$  and row element  $i$  can be plotted as a point  $N_j$  whose coordinates are the  $h$  first elements of the  $j$ th row of  $P\Delta^{1/2}$  and a point  $M_i$  whose coordinates are the  $h$  first elements of the  $i$ th row of  $X^+P$ .

The loss function is

$$\|X^+ - X^+_h\|^2 = \text{Tr}[(X^+ - X^+_h)'(X^+ - X^+_h)] = \sum_{l=h+1}^r \delta_l^2.$$

### 1.4.3. Highlighting the intrastructure

The last step is called the study of the intra-structure or of the reproducibility of the compromise.

The original observed tables are projected separately onto the compromise to highlight which one fits the compromise best and whether some of them share the same patterns. This step is often summed up by putting together and comparing several figures whose coordinates can be computed with the following steps:

1. the projection on the first plane of principal axes of the compromise of the principal axes of the individual PCA of each observed table.
2. projection of the columns of the original observed tables  $X_k$  on the compromise  $X^+$ , *i.e.* treating them as complementary columns in the non-centered non-scaled PCA of the compromise.
3. projection of the rows of the original observed tables  $X_k$  on the compromise  $X^+$ , *i.e.* treating them as complementary rows in the non-centered non-scaled PCA of the compromise.
4. the first principal components plane of the compromise of the principal components of the individual PCA of each observed table.

To summarize, using the partial triadic analysis method as a space-time data analysis tools may improve the

knowledge of the dynamics of the linear relationships between individuals and variables.

We now proceed to the detailed analysis of an example.

## 2. Biological background

### 2.1. Aims and scope of the study

Micro-organisms, especially phytoplankton species, may be considered as potential indicators of local and more global changes in aquatic ecosystems and may thus constitute an excellent biomarker of water quality. Assessing biological, chemical and physical influence on phytoplankton regulation is a key process for understanding population structure and dynamics, population diversity and succession, and to propose, if necessary and when possible, a human-hand control before any excessive algal proliferation could occur. Such issues are of main concern to help scientists and water managers to make it possible for large-body reservoirs, lakes and ponds, to reach the “good ecological state” recommended by the Water Framework Directive (WFD) by 2015.

One goal of this statistical analysis was to identify both the spatial and temporal phytoplankton structure and dynamics in the Reservoir Marne, located in France in the Seine catchment area, see Figure 1. The Reservoir Marne is often referred to as one of the largest reservoirs in Western Europe: its area is equal to 48 km<sup>2</sup>. In 2006, *i.e.* the first year of the project, the reservoir was sampled once a month in March and April and then once every two weeks between May and September.

To assess spatial heterogeneity, six stations, see Figure 2, were investigated at a depth of three meters. Stations K and J correspond to two nautical basins, station G receives water from two “feeding” channels, stations I and H are close to the restitution channels and finally station N is more or less in the middle of the reservoir.

This sampling strategy was chosen in order to obtain the best picture of the structure of the microbial organisms considering also human feasibility in terms of sampling and analysis. More details on the experimental protocol and the sampling strategy can be found in Rolland *et al.* (2009).

### 2.2. Selection of species

Among the 282 phytoplankton species originally identified in the Reservoir Marne, 112 species were selected to be included in the multiway analysis, see 2 for

details. The criterion for selection was that the relative abundance and biovolume of the species would be greater than a given value set to 0.01%. It is a commonly used pattern of selection in phytoplankton community ecology studies.

Table 2. Phytoplankton family counts and color.

Phytoplankton Family	Total Species	Selected Species <sup>1</sup>	Color
Cryptomonads	7	6	red
Cyanobacteria	33	18	blue
Desmidiates	26	4	dark red
Diatoms	46	14	magenta
Dinoflagellates	15	6	cyan
Euglenoids	16	4	dark grey
Golden Algae	22	13	orange
Green Algae	111	47	green
Yellow-green Algae	6	1	black

## 3. Statistical analysis

### 3.1. Features of the dataset

Since the reservoir is not uniform, as revealed by the phytoplankton abundance and structure which greatly vary from one sampling station to another, and since a clear seasonal dynamic was observed for the different microbial communities with main differences on a vertical scale, multiway analyses, such as partial triadic analysis (PTA), a STATIS-like method, (Lavit *et al.* 1994), multiple coinertia analysis, (Chessel and Hanafi 1996), or statico, (Thioulouse *et al.* 2004), are well suited statistical tools to get a clear representation of the spatial and temporal organization of the assemblages of the different species.

These analyses can be performed using the *ade4*, (Chessel *et al.* 2004, Dray *et al.* 2007) package for the R software, (R Development Core Team 2009), and several customized functions written by the first author and available with the dataset. Additionally, in order to shed light on the associations between the species and the axes constructed by the partial triadic analysis, the Kendall correlation coefficient between the biovolume of the phytoplankton species and their factorial coordinates on the first three axes of the compromise were computed.

This correlation analysis was also performed between the biovolume of the phytoplankton species and each station during the analysis involving the reproducibility of the compromise among the 6 sampling stations.

<sup>1</sup> The selection of species was carried out following the rules detailed in Section 2 of the article.

### 3.2. Partial triadic analysis

Let  $X_G$ ,  $X_H$ ,  $X_I$ ,  $X_J$ ,  $X_K$  and  $X_N$  the matrices of biovolumes values whose 112 columns are the species and 11 rows are the sampling times. Each of these matrices is centered and scaled. Thanks to Figure 2, one easily understands why this dataset turn out to be a cube of data.



Figure 1. Location of the reservoir in France



Figure 2. Location of the sampling stations in the reservoir

We assume that all the tables  $X_G$ ,  $X_H$ ,  $X_I$ ,  $X_J$ ,  $X_K$  and  $X_N$  share the same column weights  $D_p$ , and that they also share the same row weights  $D_n$ , which leads, in our case, to six duality diagrams  $(X_G, D_p, D_n)$ ,  $(X_H, D_p, D_n)$ ,  $(X_I, D_p, D_n)$ ,  $(X_J, D_p, D_n)$ ,  $(X_K, D_p, D_n)$ ,  $(X_N, D_p, D_n)$ .

#### 3.2.1. Interstructure analysis

This is the first step of the analysis and is intended to provide a global description of the sampling points as a

function of the typology of the sampling dates, extracting the information common to all sampling dates. The objects of this analysis are the variables by samples. It consists of the comparison of the structure of the 6 different sub-matrices (stations) and the identification of the stations sharing a similar annual structure. The computation of the  $R_V$  coefficient matrix between stations allows the comparison of the stations and the representation of the proximity between stations. A byproduct of this step is to attribute a weight to each station sub-matrix.

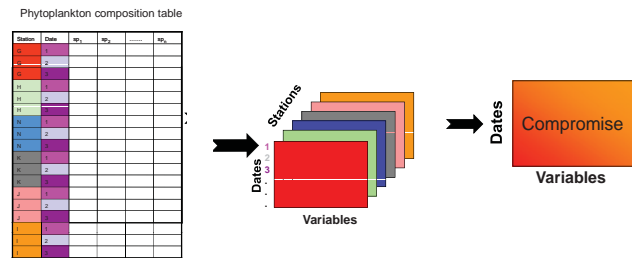


Figure 2. Phytoplankton data composition via a PTA

We begin by creating a ktab object that is designed to match the structure of the cube of data and allows the use of the functions of the ade4 package to perform partial triadic analysis.

```
> library(ade4)
> ktabbiovol3mscaled <- ktab.list.df(
  lapply(rapply(lapply(lapply(split(
    cbind(codes$Lettre,biovol3m), codes$Lettre),
    function(x) {return(x[,-1])}),scale),
    function(XX) {apply(XX, c(1,2), function(l1)
      {if(!is.nan(l1)){l1} else {0}})},
    how="replace"),as.data.frame))
> row.names(ktabbiovol3mscaled) <-
  c("06/03/2006", "04/04/2006", "03/05/2006", "07/
  06/2006", "28/06/2006", "04/07/2006", "24/07/200
  6", "07/08/2006", "22/08/2006", "05/09/2006", "19
  /09/2006")
```

We use the following R code to compute the C matrix:

```
> RV.comp <- function(mat1,mat2)
  {return(sum(diag(t(as.matrix(mat1)))%*%as.matr
  ix(mat2)))/sqrt(sum(diag(t(as.matrix(mat1)))%*
  %*%as.matrix(mat1))*sum(diag(t(as.matrix(mat2)
  ))%*%as.matrix(mat2))))}
> (RVs <- matrix(NA,nrow=6,ncol=6,
  dimnames=list(c("G","H","I","J","K",
  "N"),c("G","H","I","J","K","N"))))
> for(ii in 1:6){for(jj in 1:6){RVs[ii,jj] <-
  RV.comp(ktabbiovol3mscaled[[ii]],ktabbiovol3m
  scaled[[jj]])}}
```

We get, in a readable format for matrices:

	G	H	I	J	K	N
G	1.000	0.322	0.358	0.219	0.284	0.334
H	0.322	1.000	0.526	0.397	0.265	0.576
I	0.358	0.526	1.000	0.415	0.234	0.549
J	0.219	0.397	0.415	1.000	0.334	0.410
K	0.284	0.265	0.234	0.334	1.000	0.301
N	0.334	0.576	0.549	0.410	0.301	1.000

This is the same matrix as the one derived using the pta function of the ade4 package:

```
pta1 <- pta(ktabbiovol3mscaled, scan = FALSE,
nf=4) pta1$RV
```

	G	H	I	J	K	N
G	1.000	0.322	0.358	0.219	0.284	0.334
H	0.322	1.000	0.526	0.397	0.265	0.576
I	0.358	0.526	1.000	0.415	0.234	0.549
J	0.219	0.397	0.415	1.000	0.334	0.410
K	0.284	0.265	0.234	0.334	1.000	0.301
N	0.334	0.576	0.549	0.410	0.301	1.000

We now perform the eigendecomposition of the cosine matrix in order to reveal the structure between the matrices:

$$C = PDP'$$

with  $P'P = I$ ,

where P is the matrix of eigenvectors of C and  $D = \text{diag}(d_1, \dots, d_6)$  the diagonal matrix of the eigenvalues sorted by decreasing absolute values  $|d_1| \geq \dots \geq |d_6|$ .

An element of a given eigenvector represents the projection of one sub-matrix on this eigenvector. Thus the studies can be represented as points in the eigenspace and their similarities visually analyzed. The matrix of the projections is computed using the formula

$$G = PD^{1/2}$$

All the elements of the cosine matrix C are positive; as a consequence the six matrices are positively correlated with the first component. The coordinates of the first eigenvector share the same sign which we choose to be positive.

