



Proposition d'analyses de co-inertie entre deux multi-tableaux verticaux: les méthodes ACIs et ACIOs.

Léonard Niéré[†], Gabriel Kissita[†] and Dominique Mizere^{‡,*}

[†]Institut supérieur de Gestion

[‡]Faculté des Sciences et Techniques. Université Marien Ngouabi-Brazzaville-Congo, BP 69

Received: July 9, 2013; Accepted: November 15, 2013

Copyright © 2013, Afrika Statistika. All rights reserved

Abstract. The analysis of Co-inertia between two tables was generalized with the simultaneous analysis several couples of triplets of tables in STATICO in the link between two vertical multi-tables. In this article, being given the context of STATICO, we propose an analysis of successive Co-inertia (ACIs). The axes of Co-inertia which are given are common to all the tables. To interpret the variables of each table, we propose the orthogonal version of this analysis of Co-inertia which we call ACIOs. We apply this method to the real data.

Résumé. L'analyse de co-inertie entre deux tableaux a été généralisée l'analyse simultanée de plusieurs couples de triplets de tableaux dans STATICO dans le lien entre deux multi-tableaux verticaux. Dans cet article, étant donné le contexte de STATICO, nous proposons une analyse de co-inertie successive (ACIs). Les axes de co-inertie qui sont déterminés sont communs à tous les tableaux. Pour interpréter les variables de chaque tableau, nous proposons la version orthogonale de cette analyse de co-inertie que nous appelons ACIOs. Nous appliquons cette méthode aux données réelles.

Key words: Data analysis; Common component; Specific weights, M-tables; STATIS; Co-inertia; Couple tables; STATICO.

AMS 2010 Mathematics Subject Classification : 62M20; 62N05.

*Corresponding author Gabriel Kissita: gakissita@yahoo.fr

Léonard Niere : lniere@yahoo.fr

Dominique Mizere : do.mizere@gmail.com

1. Introduction

L'analyse de co-inertie (Chessel et Mercier, 1993; Dolédec et Chessel, 1994) de deux tableaux de données portant sur les mêmes objets donne lieu à deux typologies distinctes de ces objets engendrés respectivement par chacun des groupes des descripteurs. Lorsque cette double étude est renouvelée à plusieurs reprises, par exemple dans le temps ou dans l'espace, on peut chercher à caractériser la stabilité des relations observées entre ces deux groupes de descripteurs.

Une première voie de réponse est explorée dans les analyses de co-inertie inter-classes et intra-classes proposées par Franquet et Chessel (1994), Franquet *et al.* (1995).

Simier *et al.* (1999) utilisent l'extension de la méthode ACT-STATIS (Analyse Conjointe de Tableaux-Structuration des Tableaux A Trois Indices de la Statistique), décrite par Lavit *et al.* (1994) aux opérateurs de co-inertie. Cette méthode, appelée STATICO (pour STATIS et co-inertie), permet de traiter simultanément une série de M couples de tableaux par une analyse triadique partielle portant sur les tableaux croisés des M analyses de co-inertie. On combine ainsi la logique de couplage de co-inertie (trouver ce qui dans deux groupes de descripteurs engendre une typologie commune des objets décrits dans chacun des tableaux). Pour plus de détails en ce qui concerne la méthode STATICO, on peut se référer à Hanafi (1997) et à Simier *et al.* (1999). En outre, pour étudier le lien entre deux multi-tableaux verticaux cubiques, Vivien (2002) et Vivien et Sabatier (2004) ont proposé une généralisation de l'analyse ACT-STATIS nommée DO-ACT.

Une deuxième voie de réponse pour analyser ce type de données est l'analyse canonique entre deux multi-tableaux verticaux (Kissita *et al.*, 2013). Compte tenu du fait que cette méthode peut présenter une certaine instabilité en présence de multi-colinéarités et aussi du nombre des individus relatifs aux sous-tableaux qui peuvent être strictement inférieur au nombre des variables du premier et du deuxième groupe, il est nécessaire de recourir à une autre approche en proposant une autre méthode dans ce contexte. Cette approche constitue une autre contribution pour l'analyse de ce type de données que nous nommons l'analyse de co-inertie successive dans le couplage des multi-tableaux verticaux (ACIs).

Dans cet article, nous reprenons exactement le contexte de STATICO, c'est-à-dire rappelons-le, on renouvelle l'analyse de co-inertie de deux tableaux de données portant sur les mêmes objets à M couples de triplets de tableaux. On peut aussi ici avoir les observations modifiées à chaque renouvellement de l'analyse. Le problème ici, est de déterminer les axes de co-inertie dans ce contexte en vue de caractériser les objets dans les systèmes communs de représentation. Ces axes de co-inertie communs à tous les tableaux peuvent être assimilés aux composantes communes comme c'était le cas dans le cadre d'un M -multi-tableaux ayant en commun des individus en ligne, se reporter à Qannari *et al.*, (2000, 2001) et Hanafiet Qanari (2008). Ces axes permettent de calculer des coefficients de corrélation au carré entre composantes partielles des sous tableaux de deux multi-tableaux et des poids spécifiques de chaque sous tableau qui sont en réalité des inerties projetées. Compte tenu du fait que les systèmes des composantes partielles de l'ACIs ne sont pas orthogonaux, nous allons proposer l'analyse de co-inertie orthogonale successive (ACIOs).

Ainsi, cet article s'organise de la manière suivante :

Au paragraphe 2, nous rappelons l'analyse canonique entre deux multi-tableaux verticaux. Tout ceci dans le seul but de montrer ses limites. Au paragraphe 3, nous proposons une autre méthode dans ce contexte. Cette méthode permet de déterminer les axes de co-inertie. Au paragraphe 4, nous proposons l'analyse de co-inertie orthogonale successive. Au paragraphe 5, nous faisons quelques liens de ces méthodes avec d'autres méthodes d'analyse de données. Enfin, pour montrer l'intérêt pratique de cette méthode, nous traitons au paragraphe 6 les données écologiques déjà traitées par Pegaz-Maucet (1980) concernant le fleuve Méaudret en France. Cette étude permet de caractériser la typologie des stations en fonction de l'évolution temporelle de la liaison entre l'espèce et l'environnement. Les programmes ont été réalisés en langage scilab et les graphiques en langage R.

2. L'analyse canonique entre deux multi-tableaux verticaux

On considère ici deux multi-tableaux $X = [X'_1|X'_2|\dots|X'_M]'$ et $Y = [Y'_1|Y'_2|\dots|Y'_M]'$ associés respectivement au groupe 1 et au groupe 2 tels que définis dans la méthode STATICO, rappelons-le. On appelle analyse canonique entre deux multi-tableaux verticaux, la recherche des vecteurs u et v respectivement de \mathbb{R}^p et \mathbb{R}^q maximisant la fonction

$$\begin{aligned} f(u, v) &= \sum_{m=1}^M cov(X_m Qu, Y_m Rv) = \sum_{m=1}^M (u' Q V_{X_m Y_m} Rv) \\ &= u' Q \left(\sum_{m=1}^M V_{X_m Y_m} \right) Rv = u' Q V_{XY} Rv = cov(X Qu, Y Rv) \end{aligned} \quad (1)$$

sous les contraintes de normes unité sur les composantes globales

$$var(X Qu) = \|c_X\|_D^2 = c'_X D c_X = u' Q \left(\sum_{m=1}^M V_{X_m} \right) Qu = u' Q V_X Qu = 1, \quad (2)$$

$$var(Y Rv) = \|c_Y\|_D^2 = c'_Y D c_Y = v' R \left(\sum_{m=1}^M V_{Y_m} \right) Rv = v' R V_Y Rv = 1 \quad (3)$$

où $V_X = \sum_{m=1}^M V_{X_m} = X' D X$ et $V_Y = \sum_{m=1}^M V_{Y_m} = Y' D Y$ les matrices des variances-covariances des multi-tableaux X et Y et, $V_{XY} = \sum_{m=1}^M V_{X_m Y_m}$ la matrice des inter-covariances entre les multi-tableaux X et Y , et $D = diag(D_m, 1 \leq m \leq M)$ la métrique des poids bloc-diagonale, rappelons-le.

