

KH. ALUDAAT

A. QUAQAZEH

Corrélation entre séries chronologiques : méthode des séries décalées

Les cahiers de l'analyse des données, tome 13, n° 4 (1988),
p. 495-502

http://www.numdam.org/item?id=CAD_1988__13_4_495_0

© Les cahiers de l'analyse des données, Dunod, 1988, tous droits réservés.

L'accès aux archives de la revue « Les cahiers de l'analyse des données » 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/>

CORRÉLATION ENTRE SÉRIES CHRONOLOGIQUES: MÉTHODE DES SÉRIES DÉCALÉES

[CORRÉL. CHRON. DÉCAL.]

*KH. ALUDAAT**

*A. QUAQAZEH**

1 Divers aspects de la corrélation entre séries chronologiques

Il est commun d'avoir à comparer l'évolution de deux séries chronologiques: c'est le cas, par exemple, en analyse financière, lorsqu'on considère simultanément les indices de deux places, ou encore le cours d'une même action sur deux places différentes.

Comme il est manifeste qu'on ne peut postuler de lois de distribution, normale ou autre, pour de telles grandeurs, le statisticien peu formé aux méthodes multidimensionnelles songe, (nous l'avons vu faire dans des études réelles), à calculer un coefficient de corrélation de rang puis à appliquer le *test* correspondant. Avec deux variables observées simultanément à 100 instants, on a deux ensembles de 100 valeurs qu'on convertit en deux suites de rangs, dans chacune desquelles figurent les 100 nombres de 1 à 100; ou encore, on a 100 couples de rangs, donnée qui se prête aux calculs classiques.

Or si, comme ce peut bien être le cas, on compare deux séries prises dans un mouvement général de hausse, on aura certainement une corrélation entre les rangs positive et indubitablement *significative*, mais dépourvue d'intérêt pour le praticien; celui-ci sera alors conduit à spéculer sur la valeur plus ou moins proche de 1 du coefficient de corrélation; ce qui, d'une part, est, dans le cadre des modèles classiques, assez délicat; et, d'autre part, ne fournit pas de réponse aux véritables questions.

En effet, pour rester dans le domaine de notre exemple, on doit distinguer entre tendance à long ou moyen terme et mode de réponse aux sollicitations instantanées; et il est *a priori* certain qu'un seul coefficient ne suffira pas à mesurer la similitude entre les évolutions de deux séries de ces deux points de

(*) Université de Jordanie

vue. Si la tendance haussière est nette sur l'ensemble de la période considérée, le coefficient de corrélation de rang sera positif; si, au contraire, (éventuellement après application d'une correction pour éliminer la tendance générale), l'une des séries est croissante et l'autre décroissante, le coefficient sera négatif.

Cependant, dans l'un et l'autre cas, les fluctuations des deux séries, (ou effets à court terme affectant la ligne de la tendance), pourront être soit indépendantes, soit corrélées positivement, soit négativement, (si ce qui survient de favorable à l'une est défavorable à l'autre); soit encore liées mais déphasées, (si ce qui affecte l'une affecte l'autre semblablement, mais avec un délai différent), en sorte que, même si la tendance à long terme était identique pour les deux séries, la liaison ne se refléterait pas dans la valeur du coefficient de corrélation.

Dans ces conditions, puisque ce que recouvre le terme de *corrélation entre séries chronologiques* pris non dans son sens numérique, mais en général, est une réalité multidimensionnelle, il apparaît nécessaire de recourir à une méthode d'analyse elle-même multidimensionnelle. C'est ce qu'on fera ici sur des cas modèles conçus pour offrir, sous une forme aussi nette que possible, des exemples des phénomènes que nous venons d'évoquer. La méthode multidimensionnelle employée étant indépendante de tout paramétrage, elle réussira sur les cas réels, nécessairement complexes et irréguliers, grâce à l'aptitude au filtrage reconnue à l'analyse des correspondances.

