LES CAHIERS DE L'ANALYSE DES DONNÉES

A. QUAQAZEH

L'analyse des séries chronologiques décalées. Principes d'interprétation sur des cas modèles

Les cahiers de l'analyse des données, tome 12, nº 4 (1987), p. 407-418

<http://www.numdam.org/item?id=CAD_1987__12_4_407_0>

Lesd Cahiers de l'Analyse des Données
Vol XII - 1987 - n°4 PP.407-418

L'ANALYSE DES SÉRIES CHRONOLOGIQUES DÉCALÉES. PRINCIPES D'INTERPRÉTATION SUR DES CAS MODELES

[MODELE CHRON. DECAL.]

A. QUAQAZEH (*)

1. Introduction des séries décalées

Il est commun d'analyser des tableaux de correspondance TxJ pour lesquels T est une suite d'instants séparés par des intervalles de durée constante, ou encore une suite de tels intervalles, qu'il s'agisse de secondes, de jours, de mois ou d'années. Chaque colonne j du tableau constitue alors une série chronologique. Dans de nombreux domaines, notamment en science économique, l'un des buts de telles analyses est de prévoir l'avenir ; ou, du moins, de formuler des hypothèses vraisemblables quant à l'évolution des séries au delà du dernier instant auquel elles sont connues.

Une voie ordinaire pour extrapoler vers l'avenir est de se placer d'abord dans le passé pour exprimer la série décalée vers l'avenir, f(t+1) en combinaison linéaire approchée de séries non décalées u(t), v(t), w(t). On en arrive ainsi à considérer le schéma général suivant. Etant donné un tableau de correspondance k(t,j), on associe à chaque colonne, ou variable de base j, un ensemble de 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 on a supposé pour simplifier les notations que les instants t considérés sont désignés par des nombres entiers successifs):

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

Il n'est pas nécessaire que l'ensemble des décalages utilisés soit le même pour toutes les variables de base.

Si on désigne par JD l'ensemble des variables de base et de leurs modalités décalées, on a ainsi un tableau TxJD ; plus précisément, il faut prendre garde que si les décalages maxima utilisés vers le passé et vers l'avenir sont respectivement p et q, le tableau a pour ensemble de lignes T', qui diffère de T par la supression

(*)Université de Jordanie

des p premiers instants et des q derniers : certaines données manquent en effet pour les valeurs décalées afférentes à ces instants. Pour tout t de T', la ligne t du tableau TxJD donne les valeurs des variables de base à l'instant t encadrées par leurs valeurs à des instants décalés par rapport à t comme on l'a choisi.

Eventuellement, à partir d'une seule variable de base on construit par décalages un véritable tableau à plusieurs colonnes dont l' analyse peut non seulement aider à la prédiction mais, d'abord, fournir pour une longue période une vue d'ensemble du rythme de la croissance ou de la décroissance.Dans certains cas, on découvrira ainsi des cycles se répétant avec plus ou moins de régularité ; et celà, non par ajustement à un modèle adopté a priori, avec tous les artefacts, aujourd'hui assez généralement reconnus, qui rendent illusoire un tel ajustement, mais parce que la succession des points t sur l'un ou l'autre des graphiques plans issus de l'analyse, aura dessiné des cycles qu'il restera à interpréter en critiquant la régularité. En particulier les variations saisonnières, souvent éliminées des séries publiées, par des méthodes de correction dont les effets ne sont pas maitrisés, apparaitront comme ayant au long des années une amplitude et une forme variables, dignes d'être étudiées.

Comme on peut le voir dans le volume Pratique de l'Analyse des Données en Economie, le point de vue adopté ici n'est pas nouveau ; l'interet des séries décalées est seulement de permettre d'étudier par une méthode multidimensionnelle même une série unique (utilisée pour créer par décalages un tableau à plusieurs colonnes); d'enrichir de références temporelles l'ensemble des variables de base; de concilier avec la considération des profils si féconde en analyse des correspondances, la prise en compte explicite des variations de masse au cours du temps ; le modèle de la croissance exponentielle jouant ici, comme nous le verrons le rôle d'un terme de référence pour lequel la trace est nulle et donc l'analyse triviale.

Nous avons initialement utilisé des séries décalées pour analyser les statistiques de l'économie monétaire de la Jordanie ; la méthode a servi ensuite à Kh. Aludaat et à A Hathout. dans le présent article il s'agira non de données réelles, mais de cas modèles étudiés par l'analyse mathématique (§2) ou simulés par le calcul sur ordinateur (§3).

2. Etudes de cas modèles

2.1. Le modèle de la croissance exponentielle

Même si, en explosion ou en implosion, les conséquences d'un processus exponentiel peuvent être désastreuses, la croissance exponentielle est, après la croissance zéro, le modèle le plus simple de l'évolution temporelle. On a le théorème suivant :

Théorème : La condition nécessaire et suffisante pour qu'une variable de base j et la modalité décalée j+1 (ou j-1) aient même profil sur T est que K(t,j) soit une fonction exponentielle de t.