The following R code computes the P and D matrices. We compare the results with those of the pta built-in function and begin with the eigenvalues of the C matrix.

```
> eigendecomp <- eigen(x=RVs, symmetric=TRUE)
> (eigenvalues <- eigendecomp$values)
[1] 2.8840258 0.8624162 0.7877673 0.5845975
0.4637148 0.4174784
> pta1$RV.eig
[1] 2.8840258 0.8624162 0.7877673 0.5845975
0.4637148 0.4174784
> eigendecomp$vectors[,1]
[1] -0.3398160 -0.4546875 -0.4533865 -
0.3928325 -0.3156219 -0.4672157
> if(any(eigendecomp$vectors[,1] < 0))
```

```
eigendecomp$vectors[,1] <- -
eigendecomp$vectors[,1]
> eigendecomp$vectors[,1]
[1] 0.3398160 0.4546875 0.4533865 0.3928325
0.3156219 0.4672157
> D <- diag(eigenvalues)
> dimnames(D) <- list(c(paste("axis",
1:6, sep="")), c(paste("axis", 1:6, sep="")))
> D
```

	axis1	axis2	axis3	axis4	axis5	Axis6
axis1	2.884	0.000	0.000	0.000	0.000	0.000
axis2	0.000	0.862	0.000	0.000	0.000	0.000
axis3	0.000	0.000	0.788	0.000	0.000	0.000
axis4	0.000	0.000	0.000	0.585	0.000	0.000
axis5	0.000	0.000	0.000	0.000	0.464	0.000
axis6	0.000	0.000	0.000	0.000	0.000	0.417

```
> P <- eigendecomp$vectors
dimnames(P) <-
list(c("G", "H", "I", "J", "K", "N"),
c(paste("axis", 1:6, sep="")))
> P
```

	axis1	axis2	axis3	axis4	axis5	axis6
G	0.340	-0.303	0.795	0.360	0.171	0.050
H	0.455	0.300	-0.007	-0.338	0.537	-0.548
I	0.453	0.315	0.073	0.085	-0.781	-0.269
J	0.393	-0.084	-0.561	0.694	0.199	0.054
K	0.316	-0.809	-0.220	-0.397	-0.171	-0.101
N	0.467	0.240	-0.021	-0.330	0.060	0.782

We now compute the matrix G.

```
> P%*%sqrt(D)
```

	axis1	axis2	axis3	axis4	axis5	axis6
G	0.577	-0.281	0.705	0.275	0.116	0.032
H	0.772	0.279	-0.007	-0.258	0.366	-0.354
I	0.770	0.292	0.064	0.065	-0.532	-0.174
J	0.667	-0.078	-0.498	0.531	0.135	0.035
K	0.536	-0.751	-0.195	-0.304	-0.117	-0.066
N	0.793	0.222	-0.019	-0.253	0.041	0.505

```
> pta1$RV.coo
```

	IS1	IS2	IS3	IS4
G	0.577	0.281	0.705	-0.275
H	0.772	-0.279	-0.007	0.258
I	0.770	-0.292	0.064	-0.065
J	0.667	0.078	-0.498	-0.531
K	0.536	0.751	-0.195	0.304
N	0.793	-0.222	-0.019	0.253

Figure 5 displays these projections of the sampling stations onto the first and second components.



3.2.2. Construction of the compromise

The second step is the construction of the compromise and its analysis involves the building of a mean matrix of maximum inertia (referred to as the compromise matrix).

Interstructure

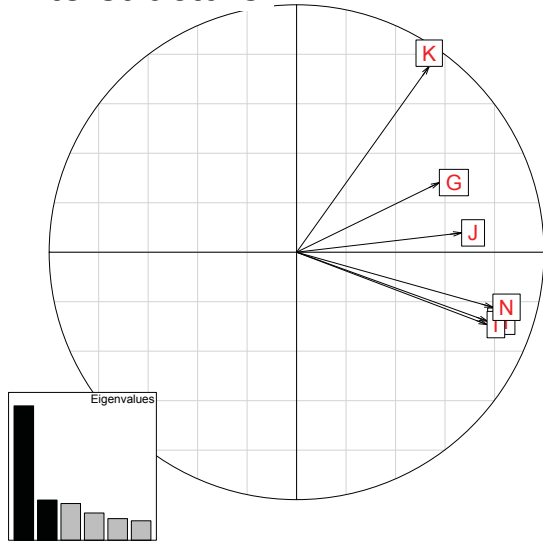


Figure 5. Temporal interstructure derived from each sampling station table and their closeness according to the Axis1 and Axis2 computed from the partial triadic analysis.

It is derived from the initial sub-matrices in proportion to the weight vector  $\alpha = (\alpha_G, \alpha_H, \alpha_I, \alpha_J, \alpha_K, \alpha_N)$  that is obtained from the PCA of the cosine matrix. The weights are obtained by re-scaling the elements of the first eigenvector of C so that the sum of their squares is equal to one. The C matrix is not centered, thus the first eigenvector of C highlights what is common to the different sub-matrices. As a consequence, sub-matrices with larger values on the first eigenvector are more similar to the other sub-matrices and therefore will have a larger weight.

```
> alpha <- eigendecomposition$vector[,1]/
sum(eigendecomposition$vector[,1]^2)
> names(alpha) <- c("G", "H", "I", "J", "K", "N")
> alpha
      G      H      I
0.3398160 0.4546875 0.4533865
      J      K      N
0.3928325 0.3156219 0.4672157
> tabw_ <- ptal$tabw
> names(tabw_) <- ptal$tab.names
> tabw_
      G      H      I
0.3398160 0.4546875 0.4533865
      J      K      N
0.3928325 0.3156219 0.4672157
```

The compromise matrix  $X^+$  is defined by:

$$X^+ = \alpha_G X_G + \alpha_H X_H + \dots + \alpha_N X_N.$$

It is a matrix of biovolume values with 112 columns – the species – and 11 rows – the sampling times – just as for the observed matrices  $X_G, X_H, X_I, X_J, X_K$  and  $X_N$ . The

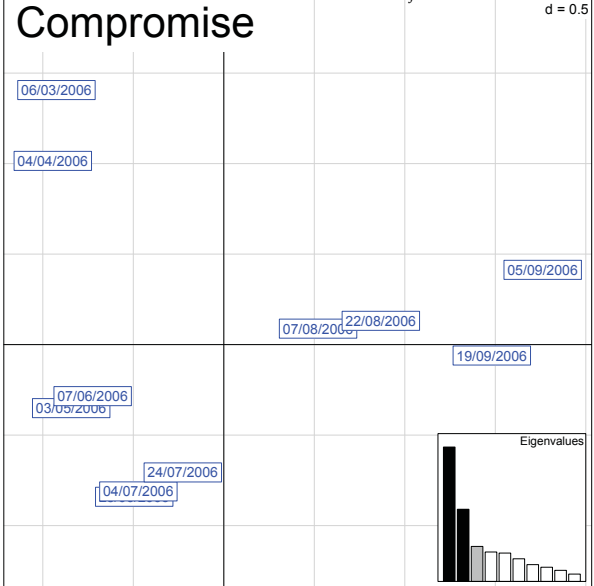


Figure 6. First and second axes of the compromise for the partial triadic analysis with the temporal evolution of the phytoplankton community.

actual value of the compromise matrix  $X^+$  was computed using the following R code and its transpose is given Appendix 1. See also Figures 6 and 7.

```
> Comp <- 0
> for(ii in 1:6){Comp<-Comp+alpha[ii]
+ *as.matrix(ktabbiovol3mscaled[[ii])
+ }}
```

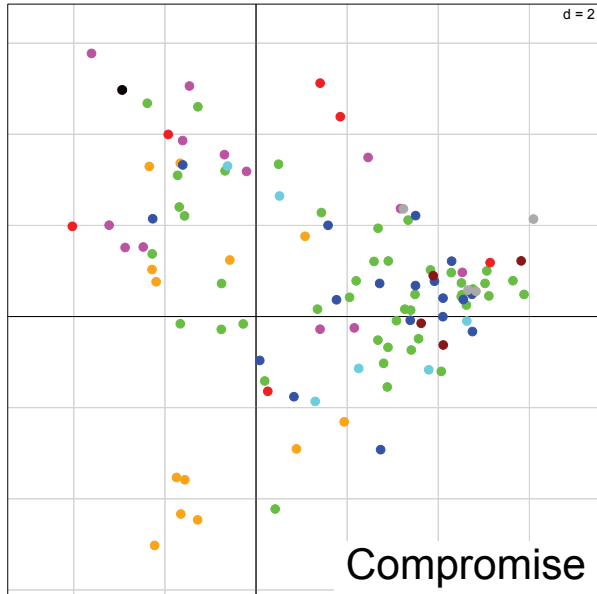
In the compromise matrix, a greater importance is thus given to stations which have a similar structure. This leads to the establishment of a common temporal typology shared by those stations; in other words, the compromise will express the spatially stable part of the annual structure.

3.2.3. How representative is the compromise matrix?

The compromise matrix  $X^+$  maximizes the inertia of the combined biovolume values matrices  $\alpha_G X_G + \alpha_H X_H + \dots + \alpha_N X_N$  under the constraint that  $\alpha_G^2 + \alpha_H^2 + \dots + \alpha_N^2 = 1$ .

It is the best aggregate of the original matrices that lies in

the blunt convex cone<sup>2</sup> spanned by the original observed matrices yet it is a desirable property to be able to assess the quality of this aggregate. A common way to achieve this aim is to compute:



**Figure 7.** First and second axes of the compromise for the partial triadic analysis with phytoplankton species colored by family and projected on the compromise of the partial triadic analysis.

$$\text{Quality of compromise} = \frac{d_1}{\text{tr}(D)} = \frac{d_1}{\text{tr}(C)}$$

```
> (eigenvalues/sum(eigenvalues))*100
[1] 48.067097 14.373604 13.129454 9.743292
7.728580 6.957973
> pta1$RV.eig/sum(pta1$RV.eig)*100
[1] 48.067097 14.373604 13.129454 9.743292
7.728580 6.957973
```

The quality of the compromise we just computed is approximately equal to .48. So we can say that the compromise “explains” 48% of the inertia of the original set of observed matrices.

