



HAL
open science

Modèles bilinéaires et polynomiaux de séries chronologiques : étude probabiliste et analyse statistique

Dominique Guegan

► **To cite this version:**

Dominique Guegan. Modèles bilinéaires et polynomiaux de séries chronologiques : étude probabiliste et analyse statistique. Modélisation et simulation. Université Joseph-Fourier - Grenoble I, 1988. tel-00330671

HAL Id: tel-00330671

<https://theses.hal.science/tel-00330671>

Submitted on 15 Oct 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THESE

présentée par

GUEGAN Dominique

pour obtenir

LE GRADE DE DOCTEUR D'ETAT ES SCIENCES MATHÉMATIQUES

DE L'UNIVERSITÉ JOSEPH FOURIER-GRENOBLE I

**MODELES BILINEAIRES ET POLYNOMIAUX DE SERIES
CHRONOLOGIQUES :**

ETUDE PROBABILISTE ET ANALYSE STATISTIQUE

Date de soutenance : 3 JUIN 1988

Composition du Jury : B. VAN CUTSEM Président
D. DACUNHA CASTELLE
A. LEBRETON
D.T. PHAM
H. TONG
J.C. WILLEMS

C'est à travers l'expérience de la vie que l'on se forge une idée de ce que l'on est. Une thèse est bien peu de choses au regard de tout ce qu'il a fallu parcourir et découvrir pour l'atteindre. Au delà de la construction mathématique que l'on tente de présenter se trouve l'être humain et le pourquoi de cette longue traversée souvent solitaire. Pourtant si j'ai commencé tard la recherche, je ne regrette pas ces années passées en Afrique et au Niger.

Je tiens à remercier ceux qui ont accepté de participer à mon jury et qui par leurs travaux scientifiques restent pour moi un exemple constant : Didier DACUNHA CASTELLE, Alain LEBRETON, Tuan D. PHAM, Howell TONG, Jans C. WILLEMS et Bernard VAN CUTSEM qui m'a fait l'amitié de présider ce jury.

C'est en revenant du Niger, en 1976, que j'ai trouvé un accueil chaleureux dans le laboratoire de Statistiques d'Orsay où j'ai entrepris quelques années plus tard cette recherche : aussi je tiens avant tout à remercier Didier DACUNHA CASTELLE qui m'a permis d'entreprendre ce travail et dont les remarques m'ont toujours été d'une grande aide.

Mais cette thèse n'aurait pas pu aboutir sans le concours de l'équipe TIM 3 de GRENOBLE au sein de laquelle je soutiens : je remercie ici Alain LEBRETON et Bernard VAN CUTSEM qui ont eu à coeur d'aplanir bon nombre de difficultés matérielles, et qui ont su être proches aux bons moments.

C'est au sein de cette équipe que j'ai rencontré Tuan PHAM avec qui j'ai cheminé depuis deux ans dans le monde des modèles bilinéaires. Il a été pour moi celui qui par son dynamisme scientifique, ses compétences certaines m'a fait le plus progresser et, peut-être aussi découvrir comment une proche collaboration pouvait être source de réussite et d'amitié. Il reste un pilier essentiel pour la soutenance de cette thèse. Je le remercie de son aide constante, de sa disponibilité sans faille, de sa gentillesse et du soin qu'il a mis à relire le manuscrit de ma thèse.

Je remercie, Howell TONG avec qui j'ai pu avoir des échanges très constructifs, Jans WILLEMS qui par l'orientation même de ses recherches m'a amené à envisager des points de vue qui m'ont permis de progresser dans ma recherche.

Je remercie Bernard PRUM sans qui cette recherche n'aurait jamais vraiment commencé, il m'a toujours remis le pied à l'étrier quand il le fallait et n'a jamais cessé de

répondre avec clarté, compétence et gentillesse aux questions que je posais. Dominique PICARD pour les longues heures de discussions que j'ai pu avoir avec elle, marque de sa curiosité scientifique incessante, T. SUBBA RAO qui fut l'un des premiers à étudier les modèles bilinéaires et qui lors de nos différentes rencontres m'a permis de préciser mon travail de recherche.

Il y a aussi tous ceux qui ont fait un bout de chemin avec moi dans l'univers si chaotiques des modèles non linéaires, ou qui par leur soutien moral ou matériel ont souvent empêché que le découragement ne se fasse sentir. Je tiens à les remercier: Patrice ASSOUAD, Gerard COLLOMB, Jean DESHAYES, Catherine HUBER, Xavier GUYON, Patrick JAKUBOWICK, Francis LALOE, Henri LEREDDE, Georges OPPENHEIM, Elisabeth de TURKEIM.

Mais je ne voudrais pas oublier ici tous mes collègues de Villetaneuse qui représentent pour moi une famille au sein de laquelle je me sens bien et qui n'ont pas manqué, en de maintes occasions, de me manifester leur soutien et leur amitié. Je les remercie tous.

Je suis particulièrement reconnaissante à Catherine SIMON pour sa disponibilité et sa gentillesse. Je la remercie pour tout le travail qu'elle a fourni pour cette frappe, ainsi que Madame CALLAREC, pour le soin et la minutie qu'elle a mis à reproduire cette thèse.

Je ne peux pas oublier tous mes amis du NIGER qui n'ont pas cessé de me soutenir et de m'encourager dans l'aboutissement de ce travail : je les en remercie.

Ma famille fut pour moi d'un secours inestimable et sous la pudeur de certains j'ai toujours su découvrir une vraie confiance : j'espère que Claude trouvera ici matière à réflexion. Jean n'a pas été épargné pendant ces longs mois, je le remercie de ses encouragements constants et de sa confiance. Quand à Jean Stéphane et Sahia Magalie je trouve qu'ils ont fait preuve de beaucoup de patience et d'amour. Je les remercie

Dominique Guegan le 15/4/88

PRESENTATION DE LA THESE

L'intérêt et le goût pour les modèles bilinéaires nous sont venus alors que nous découvrons les séries chronologiques linéaires et que nous avons eu l'occasion d'apprécier les limites de leur champ d'application.

En 1979 et en 1981 deux séminaires furent organisés par D.DACUNHA-CASTELLE et R. AZENCOTT à l'Ecole Normale Supérieure, dans le but d'étudier dans un premier temps les séries chronologiques scalaires, leur cadre hilbertien, les techniques d'identification, les problèmes de tests, d'estimation, de prévisions, leur champ d'application, et dans un deuxième temps les séries vectorielles, la représentation markovienne et ses divers prolongements. Nous avons pour notre part présenté un premier séminaire (1979) sur les tests d'adéquation de modèles : cet exposé mettait en évidence les différents tests utilisés pour différencier des modèles ou pour valider le choix de modèles ajustés. Nous avons ensuite fait un exposé introductif, lors du deuxième séminaire, sur les séries chronologiques vectorielles et la démarche suivie par H. AKAIKE pour obtenir la représentation markovienne de ces séries (1981). C'est à cette époque que nous avons eu l'occasion de traiter des données météorologiques. Nous avons alors tenté d'utiliser des méthodes du type BOX JENKINS en vue de leur identification. Cela s'est vite avéré une gageure dans la mesure où ces données très irrégulières constituées de pics suivis de longues périodes calmes ne présentaient pas l'aspect "classique" des séries linéaires que nous avons eu l'occasion d'étudier ces deux dernières années.

Nous avons alors eu la chance de suivre une série de conférences traitant de la théorie des systèmes stochastiques, du problème de l'identification, des méthodes de contrôle et de

filtrage récursif linéaire et non linéaire, à partir du point de vue des automaticiens. Ces conférences organisées aux ARCS en 1980 par M. HAZEWINDEL et J.C. WILLEMS nous ont permis en particulier de découvrir les modèles bilinéaires. Cela a été l'occasion de nombreuses discussions sur le sujet avec des chercheurs tels E.J. HANNAN, M.DEISTLER, J.C. WILLEMS, P.CAINES, etc...

A la suite des travaux de C.W.J GRANGER et A.P. ANDERSEN (1978), et T. SUBBA RAO (1981), nous nous sommes intéressée à la forme stochastique des modèles bilinéaires. Comme d'autres chercheurs à partir de cette approche directe des modèles, nous en avons étudié les propriétés probabilistes, essentiellement à partir de cas particuliers à cause de la complexité structurale des modèles en question.

Nous avons tout d'abord étudié l'existence, l'inversibilité, l'expression de la fonction de vraisemblance d'un modèle superdiagonal à un terme (1981, p. 244) ; puis afin de pallier aux limites du cadre hilbertien pour ce type de modèles, nous avons introduit les chaos de Wiener comme cadre d'étude plus adéquat permettant de prendre en compte la structure complète des modèles considérés (1983, p. 248). Ceci nous a permis en étudiant les processus réguliers et singuliers de fournir une première typologie des modèles non linéaires. Nous avons alors abordé le problème de l'ergodicité pour des cas particuliers (1983, p.252, 1984) problème aussi abordé par H. TONG (1981) sur un cas particulier. C'est à cette époque que des théorèmes plus généraux sur l'existence des modèles, la fonction de covariance et les moments d'ordre supérieur ont vu le jour (M. BHASKARA RAO, T. SUBBA RAO et A.M. WALKER, 1983), et que l'approche spectrale a été abordée (T. SUBBA RAO et M.M. GABR, 1984). Mais l'étude statistique de ces modèles était encore peu développée et souvent limitée à des considérations générales qui ne pouvaient aboutir par manque de résultats essentiels, comme l'inversibilité, ou l'existence de certains moments. Néanmoins T. PHAM DINH et L.TAT TRAN (1981) fournissent

pour un modèle particulier, des estimateurs consistants des paramètres par la méthode des moindres carrés, et nous mêmes, estimons par la méthode des moments les paramètres de certains modèles (1984, p. 256).

L'année 1985 représente un tournant essentiel dans l'étude des modèles bilinéaires et une nouvelle orientation se dessine. En effet, T. PHAM DINH établit la représentation markovienne des modèles bilinéaires superdiagonaux. Il fournit des conditions d'existence, d'unicité et de minimalité de la représentation. Cette représentation lui permet aussi d'obtenir l'expression de la fonction de covariance et des moments d'ordre supérieur des modèles bilinéaires correspondants. Nous montrons alors que cette représentation s'étend difficilement au cas des modèles sous diagonaux, et nous introduisons la forme stochastique des modèles polynomiaux en les entrées, affines en l'état ainsi définis en 1979 par E.D. SONTAG, (1986, p. 277). T. PHAM DINH montrera qu'ils sont la représentation markovienne de l'ensemble des modèles bilinéaires. Nous en donnons les conditions d'existence et d'unicité et nous fournissons l'expression de la fonction de covariance et des moments d'ordre supérieur (1987, p. 281). Parallèlement T. PHAM DINH (1986) établit des conditions d'ergodicité géométrique pour ces modèles, résolvant ainsi, grâce aux techniques propres aux chaînes de Markov, un problème que l'approche directe n'avait pas permis d'atteindre pour tous les modèles. Puis nous établissons avec T. PHAM DINH des conditions de minimalité d'une telle représentation, (1987, p. 301). C'est grâce à l'approche markovienne que l'étude statistique des modèles bilinéaires va connaître quelques développements récents. En effet, nous obtenons avec T. PHAM DINH l'inversibilité de la classe des modèles diagonaux (1987, p. 305), ce qui nous permet d'obtenir, pour l'ensemble des paramètres des modèles appartenant à cette classe, des estimateurs consistants par la méthode des moindres carrés (1987, p. 305).

Certes l'étude des modèles bilinéaires est loin d'être achevée. De nombreux problèmes restent ouverts, mais ces deux dernières années l'orientation prise a permis de résoudre des problèmes restés jusqu'à présent sans solution. Néanmoins on peut se demander s'il est possible de résoudre le problème de l'identification des modèles bilinéaires comme ce fut le cas pour les séries chronologiques linéaires, dans la mesure où la classe des modèles considérés est très riche et très complexe. Par contre la théorie des tests qui jusqu'à présent a été peu étudiée, doit permettre d'aboutir assez rapidement à quelques résultats : outre les tests de linéarité pour lesquels quelques procédures ont été proposées, on a besoin de tests portant sur le choix des modèles (la classe et l'ordre), ainsi que de tests de validité pour les modèles déjà ajustés.

Le document que nous présentons ici permet de faire le point sur l'état actuel des recherches concernant les modèles bilinéaires, ainsi que sur les différents points non encore démontrés et indispensables pour obtenir une meilleure connaissance de ces modèles. Ainsi outre les résultats de l'auteur dont les articles correspondants se trouvent à la fin de la thèse aux pages indiquées, on trouve d'autres résultats présentés sans preuves. Celles-ci pourront être consultées dans les articles cités en référence.

La thèse est composée de sept parties, chacune d'elle comprend une introduction. La bibliographie est générale et se trouve à la fin du septième chapitre.

Le chapitre I présente d'un point de vue historique quelques uns des processus non linéaires étudiés dans la littérature et en ce qui concerne les modèles bilinéaires précise les points de vue abordés, les outils utilisés, les concepts étudiés, les problèmes ouverts.

Le chapitre II permet de découvrir les modèles bilinéaires d'un point de vue expérimental. En effet, avant d'en faire une étude théorique nous avons voulu donner une idée "physique" de ces modèles. On présente pour cela un certain nombre de simulations correspondant à des modèles simples et permettant de mettre en valeur les principales caractéristiques de ces modèles.

Le chapitre III se place dans une perspective assez générale et développe outils et concepts nécessaires pour l'analyse des modèles non linéaires : chaos de Wiener, inversibilité, ergodicité, régularité, singularité, polyspectre, processus d'innovation .

Le chapitre IV, après avoir donné une première représentation des modèles bilinéaires à partir des chaos de Wiener, expose dans une revue d'ensemble les premiers travaux concernant la structure probabiliste du modèle bilinéaire tel qu'il fut introduit par C.W.J. GRANGER et A.P. ANDERSEN.

Le chapitre V traite des représentations markoviennes des modèles bilinéaires et de leurs propriétés : existence, unicité, minimalité, inversibilité. Elles sont utilisées pour démontrer certaines propriétés des modèles bilinéaires comme l'ergodicité géométrique, le calcul de la fonction de covariance et des moments d'ordre supérieur.

Le chapitre VI aborde l'inférence statistique des modèles bilinéaires : on y traite de quelques tests de linéarité ainsi que de l'estimation des paramètres des modèles à partir de différentes méthodes.

Le chapitre VII permet de faire le point sur les connaissances actuelles et propose un certain nombre de voies de recherche.

TABLE DES MATIERES

CHAPITRE I INTRODUCTION

- I - APERCU DES DIFFERENTES CLASSES DE MODELES NON LINEAIRES
 CONSIDEREES DANS LA LITTERATURE
 - 1. Les modèles autorégressifs non linéaires
 - 2. Les modèles autorégressifs avec coefficients aléatoires
 - 3. Les modèles "S.D.M"
 - 4. Les modèles bilinéaires

- II - CONCEPTS

- III - DIFFERENTES APPROCHES POUR L'ETUDE DES MODELES BILINEAIRES
 - 1. Les chaos de Wiener
 - 2. L'approche directe
 - 3. La représentation markovienne
 - 4. La représentation affine d'état

- IV - INFERENCE STATISTIQUE
 - 1. Estimation des paramètres
 - 2. Tests de linéarité

- V - PROBLEMES OUVERTS CONCERNANT LES MODELES BILINEAIRES

CHAPITRE II ETUDE EXPERIMENTALE DES MODELES BILINEAIRES

I - INTRODUCTION

1. Modèles présentant des "périodicités" (faible et forte composante bilinéaire).
2. Modèles apériodiques (un ou deux termes autorégressifs, composante bilinéaire faible et forte).
3. Modèles apériodiques (avec un terme autorégressif, un terme moyenne mobile, une composante bilinéaire faible et forte).

II - ETUDE PRELIMINAIRE

1. Distributions Marginales.
 - a - Distributions univariées.
 - b - Distributions bivariées.
2. Comportements des trajectoires.

III - ETUDE DU PHENOMENE EXPLOSIF.

IV - CONCLUSION

CHAPITRE III OUTILS ET CONCEPTS POUR L'ANALYSE DE MODELES NON
LINEAIRES

I - INTRODUCTION

II - CHAOS DE WIENER. RAPPELS ET NOTATIONS

1. Produits tensoriels
2. Produits tensoriels symétriques
3. Chaos de Wiener

III - CONCEPTS FONDAMENTAUX

1. Inversibilité
2. Ergodicité et ergodicité géométrique
3. Régularité et Singularité

IV - POLYSPECTRE

V - ESPACES D'INNOVATIONS DE PROCESSUS

CHAPITRE IV ETUDE PROBABILISTE DU MODELE BILINEAIRE
REVUE D'ENSEMBLE DES PREMIERS TRAVAUX

- I - INTRODUCTION
- II - DECOMPOSITION FORMELLE DE PROCESSUS BILINEAIRES SUR LES
CHAOS DE WIENER
1. Décomposition de Volterra du modèle bilinéaire général
 2. Exemples
 3. Application aux transformations non linéaires
instantanées de processus linéaires
- III - THEOREMES D'EXISTENCE POUR DES MODELES BILINEAIRES
SUPERDIAGONAUX
1. Expression vectorielle des modèles bilinéaires
superdiagonaux
 2. Théorème d'existence de processus bilinéaires
superdiagonaux
 3. Théorèmes d'existence pour des cas particuliers de
modèles bilinéaires
 - a. modèle bilinéaire $BL(p,0,p,1)$
 - b. modèle bilinéaire diagonal à un terme $BL(1,0,1,1)$
 - c. modèles bilinéaires superdiagonaux sans partie
autorégressive et sans partie moyenne mobile
 $BL(0,0,p,q)$
- IV - INVERSIBILITE DES MODELES BILINEAIRES
1. Inversibilité au sens de GRANGER ANDERSEN
 2. Inversibilité au sens fort
- V - STATIONNARITE AU SECOND ORDRE ET FONCTION DE COVARIANCE
- VI - MOMENTS D'ORDRE SUPERIEUR DES PROCESSUS BILINEAIRES

CHAPITRE V REPRESENTATIONS MARKOVIENNES ET APPLICATIONS**I - INTRODUCTION****II - CONSTRUCTION D'UNE REPRESENTATION MARKOVIENNE****A. Représentation Markovienne des modèles bilinéaires**

1. Construction d'une représentation markovienne bilinéaire
2. Application aux modèles bilinéaires superdiagonaux
3. Représentation ℓ -markovienne des modèles sous diagonaux

B. Modèles affines en l'état polynomiaux en les entrées et modèles autorégressifs généralisés à coefficients aléatoires**III - CONDITION D'EXISTENCE DES REPRESENTATIONS MARKOVIENNES****IV - PROPRIETES DES REPRESENTATIONS MARKOVIENNES**

1. Minimalité des représentations Markoviennes
2. Inversibilité d'une représentation Markovienne

V - APPLICATION DE LA REPRESENTATION MARKOVIENNE A L'ETUDE DE MODELES BILINEAIRES

1. Ergodicité et ergodicité géométrique pour les représentations Markoviennes
2. Calcul des moments des représentations Markoviennes
 - a. Modèles superdiagonaux
 - b. Modèles polynomiaux en les entrées, affines en l'état
3. Application au modèle ARCH(p)
4. Prédiction

VI - CONCLUSION

CHAPITRE VI INFERENCE STATISTIQUE ET APPLICATIONS

I - INTRODUCTION

II - ESTIMATION DES PARAMETRES DES MODELES BILINEAIRES

1. Méthode générale d'estimation
2. Mise en oeuvre de la méthode d'estimation pour certains modèles bilinéaires
3. Estimation par la méthode des moments
4. Estimation des paramètres du modèle ARCH(p)

III - TESTS

1. Tests de linéarité
2. Tests d'adéquation de modèles

CHAPITRE VII PROBLEMES OUVERTS

I - INTRODUCTION

II - PROPRIETES PROBABILISTES DES MODELES BILINEAIRES

III - IDENTIFICATION DES MODELES BILINEAIRES

IV - PROLONGEMENTS STATISTIQUES

BIBLIOGRAPHIE GENERALE

ARTICLES DE L'AUTEUR PUBLIES SE TROUVANT EN APPENDICE (p.244-315)

- Etude d'un modèle non linéaire, le modèle superdiagonal d'ordre un. C.R.A.S. t.293. Série I. 95-98 (1981). (page 244).
- Cadre d'étude pour des modèles non linéaires. C.R.A.S. t.296. Série I. 167-170 (1983). (page 248).
- Une condition d'ergodicité pour des modèles à temps discret. C.R.A.S. t.297. Série I. 537-540 (1983). (page 252).
- Tests de modèles non linéaires. Proceedings of the 3rd Franco-Belgian Meeting of Statisticians, november 1982. Publications de St-Louis. 45-65 (1984). (page 256).
- Représentation ℓ -markovienne et existence d'une représentation affine en l'état des modèles bilinéaires. C.R.A.S. t. 302. Série I 279-282 (1986). (page 277).
- Different representations for bilinear models. J.T.S.A. Vol. 8. n°4. 389-408. (1987). (page 281).
- Minimalité et inversibilité des modèles bilinéaires à temps discret. C.R.A.S. t.304. Série I. 159-162 (1987). (en collaboration avec T. PHAM DINH). (page 299).
- A note on the estimation of the parameters of the diagonal bilinear model by the least squares method. (à paraître dans Scand. Journ. of Stat. Theor. and Appl.). (en collaboration avec T. PHAM DINH). (page 305).

CHAPITRE I

INTRODUCTION

I - APERCU DES DIFFERENTES CLASSES DE MODELES NON LINEAIRES CONSIDEREES DANS LA LITTERATURE

Jusqu'à une époque récente, la plus grande partie des travaux traitant de l'analyse des séries chronologiques concernait essentiellement les modèles linéaires, et, implicitement ou explicitement, les modèles gaussiens. La théorie des modèles linéaires comme les modèles autorégressifs, les modèles moyenne mobile ou les modèles autorégressifs - moyenne mobile (ARMA) a été bien développée et d'excellents travaux la concernant ont vu le jour.

Le livre de G.E.P. BOX et G.M. JENKINS (1970), et les travaux de H. AKAIKE (1974 a) pour la classification structurale par exemple, suffisent à démontrer la maturité atteinte dans la modélisation des séries chronologiques linéaires. La classe de modèles qu'ils ont étudiée a eu un très grand succès dans de nombreux domaines. Néanmoins ces modèles sont linéaires, or dans de nombreuses disciplines, on rencontre des processus générés de manière non linéaire ou des processus présentant des relations non linéaires entre les variables. D'où la nécessité d'étendre la classe des modèles linéaires.

Dès que la contrainte linéaire est levée, la construction d'un modèle non linéaire nécessite quelques précautions. Il va donc s'agir de construire des modèles qui, comme dans le cas linéaire,

ne doivent pas conduire à des solutions explosives, permettent de faire des prévisions fiables, collent à la réalité en essayant de prendre en compte les disparités que le modèle linéaire exclut.

Un premier point de vue va consister à ne considérer que des modèles dont la structure sous jacente ne sera plus linéaire. Un autre point de vue est celui qui prend en compte des modèles évolutifs ; ce sont des modèles non stationnaires pour lesquels la moyenne et les covariances dépendent du temps, situations assez bien adaptées à des problèmes issus de l'économie, de l'hydrologie, de la météorologie. Nous nous intéressons dans le présent travail au premier point de vue. Une façon d'approcher le concept de non linéarité dans un modèle est de considérer la notion de prédiction pour le modèle dont on dispose. En effet, étant donné un ensemble d'observations, dans le cas des modèles linéaires, le prédicteur linéaire des moindres carrés est aussi le prédicteur des moindres carrés, cf. E.J. HANNAN (1979), situation que l'on ne retrouve pas dans le cas des modèles non linéaires. La notion de prédicteur étant essentielle pour l'étude d'un modèle, on sera alors amené à définir d'autres prédicteurs, (Cf. P. MASANI et N. WIENER, 1959, M. ROSENBLATT, 1979 et L.A. SHEPP, D. SLEPIAN et A.D. WYNER, 1980.)

Historiquement, c'est l'étude des séries économiques qui a conduit à considérer des modèles non linéaires issus des séries chronologiques. Une façon naturelle d'introduire la non linéarité a consisté à prendre une généralisation du modèle autorégressif d'ordre 1 sous forme de modèles polynômiaux du type suivant :

$$X(t) = \sum_i \sum_j a_{ij} X(t-i)^j + \epsilon(t) .$$

Malheureusement on tombe vite sur des modèles explosifs. Les conditions à imposer pour éviter ce genre de situations sont souvent contraignantes dans le cas des modèles polynômiaux, aussi l'étude de ces modèles n'a-t-elle pas été poursuivie, et d'autres orientations ont été envisagées.

1. Les modèles autorégressifs non linéaires

En 1978 D.A. JONES introduisit la classe de modèles suivante :

$$X(t) = \varphi(X(t-1), X(t-2), \dots) + \epsilon(t)$$

où φ est une certaine fonction non linéaire en les données X et où $\epsilon(t)$ est un bruit blanc. L'idée était alors d'utiliser des modèles un peu plus généraux que les modèles autorégressifs, mais ayant moins de paramètres à estimer. Certaines propriétés de ces modèles étudiés en tant que processus de Markov, (irréductibilité, récurrence, ergodicité géométrique) furent développées par P. DOUKHAN et M. GHINDES (1980). Les mêmes auteurs développèrent une théorie de l'estimation des paramètres de ces modèles, et en 1983 K.K. AASE a fourni un schéma récursif pour estimer les paramètres de cette classe de modèles.

Des domaines comme la biologie, l'hydrologie, la météorologie qui présentent des phénomènes proches de la théorie des oscillations non linéaires ont conduit à envisager des modèles autorégressifs non linéaires particuliers permettant de prendre en compte des comportements de cycle limite.

Une première approche a été menée par T. OZAKI (1980) et V. HAGGAN et T. OZAKI (1981). Ils ont introduit les modèles exponentiels autorégressifs : c'est une classe de modèles qui exhibent certains traits bien connus de la théorie des vibrations non linéaires, c'est un modèle univarié défini de la manière suivante :

$$X(t) = \sum_{j=1}^k A_{t,j} X(t-j) + \epsilon(t)$$

où $\epsilon(t)$ est un bruit blanc gaussien, et où les coefficients $A_{t,j}$ sont une fonction exponentielle de $X^2(t-1)$:

$$A_{t,j} = a_j + b_j \exp(-c^2 X^2(t-1)) .$$

Une interprétation physique possible pour ces modèles est de considérer que la réponse de fréquence instantanée $\{A_{t,j}\}$ est contrôlée par le flux $X^2(t-1)$ représentant l'énergie au temps t .

Les auteurs précités ont donné une étude théorique de ces modèles (stationnarité, ergodicité). Leurs résultats ont été étendus en 1982 par T. OZAKI au cas où les coefficients $A_{t,j}$ sont des fonctions polynomiales type Hermite. T. OZAKI (1985 a) a utilisé ces modèles en vue d'estimer des modèles de stockage : type écoulement journalier des rivières ou hauteur des précipitations journalières.

En même temps, une autre approche fut menée par H. TONG et K.S. LIM (1980), voir aussi H. TONG (1983). Ils ont introduit les modèles autorégressifs à seuil. Il s'agit de modèles autorégressifs par morceaux avec des seuils d'excitation modélisés de la manière suivante : si $\{r_0, r_1, \dots, r_\ell\}$ représente un sous-ensemble linéairement ordonné de nombres réels tels que $r_0 < r_1 < \dots < r_\ell$ où r_0 et r_ℓ varient de $-\infty$ à $+\infty$ respectivement et si $R_i = [r_{i-1}, r_i]$ où R_1, R_2, \dots, R_ℓ définissent une partition de \mathbb{R} , alors le processus $X(t)$ est tel que :

$$X(t) = a_0^{(j)} + \sum_{i=1}^k a_i^{(j)} X(t-j) + \epsilon^{(j)}(t)$$

où $X(t-j) \in R_j$, $j = 1, \dots, \ell$, et où $\epsilon^{(j)}(t)$ est un bruit blanc avec $\epsilon^{(j)}(t)$ et $\epsilon^{(j')}(t)$ indépendants pour $j \neq j'$. Ces nombres r_1, \dots, r_ℓ sont appelés les seuils. Ces modèles se présentent donc comme un système d'équations aux différences linéaires par

morceaux. L'étude probabiliste de ces modèles est encore peu développée. Par le type de non linéarité qu'ils introduisent, ils permettent de modéliser des phénomènes avec transition de phase, ou avec changement d'état. K.S. LIM et H. TONG (1983) ont utilisé ces modèles pour étudier certaines séries écologiques. H. TONG, B. THANOON et G. GUDHINDSON (1985) ont utilisé ces modèles pour l'étude de l'écoulement des rivières alimentées par des précipitations d'eau, ou des chutes de neige, avec présence de glaciers.

2. Les modèles autorégressifs avec coefficients aléatoires

Dans une récente monographie D.F. NICHOLLS et B.G. QUINN (1982) ont étudié les modèles autorégressifs à coefficients aléatoires qui englobent comme cas particuliers les modèles autorégressifs vectoriels. Ces modèles markoviens ont pour représentation générale :

$$X(t) = A(t)X(t-1) + \zeta(t)$$

où $A(t)$ et $\zeta(t)$ sont respectivement une matrice et un vecteur polynomiaux en $\epsilon(t)$ indépendants l'un de l'autre, $\epsilon(t)$, $t \in \mathbb{Z}$ étant un bruit blanc. Les auteurs donnent des conditions d'existence de tels modèles stationnaires et proposent plusieurs méthodes d'estimation des paramètres. P.D. FEIGIN et R.L. TWEEDIE (1985) considèrent ces modèles à partir de leur caractère markovien et en étudient les propriétés d'ergodicité ainsi que l'existence de moments d'ordre arbitraire.

3. Les modèles "S.D.M."

M.B. PRIETLEY (1980) développe un autre modèle : le modèle dépendant d'un état ("State Dependant model" en anglais). L'idée

générale est de permettre aux coefficients d'un modèle ARMA de devenir dépendants d'un nombre fini de valeurs passées de X et d'un nombre fini de valeurs passées de ϵ , ce qu'il regroupe sous la dénomination de vecteur d'état, sans que cette notion recouvre la notion habituelle considérée dans une représentation d'état ; On peut d'ailleurs montrer que la propriété de minimalité en particulier n'intervient pas dans la définition. Ce modèle a été utilisé pour ajuster des séries de données et fournir des prévisions, (cf. V. HAGGAN, S.M. HERAVI et M.B. PRIESTLEY, 1984, et P.A. CARTWRIGHT, 1985).

On peut remarquer que la présentation générale de ces modèles proposée par M.B. PRIESTLEY permet de considérer certains modèles dont nous venons de parler comme des cas particuliers des modèles "S.D.M.", à savoir les modèles autorégressifs exponentiels et les modèles autorégressifs à seuil, sans que cela permette une étude plus approfondie de leurs propriétés respectives. D'autres modèles rentrent aussi dans ce cadre, ce sont les modèles bilinéaires dont nous allons maintenant parler.

4. Les modèles bilinéaires

D'un point de vue algébrique la linéarité d'un modèle ou d'un système se traduit par les équations où n'interviennent que des sommes ; la non linéarité va se traduire en introduisant dans les équations des multiplications entre les variables.

De ce point de vue, les premiers travaux sont dus à N. WIENER (1958) qui introduisit une classe très générale de modèles non linéaires : les développements en série de Volterra. Si $X(t)$ représente l'entrée du système et $Y(t)$ la sortie, de tels modèles ont pour expression

$$\begin{aligned}
Y(t) = & \sum_{i=0}^{\infty} a_i X(t-i) + \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} a_{ij} X(t-i) X(t-j) \\
& + \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} a_{ijk} X(t-i) X(t-j) X(t-k) + \dots
\end{aligned}$$

D.R. BRILLINGER (1970) a entrepris une analyse probabiliste de cette classe de modèles à l'aide des polyspectres. Le caractère très général de ces modèles n'a pas permis de les utiliser de manière très efficace sous cette forme. Néanmoins R.W. BROCKETT (1976) (dans le cas déterministe) et P. BOURDON (1982) (dans le cas stochastique) montrent que sur tout intervalle de temps fini on peut approcher un modèle non linéaire satisfaisant des conditions de régularité générale, avec un degré de précision arbitraire, par un modèle bilinéaire. Mais ce résultat n'est valable que sur un intervalle de temps fini, et l'approximation bilinéaire ne permet pas de reproduire certains effets, comme par exemple l'existence de cycles limites ou des phénomènes d'hystérésis.

Une classe de modèles, qui est en quelque sorte un cas particulier des développements de Volterra, a alors été considérée par les automaticiens : il s'agit des modèles bilinéaires qui ont la particularité d'être linéaires par rapport à chacune des variables (entrées et sorties) quand on considère l'autre constante. L'autre particularité de ces modèles est qu'ils se présentent comme une extension naturelle des modèles linéaires avec un nombre fini de paramètres. Enfin ils se sont bien adaptés à un certain nombre de situations pratiques. En effet en théorie du contrôle, ces modèles ont été largement étudiés et utilisés, que ce soit pour l'étude des populations dynamiques en biologie (R.R. MOLHER et R.E. RINK, 1969), pour établir des lois en chimie (R.R. MOLHER, 1970), en économie pour étudier l'accroissement des économies nationales, ou dans diverses branches de l'ingeneering.

La théorie concernant ces modèles est maintenant bien connue d'un point de vue déterministe. Les principales propriétés concernent la théorie de la réalisation, la décomposition canonique sur l'espace d'état, la controllabilité, l'observabilité, les propriétés d'invariance, la stabilité. En temps discret, on peut se rapporter aux travaux de A. RUBERTI, A. ISIDORI et P. d'ALESSANDRO (1972), R.R. MOLHER (1973), C. BRUNI, G.D. Di PILLO et G. KOCH (1974), H.J. SUSSMANN (1976), D.N.P. MURPHY (1979), Y. FUNAHASHI (1979), M. FLIESS et D. NORMAND-CYROT (1980),...

C'est à partir de la démarche des automaticiens, que C.W.J. GRANGER et A.P. ANDERSEN (1978 a) ont introduit la classe des modèles bilinéaires analogue à celle des automaticiens où l'entrée déterministe a été remplacée par une suite de variables aléatoires indépendantes équidistribuées non observées, supposées généralement de lois gaussiennes. Les modèles considérés ont alors pour expression :

$$(1.1.4) \quad X(t) = \sum_{i=1}^p a_i X(t-i) + \sum_{j=1}^q c_j \epsilon(t-j) + \epsilon(t) + \sum_{j=1}^P \sum_{k=1}^Q b_{jk} X(t-j) \epsilon(t-k).$$

De nombreux travaux concernant les propriétés probabilistes de ces modèles ont maintenant vu le jour, l'approche statistique n'en est encore qu'à ses débuts et concernent essentiellement des cas particuliers. On peut par exemple se rapporter aux travaux de T. SUBBA RAO (1981, 1983), T. PHAM DINH et L. TAT TRAN (1981), T. SUBBA RAO et M.M. GABR (1980, 1984), D. GUEGAN (1981, 1982, 1983, 1984, 1986, 1987), E.J. HANNAN (1982), B.G. QUINN (1982), M. BHASKARA RAO, T. SUBBA RAO et A.M. WALKER (1983), T. PHAM DINH (1985, 1986), D. GUEGAN et T. PHAM DINH (1987), ... Ces modèles ont déjà été utilisés dans différents domaines, en particulier en économie (C.W.J. GRANGER et A.P. ANDERSEN, 1978 a), et en dynamique des populations (T. SUBBA RAO et M.M. GABR, 1984). En étudiant les trajectoires de ces modèles, (cf. chapitre II), on est aussi amené

à penser qu'ils pourront être utiles dans l'étude de phénomènes tels que, par exemple, la vitesse du vent, les éruptions volcaniques.

Récemment B.K. STENSHOLT et D. TJOSTHEIM (1987) ont établi l'existence des modèles bilinéaires "multiples" et ont donné l'expression de la fonction de covariance de ces modèles, B.K. STENSHOLT et T. SUBBA RAO (1987) ont abordé le problème de l'estimation des paramètres de ces modèles.

Le présent travail consiste à présenter une synthèse des modèles bilinéaires à partir de différents points de vue.

Nous introduisons maintenant quelques concepts et certains problèmes propres aux probabilistes et aux statisticiens, nécessaires pour traiter le sujet qui nous intéresse.

II. CONCEPTS

L'un des objectifs du statisticien est, à partir de l'étude de collections d'observations de trouver un modèle qui approxime le mieux possible le vrai processus aléatoire sous-jacent à ces observations. Pour mener cette étude le statisticien a besoin de la notion de modèles de probabilité et donc de processus stochastiques.

Rappelons qu'un processus stochastique est défini comme une famille de variables aléatoires $X(t)$ où t est un paramètre appartenant à un certain ensemble T . Si $T = \mathbb{Z}$, ou $T = \mathbb{N}$ alors $X(t)$, $t \in T$ est dit processus à temps discret. Si $T = (-\infty, +\infty)$, alors $X(t)$, $t \in T$ est appelé processus à temps continu. L'espace d'état du processus $X(t)$ est l'espace dans lequel toutes les valeurs possibles de $X(t)$ se trouvent.

Dans la suite de ce travail, nous ne considérons que des processus à temps discret (en fait à temps dans \mathbb{Z}) et les séries que l'on étudie sont des collections d'observations d'un processus aléatoire effectuées séquentiellement dans le temps. Pour étudier ces séries, nous allons utiliser plusieurs notions probabilistes fondamentales.

Tout d'abord la stationnarité. Intuitivement, un processus stochastique $X(t)$, $t \in \mathbb{Z}$ est stationnaire si les propriétés statistiques du processus ne changent pas avec le temps. On distingue deux notions de stationnarité : la stationnarité stricte et la stationnarité d'ordre m , $m > 0$, $m \in \mathbb{Z}$.

$X(t)$, $t \in \mathbb{Z}$ est strictement stationnaire si pour toute famille finie d'instants $t_1, \dots, t_k \in \mathbb{Z}$, et pour tout entier r , les lois jointes de $(X(t_1), \dots, X(t_k))$ et de $(X(t_1+r), \dots, X(t_k+r))$ sont les mêmes.

On dit que $X(t)$, $t \in \mathbb{Z}$ est stationnaire au second ordre ($m=2$), si $X(t)$, $t \in \mathbb{Z}$ est un processus du second ordre dont la moyenne et la covariance sont invariantes par translation dans le temps.

Ainsi pour un processus du second ordre, la stationnarité stricte implique la stationnarité à l'ordre deux. Par contre, il existe des processus stationnaires à l'ordre deux, et qui ne sont pas stationnaires stricts. Par exemple une suite $X(n)$ de variables aléatoires indépendantes de même moyenne et de même variance est toujours stationnaire à l'ordre deux, mais $X(n)$ n'est stationnaire stricte que si les $X(n)$ ont même loi pour tout n . Dans le cas gaussien ces deux notions coïncident.

On constate donc que la stationnarité est une notion fondamentale dans l'étude des processus stochastiques car elle est

intimement liée aux notions de covariance et de spectre, propriétés qui permettent dans le cas linéaire de déterminer totalement une série chronologique. La stationnarité assure aussi que les séries considérées n'explorent pas.

Une autre notion importante dans l'étude des processus stochastiques est l'inversibilité. Dans le cas des modèles bilinéaires, plusieurs définitions en ont été données. Une première formulation assez générale pour être appliquée autant aux modèles linéaires qu'aux modèles bilinéaires a été donnée par C.W.J. GRANGER et A.P. ANDERSEN (1978 b). Intuitivement elle correspond à l'idée d'extraire le bruit $\epsilon(t)$ à partir des observations passées du processus. On peut considérer la définition plus précise suivante : on dira qu'un processus $X(t)$, $t \in \mathbb{Z}$ est inversible si $\epsilon(t)$ est mesurable par rapport à la tribu $\sigma(X(s), s \leq t)$. Une définition équivalente à celle-ci a été utilisée dans la littérature par différents auteurs, (cf. T. PHAM DINH et L. TAT TRAN, 1981, B.G. QUINN, 1982, D. GUEGAN et T. PHAM DINH, 1987 a,b).

En fait la notion d'inversibilité est très importante dans l'étude des processus stochastiques. Elle permet de faire des prévisions et dans certains cas, elle est utile pour estimer les paramètres des modèles.

Notons que si cette notion est assez facilement réalisée pour les modèles linéaires, elle peut n'être jamais réalisée dans le cas de modèles non linéaires, ou ne l'être que pour des sous classes bien particulières.

Deux autres notions jouent aussi un rôle important dans l'étude des processus stochastiques : il s'agit de l'ergodicité et du mélange.

La notion d'ergodicité pour un processus $X(t)$, $t \in \mathbb{Z}$ défini sur un certain espace de Probabilité est essentiellement

liée à une propriété de moyenne pour l'ensemble des trajectoires du processus $X(t)$, ce qui implique en particulier l'existence d'une loi invariante pour le processus considéré. Cette notion est utile pour utiliser des théorèmes concernant la loi des grands nombres.

La notion de mélange, elle, permet, dans le cas de suites de variables dépendantes, d'obtenir la "presque indépendance" de variables d'indices éloignés. Cette notion est donc aussi très utile et permet d'étendre à des variables dépendantes des résultats démontrés dans le cas de variables indépendantes. Cette notion qui est beaucoup plus forte que la notion d'ergodicité nécessite d'autres hypothèses que la simple stationnarité, elle sera rarement atteinte directement dans le cas des modèles bilinéaires. Elle est particulièrement importante pour utiliser des théorèmes central limite.

Notons que pour démontrer les propriétés de stationnarité, d'ergodicité et de mélange, il est possible d'utiliser des outils du type de ceux employés par R.L. TWEEDIE qui sont alors des techniques markoviennes.

III- DIFFERENTES APPROCHES POUR L'ETUDE DES MODELES BILINEAIRES

Les modèles bilinéaires tels qu'ils ont été introduits par C.W.J. GRANGER et A.P. ANDERSEN (1978 a) ont fait l'objet de nombreuses études. Ils se présentent comme une extension naturelle des modèles linéaires, aussi a-t-on tenté de les étudier à l'aide des outils utilisés classiquement en théorie du modèle linéaire : covariance, spectre. Or si toute série chronologique linéaire est parfaitement déterminée par sa fonction de covariance, il n'en va malheureusement plus de même pour les modèles bilinéaires, et non linéaires en général. L'outil hilbertien est totalement insuffisant pour étudier les modèles bilinéaires. Il ne permet pas de les

différentier des modèles linéaires, ni de les différentier entre eux, d'où la nécessité d'utiliser d'autres outils permettant de les étudier. Par contre en ce qui concerne l'estimation des paramètres, on applique les mêmes méthodes que dans le cas linéaire. Les $\epsilon(t)$ sont toujours supposés gaussiens donc les techniques du maximum de vraisemblance et des moindres carrés sont toujours les mêmes et les deux méthodes coïncident encore.

A cause de la complexité de la structure profonde des modèles bilinéaires, on rencontre des problèmes concernant la définition de ces modèles, aussi diverses approches ont-elles été utilisées pour les étudier. Certaines ont conduit à des résultats ne concernant que des cas particuliers, d'autres ont permis d'obtenir récemment des résultats théoriques plus globaux.

1. Les chaos de Wiener

Les chaos de Wiener développés par J. NEVEU (1968) semblent être un bon cadre pour étudier les modèles bilinéaires (cf. D. GUEGAN, 1982 b, 1983 a). On peut obtenir la représentation globale des modèles bilinéaires sur chacune des composantes du chaos, et pour des cas particuliers en donner les conditions d'existence. On obtient ainsi un cadre général dans lequel les modèles bilinéaires s'intègrent bien. Néanmoins l'algèbre sous-jacente aux chaos est très compliquée et ne permet pas d'aboutir à des résultats spectaculaires. Au lieu d'utiliser les chaos de Wiener, on peut utiliser l'algèbre des polynômes d'observations qui a des propriétés moins intéressantes mais plus facilement manipulables. On est conduit alors à la représentation de Volterra à partir de laquelle on peut construire des tests permettant d'identifier ces modèles ou de les différentier composantes par composantes, en utilisant en particulier la théorie des polyspectres (cf. D.R. BRILLINGER, 1970), ou les décompositions sous forme de processus d'innovations (cf. D. GUEGAN, 1982 b, 1983 a).

2. L'approche directe

C'est à partir de la représentation introduite par C.W.J. GRANGER et A.P. ANDERSEN (1978 a), que l'étude du modèle bilinéaire fut entreprise. C'est celle qui a donné lieu au plus grand nombre de publications.

L'attention des chercheurs s'est tout d'abord focalisée sur l'existence de tels modèles stationnaires (stricts ou au second ordre) : de l'étude des cas particuliers à l'étude des modèles les plus généraux, on trouve les travaux de T. PHAM DINH et L. TAT TRAN (1981), D. GUEGAN (1981 a), B.G. QUINN (1982 b), E.J. HANNAN (1982), T. SUBBA RAO (1981), M. BHASKARA RAO, T. SUBBA RAO et A.M. WALKER (1983), S.I. AKAMANAM, M. BHASKARA RAO et K. SUBRAMANYAM (1986). Certains auteurs se sont intéressés au problème de l'inversibilité : T. PHAM DINH et L. TAT TRAN (1981), D. GUEGAN (1981 a), B.G. QUINN (1982 a); d'autres en ont étudié l'ergodicité : H. TONG (1981), S.I. AKAMANAM, M. BHASKARA RAO et S. SUBRAMANYAM (1986). L'étude de la covariance fut menée par T. SUBBA RAO (1981), celle des moments d'ordre supérieur par D. GUEGAN (1984 a). T. SUBBA RAO (1983) a d'autre part présenté une analyse bispectrale de ces modèles. Sur un cas particulier, S.R. WANG, H.Z. AN et H. TONG (1983) ont montré que la densité de probabilité du modèle pouvait ne pas être bornée sans contraintes particulières.

Malgré l'ampleur des travaux sur le sujet, beaucoup de domaines abordés ne le sont encore qu'à partir de modèles simples à cause de la complexité du modèle général.

3. La représentation markovienne

Dans la théorie des modèles linéaires, la représentation markovienne joue un rôle important, en particulier dès qu'il s'agit de faire du filtrage de Kalman. H. AKAIKE (1974 a) a montré que tout processus autorégressif moyenne mobile admettait une représentation markovienne et inversement. (Pour une revue voir D. GUEGAN, 1981 b). Ce résultat a permis de considérer le problème de l'identification de ces processus sous un angle différent. Une démarche similaire a été menée pour les modèles bilinéaires. T. PHAM DINH (1985) a établi la représentation markovienne pour une sous-classe des modèles bilinéaires : les modèles superdiagonaux, généralisant ainsi les travaux de H. AKAIKE. D. GUEGAN (1986, 1987) a montré que cette représentation markovienne ne pouvait pas se généraliser aux modèles sousdiagonaux de façon naturelle, et parallèlement T. PHAM DINH (1987), en augmentant suffisamment l'espace d'état, a établi la représentation markovienne pour l'ensemble des modèles bilinéaires. Ces représentations markoviennes ont permis d'aborder souvent de façon efficace et élégante, des problèmes traités (pour tout ou partie des modèles) à partir de la représentation univariée, souvent avec difficulté, à savoir l'existence d'une solution stationnaire unique (T. PHAM DINH, 1985), la minimalité de la représentation (T. PHAM DINH, 1985), l'ergodicité (D. GUEGAN, 1983 b, 1984 b ; T. PHAM DINH, 1987), l'inversibilité (D. GUEGAN et T. PHAM DINH, 1987 a,b), le calcul des moments (T. PHAM DINH, 1985).

4. La représentation Affine d'état

En théorie du contrôle, E.D. SONTAG (1979) a introduit les modèles à réponse polynomiale. Il démontra à l'aide de la géométrie algébrique l'équivalence entre les modèles bilinéaires et les modèles à réponse polynomiale dans la mesure où ces modèles sont réalisables. D. GUEGAN (1986, 1987) s'inspirant de ce résultat,

introduisit les modèles polynomiaux en les entrées, affines en l'état dans lesquels le contrôle déterministe est remplacé par un bruit aléatoire $\epsilon(t)$, supposé bruit blanc gaussien. La représentation ainsi obtenue a l'intérêt de se présenter sous forme d'une représentation d'état markovienne, les coefficients étant soit des matrices polynomiales en $\epsilon(t)$, soit des polynômes en $\epsilon(t)$, de degré fini. Cette représentation est en fait identique à celle introduite par T. PHAM DINH (1986). D. GUEGAN (1987) établit l'existence de cette représentation, étudie sa fonction de covariance, donne des conditions d'existence pour les moments d'ordre supérieur ainsi qu'une méthode récursive de calcul de tous les moments d'ordre pair. Des conditions pour que la représentation minimale soit atteinte sont obtenues par D. GUEGAN et T. PHAM DINH (1987 a).

IV - INFERENCE STATISTIQUE

Les propriétés statistiques concernant les modèles bilinéaires, quelque soit le type de représentation que l'on considère, ne sont pas encore très développées. La plupart des résultats qui existent concernent des cas particuliers de modèles, ou des sous-classes de modèles.

1. Estimation des paramètres

On peut considérer que trois méthodes ont été essentiellement utilisées pour l'estimation des paramètres des modèles bilinéaires. Il s'agit de la méthode par maximum de vraisemblance, de la méthode des moindres carrés conditionnels et de la méthode des moments.

Dans le cas de la méthode par maximum de vraisemblance, on considère la distribution jointe de la suite des $\epsilon(t)$ (soit le produit de gaussiennes marginales), et on obtient la fonction de

vraisemblance après avoir introduit le Jacobien approprié qui provient du passage des $\epsilon(t)$ (le bruit blanc) aux $X(t)$ (les observations). La maximisation de cette fonction est alors obtenue soit par la méthode de Newton-Raphson, soit par la méthode du gradient. Cette méthode est utilisée par M.M. GABR et T. SUBBA RAO (1981) pour estimer des paramètres dans le cas de modèles bilinéaires univariés sans partie moyenne mobile. Il s'agit d'une méthode empirique. Les paramètres obtenus peuvent être consistants, cependant la distribution asymptotique des estimateurs n'est pas atteinte. Une des difficultés principales étant due au manque de propriétés sur les moments de ces modèles.

L'estimation des paramètres par la méthode des moindres carrés a été utilisée d'une part par T. PHAM DINH et L. TAT TRAN (1981) pour estimer les paramètres d'un modèle bilinéaire à un terme, d'autre part par D. GUEGAN et T. PHAM DINH (1987 b) pour estimer les paramètres des modèles diagonaux à partir de la représentation markovienne obtenue par T. PHAM DINH (1985). Cette méthode consiste essentiellement à minimiser la somme des carrés des erreurs sur un certain espace de paramètres contenant la vraie valeur du paramètre. Elle ne peut être utilisée évidemment que dans le cas de modèles inversibles. Les paramètres estimés obtenus sont convergents, mais leur normalité asymptotique n'a pas pu être atteinte, les dérivées du bruit par rapport aux paramètres n'étant pas de carré intégrable. Dans le cas de modèles particuliers, D. GUEGAN (1984 a) a pu obtenir l'estimation des paramètres par la méthode des moments. Les paramètres estimés sont convergents et ont une distribution asymptotique normale. Cette méthode qui peut se généraliser aux cas de modèles plus complexes est néanmoins très lourde à manier. On retrouve le même type de démarche dans l'article de P.M. ROBINSON (1977) qui s'est intéressé à un modèle non linéaire moyenne mobile avec un terme de la forme $\epsilon(t)\epsilon(t-1)$.

2. Tests de linéarité

Dans toutes les discussions précédentes, il a toujours été admis implicitement que le modèle étant choisi, son ordre était connu. En pratique ajuster des procédures à partir d'observations signifie en fait que l'on dispose d'un critère pour déterminer le modèle et l'ordre de ce modèle.

En fait les tests dont on dispose ne permettent pas de dire avec quel type de modèles il faut travailler. Il s'agit plutôt de tests de validité qui permettent de confirmer le choix d'un modèle. Dans le cas linéaire ce type de tests existe, mais il n'est pas encore très étudié dans le cas non linéaire. Pour ces modèles, les seuls types de tests dont on dispose sont des tests de linéarité. En 1980, à partir de l'analyse spectrale, T. SUBBA RAO et M.M. GABR en ont proposé un. Ce test est basé sur un résultat concernant le spectre d'ordre supérieur, établi par D.R. BRILLINGER (1965). Sachant que le rapport

$$|h_k(w_1, w_2, \dots, w_k)|^2 \{h(w_1)h(w_2) \dots h(w_k)h(w_1+w_2+\dots+w_k)\}^{-1}$$

est constant sur tout intervalle w_1, w_2, \dots, w_k , où h_k représente la densité spectrale d'ordre k , et h la densité spectrale d'un processus stationnaire possédant les moments appropriés, les auteurs établissent leur test dans le cas $k = 2$. M.J. HINNICH (1982) modifie ce test en prenant en compte les difficultés rencontrées pour obtenir des estimateurs de $h(w)$ près de zéro. Notons que ce test n'est pas spécifique aux modèles bilinéaires, et qu'il ne peut pas s'appliquer aux modèles qui ont des distributions symétriques (cf. J. PEMBERTON et H. TONG, 1981).

D. GUEGAN (1984 a), propose un autre type de test. A partir de cas particuliers de modèles bilinéaires, elle établit un test basé sur les moments des modèles, généralisant ainsi aux modèles non linéaires, le test de M.S. BARTLETT (1946) et M.H. QUENOUILLE

(1949). Cette démarche permet de tester l'hypothèse d'un modèle appartenant aux deux premières composantes du chaos contre l'hypothèse d'un modèle appartenant au chaos tout entier. On montre que le test ainsi obtenu est assez performant.

Plus récemment différents types de tests ont été proposés permettant, soit de tester l'hypothèse linéarité contre l'hypothèse non linéarité, par exemple à partir du test de non-additivité de J.W. TUKEY (1949), (cf. D.M. KEENAN, 1985 et R.S. TSAY, 1986), soit de détecter des modèles non linéaires spécifiques, (cf. A.I. Mc LEOD et W.L. LI, 1983, J.D. PETRUCCELLI et N. DAVIES, 1986). Ces derniers construisent un test qu'ils appellent test de Portmanteau et qui est en fait un test proche des techniques de rupture de régression.

En ce qui concerne l'ordre du modèle, si maintenant le problème est bien résolu pour les modèles linéaires, grâce en particulier au critère d'AKAIKE (1974 b), on ne dispose pas encore de résultats similaires dans le cas des modèles bilinéaires. T. SUBBA RAO et M.M. GABR (1984) ont établi des programmes permettant d'utiliser le critère d'AKAIKE pour déterminer pas à pas l'ordre de la partie bilinéaire des modèles étudiés.

V - PROBLEMES OUVERTS CONCERNANT LES MODELES BILINEAIRES

Un grand nombre de problèmes concernant tant l'approche probabiliste que l'approche statistique des modèles bilinéaires, restent non résolus. Un certain nombre de résultats sont à généraliser.

Certaines propriétés probabilistes plus approfondies devraient permettre de meilleurs résultats dans le domaine de l'inférence statistique, à savoir la stabilité, l'ergodicité, les propriétés de mélange, l'inversibilité, l'existence et le calcul

des moments d'ordre supérieur. Autant de propriétés qui permettraient d'utiliser des théorèmes central limite, pour améliorer en particulier les résultats concernant les propriétés des estimateurs des paramètres (cf. C. VILLEGAS, 1976).

Précisément, en ce qui concerne l'estimation des paramètres, les méthodes que nous avons rappelées doivent être développées et adaptées. On peut penser que la méthode des moindres carrés peut être généralisée à d'autres modèles à condition d'obtenir l'inversibilité. Certains travaux à ce propos peuvent être certainement utilisés, par exemple ceux de P.M. ROBINSON (1983), L.A. KLIMKO et P.I. NELSON (1978). D. TJOSTHEIM (1986) a établi aussi des conditions pour avoir la consistance forte et la normalité asymptotique d'estimateurs des moindres carrés et d'estimateurs du maximum de vraisemblance dans le cas de modèles non linéaires. Ces travaux pour l'instant n'ont pas encore pu être adaptés au cas des modèles bilinéaires.

La méthode des moments, applicable dans le cas où l'on connaît le calcul des moments permet dans le cas linéaire d'identifier un modèle à l'aide des deux premiers moments. Dans le cas non linéaire, malgré la complexité des calculs elle permet de disposer de premiers estimateurs. Néanmoins il faudrait pouvoir connaître la loi du modèle pour déterminer les paramètres de manière unique. Cette méthode en première approche peut être utilisée dans le cas des modèles polynomiaux affines en l'état.

En ce qui concerne les tests, le problème de l'identifiabilité du modèle reste posé, aucun test pour l'instant ne permet d'identifier un modèle bilinéaire.

Nous avons vu que les seuls tests dont on dispose pour l'instant sont des tests de linéarité. On peut espérer établir un test de non linéarité plus sensible aux déviations de type bilinéaires. Le test proposé par J.D. PETRUCCELLI et N. DAVIES

(1986) est un test de non linéarité mais il est plus sensible au problème des seuils, à la rupture dans l'autorégression.

Il faut maintenant s'orienter vers des tests de validité : un modèle bilinéaire étant proposé et ajusté, il faut trouver un test permettant de le valider.

On dispose aussi de quelques tests de discrimination entre plusieurs types de modèles non linéaires, cf. D. GUEGAN, 1984 a, et K. LI, 1984 ; ce dernier utilise les autocorrélations des processus carrés.

En ce qui concerne la prédiction, domaine maintenant bien connu dans le cas des séries linéaires, le problème n'a pour ainsi dire pas été abordé dans le cas des modèles bilinéaires. C'est un domaine pourtant privilégié pour le statisticien qui cherche à adapter des séries en vue de faire des prévisions. Dans le cas des modèles bilinéaires, c'est la prévision au delà d'un pas qui pose problème. En fait ce problème est intimement lié à celui de l'inversibilité, car ce qui est important c'est de pouvoir exprimer le bruit à partir des observations et du bruit passés.

Notons que la notion de prédiction est d'autant plus importante qu'elle peut permettre de différencier un modèle linéaire d'un modèle non linéaire, (cf. E.J. HANNAN, 1979).

De nombreux articles traitant de la prédiction linéaire et non linéaire peuvent fournir des bases de réflexion, nous pensons plus précisément aux articles de P. MASANI et N. WIENER (1959), M. ROSENBLATT (1979) et L.A. SHEPP, D. SLEPIAN et A.D. WYNER (1980).

CHAPITRE IIETUDE EXPERIMENTALE DES MODELESBILINEAIRESI - INTRODUCTION

Une approche naturelle pour commencer l'étude d'un modèle est d'en faire une description approfondie afin d'avoir une première idée de son comportement .

Nous nous intéressons au comportement des trajectoires. Vu la complexité du modèle, l'approche analytique ne permet pas d'aboutir, nous avons donc recours à des simulations. Pour une approche expérimentale, il n'est pas possible de prendre trop de modèles, nous avons donc décidé d'étudier uniquement des modèles bilinéaires d'ordre 1 avec ou sans terme moyenne mobile de la forme suivante :

$$(2.1.1) \quad X(t) = a_1X(t-1) + a_2X(t-2) + b_1X(t-1)\epsilon(t-1) + \epsilon(t)$$

$$(2.1.2) \quad X(t) = a_1X(t-1) + b_1X(t-1)\epsilon(t-1) + \epsilon(t)$$

$$(2.1.3) \quad X(t) = a_1X(t-1) + c_1\epsilon(t-1) + b_1X(t-1)\epsilon(t-1) + \epsilon(t)$$

Nous avons commencé par étudier dix modèles du type précédent pour avoir une première idée des phénomènes que l'on pouvait attendre. Le principal phénomène qui est apparu est le caractère explosif que nous analyserons dans une deuxième partie.

Pour mener cette étude, nous avons choisi les paramètres

pour assurer la stationnarité et l'inversibilité des modèles, en se servant des résultats des chapitres suivants. D'autre part, on s'est efforcé de créer des séries avec des comportements périodiques ou non en jouant sur la position des racines du polynôme autorégressif des modèles. On a essayé de donner des importances différentes aux diverses composantes bilinéaire et linéaire des modèles. Ceci nous a conduit au choix des dix modèles suivants :

1. Modèles présentant des "périodicités" (faible et forte composante bilinéaire)..

$$\text{(Série 1)} \quad X(t) = 1,55 X(t-1) - 0,81 X(t-2) + 0,1 X(t-1)\epsilon(t-1) + \epsilon(t)$$

$$\text{(Série 2)} \quad X(t) = 1,55 X(t-1) - 0,81 X(t-2) + 0,4 X(t-1)\epsilon(t-1) + \epsilon(t)$$

$$\text{(Série 3)} \quad X(t) = 0,9 X(t-1) - 0,81 X(t-2) + 0,4 X(t-1)\epsilon(t-1) + \epsilon(t)$$

2. Modèles apériodiques (un ou deux termes autorégressifs, composante bilinéaire faible et forte).

$$\text{(Série 4)} \quad X(t) = 0,86 X(t-1) - 0,25 X(t-2) + 0,6 X(t-1)\epsilon(t-1) + \epsilon(t)$$

$$\text{(Série 5)} \quad X(t) = 0,5 X(t-1) + 0,85 X(t-1)\epsilon(t-1) + \epsilon(t)$$

$$\text{(Série 6)} \quad X(t) = 0,2 X(t-1) + 0,3 X(t-1)\epsilon(t-1) + \epsilon(t)$$

$$\text{(Série 7)} \quad X(t) = 0,2 X(t-1) + 0,55 X(t-1)\epsilon(t-1) + \epsilon(t)$$

3. Modèles apériodiques (avec un terme autorégressif, un terme moyenne mobile, une composante bilinéaire faible et forte).

$$\text{(Série 8)} \quad X(t) = 0,5 X(t-1) + 0,5 \epsilon(t-1) + 0,85 X(t-1)\epsilon(t-1) + \epsilon(t)$$

$$\text{(Série 9)} \quad X(t) = 0,2 X(t-1) + 0,5 \epsilon(t-1) + 0,3 X(t-1)\epsilon(t-1) + \epsilon(t)$$

$$\text{(Série 10)} \quad X(t) = 0,2 X(t-1) + 0,5 \epsilon(t-1) + 0,55 X(t-1)\epsilon(t-1) + \epsilon(t)$$

L'examen des trajectoires de ces modèles révèle souvent l'existence d'explosions. Nous avons alors essayé de discerner le rôle des paramètres quant à ce phénomène. Ceci constituera la deuxième partie de l'étude.

Les simulations présentées dans les paragraphes suivants sont obtenues à partir d'un bruit blanc gaussien centré de variance 1. Elles sont soit de longueur 1000, soit de longueur de longueur 5000.

II - ETUDE PRELIMINAIRE.

On commence par regarder le comportement des distributions marginales univariées, i.e., la distribution des $X(t)$, puis bivariées, i.e., la distribution conjointe $(X(t), X(t+1))$. On regarde ensuite les comportements des trajectoires.

1 - Distributions Marginales.

a - Distributions univariées.

Les caractéristiques marginales de ces séries (moyenne, écart-type, quantiles) sont données dans le tableau 2.1, et les histogrammes correspondants sont donnés respectivement par les figures 1.1 à 1.10.

On peut constater que pour les séries 2, 4, 5 et 8 il existe un grand pourcentage de valeurs extrêmes. Cela se traduit entre autre par un écart-type très important par rapport à l'interquartile, et par une augmentation de la moyenne par rapport à la médiane qui reste comparable aux autres séries. Nous avons cherché l'existence possible d'une densité bimodale mais nous n'avons pas pu la mettre en évidence peut-être à cause du faible nombre de répétitions des simulations. Toutefois, pour la série 8, la forme de l'histogramme, cf. fig. 8.1, peut laisser penser à l'existence d'un maximum local de la densité aux alentours de 14. Notons que les séries 2, 4, 5 et 8 sont toutes à forte composante bilinéaire. La série 8 est la même que la série 5 mais comporte un terme moyenne mobile important en plus.

