REVUE FRANÇAISE D'AUTOMATIQUE, INFORMATIQUE, RECHERCHE OPÉRATIONNELLE. RECHERCHE OPÉRATIONNELLE

MICHEL TENENHAUS

BERNARD PRIEURET

Analyse des séries chronologiques multidimensionnelles

Revue française d'automatique, informatique, recherche opérationnelle. Recherche opérationnelle, tome 8, n° V2 (1974), p. 5-16

<http://www.numdam.org/item?id=RO_1974__8_2_5_0>

R.A.I.R.O.
(8e anneé, mai 1974, V-2, p. 5 à 16)

ANALYSE DES SERIES CHRONOLOGIQUES MULTIDIMENSIONNELLES

par Michel TENENHAUS [1] et Bernard PRIEURET [2] [3]

Résumé. — *Les éléments d'une population évoluant au cours du temps sont décrits au moyen de différentes caractéristiques.*

Nous allons étudier l'évolution simultanée de ces séries chronologiques en mettant en évidence une « évolution commune » et en tenant compte de cette évolution, nous allons résumer l'ensemble de ces caractéristiques par un nombre plus restreint de facteurs.

Cette série chronologique multidimensionnelle sera représentée dans un espace de faible dimension en déformant le moins possible les trajectoires.

Nous pourrons ainsi, entre autres, visualiser l'évolution du phénomène. Par une méthode des moindres carrés, nous chercherons la variété linéaire la plus proche des trajectoires.

Une des dimensions sera interprétée comme la tendance de la série chronologique, les autres dimensions seront un « résumé » des caractéristiques dégagées de l'influence du temps.

Nous obtiendrons donc l'évolution générale au cours du temps ainsi que la description instantanée des liaisons entre les caractères.

INTRODUCTION

L'évolution temporelle d'une population est étudiée suivant p caractères quantitatifs $X_1, \dots X_p$.

Soit le vecteur $X_t = \begin{bmatrix} X_{1t} \\ X_{pt} \end{bmatrix}$ composé des p caractères $X_1, \dots, X_p$ à l'instant t.

Nous appelons série chronologique multidimensionnelle la suite des X_t.

Leur étude présente deux aspects :

— la recherche des liaisons entre les X_t (liaison temporelle),

— la recherche des liaisons entre les X_i (liaison entre les caractères en tenant compte de l'évolution temporelle).

(1) Professeur au C.E.S.A.

(2) Statisticien chez Nielsen.

(3) Les auteurs étaient professeurs à l'Université d'Ottawa lorsqu'ils écrivirent cet article.

Liaison entre les X_t

L'étude des liaisons entre les X_t est basée sur l'idée suivante : dans une série chronologique les termes successifs dépendent fortement des précédents; il est avantageux de les considérer comme générés par une suite de termes indépendants les uns des autres.

Par des considérations mathématiques on est conduit au modèle :

$$X_t + \emptyset_1 X_{t-1} + \dots + \emptyset_p X_{t-p} = \varepsilon_t + \Theta_1 \varepsilon_{t-1} + \dots + \Theta_q \varepsilon_{t-q}$$

c'est-à-dire X_t dépend des p valeurs précédentes $X_{t-1}, \dots, X_{t-p}$ et de $q + 1$ valeurs de termes indépendants.

Si $p = 0$ c'est-à-dire si $X_t = \varepsilon_t + \dots + \Theta_q \varepsilon_{t-q}$ on a le modèle moyenne-mobile d'ordre q.

Si $q = 0$ c'est-à-dire si $X_t + \emptyset_1 X_{t-1} + \dots + \emptyset_p X_{t-p} = \varepsilon_t$ on a le modèle autorégressif d'ordre p.

Si p et q sont différents de zéro on a le modèle autorégressif moyenne-mobile d'ordre (p, q).

Ces modèles ont été étudiés dans le cas multidimensionnel particulièrement par M. H. Quenouille [2]. Ils mettent en évidence la structure temporelle de la série chronologique ainsi que la liaison, si elle existe, entre les X_{it}.