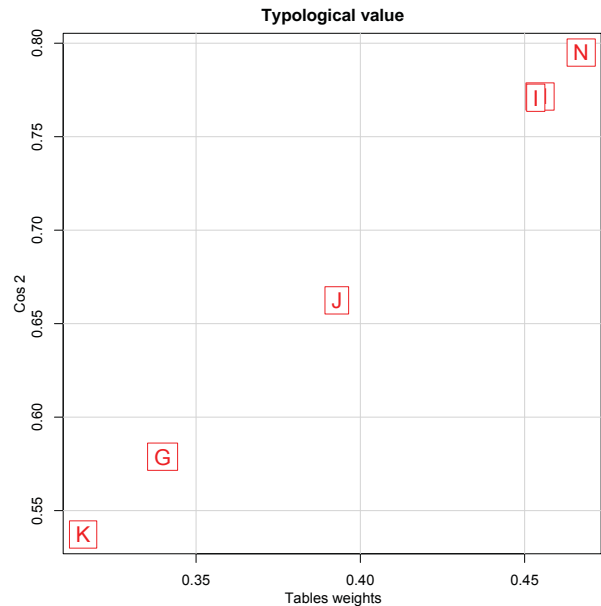
The  $\cos^2$  is an indicator of how much the compromise expresses the information contained in each sub-matrix. The following pta built-in function computes the  $\cos^2$  between the six sub-matrices and the compromise.

```
> pta1$cos2
      1      2      3
```

<sup>2</sup> A blunt convex cone is a subset of non-zero elements of a vector space that is closed under linear combinations with positive coefficients.

```
0.5788371 0.7714891 0.7708412
      4      5      6
0.6624583 0.5372276 0.7949918
```

Table weights and  $\cos^2$  for each of the observed original matrices can be plotted together as on Figure 8.



**Figure 8.** Table and  $\cos^2$  Weights.

### 3.2.4. Analyzing the compromise

We now proceed to the analysis of the compromise matrix  $X^+$ . The duality diagram to be used is  $(X^+, D_n=I_n, D_p=(1/p)I_p)$  which amounts to a non-centered non-scaled PCA of the transpose of the compromise matrix  $X^+$ . Since in our case  $112=p>n=11$ , we need to compute the eigendecomposition of the matrix  $(1/p)X^+X^+$ . Let  $D_{da}$  and  $P_{da}$  be this decomposition:

$$\frac{1}{112}X^+X^+ = P_{da}D_{da}P_{da}' \text{ with } k, k, j, m,$$

```
> Compde <- eigen(Comp%*%t(Comp))/112
> Compde$values[1:10]
[1] 9.8226486 5.2754266 2.5594261 2.1513154
2.0738068 1.6547172 1.2283484 1.0386681
0.8110492 0.5368458
> dudi.pca(t(Comp), scannf=FALSE, center =
FALSE, scale = FALSE)$eig
[1] 9.8226486 5.2754266 2.5594261 2.1513154
2.0738068 1.6547172 1.2283484 1.0386681
0.8110492 0.5368458
> pta1$eig
[1] 9.8226486 5.2754266 2.5594261 2.1513154
2.0738068 1.6547172 1.2283484 1.0386681
0.8110492 0.5368458
```

As claimed, we find the same eigenvalues for the compromise analysis with any of the three methods we

used.

```
> Compde$values[1:10] / sum(
+ Compde$values[1:10]) * 100
[1] 36.176184 19.429057 9.426202 7.923156
7.637697 6.094217 4.523928 3.825348
2.987042 1.977168
```

The first two axes of the decomposition amount to 56 % of the total inertia of the compromise. The row normed scores are given by the matrix  $P_{da}$  or `pta1$li`.

We now compute the coordinates of the sampling dates on the first three axes of the eigendecomposition of the compromise using the formula, see Table :

$$G_{da} = P_{da} D_{da}^{1/2}.$$

```
> Pda <- Compde$vectors[,1:3]
> Dda <- diag(Compde$values[1:3])
> Gda <- Pda %*% sqrt(Dda)
> dimnames(Gda) <- dimnames(pta1$li)
> Gda
```

	Axis1	Axis2	Axis3
06/03/2006	-0.925	1.406	0.151
04/04/2006	-0.945	1.015	-0.085
03/05/2006	-0.843	-0.351	0.170
07/06/2006	-0.725	-0.285	0.565
28/06/2006	-0.495	-0.841	0.254
04/07/2006	-0.472	-0.811	-0.062
24/07/2006	-0.227	-0.706	-0.471
07/08/2006	0.521	0.088	-0.944
22/08/2006	0.868	0.131	-0.617
05/09/2006	1.762	0.413	0.321
19/09/2006	1.481	-0.060	0.718

`pta1$li`

	Axis1	Axis2	Axis3
06/03/2006	-0.925	1.406	0.151
04/04/2006	-0.945	1.015	-0.085
03/05/2006	-0.843	-0.351	0.170
07/06/2006	-0.725	-0.285	0.565
28/06/2006	-0.495	-0.841	0.254
04/07/2006	-0.472	-0.811	-0.062
24/07/2006	-0.227	-0.706	-0.471
07/08/2006	0.521	0.088	-0.944
22/08/2006	0.868	0.131	-0.617
05/09/2006	1.762	0.413	0.321
19/09/2006	1.481	-0.060	0.718

The coordinates of the species on the first three axes of the eigendecomposition of the compromise are given by,

see Table :

$$F_{es} = X^+ P_{da}$$

```
> (Fes <- t(Comp)%*%Pda)
> pta1$co
```

The projection matrix  $M_{es}$  is thus given by:

$$M_{es} = P_{da}.$$

Note that the column normed scores are given by the matrix  $F_{es} D_{da}^{-1/2}$  or `pta1$c1`.

```
> (Fes%%solve(sqrt(Dda)))
> pta1$c1
```

Appendix 2 shows the coordinates of the species that are the result of these computations. One can find “extreme” species on both axes:

```
> which.min(Ges[,1])
Cryptomonas.rostratiformis
62
> res.PCAc[which.min(Ges[,1]),]
Dim.1 Dim.2
-0.76752258 0.08528029
> which.max(Ges[,1])
Trachelomonas.volvocina
111
> res.PCAc[which.max(Ges[,1]),]
Dim.1 Dim.2
0.52727273 0.38181818
> which.min(Ges[,2])
Dinobryon.divergens
50
> res.PCAc[which.min(Ges[,2]),]
Dim.1 Dim.2
0.09090909 -0.70909091
> which.max(Ges[,2])
Fragilaria.ulna.var.angustissim
96
> res.PCAc[which.max(Ges[,2]),]
Dim.1 Dim.2 Dim.3
-0.5871807 0.1100964 -0.3669879
```

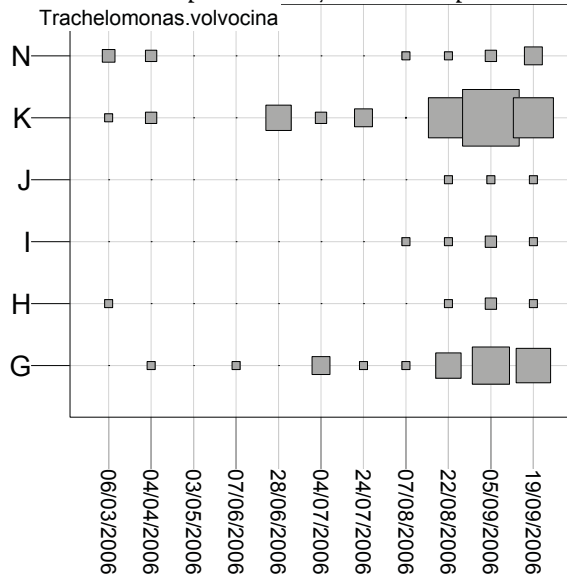
One can then easily produce Figures 9, 10, 11 and 12.

Some insights on the meaning of the two axes of the compromise can be found using the Kendall correlation coefficient to spotlight associations between the first two axes and the species, see Figures 13, 14, and 15 .

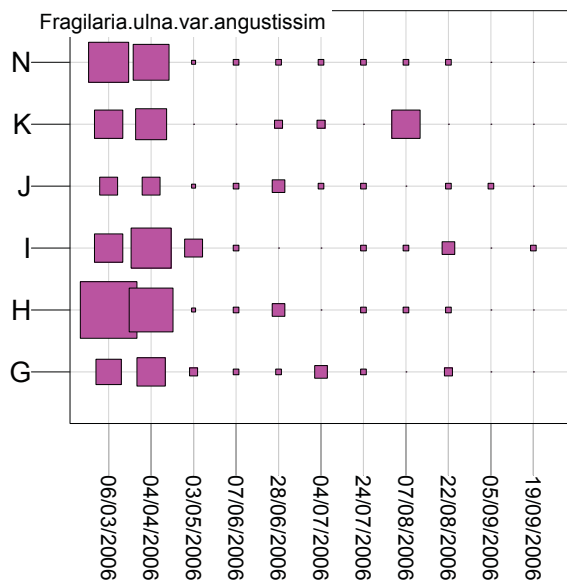
```
> library(SuppDists)
> qKendall(c(0.005,0.995),N=nrow(Gda))
[1] -0.5636364 0.5636364
> qKendall(c(0.025,0.975),N=nrow(Gda))
[1] -0.4545455 0.4545455
> res.PCAc <- Ges[,1:2]
> for(j in 1:2) {
+ for(k in Ges[[1]]) {
+ res.PCAc[k,j] <-cor(Comp[,k],
+ Gda[,j], method=c("kendall"))
+ }}
```

3.2.5. Highlighting the infrastructure

Finally, the last step is called the study of the infrastructure or of the reproducibility of the compromise. The



**Figure 9.** Biovolumes per station and dates of the phytoplankton specie lying on the outmost right on Figure 7.

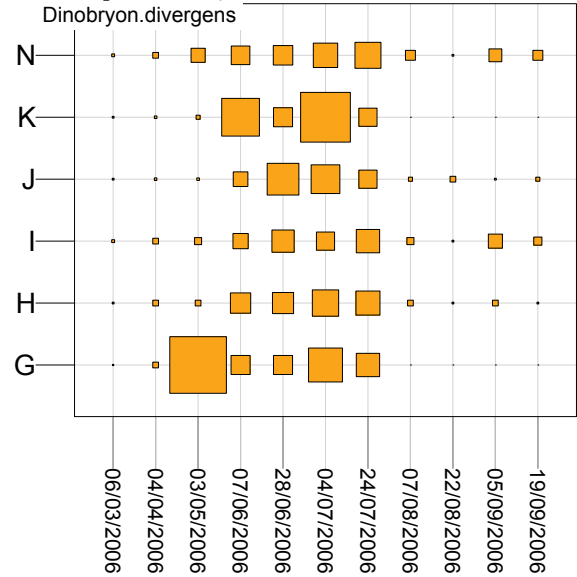


**Figure 10.** Biovolumes per station and dates of the phytoplankton specie lying on the outmost left on Figure 7.

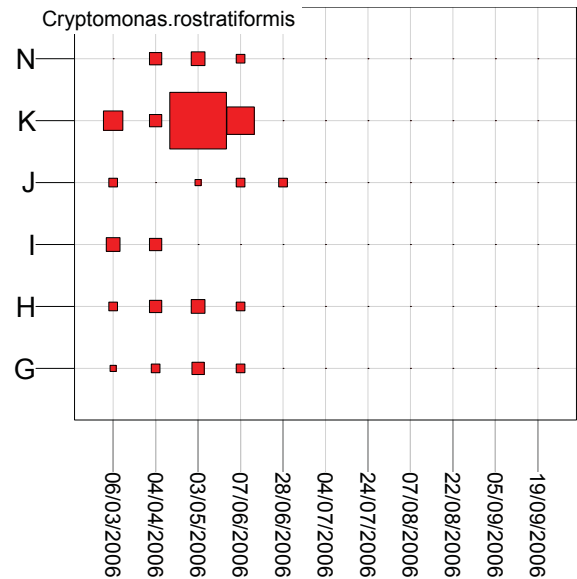
six original observed subsets are projected separately onto the compromise to highlight which station fits the compromise and whether some of them share the same patterns. This step can be summed up by putting together and comparing several figures, as shown in Figure 16.

