REVUE DE STATISTIQUE APPLIQUÉE

MOHAMED HANAFI
ROGER LAFOSSE

Généralisations de la régression simple pour analyser la dépendance de K ensembles de variables avec un $K+1$ème

Revue de statistique appliquée, tome 49, n° 1 (2001), p. 5-30
<http://www.numdam.org/item?id=RSA_2001__49_1_5_0>

Rev. Statistique Appliquée, 2001, **XLIX** (1), 5-30

GÉNÉRALISATIONS DE LA RÉGRESSION SIMPLE POUR ANALYSER LA DÉPENDANCE DE *K* ENSEMBLES DE VARIABLES AVEC UN *K* + 1ème

Hanafi Mohamed, Lafosse Roger

E.N.I.T.I.A.A., rue de la Géraudière B.P. 82225 44322 Nantes Cedex 3

hanafi@enitiaa-nantes.fr

Lab. de Statistique et Probabilités, UMR C55830 Université Paul Sabatier,

118 rte de Narbonne, 31062 Toulouse Cedex 4 lafosse@cict.fr

RÉSUMÉ

Une analyse de dépendance simultanée considérée entre une matrice et K matrices (Lafosse & Hanafi, 1997) peut se présenter comme une analyse factorielle de deux tableaux, grâce à un projecteur particulier. Les régressions simples simultanées qui s'ensuivent permettent d'évaluer la contribution à cette dépendance de chaque ligne de chaque matrice. Les coefficients de régression simultanée sont globalement solution d'un problème d'optimisation nouvellement posé.

Mots-clés : *corrélation linéaire, analyses factorielles, concordance entre matrices, moindres carrés, biplot.*

ABSTRACT

By introducing a projector, the simultaneous dependency analysis considered between a matrix and K matrices by Lafosse & Hanafi (1997) boils down to an analysis between two matrices. Some simultaneous simple regressions then lead to a calculus of the contributions to this dependency for every rows of each matrix. The simultaneous calculus of the regression coefficients now comes from the solution of one optimization problem.

Keywords : *linear correlation, factor analysis, agreement between matrices, least squares, biplot.*

1. Introduction

Cazes & coll. (1977) ont proposé une étude de la dépendance simultanée entre une variable qualitative Y et K autres variables qualitatives. Cette étude revenait à définir un codage des modalités de la variable Y fortement dépendant de chacune des K variables qualitatives. Leur analyse n'est pas une analyse canonique classique entre deux ensembles d'indicatrices des modalités, car chacun des K sous-ensembles

d'indicatrices de modalités associé respectivement à chacune des K variables est muni de sa propre semi-métrique de Mahalanobis, de sorte que la particularité de chaque sous-ensemble est bien prise en compte. Leur analyse est cependant une analyse canonique de deux tableaux d'indicatrices, en convenant que la métrique relative à l'ensemble des modalités des K variables est la métrique bloc diagonale dont les blocs sont ces semi-métriques de Mahalanobis.

L'analyse de Tucker (1958) de deux tableaux est une analyse canonique si l'on convient que pour l'obtenir il suffit de remplacer les métriques de Mahalanobis par les métriques identité (par exemple, Lafosse (1997)), ou encore si l'on considère que les composantes de Tucker des tableaux formés des composantes principales réduites, obtenus par les deux ACP respectives, sont les composantes de l'analyse canonique classique des deux tableaux (Ten Berge, 1977).

Considérons l'analyse de Tucker généralisée visant une étude de la dépendance simultanée d'un tableau $\boldsymbol{Y}$ avec K autres tableaux $\boldsymbol{X}_i, i = 1, ..., K$, et qui reprend la démarche de Cazes *& coll.*. La métrique bloc diagonale dont il est question ci-dessus devient la métrique identité, de sorte que nous sommes ramenés à l'analyse de Tucker de $\boldsymbol{Y}$ et du tableau $\boldsymbol{X}$ concaténé des K tableaux. Il n'y a alors plus prise en compte de la partition en K sous- ensembles du tableau $\boldsymbol{X}$, comme cela se produisait chez Cazes à cause des métriques. La partition en K sous-ensembles n'est en fait pas considérée dans cette analyse de Tucker.

Lafosse et Hanafi (1997) ont proposé une généralisation de Tucker visant la même étude, où les composantes successives de $\boldsymbol{Y}$ définies s'associent chacune à K composantes partenaires trouvées dans chacun des K tableaux. La première composante de cette analyse est cependant la première composante de la «généralisation» précédente. Cette première solution a en fait la particularité de pouvoir s'associer à toute partition du tableau concaténé, et donc en particulier à la partition de $\boldsymbol{X}$ considérée. Ensuite, les autres solutions successives obtenues dans cette analyse dépendent de la partition, bien que les métriques soient les métriques identité.

Dans ce papier nous réécrivons cette dernière analyse avec des métriques quelconques, de sorte que la prise en compte de la partition peut aussi provenir des métriques. Cette analyse qui est une étude de dépendance entre triplets statistiques est nommée analyse CONCOR.

Comme nous l'avons souligné ci-dessus, l'analyse de Cazes & coll. est une analyse factorielle de deux tableaux avec un choix particulier de métriques. L'analyse CONCOR peut quant à elle se ramener à une analyse factorielle de deux tableaux de la manière suivante : une fois les termes de l'analyse obtenus, il devient possible de proposer un projecteur $\boldsymbol{P}$ de sorte que l'analyse factorielle de $\boldsymbol{Y}$ et de $\boldsymbol{XP}$ redonnent ces termes. C'est cette propriété qui fonde la suite de l'article. En effet, pouvant se ramener à la situation de deux tableaux, il devient possible de faire le calcul des contributions partielles à la dépendance simultanée des lignes de chaque tableau. Cela revient à définir des régressions simples simultanées, généralisant à plus de deux tableaux un calcul proposé pour deux tableaux (Lafosse 1997). Une évaluation des participations des colonnes est alors envisageable.

La régression simple simultanée entre deux tableaux a été proposée pour répondre à un problème d'ajustement entre nuages d'individus (Lafosse, 1985)

ou pour proposer une matrice estimée caractérisant la participation d'un tableau à la dépendance (image concordante, Lafosse, 1997). Dans ce papier, l'obtention simultanée de l'ensemble des coefficients de régression simple est solution d'un problème d'optimisation. Le problème d'invariance posé vient en prolongement d'un problème d'invariance par rotation considéré notamment par Green (1952), Cliff (1966) et Ten Berge (1977). On suggère que ce qui caractérise la participation d'un tableau à la dépendance doit être invariant par transformation linéaire de l'autre tableau, et pas seulement par isométrie.

Un programme de l'analyse CONCOR avec les développements donnés dans cet article a été écrit en langage MATLAB (Lafosse, 1998). Pour quelques autres écritures on peut s'adresser aux deux auteurs.

2. (K+1)-uples de l'analyse CONCOR : Rappels et compléments

2.1. Définition des composantes

On rappelle la définition des $K+1$ uples définis en analyse de la concordance d'un tableau avec K tableaux (Lafosse et Hanafi, 1997). Les écritures sont ici plus générales, $K+1$ métriques euclidiennes respectives quelconques venant maintenant à la place des métriques identités implicitement considérées. Le calcul lui-même mérite alors d'être précisé.

Soient $\boldsymbol{Y}$ un tableau centré $n \times q$ et K tableaux centrés $\boldsymbol{X_1}, \boldsymbol{X_2}, \ldots, \boldsymbol{X_i}, \ldots, \boldsymbol{X_K}$ de dimensions respectives $n \times m_i$.

On note $\boldsymbol{X} = [\boldsymbol{X_1 X_2} \ldots \boldsymbol{X_K}]$ le tableau $n \times m$, avec $m = \sum m_i$, formé par la concaténation des K tableaux $\boldsymbol{X_i}$, qui alors peuvent être considérés comme des sous-ensembles de variables de l'ensemble $\boldsymbol{X}$.

La matrice diagonale $\boldsymbol{D}$ représente la métrique des poids affectés à chacun des n individus, la somme des poids valant 1.

La matrice $\boldsymbol{N}$ désigne la métrique euclidienne considérée dans $\mathcal{R}^q$ et les matrices $\boldsymbol{M_i}$ celles respectivement considérées dans les espaces $\mathcal{R}^{m_i}$.

Un couple tel que le couple ($\boldsymbol{Y}$, $\boldsymbol{N}$) fait référence à l'ensemble des vecteurs lignes de $\boldsymbol{Y}$ constitutif d'un nuage de points de $\mathcal{R}^q$ représenté avec la métrique $\boldsymbol{N}$. Ainsi parlera-t-on du nuage ($\boldsymbol{Y}$, $\boldsymbol{N}$) et de relation entre nuages (plus de détails dans Hanafi, 1997, chapitre 2).

Un vecteur $\boldsymbol{x}$ est considéré comme vecteur colonne dans sa notation matricielle et le vecteur ligne transposé est noté $\boldsymbol{x}'$. On note $(\boldsymbol{b'_j}, \boldsymbol{a'_{1j}}, \ldots, \boldsymbol{a'_{ij}}, \ldots, \boldsymbol{a'_{Kj}})$ le j-ème $K+1$ uple d'axes de $\mathcal{R}^q \times \mathcal{R}^{m_1} \times \mathcal{R}^{m_2} \times \ldots \times \mathcal{R}^{m_K}$ définis dans l'analyse CONCOR de $\boldsymbol{Y}$ avec les K tableaux $\boldsymbol{X_1}, \boldsymbol{X_2}, \ldots, \boldsymbol{X_i}, \ldots, \boldsymbol{X_K}$, $j = 1, 2, \ldots, S$.

Le système $\{\boldsymbol{b_j}\}_{j=1,.,S}$ est formé de S vecteurs $\boldsymbol{N}$-orthonormés.

Les K systèmes $\{\boldsymbol{a_{ij}}\}_{j=1,.,S}$ $i = 1, \ldots, K$, sont chacun formés de S vecteurs respectivement $\boldsymbol{M_i}$-orthonormés. Il peut exister pour j fixé des vecteurs $\boldsymbol{a_{ij}}$ nuls, mais au moins l'un d'entre eux est non nul. La valeur de S qui correspond au nombre de solutions trouvées est au moins égale à sup(rang($\boldsymbol{Y'DX_i}$)).

Pour j fixé, une composante $\boldsymbol{X_i M_i a_{ij}}$ est dite composante *partenaire* dans le tableau $\boldsymbol{X_i}$ de la composante $\boldsymbol{YNb_j}$. Les S composantes $\boldsymbol{YNb_j}$ constituent *les composantes principales de l'analyse CONCOR*. Un j-ème $K+1$-uple $(\boldsymbol{YNb_j}, \boldsymbol{X_1 M_1 a_{1j}}, \ldots, \boldsymbol{X_i M_i a_{ij}}, \ldots, \boldsymbol{X_K M_K a_{Kj}})$ de composantes est donc formé de la j-ème composante principale et de ses K composantes partenaires.