Preuve : dire que j et j+1 ont même profil revient à dire que :

$\forall t \in T : k(t,j+1)/k(t,j) = k(t+1,j)/k(j) = k(j+1)/k(j) = \text{Constante}$;

or cette formule est celle de la fonction exponentielle.

Il résulte de cette formule que si k(t,j) est exponentielle en t, un tableau formé seulement de la variable j et d'un ensemble de ses modalités décalées aura trace nulle ; et que son analyse ne fournira donc aucun facteur non trivial. Réciproquement, si la trace est non nulle, la variable de base s'écarte du modèle exponentiel. L'interprétation de cet écart se fera selon les principes usuels, en se référant à la formule de reconstitution : la proximité d'un instant t avec les modalités de décalage vers l'avenir signifiant notamment que la croissance est plus rapide que la moyenne, au sens de la dérivée logarithmique;(en effet, t est caractérisé non par des valeurs absolues, mais par les valeurs relatives, dans son profil de la valeur à cet instant et aux instants qui l'encadrent).

Si l'on part d'un ensemble de variables de base qui sont toutes des exponentielles (différant entre elles par la constante de croissance), l'introduction de modalités décalées ne créera aucun nouveau profil de colonne ; et ne pourra modifier l'analyse qu'en modifiant les pondérations relatives des variables de base.

2.2. Le modèle périodique sinusoïdal

Supposons maintenant que la variable j est une fonction sinusoïdale du temps :

$$k(t,j) = Cj + \sin (vj.\ t) ;$$

formule où un terme constant a été introduit afin que k(t,j) soit toujours positif. On a :

$$k(t,j+h) = Cj + \cos(vj.h).\ \sin(vj.\ t) + \sin(vj.h).\cos(vj.\ t) ;$$

d'où il résulte les colonnes construites à partir de j par décalage sont toutes des combinaisons linéaires de 3 fonctions : la constante 1, sin(vj. t) et cos(vj. t). Donc, à partir d'un ensemble de variables de base qui sont toutes des fonctions sinus (ou cosinus) de même période et de phase quelconque, on ne pourra obtenir que 2 facteurs non triviaux. La formule de reconstitution de transition montre que ces facteurs sont des combinaisons linéaires des 2 fontions de base :

$$(Cs + \sin(vj.\ t))/k(t) ; \quad (Cc + \cos(vj.\ t))/k(t) ;$$

où les constantes C sont choisies pour que les fonctions de base soient de moyenne nulle.

Si, en particulier la fonction de marge k(t) est constante, ces fonctions de base ne sont autres que sin(vj.t) et cos(vj.t). Ce cas est réalisé par exemple si on pose :

vj = 2p / ((2.p)+1) ;

(où p est un entier quelconque) et que l'on considère, avec la série de base, p modalités décalées en avance et p en retard ; en effet, dans la somme de chaque ligne les termes trigonométriques ont une somme nulle. Dans le plan (1,2) les points t se placent donc sur une ellipse, centrée à l'origine. En général, si k(t) n'est pas constant, cette figure est déformée, mais reconnaissable, comme on peut le voir par des expériences de calcul numérique (cf §3).

Quand les variables de base sont des fonctions trigonométriques de périodes différentes,on a pour chaque période 2 fonctions de base ; et le nombre des facteurs est au plus égal au double du nombre des périodes distinctes, quel que soit le nombre de séries décalées qu'on introduit.

2.3. Le modèle polynomial

Si K(t,j) est un polynôme en t de degré <=p, il en sera de même pour toutes les modalités décalées de j. Donc si toutes les variables de base sont des polynômes en t de degré <=p, le nombre maximum de facteurs non triviaux sera p; et ces facteurs seront tous des polynômes, à ceci près que, comme dans le cas de sin et cos, il faut tenir compte de la division par K(t); qui est lui-même un polynôme et peut éventuellement être constant. Tous les raisonnements faits ici sont classiques en analyse des correspondances pour l'étude de l'effet Guttman et de ses généralisations (cf. Traité, TII).

2.4. Combinaisons de modèles

Par ce titre, nous voulons évoquer toute sorte de produits et combinaisons linéaires de fonctions classiques dont nous étudierons une famille d'exemples par des calculs de simulation au § 3. Ces exemples n'ont pas seulement un intérêt esthétique, ils habituent le géomètre à reconnaître de quelles relations mathématiques approchées résultent les formes de graphiques issues de l'analyse de données réelles.