From a mathematical point of view, one needs to

compute the projection of the original observed tables  $X_G, X_H, X_I, X_J, X_K$  and  $X_N$  on the compromise, by treating them as complementary tables in the non-centered non-



**Figure 11.** Biovolumes per station and dates of the phytoplankton specie lying lower left on Figure 7.



**Figure 12.** Biovolumes per station and dates of the phytoplankton specie lying upper left on Figure 7.

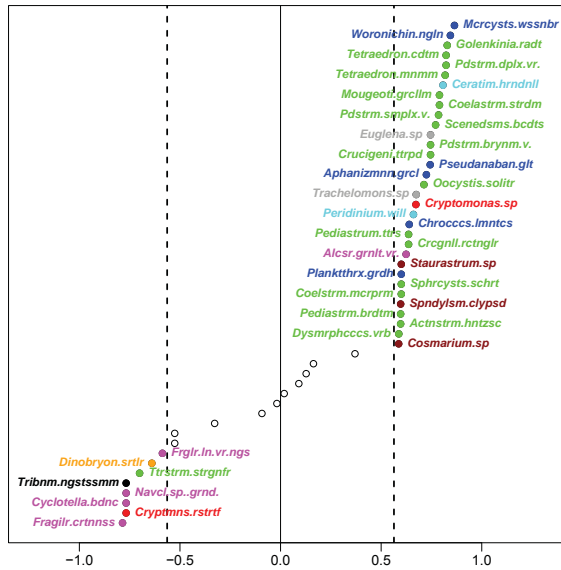
scaled PCA of the compromise. Table 1 teaches us how to deal with complementary individuals or variables.

As for the complementary rows, for instance the rows of the  $X_G$  matrix, we compute their coordinates using the formula:

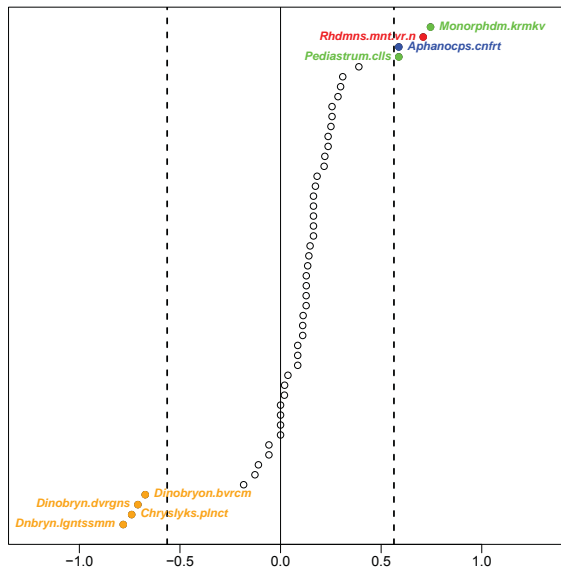
$$F_{es,N} = X_G 'P_{da} .$$

The same can be done for the rows of tables  $X_H, X_I, X_J,$

$X_K$  and  $X_N$ . The following R code deals with the rows of  $X_G$ .



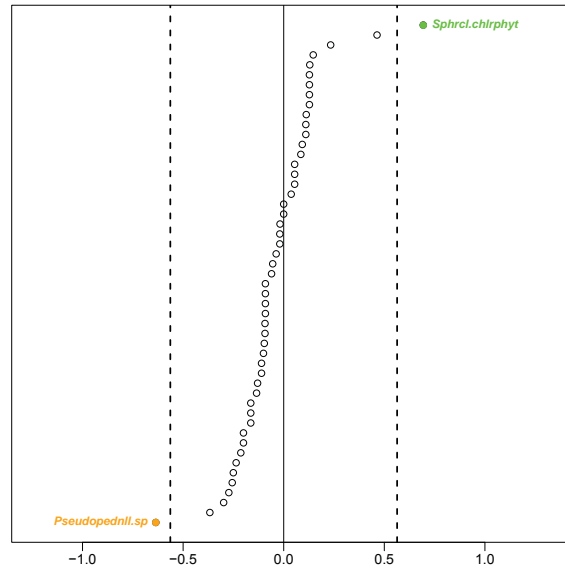
**Figure 13.** Significant Kendall correlations, at a 1% level, between the 1<sup>st</sup> axis of the compromise and the phytoplankton species biovolumes.



**Figure 14.** Significant Kendall correlations, at a 1% level, between the 2<sup>nd</sup> axis of the compromise and the phytoplankton species biovolumes.

```
> FGes <- t(ktabbiovol3mscaled[[
+ "G"]])%*%Pda
> FGes
> pta1$Tco[1:112,]
```

This leads to the six plots in the third column of Fig. 16.



**Figure 15.** Significant Kendall correlations, at a 1% level, between the 3<sup>rd</sup> axis of the compromise and the phytoplankton species biovolumes.

As for the complementary columns, for instance the columns of the  $X_G$  matrix, we compute their coordinates using the formula:

$$G_{es,N} = (1/p)X_G P_{da} D_{da}^{-1/2}.$$

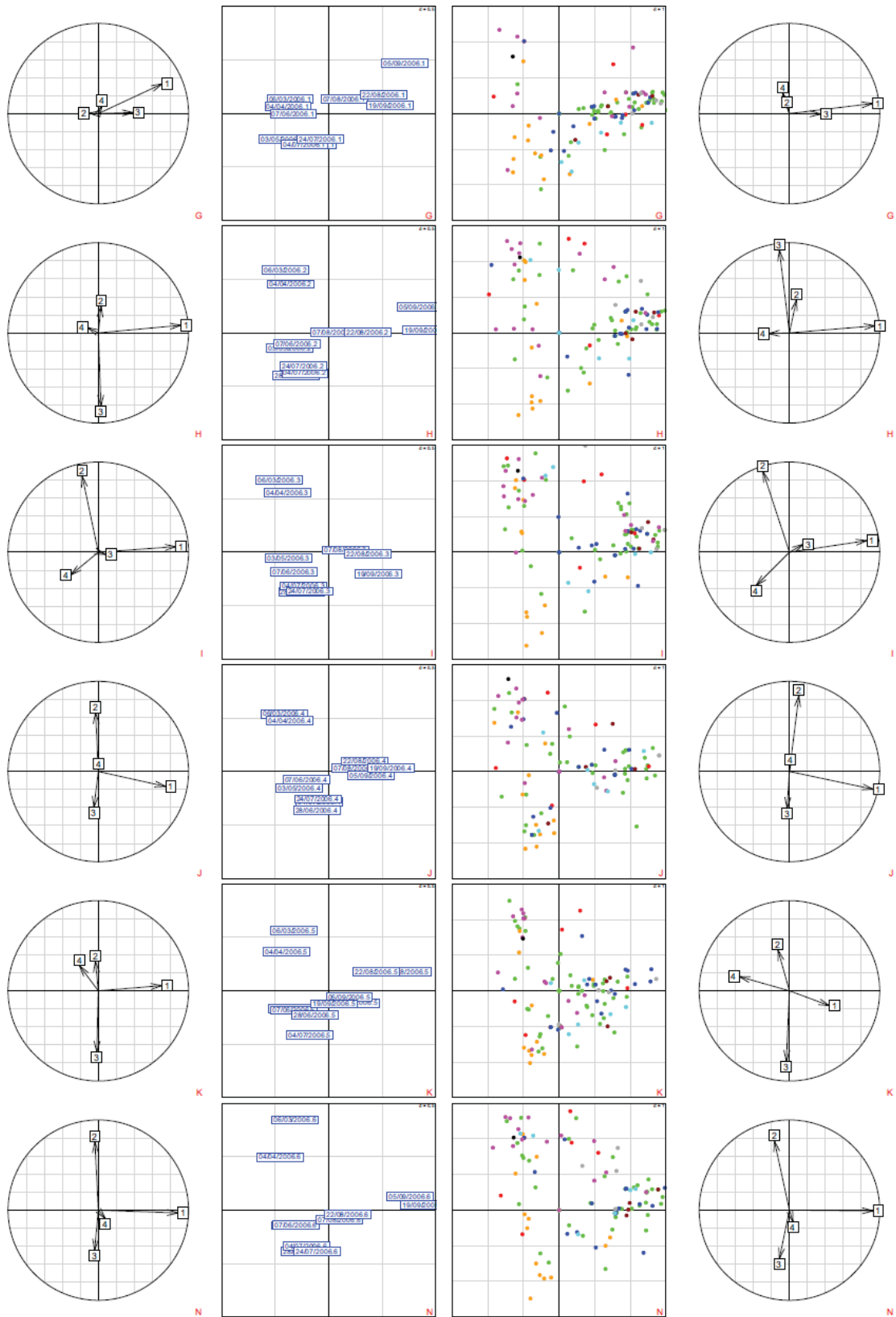
The R code to derive the coordinates of the sampling times on the first three axes of the compromise is:

```
> GdaG <- 1/112*as.matrix(
+ ktabbiovol3mscaled[["G"]])%*%Fes
+ %*%solve(sqrt(Dda))
> GdaG
> pta1$Tli[1:11,]
```

This leads to the six plots of the second column of Figure 16. The plots on the first and the fourth columns of Figure 16 are the projection on the first plane of principal axes of the compromise of the principal axes of each of the six matrices  $X_G$ ,  $X_H$ ,  $X_I$ ,  $X_J$ ,  $X_K$  and  $X_N$  and on the first plane of principal components of the compromise of the principal components of each of these matrices.