On précise maintenant le calcul de ces termes. Le critère optimisé en concordance (Lafosse et Hanafi, 1997), est ici considéré sous contraintes de norme des vecteurs $\boldsymbol{a_{ij}}$ et $\boldsymbol{b_j}$ avec les métriques euclidiennes respectives choisies, de sorte que l'optimum λ_j^2 obtenu lors du calcul d'un $K+1$-uple j vaut :

$$\lambda_j^2 = \sum_{i=1}^{K} \mathrm{cov}^2(\boldsymbol{X_i M_i a_{ij}}, \boldsymbol{YNb_j}). \tag{1}$$

On note $\boldsymbol{M}$ la métrique euclidienne de $\mathcal{R}^m$ formée des blocs diagonaux $\boldsymbol{Mi}$, les autres blocs étant nuls. Avec la convention que les matrices $\sum_{h=1}^{j-1} \boldsymbol{M_i a_{ih} a'_{ih}}$, $i = 1, \ldots, K$, sont nulles quand $j = 1$, une composante $\boldsymbol{YNb_j}$ est obtenue en calculant les deux vecteurs $\boldsymbol{b_j}$ et $\boldsymbol{a_j}$, où $\boldsymbol{b_j}$ est $\boldsymbol{N}$-normé dans $\mathcal{R}^q$ et $\boldsymbol{a_j}$ est $\boldsymbol{M}$-normé dans $\mathcal{R}^m$, associés à la plus grande valeur singulière λ_j de la matrice :

$$[(\boldsymbol{X_1} - \boldsymbol{X_1} \sum_{h=1}^{j-1} \boldsymbol{M_1 a_{1h} a'_{1h}}) \ldots (\boldsymbol{X_i} - \boldsymbol{X_i} \sum_{h=1}^{j-1} \boldsymbol{M_i a_{ih} a'_{ih}})$$
$$\ldots (\boldsymbol{X_K} - \boldsymbol{X_K} \sum_{h=1}^{j-1} \boldsymbol{M_K a_{Kh} a'_{Kh}})]' \boldsymbol{DY} \tag{2}$$

Le vecteur $\boldsymbol{b_j}$ est ainsi égal au vecteur $\boldsymbol{N^{-1/2} v_j}$, avec $\boldsymbol{v_j}$ vecteur propre $\boldsymbol{I_q}$-normé associé à la plus grande valeur propre λ_j^2 de la matrice symétrique :

$$\boldsymbol{N^{1/2} Y' D} [\sum_{\boldsymbol{h=1}}^{\boldsymbol{K}} \boldsymbol{X_i} (\boldsymbol{I_{mi}} - \sum_{\boldsymbol{h=1}}^{\boldsymbol{j-1}} \boldsymbol{M_i a_{ih} a'_{ih}}) \boldsymbol{M_i X'_i}] \boldsymbol{DY N^{1/2}}. \tag{3}$$

Les K composantes partenaires d'une j-ème composante principale se déduisent de cette composante principale par les relations :

$$\boldsymbol{X_i M_i} (\boldsymbol{I_{mi}} - \sum_{h=1}^{j-1} \boldsymbol{a_{ih} a'_{ih} M_i}) \boldsymbol{X'_i DYNb_j} = \mathrm{cov}(\boldsymbol{X_i M_i a_{ij}}, \boldsymbol{YNb_j}) \boldsymbol{X_i M_i a_{ij}},$$
$$i = 1, \ldots, K.$$

La moyenne pondérée $\boldsymbol{XMa_j}$ des composantes $\boldsymbol{X_i M_i a_{ij}}$ partenaires de $\boldsymbol{YNb_j}$ est donnée par la relation

$$\boldsymbol{XMa_j} = \sum_{i=1}^{K} p_{ij} \boldsymbol{X_i M_i a_{ij}}, \tag{4}$$

Avec les poids p_{ij}

$$p_{ij} = \text{cov}(\boldsymbol{X_i M_i a_{ij}}, \boldsymbol{YNb_j})/\lambda_j. \tag{5}$$

On a ainsi :

$$\lambda_j^2 = \text{cov}^2(\boldsymbol{XMa_j}, \boldsymbol{YNb_j}) \text{ avec : } \sum_{i=1}^{K} p_{ij}^2 = 1.$$

Un poids p_{ij} mesure la valeur relative de la participation du tableau $\boldsymbol{X_i}$ lors du calcul du $K+1$ uple j. La composante moyenne $\boldsymbol{XMa_j}$ s'associe à la composante principale $\boldsymbol{YNb_j}$ par la relation

$$\boldsymbol{YNY'DXMa_j} = \lambda_j \boldsymbol{YNb_j}, \tag{6}$$

et d'après (5) et (4), le vecteur $\boldsymbol{a_j}$ est le concaténé les vecteurs $p_{ij}\boldsymbol{a_{ij}}$

$$\boldsymbol{a_j'} = [p_{1j}\boldsymbol{a_{1j}'}, \ldots, p_{ij}\boldsymbol{a_{ij}'}, \ldots, p_{Kj}\boldsymbol{a_{Kj}'}]. \tag{7}$$

Remarques

La démarche en analyse CONCOR vise l'obtention d'un découpage de dépendance entre tableaux. La solution proposée est associée à deux optimisations indépendantes sous contraintes de norme, les deux critères respectifs étant :

$$f(\boldsymbol{u}, \boldsymbol{v}) = \text{cov}^2(\boldsymbol{XMu}, \boldsymbol{YNv})$$

et

$$g(\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_K}, \boldsymbol{f}) = \sum_{\boldsymbol{i}=1}^{\boldsymbol{K}} \text{cov}^2(\boldsymbol{X_i M_i u_i}, \boldsymbol{YNf}).$$

Pour une solution j de l'analyse, les deux optimisations respectives ont mené à l'optimum :

$$\text{cov}^2(\boldsymbol{XMa_j}, \boldsymbol{YNb_j}) = \sum_{\boldsymbol{i}=1}^{\boldsymbol{K}} \text{cov}^2(\boldsymbol{X_i M_i a_{ij}}, \boldsymbol{YNb_j}).$$

L'égalité induit le découpage et fonde les développements donnés dans ce papier (section 2.2). Cependant on ne va pas se servir directement des valeurs de critère optimisé pour constituer les mesures de l'analyse. Notant ρ_j les corrélations linéaires $\rho(\boldsymbol{YNb_j}, \boldsymbol{XMa_j})$, et ρ_{ij} les coefficients de corrélation $\rho(\boldsymbol{YNb_j}, \boldsymbol{X_i M_i a_{ij}})$, les S égalités suivantes, qui se déduisent des égalités précédentes entre covariances, seront

celles préférées par la suite (section 3) pour décomposer l'importance relative des composantes partenaires :

$$\rho_j^2 \mathrm{var}(\boldsymbol{XMa_j}) = \sum_{i=1}^{K} \rho_{ij}^2 \ \mathrm{var}(\boldsymbol{X_i M_i a_{ij}}), j = 1, \ldots S,$$

précisant ainsi le type de compromis défini par la composante moyenne $\boldsymbol{XMa_j}$.

De telles « variances expliquées » définissent des parts d'inertie impliquées dans la dépendance mesurée par des corrélations linéaires. Ces variances expliquées ne traduisent des parts de variance des variables des tableaux que lorsque les métriques sont métriques identité. L'égalité globale suivante, qui se déduit des précédentes, traduit ainsi une décomposition de l'importance relative des tableaux $\boldsymbol{X_i}$ quant à la dépendance simultanée avec $\boldsymbol{Y}$:

$$\sum_{j=1}^{S} \rho_j^2 \mathrm{var}(\boldsymbol{XMa_j}) = \sum_{i=1}^{K} \sum_{j=1}^{S} \rho_{ij}^2 \mathrm{var}(\boldsymbol{X_i M_i a_{ij}}).$$

Les K cumuls :

$$\sum_{j=1}^{S} \rho_{ij}^2 \mathrm{var}(\boldsymbol{X_i M_i a_{ij}}), i = 1, \ldots, K,$$

peuvent donc être plus judicieux à considérer que les cumuls de covariances pour désigner le tableau jugé le plus important. Un résultat de simulation faite dans un contexte de reconnaissance de structure va dans ce sens (Lafosse 1999b).

Cependant, la distinction entre critères qui sous-tendent un découpage, et mesures par les variances expliquées, est inutile en analyse canonique ou en analyse de type ACPVI, puisqu'alors les covariances sont elles-mêmes des variances expliquées (ces contextes particuliers étant introduits en choisissant pour métriques des métriques de Mahalanobis).

Les écritures avec des métriques quelconques, permettent de définir l'analyse CONCOR dans toute sa généralité. Cependant métriques identité et (semi-)métriques de Mahalanobis constituent les choix les plus évidents dans ce contexte. Une autre façon de supprimer l'information sur la structure interne à un tableau, que par le choix de Mahalanobis, consiste à remplacer les variables initiales par un système de variables non corrélées et de variances 1, engendrant cependant le même sous-espace que les variables initiales. Ce qui avait été relevé par Ten Berge (1977, 1988) et Van de Geer (1984), pour l'analyse SUMCOR de Horst (1961), reste valide en analyse CONCOR. Un programme d'analyses CONCOR correspondant seulement aux deux types de métriques indiqués peut donc être simplifié, puisque finalement, avec cette substitution faite au départ et aisément produite depuis le calcul des composantes principales réduites d'ACP, seules les métriques identité peuvent alors être considérées.

2.2. Couples monogames

On définit maintenant des couples de composantes qui sont appelés *couples monogames* de l'analyse CONCOR. Ce sont eux qui réalisent le découpage de relation entre un tableau et K tableaux. La notion de couples monogames a été proposée en analyse de concordance de deux tableaux (Lafosse, 1997). Plus généralement en analyse CONCOR, les couples monogames définis sont associés à un couple de tableaux obtenu après le calcul d'un projecteur.

Les corrélations croisées suivantes sont nulles pour tous les indices h et j, $h \neq j$, de $[1, S]$:

$$\rho(\boldsymbol{XMa_h}, \boldsymbol{YNb_j}) = 0. \tag{8}$$

Les S couples $(\boldsymbol{XMa_j}, \boldsymbol{YNb_j})$ sont donc tels qu'une composante principale $\boldsymbol{YNb_j}$ n'est corrélée qu'à sa composante partenaire compromis $\boldsymbol{XMa_j}$, dans la mesure où elle est de corrélation nulle avec les $S-1$ autres composantes partenaires compromis, tout comme sont nulles les corrélations de $\boldsymbol{XMa_j}$ avec les $S-1$ composantes $\boldsymbol{YNb_h}$, $h \neq j$.

De plus le système N-orthonormé $\{\boldsymbol{b_j}\}$ induit un découpage de l'inertie du nuage $(\boldsymbol{Y}, \boldsymbol{N})$ alors que le système M-orthonormé $\{\boldsymbol{a_j}\}$ induit un découpage de l'inertie du nuage $(\boldsymbol{X}, \boldsymbol{M})$.

Par suite on considère que les S couples $(\boldsymbol{XMa_j}, \boldsymbol{YNb_j})$ sont isolables les uns des autres et constituent les couples monogames de l'analyse. Ces couples vérifient les relations suivantes :

$$\boldsymbol{Y'DXMa_j} = \lambda_j \boldsymbol{b_j}, \tag{9}$$

$$\boldsymbol{YNY'DXMa_j} = \lambda_j \boldsymbol{YNb_j}.$$

Mais, excepté le cas où $K = 1$, dès que $j > 1$ et à cause des déflations à chaque pas des tableaux $\boldsymbol{X_i}$, on a :

$$\boldsymbol{X'DYNb_j} \neq \lambda_j \boldsymbol{a_j},$$

de sorte que les couples $(\boldsymbol{XMa_j}, \boldsymbol{YNb_j})$ ne sont pas les couples monogames des deux nuages $(\boldsymbol{Y}, \boldsymbol{N})$ et $(\boldsymbol{X}, \boldsymbol{M})$.