Considérons la formule :

k(t,j) = exp(t/Tj) . (cj+ aj sin(vj. t) + bj + cos(vj. t)) ;

il est clair que toutes les modalités décalées de j sont comptrises dans la même formule, les valeurs des constantes a, b, c étant seules modifiées. Quant aux facteurs ils sont, ici encore, des combinaisons linéaires de moyenne nulle des quotients k(t,j)/k(j).Donc, si toutes les p variables de base sont de la même forme, il y a au plus 2p facteurs, à supposer que ces variables diffèrent quant au couple (T,v).

Voici un modèle d'une certaine complexité:

$k(t,j) = (1 + e.t)\ (1+ a.\sin(v.t));$

en laissant fixe v, et modifiant seulement les constantes e et a, ou en créant des modalités décalées, on n'obtiendra que des combinaisons linéaires des 6 fonctions :

$\{1,\ t,\ \sin(v.t),\ \cos(v.t),\ t.\sin(v.t),\ t.\cos(v.t)\};$

la marge k(t) est aussi de cette forme; et les facteurs sont de telles combinaisons, divisées par k(t), et satisfaisant à la condition d'avoir moyenne nulle. Il y a donc au plus 5 facteurs.

On peut, pour la ligne tracée par la suite des t dans les graphiques plans, concevoir des formes type qu'on retouvera sur les simulations du §3, et aussi sur des données réelles.

On sait que si le plan est rapporté à deux axes pour lesquels les facteurs sont des combinaisons linéaires de sin(v.t) et cos(v.t), le point t décrit une ellipse. La division par k(t) peut produire une déformation sensible.

Supposons, pour simplifier le discours, que les deux facteurs sont exactement sin et cos ; le point t décrit alors un cercle de rayon 1. Du fait du coefficient (1/k(t)), la distance du point à l'origine varie avec t ; soit, en particulier, le cas :

$k(t) = 1 + e\ .\ t.\cos(v.t)\ ;\ 1/k(t) \approx 1 - e\ .\ t.\cos(v.t)\ ;$

quand vt varie de 2p, le point t fait un tour, d'autant plus près du cercle de rayon 1 que t est plus petit en valeur absolue; et tantôt à l'extérieur, tantôt à l'intérieur de ce cercle selon le signe de cos(v.t).

Soit maintenant le cas où les deux facteurs sont :

$t.\ \sin(v.t)\ \ ;\ \ t.\ \cos(v.t)\ ;$

la distance à l'origine est égale à t, en valeur absolue; elle varie donc à chaque tour; on a une spirale.

Si l'on ajoute aux deux facteurs sin et cos, le facteur t, on a pour figure tridimensionnelle une spirale tracée sur les deux nappes d'un cône de révolution et passant par le sommet, à l'origine pour t=0.

En réalité, les facteurs ne coïncident généralement pas avec les fonctions de base ; ils en sont des combinaisons linéaires (à la division prèes par k(t), qui est de la même forme. De ce fait, les situations simples que nous venons de décrire se combinent entre elles ; en particulier, on verra au §3.2 des vues perspectives de la spirale tracée sur un cone.

3. Expérimentation numérique

Nous considérerons successivement deux exemples. Sur le premier qui ne produit que deux facteurs, avec un cycle elliptique, nous expliquerons comment on a procédé dans les calculs de simulation. Le deuxième exemple, plus complexe, produira les figures évoquées au §2.4.

3.1. Exemple de cycle elliptique

Avant de procéder à l'analyse factorielle et à l'affichage des résultats, on crée le tableau en deux étapes : d'abord, création des variables de base ; ensuite création des modalités de décalage.

Comme on le voit sur la copie d'écran du dialogue de créations des variables de base, la formule proposée est le produit de 3 facteurs: un terme trigonométrique, un terme linéaire, un terme exponentiel. Les paramètres entiers,. dont le choix est proposé à l'utilisateur, ont été renfermés dans des limites telles qu'on évite la création de valeurs négatives ou de nombres astronomiques ! Les variables de base sont simplement désignées par une capitale suivie de la lettre v, initiale de variable ; les individus, ou instants t sont désignés par les nombres de 001 à 200. Les valeurs des paramètres, choisies en dialogue par l'utilisateur sont conservées dans un fichier texte. Dans notre cas, il n'y a pas de terme linéaire (K=0); les deux variables Av et Bv ont même constante de temps Te dans l'exponetielle et même période Tc dans le terme trigonométrique : on est donc dans un cas cyclique bidimensionnel très simple.

On utilise ensuite le programme "decaler", destiné à adjoindre des modalitées décalées à tout tableau (quelle qu'en soit l'origine : naturelle ou artificielle). Ici, on demande pour chacune des variables 5 modalités en avance et 5 en retard. Le nombre 5 est choisi parce que 10 instants t successifs couvrent à peu près une période du cycle. Le programme affiche à l'écran les modalités décalées, au fur et à mesure qu'il les crée : après le numéro de l'instant, on a une ligne pour les modalités décalées de chaque variable. Sur la copie d'écran, on remarque que, du fait du décalage, les mêmes valeurs se reproduisent sur des lignes obliques ; de plus la suite des instants se termine à 195, parce qu'on a créé 200 valeurs et demandé des modalités décalées jusqu'à A+5 et B+5.