Liaison entre les X_i

Cet aspect de l'étude des séries chronologiques est moins développé que le précédent (signalons l'article de T. W. Anderson [1] sur l'analyse factorielle des séries chronologiques multidimensionnelles) et fait l'objet de cet article.

Nous allons étudier l'évolution simultanée des p séries chronologiques en mettant en évidence une « évolution commune » et en tenant compte de cette évolution nous allons résumer l'ensemble de ces p caractéristiques par un nombre plus restreint de facteurs.

Cette série chronologique multidimensionnelle sera représentée dans un espace de faible dimension en déformant le moins possible les trajectoires. Nous pourrons ainsi, entre autres, visualiser l'évolution du phénomène. Par une méthode des moindres carrés nous chercherons la variété linéaire la plus proche des trajectoires.

Une des dimensions sera interprétée comme la tendance de la série chronologique, les k autres dimensions seront un « résumé » des p caractéristiques dégagées de l'influence du temps.

Nous obtiendrons donc l'évolution générale au cours du temps ainsi que la description instantanée des liaisons entre les caractères.

1. LES DONNEES ET LEUR FORMALISATION

1.1. L'évolution d'une population finie E est étudiée vis-à-vis de p caractères réels au cours du temps.

Soit T l'ensemble fini des temps d'observations, soient $X_1, ..., X_p$ les caractères.

Appelons X l'application de $T \times E$ dans R^{p+1} définie par

$$X(t, e) = \begin{pmatrix} X_1(t, e) \\ . \\ . \\ . \\ X_p(t, e) \\ X_{p+1}(t, e) \end{pmatrix}$$

où $X_i(t, e)$ représente la valeur de la variable X_i sur l'observation e considérée au temps t et où $X_{p+1}(t, e) = t$.

La trajectoire associée à l'élément e de la population E est :

$$\{ X(t, e) ; t \in T \}; \text{ elle est notée } X(T, e).$$

La coupe instantanée à t, c'est-à-dire la population E étudiée à l'instant t, est :

$$\{ X(t, e) ; e \in E \}; \text{ elle est notée } X(t, E).$$

1.2. Soit $f_1, ..., f_{p+1}$ une base d'un espace vectoriel réel V de dimension $p + 1$.

Le vecteur de V associé à l'observation (t, e), vecteur que nous notons encore $X(t, e)$, s'écrit :

$$X(t, e) = \sum_{i=1}^{p+1} X_i(t, e) f_i$$

Posons $E_t = \{ t \} \times E$; à E_t est associé l'hyperplan $H_t = H_0 + t \cdot f_{p+1}$ où H_0 est le sous-espace vectoriel de V engendré par $f_1, ..., f_p$.

Nous supposons que V est muni d'un produit scalaire définie par une matrice Q définie positive :

Notons $\langle x, y \rangle = x'Qy$ et $d^2(x, y) = (x - y)'Q(x - y)$.

Le rôle particulier joué par la variable temps nous conduit à supposer :

$$\langle f_{p+1}, f_i \rangle = \begin{cases} 0 & i = 1, ..., p \\ 1 & i = p + 1 \end{cases}.$$

2. REPRESENTATION DE LA SERIE CHRONOLOGIQUE MULTIDIMENSIONNELLE DANS UNE VARIETE LINEAIRE

2.1. Notre but est de représenter la série chronologique multidimentionnelle dans une variété linéaire H de V.

Nous supposons que la dimension $k+1$ de H est inférieure à $p+1$ et que H n'est pas parallèle à H_0.

La coupe instantanée $X(t, E)$ est un sous-ensemble de H_t, on désire que sa représentation dans H soit encore un sous-ensemble de H_t.

Une représentation naturelle de $X(t, E)$ est le sous-ensemble de $H \cap H_t$ définie par : $P_{H \cap H_t}(X(t, E)) = \{ P_{H \cap H_t}(X(t, e)), e \in E \}$ où $P_{H \cap H_t}(X(t, e))$ représente la projection Q-orthogonale de $X(t, e)$ dans $H \cap H_t$.