On note $\boldsymbol{B}$ la matrice $q \times S$ constituée des vecteurs colonnes $\boldsymbol{b_j}$ et $\boldsymbol{A}$ la matrice $m \times S$ constituée des vecteurs colonnes respectifs $\boldsymbol{a_j}$. On note Δ la matrice diagonale $S \times S$ contenant sur la diagonale les valeurs respectives $\lambda_j, j = 1, \ldots, S$. D'après (9), on a :

$$\boldsymbol{Y'DXMA} = \boldsymbol{B}\Delta. \tag{10}$$

La matrice $\boldsymbol{AA'M}$ est celle du projecteur sur le sous-espace de dimension S de $\mathcal{R}^m$ impliqué dans la dépendance simultanée avec $\boldsymbol{Y}$. La relation établie au départ

entre $\boldsymbol{Y}$ et $\boldsymbol{X}$ est restreinte maintenant à une relation entre les individus de $\boldsymbol{Y}$ et ceux de $\boldsymbol{X}$ projetés sur ce sous-espace.

Propriété fondamentale 2.2. – *Les couples* $(\boldsymbol{XMa_j}, \boldsymbol{YNb_j})$ *sont les couples monogames de l'analyse de concordance des triplés statistiques* $(\boldsymbol{Y}, \boldsymbol{N}, \boldsymbol{D})$ *et* $(\boldsymbol{XMAA'}, \boldsymbol{M}, \boldsymbol{D})$.

Comme $\boldsymbol{AA'Ma_j} = \boldsymbol{a_j}$, la relation (9) peut encore s'écrire :

$$\boldsymbol{Y'DXMAA'Ma_j} = \lambda_j \boldsymbol{b_j}, \tag{11}$$

alors que, d'après (10) et sachant l'orthogonalité des vecteurs $\boldsymbol{b_j}$, on a

$$(\boldsymbol{XMAA'})'\boldsymbol{DYNb_j} = \boldsymbol{AA'MX'DYNb_j} = \boldsymbol{A\Delta B'Nb_j} = \lambda_j \boldsymbol{a_j}. \tag{12}$$

Cela prouve que les S couples de vecteurs $(\boldsymbol{b_j}, \boldsymbol{a_j})$ sont ceux associés aux S valeurs singulières de la décomposition en valeurs singulières de la matrice :

$$\boldsymbol{Y'DXMAA'}.$$

L'analyse sera désormais conduite depuis le découpage de la relation entre les tableaux $\boldsymbol{Y}$ et $\boldsymbol{XMAA'}$, découpage défini par les S couples $(\boldsymbol{XMa_j}, \boldsymbol{YNb_j})$.

Par construction, le nombre de solutions S est donc inférieur ou égal au rang de $\boldsymbol{Y'DX}$. Quand aucun partenaire n'est nul, $S = \sup(\mathrm{rang}(\boldsymbol{Y'DX_i}))$.

3. Régressions simultanées

3.1. Introduction

Utilisant par la suite des notions notamment définies dans Lafosse (1997), cette introduction en constitue un rappel original produit depuis un problème introduit par Green (1952), plus connu sous le nom de problème de Cliff (1966).

Soient $\boldsymbol{X}$ $n \times m$ et $\boldsymbol{Y}$ $n \times q$ deux matrices centrées. Pour obtenir un ajustement de $\boldsymbol{Y}$ à $\boldsymbol{X}$, Cliff pose le problème (pour $m = q$) de la recherche d'une transformation $\boldsymbol{R}$, application de $\mathcal{R}^q$ dans $\mathcal{R}^m$ vérifiant $\boldsymbol{R'R} = \boldsymbol{I_q}$ et maximisant le critère

$$f_1(\boldsymbol{R}) = \mathrm{tr}[\boldsymbol{X'DYR'}].$$

Soit $\boldsymbol{U}\,\Delta\boldsymbol{V'}$ une décomposition en valeurs singulières usuelle de $\boldsymbol{X'DY}$.

Une solution est $\boldsymbol{R'} = \boldsymbol{VU'}$ (Cliff, et plus de détails quand $m \neq q$ dans Ten Berge, 1977). L'intensité de la relation entre $\boldsymbol{X}$ et le tableau transformé pourrait alors s'évaluer par

$$\mathrm{tr}(\boldsymbol{X'DYVU'}) = \mathrm{tr}[\boldsymbol{U\,\Delta\,U'}] = \mathrm{tr}[\boldsymbol{\Delta}],$$

mais le critère f_1 demande à être relativisé.

Le problème de Cliff peut ainsi être vu comme celui de la maximisation d'un critère g, équivalent au précédent quand $\boldsymbol{R}'\boldsymbol{R} = \boldsymbol{I_q}$:

$$g_1(\boldsymbol{R}) = \cos^2_D(\boldsymbol{X}, \boldsymbol{YR}') = \mathrm{tr}^2[\boldsymbol{X}'\boldsymbol{DYR}']/\{\mathrm{tr}[\boldsymbol{X}'\boldsymbol{DX}]\mathrm{tr}[(\boldsymbol{YR}')'\boldsymbol{DYR}']\}.$$

La valeur du critère g_1 obtenue au maximum est indépendante de toute rotation orthogonale préliminaire faite sur $\boldsymbol{Y}$.

On suppose maintenant que le problème vise à définir une mesure du lien entre $\boldsymbol{X}$ et $\boldsymbol{Y}$ pouvant mieux caractériser la participation de $\boldsymbol{X}$ à ce lien, parce qu'elle est aussi indépendante d'affinités orthogonales faites sur $\boldsymbol{Y}$. La solution de Cliff ne serait alors qu'une première étape à franchir (cette première étant en fait déjà une seconde, le centrage initial des tableaux revenant à effectuer une famille de translations optimale au sens des moindres carrés). En effet, réaliser des affinités orthogonales dans les directions des axes définis par les colonnes de $\boldsymbol{U}$ ne modifie pas le problème de Cliff et sa solution. La question de ces affinités peut donc se traiter une fois la rotation de Cliff adoptée. On considère ainsi la famille des transformations $\boldsymbol{C}' = \boldsymbol{VU}'\boldsymbol{U}\Delta_C\boldsymbol{U}'$ avec le critère à optimiser

$$h(\Delta_C) = \cos^2(\boldsymbol{X}, \boldsymbol{YC}').$$

Il s'agit alors de trouver une matrice diagonale Δ_C à éléments diagonaux strictement positifs maximisant

$$h(\Delta_C) = \mathrm{tr}^2(\Delta\Delta_C)/\{\mathrm{tr}(\boldsymbol{X}'\boldsymbol{DX})\mathrm{tr}(\Delta_Y\Delta_C^2)\}. \qquad (*)$$

Dans l'expression précédente la matrice diagonale Δ_Y contient sur la diagonale les valeurs $\mathrm{var}(\boldsymbol{Yv_j})$, alors que la matrice diagonale Δ contient sur la diagonale les valeurs $\mathrm{cov}(\boldsymbol{Xu_j}, \boldsymbol{Yv_j})$.

Soit $\boldsymbol{v}$ le vecteur dont les composantes sont les éléments diagonaux de Δ_C,$\boldsymbol{u}$ le vecteur dont les composantes sont les éléments diagonaux de Δ.

La quantité $(*)$ est équivalente à la quantité

$$\frac{< \boldsymbol{u}, \boldsymbol{v} >^2}{< \boldsymbol{v}, \boldsymbol{v} >_{\Delta_Y}}.$$

Maximiser le critère h revient ainsi à maximiser

$$\frac{< \boldsymbol{u}, \boldsymbol{v} >^2}{< \boldsymbol{v}, \boldsymbol{v} >_{\Delta_Y}}.$$

On pose $\tilde{\boldsymbol{v}} = \Delta_Y^{\frac{1}{2}}\boldsymbol{v}$.

Sous contrainte que $\tilde{\boldsymbol{v}}$ soit normé, le problème revient à maximiser le critère g_2

$$g_2(\tilde{\boldsymbol{v}}) = \tilde{\boldsymbol{v}}'\Delta_Y^{-\frac{1}{2}}\boldsymbol{u}\boldsymbol{u}'\Delta_{\boldsymbol{Y}}^{-\frac{1}{2}}\tilde{\boldsymbol{v}} = (\boldsymbol{u}'\Delta_Y^{-\frac{1}{2}}\tilde{\boldsymbol{v}})^2.$$

Une solution est alors obtenue pour

$$\tilde{\boldsymbol{v}} = \Delta_Y^{-\frac{1}{2}} \boldsymbol{u} / ||\Delta_Y^{-\frac{1}{2}} \boldsymbol{u}||.$$

Ainsi $\boldsymbol{v} = \Delta_Y^{-1} \boldsymbol{u} / ||\Delta_Y^{-\frac{1}{2}} \boldsymbol{u}||$ et comme le problème est invariant par homothétie, $\boldsymbol{v} = \Delta_Y^{-1} \boldsymbol{u}$ est aussi une solution, de sorte que $\Delta_C = \Delta_Y^{-1} \Delta$ est une solution de maximisation du critère h.

La matrice Δ_C a donc sur la diagonale les coefficients des régressions simples des composantes $\boldsymbol{X u_j}$ sur $\boldsymbol{Y v_j}, j = 1, \ldots, r$.

La matrice transformée $\boldsymbol{YV} \Delta_Y^{-1} \Delta \boldsymbol{U}'$ caractérise maintenant d'une façon plus indépendante de $\boldsymbol{Y}$ la participation de $\boldsymbol{X}$ au lien de dépendance établi entre $\boldsymbol{X}$ et $\boldsymbol{Y}$. C'est une matrice qui a été définie la première fois en terme d'ajustement entre nuages d'individus dans Lafosse (1985) et qui a été nommée image concordante dans Lafosse (1997).

Pour la solution $\Delta_C = \Delta_Y^{-1} \Delta$ le critère h devient

$$h(\boldsymbol{C}) = \cos^2(\boldsymbol{X}, \boldsymbol{YV} \Delta_Y^{-1} \Delta \boldsymbol{U}') = \mathrm{tr}(\Delta^2 \Delta_Y^{-1}) / \mathrm{tr}[\boldsymbol{X}' \boldsymbol{D} \boldsymbol{X}].$$

Ce rapport entre somme de variances expliquées et variance totale qui traduit la quantité relative d'inertie de $\boldsymbol{X}$ impliquée dans la liaison entre $\boldsymbol{X}$ et $\boldsymbol{Y}$ quand les métriques sont les métriques identité avait été nommé LAI$(\boldsymbol{X} \ / \ \boldsymbol{Y})$, pour « Linear Agreement Index ». Chaque variance expliquée est décomposable en contributions partielles, contributions des individus de $\boldsymbol{X}$ à la dépendance due à la corrélation linéaire au carré correspondante.

Depuis la régression simultanée de $\boldsymbol{Y}$ sur $\boldsymbol{X}$, on peut aussi définir les contributions partielles des individus de $\boldsymbol{Y}$, contributions à la dépendance due à la même corrélation. Ces deux ensembles de contributions conduisent à associer deux biplots pour définir des quadriplots (Lafosse, 1999a).

L'analyse CONCOR est une généralisation des analyses factorielles de deux tableaux, que l'on retrouve pour $K = 1$. Quand la généralisation induit des problèmes particuliers, on situe par la suite cette particularité par rapport au cas « $K = 1$ ». Sinon, les démonstrations déjà produites pour $K = 1$ ne sont pas reprises.

3. 2. Régression de $\boldsymbol{Y}$ sur les tableaux $\boldsymbol{X_i}$

3.2.1. Image concordante

Sans que de nouvelles justifications soient nécessaires puisqu'on s'est ramené dans la section 2.2 à une dépendance entre les deux triplés $(\boldsymbol{Y}, \boldsymbol{N}, \boldsymbol{D})$ et $(\boldsymbol{XMAA}', \boldsymbol{M}, \boldsymbol{D})$, on définit une image concordante $\boldsymbol{Y_A}$ depuis des régressions simples simultanées considérées entre composantes des couples monogames $(\boldsymbol{XMa_j}, \boldsymbol{YNb_j}), j = 1, , S$.