Séries	Série 1	Série 2	Série 3	Série 4	Série 5	Série 6	Série 7	Série 8	Série 9	Série 10
Caractéristiques										
Valeur Maximale	30,04	298,40	16,01	415,30	702,96	9,14	31,25	840	13,36	41,11
Valeur Minimale	-11,72	-1475,80	-12,50	-11,41	-26,99	-2,81	-3,41	-37,47	-3,08	-6,16
Moyenne	-0,65	24,19	-0,50	4,27	5,65	0,45	0,83	6,38	0,48	0,87
Ecart type	4,14	303,60	3,74	27,49	43,56	1,23	2,00	52,05	1,56	2,57
Quantiles :										
1%	-7,01	-890,91	-9,10	-4,41	-11,69	-1,89	-2,30	-14,70	-2,44	-2,97
5%	-4,84	-146,64	-5,60	-2,13	-4,04	-1,37	-1,36	-5,07	-1,66	-1,73
10%	-3,79	-24,38	-3,84	-1,40	-1,40	-1,02	-0,94	-1,73	-1,23	-1,20
25%	-2,14	-5,52	-1,51	-0,55	-0,40	-0,42	-0,31	-0,64	-0,61	-0,51
50% Médiane	0,22	0,78	-0,27	0,77	0,81	0,39	0,57	0,61	0,38	0,48
75%	2,82	9,86	2,34	2,56	2,56	1,19	1,56	2,73	1,28	1,69
90%	5,44	50,88	5,24	5,85	6,68	1,99	2,67	7,21	2,33	3,25
95%	7,54	229,22	7,30	10,22	11,47	2,64	3,85	12,86	3,32	4,90
99%	10,04	1042,50	9,18	18,96	23,79	3,20	5,26	24,19	4,39	6,98
Interquantile	4,96	153,89	3,85	3,11	2,96	1,62	1,87	3,37	1,90	2,21

Tableau 2.1

HISTOGRAM OF SERIE GROUPED BY GL

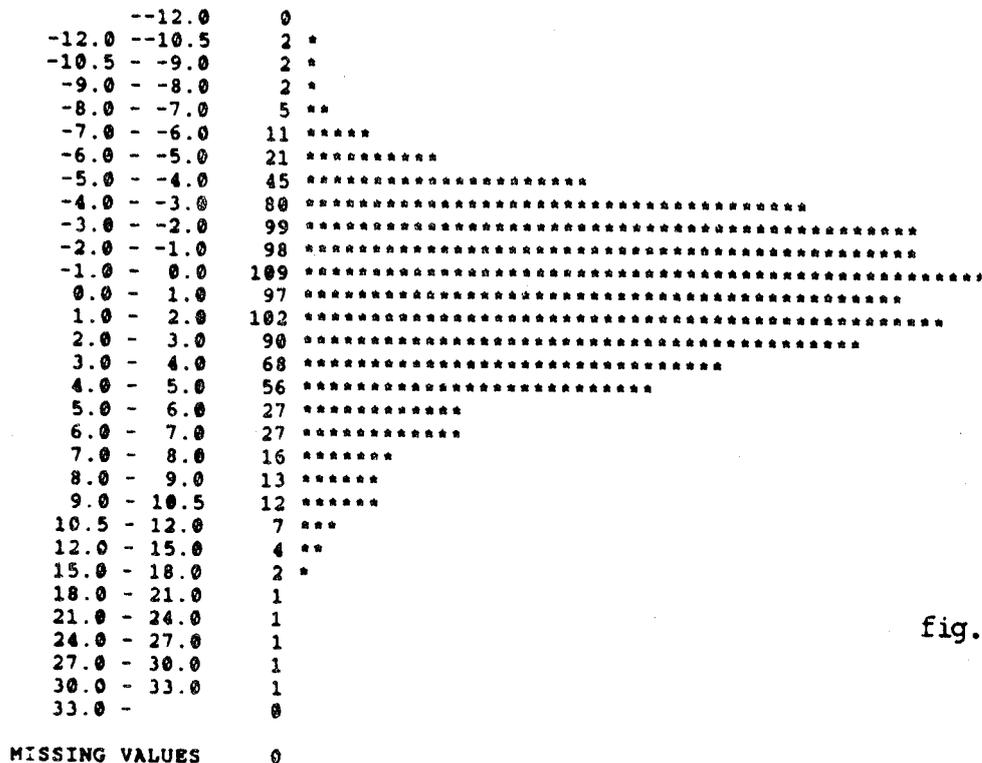


fig. 1.1

SCALE: 1 ASTERISK REPRESENTS 2.18 UNIT(S)

SERIE 2 35 CLASSES

HISTOGRAM OF SERIE GROUPED BY GL

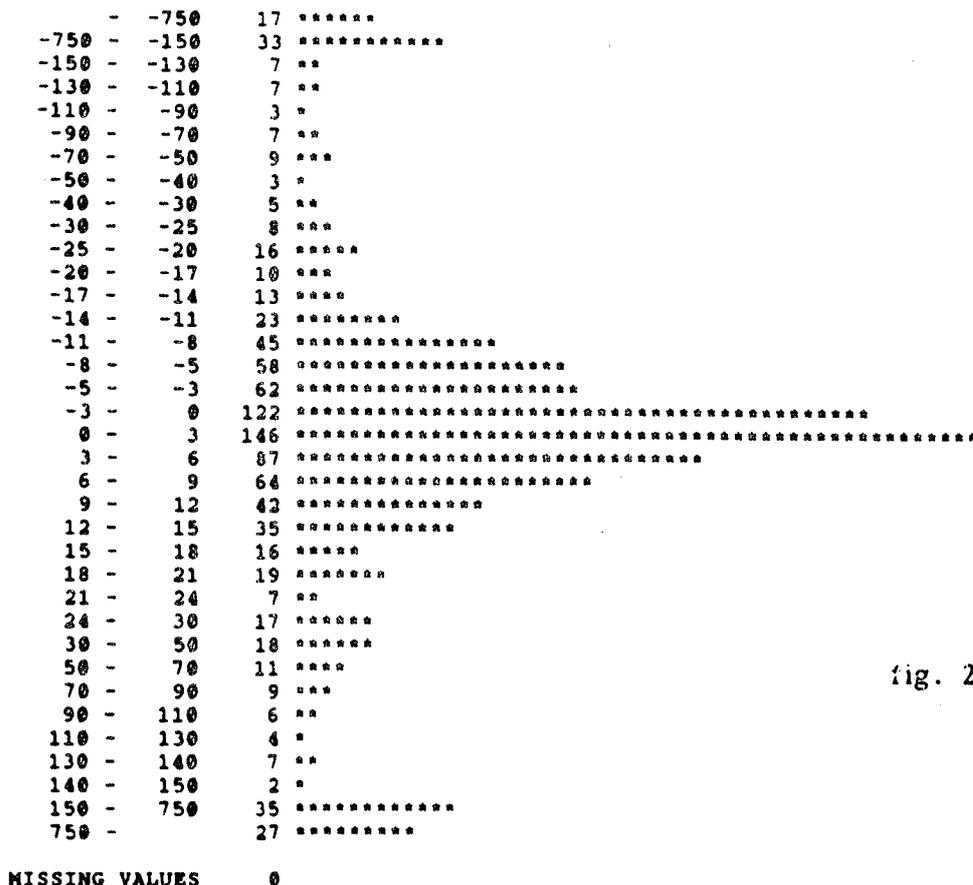


fig. 2.1

SCALE: 1 ASTERISK REPRESENTS 2.92 UNIT(S)

SERIE 3 24 CLASSES

HISTOGRAM OF SERIE GROUPED BY GL

--13.00	0
-13.00 --11.00	3 *
-11.00 --9.00	8 ***
-9.00 --8.00	5 **
-8.00 --7.00	12 ****
-7.00 --6.00	16 *****
-6.00 --5.00	19 *****
-5.00 --4.00	29 *****
-4.00 --3.00	46 *****
-3.00 --2.00	63 *****
-2.00 --1.00	111 *****
-1.00 --0.00	142 *****
0.00 --1.00	135 *****
1.00 --2.00	126 *****
2.00 --3.00	85 *****
3.00 --4.00	54 *****
4.00 --5.00	37 *****
5.00 --6.00	35 *****
6.00 --7.00	15 *****
7.00 --8.00	25 *****
8.00 --9.00	14 *****
9.00 --11.00	12 *****
11.00 --13.00	3 *
13.00 --15.00	3 *
15.00 --	2 *
MISSING VALUES	0

fig. 3.1

SCALE: 1 ASTERISK REPRESENTS 2.84 UNIT(S)

SERIE 4 35 CLASSES

HISTOGRAM OF SERIE GROUPED BY GL

--30.00	0
-30.00 --27.00	0
-27.00 --25.00	0
-25.00 --22.00	0
-22.00 --19.00	0
-19.00 --16.00	0
-16.00 --13.00	0
-13.00 --10.00	1
-10.00 --9.00	0
-9.00 --8.00	0
-8.00 --7.00	0
-7.00 --6.00	1
-6.00 --5.00	5 *
-5.00 --4.00	8 **
-4.00 --3.00	7 **
-3.00 --2.00	38 *****
-2.00 --1.00	94 *****
-1.00 --0.00	204 *****
0.00 --1.00	188 *****
1.00 --2.00	149 *****
2.00 --3.00	90 *****
3.00 --4.00	59 *****
4.00 --5.00	32 *****
5.00 --6.00	28 *****
6.00 --7.00	11 ***
7.00 --8.00	11 ***
8.00 --9.00	13 ***
9.00 --10.00	9 **
10.00 --13.00	18 ****
13.00 --16.00	10 **
16.00 --19.00	5 *
19.00 --22.00	3 *
22.00 --25.00	1
25.00 --27.00	0
27.00 --30.00	1
30.00 --	14 ***
MISSING VALUES	0

fig. 4.1

SCALE: 1 ASTERISK REPRESENTS 4.08 UNIT(S)

SERIE 5 27 CLASSES

HISTOGRAM OF SERIE GROUPED BY GL

--26.00	1
-26.00 --23.00	0
-23.00 --20.00	2
-20.00 --17.00	3 *
-17.00 --14.00	1
-14.00 --11.00	6 *
-11.00 --8.00	10 **
-8.00 --5.00	17 ****
-5.00 --4.00	12 ***
-4.00 --3.00	4 *
-3.00 --2.00	17 ****
-2.00 --1.00	70 *****
-1.00 --0.00	187 *****
0.00 --1.00	210 *****
1.00 --2.00	153 *****
2.00 --3.00	79 *****
3.00 --4.00	58 *****
4.00 --5.00	33 *****
5.00 --8.00	56 *****
8.00 --11.00	27 *****
11.00 --14.00	16 *****
14.00 --17.00	5 *
17.00 --20.00	9 **
20.00 --23.00	4 *
23.00 --26.00	3 *
26.00 --29.00	2
29.00 --30.00	0
30.00 -	15 ****
MISSING VALUES	0

fig. 5.1

SCALE: 1 ASTERISK REPRESENTS 4.20 UNIT(S)

SERIE 6 27 CLASSES

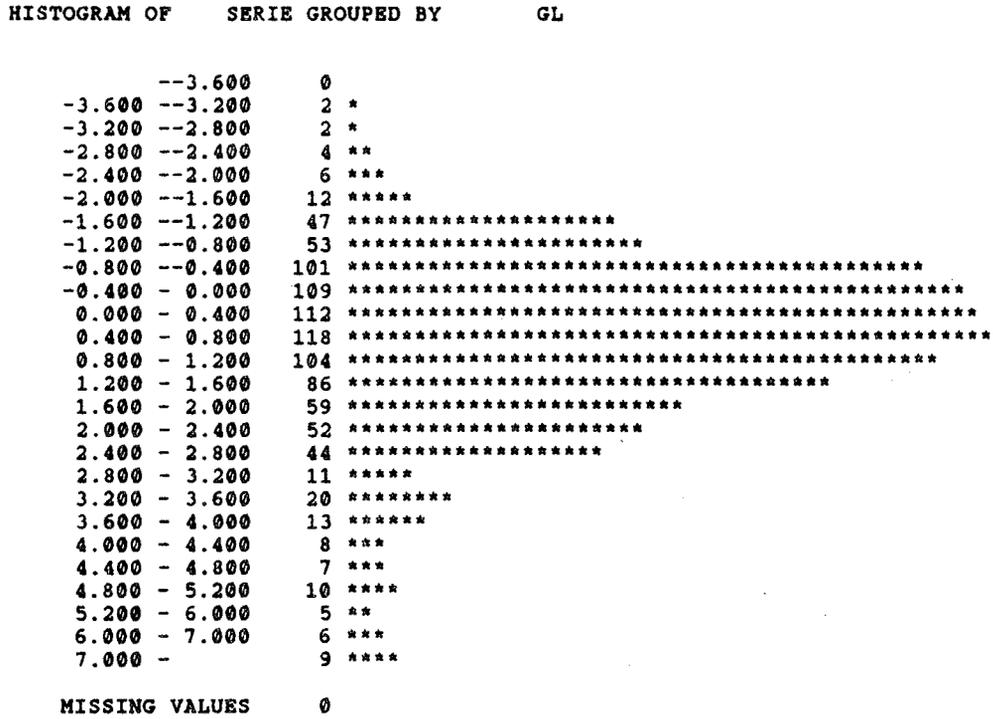
HISTOGRAM OF SERIE GROUPED BY GL

--4.4	0
-4.4 --4.0	0
-4.0 --3.6	0
-3.6 --3.2	0
-3.2 --2.8	1
-2.8 --2.4	1
-2.4 --2.0	5 **
-2.0 --1.6	17 *****
-1.6 --1.2	47 *****
-1.2 --0.8	68 *****
-0.8 --0.4	117 *****
-0.4 --0.0	121 *****
0.0 --0.4	124 *****
0.4 --0.8	142 *****
0.8 --1.2	108 *****
1.2 --1.6	103 *****
1.6 --2.0	47 *****
2.0 --2.4	31 *****
2.4 --2.8	29 *****
2.8 --3.2	19 *****
3.2 --3.6	11 ****
3.6 --4.0	4 *
4.0 --4.4	1
4.4 --4.8	1
4.8 --5.2	1
5.2 --5.6	0
5.6 --6.0	0
6.0 -	2 *
MISSING VALUES	0

fig. 6.1

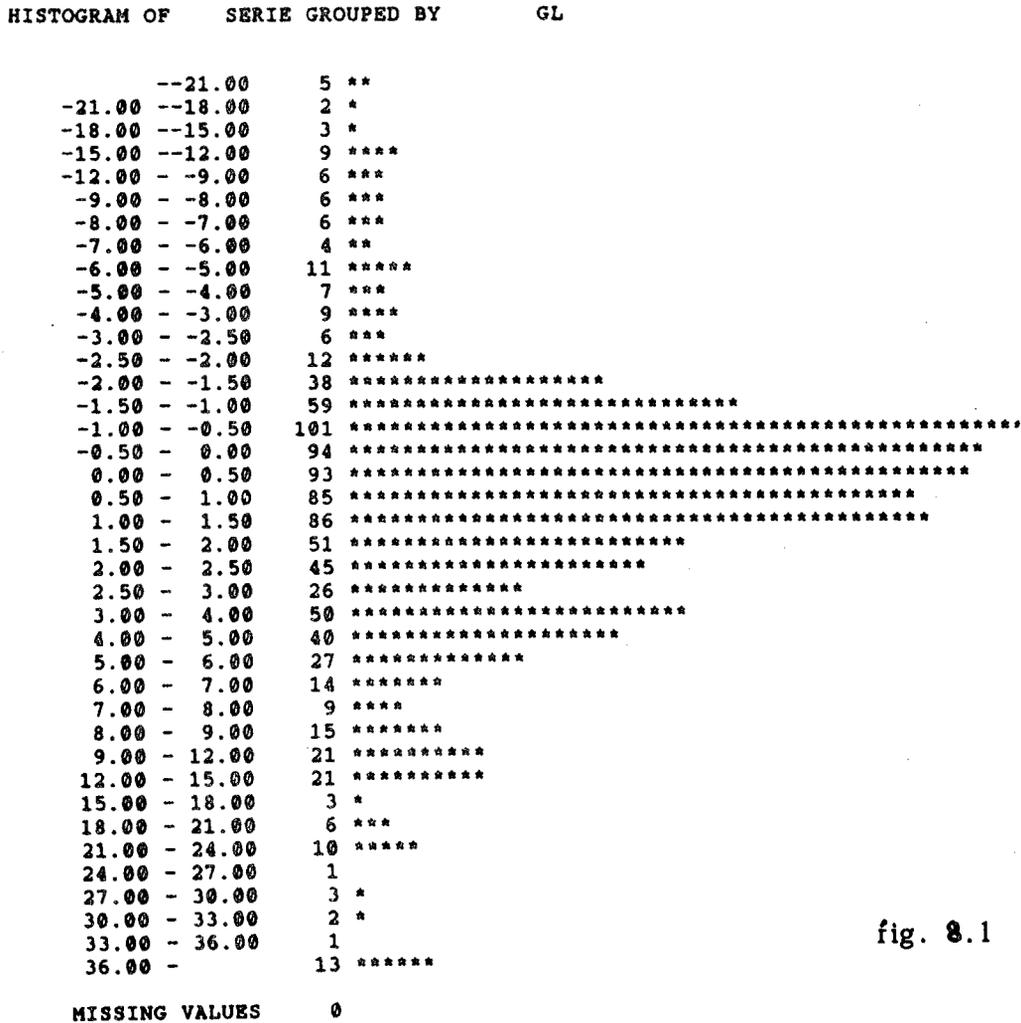
SCALE: 1 ASTERISK REPRESENTS 2.84 UNIT(S)

SERIE 7 25 CLASSES



SCALE: 1 ASTERISK REPRESENTS 2.36 UNIT(S)
 SERIE 8 33 CLASSES

fig. 7.1



SCALE: 1 ASTERISK REPRESENTS 2.02 UNIT(S)

fig. 8.1

SERIE 9 22 CLASSES

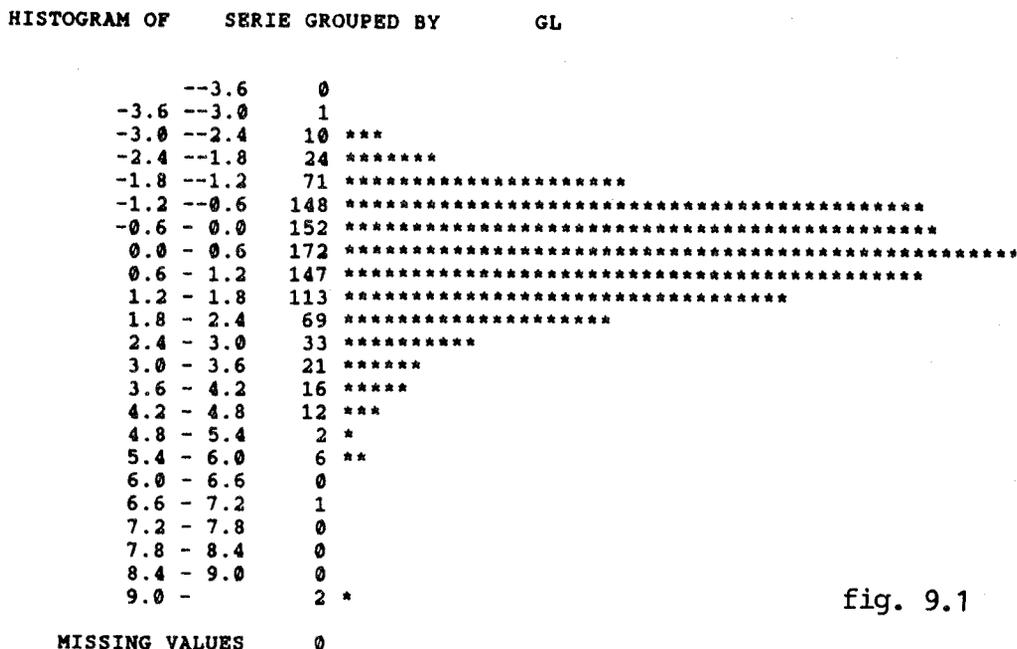


fig. 9.1

SCALE: 1 ASTERISK REPRESENTS 3.44 UNIT(S)

SERIE 10 36 CLASSES

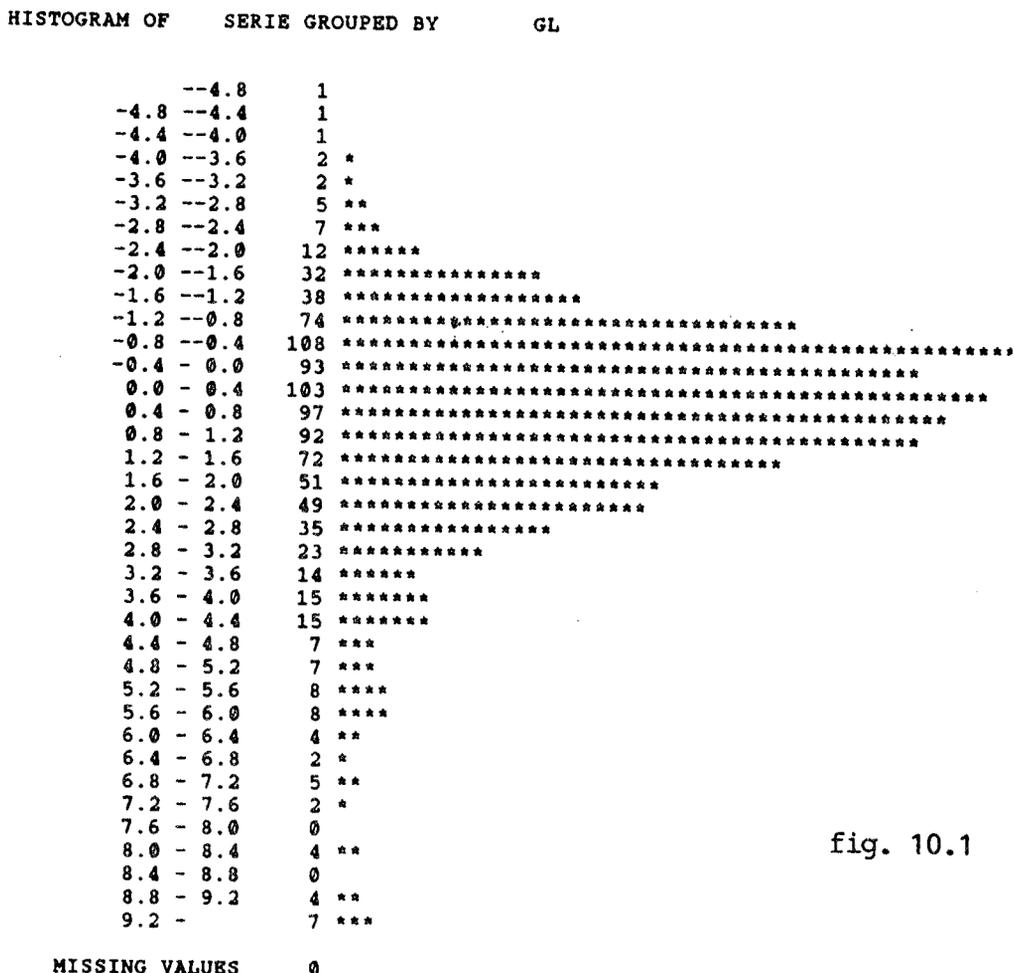


fig. 10.1

SCALE: 1 ASTERISK REPRESENTS 2.16 UNIT(S)

b - Distributions bivariées.

Les distributions bivariées, pour 5000 points des dix précédentes séries sont présentées dans les figures 1.2 à 10.2. Nous avons fait ces études pour rechercher l'existence possible d'oscillations entretenues du type cycle limite que l'on a pu constater dans certains modèles autorégressifs à seuil. Nous n'avons pas trouvé ce genre de phénomènes pour nos séries bilinéaires. Par ailleurs nous avons constaté des nuages de points nettement non gaussiens, par exemple pour les séries 5 à 10 (fig. 5.2 à 10.2). Notons que les séries 5 et 8 ont une forte composante bilinéaire (beaucoup de valeurs extrêmes, cf. tableau 2.1), par contre les séries 6, 7, 9 et 10 ont une partie bilinéaire faible (peu de valeurs extrêmes) ,et leur partie autorégressive est très faible. Il semble donc que c'est l'importance relative entre la composante autorégressive et la composante bilinéaire qui produise la forme particulière de ces nuages.

$$X(t) = 1.55 X(t-1) - .81 X(t-2) + .1 X(t-1)e(t-1) + e(t)$$

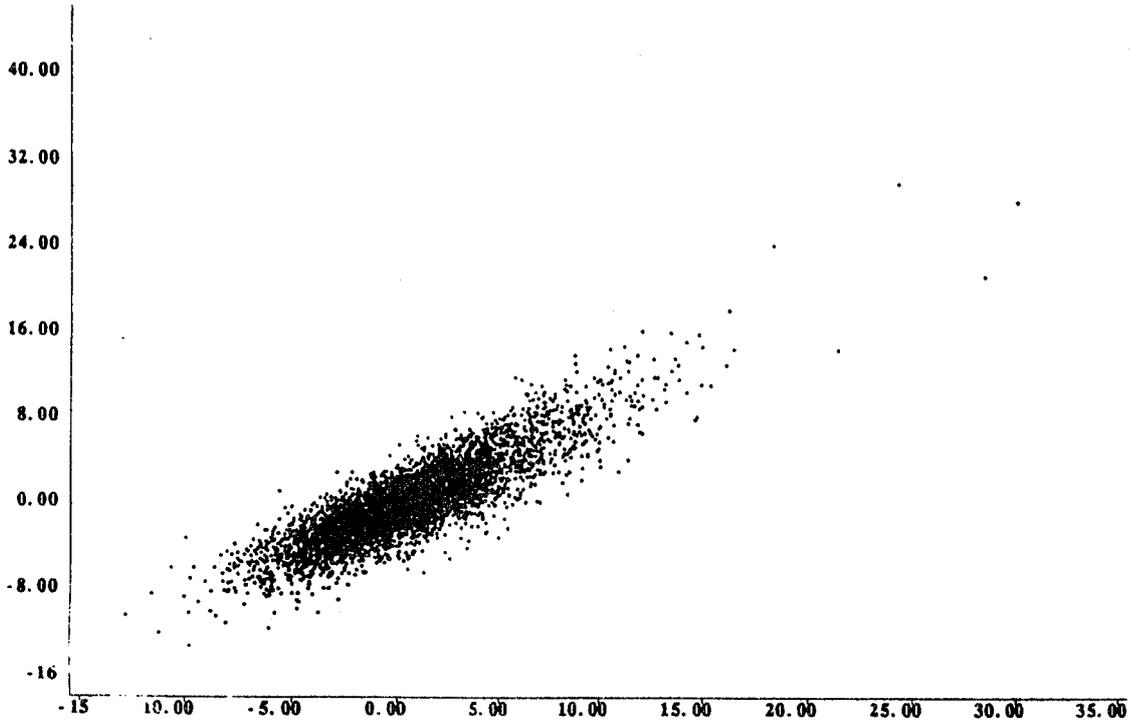


fig. 1.2

$$X(t) = 1.55 X(t-1) - .81 X(t-2) + .4 X(t-1)e(t-1) + e(t)$$

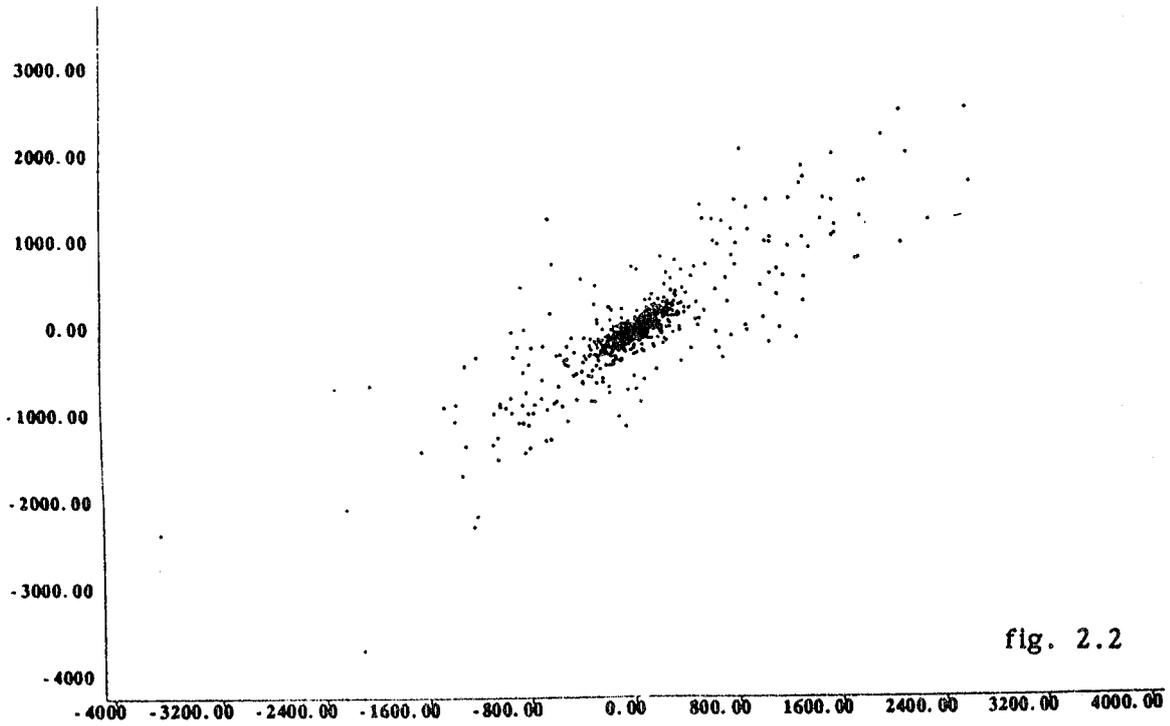


fig. 2.2

$$X(t) = .9 X(t-1) - .81 X(t-2) + .4 X(t-1)e(t-1) + e(t)$$

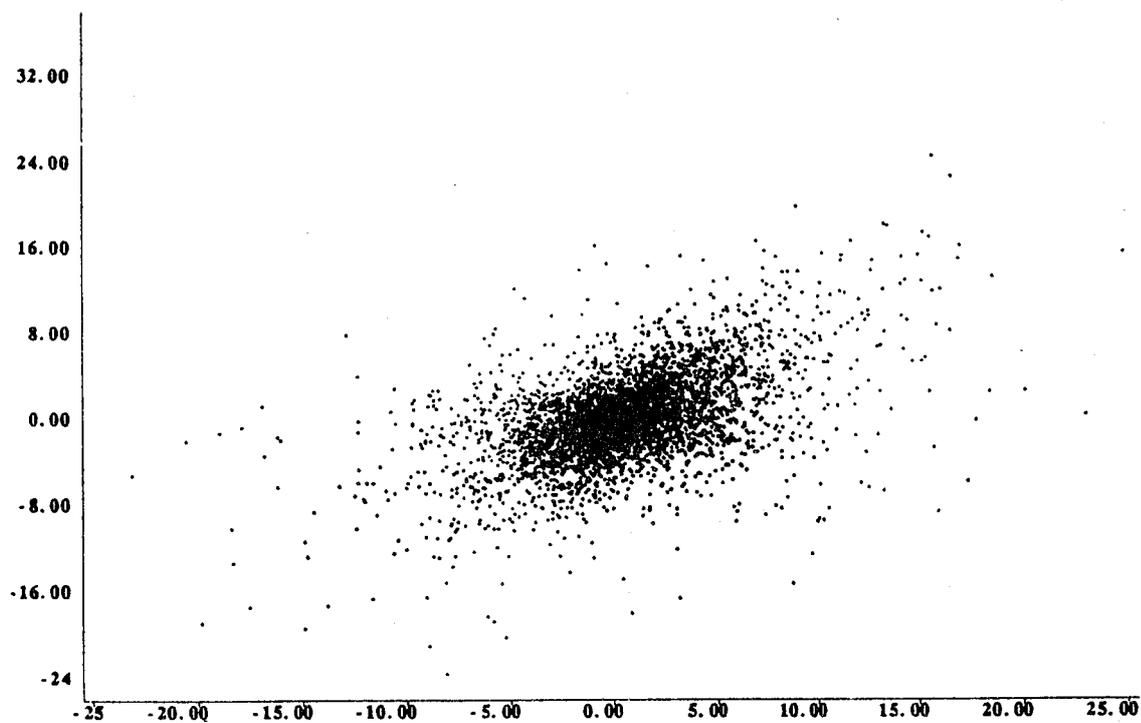


fig. 3.2

$$X(t) = .86 X(t-1) - .25 X(t-2) + .6 X(t-1)e(t-1) + e(t)$$

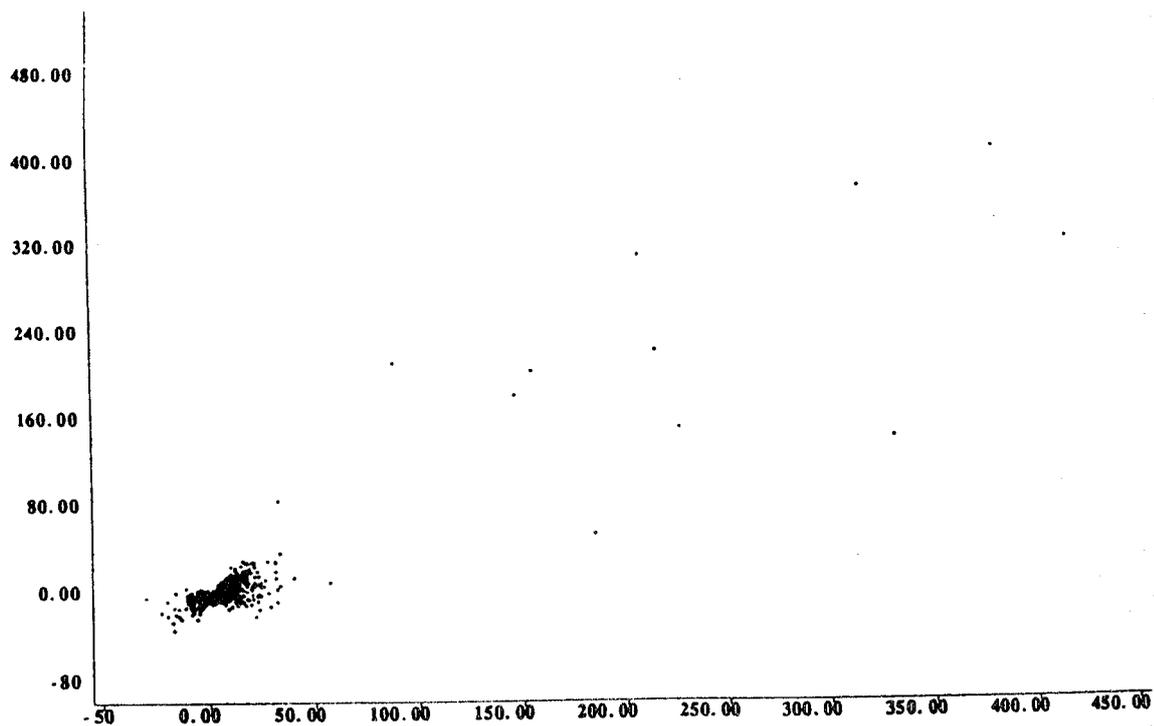


fig. 4.2

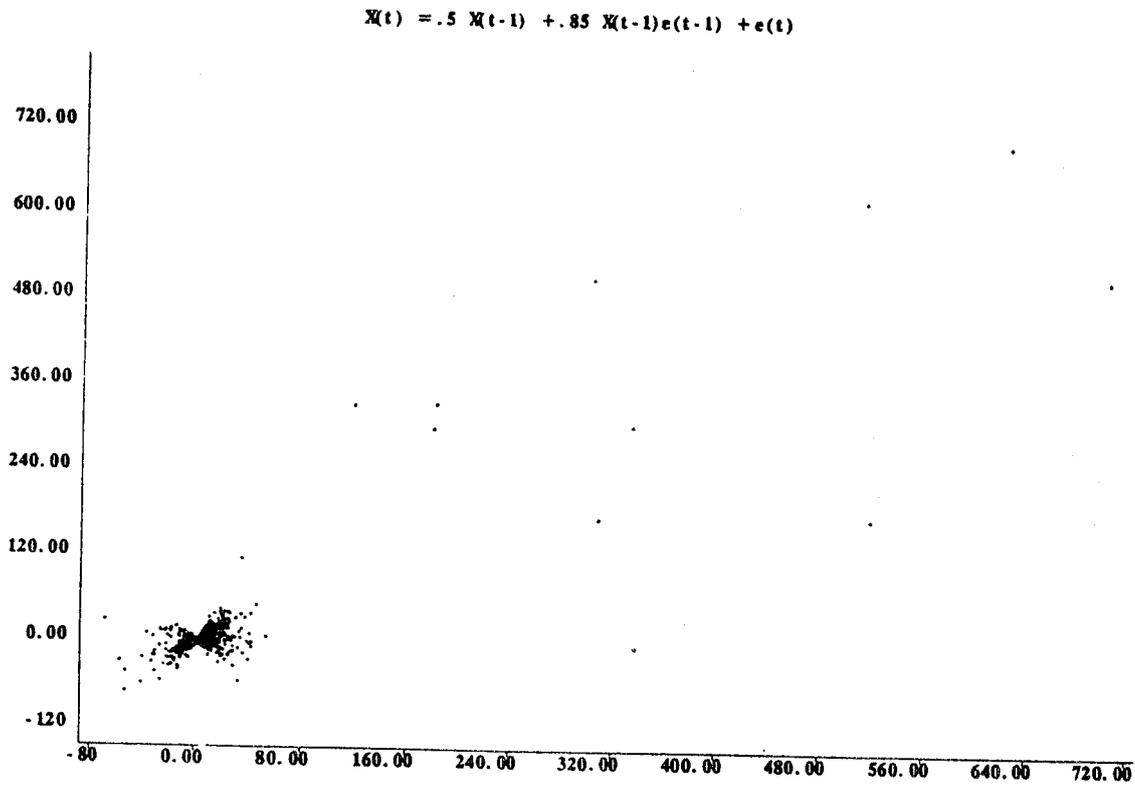


fig. 5.2

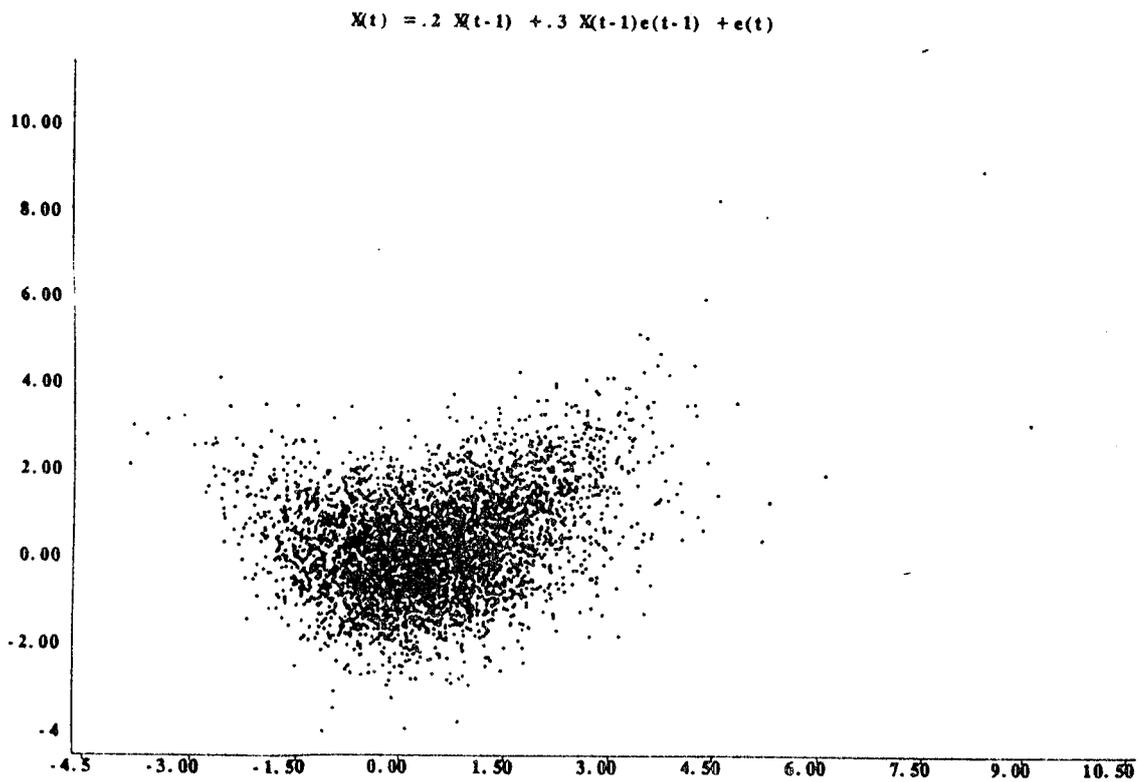


fig. 6.2

$$X(t) = .2 X(t-1) + .55 X(t-1)e(t-1) + e(t)$$

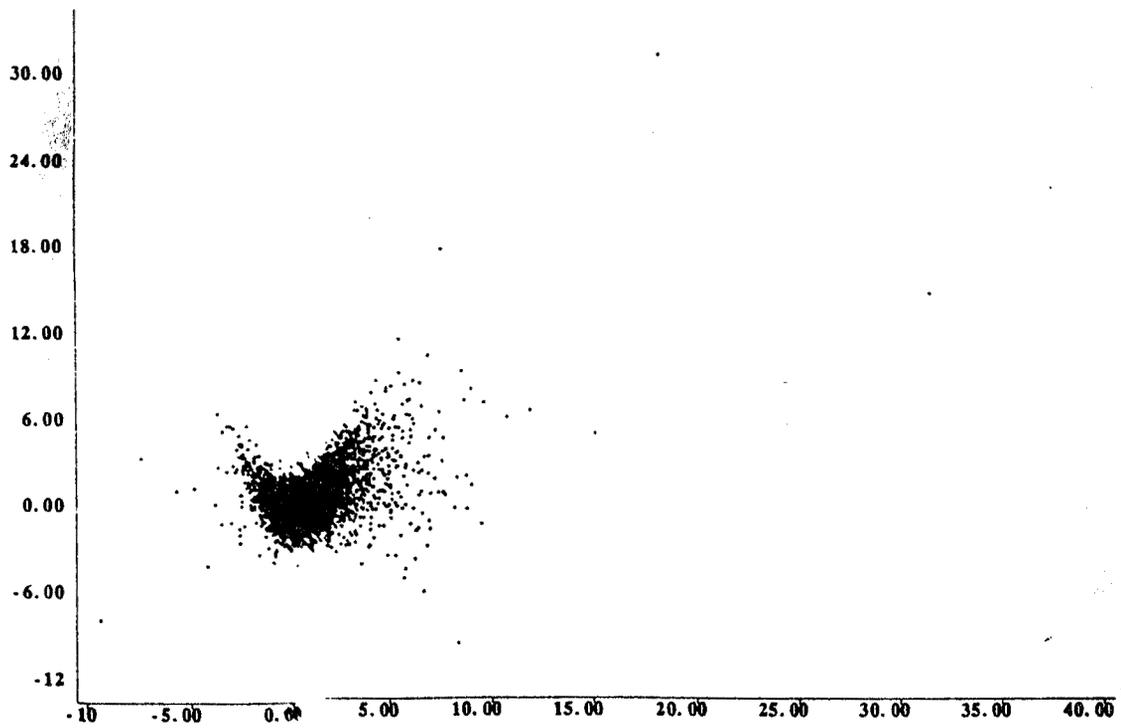


fig. 7.2

$$X(t) = .5 X(t-1) + .85 X(t-1)e(t-1) + e(t) + .5 e(t-1)$$

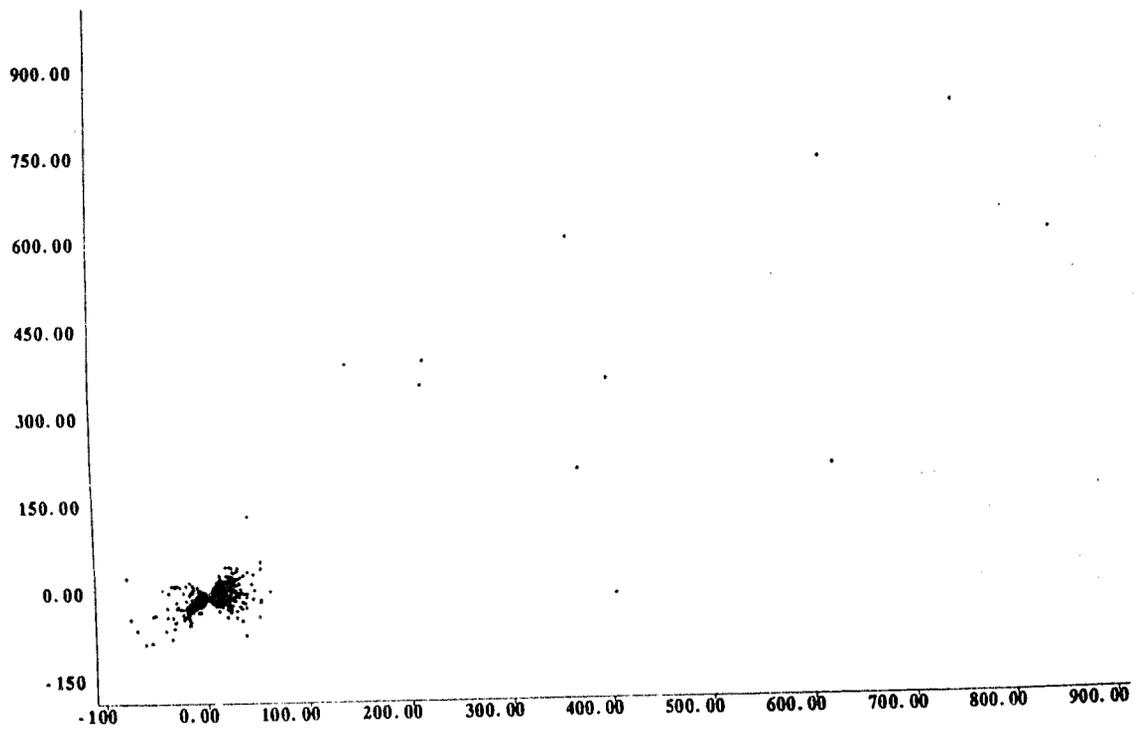


fig. 8.2

$$X(t) = .2 X(t-1) + .3 X(t-1)e(t-1) + e(t) + .5 e(t-1)$$

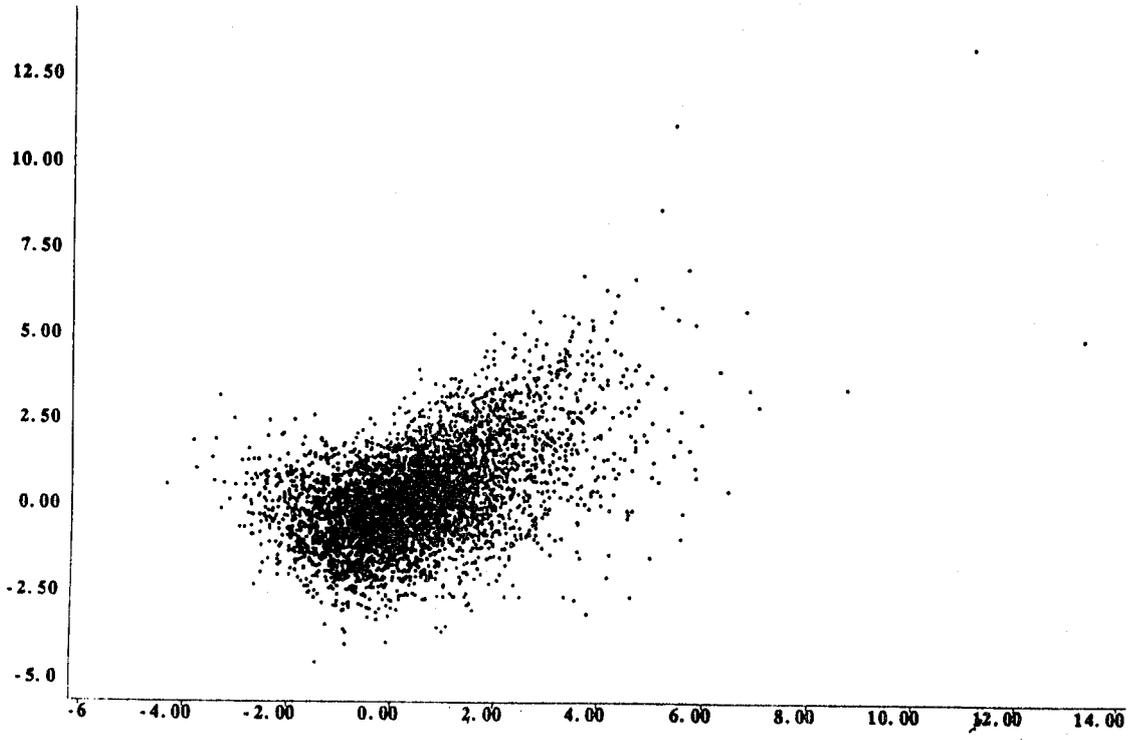


fig. 9.2

$$X(t) = .2 X(t-1) + .55 X(t-1)e(t-1) + e(t) + e(t-1)$$

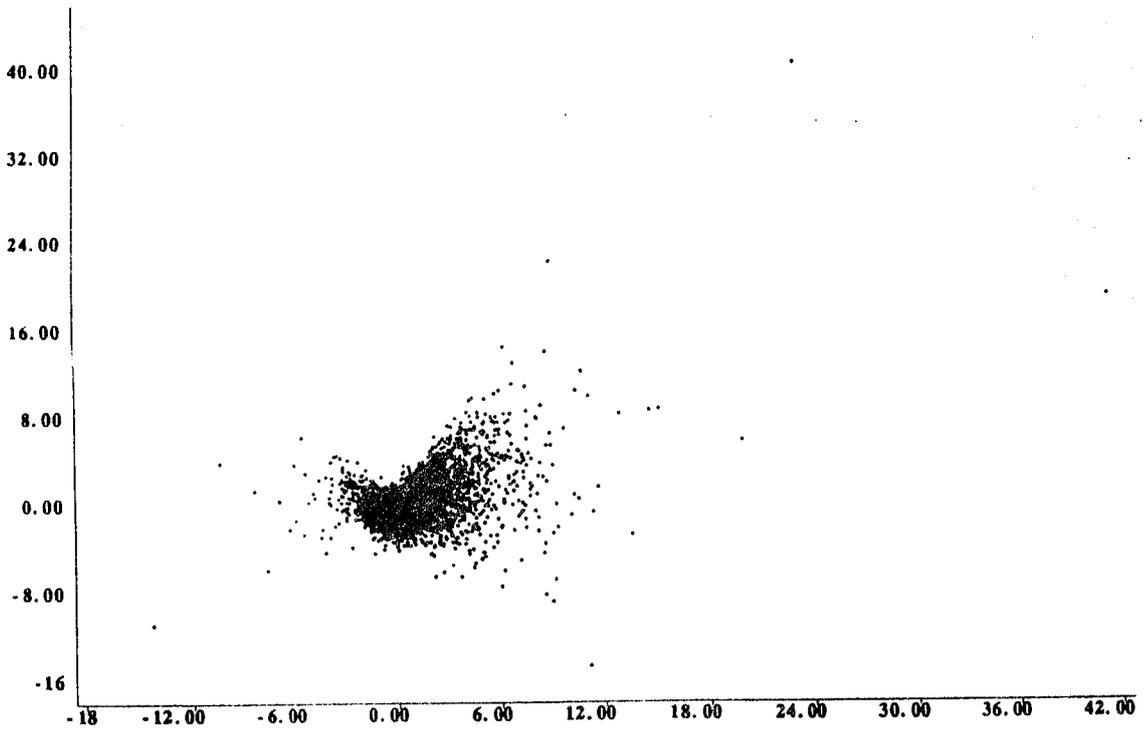


fig. 10.2

2 - Comportement des trajectoires.

On a déjà remarqué le comportement explosif des trajectoires de certains processus bilinéaires , cf. T. SUBBA RAO (1981). Nous voulons ici faire une étude plus systématique de ce phénomène pour déceler les facteurs qui provoquent les explosions, qui influent sur le type, la fréquence et l'importance de ces explosions. Les trajectoires des 10 modèles sur une longueur de 1000 points sont présentées dans les figures 1.3 à 10.3. Cela nous permet de déceler les séries explosives.

Les séries explosives sont les série 2, 4, 5 et 8 dont on a déjà remarqué la forte proportion de valeurs extrêmes (cf. § II.1.a). Ce sont aussi des séries avec un terme bilinéaire important.

$$X(t) = 1.55 X(t-1) - 0.81 X(t-2) + 0.1 X(t-1)e(t-1) + e(t)$$

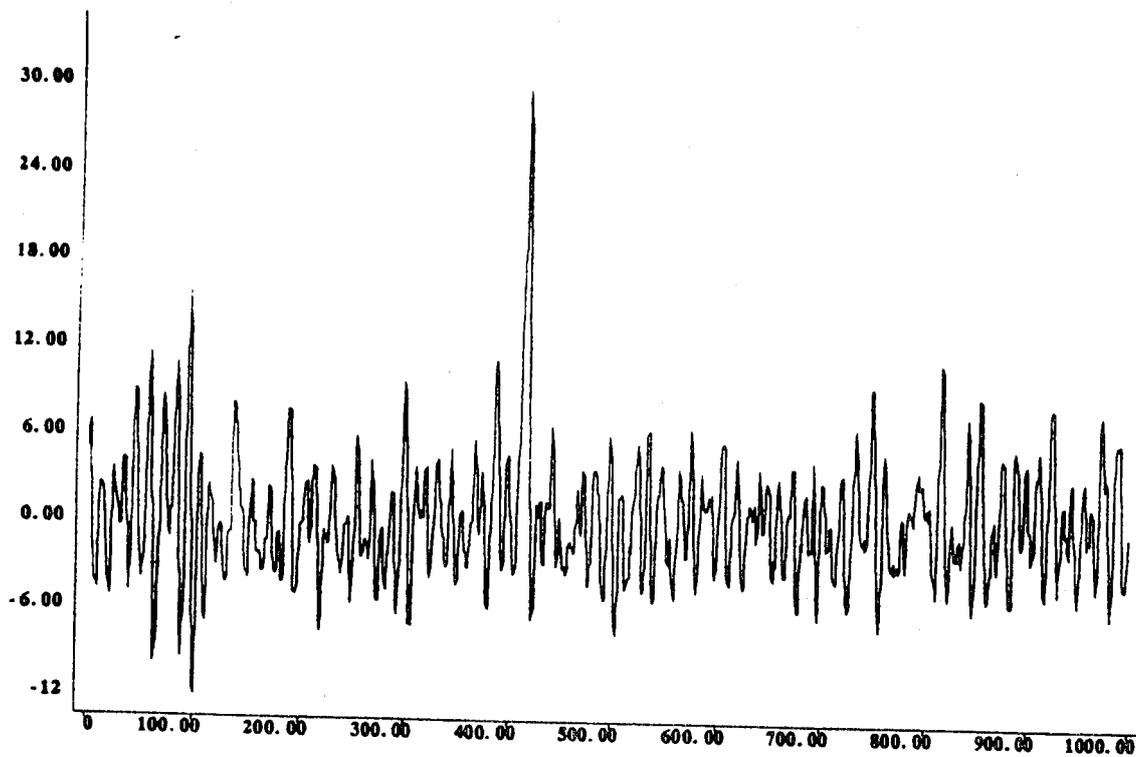


fig. 1.3

$$X(t) = 1.55 X(t-1) - 0.81 X(t-2) + 0.4 X(t-1)e(t-1) + e(t)$$

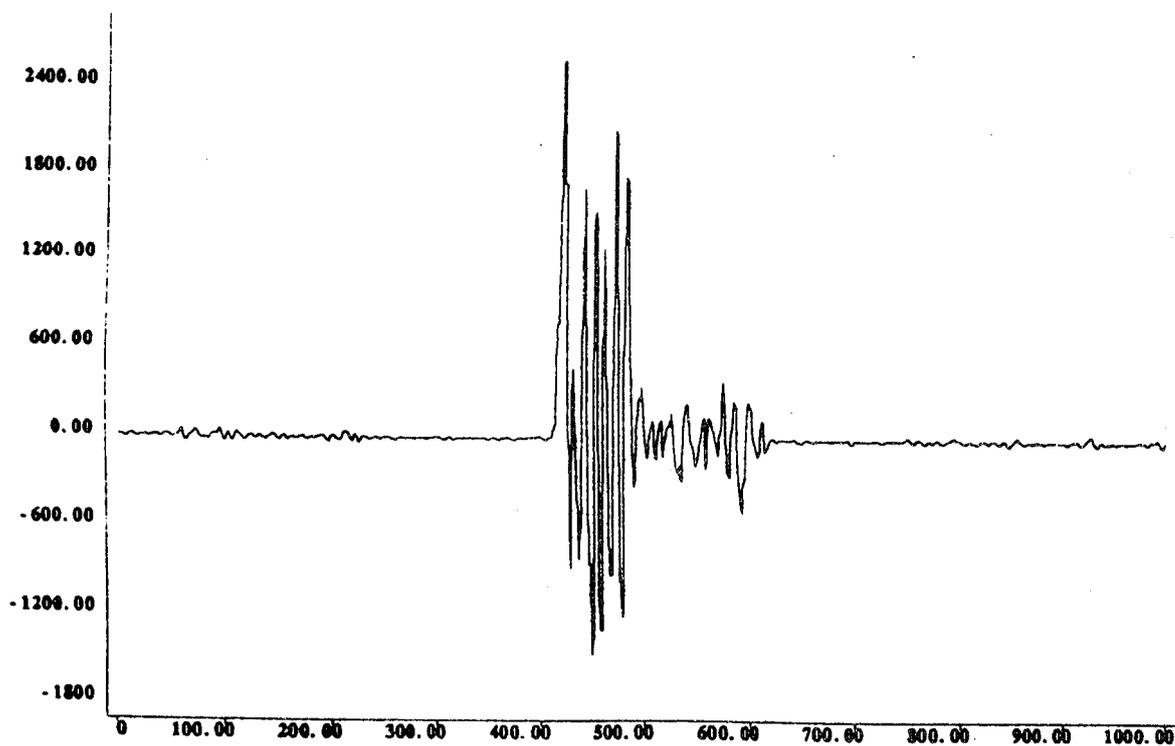


fig. 2.3

$$X(t) = 0.9 X(t-1) - 0.81 X(t-2) + 0.4 X(t-1)e(t-1) + e(t)$$

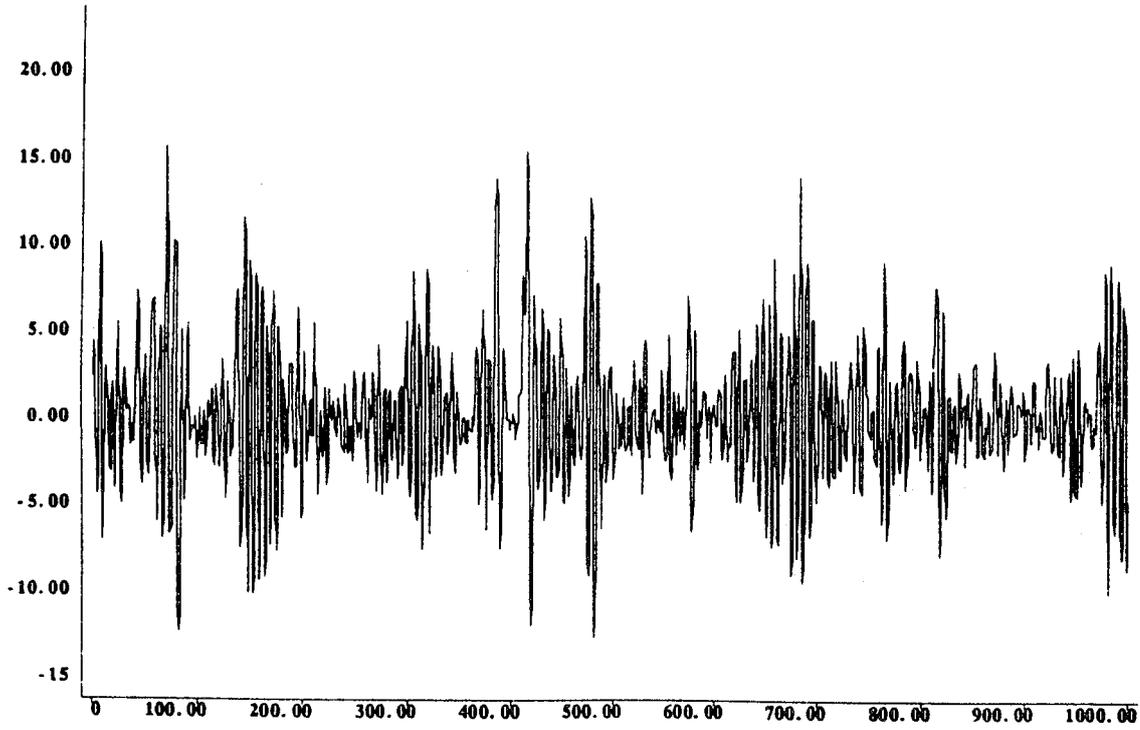


fig. 3.3

$$X(t) = 0.86 X(t-1) - 0.25 X(t-2) + 0.60 X(t-1)e(t-1) + e(t)$$

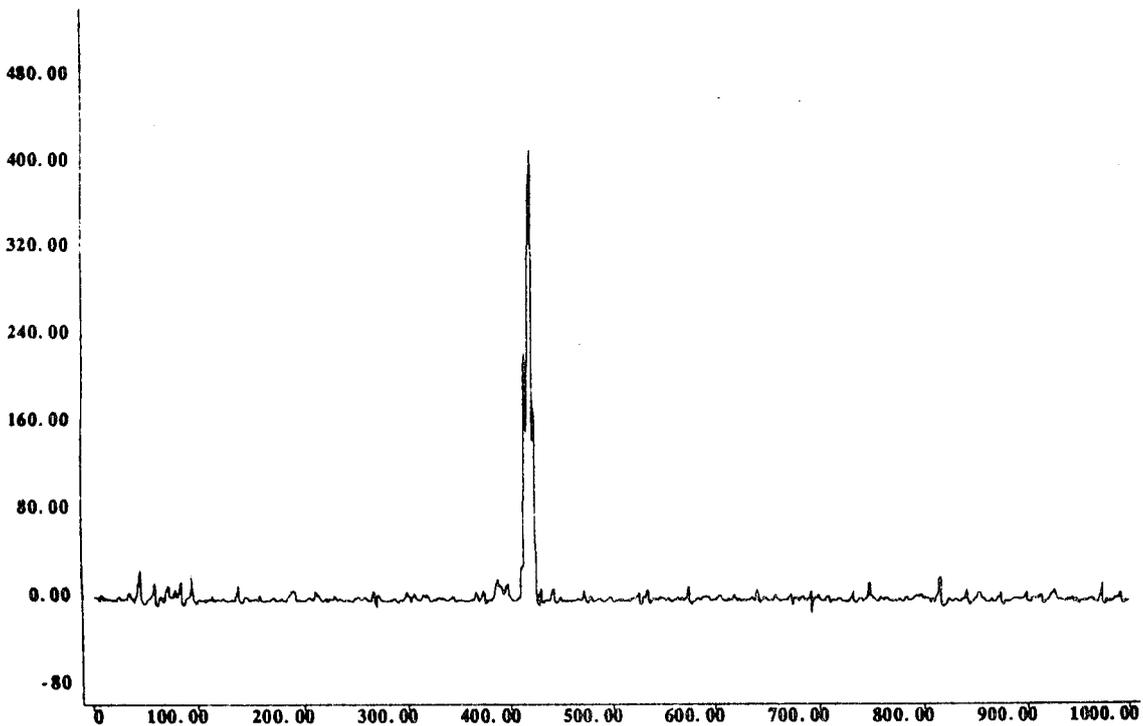


Fig. 4.3

$$X(t) = 0.5 X(t-1) + 0.85 X(t-1)e(t-1) + e(t)$$

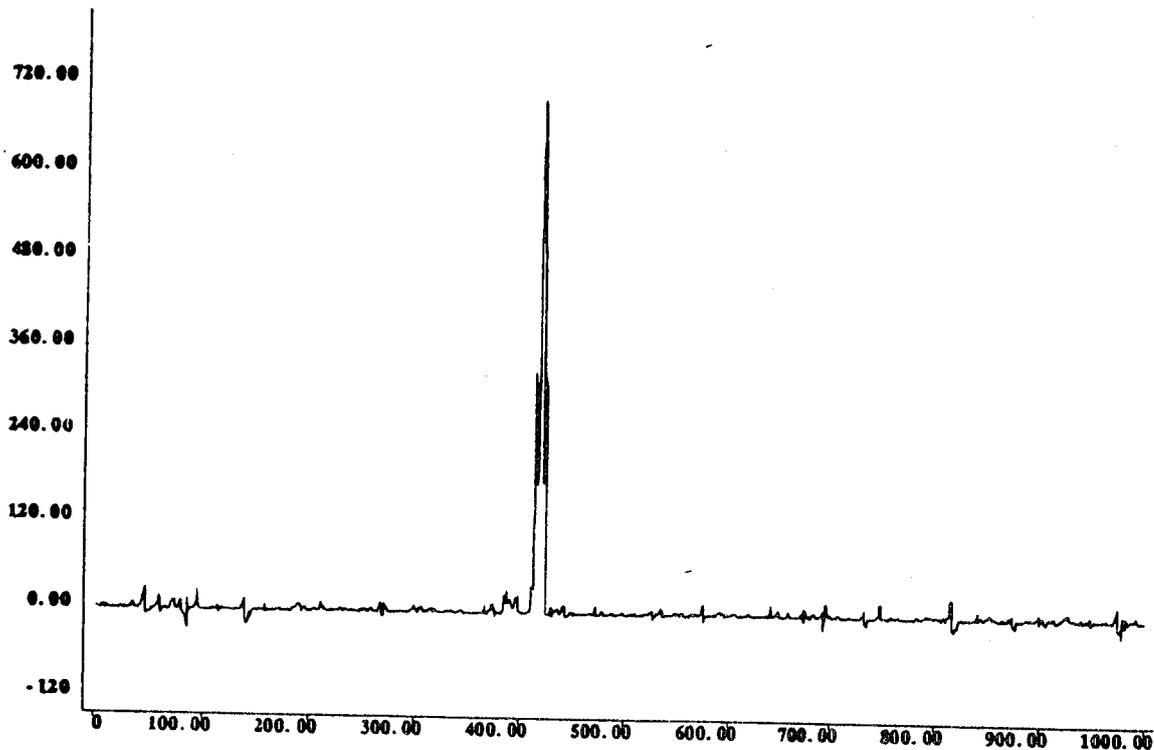


fig. 5.3

$$X(t) = 0.2 X(t-1) + 0.3 X(t-1)e(t-1) + e(t)$$

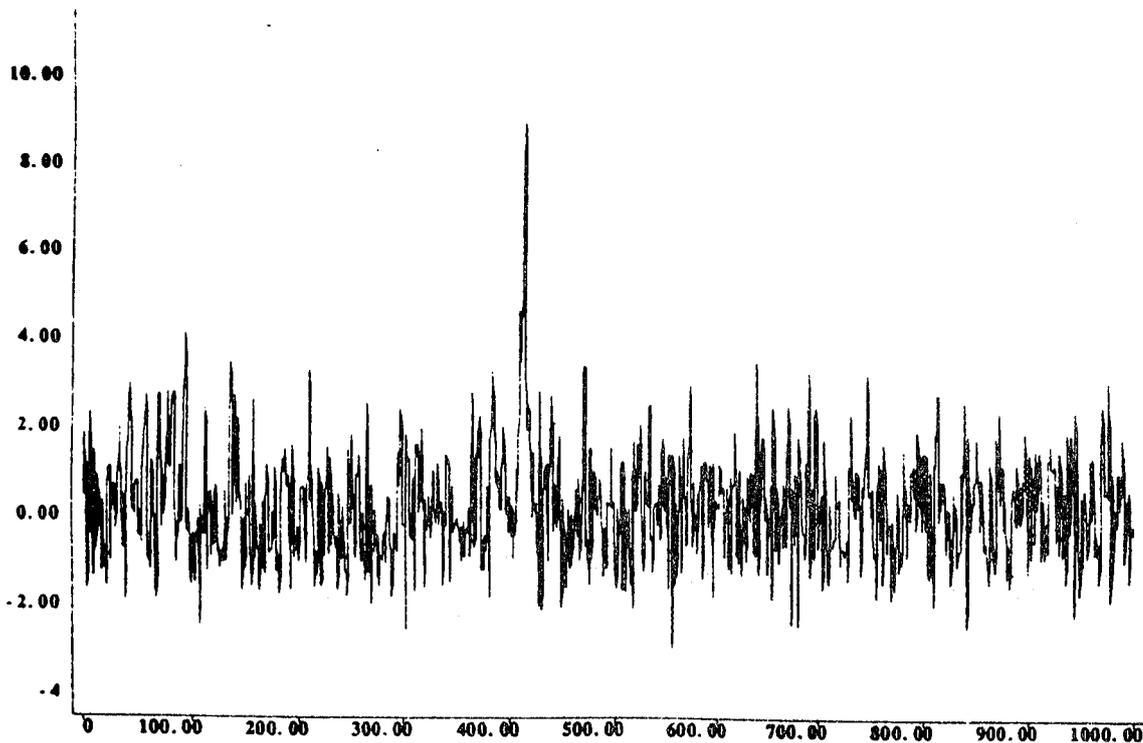


fig. 6.3

$$X(t) = 0.2 X(t-1) + 0.55 X(t-1)e(t-1) + e(t)$$

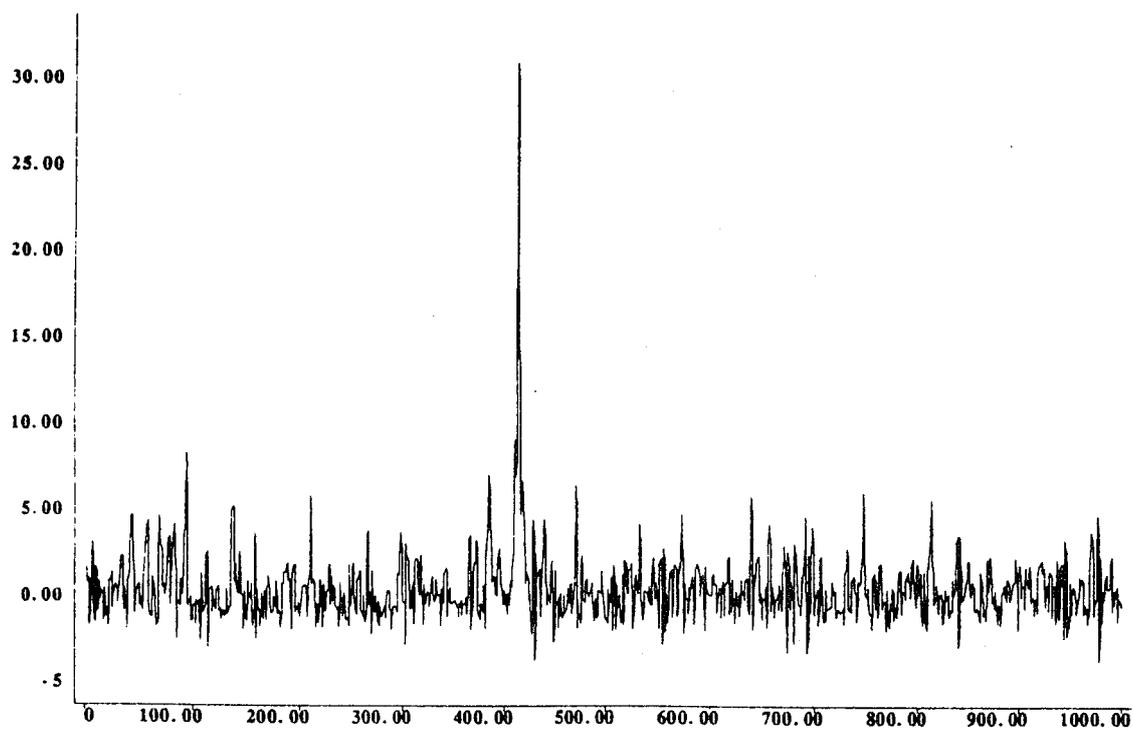


fig. 7.3

$$X(t) = 0.5 X(t-1) + 0.5 e(t-1) + 0.85 X(t-1)e(t-1) + e(t)$$

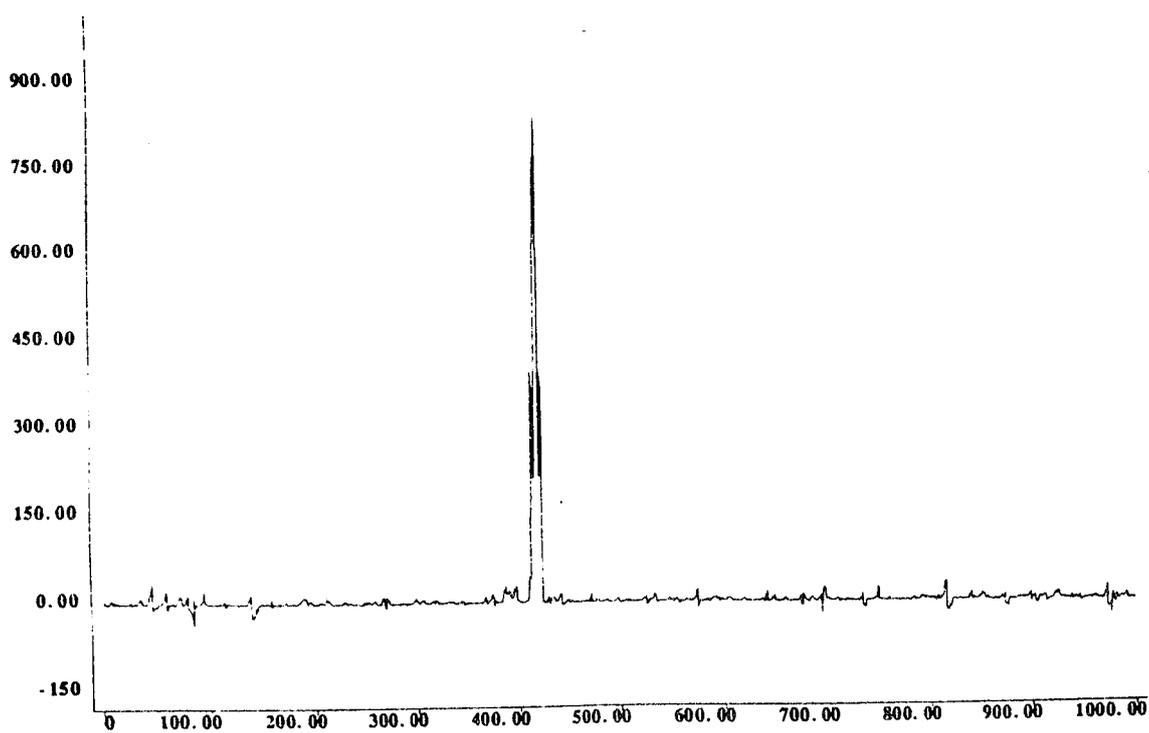


fig. 8.3

$$X(t) = 0.2 X(t-1) + 0.5 e(t-1) + 0.3 X(t-1)e(t-1) + e(t)$$

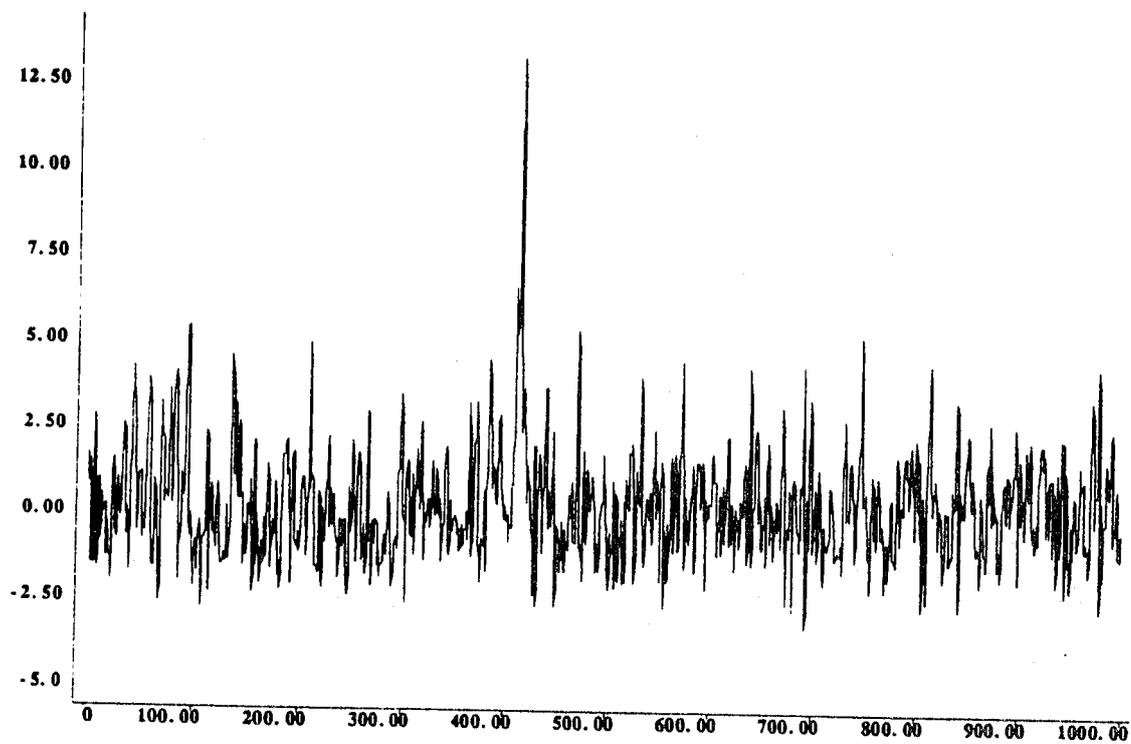


fig. 9.3

$$X(t) = 0.2 X(t-1) + 0.5 e(t-1) + 0.55 X(t-1)e(t-1) + e(t)$$

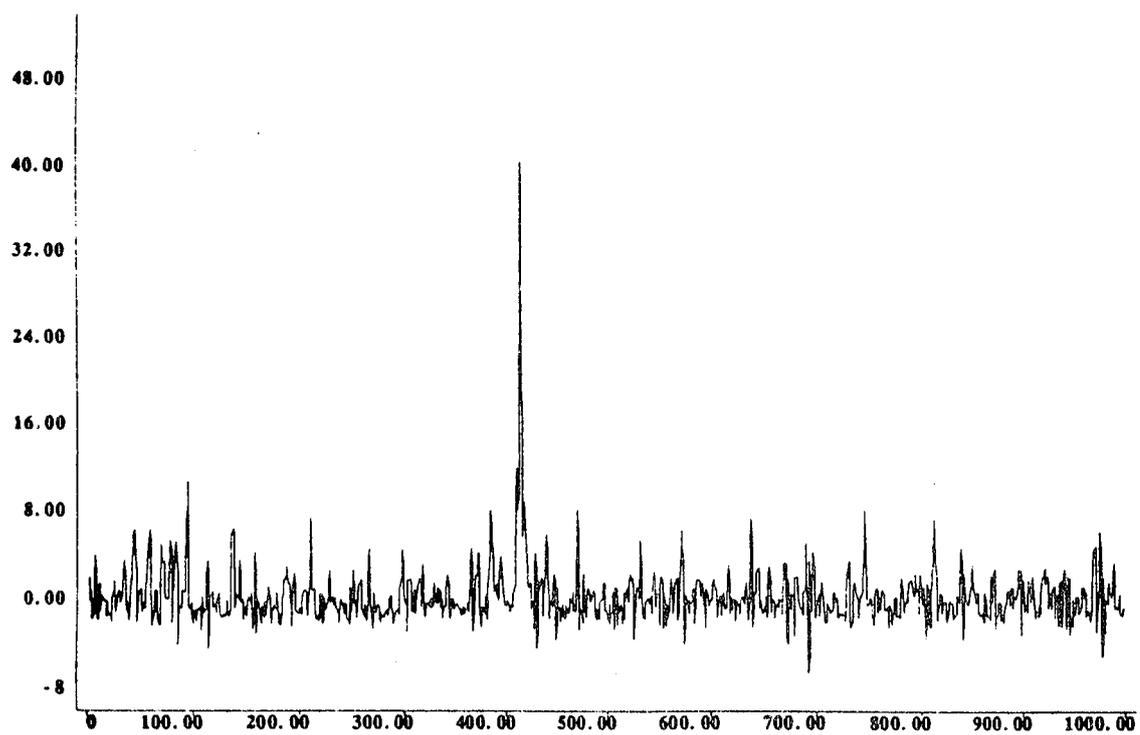


fig. 10.3

Considérons pour l'instant le modèle bilinéaire d'ordre 1 défini par l'équation (2.1.2). Ce modèle peut se réécrire sous la représentation markovienne suivante, cf. chapitre V :

$$(2.2.1) \quad \begin{cases} X(t) = Z(t-1) + \epsilon(t) \\ Z(t) = [a_1 + b_1 \epsilon(t)] Z(t-1) + [a_1 + b_1 \epsilon(t)] \epsilon(t) \\ \quad = A(t) Z(t-1) + B(t) \end{cases}$$

Cette représentation met en évidence le rôle du facteur multiplicatif $a_1 + b_1 \epsilon(t)$ dans la dynamique du processus. Nous pensons que ce facteur est responsable de l'explosion observée sur ces trajectoires. Nous avons donc tracé, pour les séries 5 et 7, sur deux graphiques parallèles les trajectoires du processus et de son facteur multiplicatif.

