REVUE DE STATISTIQUE APPLIQUÉE

J. L. BEFFY

Application de l'analyse en composantes principales à trois modes pour l'étude physico-chimique d'un écosystème lacustre d'altitude : perspectives en écologie

Revue de statistique appliquée, tome 40, n° 1 (1992), p. 37-56

<http://www.numdam.org/item?id=RSA_1992__40_1_37_0>

Rev. Statistique Appliquée, 1992, **XXXX** (1), 37-56

APPLICATION DE L'ANALYSE EN COMPOSANTES PRINCIPALES A TROIS MODES POUR L'ÉTUDE PHYSICO-CHIMIQUE D'UN ÉCOSYSTÈME LACUSTRE D'ALTITUDE : PERSPECTIVES EN ÉCOLOGIE

J.L. BEFFY

Laboratoire d'Ecologie des Eaux Douces, URA 1451, Université Lyon I, 69622 Villeurbanne Cedex

RÉSUMÉ

L'étude de variables physico-chimiques mesurées dans un lac d'altitude oligotrophe, le Brévent (Haute-Savoie), sur un cycle annuel constitue une partie de la thèse de CHACORNAC (1986). La présence d'une forte couche de glace pendant huit mois se traduit au niveau des variables par une évolution spatio-temporelle complexe. L'application de l'analyse en composantes principales à trois modes permet de mettre en évidence de façon synthétique les phénomènes interactifs contribuant pour une large part à cette complexité. Les résultats obtenus nous incitent à proposer un emploi plus intensif en Ecologie.

Mots-clés : *analyse en composantes principales à trois modes, variables physico-chimiques, écosystème lacustre d'altitude, structures spatio-temporelles, applications en Ecologie.*

ABSTRACT

Physical and chemical parameters were measured by CHACORNAC (1986) in an oligotrophic high-mountain lake, the Brévent lake (Haute-Savoie), from 1983 to 1984. The presence of a thick ice-cover during eight months involves a complex spatio-temporal evolution of the parameters. The application of three-mode principal component analysis brings to the fore the interaction between the spatial and temporal patterns of these parameters. The results obtained on this particular case allow to discuss a more intensive use in ecological research.

Key-words : *three-mode principal component analysis, physical and chemical parameters, high-mountain lake ecosystem, spatial and temporal patterns, use in Ecology.*

Introduction

Le besoin d'une meilleure compréhension du fonctionnement physico-chimique global des écosystèmes lacustres d'altitude conduit à la mise en œuvre d'études que l'on peut classer selon deux types d'approches. Une première approche, monosystémique, consiste à suivre sur un cycle, annuel en général, la

physico-chimie d'un seul lac d'altitude. Une seconde approche, plurisystémique, considère un ensemble d'écosystèmes lacustres d'altitude afin d'établir une typologie. La thèse de J.M. CHACORNAC (1986) appréhende le problème de façon progressive en adoptant consécutivement les deux points de vue. Dans cet article, nous ne nous intéresserons qu'à la première partie de la démarche.

Les données recueillies constituent un cube du type espace x variables x temps que l'auteur choisit d'arranger classiquement sous la forme d'une matrice du type (espace x temps) x variables pour appliquer une ACP normée. La représentation graphique du plan factoriel des relevés sur les deux premières composantes conduit l'auteur à distinguer six périodes successives partitionnant le cycle annuel. Il justifie cette subdivision par le tracé d'ellipses regroupant les relevés associés à chacune des six périodes. Le chevauchement très marqué de ces ellipses sur les deux composantes rend peu probant ce découpage saisonnier.

Cette inadaptation à restituer de l'information n'étonne pas dans la mesure où elle révèle simplement que la trop grande généralité de la méthode ne convient pas dès lors que le niveau de complexité des mécanismes structurants devient trop élevé. La mise en relation de ce décalage avec le développement de méthodes d'analyse de cubes de données par les statisticiens a conduit récemment des biométriciens à présenter, en hydrobiologie et en limnologie notamment, des propositions méthodologiques plus adaptées (mise en œuvre de l'analyse triadique partielle dans THIOULOUSE & CHESSEL, 1987 et DOLEDEC, 1988 pour l'étude de cours d'eau ; CENTOFANTI *et al.*, 1989 pour l'étude d'un lac réservoir).

Notre propos s'inscrit dans cette logique en cherchant à montrer comment une méthode statistique originale en écologie, l'ACP 3-modes ou analyse triadique complète (évoquée comme alternative dans CENTOFANTI *et al. (op. cit.)* et introduite dans KROONENBERG, 1989), constitue un outil très adapté pour analyser la covariation spatio-temporelle de variables physico-chimiques, descriptives du fonctionnement d'un écosystème particulier, dès lors qu'elle explique une part importante de la variabilité des mesures.

Présentation du lac

Le lac du Brévent, situé à une altitude de 2 127 m. dans le département de Haute-Savoie, se caractérise, au niveau morphométrique, par une superficie modeste (2,95 ha) d'où un périmètre (0.9 km) et un volume ($2.5.10^5$ m^3) en rapport, malgré une profondeur maximale élevée (20 m) pour une profondeur moyenne importante (8.6 m). Les apports hydrologiques au lac sont essentiellement concentrés sur la période de dégel compte-tenu de la faible et ponctuelle (mois de Juin et Juillet) alimentation du lac par ses émissaires et tributaires principaux. La masse d'eau peut être subdivisée en 3 couches superposées : la frange littorale de 0 à -0.5 m ; la zone littorale de -0.5 m à -10 m et la zone profonde de -10 m à -20 m. La caractéristique climatique majeure d'un tel lac de haute montagne consiste en la persistance d'une épaisse couche de glace (atteignant plus de 2,3 m au plus fort de l'hiver) pendant une grande partie de l'année (environ 250 jours, de mi-novembre à mi-juillet). La présence d'une telle couche a pour principale conséquence de plonger le lac dans une obscurité presque complète inhibant toute activité photosynthétique.

Traitement des données

En fin d'article (paragraphe intitulé **Annexe**), sont consignés un exposé mathématique de l'ACP 3-modes et de quelques unes de ses propriétés (partie **I Méthodologie**) ainsi qu'une présentation de l'algorithme mis en œuvre (partie **II Algorithme**). Cette double explicitation est motivée par l'existence de présentations théoriques antérieures qui bien que multiples sont essentiellement de langue anglaise (KROONENBERG & DE LEEUW, 1980; KROONENBERG, 1983b; van der KLOOT & KROONENBERG, 1985; TEN BERGE *et al.*, 1987).

L'étude physico-chimique du lac du Brévent a conduit au plan d'observation suivant : 14 variables ont été mesurées à 8 profondeurs situées à la verticale du point le plus profond («point milieu») pour 16 dates échelonnées entre juin 1983 et juillet 1984. L'ensemble des relevés effectués constitue un cube de données du type profondeurs (premier mode) x variables (second mode) x dates (troisième mode).

Une transformation classique appliquée à des variables mésologiques consiste à normaliser les variables afin de les rendre comparables (homogénéisation des variances). Le problème qui se pose alors est de se donner une référence pour réaliser une telle opération : est-il préférable de normaliser pour chacune des dates ou pour l'ensemble ? Dans le premier cas, on cherche à uniformiser les dates en supprimant la variabilité d'une date à l'autre alors même qu'elle devrait être expliquée. Pour l'étude en question, outre les considérations d'ordre méthodologique, la spécificité des données recueillies nous amène à opter pour le second cas. En effet, l'absence de Magnésium en novembre 1983, quelle que soit la profondeur, et la constatation du même phénomène pour l'Ammoniaque en décembre 1983 et les Phosphates en novembre et décembre 1983, rendent nulles les variances intra-date correspondantes.