On note Δ_b la matrice diagonale des S coefficients de régression, c'est-à-dire des valeurs :

$$\mathrm{cov}(\boldsymbol{YNb_j}, \boldsymbol{XMa_j})/\mathrm{var}(\boldsymbol{XMa_j}). \tag{13}$$

Ramenant à 1 les variances des composantes compromis $\boldsymbol{XMa_j}$ on range par colonnes ces composantes, une fois réduites, dans une matrice nommée $\boldsymbol{K}$. La régression simple simultanée de $\boldsymbol{YNB}$ sur $\boldsymbol{XMA}$ conduit à définir l'image de $\boldsymbol{Y}$ concordante avec les tableaux $\boldsymbol{X_i}$ par la matrice $n \times q$:

$$\boldsymbol{Y_A} = \boldsymbol{XMA}\Delta_b\boldsymbol{B}' = \boldsymbol{KK'DY}. \tag{14}$$

L'indice global LAI mesure l'intensité de la participation de $(\boldsymbol{Y}, \boldsymbol{N})$ à la dépendance simultanée :

$$\begin{aligned}\mathrm{LAI}[(\boldsymbol{Y}, \boldsymbol{N})/(\boldsymbol{X_i}, \boldsymbol{M_i})] &= \mathrm{tr}(\boldsymbol{Y'_A DY_A N})/\mathrm{tr}(\boldsymbol{Y'DYN}) \\ &= \sum_{i=1}^{S} \rho_j^2 \ \mathrm{var}\ (\boldsymbol{YNb_j})/\mathrm{tr}(\boldsymbol{Y'DYN}). \end{aligned} \tag{15}$$

Aucune difficulté particulière n'apparaît quand $K > 1$ pour définir les contributions partielles des individus et les participations partielles des variables du tableau $\boldsymbol{Y}$ (par $\mathrm{diag}(\boldsymbol{Y'DY_ANb_jb'_j})$).

3.2.2. Anti-image et discordance

Pour $K = 1$, le résidu de régression simple simultanée est décomposée en deux parties, l'anti-image qui a été définie par Guttman et la discordance qui traduit la non adéquation entre systèmes de covariations internes. Notant

$$\boldsymbol{P_{XMA}} = \boldsymbol{XMA}[(\boldsymbol{XMA})'\boldsymbol{DXMA})]^{-1}(\boldsymbol{XMA})'\boldsymbol{D}$$

le projecteur sur le sous-espace engendré par les colonnes de $\boldsymbol{XMA}$, cette anti-image est ici définie par

$$\boldsymbol{Y_N} = \boldsymbol{Y} - \boldsymbol{P_{XMA}}\boldsymbol{Y}. \tag{16}$$

La matrice $\boldsymbol{Y_D}$ étant l'image discordante définie par la suite, les trois matrices nulles suivantes

$$\boldsymbol{Y'_N DY_A} = \boldsymbol{Y'_N DY_D} = \boldsymbol{Y'_N DXMAA'} = \boldsymbol{0}, \tag{17}$$

rapellent que l'anti-image définie est indépendante des matrices impliquées dans la dépendance simultanée.

L'image discordante est définie par :

$$\boldsymbol{Y_D} = \boldsymbol{P_{XMA}}\boldsymbol{Y} - \boldsymbol{Y_A}. \tag{18}$$

L'existence de cette image provient de la prise en compte de la structure de $\boldsymbol{X}$. On nomme ici structure, l'ensemble des liens de corrélations entre les variables de $\boldsymbol{X}$. Considérant que c'est celle captée dans l'ACP de $\boldsymbol{X}$ par les composantes principales, la structure de $\boldsymbol{X}$ est donc invariante par isométries.

Quand toutes les métriques sont les métriques identité, la matrice $\boldsymbol{Y_D}$ traduit une différence de structure interne, induite par l'existence de liens internes dans chacun des $\boldsymbol{K}$ tableaux différents de ceux de $\boldsymbol{Y}$, alors que $\boldsymbol{Y_A}$ traduit l'analogie de la structure du nuage $(\boldsymbol{Y}, \boldsymbol{N})$ avec les structures des nuages $(\boldsymbol{X_i}, \boldsymbol{M_i})$.

Quand $K = 1$, alors $\boldsymbol{P_{XMA}Y} = \boldsymbol{P_X Y}$ (Lafosse, 1997), et pour un axe j de l'analyse (interprétable depuis les biplots de concordance), l'image discordante :

$$\boldsymbol{Y_D N b_j} = \boldsymbol{P_{XMA} Y N b_j} - \boldsymbol{Y_A N b_j}$$

a, dans le repère des axes de l'analyse, pour coordonnées :

$$\begin{aligned}\boldsymbol{Y_D N b_j} &= \boldsymbol{X(X'DX)^{-1}X'DYN b_j} - \boldsymbol{XMA}\Delta_b \boldsymbol{B'N b_j} \\ &= \lambda_j \boldsymbol{X(X'DX)^{-1} a_j} - \boldsymbol{XMA}\Delta_b b_{\delta j}\end{aligned}$$

où $\boldsymbol{b_{\delta j}}$ est vecteur colonne nul exepté le 1 en position j. Finalement

$$\boldsymbol{Y_D N b_j} = \lambda_j [\boldsymbol{X(X'DX)^{-1} a_j} - \boldsymbol{XM a_j}/||\boldsymbol{XM a_j}||^2].$$

Pour $K = 1$ quand les deux métriques sont les métriques identité, la discordance est nulle si les tableaux ont la même structure interne relativement à un axe j (alors $\boldsymbol{a_j} = \boldsymbol{b_j}$ est un axe principal de l'ACP de $\boldsymbol{X}$ et on vérifie bien que $\boldsymbol{Y_D b_j} = \boldsymbol{0}$). Quand $K > 1$, quand toutes les métriques sont les métriques identité, si un axe j est axe d'ACP pour chacun des tableaux respectifs, alors on vérifie facilement que la composante principale $\boldsymbol{Y b_j}$ de $\boldsymbol{Y}$ est proportionnelle à la composante principale $\boldsymbol{X a_j}$ de $\boldsymbol{X}$.

D'après (14), on a alors

$$\boldsymbol{P_{XA} Y b_j} = \boldsymbol{P_X Y b_j},$$

car la projection globale est totalement et uniquement réalisée par la projection partielle $\boldsymbol{X a_j (X a_j)' D}$ sur la composante principale $\boldsymbol{Y b_j}$. Par suite $\boldsymbol{Y_D b_j} = \boldsymbol{0}$.

Avec $K = 1$ et $\boldsymbol{M}$ la métrique de Mahalanobis, $||\boldsymbol{XM a_j}|| = \boldsymbol{1}$, et la discordance est encore nulle (Lafosse, 1997). La différence d'une structure interne avec une autre structure ne peut s'établir quand l'information sur cette dernière a été supprimée.

Quand $\boldsymbol{P_{XMA} Y} \neq \boldsymbol{P_X Y}$ (ce qui se produit donc en général si $K \neq 1$) quand les métriques $\boldsymbol{M_i}$ sont celles respectives de Mahalanobis, la discordance peut être non nulle. Il ne suffit donc pas cette fois de supprimer chacune des structures internes aux tableaux $\boldsymbol{X_i}$ pour trouver une discordance $\boldsymbol{Y_D}$ nulle. La suppression de chaque structure interne n'équivaut pas ici à la suppression de toute notion de structure sur $\boldsymbol{X}$. Cette discordance provient d'une différence entre les sous espaces vectoriels

engendrés pour chaque i par les variables recpectives $\{\boldsymbol{X_i M_i a_{ij}} j = 1, \ldots, S\}$. Cela se produit quand les dimensions de ces sous-espaces sont différents.

3.3. Régressions des tableaux $\boldsymbol{X_i}$ sur $\boldsymbol{Y}$

Dans le cas de deux tableaux $\boldsymbol{Y}$ et $\boldsymbol{X_1}(K = 1)$, la participation de $\boldsymbol{Y}$ à la dépendance entre $\boldsymbol{Y}$ et $\boldsymbol{X_1}$ se définit comme se définit la participation de $\boldsymbol{X_1}$. Cette répétition de définition n'est utile que pour juger de situations particulières créées par l'emploi de métriques particulières, et n'est pas utile dans le contexte général de métriques quelconques.

Une situation nouvelle est créée dans le cas de la concordance entre un tableau $\boldsymbol{Y}$ et plusieurs tableaux $\boldsymbol{X_i}(K > 1)$, l'apport spécifique de chaque tableau $\boldsymbol{X_i}$ à la dépendance simultanée devant être ici précisé.

3.3.1. Images concordantes synthétiques et partielles

On continue de développer l'analyse CONCOR en se fondant sur les couples monogames $(\boldsymbol{XMa_j}, \boldsymbol{YNb_j})$, pour obtenir cette fois les contributions de chaque tableau $\boldsymbol{X_i}$ aux trois notions de concordance, discordance et anti-image.

Les régressions simples simultanées maintenant considérées permettent de définir une matrice synthétisant les K images concordantes relatives aux K tableaux $\boldsymbol{X_i}$.

On note $\boldsymbol{D_a}$ la matrice diagonale des S coefficients de régression de la régression simple simultanée de $\boldsymbol{XMA}$ sur $\boldsymbol{YNB}$, c'est-à-dire des valeurs :

$$\operatorname{cov}(\boldsymbol{YNb_j}, \boldsymbol{XMa_j})/\operatorname{var}(\boldsymbol{YNb_j}). \tag{19}$$

Vu qu'une composante $\boldsymbol{XMa_j}$ constitue un compromis entre les composantes partenaires du $K+1$ uple j, on se sert de la régression simple de $\boldsymbol{XMa_j}$ sur $\boldsymbol{YNb_j}$ pour synthétiser sur la dimension j la participation à la concordance des tableaux $\boldsymbol{X_i}$.

On note $\boldsymbol{L}$ la matrice ayant en colonnes les composantes $\boldsymbol{YNb_j}$, après les avoir cependant réduites. Les S régressions simultanément réalisées définissent ainsi une matrice synthétique $n \times m$:

$$\boldsymbol{X_A} = \boldsymbol{YNBD_a A'} = \boldsymbol{LL'DXMAA'}. \tag{20}$$

Cette matrice est la synthèse d'images partielles $\boldsymbol{X_{Ai}}$ traduisant les contributions de chaque tableau à la concordance avec $\boldsymbol{Y}$. Ces images partielles sont les K sous-blocs de $\boldsymbol{X_A}$ de dimensions respectives $n \times m_i$. La propriété suivante montre que ces images partielles peuvent être obtenues à partir de régressions simples simultanées, sans passer par l'image synthétique.

On note $\boldsymbol{A_i}$ la matrice $m_i \times S$ ayant pour colonnes les vecteurs normés $\mathbf{a_{ij}}, j = 1, \ldots, S$.