Sur les graphiques représentant les facteurs multiplicatifs, fig. 5.4 et 7.4, on peut remarquer un décrochement juste avant l'explosion de la série, cf. fig. 5.5 et 7.5. Dans le cas de la série 5, on constate que le facteur multiplicatif dépasse alors le seuil 1, tandis que pour la série 7 le dépassement est plus faible et ne dépasse pas 1.

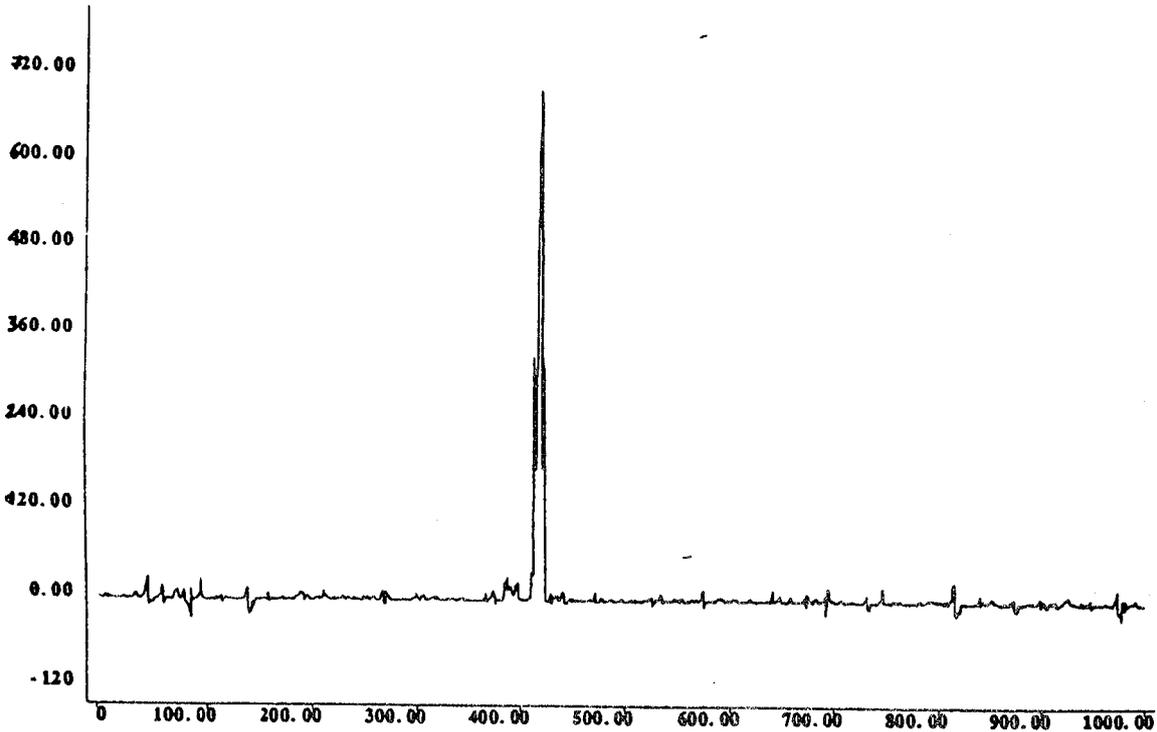
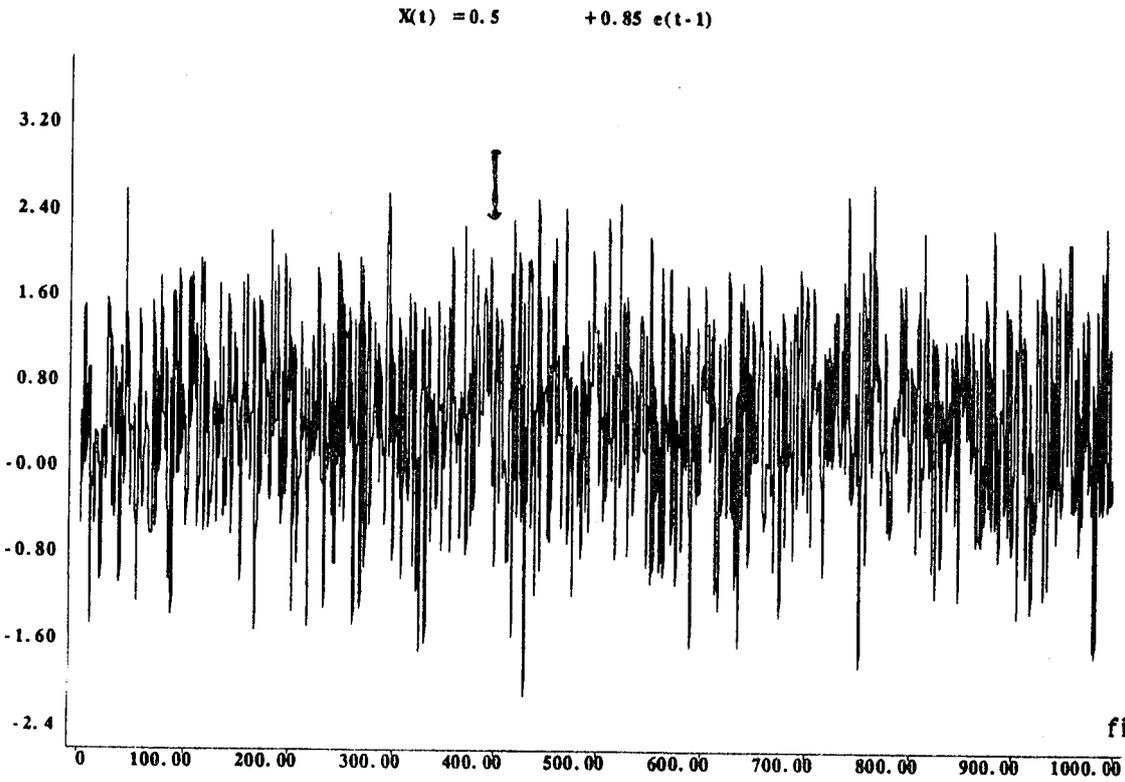


fig. 5.5

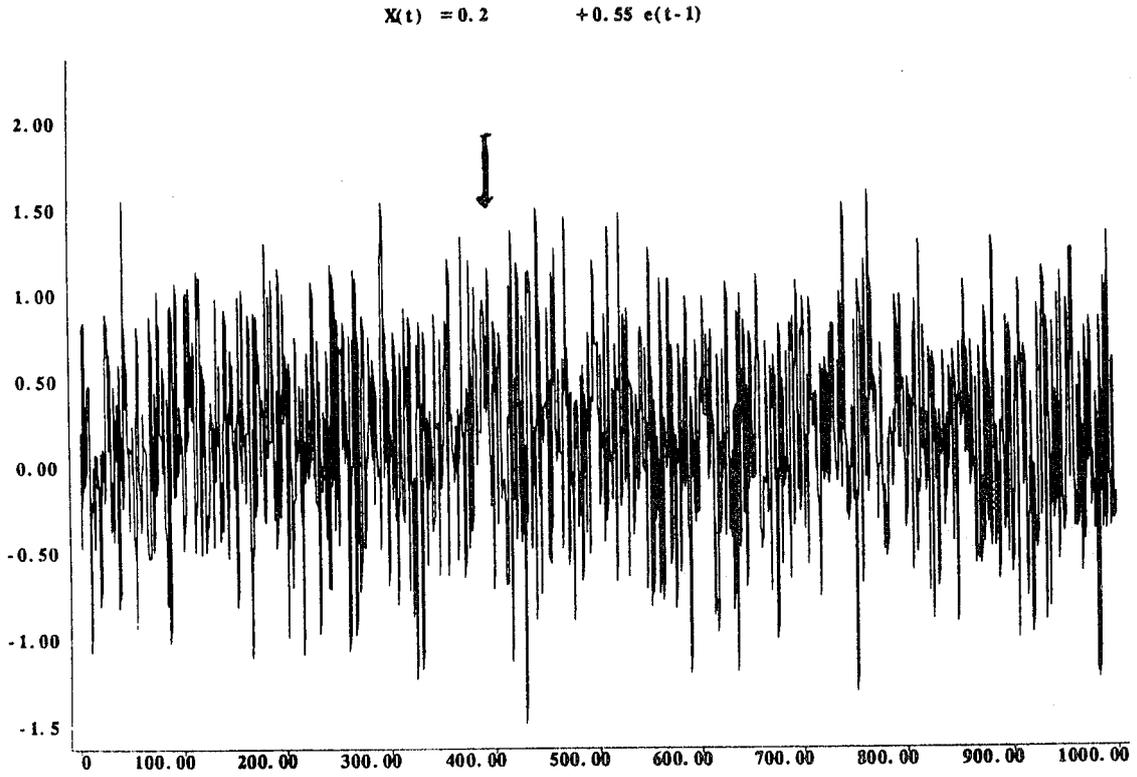


fig. 7.4

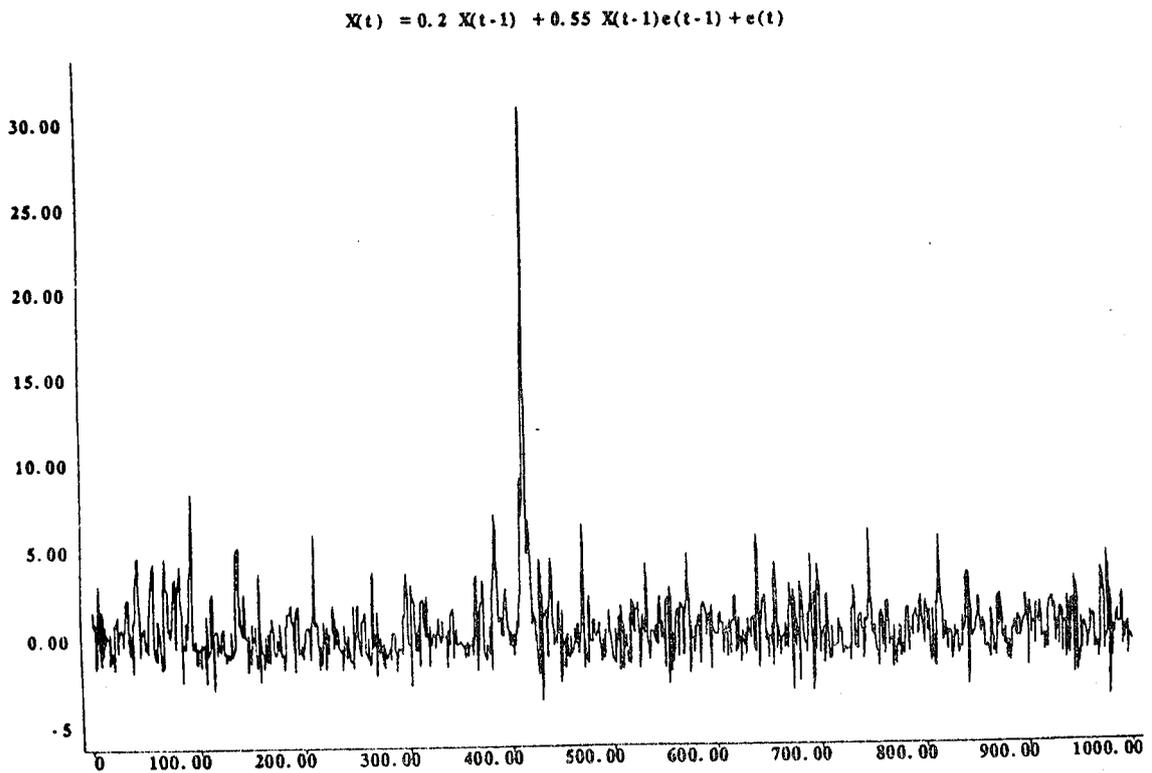


fig. 7.5

3. RÉCAPITULATION :

Dans le tableau suivant, nous résumons les remarques faites précédemment pour les 10 séries:

Séries	Distributions marginales	Distributions conjointes	Explosions	Remarques
Série 1	Asymétrique Faible proportion de valeurs extrêmes	Nuage de points de forme "ovale"	Explosion faible de valeur positive	Racines du polynôme AR proches de l'unité. Sans partie MA. Faible comp. bil.
Série 2	Symétrique Forte proportion de valeurs extrêmes	Nuage de points de forme "ovale", plus concentré en zéro	Explosion très forte de valeur positive	Racines du polynôme AR proches de l'unité. Sans partie MA. Forte comp. bil.
Série 3	Symétrique Faible proportion de valeurs extrêmes	Nuage de points de forme "ovale"	Pas d'explosion Comportement périodique	Racines du polynôme AR proches de l'unité. Sans partie MA. Faible comp. bil.
Série 4	Asymétrique Forte proportion de valeurs extrêmes	Nuage de points très concentré en forme de "papillon"	Explosion forte de valeur positive	Deux termes AR. Pas de partie MA. Forte comp. bil.
Série 5	Asymétrique Forte proportion de valeurs extrêmes	Nuage de points très concentré en forme de "papillon"	Explosion très forte de valeur positive	Pas de partie MA. Forte comp. bil.
Série 6	Asymétrique Faible proportion de valeurs extrêmes	Nuage de points en forme de "bol"	Explosion très faible de valeur positive	Pas de partie MA. Faible comp. bil.
Série 7	Asymétrique Longue queue Forte proportion de valeurs extrêmes	Nuage de points en forme de "bol"	Explosion faible de valeur positive	Pas de partie MA. Comp. bil. moyenne
Série 8	Asymétrique. Longue queue Très forte proportion de valeurs extrêmes	Nuage de points très concentré en forme de "papillon"	Très forte explosion de valeur positive	Un terme MA. Forte comp. bil.
Série 9	Asymétrique Faible proportion de valeurs extrêmes	Nuage de points en forme de "bol"	Explosion très faible de valeur positive	Un terme MA. Faible comp. bil.
Série 10	Asymétrique Longue queue Forte proportion de valeurs extrêmes	Nuage de points en forme de "bol"	Explosion faible de valeur positive	Un terme MA. Comp. bil. moyenne

Tableau 2.2

III . ETUDE DU PHENOMENE EXPLOSIF

Pour des raisons matérielles, nous allons nous limiter au modèle d'ordre 1 avec ou sans terme moyenne mobile, eq. (2.1.2) et (2.1.3) . Le tableau 3.1 ci-dessous indique les différents choix de paramètres utilisés dans les simulations . La case correspondante dans ce tableau contient le numéro de la figure, et dans la parenthèse la valeur du coefficient de corrélation entre $A(t)$ et $B(t)$:

b	a	c		
		0	0,95	- 0,95
0,3	0,2	11 (0,49)	12 (0,19)	13 (0,34)
	0,8	14 (0,33)	15 (0,14)	16 (0,44)
	- 0,8	17 (-0,20)	18 (-0,13)	19 (= 0)
- 0,3	0,2	20 (-0,49)	21 (-0,19)	22 (-0,34)
	0,8	23 (- 0,33)	24 (- 0,14)	25 (- 0,44)
	- 0,8	26 (0,20)	27 (0,13)	28 (+ 0)
0,55	0,2	29 (0,56)	30 (0,31)	31 (0,63)
	- 0,2	32 (=0)	33 (+ 0)	34 (+0)
	0,8	35 (0,71)	36 (0,39)	37 (0,83)
	- 0,8	38 (-0,35)	39 (- 0,20)	40 (-0,43)
- 0,55	0,2	41 (- 0,56)	42 (- 0,31)	43 (-0,63)
	- 0,2	44 (0)	45 (0)	46 (0)
	0,5	47 (- 0,60)	48 (-0,34)	49 (1)
	0,8	50 (- 0,71)	51 (- 0,39)	52 (-0,83)
	- 0,8	53 (0,35)	54 (0,20)	55 (0,43)
0,85	0,2	56 (0,66)	57 (0,45)	58 (0,88)
	- 0,2	59 (0,31)	60 (0,22)	61 (0,30)
	0,5	62 (0,72)	63 (0,53)	64 (1)
	- 0,5	65 (- 0,70)	66 (- 0,37)	67 (- 0,42)
- 0,85	0,2	68 (- 0,66)	69 (- 0,45)	70 (- 0,88)
	-0,2	71 (- 0,31)	72 (- 0,22)	73 (- 0,30)
	0,5	74 (- 0,72)	75 (- 0,53)	76 (- 1)
	- 0,5	77 (0,70)	78 (0,37)	79 (0,42)

Tableau 3.1

Nous nous intéressons aux points suivants : type de l'explosion (valeurs positives, ou négatives, ou alternées positives et négatives), importance et fréquence des explosions. Nous nous efforçons de déterminer le rôle des paramètres du modèle sur les caractéristiques précédentes de l'explosion. Nous pensons que les signes respectifs des paramètres autorégressif et bilinéaire jouent un rôle ainsi que l'importance relative de leurs valeurs absolues. Par ailleurs, la présence ou l'absence du terme moyenne mobile peut avoir une certaine influence sur ces caractéristiques. D'autre part, à partir de l'équation (2.2.1), on s'attend à ce que l'explosion ait lieu si $A(t)$ est grand par rapport à 1, mais pour que l'explosion puisse durer et avoir une amplitude importante il faut que l'effet de $A(t)$ ne soit pas annulé par celui de $B(t)$. Cet argument conduit à considérer la corrélation entre les processus $A(t)$ et $B(t)$. Si $A(t)$ et $B(t)$ sont corrélés positivement, on peut s'attendre à ce que le processus explose plus souvent et plus fortement, au contraire si $A(t)$ et $B(t)$ sont corrélés négativement, l'effet de $B(t)$ tend à diminuer celui de $A(t)$, ce qui tend à produire des explosions alternées positives et négatives avec une amplitude beaucoup plus faible qui se distinguent des oscillations. Le choix des paramètres précédents a été partiellement conçu de façon à avoir un coefficient de corrélation faible et fort de signe positif ou non, les valeurs des corrélations sont indiquées dans le tableau 3.1. Les résultats sont résumés dans le tableau 3.2. Des commentaires sur l'apparition des explosions se trouvent après les simulations p. 86 .

b	a	c		
		0	0,95	- 0,95
0,3	0,2	Pas d'explosion	Pas d'explosion	Pas d'explosion
	0,8	Explosion positive moyenne	Explosion positive forte	Explosion positive faible
	- 0,8	Pas d'explosion	Pas d'explosion	Pas d'explosion
-0,3	0,2	Pas d'explosion	Pas d'explosion	Pas d'explosion
	0,8	Explosion négative faible	Explosion négative faible	Pas d'explosion
	-0,8	Explosion alternée faible	Pas d'explosion	Explosion alternée moyenne
0,55	0,2	Explosion positive faible	Explosion positive faible	Pas d'explosion
	-0,2	Pas d'explosion	Explosion positive faible	Pas d'explosion
	0,8	Explosion positive très forte	Explosion positive très forte	Explosion positive très forte
	-0,8	Explosion alternée faible avec périodicité	Explosion alternée faible avec périodicité	Explosion alternée faible avec périodicité
- 0,55	0,2	Pas d'explosion	Pas d'explosion	Pas d'explosion
	-0,2	Pas d'explosion	Pas d'explosion	Explosion alternée faible
	0,5	Explosion négative faible	Explosion négative faible	Pas d'explosion
	0,8	Explosions alternée et négative faibles	Explosions alternée et négative moyennes	Explosions alternée et négative faibles
	-0,8	Explosion alternée forte	Explosion alternée faible	Explosion alternée très forte
0,85	0,2	Explosion positive forte	Explosion positive forte	explosion positive moyenne
	-0,2	Explosion positive faible	Explosion positive faible	Pas d'explosion
	0,5	Explosion positive très forte	Explosion positive très forte	Explosion positive forte
	-0,5	Explosion négative faible	Explosion alternée faible	Explosion alternée faible
-0,85	0,2	Pas d'explosion	Pas d'explosion	Explosion positive faible
	- 0,2	Explosion alternée faible	Explosion alternée faible	Explosion alternée moyenne
	0,5	Explosion négative faible	Explosion alternée faible	Explosion alternée moyenne
	-0,5	Explosion alternée forte	Explosion alternée moyenne	Explosion alternée très forte

Tableau 3.2

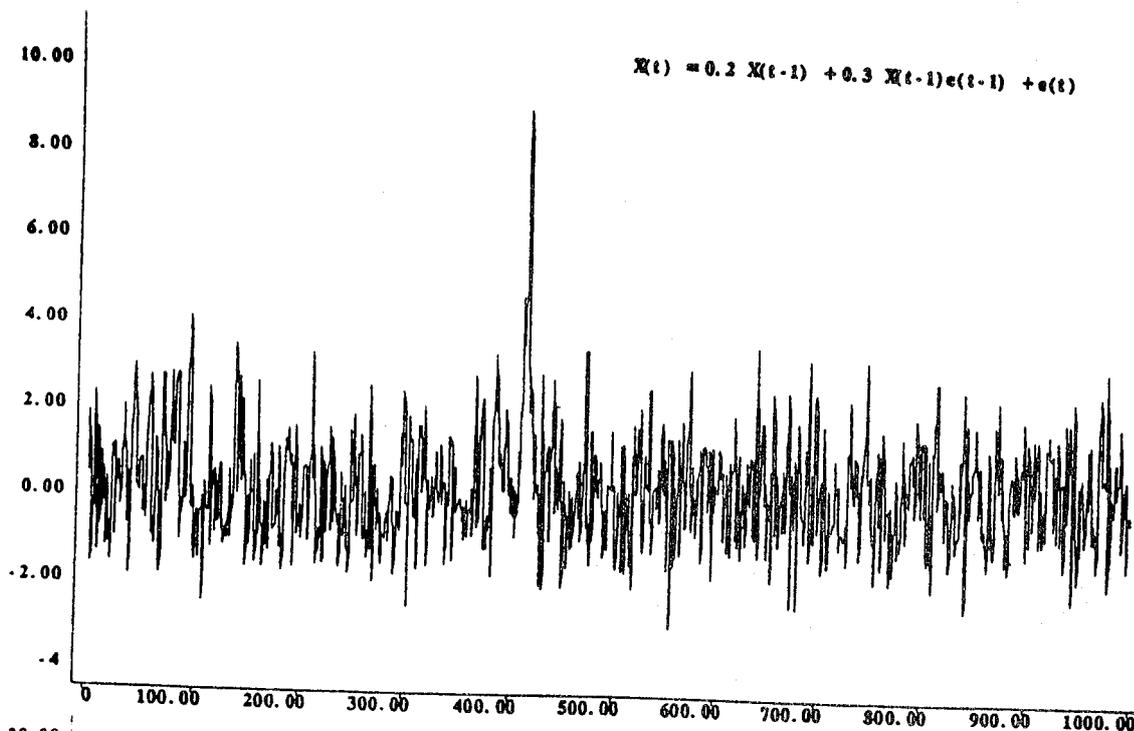


fig. 11

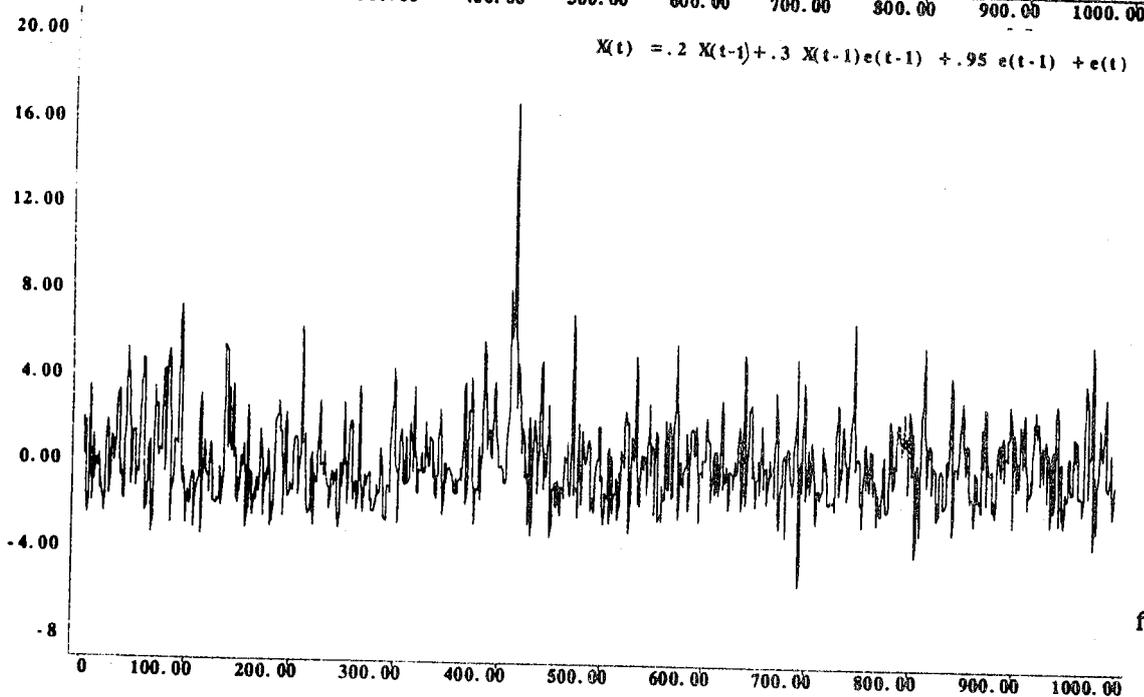


fig. 12

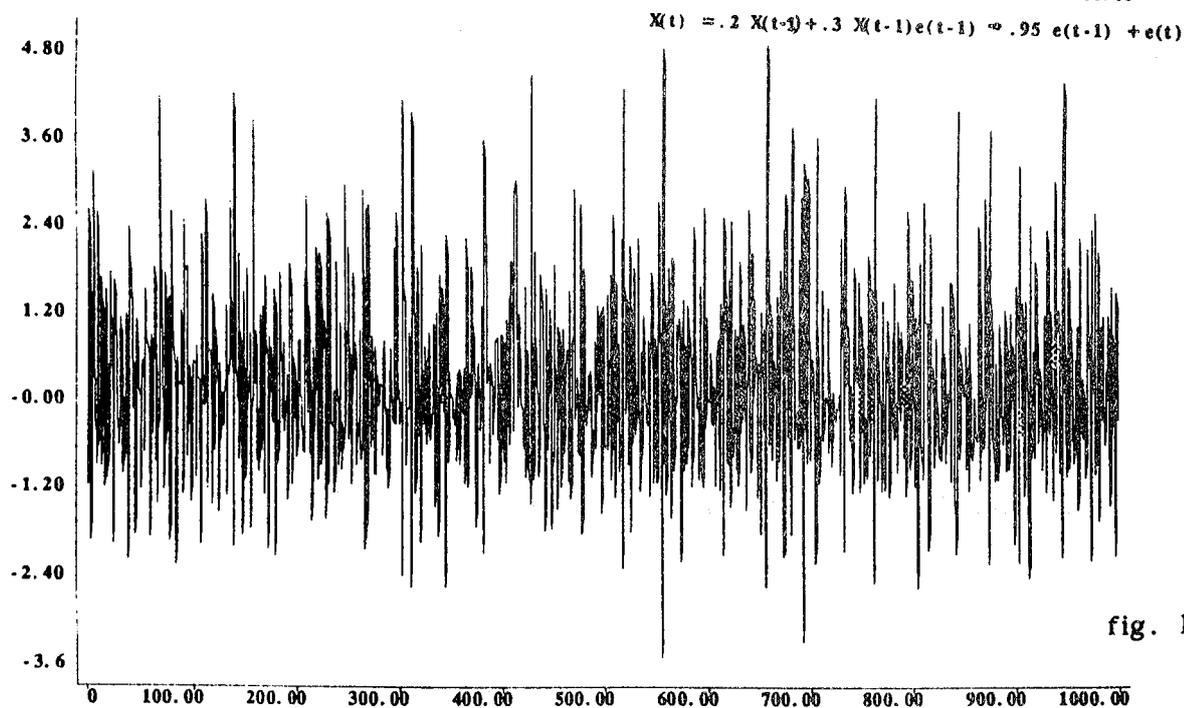
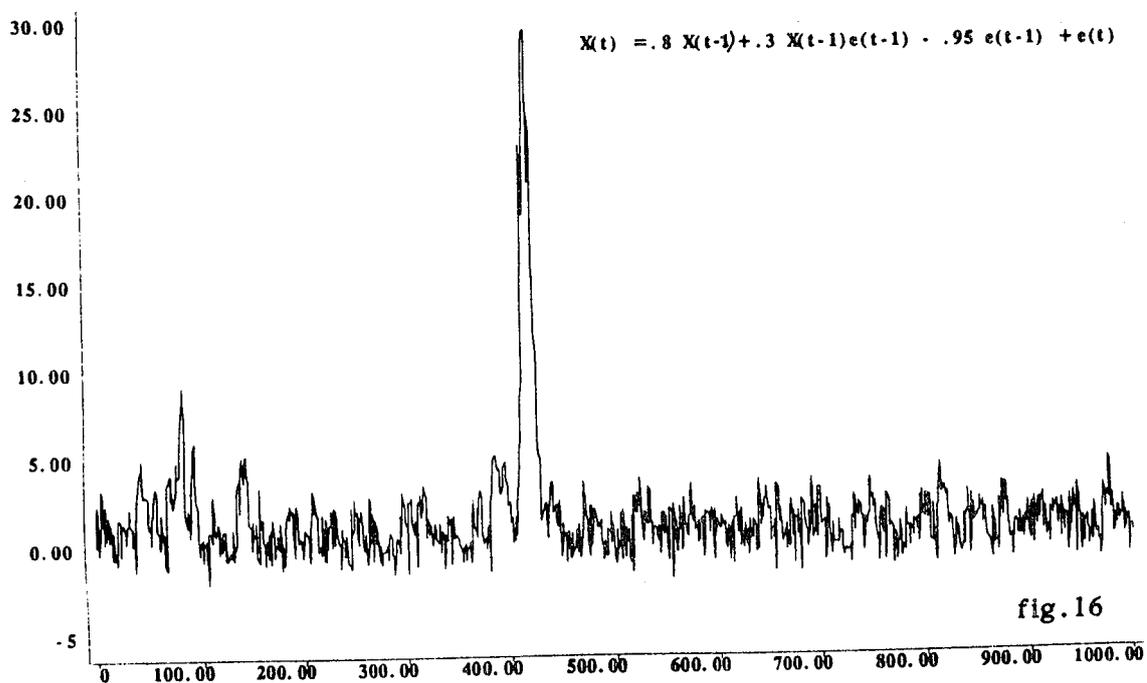
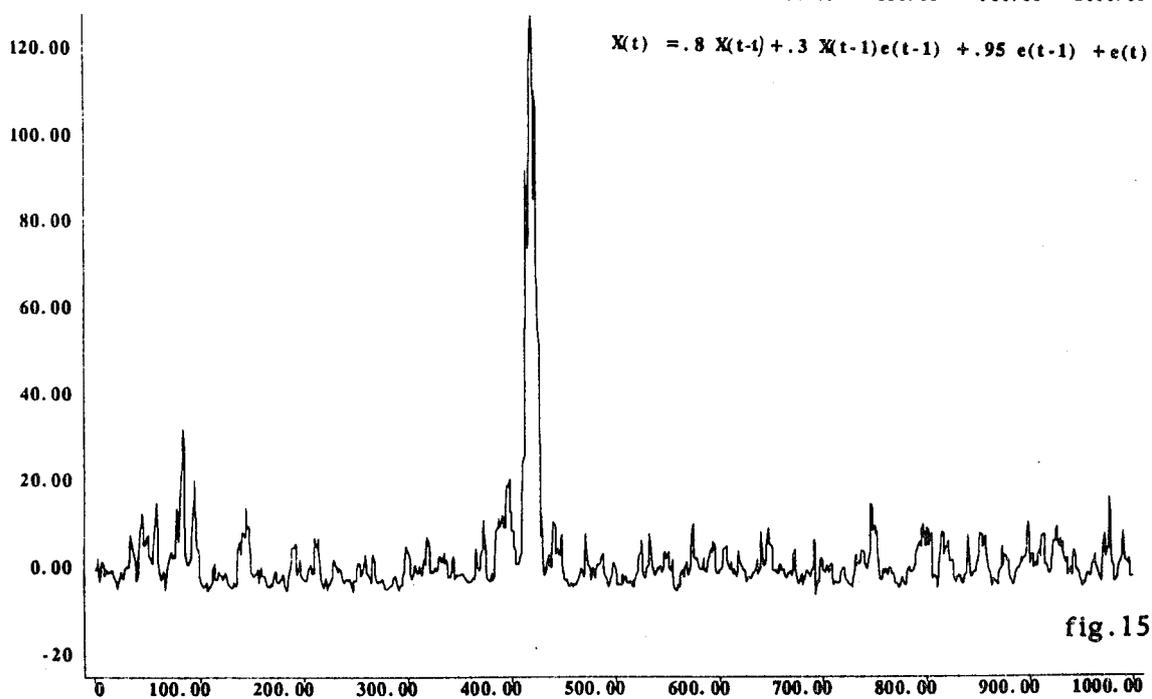
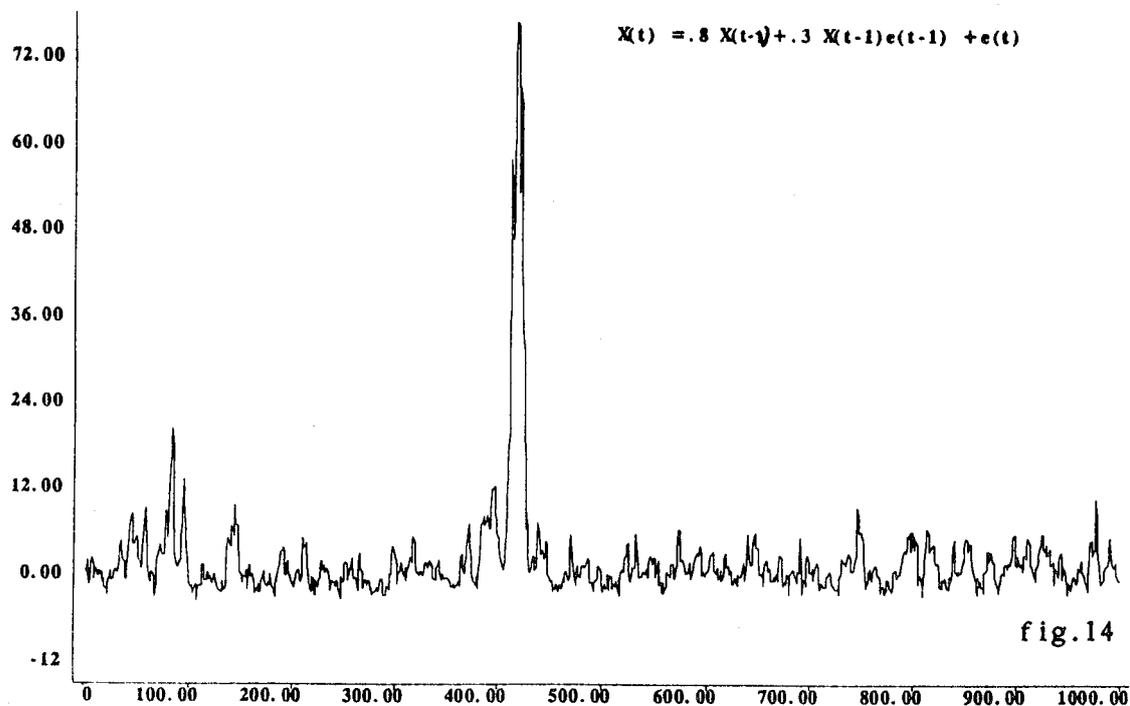
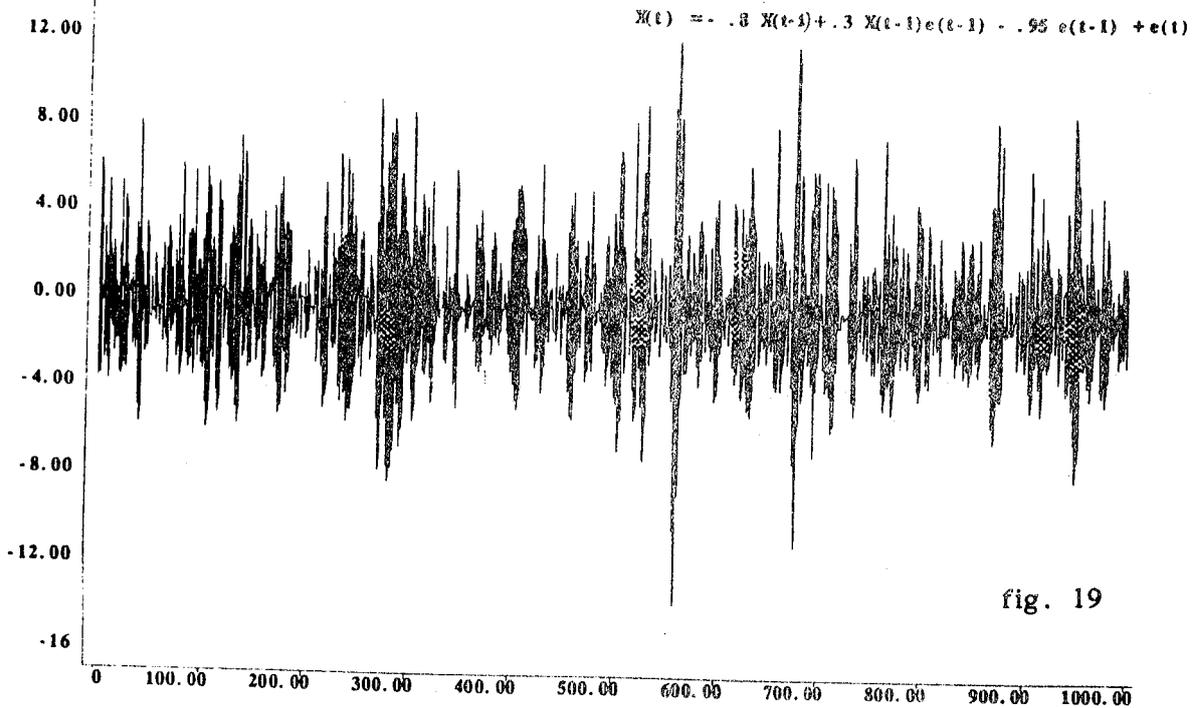
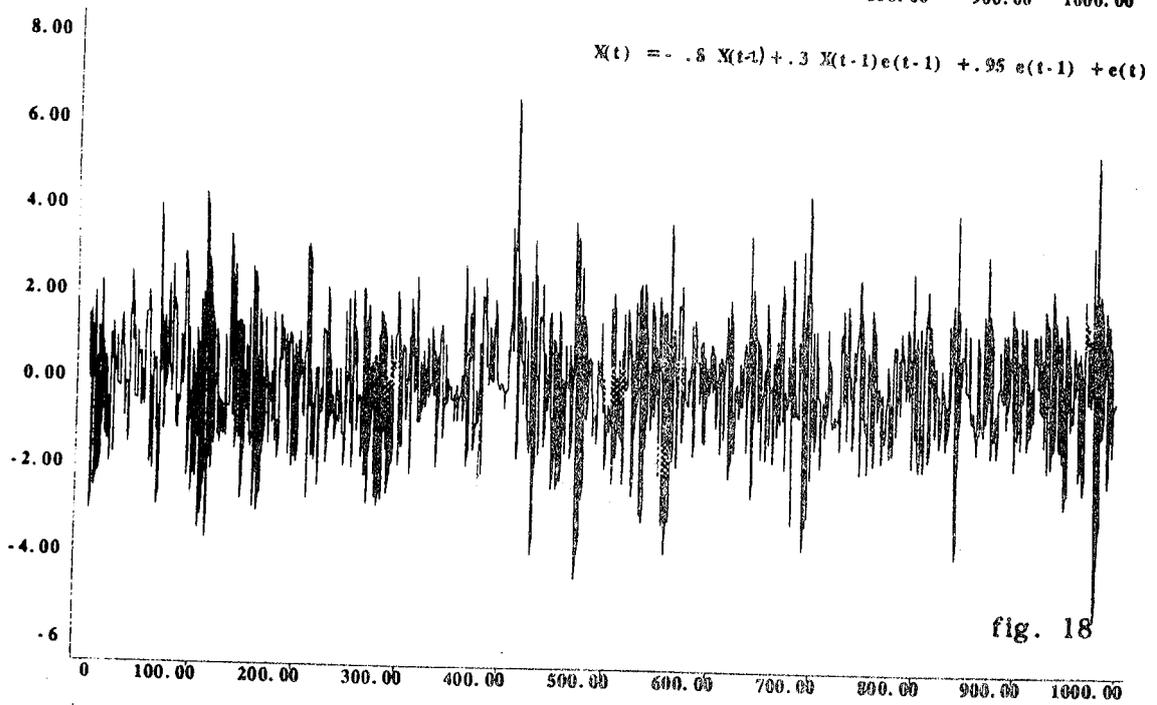
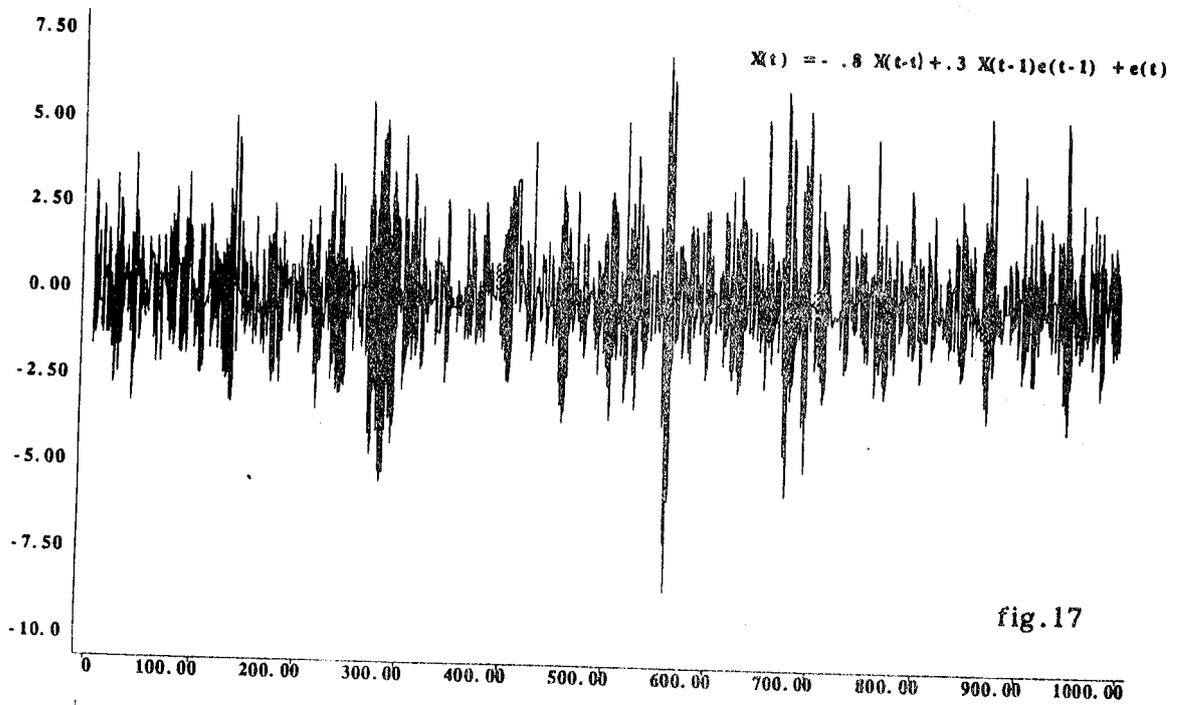