Les résultats sont ceux qu'on attendait. On remarque que les modalités décalées de chaque variable se rangent sur des ellipses ; et se referment à peu près : A+5 est sur A-5; et B+5 sur B-5, compte tenu du choix de 5 d'après la période. Quant aux instants t, leurs numéros s'écrasent en dessinant une ellipse : mais si l'on observe l'affichage des numéros, on les voit se placer successivement en décrivant 10 par 10 des polygones semblables à ceux des modalités de Av ou de Bv, avec un lent décalage; de sorte que finalement, toute l'ellipse est recouverte. Ici encore est en cause le fait qu'avec les constantes choisies la phase se reproduit à peu près quand t augmente de 10. Tous ces

```
faut il poursuivre O ou N O
le nom de base du fichier a creer est eco:ser
ce nom est il confirme O ou N O
nombre cart de valeurs v(t) a creer par variable = 200
nombre de variables a creer = 2
ces nombres sont ils confirmes O ou N O
  Av
Av(t)=A*(1+S*sin(t/Tc)+C*cos(t/Tc))*(1+K*t)*exp(t/Te)
valeur (de 1 a 10) de A = 1
valeur (de 10 a 100) de 10*Tc = 16
valeur (de -50 a 50) de 100*S = 40
valeur (de -50 a 50) de 100*C = -20
valeur (de -8 a 50) de 9*cart*K = 0
valeur (de -40 a 40) de 10*cart/Te = 10
ces nombres sont ils confirmes O ou N O
  Bv
Bv(t)=A*(1+S*sin(t/Tc)+C*cos(t/Tc))*(1+K*t)*exp(t/Te)
valeur (de 1 a 10) de A = 2
valeur (de 10 a 100) de 10*Tc = 16
valeur (de -50 a 50) de 100*S = 20
valeur (de -50 a 50) de 100*C = 15
valeur (de -8 a 50) de 9*cart*K = 0
valeur (de -40 a 40) de 10*cart/Te = 10
ces nombres sont ils confirmes O ou N O
```

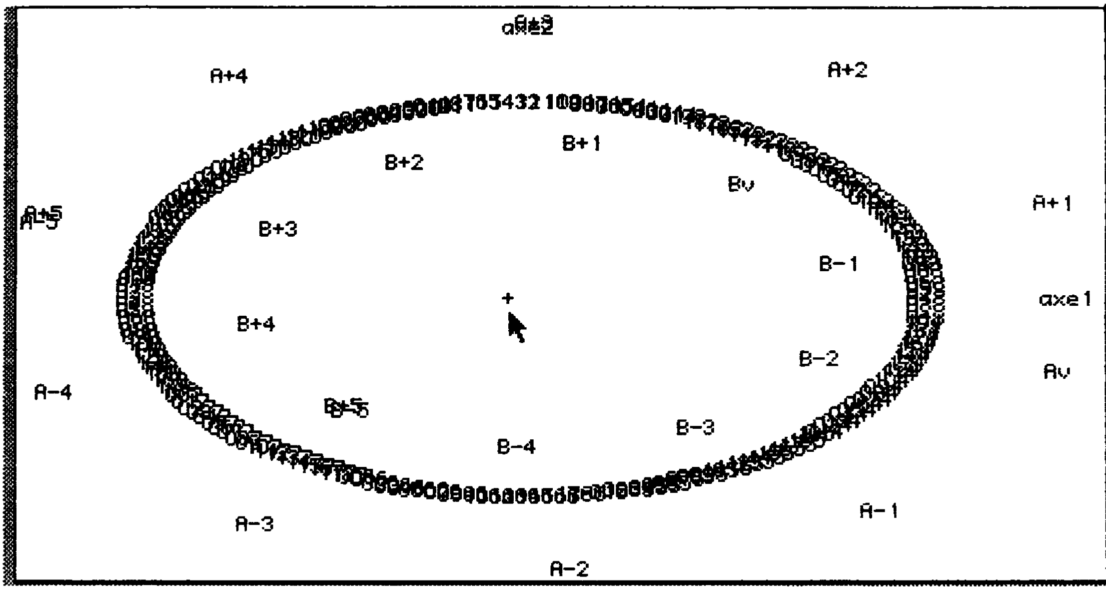
TML Pascal

```
le fichier des donnees est eco:ser
ce nom est il confirme oui(O) ou non(N) O
fichier cree par calculs en fonction du temps (cf param)
nombre de series decalees a creer avant et apres   Av = 5
ce nombre est il confirme O ou N O
nombre de series decalees a creer avant et apres   Bv = 5
ce nombre est il confirme O ou N O
```