La première étape de l'ACP 3-modes, dite d'initialisation (voir la partie **II Algorithme** de l'annexe), consiste en la réalisation d'une ACP classique sur chacune des trois matrices symétriques générées pour chaque mode, à savoir la matrice des corrélations entre variables et deux matrices de covariations sans signification précise entre profondeurs pour l'une et entre dates pour l'autre. L'examen des valeurs propres obtenues à l'issue de chacune des trois diagonalisations nous conduit à ne conserver que 2 composantes pour le premier mode (82% de la variabilité totale), 4 composantes pour le second mode (64,5% de la variabilité totale) et 4 composantes pour le troisième mode (63% de la variabilité totale). Il s'ensuit que chaque donnée expérimentale, après transformation, est reconstituée par la somme de 32 ($2 \times 4 \times 4$) termes.

La seconde étape de l'ACP 3-modes, dite d'itération (voir la partie **II Algorithme** de l'annexe), consiste à rendre identique, pour chaque mode, la variabilité prise en compte par les composantes conservées. Cette variabilité représente 52.7% de la variabilité totale ce qui reste convenable. Consécutivement aux trois ensembles de composantes, la méthode génère une matrice noyau qui va permettre de mesurer l'interaction entre les composantes spécifiques à chacun des trois modes. A chaque combinaison (triplet) associant trois composantes, une par mode, correspond une valeur de la matrice noyau qui indique dans quelle proportion la structure ternaire décrite par cette combinaison participe à la variabilité globale

des données (voir la partie **I Méthodologie** de l'annexe pour les développements théoriques).

A L'examen du pourcentage de variabilité pris en compte par le modèle pour les éléments de chaque mode (fig.1), trois commentaires peuvent être faits :

– Pour le mode spatial (fig.1A), les profondeurs correspondant aux eaux proches de la surface (-2.5 m) ou du fond (- 19 m) caractérisées par les plus fortes variabilités sont les mieux reconstituées (deux meilleurs pourcentages)

– Pour le mode des variables (fig. 1B), les orthophosphates, le pH, la température et l'oxygène dissous sont les variables les mieux reconstituées ce qui signifie que leur covariation spatio-temporelle structure l'essentiel du jeu de données

– Pour le mode temporel (fig. 1C), les dates de Juin et Juillet 1983 caractérisées par les plus fortes variabilités sont, là encore, les mieux reconstituées (trois meilleurs pourcentages) alors que parallèlement des dates à forte variabilité (Août 1983) le sont mal et des dates de variabilité moyenne ou faible (mi-Octobre 1983 ; Novembre 1983) le sont assez bien.

La signification des composantes spatiales synthétiques est évidente lorsqu'on procède à une représentation graphique (fig.2A) : la première correspond à une homogénéisation des profondeurs (P1) alors que la seconde exprime une stratification régulière des profondeurs de la surface vers le fond (P2). Les traits dominants de l'évolution saisonnière de chaque variable bien prise en compte par le modèle se caractérisent donc par la réalisation d'un de ces deux cas de figure.

A l'opposé, les composantes synthétiques associées aux deux autres modes ne semblent pas pouvoir s'analyser indépendamment les unes des autres et nécessitent le dépouillement de la matrice noyau (fig.2B et 2C).

L'arrangement de la matrice noyau consistant à mettre en lignes les composantes synthétiques associées aux profondeurs et en colonnes celles associées aux deux autres modes permet de mettre en évidence deux phénomènes majeurs. D'une part, on constate que lorsque l'on a une homogénéisation des profondeurs (P1), les triplets pour lesquels l'interaction est la plus forte couplent la composante temporelle synthétique la plus structurante avec la composante physico-chimique synthétique la plus structurante et cette association entre composantes de même niveau hiérarchique se répète pour les autres niveaux (triplets 2, 4, 5 et 7 de la figure 3). D'autre part, dans le cas d'une stratification verticale régulière des profondeurs (P2), seules les deux premières composantes physico-chimiques synthétiques et les deux premières composantes temporelles synthétiques sont concernées (triplets 1, 3 et 6 de la figure 3) .

La synthèse des résultats de la matrice noyau nous conduit à définir pour les variables quatre types de covariation sur un cycle annuel.

• 1er type associé à la composante synthétique V1

Cette composante oppose l'oxygène dissous (variable la mieux prise en compte) aux orthophosphates, silicates ainsi qu'à l'ammoniaque, l'alcalinité et le calcium. Elle est la composante associée aux variables qui intervient dans les structures ternaires les plus fortes (triplets 1, 2 et 3 de la figure 3). Les deux

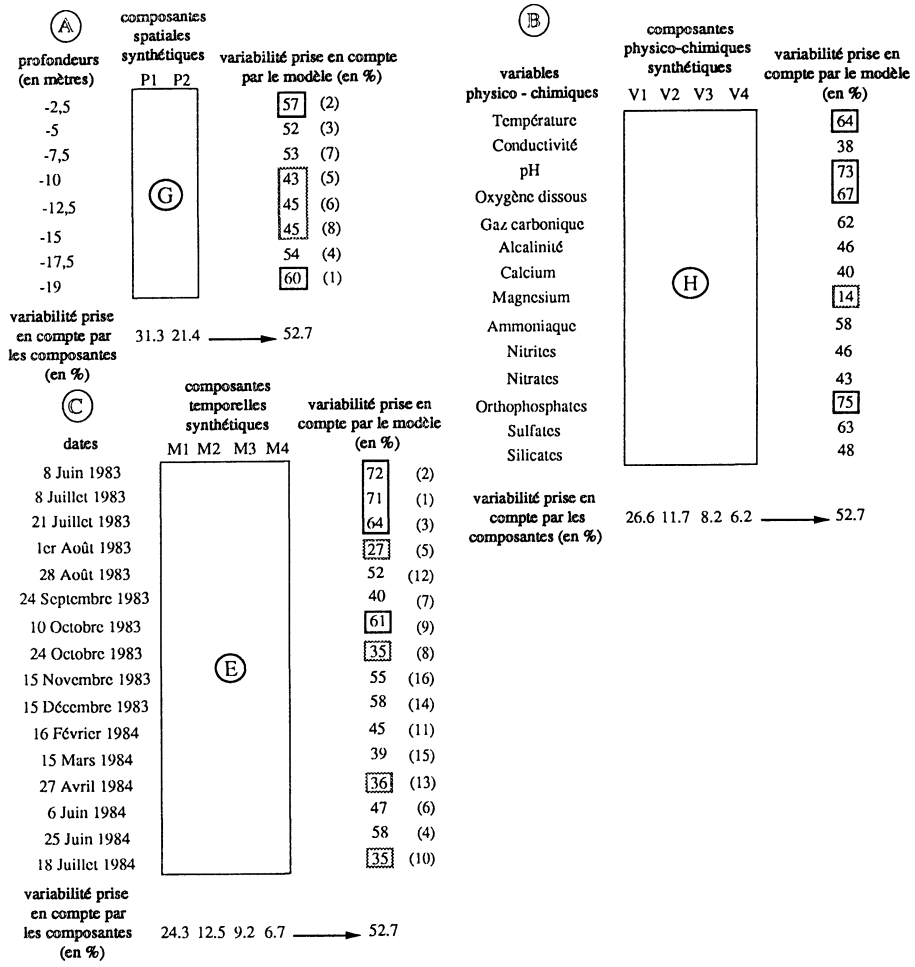

FIGURE 1

A - Qualité de la reconstitution de la variabilité de chaque profondeur par les deux composantes synthétiques conservées (les valeurs extrêmes sont encadrées). Le classement des profondeurs dans la hiérarchie des variabilités décroissantes est indiqué entre parenthèses. B - Qualité de la reconstitution de la variabilité de chaque variable par les quatre composantes synthétiques conservées (les valeurs extrêmes sont encadrées). Le classement des variables dans la hiérarchie des variabilités décroissantes n'est pas indiqué dans la mesure où la variabilité de chaque variable a été ramenée à l'unité. C - Qualité de la reconstitution de la variabilité de chaque date par les quatre composantes synthétiques conservées (les valeurs extrêmes sont encadrées). Le classement des dates dans la hiérarchie des variabilités décroissantes est indiqué entre parenthèses.