Propriété 3.3.1. — *Les K matrices obtenues par régression simple simultanée de $\boldsymbol{X_i M_i A_i}$ sur $\boldsymbol{YNB}, i = 1, \ldots, K$, sont les images concordantes partielles $\boldsymbol{X_{Ai}}$ constituant l'image concordante synthétique $\boldsymbol{X_A}$.*

On note par $\boldsymbol{P_i}$ la matrice diagonale contenant sur la diagonale les normes p_{ij} des sous-vecteurs $p_{ij}\boldsymbol{a_{ij}}$ de $\boldsymbol{a_j}, j = 1, \ldots, S$. D'après (20), l'image partielle exprimant la concordance de $\boldsymbol{X_i}$ avec $\boldsymbol{Y}$ est définie par :

$$\boldsymbol{X_{Ai}} = \boldsymbol{YNBD_aP_iA_i'} = \boldsymbol{YNBD_{ai}A_i'}. \tag{21}$$

Or la matrice diagonale $\boldsymbol{D_{ai}} = \boldsymbol{D_aP_i}$ contient les valeurs :

$$[\mathrm{cov}(\boldsymbol{YNb_j}, \boldsymbol{XMa_j})/\mathrm{var}(\boldsymbol{YNb_j})] \times [\mathrm{cov}(\boldsymbol{X_iM_ia_{ij}}, \boldsymbol{YNb_j})/\mathrm{cov}(\boldsymbol{YNb_j}, \boldsymbol{XMa_j})]. \tag{22}$$

Ce sont donc les coefficients de régression linéaire simple de $\boldsymbol{X_iM_ia_{ij}}$ sur $\boldsymbol{YNb_j}$:

$$\mathrm{cov}(\boldsymbol{X_iM_ia_{ij}}, \boldsymbol{YNb_j})/\mathrm{var}(\boldsymbol{YNb_j}). \tag{23}$$

Donc tout se passe comme si on avait fait ce calcul de l'image partielle en partant d'une régression simple simultanée basée sur les S couples $(\boldsymbol{X_iM_ia_{ij}}, \boldsymbol{YNb_j})$.

Mais ce faisant directement, notre démarche n'aurait pu être justifiée, car n'étant pas monogames, ces couples ne peuvent être considérés séparément. En effet, $\boldsymbol{YNb_j}$ est corrélée aux composantes $\boldsymbol{X_iM_ia_{ih}}$, pour $h < j$. Par suite, on retrouve ci-après pour les images concordantes partielles les propriétés usuelles des images concordantes.

3.3.2 Décomposition des contributions de $\boldsymbol{X_i}$

On pose $\boldsymbol{X_{Ti}} = \boldsymbol{X_i} - \boldsymbol{X_{Ai}}$ et on considère la décomposition :

$$\boldsymbol{X_i} = \boldsymbol{X_{Ai}} + \boldsymbol{X_{Ti}}. \tag{24}$$

Propriété 3.3.2. — *L'inertie totale de* $\boldsymbol{X_i}$ *se décompose en somme de deux contributions totales, celle de* $\boldsymbol{X}_{A_i}$ *et celle de* $\boldsymbol{X}_{T_i}$ *. La somme des contributions relatives à la concordance est égale à la somme des variances expliquées de la régression simple simultanée qui définit l'image partielle concordante* $\boldsymbol{X}_{A_i}$.

D'après (21), pour chaque i on a :

$$\mathrm{tr}(\boldsymbol{X_i'DX_{Ai}M_i}) = \mathrm{tr}(\boldsymbol{X_i'DYNBD_{ai}A_i'M_i}) = \mathrm{tr}[(\boldsymbol{X_iM_iA_i})'\boldsymbol{DYNB}]\boldsymbol{D_{ai}}. \tag{25}$$

Soit encore, d'après (22) et (23) :

$$\mathrm{tr}(\boldsymbol{X_i'DX_{Ai}M_i}) = \sum_{j=1}^{S} \mathrm{cov}^{\boldsymbol{2}}(\boldsymbol{X_iM_ia_{ij}}, \boldsymbol{YNb_j})/\mathrm{var}(\boldsymbol{YNb_j}). \tag{26}$$

On a donc :

$$\mathrm{tr}(\boldsymbol{X_i'DX_{Ai}M_i}) = \sum_{j=1}^{S} \rho_{ij}^2 \mathrm{var}(\boldsymbol{X_iM_ia_{ij}}), \tag{27}$$

et aussi :

$$\begin{aligned}\mathrm{tr}(\boldsymbol{X}'_{\boldsymbol{Ai}}\boldsymbol{DX}_{\boldsymbol{Ai}}\boldsymbol{M_i}) &= \mathrm{tr}[(\boldsymbol{YNBD}_{\boldsymbol{ai}}\boldsymbol{A}'_{\boldsymbol{i}})'\boldsymbol{DYNBD}_{\boldsymbol{ai}}\boldsymbol{A}'_{\boldsymbol{i}}\boldsymbol{M_i}] \\ &= \mathrm{tr}[\boldsymbol{D}_{\boldsymbol{ai}}(\boldsymbol{YNB})'\boldsymbol{DYNBD}_{\boldsymbol{ai}}], \end{aligned} \tag{28}$$

Et ainsi on a :

$$\mathrm{tr}(\boldsymbol{X}'_{\boldsymbol{Ai}}\boldsymbol{DX}_{\boldsymbol{Ai}}\boldsymbol{M_i}) = \sum_{j=1}^{S} \rho_{ij}^2 \mathrm{var}(\boldsymbol{X_i M_i a_{ij}}). \tag{29}$$

De (27) et (29) on déduit alors :

$$\mathrm{tr}(\boldsymbol{X}'_{\boldsymbol{i}}\boldsymbol{DX}_{\boldsymbol{Ai}}\boldsymbol{M_i}) = \mathrm{tr}(\boldsymbol{X}'_{\boldsymbol{Ai}}\boldsymbol{DX}_{\boldsymbol{Ai}}\boldsymbol{M_i}). \tag{30}$$

Comme d'après (24) on a :

$$\boldsymbol{X}'_{\boldsymbol{Ai}}\boldsymbol{DX_i M_i} = \boldsymbol{X}'_{\boldsymbol{Ai}}\boldsymbol{DX}_{\boldsymbol{Ai}}\boldsymbol{M_i} + \boldsymbol{X}'_{\boldsymbol{Ai}}\boldsymbol{DX}_{\boldsymbol{Ti}}\boldsymbol{M_i}, \tag{31}$$

on obtient :

$$\mathrm{tr}(\boldsymbol{X}'_{\boldsymbol{Ai}}\boldsymbol{DX}_{\boldsymbol{Ti}}\boldsymbol{M_i}) = \boldsymbol{0}. \tag{32}$$

En usant encore de (24), on obtient alors :

$$\mathrm{tr}(\boldsymbol{X}'_{\boldsymbol{Ti}}\boldsymbol{DX}_{\boldsymbol{Ti}}\boldsymbol{M_i}) = \mathrm{tr}(\boldsymbol{X}'_{\boldsymbol{i}}\boldsymbol{DX}_{\boldsymbol{Ti}}\boldsymbol{M_i}). \tag{33}$$

Comme d'après (24) on a :

$$\boldsymbol{X}'_{\boldsymbol{i}}\boldsymbol{DX_i M_i} = \boldsymbol{X}'_{\boldsymbol{i}}\boldsymbol{DX}_{\boldsymbol{Ai}}\boldsymbol{M_i} + \boldsymbol{X}'_{\boldsymbol{i}}\boldsymbol{DX}_{\boldsymbol{Ti}}\boldsymbol{M_i}. \tag{34}$$

Finalement pour chaque i, l'égalité (24) est relative à une décomposition de l'inertie

$$\mathrm{tr}(\boldsymbol{X}'_{\boldsymbol{i}}\boldsymbol{DX_i M_i}) = \mathrm{tr}(\boldsymbol{X}'_{\boldsymbol{Ai}}\boldsymbol{DX}_{\boldsymbol{Ai}}\boldsymbol{M_i}) + \mathrm{tr}(\boldsymbol{X}'_{\boldsymbol{Ti}}\boldsymbol{DX}_{\boldsymbol{Ti}}\boldsymbol{M_i}). \tag{35}$$

Si tout semble donc se passer comme en concordance d'un tableau avec un autre tableau, on remarque cependant qu'ici les matrices $\boldsymbol{X}'_{\boldsymbol{i}}\boldsymbol{DX}_{\boldsymbol{Ai}}$ et $\boldsymbol{X}'_{\boldsymbol{i}}\boldsymbol{DX}_{\boldsymbol{Ti}}$ ne sont pas symétriques.

En fait, beaucoup de décompositions en somme de type (24) mènent à un découpage d'inertie de type (35). Mais celui proposé a été justifié depuis le découpage initial de relation fondé sur les couples monogames de l'analyse CONCOR.

Les participations globales à la concordance simultanée des variables de $\boldsymbol{X}$ s'évaluent par les termes diagonaux :

$$\mathrm{diag}(\boldsymbol{X}'\boldsymbol{DX}_{\boldsymbol{A}}) = \mathrm{diag}(\boldsymbol{X}'\boldsymbol{DLL}'\boldsymbol{DXMAA}'). \tag{36}$$

Quand $\boldsymbol{X} = \boldsymbol{XMAA'}$, les participations se décomposent par axe, leurs valeurs positives étant pour un axe j égales à (Lafosse, 1997) :

$$\text{diag}(\boldsymbol{AA'MX'DLL'DXMa_j a_j'}) = \rho_j^2 \text{ var } (\boldsymbol{XMa_j}) \text{ diag } (\boldsymbol{a_j a_j'}). \qquad (37)$$

Mais quand $\boldsymbol{X} \neq \boldsymbol{XMAA'}$, des valeurs négatives apparaissent, correspondant à quelques covariances négatives entre les colonnes de $\boldsymbol{X}$ et celles de $\boldsymbol{X_A}$. Cela traduit une difficulté à être capable d'établir aussi clairement que l'on voudrait la dépendance simultanée de $\boldsymbol{Y}$ avec chaque tableau $\boldsymbol{X_i}$: les signes des vecteurs $\boldsymbol{a_{ij}}$ sont bien, par construction, fonctions du signe du vecteur $\boldsymbol{b_j}$ de façon à retrouver les individus des tableaux respectifs projetés sur les axes respectifs dans des dispositions analogues. Mais cela étant, le découpage correspondant de la variabilité de $\boldsymbol{X}$ ne peut être toujours parfait comme pour $K = 1$. Cependant les valeurs négatives sont souvent assez faibles parce que souvent associées à des axes porteurs d'une faible inertie et parce que les dimensions supplémentaires aux dimensions $\boldsymbol{a_{ij}}$ leur étant orthogonales, la correction qu'elles induisent est nulle en moyenne pour chaque axe j. On tient compte de ces signes négatifs sur les biplots en modifiant alors les valeurs des signes correspondants des cosinus directeurs des flèches ($\boldsymbol{a_j}$ devenant ainsi $\boldsymbol{d_j}$) sans changer les coordonnées des individus de sorte que les participations deviennent égales à

$$\text{diag}(\boldsymbol{X'DLL'DXMa_j d_j'}) = \textbf{abs}[\text{diag}(\boldsymbol{X'DLL'DXMa_j a_j'})].$$

Cela induit une redondance des participations par rapport aux inerties, qui traduit l'idée qu'il est nécessaire de « se répéter » pour arriver à prendre en compte toutes les relations qu'ont les variables avec les contributions des individus.

3.3.3. Discordances et anti-images. Indices.

L'anti-image relative à $\boldsymbol{X_i}$ définit la part de contributions de $\boldsymbol{X_i}$ n'ayant aucune influence dans l'analyse parce que non corrélée à $\boldsymbol{Y}$.

Notant $\boldsymbol{P_Y} = \boldsymbol{Y}(\boldsymbol{Y'DY})^{-1}\boldsymbol{Y'D}$ le projecteur sur le sous-espace engendré par toutes les colonnes de $\boldsymbol{Y}$, on définit ainsi les anti-images partielles (relatives aux tableaux $\boldsymbol{X_i}$) par :

$$\boldsymbol{X_{Ni}} = \boldsymbol{X_i} - \boldsymbol{P_Y X_i}. \qquad (38)$$

Ces anti-images décomposent par blocs l'anti-image synthétique

$$\boldsymbol{X_N} = \boldsymbol{X} - \boldsymbol{P_Y X}. \qquad (39)$$

Les images concordantes partielles obtenues dans l'analyse de la concordance du tableau $\boldsymbol{Y}$ avec les tableaux $\boldsymbol{P_Y X_i}$ sont encore les images $\boldsymbol{X_{Ai}}$ précédemment définies. Il suffit de remplacer les $\boldsymbol{X_i}$ par les tableaux $\boldsymbol{P_Y X_i}$ pour le constater, les calculs faits avec la matrice donnée en (2) restant alors inchangés.