La maximisation de (1) sous les contraintes (2) et (3) conduit à l'ordre s ($s = 1, \dots, \min(p, q)$) aux équations stationnaires dans les espaces des facteurs $(\mathbb{R}^p)^*$ et $(\mathbb{R}^q)^*$

$$V_X^{-1} V_{XY} V_Y^{-1} V_{YX} a_s = \alpha_s^2 a_s \quad (4)$$

$$V_Y^{-1} V_{YX} V_X^{-1} V_{XY} b_s = \alpha_s^2 b_s \quad (5)$$

ou encore

$$\left(\sum_{m=1}^M V_{X_m} \right)^{-1} \left(\sum_{m=1}^M V_{X_m Y_m} \right) \left(\sum_{m=1}^M V_{Y_m} \right)^{-1} \left(\sum_{m=1}^M V_{Y_m X_m} \right) a_s = \alpha_s^2 a_s, \quad (6)$$

$$\left(\sum_{m=1}^M V_{Y_m}\right)^{-1} \left(\sum_{m=1}^M V_{Y_m X_m}\right) \left(\sum_{m=1}^M V_{X_m}\right)^{-1} \left(\sum_{m=1}^M V_{X_m Y_m}\right) b_s = \alpha_s^2 b_s, \quad (7)$$

où $a_s = Qu_s$ et $b_s = Rv_s$, et donc $u_s = Q^{-1}a_s$ et $v_s = R^{-1}b_s$ les axes de co-inertie.

Dans l'espace des variables, on trouve les équations stationnaires suivantes :

$$P_X P_Y c_{X,s} = \alpha_s^2 c_{X,s} \quad (8)$$

$$P_Y P_X c_{Y,s} = \alpha_s^2 c_{Y,s} \quad (9)$$

où $c_{X,s} = XQu_s = Xa_s$ et $c_{Y,s} = YRv_s = Yb_s$ sont les composantes synthétiques associées aux tableaux X et Y respectivement, P_X et P_Y désignant les projecteurs D -orthogonaux sur les sous-espaces de \mathbb{R}^n engendrés respectivement sur les colonnes de X et de Y .

Dans l'analyse canonique entre deux multi-tableaux verticaux X et Y , l'inversion des matrices de variance-covariances peut être impossible au cas où les tableaux X_m et Y_m ne sont pas de plein rang et dans le cas où le nombre des individus est strictement inférieur au nombre de variables des tableaux X_m et Y_m . Dans ce cas, le calcul des facteurs canoniques et des variables canoniques associées devient très délicat.

Après avoir montré les limites de cette approche dans l'analyse simultanée des M couples de triplets de tableaux surtout en écologie dans la recherche spatio-temporelle de la relation faune et milieu où le nombre de descripteurs est souvent très grand par rapport au nombre d'individus, il est nécessaire de proposer une autre méthode adaptée à cette problématique.

3. L'analyse de co-inertie entre deux multi-tableaux verticaux (ACIs)

On considère M couples de triplets $\{(X_m, Q, D_m), (Y_m, R, D_m)\}$, pour $m = 1, \dots, M$, X_m est un tableau de dimensions (n_m, p) dont les lignes des observations des n_m individus et dont les colonnes sont les mesures de p variables sur ces n_m individus, Y_m est un tableau de dimensions (n_m, q) portant les mêmes individus que le tableau X_m . On se place donc dans le même contexte que celui de la méthode STATICO et qui est aussi celui de l'analyse canonique entre deux multi-tableaux verticaux.

L'analyse de la co-inertie entre deux multi-tableaux verticaux (ACIs) (la lettre s fait référence à successive) avec le critère de covariance, permet de caractériser la stabilité des relations observées entre ces deux groupes de descripteurs. Elle permet de déterminer les axes de co-inertie u du premier groupe et v du second groupe, vecteurs respectivement de \mathbb{R}^p et \mathbb{R}^q réalisant le maximum de la fonction f

$$f(u, v) = \sum_{m=1}^M cov(X_m Qu, Y_m Rv) \quad (10)$$

sous les contraintes de normalisation :

$$\|u\|_Q^2 = \|v\|_R^2 = 1 \quad (11)$$

La fonction (10) peut aussi se mettre sous la forme

$$f(u, v) = \sum_{m=1}^M u' Q V_{X_m Y_m} R v = u' Q \left(\sum_{m=1}^M V_{X_m Y_m} \right) R v. \quad (12)$$

Proposition 1. *Les axes de co-inertie u et v de l'ACIs vérifient les équations stationnaires*

$$\left(\sum_{m=1}^M V_{X_m Y_m} \right) R \left(\sum_{m=1}^M V_{Y_m X_m} \right) Qu = \alpha^2 u \quad (13)$$

$$\left(\sum_{m=1}^M V_{Y_m X_m} \right) Q \left(\sum_{m=1}^M V_{X_m Y_m} \right) Rv = \alpha^2 v. \quad (14)$$

En effet, la maximisation de (12) sous les contraintes de normalisation (11) définies précédemment, revient à maximiser le lagrangien de f défini par la fonction ci-dessous

$$L = u'Q \left(\sum_{m=1}^M V_{X_m Y_m} \right) Rv + \alpha_1(1 - u'Qu) + \alpha_2(1 - v'Rv),$$

où α_1 et α_2 sont les multiplicateurs de Lagrange associés aux contraintes de (12).