TML Pascal

```
 3309  3622  3515  3030  2352  1741  1435  1558  2073  2793  3451
 6166  5722  5006  4297  3874  3911  4407  5188  5971  6468  6497
189
 3622  3515  3030  2352  1741  1435  1558  2073  2793  3451  3802
 5722  5006  4297  3874  3911  4407  5188  5971  6468  6497  6051
190
 3515  3030  2352  1741  1435  1558  2073  2793  3451  3802  3714
 5006  4297  3874  3911  4407  5188  5971  6468  6497  6051  5307
191
 3030  2352  1741  1435  1558  2073  2793  3451  3802  3714  3221
 4297  3874  3911  4407  5188  5971  6468  6497  6051  5307  4553
192
 2352  1741  1435  1558  2073  2793  3451  3802  3714  3221  2512
 3874  3911  4407  5188  5971  6468  6497  6051  5307  4553  4087
193
 1741  1435  1558  2073  2793  3451  3802  3714  3221  2512  1858
 3911  4407  5188  5971  6468  6497  6051  5307  4553  4087  4097
194
 1435  1558  2073  2793  3451  3802  3714  3221  2512  1858  1515
 4407  5188  5971  6468  6497  6051  5307  4553  4087  4097  4596
195
 1558  2073  2793  3451  3802  3714  3221  2512  1858  1515  1620
 5188  5971  6468  6497  6051  5307  4553  4087  4097  4596  5409
faut il traiter un autre tableau O ou N N
```

détails valent d'être remarqués, parce qu'ils préparent à comprendre l'exemple du §3.2, et ensuite à déceler des cycles dans des données réelles.



3.2. Exemple de spirale tracée sur un cône

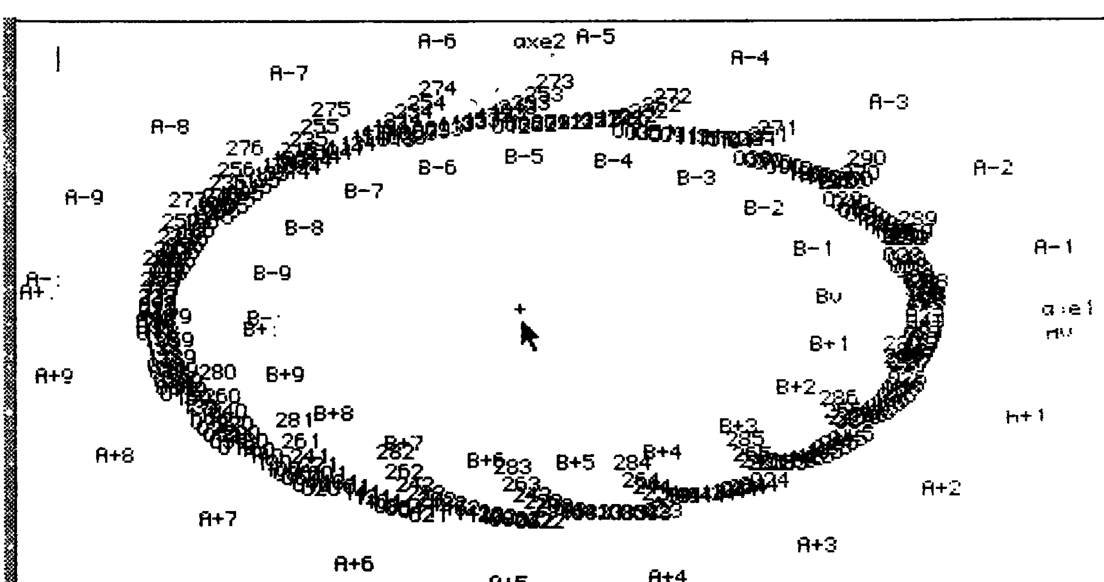
Pour décrire cet exemple, nous ne reproduisons pas le dialogue de création déjà expliqué au §3.1, mais seulement le fichier des parmètres, en y ajoutant les taux et valeurs propres pris sur le listage d'analyse des correspondances. Nous commenterons ce texte avant de considérer les graphiques plans.

```
ex:serparam
fichier cree par calculs en fonction du temps avec decalage
parametres du fichier ex:serzz
cart =  300  :  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     32      40     -20     -8       23
 Bv      2     32      20     -15     -8       23
A chacune des variables Av et Bv on a associe 10 series
decalees en avance et 10 en retard et on a obtenu ce qui suit
trace   :   5.447e-2
rang    :       1      2      3      4      5
lambda  :     282    260      3      0      0     e-4
taux    :    5177   4767     55      1      1     e-4
cumul   :    5177   9944   9999   9999  10000     e-4
```

3.2.1. Les paramètres du modèle

Pour plus de netteté dans le tracé de la courbe des t, on a posé cart=300, la période des fonctions trigonométriques a été doublée (10Tc=32) et on a créé 10 séries décalées avant et après chacune des variables de base, qui, ici encore ne sont que 2, Av et Bv. Les modalités A+10 et A-10 se trouvent notées A+: et A-:, parce que dans le code des caractères le signe ":" vient immédiatement après le chiffre "9". L'utilisation simultanée d'un facteur linéaire (1+K*t) et d'une exponentielle exp(t/Te) n'apporte pas de changement essentiel relativement au cas de la fin du §2.4 . Les fontions décalées sont, ici comme là, dans un espace de dimension 6: il a seulement 5 facteurs non triviaux. Le terme exponentiel a été choisi pour compenser à peu près la décroissance en fonction de t qui résulte d'un coefficient K négatif.