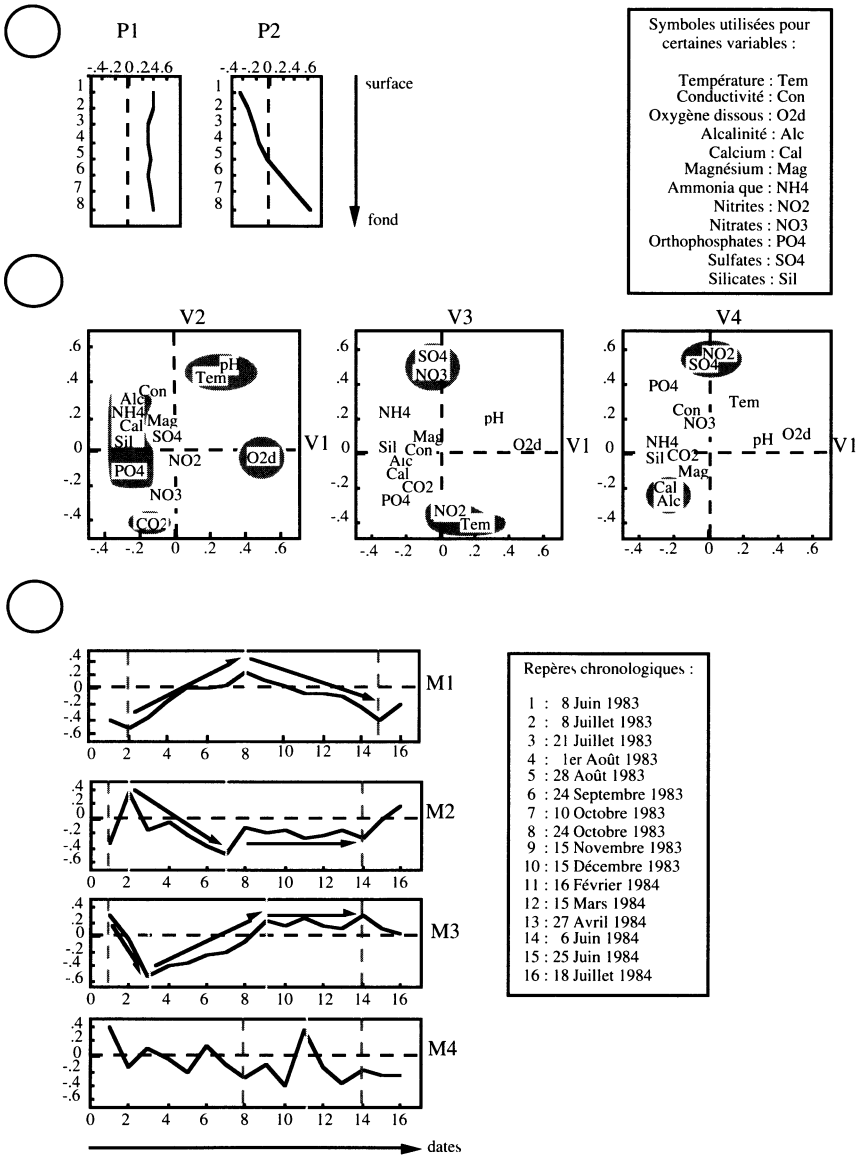

FIGURE 2

A - Représentation graphique des deux composantes spatiales synthétiques retenues (P1 et P2). B - Représentation graphique des quatre composantes physico-chimiques synthétiques retenues (V1, V2, V3 et V4). C - Représentation graphique des quatre composantes temporelles synthétiques retenues (M1, M2, M3 et M4).

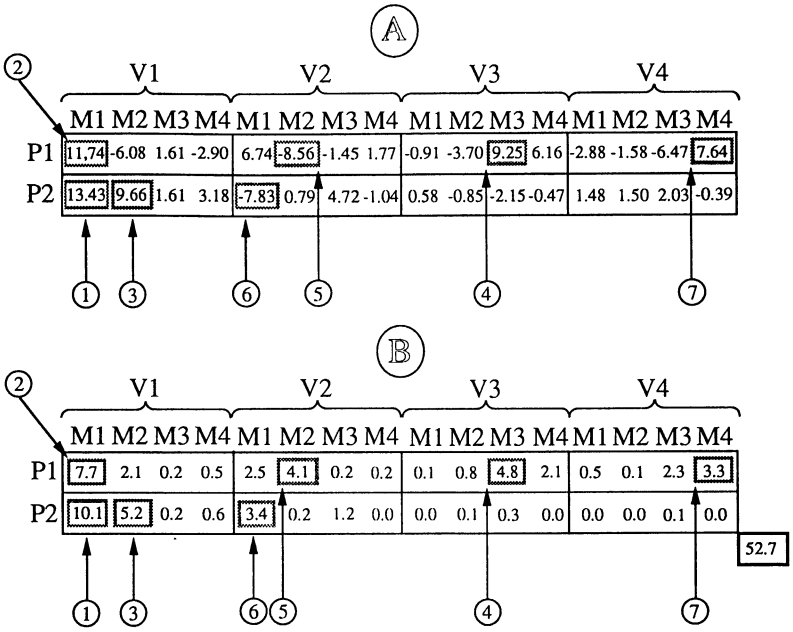

FIGURE 3

A - Matrice noyau : les valeurs absolues les plus élevées sont encadrées et leur classement dans la hiérarchie des valeurs décroissantes est indiqué. B - Matrice des carrés des valeurs de la matrice noyau divisés par la somme totale des carrés des valeurs du cube initial après transformation (en sommant ces valeurs on retrouve le pourcentage de la variabilité totale prise en compte par l'ACP 3-modes).