fig. 13





$$X(t) = .2 X(t-1) - .3 X(t-1)e(t-1) + e(t)$$

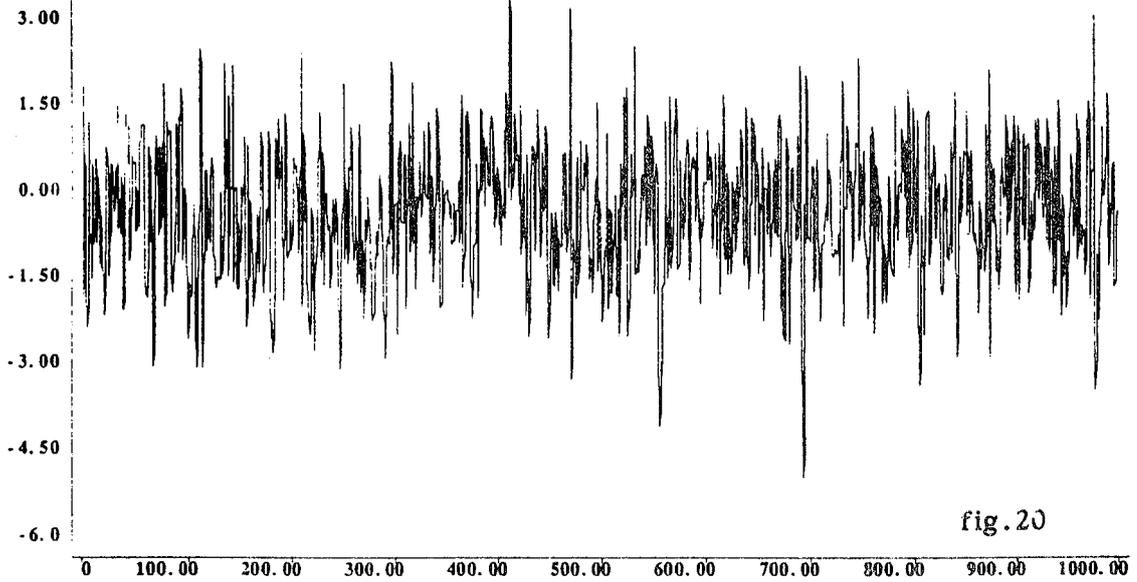


fig. 20

$$X(t) = .2 X(t-1) - .3 X(t-1)e(t-1) + .95 e(t-1) + e(t)$$

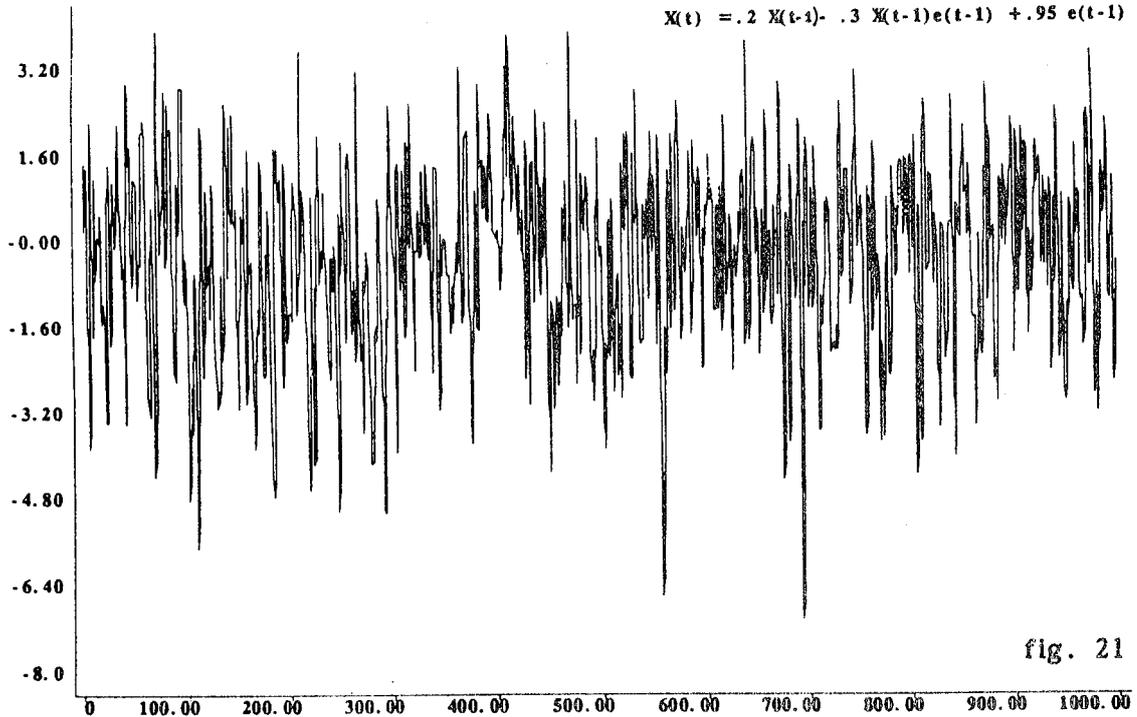


fig. 21

$$X(t) = .2 X(t-1) - .3 X(t-1)e(t-1) - .95 e(t-1) + e(t)$$

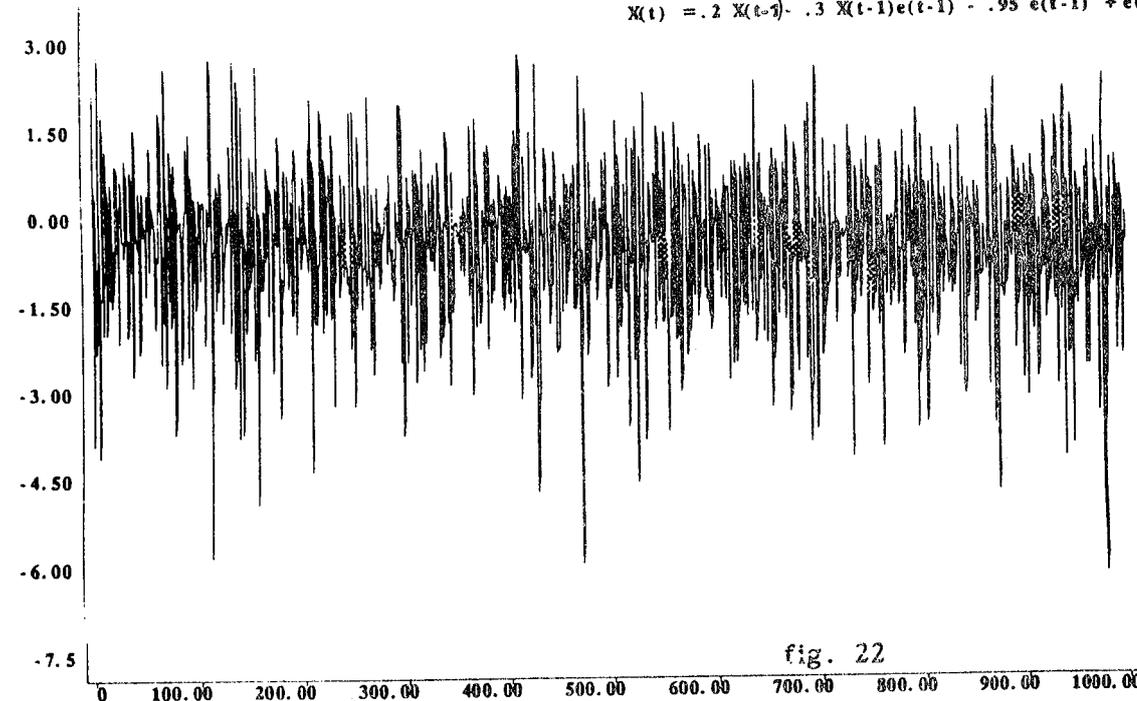
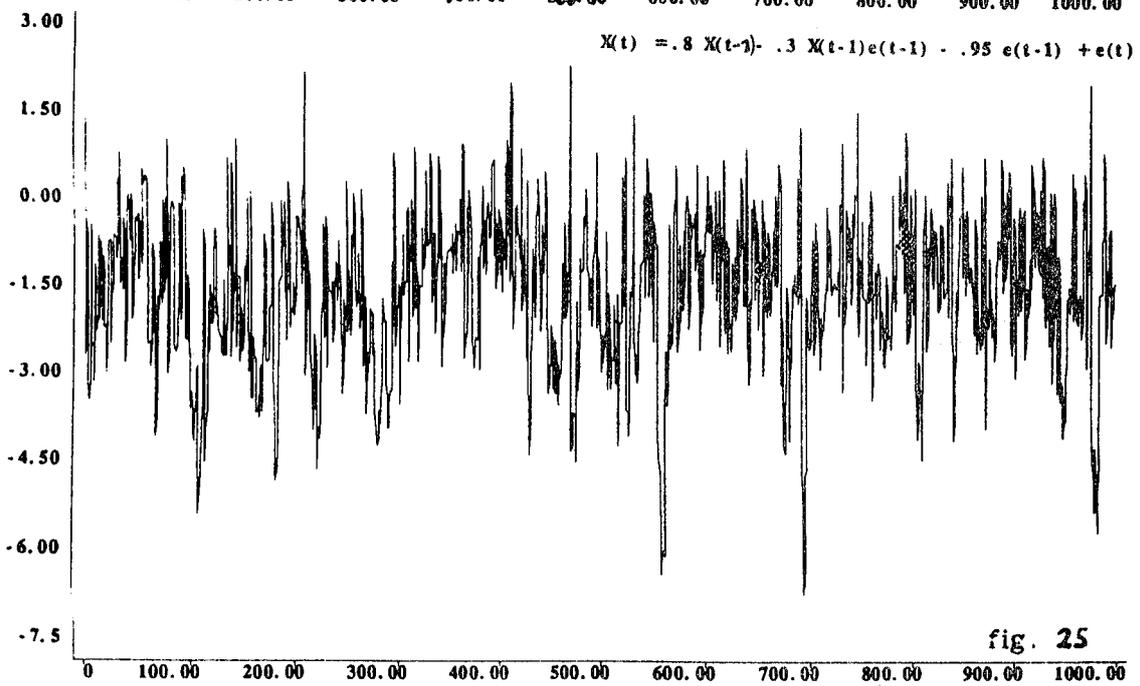
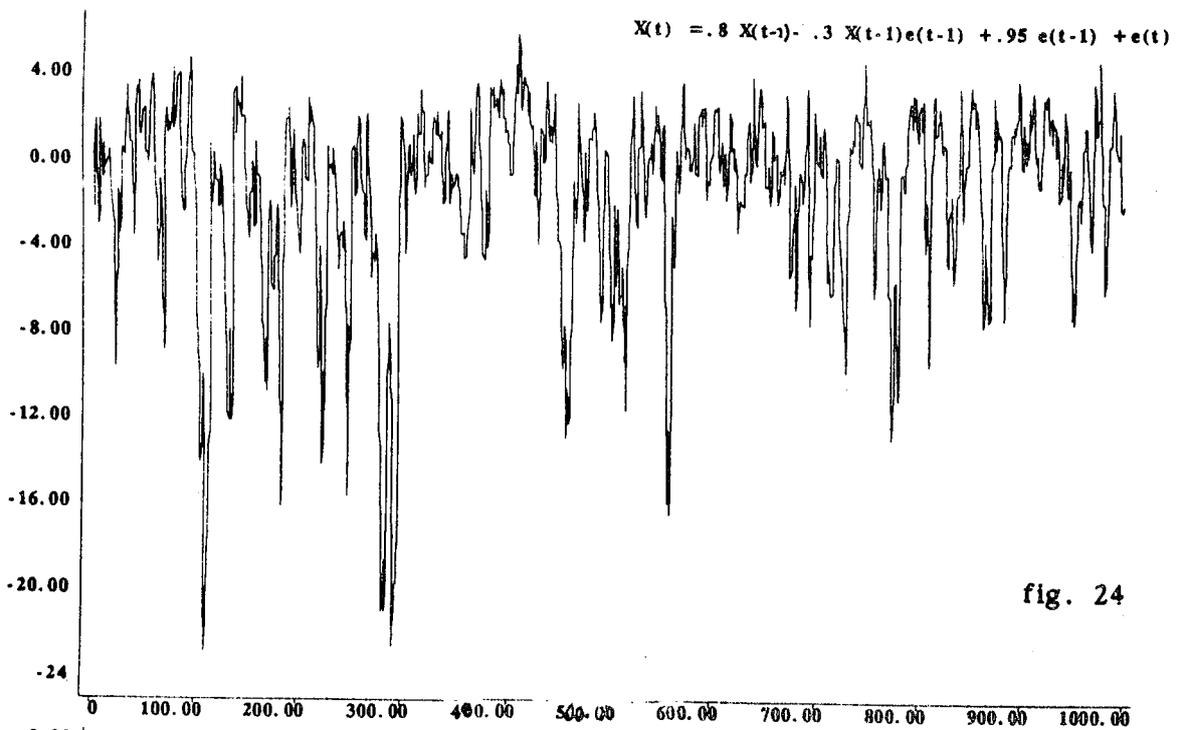
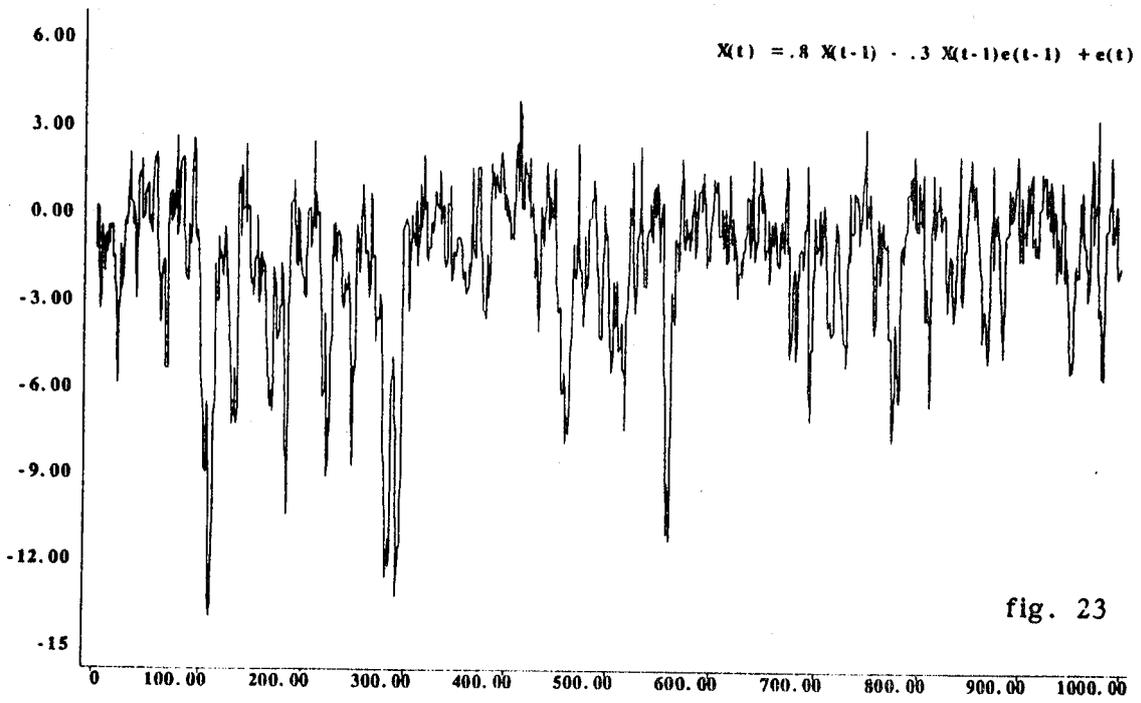
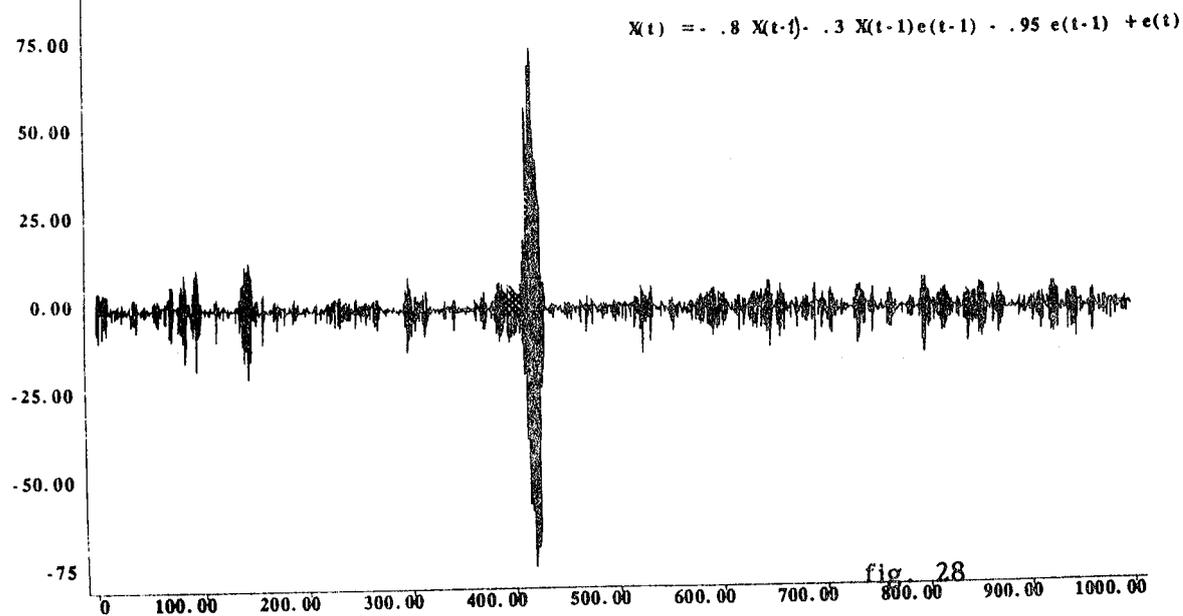
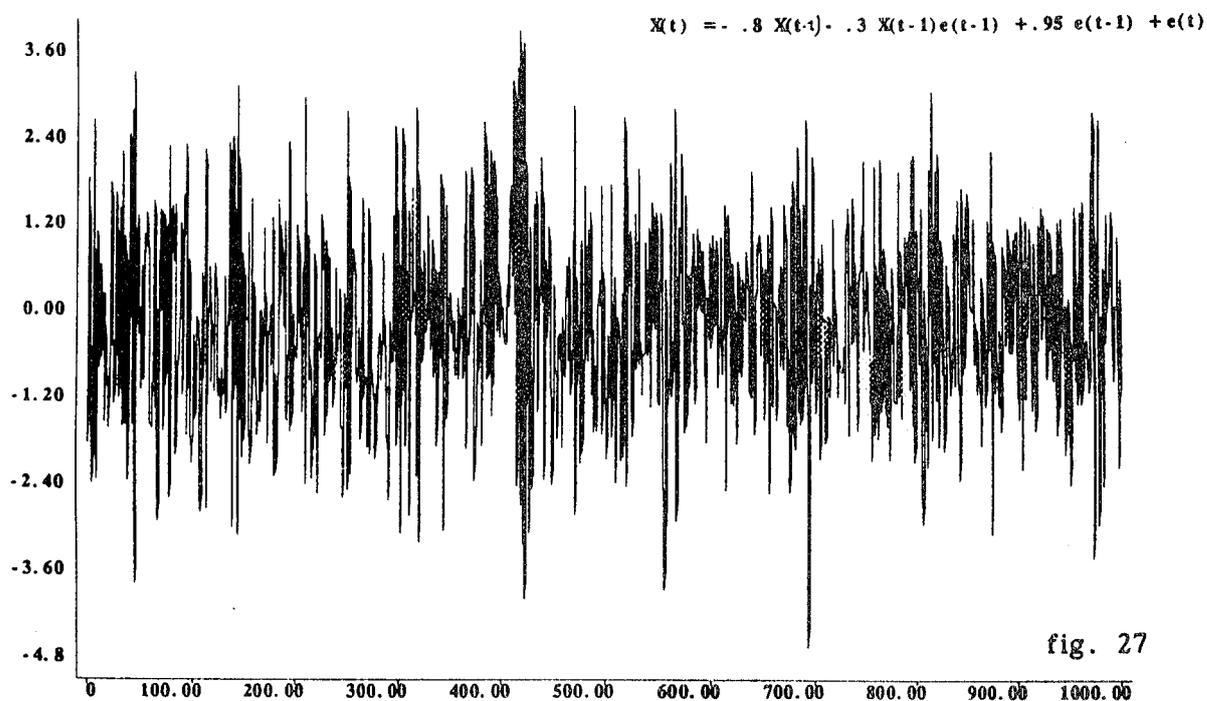
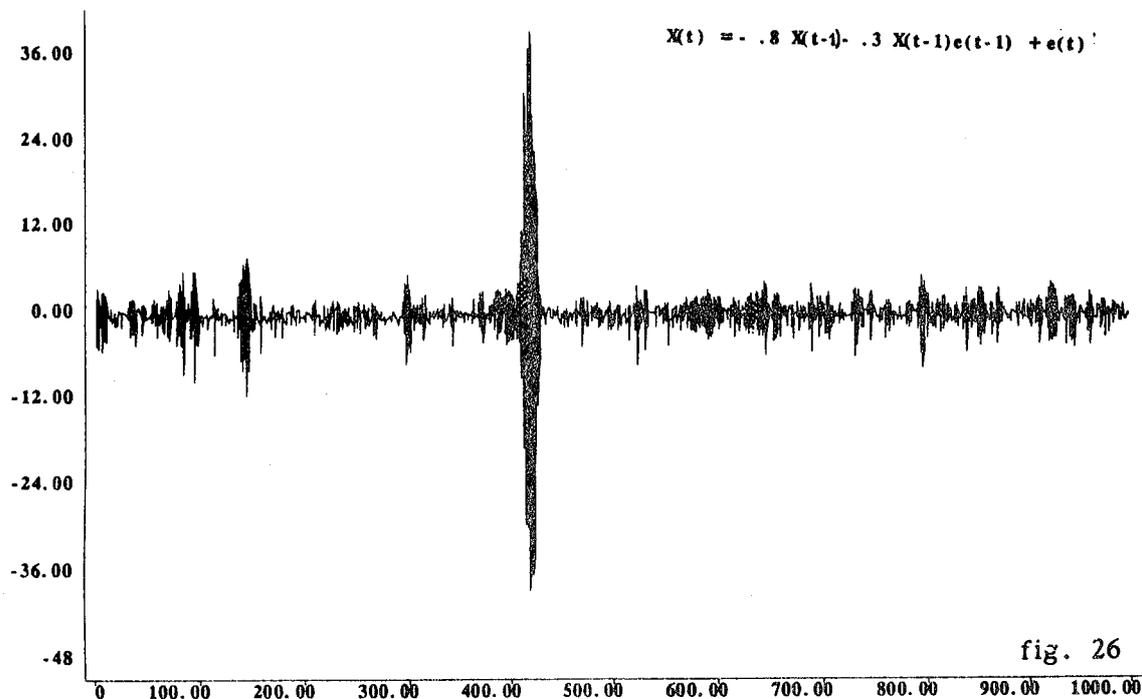
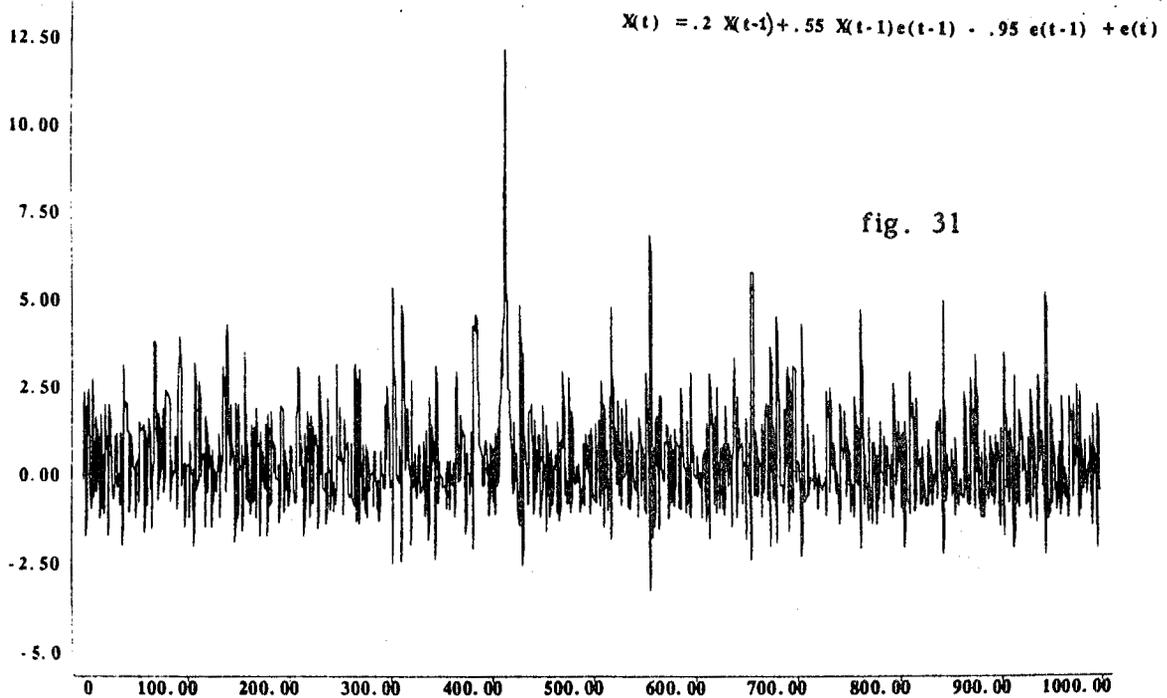
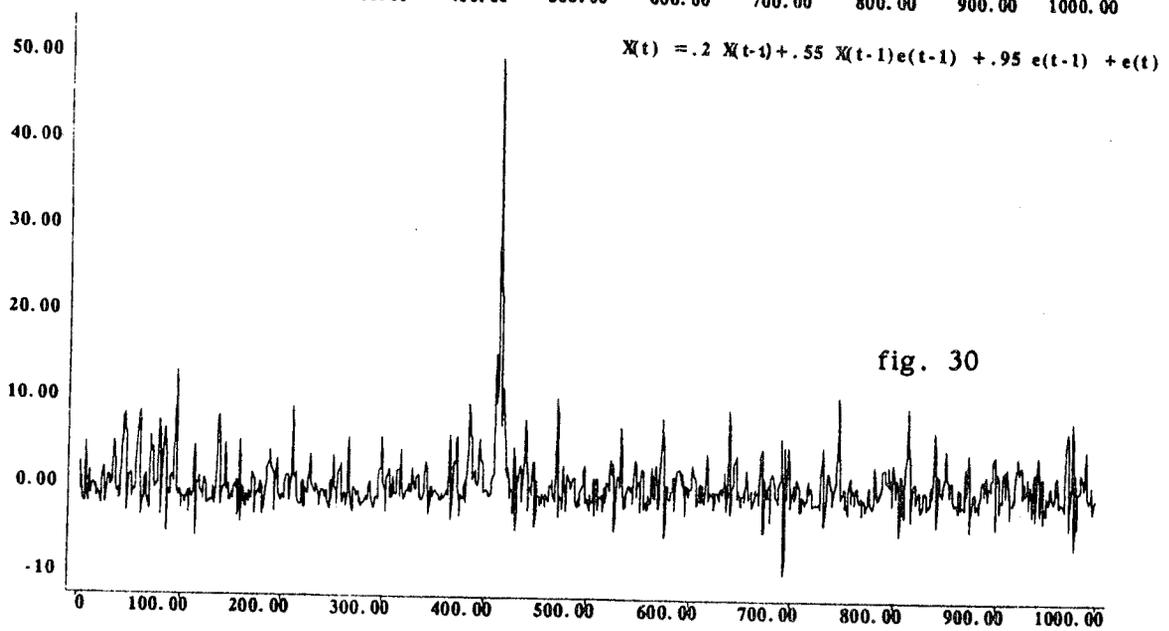
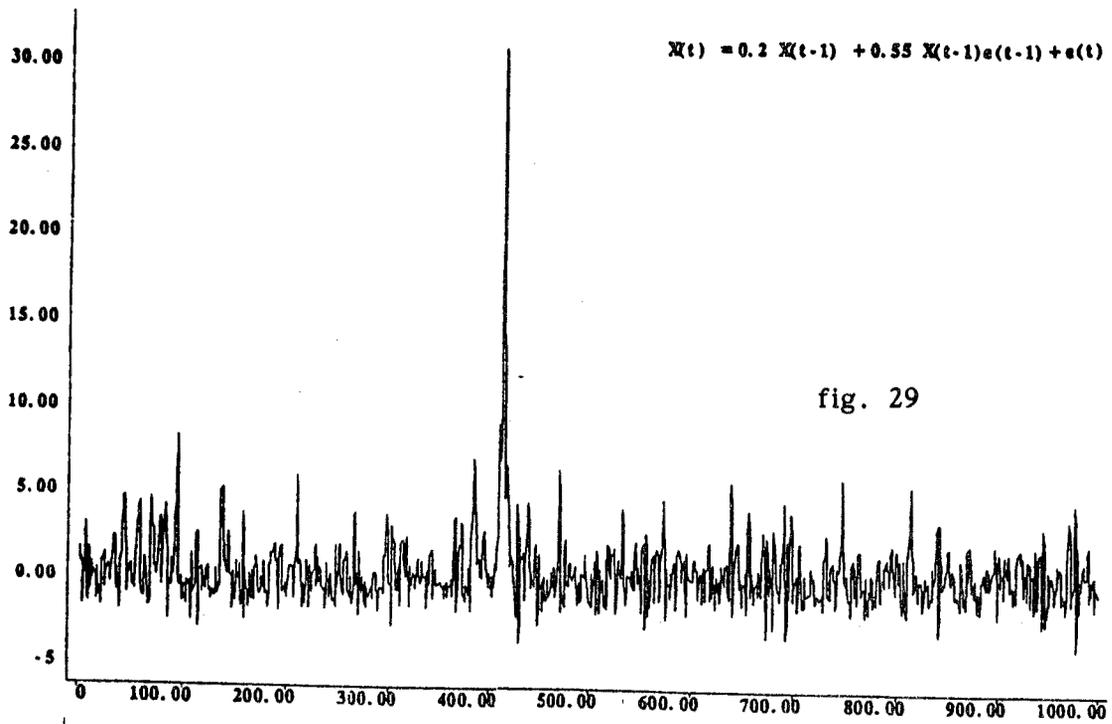
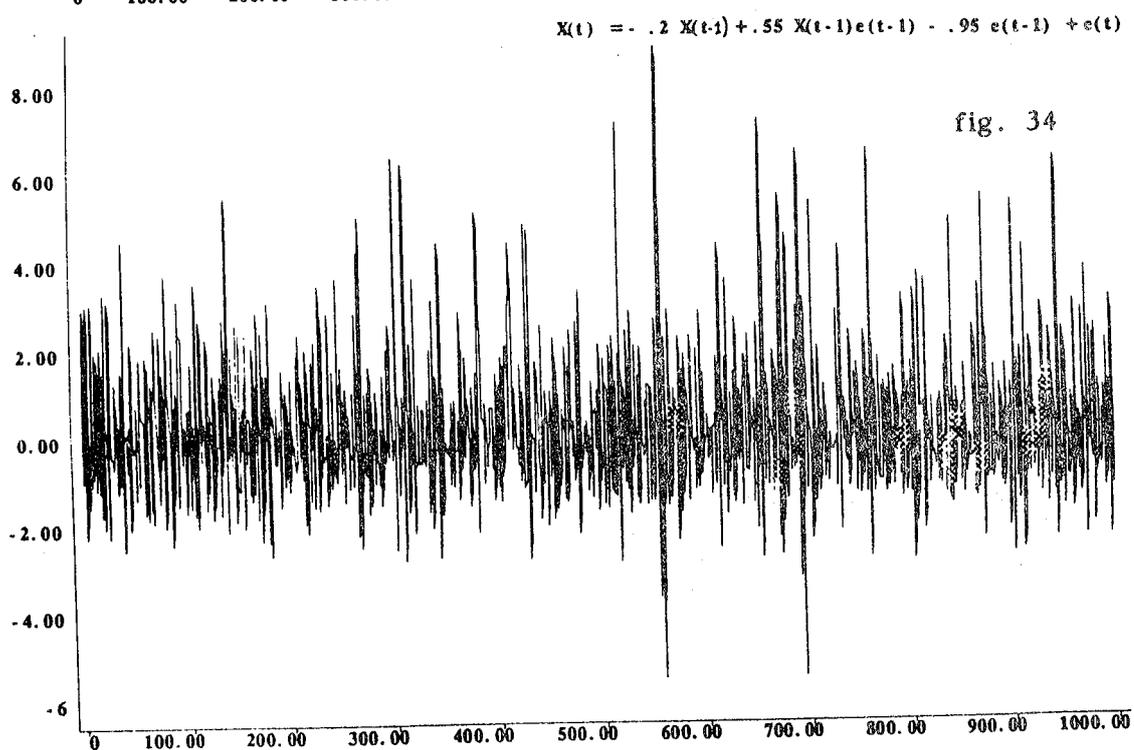
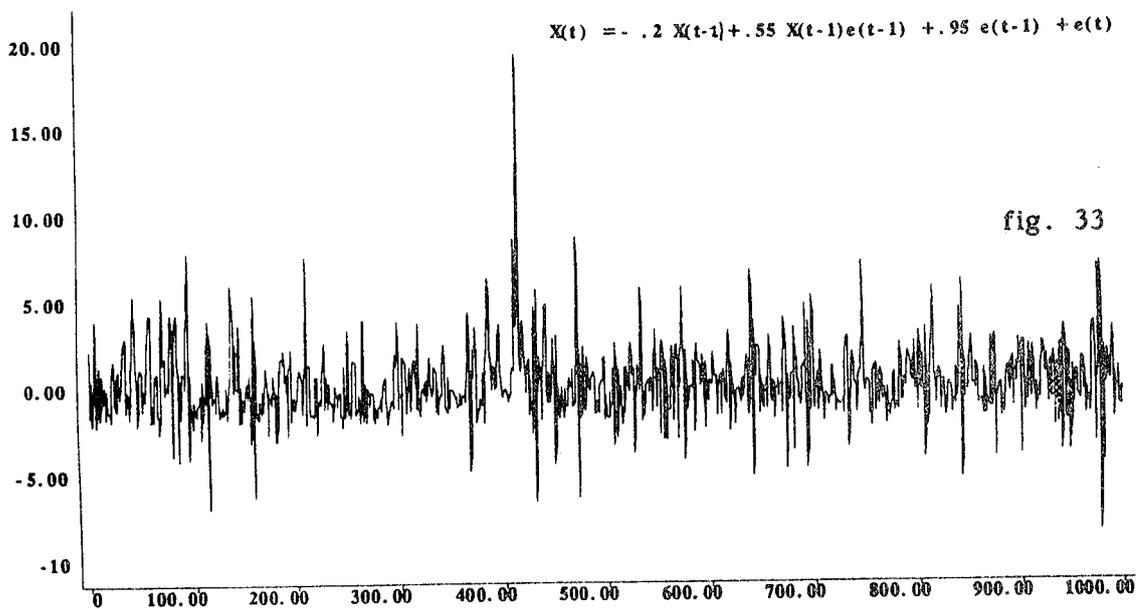
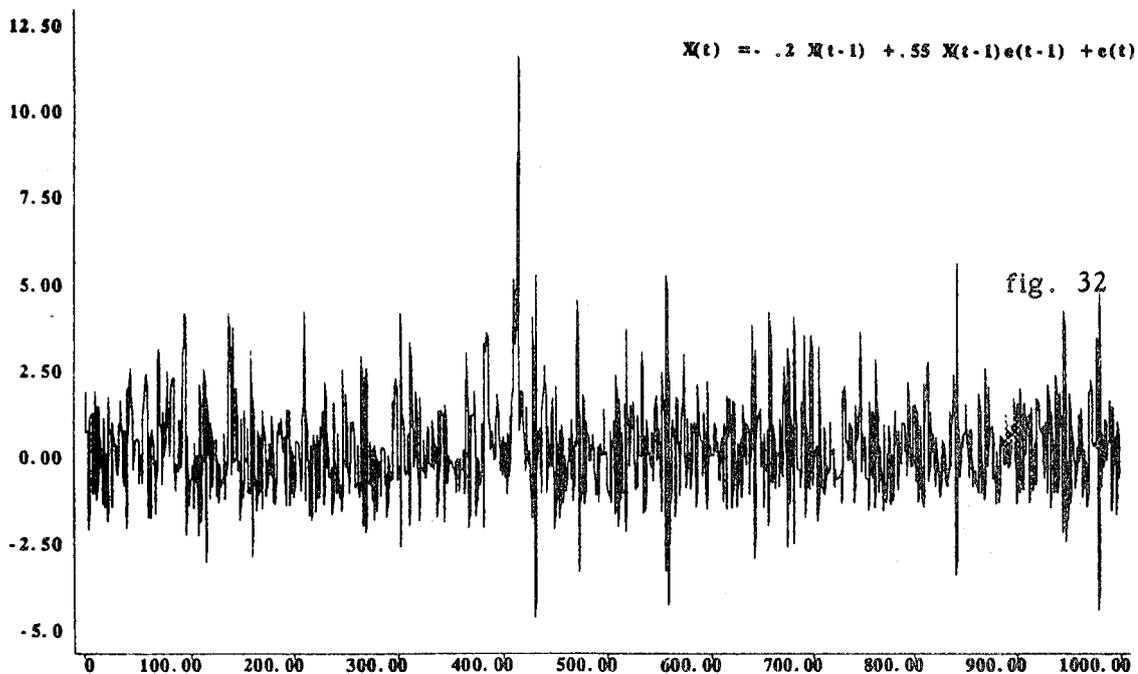


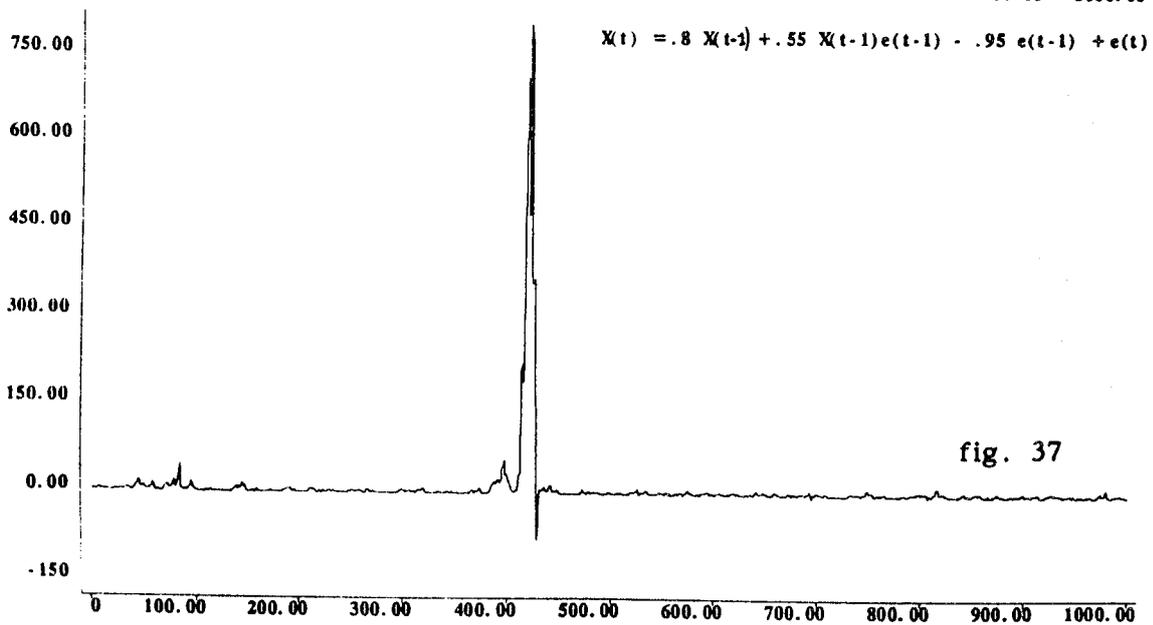
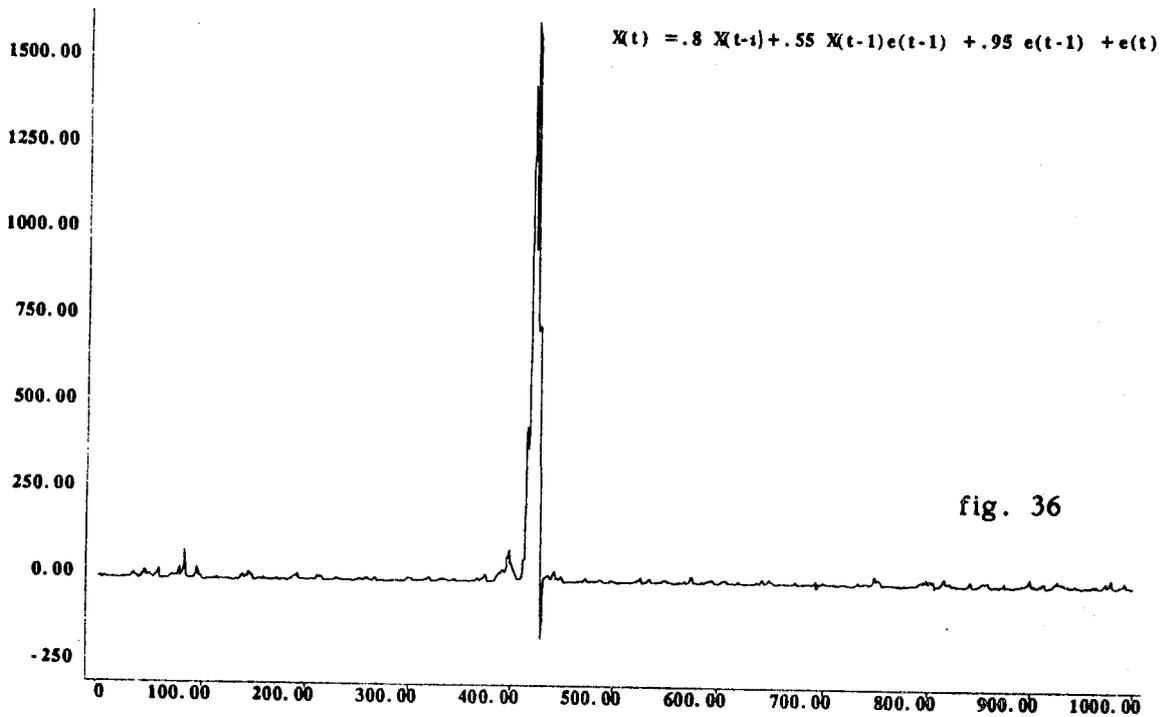
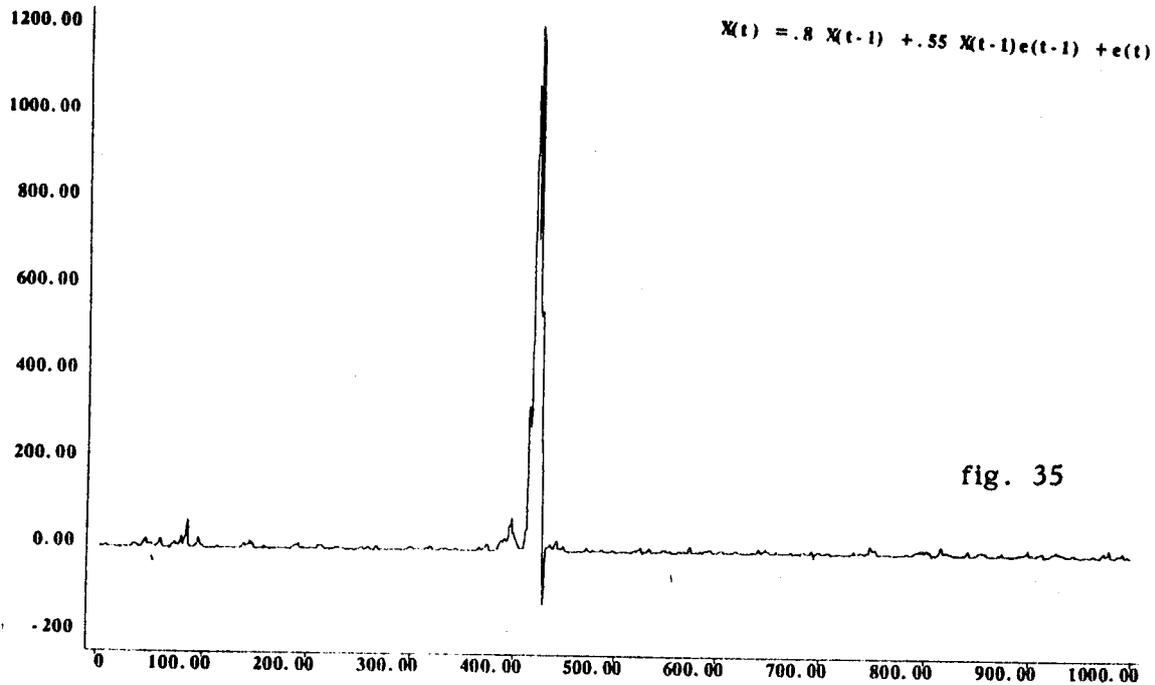
fig. 22

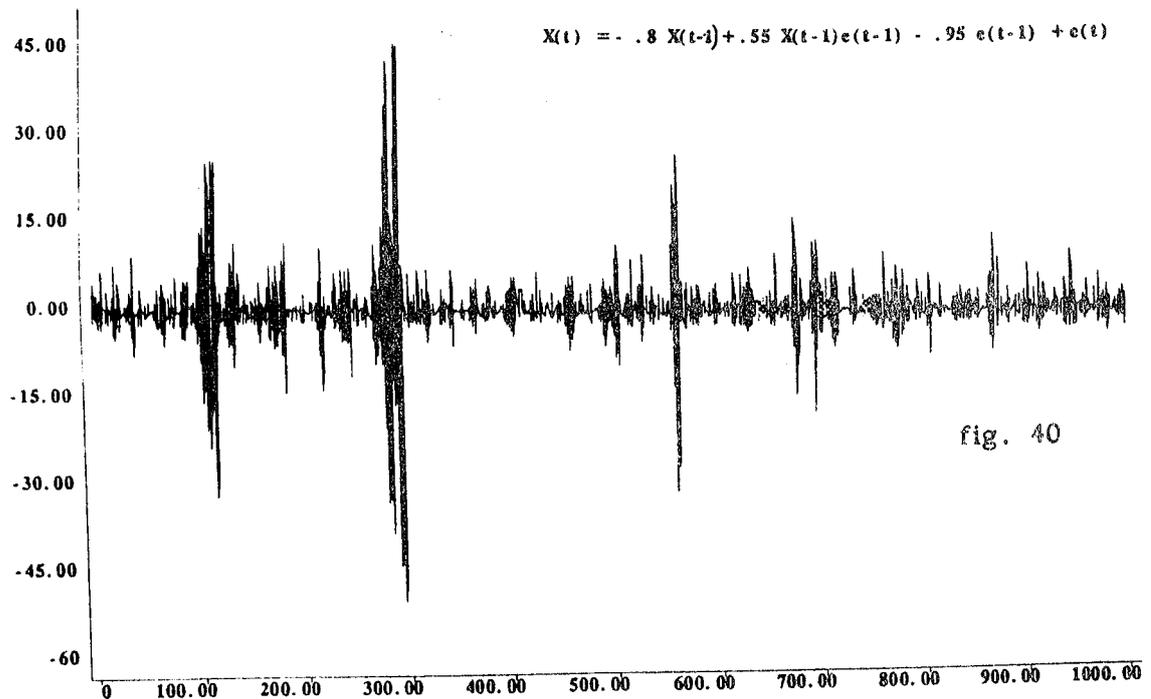
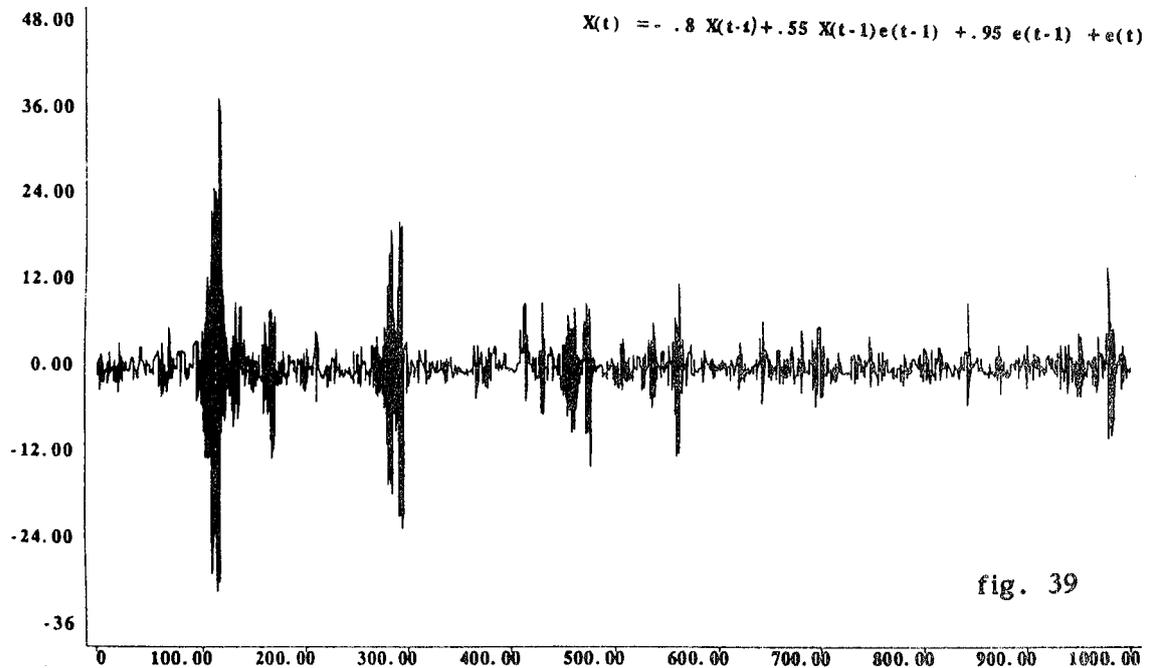
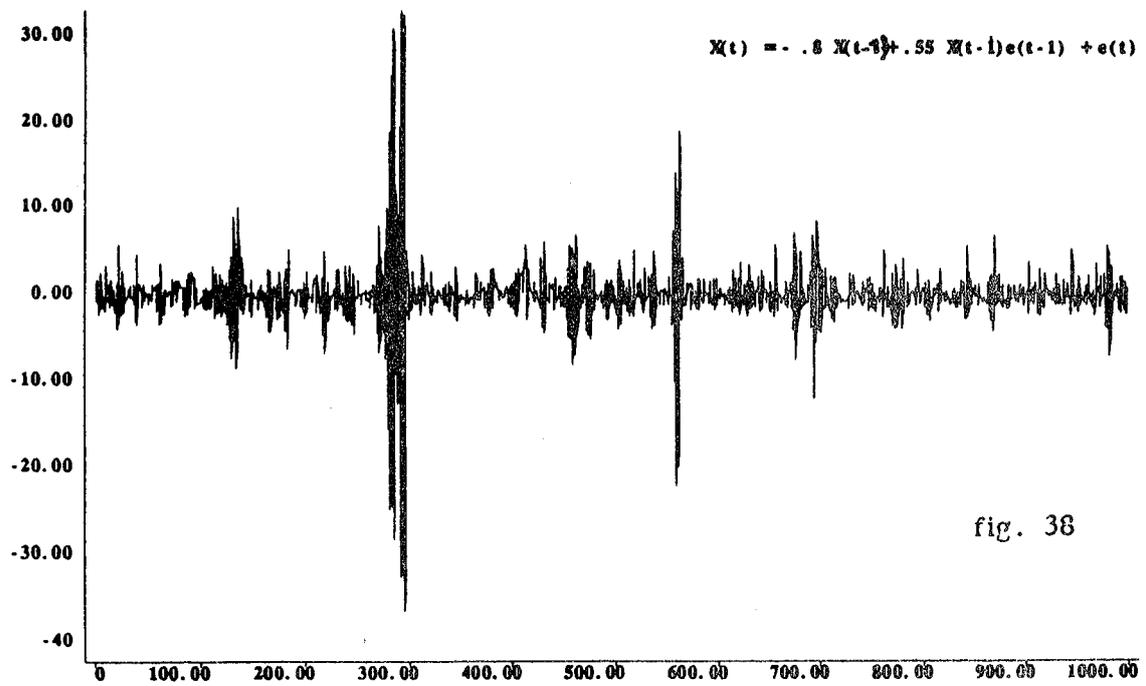


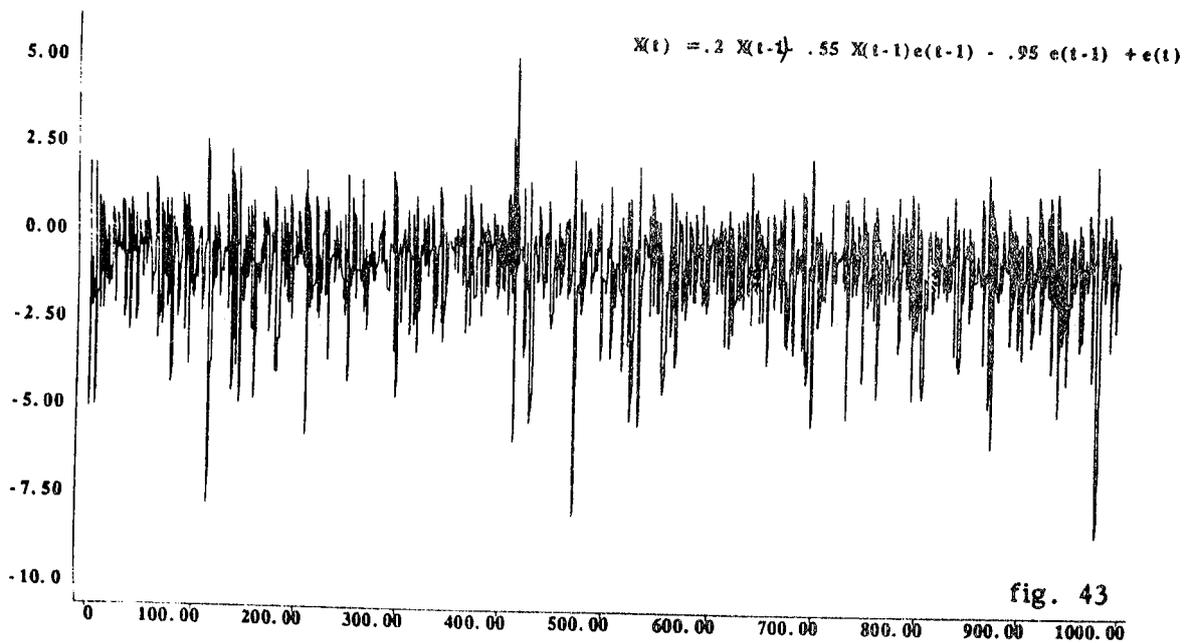
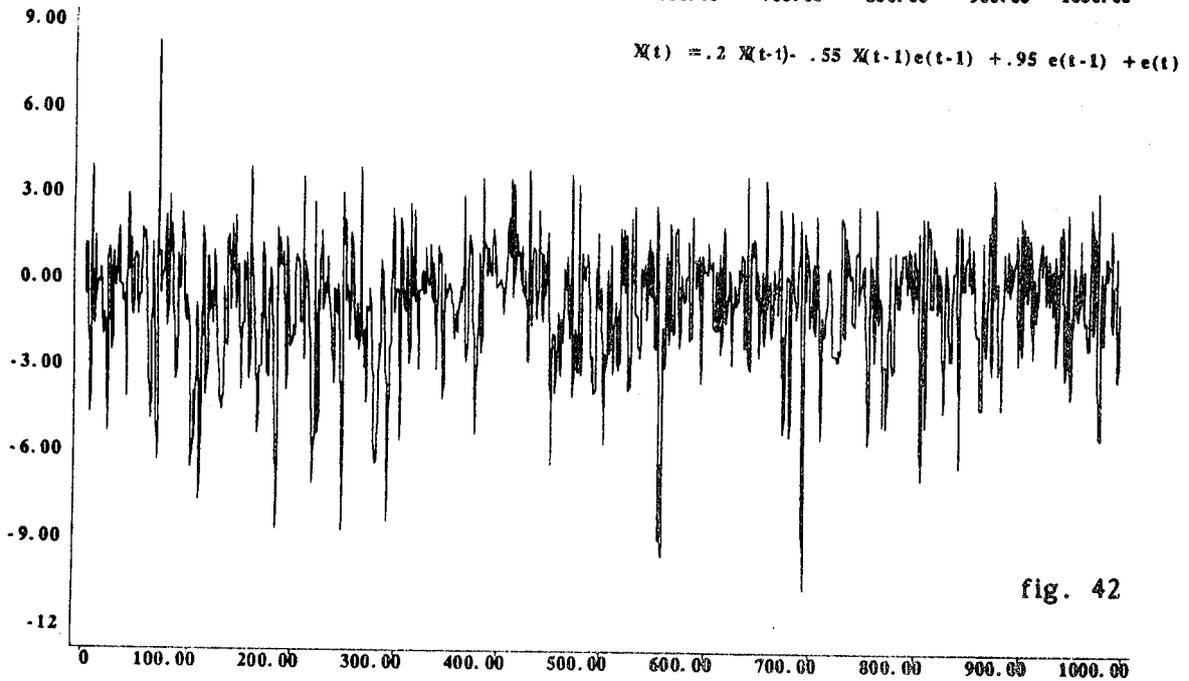
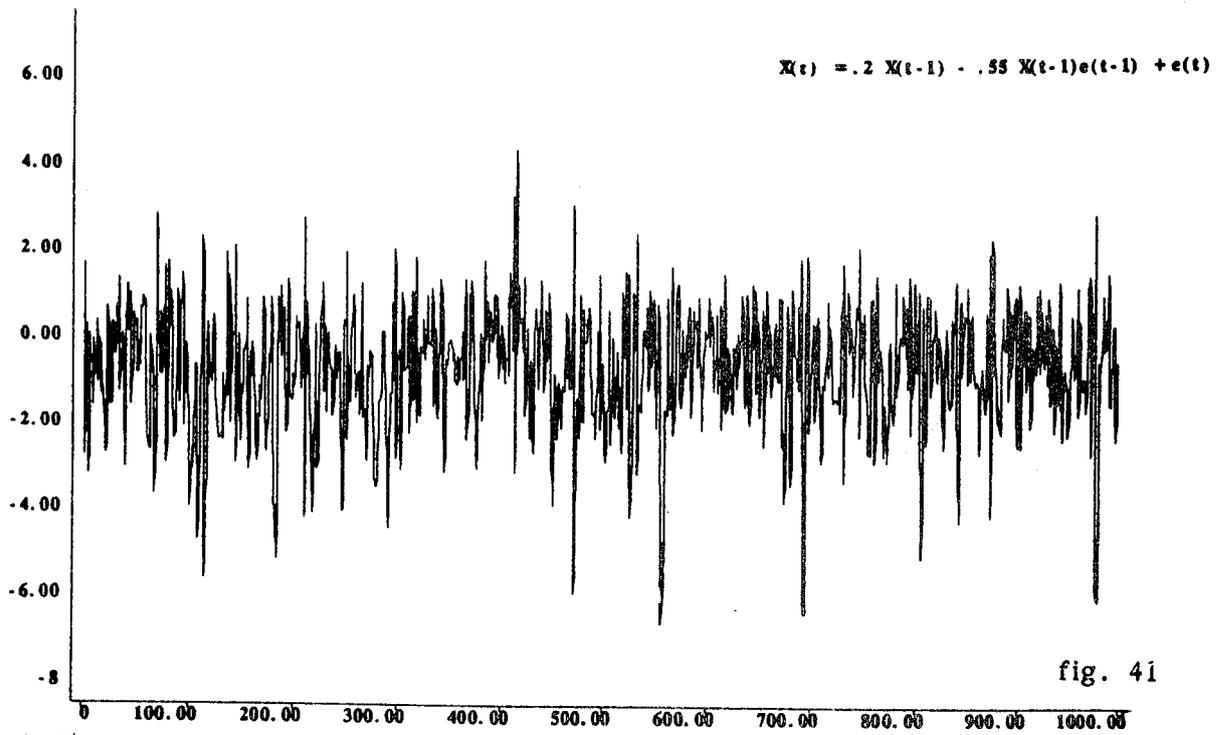


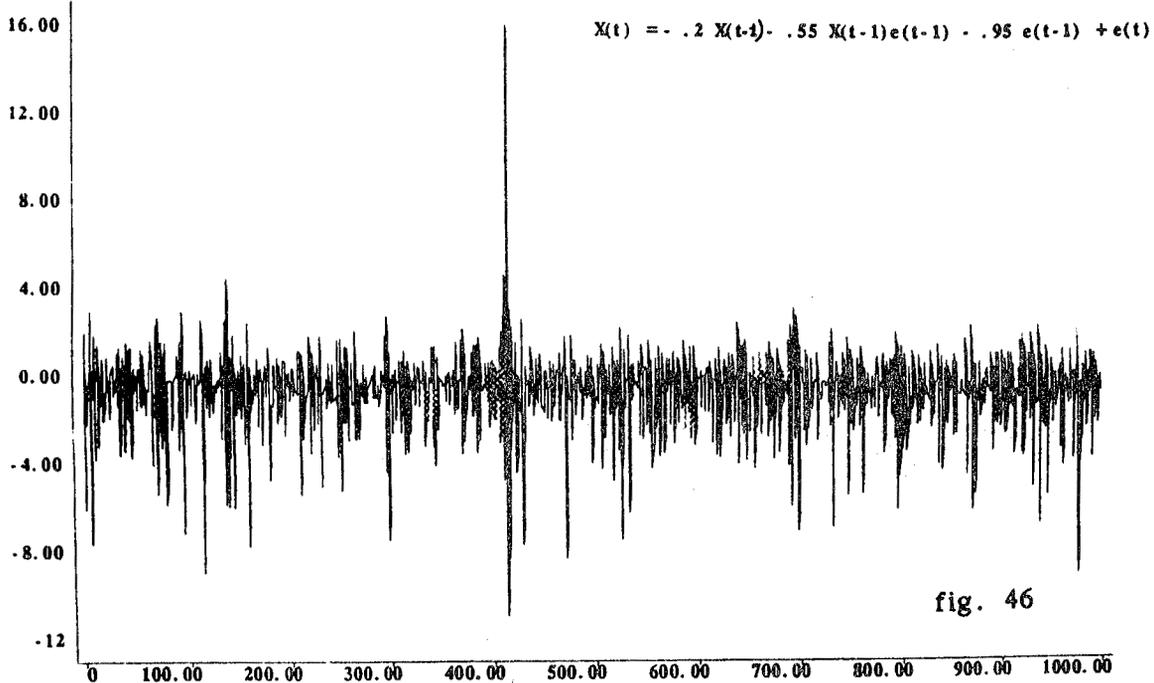
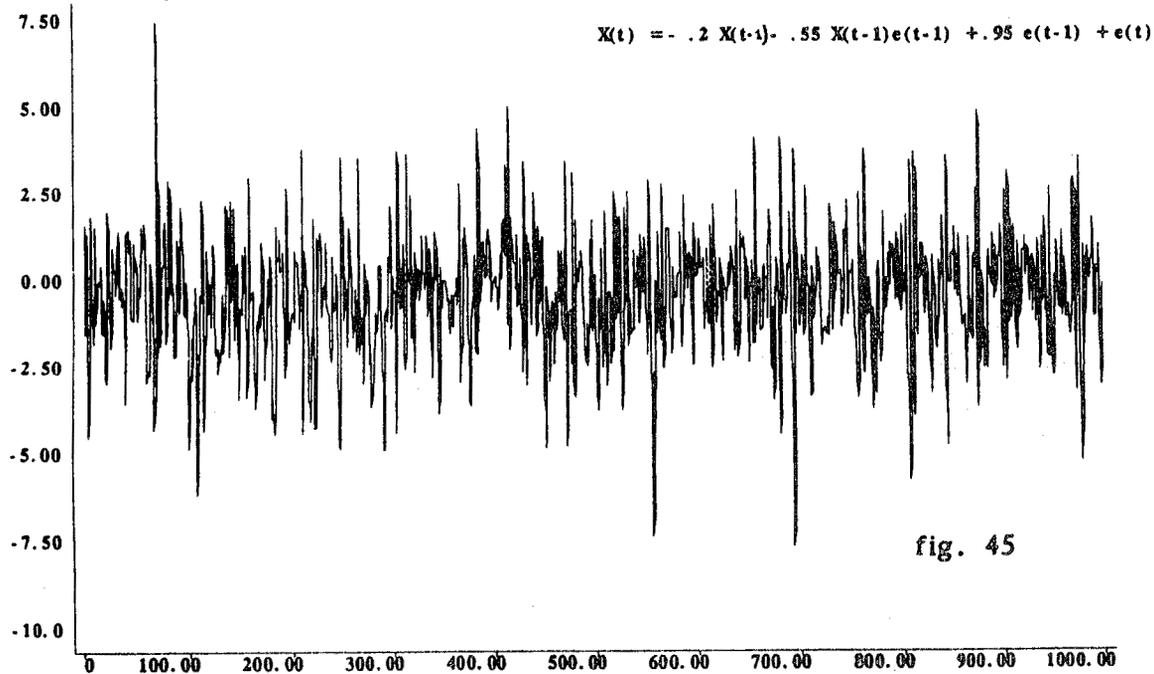
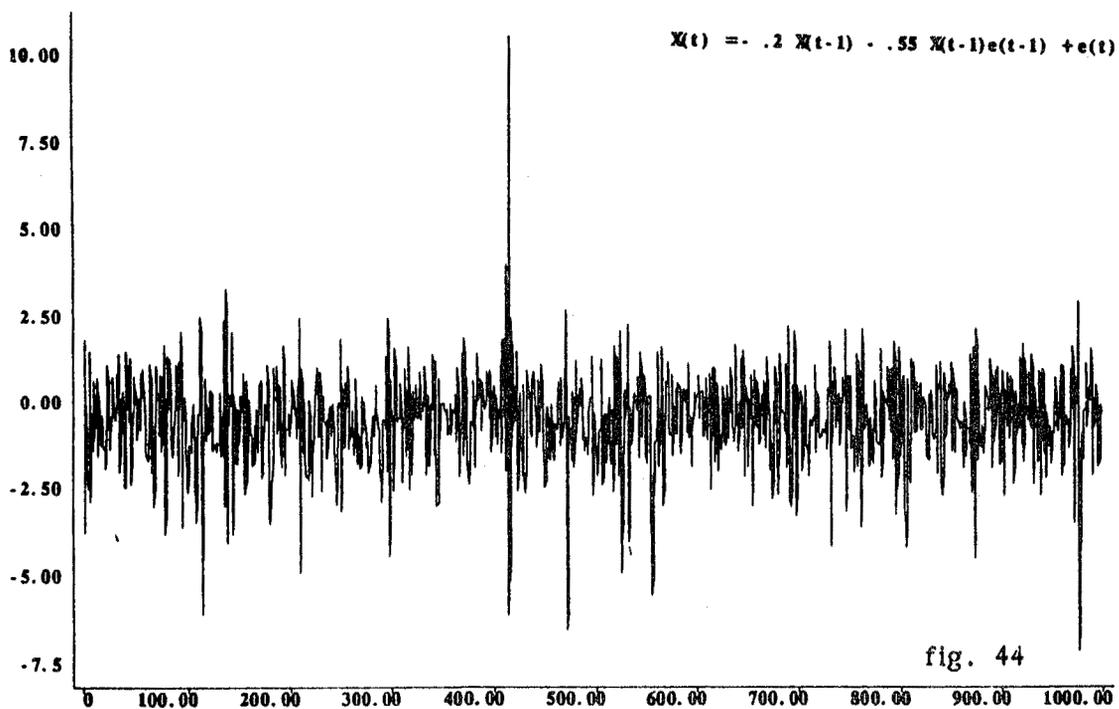


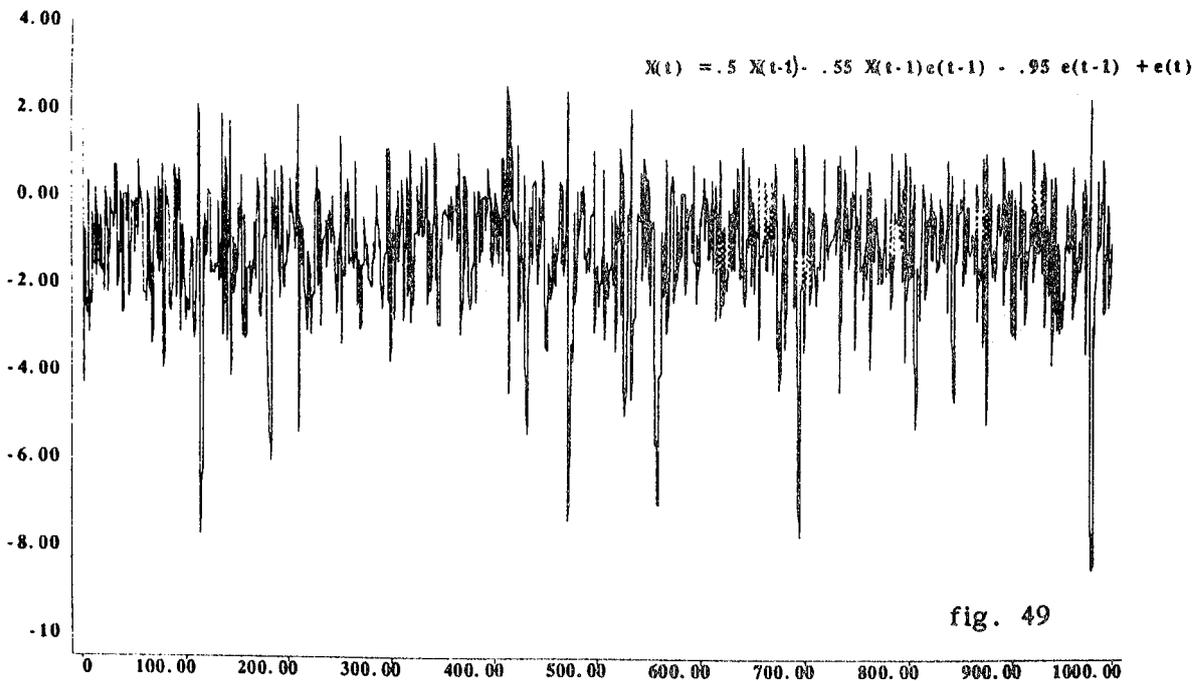
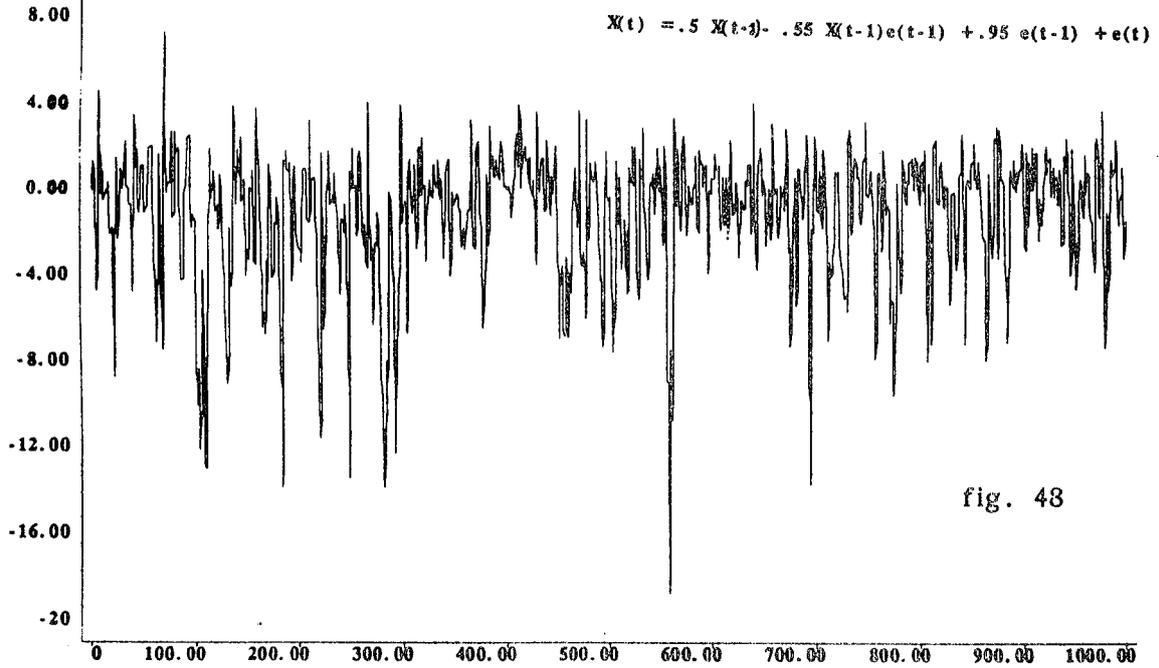
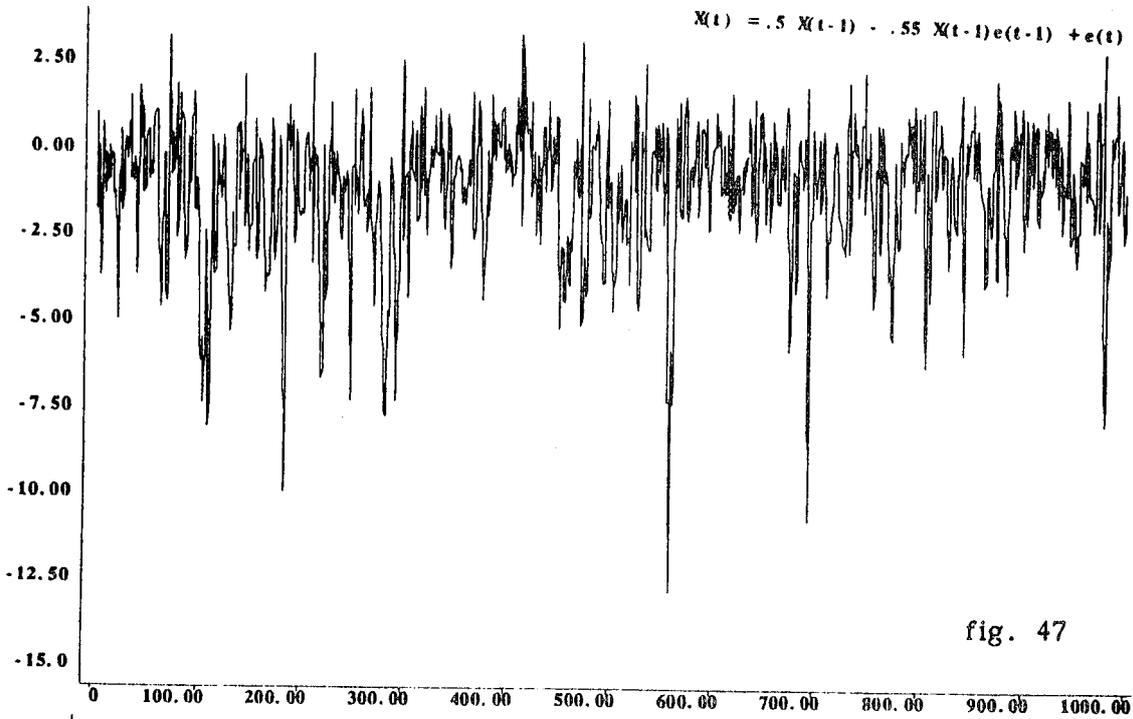


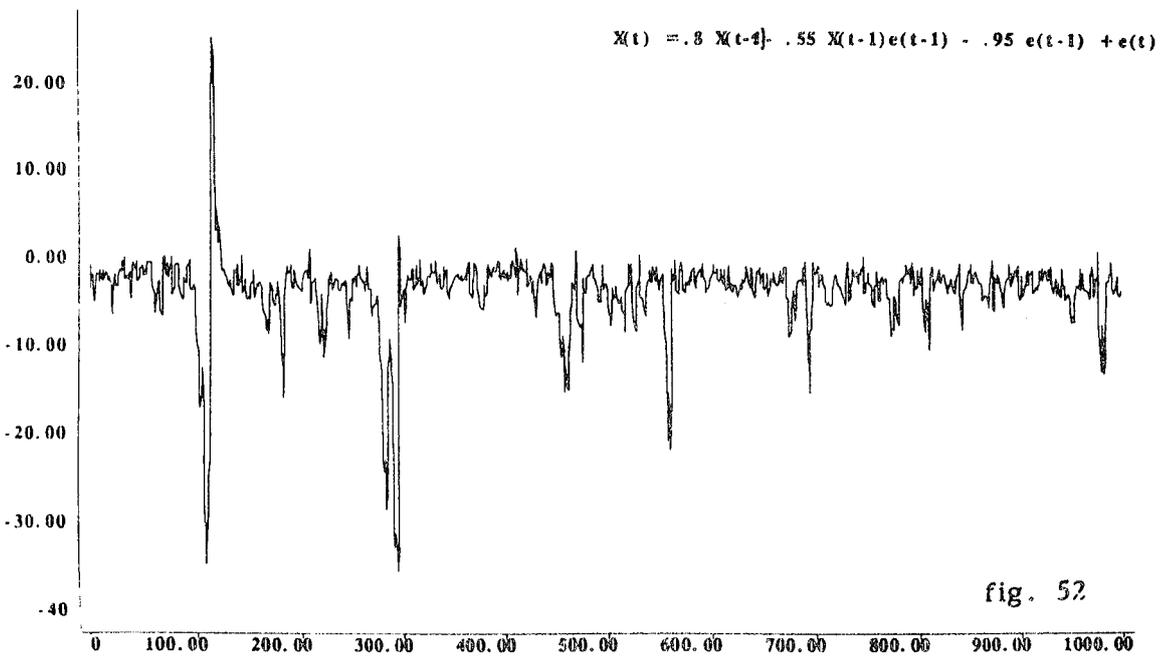
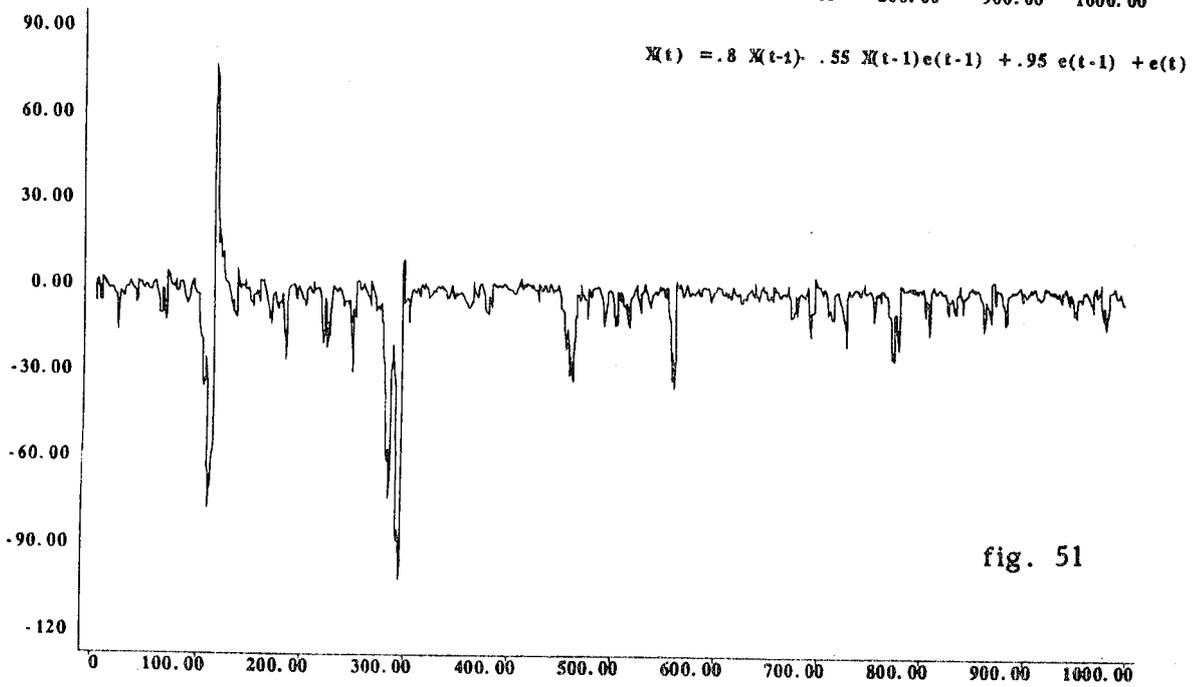
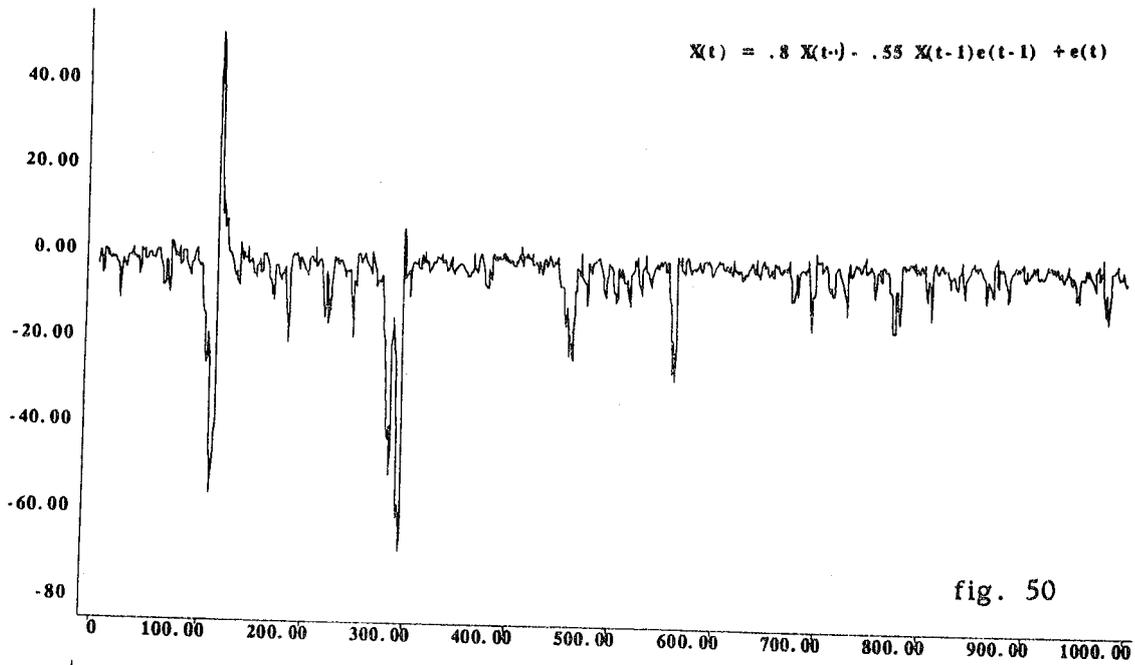