La représentation dans H de la trajectoire associée à l'élément e est définie par

$$R_H(X(T, e)) = \{ P_{H \cap H_t}(X(t, e)), t \in T \}.$$

2.2. Une variété linéaire est somme directe d'un vecteur a et d'un sous-espace vectoriel $\mathcal{K}$ de V : $H = \{ a \} + \mathcal{K}$. Un vecteur a de H et une base de $\mathcal{K}$ peuvent être choisis de manière à simplifier les calculs.

Les propositions suivantes précisent ce choix.

Proposition 2.2.1.

Une variété linéaire H de dimension $k+1$ non parallèle à H_0 peut être définie par un vecteur a appartenant à $H \cap H_0$ et une base $(u_1, ..., u_k, v)$ de $\mathcal{K}$ où $(u_1, ..., u_k)$ est une base Q-orthonormale de $U = \mathcal{K} \cap H_0$ et v un vecteur de $\mathcal{K} \cap H_1$.

Preuve :

1) a étant un vecteur quelconque de H; il peut être choisi dans $H \cap H_0$.

2) La dimension de $U = \mathcal{K} \cap H_0$ est k :

Cela provient de la relation

$$\dim \mathcal{K} + \dim H_0 = \dim (\mathcal{K} + H_0) + \dim (\mathcal{K} \cap H_0),$$

sachant que $\dim \mathcal{K} = k+1$, $\dim H_0 = p$ et $\dim (\mathcal{K} + H_0) = p+1$ puisque $\mathcal{K}$ n'est pas contenu dans H_0.

Il est donc possible de choisir une base Q-orthonormale de $U : u_1, ..., u_k$.

3) Un vecteur v appartenant à $\mathcal{K} \cap H_1$ est de la forme $f_{p+1} + \sum_{i=1}^{p} \alpha_i f_i$ alors que les u_i appartenant à $\mathcal{K} \cap H_0$ sont de la forme $\sum_{i=1}^{p} \alpha_i f_i$. Le vecteur v est donc indépendant des $u_1 ... u_k$; $(u_1, ..., u_k, v)$ forme bien une base de $\mathcal{K}$.

Proposition 2.2.2.

Si H est définie comme dans la proposition 2.2.1

$$H \cap H_t = \{ h_t \} + U \qquad \text{où} \qquad h_t = a + t \cdot v$$

Preuve :

$$\begin{aligned} H \cap H_t &= (\{ a \} + \mathcal{K}) \cap (t \cdot f_{p+1} + H_0) \\ &= \{ h_t \} + \mathcal{K} \cap H_0 = \{ h_t \} + U \text{ où } h_t \text{ est un vecteur de } H \cap H_t. \end{aligned}$$

Montrons que $a + t \cdot v \in H \cap H_t$

$a + t \cdot v \in H$ puisque $v \in \mathcal{K}$

$a + t \cdot v \in H_t$ puisque $\langle a + t \cdot v, f_{p+1} \rangle = t$.

$(a \in H_0 \Rightarrow \langle a, f_{p+1} \rangle = 0, \qquad v \in H_1 \Rightarrow \langle v, f_{p+1} \rangle = 1)$

On peut donc prendre $h_t = a + t \cdot v$.

3. MEILLEURE VARIETE LINEAIRE AU SENS DES MOINDRES CARRES

3.1. Distance entre une trajectoire $X(T, e)$ et sa représentation dans $H : R_H(X(T, e))$

Cette distance est définie par

$$D^2(X(T, e), R_H(X(T, e))) = \sum_{t \in T} d^2(X(t, e), P_{H \cap H_t}(X(t, e)))$$

3.2. Meilleure variété linéaire au sens des moindres carrés

C'est la variété linéaire H minimisant la quantité

$$\sum_{e \in E} D^2(X(T, e), R_H(X(T, e)))$$

REMARQUES

1) Si T se réduit à un seul élément on retrouve le critère utilisé en analyse des composantes principales.

2) Si $p = 1$ on retrouve la régression simple entre la variable dépendante X_t et la variable temps.

Le théorème suivant permet de construire la meilleure variété linéaire, de dimension $k + 1$, au sens des moindres carrés.