En dérivant le lagrangien respectivement par rapport à u et v , on obtient les équations normales suivantes :

$$\frac{\partial L}{\partial u} = Q \left(\sum_{m=1}^M V_{X_m Y_m} \right) Rv - 2\alpha_1 Qu = 0 \quad (15)$$

$$\frac{\partial L}{\partial v} = R \left(\sum_{m=1}^M V_{Y_m X_m} \right) Qu - 2\alpha_2 Rv = 0 \quad (16)$$

$$\frac{\partial L}{\partial \alpha_1} = 1 - u'Qu = 0 \quad (17)$$

$$\frac{\partial L}{\partial \alpha_2} = 1 - v'Rv = 0 \quad (18)$$

Par suite, on trouve les relations suivantes

$$Q \left(\sum_{m=1}^M V_{X_m Y_m} \right) Rv = 2\alpha_1 Qu, \quad (19)$$

$$R \left(\sum_{m=1}^M V_{Y_m X_m} \right) Qu = 2\alpha_2 Rv. \quad (20)$$

En pré-multipliant à gauche respectivement (15) par u' et (16) par v' , on montre aisément que

$$u'Q \left(\sum_{m=1}^M V_{X_m Y_m} \right) Rv = 2\alpha_1 = 2\alpha_2 = \alpha. \quad (21)$$

Par conséquent, on trouve les relations de transition suivantes

$$\left(\sum_{m=1}^M V_{X_m Y_m} \right) Rv = \alpha u, \quad (22)$$

$$\left(\sum_{m=1}^M V_{Y_m X_m} \right) Qu = \alpha v. \quad (23)$$

En combinant les relations (22) et (23), on trouve les relations (13) et (14) recherchées. Comme $V_{XY} = X'DY = \sum_{m=1}^M V_{X_m Y_m}$, ainsi la fonction à maximiser devient

$$\begin{aligned} f(u, v) &= \sum_{m=1}^M u'QV_{X_m Y_m}Rv = u'Q \left(\sum_{m=1}^M V_{X_m Y_m} \right) Rv \\ &= u'QV_{XY}Rv = cov(XQu, YRv) \end{aligned}$$

Les équations stationnaires définies aux (13) et (14) dans la proposition 1 ci-dessus deviennent :

$$V_{XY}RV_{YX}Qu = \alpha^2 u, \tag{24}$$

$$V_{YX}QV_{XY}Rv = \alpha^2 v. \tag{25}$$

Les relations de transition (22) et (23) peuvent en s'écrire de la manière suivante :

$$V_{XY}Rv = \alpha u, \tag{26}$$

$$V_{YX}Qu = \alpha v. \tag{27}$$

Les équations stationnaires (24) et (25) sont celles trouvées par [Chessel et Mercier \(1993\)](#) dans l'analyse de co-inertie entre deux triplets de tableaux (X, Q, D) et (Y, R, D) .

Ainsi, l'analyse de co-inertie entre deux multi-tableaux verticaux n'est rien d'autre que l'analyse de co-inertie entre les triplets (X, Q, D) et (Y, R, D) . C'est vraiment une généralisation de l'analyse de co-inertie de deux multi-tableaux verticaux.

Ainsi, α^2 est la plus grande valeur propre de $V_{XY}RV_{YX}Q$, et α est un maximum local ou global sur l'espace défini par les contraintes. En pratique, il n'est pas obligé de diagonaliser les deux matrices qui sont définies dans (24) et (25). On peut seulement diagonaliser l'une des deux matrices et déterminer l'autre valeur u et/ou v en utilisant les relations (26) et (27).

Le critère maximisé dans (10) ou encore $cov(XQu, YRv)$ à l'ordre un sous les contraintes de norme unité pour les axes u et v est celui maximisé dans la régression PLS d'ordre un de [Wold \(1985\)](#). La fonction qui est définie en (10) peut encore s'écrire sous la forme suivante :

$$cov^2(XQu, YRv) = var(XQu)var(YRv)cor^2(XQu, YRv).$$

Ce qui veut dire que dans la recherche de la liaison entre les deux multi-tableaux verticaux il est aussi question d'étudier la structure interne de deux multi-tableaux verticaux.

Les matrices permettant de trouver les axes de co-inertie u et v ne sont pas symétriques, il n'est pas possible de construire des bases orthogonales à partir de ces vecteurs.

Proposition 2. *Les équations stationnaires à l'ordre s de l'analyse de co-inertie entre deux multi-tableaux verticaux définies par les relations (24) et (25) deviennent*

$$Q^{1/2}V_{XY}RV_{YX}Q^{1/2}c_s = \alpha_s^2 c_s \tag{28}$$

$$R^{1/2}V_{YX}QV_{XY}R^{1/2}d_s = \alpha_s^2 d_s \tag{29}$$

En effet, pour construire ces bases orthogonales, on pose respectivement $u_s = Q^{-1/2}c_s$ et $v_s = R^{-1/2}d_s$ dans les relations (24) et (25) respectivement où s est l'ordre de la solution ($s = 1, \dots, r; r \leq \min(p, q)$) et en pré-multipliant respectivement la première équation par $Q^{1/2}$ et la deuxième par $R^{1/2}$, il vient les équations stationnaires définies au (28) et (29).

Les vecteurs c_s et d_s sont normés. En effet, vérifions-le pour les vecteurs c_s . Les vecteurs u_s étant Q -normés ou encore $u'_s Q u_s = 1$, compte tenu de ce que $u_s = Q^{-1/2}c_s$, on a $u'_s Q u_s = c'_s c_s = 1$. On montre de la même manière que $d'_s d_s = 1$.

Par conséquent, les bases $\{c_s\}_s$ et $\{d_s\}_s$ sont orthonormées dans \mathbb{R}^p et \mathbb{R}^q . On en déduit que la base $\{u_s\}_s$ (respectivement $\{v_s\}_s$) sont Q -orthonormée dans \mathbb{R}^p (respectivement R -orthonormée dans \mathbb{R}^q).

Dans l'espace des variables, les équations aux valeurs propres (24) et (25) à l'ordre s peuvent s'écrire de la façon suivante :

Proposition 3. *Les équations stationnaires à l'ordre s de l'analyse de co-inertie entre deux multi-tableaux verticaux dans l'espace des variables sont définies par les relations*

$$W_X D W_Y D c_{X,s} = \alpha_s^2 c_{X,s} \quad (30)$$

$$W_Y D W_X D c_{Y,s} = \alpha_s^2 c_{Y,s} \quad (31)$$

où $c_{X,s} = X Q u_s$ et $c_{Y,s} = Y R v_s$ sont des vecteurs de \mathbb{R}^n et $W_X = X Q X'$ et $W_Y = Y R Y'$ sont des matrices de produits scalaires entre individus de X et Y respectivement.

Les composantes partielles $c_{X_{m,s}} = X_m Q u_s$ et $c_{Y_{m,s}} = Y_m R v_s$ ne sont pas D_m -orthogonales. En revanche, les axes de co-inertie u_s et v_s sont Q -orthogonaux respectivement.

Afin que la variation expliquée de chaque tableau soit différent à chaque étape, nous allons définir une autre façon d trouver les axes de co-inertie dans le couplage de deux multi-tableaux verticaux.

4. L'analyse de co-inertie orthogonale entre deux multi-tableaux verticaux : la méthode ACIOs

Dans ce paragraphe, nous allons définir une nouvelle méthode qui permet de déterminer les axes de co-inertie dans le couplage de deux multi-tableaux verticaux : l'analyse de co-inertie successive orthogonale entre deux multi-tableaux verticaux (ACIOs). Le terme "orthogonale" fait référence à l'orthogonalité des composantes synthétiques partielles pour les deux groupes de variables obtenues à l'ordre strictement supérieur à un.

L'objet de cette méthode est identique au précédent, ou encore d'étudier la stabilité de la liaison entre deux multi-tableaux verticaux en expliquant au mieux les tableaux individuellement au niveau des individus et des variables.

4.1. Solution d'ordre 1 de l'ACIOs

A l'ordre un, l'analyse de co-inertie successive orthogonale entre multi-tableaux verticaux est confondue à l'analyse de co-inertie successive entre deux multi-tableaux verticaux.

