Baku Mathematical Journal
2025, Vol. 4, No 1, PP. 48-68
https://doi.org/10.32010/j.bmj.2025.05

CO-INERTIA ANALYSIS OF DATA STRUCTURED IN GROUPS OF INDIVIDUALS

R.O. MALOUATA*, M. KOUKOUATIKISSA DIAFOUKA, G.C. LOUZAYADIO

Received: 14.06.2024 / Revised: 29.01.2025 / Accepted: 17.02.2025

Abstract. *This paper presents an overview of methods for the analysis of data structured in N multiblocks of variables partitioned in M multigroups of individuals. More specifically, successive generalized co-inertia analysis (SGCIA) and its dual method, which are two unifying approaches for multiblock data analysis and multigroup data analysis. Examples are given to illustrate the use of the proposed methods.*

Keywords: successive generalized co-inertia analysis, multiblock data analysis, multigroup data analysis

Mathematics Subject Classification (2020): 62H25, 62H30, 62H35

1. Introduction

The main purpose of this article is to generalize the multiblock data analysis methods and the multigroup data analysis methods. A mutiblock is a partition of columns structured in blocks. Each block is a data matrix whose variables are measured on the same number of individuals. A mutigroup is a partition of rows structured in groups. Each group is a data matrix whose variables are measured on different groups of individuals.

These two general classes of methods have two special cases. Canonical correlation analysis [6] is the seminal paper for the first family and Tucker's interbattery factor analysis [19] for the second one. When we consider a data set structured in blocks of variables,

* Corresponding author.

Rodnellin O. Malouata
Marien Ngouabi University, Brazzaville, Congo
E-mail: onesimero@gmail.com

Michel Koukouatikissa Diafouka
Marien Ngouabi University, Brazzaville, Congo
E-mail: diafdiak@gmail.com

Gélin C. Louzayadio
Marien Ngouabi University, Brazzaville, Congo
E-mail: gelinlouzayadio@gmail.com

the criterion of interbattery factor analysis has been extended to multiple co-inertia analysis [2]. However, the criterion of canonical correlation analysis has been extended to the generalized canonical correlation analysis [1], [5], [7]. For the case of two data matrices, interbattery factor analysis is an important case in point. An important difference between (generalized) interbattery factor analysis and generalized canonical correlation analysis is that the former does not only focus on optimally describing the relationship between sets of variables, but in addition requires that the variance within sets of variables is explained well by the components used. Morever, methods of regularization of generalized canonical correlation analysis [15]-[17] have been proposed. These methods are a framework for modeling linear relationships between several blocks of variables observed on the same set of individuals. Another computational method for measuring the common structure between two data matrices can be found in [10]. In [10], They maximize the following criterion

$$f(u,v) = \left[\sum_{h=1}^{p} cov^2(Yv, x_h)\right]\left[\sum_{l=1}^{q} cov^2(Xu, y_l)\right] \tag{1}$$

subject to the normalization constraints (2). This criterion is equivalent to maximize

$$f(u,v) = (u'Ku)(v'Hv), \tag{2}$$

where $K = V_{XY}V_{YX}$ and $H = V_{YX}V_{XY}$ are two positive semidefinite symmetric matrices.

Several generalizations of canonical correlation analysis and interbattery factor analysis have been proposed for handling situations with more than two sets of variables [1], [2], [4], [5], [7], [10], [11], [15], [16].

In the case of the multigroup framework, when the same set of variables is observed on different groups of individuals, the Partial Triadic Analysis (PTA) of [18] which is one of the simplest analyses of the STATIS family and the multigroup analysis of [9] can be seen as the principal component analysis (PCA) [12] of a series of PCAs.

To study the stability of relationships between several pairs of matrices, Simier and others [14] have proposed the STATICO method. It is well known that the weighting coefficients of the compromise may be contrary sign in some cases. For this reason, alternatives have been proposed which maximize the sum of covariances and the sum of squared covariances between the components, with orthonormality constraints on the components. For instance, Kissita and others [8] have proposed CIAs3, which maximizes

$$f(u,v) = \sum_{i=1}^{M}(u^{'}K_iu)\sum_{i=1}^{M}(v^{'}H_iv) \tag{3}$$

subject to the constraints $||u|| = ||v|| = 1$, with $K_i = V_{X_iY_i}V_{Y_iX_i}$ and $H_i = V_{Y_iX_i}V_{X_iY_i}$. Maximizing (1) offers a method of analyzing relationships between two partitioned matrices $X = [X_1', \cdots, X_M']'$ and $Y = [Y._1', \cdots, Y_M']'$, centered column wise and measured on p and q variables.

Eslami and others [3] proposed a approach multiblock/multigroup situation. But this approach is a multiblock/multigroup PCA. The idea of having the SGCIA method is to provide on the one hand multigroup operators which are symmetric and positive semidefinite for investigating the relationships between pairs of multiblocks structured in multigroups, on the other hand systems of the orthogonal vectors for the representation of the groups of individuals and blocks of variables. When the data set is partitioned in several multiblocks, we propose a dual method of SGCIA.

Finally, we will conclude this paper with a detailed analyses of a practical example where many of the special cases are explored. This paper is organized as follows: In section 2, we will propose the SGCIA method and its dual method. In section 3 and 4, an overview of applications of SGCIA for several multiblock and multigroup data analysis is given.

2. Methods

In this paper, we consider a data supermatrix X structured in multigroups (partition of rows) or in multiblocks (partition of columns). Rows of X are related to individuals and columns to variables.

2.1. Successive generalized co-inertia analysis

In the multiblock framework, we consider $X = [X_1, \ldots, X_j, \ldots, X_N]$ a column partition. Each $n \times p_j$ data matrix $X_j = [X'_{1j}, \ldots, X'_{ij}, \ldots, X'_{Mj}]'$ is called a multigroup. In this subframework, the same set of variables is observed on different groups of individuals. Each $n_i \times p_j$ data submatrix X_{ij} centered column wise is called a group. The number of individuals of each group can differ from one group to another. Finally, $X = [X_{ij}]_{i,j}$ is a supermatrix having $n = \sum_{i=1}^{M} n_i$ rows and $p = \sum_{j=1}^{N} p_j$ columns.

Definition 1. *The successive generalized co-inertia analysis (SGCIA) consists of finding components $X_j u_j$, where u_j are loading vectors, summarizing a community of structures of the data matrices X_j related to each of the sets covariances. Thus, In this way SG-CIA puts more emphasis on describing sets covariance than does multiblock/multigroup PCA of [3]. SGCIA for multiblock and multigroup data analyses is based on a single optimization problem. The core optimization problem considered in this paper is defined as follows:*

$$\textit{Maximise} \quad f(u_1, \cdots, u_N) = \left(\sum_{i=1}^{M} u_1^{'} K_{i1} u_1 \right) \left[\prod_{j=2}^{N} \left(\sum_{i=1}^{M} u_j^{'} H_{ij} u_j \right) \right] \tag{4}$$

$$\textit{subject to the constraints} \quad ||u_j|| = 1, \quad j = 1, \ldots, N,$$

where $K_{i1} = V_{X_{i1}X_{i2}} V_{X_{i2}X_{i1}}$ and $H_{ij} = V_{X_{ij}X_{ij-1}} V_{X_{ij-1}X_{ij}}$ are symmetric and positive semidefinite matrices. K_{i1} is a matrix which allows to investigate the relationships between variables of the data submatrices X_{i1} and X_{i2}. H_{ij} is a matrix which allows to investigate the relationships between variables of the data submatrices X_{ij} and X_{ij-1} and X_j and X_{j-1}.