Notons :

$$g_t = |E|^{-1} \sum_{e \in E} X(t, e) \qquad ; \qquad \bar{g}_t = |T|^{-1} \sum_{t \in T} g_t$$

$$\bar{t} = |T|^{-1} \sum_{t \in T} t \qquad \text{(où } |E| \text{ représente le cardinal de } E)$$

3.3. Théorème

La meilleure variété linéaire (au sens des moindres carrés) $H = a + \mathcal{K}$ de dimension $k + 1$ est définie par :

1) le vecteur a :

$$a = \bar{g}_t - \bar{t}v$$

où

$$v = \frac{\sum_t (t - \bar{t})(g_t - \bar{g}_t)}{\sum_t (t - \bar{t})^2}$$

2) le sous-espace vectoriel $\mathcal{K}$ engendré par les vecteurs $u_1, \ldots, u_k, v$ où $u_1, \ldots, u_k$ sont les k vecteurs propres associés aux k plus grandes valeurs propres de la matrice SQ, S étant la matrice définie par :

$$S = \sum_{t \in T} \sum_{e \in E} (X(t, e) - a - tv)(X(t, e) - a - tv)'$$

Preuve

3.3.1. Position du problème

Une variété linéaire H pouvant être définie (cf. § 2) à l'aide des vecteurs $a \in H_0, v \in H_1$ et $u_1 \ldots u_k$ vecteurs Q-orthonormés de H_0, la recherche de la meilleure variété linéaire se ramène au problème suivant :

Minimiser la fonction

$$L(a, v, u_1, \ldots, u_k) = \sum_{e \in E} D^2(X(T, e), R_H(X)T, e)))$$

sous les contraintes : $a \in H_0$; $v \in H_1$; $u_1 \ldots u_k$ vecteurs Q-orthonormés de H_0.

3.3.2. Expressions explicites de $L(a, v, u_1, \ldots, u_k)$

3.3.2.1. *Calcul de* $d^2(X(t, e), P_{H \cap H_t}(X(t, e)))$

$H \cap H_t = \{ h_t \} + U$ et $u_1, \ldots, u_k$ est une base Q-orthonormale de U.

$$P_{H \cap H_t}(X(t, e)) = h_t + \sum_{i=1}^{k} u_i' Q(X(t, e) - h_t) u_i$$

d'où

$$d^2(X(t, e), P_{H \cap H_t}(X(t, e))) \\ = [X(t, e) - P_{H \cap H_t}(X(t, e))]' Q [X(t, e) - P_{H \cap H_t}(X(t, e))]$$

ce qui s'écrit

$$(X(t, e) - h_t)' Q (X(t, e) - h_t) - \sum_{i=1}^{k} u_i' Q (X(t, e) - h_t)(X(t, e) - h_t)' Q u_i$$

ou encore

$$(X(t, e) - h_t)' Q (X(t, e) - h_t) - (X(t, e) - h_t)' Q \sum_{i=1}^{k} u_i u_i' Q (X(t, e) - h_t).$$

d'où

3.3.2.2. *Expressions explicites de $L(a, v, u_1, ..., u_k)$*

$$L(a, v, u_1, ..., u_k) = \sum_{t \in T} \sum_{e \in E} (X(t, e) - a - tv)' Q (X(t, e) - a - tv) \\ - \sum_{t \in T} \sum_{e \in E} \sum_{i=1}^{k} u_i' Q (X(t, e) - a - tv)(X(t, e) - a - tv)' Q u_i \quad (1)$$

et d'après la deuxième expression

$$L(a, v, u_1, ..., u_k) = \\ \sum_{t \in T} \sum_{e \in E} ((X(t, e) - a - tv)' \left[Q - Q \sum_{i=1}^{k} u_i u_i' Q \right] (X(t, e) - a - tv)) \quad (2)$$

3.3.3. Expression des contraintes

Les contraintes s'écrivent $a \in H_0 : g_1(a) = \langle f_{p+1}, a \rangle = 0$

$v \in H_1 : g_2(v) = \langle f_{p+1}, v \rangle - 1 = 0$

$u_1 ..., u_k$ base orthonormale de $H_0 : g_3(u_i, u_j) = \langle u_i, u_j \rangle = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$

3.3.4. Recherche du minimum

Il s'agit de minimiser $L(a, v, u_1, ..., u_k)$ sous les contraintes

$$g_1(a) = 0 \qquad g_2(v) = 0 \qquad g_3(u_i, u_j) = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$$

3.3.4.1. *Calcul de a^* et v^* minimisant $L(a, v, u_1, ..., u_k)$*

Utilisons la deuxième expression de L et posons $M = Q - Q \sum_{i=1}^{k} u_i u_i' Q$

L s'écrit : $L(a, v, u_1, ..., u_k) = \sum_{t \in T} \sum_{e \in E} (X(t, e) - a - tv)' M (X(t, e) - a - tv)$

La minimisation de $L(a, v, u_1, \ldots, u_k)$ en fonction de a et v sous les contraintes $g_1(a) = 0$ et $g_2(v) = 0$ s'effectue en utilisant la méthode des multiplicateurs de Lagrange.

Introduisons la fonction

$$F(\lambda_1, \lambda_2, a, v, u_1, \ldots, u_k) = L(a, v, u_1, \ldots, u_k) - \lambda_1 g_1(a) - \lambda_2 g_2(v)$$

et annulons les dérivées premières :

$$\frac{\partial F}{\partial a}(\lambda_1^*, \lambda_2^*, a^*, v^*, u_1, \ldots, u_k) = \sum_{t \in T} \sum_{e \in E} -(X(t, e) - a^* - tv^*)'M - \lambda_1^* f'_{p+1} Q = 0$$

$$\frac{\partial F}{\partial v}(\lambda_1^*, \lambda_2^*, a^*, v^*, u_1, \ldots, u_k) = \sum_{t \in T} \sum_{e \in E} -t(X(t, e) - a^* - tv^*)'M - \lambda_2^* f'_{p+1} Q = 0$$

Remarquant que

$$Mx = Q\left[I - \sum_{i=1}^{k} u_i u_i' Q\right] x = Q[x - P_U(x)] = QP_{U^\perp}(x),$$

la multiplication par f_{p+1} des deux équations donne (puisque $Mf_{p+1} = Qf_{p+1}$, $X'(t, e)Qf_{p+1} = t$, $a'Qf_{p+1} = 0$ et $v'Qf_{p+1} = 1$) $\lambda_1^* = \lambda_2^* = 0$

A l'optimum on a donc :

$$\sum_{t \in T} \sum_{e \in E} (X(t, e) - a^* - tv^*)'M = 0$$

$$\sum_{t \in T} \sum_{e \in E} t(X(t, e) - a^* - tv^*)'M = 0$$

M étant une matrice semi-définie positive ($x'Mx = d^2(x, P_U(x))$) dont le noyau est U, ces deux égalités ont lieu lorsque :

$\sum_{t \in T} \sum_{e \in E} (X(t, e) - a^* - tv^*)$ et $\sum_{t \in T} \sum_{e \in E} t(X(t, e) - a^* - tv^*)$ appartiennent à U.

U étant un sous-espace vectoriel, il contient l'origine. On choisit :

$$\sum_{t \in T} \sum_{e \in E} (X(t, e) - a^* - tv^*) = 0$$

et

$$\sum_{t \in T} \sum_{e \in E} t(X(t, e) - a^* - tv^*) = 0$$

d'où

$$a^* = \bar{g}_t - \bar{t} \cdot v^*$$

et

$$v^* = \frac{\sum_t (t - \bar{t})(g_t - \bar{g}_t)}{\sum_t (t - \bar{t})^2}$$

les calculs précédents ($\lambda_1^* = \lambda_2^* = 0$) montrent que la minimisation de L sous les contraintes $g_1(a) = g_2(v) = 0$ est équivalente à la minimisation de L sans contrainte.

La résolution de $\frac{\partial L}{\partial a} = 0$ et de $\frac{\partial L}{\partial v} = 0$ donne les mêmes résultats a^* et v^* que précèdemment. Vérifions maintenant qu'on obtient bien le minimum.

L est une fonction convexe :

$(a, v) \mapsto (X(t, e) - a - tv)'M(X(t, e) - a - tv)$ composée d'une fonction linéaire et d'une fonction convexe (M est semi-définie positive) est une fonction convexe.