4.2. Solution à l'ordre s de l'ACIOs

A l'ordre s les solutions sont définies de la manière suivante : les axes de co-inertie d'ordre s , avec $s = 2, \dots, \min(p, q)$, de l'ACIOs sont les axes d'ordre 1 de l'analyse de co-inertie dans le couplage de deux multi-tableaux verticaux des triplets $(X^{(s-1)}, Q, D)$ et $(Y^{(s-1)}, R, D)$ où $X^{(s-1)} = [X_1^{(s-1)'}|X_2^{(s-1)' }|\dots|X_M^{(s-1)'}]'$ et $Y^{(s-1)} = [Y_1^{(s-1)' }|Y_2^{(s-1)' }|\dots|Y_M^{(s-1)' }]'$ avec $X_m^{(s-1)} = (I_{n_m} - P_{c_{X_m, s-1}})X_m^{(s-2)}$ et $Y_m^{(s-1)} = (I_{n_m} - P_{c_{Y_m, s-1}})Y_m^{(s-2)}$, $P_{c_{X_m, s-1}}$ et $P_{c_{Y_m, s-1}}$ les projecteurs D_m -orthogonaux respectivement sur les sous-espaces engendrés par les composantes synthétiques $c_{X_m, s-1} = X_m^{(s-1)}Qu_s$ et $c_{Y_m, s-1} = Y_m^{(s-1)}Rv_s$. Les relations de transition et les équations stationnaires concernant cette méthode sont données par

$$V_{X^{(s-1)}Y^{(s-1)}}Rv_s = \alpha_s u_s, \tag{32}$$

$$V_{Y^{(s-1)}X^{(s-1)}}Qu_s = \alpha_s v_s, \tag{33}$$

$$V_{X^{(s-1)}Y^{(s-1)}}RV_{Y^{(s-1)}X^{(s-1)}}Qu_s = \alpha_s^2 u_s, \tag{34}$$

$$V_{Y^{(s-1)}X^{(s-1)}}QV_{X^{(s-1)}Y^{(s-1)}}Rv_s = \alpha_s^2 v_s. \tag{35}$$

Le principe de cette façon de faire consiste à chaque étape à ne retenir que les axes de co-inertie u_s et v_s qui correspondent à la plus grande valeur propre et aux variables synthétiques partielles associées.

Proposition 4. *Les systèmes des axes de co-inertie $\{u_s\}_s$ et $\{v_s\}_s$ sont Q -orthogonaux et R -orthogonaux respectivement.*

En effet, en posant pour tout $m = 1, \dots, M$, $X_m^{(0)} = X_m$ et $Y_m^{(0)} = Y_m$, la transposée de l'équation de transition (35) de l'analyse de co-inertie orthogonale entre deux multi-tableaux à l'ordre s peut s'écrire

$$\alpha_s u'_s = v'_s RV_{Y^{(s-1)}X^{(s-1)}}. \tag{36}$$

En multipliant à droite la relation (32) par Qu_l pour tout $l = 1, \dots, s - 1$ dans les deux membres, on aboutit à la relation suivante :

$$\begin{aligned} \alpha_s u'_s Qu_l &= v'_s RV_{Y^{(s-1)}X^{(s-1)}}Qu_l = v'_s RY^{(s-1)}DX^{(s-1)}Qu_l \\ &= v'_s RY^{(s-1)}D \left[u'_l QX_1^{(s-1)' }|\dots|u'_l QX_m^{(s-1)' }|\dots|u'_l QX_M^{(s-1)' } \right] \end{aligned}$$

Or $X_m^{(s-1)}Qu_l = \left(\prod_{h=l+1}^{s-1} P_{c_{X_m, h}}^\perp \right) P_{c_{X_m, l}}^\perp X_m^{(l-1)}Qu_l = 0$ pour tous $l = 1, \dots, s - 1$ et $m = 1, \dots, M$, puisque $P_{c_{X_m, l}}^\perp c_{X_m, l} = 0$.

Comme $\alpha_s \neq 0$, nous avons $u'_s Qu_l = 0$, d'où l'orthogonalité des axes de co-inertie u_s .

Une preuve analogue à celle permettant d'établir l'orthogonalité des axes u_s permet de montrer l'orthogonalité des axes v_s .

Cette méthode est appelée analyse de co-inertie orthogonale successive (ACIOs) entre deux multi-tableaux verticaux. Elle présente plus d'intérêt par rapport à l'analyse de co-inertie entre deux multi-tableaux verticaux, en ce sens que l'ACIOs est interprétée au niveau des individus et au niveau des variables, mais ces deux méthodes sont confondues à l'ordre 1.

5. Liens avec d'autres méthodes d'analyse de données

Dans ce paragraphe, nous allons établir les liens entre les méthodes de couplage de multi-tableaux verticaux avec les méthodes classiques d'analyse de données.

1. On remarque que l'analyse de co-inertie entre deux multi-tableaux verticaux est une analyse en composantes communes et poids spécifiques. En effet, aux tableaux X_m et Y_m pour tout $m = 1, \dots, M$ sont respectivement associés les axes de co-inertie u et v qui ne dépendent pas de m .
2. L'analyse de co-inertie entre deux multi-tableaux verticaux est très proche de STATICO au niveau de l'analyse de la moyenne, car l'analyse de la moyenne est une analyse de co-inertie du triplet moyen $(\sum_{m=1}^M \alpha_m X'_m D_m Y_m, Q, R)$ qui est définie à l'étape précédente de la définition de la moyenne en suivant la démarche de STATICO. Les pondérations α_m vérifient la contrainte de normalisation $\sum_{m=1}^M \alpha_m^2 = 1$, à un coefficient près, à l'infrastructure dans STATIS duale, puisque l'analyse du compromis correspond à l'ACP du multi-tableau vertical X muni des métriques Q et $D = \text{diag}(\sqrt{\alpha_m} D_m, m = 1, \dots, M)$.
3. Dans les espaces $(\mathbb{R}^n)^*$ et $(\mathbb{R}^q)^*$ les équations stationnaires (24) et (25) s'écrivent respectivement $QV_{XY}RV_{YX}a = \alpha^2 a$ et $RV_{YX}QV_{XY}b = \alpha^2 b$, où $a = Qu$ et $b = Rv$.
4. Si $Q = I_p$ et $R = I_q$, l'analyse définie au (24) et (25) devient l'analyse inter-batterie de Tucker (1958xxx) des multi-tableaux verticaux X et Y . Les équations stationnaires deviennent $V_{XY}V_{YX}u = \alpha^2 u$ et $V_{YX}V_{XY}v = \alpha^2 v$.
5. Si les deux multi-tableaux X et Y sont confondus ($X = Y$), le critère de l'analyse de co-inertie devient le critère de l'analyse en composantes principales d'un multi-tableau vertical. En effet, l'analyse en composantes principales d'un multi-tableau vertical est la recherche d'un vecteur u de \mathbb{R}^p maximisant la fonction

$$f(u) = \sum_{m=1}^M \text{var}(c_m) = \sum_{m=1}^M \text{var}(X_m Qu) = u' Q V_X Qu \quad (37)$$

sous la contrainte de norme unité

$$\|u\|_Q^2 = 1 \quad (38)$$

où $c_m = X_m Qu$ est la composante synthétique correspondant au tableau X_m et $V_X = \sum_{m=1}^M V_{X_m} = X' D X$ la matrice des variances-covariances de X . $V_m = X'_m D_m X_m$ est la matrice des variance-covariances du tableau X_m et $D_m = \frac{1}{n_m} I_{n_m}$ la matrice des poids des individus du tableau X_m et $D = \text{diag}(D_m, m = 1, \dots, M)$ pour tout $m = 1, \dots, M$.