Les images partielles discordantes sont alors définies par $\boldsymbol{X_{Di}} = \boldsymbol{P_Y X_i} - \boldsymbol{X_{Ai}}$ et elles décomposent par blocs l'image discordante synthétique :

$$\boldsymbol{X_D} = \boldsymbol{P_Y X} - \boldsymbol{X_A} \qquad (40)$$

Finalement et comme en concordance d'un tableau avec un autre, on décompose chaque tableau $\boldsymbol{X_i}$ de dimension $n \times m_i$ en trois tableaux de même dimension :

$$\boldsymbol{X_i} = \boldsymbol{X_{Ai}} + \boldsymbol{X_{Di}} + \boldsymbol{X_{Ni}}, \tag{41}$$

Cette décomposition correspond bien à un découpage des contributions de $\boldsymbol{X_i}$ en trois ensembles de contributions, des contributions partielles des individus de $\boldsymbol{X_i}$ à la dépendance simultanée avec $\boldsymbol{Y}$, à la discordance et aux anti-images :

$$\mathrm{tr}(\boldsymbol{X_i'DX_iM_i}) = \mathrm{tr}(\boldsymbol{X'_{Ai}DX_{Ai}M_i}) + \mathrm{tr}(\boldsymbol{X'_{Di}X_{Di}M_i}) + \mathrm{tr}(\boldsymbol{X'_{Ni}DX_{Ni}M_i}). \tag{42}$$

On a finalement :

$$\boldsymbol{X} = [\boldsymbol{X_{A1}X_{A2}}\ldots\boldsymbol{X_{Ai}}\ldots\boldsymbol{X_{AK}}] + [\boldsymbol{X_{D1}X_{D2}}\ldots\boldsymbol{X_{Di}}\ldots\boldsymbol{X_{DK}}] + [\boldsymbol{X_{N1}X_{N2}}\ldots\boldsymbol{X_{Ni}}\ldots\boldsymbol{X_{NK}}]. \tag{43}$$

Indices partiels et synthétiques

D'après (35) on peut proposer pour indice partiel $\mathrm{LAI_P}$ de mesure de l'intensité de la participation partielle du tableau $(\boldsymbol{X_i}, \boldsymbol{M_i})$ à la concordance avec $(\boldsymbol{Y}, \boldsymbol{N})$:

$$\mathrm{LAI_P}[(\boldsymbol{X_i}, \boldsymbol{M_i})/(\boldsymbol{Y}, \boldsymbol{N})] = \mathrm{tr}(\boldsymbol{X'_{Ai}DX_{Ai}M_i})/\mathrm{tr}(\boldsymbol{X_i'DX_iM_i}). \tag{44}$$

Le numérateur étant décomposable en somme de variances expliquées (29), rangeant ces variances par ordre décroissant, on peut fonder sur cet ordre l'établissement de résumés concernant la participation du tableau $\boldsymbol{X_i}$ à la concordance.

D'après (30), cet indice est interprétable comme un cosinus mesurant la proximité entre $\boldsymbol{X_i}$ et l'image concordante partielle de $\boldsymbol{X_i}$ avec $\boldsymbol{Y}$:

$$\begin{aligned}\cos^2_{(D,M_i)}(\boldsymbol{X_i}, \boldsymbol{X_{Ai}}) &= \mathrm{tr}^{\boldsymbol{2}}(\boldsymbol{X_i'DX_{Ai}M_i})/\mathrm{tr}(\boldsymbol{X_i'DX_iM_i})\mathrm{tr}(\boldsymbol{X_{Ai'}DX_{Ai}M_i}) \\ &= \mathrm{LAI_P}[(\boldsymbol{X_i}, \boldsymbol{M_i})/(\boldsymbol{Y}, \boldsymbol{N})].\end{aligned}$$

De la même façon que par (44) et pour chaque tableau, peuvent se définir un indice partiel de discordance et un indice partiel de l'anti-image.

Les indices partiels (44) indiquent les contributions relatives de chaque tableau $\boldsymbol{X}_i$ dans leur relation de concordance avec $\boldsymbol{Y}$.

La moyenne pondérée suivante de ces indices constitue alors une mesure globale de l'importance de l'image concordante synthétique $\boldsymbol{X_A}$, puisque l'on a :

$$\sum_{i=1}^{K} \frac{\mathrm{tr}(\boldsymbol{X_i'DX_iM_i})}{\mathrm{tr}(\boldsymbol{X'DXM})}\mathrm{LAI_P}[(\boldsymbol{X_i}, \boldsymbol{M_i})/(\boldsymbol{Y}, \boldsymbol{N})] = \frac{\mathrm{tr}(\boldsymbol{X_A'DX_AM})}{\mathrm{tr}(\boldsymbol{X'DXM})} = \mathrm{LAI}[\boldsymbol{XMAA'}, \boldsymbol{M})/(\boldsymbol{Y}, \boldsymbol{N})]. \tag{45}$$

De même, les moyennes pondérées par les mêmes poids des indices partiels de discordance (resp. des indices partiels de l'anti-image) conduisent à une mesure globale de l'importance de la discordante synthétique (resp. à une mesure globale de l'importance de l'anti-image synthétique).

3.4. Graphiques juxtaposés

Quand plusieurs matrices sont analysées simultanément, les possibilités de graphiques juxtaposés deviennent nombreuses. L'objectif est ici d'apporter des arguments en faveur d'un nombre assez limité d'entre elles.

Par exemple, la représentation des contributions des individus-lignes de $\boldsymbol{X_{Di}}$ par sous-espace pourrait se faire en réalisant l'ACP du triplet statistique $(\boldsymbol{X_{Di}}, \boldsymbol{M_i}, \boldsymbol{D})$.

Mais il est en général judicieux de juxtaposer les graphiques des trois images dans les repères des axes de concordance, cette juxtaposition étant justifiée par la propriété 3.4 suivante, qui décompose axe par axe la relation (42). L'interprétation des axes est celle qui aura été établie en concordance avec les biplots.

Propriété 3.4. — *L'inertie du nuage* $(\boldsymbol{X_i}, \boldsymbol{M_i})$ *projeté sur un axe* $\mathbf{a_{ij}}$ *est égale à la somme des inerties des projetés sur cet axe de chacun des trois nuages* $(\boldsymbol{X_{Ai}}, \boldsymbol{M_i})$, $(\boldsymbol{X_{Di}}, \boldsymbol{M_i})$ *et* $(\boldsymbol{X_{Ni}}, \boldsymbol{M_i})$.

Le coefficient ρ_{ij} est la corrélation linéaire entre $\boldsymbol{X_{Ai}M_ia_{ij}}$ et $\boldsymbol{YNb_j}$. On remarque d'abord que l'on a, d'après (21), (23) et sachant l'orthonormalité des $\mathbf{a_{ij}}$ pour i fixé :

$$\begin{aligned}
\boldsymbol{a'_{ij}M_iX'_iDX_{Ai}M_ia_{ij}} &= \boldsymbol{a'_{ij}M_iX'_iDYNb_j}\mathrm{cov}(\boldsymbol{X_iM_ia_{ij}}, \boldsymbol{YNb_j})/\mathrm{var}(\boldsymbol{YNb_j}) \\
&= \mathrm{cov}^{\boldsymbol{2}}(\boldsymbol{X_iM_ia_{ij}}, \boldsymbol{YNb_j})/\mathrm{var}(\boldsymbol{YNb_j}) \\
&= \rho^2_{ij}\mathrm{var}(\boldsymbol{X_iM_ia_{ij}}) \\
&= \mathrm{var}(\boldsymbol{X_{Ai}M_ia_{ij}})
\end{aligned}$$

Soit finalement :

$$\boldsymbol{a'_{ij}M_iX'_iDX_{Ai}M_ia_{ij}} = \boldsymbol{a'_{ij}M_iX'_{Ai}DX_{Ai}M_ia_{ij}}. \tag{46}$$

Or d'après (34) et (41) et puisque $\boldsymbol{X'_{Ni}DX_{Ai}}$ est une matrice nulle, on a :

$$\boldsymbol{a'_{ij}M_iX'_iDX_{Ai}M_ia_{ij}} = \boldsymbol{a'_{ij}M_iX'_{Ai}DX_{Ai}M_ia_{ij}} + \boldsymbol{a'_{ij}M_iX'_{Di}DX_{Ai}M_ia_{ij}}. \tag{47}$$

On en déduit avec (46) que :

$$\boldsymbol{a'_{ij}M_iX'_{Di}DX_{Ai}M_ia_{ij}} = \mathbf{0}. \tag{48}$$

Par suite, puisque $\boldsymbol{X'_{Ni}DX_{Di}} = \boldsymbol{0}$, on a aussi :

$$\boldsymbol{a'_{ij}M_iX'_iDX_{Di}M_ia_{ij}} = \boldsymbol{a'_{ij}M_iX'_{Di}DX_{Di}M_ia_{ij}}. \tag{49}$$

Finalement, comme :

$$\boldsymbol{a'_{ij}M_iX'_iDX_iM_ia_{ij}} = \boldsymbol{a'_{ij}M_iX'_iDX_{Ai}M_ia_{ij}} + \boldsymbol{a'_{ij}M_iX'_iDX_{Di}M_ia_{ij}} + \boldsymbol{a'_{ij}M_iX'_iDX_{Ni}M_ia_{ij}},$$

on a bien :

$$\boldsymbol{a'_{ij}M_iX'_iDX_iM_ia_{ij}} = \boldsymbol{a'_{ij}M_iX'_{Ai}DX_{Ai}M_ia_{ij}} + \boldsymbol{a'_{ij}M_iX'_{Di}DX_{Di}M_ia_{ij}} + \boldsymbol{a'_{ij}M_iX'_{Ni}DX_{Ni}M_ia_{ij}},$$

Chacun des trois graphiques à juxtaposer est donc obtenu par projection des nuages respectifs $(\boldsymbol{X_{Ai}}, \boldsymbol{M_i}), (\boldsymbol{X_{Di}}, \boldsymbol{M_i})$ et $(\boldsymbol{X_{Ni}}, \boldsymbol{M_i})$ dans un sous-espace commun, engendré par un sous-ensemble de deux ou trois vecteurs du système d'axes $\{\boldsymbol{a_{ij}}\} j = 1, \ldots, S$.

Cette complémentarité entre inerties pour les graphiques juxtaposés peut être complétée, comme en analyse de concordance d'un tableau avec un autre, par une complémentarité de mesures partielles de l'intensité par axe de chacune des trois notions de concordance, discordance et anti- image.

Deux types de graphiques peuvent donc être préconisés : un premier type juxtapose les $K + 1$ biplots de concordance dans les repères qui se correspondent et qui permettent d'interpréter les axes (Lafosse, 1999a), ou juxtapose dans ces repères les individus des $K + 1$ images discordantes, ou encore des $K + 1$ anti-images; avec le deuxième type, images concordantes, discordance et anti-image sont juxtaposés en se servant du même repère.