Mais ici, cette ligne se compose de petits arcs séparés dont il faut expliquer la genèse.



3.2.2 .Le tableau des valeurs propres

Les valeurs propres 1et 2 sont prédominantes ; 13 ne représente que 0,55% de la trace ; et 14, 15 font environ 1/10000. Il vaudra cependant la peine de représenter graphiquement les facteurs 3 à 5 qui produisent une image conforme à ce qu'annonçait le § 2.4.

3.2.3. Le plan des axes 1 et 2

La figure ressemble à celle du § 3.1 : les deux variables de base, avec leurs modalités décalées, dessinent deux ellipses, entre lesquelles se place la ligne des instants t.

On a dit au § 4.1 qu'à chaque cycle 10 points s'affichaient, en formant une sorte de polygone régulier inscrit dans une ellipse. Il en est de même ici ; mais l'ellipse se déforme de cycle en cycle sous l'effet des termes complémentaires qui perturbent les facteurs, ainsi qu'on l'a expliqué au § 2.4. On peut suivre cette déformation avec assez de précision, au moins pour les deux derniers cycles (de 251 à 270; et de 271 à 290), sur un graphique agrandi du fait de la suppression du nuage des modalités. (plan 2,1).

3.2.4. Le plan des axes 4 et 5

On a la spirale annoncée au § 2.2 ; de rayon modéré initialement, (à partir de t=010) les spires se resserrent au tour du centre; puis s'agrandissent jusquà dépasser largement leur taille initiale. Il résulte de ce mouvement que la zone centrale est illisible : on y reconnait seulement des sortes de rayons ; tandis qu'à la périphérie on peut aisément suivre les deux derniers cycles, de 251 à 290.