$(a, v) \mapsto \sum_{t \in T} \sum_{e \in E} (X(t, e) - a - tv)'M(X(t, e) - a - tv)$ est donc une fonction convexe. Donc pour tout a et v annulant les dérivées premières la valeur de la fonction est minimum.

3.3.4.2. *Recherche de $u_1^* \ldots u_k^*$ minimisant $L(a^*, v^*, u_1, \ldots, u_k)$*

Appelons $R(u_1, \ldots, u_k) = L(a^*, v^*, u_1, \ldots, u_k)$.

Minimiser R revient, d'après l'expression (1) du paragraphe 3.3.2.2. à maximiser

$$\sum_{t \in T} \sum_{e \in E} \sum_{i=1}^{k} u_i'Q(X(t, e) - h_t^*)(X(t, e) - h_t^*)'Qu_i$$

sous les contraintes $u_i'Qu_j = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$

$$(h_t^* = a^* + tv^*)$$

Notons S la matrice $\sum_{t \in T} \sum_{e \in E} (X(t, e) - h_t^*)(X(t, e) - h_t^*)'$ et $A = QSQ$.

Le problème consiste à maximiser $\sum_{i=1}^{k} u_i'Au_i$ sous les contraintes

$$u_i'Qu_j = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$$

On sait (voir C. R. Rao [3], page 51) que ce maximum est atteint pour les k vecteurs propres correspondants aux k plus grandes valeurs propres de A

dans la métrique Q (c'est-à-dire aux k plus grandes valeurs propres de $Q^{-1}A$ dans les métriques I).

$u_1^*, u_2^*, \ldots, u_k^*$ sont les k vecteurs propres de $Q^{-1}A = SQ$ correspondants aux k plus grandes valeurs propres.

4. QUALITE DE LA REPRESENTATION DE LA SERIE CHRONOLOGIQUE DANS LA VARIETE LINEAIRE DES MOINDRES CARRES

4.1. Tendance de la série et qualité de cette tendance

Les vecteurs $a^* = \bar{g}_t - \bar{t} \cdot v^*$ et $v^* = \dfrac{\sum_t (t - \bar{t})(g_t - \bar{g}_t)}{\sum_t (t - \bar{t})^2}$ obtenus au paragraphe précédent sont aussi solution de

$$\min_{a,v} \sum_t d^2(a + tv, g_t) = \min_{a,v} \sum_t (a + tv - g_t)' Q (a + tv - g_t)$$

On a donc trouvé la droite des moindres carrés des points $\{ t, g_t \}_{t \in T}$. La droite $\{ x \mid x = a^* + \rho \cdot v^*, \rho \in R \}$ résume l'évolution de la série chronologique au cours du temps. On peut l'appeler tendance de la série chronologique. La décomposition de la somme des carrés s'écrit :

$$\sum_t d^2(g_t, \bar{g}_t) = \sum_t d^2(g_t, h_t^*) + \sum_t d^2(h_t^*, \bar{g}_t) \quad \text{où} \quad h_t^* = a^* + t \cdot v^*$$

il est donc naturel de mesurer la qualité de cette tendance, par le rapport :

$$\frac{\sum_t d^2(h_t^*, \bar{g}_t)}{\sum_t d^2(g_t, \bar{g}_t)}$$

4.2. Qualité de la représentation des coupes $X(t, E)$

Les vecteurs $u_1^*, \ldots, u_k^*$ engendrent le sous-espace vectoriel U de dimension k contenu dans H_0. C'est dans ce sous-espace qu'on obtient la meilleure représentation des coupes $X(t, E)$ centrées sur h_t^* : $\{ X(t, e) - h_t^*, e \in E \} \subset H_0$.

On a la décomposition de la somme des carrés pour une coupe instantanée :

$$\sum_{e \in E} d^2(X(t, e), h_t^*) = \sum_{e \in E} d^2(X(t, e), P_{\{h_t^*\} + U^\perp}(X(t, e)) + \sum_{e \in E} d^2(X(t, e), P_{\{h_t^*\} + U}(X(t, e)))$$

dont l'interprétation en terme d'inertie est :

$$I(X(t, E) \mid h_t^*) = I(X(t, E) \mid \{ h_t^* \} + U^\perp) + I(X(t, E) \mid \{ h_t^* \} + U)$$

$I(X(t, E) \mid h_t^*)$ mesure la dispersion de la coupe instantanée autour de h_t^*.