4. Application

4.1. Présentation des données

On reprend ici des données d'écologie marine largement décrites par diverses approches multivariées dans Fromentin *et al.* (1993). On pourra donc s'y référer pour une connaissance plus complète. Les 68 relevés successifs effectués en s'éloignant de la côte, sont ici répartis en 5 zones, reprenant en partie une subdivision adoptée par Fromentin. Ces relevés constituent les 68 lignes des deux tableaux de départ.

Zone 0 : relevés de 1 à 16, zone côtière.

zone 1 : relevés de 17 à 25, zone de divergence D1.

Zone 2 : relevés de 26 à 30, zone de convergence.

Zone 3 : relevés de 31 à 41, zone frontale.

Zone 4 : relevés de 42 à 46, zone de divergence D2.

Zone 5 : relevés de 47 à 68, zone centrale.

En fait, la zone côtière 0 a été supprimée, suite à une étude préliminaire. En effet, cette zone s'est révélée très prépondérante par rapport à toutes les autres quant à la dépendance créée entre les deux tableaux. Cette première étude pouvant donc préciser surtout le rôle spécifique de la zone 0, nous passons ici directement à une deuxième étude relative aux autres zones. On a cependant gardé les numéros initiaux des relevés.

Quatre paramètres hydrologiques ont été mesurés pour chaque relevé : température, salinité, fluorescence et densité. Ils définissent un tableau de dimension 52×4 qui a été centré et réduit pour donner un tableau $\boldsymbol{Y}$.

Dans le même temps, le dénombrement de 24 espèces est réalisé à partir d'une même masse d'eau et du plancton chaque fois recueilli, constituant les données biologiques. Le logarithme (+1) de ces effectifs conduit à un tableau $\boldsymbol{X}$ centré et réduit de dimension 52×24. Les espèces sont :

Acartia (Acar)	Copepodits 5 (Cop5)	Nauplii (Naup)	Gasteropods larvae (Gast)
Adults of Calanus (Cala)	Clausocalanus A (ClauA)	Oithona (Oith)	Eggs of crustaceans (Eggs)
Copepodits1 (Cop1)	Clausocalanus B (ClauB)	Acanthaires (Acan)	Ostracods (Ostr)
Copepodits 2 (Cop2)	Clausocalanus C (ClauC)	Cladocerans (Clad)	Pteropods (Pter)
Copepodits 3 (Cop3)	Adults of Centropages (CenA)	Echinoderms larvae (Echi)	Siphonophores (Siph)
Copepodits 4 (Cop4)	Juvenils of Centropages (CenJ)	Decapods larvae (Deca)	Bells of calycophore(Bell)

Ici, les 24 espèces biologiques, qui ont toutes le même poids sachant la réduction des variables, ont été partitionnées en 3 groupes définis d'après l'importance de densité maximale observée sur les 68 relevés. Le groupe I est ainsi constitué de 8 espèces qui en nombre ne dépasse jamais l'effectif 200 par relevé (par exemple sans jamais dépasser 8 pour Pteropods). Le groupe III est constitué de 8 espèces pouvant aller jusqu'à l'effectif 10 000 sur un seul relevé et le groupe II est constitué de 8 espèces intermédiaires de ce point de vue. Les 3 groupes définissant les 3 tableaux $\boldsymbol{X_i}$ sont alors :

G I	Ostr	Acan	Siph	Clad	Echi	Pter	Gast	Bell
G II	Cop1	Cop2	Cop3	Cop4	Cop5	Naup	Eggs	Deca
G III	Acar	Cala	ClauA	ClauB	ClauC	CenA	CenJ	Oith

4.2. Une analyse CONCOR

Une des analyses menées par Fromentin est l'analyse canonique des 2 tableaux initiaux formés des 68 relevés, permettant l'étude de dépendance sans prise en compte des interactions entre espèces ni des corrélations entre paramètres hydrologiques.

La dépendance simultanée des quatre paramètres hydrologiques avec ces 3 groupes biologiques est ici étudiée en prenant en compte les interactions possibles que les espèces peuvent avoir quant à leurs présence-absences à l'intérieur de chaque groupe $\boldsymbol{X_i}, i = 1, \ldots, 3$. C'est dire que les trois métriques correspondantes $\boldsymbol{M_i}$ choisies sont les métriques identité.

De même, notre problème n'étant pas de définir des combinaisons linéaires des paramètres hydrologiques susceptibles de servir au mieux d'explication pour la présence des espèces, mais de détailler la dépendance entre tableaux, les raisons de ne pas tenir compte des fortes corrélations qui existent entre ces paramètres semblent alors manquer. La métrique $\boldsymbol{N}$ est ainsi choisie égale à $\boldsymbol{I}$.

TABLEAU 1
Contributions par axes.

	hydrologiques				Biologique GI				Biologique GII				Biologique GIII		
axe	conc	disc	anti		conc	disc	anti		conc	disc	anti		conc	disc	anti
1	582	53	159		49	1	17		74	1	26		168	3	58
2	97	4	61		59	1	39		110	1	72		84	1	55
3	12	1	31		3	1	7		11	3	27		15	5	37
4	0	0	0		5	2	35		3	1	20		1	0	3
	691	58	252		117	4	98		198	7	145		268	9	154
	1000				219				350				431		

Les contributions sont exprimées en millièmes. La somme des contributions des 3 groupes vaut 1000. Ces mesures sont explicitées en section 3.2 pour les mesures hydrologiques (les 3 colonnes correspondant respectivement aux 3 tableaux de coordonnées (14), (18) et (16)) et en section 3.3 pour les mesures biologiques (tableaux de coordonnées (20), (40) et (39)).

On constate d'après la table 1 des contributions que les axes 3 et 4 apportent une information relative négligeable en ce qui concerne la dépendance entre tableaux (colonnes **conc** comme concordance). En fait il ne peut en être tellement autrement, sachant que les 4 dimensions de départ du tableau $\boldsymbol{Y}$ pourraient être ramenées à 2, avec une ACP par exemple, sans perte sensible d'information.

Les espèces les plus dépendantes des paramètres hydrologiques sont les espèces les plus denses si on se réfère aux pourcentages 43.1 %, 35 % et 21.9 %.

Les faibles discordances calculées sur les 3 groupes biologiques indiquent que l'analyse serait peu changée si on supprimait l'information sur les corrélations entre paramètres hydrologiques (ce qui induirait la nullité des discordances sur les 3 groupes). Si on se réfère au graphique correspondant de la figure 1 on voit en effet que la structure des 4 variables est assez bien résumable par 2 variables réduites très peu corrélées, comme température et densité. Ainsi donc l'introduction de la métrique de Mahalanobis sur $\boldsymbol{Y}$ ne changerait pas tellement ce résumé.

Par ailleurs, le pourcentage de 5.8 % est la mesure globale de perte de dépendance entre tableaux due à la prise en compte des interactions entre espèces à l'intérieur de chacun des 3 groupes. Il apparaît plutôt faible aussi, et on pourrait penser que cela est partiellement dû à notre façon de constituer les trois groupes, dont le but n'était pas de rassembler entre elles les espèces pouvant interagir. Or l'analyse faite en rassemblant

en un seul groupe toutes les espèces ensemble donne un pourcentage de $5.9\,\%$. Tout semble indiquer alors que notre répartition des espèces en 3 groupes n'a pas détruit l'essentiel des interactions entre espèces.

C'est pour les relevés 20, 32 et 49 que l'interaction entre espèces vient le plus contrarier la dépendance espèces - paramètres hydrologiques (contributions à la discordance observée plus élevées pour les 3 lignes correspondantes de $\boldsymbol{Y_D}$). L'analyse faite avec un seul groupe réunissant les 3 groupes désigne aussi d'abord ces 3 relevés.

Les cumuls de contributions à la discordance $\boldsymbol{Y_D}$ (table 2) précisent que la zone 2 est celle pour laquelle l'interaction entre les espèces n'a pas d'influence sur la dépendance. Les mesures ci-après sont relativisées par le nombre de relevés dans chaque zone.

La figure 1 détaille la dépendance dans les dimensions 1 et 2. Les 4 facteurs hydrologiques ont un rôle également important quant à leur influence sur les espèces puisque les longueurs de flèches sont sensiblement égales (biplot Y en haut à gauche). Sur les 3 autres biplots les plus grandes longueurs de flèches indiquent les espèces les plus dépendantes des facteurs hydrologiques, la même échelle étant utilisée pour les trois graphiques. Par construction, les directions et sens des flèches sur ces trois biplots sont associables aux mêmes directions considérées sur le biplot Y (et par là associables entre elles). Ainsi la présence des espèces Clausocalanus C et Decapods est fortement liée à une densité de l'eau élevée. Ainsi les relevés de zone 5 associés à une température et une salinité élevées se caractérisent par une absence relative de la plupart des espèces. L'interprétation des axes est donnée par les flèches plutôt que par les points. Les 4 nuages de points, codés par l'appartenance des relevés à leur zone, se ressemblent. Ils seraient identiques si la dépendance était totale (avec des corrélations calculées égales à 1). Les relevés les plus éloignés de l'origine sont ceux qui contribuent le plus à la dépendance. Les cercles tracés servent à faciliter la comparaison de tels éloignements. L'échelle, seulement indiquée pour le premier groupe (subset1), est celle aussi adoptée pour les 2 autres biplots des 2 autres groupes (subset 2 et 3). Le nuage le plus ramassé autour de l'origine désigne ainsi le Groupe I comme celui participant le moins à la dépendance.

L'analyse de dépendance avec Y faite en réunissant les trois groupes en un même groupe présente une forte analogie : on a vérifié que le graphe unique où toutes les espèces sont alors représentées correspond assez à la superposition des trois graphes associés aux trois groupes ici obtenus. En fait cela se produit parce que nos données sont bien particulières : on a déjà remarqué que la dépendance entre espèces et les paramètres hydrologiques était captée pour l'essentiel par des espaces de dimension 2, sachant la structure des paramètres; donc rien n'est changé à ce sujet quand on divise les espèces en trois groupes, la dépendance des espèces s'établissant sensiblement avec le même sous-espace de dimension 2 dans les deux approches. L'intérêt de la partition est ici moins exploitable à cause de cette pauvreté en dimensions.

La figure 2 décrit les contributions des relevés à la discordance, en usant des mêmes repères respectifs utilisés en concordance. L'interprétation des axes est donc inchangée : c'est celle donnée par les flèches sur les biplots de concordance. Les quatre graphiques peuvent être respectivement associés aux 4 biplots précédents, explicitant des contributions complémentaires.