Definition 2. *The second criterion (SGCIA) is formulated as follows:*
Maximize

$$f(u_1, \cdots, u_N) = \left(\sum_{i=1}^{M} (u_1^{'} Q_1 K_{i1} Q_1 u_1) \right) + \left[\prod_{j=2}^{N} \left(\sum_{i=1}^{M} (u_j^{'} Q_j H_{ij} Q_j u_j \right) \right] \tag{5}$$

subject to the same normalization constraints of criterion (4).

Definitions 1 and 2 are equivalent to the optimum.

In what follows, we propose only the SGCIA3 solution, given that the SGCIA4 solution is identical to the optimum of the SGCIA3 solution. We call this SGCIA method.

To simplify the presentation, the metrics implicitly considered in individual spaces are the identity metrics. However, other metrics could also be used, as is done in co-inertia analysis.

The following Lagrangian function related to optimization problem (4) is considered:

$$L(u_1, \cdots, u_N, \alpha_1, \ldots, \alpha_N) =$$

$$= \left(\sum_{i=1}^{M} u_1^{'} K_{i1} u_1 \right) \left[\prod_{j=2}^{N} \left(\sum_{i=1}^{M} u_j^{'} H_{ij} u_j \right) \right] + \alpha_1 (1 - u_1^{'} u_1) + \sum_{j=2}^{N} \alpha_j (1 - u_j^{'} u_j), \tag{6}$$

where α_j, $j = 1, \cdots, N$, are the Lagrange multipliers.

The following proposition specifies the role of the vectors u_1 and u_j in the criterion to be maximized.

Property 1. *If we set* $r_{u_1} = \sum_{i=1}^{M} (u_1^{'} K_{i1} u_1)$ *and* $r_{u_j} = \sum_{i=1}^{M} (u_j^{'} H_{ij} u_j)$ *for all* $(j = 2, \cdots, N)$*, partial co-inertia axes* u_1 *and* u_j *for all* $(j = 2, \cdots, N)$ *from SGCIA verify the stationary equations*

$$\left(\sum_{i=1}^{M} K_{i1} \right) u_1 = r_{u_1} u_1, \tag{7}$$

$$\left(\sum_{i=1}^{M} H_{ij} \right) u_j = r_{u_j} u_j, \tag{8}$$

$$\alpha = f(u_1, \cdots, u_N) = \prod_{j=1}^{N} r_{u_j}. \tag{9}$$

u_1 *and* u_j *are eigenvectors of the* $\sum_{i=1}^{M} K_{i1}$ *and* $\sum_{i=1}^{M} H_{ij}$ *matrices respectively, related to the largest eigenvalues* r_{u_1} *and* r_{u_j}*.*

Proof. We may also consider the derivative L'. Canceling the derivatives of the Lagrangian function with respect to u_j and α_j yields the following stationary equations:

$$\frac{1}{2}\frac{\partial L}{\partial u_1} = \left(\prod_{j=2}^{N} r_{u_j}\right)\sum_{i=1}^{M} K_{i1}u_1 - \alpha_1 u_1 = 0,$$

$$\frac{1}{2}\frac{\partial L}{\partial u_j} = \left(\prod_{h=1,h\neq j}^{N} r_{u_h}\right)\sum_{i=1}^{M} H_{ij}u_j - \alpha_j u_j = 0, \quad j = 2,\cdots,N,$$

$$\frac{\partial L}{\partial u_j} = 1 - u_j^{'}u_j = 0, \quad j = 1,\cdots,N.$$

By pre-multiplying relations (7) and (8) by $u_1^{'}$ and $u_j^{'}$ respectively, and taking into account equalities (9), we find relation (6):

$$\alpha_1 = \alpha_2 = \cdots = \alpha_j = \cdots \alpha_N = \alpha = f(u_1,\cdots,u_N) = \prod_{j=1}^{N} r_{u_j}.$$

Taking into account relation (6) in (7) and (8), It yield the stationary equations (4) and (5). ◀

Having determined the solutions of order 1, which we denote $u_{1,1}$ and $u_{j,1}$, we determine the solutions of order greater than 1.

The co-inertia axes $u_j^{(s)}$ (respectively $u^{(s)} = [u_1^{(s)'}|\cdots|u_j^{(s)'}|\cdots|u_N^{(s)'}]'$ the block vector of $\mathbb{R}^p$) are orthornormal (respectively orthogonal). On the other hand, the $c_{X_{ij}}^{(s)} = X_{ij}u_j^{(s)}$ components are not D_i-orthogonal. To obtain this orthogonality property for the synthetic components, we set $X_{ij}^{(0)} = X_{ij}$, for all $i = 1,\cdots,M$ and $j = 1,\cdots,N$ and

$$X_{ij}^{(s-1)} = P^{\perp}_{c_{X_{ij}}^{(s-1)}} X_{ij}^{(s-2)},$$

with

$$P^{\perp}_{c_{X_{ij}}^{(s-1)}} = I_{n_i} - P_{c_{X_{ij}}^{(s-1)}} \quad \text{and} \quad P_{c_{X_{ij}}^{(s-1)}} = \frac{c_{X_{ij}}^{(s-1)} c_{X_{ij}}^{(s-1)'} D_i}{\|c_{X_{ij}}^{(s-1)}\|^2_{D_i}}$$

the D_i-orthogonal projector onto the subspace of $c_{X_{ij}}^{(s-1)} = X_{ij}^{(s-2)}u_j^{(s-1)}$. The following proposition specifies the role of the vectors $u_1^{(s)}$ and $u_j^{(s)}$ in the criterion to be maximized.