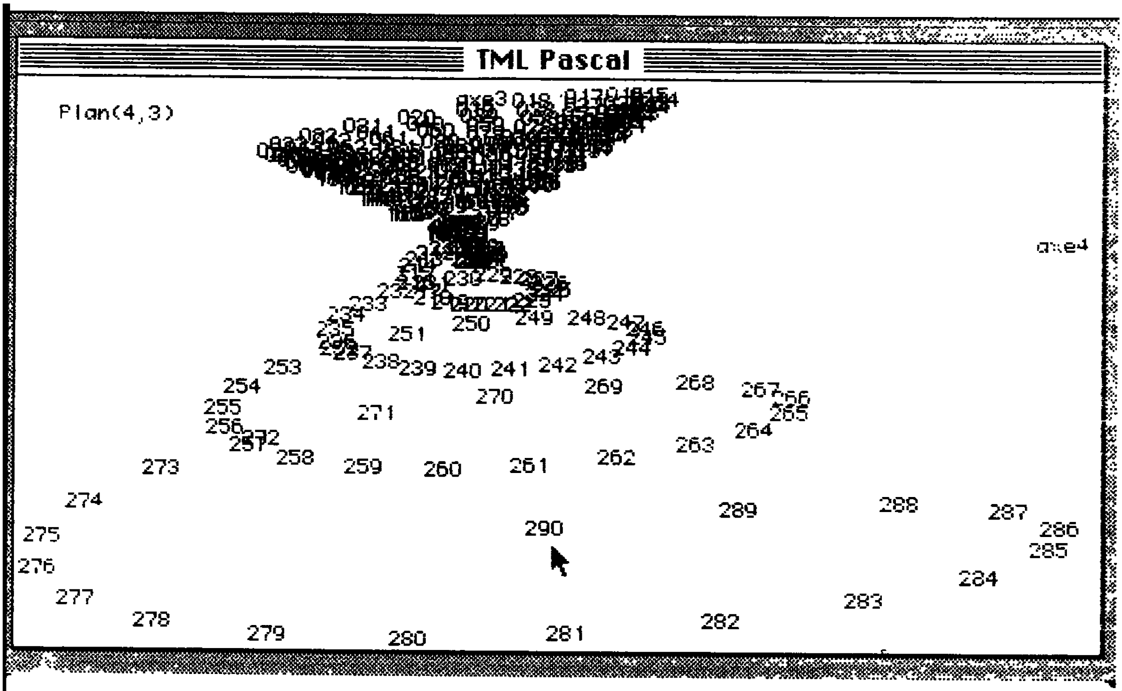
3.2.5. L'espace des axes 3, 4 et 5

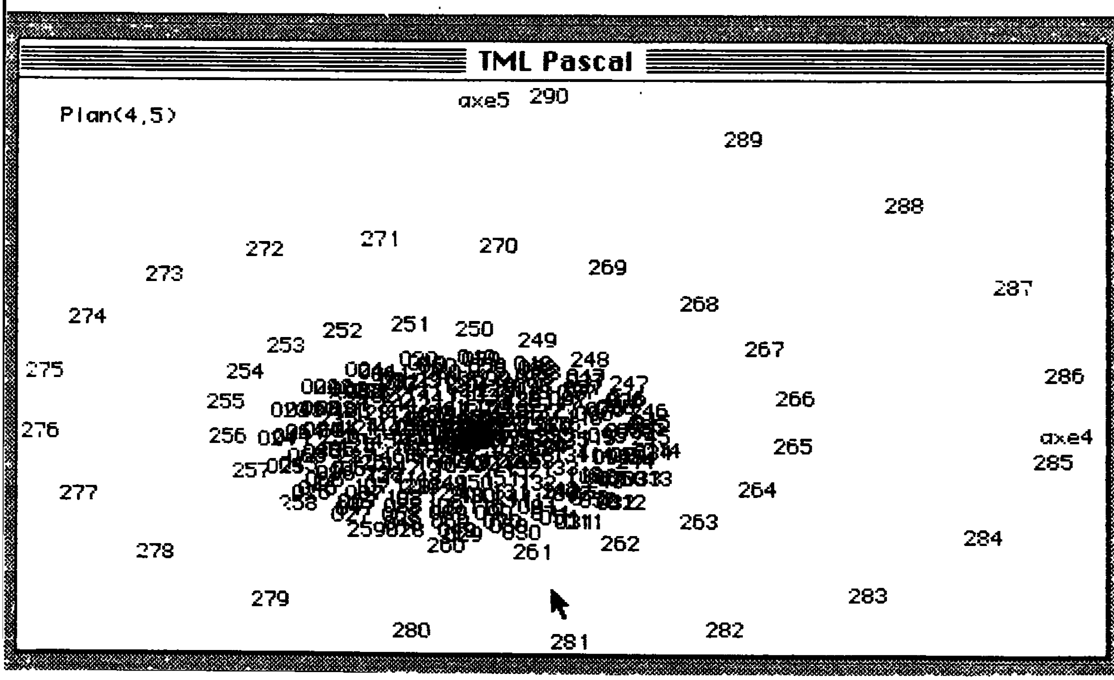
Si l'on regarde simultanément le plan (4,5) et, au dessus de celui-ci, le plan (4,3) on a une épure de cet espace : (4,5) étant la projection sur le plan horizontal et (4,3) la vue frontale ; avec pour axe vertical l'axe 3. Le dessin d'une spirale sur un cône se comprend bien sur cette épure ; même s'il y manque la cinématique de l'affichage rapide des points à l'écran de t=010 à t=290. Les premières spires, du côté (F3>0) sont très resserrées en vue frontale ; mais les dernières spires (F3<0) se voient parfaitement.