valeurs les plus élevées sont associées aux triplets **(P2, V1, M1)** et **(P1, V1, M1)** et sont toutes deux positives. De plus, les dates les mieux prises en compte par la composante M1 ont des coordonnées négatives (dates 1, 2, 3 et 15). Pour respecter le signe de l'interaction symbolisée par le triplet **(P2, V1, M1)**, on doit associer pour les dates en question l'oxygène dissous à un profil vertical de type - P2 et les cinq autres variables opposées sur V1 avec un profil de type P2. Le retour aux données transformées montre bien un gradient surface-fond de diminution (respectivement d'augmentation) des teneurs en oxygène dissous (respectivement des autres variables) très net en Juin et Juillet 1983 ainsi qu'en Juin 1984. Pour respecter le signe de l'interaction symbolisée par le triplet **(P1, V1, M1)** et pour les dates de coordonnée positive sur la composante M1 (dates 8 et 9), il faut associer l'oxygène dissous avec un profil vertical de type P1 et les cinq autres variables avec un profil de type -P1. L'examen des données transformées montre qu'en effet fin octobre et mi-novembre 1983 la distribution verticale des concentrations en oxygène dissous (respectivement des cinq variables) correspond à une uniformisation à une valeur supérieure (respectivement inférieure) à la moyenne annuelle. L'interaction symbolisée par le triplet **(P2, V1, M2)** est positive et les

composantes M1 et M2 ont en commun de mettre en évidence les dates 1 et 2 donc ce n'est pas par rapport à ces dates que l'on doit chercher une différenciation entre **(P2, V1, M1)** et **(P2, V1, M2)**. Par contre, c'est pour la date 7 de plus forte coordonnée (négative) sur la composante M2 que l'on doit associer un profil de type -P2 à l'oxygène dissous et un profil de type P2 aux cinq autres variables. Or, à cette date, on constate bien une diminution (respectivement une augmentation) des teneurs en oxygène dissous (respectivement des cinq variables) de la surface vers le fond mais selon un gradient irrégulier, d'où une interaction moins forte en valeur absolue que celle symbolisée par le triplet **(P2, V1, M1)**.

• 2ème type associé à la composante synthétique V3

Cette composante distingue les sulfates et nitrates (plus fortes coordonnées positives) des nitrites et de la température (plus fortes coordonnées négatives). Elle intervient essentiellement par la structure ternaire symbolisée par le triplet numéro 4 sur la figure 3. La valeur positive associée au triplet **(P1, V3, M3)** et les coordonnées toutes positives des profondeurs sur P1 impliquent, en théorie, d'une part la combinaison des sulfates et nitrates avec les dates 1 et 14; d'autre part la combinaison des nitrites et de la température avec la date 3. Or, la plus forte coordonnée en valeur absolue sur M3 est celle de la date 3, c'est pourquoi c'est la deuxième combinaison qui doit être privilégiée. L'examen des données transformées montre bien une parfaite homogénéisation de la température fin Juillet 1983 (brève isothermie estivale).

• 3ème type associé à la composante synthétique V2

Cette composante sépare le pH (plus forte coordonnée positive) et la température (dans une moindre mesure puisqu'elle est bien prise en compte par V3) du gaz carbonique (plus forte coordonnée négative). Les structures ternaires les plus fortes auxquelles elle participe correspondent aux triplets numérotés 5 et 6 sur la figure 3. Analysons chacun d'entre eux. La valeur négative caractérisant le triplet **(P2, V2, M1)** traduit pour les dates 2 et 15 (coordonnées négatives) une distribution des concentrations selon un gradient qui oppose le pH au gaz carbonique. On constate effectivement que les concentrations ou valeurs du pH (respectivement du gaz carbonique) sont moindres (repectivement plus élevées) aux faibles profondeurs début Juillet 1983 alors que le phénomène est symétrique fin Juin 1984. La valeur négative associée au triplet **(P1, V2, M2)** et les coordonnées toutes positives des profondeurs sur P1 impliquent, en théorie, d'une part l'association du pH avec les dates 1 et 7 (plus fortes coordonnées négatives) et d'autre part, l'association du gaz carbonique avec la date 2 (plus forte coordonnée positive). Or, on constate une homogénéisation de la valeur du pH à une valeur supérieure à la moyenne début Juin 1983 et mi-octobre 1983 et une homogénéisation du même type pour le gaz carbonique début Juillet 1983.

• 4ème type associé à la composante synthétique V4

Cette composante prend en compte les nitrites et sulfates (plus fortes coordonnées positives). Elle intervient essentiellement par la structure ternaire symbolisée par le triplet numéro 7 de la figure 3. La valeur positive associée au triplet **(P1, V4, M4)** et les coordonnées positives sur P1 impliquent l'association des nitrites et sulfates avec les dates 1 et 11 (plus fortes coordonnées positives sur M4)

pour respecter le signe. Or, début Juin 1983 et en Février 1984, on constate effectivement une uniformisation très ponctuelle des concentrations des deux variables sur l'ensemble des profondeurs à un niveau supérieur à la moyenne.

La figure 4 résume l'ensemble des phénomènes majeurs pris en compte par le modèle.

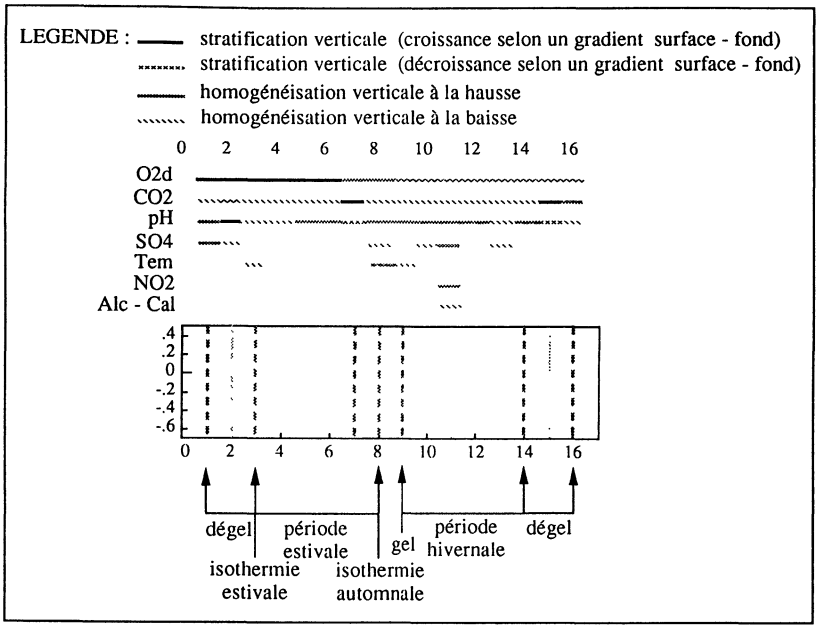

FIGURE 4
Synthèse générale.

Conclusions et perspectives

Nous conclurons par des considérations d'ordre pratique et méthodologique.

1) Initialement introduite sous la dénomination anglaise de "three-mode factor analysis" et formulée dans le contexte des sciences du comportement par TUCKER (1963), qui a developpé par la suite la description mathématique et la mise en œuvre informatique (TUCKER, 1964 et 1966), l'ACP 3-modes appartient à la famille d'analyses par estimation de modèle sans prise en compte de facteurs externes (endomodélisation). Le terme d'ACP 3-modes est à la fois plus explicite et plus impropre que l'appellation première : il précise bien la filiation d'une telle approche qui se présente dans ses fondements comme une tentative de généralisation de l'ACP classique ("two-mode factor analysis"), ou plutôt de la décomposition en valeurs singulières, appliquée sur les tableaux bidimensionnels mais au niveau mathématique le passage de deux à trois dimensions ne correspond pas à une simple extension (LEVIN, 1965). En effet, dans le cas bidimensionel

la décomposition en valeurs singulières est une décomposition canonique dans la mesure où les composantes sont communes aux deux modes, ce qui n'est plus le cas lorsque l'on passe à trois modes. Cependant, certains auteurs ont tenté une généralisation de la méthode de TUCKER à l'analyse de tableaux à $n(n \geq 2)$ modes (proposée par KAPTEYN *et al.*, 1984 et développée par POLIT, 1986; d'AUBIGNY & POLIT, 1989). Cette généralisation repose sur l'idée qu'une matrice de données Z peut être interprétée comme la représentation matricielle d'un tenseur z élément de l'espace des formes n-linéaires défini sur le produit cartésien des espaces de représentation associé à chacun des n modes. A chacun des espaces de représentation est associé une métrique (identité pour les trois modes dans le modèle TUCKER3) qui le munit d'un produit scalaire. L'objectif de l'ACP n-modes devient alors la recherche d'un sous-espace du produit cartésien tel que la projection de z sur ce sous-espace soit optimal. Dans cette approche, l'extension de deux à trois modes se fait directement.