Property 2. *At order s, the co-inertia axes $u_1^{(s)}$ and $u_j^{(s)}$ ($j = 2,\cdots,N$) verify the stationary equations*

$$\left(\sum_{i=1}^{M} K_{i1}^{(s-1)}\right) u_1^{(s)} = r_{u_{1,s}} u_1^{(s)},$$

$$\left(\sum_{i=1}^{M} H_{ij}^{(s-1)}\right) u_j^{(s)} = r_{u_{j,s}} u_j^{(s)}, \tag{10}$$

$$\alpha = f(u_1^{(s)}, \cdots, u_N^{(s)}) = \prod_{j=1}^{N} r_{u_j^{(s)}},$$

where $K_{i1}^{(s-1)} = X_{i1}^{(s-1)'} D_i X_{i2}^{(s-1)} X_{i2}^{(s-1)'} D_i X_{i1}^{(s-1)}$ *and*

$$H_{ij}^{(s-1)} = X_{ij}^{(s-1)'} D_i X_{ij-1}^{(s-1)} X_{ij-1}^{(s-1)'} D_i X_{ij}^{(s-1)}.$$

Property 3. *For $s = 1, \cdots, min(p_j)$ and $j = 1, \cdots, N$, the co-inertia axes $u_j^{(s)}$ are orthogonal.*

Proof. We only show the orthogonality of the $u_j^{(s)}$ axes, since the orthogonality of the $u_1^{(s)}$ axes can be demonstrated in the same way. Multiplying the left-hand side of relation (10) by the transpose of $Q_j u_j^{(t)}$ for all $t = 1, \cdots, s-1$, we obtain

$$r_{u_j^{(s)}} u_j^{(s)'} u_j^{(t)} = u_j^{(s)'} \left(\sum_{i=1}^{M} X_{ij}^{(s-1)'} D_i X_{ij-1}^{(s-1)} X_{ij-1}^{(s-1)'} D_i X_{ij}^{(s-1)}\right) u_j^{(t)} = 0$$

because

$$X_{ij}^{(s-1)} u_j^{(t)} = \left(\prod_{d=t}^{s-1} P^{\perp}_{c_{X_{ij}}^{(d)}}\right) X_{ij}^{(t-1)} u_j^{(t)} =$$

$$= P^{\perp}_{c_{X_{ij}}^{(s-1)}} P^{\perp}_{c_{X_{ij}}^{(s-2)}} \cdots P^{\perp}_{c_{X_{ij}}^{(t+1)}} P^{\perp}_{c_{X_{ij}}^{(t)}} c_{X_{ij}}^{(t)} = 0$$

and $P^{\perp}_{c_{X_{ij}}^{(t)}} c_{X_{ij}}^{(t)} = 0$, for all $t = 1, \cdots, s-1$, $i = 1, \cdots, M$ and $j = 1, \cdots, N$.

As $r_{u_{j,s}} \neq 0$, we obtain $u_j^{(s)'} u_j^{(t)} = 0$. ◀

The orthogonality of the components $(c_{X_{ij}}^{(s)})_s$ for all $i = 1, \cdots, M$, $j = 1, \cdots, N$ and $s = 1, \cdots, \min(p_j)$ allows to study the internal structures of each of the matrices. If the X_{ij} groups are reduced, the coordinates of the variables in the plane given by $c_{X_{ij}}^{(s)}$ and $c_{X_{ij}}^{(t)}$ are the correlations between the X_{ij} variables and the $c_{X_{ij}}^{(s)}$ and $c_{X_{ij}}^{(t)}$ components. These pictures of the variables allow to interpret the components of each X_{ij} group. To represent the variables in the X_{i1} group, proceed in the same way as above. It is also possible to use the additional elements technique to represent the variables in each of the groups X_{ij} by projecting the rows of the H_{ij} matrices onto the co-inertia axes $u_j^{(s)}$ and $u_j^{(t)}$ respectively. In the same way, we project the variables of the K_{i1} matrices to represent the variables of the X_{i1} groups on the co-inertia axes $u_1^{(s)}$ and $u_1^{(t)}$.

Taking into account the orthogonality of the co-inertia axes $u_1^{(s)}$ and $u_j^{(s)}$, we can project the individuals of the groups $X_{i,1}$ and $X_{i,j}$ for all $j = 2, \cdots, N$ respectively in the planes defined by $(u_1^{(s)}, u_1^{(t)})$ and $(u_j^{(s)}, u_j^{(t)})$. But the coordinates of these projections are not exactly given by the components of the vectors $c_{X_{i1}}^{(s)}$ and $c_{X_{i1}}^{(t)}$ and/or $c_{X_{ij}}^{(s)}$ and $c_{X_{ij}}^{(t)}$ due to the bias caused by deflations on the tables.

At order s $(s = 1, \cdots, r)$, the specific weights associated with the pairs of groups X_{ij} and X_{ik} for all $i = 1, \cdots, M$ and $j, k = 1, \cdots, N$ are respectively defined by $\rho_{X_{ij}}^{(s)} = var(X_{ij}u_j^{(s)})$ and $\rho_{X_{ik}}^{(s)} = var(X_{ik}u_k^{(s)})$. These weights define the projected inertia of the clouds of individuals associated with the tables X_{ij} and X_{ik} on the co-inertia axes u_j and u_k respectively, $r \leq min(p_j)$. These weights characterize the stability of each group of variables. We associate with the groups X_{ij} and X_{ik} are the numbers $\rho_{X_{ijk}}^{(s)} = cor^2(X_{ij}u_j^{(s)}, X_{ik}u_k^{(s)})$, which are the squares of the correlation coefficients. On the other hand, these coefficients characterize the stability of the relationship between groups X_{ij} and X_{ik} for all $i = 1, \cdots, M$ for multigroups X_j and X_k with $j \neq k$.

Suppose outer vectors, for $s \geq 2$, $u_{1,s}$ and $u_{j,s}$ have been constructed. We now consider the different special cases which give this generalization and powerfulness of the optimization problem (3) for multigroup and multigroup data analysis.

2.1.1. SGCIA is a PCA [12]

SGCIA is a PCA for covariance matrix with special structure. Let us consider Σ a $p \times p$ block diagonal matrix whose principal diagonal can be expressed in matrices $\sum_{i=1}^{M} K_{i1}$ and $\sum_{i=1}^{M} H_{ij}$. From the stationary equations (4) and (5), suppose Σ is the block diagonal matrix

$$\Sigma = \begin{pmatrix} \sum_{i=1}^{M} K_{i1} & & & 0 \\ & \sum_{i=1}^{M} H_{i2} & & \\ & & \ddots & \\ 0 & & & \sum_{i=1}^{M} H_{iN} \end{pmatrix}$$

Setting $u_s = [u'_{1,s}, \ldots, u'_{j,s}, \ldots, u'_{N,s}]'$ a block vector and $\Lambda = diag(r_{u_{j,s}}, j = 1, \ldots, N)$ eigenvalues diagonal matrix, we observe that

$$\Sigma u_s = u_s \Lambda.$$

Since the sth principal component $\xi_s = Xu_s = \sum_{j=1}^{N} X_j u_{j,s}$ is a linear combination of the multigroup matrices X_j or the sum of the components $\xi_{j,s} = X_j u_{j,s}$, the set of principal components contains the linear combinations of the groups X_{ij} or the sums of the components $\xi_{ij,s} = X_{ij}u_{j,s}$. This principal component explains a proportion

$$\frac{\alpha_s}{\rho}, \quad \text{where} \quad \alpha_s = \prod_{j=1}^{N} r_{u_{j,s}} \quad \text{and} \quad \rho = \sum_{s=1}^{\min(p_j)} \alpha_s$$

of the total population variation.