Proposition 5. *La solution u du problème (37) sous la contrainte (38) est donnée par l'équation suivante*

$$V_X Qu = \theta u, \quad (39)$$

ou encore

$$\left(\sum_{m=1}^M V_m \right) Qu = \theta u.$$

La relation (39) est équivalente à l'ACP du triplet (X, Q, D) .

En revanche, dans l'espace des variables, la relation (39) devient $W_X D c_X = \theta c_X$ où $W_X = X Q X'$ est la matrice des produits scalaires entre individus du tableau X , rappelons-le.

L'analyse en composantes principales d'un multi-tableau vertical X est très proche, à un coefficient près, à l'intra-structure dans STATIS duale, puisque l'analyse du compromis correspond à l'ACP du multi-tableau vertical X muni des métriques Q et $D = \text{diag}(\sqrt{\alpha_m} D_m, m = 1, \dots, M)$. Les axes principaux u de cette ACP sont considérés comme des axes communs de représentation.

L'analyse du multi-tableau Y peut être réalisée par l'ACP du triplet (Y, R, D) .

6. Si $Q = V_X^{-1}$ et $R = V_Y^{-1}$, l'analyse de co-inertie entre deux multi-tableaux verticaux se confond avec l'analyse canonique entre deux multi-tableaux verticaux.
7. Les poids spécifiques à l'ordre s associés aux tableaux X_m et Y_m ($m = 1, \dots, M$) sont respectivement définis par $\rho_{X_{m,s}} = \frac{\text{var}(X_m Q u_s)}{\text{var}(X Q u)} \times 100$ et $\rho_{Y_{m,s}} = \frac{\text{var}(Y_m R v_s)}{\text{var}(Y R v)} \times 100$. Ces poids sont en fait des inerties projetées des tableaux du premier et deuxième groupes respectivement sur les axes de co-inertie u_s et v_s . Pour estimer la stabilité de la liaison entre ces deux multi-tableaux, on détermine les quantités $\rho_{X_m Y_{m,s}} = \text{cor}^2(c_{X_{m,s}}, c_{Y_{m,s}})$ pour tous $m = 1, \dots, M$ et $s = 1, \dots, r$ qui sont des corrélations au carré entre composantes synthétiques $c_{X_{m,s}}$ et $c_{Y_{m,s}}$. Compte tenu du fait que la base $\{u_s\}_s$ est Q -orthonormée dans \mathbb{R}^p , on peut donc représenter les individus du premier groupe dans le repère d'axes de co-inertie u_s et u_l ($l \neq s$). Les coordonnées des individus relatifs aux tableaux X_m ($m = 1, \dots, M$) dans le repère (u_s, u_l) sont les composantes des vecteurs $c_{X_{m,s}} = X_m Q u_s$ et $c_{X_l} = X_l Q u_l$. Il est possible aussi de représenter dans le même repère les individus correspondant au multi-tableau vertical X . Il en est de même en ce qui concerne la représentation des individus des tableaux Y_m ($m = 1, \dots, M$) et du multi-tableau vertical Y qui peuvent être également représentés dans le repère (v_s, v_l) de \mathbb{R}^q .
8. Si les deux multi-tableaux X et Y sont confondus ($X = Y$), le critère de l'analyse de co-inertie orthogonale successive (ACIOs) devient le critère de l'analyse en composantes principales orthogonales (ACPOs) d'un multi-tableaux vertical. Cette analyse revient à déterminer à l'étape s le vecteur u_s de \mathbb{R}^p maximisant la fonction

$$f(u_s) = \sum_{m=1}^M \text{var}(X_m^{(s-1)} Q u_s) = \text{var}(X^{(s-1)} Q u_s) = u_s' Q V_{X^{(s-1)}} Q u_s \quad (40)$$

sous la contrainte unité

$$\|u_s\|_Q^2 = 1 \quad (41)$$

L'ACPOs est équivalente à l'ACP du triplet $(X^{(s-1)}, Q, D)$. A l'ordre 1, l'ACPOs est confondue à l'ACPs. L'ACPOs a plus d'intérêt que l'ACPs, car elle est interprétable au niveau des individus et admet des représentations séparées au niveau des variables de chaque sous tableau de X . Puisque les systèmes des composantes partielles $\{c_{X_{m,s}}\}_s$ et des axes $\{u_s\}$ sont respectivement D_m -orthogonaux et Q -orthogonaux. Ce qui n'est pas le cas de l'ACPs qui ne dispose que d'axes Q -orthogonaux. Une ACPs orthogonale successive du multi-tableau Y peut être effectuée par l'ACP du triplet $(Y^{(s-1)}, R, D)$.

9. La stabilité du lien entre les tableaux X_m et les tableaux Y_m peut être réalisée aussi par la projection des axes des analyses séparées sur des axes de co-inertie des méthodes ACIs et ACIOs.

6. Application

Dans ce paragraphe, nous allons d'abord préciser les données qui seront appliquées à l'ACIs et à l'ACIOs, et après, nous allons faire un commentaire des résultats obtenus.

6.1. Les données

Les données qui sont dans cet exemple ont déjà été traitées par Pegaz-Maucet (1980) par Hanafi (1997), Blanc *et al.* (1998) et aussi par Thioulouse *et al.* (2004). Ces données écologiques ont été mesurées sur la rivière Méaudret en France. Six (6) stations réparties dans la rivière Méaudret ont été visitées chacune une fois par saison (1-Printemps, 2-Eté, 3-Automne, 4-Hiver).

A chaque fois ont été mesurés dans chaque station, 10 paramètres physico-chimiques du cours d'eau (Temp = température, Dédi = débit, pH = conductivité, Oxyg = oxygène, BDO5 = demande d'oxygène pour la biologie, Oxyd = oxydabilité, Ammo = ammoniacque, Nitra = nitrates, Phosp = phosphates) et identifiées les 13 espèces (éphéméroptères) présentes (Eda = ephemera, Bsp = baetis sp, Brh = baetis rhodani, Bni = baetis niger, Bpu = baetis pumilus, Cen = centroptilum, Ecd = eclyonorus, Rhi = rihlogena, Hla = habrophlebialauta, Hab = habroletoides modesta, Par = paraletophlebia, Cae = caenis, Eig = ephemerella ignita). On obtient donc un total de 24 relevés (6 stations x 4 saisons). Le tableau faunistique X de dimension (23, 24) est constitué à l'aide de quatre tableaux X_m de dimension (6, 13). De même, le tableau mésologique Y de dimension (24, 10) est formé de quatre tableaux Y_m de dimension (6, 10). Les deux tableaux X_m (faune) et Y_m (milieu) correspondent à la même saison m ($m = 1, 2, 3, 4$). L'objectif de cette étude est de décrire l'évolution de la relation faune-milieu d'une saison à l'autre.

6.2. L'analyse des résultats

Les espèces sont centrées par saison et les variables environnementales sont centrées puis normalisées globalement (Bouroche, 1975). Cette normalisation globale permet de prendre en compte la variance inter-saison. Chacun de ces tableaux correspond à une saison rappelons-le et à un triplet (X_m, Q, D_m) pour la faune et (Y_m, R, D_m) pour le milieu.