2) Toutes les méthodes initialement développées dans le cadre bidimensionnel ne peuvent être employées sur des cubes de données sans une réécriture matricielle nécessitant la combinaison de deux modes sur les lignes ou les colonnes (ACP, AFC,...). Cette contrainte entraine par essence une asymétrie dans l'importance accordée à chacun des trois modes : un mode élémentaire est considéré sur le même pied d'égalité que le mode généré par le produit cartésien des deux autres modes élémentaires ("combination mode"). Il s'ensuit que les structures mises en évidence pour ce type de mode privilégient les structures du mode élémentaire dominant et masquent ou atténuent les structures propres au mode élémentaire moins influent. Ainsi, une partie de l'information contenue dans les données se trouve tronquée par l'organisation de celles-ci. L'ACP 3-modes ne se dispense pas d'un tel arrangement des données mais l'analyse simultanée des trois matrices bidimensionnelles associées a chacune des arêtes du cube de données remédie à l'inconvénient évoqué en le généralisant aux trois modes : la systématisation du procédé rend caduque l'asymétrie de son effet.

Le modèle présente un *a priori* en rejetant notamment les effets principaux (propres aux éléments de chaque mode d'observation) et les interactions de premier ordre (entre deux éléments appartenant à deux modes d'observation différents) que l'on pourrait obtenir en appliquant une analyse de variance à trois facteurs contrôlés ("three-way ANOVA"). Cependant, l'exemple proposé permet d'affiner ce postulat théorique dans la mesure où la composante synthétique P1 (compromis des profondeurs) apparait comme un effet principal propre au mode spatial. En conséquence, une structure ternaire générée par le modèle peut se décomposer en produit d'un effet principal par une interaction de premier ordre qui ne sont donc pas totalement éliminés par le modèle.

3) Si le recours à l'ACP 3-modes obéit à un objectif essentiel de condensation de l'information contenue dans les données, diverses aides peuvent néanmoins être employées (cf. KROONENBERG & BROUWER, 1985) pour dépouiller les résultats (étude de la qualité de l'ajustement du modèle aux données pour chaque élément des trois modes, analyse de l'écart au modèle (résidu) pour chaque donnée, analyse de variance sur les carrés des résidus pour détecter éventuellement la persistance de l'influence d'un mode) ou affiner l'interprétation (étude des liens entre les éléments du mode espace et du mode des variables par leur représentation si-

multanée ("joint plot") pour chaque composante temporelle synthétique conservée, calcul des covariations ("latent covariation matrix") entre tous les couples (composante physico-chimique synthétique - composante temporelle synthétique) sur les toutes les composantes spatiales synthétiques conservées). Ces aides, tant graphiques que numériques, sont disponibles mais n'ont pas été exploitées pour faciliter l'interprétation dans la mesure où les résultats exposés expriment de façon suffisamment claire les structures recherchées.

4) Le peu d'attention dont l'ACP 3-modes a fait l'objet de la part des statisticiens français (non testée dans ESCOUFIER et al, 1985, par exemple), interrompu depuis peu par des études comparatives avec d'autres méthodes dites multitableaux (LAVIT, 1988; CARLIER *et al.*, 1989), contraste avec la multiplicité tant des développements théoriques (en particulier les modèles PARAFAC de HARSHMAN, 1970 et 1972) que des domaines d'application (revue dans KROONENBERG, 1983a), exception faite des sciences de la nature (citons tout de même HOHN & FRIBERG, 1979 en Géologie; BASFORD *et al.*, 1990 en Agronomie et KROONENBERG en Hydrobiologie, 1989), dans les pays anglo-saxons. Réciproquement, peu d'auteurs de ces pays, ou de pays tiers, ont cherché à positionner l'ACP 3-modes et ses variantes par rapport à d'autres méthodes afin de surmonter le cloisonnement existant entre l'école française et l'école anglo-saxonne. Cependant, des comparaisons récentes ont été effectuées par le biais d'applications conjointes sur les mêmes données (KROONENBERG en Auxologie, 1987; BOVE & DI CIACCIO en Economie, 1989) ou dans le cadre de tentatives de classification et hiérarchisation de diverses méthodes multitableaux (dont STATIS et l'AFM) à partir d'un critère commun (la fonction de perte dans KIERS, 1988a et b).

A l'écart des champs d'investigation, l'Ecologie est caractérisée par le retard accumulé au niveau de la pratique exploratoire comme l'atteste l'absence de citation dans des synthèses récentes sur les développements numériques (ouvrages de DIGBY & KEMPTON, 1987; JONGMAN *et al.*, 1987). Pourtant, les potentialités de l'ACP 3-modes laissent à penser que son usage en Ecologie devrait devenir familier et ce, d'autant plus, que les perspectives paraissent prometteuses. En effet, des objectifs concrets de nature apparemment différente expriment les formes multiples que revêt un unique objectif théorique : la mise en évidence des covariations spatio-temporelles majeures des descripteurs mesurés (espèces animales ou végétales d'un peuplement; stades de développement d'une population; variables quantitatives, semi-quantitatives ou qualitatives descriptives d'un milieu;...).

5) Reconnaître que dans les domaines d'application courante les modèles de ce type sont «clairement des aides puissantes pour la recherche empirique» (HATTIE *et al.*, 1984) signifie notamment que l'ACP 3-modes se présente comme un outil exploratoire dont la plasticité permet de préconiser un emploi plus systématique qui n'implique pas, pour autant, la généralisation de sa validité. En effet, comme l'indique un spécialiste de l'ACP 3-modes en guise de bilan provisoire de son activité, «une observation intéressante tirée à l'issue d'un grand nombre d'applications est que la solution par les composantes principales semble déjà ne produire des structures si simples que si elles sont présentes dans les données et que si elles sont compatibles avec le modèle employé» (KROONENBERG, 1984); ce que SNYDER *et al.* (1984) expriment plus succinctement en écrivant que «la forme de l'information résulte directement du choix du filtre». Ces réflexions

indiquent bien que «l'on ne laisse pas les données parler pour elles-mêmes» (voir GOULD (1981) pour un avis opposé) et qu'une réalité complexe ne peut jamais être abordée qu'à partir d'hypothèses implicites et/ou explicites dont on cherche une confirmation empirique en recourant à des méthodes qui «sont simplement des véhicules pour l'exercice de la preuve» (RYCHLAK, 1981 in HATTIE *et al.*, *op. cit.*).

ANNEXE

I. Méthodologie

Soit un cube de données expérimentales Z caractérisé par trois modes d'observation ("observational modes") définis comme des «ensembles d'indices par lesquels les données peuvent être classées» (TUCKER, 1964). Le postulat de l'ACP 3-modes, dérivé de celui de l'ACP classique, consiste à supposer que les modes d'observation peuvent être synthétisés par des modes plus fondamentaux qui s'en déduisent ("intrinsic modes" (TUCKER, 1963); "derivational modes" (TUCKER, 1966)) : chaque élément d'un mode réduit exprime une structure particulière existant entre tout ou partie des éléments du mode d'observation associé.