Special case. If $M = N = 1$, the super multigroup is reduced to a single group $X_{11} = X$ and the SGCIA is reduced to the analysis of the triplet (X, I_p, D).

2.1.2. SGCIA is a interbattery factor analysis [19]

Clearly, by setting $M = 1$ and $N = 2$ in the SGCIA criterion, we obtain the interbattery factor analysis. Since the function can be written

$$f(u_1, u_2) = (u_1' K_{11} u_1)(u_2' H_{12} u_2).$$

When we have a table X_{11}, the search for a component $X_{11}u_1$ synthesizing the system of covariations of the variables x_{1l}^1 of a table $X_{11} n \times p_1$ is done by principal component analysis, using the criterion optimization problem:

$$f(u_1) = \sum_{l=1}^{p_1} Cov^2(X_{11}u_1, x_{1l}^1) = u_1' K_{11} u_1.$$

When we have two tables X_{11} and X_{12}, the information on the score analogy between X_{11} and X_{12} is contained in the variance-covariance matrix $X_{11}' D X_{12}$. The $X_{11}u_1$ and $X_{12}u_2$ components synthesizing this information are obtained from the singular value decomposition defined by:

$$X_{11}' D X_{12} = U \Delta \tilde{U}'$$

with $U \quad p_1 \times r$ and $\tilde{U} \quad p_2 \times r$ two matrices such that $U'U = U\tilde{U}\tilde{U}' = I_{n_1}$ and $\Delta \quad r \times r$ a diagonal block matrix where $r$is the rank of the matrix $X_{11}' D X_{12}$.

The orthonormal base systems $\{u_{1s}\}_{s=1,\dots,r}$ and $\{u_{2s}\}_{s=1,\dots,r}$ being respectively formed by the columns of U and $\tilde{U}$, then the vectors u_{1s} and u_{2s} for $s = 1 \dots, r$ which verify the following relations

$$K_{11} u_{1s} = r_{u_s} u_{1s} \quad \text{and} \quad H_{12} u_{2s} = r_{u_s} u_{2s}$$

are solutions of the function

$$f(u_{1s}, u_{2s}) = (u_{1s}' K_{11} u_{1s})(u_{2s}' H_{12} u_{2s}),$$

where the positive value $r_{u_s} = Cov(X_{11}u_{1s}, X_{12}u_{2s})$ constitutes the $s^{textith}$ diagonal of Δ. The vectors u_{1s} are singular to the left of $X_{11}' D X_{12}$ and u_{2s} are singular to the right.

2.1.3. SGCIA is a SCIA3 [8]

If M is arbitrary and $N = 2$, the super multigroup $T = [X_{ij}]$ reduces to two multigroups and the SGCIA reduces to the SCIA3, confirming that the SGCIA is a generalization of the CIAs3 method proposed by [8]. Since the function can be written

$$f(u_1, u_2) = \sum_{i=1}^{M} (u_1^{'} K_{i1} u_1) \sum_{i=1}^{M} (u_2^{'} H_{i2} u_2)$$

subject to the constraints $||u_1|| = ||u_2|| = 1$, with $K_{i1} = V_{X_{i1}X_{i2}}V_{X_{i2}X_{i1}}$ and $H_{i2} = V_{X_{i2}X_{i1}}V_{X_{i1}X_{i2}}$.

By setting $r_{u_1} = \sum_{i=1}^{M}(u_1^{'}K_{i1}u_1)$ and $r_{u_2} = \sum_{i=1}^{M}(u_2^{'}H_{i2}u_2)$, we get $\alpha = r_{u_1}r_{u_2}$. Thus, relations (4) and (5) yield the following stationary equations:

$$\left(\sum_{i=1}^{M}K_{i1}\right)u_1 = r_{u_1}u_1,$$
$$\left(\sum_{i=1}^{M}H_{i2}\right)u_2 = r_{u_2}u_2.$$

We obtain:

- u_1 is the eigenvector of the matrix $\sum_{i=1}^{M}K_{i1}$ related to the largest eigenvalue r_{u_1},
- u_2 is the eigenvector of the matrix $\sum_{i=1}^{M}H_{i2}$ related to the largest eigenvalue r_{u_2}.

2.1.4. SGCIA is a multiple co-inertia analysis (MCOIA) [2]

If $M = 1$ and N is arbitrary, the super multiblock $T = [X_{ij}]$ becomes a N horizontal matrices with general element the block X_{1j} of dimension (n_1, p_j) for all $j = 1, \cdots, N$. The SGCIA method becomes a multiple co-inertia analysis proposed by [2] whose stationary equations are:

$$(K_{11}^{(s-1)})u_1^{(s)} = r_{u_{1,s}}u_1^{(s)},$$
$$(H_{1j}^{(s-1)})u_j^{(s)} = r_{u_{j,s}}u_j^{(s)}.$$

2.1.5. SGCIA is a Concor method [10]

If, instead of N multigroups, we have $N+1$ multigroups, of which the first X_0 multigroup is the reference multigroup Y and the others form the super-multigroup T made up of N multigroups, the SGCIA method adopts the approach of the Concor analysis proposed by [10] and is equivalent to maximizing the function

$$f(v, u_1, \cdots, u_N) = \left(\sum_{i=1}^{M}(v^{'}K_{iY}v)\right)\left[\prod_{j=1}^{N}\left(\sum_{i=1}^{M}(u_j^{'}H_{ij}u_j\right)\right]$$

subject to the constraints $u_j^{'}u_j = 1$ for all $j = 1, \cdots, N$ and $v^{'}v = 1$ with $K_{iY} = Y_i^{'}D_iX_{i1}X_{i1}^{'}D_iY_i$, $Y_i = X_{i0}$ and v a vector of $\mathbb{R}^q$.

This maximization problem leads to the order s to the stationary equations

$$\sum_{i=1}^{M}K_{iY}v^{(s)} = r_{v_s}v^{(s)},$$

$$\sum_{i=1}^{M}H_{ij}u_j^{(s)} = r_{u_{j,s}}u_j^{(s)}.$$

The result is the CONCOR analysis proposed by [10], which is an extension of MCOIA.

2.2. Dual Successive generalized co-inertia analysis

In the previous subsection we proposed SGCIA, in this subsection we will propose the dual method of SGCIA. The Dual Successive generalized co-inertia analysis (DSGCIA) is similar to SGCIA, but SGCIA method does not have the same solution.

When we have M multiblocks X_i (we cut the table T into rows), to study the internal structure between these M multiblocks, we seek to define $X_i^{'} D_i v_i$ components, where v_i are D_i-normed vectors for all $i = 1, \cdots, M$ synthesizing a community of structures of multiblocks X_i relative to each of the systems of internal proximities. The aim is to simultaneously want the $X_{1j}^{'} D_1 v_1$ components characterize the proximity systems of the individuals in the X_{2j} tables, and the $X_{2j}^{'} D_2 v_2$ components characterize the proximity systems of the individuals in the X_{1j} tables.