Dans un premier temps, une ACP centrée est réalisée sur chacun de deux multi-tableaux X et Y afin d'avoir un aperçu de la position des variables (faune et milieu) et des stations aux différentes saisons sur les premiers plans principaux.

L'ACPs du multi-tableau faunistique (Figure 1) fournit dans le premier plan principal (65,01%) avec (46,50%) pour le premier axe et (18,50%) pour le second axe, une image des stations et de descripteurs. Un effet de taille de l'axe 1 est observé, les espèces Eda, Brh sont présentes dans les stations S6 en Automne et en Hiver. Ces deux stations s'opposent sur l'axe 1 à la station S4 en Hiver. Dans la station S1 pour toutes les saisons, on retrouve les espèces Bpu et Hab. Ces dernières s'opposent sur l'axe 2 par les espèces Eig et Ecd et qui sont dans les stations S5 en Automne et la station S6 au Printemps.

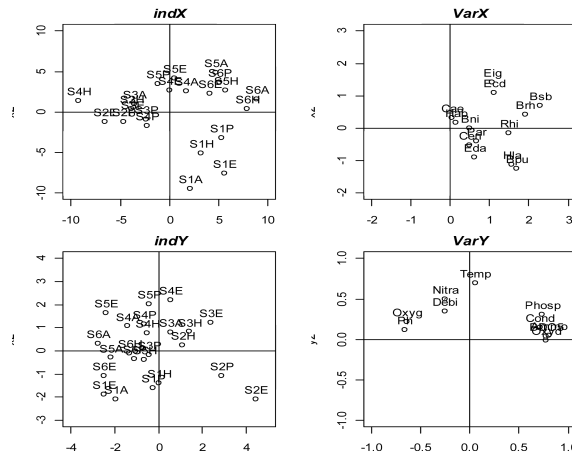


Figure 1. Représentation des stations (indX et indY) et des variables (VarX et VarY) respectivement sur les deux premiers axes et composantes principales des ACPs séparées des multi-tableaux X et Y. Les stations sont notées ici par SNV (N représente le numéro de la station et V représente la saison).

En revanche, pour ce qui est de l'ACPs du multi-tableau des variables environnementales, les deux premiers axes représentent à eux seuls plus de 73,50% de l'inertie totale. Mais le premier axe est le plus informatif (58,05%). L'axe 1 oppose les variables Oxygène et le pH d'une part et les variables Conductivité de l'eau, le Phosphate, l'Ammoniaque et l'Oxydabilité d'autre part. La température de l'eau est la plus élevée dans la station 4 (S4) en Eté. La station 2 en Eté est opposée à la station 1 pour l'Eté et l'Automne. Dans la station 5 au Printemps, on trouve du Nitrate et un Débit fort.

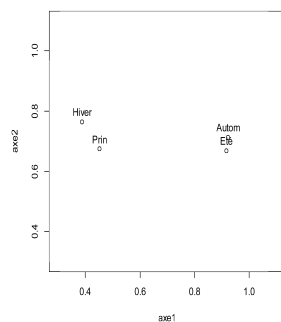


Figure 2. Représentation des saisons sur les deux premiers axes des corrélations au carré entre composantes synthétiques partielles faune-milieu pour l'ACIS.

Après avoir procédé à ces deux ACPs séparées, il est nécessaire dans un deuxième temps de faire des ACPOs pour avoir un aperçu des positions des stations et des descripteurs par saison. Au niveau faunistique (Figure (3X2)) un effet de taille de l'axe 1 est observé. En revanche au niveau du multi-tableau des variables environnementales (Figure (4Y1 et 4Y2)), l'axe 1 oppose les variables Oxygène et le pH d'une part en Été et en Automne. C'est un axe de pollution. L'axe 2 est caractérisé pour toutes les saisons par la température, le nitrate et le débit. Les positions des stations sont les mêmes que celles retrouvées dans les deux ACPs.

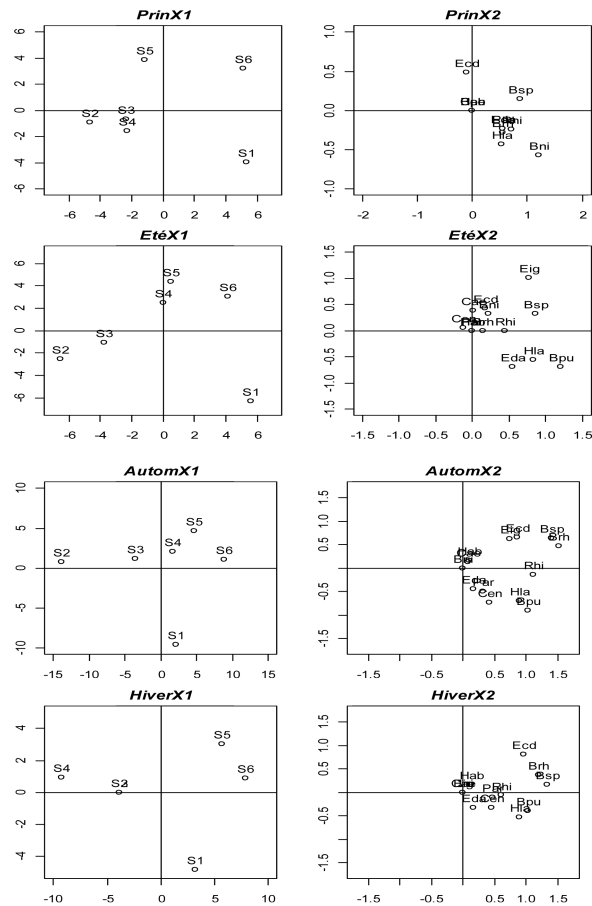


Figure 3. Position des stations (Fig.3X1) et des espèces (Fig.3X2) par saison pour les deux premiers axes pour le multi-tableau X de l'ACPOs.

Enfin, après ces ACP (ACP et ACPO) une analyse globale dans la recherche du lien entre les espèces et le milieu peut être effectuée.

L'ACI ne fournit que des axes au niveau des individus (Figure 5). On tient à signaler que nous ne faisons pas de l'interprétation une priorité, mais le but est de montrer la démarche à

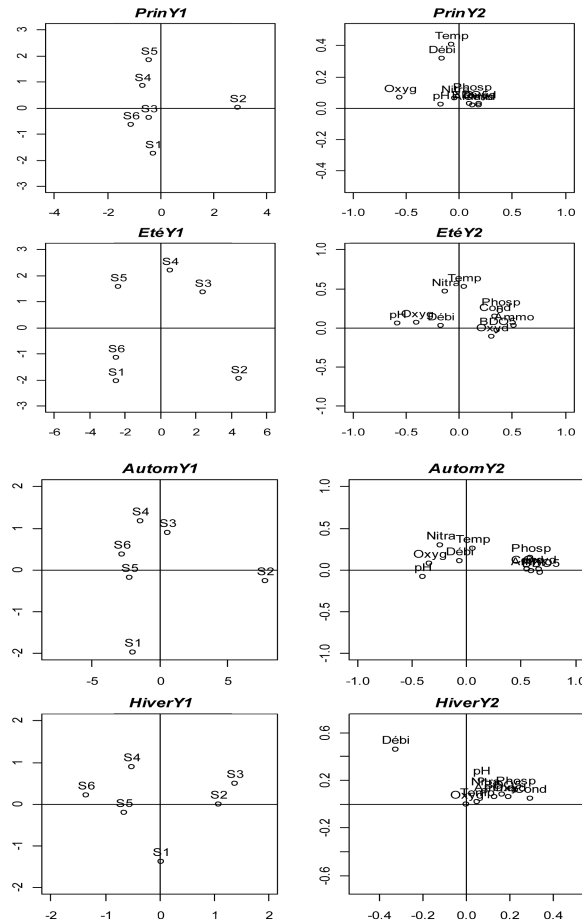


Figure 4. Représentation des stations (Fig. 4Y1) et des variables du milieu (Fig. 4Y2) par saison pour les deux premiers axes pour le multi-tableau Y de l'ACPOS.