Chaque valeur $z_{ijk} (i = 1,1; j = 1,m$ et $k = 1,n)$ du cube Z peut être reconstituée selon le modèle multiplicatif TUCKER3 qui se formalise comme suit :

$$z_{ijk} = \sum_{p=1}^{s} \sum_{q=1}^{t} \sum_{r=1}^{u} g_{ip} h_{jq} e_{kr} c_{pqr} \tag{1}$$

où les éléments $g_{ip} (p = 1, s)$, $h_{jq} (q = 1, t)$ et $e_{kr} (r = 1, u)$ sont respectivement les composantes des vecteurs orthonormés $\vec{g}_i$, $\vec{h}_j$ et $\vec{e}_k$ ce qui se traduit par les égalités suivantes :

$$\sum_i g_{ip} g_{ip'} = 1 \text{ si } p = p' \text{ et } 0 \text{ sinon} \tag{2}$$

$$\sum_j h_{jp} h_{jp'} = 1 \text{ si } q = q' \text{ et } 0 \text{ sinon} \tag{3}$$

$$\sum_k e_{kr} e_{kr'} = 1 \text{ si } r = r' \text{ et } 0 \text{ sinon} \tag{4}$$

De la formule (1) on peut tirer l'égalité suivante :

$$\sum_i \sum_j \sum_k z_{ijk}^2 =$$

$$\sum_i \sum_j \sum_k \left(\sum_{p=1}^{s} \sum_{p'=1}^{s} \sum_{q=1}^{t} \sum_{q'=1}^{t} \sum_{r=1}^{u} \sum_{r'=1}^{u} g_{ip} g_{ip'} h_{jq} h_{jq'} e_{kr} e_{kr'} c_{pqr} c_{p'q'r'} \right)$$

qui, en utilisant la propriété d'orthonormalité des vecteurs $\vec{g}_i$, $\vec{h}_j$ et $\vec{e}_k$ explicitée en (2), (3) et (4), se réduit à :

$$\sum_i \sum_j \sum_k z_{ijk}^2 = \sum_p \sum_q \sum_r c_{pqr}^2 \tag{5}$$

Le cube C, de terme général c_{pqr}, appelé matrice noyau ("core matrix"), est une réécriture du cube de données initial auquel il est lié par l'égalité fondamentale (5) qui indique que la somme des carrés des valeurs du cube Z est préservée. Chaque élément c_{pqr} est associé à une combinaison, appelée triplet et notée (p, q, r), de trois composantes (une par mode), et mesure la part relative de la structure ternaire symbolisée par le triplet dans la reconstitution des valeurs observées. Ainsi, l'égalité (1) exprime que toute valeur observée associée au triplet (i, j, k) est estimée par la superposition des structures ternaires dominantes, chacune participant de façon différenciée (les produits du type $g_{ip}h_{jq}e_{kr}$ sont alors assimilables à des pondérations).

Les matrices $G = (g_{ip})_{i=1,l;p=1,s}$, $H = (h_{jq})_{j=1,m;q=1,t}$ et $E = (e_{kr})_{k=1,n;r=1,u}$ sont les matrices des vecteurs propres orthonormés correspondant respectivement :

– aux s valeurs propres non nulles de la matrice $P = \{p_{ii'}\}$ avec $p_{ii'} = \sum_j \sum_k z_{ijk} z_{i'jk}$ soit $P = (Z_I^{JK})(Z_I^{JK})'$ où Z_I^{JK} est la matrice de terme général

$$z(i,(j,k)) = z_{ijk} \tag{6a}$$

– aux t valeurs propres non nulles de la matrice $Q = \{q_{jj'}\}$ avec $q_{jj'} = \sum_i \sum_k z_{ijk} z_{ij'k}$ soit $Q = (Z_J^{IK})(Z_J^{IK})'$ où Z_J^{IK} est la matrice de terme général

$$z(j,(i,k)) = z_{ijk} \tag{6b}$$

– aux u valeurs propres non nulles de la matrice $R = \{r_{kk'}\}$ avec $r_{kk'} = \sum_i \sum_j z_{ijk} z_{ijk'}$, soit $R = (Z_K^{IJ})(Z_K^{IJ})$ où Z_K^{IJ} est la matrice de terme général $z(k,(i,j)) = z_{ijk}$ (6c)

La relation (1) s'écrit en notation matricielle :

$$Z = GC(H' \otimes E') \tag{7}$$

où $\otimes$ symbolise le produit tensoriel.

L'égalité (7) n'est valide que dans le cas où l'on conserve toutes les composantes associées à chacun des 3 modes ce qui n'est pas réaliste. L'objectif du modèle consiste donc en la minimisation de la fonction de perte d'information dite des moindres carrés ("mean-squared loss function") notée f et définie comme suit :

$$f(G,H,E,C) = \| Z - \widetilde{Z} \|^2 = \| Z - GC(H' \otimes E') \|^2 \tag{8}$$

où $\| \quad \|$ désigne la norme euclidienne

Elle exprime simplement une mesure de l'erreur commise dans la reconstitution de la matrice Z, notée $\widetilde{Z}$, lorsque l'on ne conserve qu'un nombre réduit de composantes associées à chaque mode (composantes synthétiques). En fixant G, H et E, la solution C de ce problème d'optimisation est obtenue en utilisant une version simplifiée d'un lemme de PENROSE (1955) et s'écrit :

$$C = G'Z(H \otimes E) \tag{9}$$

Soient s_c, t_c et u_c le nombre de composantes conservées respectivement pour le premier, second et troisième mode et p^*, q^* et r^* les indices associés. Le modèle s'écrit alors :

$$\widetilde{z}_{ijk} = \sum_{p^*=1}^{s_c} \sum_{q^*=1}^{l_c} \sum_{r^*=1}^{u_c} g_{ip^*} h_{jq^*} e_{kr^*} c_{p^*q^*r^*} \tag{10}$$

Posons $\forall j = 1, m$ $z^b_{.j.} = \frac{1}{ln} \sum_i \sum_k z^b_{ijk}$ et $s^2_{.j.} = \frac{1}{ln} \sum_i \sum_k (z^b_{ijk} - z^b_{.j.})^2$ où les z^b_{ijk} correspondent aux données brutes.