Furthermore, it is the case that the components $X_{i-1j}^{'} D_{i-1} v_{i-1}$ characterize the systems of proximities of the individuals of the tables X_{ij}, and that the components $X_{ij}^{'} D_i v_i$ characterize the systems of proximities of the individuals of the tables X_{i-1j} for all $i = 2, \cdots, M$ and $j = 1, \cdots, N$.

Definition 3. *The dual SGCIA consists of searching for vectors v_i of $\mathbb{R}^{n_i}$ by maximizing the function*

$$f(v_1, \cdots, v_M) = \left(\sum_{j=1}^{N} (v_1^{'} D_1 L_{1j} D_1 v_1) \right) \left[\prod_{i=2}^{M} \left(\sum_{j=1}^{N} (v_i^{'} D_i P_{ij} D_i v_i \right) \right] \tag{11}$$

subject to the constraints les contraintes de normalization

$$v_i^{'} D_i v_i = 1, \quad \text{for all} \quad i = 1, \cdots, M, \tag{12}$$

where $L_{1j} = X_{1j} X_{2j}^{'} X_{2j} X_{1j}^{'}$ denotes a symmetric positive semidefinite matrix of dimension (n_1, n_1) for all $j = 1, \cdots, N$. These matrices are respectively used to describe the proximities between individuals in the X_{1j} and X_{2j} blocks of the X_1 and X_2 multiblocks. $P_{ij} = X_{ij} X_{i-1j}^{'} X_{i-1j} X_{ij}^{'}$ denotes a symmetric positive semidefinite matrix (n_i, n_i) for all $i = 2, \cdots, M$ and $j = 1, \cdots, N$.These matrices are also used to describe the proximities between individuals in the blocks X_{i-1j} and X_{ij} associated respectively with the multiblocks X_{i-1} and X_i. $X_i = [X_{i1} | \cdots | X_{ij} | \cdots | X_{iN}]$, the multiblocks extracted from T of dimension (n_i, p).

Maximization (11) subject to constraints (12) leads for all $i = 2, \cdots, M$ to the stationary equations

$$\left(\sum_{j=1}^{N} L_{1j} \right) D_1 v_1 = r_{v_1} v_1,$$

$$\left(\sum_{j=1}^{N} P_{ij} \right) D_i v_i = r_{v_i} v_i,$$

$$\beta = f(v_1, \cdots, v_M) = \prod_{i=1}^{M} r_{v_i}.$$

v_1 and v_i are respectively eigenvectors of the matrices $\sum_{j=1}^{N} L_{1j} D_1$ and $\sum_{j=1}^{N} P_{ij} D_i$, associated with the eigenvalues r_{v_1} and r_{v_i}.

Definition 4. *The Dual SGCIA can also be obtained by maximizing the*

$$f(v_1, \cdots, v_M) = \left(\sum_{j=1}^{N} (v_1^{'} D_1 L_{1j} D_1 v_1) \right) + \left[\prod_{i=2}^{M} \left(\sum_{j=1}^{N} (v_i^{'} D_i P_{ij} D_i v_i \right) \right]$$

subject to the constraints:

$$v_i^{'} D_i v_i = 1, \quad \textit{for all} \quad i = 1, \cdots, M.$$

The criteria defined in definitions 3 and 4 are equivalent to the optimum.

To find solutions of order greater than one, we proceed in the same way as for SGCIA. The special cases of this method lead to well-known methods such as SGCIA.

3. Main Results

In this section we present two methods for analyzing two multigroups: SCIA3 and SGCIA. The principle of each method is briefly explained, and the result obtained on the example data set is detailed.

To demonstrate the greater suitability of the SGCIA, we reanalyze the datasets that have been acquired by [13] and which serve as an illustration in SCIA3 [8].

Specifically, we reanalyze two data matrices: one data matrix X with 24 rows and 13 columns, containing the ephemeroptera species and one data matrix Y with 24 rows and 10 columns, containing the environmental variables.

The rows of both matrices correspond to 6 sampling stations ordered upstream-downstream along a small stream, the Méaudret of France. These 6 stations are sampled 4 times, in Spring, Summer, Autumn and Winter.

The 24×13 data matrix X is partitioned in four 6×13 data matices X_i. The 13 columns of the species data table correspond to 13 Ephemeroptera species, which are known to be highly sensitive to water pollution. These species are as follows: Eda=Ephemera, Bsp=Baetis sp, Brh = Baetis rhodani, Bni = Baetis niger, Bpu = baetis pumilus, Cen = centroptilum, Ecd = Ecdyonurus, Rhi = Rhihrogena, Hla = Habrophlebialauta, Hab = Habroletoides modesta, Par = Paraletophlebia, Cae = Caenis, Eig = Epheme - rella ignita.

In addition, 24×10 data matrix Y is partitioned in four 6×10 data matices Y_i. The 10 environmental variables are physico-chemical measures: Temp=water temperature, flow, pH, Cond=conductivity, Oxyg=oxygen, BDO5=biological oxygen demand, Oxyd=oxidability, Ammo=ammonium, Nitr=nitrates and Phos=phosphates.

Species are centered by season and environmental variables are centered and then normalized globally. This global normalization allows intra-season variance to be taken into account. Each of these tables corresponds to a season and a triplet (X_i, I_p, D_i) for the fauna and (Y_i, I_q, D_i) for the environment.

The problem is to investigate the stability relationships between Ephemeroptera species distribution and the quality of water in the station typology.

3.1. Successive co-inertia analysis 3 (SCIA3)

We consider the orthogonal version of SCIA3 called SOCIA.

Table 1 contains the squared correlations between the partial linear combinations of variables (species of fish) and environmental variables of order 1 and 2 for the SOCIA3 method. These squared correlations enable to describe the evolution of species-environment relationships. Constancy of these squared coefficients of correlation enables to conclude the stability of the relationship.

Table 1. *Squared correlations between linear combinations of the variables in fish abundance and environmental variables to order 1 and 2 for SOCIA3*

	Seasons			
Methods	Spring	Summer	Autumn	Winter
SOCIA3	0.371	0.877	0.935	0.329
	0.436	0.715	0.817	0.404

It emerges from Table 1, a same evolution of species-environmemt relationships for Summer and Automn concerning SOCIA3 method (Confer the graph of Figure 1).