As a consequence, one needs to first perform a separate PCA analysis of the matrices  $X_G$ ,  $X_H$ ,  $X_I$ ,  $X_J$ ,  $X_K$  and  $X_N$ . The sepan function of the ade4 package is designed to perform this task. We provide now the R code for these

Figure 16. Reproducibility of the compromise for each of the six stations on the 1st and 2nd axes.



computations and compare the results with those of the `sepan` and `pta` functions.

```
> Gde <- eigen(as.matrix(
+ ktabbiovol3mscaled[["G"]])%*%
+ t(as.matrix(ktabbiovol3mscaled[["G"]]))/112)
> Gde$values[1:10]
> sepan(ktabbiovol3mscaled,
+ nf=3)$Eig[1:10]

> PdasepG <- Gde$vectors[,1:4]
> DdasepG <- diag(Gde$values[1:4])
> GdasepG <- PdasepG%*%sqrt(DdasepG)
> dimnames(GdasepG)[[1]] <-
+ dimnames(pta1$li)[[1]]
> dimnames(GdasepG)[[2]] <-
+ c(dimnames(pta1$li)[[2]], "Axis4")
> GdasepG
> sepan(ktabbiovol3mscaled,
+ nf=4)$Li[1:11,]
> PdasepG
> GdasepG%*%solve(sqrt(DdasepG))
> sepan(ktabbiovol3mscaled,
+ nf=4)$L1[1:11,]

> FessepG <- t(ktabbiovol3mscaled[["G"]])%*%PdasepG
> FessepG
> sepan(ktabbiovol3mscaled,
+ nf=4)$Co[1:112,]
> FessepG%*%solve(sqrt(DdasepG))
> sepan(ktabbiovol3mscaled,
+ nf=4)$C1[1:112,]
```

#### 4. Results

The three successive steps of the partial triadic analysis successfully revealed the (1) proximity between stations, (2) pattern of phytoplankton dynamics through the sampling times of the one-year study, and (3) stability and reproducibility of this pattern for each station of the lake.

Investigations began with the spatial homogeneity of the Reservoir Marne through time. A rather good interstructure was revealed by the partial triadic analysis: on the one hand, the arrows on Figure 5 head towards the same direction and on the other hand the  $\cos^2$  are above average values for each table (see Figure 8), which is tantamount to a good representation of each of the tables on the compromise table whose transpose is given in Appendix 1.

This interstructure enables us to look at the compromise, a common temporal structure derived from each sampling station table (see Figure 6 and Figure 7) and check for its reproducibility among the 6 sampling stations (G, H, I, J, K and N), **Error! Reference source not found.**

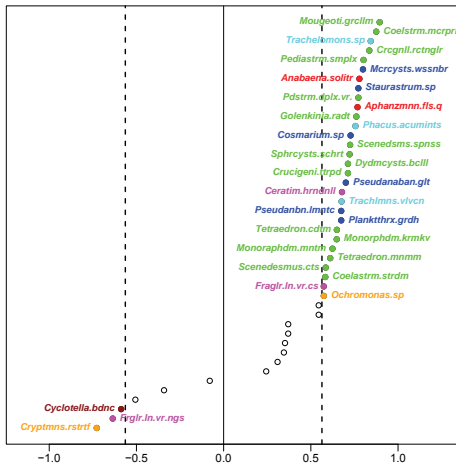
The links highlighted by the interstructure between stations are sensible from a biological point of view if one looks at the different features of the stations such as maximum depth, distance from the shore, stream velocity and so on, depicted in Section 1. As one can spot on the first two axes of the compromise showing the temporal evolution of the phytoplankton community (Figure 4), the phytoplankton community structure does not change between winter and spring. Then, after some time spent growing, the community structure undergoes some changes between late spring and summer.

The projection of variables, *i.e.* phytoplankton species, on the first two axes of the compromise, shown in Figure 7, enables us to spot associations between the species themselves. Additionally, the species were colored according to the phytoplankton family to which they belong, see Table 2 for the meaning of the colors, thus allowing the detection of a common pattern of behavior for a given phytoplankton family.

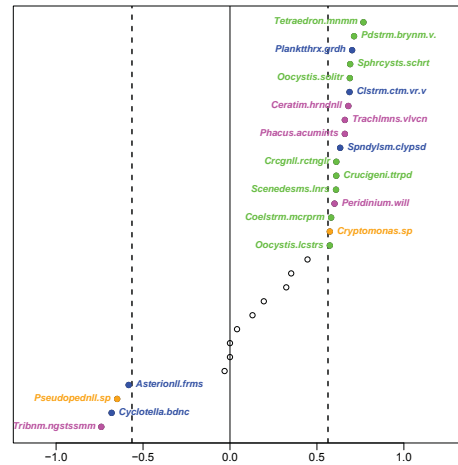
Comparing the coordinates of the species on the first two axes of the compromise, Appendix 2 and Figure 7, and the coordinates of the sampling times, Figure 6, highlights which ones of the species are specific to some season of the year. Such an analysis can be completed by looking at the values of the biovolumes per station and dates which can be easily plotted for any of the species. It can be done for all of the species and is reported for four of them, “*Trachelomonas volvocina*”, “*Fragilaria ulna var angustissim*”, “*Dinobryon divergens*” and “*Cryptomonas rostratiformis*”, in Figures 9-12.

A more precise account on the links between the biovolume of the species and their factorial coordinates on the first three axes of the compromise was derived using Kendall correlation coefficients. These correlation coefficients were plotted in Figures 13-15 for the species for which the association was significant at a 1% level with any of the three axes. Limits of significance, at a 1% level, were reported on the graphics and whenever an association was found to be significant, at a 1% level, between the biovolume of one of the species and an axis, the name of the specie was added to the graph.

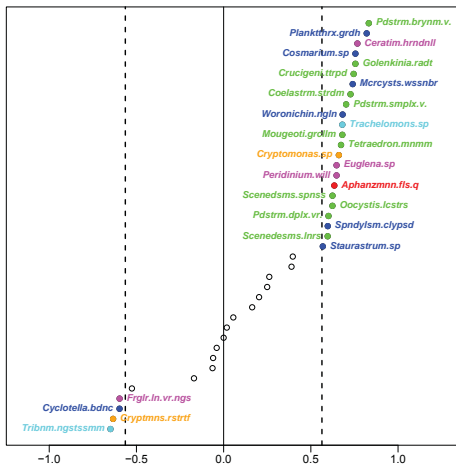
Two of the four previously quoted species, “*Cryptomonas rostratiformis*” (*Cryptmns.rstrtf*) and “*Fragilaria ulna var angustissim*” (*Frglr.ln.vr.ngs*) show a significant, at a 1% level, association with the first axis. Additionally, the “*Trachelomonas volvocina*” shows a significant, at a 5% level, association with the first axis. “*Dinobryon divergens*” shows a significant, at a 1% level, association with the second axis. The previous correlation analysis can also be performed between the biovolume of the



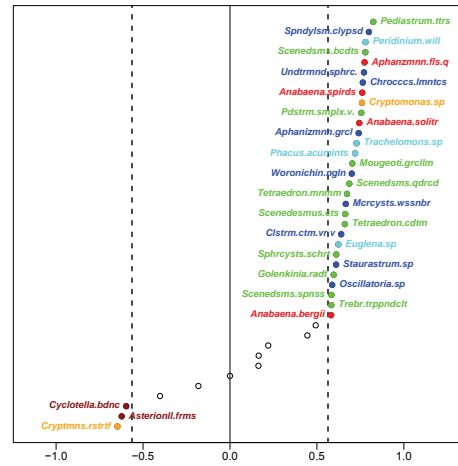
**Figure 17.** Significant Kendall correlations, at a 1% level, between the 1<sup>st</sup> axis of the G station and the phytoplankton species.



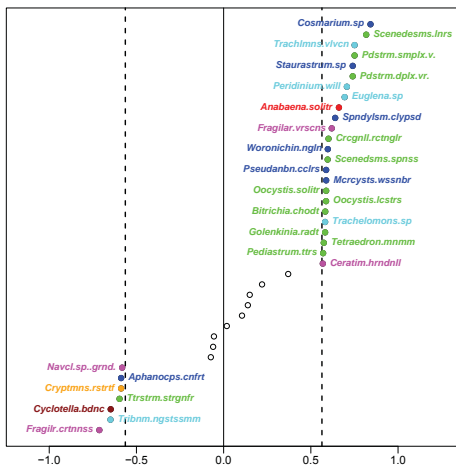
**Figure 20.** Significant Kendall correlations, at a 1% level, between the 1<sup>st</sup> axis of the J station and the phytoplankton species.



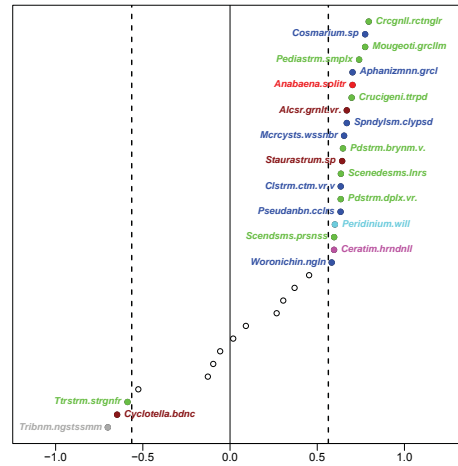
**Figure 18.** Significant Kendall correlations, at a 1% level, between the 1<sup>st</sup> axis of the H station and the phytoplankton species.



**Figure 21.** Significant Kendall correlations, at a 1% level, between the 1<sup>st</sup> axis of the K station and the phytoplankton species.



**Figure 19.** Significant Kendall correlations, at a 1% level, between the 1<sup>st</sup> axis of the I station and the phytoplankton species.



**Figure 22.** Significant Kendall correlations, at a 1% level, between the 1<sup>st</sup> axis of the N station and the phytoplankton species.



phytoplankton species and each of the station during the step of analysis of the reproducibility of the compromise among the 6 sampling stations using Figures 17-22. Such an analysis helps us to spot patterns of associations between species, stations and dates helping freshwater scientists identify species that should be good markers of a “good ecological state”.

## References

- Abdi, H. and Valentin, D. 2007. The STATIS Method. In: Salkind, N. (Ed.). *Encyclopedia of Measurement and Statistics*. Thousand Oaks: Sage.
- Aliaume, C., Monteiro, C., Louis, M., Lam Hoai, T., and Lasserre, G. 1993. Organisation spatio-temporelle des peuplements ichthyologiques dans deux lagunes côtières: au Portugal et en Guadeloupe. *Oceanologica Acta* 16 :291–301.
- Anneville, O., Ginot, V., Druart, J.C., Angeli, N., 2002. Long-term study (1974–1998) of seasonal changes in the phytoplankton in Lake Geneva: a multi-table approach. *Journal of Plankton Research* 24, no. 10:993–1007.
- Bertrand, F., Maumy, M., Fussler, L., Kobes, N., Savary, S. and Grosman, J. 2007. Using Factor Analyses to Explore Data Generated by the National Grapevine Wood Diseases Survey. *Case Studies in Business, Industrial or Governmental Statistics* 1, no. 2:183–202.
- Blanc, L. and Beaudou, D. 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., Chessel, D. and Dolédec, S. 1998. Étude de la stabilité temporelle des structures spatiales par analyses d’une série de tableaux de relevés faunistiques totalement appariés. *Bulletin Français de la Pêche et de la Pisciculture* 348 :1–21.
- Cadet, P., Masse, D. and Thioulouse, J. 2005. Relationships between plant-parasitic nematode community, fallow duration and soil factors in the Sudano-Sahelian area of Senegal. *Agriculture, Ecosystems and Environment* 108:302–317.
- Carassou, L. and Ponton, D., 2007. Spatio-temporal structure of pelagic larval and juvenile fish assemblages in coastal areas of New Caledonia, southwest Pacific *Marine Biology* 150, no. 4:697–711.
- Centofanti, M., Chessel, D. and Dolédec, S. 1989. Stability of a spatial structure and compromise in a multiway matrices statistical analysis: application to physical and chemical properties in a reservoir lake. *Journal of Water Science* 2: 71–93.
- In a word, the commonly used classification of phytoplankton species does not appear to be relevant to characterize either the season or the features of the stations. The use of the partial triadic analysis provides an alternative solution by showing significant associations.
- Chaya, C., Perez-Hugalde, C., Judez, L., Wee, C.S. and Guinard, J.X. 2004. Use of the STATIS method to analyze time-intensity profiling data. *Food Quality and Preference*, 15 no. 1: 3–12.
- Chessel, D. and Hanafi, M. 1996. Analyses de la co-inertie de K nuages de points. *Revue de Statistique Appliquée* 44: 35–60.
- Chessel, D. Dufour, A.-B. and Thioulouse, J. 2004. The ade4 package I: One-table methods, *RNews* 4: 5–10.
- Coquet, R., Troxler, L. and Wipff, G. 1996. The STATIS method: Characterization of conformational states of flexible molecules from molecular dynamics simulation in solution. *Journal of Molecular Graphics* 14: 206–212.
- Dolédec, S. 1988. Les analyses multi-tableaux en écologie factorielle. II Stratification longitudinale de l’Ardèche à partir des descripteurs physico-chimiques. *Acta OEcologica, OEcologia Generalis* 9: 119–135.
- Dolédec, S. and Chessel, D., 1991. Recent developments in linear ordination methods for environmental sciences, *Advances in Ecology* 1: 133-155.
- Dray, S., Dufour, A.-B. and Chessel, D., 2007. The ade4 package II: Two-table and K-table methods, *RNews*, 7: 47–52.
- Ernoul, A., Freire-Diaz, S., Langlois, E. and Alard, D., 2006. Are similar landscapes the result of similar histories? *Landscape Ecology* 21: 631–639.
- Escoufier, Y. 1973. Le traitement des variables vectorielles. *Biometrics* 29: 750–760.
- Escoufier, Y., 1980. L’analyse conjointe de plusieurs matrices de données. In: Jolivet et al., eds., *Biométrie et Temps*, Paris: Société Française de Biométrie. 59–76.
- Escoufier, Y., 1987. The duality diagram: a means for better practical applications. In: Legendre P. & Legendre L., eds., *Development in numerical ecology*, Nato advanced Study Institute Serie G (Ecological Sciences), Berlin: Springer-Yerlag.
- Gaertner, J.C., 2000. Seasonal organization patterns of demersal assemblages in the Gulf of Lions (north-western Mediterranean Sea). *Journal of the Marine Biological Association of the United Kingdom*, 80 no. 5:

- 777–783.
- Gaertner, J.C., Chessel, D. and Bertrand, J., 1998. Stability of spatial structures of demersal assemblages: a multitable approach. *Aquatic Living Resources*, 11 no. 2: 75–85.
- Gaillard, I., Durbec, J.P., Beliaeff, B. and Sabatier, R. 2003. Phytoplankton Ecology along French coasts: inter-site comparison. *Comptes Rendus Biologies*, 326 no. 9: 853–863.
- Holmes, S. 2008. Multivariate data analysis: The French way. In Deborah Nolan and Terry Speed, eds., *Probability and Statistics: Essays in Honor of David A. Freedman*. Institute of Mathematical Statistics, Beachwood, 219–233.
- Jaffrenou, P. A., 1978. *Sur l'analyse des familles finies de variables vectorielles. Bases algébriques et application à la description statistique*. Thèse 3<sup>ème</sup> cycle. Université des Sciences et Technique du Languedoc, Montpellier-II.
- Jiménez, J.J., Decaens, T. and Rossi, J.P., 2006. Stability of the spatiotemporal distribution and niche overlap in neotropical earthworm assemblages. *Acta Oecologica – International Journal of Ecology* 30: 299–311.
- Kherif, F., Poline, J.-P., Mériaux, S., Benali, H., Flandin, G. and Brett, M., 2003. Group analysis in functional neuroimaging: selecting subjects using similarity measures. *NeuroImage* 20: 2197–2208.
- Kiers H. A. L. 1991. Hierarchical relations among three-way methods, *Psychometrika* 56: 449–470.
- Korth B. and Tucker L.R. 1976. Procrustes matching by congruence coefficients. *Psychometrika* 41:531–535.
- Kroonenberg, P. M. 1989. The analysis of multiple tables in factorial ecology. III. Three mode principal component analysis: analyse triadique complète, *Acta oecologica. Oecologia generalis*, 10 no. 3: 245–256.
- L'Hermier des Plantes H., 1976. *Structure of Tables with Three Indices. Theory and Application of a Joint Analysis Method*. Thèse 3<sup>ème</sup> cycle. Université des Sciences et Technique du Languedoc, Montpellier-II.
- Lavit Ch. 1988. *Analyse Conjointe de Tableaux Quantitatifs*, Masson, Paris.
- Lavit Ch., Escouffier Y., Sabatier R. and Traissac P. 1994. The ACT (STATIS method), *Computational Statistics and Data Analysis*, 18:97–119.
- Lazraq, A., Hanafi M., Cléroux, R., Allaire, J. and Lepage, Y. 2008. Une approche inférentielle pour la validation du compromis de la méthode STATIS, *Journal de la société française de statistique*, 149 no. 1 :97-109
- Leibovici, B. 1993. *Facteurs à mesures répétées et analyses factorielles : application à un suivi épidémiologique*. Thèse de doctorat, Université Montpellier II.
- Lekve, K., Stenseth, N. Chr., Gjøsæter, J. and Dolédec, S. 2002. Species richness and environmental conditions of fish along the Norwegian Skagerrak coast. *ICES Journal of Marine Science*, 59 no. 4:757–769.
- Licandro, P. and Ibanez, F. 2000. Changes of zooplankton communities in the Gulf of Tigullio (Ligurian Sea, Western Mediterranean) from 1985 to 1995. Influence of hydroclimatic factors. *Journal of Plankton Research*, 22 no. 12:2225–2253.
- Lobry, J., Lepage, M. and Rochard, E. 2006. From seasonal patterns to a reference situation in an estuarine environment: example of the small fish and shrimp fauna of the Gironde estuary (SW France). *Estuarine, Coastal and Shelf Science*, 70 no.1–2:239–250.
- Martin, N., Molimard, P., Spinnler, H.E. and Schlich, P. 2000. Comparison of odour sensory profiles performed by two independent trained panels following the same descriptive analysis procedures. *Food Quality and Preference*, 11 no. 6: 487–495.
- Meyners, M. 2002. Methods to analyse sensory profiling data – a comparison. *Food Quality and Preference*, 14 no. 5–6: 507–514.
- Meyners, M., Kunert, J. and Qannari, E.M. 2000. Comparing generalized procrustes analysis and STATIS. *Food Quality and Preference*, 11: 77–83.
- Muiño, R., Carrera, P., Petitgas, P., Beare, D.J., Georgakarakos, S., Haralambous, J., Iglesias, M., Liorzou, B., Massé, J. and Reid, D.G. 2003. Consistency in the correlation of school parameters across years and stocks. *ICES Journal of Marine Science*, 60 no. 1: 164–175.
- Pavoine, S., Blondel J., Baguette, M. and Chessel, D. 2007. A new technique for ordering asymmetrical three-dimensional data sets in ecology. *Ecology*, 88 no. 2: 512–523.
- Perez, Aparicio J., Toledano, M. and Lafuente, R. 2007. Descriptive sensory analysis in different classes of orange juice by a robust free-choice profile method. *Analytica Chimica Acta*, 595 no. 1–2: 238–247.
- Perez-Hugalde, C., Judez, L., Litago, J., Yuste, J. and Fuentes-Pila, J. 2004. Statistical procedure for clonal preselection of *Vitis vinifera* L. cv. Tempranillo in the Duero Valley, Spain. *American Journal of Enology and Viticulture*, 55 no. 4: 335–345.
- Perrin, L., Symoneaux, R., Maitre, I., Asselin, C., Jourjon, F. and Pages, J. 2008. Comparison of three sensory Mapping (R) procedure: case of methods for use with the ten wines from Loire valley. *Food Quality and Preference*, 19 no. 1: 1–11.
- Qannari, E.M., Wakeling, I. and MacFie, H.J.H. 1995. A hierarchy of models for analyzing sensory data. *Food*

- and *Quality Preference* 6 no. 4: 309–314.
- R Development Core Team 2009. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN: 3-900051-07-0, url: <http://www.R-project.org>.
- Robert, P. and Escoufier, Y. 1976. A unifying tool for linear multivariate statistical methods: The RV-coefficient, *Applied Statistics*, 25: 257–265.
- Rolland, A., Bertrand, F., Maumy, M. and Jacquet, S. 2009. Assessing phytoplankton structure and spatio-temporal dynamics in a freshwater ecosystem using a powerful multiway statistical analysis, *Water Research*, 43 no. 13: 3155–3168.
- Rossi, J.P., 2003. The spatio-temporal pattern of a tropical earthworm species assemblage and its relationship with soil structure. *Pedobiologia*, 47 no. 5–6: 497–503.
- Schlich, P. 1996. Defining and validating assessor compromises about product distances and attribute correlations. In: Naes, T., Risvik, E. (Eds.), *Multivariate Analysis of Data Sensory Sciences*. Elsevier, New York, 229–306.
- Simier, M., Blanc L., Aliaume, C., Diouf, P.S. and Albaret J.J. 2004. Spatial and temporal structure of fish assemblages in an “inverse estuary”, the Sine Saloum system (Senegal). *Estuarine, Coastal and Shelf Science*, 59 no. 1: 69–86.
- Simier, M., Blanc L., Pellegrin, F. and Nandris, D. 1999. Approche simultanée de K couples de tableaux : application à l'étude des relations pathologie végétale – environnement, *Revue de statistique appliquée*, 47 no. 1: 31–46.
- Stanimirova, I., Walczak, B., Massart, D.L., Simeonov, V., Sabyd C.A. and di Crescenzo, E. 2004. STATIS, a three-way method for data analysis. Application to environmental data. *Chemometrics and Intelligent Laboratory Systems* 73: 219–233.
- Thioulouse, J. and Chessel, D. 1987. Les analyses multitableaux en écologie factorielle. I. De la théorie d'état à la typologie de fonctionnement par l'analyse triadique, *Acta oecologica. Oecologia generalis*, 8: 463–480.
- Thioulouse, J., Simier M. and Chessel, D. 2004. Simultaneous analysis of a sequence of paired ecological tables, *Ecology*, 85 no. 1: 272–283.
- Tucker, L. R., 1966. Some mathematical notes on three-mode factor analysis, *Psychometrika*, 31: 279–311.