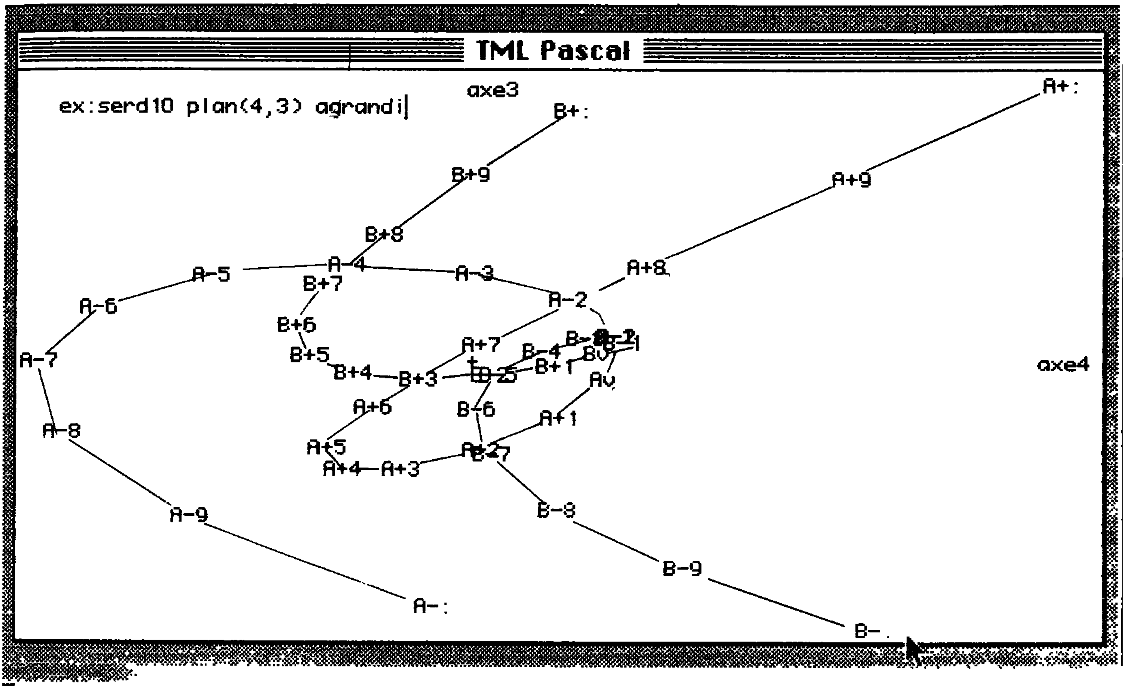
4. Conclusion

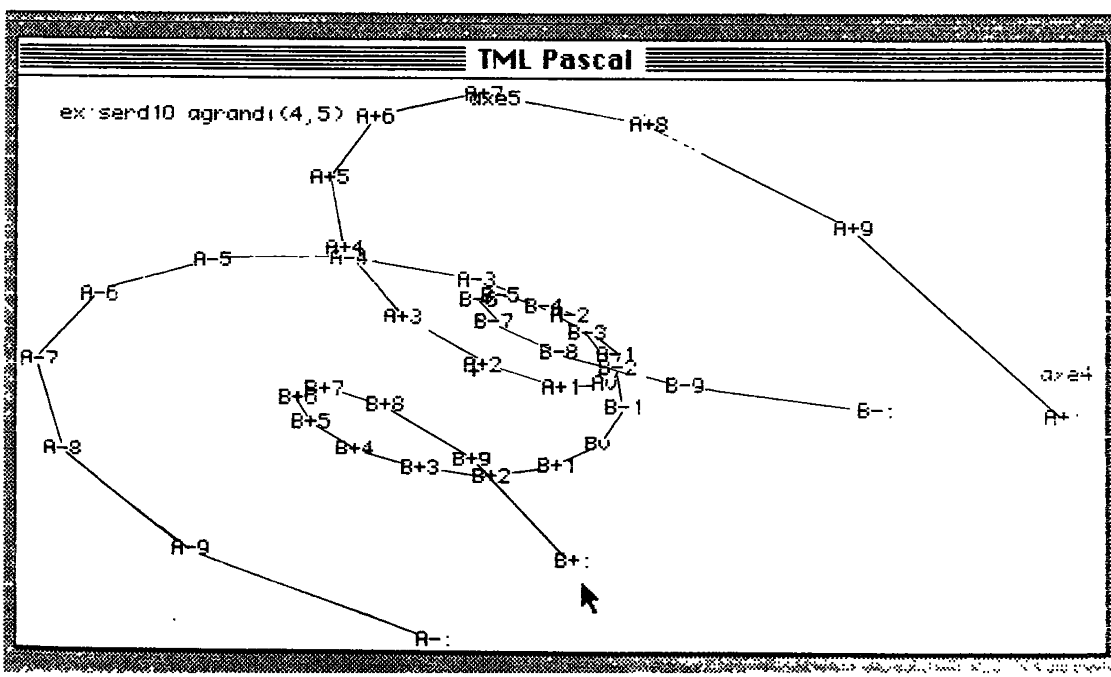
Plus encore que le plaisir de retrouver, au terme d'un calcul de simulation et d'une analyse de correspondance, le dessin en perspective d'une structure géométrique, nous cherchons ici une préparation à l'analyse des séries chronologiques réelles dans toute leur richesse.

Remerciements : L'auteur exprime toute sa gratitude aux Services Français des Relations Culturelles, qui lui ont permis de séjourner à Paris pour avancer dans ses recherches.


TML Pascal
Plan(4,3)
axe3
axe4


TML Pascal
Plan(4,5)
axe5
axe4


TML Pascal
ex:serd10 plan(4,3) agrandi
axe3
axe4
B+:
B+9
B+8
B+7
B+6
B+5
B+4
B+3
B+1
Bv
B-4
B-6
B-7
B-8
B-9
B-.
A+:
A+9
A+8
A+7
A+6
A+5
A+4
A+3
A+2
A+1
Av
A-2
A-3
A-4
A-5
A-6
A-7
A-8
A-9
A-:


TML Pascal
ex:serd10 agrandi(4,5)
axe5
axe4
A+7
A+6
A+8
A+5
A+9
A+4
A+3
A+2
A+1
Av
A+·
A-3
A-4
A-5
A-6
A-7
A-8
A-9
A-:
B-5
B-4
B-2
B-3
B-7
B-8
B-1
B-9
B-:
B+7
B+8
B+5
B+4
B+3
B+9
B+2
B+1
Bv
B+: