

MICHEL CHAVANCE

**Choix des variables les plus informationnelles
dans un tableau à triple entrée**

Revue de statistique appliquée, tome 26, n° 1 (1978), p. 35-44

http://www.numdam.org/item?id=RSA_1978__26_1_35_0

© Société française de statistique, 1978, tous droits réservés.

L'accès aux archives de la revue « *Revue de statistique appliquée* » (<http://www.sfds.asso.fr/publicat/rsa.htm>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

CHOIX DES VARIABLES LES PLUS INFORMATIONNELLES DANS UN TABLEAU A TRIPLE ENTRÉE

Michel CHAVANCE

Attaché de Recherche, INSERM U 88
Méthodologie Informatique et Statistique en Médecine
CHU Pitié-Salpêtrière, 91, Bd de l'hôpital
75634 Paris cédex 13

I. INTRODUCTION

On s'intéresse de plus en plus à l'analyse des données contenues dans des tableaux à triple entrée [9], appelées souvent données évolutives lorsque la troisième dimension est le temps. Nous n'envisagerons pas dans cet article l'aspect spécifique des données évolutives (troisième dimension ordonnée). Différentes méthodes ont été proposées pour décrire les données contenues dans de tels tableaux en généralisant l'analyse en facteurs communs et spécifiques [10, 3, 6], l'analyse factorielle des correspondances [4], l'analyse en composantes principales [1]. Le problème auquel nous nous sommes attachés est voisin : étant donné un ensemble de variables mesurées sur plusieurs sujets et dans diverses situations, quelles variables choisir pour simplifier ultérieurement le recueil des données ? L'analyse en composantes principales permet de répondre à cette question dans le cas d'un tableau à double entrée soit en utilisant l'algorithme proposé par J.M. BRAUN [2], soit en recherchant les variables les plus corrélées avec les premières composantes. On peut envisager de se ramener à cette situation en juxtaposant les tableaux, c'est-à-dire en considérant les mesures effectuées sur un même sujet dans différentes situations comme provenant de sujets distincts. Mais il n'est pas possible de démêler aisément dans les résultats : 1) ce qui traduit les modifications globales en fonction des situations de recueil des données et, 2) ce qui traduit les différences inter individuelles. Il est donc préférable pour répondre à la première de ces deux questions d'analyser le tableau à double entrée variables x situations obtenu par moyennage sur l'ensemble des sujets comme cela a été proposé par BOUROCHE [1] et d'utiliser pour répondre à la deuxième question, une méthode qui prenne en compte les trois dimensions du tableau initial.

II. ALGORITHME POUR LA RECHERCHE DES VARIABLES QUI PERMETTENT LA MEILLEURE DESCRIPTION D'UNE POPULATION DANS PLUSIEURS SITUATIONS

L'algorithme proposé pour choisir les variables les plus informationnelles dans l situations consiste d'abord à effectuer une analyse en composantes principales pour chaque situation, puis au vu de ces ACP à fixer le nombre q de composantes

que l'on souhaite conserver (le même pour chaque situation), enfin, à chercher les q variables les mieux corrélées en moyenne avec les l espaces définis par les q premières composantes principales de chaque analyse.

Nous appellerons J l'ensemble des p variables et T l'ensemble des l situations où elles sont mesurées. L'algorithme n'exige pas que la troisième dimension soit ordonnée comme l'est, en particulier, le temps.

Soit X^t le tableau $n^t \times p$ des données dans la situation t . Rappelons que l'analyse en composantes principales revient à effectuer une transformation pour obtenir le tableau $V^t = (v_1^t, \dots, v_p^t)$ où v_j^t est le vecteur centré (ACP) ou centré réduit (ACP normée) des valeurs prises par la variable j pour les n^t sujets étudiés dans la situation t puis à rechercher les valeurs et vecteurs propres de la matrice d'inertie $V^{t'} \cdot D^t \cdot V^t$ avec $D^t = 1/n^t \cdot I$. La k -ième composante principale c_k^t est le vecteur propre associé à la k -ième valeur propre λ_k^t de $V^t \cdot D^t \cdot V^{t'}$:

$$V^t \cdot D^t \cdot V^{t'} \cdot c_k^t = \lambda_k^t \cdot c_k^t$$

avec : $c_k^{t'} \cdot D^t \cdot c_k^t = \lambda_k^t$.