2 Séries chronologiques décalées et modèles de corrélation

L'analyse des séries chronologiques décalées a déjà fait l'objet de plusieurs articles publiés dans *C.A.D.*, (cf. 1, 2, 3). En bref, partant d'un tableau de correspondance $T \times J$ pour lequel l'ensemble T des lignes est un ensemble d'instant successifs, séparés par des intervalles de durée constante, on associe à chaque colonne, ou variable j , une ou plusieurs modalités décalées, notées simplement $j+h$ ou $j-h$, selon que le décalage est effectué vers l'avenir ou vers le passé, et définies par la formule suivante, (où, pour simplifier les notations, les instants t sont assimilés à des entiers successifs):

$$k(t, j+h) = k(t+h, j) \quad ; \quad k(t, j-h) = k(t-h, j) ;$$

sans qu'il soit nécessaire que l'ensemble des décalages utilisés, $+h$ ou $-h$, soit le même pour toutes les variables j .

Dans (3), la méthode est appliquée à des tableaux dont les colonnes sont engendrées par calcul, afin de découvrir, sur ces cas modèle, des principes d'interprétation valant pour des données réelles. On reprend ici le même programme de génération de séries chronologiques, afin de créer des couples de variables présentant des corrélations de toute forme; et on juxtapose à ces séries des séries décalées, avant de procéder à l'analyse. Bien qu'il soit souhaitable, et, d'ailleurs, facile, de procéder à de très nombreux essais, auxquels nous

```

parametres du fichier B:AS2ww
cart = 120 ; carj = 2
v(t)=A*(1+S*sin(t/Tc)+C*cos(t/Tc))*(1+K*t)*exp(t/Te)
var      A   10Tc   100S   100C   9cartK  10cart/Te
Av       1    20    20     0     0     16
Bv       5    20     0     20    0    -16

résultats de l'analyse factorielle pour B:AS2ww
trace   :  1.748e-1
rang    :   1     2     3     4     5
lambda  :  1519   100   99    15   14   e-4
taux    :  8690   575   564   89   82   e-4
cumul   :  8690  9265  9829  9918 10000  e-4

```

engageons le lecteur à se livrer par lui-même, on se bornera ici à trois exemples, désignés par les sigles AS1, AS2 et AS3.

Sans reprendre les développements mathématiques de (3), il nous suffira de commenter le listage des paramètres produit par le programme 'creserie', ainsi que le tableau des valeurs propres issu de l'analyse du tableau avec séries décalées, pour décrire nos exemples, tout en rappelant quelques principes.

Dans tous nos exemples, on s'est borné à considérer deux séries, chacune calculée comme un produit de 2 facteurs au plus: une fonction sinusoidale, (de période $2\pi \times Tc$), du temps t , et une exponentielle, (dont la constante de temps est notée Te). La valeur de Tc est toujours 2, ce qui correspond à une période voisine de 12; et Te est choisi de telle sorte que, de l'instant initial 1 à l'instant final $cart$, la valeur de la série soit multipliée par environ 5, (ou divisée, si l'on donne à Te le signe -).

Dans AS1 et AS2, les variables de base, Av et Bv , sont l'une croissante, l'autre décroissante. Afin de simplifier au maximum la structure, on a choisi le coefficient A par lequel débute la formule de telle sorte que, sur l'ensemble de la période considérée, les deux séries aient sensiblement le même poids total; profitant de ce que le facteur exponentiel varie de 1 à 5, on a donné ces valeurs aux coefficients des séries Av et Bv . Dans AS3, il n'y a pas d'exponentielle.