Correspondence: [fbertran@math.unistra.fr](mailto:fbertran@math.unistra.fr).

## Appendix 1. Transpose of the compromise matrix

	06/03	04/04	03/05	07/06	28/06	04/07	24/07	07/08	22/08	05/09	19/09
<i>Actinastrum.hantzschii</i>	-0.24	-0.24	-0.24	-0.24	-0.24	-0.24	-0.24	-0.23	-0.24	0.89	1.27
<i>Ankyra.lanceolata</i>	3.44	0.46	0.48	-0.88	-0.88	-0.88	-0.88	-0.88	-0.88	0.28	0.63
<i>Pediastrum.cells</i>	2.69	2.04	-0.53	-0.53	-0.53	-0.53	-0.53	-0.53	-0.53	-0.53	-0.53
<i>Chlamydomonas.sp</i>	0.68	2.06	-1.05	3.04	-1.05	0.57	-0.70	-1.08	-1.08	-0.32	-1.08
<i>Chlorella.vulgaris</i>	-0.48	1.34	-2.59	-1.56	-0.68	-1.00	-0.09	3.55	-0.73	1.86	0.39
<i>Undetermined.chlorophytes</i>	3.45	2.49	-1.55	-1.96	-1.82	-0.98	0.29	2.50	-0.06	-0.83	-1.54
<i>Oval.chlorophytes</i>	-1.19	1.94	-2.72	2.37	3.04	-0.68	-0.71	-1.08	-0.86	1.02	-1.15
<i>Spherical.chlorophytes</i>	-1.28	-1.28	-1.28	1.52	1.21	-1.28	-1.28	-1.28	-1.28	3.61	2.65
<i>Coelastrum.astroideum</i>	-1.09	-1.09	-1.09	-1.09	-0.47	-0.89	-0.41	-0.50	1.55	3.80	1.29
<i>Coelastrum.microporum</i>	-0.95	-0.86	-0.92	-0.89	-0.95	-0.85	0.10	0.15	-0.17	5.83	-0.49
<i>Coelastrum.reticulum</i>	-0.64	-0.64	-0.64	-0.64	-0.64	-0.64	-0.64	-0.64	0.49	-0.64	5.23
<i>Crucigenia.tetrapedia</i>	-1.12	-1.33	-1.34	-1.28	-1.24	-1.27	-0.64	0.84	0.61	3.43	3.34
<i>Crucigeniella.rectangularis</i>	-1.32	-1.17	-1.32	-1.32	-1.16	-1.30	0.54	-0.10	3.35	1.95	1.83
<i>Dictyosphaerium.pulchellum</i>	0.74	0.86	-0.77	-0.77	0.32	0.67	-0.41	0.35	0.34	-0.77	-0.56
<i>Dydimocystis.bicellularis</i>	-1.17	2.50	-1.44	-1.41	-0.81	-1.01	-1.32	-0.95	0.15	2.19	3.27
<i>Dysmorphococcus.variabilis</i>	-0.64	-0.64	-0.64	-0.64	-0.64	-0.64	-0.64	-0.64	-0.64	5.23	0.49
<i>Eiakatothrix.gelatinosa</i>	-1.23	1.89	1.56	0.21	-0.21	0.44	-0.94	0.60	-0.36	-1.14	-0.81
<i>Little.green.flagellates</i>	2.45	4.90	-1.35	-1.05	-0.53	-0.41	-1.20	-0.21	-1.18	0.06	-1.48
<i>Golenkinia.radiata</i>	-1.31	-1.31	-1.31	-1.31	-1.11	-0.20	-0.70	1.47	1.53	4.37	-0.12
<i>Hyaloraphidium.contortum</i>	-0.05	3.78	0.24	0.97	-0.22	-0.65	-1.79	-0.66	-1.42	0.81	-1.02
<i>Korshikoviella.sp</i>	-0.10	-0.10	0.95	-0.10	-0.10	-0.10	-0.10	-0.10	-0.10	-0.10	-0.10
<i>Monoraphidium.kormakovae</i>	0.01	3.61	-1.42	-0.35	-1.13	-1.44	-1.05	0.30	0.89	0.92	-0.34
<i>Monoraphidium.minutum</i>	-2.14	2.59	-1.97	-2.03	0.29	0.19	-0.68	0.79	1.40	1.43	0.12
<i>Mougeotia.gracillima</i>	-0.96	-0.96	-0.96	-0.96	-0.96	-0.96	-0.79	1.09	-0.84	2.07	4.21
<i>Oocystis.lacustris</i>	-1.70	-1.69	-1.61	1.40	-0.19	-0.98	-0.90	2.12	0.49	1.29	1.76
<i>Oocystis.solitaria</i>	-1.22	-1.22	-1.22	-1.22	-0.10	0.28	1.33	0.58	0.13	2.13	0.53
<i>Pediastrum.biradiatum</i>	-0.44	-0.44	-0.44	-0.44	-0.44	-0.44	-0.44	-0.44	1.70	1.37	0.49
<i>Pediastrum.boryanum.var.longico</i>	-1.78	-1.61	-1.77	-0.92	0.20	-0.52	0.72	1.47	0.14	1.79	2.28
<i>Pediastrum.duplex.var.gracillim</i>	-1.32	-1.32	-1.32	-1.31	-0.65	-0.64	-1.30	0.06	2.90	2.80	2.10
<i>Pediastrum.simplex</i>	-1.67	-0.52	-1.68	-1.68	-1.45	-1.46	1.28	1.43	0.81	4.34	0.60
<i>Pediastrum.simplex.var.sturmii</i>	-1.30	-1.30	-1.30	-1.01	-1.29	-1.02	-0.98	0.63	4.17	1.36	2.05
<i>Pediastrum.tetras</i>	-0.62	-1.53	-1.60	-0.80	-0.24	-0.42	-0.66	0.90	3.80	1.06	0.13
<i>Phacotus.lendneri</i>	-1.90	-1.70	-1.84	-1.63	3.21	4.14	0.01	0.19	0.55	-1.06	0.04
<i>Scenedesmus.acutus</i>	-0.97	0.76	-0.64	-1.24	-0.98	-1.27	-0.85	2.56	0.19	0.50	1.93
<i>Scenedesmus.bicaudatus</i>	-0.88	-1.05	-0.94	-0.86	1.27	-0.72	-0.22	-0.22	0.69	2.09	0.84
<i>Scenedesmus.linearis</i>	-1.44	-1.41	-1.52	-1.47	-1.35	2.78	1.30	0.55	-0.31	2.46	0.40
<i>Scenedesmus.opoliensis</i>	2.36	-1.00	-1.13	-1.06	-1.00	-0.31	-0.29	0.66	0.37	1.96	-0.55
<i>Scenedesmus.parisiensis</i>	-1.02	-1.02	-0.98	-1.02	-0.99	-1.02	-0.86	2.87	1.95	-1.02	3.11
<i>Scenedesmus.quadricauda</i>	-1.04	-1.16	-1.41	-1.19	-0.74	-1.28	-1.21	1.24	1.80	3.29	1.70
<i>Scenedesmus.spinossus</i>	-0.96	-1.14	-1.10	-1.05	-1.12	-1.09	-1.10	0.65	2.98	1.47	2.47
<i>Sphaerocystis.schroeteri</i>	-1.37	-1.16	-1.06	-0.42	-1.31	-0.96	-0.77	2.83	3.30	1.44	-0.51
<i>Tetraedron.caudatum</i>	-1.19	-1.19	-1.19	-0.16	-0.01	0.04	-0.36	0.15	0.96	1.41	1.54
<i>Tetraedron.minimum</i>	-1.23	-1.31	-1.30	-1.28	-1.26	-1.24	-0.65	-0.42	1.05	2.52	5.11
<i>Tetraedron.triangularare</i>	-0.77	-0.77	-0.77	0.07	-0.77	2.40	0.07	1.80	-0.77	-0.77	0.28
<i>Tetrastrum.staurogeniaforme</i>	1.16	2.48	0.22	-0.48	-0.48	-0.48	-0.48	-0.48	-0.48	-0.48	-0.48
<i>Treubaria.triappendiculata</i>	-0.46	-0.09	-1.07	-0.89	-1.07	0.26	-1.07	-0.35	2.06	2.94	-0.27
<i>Bitrichia.chodatii</i>	-1.31	-1.49	-1.49	-1.49	0.66	2.38	0.94	-0.56	1.01	0.54	0.84
<i>Chrysolykos.planctonicus</i>	-0.95	-0.95	-0.95	-0.46	4.47	2.50	0.10	-0.95	-0.95	-0.95	-0.95
<i>Dinobryon.bavaricum</i>	-1.28	-1.26	-0.27	-1.26	2.04	3.11	3.42	-0.84	-1.24	-1.24	-1.17
<i>Dinobryon.divergens</i>	-1.79	-1.62	-0.12	1.01	2.19	3.57	2.44	-1.45	-1.78	-0.99	-1.46
<i>Dinobryon.elegantissimum</i>	-1.49	-1.47	-0.74	-1.14	4.42	0.13	3.93	-0.16	-0.87	-1.43	-1.18
<i>Dinobryon.sertularia</i>	0.68	1.55	-0.81	3.15	-0.83	0.43	-0.83	-0.83	-0.83	-0.83	-0.83
<i>Dinobryon.sociale.var.americanu</i>	2.05	-0.84	0.11	3.51	-0.84	-0.84	0.21	-0.84	-0.84	-0.84	-0.84
<i>Dinobryon.sociale.var.stipitatu</i>	-1.30	-1.19	-0.11	-0.99	0.55	0.74	2.81	-1.27	-1.21	-1.10	3.08
<i>Erkenia.subaquaeciliata</i>	1.05	4.42	-2.19	-1.32	-0.32	0.11	-0.77	2.20	-0.95	-0.64	-1.58
<i>Kephyrion.sp</i>	-0.54	-1.22	-0.95	0.10	5.02	2.62	-0.64	-1.66	-1.29	-0.64	-0.79