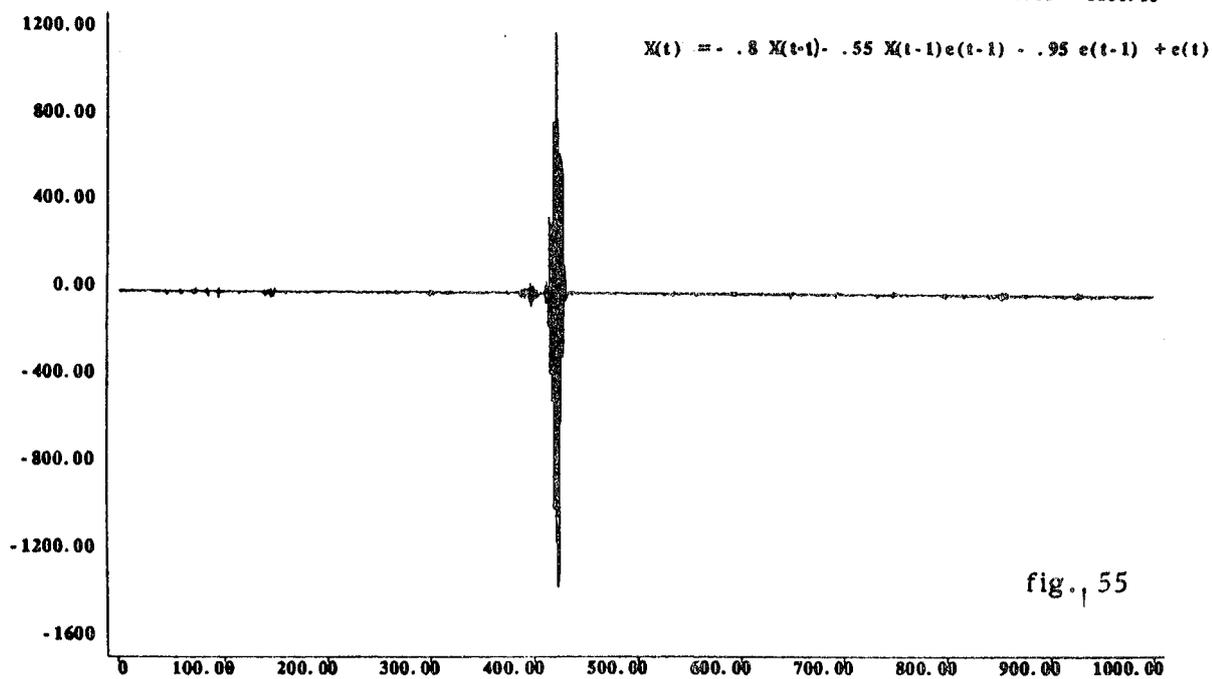
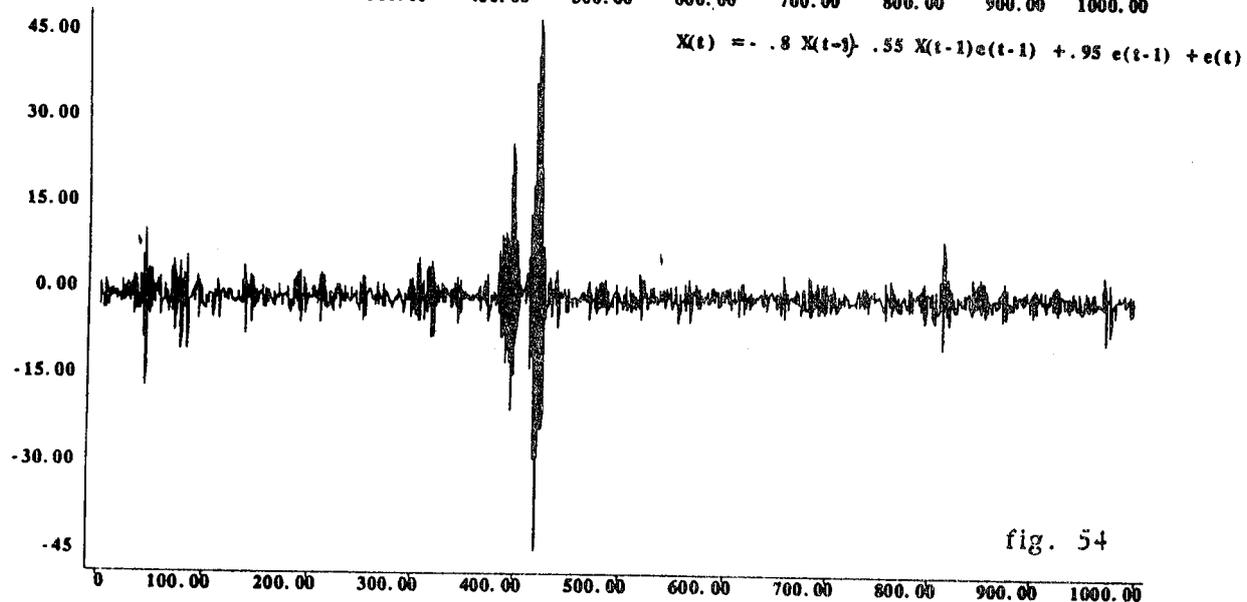
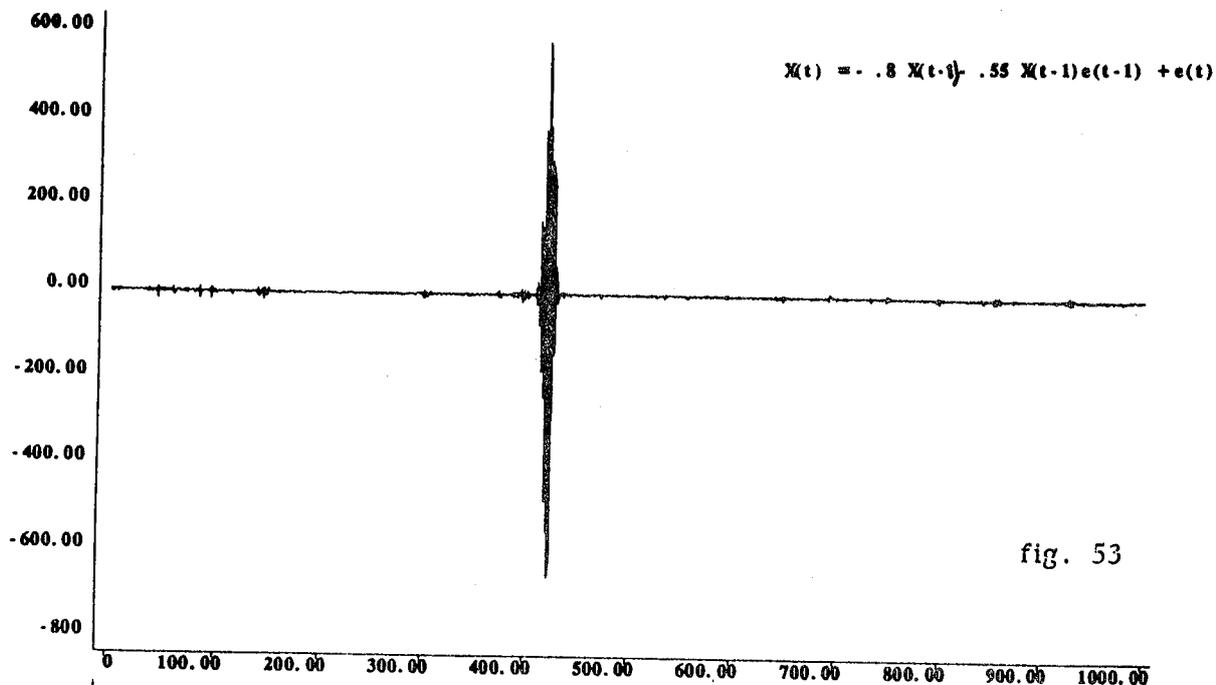


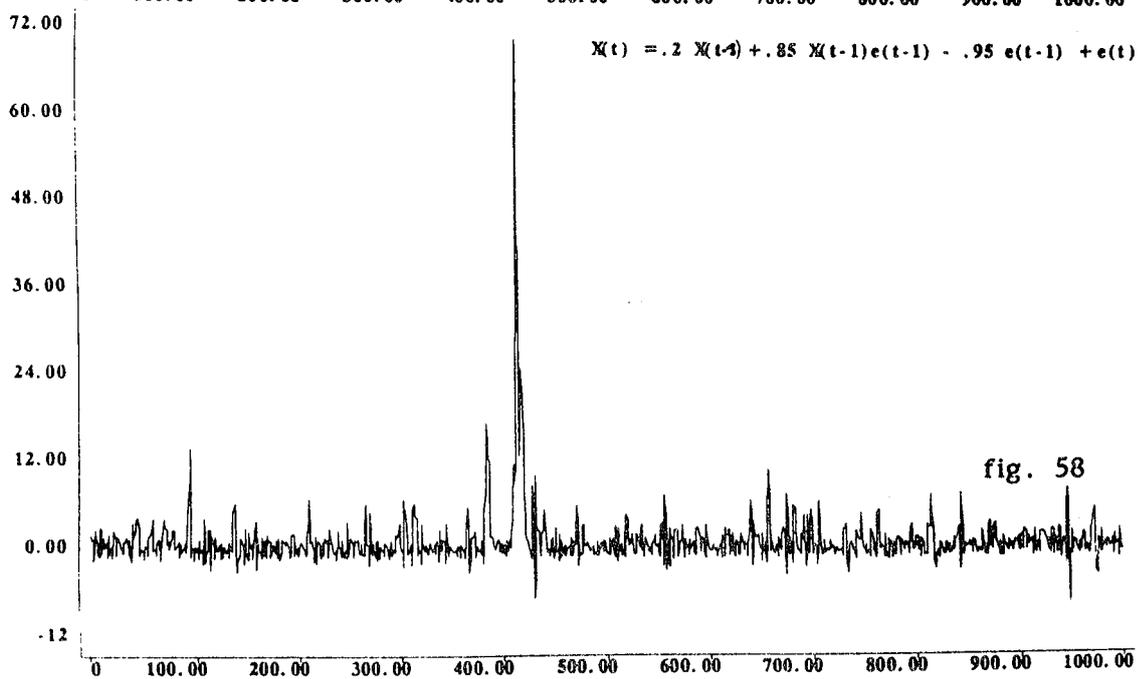
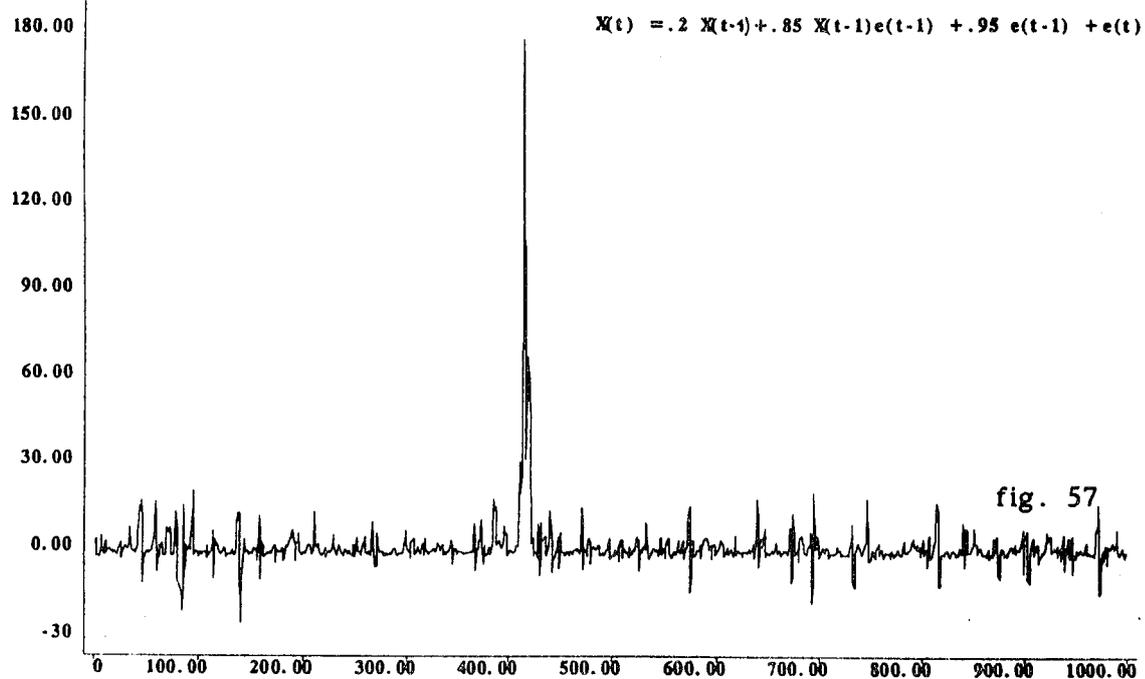
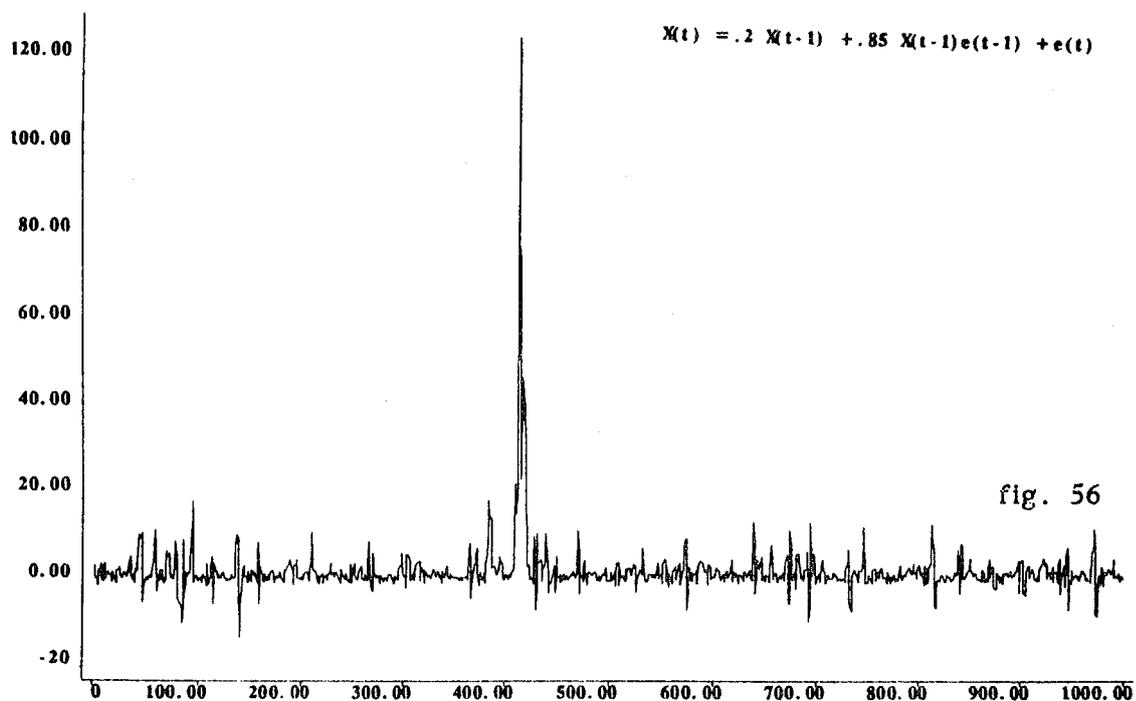


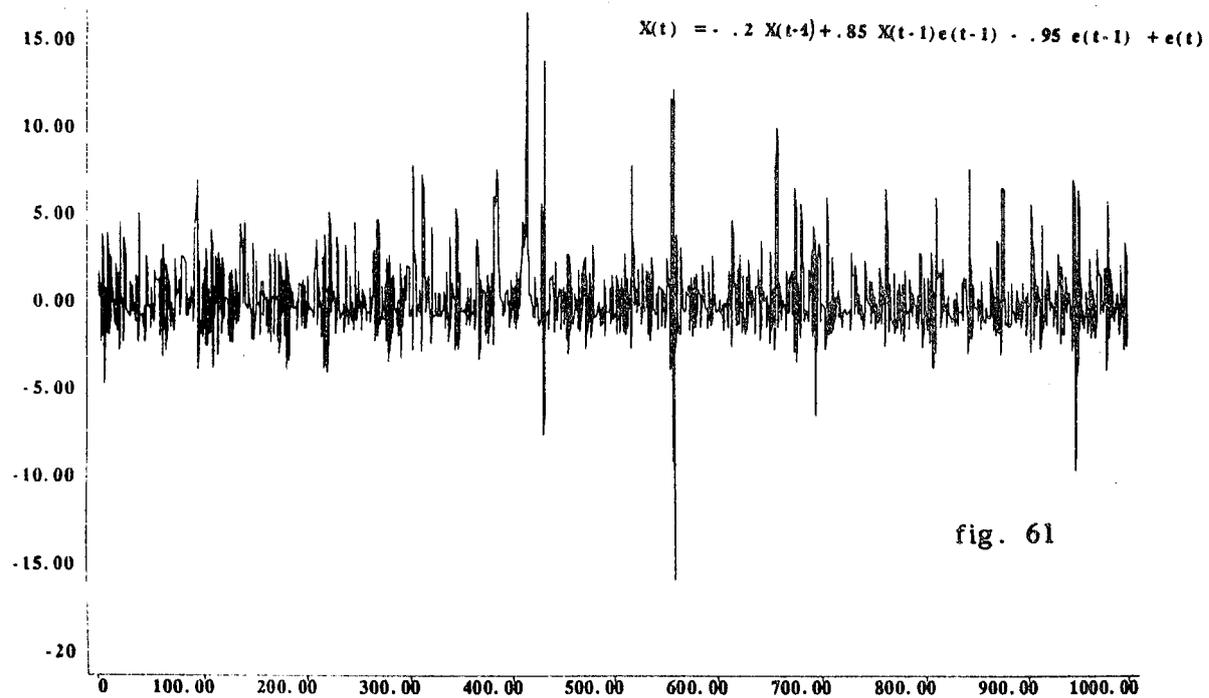
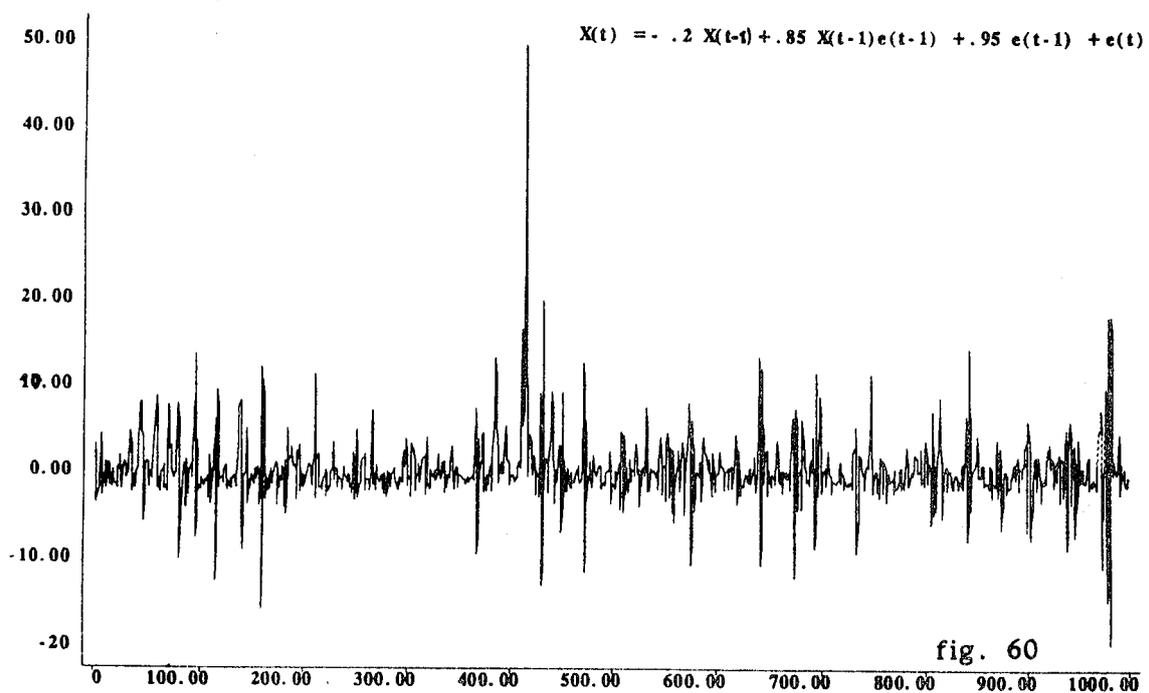
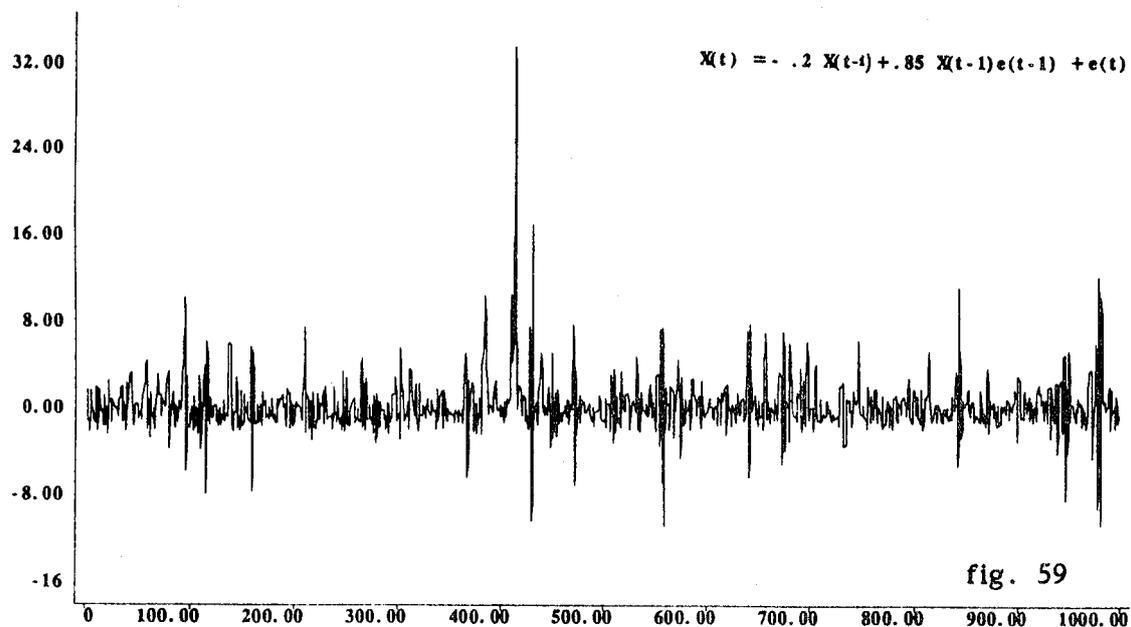


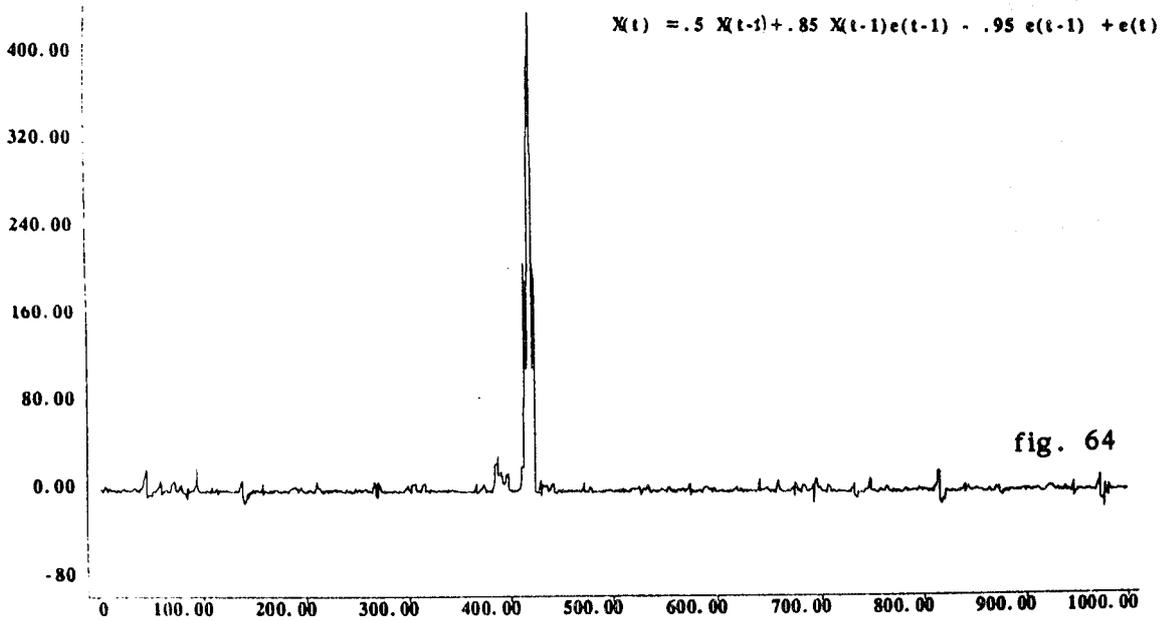
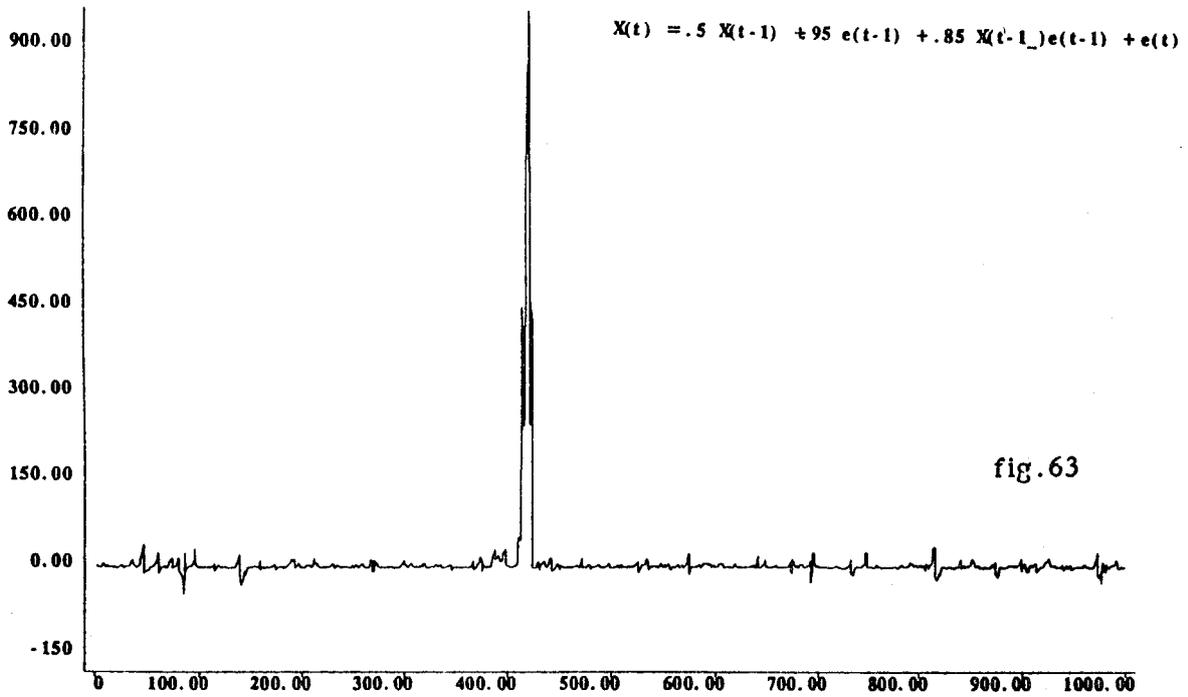
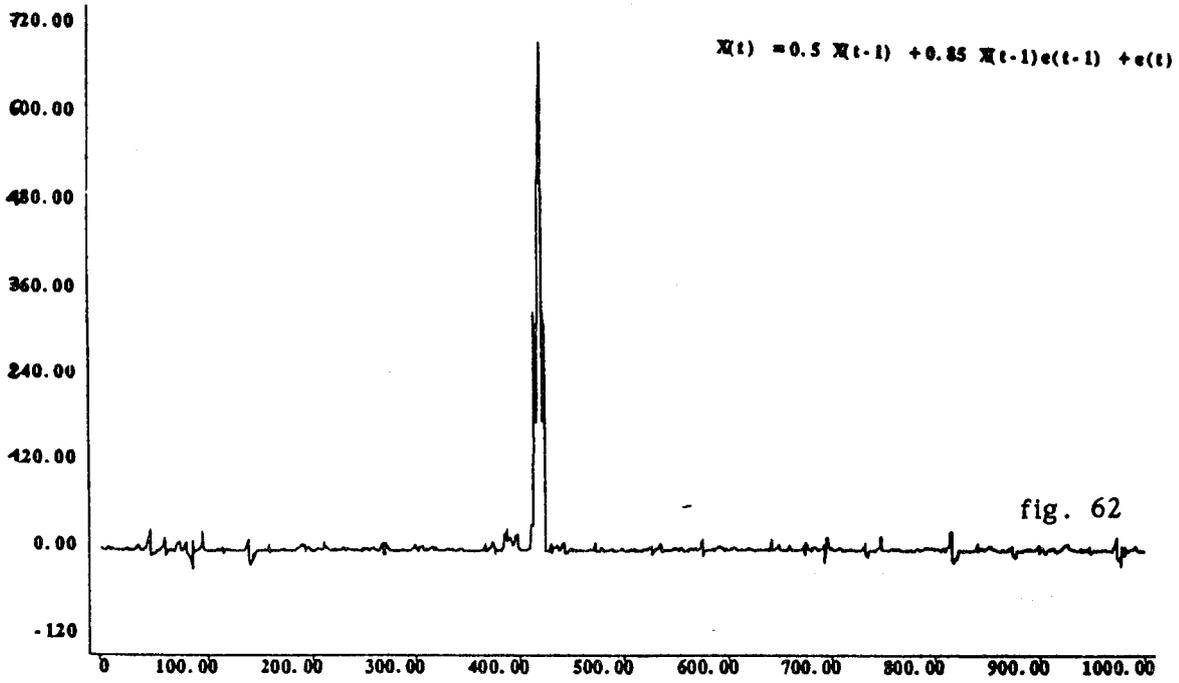


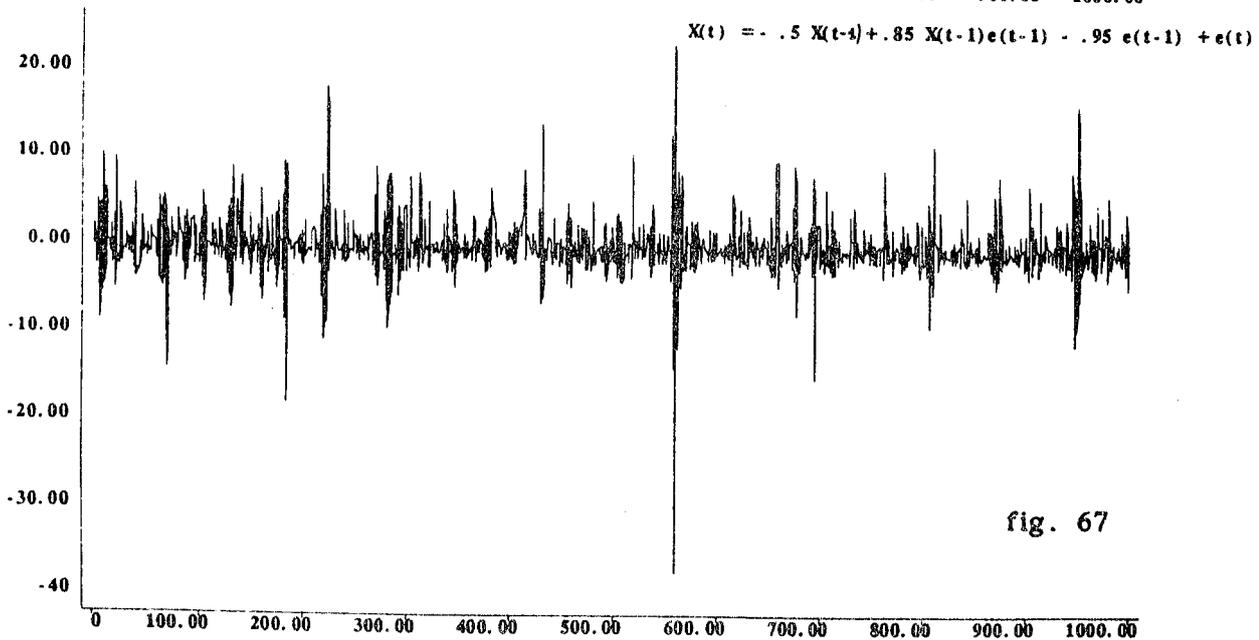
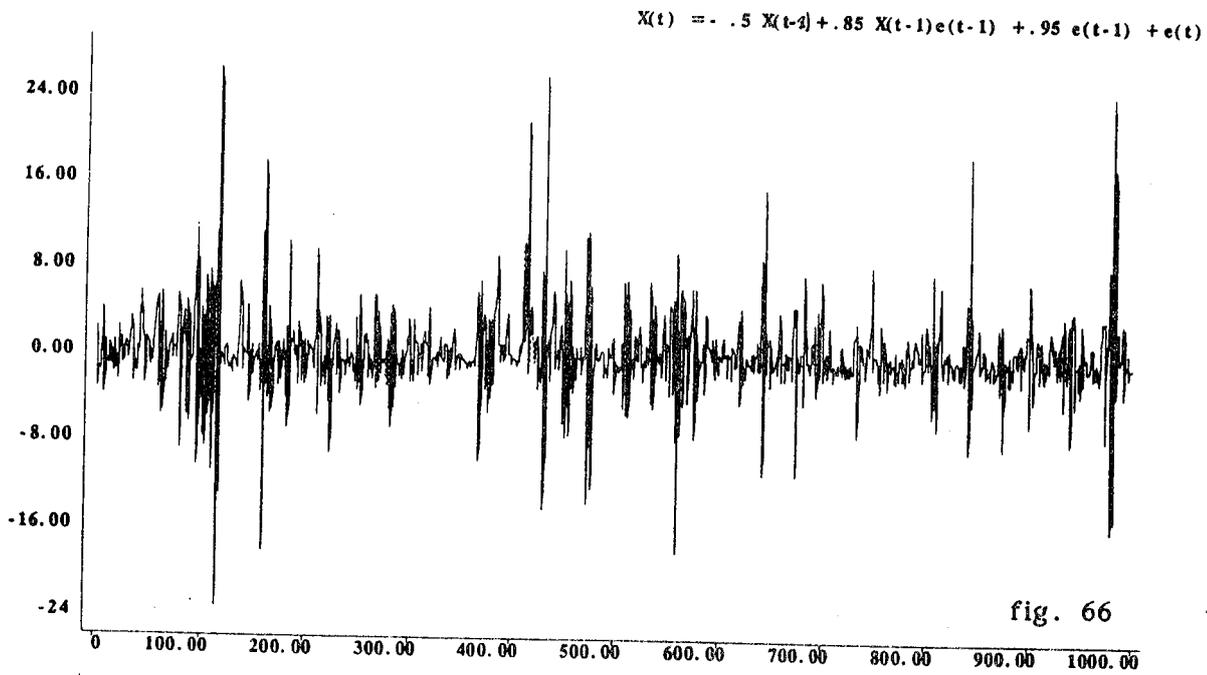
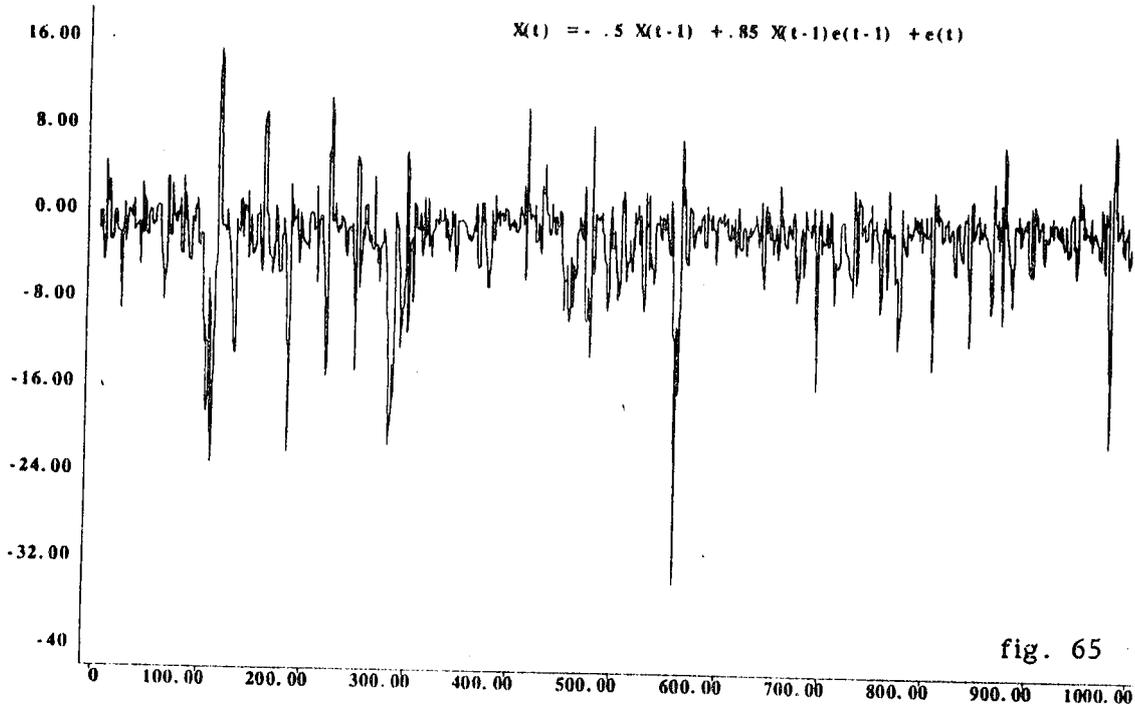


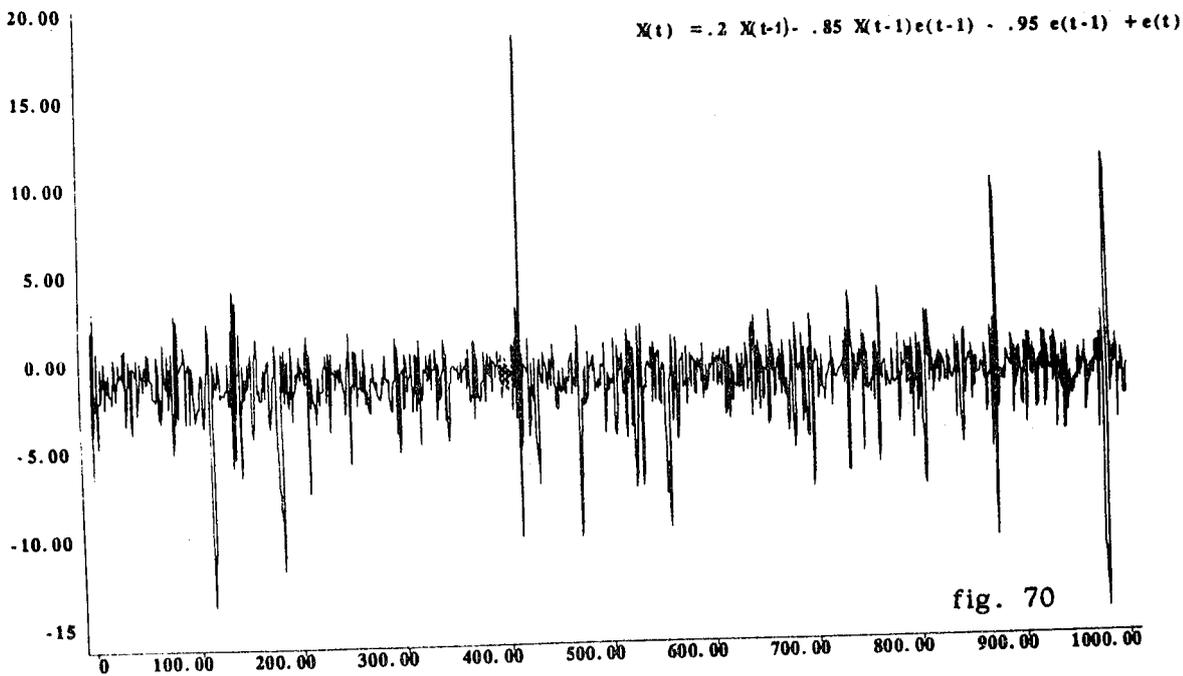
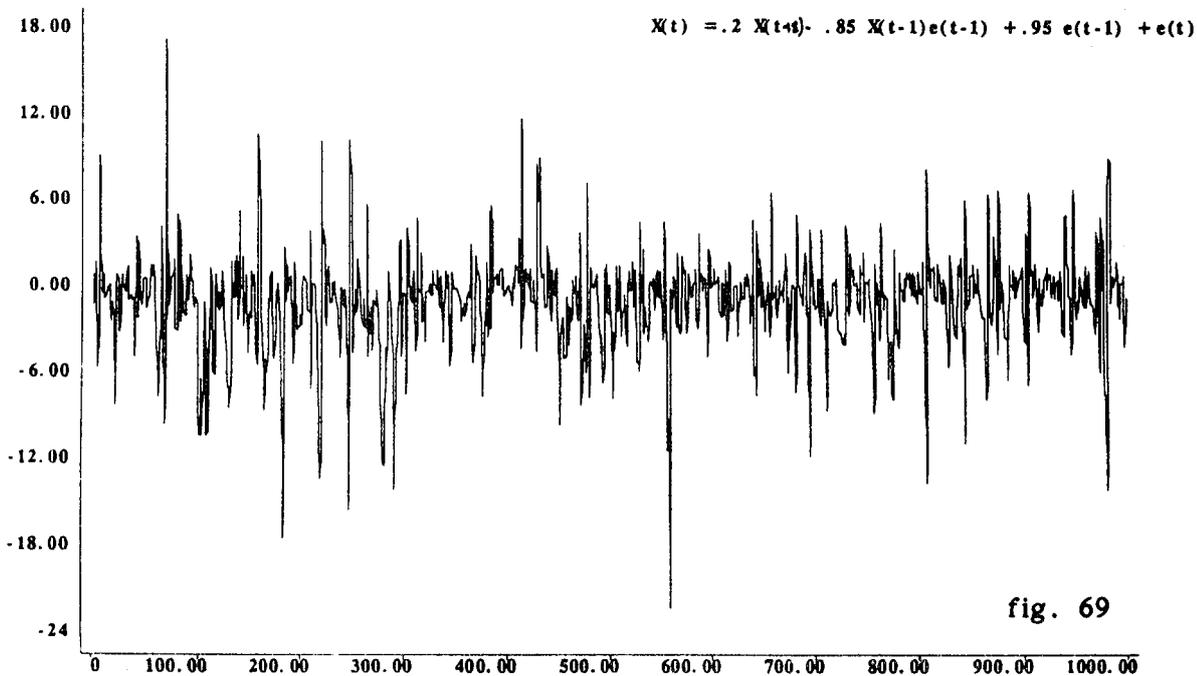
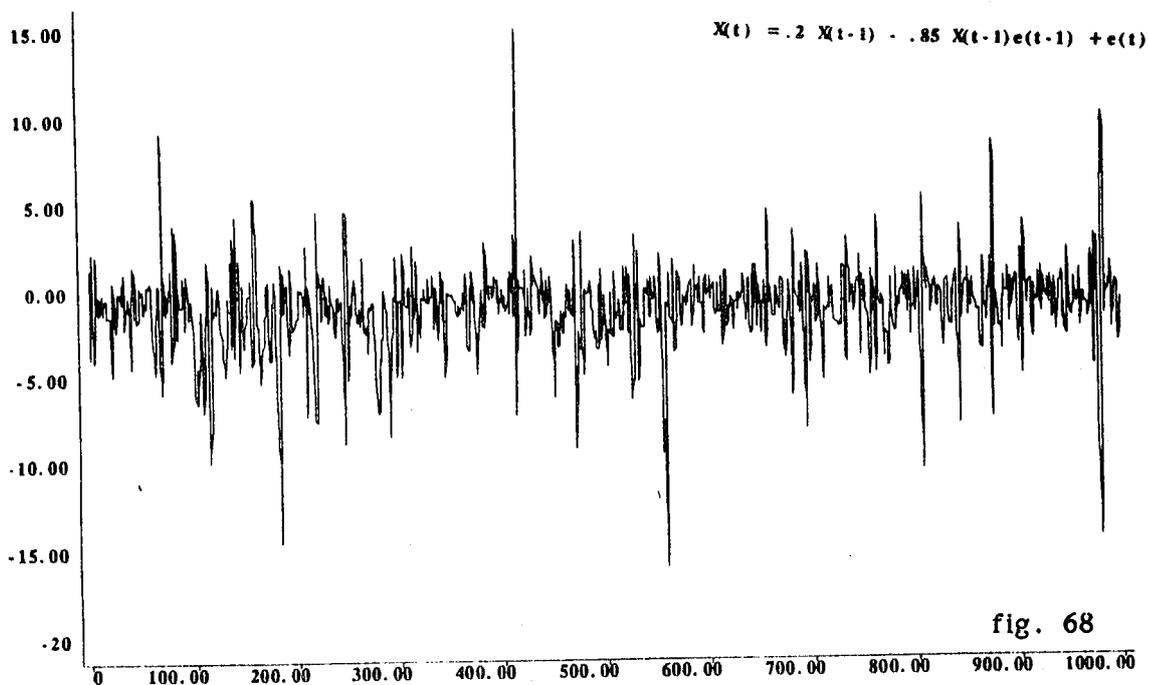


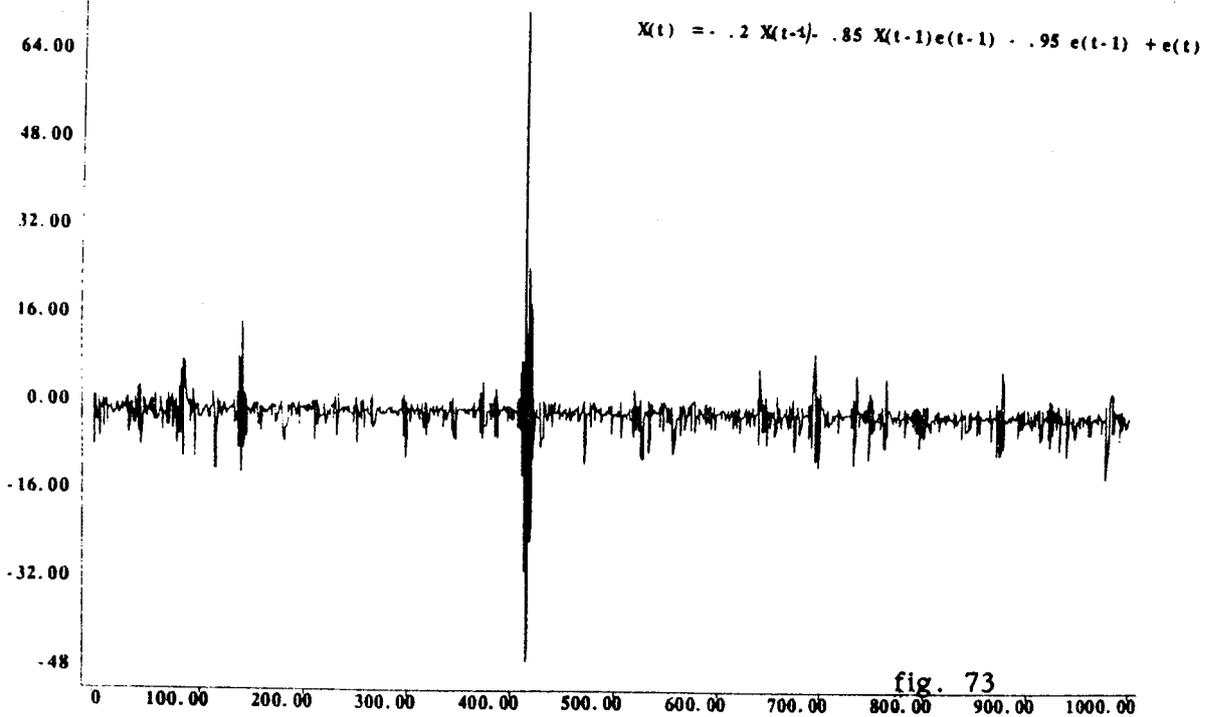
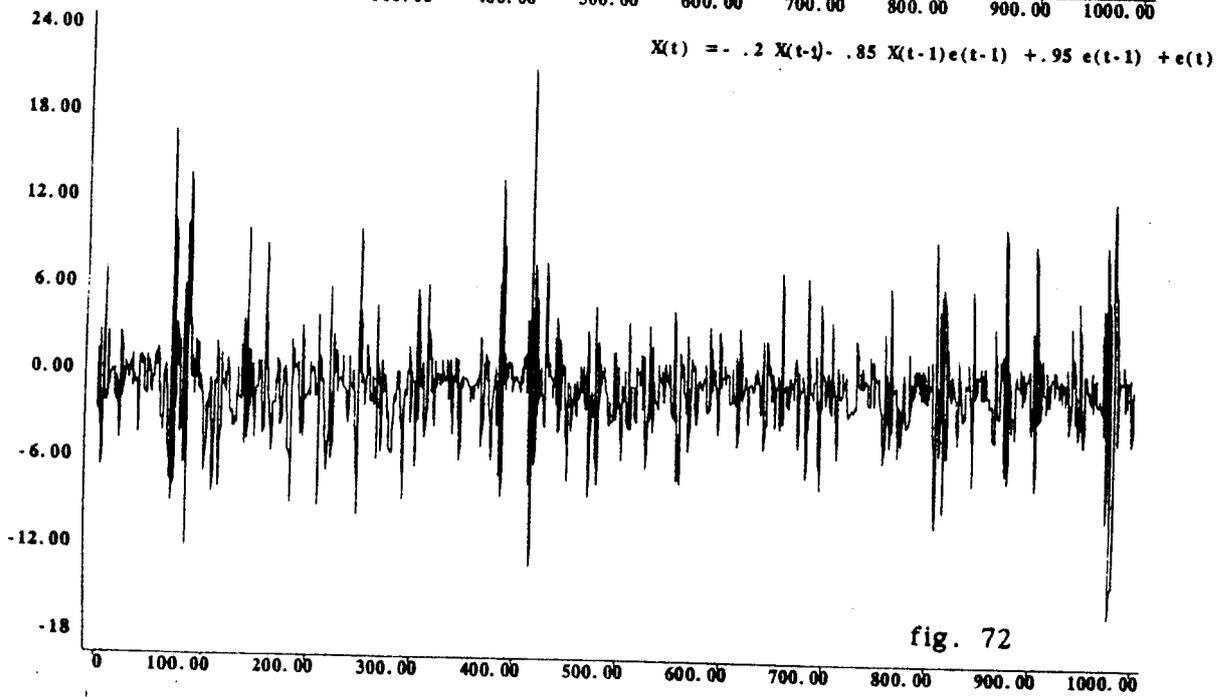
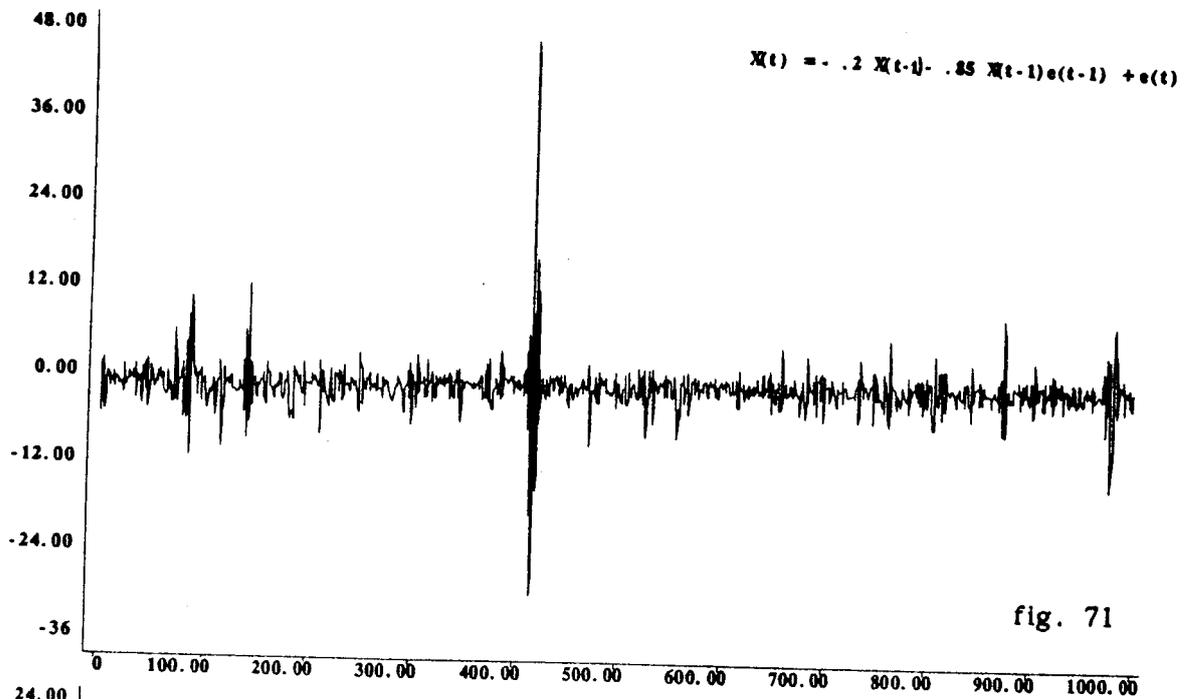


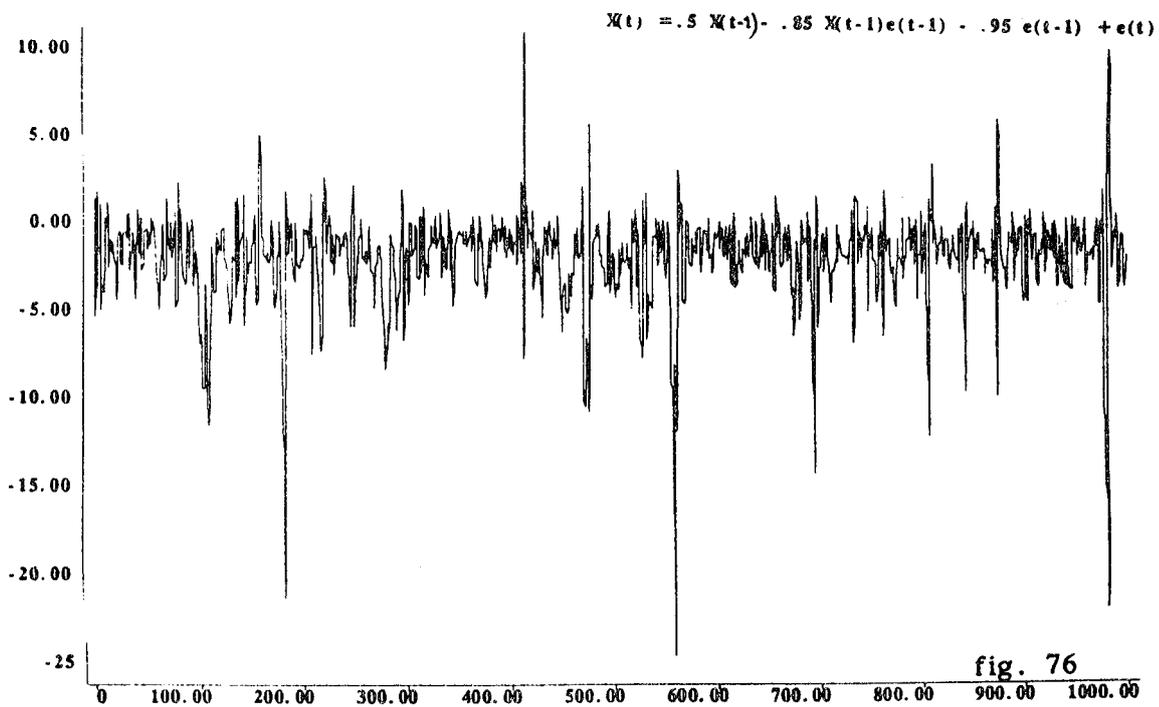
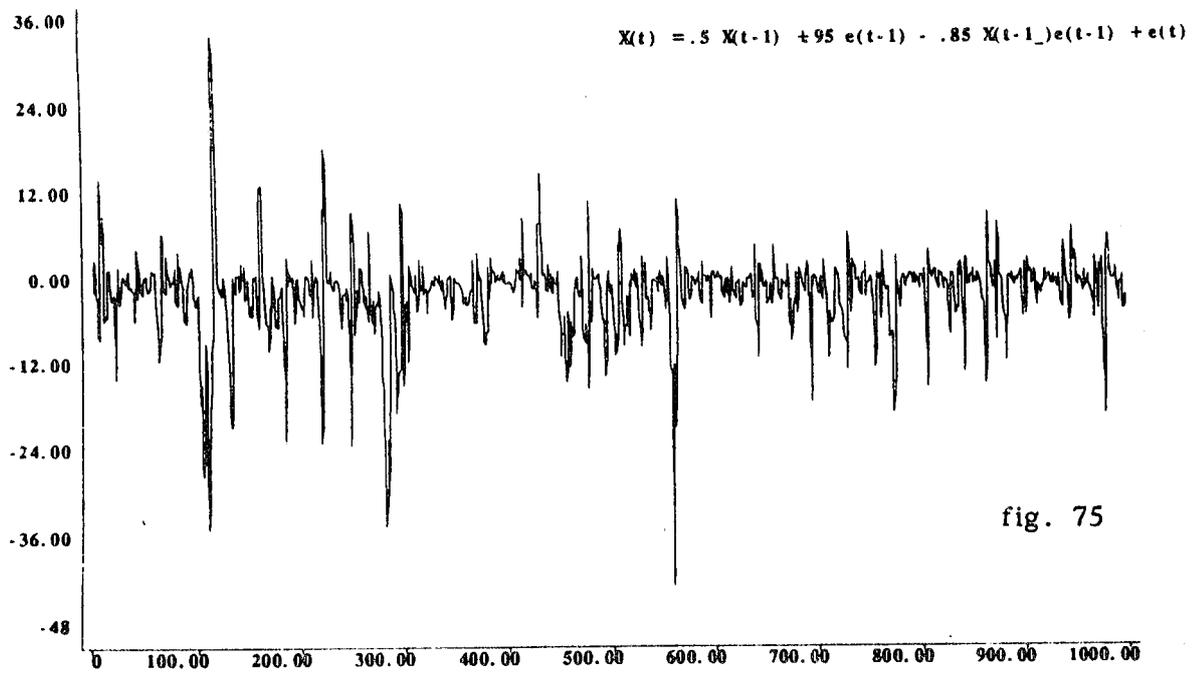
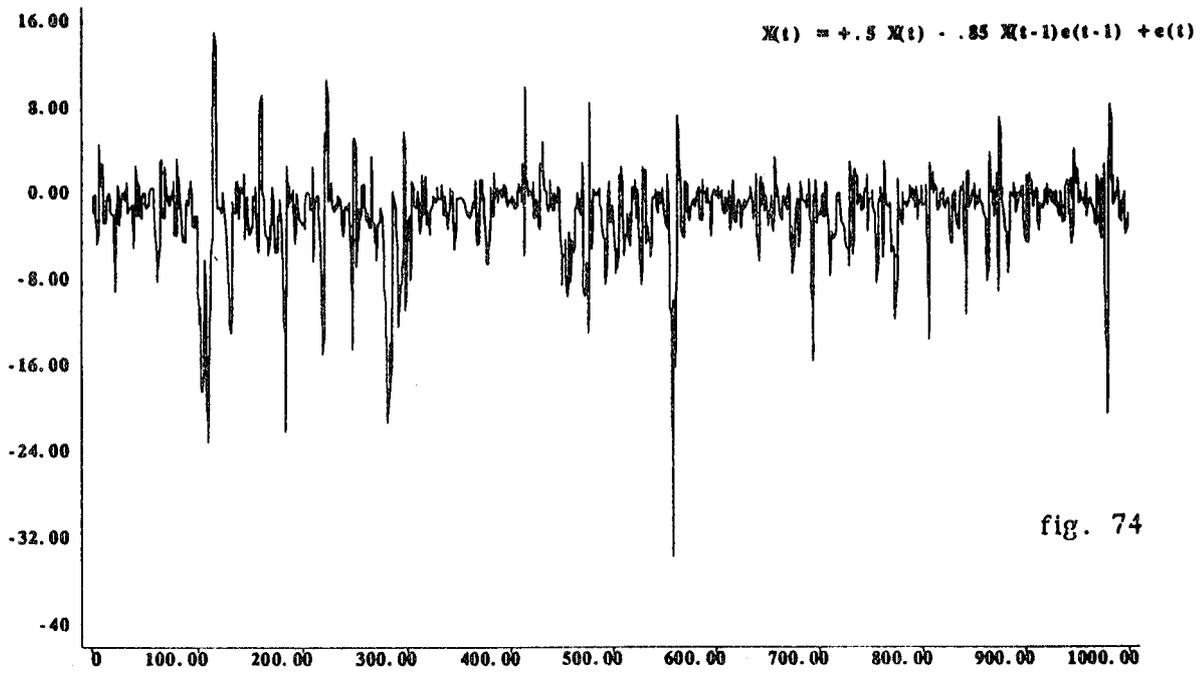


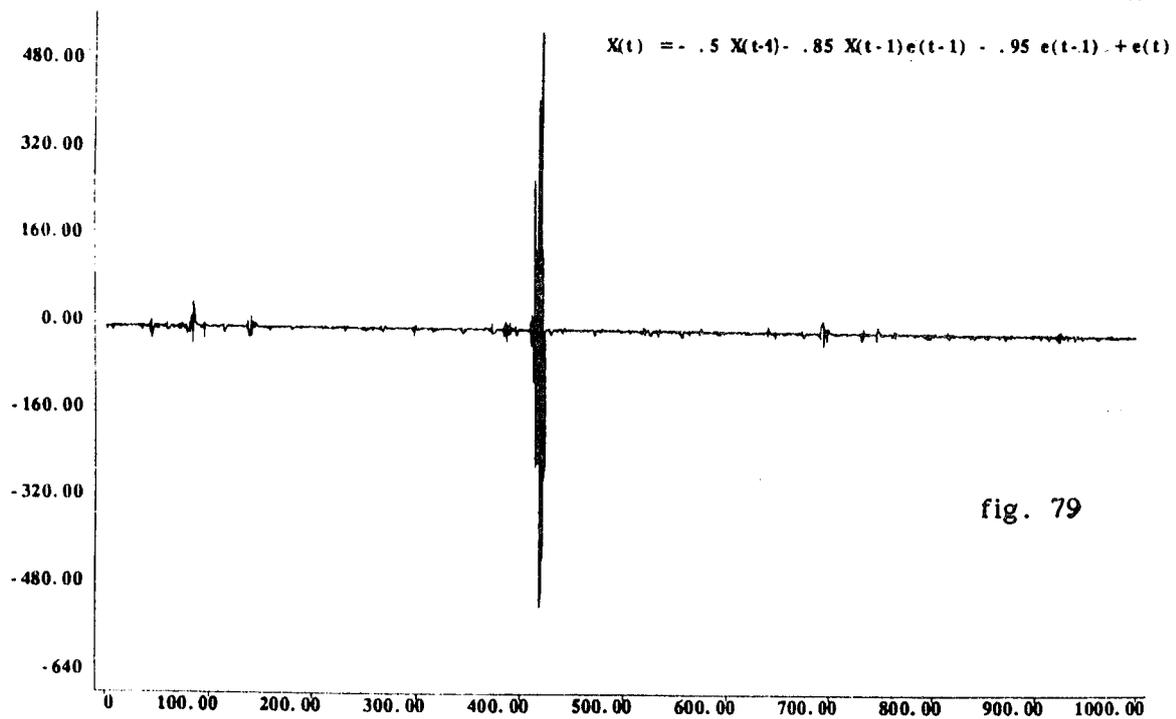
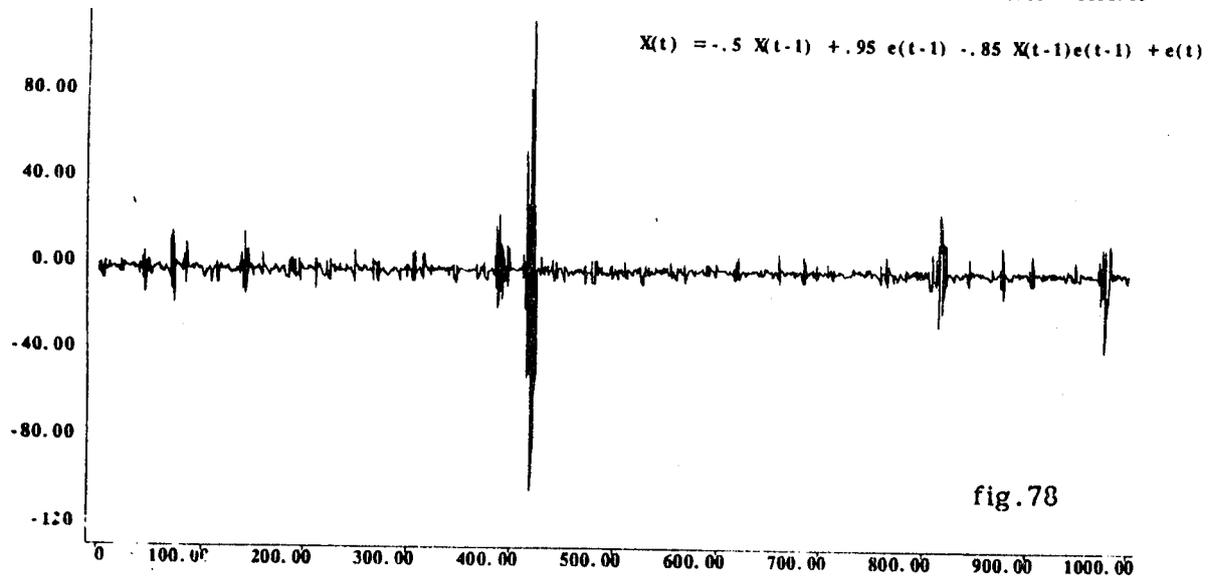
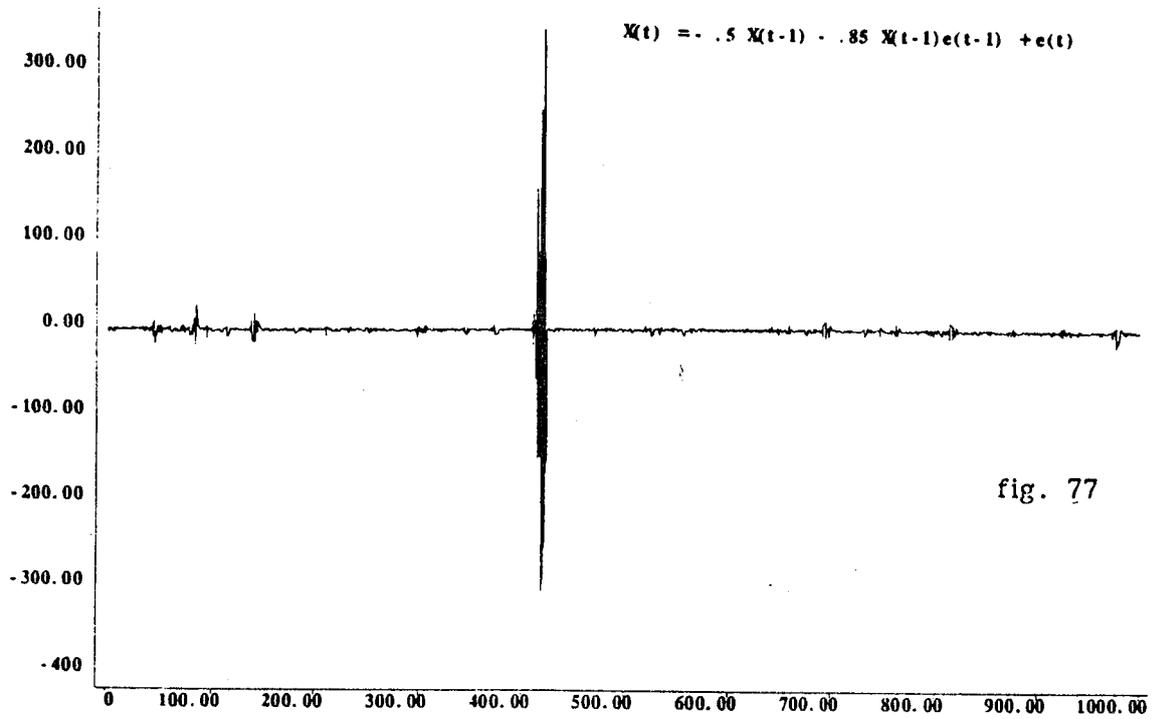












Quelques commentaires sur les simulations précédentes :

1 - Le signe des coefficients est très important :

Si a_1 et b_1 sont de même signe les effets se renforcent, les explosions sont fortes.