adopter lors de la mise en oeuvre de cette méthode. C'est pour cette raison que nous avons seulement analysé ces données sur les deux premiers axes. L'idéal aurait été de faire une analyse plus affinée sur tous les dix axes. Dix étant le nombre de solutions à calculer et qui correspond au rang des matrices $V_{X_m Y_m} = X'_m D_m Y_m$. Le Tableau 1 contient les poids qui concourent à la détermination du compromis dans la méthode STATICO à titre de rappel (Thioulouse *et al.*, 2004). Le Tableau 2 contient des pourcentages d'inerties projetées (poids spécifiques) de chaque tableau (saison) sur les deux premiers axes. Sur les deux premiers axes des méthodes (ACIs) et (ACIOs) et pour le premier multi-tableau, l'Automne présente des pourcentages d'inertie qui sont les plus élevés. Ce n'est pas le cas pour le deuxième multi-tableau. Pour ce dernier et sur le deuxième axe les pourcentages d'inertie sont faibles. On remarque aussi des pourcentages d'inertie moyennement élevés en Eté pour le deuxième multi-tableau (milieu).

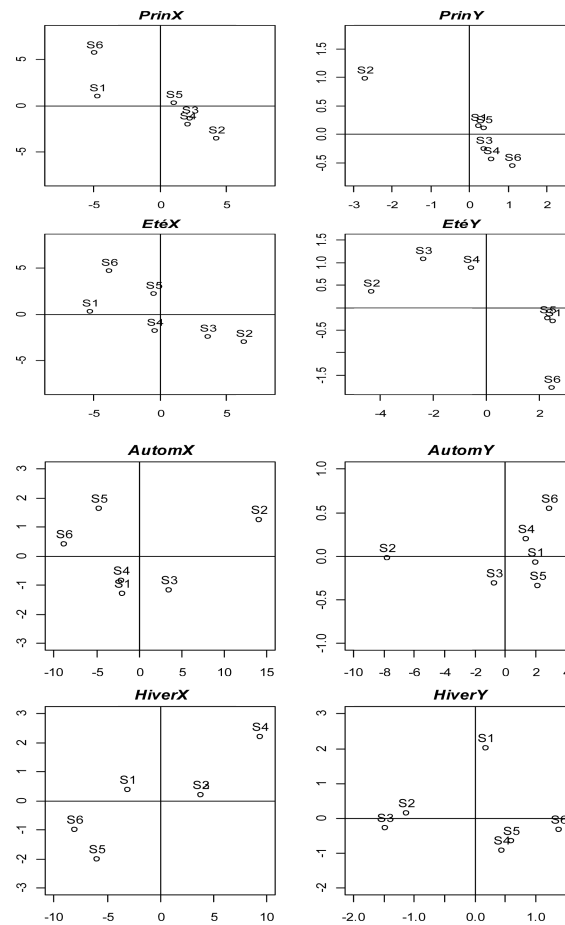


Figure 5. Position des stations par saison pour les deux premiers axes pour les deux multi-tableaux de l'ACIs.

En revanche, le Tableau 3 contient les corrélations au carré entre combinaisons linéaires partielles des variables (espèces des poissons) et des variables du milieu à l'ordre 1 et 2 pour les méthodes ACIs et ACIOs. Ces corrélations au carré permettent de décrire l'évolution de la relation faune-milieu. Une constance de ces coefficients de corrélation au carré permet de conclure à la stabilité du lien. Il ressort du Tableau 2 une même évolution de la relation faune-milieu pour les saisons Eté et Automne concernant les méthodes ACIs et ACIOs (confer le graphe de la Figure 2). Ce constat ne se retrouve pas pour l'Hiver et le Printemps qui se distinguent trop des autres saisons concernant les deux méthodes. Cette dernière situation confirme bien les résultats de la méthode STATICO (Tableau 1) qui montrent aussi le comportement particulier de la saison Hiver par rapport aux autres saisons.

Concernant l'ACIs, l'axe 1 représente 87,33% de l'inertie totale, tandis que l'axe 2 à une inertie de 8,23% de l'inertie totale (Figure 5). Ainsi nous choisissons d'expliciter les deux

Dans les deux premiers axes de l'ACIOs, on a représenté les stations (Figure 6X1) et les espèces (Figure 6X2). L'ACIOs détermine simultanément deux systèmes d'axes orthogonaux au niveau des individus et des variables. Il ressort de ces graphiques que, d'une saison à l'autre, une organisation générale se retrouve plus ou moins au niveau des stations et des espèces. On note globalement un effet de taille au niveau de l'axe 1 en ce qui concerne les espèces. L'axe 1 opposant d'une part la station S2 et la station S6 d'autre part pour toutes les saisons.

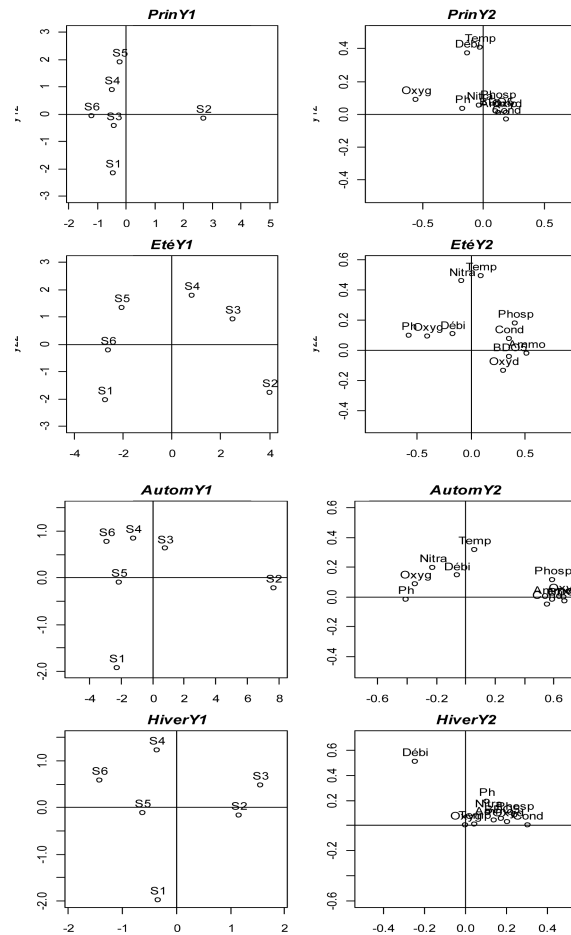


Figure 7. Représentation des stations (Fig.7Y1) et des variables du milieu (Fig.7Y2) par saison pour les deux premiers axes pour le multi-tableau Y de l'ACIOs.