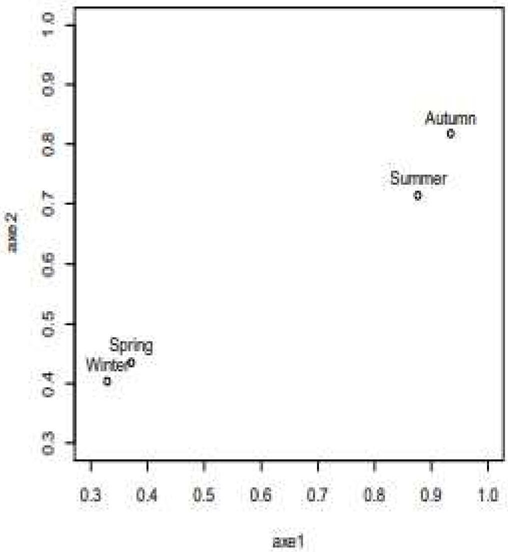

Fig. 1. Position of the seasons on the first two axes of squared correlations between partial synthetic components species-environment for SOCIA3

This observation does not find oneself in Winter and Spring which differ too much from other seasons on this method. The last situation confirms good results from other methods of co-inertia analysis cited above.

Table 2 contains the percentages of projected inertia of each table on the first two axes.

Table 2. *Percentages of projected inertia (specific weights) of each season on the first two axes*

Methods	Seasons			
	Spring	Summer	Autumn	Winter
SOCIA3 (X)	11.100	14.292	43.947	30.658
	9.521	32.679	47.328	10.470
SOCIA3 (Y)	6.725	31.022	57.926	4.324
	22.657	47.301	18.949	11.091

On the first two axes according to SOCIA3 method for the first multi-table, Autumn has projected inertia percentages are highest. Regarding the second multi-table corresponding to the environmental variables, it is rather than on the first axis of high percentages of projected inertia for Autumn found. But on the second axis, it is the Summer that has the largest percentage. We find perfectly the same results than previous methods.

In the first two axes of SOCIA3, we show the stations (Fig2X1) and species (Fig2X2). The SOCIA3 method determines simultaneously two sets of orthogonal axes at the individuals level and variables level. It follows from these graphics, any season, a general organization finds again more or less at stations and species. It notes an overall size effect at axis 1 regarding the species. The axis 1 opposes on the one hand station S6 and station S2 on the other hand for all seasons. We can find for all seasons more or less in the station S6 species Baetis sp and Baetis Rhodani. In the Spring, station S6 is characterized by high temperatures and flow (Fig3Y1 and Fig3Y2).

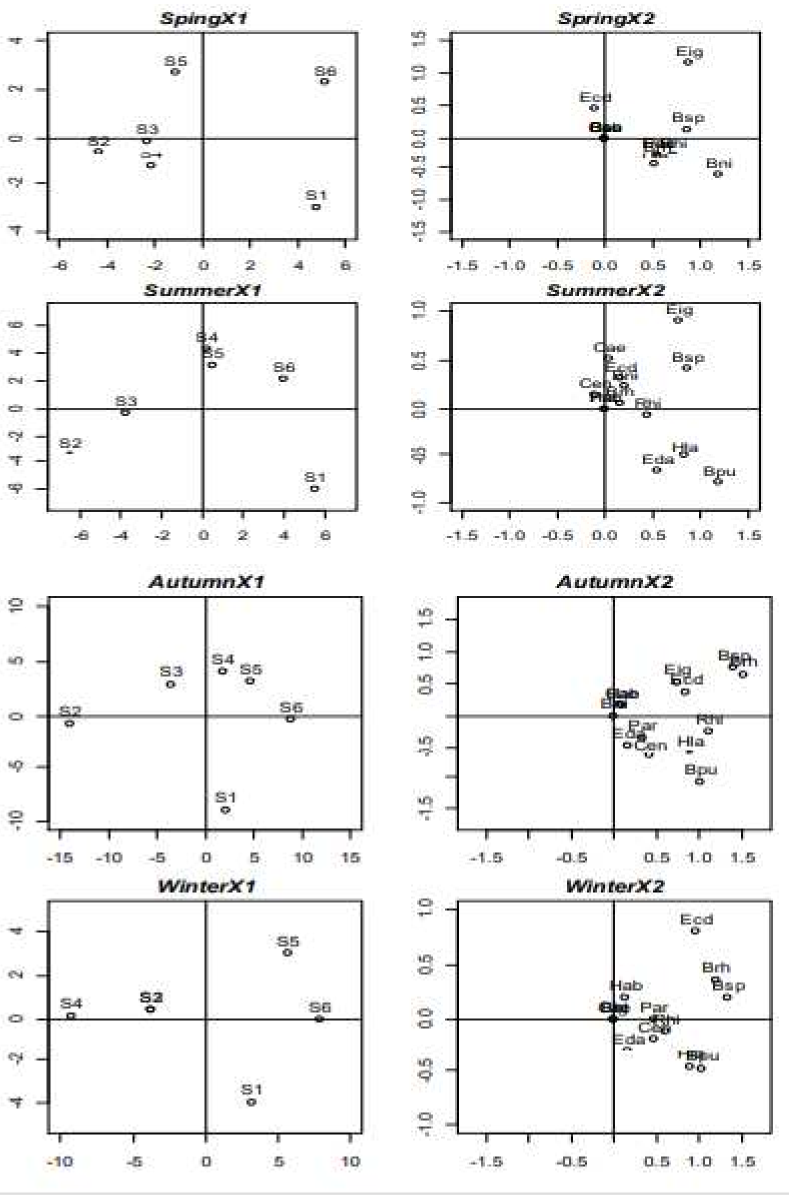

Fig. 2. Position of the stations (Fig2X1) and the species (Fig2X2) per season on the first two axes for multigroup X of SOCIA3

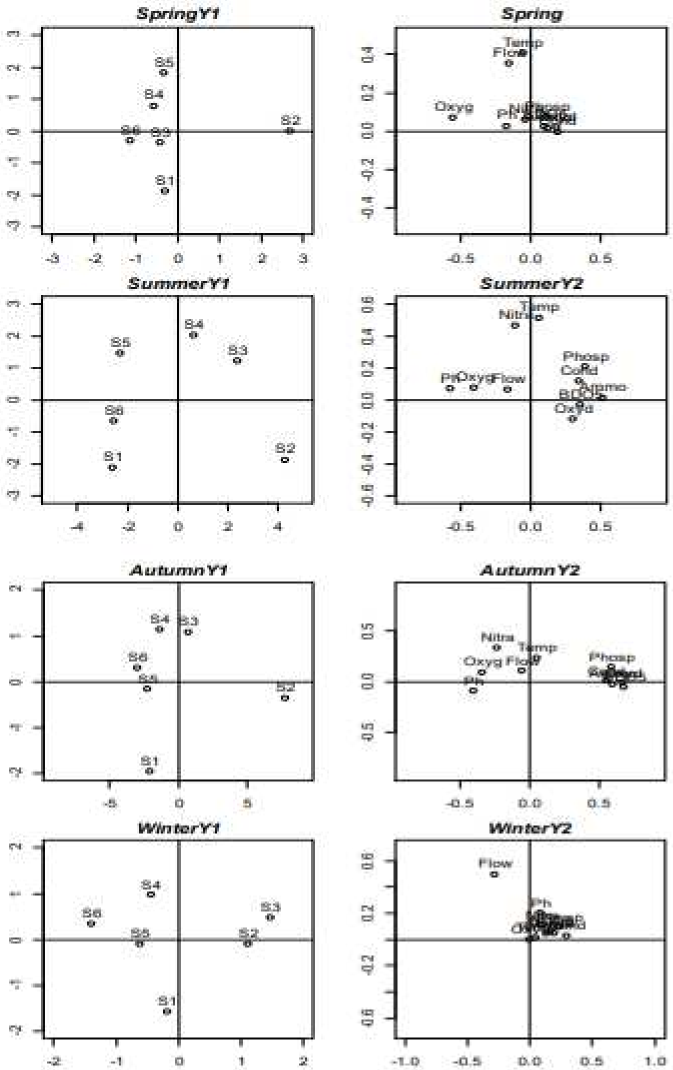

Fig. 3. Representation of the stations (Fig3Y1) and the environmental variables (Fig3Y2) per season on the first two axes for multigroup Y of SOCIA3