Si a_1 et b_1 sont tous les deux positifs, l'explosion est positive.

Si a_1 et b_1 sont tous les deux négatifs, les explosions sont alternées et légèrement moins fortes que dans le cas précédent.

Si a_1 et b_1 sont de signe opposé, il y a un effet de compensation : il n'y a pas d'explosions ou bien des explosions faibles. L'amplitude des oscillations varie dans le même sens que la valeur absolue de a_1 et b_1 .

2 - Le paramètre c_1 renforce ou diminue les tendances suivant le signe qu'il a par rapport au signe des autres paramètres.

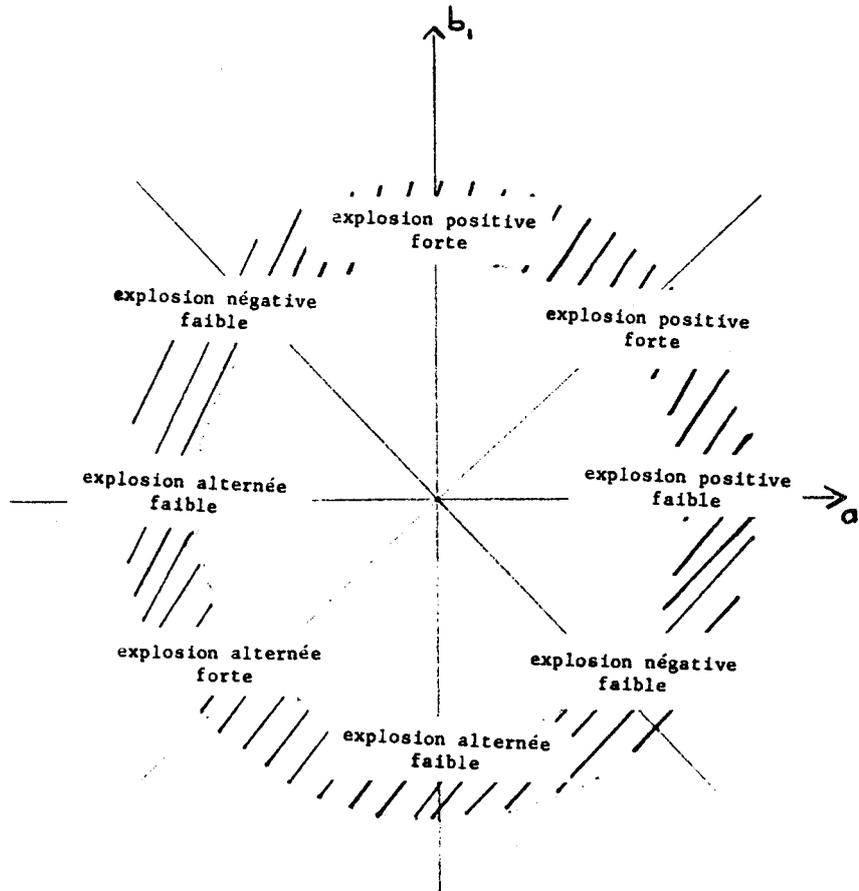
Quand a_1 et b_1 sont tous les deux positifs, si c_1 est positif, il renforce la tendance, si c_1 est négatif il diminue ou il inverse la tendance.

Quand a_1 et b_1 sont tous les deux négatifs, c_1 ne semble pas avoir une grande influence.

Quand a_1 et b_1 sont de signe contraire, c_1 n'intervient que si a_1 est très faible, et dans ce cas il renforce ou diminue la tendance suivant qu'il ait le même signe ou non que b_1 .

3 - Dans le cas où le coefficient de corrélation entre $A(t)$ et $B(t)$ est nul, on constate qu'il n'y a pas d'explosion, les effets semblent alors s'annuler.

4 - Dans le cas où c_1 est voisin de zéro, on peut représenter schématiquement le comportement des trajectoires par le graphe suivant:



L'anneau foncé correspond à la zone proche de la non stationnarité pour le modèle (2.1.2). On peut alors distinguer pour ce modèle trois grandes zones assez bien définies avec des phénomènes explosifs moins nets et plus perturbés aux alentours des frontières ainsi déterminées.

Remarques :

1. En dehors de l'explosion il peut se trouver des perturbations importantes, ou même des explosions secondaires qui n'apparaissent pas sur le graphique à cause du choix de l'échelle ajustée pour que le dessin contienne les valeurs maximales. Pour cette raison on a tracé les trajectoires des dix séries étudiées au paragraphe II avec une nouvelle échelle beaucoup plus importante, la même pour toutes. Ces séries sont représentées par les fig. 1.6 à 10.6.

2. Pour avoir une idée de la fréquence des explosions on a aussi tracé ces dix mêmes séries sur une longueur de 5000 points, cf. fig. 1.7 à 10.7.

L'explosion, pour ces séries, ne semble pas très fréquente et l'intervalle entre deux apparitions d'explosion peut être très variable.

$$X(t) = 1.55 X(t-1) - .81 X(t-2) + .1 X(t-1)e(t-1) + e(t)$$

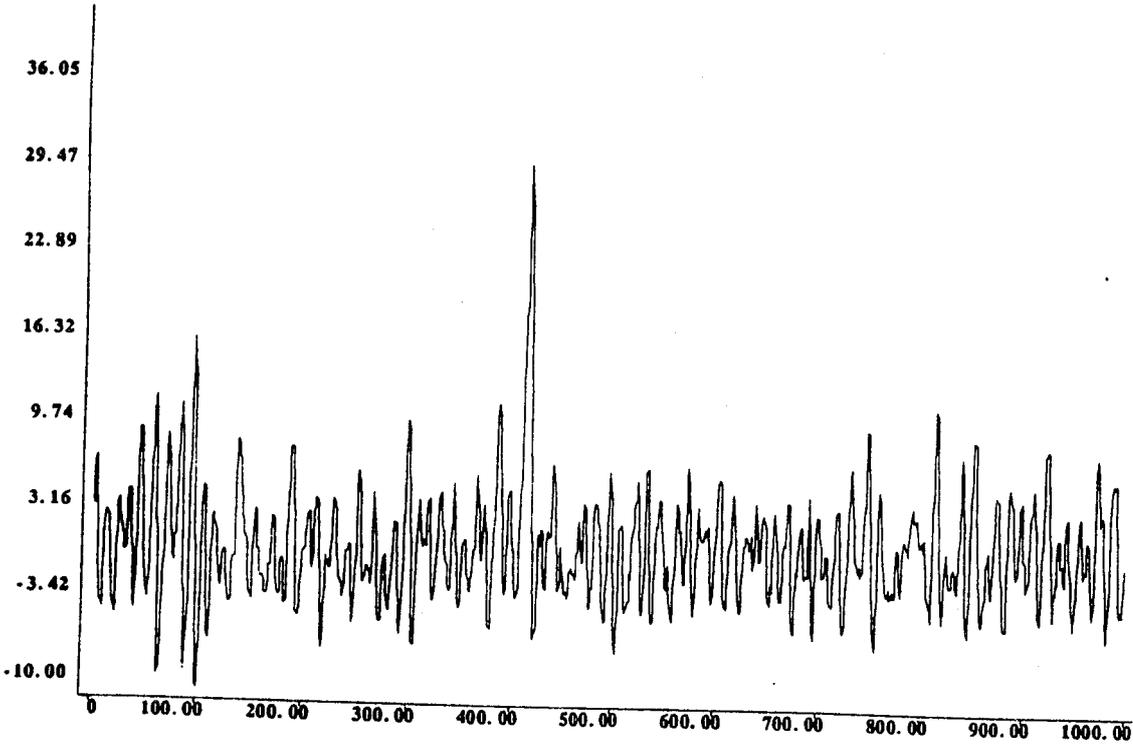


fig. 1.6

$$X(t) = 1.55 X(t-1) - .81 X(t-2) + .4 X(t-1)e(t-1) + e(t)$$

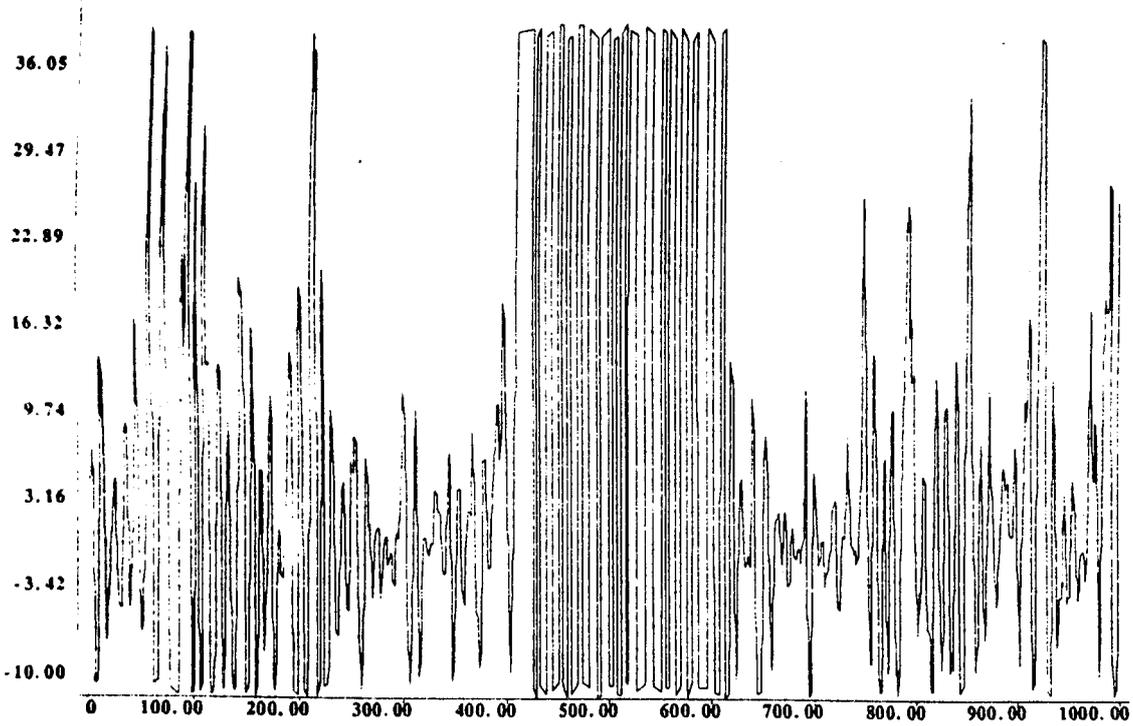


fig. 2.6

$$X(t) = .9 X(t-1) - .81 X(t-2) + .4 X(t-1)e(t-1) + e(t)$$

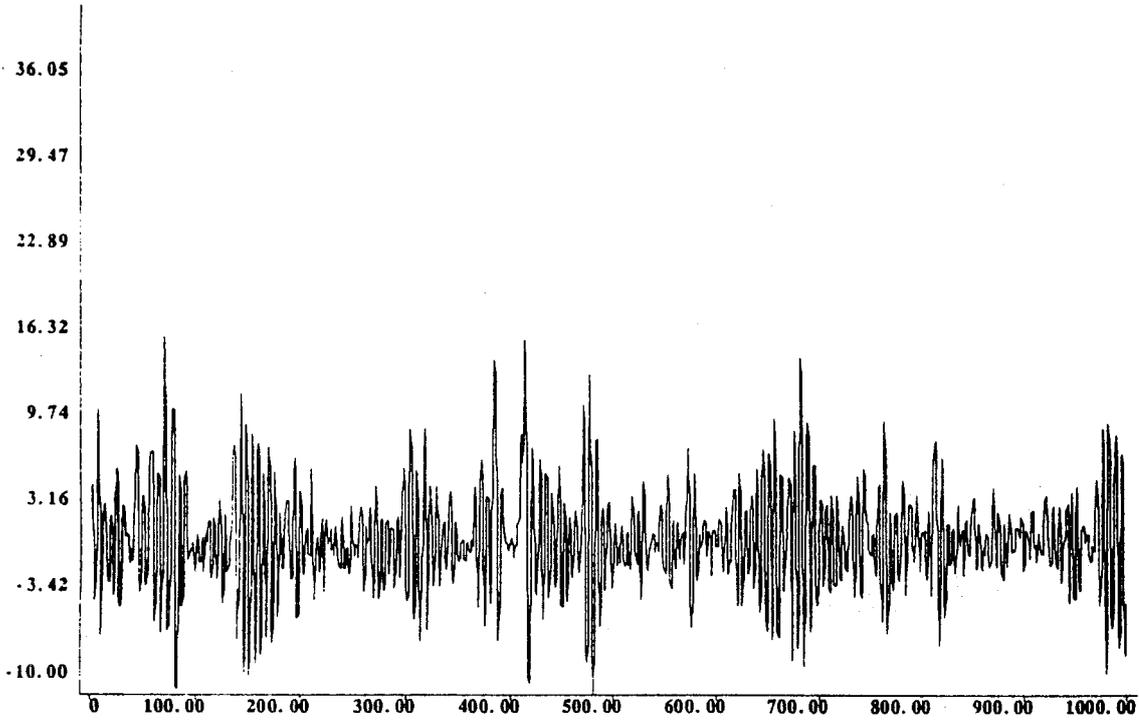


fig. 3.6

$$X(t) = .86X(t-1) - .25 X(t-2) + .6 X(t-1)e(t-1) + e(t)$$

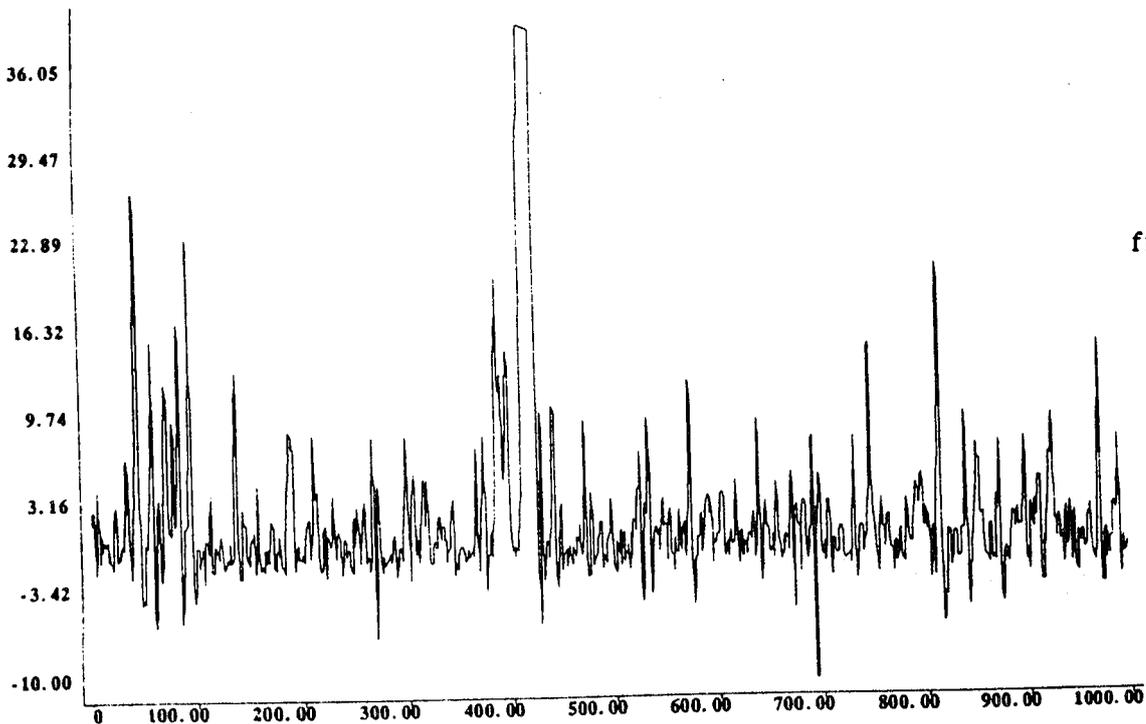


fig. 4.6

$$X(t) = .5 X(t-1) + .85 X(t-1)e(t-1) + e(t)$$

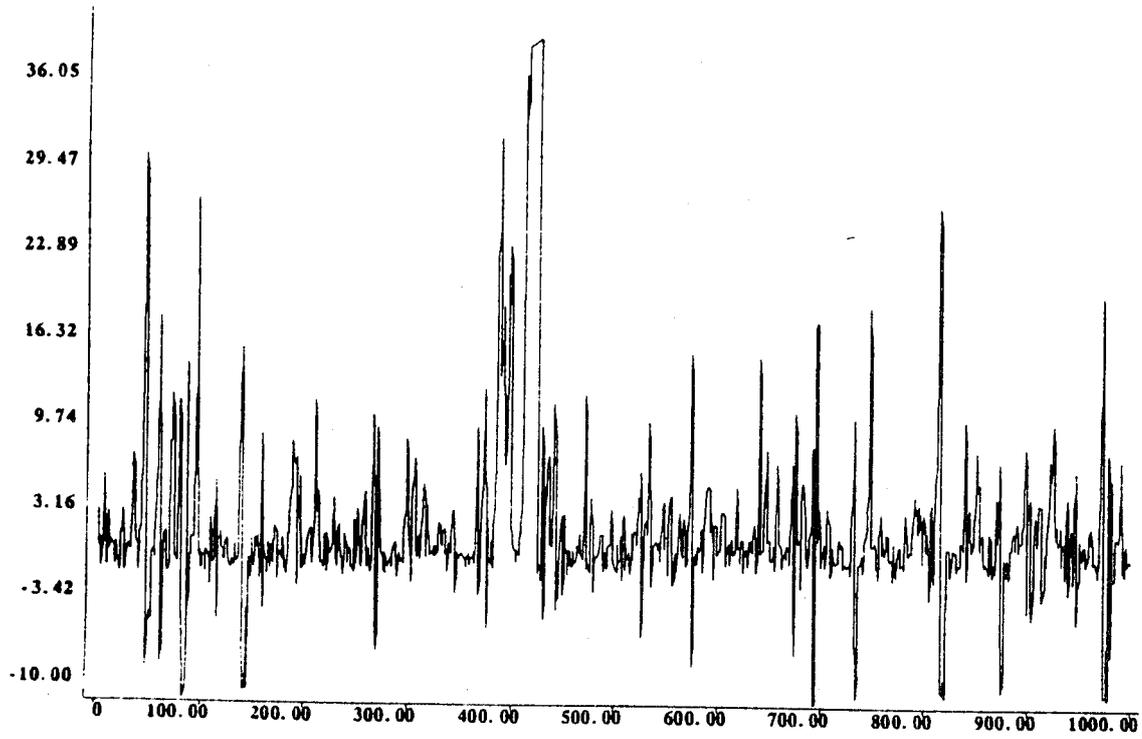


fig. 5.6

$$X(t) = .2 X(t-1) + .3 X(t-1)e(t-1) + e(t)$$

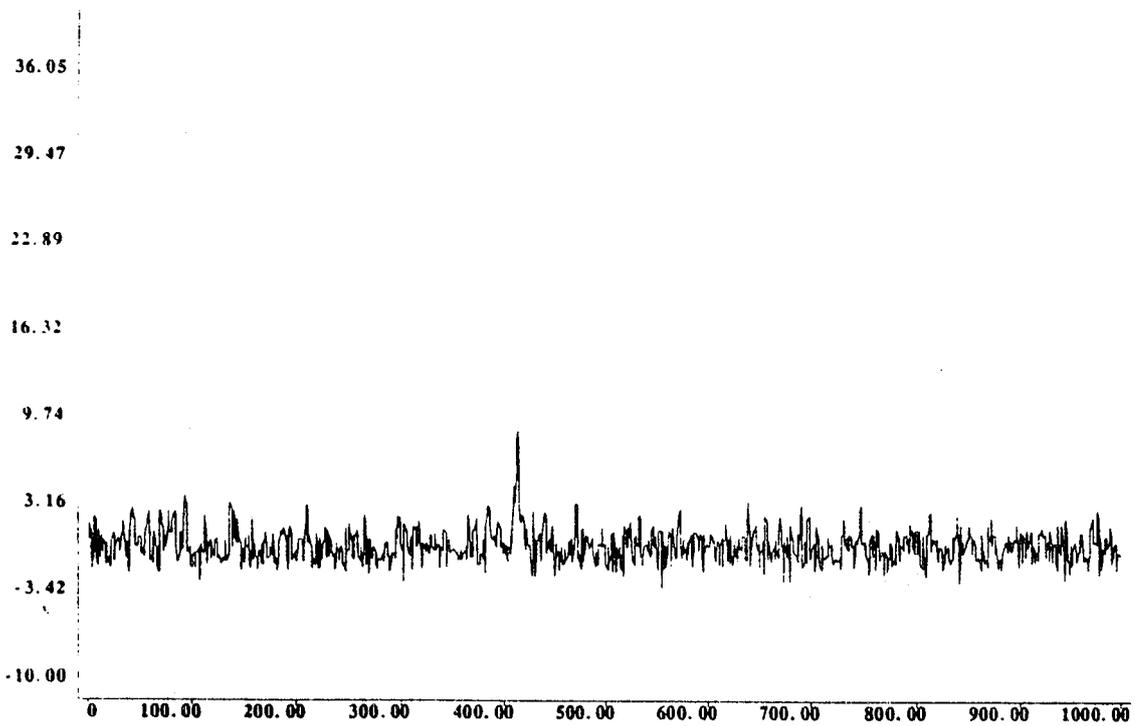


fig. 6.6

$$X(t) = .2 X(t-1) + .55 X(t-1)e(t-1) + e(t)$$

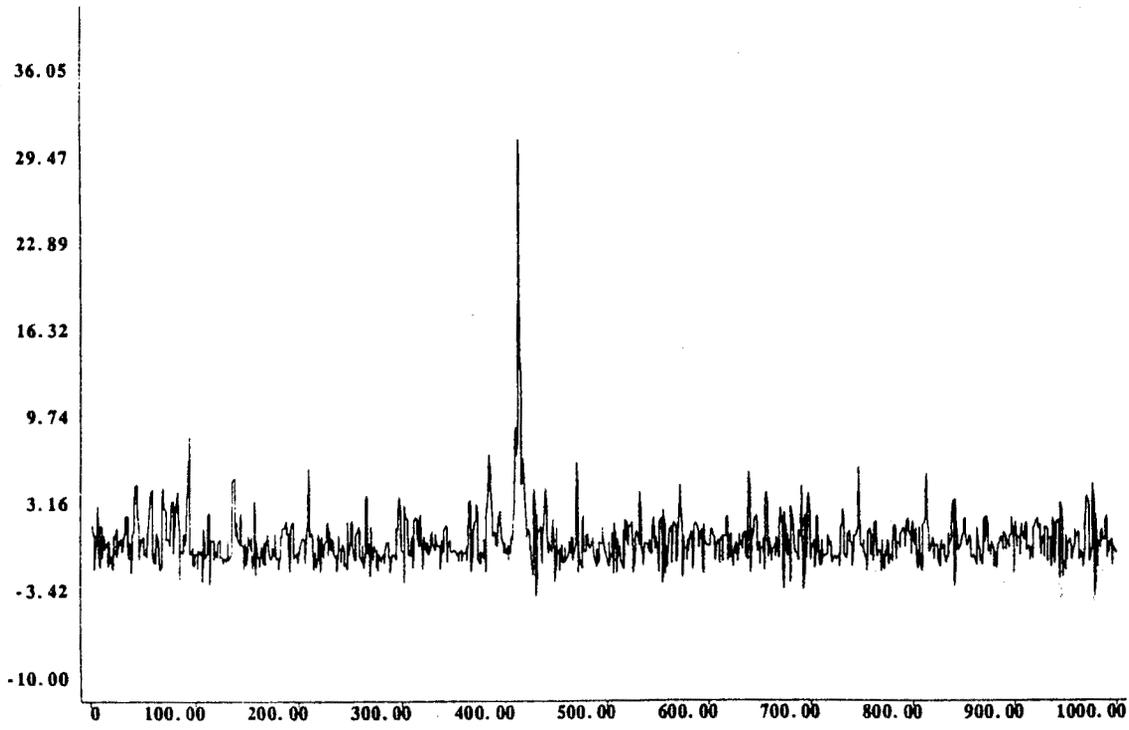


fig. 7.6

$$X(t) = .5 X(t-1) + .5 e(t-1) + .85 X(t-1)e(t-1) + e(t)$$

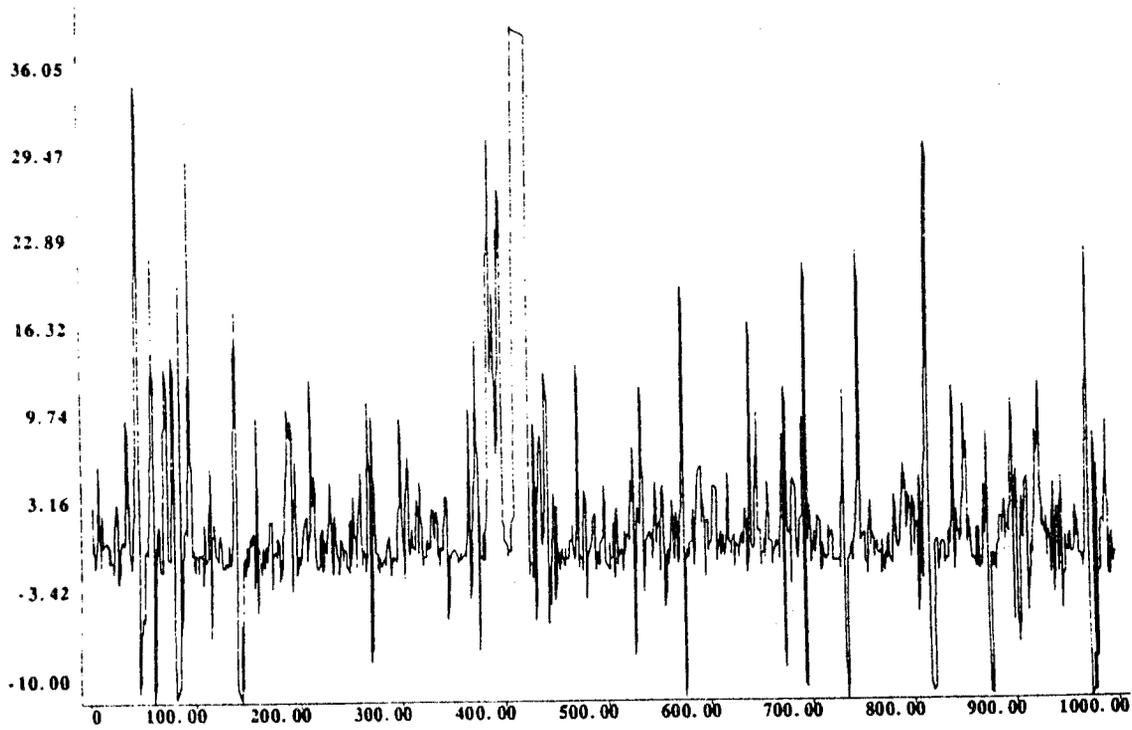


fig. 8.6

$$X(t) = .2 X(t-1) + .5 e(t-1) + .3 X(t-1)e(t-1) + e(t)$$

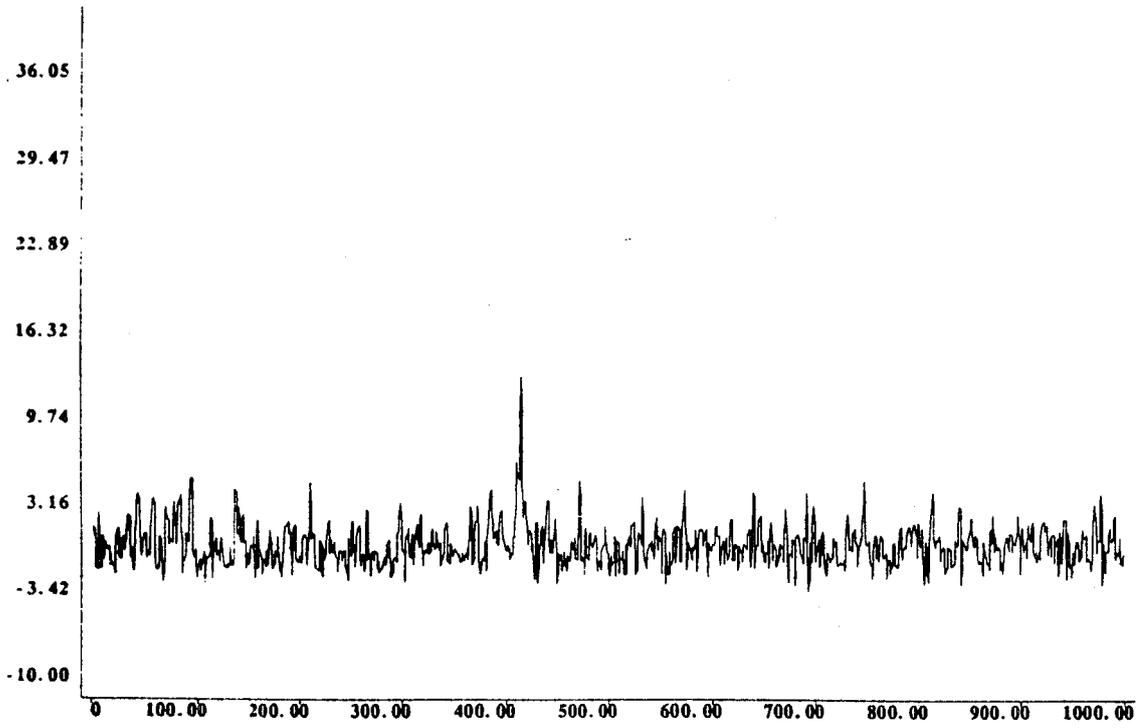


fig. 9.6

$$X(t) = .2 X(t-1) + .5 e(t-1) + .55 X(t-1)e(t-1) + e(t)$$

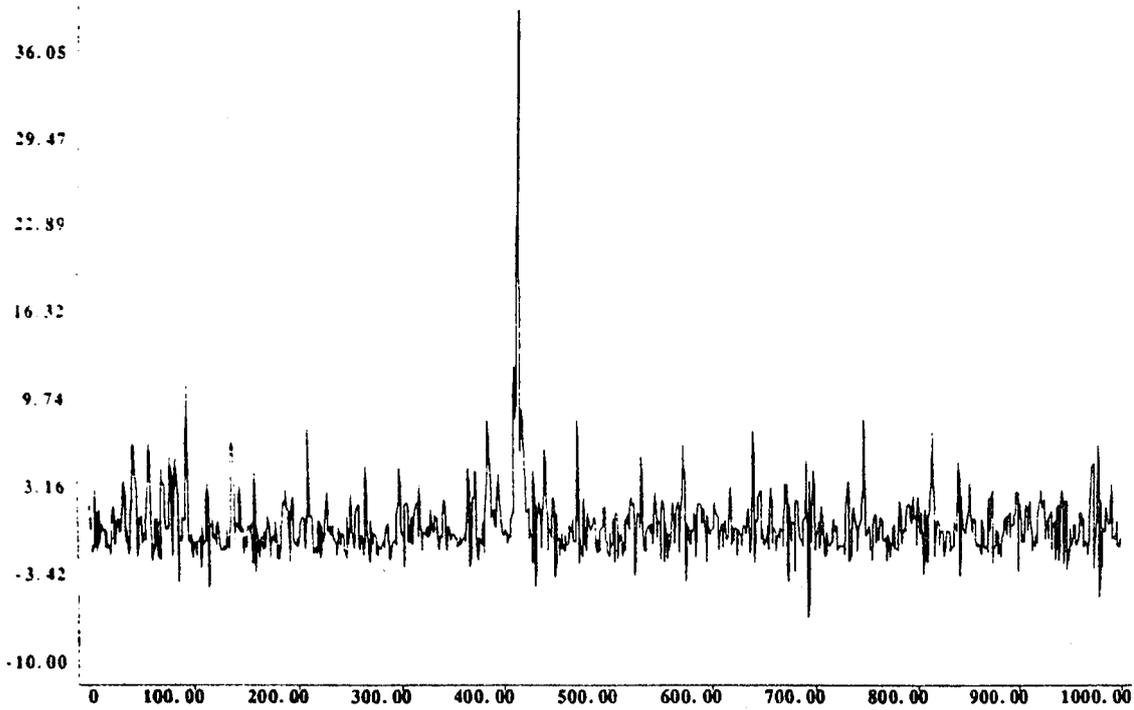


fig. 10.6

Serie 1

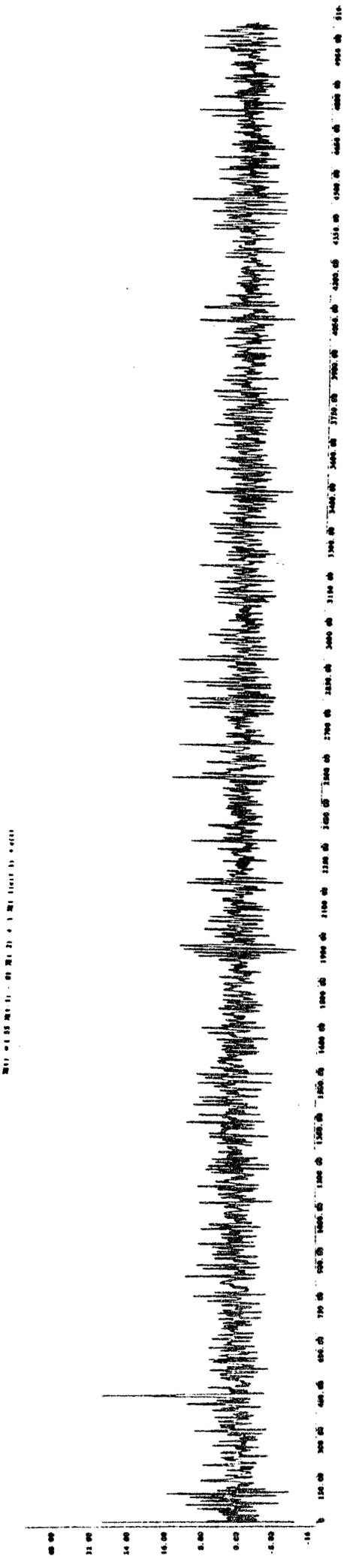


Fig. 1.7

Serie 2

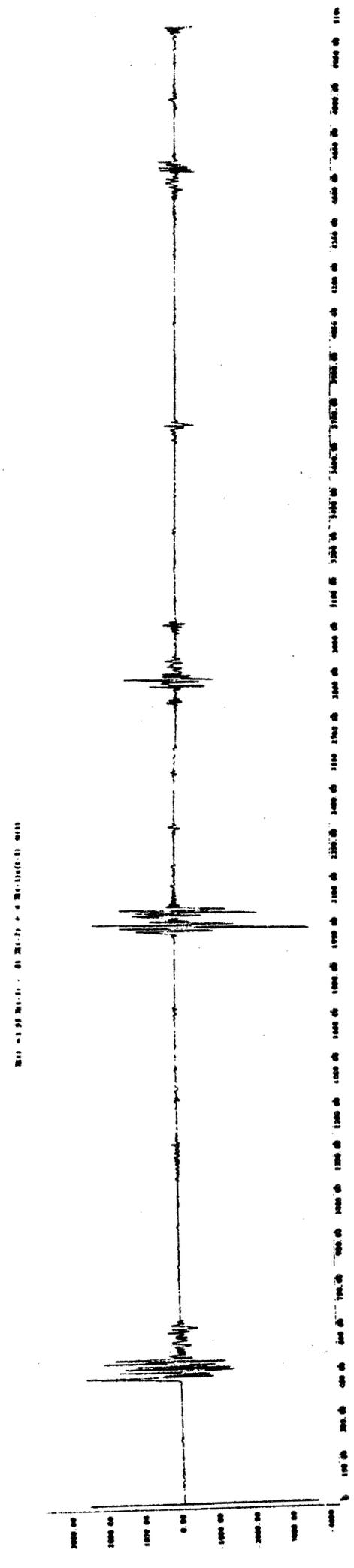


Fig. 2.7

Serie 3

MI 11 - 0.01 11 - 01 01 21 - 0.01 11 01 01 11

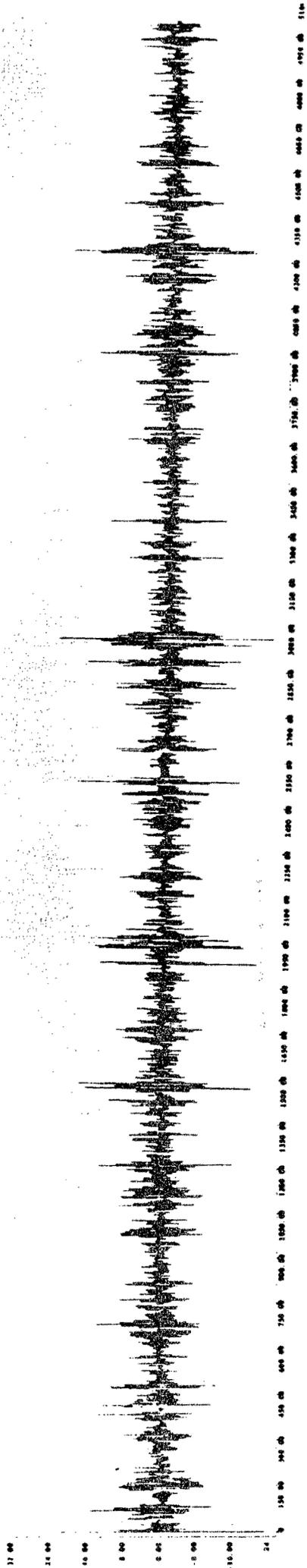


Fig. 3.7

Serie 4

MI 11 - 0.01 11 - 01 01 21 - 0.01 11 01 01 11

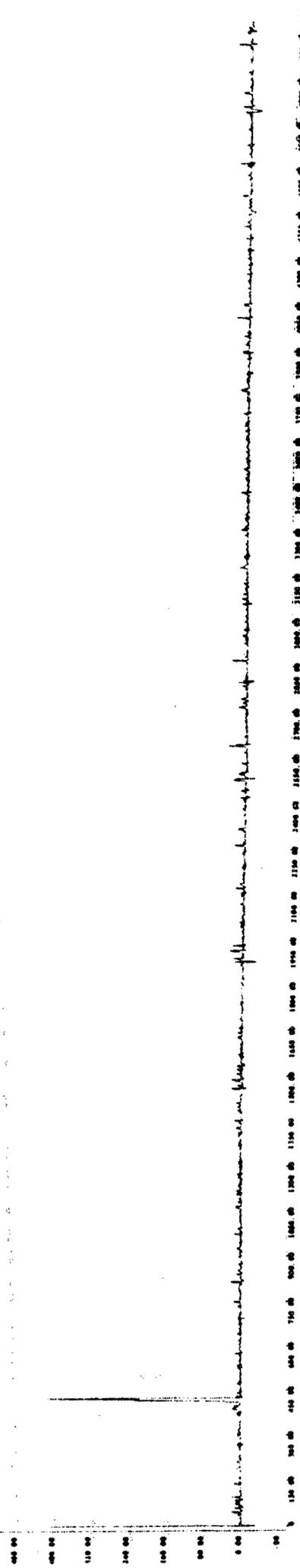


Fig. 4.7

Serie 5

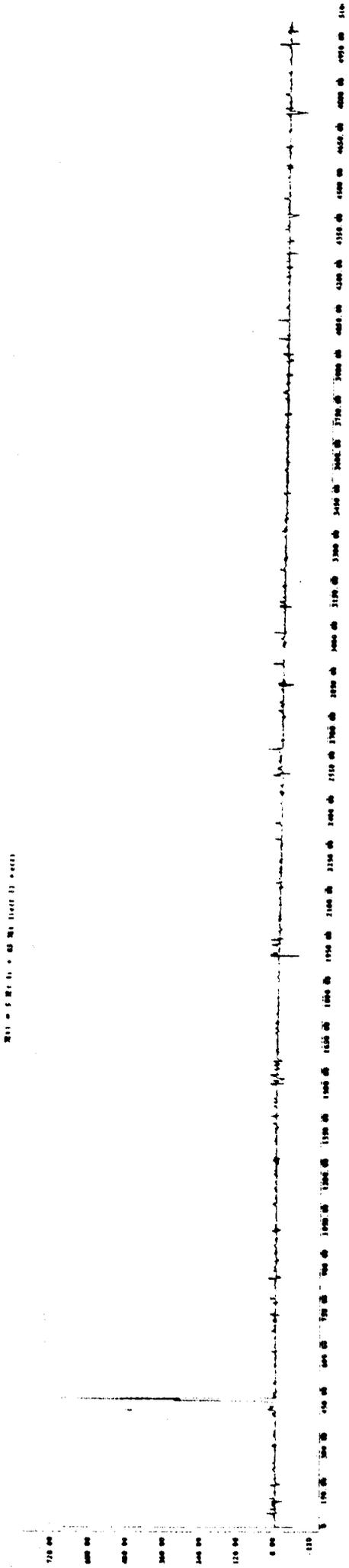


Fig. 5.7

Serie 6

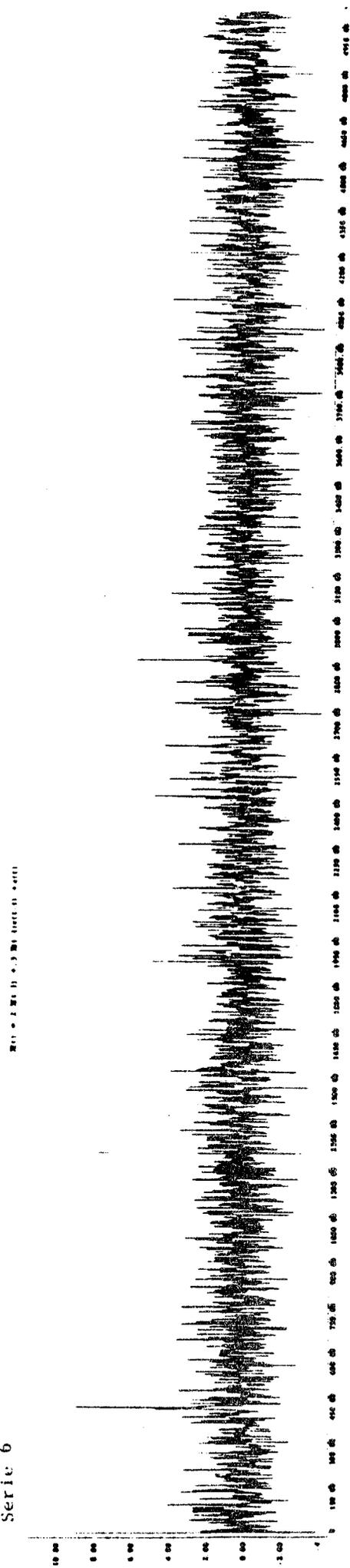


Fig. 6.7

Serie 7

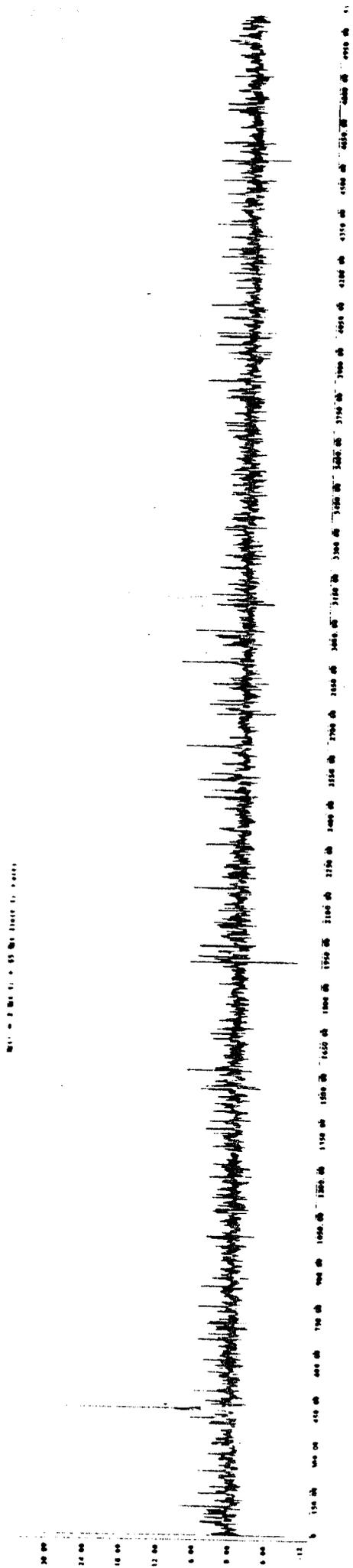


Fig. 7.7

Serie 8

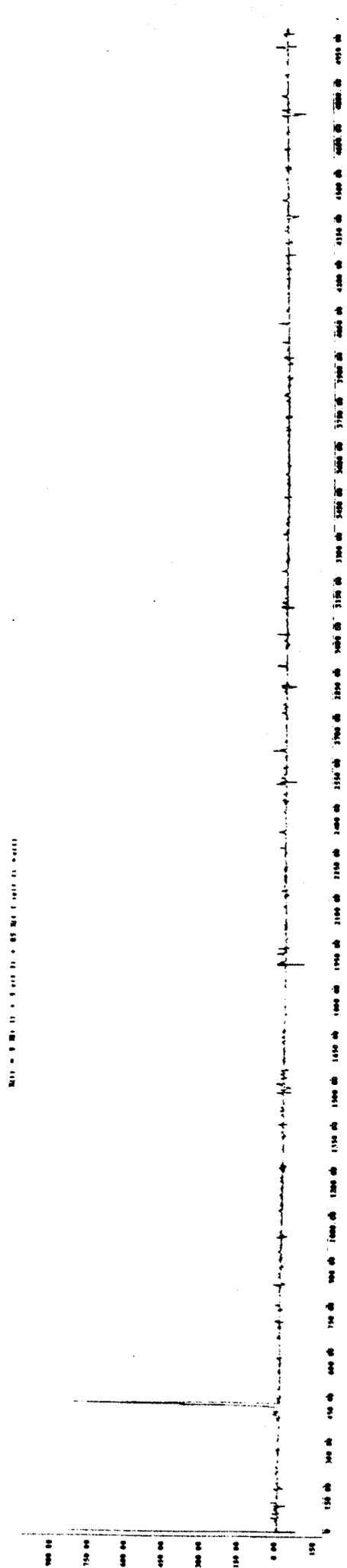


Fig. 8.7

Serie 9

W11 - 2 W11 (1) - 65 000 (1) - 85 W11 (1) - 4000 (1)

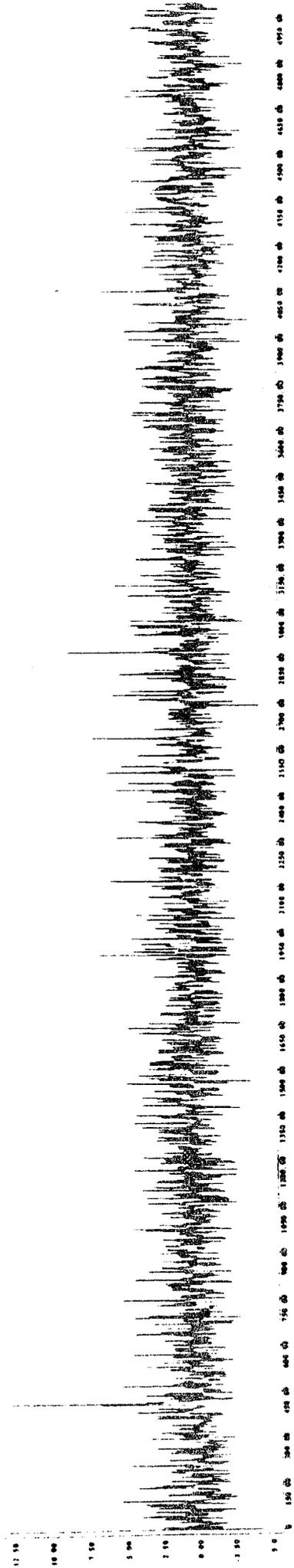


Fig. 9.7

Serie 10

W11 - 3 W11 (1) - 500 (1) - 655 W11 (1) - 800 (1)

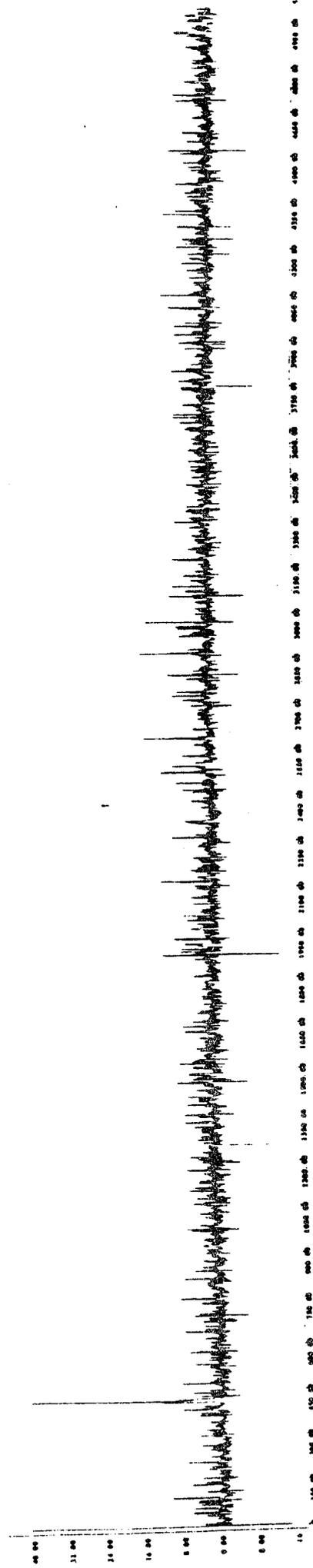


Fig. 10.7

IV - CONCLUSION

Cette présentation bien que rapide permet de mettre en évidence des éléments déterminants pour l'investigation des modèles bilinéaires.

Partant de trois types de modèles simples, nous avons pu montrer des comportements typiques de ces modèles, modèles avec effets cycliques plus ou moins perturbés, ou modèles à faible explosion, ou modèles à forte explosion avec des périodes de perturbation plus ou moins accentuées. Il ressort aussi de l'étude de ces trajectoires qu'une caractéristique importante de ces modèles est la présence d'explosions dont le phénomène peut être plus ou moins tempéré, amplifié, et dont l'apparition en est plus ou moins fréquente.

Il est à noter que la présentation des modèles telle qu'elle a été faite ici peut être, bien évidemment, étendue à des modèles d'ordre supérieur. Ces études peuvent conduire à une connaissance plus profonde des trajectoires des modèles considérés.

Si l'étude statistique descriptive de ces modèles nous permet d'avoir une première idée de leur comportement (histogramme, caractéristiques diverses), l'étude de la loi conjointe est aussi une démarche très utile car elle apporte sur le plan descriptif un complément indispensable à l'étude de l'évolution des trajectoires de tels modèles. Nous avons ainsi vu apparaître différentes formes de densités conjointes.

Notre attention s'est ensuite portée plus spécialement sur le phénomène d'explosion. Nous avons pu, dans le modèle d'ordre 1 discerner le rôle des coefficients dans ce phénomène. La représentation markovienne a permis de voir plus clair sur le

caractère explosif du processus. Le facteur multiplicatif $A(t)$ et son interaction avec le terme additif $B(t)$ nous a semblé un élément important pour l'apparition des explosions, leurs signes et leurs amplitudes.

Le caractère explosif des processus est assez typique des processus bilinéaires et on peut penser à utiliser ce genre de modèles pour modéliser des phénomènes physiques comme par exemple les éruptions volcaniques et le phénomène des nuits ardentes.

Note : L'ensemble des simulations de ce chapitre a été effectué par la méthode GENSTAT.

CHAPITRE III

OUTILS ET CONCEPTS POUR L'ANALYSE DE MODELES NON LINEAIRES

I - INTRODUCTION

La théorie classique du modèle linéaire est une théorie hilbertienne. On considère généralement des modèles stationnaires au second ordre $X(t)$ réguliers, et on sait que pour ces modèles on dispose de la décomposition de Wold :

$$X(t) = \sum_{k=0}^{\infty} a_k \epsilon(t-k) ,$$

où les $\epsilon(t)$ sont des variables aléatoires non corrélées. Rappelons que l'on appelle bruit blanc une suite de variables aléatoires de même variance, centrées et décorréélées, donc indépendantes dans le cas gaussien.

Toute procédure statistique sur les modèles linéaires ne se fait qu'à partir des covariances et des quantités déduites de celles-ci alors que pour les modèles non linéaires il faut aller plus loin et faire intervenir d'autres structures que les moments

d'ordre deux. Il en est ainsi des propriétés comme la singularité, la régularité où il est nécessaire de renforcer la définition en faisant intervenir la structure complète du modèle. La démarche qui consiste à ne considérer que les moments d'ordre deux, conduit à oublier d'autres structures dans le cas de modèles non linéaires. Ainsi le processus

$$X(t) = \epsilon(t) + \alpha \epsilon(t-1)\epsilon(t-2)$$

où $\epsilon(t)$ est un bruit blanc gaussien, en tant que processus au second ordre est un bruit blanc, alors que ses observations sont fortement dépendantes.

Afin d'acquérir une meilleure connaissance de la structure profonde des modèles non linéaires, il est nécessaire de se doter d'outils plus généraux que ceux utilisés en théorie des modèles linéaires. Nous allons présenter dans ce chapitre quelques uns de ces outils.

On va tout d'abord sortir du cadre de l'espace linéaire $H_t(X)$, sous espace de L^2 engendré par les $X(u)$, $u \leq t$, puis considérer l'espace L^2 en tant qu'espace de fonctions mesurables par rapport à ϵ , isomorphe au chaos de Wiener \mathcal{H} et montrer comment les modèles non linéaires se fondent bien dans cet espace. Ensuite on introduira les notions de régularité faible et forte, de singularité faible et forte, afin de faire une typologie des processus définis sur l'espace \mathcal{H} , (cf. D. GUEGAN, 1981 a, 1982 a). Les concepts utiles pour les modèles non linéaires tels que l'inversibilité et l'ergodicité seront considérés. Puis nous aborderons l'analyse spectrale pour ces modèles en introduisant le polyspectre et quelques unes de ses propriétés fondamentales. Nous généraliserons la représentation de Wold pour les processus linéaires réguliers.

II - CHAOS DE WIENER - RAPPELS ET NOTATIONS

Dans ce paragraphe, nous faisons une présentation rapide des chaos de Wiener à partir du livre de J. NEVEU (1968). Soit $\epsilon(t)$, $t \in \mathbb{Z}$ un bruit blanc gaussien, $\mathcal{F}(\epsilon)$ la σ -algèbre engendrée par tous les $\epsilon(t)$, et $\mathcal{F}_t(\epsilon)$ celle engendrée par les $\epsilon(u)$, $u \leq t$. Nous allons mettre une structure de chaos de Wiener sur l'espace $L^2(\mathcal{F}(\epsilon))$. Pour cela nous allons définir un isomorphisme entre $L^2(\mathcal{F}(\epsilon))$ et le chaos \mathcal{H} construit à l'aide des $\epsilon(t)$, $t \in \mathbb{Z}$. Nous avons besoin de quelques propriétés élémentaires sur les produits tensoriels et les produits symétriques.

1. Produits tensoriels

Soient F, G, H , trois espaces vectoriels sur \mathbb{R} .

Définition 3.2.1.

Le produit tensoriel $F \otimes G$ est le quotient de l'espace produit $F \times G$ par les relations d'équivalence

$$(af, g) = (f, ag) = a(f, g)$$

$$(f+f', g) = (f, g) + (f', g)$$

$$(f, g+g') = (f, g) + (f, g')$$

où $a \in \mathbb{R}$, $f, f' \in F$, $g, g' \in G$.

Cet espace dont les éléments sont notés $\sum a_k f_k \otimes g_k$ (sommes finies), factorise les applications bilinéaires de $F \times G$ dans H , en applications linéaires de $F \otimes G$ dans H . Si F et G sont des espaces de fonctions, définis sur deux espaces X et Y respectivement, le produit tensoriel de $F \otimes G$ est isomorphe à l'espace des fonctions

sur $X \otimes Y$ par l'application

$$\sum a_k f_k \otimes g_k \longrightarrow \sum a_k f_k(x) g_k(y)$$

L'injectivité de celle-ci se démontre à l'aide d'une représentation où les g_k sont indépendants.

D'autre part si F et G sont des espaces vectoriels préhilbertiens séparés pour des produits scalaires notés $\langle \cdot, \cdot \rangle_1$, et $\langle \cdot, \cdot \rangle_2$, alors $F \otimes G$ est un espace vectoriel préhilbertien séparé pour le produit scalaire engendré par la formule

$$\langle f \otimes g, f' \otimes g' \rangle = \langle f, f' \rangle_1 \cdot \langle g, g' \rangle_2$$

Dans le cas hilbertien, on note $F \hat{\otimes} G$ l'espace de Hilbert complété. Soit A_1 et A_2 deux espaces mesurés, on a alors l'identification par l'isomorphisme précédemment défini :

$$L^2(A_1 \otimes A_2) = L^2(A_1) \hat{\otimes} L^2(A_2) .$$

2. Produits tensoriels symétriques

Notons $H^{\otimes n}$ la puissance tensorielle $n^{\text{ième}}$ du Hilbert H complété, (pour simplifier nous supprimerons la notation $\hat{\cdot}$), et soit σ_n le groupe des permutations de $(1, 2, \dots, n)$. L'application u_σ définie par :

$$u_\sigma(f_1 \otimes f_2 \otimes \dots \otimes f_n) = f_{\sigma(1)} \otimes f_{\sigma(2)} \otimes \dots \otimes f_{\sigma(n)}$$

pour tout $f_j \in H$, $1 \leq j \leq n$, est une isométrie de $H^{\otimes n}$.

Définition 3.2.2.

La $n^{\text{ième}}$ puissance tensorielle symétrique de H , notée $H^{\otimes n}$ est le sous-espace vectoriel de $H^{\otimes n}$ des éléments invariants par tous les u_{σ} .

La projection de $H^{\otimes n}$ sur $H^{\otimes n}$ est l'opérateur $\frac{1}{n!} \sum_{\sigma} u_{\sigma}$.

Par convention, on pose :

$$f_1 \otimes f_2 \otimes \dots \otimes f_n = (n!)^{-1/2} \sum_{\sigma_n} f_{\sigma(1)} \otimes f_{\sigma(2)} \otimes \dots \otimes f_{\sigma(n)},$$

et $f^{\otimes n} = (n!)^{-1/2} f^{\otimes n}$,

de sorte que pour le même produit scalaire, on a :

$$\langle f_1 \otimes \dots \otimes f_n, g_1 \otimes \dots \otimes g_n \rangle = \sum_{\sigma_n} \langle f_1, g_{\sigma(1)} \rangle \dots \langle f_n, g_{\sigma(n)} \rangle$$

et $\langle f^{\otimes n}, g^{\otimes n} \rangle = n! \langle f, g \rangle^n$.

Dans ces conditions si (f_i) est une base orthonormée de H , une base orthonormée de $H^{\otimes n}$ est donnée par les produits :

$$\prod_{i \in I} (n_i!)^{-1/2} \otimes_{i \in I} f_i^{\otimes n_i} \quad \text{où } I \text{ est finie, } \sum_{i \in I} n_i = n.$$

Le produit tensoriel symétrique étant commutatif et distributif pour l'addition, on a :

$$(f+g)^{\otimes n} = \sum \binom{n}{k} f^{\otimes k} \otimes g^{\otimes (n-k)}.$$

Et de façon générale :

$$\left(\sum_{i=1}^n \lambda_i f_i\right)^{\otimes n} = \sum \frac{n!}{\prod_i n_i!} \prod_i \lambda_i^{n_i} \otimes f_i^{\otimes n_i} .$$

les λ_j étant des nombres de signe indépendants.

On en déduit, dans l'espace $\bigoplus_{n \geq 0} H^{\otimes n}$, que si on pose :

$$\exp \otimes f = \sum \frac{1}{n!} f^{\otimes n}$$

(On a la convergence immédiate dans la somme directe), alors :

$$\langle \exp \otimes f, \exp \otimes g \rangle = \exp \langle f, g \rangle$$

et :

$$\exp \left(\otimes \sum_j u_j f_j \right) = \sum \prod_j \frac{u_j^{n_j}}{n_j!} \otimes_j f_j^{\otimes n_j} .$$

3. Chaos de Wiener

Nous pouvons maintenant définir le chaos de Wiener \mathcal{H} .

Définition 3.2.3.

On appelle chaos de Wiener \mathcal{H} , la somme directe orthogonale des $H^{\otimes n}(\epsilon)$, sous espaces fermés engendrés par les combinaisons linéaires des produits symétriques formels

$\epsilon(t_1) \otimes \epsilon(t_2) \otimes \dots \otimes \epsilon(t_n)$ où $t_1, \dots, t_n \in \mathbb{Z}$:

$$\mathcal{H} = \bigoplus_{n \geq 0} H^{\otimes n}(\epsilon)$$

$H^{\otimes n}$ est appelée la $n^{\text{ième}}$ composante du chaos \mathcal{H} .

$H^{\otimes 0}$ est l'espace des constantes et on peut remarquer que $H^{\otimes 1} = H(\epsilon)$ est l'espace linéaire classique engendré par $\epsilon(t), t \in \mathbb{Z}$.

On identifie le chaos \mathcal{H} avec $L^2(\mathcal{F}(\epsilon))$ à partir du résultat suivant:

Théorème 3.2.1. (Wiener-Segal)

Il existe un isomorphisme unique φ de la somme directe hilbertienne $\bigoplus_{n \geq 0} H^{\otimes n}$ sur l'espace $L^2(\mathcal{F}(\epsilon))$, qui pour tout $X \in H$, est tel que :

$$\varphi(\exp \otimes X) = \exp\left[X - \frac{1}{2} E(X^2)\right],$$

à condition de poser :

$$\exp \otimes X = \sum_{n \geq 0} \frac{1}{n!} X^{\otimes n} \quad \text{dans} \quad \bigoplus_{n \geq 0} H^{\otimes n}.$$

L'application φ que l'on vient de définir est telle que :

$$\varphi\left(\bigotimes_i \epsilon(t_i)^{\otimes n_i}\right) = \prod_i \sigma_i^{n_i} h_{n_i} \left(\frac{\epsilon(t_i)}{\sigma_i}\right)$$

où $\sigma_i = \sqrt{E[\epsilon(t_i)^2]}$, et où h est un polynôme d'Hermite, les

$\epsilon(t_i)$ étant orthogonaux dans H .

Les polynômes d'Hermite sont des polynômes orthogonaux pour la mesure $(2\pi)^{-1/2} e^{-1/2 x^2} dx$, qui peuvent être aussi définis par la fonction génératrice :

$$\exp\left(ux - \frac{1}{2} u^2\right) = \sum_{n \geq 0} \frac{u^n}{n!} h_n(x), \quad (u, x) \in \mathbb{R}^2.$$

Ils ont encore pour expression

$$h_n(x) = \sum_{0 \leq m \leq \lfloor \frac{n}{2} \rfloor} \frac{n!}{(n-2m)! 2^m m!} (-1)^m x^{n-2m} .$$

En particulier, si les t_i , $1 \leq i \leq k$, sont tous distincts, on a :

$$\varphi(\epsilon(t_1) \otimes \epsilon(t_2) \otimes \dots \otimes \epsilon(t_k)) = h_1\left(\frac{\epsilon(t_1)}{\sigma_1}\right) \dots h_k\left(\frac{\epsilon(t_k)}{\sigma_k}\right) ,$$

et si les $\epsilon(t_i)$ forment une base orthonormée de $H(\epsilon)$, une base orthonormée de $L^2(\mathcal{F}(\epsilon))$ est donnée par les produits :

$$\prod_{i \in I} (n_i!)^{-1/2} h_{n_i}\left(\frac{\epsilon(t_i)}{\sigma_i}\right) ,$$

(où, sauf un nombre fini, tous les n_i sont nuls).

Ainsi la structure tensorielle algébrique de l'espace de Hilbert $\mathcal{H} = \bigoplus_{n \geq 0} H^{\otimes n}(\epsilon)$, construction algébrique "abstraite" se transporte à l'espace "concret" $L^2(\mathcal{F}(\epsilon))$ sur lequel nous travaillerons désormais.

On a alors le résultat suivant : Etant donné un processus $X(t)$, $\mathcal{F}_t(\epsilon)$ -mesurable, de carré intégrable, il possède une décomposition en fonction des polynômes d'Hermite sur les composantes du chaos :

$$(3.2.1) \quad X(t) = \sum_{k > 0} \sum_{t_1 < t} \dots \sum_{t_k < t} \alpha_{n_1, \dots, n_k, t_1, \dots, t_k} .$$

$$h_{n_1}(\epsilon(t_1)) \dots h_{n_k}(\epsilon(t_k))$$

où les coefficients $\alpha_{n_1, \dots, n_k, t_1, \dots, t_k}$ sont des réels

caractérisant le processus $X(t)$.

III - CONCEPTS FONDAMENTAUX

Nous présentons certains concepts utiles dans l'étude des modèles non linéaires comme l'inversibilité et l'ergodicité. Les notions de régularité et de singularité nous permettront de proposer une première classification des modèles non linéaires. Pour cela nous considérons des processus $X(t)$ appartenant à $L^2(\mathcal{F}(\epsilon))$.

1. Inversibilité

Nous avons vu au chapitre 1 que ce concept était fondamental d'une part pour estimer les paramètres d'un modèle, d'autre part pour faire des prévisions. Dans la littérature plusieurs définitions en ont été données.

En ce qui concerne les modèles bilinéaires une première approche fut introduite par C.W.J. GRANGER et A.P. ANDERSEN (1978 b) : étant donné un processus stochastique $X(t)$, $t \in \mathbb{Z}$, celui-ci est inversible s'il est possible d'estimer la suite $\epsilon(t)$ à partir des valeurs passées de $X(t)$ avec une exacte connaissance du modèle générateur.

Cette approche qui est très générale peut se préciser de la manière suivante. (Notons qu'elle peut autant s'appliquer aux modèles linéaires qu'aux modèles non linéaires). Posons :

$$X(t) = f(X(t-i), \epsilon(t-j)), \quad i = 1, \dots, k, \quad j = 1, \dots, \ell) + \epsilon(t)$$

où $\epsilon(t)$ est une suite de variables aléatoires indépendantes.

Soit $\hat{\epsilon}(t)$ un estimateur de $\epsilon(t)$ g n r  par :

$$\hat{\epsilon}(t) = X(t) - f(X(t-i), \hat{\epsilon}(t-j), i = 1, \dots, k, j = 1, \dots, \ell) \quad \text{o }$$

$$\hat{\epsilon}(t-j) = 0 \quad \text{pour } j > t. \quad \text{Alors le processus } X(t), t \in \mathbb{Z} \text{ est}$$

inversible si $\lim_{t \rightarrow \infty} E[(\epsilon(t) - \hat{\epsilon}(t))^2] = 0$ quand le mod le et les param tres sont compl tement connus. On peut remarquer que cette condition implique que le processus est inversible si la variance de l'erreur $e(t) = \epsilon(t) - \hat{\epsilon}(t)$, obtenue en estimant $\epsilon(t)$ en fonction d'un nombre fini de valeurs pass es et pr sentes de $X(t)$, tend vers z ro quand ce nombre devient infini, et ceci conditionnellement aux valeurs d'origine.