Si l'on souhaite conserver les q premiers facteurs de chacune des l analyses effectuées, les variables les plus informationnelles sont celles qui permettent le mieux de reconstituer les projections du nuage des individus dans les espaces à q dimensions ainsi définis. L'algorithme recherche donc les q variables les plus proches en moyenne de ces l espaces, c'est-à-dire les variables qui permettent au mieux de reconstituer dans les espaces \mathbf{R}^{n^t} munis du produit scalaire associé à la matrice $D^t = 1/n^t \cdot I$ les hyperplans définis par les q vecteurs $c_1^t, c_2^t, \dots, c_q^t$.

La méthode utilisée est analogue à celle suivie pour la sélection des variables explicatives en régression multiple : la proximité entre un vecteur et un sous espace vectoriel est appréciée à l'aide du carré du coefficient de corrélation multiple entre ce vecteur et une base du sous-espace considéré. Mais ici, au lieu de rechercher le sous espace *le plus proche de la variable à expliquer* on recherche les sous espaces *les plus proches des sous espaces engendrés par les q premières composantes de chaque analyse* :

$$f_1(j) = \sum_{t \in T} v_j^{t'} \cdot D^t \cdot U^t \cdot U^{t'} \cdot D^t \cdot v_j^t = \sum_{t \in T} \sum_{k=1}^q \rho_{jk}^{t2}$$

en appelant U^t le tableau $n^t \times q$ des vecteurs unitaires des q premières composantes principales dans la situation t et ρ_{jk}^t le coefficient de corrélation linéaire entre la composante k et le vecteur variable v_j^t dans la situation t . Puis on cherche la variable j_1 qui maximise cette quantité :

$$f_1(j_1) = \max_{j \in J} f_1(j)$$

Au deuxième pas, on cherche la variable j_2 la plus proche des hyperplans orthogonaux à j_1 définis par les q premières composantes de chaque analyse. On l'obtient en projetant pour chaque situation t les vecteurs $v_j^t (j \in J)$ sur l'hyperplan de \mathbf{R}^{n^t} orthogonal à $v_{j_1}^t$ puis en calculant la somme des normes des projections de ces nouveaux vecteurs sur les hyperplans définis par les q premières composantes de chaque analyse.

Plus généralement, au pas $m + 1$ ($m + 1 \leq q$), on cherche la variable j_{m+1} la plus proche des hyperplans orthogonaux aux vecteurs correspondant aux m variables déjà choisies et définis par les q premières composantes de chaque analyse. On l'obtient en projetant pour chaque situation t les vecteurs v_j^t ($j \in J$) sur l'hyperplan de \mathbb{R}^n D^t orthogonal à $v_{j_1}^t, v_{j_2}^t, \dots, v_{j_m}^t$ au moyen de l'opérateur Π_m^t :

$$\Pi_m^t = \Pi_{m-1}^t - \frac{v_{j_m} \cdot \Pi_{m-1}^t \cdot \Pi_{m-1}^t \cdot v_{j_m}'}{(\Pi_{m-1}^t \cdot v_{j_m} \cdot v_{j_m}' \cdot \Pi_{m-1}^t)^2}$$

où Π_{m-1}^t est l'opérateur de projection sur l'hyperplan de \mathbb{R}^n D^t orthogonal à $v_{j_1}^t, \dots, v_{j_{m-1}}^t$ (avec $\Pi_0^t = I$). Puis en calculant la somme des normes des projections de ces nouveaux vecteurs sur les hyperplans définis par les q premières composantes de chaque analyse :

$$f_{m+1}(j) = \sum_{t \in T} v_j^t \cdot \Pi_m^t \cdot D^t \cdot U^t \cdot U^{t'} \cdot D^t \cdot \Pi_m^t \cdot v_j^t$$

la variable cherchée j_{m+1} est celle qui maximise cette quantité :

$$f_{m+1}(j_{m+1}) = \max_{j \in J} f_{m+1}(j)$$

III. COMPARAISON AVEC LA METHODE DE RECHERCHE D'INDICATEURS DE J.M. BRAUN

J.M. BRAUN [2] a proposé une méthode permettant de choisir un nombre limité q de variables parmi un ensemble de p variables se présentant sous la forme de vecteurs de \mathbb{R}^n ; par exemple son algorithme permet de déterminer, pour les chroniques mensuelles des consommations d'électricité en France quelles branches industrielles fournissent le meilleur résumé de l'information (les variables sont alors les branches industrielles et les sujets les mois de relevé). Il s'agit d'un problème plus simple que celui que nous avons abordé puisqu'il revient à se limiter à une seule situation (c'est-à-dire un indice économique : consommation d'électricité), cependant, nous verrons que la généralisation à plusieurs situations est possible (on pourrait par exemple ajouter une troisième dimension au tableau évoqué plus haut en considérant plusieurs indices économiques).

La méthode utilise la notion d'opérateur linéaire associé à un ensemble de variables et de distances entre opérateurs. Aux variables centrées réduites (v_j) $j \in J$, vecteurs de \mathbb{R}^n est associé l'opérateur V de \mathbb{R}^n dans \mathbb{R}^n :

$$V(y) = \frac{1}{p} \sum_{j=1}^p (v_j' \cdot D \cdot y) v_j$$

où $D = \frac{1}{n} I$ est la matrice associée ou produit scalaire covariance. Cet opérateur a

pour vecteurs propres les composantes principales (c_j) , $j = 1, p$ obtenues par l'analyse du tableau des (v_j) , $j = 1, p$ et par conséquent les ensembles (c_j) , $j = 1, p$ et (v_j) , $j = 1, p$ ont le même opérateur associé.

Ayant défini une distance entre opérateurs à partir du produit scalaire :

$$\langle U, V \rangle = \sum_{k=1}^n U(e_k)' \cdot D \cdot V(e_k)$$

où U et V sont des opérateurs linéaires D symétriques de \mathbb{R}^n et (e_k) une base D -orthonormée de \mathbb{R}^n , J.M. BRAUN recherche les variables fournissant le meilleur résumé au sens de la plus petite distance entre les opérateurs associés à l'ensemble des variables et au sous-ensemble résumé. Il montre que la variable fournissant le meilleur résumé est celle qui maximise :

$$g_1(j) = \sum_{k=1}^p \lambda_k \rho_{jk}^2 = v_j' \cdot D \cdot C \cdot C' \cdot D \cdot v_j$$

où λ_k est la valeur propre associée à la k -ième composante principale c_k obtenue par l'analyse du tableau des (v_j) $j \in J$, ρ_{jk} le coefficient de corrélation entre la variable j et la composante c_k , et C le tableau des composantes principales. Une fois déterminé l'ensemble M des m variables fournissant le meilleur résumé la $(m+1)$ ième est celle qui maximise :

$$g_{m+1}(j) = \frac{1}{p} \sum_{k=1}^p \lambda_k \rho_{jk}^2 - \frac{1}{m+1} \sum_{i \in M} r_{ij}^2$$

où r_{ij} est le coefficient de corrélation linéaire entre les variables i et j .

La quantité $\frac{1}{p} \sum_{k=1}^p \lambda_k \rho_{jk}^2$ est constante. La $(m+1)$ ième variable est donc

celle qui minimise $\frac{1}{m+1} \sum_{i \in M} r_{ij}^2$ c'est-à-dire la variable la moins corrélée aux m variables précédemment choisies.

Si l'on prend comme opérateur associé à un ensemble de vecteurs de \mathbb{R}^n non plus celui proposé par J.M. BRAUN mais la projection D -orthogonale sur le sous-espace engendré par ces vecteurs, on remarque que les v_j ($j \in J$) et leurs composantes principales ont également même opérateur associé. Etant donnés deux sous-espaces vectoriels de \mathbb{R}^n de dimension p_1 et p_2 , φ_k ($k=1, p_1$) et ψ_l ($l=1, p_2$) des bases de ces sous-espaces on montre aisément que la distance entre les opérateurs U et V qui leur sont associés est :

$$d^2(U, V) = p_1 + p_2 - 2 \sum_{k=1}^{p_1} \sum_{l=1}^{p_2} (\varphi_k' \cdot D \cdot \psi_l)^2$$

On en déduit que la distance entre l'opérateur U associé à p variables et l'opérateur V associé à l'une d'elles est indépendante de j

$d^2(U, V) = p + 1 - 2 = p - 1$ (si les p variables engendrent un espace de dimension p).

Il n'est donc pas possible de déterminer une variable résumé qui minimise cette distance. On peut par contre déterminer de façon ascendante les q variables-résumé qui minimisent la distance entre l'opérateur qui leur est associé et celui

associé aux q premières composantes principales c_k de vecteurs unitaires u_k . La première variable résumé j est celle qui maximise :

$$\sum_{k=1}^q (u'_k \cdot D \cdot v_j)^2 = \sum_{k=1}^q \rho_{jk}^2 = v'_j \cdot D \cdot U \cdot U' \cdot D \cdot v_j$$

où U est le tableau $n \times p$ des vecteurs unitaires des p composantes principales.

Une fois déterminé l'ensemble M des m meilleures variables-résumé, la $(m + 1)$ ième est celle dont la composante w_{j_m} orthogonale au sous-espace engendré par les vecteur $(v_j)_{j \in J}$ maximise.

$$\sum_{k=1}^q \left(\frac{u'_k \cdot D \cdot w_{j_m}}{\|w_{j_m}\|} \right)^2$$

On retrouve, en ce limitant à une seule situation ($\text{card } T = 1$) un critère qui au premier pas est identique à celui utilisé dans notre algorithme mais qui en diffère aux pas suivants en raison du terme $\|w_{j_m}\|^2$ au dénominateur. On peut montrer que ce nouveau critère conduit à retenir la variable dont la composante orthogonale au sous-espace engendré par les vecteurs $(v_j)_{j \in M}$ est la plus proche de l'espace engendré par les q premières composantes principales et ce, même si cette composante est faible, c'est-à-dire même si la variable considérée est fortement corrélée aux variables déjà choisies ; ce nouveau critère est donc peu satisfaisant.

La méthode de J.M. BRAUN peut être généralisée à 1 situations en effectuant 1 analyses en composantes principales et en minimisant la somme des distances entre opérateurs sur l'ensemble des 1 situations : au pas $m + 1$ on cherche la variable j_{m+1} qui maximise :

$$h_{m+1}(j) = \sum_{t \in T} \left[\frac{1}{p} \sum_{k=1}^p \lambda_k^t \rho_{jk}^{t2} - \frac{1}{m+1} \sum_{i \in M} r_{ij}^{t2} \right]$$

où pour la situation t , λ_k^t est la k^e valeur propre, ρ_{jk}^t le coefficient de corrélation entre la k^e composante principale et la variable j et r_{ij}^t le coefficient de corrélation entre les variables i et j , M désignant l'ensemble des m variables choisies aux pas précédents.

On voit que cet algorithme revient à choisir à chaque pas la variable la moins corrélée avec les variables déjà choisies tandis que notre algorithme revient à choisir à chaque pas la variable dont les projections dans les hyperplans orthogonaux aux hyperplans définis par les variables déjà choisies sont les plus corrélées avec les espaces définis par les premières composantes principales.

Les deux algorithmes peuvent se comparer de différents points de vue.

1. Volume des données à manipuler

L'algorithme de J.M. BRAUN exige le calcul et la mémorisation des p coefficients de corrélation entre les variables et *chacun* des facteurs de chaque analyse. Si le nombre de variables, et donc le nombre de facteurs, sont importants, la procédure devient très lourde. Au contraire notre algorithme n'utilise que les coefficients de corrélations entre les variables et les q premiers facteurs de chaque analyse (si q est le nombre de variables-résumé recherché). L'allègement est appréciable.

2. Nombre de variables-résumé obtenu

L'algorithme de J. M. BRAUN permet de modifier ou de fixer a posteriori le nombre de variables-résumé recherché. En effet, il fournit en un seul passage une suite d'ensembles de variables résumé dont le cardinal croît de 1 à p. Au contraire notre algorithme suppose que soit d'abord déterminé le nombre q de variables-résumé souhaité et il ne travaille que sur les q premières composantes de chaque analyse. Cette contrainte correspond en fait à une démarche naturelle : analyser les résultats des analyses en composantes principales effectuées et fixer l'importance de la réduction de données recherchée.

3. Signification des résultats obtenus

L'algorithme de J.M. BRAUN fournit un ensemble de q variables-résumé tel que les q composantes principales qui leur sont associées soient aussi proches que possible (selon un critère donné) des composantes principales calculées à partir de l'ensemble des variables. Notre algorithme fournit un ensemble de q variables résumé permettant de reconstituer le mieux possible les espaces définis par les q premières composantes de chaque analyse. Dans ce cas, il est théoriquement possible que les variables choisies soient moins bien corrélées avec la première composante qu'avec les suivantes. Rappelons cependant que dans une analyse en composantes principales, la somme des carrés des corrélations entre une composante et chaque variable est égale à la valeur propre associée à cette composante. Si la première valeur propre est nettement supérieure aux suivantes, le danger évoqué plus haut est fort improbable. De plus, la méthode de J.M. BRAUN privilégie également les premières composantes en faisant intervenir les valeurs propres dans le calcul. Les deux méthodes doivent donc donner des résultats analogues : même si les variables choisies ne sont pas les mêmes, les valeurs des critères de décision qui leur sont associées doivent être proches. C'est ce que l'on vérifiera dans l'application suivante.

IV. APPLICATION A L'ANALYSE AUTOMATIQUE DE L'EEG – CHOIX DES PARAMETRES SPECTRAUX LES PLUS INFORMATIONNELS

L'analyse spectrale par transformée de Fourier rapide est l'une des méthodes les plus utilisées en analyse automatique de l'électroencéphalogramme. Il faut cependant définir des paramètres permettant de décrire les spectres de puissance ainsi obtenus. De multiples propositions ont été formulées en ce sens, nous en avons retenu 19 et nous avons cherché à déterminer les paramètres les plus informationnels pour la description de l'électroencéphalogramme de l'enfant normal. Les 19 paramètres sont les suivants :

- DEL puissance de la bande delta (1,1 – 3,3 Hz)
- T puissance de la bande theta (4,4 – 6,6 Hz)
- ALP puissance de la bande alpha (7,7 – 12,1 Hz)
- B1 puissance de la bande beta 1 (13,2 – 16,5 Hz)
- B2 puissance de la bande beta 2 (17,6 – 24,2 Hz)
- B3 puissance de la bande beta 3 (25,3 – 33,0 Hz)
- PA1 puissance du pic alpha (1 raie de fréquence)

- PA3 puissance du pic alpha (3 raies de fréquence)
- PB1 puissance du pic beta (1 raie de fréquence)
- PB3 puissance du pic beta (3 raies de fréquence)
- T/A rapport de la puissance de la bande theta sur la puissance de la bande alpha
- FA fréquence du pic alpha
- FB fréquence du pic beta
- ACT activité de HJROTH [7]
- MOB mobilité de HJORTH [7]
- MBD mobilité de la dérivée [5]
- CX complexité de HJORTH [7]
- CX1 complexité d'ordre 1 [8]
- K2 carré du coefficient de résonance [5].

Les enfants sur qui l'étude a porté provenaient de trois classes d'âge 6-8 ans (n = 87), 10-12 ans (n = 82), 14-16 ans (n = 76). Leurs électroencéphalogrammes ont été enregistrés sur quatre dérivations (fronto-centrales droite et gauche, centro-occipitales droite et gauche) et dans 3 états : repos yeux fermés, repos yeux ouverts et hyperpnée (respiration profonde et rapide). Etant donné le caractère symétrique de l'électroencéphalogramme chez le sujet normal, les dérivations droite et gauche n'ont pas été analysées séparément. En raison des artefacts oculaires, l'analyse des dérivations antérieures au repos yeux ouverts a été abandonnée. Au total, ce sont donc 5 situations qui sont étudiées pour chaque classe d'âge.

A. Comparaison des deux algorithmes

Nous avons appliqué l'algorithme de J.M. BRAUN (algorithme 1) et le nôtre (algorithme 2) aux données provenant des 82 enfants de la classe d'âge 10-12 ans. Les trois variables les plus informationnelles selon l'algorithme de J.M. BRAUN sont :

1. l'activité (ACT),
2. la complexité (CX),
3. la puissance du pic beta (PB3) ;

et selon notre algorithme :

1. l'activité (ACT),
2. la mobilité (MOB),
3. la complexité d'ordre 1 (CX1).

On remarque que pour les deux algorithmes l'activité est la variable la plus informationnelle et que les résultats divergent ensuite (malgré l'analogie de nom, la complexité et la complexité d'ordre 1 sont deux paramètres distincts). Le tableau I permet de comparer plus précisément les résultats. On constate qu'au pas numéro 1 (choix de la variable la plus informationnelle) les deux variables pour lesquelles le critère de décision est le plus élevé sont les mêmes dans chaque algorithme et qu'elles figurent au même rang. Au pas numéro deux (choix de la deuxième variable la plus informationnelle), les trois variables pour lesquelles le critère de décision est le plus élevé sont les mêmes dans chaque algorithme, mais les rangs des deux premières sont permutés. Ce résultat est conforme à ce qui a été dit plus haut : bien que les variables choisies par chaque algorithme ne soient pas

les mêmes, les valeurs des critères de décision qui leur sont associées sont proches. Au pas numéro trois les résultats ne sont plus comparables puisque les deux premières variables choisies ne sont pas les mêmes dans chaque algorithme.

TABLEAU I

Comparaison des résultats des deux algorithmes aux deux premiers pas
Variables pour lesquelles le critère de décision est le plus élevé pour
l'algorithme de J.M. BRAUN (algorithme 1) et pour le nôtre
(algorithme 2)

	Algorithme 1		Algorithme 2	
	variable	critère	variable	critère
Pas n° 1	ACT	1,57	ACT	4,80
	ALP	1,52	ALP	4,44
	PA3	1,45	DEL	4,38
Pas n° 2	CX	0,94	MOB	3,68
	MOB	0,79	CX	3,61
	PB3	0,78	PB3	2,96

B. Recherche des variables plus informationnelles

Nous avons ensuite travaillé sur l'ensemble des trois classes d'âge. L'analyse du tableau 19×15 des centres de gravité (19 variables et 5×3 situations) permet de mettre en évidence un premier facteur qui exprime 47,3 % de l'inertie et oppose les situations où le rythme alpha est présent à celles où il est absent. C'est évidemment la puissance de la bande alpha qui est la plus corrélée à cette composante ($\rho = 0,96$). Le deuxième facteur, qui exprime 31,1 % de l'inertie oppose les tracés lents aux tracés rapides, il traduit en particulier la maturation des tracés et la sensibilité à l'hyperpnée. Les variables les mieux corrélées avec ce facteur sont la mobilité ($\rho = 0,93$) et le rapport des puissances dans les bandes theta et alpha ($\rho = -0,96$). Le troisième facteur exprime 13,4 % de l'inertie, il est bien corrélé avec la puissance dans la bande beta 3 ($\rho = 0,81$). Le quatrième facteur n'exprime que 3,1 % de l'inertie.

Nous avons ensuite effectué 15 analyses en composantes principales afin de pouvoir utiliser notre algorithme. Il a été décidé de conserver les trois premiers facteurs de chaque analyse qui expriment de 70 à 76 % de l'inertie. Les trois variables les plus informationnelles selon l'algorithme sont :

1. l'activité $\left(\sum_k \sum_t \rho_k^t{}^2 = 14,3 \right)$
2. la mobilité $\left(\sum_k \sum_t \rho_k^t{}^2 = 13,2 \right)$
3. la complexité d'ordre 1 $\left(\sum_k \sum_t \rho_k^t{}^2 = 12,2 \right)$.

L'ensemble de ces résultats conduit donc à proposer de décrire le spectre de puissance de l'électroencéphalogramme d'enfant au moyen de quatre paramètres : l'activité, la mobilité, la complexité d'ordre un et la puissance de la bande alpha. On peut envisager d'utiliser ces paramètres dans un but de recherche ou d'enseignement et même de les calculer systématiquement, à l'aide d'un mini-ordinateur, afin de les utiliser en routine dans un laboratoire d'électroencéphalographie.

V. CONCLUSION

Diverses méthodes peuvent être envisagées pour reconnaître les variables les plus informationnelles dans un tableau à triple entrée. L'algorithme proposé ici repose sur l'analyse en composantes principales des matrices qui composent ce tableau et il peut être comparé à une généralisation de l'algorithme proposé par J.M. BRAUN pour la recherche d'indicateurs dans les séries chronologiques multiples qui repose également sur l'analyse en composantes principales. La méthode que nous proposons suppose que les résultats des analyses en composantes principales soient étudiés d'abord afin de déterminer le nombre de variables à rechercher. Elle présente l'avantage que le volume des données à manipuler est ainsi plus faible que dans l'algorithme de J.M. BRAUN. Par contre, celui-ci permet à l'utilisateur de choisir a posteriori ou de modifier le nombre des variables retenues. Les deux méthodes donnent cependant des résultats analogues : même lorsque les variables choisies ne sont pas les mêmes, les valeurs des critères de décision qui leur sont associés sont proches.

REFERENCES

- [1] BOUROCHE J.M. — Analyse des données ternaires : la double analyse en composantes principales. Thèse de doctorat de 3^{ème} cycle, Université de Paris VI, 1975.
- [2] BRAUN J.M. — Séries chronologiques multiples : recherche d'indicateurs. *Revue de Stat. Appliquée*, XXI, 1, 1973, 81-108.
- [3] CARROLL J.D., CHANG J.J. — Analysis of individual difference in multidimensional scaling via a N-way generalization of "Eckart-Young" decomposition. *Psychometrika*, 35, 3, 1970 : 283-319.
- [4] CHEVALIER A. — Analyse factorielle évolutive. Thèse de 3^{ème} cycle, Université de Paris VI, 1977.
- [5] GOLDBERG P., ETEVENON P. — Analyse spectrale statistique du signal EEG - Calcul de nouvelles données spectrales caractéristiques. *Rev. Informatique Méd.*, 1973, 4(1) : 23-30.
- [6] HARSHMAN R. — PARAFAC 2 : mathematical and technical note. *Working paper in phonetics*, n° 22, UCLA, 1972.
- [7] HJORTH B. — EEG analysis based on time domain properties. *Electroenceph. Clin. Neurophysiol.*, 1970, 29 : 306-310.

- [8] HJORTH B. – The physical significance of time domain descriptors in EEG analysis. *Electroenceph. Clin. Neurophysiol.*, 1973, 34 : 321-325.
- [9] NAKACHE J.P. – Multidimensional analysis of three-way of data. Proceedings of MEDCOMP 77 (à paraître).
- [10] TUCKER L.R. – Some mathematical notes on three mode factor analysis. *Psychometrika*, 31, 1966 : 279-311.