TABLEAU 2

Sensibilité relative de la dépendance étudiée aux interactions entre espèces.

zone 1	zone 2	zone 3	zone 4	zone 5
.24	.07	.25	.18	.25

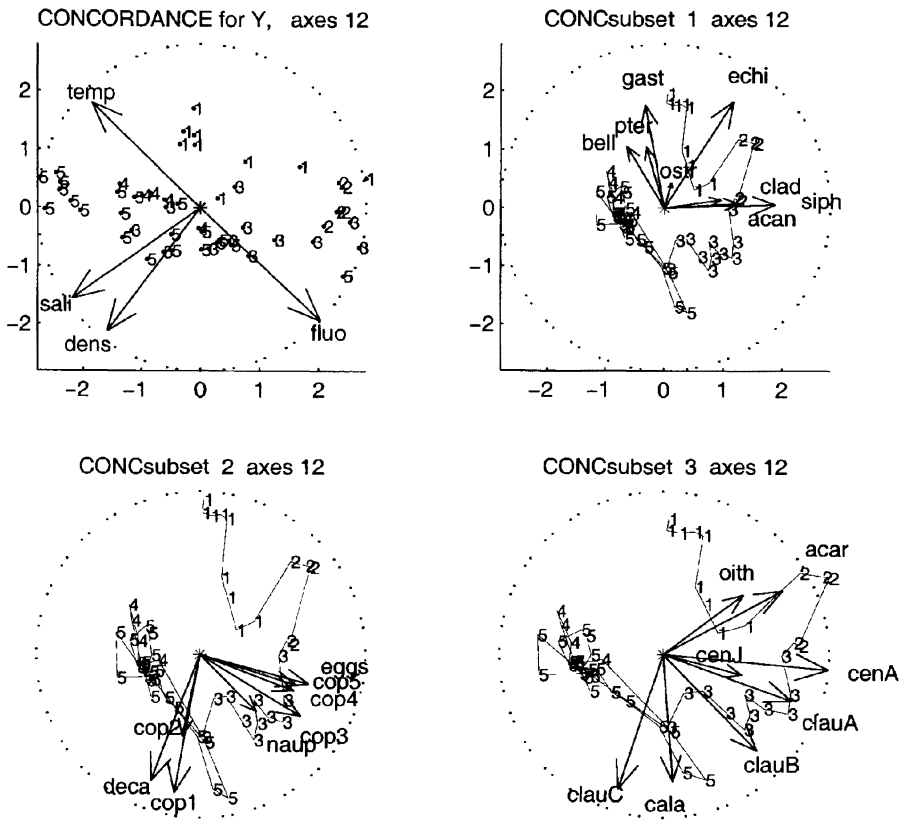

FIGURE 1

Les longueurs de flèches sont indiquées en section 3.2.1 pour le biplot Y et formule (37) pour les 3 autres biplots. Les vecteurs b_1 et b_2 pour Y, a_{i1} et a_{i2}, avec $i = 1, 2$ et 3 respectivement pour les 3 autres biplots, contiennent les cosinus directeurs des flèches. Les points correspondent aux coordonnées des lignes de Y_A projetées dans le repère (b_1, b_2) pour le biplot Y et aux coordonnées des lignes de X_{A_i} dans les repères respectifs $(a_{i1}, a_{i2}), i = 1, 2$ et 3, pour les trois autres biplots.

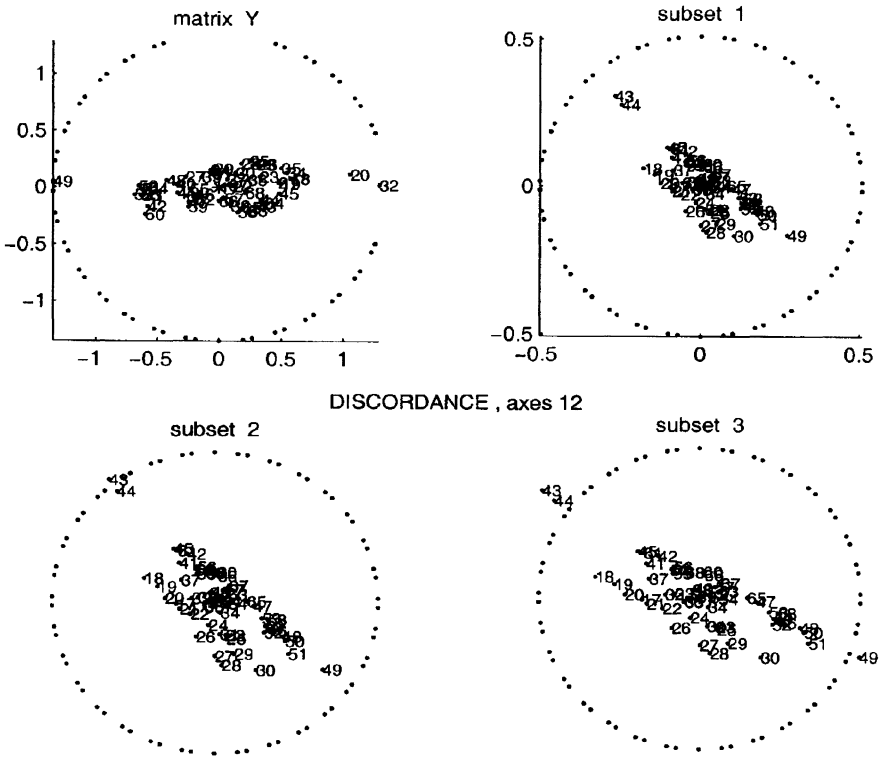

FIGURE 2

Coordonnées des lignes de Y_D dans le repère (b_1, b_2) pour le plot Y et coordonnées des lignes de X_{Di} dans les repères respectifs (a_{i1}, a_{i2}), $i = 1$, 2 et 3, pour les trois autres plots.

La discordance peut paraître assez négligeable. Cependant l'allongement des nuages traduit une cohérence trouvée sur l'ensemble des relevés, qui peut être prise en compte. L'allongement horizontal observé sur le graphique Y indique que les espèces moins dépendantes des paramètres hydrologiques pour la raison qu'elles tiendraient compte de la présence ou de l'absence d'autres espèces seraient plutôt celles constitutives de l'axe 1, comme centropages adultes. Les autres trois nuages sont tous obliquement allongés dans la même direction. Cette répétition est naturelle, la discordance caractérisant quand $K > 1$ et pour des axes relatifs à des corrélations élevées une façon commune pour les trois nuages de se différencier de Y. C'est l'interaction entre température et fluorescence qui est ici désignée comme un facteur de réduction de la dépendance entre paramètres hydrologiques et les trois groupes.

La figure 3 révèle les contributions à l'indépendance entre espèces et paramètres hydrologiques pour les axes 1 et 2. Ces axes sont encore les mêmes axes respectifs précédents et les contributions respectives sont à nouveau complémentaires. Le fait que de mêmes relevés (comme 17 ou 38) apparaissent loin de l'origine sur les 4

graphiques signifient que la contribution à l'indépendance de ces relevés n'est pas liée à l'appartenance à un groupe d'espèces particulier. Plus globalement, l'ensemble des mesures caractérisent une façon originale pour chaque groupe d'être assez indépendant des paramètres biologiques, et cela pourrait déjà suffire à justifier la partition. Celle du groupe 1 se fait plutôt selon l'axe 2, grâce à la présence de Gastéropodes, Echinoderms et Bells dans les zones 1-2 et de leur absence dans la zone 3. Celle du groupe 2 se fait aussi plus selon l'axe 2, grâce notamment la présence de Decapods et Copepodits 1 dans les zones 3 et 5 et leurs absences des zones 1 et 2. Celle du groupe 3 est plus diffuse, associée aux deux axes simultanément, des espèces étant présentes presque dans toutes les zones.

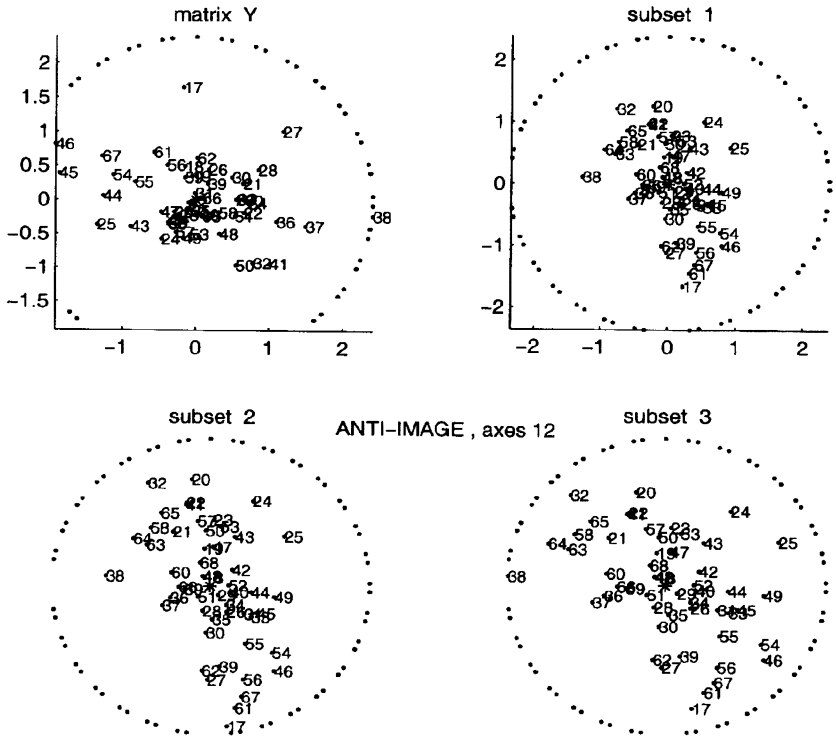

FIGURE 3

Coordonnées des lignes de Y_N dans le repère (b_1, b_2) pour le plot Y et coordonnées des lignes de X_{Ni} dans les repères respectifs (a_{i1}, a_{i2}), $i = 1$, 2 et 3, pour les trois autres plots.

Références

CAZES P., BAUMERDER A., BONNEFOUS S. & PAGES J.P. (1977), Codage et analyse des tableaux logiques. Introduction à la pratique des variables qualitatives. *Cahiers du BURO*, **27**, Paris VI.

CLIFF N. (1966), Orthogonal rotation to congruence. *Psychometrika*, **31**, 33-42.

FROMENTIN J.M., IBANEZ F. & LEGENDRE P. (1993), A Phytosociological Method for interpreting Plankton data. *Mar.Ecol.Progr.*Series 93 : 285-306.

GREEN B.F.(1952), The Orthogonal Approximation of an Oblique Structure in factor analysis. *Psychometrika*, **17**, 429-440.

GUTTMAN L. (1953), Image theory for the structure of quantitative variates. *Psychometrika*, **18**, 277-296.

HANAFI M. (1997), Structure de l'ensemble des analyses multivariées des tableaux de données à trois entrées : Eléments théoriques et appliqués. *Thèse*, Univ. Claude Bernard, Lyon I.

HORST (1961), Relation among m sets of variables. *Psychometrika*, **38**, 433-451.

LAFOSSE R. (1985), Une nouvelle analyse procrustéenne de deux tableaux. *Data Analysis and Informatics, IV (Diday & coll. eds). Elsevier Science Pub.* North Holland, 407-414.

LAFOSSE R. (1997), Analyse de concordance de deux tableaux : monogamies, simultanéités et découpages. *Rev. Stat. Appliquée*. vol. 45. n°3. 45-72.

LAFOSSE R. & HANAFI M. (1997), Concordance d'un tableau avec K tableaux : définition de $K+1$ uples synthétiques. *Rev. Stat. Appliquée*. vol 45 n°4. 111-126.

LAFOSSE (1998), Programme de l'analyse CONCOR en langage MATLAB. version 1.3. http ://www.mathworks.com/support/ftp/statv5.shtml ou mel à : lafosse@cict.fr.

LAFOSSE (1999a), Analyses of some relation between arrays and graphics. *Proc. Int. Conf. on Probability and Statistics and their Applications* (Hanoï).

LAFOSSE (1999b), Analysis of concordance between matrices and proposals for selecting variables. *IX Int. Symp. on Applied Stochastic Models and Data Analysis* (Lisbonne).

TEN BERGE J.M.F. (1977), Optimizing Factorial Invariance. *Thesis Groningen*.

TEN BERGE J.M.F. (1977), Orthogonal procrustes rotation for two or more matrices. *Psychometrika*, **42**, **2**, 267-273.

TEN BERGE J.M.F. (1988), Generalized approaches to the maxbet problem and the maxdiff problem, with applications to canonical correlations. *Psychometrika*, **52**, **4**, 487-494.

TUCKER L.R. (1958), An interbattery method of factor analysis. *Psychometrika*, **23**, 111-136.

VAN DE GEER P. (1984), Linear relations among K sets of variables. *Psychometrika*, **49**,**1**, 79-94.