Une d finition pr cise de l'inversibilit  est alors la suivante :

D finition 3.3.1.

Un processus stochastique $X(t)$, $t \in \mathbb{Z}$ est inversible si $\epsilon(t)$ est mesurable par rapport   la tribu $\sigma(X(s), s \leq t)$.

On a aussi la d finition  quivalente suivante, (cf. D. GUEGAN, T. PHAM DINH, 87) :

D finition 3.3.2.

Un processus stochastique $X(t)$, $t \in \mathbb{Z}$ est inversible si et seulement si il existe une suite $\hat{\epsilon}(n)$ fonction de $X(1), \dots, X(n)$ uniquement telle que $\hat{\epsilon}(n) - \epsilon(n) \rightarrow 0$ en probabilit  quand $n \rightarrow \infty$.

Cette d finition est en particulier utilis e par T. PHAM DINH et L. TAT TRAN (1981), B.G. QUINN (1982 a), D. GUEGAN et T. PHAM DINH (1987 a,b).

2. Ergodicité

La notion d'ergodicité pour un processus $X(t)$, $t \in \mathbb{Z}$ défini sur un certain espace de probabilité est essentiellement liée à une propriété de moyenne pour l'ensemble des trajectoires du processus $X(t)$, $t \in \mathbb{Z}$, ce qui implique en particulier l'existence d'une loi invariante pour le processus considéré. Cette notion est utile pour utiliser des théorèmes concernant la loi des grands nombres.

C'est essentiellement à partir de la forme markovienne des modèles bilinéaires que cette propriété a été étudiée pour ces modèles, et les techniques utilisées sont celles qui ont été développées par R.L.TWEEDIE (1975, 1983 a).

Nous rappelons ici les principaux résultats utiles dans le cas des modèles qui nous intéressent.

Soit $X(t)$, $t = 0, 1, \dots$, une chaîne de Markov définie sur un espace métrique localement compact E ayant pour densité de transition :

$$P(x, A) = \mathbb{P}(X(t) \in A \mid X(t-1) = x), \quad x \in E, A \in \mathcal{E}$$

où \mathcal{E} est un élément de la tribu borélienne de E .

Rappelons qu'une chaîne de Markov $X(t)$, $t = 0, 1, \dots$, est φ irréductible si, pour une mesure φ quelconque définie sur

$$(E, \mathcal{E}) : \quad \sum_{t>0} P^t(x, A) > 0 \quad \forall x \in E, \quad \text{où}$$

$$P^t(x, A) = \mathbb{P}(X(t) \in A \mid X(0) = x) .$$

On a alors la définition suivante, si $\|\cdot\|$ représente la norme en variation :

Définition 3.3.3.

Une chaîne de Markov φ -irréductible est Harris ergodique s'il existe une mesure de probabilité π sur \mathcal{E} telle que pour tout $x \in \mathcal{E}$:

$$\|P^t(x, \cdot) - \pi(\cdot)\| \longrightarrow 0 \text{ quand } t \rightarrow \infty .$$

Cette définition d'ergodicité nécessite donc la notion de φ -irréductibilité et une des difficultés techniques dans les applications est d'établir cette propriété.

Dans le cas où une chaîne est φ -irréductible, on dispose d'un théorème établi par R.L. TWEEDIE permettant d'obtenir l'ergodicité d'une telle chaîne.

Théorème 3.3.1. (R.L. TWEEDIE, 1975)

Ω étant un espace d'état, soit $X(t)$, $t = 0, 1, \dots$ une chaîne de Markov φ -irréductible avec une probabilité de transition fortement continue.

Une condition suffisante d'ergodicité est qu'il existe un compact K sur \mathcal{E} avec $\varphi(K) > 0$ et une constante $c > 0$ telle que

$$\gamma_x \leq -c \quad \forall x \notin K$$

où

$$\gamma_x = E(\|X(t+1)\| - \|X(t)\| \mid X(t) = x) .$$

Dans le cas des processus bilinéaires simples, D. GUEGAN (1983) obtient la condition suffisante d'ergodicité en établissant l'existence d'un compact $K \subset E$, et d'une fonction mesurable non négative g sur E vérifiant les conditions suivantes :

- (i) $\int_E P(x, dy) g(y) \leq g(x) - 1 \quad \forall x \in K .$
- (ii) $\int_E P(x, dx) g(y) = \lambda(x) \leq B < \infty \quad \forall x \in K \quad \forall B > 0 \text{ fixé .}$

Les résultats auxquels conduisent le théorème 3.3.1. peuvent être améliorés afin de démontrer l'ergodicité géométrique des chaînes considérées. La notion d'ergodicité géométrique qui est une notion propre aux processus markoviens est importante car elle englobe les propriétés comme Harris-réurrence, ergodicité, et mélange.

Rappelons qu'une chaîne ergodique $X(t)$ est dite géométriquement ergodique, s'il existe $\rho < 1$ tel que :

$$\|P^t(x, \cdot) - \pi(\cdot)\| = O(\rho^t) \text{ pour } \pi\text{-presque tout } x ,$$

$\|\cdot\|$ désignant la norme en variation.

On dispose alors du résultat suivant (théorème 3.3.2) établi par E. NUMMELIN et P. TUOMINEN (1982) et R.L. TWEEDIE (1983 a). La chaîne est supposée φ -irréductible, et faiblement de Feller (i.e. pour chaque fonction continue bornée g sur E , la fonction de x définie par $E[g(X(t)) \mid X(t-1) = x]$ est aussi continue, (cf. D. REVUZ, 1975)). On définit les petits ensembles de la manière suivante : A est dit petit si A est relativement compact et $\varphi(A) > 0$ (cf. E. NUMMELIN et P. TUOMINEN, 1982, p. 188). On a alors le théorème suivant :

Théorème 3.3.2. (R.L. TWEEDIE, 1983 a)

Supposons que $X(t)$, $t \geq 0$ soit une chaîne de Feller, qu'il existe une mesure φ et un ensemble compact A avec $\varphi(A) > 0$ tel que :

- (i) $X(t)$, $t \geq 0$ soit φ -irréductible
- (ii) il existe une fonction non négative continue $g : E \rightarrow \mathbb{R}$ satisfaisant $g(x) \geq 1 \quad \forall x \in A$, et pour $\delta > 0$ telle que

$$E[g(X(t)) \mid X(t-1) = x] \leq (1-\delta)g(x), \quad x \in A^c.$$

Alors $X(t)$, $t \geq 0$ est géométriquement ergodique.

Ce résultat sert dans le cas des modèles autorégressifs généralisés à coefficients aléatoires $X(t)$, $t \in \mathbb{Z}$ définis par les équations (5.2.6) : T. PHAM DINH (1986) a démontré, la régularité absolue de ces processus. Cette dernière notion est définie de la façon suivante. Soit $Y(t)$, $t \in \mathbb{Z}$ un processus vectoriel stationnaire, on pose :

$$\Delta_n = \sup_{h \in S_n} E[h \mid Y(s), s < 1] - E[h]$$

où S_n représente l'ensemble des fonctions aléatoires mesurables pour la tribu engendrée par $Y(t)$, $t \geq n$, et bornées par 1 en valeur absolue, alors le processus $Y(t)$ est dit absolument régulier si $E[\Delta_n]$ tend vers zéro quand n tend vers l'infini.

Notons que la régularité absolue, avec une vitesse de convergence exponentielle vers zéro de $E[\Delta_n]$, est équivalente à l'ergodicité géométrique pour des processus markoviens.

3 - Régularité et Singularité

Les notions de régularité et singularité vont permettre de préciser la structure des modèles non linéaires.

Définition 3.3.4.

Etant donné un processus $X(t)$, $t \in \mathbb{Z}$

$X(t)$ est régulier linéaire si $H_{-\infty}(X) = \{0\}$

$X(t)$ est régulier fort si $\mathcal{F}_{-\infty}(X) = \{\emptyset, \Omega\}$ où

$$\mathfrak{F}_{-\infty}(X) = \bigwedge_t \mathfrak{F}_t(X) .$$

Définition 3.3.5.

Etant donné un processus $X(t)$, $t \in \mathbb{Z}$:

$X(t)$ est singulier linéaire si $H_{-\infty}(X) = H_{+\infty}(X) = H$

$X(t)$ est singulier fort si $\mathfrak{F}_{-\infty}(X) = \mathfrak{F}_{+\infty}(X)$ où

$$\mathfrak{F}_{+\infty}(X) = \bigvee_t \mathfrak{F}_t(X) .$$

On en déduit qu'un processus stationnaire au second ordre $X(t)$, $t \in \mathbb{Z}$ est non singulier linéaire si et seulement si $H_{-\infty}(X) \neq H_{+\infty}(X)$, ce qui n'est possible que si $H_{t+1}(X) \neq H_t(X)$, de même un processus strictement stationnaire $X(t)$, $t \in \mathbb{Z}$ est non singulier fort si et seulement si

$$\mathfrak{F}_t(X) \neq \mathfrak{F}_{t-1}(X) , \forall t \in \mathbb{Z} .$$

Dans le cas linéaire, on dispose de critères, de conditions nécessaires et suffisantes permettant de dire qu'un processus est régulier, (par exemple que la densité spectrale est continue), ou qu'il possède une décomposition de Wold suivant son processus d'innovations, ou singulier, cf. I. IBRAJIMOV et Y. ROZANOV, 1974.

Etant donné un processus $X(t)$ on peut donc considérer son processus d'innovation, au sens faible défini par $X(t) - \tilde{X}(t|t-1)$ où $\tilde{X}(t|t-1)$ représente la prédiction linéaire, ou au sens fort défini par $X(t) - E[X(t) | \mathfrak{F}_{t-1}(\epsilon)]$, et l'on sait que dans le cas linéaire tout processus régulier peut être exprimé comme une combinaison linéaire faisant intervenir le processus d'innovations.

Dans le cas non linéaire on peut se demander si un résultat analogue peut être obtenu. Une démarche possible est la suivante : Considérons la loi conditionnelle de $X(t)$ sachant

$X(t-1), X(t-2), \dots$, et soit $F(x|X(t-1), X(t-2), \dots)$ la fonction de répartition de cette loi conditionnelle, ainsi que $U(t)$, $t \in \mathbb{Z}$, une suite de variables aléatoires indépendantes de lois uniformes. On peut montrer que le processus $Y(t)$ défini par $Y(t) = F(U(t), Y(t-1), Y(t-2), \dots)$ a la même loi que X , (car $U(t)$ est uniforme, et les lois conditionnelles sont les mêmes d'après le théorème de C.T. IONESCU TULCEA, (cf. M. ROSENBLATT, 1971)). On peut donc construire un processus bruit blanc (qui est ici $U(t)$) mesurable par rapport à celui-ci et qui a la même loi. Remarquons que la construction n'est pas unique et que $U(t)$ n'est pas forcément un processus d'innovations.

Ainsi il n'est pas possible dans le cas non linéaire de disposer de critères permettant de dire qu'un processus est régulier. On doit donc travailler cas par cas. En particulier s'il est facile de voir que la régularité forte est plus forte que la régularité faible, et que la singularité faible est plus forte que la singularité forte, l'inverse n'est pas vrai comme on peut le constater dans l'exemple suivant. Considérons le processus $X(t)$ défini par

$$X(t) = \epsilon(t) + b_{21}X(t-2)\epsilon(t-1)$$

où $\epsilon(t)$ est un bruit blanc gaussien, de variance σ^2 , et où $b_{21}^2\sigma^2 < 1$, c'est un processus bruit blanc. Il est donc régulier linéaire, mais il est aussi régulier au sens fort, car $\mathcal{F}_{-\infty}(X) \subseteq \mathcal{F}_{-\infty}(\epsilon) = \{0, \Omega\}$, (cf. D. GUEGAN, 1981 a).

Ainsi parmi les processus stationnaires au second ordre, il existe :

- (1) des processus réguliers linéaires et singuliers forts.
- (2) des processus singuliers forts et non singuliers linéaires.

Nous allons construire un exemple pour chacune de ces classes, (cf. D. GUEGAN, 1981 a):

Classe (1) : Soit ϵ une variable $N(0,1)$, et soit le processus $X(t)$ défini par :

$$X(t) = \epsilon^{-t} - E(\epsilon^{-t}) \quad t < 0$$

Si $K_t = \overline{\text{sp}\{X(u), \text{ et } t \leq u < 0\}}$, alors $\dim(K_t) = t$, car la variable aléatoire n'est pas combinaison linéaire de $1, \epsilon, \epsilon^2, \dots, \epsilon^{-t-1}$. Posons alors $I_t = K_t \ominus K_{t-1}$, où K_t et K_{t-1} sont des espaces complémentaires dans $L^2(\mathcal{F}(\epsilon))$, alors $\dim(I_t) = 1$. Le polynôme ϵ^{-t} appartient à K_t et une base de I_t est constituée des polynômes d'Hermite d'ordre t . Par construction le processus $X(t)$ est non corrélé et centré, donc $X(t)$ est régulier linéaire. Or $\forall t$, $\mathcal{F}_t(X) = \mathcal{F}(\epsilon)$ donc $\mathcal{F}_{-\infty}(X) = \mathcal{F}(\epsilon) = \mathcal{F}$: la σ -algèbre associée au processus $X(t)$ est constante donc $X(t)$ est aussi singulier fort.

Classe (2) : Soit ϵ la variable définie comme précédemment, φ une variable aléatoire indépendante de ϵ uniformément répartie sur $[0, 2\pi]$, et w un réel fixé ($\neq 0, \text{ mod. } \pi$). On considère le processus $X(t)$ non stationnaire strict défini par :

$$\begin{cases} X(t) = \epsilon \sin(wt + \varphi) & t \neq 1 \\ X(1) = (\epsilon^2 - 1) \sin(w + \varphi) \end{cases} .$$

Comme $\forall t$, $\mathcal{F}_t(X) = \mathcal{F}(\epsilon)$ ce processus est singulier fort, mais il n'est pas régulier linéaire car $H_t(\epsilon) \neq H$, à cause de la présence de ϵ^2 dans l'équation de $X(t)$ pour $t = 1$.

On peut aussi avoir $H_t \neq H_{t+1}$ sans que $\mathcal{F}_t(\epsilon)$ soit strictement contenu dans $\mathcal{F}_{t+1}(\epsilon)$.

Considérons l'exemple suivant : Soit le processus $\epsilon(t)$ tel que :

$$\begin{cases} \epsilon(t) \text{ suit une loi normale centrée réduite si } t \leq 0 \\ \epsilon(t) = \frac{1}{\sqrt{2}} (\epsilon^2(t) - 1) & t > 0 \end{cases}$$

Ce processus est stationnaire au second ordre, mais il n'est pas stationnaire en loi. C'est un processus non corrélé car $\epsilon(t)$ est un processus non corrélé, mais $\forall t > 0, \mathcal{F}_t(\epsilon) = \mathcal{F}_0(\epsilon)$.

Ainsi donc, si nous considérons des processus $X(t), t \in \mathbb{Z}$, de $L^2(\mathcal{F}_t(\epsilon))$, stationnaires, ils peuvent être stationnaires au second ordre ou, stationnaires au second ordre et non stationnaires en loi. Parallèlement, ils peuvent être réguliers ou singuliers. Mais notons qu'un processus non singulier n'est pas équivalent à un processus régulier, cette notion étant plus faible. En effet la régularité implique la non singularité, mais la non singularité n'implique pas la régularité.

IV - POLYSPECTRE

Tout processus $X(t), t \in \mathbb{Z}$, stationnaire au second ordre, admet une représentation spectrale. Ce résultat est dû à l'origine à H. CRAMER (cf. H. CRAMER et M. LEADBETTER, chapitre 7, 1955). On peut trouver des approches différentes de ce résultat dans la littérature qui sont équivalentes (cf. J.L. DOOB, chapitre X, 1953). La représentation spectrale d'un processus $X(t) \in L^2(\mathcal{F}(\epsilon))$ est donnée par

$$(3.4.1) \quad X(t) = \int_{-\pi}^{\pi} e^{i\lambda t} \varphi(d\lambda) X(0) = \int_{-\pi}^{\pi} e^{i\lambda t} \tau(d\lambda)$$

où $\tau(d\lambda) = \varphi(d\lambda) X(0)$ est une mesure σ -additive sur $[-\pi, +\pi]$ à valeurs dans $L^2(\mathcal{F}(\epsilon))$ telle que :

$$E|\tau(\Delta)|^2 \leq E|X(0)|^2 \text{ pour tout } \Delta \text{ borélien de } [-\pi, +\pi],$$

$$\text{et } E[\tau(\Delta)\tau(\Delta')] = 0 \text{ pour tous boréliens disjoints de } [-\pi, +\pi].$$

Si l'on pose $\mathcal{F}(\Delta) = E|\tau(\Delta)|^2$, pour tout Δ borélien de $[-\pi, +\pi]$, alors $\mathcal{F}(\Delta)$ est appelée la mesure spectrale du processus sur $L^2(\mathcal{F}_t(\epsilon))$. $\mathcal{F}(\Delta)$ permet de calculer les covariances du processus $X(t) : c(s) = E[X(t)X(t+s)], \forall s \geq 0$ qui ne dépendent

pas de t , $c(s)$ est le $s^{\text{ième}}$ coefficient de Fourier de $\mathcal{F}(\cdot)$, et l'on a la relation suivante :

$$c(s) = \int_{-\pi}^{\pi} e^{i\lambda t} F(d\lambda) .$$

Mais $\mathcal{F}(\Delta)$ ne donne qu'une approche au second ordre du processus $X(t)$, $t \in \mathbb{Z}$, et ne suffit pas à caractériser les processus non linéaires. D'où la nécessité d'introduire le polyspectre, étudié pour la première fois par A. BLANC LAPIERRE et R. FORTET (1953).

A partir de maintenant on suppose que le processus $X(t)$, $t \in \mathbb{Z}$ est stationnaire en loi.

Définition 3.4.1.

On appelle fonction des moments d'ordre k :

$$\begin{aligned} M(t_1, \dots, t_{k-1}) &= E(X(t)X(t_1)\dots X(t_{k-1})) \\ &= E(X(t)X(t+t_1)\dots X(t+t_{k-1})). \end{aligned}$$

Cette fonction ne dépend pas de t à cause de la stationnarité, et est moins utile que la fonction de cumulants, car ces derniers fournissent une mesure de la dépendance entre les observations du processus et sont donc plus facilement interprétables. De plus les fonctions de cumulants généralement tendent vers 0 à l'infini contrairement aux fonctions de moments.

Définition 3.4.2.

Le cumulants joint de k variables aléatoires $X(1), \dots, X(k)$, éventuellement identiques est le coefficient de (t_1, \dots, t_k) dans le développement de $\log \varphi(t_1, \dots, t_k)$, $\varphi(t_1, \dots, t_k)$ étant la fonction caractéristique du vecteur $(X(1), \dots, X(k))$.

La fonction de cumulants $C(t_1, t_2, \dots, t_k)$ est définie comme le cumulants de $(X(t), X(t+t_1), \dots, X(t+t_{k-1}))$, ce dernier est indépendant de t par stationnarité. La fonction de cumulants présente dans certains cas des relations de symétrie que nous expliciterons par la suite (cf. D.R. BRILLINGER et H. ROSENBLATT, 1966 b). La transformée de Fourier du cumulants d'ordre k est appelée polyspectre.

Définition 3.4.3.

La densité polyspectrale de cumulants d'ordre k est définie par :

$$(3.4.2) \quad f(\lambda_1, \dots, \lambda_{k-1}) = \left(\frac{1}{2\pi}\right)^{k-1} \sum_{t_1=-\infty}^{+\infty} \dots \sum_{t_{k-1}=-\infty}^{+\infty} C(t_1, \dots, t_{k-1}) \cdot e^{-i(\lambda_1 t_1 + \dots + \lambda_{k-1} t_{k-1})}$$

si la série est convergente.

Une condition suffisante pour l'existence de (3.4.2) est que les cumulants $C(t_1, \dots, t_{k-1})$ soient absolument sommables, i.e.

$$\sum_{t_1=-\infty}^{+\infty} \dots \sum_{t_{k-1}=-\infty}^{+\infty} |C(t_1, \dots, t_{k-1})| < \infty .$$

On a alors la relation réciproque :

$$(3.4.3) \quad C(t_1, \dots, t_{k-1}) = \int e^{i\lambda_1 t_1 + \dots + i\lambda_{k-1} t_{k-1}} \cdot f(\lambda_1, \dots, \lambda_{k-1}) d\lambda_1, \dots, d\lambda_{k-1}$$

On peut alors généraliser la notion de mesure spectrale d'ordre deux à l'ordre supérieur, cf. M. ROSENBLATT, 1985. Par commodité nous donnons la présentation suivante :

Définition 3.4.4.

Soit $C(t_1, \dots, t_{k-1})$ une fonction de cumulant d'ordre k .
Supposons qu'il existe une mesure sur $[-\pi, +\pi]^{k-1}$ telle que

$$C(t_1, \dots, t_{k-1}) = \int e^{i\lambda_1 t_1 + \dots + i\lambda_{k-1} t_{k-1}} dF(\lambda_1, \dots, \lambda_{k-1}) .$$

Alors on appelle F , la mesure spectrale d'ordre k .

Quel lien existe-t-il entre la mesure aléatoire $\tau(d\lambda)$ qui intervient dans la représentation spectrale du processus $X(t)$ et la mesure spectrale F ? On a vu que dans le cas des processus au second ordre, il existe une relation entre $d\tau$ et la mesure de densité spectrale de covariance, nous allons généraliser ce résultat

Proposition 3.4.1.

On suppose que la fonction de cumulant est absolument intégrable.

Soit $\hat{\mathcal{S}}$ l'espace vectoriel des fonctions indéfiniment différentiables définies sur $[-\pi, +\pi]$ à valeurs complexes.

Alors:

$$\text{cum} \left\{ \int \hat{\varphi}_1(\lambda) d\tau(\lambda), \dots, \int \hat{\varphi}_k(\lambda) d\tau(\lambda) \right\} =$$

$$\int \hat{\varphi}_1(\lambda_1) \dots \hat{\varphi}_{k-1}(\lambda_{k-1}) \hat{\varphi}_k \left(-\sum_{i=1}^{k-1} \lambda_i \right) f(\lambda_1, \dots, \lambda_{k-1}) d\lambda_1 \dots d\lambda_{k-1} .$$

$$\forall \hat{\varphi}_1, \dots, \forall \hat{\varphi}_k \in \hat{\mathcal{S}} .$$

Pour établir ce résultat, on définit une forme multilinéaire de cumulants, qui opère sur l'espace D des fonctions à valeurs sur \mathbb{Z} , nulles sauf en un nombre fini de points, par :

$$\forall \varphi_1, \forall \varphi_2, \dots, \forall \varphi_k \in D :$$

$$K(\varphi_1, \varphi_2, \dots, \varphi_k) = \text{cum} \left\{ \sum_t \varphi_1(t) X(t), \dots, \sum_t \varphi_k(t) X(t) \right\} .$$

Cette forme multilinéaire K peut-être prolongée par continuité en une forme linéaire continue sur l'espace \mathcal{S} des suites à décroissance rapide. Ce dernier espace étant défini par :

$$\mathcal{S} = \{ \varphi : |\varphi|_{n_1, \dots, n_k} = \text{Sup} |t_1^{n_1} \dots t_k^{n_k} \varphi(t_1, \dots, t_k)| < \infty, \forall n_1, \dots, n_k \}$$

muni de la famille des semi-normes $|\varphi|_{n_1, \dots, n_k}$.

(K est une distribution tempérée, cf. L. SCHWARTZ, 1965).

On peut alors définir la transformation de Fourier de K par la formule :

$$\langle \mathcal{F}K, \hat{\varphi} \rangle = \langle K, \varphi \rangle$$

où φ est la transformée de Fourier ordinaire de la fonction $\hat{\varphi}$. Cette transformée de Fourier est une isométrie entre \mathcal{S} et $\hat{\mathcal{S}}$.

On a :

$$\hat{\varphi}(-\lambda_2, \dots, -\lambda_k, \lambda_2, \dots, \lambda_k) = \sum_{t_1, \dots, t_k} e^{i[\lambda_2(t_2 - t_1) + \dots + \lambda_k(t_k - t_1)]} \varphi(t_1, \dots, t_k)$$

(3.4.4)

$$= \sum_{u_1, \dots, u_{k-1}} e^{i(u_1 \lambda_2 + \dots + u_{k-1} \lambda_k)} \sum_t \varphi(t + u_1, \dots, t + u_{k-1})$$

Par suite :

$$\langle \mathcal{F}K, \hat{\varphi} \rangle = \langle K, \varphi \rangle = \frac{1}{(2\pi)^k} \sum_{t_i} C(t_2 - t_1, \dots, t_k - t_1) \varphi(t_1, \dots, t_k)$$

(3.4.5)

$$= \frac{1}{(2\pi)^{k-1}} \sum_{u_i} C(u_1, \dots, u_{k-1}) \frac{1}{2\pi} \sum_t \varphi(t + u_1, \dots, t + u_{k-1}).$$

En se servant de (3.4.3) on obtient :

$$\langle K, \varphi \rangle = \int f(\lambda_2, \dots, \lambda_k) \sum_{u_i} \sum_t e^{i(u_1 \lambda_2 + \dots + u_{k-1} \lambda_k)} \varphi(t+u_1, \dots, t+u_{k-1}) d\lambda_2 \dots d\lambda_k.$$

Alors en utilisant (3.4.4), on a :

$$\langle \mathbb{F}K, \hat{\varphi} \rangle = \int f(\lambda_2, \dots, \lambda_k) \hat{\varphi}(-\lambda_2 - \dots - \lambda_k, \lambda_2, \dots, \lambda_k) d\lambda_2, \dots, d\lambda_k.$$

En prenant pour $\hat{\varphi}$: $\hat{\varphi}(\lambda_1, \dots, \lambda_k) = \hat{\varphi}_1(\lambda_1) \dots \hat{\varphi}_k(\lambda_k)$, on obtient le résultat annoncé.

Le résultat précédent va permettre de trouver très vite des relations de symétrie qui ne sont pas toujours évidentes à trouver à partir de la mesure spectrale telle qu'elle est habituellement définie. Ainsi :

$$\int \hat{\varphi}_1(\lambda_1) \dots \hat{\varphi}_{k-1}(\lambda_{k-1}) \hat{\varphi}_k(-\sum_{i=1}^{k-1} \lambda_i) f(\lambda_1, \dots, \lambda_{k-1}) d\lambda_1 \dots d\lambda_{k-1} =$$

$$\int \hat{\varphi}_k(\lambda_1) \dots \hat{\varphi}_{k-1}(\lambda_{k-1}) \hat{\varphi}_1(-\sum_{i=1}^{k-1} \lambda_i) f(\lambda_1, \dots, \lambda_{k-1}) d\lambda_1 \dots d\lambda_{k-1}.$$

Ce qui donne la relation suivante :

$$f(\lambda_1, \dots, \lambda_{k-1}) = f(\lambda_1, \dots, \lambda_{j-1}, -\sum_{i \neq j} \lambda_i, \lambda_{j+1}, \dots, \lambda_k).$$

C'est une fonction qui est symétrique par rapport à ses arguments.

Pour une étude des cumulants, on peut consulter aussi V.P. LEONOV et A.M. SHIRYAEV (1959), A.M. SHIRYAEV (1960), D.R. BRILLINGER (1975), D.R. BRILLINGER et M. ROSENBLATT (1966 b) et M. ROSENBLATT (1980, 1983). Notons que dans la littérature la notion de cumulants intervient souvent sous la dénomination de semi-invariants. L'estimation des polyspectres et l'étude des

propriétés asymptotiques des estimateurs a fait l'objet de nombreux travaux, on peut voir par exemple, D.R. BRILLINGER (1965), M. ROSENBLATT et J.W. VAN NESS (1965), D.R. BRILLINGER et M. ROSENBLATT (1966 a), J.W. VAN NESS (1966).

Remarque

Nous venons de voir que le polyspectre d'ordre k permettait d'obtenir une représentation spectrale d'ordre k des modèles de $L^2(\mathcal{F}(\epsilon))$, il va permettre aussi de les analyser. Soit $X(t)$, $t \in \mathbb{Z}$ un processus stationnaire linéaire tels que tous ses moments d'ordre k existent. Soit $f_k(\cdot)$ la densité spectrale de cumulant d'ordre k , et f sa densité spectrale, alors $\forall \lambda_1, \lambda_2, \dots, \lambda_k$, le rapport :

$$|f_k(\lambda_1, \dots, \lambda_k)|^2 \{f(\lambda_1)f(\lambda_2)\dots f(\lambda_k)f(\lambda_1+\lambda_2+\dots+\lambda_k)\}^{-1}$$

est constant. Ce résultat va permettre d'établir des tests permettant de différencier les modèles linéaires des modèles non linéaires (cf. T. SUBBA RAO et M.M. GABR, 1980, M.J. HINNICH, 1982), en particulier à l'aide de la densité bispectrale, densité spectrale de cumulant d'ordre 2. Il n'y a d'ailleurs aucune raison de s'arrêter à la densité spectrale d'ordre deux, si ce n'est pour des raisons pratiques. En effet si nous considérons par exemple les deux modèles non linéaires suivants appartenant à $L^2(\mathcal{F}_t(\epsilon))$:

$$X(t) = \epsilon(t) + \alpha \epsilon(t-1)\epsilon(t-2)$$

$$\text{et } Y(t) = \epsilon(t) + \beta X(t-2)\epsilon(t-1)$$

leurs moments d'ordre trois sont nuls et donc cet ordre n'est pas suffisant pour caractériser et identifier ces modèles, il est alors nécessaire de passer à l'ordre supérieur, cf. D. GUEGAN (1984 a), voir aussi J. PEMBERTON et H. TONG (1981).

Une interprétation des processus réversibles dans le temps peut être donnée à l'aide des cumulants. On sait que les processus

linéaires stationnaires au second ordre sont réversibles dans le temps, (essentiellement dans le cas des processus gaussiens d'ailleurs), les modèles non linéaires le seraient s'ils conservaient leur structure de probabilité quand on inverse le temps, c'est à dire si la structure de probabilité de $X(t)$ était la même que celle de $X(-t)$. Si cela est le cas, on a alors :

$$(3.4.6) \quad c(t_1, t_2, \dots, t_k) = c(-t_1, -t_2, \dots, -t_k)$$

et la partie imaginaire de la densité spectrale de cumulant d'ordre k , $f_k(\lambda_1, \lambda_2, \dots, \lambda_k)$ est identiquement nulle. Inversement si la partie imaginaire de tous les spectres d'ordre supérieur est identiquement nulle alors (3.4.6) est vérifiée. On a alors le résultat suivant :

"Si un processus $X(t)$, $t \in \mathbb{Z}$ est déterminé par tous ses moments et si la relation (3.4.6) est vérifiée, alors il est inversible dans le temps".

Cette propriété peut être entre autre très utile dans l'étude des processus markoviens.

V - ESPACES D'INNOVATIONS DE PROCESSUS

On ne dispose pas dans le cas des modèles non linéaires réguliers de décomposition semblable à la décomposition de Wold des processus du second ordre réguliers, ou des processus gaussiens cf. T. HIDA, 1980.

On va donner ici un résultat propre à certaines classes de modèles bilinéaires, cf. D. GUEGAN, 1982 b, 1983 : Etant donné un processus stationnaire $X(t)$, $t \in \mathbb{Z}$, $\mathcal{F}_t(\epsilon)$ -mesurable, il lui correspond une décomposition en moyenne mobile infinie d'éléments ne dépendant que des puissances d'un seul ϵ . Pour cela gardons les notations du paragraphe II.

On considère la projection orthogonale :

$$E(\cdot / \mathcal{F}_t(\epsilon)) : L^2(\mathcal{F}_{t+1}(\epsilon)) \longrightarrow L^2(\mathcal{F}_t(\epsilon)) .$$

et on définit l'espace d'innovation \mathcal{N}_{t+1} comme son noyau

$$\mathcal{N}_{t+1} = L^2(\mathcal{F}_{t+1}(\epsilon)) \ominus L^2(\mathcal{F}_t(\epsilon)) .$$

Une caractérisation des éléments de \mathcal{N}_{t+1} est la suivante :

Théorème 3.5.1. (D.GUEGAN, 1983, p. 251)

$\forall X \in \mathcal{N}_{t+1}$, on a les deux représentations suivantes équivalentes.

$$(3.5.1) \quad X = \sum_{k=1}^{\infty} Q_k(\epsilon(t+1)^k - m_k) \quad \text{où } Q_k \in L^2(\mathcal{F}_t(\epsilon))$$

et où m_k est le moment d'ordre k de $\epsilon(t)$.

$$(3.5.2) \quad X = \varphi\left(\sum_{k=1}^{\infty} R_k \otimes \epsilon(t+1)^{\otimes k}\right) \quad \text{où } R_k \in L^2(\mathcal{F}_t(\epsilon))$$

où φ est l'application définie au paragraphe II.

Preuve

Si $X \in L^2(\mathcal{F}_{t+1}(\epsilon))$, alors $X = \sum_{k=0}^{\infty} P_k(\epsilon(t+1))$ où P_k appartient à l'espace dénombrable \mathcal{P}_{t+1} des polynômes en $\epsilon(t+1), \epsilon(t), \dots$ qui est dense dans $L^2(\mathcal{F}_{t+1}(\epsilon))$, donc $P_k(\epsilon(t+1)) = Q_k(\epsilon(t+1)^k - m_k) \quad \forall k$, où Q_k est $\mathcal{F}_t(\epsilon)$ -mesurable et $m_k = E[\epsilon^k(t+1) | \mathcal{F}_t(\epsilon)]$, d'où (3.5.1).

D'autre part, considérons le développement de X sur les composantes du chaos, alors :

$$\begin{aligned}
X &= \varphi \left(\sum_{k=0}^{\infty} \sum_{t_1 < t} \dots \sum_{t_k < t} \alpha_{n_1 \dots n_k, t_1 \dots t_k} \epsilon^{\otimes n_1}(t_1) \otimes \dots \otimes \epsilon^{\otimes n_k}(t_k) \right) \\
&= \varphi \left(\sum_{k=0}^{\infty} \sum_{t_1 < t} \dots \sum_{t_k < t} \alpha_{m_1 \dots m_k, t_1 \dots t_k} \epsilon^{\otimes m_1}(t_1) \otimes \dots \otimes \epsilon^{\otimes m_k}(t_k) \otimes \epsilon^{\otimes k}(t+1) \right) \\
&= \varphi \left(\sum_{k=0}^{\infty} R_k(\epsilon(t)) \otimes \epsilon^{\otimes k}(t+1) \right)
\end{aligned}$$

où $R_k(\cdot) \in L^2(\mathcal{F}_t(\epsilon))$ et $E[\epsilon(t)^{\otimes n}] = 0$ pour $n > 0$ d'où l'expression (3.5.2).

φ étant un isomorphisme, on a l'équivalence des deux définitions. \square

Considérons l'opérateur de translation \mathcal{B} défini sur $L^2(\mathcal{F}(\epsilon))$, de la manière suivante :

Définition 3.5.1.

1. $\forall U \in L^2(\mathcal{F}(\epsilon))$ tel que :

$$U = \sum_{k=0}^{\infty} \sum_{t_1 < t} \dots \sum_{t_k < t} \alpha_{k, t_1, \dots, t_k} \epsilon(t_1) \otimes \dots \otimes \epsilon(t_k)$$

L'opérateur \mathcal{B} défini de $L^2(\mathcal{F}_t(\epsilon))$ dans $L^2(\mathcal{F}_{t-1}(\epsilon))$ est tel que :

$$\mathcal{B}U = \sum_{k=0}^{\infty} \sum_{t_1 < t} \dots \sum_{t_k < t} \alpha_{k, t_1, \dots, t_k} \epsilon(t_1-1) \otimes \dots \otimes \epsilon(t_k-1)$$

2. $\forall X \in L^2(\mathcal{F}(\epsilon))$ tel que :

$$X = \lim_{L^2} h_n(\epsilon(t_1), \dots, \epsilon(t_k))$$

(limite de polynômes de degré n en $\epsilon(t_1), \dots, \epsilon(t_k)$),
 l'opérateur \mathcal{B} , en tant qu'opérateur de $L^2(\mathcal{F}(\epsilon))$ dans $L^2(\mathcal{F}(\epsilon))$
 est tel que :

$$\mathcal{B}X = \lim_{L^2} h_n(\epsilon(t_1-1), \dots, \epsilon(t_k-1)).$$

On peut remarquer que ces deux définitions sont équivalentes. On a alors la proposition suivante :

Proposition 3.5.1.

1. $\forall s, \forall t \quad \mathcal{B}^{-s} \mathcal{N}_t = \mathcal{N}_{t+s}$

2. $L^2(\mathcal{F}_0(\epsilon)) = \bigoplus_{t \leq 0} \mathcal{N}_t$

3. Pour tout processus strictement stationnaire $X(t)$ défini dans $L^2(\mathcal{F}_t(\epsilon))$, il existe une décomposition en moyenne mobile infinie sur les éléments de \mathcal{N}_t :

$$X(t) = \sum_{s \leq 0} \sum_{i,k} \lambda_{iks} \zeta_{ik}(t-s)$$

où ζ_{ik} est une base de \mathcal{N}_0 , $\lambda_{iks} \in \mathbb{R}$.

Preuve

La condition (1) découle de la stationnarité .

La propriété (2) provient de la définition de l'opérateur de projection $E(\cdot | \mathcal{F}_t(\epsilon))$.

Pour obtenir la propriété (3), on note :

$\zeta_{ik} = u_i \epsilon^k(0)$ les éléments de la base de \mathcal{N}_0 ,
 alors $\mathcal{B}^{-s} \zeta_{ik} = \zeta_{ik}(s)$ est une base de \mathcal{N}_s . Soit

$$X(0) \in L^2(\mathcal{F}_0(\epsilon)), \quad X(0) = \sum_{s \leq 0} \omega_s \quad \text{avec } \omega_s \in \mathcal{N}_s, \quad \text{et}$$

$$\omega_s = \sum_{i,k} \lambda_{iks} \zeta_{ik}(s) , \lambda_{iks} \in \mathbb{R} .$$

Or $X(t) = \beta^{-t} X(0)$, donc

$$X(t) = \sum_{s \leq 0} \sum_{i,k} \lambda_{iks} \zeta_{ik}(t-s) . \quad \square$$

Il est important de remarquer que l'espace \mathcal{N}_t introduit ici joue le même rôle que l'espace d'innovation dans le cas linéaire, mais que malheureusement il est de dimension infinie. Les éléments de \mathcal{N}_t sont cependant non corrélés avec des variables aléatoires ne dépendant que des observations passées du processus, comme dans le cas linéaire.

CHAPITRE IV

ETUDE PROBABILISTE DU MODELE BILINEAIRE

Revue d'ensemble des premiers travaux

I - INTRODUCTION

Dans ce chapitre on passe en revue les premières études sur la structure probabiliste (moments notamment) du modèle bilinéaire d'ordre (p, q, P, Q) , noté $BL(p, q, P, Q)$ défini de la manière suivante : soit $X(t)$ un processus stochastique défini sur un certain espace de probabilité (Ω, \mathcal{F}, P) , pour tout $t \in \mathbb{Z}$ par la relation suivante :

$$(4.1.1) \quad X(t) = \sum_{i=1}^p a_i X(t-i) + \epsilon(t) + \sum_{j=1}^q c_j \epsilon(t-j) + \sum_{j=1}^Q \sum_{k=1}^P b_{jk} X(t-k) \epsilon(t-j)$$

où $(a_i, 1 \leq i \leq p)$, $(c_j, 1 \leq j \leq q)$, $(b_{jk}, 1 \leq j \leq Q, 1 \leq k \leq P)$ sont des constantes et où $(\epsilon(t), t \in \mathbb{Z})$ est une suite de variables aléatoires indépendantes équidistribuées, gaussiennes, centrées, de variance finie σ^2 , définies aussi sur l'espace de probabilité (Ω, \mathcal{F}, P) .

Les résultats obtenus à partir de la représentation markovienne du processus bilinéaire (4.1.1) seront abordés dans le chapitre suivant.

On donne tout d'abord la décomposition formelle des modèles bilinéaires sur les chaos de Wiener. C'est en fait la voie

d'attaque naturelle pour l'étude de ces modèles, malheureusement ce développement qui permet d'obtenir la décomposition du processus sur chacune des composantes du chaos n'est que formel car pour l'instant on ne sait pas obtenir la convergence des puissances tensorielles qui apparaissent dans la décomposition. Dans l'expression (3.2.1) que nous avons obtenue d'un processus stationnaire appartenant à $L^2(\mathcal{F}(\epsilon))$, nous avons utilisé une décomposition en fonction des polynômes d'Hermite, de $\epsilon(t_1), \dots, \epsilon(t_k)$. Or une telle décomposition peut aussi se calculer à partir des polynômes ordinaires plus faciles à obtenir dans la pratique, ce qui conduit alors à la représentation de Volterra. C'est cette seconde démarche que nous utilisons ici : nous donnons tout d'abord un résumé d'une partie de l'article de D. GUEGAN (1987) qui fournit la décomposition du modèle bilinéaire général sur les composantes du chaos, puis nous développons quelques cas particuliers pour lesquels nous avons des résultats de convergence.

Dans un deuxième temps, nous rappelons des résultats concernant uniquement la structure probabiliste du modèle à partir de la représentation (4.1.1). Les principaux thèmes qui ont retenu l'attention des chercheurs sont l'existence et l'unicité d'une solution stationnaire, l'inversibilité, l'ergodicité, la fonction de covariance, l'existence des moments d'ordre supérieur et les méthodes de calcul appropriées. A cause de la complexité structurale du modèle, peu de résultats sont généraux. On peut distinguer trois classes de modèles bilinéaires : les modèles bilinéaires diagonaux, les modèles bilinéaires superdiagonaux et les modèles bilinéaires sousdiagonaux. La plupart des résultats dont on dispose actuellement ne concernent en fait qu'une ou deux de ces sous classes. Par exemple si l'existence d'une solution stationnaire unique est un problème résolu pour certains modèles superdiagonaux, les résultats sont parcellaires pour l'ensemble des modèles superdiagonaux et diagonaux, et l'on ne dispose à l'heure actuelle d'aucun résultat sur ce problème pour les modèles sousdiagonaux. Les principaux résultats présentés sur ce sujet sont

dus aux chercheurs suivants : T. SUBBA RAO (1981), M. BHASKARA RAO, T. SUBBA RAO et A.M. WALKER (1983), T. SUBBA RAO et M.M. GABR (1984), S.I. AKAMANAM, M. BHASKARA RAO et S. SUBRAMANYAM (1986). T. PHAM DINH et L. TAT TRAN (1981) ont fait une étude complète du modèle diagonal d'ordre 1. Nous avons pour notre part contribué à l'étude du modèle superdiagonal d'ordre 1, (cf. D. GUEGAN 1981 a), ainsi qu'à celle d'autres modèles simples dans divers documents, (cf. D. GUEGAN, 1981 b, 1984 a).

II - DECOMPOSITION FORMELLE DE PROCESSUS SUR LES CHAOS DE WIENER

1. Décomposition de Volterra du modèle bilinéaire général

Pour obtenir la décomposition formelle du processus bilinéaire sur les composantes du chaos \mathcal{H} , nous allons réécrire les équations (4.1.1) sous forme matricielle. Sans perdre de généralité, nous posons pour simplifier les calculs $p = \max(p, q, P, Q)$, et nous ajoutons autant de zéros que nécessaires dans les matrices que nous définissons. L'équation (4.1.1) se réécrit de la manière suivante :

$$(4.2.1) \quad \begin{cases} \tilde{X}(t) = A\tilde{X}(t-1) + C\epsilon(t) + \sum_{j=1}^p C_j \epsilon(t-j) + \sum_{j=1}^p B_j \tilde{X}(t-1) \epsilon(t-j) \\ X(t) = H\tilde{X}(t) \end{cases}$$

où $\tilde{X}(t) = [X(t), \dots, X(t-p)]$

$C_j = [c_j \ 0 \ \dots \ 0]$ $0 \leq j \leq p$, $C_0 = C = I$ où

I est la matrice identité

$H = [1, 0, \dots, 0]'$

ne peut pas en obtenir de conditions simples d'existence. Dans certains cas, on obtient des conditions suffisantes permettant de se ramener à (4.2.2.), et dans des cas particuliers on sait démontrer rigoureusement que la représentation (4.2.2) est la bonne.

2. Exemples

Nous allons donner la décomposition formelle de quelques modèles bilinéaires simples qui furent les premiers étudiés dans la littérature.

Notons tout d'abord que les modèles linéaires appartiennent à la première composante du chaos $H_t^{\otimes 1}(\epsilon)$, leur décomposition est donnée par la décomposition de Wold.

D'autre part les modèles non linéaires du type moyenne mobile, composés uniquement de termes polynomiaux en $\epsilon(t), \epsilon(t-1), \dots$, de degré fini d , sont déjà sous forme de décomposition de Volterra, ils appartiennent à un nombre fini de composantes du chaos

$$H_t^{\otimes 1}(\epsilon) + H_t^{\otimes 2}(\epsilon) + \dots + H_t^{\otimes d}(\epsilon) \quad \forall d$$

Un modèle bilinéaire même simple qui comporte un seul terme de la forme $b_{jk}X(t-k)\epsilon(t-j)$ ne peut plus être développé sur un nombre fini de composantes du chaos, mais sur le chaos tout entier $L^2(\mathcal{F}(\epsilon))$. Par exemple, en appliquant la formule (4.2.2), on peut obtenir facilement le développement de Volterra de modèles simples.

Pour le modèle diagonal d'ordre 1

$$(4.2.3) \quad X(t) = b_{11}X(t-1)\epsilon(t-1) + \epsilon(t) \\ \text{ssi } 0 < |b_{11}| < 1, \quad X(0) = 1, \quad t \in \mathbb{N}.$$

On a alors la représentation suivante :

$$X(t) = \epsilon(t) + \sum_{n \geq 1} b_{11}^n \sum_{j=1}^{n-1} \epsilon(t-j) [\epsilon(t-j) + \epsilon(t-j+1)] .$$

Pour le modèle superdiagonal d'ordre 1

$$(4.2.4) \quad X(t) = \epsilon(t) + b_{21} X(t-2) \epsilon(t-1)$$

ssi $0 < |b_{21}| < 1$, $X(0) = 1$, $t \in \mathbb{N}$.

On a alors la représentation :

$$X(t) = \epsilon(t) + \sum_{n \geq 1} b_{21}^n \epsilon(t-2n) \prod_{j=1}^n \epsilon(t-2j+1) .$$

Pour le modèle sous diagonal d'ordre 1 :

$$(4.2.5) \quad X(t) = \epsilon(t) + b_{12} X(t-1) \epsilon(t-2)$$

ssi $0 < |b_{12}| < 1$, $X(0) = 1$, $t \in \mathbb{N}$.

On a alors la représentation :

$$X(t) = \epsilon(t) + b_{12} \epsilon(t-2) \epsilon(t-1) + \sum_{n \geq 2} b_{12}^n \epsilon(t-n-1) \epsilon(t-n)^2 \prod_{j=1}^{n-2} \epsilon(t-j-1)$$

Dans ces cas particuliers, nous avons donné des conditions nécessaires et suffisantes d'existence de ces décompositions qui portent sur les coefficients b_{11} , b_{12} et b_{21} . Mais dès que le nombre de termes dans le modèle augmente, on obtient très vite des équations sur les coefficients du modèle qui ne conduisent pas à des conditions pouvant s'utiliser facilement ou tout simplement se généraliser.

Pour ces modèles on retrouve les conditions étudiées par T. PHAM DINH et L. TAT TRAN (1981) pour le modèle (4.2.3), par D. GUEGAN (1981 a et b) pour les modèles (4.2.4) et (4.2.5) : conditions qui apparaissent maintenant comme des cas particuliers de la condition donnée par T. PHAM DINH (1985) quand on passe à la représentation markovienne.

3. Application aux transformations non linéaires instantanées de processus linéaires

Si l'on part d'un processus linéaire et que l'on fait une transformation non linéaire de ce processus, on va tomber sur un processus non linéaire, (cf. C.W.J. GRANGER et A.P. ANDERSEN, 1978 c). Notons que la décomposition de Volterra peut être appliquée à un processus non gaussien. Dans le cas d'un processus non linéaire obtenu par une transformation à partir d'un modèle linéaire, on peut obtenir cette décomposition dans le cas particulier où on utilise une transformation puissance. D'ailleurs une transformation quelconque peut toujours être approchée par une transformation polynôme qui est une combinaison linéaire de fonctions puissance. Considérons l'exemple suivant, (cf. D. GUEGAN 1982 b, 1983). Soit $X(t)$ un processus strictement stationnaire régulier appartenant à $H_t^{\otimes 1}(\epsilon)$, alors le processus $U(t) = (X(t))^n$ appartient à $H_t^{\otimes n}(\epsilon)$, à cause de la décomposition de Wold, et

$$U(t) = \sum_{t_1 < t} \dots \sum_{t_n < t} \alpha_{n, t_1, \dots, t_n} \epsilon(t_1) \otimes \dots \otimes \epsilon(t_n) .$$

Parmi les processus $X(t)$, $t \in \mathbb{Z}$ précédents, considérons le processus autorégressif d'ordre 1 :

$$X(t) = \alpha X(t-1) + \epsilon(t) \text{ avec } |\alpha| < 1 .$$

Nous nous intéressons aux propriétés de la transformation carré du processus $X(t)$:

$$U(t) = (X(t))^2$$

D'après ce qui précède, $U(t)$ est un processus strictement stationnaire, qui appartient au deuxième chaos. En tant que processus au second ordre $U(t)$ est un processus autorégressif, mais le bruit n'est plus indépendant. On a la représentation suivante :

$$U(t) = \alpha^2 U(t-1) + \eta(t)$$

où $\eta(t)$ est un processus stationnaire qui appartient à $H_t^{\otimes 2}(\epsilon)$:

$$\eta(t) = 2 \sum_{j=1}^{\infty} \alpha^j \epsilon(t-j) \epsilon(t) + \epsilon(t)^2,$$

D'autre part : $E(\eta(t)) = 1$, $\text{var}(\eta(t)) = \frac{2(1+\alpha^2)}{1-\alpha^2}$ et

$$\text{cov}(\eta(t), \eta(t+k)) = 0 \quad \forall k \geq 1.$$

Le processus $U(t)$ a alors pour décomposition de Wold :

$$U(t) = \sum_{k=0}^{\infty} \alpha^{2k} \eta(t-k).$$

$\eta(t)$ est le processus d'innovation du processus $U(t)$ et

$$\sum_{k=0}^{\infty} \alpha^{2k} < \infty.$$

III - THEOREMES D'EXISTENCE POUR DES MODELES BILINEAIRES SUPERDIAGONAUX

Si l'on considère l'expression (4.1.1) définissant le modèle bilinéaire d'ordre (p, q, P, Q) , on constate que le premier terme du membre de droite peut être considéré comme la partie autorégressive du processus $X(t)$, $t \in \mathbb{Z}$, le troisième terme comme sa partie moyenne mobile, et le quatrième terme comme la contribution purement bilinéaire du processus $X(t)$, $t \in \mathbb{Z}$. En fonction du quatrième terme dans le membre de droite de l'expression (4.1.1), on peut alors différentier trois sous classes de modèles bilinéaires :

- Les modèles bilinéaires diagonaux pour lesquels
- (H₀) $b_{jk} = 0$ pour tout $k \neq j$.
- Les modèles bilinéaires superdiagonaux pour lesquels :
- (H) $b_{jk} = 0$ si $k < j$
- Les modèles bilinéaires sous diagonaux pour lesquels
- (H₁) $b_{jk} = 0$ si $k > j$.

Dans ce paragraphe, nous allons donner des conditions d'existence et de stationnarité des modèles superdiagonaux. La

classe des modèles sousdiagonaux (excluant les modèles diagonaux) n'est pas abordée dans ce chapitre car sa structure est trop complexe et l'approche utilisée ici ne permet pas d'obtenir des résultats satisfaisants.

1.- Expression vectorielle des modèles bilinéaires superdiagonaux

Posons $r = \max(p, P)$
 $s = \max(P, Q)$
 et $u = \max(q, s)$

$$A_{(r,r)} = \begin{bmatrix} a_1 & a_2 & \dots & a_{p-1} & \overbrace{a_p \ 0 \ \dots \ 0}^{r-p} \\ 1 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 1 & 0 & 0 & \dots & 0 \end{bmatrix}$$

$$C_0 = [1 \ 0 \ \dots \ 0]'_{(1,r)}$$

$$C_j = [c_j \ 0 \ \dots \ 0]' \quad j = 1, \dots, q.$$

= $\underline{0}$ pour tout $j = q+1, q+2, \dots, u$ quand $q < s$.

$$B_j = \begin{bmatrix} b_{jj} & b_{j+1,j} & \dots & b_{p,j} & \overbrace{0 \ 0 \ \dots \ 0}^{r-p+j-1} \\ 1 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 1 & 0 & 0 & \dots & 0 \end{bmatrix} \quad \text{pour } j = 1 \dots s$$

= $\underline{0}$ pour $j = s+1, s+2, \dots, u$ quand $q > s$.

$$\tilde{X}_{(1,r)}(t) = [X(t), X(t-1), \dots, X(t-r+1)]', \quad t \in \mathbb{Z}$$

où $\underline{0}$ est la matrice d'ordre approprié dans laquelle chaque entrée

est nulle, et où "''" est utilisé pour désigner l'opération de transposition d'une matrice.

Théorème 4.3.1.

Si $X(t)$, $t \in \mathbb{Z}$ et $\epsilon(t)$, $t \in \mathbb{Z}$ sont deux processus stochastiques vérifiant l'équation (4.1.1) avec la condition (H), alors :

$$(4.3.1) \quad \tilde{X}(t) = A\tilde{X}(t-1) + C_0\epsilon(t) + \sum_{j=1}^u C_j\epsilon(t-j) + \sum_{j=1}^u B_j\tilde{X}(t-j)\epsilon(t-j) \quad \text{p.s.}$$

pour tout $t \in \mathbb{Z}$.

On cherche à résoudre le problème suivant : étant donné $\epsilon(t)$, $t \in \mathbb{Z}$, une suite de variables aléatoires indépendantes équidistribuées gaussiennes centrées, de variance finie $\sigma^2 < \infty$, et les matrices réelles A (r, r), B_j (r, r), $j = 1, \dots, u$, C_0 ($1, r$), C_j ($r, 1$), $j = 1, \dots, u$, existe-t-il un processus stationnaire unique $\tilde{X}(t)$, ($r, 1$), $t \in \mathbb{Z}$ satisfaisant (4.3.1) ?

2. Théorème d'existence des processus bilinéaires superdiagonaux

Le résultat que nous énonçons maintenant fut établi par S.I. AKAMANAM, M. BHASKARA RAO et K. SUBRAMANYAM (1986). Ce résultat généralise un précédent résultat établi en 1983 par M. BHASKARA RAO, T. SUBBA RAO et A.M. WALKER, pour des processus $X(t)$, $t \in \mathbb{Z}$ vérifiant l'équation (4.1.1) avec la condition (H), sans partie moyenne mobile. S.I. AKAMANAM, M. BHASKARA RAO et S. SUBRAMANYAM mettent ainsi en évidence le fait que la partie moyenne mobile ne joue aucun rôle quant à l'établissement de l'existence de tels modèles. Ils ont d'autre part établi l'unicité de la solution. Le théorème est le suivant :

Théorème 4.3.2. (AKAMANAM, BHASKARA RAO, SUBRAMANYAM, 1986)

Soit $\epsilon(t)$, $t \in \mathbb{Z}$ une suite de variables aléatoires indépendantes équidistribuées centrées et de variance σ^2 définie sur un certain espace de probabilité (Ω, \mathcal{F}, P) , $(u+1)$ matrices d'ordre (r, r) : A, B_1, \dots, B_u , et les relations :

$$\Gamma_1 = A \otimes A + \sigma^2(B_1 \otimes B_1)$$

$$\begin{aligned} \Gamma_j = \sigma^2 [& B_j \otimes (A^{j-1}B_1 + A^{j-2}B_2 + \dots + AB_{j-1}) \\ & + (A^{j-1}B_1 + A^{j-2}B_2 + \dots + AB_{j-1}) \otimes B_j \\ & + B_1 \otimes B_j] \quad \text{pour } j = 1, \dots, u \end{aligned}$$

où \otimes est le symbole de Kronecker du produit des matrices.

Si toutes les valeurs propres de la matrice

$$M_{(r^2u, r^2u)} = \begin{bmatrix} \Gamma_1 & \Gamma_2 & \dots & \Gamma_{u-1} & \Gamma_u \\ I_r^2 & 0 & \dots & 0 & 0 \\ 0 & I_r^2 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & I_r^2 & 0 \end{bmatrix}$$

ont un module plus petit que l'unité, I_n étant la matrice unité d'ordre (n, n) , $C_0(1, r)$ et $C_j(1, r)$, $1 \leq j \leq u$, des vecteurs colonne donnés, alors il existe un processus strictement stationnaire $\tilde{X}(t)$, $t \in \mathbb{Z}$, satisfaisant l'équation (4.3.1) pour tout $t \in \mathbb{Z}$.

D'autre part, si un processus $U(t)$, $t \in \mathbb{Z}$ vérifie le modèle bilinéaire (4.3.1), alors :

$$U(t) = \tilde{X}(t), \text{ p.s. pour tout } t \in \mathbb{Z}.$$

On peut remarquer que dès que l'on a l'existence du processus $\tilde{X}(t)$, on a aussi l'ergodicité du processus. Pour la démonstration de ce théorème on pourra se reporter à l'article de M. BHASKARA RAO, T. SUBBA RAO et A.M. WALKER (1983, p. 106-109).

Ce théorème généralise des résultats qui ont été établis pour des sous classes de ces modèles, ou pour des cas particuliers. Néanmoins il ne fournit qu'une condition suffisante d'existence et d'unicité des modèles vérifiant (4.1.1) avec la condition (H), sans exhiber l'expression du processus lui-même. On va voir maintenant que l'on a des résultats plus complets dans des cas particuliers.

3.- Théorèmes d'existence pour des cas particuliers de modèles bilinéaires superdiagonaux

a. Modèle bilinéaire BL(p,0,p,1)

Considérons le modèle superdiagonal sans partie moyenne mobile BL(p,0,p,q) qui a pour expression :

$$(4.3.2) \quad X(t) = \sum_{i=1}^p a_i X(t-i) + \epsilon(t) + \sum_{j=1}^q \sum_{k=1}^p b_{jk} X(t-k) \epsilon(t-j) .$$

Il peut se réécrire comme le modèle (4.1.1) sous forme vectorielle, les matrices sont telles que : A est d'ordre (p,p), C₀ est d'ordre (1,p), B_j, 1 ≤ j ≤ q est d'ordre (p,p) et on pose p = P dans la dernière colonne.

On obtient alors le théorème suivant :

Théorème 4.3.3.

Si $X(t)$, $t \in \mathbb{Z}$ et $\epsilon(t)$, $t \in \mathbb{Z}$ sont deux processus stochastiques vérifiant l'équation (4.3.2), alors

$$(4.3.3) \quad \tilde{X}(t) = A \tilde{X}(t-1) + C_0 \epsilon(t) + \sum_{j=1}^q B_j \tilde{X}(t-j) \epsilon(t-j) \quad \text{p.s.}$$

pour tout $t \in \mathbb{Z}$.

Notons que dans l'expression de ces modèles, les $X(t)$ qui

apparaissent dans le premier et le troisième termes du membre de droite de l'expression (4.3.2) présentent les mêmes retards. On se pose alors le même problème que précédemment, à savoir l'existence et l'unicité d'un processus stationnaire $\tilde{X}(t)$, $t \in \mathbb{Z}$ vérifiant l'équation (4.3.3). M. BHASKARA RAO, T. SUBBA RAO et A.M. WALKER (1983) ont résolu ce problème pour une classe particulière de modèles bilinéaires superdiagonaux, les modèles $BL(p,0,p,1)$. Ils donnent alors des conditions nécessaires et suffisantes d'existence des modèles ainsi que l'expression du processus limite.

Théorème 4.3.4. (BHASKARA RAO, SUBBA RAO, WALKER, 1983)

Soit $\epsilon(t)$, $t \in \mathbb{Z}$ une suite de variables aléatoires indépendantes équidistribuées centrées et de variance $\sigma^2 < \infty$ définies sur un certain espace de probabilité (Ω, \mathcal{F}, P) , A et B_1 deux matrices d'ordre (p,p) telles que :

$$(H_2) \quad \rho(A \otimes A + \sigma^2 B_1 \otimes B_1) = \lambda < 1$$

où $\rho(M)$ désigne la plus grande valeur propre de la matrice M , et C_0 un vecteur colonne de dimension p .

Alors la série de vecteurs aléatoires :

$$\sum_{r \geq 1} \prod_{j=1}^r (A + B_1 \epsilon(t-j)) C_0 \epsilon(t-r)$$

converge absolument presque sûrement, et en moyenne pour tout $t \in \mathbb{Z}$.

Si

$$(4.3.4) \quad \tilde{X}(t) = C_0 \epsilon(t) + \sum_{r \geq 1} \prod_{j=1}^r (A + B_1 \epsilon(t-j)) C_0 \epsilon(t-r), \quad t \in \mathbb{Z}$$

alors le processus $\tilde{X}(t)$, $t \in \mathbb{Z}$ est un processus strictement stationnaire vérifiant le modèle bilinéaire :

$$(4.3.5) \quad \tilde{X}(t) = A\tilde{X}(t-1) + C_0 \epsilon(t) + B_1 \tilde{X}(t-1) \epsilon(t-1) \quad \text{p.s.}$$

pour tout $t \in \mathbb{Z}$.

Inversement si le processus $\tilde{X}(t)$, $t \in \mathbb{Z}$ est strictement stationnaire et vérifie l'équation (4.3.5), pour tout $t \in \mathbb{Z}$,

pour une suite de variables aléatoires, indépendantes équadistribuées $\epsilon(t)$, $t \in \mathbb{Z}$ centrées, et de variance $\sigma^2 < \infty$, et A , B_1 , C_0 étant des matrices respectivement d'ordre (p,p) , (p,p) et $(p,1)$ vérifiant la condition (H_2) , alors $\tilde{X}(t)$ est donné par l'équation (4.3.4) p.s., pour tout $t \in \mathbb{Z}$.

Remarques

1. Ce théorème est plus général que le théorème 4.3.2 car il donne l'expression du processus limite, mais il ne concerne qu'une sous classe de modèles superdiagonaux.

2. Apparaissent comme cas particulier de cette sous classe de modèles, les modèles bilinéaires à un terme, $BL(1,0,1,1)$ et $BL(0,0,2,1)$ respectivement étudiés par T. PHAM DINH et L. TAT TRAN (1981), et par D. GUEGAN (1981 a), à partir de leur représentation scalaire. Ces modèles vérifient la représentation vectorielle (4.3.5) avec les matrices appropriées.

Pour le modèle bilinéaire $BL(1,0,1,1)$, il suffit de prendre :

$$A = \begin{bmatrix} a & 0 \\ 0 & 0 \end{bmatrix} \quad B = \begin{bmatrix} b_{11} & 0 \\ 0 & 0 \end{bmatrix} \quad C = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\tilde{X}(t) = [X(t) \quad X(t-1)]'$$

la condition (H_2) devient alors $a^2 + \sigma^2 b_{11}^2 < 1$.

Pour le modèle bilinéaire $BL(0,0,2,1)$, on pose :

$$A = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad B = \begin{bmatrix} 0 & b_{21} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad C = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

$$\tilde{X}(t) = [X(t) \quad X(t-1) \quad X(t-2)]'$$

la condition (H_2) devient : $b_{21}^2 \sigma^2 < 1$.

On retrouve alors les résultats décrits par les auteurs des articles cités. Mais ces résultats ont été établis à partir d'une autre démarche et ont conduit à des résultats plus précis et parfois plus forts. Nous donnons maintenant ces résultats.

b. Le Modèle Bilinéaire Diagonal à un terme BL(1,0,1,1)

Ce modèle est étudié par T. PHAM DINH et L. TAT TRAN (1981). Il a pour expression :

$$(4.3.6) \quad X(t) = a_1 X(t-1) + \epsilon(t) + b_{11} X(t-1) \epsilon(t-1) .$$

Théorème 4.3.5. (PHAM DINH et TAT TRAN, 1981)

Soit une suite de variables aléatoires indépendantes équadistribuées $\epsilon(t)$, $t \in \mathbb{Z}$, centrées, de variance $\sigma^2 < \infty$, admettant des moments d'ordre 4 finis, définis sur un certain espace de probabilité (Ω, \mathcal{F}, P) , alors une condition nécessaire et suffisante d'existence d'un processus strictement stationnaire de carré intégrable $X(t)$, $t \in \mathbb{Z}$, vérifiant (4.3.6) est $a_1^2 + b_{11}^2 \sigma^2 < 1$.

Si cette condition est vérifiée, alors $X(t)$ est unique et est donné par l'expression :

$$X(t) = \epsilon(t) + \sum_{r \geq 1} \left[\prod_{k=1}^r (a_1 + b_{11} \epsilon(t-k)) \right] \epsilon(t-r)$$

qui converge fortement et en moyenne quadratique.

Dans ce cas particulier, on obtient une condition nécessaire et suffisante d'existence du processus $X(t)$, on a donc

un résultat plus fort que dans le cas général qui ne fournit qu'une condition suffisante.

c. Modèles Bilineaires superdiagonaux sans partie autorégressive et sans partie moyenne mobile BL(0,0,p,q)

Un premier résultat a été établi pour le modèle bilinéaire superdiagonal à un terme BL(0,0,2,1). Soit le processus $X(t)$, $t \in \mathbb{Z}$ défini par l'équation :

$$(4.3.7) \quad X(t) = \epsilon(t) + b_{21}X(t-2)\epsilon(t-1)$$

Théorème 4.3.6. (GUEGAN, 1981, p. 244)

Etant donné une suite de variables aléatoires indépendantes équidistribuées $\epsilon(t)$, $t \in \mathbb{Z}$ centrées, de variance $\sigma^2 < \infty$ définies sur un certain espace de probabilité (Ω, \mathcal{F}, P) , une condition nécessaire et suffisante d'existence d'un processus strictement stationnaire, de carré intégrable $X(t)$, $t \in \mathbb{Z}$ défini par (4.3.7) est $b_{21}^2 \sigma^2 < 1$.

Sous cette condition, le processus $X(t)$, $t \in \mathbb{Z}$ est unique et est donné par l'équation

$$X(t) = \epsilon(t) + \sum_{j \geq 1} b_{21}^j \epsilon(t-2j) \prod_{k=1}^j \epsilon(t-2k+1)$$

qui converge en moyenne quadratique.

Parallèlement E.J. HANNAN (1982), et B.G. QUINN (1982 b) ont abordé le problème de l'existence de processus bilinéaires superdiagonaux, sans partie moyenne mobile, ni partie autorégressive. On considère le processus un peu plus général $X(t)$, $t \in \mathbb{Z}$ défini par l'équation :

$$(4.3.8) \quad X(t) = \epsilon(t) + \sum_{j=1}^Q \sum_{k=2}^P b_{jk} X(t-k) \epsilon(t-j) .$$

E.J. HANNAN établit le résultat suivant :

Théorème 4.3.7. (HANNAN, 1982)

Etant donné une suite de variables aléatoires indépendantes équidistribuées, gaussiennes, centrées, de variance $\sigma^2 < \infty$, définies sur un certain espace de probabilité, une condition nécessaire et suffisante pour qu'il existe un processus stationnaire au second ordre $X(t)$, $t \in \mathbb{Z}$ vérifiant (4.3.8) est que :

$$\sum_1^{\infty} \sigma^{2n} \sum_r d_r (j_1, \dots, j_n)^2 < \infty$$

où la somme \sum_r est faite pour $0 < j_1 < \dots < j_n < \infty$, sauf quand $n = 1$, alors on somme pour $0 \leq j_1 < \infty$. Sous cette condition, $X(t)$ est donné par l'expression :

$$X(t) = \sum_{n=1}^{\infty} \sum_r d_r (j_1, \dots, j_n) \epsilon(t-j_1) \dots \epsilon(t-j_n)$$

qui converge en moindres carrés.