Compte-tenu de la transformation appliquée aux données traitées dans l'article (les z_{ijk} intervenant dans les formules (1), (5) à (9) sont les valeurs initiales centrées et réduites), la reconstitution de chaque donnée brute se formalise par :

$$z^b_{ijk} = s_{.j.} \widetilde{z}_{ijk} + z^b_{.j.} \tag{11}$$

Les propriétés majeures du modèle sont les suivantes :

1) La contribution d'un triplet (p_0, q_0, r_0) à la variabilité globale du tableau Z est mesurée en % par :

$$100 \frac{c^2_{p_0 q_0 r_0}}{\sum_i \sum_j \sum_k z^2_{ijk}} = 100 \frac{c^2_{p_0 q_0 r_0}}{\sum_p \sum_q \sum_r c^2_{pqr}} = 100 \frac{\left(\sum_i \sum_j \sum_k g_{ip_0} h_{jq_0} e_{kr_0} z_{ijk} \right)^2}{\sum_i \sum_j \sum_k z^2_{ijk}}$$

2) La contribution d'un triplet (p_0, q_0, r_0) à la variabilité prise en compte par le modèle est mesurée en % par :

$$100 \frac{c^2_{p_0 q_0 r_0}}{\sum_{p^*} \sum_{q^*} \sum_{r^*} c^2_{p^*q^*r^*}}$$

3) La contribution d'une composante, disons p_0 associée au premier mode, à la variabilité globale est mesurée en % par :

$$100\frac{\sum_{q^*}\sum_{r^*} c^2_{p_0q^*r^*}}{\sum_i\sum_j\sum_k z^2_{ijk}} = 100\,\frac{\sum_{q^*}\sum_{r^*} c^2_{p_0q^*r^*}}{\sum_p\sum_q\sum_r c^2_{pqr}} = 100\,\frac{\sum_{q^*}\sum_{r^*}\left(\sum_i\sum_j\sum_k g_{ip_0}h_{jq^*}e_{kr^*}z_{ijk}\right)^2}{\sum_i\sum_j\sum_k z^2_{ijk}}$$

4) La contribution d'une composante, toujours p_0, à la variabilité prise en compte par le modèle est mesurée en % par :

$$100\frac{\sum_{q^*}\sum_{r^*} c^2_{p_0q^*r^*}}{\sum_{p^*}\sum_{q^*}\sum_{r^*} c^2_{p^*q^*r^*}} = 100\,\frac{\sum_{q^*}\sum_{r^*}\left(\sum_i\sum_j\sum_k g_{ip_0}h_{jq^*}e_{kr^*}z_{ijk}\right)^2}{\sum_{p^*}\sum_{q^*}\sum_{r^*} c^2_{p^*q^*r^*}}$$

5) Soit i_0 un élément du premier mode. Si $\tilde{z}_{i_0jk/p_0}$ indique dans quelle proportion intervient la composante p_0 dans la reconstitution de la valeur z_{i_0jk}, alors la contribution de cette composante à la variabilité de l'élément i_0 est mesurée en % par :

$$100\frac{\sum_j\sum_k \tilde{z}^2_{i_0jk/p_0}}{\sum_j\sum_k z^2_{i_0jk}} = 100\,\frac{\sum_j\sum_k\left(\sum_{q^*}\sum_{r^*} g_{i_0p_0}h_{jq^*}e_{kr^*}c_{p_0q^*r^*}\right)^2}{\sum_j\sum_k z^2_{i_0jk}}$$

$$= 100\,g^2_{i_0p_0}\,\frac{\sum_j\sum_k\left(\sum_{q^*}\sum_{r^*} h_{jq^*}e_{kr^*}c_{p_0q^*r^*}\right)^2}{\sum_j\sum_k z^2_{i_0jk}}$$

$$= 100\,g^2_{i_0p_0}\,\frac{\sum_{q^*}\sum_{r^*} c^2_{p_0q^*r^*}}{\sum_j\sum_k z^2_{i_0jk}}$$

II. ALGORITHME

Il repose sur la méthode d'itération simultanée de BAUER-RUTISHAUSER (RUTISHAUSER, 1969) dont l'emploi répété fonde la technique, dite des moindres carrés alternés, de minimisation de la fonction f. En reportant la formule 9 dans la formule 8 et en développant celle-ci, on aboutit au problème de la maximisation de la fonction $p(G,H,E) = \text{tr } G'Z(HH' \otimes EE')Z'G$ où tr symbolise la trace. La solution consiste en l'obtention simultanée des trois matrices G, H et E de vecteurs propres. La transcription algorithmique ("alternating least squares algorithm" encore appelé TUCKALS3 et exposé dans KROONENBERG & DE LEEUW, *op. cit.*) se compose de deux parties :

• une première étape d'initialisation des matrices de vecteurs propres : les matrices G_0, H_0 et E_0 sont les matrices de vecteurs propres obtenues respectivement par la diagonalisation des matrices P, Q et R (voir formules 6a, 6b et 6c).

• une seconde étape d'itération explicitée pour une itération $a+1$ constituée de 3 pas :

* pas associé au mode 1 : calcul de

$$P_a = Z(H_a H_a' \otimes E_a E_a')Z' \quad \text{où} \quad Z = Z_I^{JK}$$
$$G_{a+1} = P_a G_a (G_a' P_a^2 G_a)^{-1/2}$$

* pas associé au mode 2 : calcul de

$$Q_a = Z(E_a E_a' \otimes G_{a+1} G_{a+1}')Z' \quad \text{où} \quad Z = Z_J^{KI}$$
$$H_{a+1} = Q_a H_a (H_a' Q_a^2 H_a)^{-1/2}$$

* pas associé au mode 3 : calcul de

$$R_a = Z(G_{a+1} G_{a+1}' \otimes H_{a+1} H_{a+1}')Z' \quad \text{où} \quad Z = Z_K^{IJ}$$
$$E_{a+1} = R_a E_a (E_a' R_a^2 E_a)^{-1/2}$$

La convergence de l'algorithme vers une solution optimale (G,H,E) pour laquelle la fonction p atteint un maximum est assurée par le recours à un lemme, dit du «point fixe», décrit et démontré par d'ESOPO (1959). Cependant, p étant définie par une expression polynomiale de degré 6, il n'est pas possible de démontrer que ce maximum est global (KROONENBERG & DE LEEUW, *op. cit.*).

Soient $\lambda_{p^*}(p^* = 1, s_c)$, $\mu_{q^*}(q^* = 1, t_c)$ et $\gamma_{r^*}(r^* = 1, u_c)$ les valeurs propres conservées à l'issue de l'étape d'initialisation alors l'inégalité

$$p(G,H,E) \leq \min\left(\sum_{p^*} \lambda_{p^*}, \sum_{q^*} \mu_{q^*}, \sum_{r^*} \gamma_{r^*}\right)$$

définit la limite supérieure de l'ajustement de la somme des carrés (lemme démontré dans TEN BERGE *et al.*, *op. cit.*) et fournit, le cas échéant, une stratégie d'amélioration progressive de l'ajustement (nouvelle mise en œuvre de l'algorithme en augmentant le nombre de valeurs propres conservées pour un ou deux modes lors de la phase d'initialisation).

Remerciements

L'exposition et l'interprétation des résultats ont largement bénéficié des critiques et recommandations formulées par P.M. KROONENBERG lors d'une présentation orale qu'il m'a été permis d'effectuer dans le cadre de la réunion de travail organisée par C. LAVIT et R. SABATIER au laboratoire de Biométrie de Montpellier le 05-07-1990.

Nous remercions vivement les deux referees pour leurs critiques et suggestions.

Note

Les programmes informatiques mis en œuvre pour l'obtention des résultats ont été écrits en Microsoft Basic. Ils correspondent dans une large mesure à une fragmentation des procédures du programme unique écrit en FORTRAN par KROONENBERG & BROUWER *(op. cit.)*. L'implantation de ces programmes s'est effectuée dans le cadre du logiciel «Analyse de Données ECOlogiques (ADECO)» de D. Chessel, J. Thioulouse, Y. Auda, J.L. Beffy et S. Dolédec (PIREN-Vallées Fluviales, URA C.N.R.S. 367, Université Lyon I).

Références

AUBIGNY (d') G. & POLIT E. (1989). Some optimality properties of the generalization of the TUCKER method to the analysis of n-way tables with specified metrics. In R. Coppi & S. Bolasco (Eds.), *Multiway data analysis*. Amsterdam, Elsevier, 39-52.