On peut trouver pour toutes les saisons plus ou moins dans la station S6 les espèces *Baetis* sp et *Baetis rhodani*. La station S5 au Printemps est caractérisée par des températures et Débit (Figure 7Y1) et (Figure 7Y2) élevés. La station 2 est caractérisée par le Phosphate et

l'Ammoniaque en Automne. A proximité du centre des repères, on trouve les espèces rares qui ne sont pas prises en compte par l'ACIOs. L'axe 2 oppose en général d'une part la station S1 et les stations S4 et S6 d'autre part.

Méthode	Saisons			
	Printemps	Eté	Automne	Hiver
STATICO	0.478	0.537	0.568	0.398

Tableau 1. Poids des tableaux définissant le tableau compromis

Méthodes	Saisons			
	Printemps	Eté	Automne	Hiver
ACIs (X)	11.99	16.375	42.496	29.136
	20.416	28.120	40.675	10.786
ACIs (Y)	6.854	30.633	57.804	4.707
	30.869	35.182	15.655	18.291
ACIOs (X)	11.990	16.375	42.496	29.136
	16.340	29.836	42.840	10.981
ACIOs (Y)	6.854	30.633	57.804	4.707
	27.398	38.882	16.002	17.715

Tableau 2. Pourcentages d'inerties projetées (poids spécifiques) de chaque saison sur les deux premiers axes.

Méthodes	Saisons			
	Printemps	Eté	Automne	Hiver
ACIs	0.451	0.916	0.921	0.388
	0.674	0.669	0.713	0.764
ACIOs	6.854	30.633	57.804	4.707
	0.451	0.916	0.921	0.388

Tableau 3. Corrélations au carré entre combinaisons linéaires des variables de l'abondance des poissons et des variables du milieu à l'ordre 1 et 2 pour les méthodes ACIs et ACIOs.

7. Conclusion

En conclusion, nous pouvons dire que la question concernant l'étude de la stabilité du lien entre les couples de tableaux vient d'être résolue par la construction des coefficients de corrélation au carré entre les composantes partielles de chaque couple de tableaux et des

inerties projetées sur les axes de co-inerties associés aux deux groupes de variables. Cette démarche très simple et intéressante ne souffre d'aucune restriction contrairement à celle proposée par STATICO, qui adopte une stratégie en trois étapes comme pour la méthodes STATIS et plus particulièrement celle de l'analyse triadique partielle (ATP). La méthode STATICO n'est pas fondée sur l'optimisation d'un critère, ce qui n'est pas intéressant. Contrairement à l'analyse canonique entre deux multi-tableaux verticaux qui détermine le lien entre les deux ensembles de variables, l'analyse de co-inertie entre deux multi-tableaux verticaux en revanche tout en déterminant le lien entre ces deux ensembles de variables étudie également la structure interne de ces ensembles de variables. Mais le seul handicap de cette dernière méthode est de ne pas fournir des ensembles de combinaisons linéaires de variables partielles et globales non corrélés, ce qui limite la représentation des variables d'origine dans les repères orthogonax. D'où l'adoption dans cet article de l'analyse de co-inertie orthogonale entre deux multi-tableaux verticaux pour lever cette difficulté. Les positions des stations et des descripteurs sont globalement conformes à celles trouvées par Thioulouse *et al.* (2004) et Hanafi (1997). En perspective dans ce contexte de lien entre deux multi-tableaux verticaux, un autre problème fondé sur la somme des carrés des covariances est en cours de réalisation.

Références

- Blanc, L., Chessel, D. and Dolédec, S., 1998. Etude de la stabilité temporelle des structures spatiales par les analyses d'une série de tableaux de relevés fauniques totalement appariés. Bouroche, J.D., 1975. Analyse de données ternaires : La double analyse en composantes principales. Thèse de 3-ème cycle, Université de Paris VI. France.
- Chessel, D. and Mercier, P., 1993. Couplage de triplets statistiques et liaisons espèces-environnement. In : Biométrie et Environnement. Lebreton, J.D. & ASSELAIN, B (Eds.) Masson, Paris. 15-44.
- Dolédec, S. and Chessel, D., 1994. Co-inertia analysis: an alternative method for studying species-environment relationships. *Freshwater Biology*. **31**, 277-294.
- Franquet, E. and Chessel, D., 1994. Approche statistique des composantes spatiales et temporelles de la relation faune-milieu. *Compte Rendu de l'Académie des Sciences Paris, Sciences de la vie*. **317**, 202-206.
- Franquet, E., Dolédec, S., Chessel, D., 1995. Using multivariate analyses for separating spatial and temporal effects within species-environment relationships. *Hydrobiologia*. 300-301, 425-431.
- Hanafi, M., 1997. Structure de l'ensemble des analyses multivariées des tableaux de données à trois entrées : Eléments théoriques et appliqués. Thèse de Doctorat unique, Université Claude Bernard de Lyon 1, France.
- Hanafi, M., Qanari, E.M., 2008. Nouvelles propriétés de l'analyse en composantes communes et poids spécifiques. *Revue de Statistiques Appliquée*. **149**(2), 75-97.
- Kissita, G., Niéré, L., Makany, R.A. and Mizère, D., 2013. Canonical correlation analysis between two vertical multi-block tables and applications to qualitative variables. *Advances and Applications in Statistics*. **34**(1), 11-38.
- Lavit, Ch., Escoufier, Y., Sabatier, R. and Traissac, P., 1994. The ACT (Statis method). *Computational Statistics and Data Analysis*. **18**, 97-119. Paris. 240 p.

- Qannari, E.M., Courcoux, Ph. and Vigneau, E., 2001. Common components and specific weights analysis performed on preference data. *Food Quality and Preference*. **12**, 365-368.
- Qannari, E.M., Wakeling, I., Courcoux, Ph. and Macfie, M.F., 2000. Defining the underlying sensory dimensions; *Food Quality and Preference*. **11**, 151-154.
- Pegaz-Maucet, D., 1980. Impact d'une perturbation d'origine organique sur la dérive des macro-invertébrés benthiques d'un cours d'eau. Comparaison avec le benthos. Thèse de 3-ème cycle. Université de Lyon I. France.
- 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*. **XLVIII** (1), 31-46.
- Thioulouse, J., Simier, M., and Chessel, D. 2004. Simultaneous analysis of a sequence of paired ecological tables. *Ecology*. **85**(1), 272-283. (c) 2004 by Ecological Society of America.
- Tucker, L.R., 1958. An inter-battery method of factor analysis. *Psychometrika*. **23**, 111-136.
- Vivien, M., 2002. Approches PLS linéaires et non linéaires pour la modélisation de multi-tableaux. Théories et applications. Thèse, Université de Montpellier I.
- Vivien, M., Sabatier, R., 2004. A generalization of STATIS-ACT strategy: DO-ACT for two multiblocks tables. *Computational Statistics and Data Analysis*. **46**, 155-171.
- Wold, H., 1985. In S., Kotz, and N.L., Johnson, (Eds). *Encyclopedia of statistical sciences*. Partial least squares. New York : Wiley. **6**, 581-591.