$I(X(t, E) \mid \{ h_t^* \} + U^\perp)$ mesure la dispersion de la représentation de la coupe instantanée dans $\{ h_t^* \} + U$.

$I(X(t, E) \mid \{ h_t^* \} + U)$ mesure la dispersion de la coupe instantanée autour de $\{ h_t^* \} + U$.

Par conséquent la somme des carrés globale s'écrit

$$\sum_{t \in T} I(X(t, E) \mid h_t^*) = \sum_{t \in T} I(X(t, E) \mid \{ h_t^* \} + U^\perp) + \sum_{t \in T} I(X(t, E) \mid \{ h_t^* \} + U)$$

et s'interprète comme suit :

$\sum_{t \in T} I(X(t, E) \mid h_t^*)$: inertie totale

$\sum_{t \in T} I(X(t, E) \mid \{ h_t^* \} + U^\perp)$: inertie expliquée par la vérité linéaire H

$\sum_{t \in T} I(X(t, E) \mid \{ h_t^* \} + U)$: inertie résiduelle.

Notons $\lambda_1, \ldots, \lambda_p$ les p valeurs propres de SQ, ordonnées par ordre décroissant. Il est facile de vérifier, puisque u_i^* est le vecteur propre associé à la valeur propre λ_i, que

$$\sum_{t \in T} I(X(t, E) \mid h_t^*) = \sum_{i=1}^{p} \lambda_i$$

$$\sum_{t \in T} I(X(t, E) \mid \{ h_t^* \} + U^\perp) = \sum_{i=1}^{k} \lambda_i$$

et

$$\sum_{t \in T} I(X(t, E) \mid \{ h_t^* \} + U) = \sum_{i=k+1}^{p} \lambda_i$$

La qualité de la représentation des coupes est donc mesurée par le rapport :

$$\frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{p} \lambda_i} \text{ pourcentage d'inertie expliquée.}$$

5. FACTEURS

Aux vecteurs $u_1^*, ..., u_k^*$ sont associés les k facteurs $Y_1 ... Y_k$ définie par $Y_i : T \times E \to R$, $Y_i(t, e) = \langle X(t, e) - h_i^*, u_i^* \rangle$

les facteurs Y_i sont centrés et de variance $\lambda_i/|E|\,|T|$.

Les p variables $X_1, ..., X_p$ sont résumées par k nouvelles variables $Y_1, ..., Y_k$. Si les $p - k$ dernières valeurs propres sont faibles, cela signifie que les variances des facteurs $Y_{k+1}, ..., Y_p$ sont faibles et que par conséquent ces derniers apportent une information plus restreinte. Les facteurs sont interprétés à l'aide de l'étude des corrélations entre les variables et les facteurs.

CONCLUSION

Seul l'aspect géométrique de l'analyse des séries chronologiques multidimensionnelles a été étudié dans cet article. Nous avons mis en évidence la dimension temporelle ainsi que les dimensions instantanées — dimensions liées aux caractères.

Nous avons également construit des indices mesurant l'adéquation du modèle à la réalité. Nous avons l'intention de poursuivre cette recherche dans deux directions :

1) développement de programmes permettant de tester cette méthode sur des données réelles (nous ne disposons actuellement que d'un programme écrit en APL).

2) Ne plus considérer les données comme formant toute la population mais formant un échantillon. Cela conduit à l'introduction d'hypothèses probabilistes (loi multi-normale) permettant de tester la signification statistique du modèle.

REFERENCES

[1] ANDERSON T. W. (1963), « The use of factor analysis in the statistical analysis of multiple time series, » *Psychometrika*, vol. 28.

[2] QUENOUILLE M. H. (1957), *The Analysis of Multiple Time-Series*, London : Griffin.

[3] RAO C.R. (1965), *Linear Statistical Inference and its Applications*, New York : J. Wiley, 1965.