## Appendix 2. Species coordinates on the first three axes of the compromise

	CS1	CS2	CS3		CS1	CS2	CS3
<i>Actinastrum.hantzschii</i>	-1.3488	-0.1638	-0.8958	<i>Ochromonas.sp</i>	0.5744	-1.2422	-1.3352
<i>Ankyra.lanceolata</i>	0.6799	-3.1994	-1.3917	<i>Pseudopedinella.sp</i>	2.3420	-3.2963	1.9237
<i>Pediastrum.cells</i>	1.7210	-3.1014	-0.1682	<i>Salpingoeca.frequentissima</i>	-1.0769	-1.7631	-1.7290
<i>Chlamydomonas.sp</i>	2.2789	-1.3767	-1.4379	<i>Cryptomonas.marsonii</i>	-0.2586	1.6385	2.9978
<i>Chlorella.vulgaris</i>	-2.6782	-1.9386	2.2460	<i>Cryptomonas.obovata</i>	1.9266	-3.9986	-0.6171
<i>Undetermined.chlorophytes</i>	1.2762	-4.6053	3.3089	<i>Cryptomonas.rostratiformis</i>	4.0307	-1.9817	-1.6792
<i>Oval.chlorophytes</i>	0.7633	0.2794	-1.7025	<i>Cryptomonas.sp</i>	-5.1383	-1.1852	2.2689
<i>Spherical.chlorophytes</i>	-3.5667	0.4855	-4.1339	<i>Rhodomonas.minuta</i>	-1.4077	-5.1246	0.6298
<i>Coelastrum.astroideum</i>	-4.5219	-0.4827	-0.5700	<i>Rhodomonas.minuta.var.nanoplanc</i>	-1.8501	-4.3879	2.4117
<i>Coelastrum.microporum</i>	-4.2851	-0.9665	-0.3253	<i>Anabaena.bergii</i>	-0.8328	1.7578	3.4732
<i>Coelastrum.reticulum</i>	-3.0824	0.0893	-2.1973	<i>Anabaena.solitaria</i>	-2.7357	2.9201	-0.0966
<i>Crucigenia.tetrapedia</i>	-5.6385	-0.7882	-0.8735	<i>Anabaena.spiroides</i>	-0.0827	0.9620	1.4603
<i>Crucigeniella.rectangularis</i>	-4.6170	-0.2535	0.9822	<i>Aphanizomenon.flos.aquae</i>	-4.2937	-1.2173	-1.4324
<i>Dictyosphaerium.pulchellum</i>	0.7570	-0.7267	0.9250	<i>Aphanizomenon.gracile</i>	-3.9170	-0.7769	1.5732
<i>Dydimocystis.bicellularis</i>	-3.3370	-2.1181	-1.8148	<i>Aphanocapsa.conferta</i>	1.6077	-3.3291	-0.5087
<i>Dysmorphococcus.variabilis</i>	-3.8289	-1.0249	-1.6833	<i>Aphanocapsa.holsatica</i>	2.2670	-2.1479	-0.5345
<i>Elakatothrix.gelatinosa</i>	1.6627	0.1565	0.5641	<i>Chroococcus.limneticus</i>	-3.4998	-0.6814	2.0138
<i>Little.green.flagellates</i>	2.3873	-4.6834	0.3322	<i>Undetermined.spherical.cyanobacteria</i>	-3.5028	-2.2170	0.2695
<i>Golenkinia.radiata</i>	-4.7603	-0.6061	1.2524	<i>Microcystis.aeruginosa</i>	-2.7138	-0.7281	1.7385
<i>Hyaloraphidium.contortum</i>	1.6863	-2.4066	-1.3186	<i>Microcystis.weissenbergii</i>	-4.7401	-0.4904	-1.2801
<i>Korshikoviella.sp</i>	0.2816	0.1600	-0.1110	<i>Oscillatoria.sp</i>	-1.7668	-0.3694	2.8353
<i>Monoraphidium.kormakovae</i>	-0.4942	-3.3456	0.7637	<i>Planktothrix.agardhii</i>	-4.5516	-0.3748	-3.1313
<i>Monoraphidium.minutum</i>	-2.1976	-0.7848	1.6877	<i>Planktothrix.rubescens</i>	-1.5819	-2.0039	-0.8615
<i>Mougeotia.gracillima</i>	-4.5035	-0.4395	-1.6214	<i>Pseudanabaena.acicularis</i>	-4.1059	-0.4058	-2.8885
<i>Oocystis.lacustris</i>	-3.4075	0.7328	-0.1347	<i>Pseudanabaena.galeata</i>	-4.1019	0.0014	-2.8950
<i>Oocystis.solitaria</i>	-2.7997	1.0246	0.7537	<i>Pseudanabaena.limnetica</i>	-3.3859	0.0810	-1.2607
<i>Pediastrum.biradiatum</i>	-2.0526	-0.4241	0.0452	<i>Woronichinia.naegeliana</i>	-4.7507	0.3277	-0.8901
<i>Pediastrum.boryanum.var.longico</i>	-4.0659	1.2017	0.2938	<i>Closterium.acutum.var.variabile</i>	-5.8198	-1.2237	-1.7545
<i>Pediastrum.duplex.var.gracillin</i>	-5.1105	-0.4543	0.0012	<i>Cosmarium.sp</i>	-4.1097	0.6225	-1.4393
<i>Pediastrum.simplex</i>	-5.0262	-0.7297	1.4724	<i>Spondylosium.clypsedra</i>	-3.6256	0.1456	3.3373
<i>Pediastrum.simplex.var.sturmii</i>	-4.7762	-0.5337	1.2137	<i>Staurastrum.sp</i>	-3.8907	-0.8983	0.0389
<i>Pediastrum.tetras</i>	-3.2649	-0.1630	1.9812	<i>Asterionella.formosa</i>	2.4743	-1.5287	-3.0705
<i>Phacotus.lendneri</i>	-0.4203	4.2227	1.0282	<i>Aulacoseira.granulata.var.angus</i>	-4.5287	-0.9708	-0.5957
<i>Scenedesmus.acutus</i>	-2.5901	-1.2113	1.1151	<i>Cyclotella.bodanica</i>	3.2234	-2.0073	-0.4283
<i>Scenedesmus.bicaudatus</i>	-2.6751	0.5162	-0.5198	<i>Cyclotella.cyclopuncta</i>	1.4606	-5.0615	2.1292
<i>Scenedesmus.linearis</i>	-2.8809	1.5439	0.9743	<i>Cyclotella.sp</i>	-2.4601	-3.4938	-0.8661
<i>Scenedesmus.opoliensis</i>	-1.4343	-2.2835	0.6622	<i>Diatoma.tenuis</i>	0.6928	-3.5553	-1.0415
<i>Scenedesmus.parisiensis</i>	-3.3948	-0.1426	1.6278	<i>Fragilaria.crotonensis</i>	2.8722	-1.5168	-0.4610
<i>Scenedesmus.quadricauda</i>	-5.0648	-1.0031	0.3205	<i>Fragilaria.ulna.var.acus</i>	1.6100	-3.8643	0.2784
<i>Scenedesmus.spinosis</i>	-4.5085	-0.7394	0.4613	<i>Fragilaria.ulna.var.angustissim</i>	3.6104	-5.7770	0.1412
<i>Sphaerocystis.schroeteri</i>	-3.4899	-0.4876	3.1532	<i>Fragilaria.virescens</i>	-2.1572	0.2473	-1.3691
<i>Tetraedron.caudatum</i>	-2.8969	0.6740	-0.3869	<i>Navicula.sp.grande.</i>	2.9476	-4.9755	-1.0642
<i>Tetraedron.minimum</i>	-5.8823	-0.4875	-2.0465	<i>Nitzschia.acicularis</i>	0.2077	-3.1873	0.4495
<i>Tetraedron.triangulare</i>	-0.1914	1.4120	1.1185	<i>Nitzschia.fruticosa</i>	-1.4056	0.2755	-0.2285
<i>Tetrastrum.staurogeniaforme</i>	1.5688	-2.2103	-0.0728	<i>Nitzschia.sp</i>	-3.1708	-2.3708	-0.6505
<i>Treubaria.triappendiculata</i>	-2.9033	-1.2208	0.4465	<i>Ceratium.hirundinella</i>	-3.7908	1.1676	0.6651
<i>Bitrichia.chodatii</i>	-1.9352	2.3099	0.5682	<i>Gymnodinium.lantzschii</i>	0.6250	-3.3045	0.8018
<i>Chrysolykos.planctonicus</i>	1.5618	3.5808	-0.5917	<i>Gymnodinium.sp</i>	-0.5162	-2.6478	1.4679
<i>Dinobryon.bavaricum</i>	1.6503	4.3326	1.1287	<i>Peridiniopsis.Elpatewskii</i>	-1.2995	1.8610	5.0755
<i>Dinobryon.divergens</i>	2.2249	5.0219	-0.4422	<i>Peridinium.inconspicuum</i>	-2.2539	1.1369	-2.1221
<i>Dinobryon.elegantissimum</i>	1.2820	4.4612	1.3891	<i>Peridinium.willei</i>	-4.6275	0.0964	0.1515
<i>Dinobryon.sertularia</i>	2.2824	-1.0338	-1.3758	<i>Euglena.sp</i>	-4.6530	-0.5840	0.3981
<i>Dinobryon.sociale.var.americanu</i>	2.1917	-0.7654	-1.6008	<i>Phacus.acuminatus</i>	-3.2359	-2.3656	-0.0081
<i>Dinobryon.sociale.var.stipitatu</i>	-0.8873	2.9036	-1.1843	<i>Trachelomonas.sp</i>	-4.8274	-0.5552	0.9606
<i>Erkenia.subaquaeciliata</i>	1.6624	-3.3661	2.4325	<i>Trachelomonas.volvocina</i>	-6.0895	-2.1411	-1.6180
<i>Kephyrion.sp</i>	1.7442	3.5304	-1.8306	<i>Tribonema.angustissimum</i>	2.9341	-4.9751	-0.5835