Station S2 is characterized by the phosphates and ammonium in Autumn. Near the center marks, we find the rare species that are not taken into account by the SOCIA3. Axis 2 opposes generally one hand station S1 and stations S4 and S6 other hand.

In contrast to the SOCIA3 method the positions of the environmental variables are generally those of the previous methods mentioned above.

3.2. Successive generalized co-inertia analysis (SGCIA)

In this subsection we apply the SGCIA method to the ecological dataset. For SGCIA this is the case where $M = 4$ and $N = 2$.

Table 3. *Squared correlations between linear combinations of the variables in fish abundance and environmental variables to order 1 and 2 for SGCIA*

	Seasons			
Methods	Spring	Summer	Autumn	Winter
SGCIA	0.375	0.878	0.934	0.330
	0.431	0.716	0.817	0.403

Tables 3 and 4 provide the same results as Tables 1 and 2. Autumn has a higher squared correlation than Summer. On the first axis, this variation is 0.056 and on the second axis, 0.101. With the SOCIA3 method, this variation is 0.058 on the first axis and 0.102 on the second axis.

Table 4. *Percentages of projected inertia (specific weights) of each season on the first two axes*

	Seasons			
Methods	Spring	Summer	Autumn	Winter
	11.100	14.292	43.947	30.658
SGCIA (X)	9.521	32.679	47.328	10.470
	6.725	31.022	57.926	4.324
SGCIA (Y)	22.657	47.301	18.949	11.091

The same results apply to SOCIA3, where the squared correlation is much higher in Autumn.

Table 4 gives a complete overview of the variability at each date for the two matrices X and Y.

Thus, for the first X matrix, on the first axis, we find the percentages of explained inertia 11.1%, 14.29%, 43.95% and 30.66% and for the second Y matrix we have: 6.72%, 31.02%, 57.93% and 4.32%. We can see that mesological variability and faunistical diversity are low in Winter and spring, while Autumn has the highest percentage. The two matrices X and Y have similar results, i.e., there is a common structure between the two matrices.

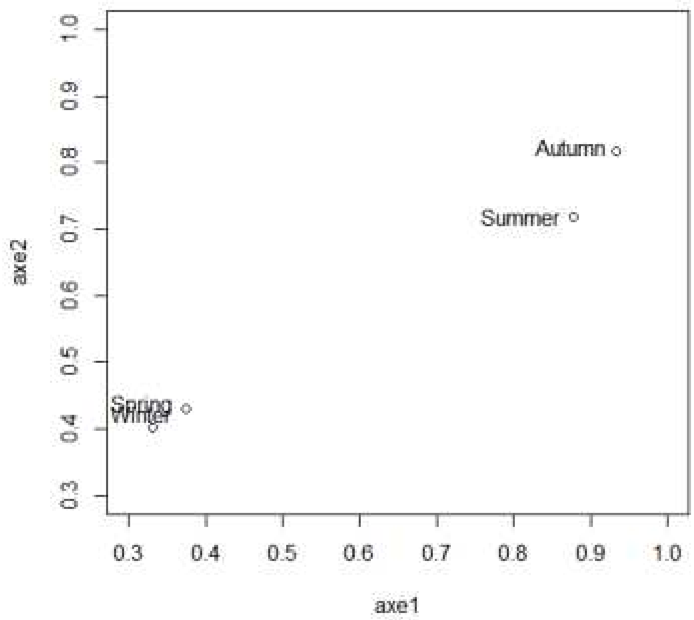

Fig. 4. Position of the seasons on the first two axes of squared correlations between partial synthetic components species-environment for SGCIA

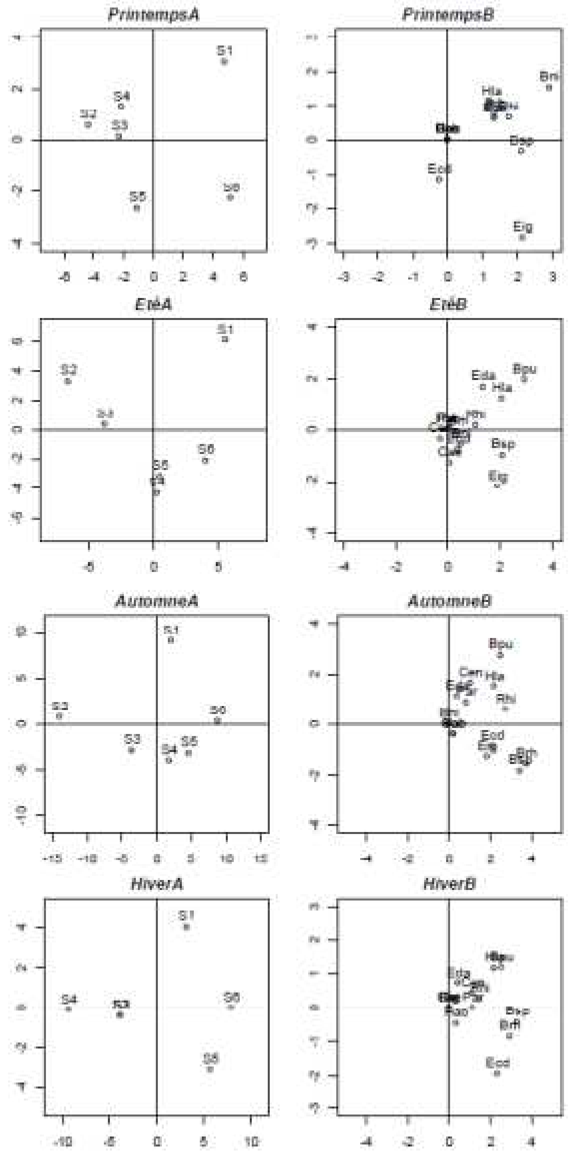

Fig. 5. Position of the stations (Fig5A) and the species (Fig5B) per season on the first two axes for multigroup X of SGCIA

Figure 5 (fig5A) shows the position of the stations on the first two co-inertia axes defined by the SGCIA. As with SOCIA3, we note an overall size effect on axis 1 (fig5B). Axis 2 is an axis of opposition between stations S2 and S6.