BASFORD K.E., KROONENBERG P.M., DELACY I.H. & LAWRENCE P.K. (1990). Multiattribute evaluation of regional cotton variety trials. *Theor. Appl. Genet.*, 79, 225-234.

BOVE G. & DI CIACCIO A. (1989). Comparisons among three factorial methods for analysing three-mode data. In R. Coppi & S. Bolasco (Eds.), *Multiway data analysis*. Amsterdam, Elsevier, 103-113.

CARLIER A., LAVIT C.H., PAGES M., PERNIN M.O. & TURLOT J.C. (1989). A comparative review of methods which handle a set of indexed data tables. In R. Coppi & S. Bolasco (Eds.), *Multiway data analysis*. Amsterdam, Elsevier, 85-102.

CENTOFANTI M., CHESSEL D. & DOLEDEC S. (1989). Stabilité d'une structure spatiale et compromis d'une analyse statistique multitableaux : application à la physico-chimie d'un lac réservoir. *Rev. Sci. Eau*, 2, 71-93.

CHACORNAC J.M. (1986). *Lacs d'altitude : métabolisme oligotrophe et approche typologique des écosystèmes*. Thèse de Doctorat, Univ. Lyon I, 214 pp.

DIGBY P.G.N. & KEMPTON R.A. (1987). *Multivariate analysis of ecological communities*. Chapman & Hall, London, 206 pp.

DOLEDEC S. (1988). Les analyses multitableaux en écologie factorielle. II. Stratification longitudinale de l'Ardèche à partir de descripteurs physico-chimiques. *Acta Œcol., Œcol. Gener.*, 9, 2, 119-135.

ESCOUFIER Y., BERNARD M.C., LAVIT C., FOUCART T., BARRE A., FICHET B., CARLIER A. & LAFAYE J.Y. (1985). Comparaison d'analyse de tableaux à trois dimensions à partir d'un exemple. *Statistique et Analyse des données*, 10, 1, 1-116.

ESOPO (d') D. (1959). A convex programming procedure. Naval Research Logistics Quarterly, 11, 33-42.

GOULD P. (1981). Letting the data speak for themselves. *Annals of the Association of American Geographers*, 71, 166-176.

HARSHMAN, R.A. (1970). Foundations of the PARAFAC procedure : Models and conditions for an "explanatory" multi-mode factor analysis. *UCLA Working Papers in Phonetics*, 16, 1-84.

HARSHMAN, R.A. (1972). PARAFAC2 : Mathematical and technical notes. *UCLA Working Papers in Phonetics*, 22, 31-44(a).

HATTIE J.A., McDONALD R.P., SNYDER C.W. & LAW H.G. (1984). Issues and perspectives in multimode analysis. In H.G. Law, C.W. Snyder, J.A. Hattie and R.P. McDonald (Eds.), *Research Methods for multimode data analysis*, New-York, Praeger, 555-564.

HOHN, M.E. & FRIBERG L.M. (1979). A generalized principal components model in petrology. *Lithos*, 12, 317-324.

JONGMAN R.H.G., TER BRAAK C.J.F. & TONGEREN O.F.R. (1987). *Data analysis in community and landscape ecology*. Pudoc Wageningen, Pays-Bas, 299 pp.

KAPTEYN A., NEUDECKER H. & WANSBEEK T. (1984). An approach to n-mode components analysis. *Psychometrika*, 51, 2, 269-275.

KIERS H.A.L. (1988a). Hierarchical relations between three-way methods. Communication aux "XX$^{\text{èmes}}$ Journées de Statistiques", Grenoble, 30 mai - 2 juin 1988.

KIERS H.A.L. (1988b). Comparison of "Anglo-Saxon" and "French" three-mode methods. *Statistique et Analyse des Données*, 13, 14-32.

KROONENBERG P.M. (1983a). Annotated bibliography of three-mode factor analysis. *British Journal of Mathematical and Statistical Psychology*, 36, 81-113.

KROONENBERG P.M. (1983b). *Three-mode principal component analysis*. DSWO Press, Leiden, 398 pp.

KROONENBERG P.M. (1984). Three-mode principal component analysis : illustrated with an example from attachment theory. In H.G. Law, C.W. Snyder, J.A. Hattie and R.P. McDonald (Eds.), *Research Methods for multimode data analysis*, New-York, Praeger, 64-103.

KROONENBERG P.M. (1987). Multivariate and longitudinal data on growing children. Solutions using a three-mode principal component analysis and some comparison results with other approaches. In J. Janssen, F. Marcotorchino & J.M. Proth (Eds.), *Data Analysis. The ins and outs of solving real problems*, New-York : Plenum, 89-112.

KROONENBERG P.M. (1989). The analysis of multiple tables in factorial ecology. III. Three-mode principal component analysis : "Analyse triadique complète". *Acta Œcol., Œcol. Gener.*, 10, 245-256.

KROONENBERG P.M. & BROUWER P. (1985). User's guide to TUCKALS3. Version 4.0. Department of Education, Leiden University, Pays-Bas.

KROONENBERG P.M. & de LEEUW J. (1980). Principal component analysis of three-mode data by means of alternating least squares algorithms. *Psychometrika*, 45, 69-97.

LAVIT C.H. (1988). *Analyse conjointe de tableaux quantitatifs*. Paris, Masson, 253 pp.

LEVIN J. (1965). Three-mode factor analysis. *Psychological Bulletin*, 64, 6, 442-452.

PENROSE R. (1955). On the best approximate solutions of linear matrix equations. Proceedings of the Cambridge Philosophical Society, 51, 406-413.

POLIT E. (1986). Une n-ACP d'un hypercube de données. Thèse de Doctorat, Univ. des Sciences Sociales de Grenoble.

RYCHLAK J.F. (1981). *A Philosophy of Science for Personality Theory*. 2nd ed. Malabar, Fla. : Krieger.

RUTISHAUSER H. (1969). Computational aspects of F.L. Bauer's simultaneous iteration method. *Numerische Mathematik*, 13, 4-13.

SNYDER C.W., LAW H.G. & HATTIE J.A. (1984). Overview of multimode analytic methods. In H.G. Law, C.W. Snyder, J.A. Hattie and R.P. McDonald (Eds.), *Research Methods for multimode data analysis*, New-York, Praeger, 2-35.

TEN BERGE J.M.F., de LEEUW J. & KROONENBERG P.M. (1987). Some additional results on principal components analysis of three-mode data by means of alternating least squares algorithms. *Psychometrika*, 52, 2, 183-191.

THIOULOUSE J. & CHESSEL D. (1987). Les analyses multitableaux en écologie factorielle. I. De la typologie d'état à la typologie de fonctionnement par l'analyse triadique. *Acta Œcol., Œcol. Gener.*, 8, 4, 463-480.

TUCKER, L.R. (1963). Implications of factor analysis of three-way matrices for measurement of change. In C.W. Harris (Ed.), *Problems in measuring change*, Madison : University of Wisconsin Press, 122-137.

TUCKER, L.R. (1964). The extension of factor analysis to three-dimensional matrices. In H. Gullikson and N. Frederiksen (eds.), *Contributions to mathematical psychology*, New-York : Holt, Rinehart and Winston, 110-119.

TUCKER, L.R. (1966). Some mathematical notes on three-mode factor analysis. *Psychometrika*, 31, 3, 279-311.

van der KLOOT W.A. & KROONENBERG P.M. (1985). External analysis with three-mode principal component models. *Psychometrika*, 50, 4, 479-494.