Avec, pour le facteur sinusoidal, une période voisine de 12, il a paru raisonnable de fixer la durée, $cart$, à 120 points, ce qui fait environ 10 périodes. Et le nombre de séries décalées créées avant et après chaque série de base a été fixé à 6 ce qui permet de couvrir la durée d'une période. Dans une étude de données réelles, 'cart' serait imposé par la durée totale des séries disponibles, ou éventuellement recadré par le praticien; quant à la période, elle n'existe généralement pas, à moins qu'on ait affaire à un phénomène cyclique, (cf. 3). En revanche, on peut, sans s'interdire de tâtonner, choisir un nombre de séries

```

parametres du fichier B:AS1lw
cart = 120 ; carj = 2
v(t)=A*(1+S*sin(t/Tc)+C*cos(t/Tc))*(1+K*t)*exp(t/Te)
var      A    10Tc   100S   100C   9cartK  10cart/Te
Av       1     20    20     0     0     16
Bv       5     20    20     0     0    -16

```

```

résultats de l'analyse factorielle pour B:AS1lw
trace : 1.749e-1
rang   : 1     2     3     4     5
lambda : 1520  102   97    15   14   e-4
taux   : 8691  584   554   88   83   e-4
cumul  : 8691  9275  9829  9917 10000 e-4

```

décalées couvrant une durée qui soit de l'ordre de celle des effets à court terme que l'on désire appréhender.

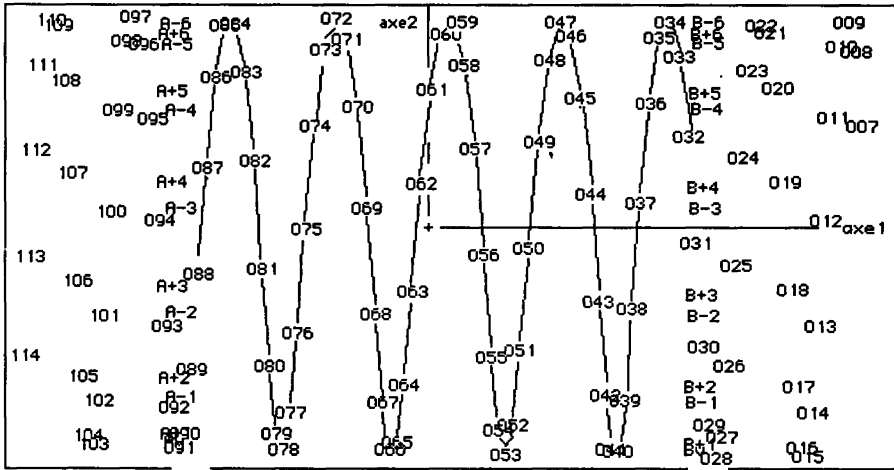
Dans toute les séries, la profondeur de modulation du facteur sinusoïdal a été fixée à 20%: de façon précise, l'un des deux coefficients S et C vaut 20 et l'autre vaut 0. Dans AS1, le facteur sinusoïdal est le même pour les deux séries Av et Bv qui sont donc en phase, répondant ensemble et de la même façon à des sollicitations extérieures présumées. Dans AS1 et AS3, Av est en cosinus et Bv en sinus; il y a donc un déphasage d'un quart de période.

Ainsi, à court terme, la corrélation entre facteurs sinusoïdaux est parfaite pour AS1: elle vaut 1 selon un calcul classique; dans les cas AS2 et AS3, la dépendance est, au déphasage près, parfaite; mais le coefficient de corrélation est nul. À long terme, la corrélation est à apprécier sur les facteurs exponentiels; dans les cas AS1 et AS2, les tendances étant opposées pour les deux variables, le coefficient de corrélation est -1. Dans le cas AS3, il n'y a pas de tendance à long terme, mais seulement un effet sinusoïdal à court terme.

Tout compris, on peut, en jouant sur la force de la tendance à long terme et la profondeur de la modulation à court terme, produire des coefficients de corrélation classiques entre séries qui aient n'importe quelle valeur, significative ou non selon l'épreuve de validité adoptée. Mais il est clair que le problème réel est autre: il s'agit de mettre en évidence, sous tous leurs aspects à long et court terme, les variations concomitantes de deux séries. Nous verrons, sur les cas modèle, comment l'analyse des séries décalées y parvient.

Auparavant, expliquons le nombre de facteurs non triviaux produits par chaque analyse. En bref, dans les cas AS1 et AS2, les séries de base et leurs décalées engendrent un espace de dimension 6, dont les fonctions de base sont, (en utilisant les notations exponentielles complexes):

$$\exp(t \cdot (\pm(1/Te) + (\pm i/Tc))) \text{ et } \exp(\pm t/Te) .$$



Exemple AS1 : plan(1, 2); les flèches signalent la place des modalités A±h et B±h

Dans le cas AS3, il n'y a que 3 fonctions de base qui sont

$$\exp(\pm i \cdot t / T_c) \quad \text{et la constante} : 1.$$

On a donc 5 facteurs dans les cas AS1 et AS2; et 2 dans le cas AS3.

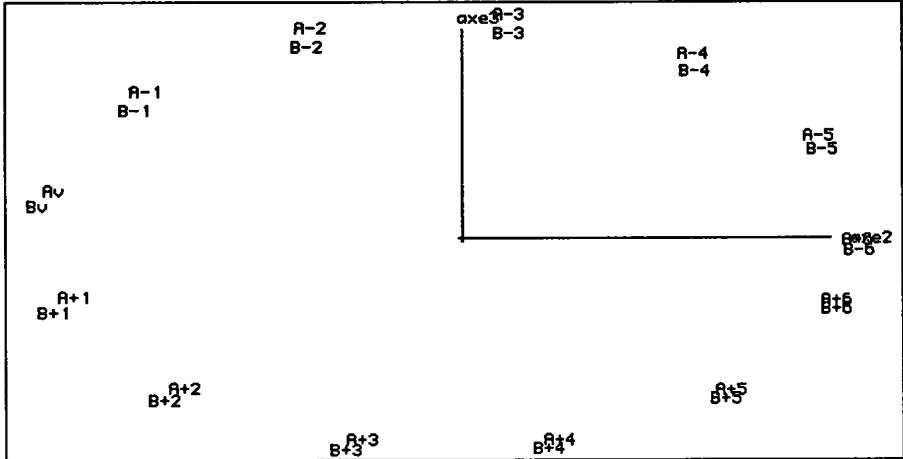
3 Résultats des analyses pour trois exemples

3.1 Exemple AS1

Ainsi qu'on l'a dit, il y a opposition entre les tendances à long terme et concordance parfaite des fluctuations à court terme. Du fait des paramètres que nous avons choisis, l'effet de la tendance à long terme prédomine: celle-ci crée le premier axe, qui porte près de 87% de l'inertie totale du nuage. Les fluctuations à court terme créent le plan (2, 3); ainsi que le plan (4, 5) dont il suffira de dire qu'il ne diffère du plan (2, 3) que par un déphasage entre les A±h et les B±h.

Sur l'axe 1 s'opposent l'avenir, (F1<0), associé aux instants de début de période, (t < 061), et aux modalités décalées ou non de la variable croissante Av; et le passé, (F1>0), associé à la fin de période et aux modalités de la variable décroissante Bv. Les modalités A±h ont des abscisses F1 quasi identiques; et de même pour les modalités B±h. Les instants, t, s'étagent régulièrement: du premier, 007, au dernier, 114. (Il faut prendre garde qu'on ne va pas de 1 à 120, parce qu'aux extrémités les informations manquent pour les séries décalées).

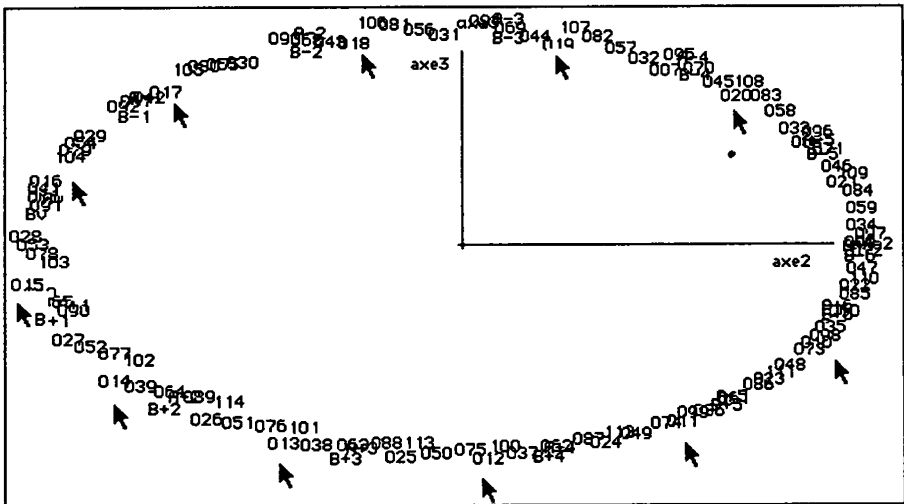
Si, dans le plan (1, 2) on joint les instants successifs, se dessine une sinusoïde; cela est dû au caractère sinusoïdal des facteurs 2 et 3.

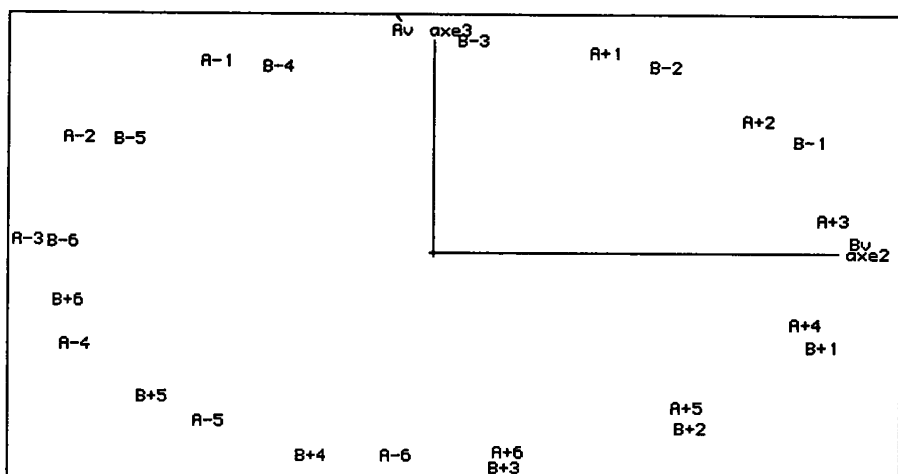


Exemple AS1 : plan (2, 3) : les cycles des modalités A±h et B±h sont en phase

En examinant le plan (2, 3) afférent à l'exemple AS1, il faut prendre garde que, sur les graphiques, on a donné aux deux axes des échelles différentes, afin d'utiliser au mieux la place disponible sur l'écran du Macintosh: mais, les facteurs étant du même ordre de grandeur, les ellipses tracées représentent en fait des cercles. Sur l'ensemble des modalités décalées, qui se succèdent dans le sens trigonométrique de -6 à +6 en passant par v, les deux variables Av et Bv sont parfaitement en phase, ce qui manifeste le parallélisme des séries, à court

Exemple AS1 : plan (2, 3) ; les flèches signalent les instants 010 à 020





Exemple AS2 : entre les modalités $A\pm h$ et $B\pm h$, le déphasage est voisin de 3

terme. Quant à l'ensemble des instants, on peut déchiffrer qu'il s'enroule dans le sens opposé; et avec une période proche de 12,5; les points 007 et 032 se plaçant entre 020 et 019.

3.2 Exemple AS2

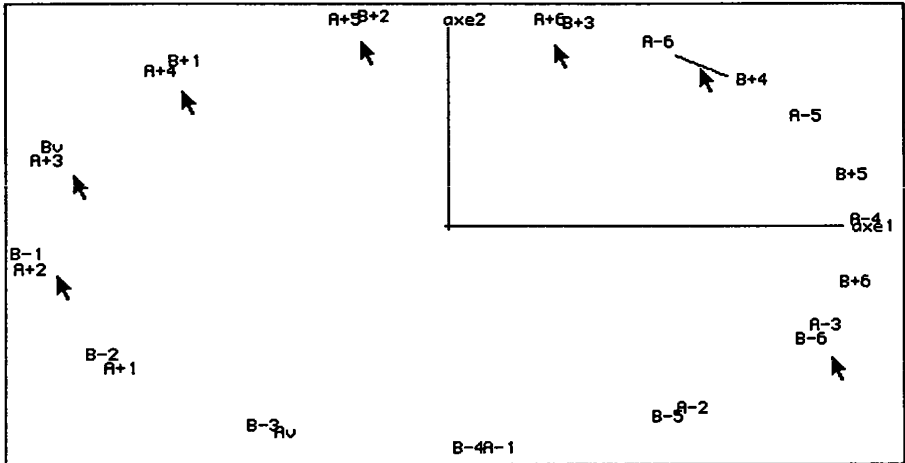
Quant à la tendance à long terme, AS2 ne diffère pas de AS1: les deux séries A_v et B_v ont même poids; celle-ci est décroissante, celle-là croissante. Aussi le facteur 1, (non représenté), est-il le même dans les deux cas.

Dans le plan (2, 3), on a un cercle, (aplatis en une ellipse pour la même raison qu'au §3.1): mais, ici, le déphasage introduit dans les données de base se retrouve dans les résultats de l'analyse.

La série A_v , modulée par un sinus, est en retard d'un quart de période sur B_v , liée au cosinus. C'est pourquoi, à l'extrémité positive de l'axe 2 on trouve B_v tout proche de $A+3$; plus précisément, $A+3$ se place entre B_v et $B-1$, car un quart de période fait un peu plus que 3.

On n'a pas représenté l'ensemble des instants t , lequel, comme dans le cas AS1, (et pour la même raison expliquée dans (1), §3.1.1), dessine un cycle parcouru en sens inverse de ceux des modalités, $A\pm h$ et $B\pm h$.

Nous répéterons que, dans le cas AS2, un calcul usuel de corrélation donnerait une valeur négative élevée, expression de l'opposition de tendance apparue sur l'axe 1; tandis que passerait inaperçu le déphasage des fluctuations, (explicable par des délais de réponse différents à l'environnement, ou encore par le fait que B_v est directement lié à A_v qu'il suit avec un retard).



Exemple AS3 : On observera le déphasage entre le cycle des $A_{\pm h}$ et celui des $B_{\pm h}$

3.3 Exemple AS3

Tous les résultats sont dans le plan (1, 2) dont il suffira de dire qu'il est quasi identique, (à une rotation près), au plan (2, 3) produit par l'exemple AS2.

```
parametres du fichier B:AS3ww
cart = 120 ; carj = 2
v(t)=A*(1+S*sin(t/Tc)+C*cos(t/Tc))*(1+K*t)*exp(t/Te)
var      A      10Tc      100S      100C      9cartK      10cart/Te
Av       2       20       20       0       0       0
Bv       2       20       0       20       0       0

résultats de l'analyse factorielle pour B:AS3ww
trace : 1.996e-2
rang : 1 2
lambda : 101 99 e-4
taux : 5059 4941 e-4
cumul : 5059 10000 e-4
```

4 Conclusion: Nous pensons avoir montré, sur des exemples de cas modèle, que l'analyse des séries décalées est apte à saisir, dans leur complexité, les phénomènes de corrélation entre séries temporelles.

Références bibliographiques

- (1) Aludaat, Kh. et Benzécrici, J.-P. : [SÉRIE MON. FR. 70-84], *CAD*, Vol XII, n° 4, (1987).
- (2) Benzécrici, J.-P. : [SÉRIE MONNAIE FRANCE], *CAD*, vol XII, n° 3, (1987).
- (3) Quaqazeh, A. : [MODÈLE CHRON. DÉCAL.], *CAD*, Vol XII, n° 4, (1987).