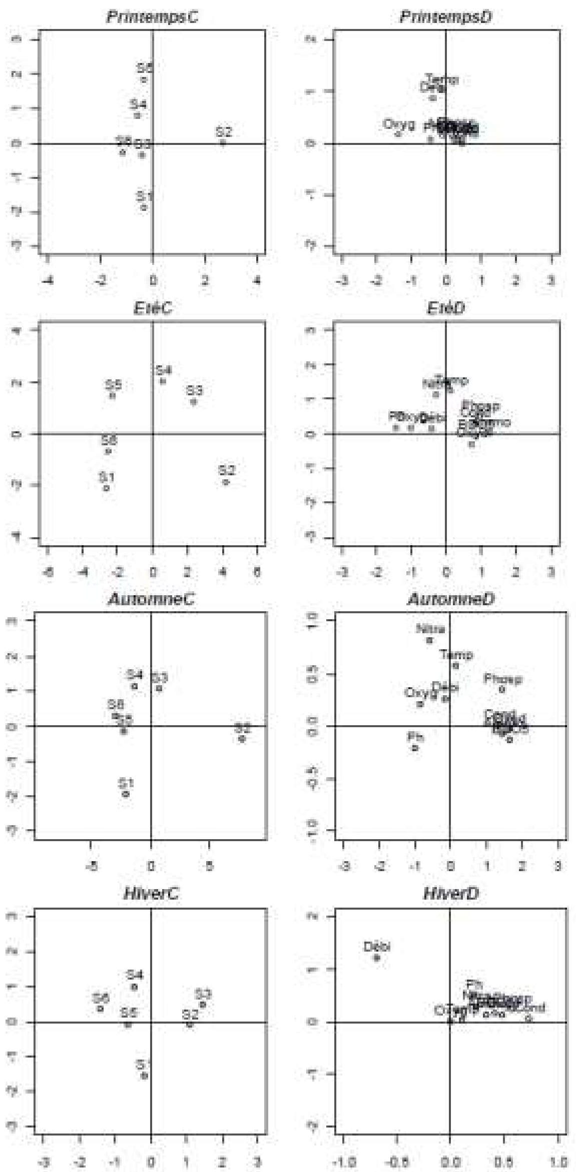

Fig. 6. Representation of the stations (Fig6C) and the environmental variables (Fig6D) per season on the first two axes for multigroup Y of SGCIA

Figure 6 (fig6C) shows the position of stations on the first two axes. As with the SOCIA3 method, a rise in Spring flows is observed at stations s4 and s5. For the SGCIA and SOCIA3 analyses, axis 1 is a pollution gradient axis and axis 2 a restoration gradient axis.

4. Conclusion

The generalizations developed allow us to find several methods for analyzing multivariate data.

The methods consist in searching for table components and co-inertia axes that are common to each multibloc or multigroup, enabling projected inertias and correlation coefficients to be calculated between table pairs.
The advantage of these methods is their simplicity. Determining the solution requires a simple diagonalization of the matrices.

The two methods presented here uncover the same features in the example data set. This is a small data set, but with strong structure, and strong structures often are clear with any method. However, the two methods used to analyze even a data set with clear structure can have advantages and drawbacks.

References

1. Carroll J.D. A generalization of canonical correlation analysis to three or more sets of variables. *Proceedings of the 76th Annual Convention of the American Psychological Association*, Washington DC, 1968, pp. 227–228.
2. Chessel D., Hanafi M. Analyses de la co-inertie de K nuages de points. *Rev. Stat. Appl.*, 1996, **44** (2), pp. 35–60.
3. Eslami A., Qannari E.M., Kohler A., Bougeard S. Analyses factorielles de données structurées en groupes d'individus. *J. Soc. Fr. Stat.*, 2013, **154** (3), pp. 44–57.
4. Hanafi M., El Hadri Z., Sahli A., Dolce P. Overcoming convergence problems in PLS path modelling. *Comput. Stat.*, 2022, **37**, pp. 2437–2470.
5. Horst P. Relations among m sets of variables. *Psychometrika*, 1961, **26**, pp. 126–149.
6. Hotelling H. Relations between two sets of variables. *Biometrika*, 1936, **28** (3-4), pp. 321–377.
7. Kettenring J.R. Canonical analysis of several sets of variables. *Biometrika*, 1971, **58** (3), pp. 433–451.
8. Kissita G., Malouata R.O., Mizère D., Makany R.A. Proposition of analyses in a vertical mulpi-table and analyses of links between two vertical multi-tables : methods (sVMA and sOVMA) and (sCIA3 et sOCIA3). *Appl. Math. Sci.*, 2013, **7** (131), pp. 6503–6525.
9. Krzanowski W.J. Principal component analysis in the presence of group structure. *Appl. Stat.*, 1984, **33** (2), pp. 164–168.
10. Lafosse R., Hanafi M. Concordance d'un tableau avec K tableaux: définition de $K+1$ uplés synthétiques. *Rev. Stat. Appl.*, 1997, **45** (4), pp. 111–126.
11. Martinez-Ruiz A., Lauro N.C. Incremental singular value decomposition for some numerical aspects of multiblock redundancy analysis. *Comput. Stat.*, 2023. https://doi.org/10.1007/s00180-023-01418-5.
12. Pearson K. On lines and planes of closest fit to systems of points in space. *The London, Edinurgh, and Dublin Philosophical Magazine and Journal of Science*, 1901, **2** (11), pp. 559-572.

13. Pegaz-Maucet D. *Impact d'une Perturbation d'Origine Organique sur la Dérive des Macro-Invertébrés d'un Cours d'Eau. Comparaison avec le Benthos.* Thése spécialité, Université Lyon I., France, 1980.
14. Simier M., Blanc L., Pellegrin F., Nandris D. Approche simultanée de K couples de tableaux: application á l'étude des relations pathologie végétale - environnement. *Rev. Stat. Appl.*, 1999, **47** (1), pp. 31–46.
15. Takane Y., Hwang H., Abdi H. Regularized multiple-set canonical correlation analysis. *Psychometrika*, 2008, **73** (4), pp. 753–775.
16. Tenenhaus A., Tenenhaus M. Regularized generalized canonical correlation analysis. *Psychometrika*, 2011, **76** (2), pp. 257–284.
17. Tenenhaus M., Tenenhaus A., Grœnen P.J.F. Regularized generalized canonical correlation analysis: a framework for sequential multiblock component methods. *Psychometrika*, 2017, **82** (3), pp. 737–777.
18. Thioulouse J., Chessel D. Les analyses multitableaux en écologie factorielle. *Acta Oecologica, Oecologia Generalis*, 1987, **8** (4), pp. 463–480.
19. Tucker L.R. An inter-battery method of factor analysis. *Psychometrika*, 1958, **23** (2), pp. 111–136.