B.G. QUINN (1982 b), s'est intéressé pour sa part à des conditions de stricte stationnarité pour un cas particulier du modèle précédent, soit :

$$(4.3.9) \quad X(t) = \epsilon(t) + b_{jk} X(t-k) \epsilon(t-j), \quad k, j > 0 \text{ et } k > j .$$

Théorème 4.3.8. (QUINN, 1982 b)

Etant donné une suite de variables aléatoires faiblement

mélangeantes $\epsilon(t)$, $t \in \mathbb{Z}$ telles que :

$$E|\log|\epsilon(t)|| < \infty \quad \text{et} \quad E|\log|\epsilon(t)||^2 < \infty ,$$

une condition nécessaire et suffisante pour qu'il existe un processus strictement stationnaire $X(t)$, $t \in \mathbb{Z}$ vérifiant (4.3.9) est :

$$\log|b_{jk}| + E \log|\epsilon(t)| < 0 .$$

IV - INVERSIBILITE DES MODELES BILINEAIRES

Comme nous l'avons précisé au chapitre I, l'inversibilité est une notion très importante dans l'étude des modèles tant linéaires que non linéaires, car elle permet d'une part de faire des prévisions, d'autre part, d'estimer dans certains cas les paramètres du modèle.

Les principales définitions de l'inversibilité ont été données au chapitre III, nous allons maintenant préciser dans quelles circonstances elles ont été utilisées dans le cadre des modèles bilinéaires.

Les résultats sont en fait assez limités et ne concernent que des cas particuliers de modèles bilinéaires.

1. Inversibilité au sens de GRANGER-ANDERSEN

On considère un processus $X(t)$, $t \in \mathbb{Z}$ vérifiant le modèle suivant :

$$(4.4.1) \quad X(t) = f(X(t-i), \epsilon(t-j), i = 1, \dots, k, j = 1, \dots, \ell) + \epsilon(t)$$

où $\epsilon(t)$, $t \in \mathbb{Z}$ est un bruit blanc. On pose $\hat{\epsilon}(t)$ un estimateur de $\epsilon(t)$ défini par :

$$\hat{\epsilon}(t) = X(t) - f(X(t-i), i = 1, \dots, k, \hat{\epsilon}(t-j), j = 1, \dots, \ell)$$

où

$$\hat{\epsilon}(t-j) = 0 \text{ pour } j > t .$$

Le modèle (4.4.1) est dit inversible si :

$$\lim_{t \rightarrow \infty} E[(\epsilon(t) - \hat{\epsilon}(t))^2] = 0$$

quand les paramètres et le modèle sont parfaitement connus.

On peut remarquer que cette condition implique que le processus est inversible si la variance de l'erreur $e(t) = \epsilon(t) - \hat{\epsilon}(t)$, obtenue en estimant $\epsilon(t)$ en fonction d'un nombre fini de valeurs passées et présentes de $X(t)$, tend vers zéro, quand ce nombre devient infini et ceci conditionnellement aux valeurs d'origine. C.W.J. GRANGER et A.P. ANDERSEN (1978 a) établissent l'inversibilité pour le modèle bilinéaire diagonal d'ordre 1 BL(0,0,1,1) défini par :

$$X(t) = \epsilon(t) + b_{11}X(t-1)\epsilon(t-1) .$$

La condition suffisante est $2b_{11}^2\sigma^2 < 1$.

Ce résultat apparaît en fait comme un cas particulier du théorème suivant établi par la suite par T. SUBBA RAO (1981), pour les modèles bilinéaires superdiagonaux BL(p,0,p,1), définis par l'équation (4.3.2) dans laquelle on a posé $q = 1$.

Théorème 4.4.1. (SUBBA RAO, 1981)

Etant donné un processus $X(t)$, $t \in \mathbb{Z}$ tel que :

$$\begin{cases} X(t) = H \tilde{X}(t) \\ \tilde{X}(t) = A \tilde{X}(t-1) + C_0\epsilon(t) + B_1 \tilde{X}(t-1)\epsilon(t-1) \end{cases}$$

où $\epsilon(t)$ est une suite de variables aléatoires indépendantes

équidistribuées, centrées, de variance $\sigma^2 < \infty$, où les matrices A, B_1, C_0 sont définies comme dans le théorème 4.3.4, et où H est un vecteur ligne de dimension p , alors une condition suffisante pour que le processus soit inversible est :

$$HB_1 E[\tilde{X}(t)\tilde{X}(t)]B_1H' < (HC_0)^2 .$$

Notons tout d'abord que A, B_1, C_0 et H étant donnés, on peut calculer explicitement $V = E[\tilde{X}(t)\tilde{X}(t)]$, (cf. Paragraphe V, Théorème 4.5.1.).

Remarque

Considérons $\hat{\epsilon}(t)$ un estimateur de $\epsilon(t)$ vérifiant l'équation

$$X(t) = HA \tilde{X}(t-1) + HB_1 \tilde{X}(t-1)\hat{\epsilon}(t-1) + HC_0\hat{\epsilon}(t)$$

et soit $e(t) = \hat{\epsilon}(t) - \epsilon(t)$. Dans sa démonstration T. SUBBA RAO utilise l'inégalité suivante, où l'on a posé $\zeta(t) = \frac{HB_1}{HC_0} \tilde{X}(t)$,

$$E[e(t)^2] \leq (E(\zeta(s))^2)^t$$

et qui ne semble pas correcte. Néanmoins en partant de la relation $e(t) = \zeta(t-1)e(t-1)$ et en suivant la démarche de C.W.J. GRANGER et A.P. ANDERSEN (1978 c) avec $\alpha x(t)$ remplacé par $\zeta(t)$, où α est défini comme dans C.W.J. GRANGER et A.P. ANDERSEN, on obtient le résultat. \square

Parallèlement, à partir d'une définition similaire de l'inversibilité, D. GUEGAN (1981 a) donne une condition suffisante d'inversibilité du modèle superdiagonal à un terme BL(0,0,2,1) défini par l'équation (4.3.7), ainsi que l'expression de $\epsilon(t)$ en fonction des valeurs passées de $X(t)$.

Théorème 4.4.2. (GUEGAN, 1981 a, p. 246)

Etant donné un processus $X(t)$, $t \in \mathbb{Z}$ défini par l'équation (4.3.7), où $\epsilon(t)$, $t \in \mathbb{Z}$ est une suite de variables aléatoires

indépendantes, centrées, de variance $\sigma^2 < \infty$, une condition suffisante d'inversibilité de ce modèle est $2b_{21}^2\sigma^2 < 1$.

Sous cette condition, le processus $\epsilon(t)$ a pour expression :

$$\epsilon(t) = X(t) + \sum_{n \geq 1} (-b_{21})^n X(t-n) \prod_{j=1}^n X(t-j-1) \quad \text{p.s.}$$

Notons que sous cette condition d'inversibilité, on peut montrer que le processus $X(t)$, $t \in \mathbb{Z}$ défini par l'équation (4.3.7) est non singulier linéaire (cf. Chapitre III, p. 114).

2. Inversibilité au sens fort

Rappelons qu'un processus $X(t)$, $t \in \mathbb{Z}$ est inversible au sens faible, s'il existe une suite $\hat{\epsilon}(n)$ fonction de $X(1), \dots, X(n)$ uniquement telle que $\hat{\epsilon}(n) - \epsilon(n) \rightarrow 0$ en probabilité, quand $n \rightarrow \infty$. Cette définition ne précisant pas comment va être construite la suite $\hat{\epsilon}(n)$.

Nous allons dans le cas du modèle étudié par T. PHAM DINH et L. TAT TRAN (1981) - qui vérifie l'équation (4.3.6) - donner une construction explicite de la suite $\hat{\epsilon}(n)$ et utiliser alors une définition forte de l'inversibilité pour obtenir une condition suffisante d'inversibilité de ce modèle.

A partir du modèle (4.3.6), $\epsilon(t)$ peut être calculé de manière récursive par :

$$(4.4.2) \quad \begin{cases} \epsilon(t) = X(t) - Z(t-1) \\ Z(t-1) = [a_1 + b_{11}\epsilon(t)]X(t) = [a_1 + b_{11}X(t)]X(t) - b_{11}X(t)Z(t-1) \end{cases}$$

Jusqu'à présent on supposait connues les valeurs des paramètres, or en fait, si on note $\theta = (a_1, b_{11})$ les paramètres, ceux-ci doivent être estimés quand on ajuste le modèle aux données.

Considérons une valeur quelconque $\tilde{\theta}$ de θ . Le bruit de l'équation récurrente (4.4.2) est alors calculé en fonction de $\tilde{\theta}$, on le note $\epsilon_{\tilde{\theta}}(t|z_0)$. On se donne maintenant une valeur arbitraire z_0 de $Z(0)$ et on définit récursivement $\epsilon_{\tilde{\theta}}(t|z_0)$, $Z(t|z_0)$ à partir des équations (4.4.2), avec $Z(0|z_0) = z_0$.

Une forme forte de l'inversibilité est de demander qu'il existe un processus stationnaire $\epsilon_{\tilde{\theta}}(t)$ tel que $\epsilon_{\tilde{\theta}}(t|z_0) - \epsilon_{\tilde{\theta}}(t) \rightarrow 0$ en probabilité quand $t \rightarrow \infty$, quelque soit z_0 .

Théorème 4.4.3. (PHAM DINH et TAT TRAN, 1981)

Le modèle (4.3.6) est fortement inversible en $\tilde{\theta} = (\tilde{a}_1, \tilde{b}_{11})$ relativement au processus observé $X(t)$, si

$$(4.4.3) \quad \tilde{b}_{11} < \exp\{-E[\log|X(t)|]\}.$$

Remarques

1. Dans cet exemple, on dispose aussi de l'expression du processus $\epsilon(t)$, $t \in \mathbb{Z}$, en fonction des observations $X(s)$, $s \leq t$:

$$\epsilon_{\tilde{\theta}}(t) = X(t) - \sum_{j \geq 1} (-\tilde{b}_{11})^{j-1} \left\{ \prod_{k=1}^j X(t-k) \right\} \{ \tilde{a}_{11} + \tilde{b}_{11} X(t-j) \} \quad \text{p.s.}$$

2. Une condition suffisante, conséquence de la condition (4.4.3), est $\tilde{b}_{11}^2 E[X^2(t)] < 1$. (On retrouve alors la condition donnée dans le théorème 4.4.1).

3. Si les variables aléatoires considérées $\epsilon(t)$ sont aussi gaussiennes, on obtient une région de l'espace où le processus est inversible :

$$2(1+a_1)\lambda^4 + 2(1-a_1)\lambda^2 - (1-a_1)^2(1+a_1) \leq 0 \quad \text{où } \lambda^2 = b_{11}^2 \sigma^2.$$

4. Dans le chapitre III nous avons donné une autre définition de l'inversibilité faible. (Définition 3.3.2). B.G. QUINN (1982 a) donne une condition nécessaire et suffisante d'inversibilité des modèles bilinéaires définis par l'équation (4.3.9), à partir de cette définition. Un processus $X(t)$, $t \in \mathbb{Z}$

est inversible si $\epsilon(t)$ est mesurable par rapport à la tribu $\sigma(X(s), s \leq t)$ engendrée par les $X(s), s \leq t$.

Théorème 4.4.4. (QUINN, 1982 a)

On suppose que le processus $X(t), t \in \mathbb{Z}$ défini par l'équation (4.3.9) satisfait $E|\log|X(t)|| < \infty$ et que les conditions du théorème 4.3.8 sont vérifiées, alors une condition nécessaire et suffisante pour que le processus soit inversible est

$$\log|b_{jk}| + E[\log|X(t)||] < 0 .$$

V - STATIONNARITE AU SECOND ORDRE ET FONCTION DE COVARIANCE

T. SUBBA RAO (1981) a établi la stationnarité asymptotique au second ordre d'une sous classe des modèles bilinéaires superdiagonaux, (les modèles $BL(p,0,p,1)$), et a fourni l'expression de sa fonction de covariance. T. SUBBA RAO et M.M. GABR (1984) généralisent ensuite ces résultats aux modèles bilinéaires $BL(p,0,p,p)$. Parallèlement S.I. AKAMANAM (1983) fournit l'expression générale de la covariance pour l'ensemble des modèles bilinéaires superdiagonaux. Nous énonçons seulement ici le résultat établi par T. SUBBA RAO (1981), pour des raisons de simplicité. Pour les autres cas, on se rapportera aux références citées.

Théorème 4.5.1. (SUBBA RAO, 1981)

Une condition suffisante pour que le processus $\tilde{X}(t)$, $t \in \mathbb{Z}$ défini par l'équation (4.3.5), où $\epsilon(t)$ est une suite de variables aléatoires indépendantes équidistribuées gaussiennes, centrées de variance $\sigma^2 < \infty$, soit stationnaire au second ordre, est donnée par :

$$\rho(A \otimes A + \sigma^2 B_1 \otimes B_1) < 1 .$$

Sous cette condition, si $\mu = E[\tilde{X}(t)]$, $V = E[\tilde{X}(t)\tilde{X}'(t)]$ et $C(s) = E[(\tilde{X}(t-s) - \mu)(\tilde{X}(t) - \mu)']$, on a les relations suivantes :

$$(i) \quad \mu = (I-A)^{-1} B_1 C_0$$

$$(ii) \quad V = AVA' + B_1 V B_1' + \Delta_1$$

où $\Delta_1 = ASB_1 + B_1 SA' + 2B_1 C_0 C_0' B_1' + C_0 C_0'$
avec $S = A\mu C_0' + B_1 C_0 C_0' + C_0 \mu' A' + C_0 C_0' B_1'$.

$$(iii) \quad C(0) = AC(0)A' + B_1 C(0)B_1' + \Delta_2$$

$$C(1) = AC(0) + \Delta_3$$

$$C(s) = AC(s-1) = A^{s-1}C(1) \quad s = 2, 3, \dots$$

$$\text{où } \Delta_2 = B_1 \mu \mu' B_1' + A \mu \mu' A' + ASB_1' + 2B_1' C_0 C_0' A' + C_0 C_0' A' + C_0 C_0' - \mu \mu'$$

$$\Delta_3 = A \mu \mu' + B_1 S - \mu \mu' .$$

Remarques

1. Les résultats donnés dans ce théorème généralisent et confirment ceux obtenus par C.W.J. GRANGER et A.P. ANDERSEN (1978 a) pour des cas particuliers.

2. Remarquons d'autre part que pour le modèle bilinéaire considéré, on a les relations suivantes

$$c(s) + a_1 c(s-1) + \dots + a_p c(s-p) = 0 \quad s > 1$$

Ces équations étant les mêmes que les équations de Yule Walker pour un processus ARMA (p,1). On met ainsi en évidence que le modèle bilinéaire BL(p,0,p,1) a même structure de covariance

qu'un modèle ARMA (p,1). Cette remarque confirme le fait que la structure de covariance ne suffit pas à caractériser les modèles bilinéaires.

S.I. AKAMANAM (1983) généralise ce résultat à l'ensemble des modèles superdiagonaux.

Théorème 4.5.2. (AKAMANAM, 1983)

Etant donné un processus $\tilde{X}(t)$, $t \in \mathbb{Z}$ défini par l'équation (4.3.1), il existe un processus ARMA (r,u), ayant pour coefficients autorégressifs a_1, \dots, a_r , et pour coefficients moyenne mobile des fonctions des matrices A, B_j, C_j , $1 \leq j \leq u$ et C_0 , qui a la même structure de covariance que le processus $\tilde{X}(t)$, $t \in \mathbb{Z}$.

VI - MOMENTS D'ORDRE SUPERIEUR DES PROCESSUS BILINEAIRES

Par cette approche il n'est pas possible d'obtenir des résultats généraux concernant l'existence de moments d'ordre supérieur. Néanmoins il existe des résultats dans des cas particuliers.

Notons que les moments n'existent pas forcément pour tous les ordres comme l'ont mis en évidence C.W.J. GRANGER et A.P. ANDERSEN (1978 a).

D. GUEGAN (1984 a) donne l'expression des moments jusqu'à l'ordre 4 de deux modèles bilinéaires simples :

$$(4.6.1) \quad X(t) = \epsilon(t) + \delta \epsilon(t-1) \epsilon(t-2)$$

$$(4.6.2) \quad X(t) = \epsilon(t) + \gamma X(t-1) \epsilon(t-2)$$

On note $m = E[X(t)]$, $c(k) = E[X(t)X(t-k)]$ $\forall k \geq 0$

$$c(k, \ell) = E[X(t)X(t-k)X(t-\ell)] \quad \forall k \geq 0, \forall \ell \geq 0$$

$$c(h, k, \ell) = E[X(t)X(t-h)X(t-k)X(t-\ell)] .$$

On a alors les propositions suivantes :

Proposition 4.6.1. (GUEGAN, 1984 a, p. 258)

Soit le processus $X(t)$, $t \in \mathbb{Z}$ défini par l'équation (4.6.1) où $\epsilon(t)$ est une suite de variables aléatoires indépendantes équidistribuées gaussiennes centrées, de variance $\sigma^2 < \infty$. C'est un processus centré, tel que $\forall p, X(t) \in L^p$ et si $v = \delta^2 \sigma^2$

$$(i) \quad c(0) = \sigma^2(1-v)$$

$$\forall k > 0 \quad c(k) = 0$$

$$(ii) \quad c(1,2) = \delta \sigma^4$$

$$\forall k \neq 1, \forall \ell \neq 2, \quad c(k, \ell) = 0$$

$$(iii) \quad c(0,0,0) = 3\sigma^4(1+2v+3v^2)$$

$$c(0,1,1) = \sigma^4(1+4v+3v^2)$$

$$c(0,2,2) = \sigma^4(1+4v+v^2)$$

$$c(0,k,k) = \sigma^4(1+v)^2 \quad \forall k \geq 3$$

$$c(1,1,3) = 2v\sigma^4$$

et $c(h, \ell, k) = 0$ pour toutes autres valeurs de h, ℓ, k .

Notons que si l'on considère le processus $X(t)$, $t \in \mathbb{Z}$:

$$X(t) = \epsilon(t) + \delta \epsilon(t-i)\epsilon(t-j), \quad i, j \text{ fixes quelconques,}$$

les résultats (ii) de la proposition(4.6.1) se généralisent facilement :

$$c(i, j) = \delta \sigma^4 \quad \text{et} \quad \forall k \neq i, \forall \ell \neq j, \quad c(k, \ell) = 0 .$$

Proposition 4.6.2. (GUEGAN, 1984 a, p. 259)

Soit le processus $X(t)$, $t \in \mathbb{Z}$ défini par l'équation (4.6.2) où $\epsilon(t)$ est une suite de variables aléatoires indépendantes équidistribuées, centrées, de variance $\sigma^2 < \infty$, alors $X(t)$, $t \in \mathbb{Z}$ est un processus centré. Si $\gamma^2 \sigma^2 < 1$, $X(t) \in L^2$ et si $\gamma^2 \sigma^2 < \frac{1}{3}$ ses moments d'ordre 4 sont finis. Sous ces conditions on a les résultats suivants. Posons $u = \gamma^2 \sigma^2$:

$$(i) \quad c(0) = \frac{\sigma^2}{1-u}$$

$$c(k) = 0 \quad \forall k > 0$$

$$(ii) \quad c(1,2) = \frac{\gamma \sigma^4}{1-u}$$

$$c(k,\ell) = 0 \quad \forall k \neq 1 \quad \forall \ell \neq 2$$

$$(iii) \quad c(0,0,0) = \frac{3\sigma^4(1+u)}{(1-u)(1-3u^2)}$$

$$c(0,1,1) = \frac{\sigma^4(1+2u)}{(1-u)^2}$$

$$c(0,2,2) = \frac{\sigma^4(1+3u)}{(1-u)(1-3u^2)}$$

$$c(0,2k-1,2k-1) = \frac{\sigma^4(1+2u^k)}{(1-u)^2}$$

$$c(0,2k,2k) = \frac{\sigma^4}{(1-u)^2} + \frac{2\sigma^4 u^{2k}}{(1-u)^2(1-3u^2)}$$

$$c(1,3,3) = \frac{2u\sigma^4}{1-u}$$

$$c(0,1,4) = \frac{2u^2\sigma^4}{1-u}$$

$$c(0,2k+1,2k+4) = \frac{2\sigma^4 u^{k+2}}{1-u}$$

$$c(1,3,4) = \frac{u\sigma^4}{1-u}$$

$c(h, k, \ell) = 0$ pour toutes autres valeurs de h, k, ℓ .

Notons que certains résultats se généralisent au processus $X(t)$, $t \in \mathbb{Z}$ défini par l'équation :

$$X(t) = \epsilon(t) + \gamma\epsilon(t-i)X(t-j)$$

et l'on a en particulier :

$$c(i, j) = \frac{\gamma\sigma^4}{1-u}, \text{ et } \forall k \neq i, \forall \ell \neq j \quad c(k, \ell) = 0.$$

Les deux propositions (4.6.1) et (4.6.2) permettent de voir que pour ces modèles un seul moment d'ordre trois est non nul ainsi que quelques moments d'ordre quatre. Ces constatations permettront d'établir un test permettant de les différencier (cf. Chapitre VI).

Des calculs de moments d'ordre 3 et 4 pour des modèles bilinéaires à deux termes peuvent aussi être trouvés dans D. GUEGAN (1981 b).

W.K. LI (1984) calcule les covariances du processus $X^2(t)$, où le processus $X(t)$, $t \in \mathbb{Z}$ est défini par l'équation (4.3.9). Il obtient alors les résultats suivants :

Proposition 4.6.3. (LI, 1984)

Soit $X(t)$, $t \in \mathbb{Z}$ un processus strictement stationnaire défini par l'équation (4.3.9) où $\epsilon(t)$ est une suite de variables aléatoires indépendantes équidistribuées gaussiennes, centrées, de variance $\sigma^2 < \infty$, alors si $|b_{kj}^h E[\epsilon(t)]^h| < 1$, et si les moments d'ordre h existent, on a les résultats suivants, en posant $r_h = \text{cov}(X^2(t), X^2(t-h))$:

(1) Si $1 \leq h \leq j$ et $h < j < k$ alors

$$(i) \quad r_h = 0 \quad \text{si } j+h \neq k$$

$$(ii) \quad r_h = 2b_{kj}^4 \sigma^6 E[X^2(t)] / (1 - b_{kj}^4 \sigma^4)$$

(2) Si $h = j < k$ alors

$$(i) \quad r_h = 2b_{kj}^2 \sigma^4 E[X^2(t)] / (1 - b_{kj}^4 \sigma^4) \quad \text{si } 2j \neq k$$

$$(ii) \quad r_h = 2b_{kj}^2 \sigma^2 E[X^2(t)]^2 \quad \text{si } 2j = k .$$

Notons enfin que S.A.O. SESAY et T. SUBBA RAO (1985) ont établi pour le modèle bilinéaire BL(1,0,1,1) les expressions des cumulants d'ordre deux, trois, quatre. Il est important de noter que le calcul des moments d'ordre supérieur et des cumulants, bien que fastidieux dès que le nombre de termes dans le modèle augmente est néanmoins important car il peut permettre d'avancer dans la connaissance des propriétés statistiques des modèles.

CHAPITRE V

REPRESENTATION MARKOVIENNE ET APPLICATIONS

I - INTRODUCTION

La représentation d'état d'un système est un concept fondamental dans la théorie moderne du contrôle. En effet dans le cas où l'hypothèse markovienne est justifiée, l'information que l'on doit stocker concernant le système peut être énormément réduite.

La notion d'état a été essentiellement développée par R.E. KALMAN et d'autres en théorie du filtrage (cf. R.E. KALMAN, P.L. FALB et M.A. ARBIB, 1969). On peut voir l'état d'un système comme une représentation condensée de l'information concernant le comportement futur du système, obtenue à partir de l'observation présente et passée du processus.

La représentation d'état d'un système linéaire invariant à temps discret utilisée en théorie du contrôle a trouvé naturellement son extension et son interprétation dans le cas d'un système stochastique où l'entrée et la sortie sont aléatoires. H. AKAIKE (1975) a montré l'existence d'une représentation markovienne d'un tel système ainsi que celle des processus ARMA. Parmi ces dernières la représentation minimale joue un rôle décisif tant en théorie de la réalisation qu'en vue d'une procédure

d'identification des systèmes linéaires stochastiques, cf. B.L. HO et R.E. KALMAN (1966), P. FAURRE (1970), J. RISSANEN (1973).

Il était alors bien normal que la représentation d'état trouve son champ d'application avec les systèmes non linéaires. Les automaticiens se sont intéressés à des systèmes linéaires en les variables d'état mais non linéaires en les entrées. De tels systèmes, appelés affines en l'état, se sont montrés particulièrement adaptés à l'étude des réacteurs nucléaires, de certains modèles de population, à des processus de transfert de chaleur, etc, ..., (cf P. WHITTLE, 1969 et R.R. MOLHER, 1973). E.D. SONTAG (1979) considère les modèles affines en l'état polynomiaux en les entrées qui sont une sous classe des modèles précédents où la non linéarité en les entrées consiste en l'introduction de termes polynomiaux. Ce sont des modèles pour lesquels les sorties à un instant donné sont des sommes de produits des entrées et d'une sortie passée. Il établit à l'aide d'argument de la géométrie algébrique l'équivalence entre ce type de modèles et les systèmes régis par une équation aux différences linéaire en les sorties et polynomiale en les entrées.

L'analogie stochastique de ces derniers systèmes contient en particulier les modèles bilinéaires introduits au chapitre précédent. Deux approches ont été considérées en vue d'obtenir une représentation markovienne des modèles bilinéaires.

T. PHAM DINH (1985) généralise la représentation markovienne aux modèles bilinéaires superdiagonaux. D. GUEGAN (1986, 1987) montre que cette représentation ne se réalise pas de façon naturelle aux modèles sousdiagonaux.

Une autre approche est alors envisagée par D. GUEGAN (1986, 1987), qui suivant les travaux de E.D. SONTAG (1979) introduit les modèles stochastiques affines en l'état polynomiaux en les entrées.

Nous verrons comment ces deux démarches peuvent être considérées comme équivalentes grâce à un résultat établi par T. PHAM DINH (1986).

Disposant alors d'une représentation markovienne pour les modèles bilinéaires, l'étude d'un certain nombre de propriétés probabilistes est rendue plus accessible grâce à la théorie dont on dispose sur les processus de Markov. La suite de ce chapitre est consacrée tout d'abord à l'existence d'une représentation markovienne pour les modèles bilinéaires, puis aux propriétés probabilistes de ces modèles, comme l'inversibilité, l'ergodicité, l'ergodicité géométrique, l'existence et le calcul de la fonction de covariance et des moments d'ordre supérieur.

II - CONSTRUCTION D'UNE REPRESENTATION MARKOVIENNE

A. Représentation markovienne des modèles bilinéaires

1. Construction d'une représentation markovienne bilinéaire

Généralisant au cas non linéaire la démarche utilisée par H. AKAIKE (1974 a) pour établir la représentation markovienne des processus ARMA, T. PHAM DINH (1985) définit l'espace des prédicteurs non linéaires \mathcal{P}_t comme le sous espace fermé engendré par les prédicteurs $\hat{X}(t+k/t) = E[X(t+k)/\mathcal{F}_t]$ où \mathcal{F}_t est la tribu engendrée par les $X(s)$, $s \leq t$, $k = 1, 2, \dots$, si $X(t)$ est un processus centré strictement stationnaire ayant ses moments d'ordre deux finis. L'espace d'innovation I_t est alors le sous espace fermé engendré par les variables aléatoires :

$$\hat{\eta}(t,k) = \hat{X}(t+k/t) - \hat{X}(t+k/t-1), \quad k = 1, 2, \dots$$

Afin de pouvoir construire une représentation bilinéaire markovienne, on fait les hypothèses suivantes :

A_0 : l'espace des prédicteurs est de dimension finie

A_1 : les éléments de \mathcal{J}_t peuvent s'exprimer sous la forme :

$$\epsilon(t)Y(t-1) + c\epsilon(t) + d(\epsilon^2(t) - \sigma^2)$$

où $\epsilon(t) = X(t) - \hat{X}(t|\mathcal{F}_{t-1})$, $Y(t-1) \in \mathcal{F}_{t-1}$ et où c , d et σ^2 sont des constantes.

A_2 : les variables aléatoires $Y(t)$ définies en A_1 appartiennent au sous espace $\mathcal{P}_t(m)$ engendré par \mathcal{P}_t et les variables aléatoires $X(t), \dots, X(t-m+1)$ pour un certain $m \geq 0$.

L'hypothèse A_0 est évidemment nécessaire pour envisager la construction de la représentation markovienne. L'hypothèse A_1 permet de prendre en compte la structure non linéaire du processus que l'on étudie. En effet :

$$\hat{X}(t|\mathcal{F}_t) = HZ(t-1) \quad \text{et} \quad \hat{Z}(t|\mathcal{F}_{t-1}) = AZ(t-1)$$

où H et A sont des matrices ligne et carrée bien choisies. Si dans le cas linéaire ces relations suffisent pour établir la relation cherchée, dans le cas non linéaire, le vecteur aléatoire $Z(t) - \hat{Z}(t|\mathcal{F}_{t-1})$ n'est pas en général un multiple de l'innovation $\epsilon(t)$. On sait seulement qu'il appartient à \mathcal{J}_t , et c'est pourquoi on précise ici la forme des éléments de \mathcal{J}_t , (la présence du terme σ^2 a pour but de centrer le modèle). L'hypothèse A_2 empêche à la variable $Y(t)$ de se ballader arbitrairement dans l'espace \mathcal{F}_t en lui imposant de rester dans un sous espace de dimension finie, ce dernier comprend des valeurs passées de $X(t)$ car on a besoin de les mettre en mémoire pour obtenir un certain type de représentation.

Soit $Z(t)$ un vecteur ayant pour composantes les éléments d'une base de $\mathcal{P}_t(m)$, alors en utilisant (A_0) , (A_1) et (A_2) , on obtient :

$$(5.2.1) \quad \begin{cases} X(t) = HZ(t-1) + \epsilon(t) \\ Z(t) = AZ(t-1) + BZ(t-1)\epsilon(t) + C\epsilon(t) + D(\epsilon^2(t) - \sigma^2) \end{cases}$$

où $\epsilon(t)$ est une suite de variables aléatoires indépendantes identiquement distribuées avec $E[\epsilon(t)|Z(t-1), \dots] = 0$,

$E[\epsilon^2(t) | Z(t-1), \dots] = \sigma^2$, et où H, A, B, C, D sont respectivement des matrices ligne, carrée, carrée, colonne et colonne.

On dit qu'une telle représentation est une représentation bilinéaire markovienne: $\epsilon(t)$ est le bruit, $Z(t)$ et ses composantes, le vecteur d'état et les variables d'état. En fait $Z(t)$ est markovien si $\epsilon(t)$ est indépendant de $Z(s)$, $s < t$. Cette indépendance est supposée vérifiée en pratique, ce qui justifie une telle appellation. Notons enfin que l'hypothèse (A_2) est encore vraie pour tout $m' > m$. Cependant si on utilise un tel m' , on obtient une représentation de dimension plus grande et on introduit d'autres contraintes sur les matrices A, B, C, D, H , (T. PHAM DINH, 1985). Ces contraintes se traduisent par l'existence d'une matrice ligne K vérifiant :

$$\begin{aligned} KA^m &= H \\ KA^j B &= 0 & 0 \leq j < m \\ KA^j C &= 0 & 0 \leq j < m-1, & KA^{m-1} C = 1 \\ KA^j D &= 0 & 0 \leq j < m. \end{aligned}$$

La plus petite valeur de m , appelée index, telle que les contraintes précédentes soient satisfaites est une caractéristique de la représentation considérée. Sous les hypothèses A_0, A_1, A_2 on montre qu'un processus $X(t)$, $t \in \mathbb{Z}$ admet une représentation bilinéaire markovienne (5.2.1). L'inverse n'est vrai que sous certaines conditions. Il faut en particulier que la représentation soit inversible pour que A_1 soit vérifiée.

2. Application aux modèles bilinéaires superdiagonaux

Considérons le modèle bilinéaire d'ordre (p, q, P, Q) défini par les équations suivantes :

$$(5.2.2) \quad X(t) = \sum_{i=1}^p a_i X(t-i) + \epsilon(t) + \sum_{j=1}^q c_j \epsilon(t-j) + \sum_{j=1}^Q \sum_{k=1}^P b_{jk} X(t-k) \epsilon(t-j)$$

où $\epsilon(t)$ est une suite de variables aléatoires indépendantes équidistribuées de moyenne zéro et de variance σ^2 , et restreignons nous au cas où $b_{jk} = 0$ si $k < j$. En changeant la définition de b_{jk} et de P , ces modèles peuvent se réécrire :

$$(5.2.3) \quad X(t) = \sum_{i=1}^p a_i X(t-i) + \epsilon(t) + \sum_{j=1}^q c_j \epsilon(t-j) + \sum_{j=1}^Q \sum_{k=0}^P b_{jk} X(t-j-k) \epsilon(t-j)$$

On a alors le théorème suivant :

Théorème 5.2.1. (PHAM DINH, 1985)

Il existe une représentation bilinéaire markovienne pour les processus bilinéaires (5.2.3) avec n variables d'état, où $n = \max(p, P+q, Q+P)$, $m = n - \max(q, Q)$, donnée par :

$$(5.2.4) \quad \begin{aligned} X(t) &= Z_m(t-1) + \epsilon(t) \\ Z_i(t) &= Z_{i+1}(t-1) \quad 1 \leq i < m \\ Z_m(t) &= Z_{m+1}(t-1) + \epsilon(t) \\ Z_{m+i}(t) &= a_i Z_{m+1}(t-1) + Z_{m+i+1}(t-1) + (a_i + c_i) \epsilon(t) \\ &\quad + \sum_{k=0}^m b_{ik} Z_{m+1}(t-1) \epsilon(t) + b_{i0} \epsilon(t)^2 \quad 1 \leq i < n-m \\ Z_n(t) &= \sum_{k=1}^{m+1} a_{n+1-k} Z_k(t-1) + (a_{n-m} + c_{n-m}) \epsilon(t) \end{aligned}$$

$$+ \sum_{k=0}^n b_{n-m,k} z_{m+1-k} (t-1)\epsilon(t) + b_{n0}\epsilon(t)^2 ,$$

où par convention a_i, c_i, b_{ik} sont égaux à zéro si i est plus grand que p, q, Q respectivement, ou si k est supérieur à P . Alors le processus $X(t), t \in \mathbb{Z}$ satisfait les hypothèses A_1 et A_2 pourvu que $\epsilon(t)$ soit indépendant de $X(t-1), \dots$, et soit inversible.

Notons que si l'on écrit la représentation markovienne (5.2.4) sous la forme (5.2.1), les matrices ont alors pour expression :

$$A_{(n,n)} = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 & \dots & 0 \\ & 0 & 1 & & & & \\ & & & 1 & & & \\ & & & a_1 & & 1 & \\ & & & \vdots & & & \\ & & & \vdots & & & \\ a_n & a_{n-1} & & a_{n-m} & & & 0 \end{bmatrix} \quad B_{(n,n)} = \begin{bmatrix} & 0 & & & 0 \\ & b_1 & & b_{10} & \\ & \vdots & & \vdots & \\ & \vdots & & \vdots & \\ & b_{n-m,m} & & b_{n-m,0} & \\ & & & & 0 \end{bmatrix}$$

$$C = [0 \dots 1 \ a_1+c_1 \ \dots \ a_{n-m} + c_{n-m}]'$$

$$D = [0 \dots 0 \ b_{10} \ \dots \ b_{n-m,0}]'$$

$$H = [1 \ \dots \ 0] .$$

Dans le cas diagonal, on a $m = 0$.

3. Représentation ℓ -markovienne des modèles sousdiagonaux

On va étudier les modèles sousdiagonaux à partir de la représentation (5.2.2) où $b_{jk} = 0$ pour tous les couples (j,k) pour lesquels $k > j$. On fait d'autre part l'hypothèse suivante :

(H₀) $b_{jj+k} = 0$ pour tous les couples (j,k) pour lesquels $j-k \neq \ell$ où ℓ est un entier fixe donné.

Avec cette hypothèse on s'intéresse uniquement aux modèles sous-diagonaux pour lesquels le décalage entre X et ϵ dans les termes produits est constant.

A partir de l'hypothèse (H₀), l'équation (5.2.2) peut se réécrire

$$(5.2.5) \quad X(t) = \sum_{i=1}^p a_i X(t-i) + \epsilon(t) + \sum_{j=1}^q c_j \epsilon(t-j) + \sum_{j=1}^Q b_{jj+\ell} X(t-j+\ell) \epsilon(t-j)$$

On cherche à obtenir une représentation semblable à celle obtenue précédemment pour ces modèles à partir de la démarche suivie par T. PHAM DINH pour les modèles superdiagonaux. Pour cela on choisit un espace des prédicteurs de dimension finie :

$$\mathcal{P}'_t = \mathcal{P}_t \cup \{X(t-m), m=0, \dots, p-\max(p, Q)+1\} \cup \\ \left\{ \epsilon(t), \sum_{m=1}^{\ell} \epsilon(t) \epsilon(t-1) \dots \epsilon(t-m), \text{ constantes} \right\}.$$

Cet espace est en fait l'espace $\mathcal{P}_t(m)$ défini précédemment augmenté d'un certain nombre de variables aléatoires, $\epsilon(t), \epsilon(t)\epsilon(t-1), \dots, \epsilon(t)\epsilon(t-1)\dots\epsilon(t-\ell)$.

Enfin on choisit comme vecteur d'état, $Z(t) = (Z_1(t), \dots, Z_n(t))$. Les $Z_1(t), \dots, Z_n(t)$ sont des éléments de \mathcal{P}'_t construits de la manière suivante : si $r = \max(Q, p-\max(q, Q))$ et $s = n-r$, on pose :

$$Z_k(t) = X(t+k-r) \quad k = 1, \dots, r \\ Z_{r+k}(t) = \sum_{i=k}^s a_i X(t+k-i) + \sum_{j=k}^s [c_j \epsilon(t+k-j) + T_{t+k, j}] \quad k = 1, \dots, s$$

où $T_{t+k,j} = b_{j,j+\ell} E[X(t+k-j+\ell)\epsilon(t+k-j)/\mathcal{F}_t]$.

En calculant $Z_k(t)$ pas à pas pour $k = 1, 2, \dots, r+s$, on obtient le théorème suivant :

Théorème 5.2.2. (GUEGAN, 1986, p. 278)

Le processus $X(t)$, $t \in \mathbb{Z}$ défini par l'équation (5.2.5) admet la représentation ℓ -markovienne suivante :

$$\begin{aligned} X(t) &= Z_r(t-1) + \epsilon(t) \\ Z_k(t) &= Z_{k+1}(t-1) \quad 1 \leq k < r \\ Z_r(t) &= Z_{r+1}(t-1) + \epsilon(t) \\ Z_{r+k}(t) &= a_{k,r+k} Z_{r+k}(t-1) + Z_{r+k+1}(t-1) + (a_k + c_k)\epsilon(t) + b_{k,k+\ell} W_t(\ell) \\ & \quad k = 1, \dots, s-1 \end{aligned}$$

$$Z_{r+s}(t) = \sum_{i=s}^n a_i Z_{r+s+1-i}(t-1) + (a_s + c_s)\epsilon(t) + b_{s,s+\ell} W_t(\ell)$$

où

$$W_t(\ell) = \sum_{1 \leq \lambda_1 + \lambda_2 \leq \ell} \sum_{n=0}^{\lambda} \sum_{i_1 \dots i_{\lambda_1}=0}^p \sum_{j_1 \dots j_{\lambda_2}=0}^Q a_{i_1} \dots a_{i_{\lambda_1}}$$

$$b_{j_1 j_1 + \ell} \dots b_{j_{\lambda_2} j_{\lambda_2} + \ell} \cdot Z_{t+n\ell+1}(t-1) \epsilon(t) \prod_{\mu=1}^m \epsilon(t+n\ell-\mu) +$$

$$P(\epsilon(t), \dots, \epsilon(t-\ell)) ,$$

avec

$$P_t(\epsilon(t), \dots, \epsilon(t-\ell)) =$$

$$\sum_{1 \leq \lambda_1 + \lambda_2 \leq \ell} \sum_{n=0}^{\lambda} \sum_{i_1 \dots i_{\lambda_1}=0}^p \sum_{j_1 \dots j_{\lambda_2}=0}^Q a_{i_1} \dots a_{i_{\lambda_1}} b_{j_1 j_1 + \ell} \dots b_{j_{\lambda_2} j_{\lambda_2} + \ell} .$$

$$\left[\sum_{k \geq n\ell - m} c_k \prod_{\mu=1}^m \epsilon(t+n\ell-\mu-k) \epsilon(t) - c_{n\ell-m} \sigma^2 \prod_{\mu=1}^m \epsilon(t+n\ell-\mu) - c_{n\ell-m} \sigma^2 \right],$$

où, par convention

$i_0 = 1$, $j_{01} = 1$, et où $m = i_1 + \dots + i_{\lambda_1} + j_1 + \dots + j_{\lambda_2}$ est un entier strictement positif et plus grand que ℓ .

Remarque

Dans l'expression de $W_t(\ell)$ apparaît un produit de termes en $\epsilon(t-j)$. Ce produit contient au plus ℓ termes, où rappelons le, ℓ représente le décalage entre l'observation X et le bruit ϵ . On a alors pensé à classifier les modèles sousdiagonaux en fonction de ℓ . De part le développement obtenu, on peut remarquer que plus grand est le décalage entre X et ϵ dans le produit $X\epsilon$, plus grand est le nombre maximum de termes dans le produit

$\prod_j \epsilon(t-j)$. C'est pourquoi on a choisi d'appeler une telle représentation, une représentation ℓ -markovienne.

B. Modèles Affines en l'état polynomiaux en les entrées et Modèles autorégressifs généralisés à coefficients aléatoires

Le Théorème 5.2.2 a permis de mettre en évidence le comportement structural des modèles sousdiagonaux : le fait qu'à un instant t il y ait accumulation des bruits sous forme d'un produit du type $\prod_{j=1}^{\ell} \epsilon(t-j)$ conduit à envisager des modèles markoviens qui sont l'analogue des modèles polynomiaux en les entrées, affines en l'état considérés par les automaticiens (D. GUEGAN, 1986).

Un processus $X(t)$, $t \in \mathbb{Z}$ est dit gouverné par un modèle polynomial en l'entrée affine en l'état, s'il est défini par les équations suivantes :

$$(5.2.6) \quad \begin{cases} (a) & Z(t+1) = A(\epsilon(t))Z(t) + H(\epsilon(t)) \\ (b) & X(t) = B(\epsilon(t))Z(t-1) + C(\epsilon(t)) \end{cases}$$

avec les hypothèses suivantes :

(H₁) $Z(t) = [Z_1(t), \dots, Z_n(t)]$ est l'état du système. $Z(0)$ est donné et est indépendant de $A(\epsilon(t))$, $H(\epsilon(t))$ pour tout $t = 1, 2, \dots$.

(H₂) $\epsilon(t)$, $t \in \mathbb{Z}$ est une suite de variables aléatoires indépendantes, équidistribuées, centrées et de variance σ^2 .

(L'hypothèse de gaussianité pour $\epsilon(t)$ n'est pas nécessaire, seule l'existence de moments finis d'ordre supérieur ou égal à 2 est nécessaire).

(H₃) A , B , H sont respectivement des matrices carrée (n,n) , ligne $(n,1)$, colonne $(1,n)$, polynomiales en $\epsilon(t)$ de degré fini.

(H₄) $C(\epsilon(t))$ est un polynôme en $\epsilon(t)$ et de degré fini. On supposera en fait que $E[C(\epsilon(t))] = 0$, mais cette hypothèse peut être levée.

Notons que ce type de représentation où n apparaît à un temps donné t que des puissances finies de $\epsilon(t)$ a l'avantage d'être markovienne contrairement à la représentation précédemment considérée, (paragraphe II.A.3).

T. PHAM DINH (1986) a établi que tout modèle bilinéaire admet une représentation markovienne (5.2.6). Pour cela il augmente suffisamment l'espace des prédicteurs. La démarche est la suivante.

Il réécrit les modèles bilinéaires (5.2.2) sous la forme suivante :

$$\begin{aligned}
 X(t) = & \sum_{i=1}^p a_i X(t-i) + \epsilon(t) + \sum_{j=1}^q c_j \epsilon(t-j) \\
 & + \sum_{j=1}^Q \sum_{k=0}^P b_{jk} X(t-j-k) \epsilon(t-j) + \sum_{j'=1}^{Q'} \sum_{k'=1}^{P'} b'_{j'k'} X(t-j') \epsilon(t-j'-k').
 \end{aligned}$$

Il rajoute autant de zéros que nécessaires dans la définition des coefficients a_i , c_j , b_{ij} , b'_{ij} quand ceux-ci ne sont pas définis. La dimension de l'espace des prédicteurs est alors $n = \max(p, P+q, P+Q, P+Q')$. On pose $m = n - \max(q, Q, Q')$ et on note $Z^{(k)}(t)$ les vecteurs dont les composantes sont définies de la manière suivante:

$$\begin{aligned}
 Z_i^{(0)}(t) &= X(t-m+i) \quad 1 \leq i \leq m \\
 Z_{m+i}^{(0)}(t) &= \sum_{k=1}^p a_k X(t+i-k) + \sum_{k=i}^{n-m} [c_k + \sum_{j=0}^P b_{jk} X(t+i-j-k)] \epsilon(t+i-k) \\
 &+ \sum_{k'=1}^{n-m} [\sum_{j'=1}^{P'} b'_{j'k'} \epsilon(t+i-j'-k')] X(t+i-k') \quad 1 \leq i \leq n-m.
 \end{aligned}$$

Alors

$$X(t) = Z_{m+1}^{(0)}(t-1) + \epsilon(t)$$

$$\begin{aligned}
 \text{et } Z^{(0)}(t) &= [A + B \epsilon(t)] Z^{(0)}(t-1) + C \epsilon(t) + D \epsilon^2(t) \\
 &+ F \epsilon^{(1)}(t-1) [Z_{m+1}^{(0)}(t-1) + \epsilon(t)]
 \end{aligned}$$

où A, B, C, D, F sont des matrices bien définies et où on a posé $\epsilon^{(k)}(t) = (\epsilon(t), \dots, \epsilon(t+k-P'))'$.

On pose encore

$$Z^{(1)}(t) = (Z^{(0)}(t), \epsilon^{(1)}(t), Z_{m+1}^{(0)}(t) \epsilon^{(1)}(t))' \quad \text{et}$$

$$Z^{(k)}(t) = (Z^{(k-1)}(t), Z^{(k-1)}(t) \otimes \epsilon^{(k)}(t))' \quad k = 1, \dots, P'$$

où (x, y) représente la concaténation des vecteurs x, y et $(x \otimes y)$ le vecteur de composantes $x_i y_j$.

Alors on obtient par récurrence la relation suivante :

$$Z^{(k-1)}(t) = P_k(t)Z^{(k)}(t-1) + Q_{k+1}(t) \quad k = 1, \dots, P',$$

où P_k et Q_k sont des matrices et des vecteurs polynomiaux en $\epsilon(t)$ de degré k et $k+1$ respectivement. Enfin si

$$Z(t) = Z^{(P')}(t) = (Z^{(P'-1)}(t), Z^{(P'-1)}(t)\epsilon(t)),$$

on obtient la relation (5.2.6) où $A(t)$ est une matrice polynomiale en $\epsilon(t)$ de degré $P'+1$, et $H(\epsilon(t))$ un vecteur polynomial en $\epsilon(t)$ de degré $P'+2$, $C(\epsilon(t))$ étant égal à $\epsilon(t)$.

T. PHAM DINH note que la relation (5.2.6) ainsi obtenue est analogue à la définition du modèle autorégressif à coefficients aléatoires étudiée par D.F. NICHOLLS et B.G. QUINN (1982), avec la différence que dans l'expression (5.2.6) la matrice aléatoire $A(\epsilon(t))$ est remplacée par une matrice indépendante de $H(\epsilon(t))$. Quand cette condition d'indépendance est supprimée, on obtient un modèle contenant les deux modèles précédents appelé modèle à coefficients aléatoires généralisés.

III - CONDITION D'EXISTENCE DE REPRESENTATIONS MARKOVIENNES

Nous allons maintenant donner les conditions d'existence des deux représentations markoviennes obtenues précédemment à savoir les représentations (5.2.1) et (5.2.6).

Dans le cas de la représentation markovienne (5.2.1) on a

la résultat suivant :

Théorème 5.3.1. (PHAM DINH, 1985)

On suppose que les $\epsilon(t)$ sont indépendants avec des moments d'ordre 4 finis. Une condition nécessaire et suffisante pour l'existence d'une solution stationnaire de (5.2.1) est que l'une des deux conditions suivantes soit vérifiée :

(i) L'équation matricielle $Q = AQA' + BQB' + \Sigma$ où Σ représente la matrice de covariance de $C\epsilon(t) + D\epsilon^2(t)$ admet une solution positive en Q .

(ii) La suite Q_k définie par :

$$Q_0 = 0$$

$$Q_{k+1} = AQ_kA' + BQ_kB'\sigma^2 + \Sigma$$

est convergente.

Alors sous les conditions ci-dessus, une solution stationnaire à (5.2.1) est donnée par :

$$(5.3.1) \begin{cases} X(t) = HZ(t-1) + \epsilon(t) \\ Z(t) = \sum_{j=0}^{\infty} [A+B\epsilon(t)] \dots [A+B\epsilon(t+1-j)] [C\epsilon(t-j) + D(\epsilon(t-j) - \sigma^2)] \end{cases}$$

avec les séries ci-dessus convergeant en moindres carrés et presque sûrement. La solution est unique si l'équation $Q = AQA' + BQB' + \sigma^2$ n'admet pas de solution positive non triviale.

On dispose d'une condition suffisante plus simple, semblable à celle établie par M. BHASKARA RAO, T. SUBBA RAO et A.M. WALKER (1983, Théorème 4.3.4) pour les modèles bilinéaires superdiagonaux définis par l'équation (5.2.2).

Théorème 5.3.2. (PHAM DINH, 1985)

Si la matrice $A^{\otimes 2} + \sigma^2 B^{\otimes 2}$ a toutes ses valeurs propres de

module plus petit que l'unité, alors la série (5.3.1) est convergente en moyenne, et si les $\epsilon(t)$ sont indépendants la série est convergente presque sûrement et définit la solution unique de (5.2.1).

Dans le cas de la représentation polynomiale en les entrées affine en l'état (5.2.6), nous posons les notations suivantes : si M et N sont deux matrices carrées, alors on a :

$$E[M(\epsilon(t)) \otimes N(\epsilon(t))] = \widetilde{M \otimes N}, \quad E[M(\epsilon(t))N(\epsilon(t))] = \widetilde{MN}$$

$$E[M(\epsilon(t))] = \widetilde{M} \quad \forall t, \quad \text{où } \widetilde{M \otimes N}, \quad \widetilde{MN} \text{ et } \widetilde{N} \text{ sont des fonctions de } \sigma.$$

Pour simplifier les notations on omet σ . \otimes représente le produit de Kronecker des matrices (cf. H. NEUDECKER, 1969). Etudions tout d'abord les conditions de stationnarité des processus $X(t)$ vérifiant les relations (5.2.6). Posons $\mu_t = E[Z(t)]$ et $W_{t,s} = E[Z(t)Z'(s)]$. On a alors le théorème suivant :

Théorème 5.3.3. (GUEGAN, 1987, p. 290)

Le processus $Z(t)$, $t \in \mathbb{Z}$, $t > 0$ défini par les équations (5.2.6) est stationnaire au second ordre si et seulement si $\mu_1 = \mu_0$ et $W_{1,1} = W_{0,0}$.

On en déduit le corollaire suivant :

Corollaire 5.3.1. (GUEGAN, 1987, p. 291)

1. Le processus $Z(t)$, $t \in \mathbb{Z}$ est stationnaire au second ordre si et seulement si $\mu = E[Z(t)]$ et $W = E[Z(t)Z'(t)]$ satisfont :

$$(5.3.2) \quad \mu = \widetilde{A}\mu + \widetilde{H}$$

$$(5.3.3) \quad \text{vec } W = \widetilde{A \otimes A} \text{ vec } W + \Delta' \quad \text{avec}$$

$$(5.3.4) \quad \Delta' = (\widetilde{A \otimes H} + \widetilde{H \otimes A})_{\mu} + \widetilde{H \otimes H} .$$

2. Le processus $X(t)$, $t \in \mathbb{Z}$ est stationnaire au second ordre si et seulement si $m = E[X(t)]$, et $V = E[X(t)X(t)']$ satisfont $m = \widetilde{B}\mu$ où μ est donnée par (5.3.2) et $\text{vec } V = \widetilde{B \otimes B} \text{vec } W + \Delta$ où W vérifie (5.3.3) et où

$$(5.3.5) \quad \Delta = 2 \widetilde{B \otimes C} + \widetilde{C \otimes C} .$$

Pour l'existence de tels processus on dispose du théorème suivant :

Théorème 5.3.4. (GUEGAN, 1986, p. 292)

On suppose que les hypothèses (H_1) - (H_4) sont vérifiées et que $E[H(\epsilon(t))] = 0$, alors une condition nécessaire et suffisante pour l'existence d'une solution stationnaire de (5.2.6) est que l'une des deux conditions suivantes soit vérifiée :

(i) L'équation matricielle $\text{vec } W = \widetilde{A \otimes A} \text{vec } W + \widetilde{H \otimes H}$ admet une solution positive en W .

(ii) La suite W_k définie par

$$W_0 = W ,$$

$$W_k = E[A(\epsilon(t))W_{k-1}A'(\epsilon(t))] + \Delta_1$$

où

$$\Delta_1 = E[A(\epsilon(t))Z(t)H'(\epsilon(t)) + H(\epsilon(t))Z'(t)A'(\epsilon(t)) + H(\epsilon(t))H'(\epsilon(t))]$$

est convergente.

Sous les conditions précédentes, une solution stationnaire de (5.2.6) est donnée par :

$$X(t) = B(\epsilon(t))Z(t-1) + C(\epsilon(t))$$

$$(5.3.2) \quad Z(t) = \sum_{j \geq 0} P_{t, j-1} H(\epsilon(t-j-1)) \quad \text{où}$$

$$P_{t, r} = \prod_{k=1}^r A(\epsilon(t-k)) \quad , \quad r > 0 .$$

La solution est unique si l'équation $W = E[A(\epsilon(t))WA'(\epsilon(t))]$ n'admet pas de solution positive non triviale.

On peut remarquer que la version de ce théorème est très semblable à celle donnée dans le cas de la représentation markovienne des modèles bilinéaires superdiagonaux. Notons que le Théorème 5.3.4 concerne les modèles centrés et qu'il peut être étendu au cas des modèles non centrés. On a alors la version suivante :

Théorème 5.3.4'. (GUEGAN, PHAM DINH, 1987, p. 301)

Une condition nécessaire et suffisante d'existence d'une solution stationnaire au second ordre de (5.2.6), c'est-à-dire l'existence de deux processus stationnaires au second ordre satisfaisant (5.2.6) est que :

l'équation $\mu = (E[A(\epsilon(t))]) \mu + E[H(\epsilon(t))]$ admette une solution en μ , et que l'une des conditions suivantes soit vraie :

(i) l'équation matricielle $\text{vec } W = \widetilde{A \otimes A} \text{vec } W + \Delta$ où

$$\Delta = E[A(\epsilon(t))\mu H'(\epsilon(t)) + H(\epsilon(t))\mu' A'(\epsilon(t)) + H(\epsilon(t))H'(\epsilon(t))]$$

admet une solution $W \geq \mu\mu'$.

(ii) la suite W_k définie par $W_0 = \mu\mu'$, $\text{vec } W_k = \widetilde{A \otimes A} \text{vec } W_{k-1} + \Delta$ est convergente.

Sous ces conditions une solution stationnaire est donnée par (5.3.2). La solution est unique si $I - E[A(\epsilon(t))]$ est

inversible et si l'équation $W = E[A(\epsilon(t))W'A'(\epsilon(t))]$ n'admet pas de solution positive non triviale.

On peut remarquer que μ et W sont respectivement la moyenne et la covariance du processus $Z(t)$ défini par les équations (5.2.6). Si on suppose que $E[H(\epsilon(t))] = 0$, alors on peut prendre $\mu = 0$ et on se retrouve dans les conditions du théorème 5.3.4. Si $\mu \neq 0$, on peut toujours se ramener au cas $\mu = 0$ en travaillant avec le processus $\tilde{Z}(t) = Z(t) - \mu$. La démonstration du théorème 5.3.4' se fait alors de la même façon que pour le théorème 5.3.4 à partir du processus $\tilde{Z}(t)$ défini par :

$$\tilde{Z}(t) = A(\epsilon(t))\tilde{Z}(t-1) + \xi(\epsilon(t))$$

où

$$\xi(\epsilon(t)) = (A(\epsilon(t)) - E[A(\epsilon(t))])\mu + H(\epsilon(t)) - E[H(\epsilon(t))]$$

est un processus centré.

On dispose comme dans le cas précédent d'une condition suffisante d'existence, pour les modèles polynomiaux en les entrées affines en l'état.

Théorème 5.3.5. (GUEGAN, 1987, p. 293)

Si les matrices \tilde{A} et $\tilde{A} \otimes \tilde{A}$ ont toutes leurs valeurs propres de module plus petit que l'unité, alors la série (5.3.2) est convergente en moyenne, et si la suite $\epsilon(t)$, $t \in \mathbb{Z}$ vérifie l'hypothèse (H_2) alors la série est convergente presque sûrement et définit une solution stationnaire unique de (5.2.6).

Les conditions données dans ce théorème généralisent les conditions obtenues par D.F. NICHOLLS et B.G. QUINN (1982) dans le cas des modèles autorégressifs à coefficients aléatoires.

IV - PROPRIETES DES REPRESENTATIONS MARKOVIENNES

Nous allons analyser un certain nombre de propriétés probabilistes des modèles bilinéaires, propriétés étudiées à partir des représentations markoviennes (5.2.1) et (5.2.6).

1.- Minimalité des Représentations Markoviennes

Dans la construction d'une représentation markovienne on a intérêt à faire en sorte que la dimension de l'espace d'état soit la plus petite possible (pour la mémorisation et les calculs). Ce problème est lié à la notion de minimalité de la représentation. Ce sont les contrôleurs qui, les premiers, ont développé cette problématique pour les systèmes dynamiques. Suivant R.E. KALMAN, P.L. FALB et M.A. ARBIB (1969), ou P.FAURRE, M.CLERGET et F. GERMAIN (1979), nous disposons de la définition suivante:

Définition 5.4.1.

Etant donné une représentation interne d'un système dynamique, celle-ci est minimale s'il n'en existe pas d'autre (représentant le même système) pour laquelle l'espace d'état soit de dimension inférieure.

H.AKAIKE (1974 c) a étendu cette notion au cas des systèmes stochastiques à partir de leur réalisation markovienne. L'obtention de la représentation minimale d'une représentation markovienne dans le cas d'un modèle ARMA est un élément déterminant dans la théorie de l'identification. On sait en particulier que dans le cas d'une représentation minimale le modèle peut être uniquement identifié à une transformation de base de l'espace d'état près, à partir de la donnée de la fonction de covariance du processus observé. Le

problème est alors de sélectionner un modèle unique parmi ceux identifiés précédemment en imposant les contraintes adéquates.

Dans le cas de la représentation markovienne d'un modèle bilinéaire on dispose de critères permettant d'assurer une propriété un peu plus faible que la minimalité définie ci-dessus, toutefois il n'y a pas de lien explicite de la minimalité avec l'unicité de la représentation markovienne.

Définition 5.4.2

On dit qu'une représentation est quasiminimale s'il n'existe pas d'autre représentation définie à partir du même bruit et dont le vecteur d'état soit la transformation linéaire du vecteur d'état d'origine avec une dimension plus petite.

Notons que si on se restreint aux représentations inversibles alors cette notion est la même que la notion de minimalité donnée dans la définition 5.4.1. On dispose alors du théorème suivant dans le cas de la représentation (5.2.1):

Théorème 5.4.1. (PHAM DINH, 1985)

On suppose que la représentation (5.2.1) est régulière (c'est-à-dire que le processus $X(t)$, $t \in \mathbb{Z}$ est donné par (5.3.1)). Soit:

n la dimension de l'espace d'état,

Q_k la suite définie dans le théorème 5.3.1, et

P_k la suite définie par :

$$P_0 = 0 ,$$

$$P_{k+1} = A'P_kA + B'P_kB\sigma^2 + H'H ,$$

C_k et L_k les matrices formées par les colonnes de $\begin{matrix} A^{n_1} & B^{m_1} & & & \\ & & \dots & & \\ & & & A^{n_r} & B^{m_r} \end{matrix} \Sigma$ et par les vecteurs lignes $\begin{matrix} HA^{n_1} & B^{m_1} & & & \\ & & \dots & & \\ & & & HA^{n_r} & B^{m_r} \end{matrix}$, où Σ est introduit dans le théorème 5.3.1, et où $n_1, m_1, \dots, n_r, m_r$ sont des entiers non négatifs tels que

$$\sum_{i=1}^n (n_i + m_i) \leq k.$$

Alors la représentation (5.2.1) est quasiminimale si et seulement si les conditions (C) et (0) sont vérifiées,

où (C) est l'une des conditions suivantes :

(C₁) Q_k est non singulière pour tout k ≤ n ,

(C₂) C_k est de rang n pour tout k ≤ n ,

et (0) est l'une des conditions suivantes :

(O₁) P_k est non singulière pour tout k ≤ n ,

(O₂) L_k est de rang n pour tout k ≤ n .

On a aussi le même type de résultat pour la représentation (5.2.8):

Théorème 5.4.2. (GUEGAN, PHAM DINH, 1987 a, p. 302)

Soit X(t), t ∈ Z le processus défini par les équations (5.2.6), p la dimension de l'espace d'état, Δ et W_k définis comme dans le théorème 5.3.4', P_k la suite définie par :

$$P_{k+1} = E[A(\epsilon(t))P_k A'(\epsilon(t))] + D$$

où

$$D = E[B'(\epsilon(t))B(\epsilon(t))] .$$

Soit enfin la décomposition de A(ε(t)) suivant les polynômes d'Hermite h_j :

$$A(\epsilon(t)) = \sum_{j=0}^r A_j h_j(\epsilon(t)) .$$

Alors la représentation (5.2.6) est quasiminimale si et seulement si l'une des conditions équivalentes de (a) et l'une des conditions équivalentes de (b) est réalisée.

(a) (i) W_k est non singulière pour k ≤ p

(ii) la matrice C_k formée par les colonnes des matrices Δ, A_{i₁} ... A_{i_n} Δ, 1 ≤ n ≤ k est de rang p, pour un k ≤ p.

(b) (i) P_k est non singulière pour k ≤ p

- (ii) la matrice L_k formée par les lignes des matrices D , $DA_{i_1} \dots A_{i_n}$, $1 \leq n \leq k$ est de rang p , par un $k \leq p$.

Notons que la décomposition de $A(\cdot)$ suivant les polynômes d'Hermite dans le théorème 5.4.2 est introduite pour faciliter la démonstration, mais que les résultats restent vrais pour toute décomposition de $A(\cdot)$ suivant une base de polynômes.

2. Inversibilité d'une représentation markovienne

Nous avons déjà vu l'intérêt de la notion d'inversibilité dans les chapitres précédents, en particulier pour l'estimation des paramètres car elle permet d'exprimer le bruit $\epsilon(t)$ à partir des observations passées du processus. Nous avons aussi noté, dans le paragraphe précédent, son intérêt par rapport à la minimalité. Nous allons établir ici l'inversibilité d'une sous-classe de modèles bilinéaires admettant la représentation markovienne (5.2.1). Pour cela nous allons utiliser la définition faible de l'inversibilité telle qu'elle a été définie au chapitre III, définition 3.3.2, et donner un algorithme permettant de construire la suite $\hat{\epsilon}(n)$. On considère un processus $X(t)$, $t \in \mathbb{Z}$, qui a pour représentation markovienne une forme un peu plus simple que l'expression (5.2.1), soit :

$$(5.4.1) \quad \begin{cases} X(t) = HZ(t-1) + \epsilon(t) \\ Z(t) = [A+B\epsilon(t)]Z(t-1) + C\epsilon(t) + D\epsilon^2(t) \end{cases}$$

où A et B sont des matrices carrées, H une matrice ligne, C et D des matrices colonne, et on suppose que ces matrices vérifient la condition " $B = DH$ ". On peut remarquer que cette relation est vérifiée pour les modèles diagonaux. Si l'on considère les modèles diagonaux définis au chapitre IV par la relation (4.1.1) avec la condition (H_0) , la représentation markovienne de ces modèles a pour matrices A, B, C, D, H (cf. théorème 5.2.1) :

$$A_{(n,n)} = \begin{bmatrix} a_1 & 1 & 0 & \dots & 0 \\ a_2 & 0 & 1 & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ a_n & 0 & 0 & \dots & \cdot \end{bmatrix} \quad B_{(n,n)} = \begin{bmatrix} b_{11} \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ b_{nn} \end{bmatrix} \quad 0$$

$$C_{(1,n)} = [a_1 + c_1 \quad \dots \quad a_n + c_n]'$$

$$D_{(1,n)} = [b_{11} \quad \dots \quad b_{nn}]'$$

$$H = [1 \quad \dots \quad 0]$$

Dans le cas des modèles que nous considérons, une construction naturelle de la suite $\hat{\epsilon}(n)$ est la suivante: étant donné une valeur arbitraire z_0 de $Z(0)$, on définit $\epsilon(t/z_0)$, $Z(t/z_0)$ récursivement à partir des équations du modèle (5.4.1), on obtient:

$$(5.4.2) \quad \begin{cases} \epsilon(t/z_0) = X(t) - HZ(t-1/z_0) \\ Z(t/z_0) = [A+B\epsilon(t/z_0)]Z(t-1/z_0) + C\epsilon(t/z_0) + D\epsilon^2(t/z_0) \end{cases}$$

avec $Z(0/z_0) = z_0$. Etant donné une norme vectorielle φ , on introduit la norme opérateur associée définie par:

$$\|K\|_{\varphi} = \sup_{\varphi(x) \leq 1} \varphi(Kx) \text{ pour une matrice carrée } K.$$

On a alors le théorème suivant:

Théorème 5.4.3. (GUEGAN, PHAM DINH, 1987 a, p. 307)

S'il existe une norme vectorielle φ telle que

$$(5.4.3) \quad E[\log \|K(t)\|_{\varphi}] < 0 \quad \text{où } K(t) = A - CH - DHX(t),$$

alors $\hat{\epsilon}(t/z_0) - \epsilon(t) \rightarrow 0$ presque sûrement quand $t \rightarrow \infty$ quelque soit z_0 .

En pratique les paramètres des matrices A,C,D,H ne sont pas connus, aussi les calculs dans le système récuratif (5.4.2) doivent être faits à l'aide de valeurs hypothétiques \hat{A} , \hat{C} , \hat{D} , \hat{H} . On obtient le théorème suivant qui renforce le théorème 5.4.3, et qui peut être prouvé de la même manière.

Théorème 5.4.3'. (GUEGAN, PHAM DINH, 1987 b, p. 308)

Soit \hat{A} , \hat{C} , \hat{D} , \hat{H} tels que pour une norme vectorielle φ :

$$E[\log \|\hat{K}(t)\|_{\varphi}] < 0 \quad \text{où} \quad \hat{K}(t) = \hat{A} - \hat{C}\hat{H} - \hat{D}\hat{H} X(t).$$

Alors la série

$$(5.4.4) \quad \sum_{j=0}^{\infty} \prod_{k=0}^{j-1} \hat{K}(t-k) [\hat{C} X(t-j) + \hat{D} X^2(t-j)]$$

converge presque sûrement, et sa différence avec $\hat{Z}(t/z_0)$ donné par (5.4.2) avec les paramètres \hat{A} , \hat{C} , \hat{D} , \hat{H} à la place de A, C, D, H tend vers zéro p.s., quand $t \rightarrow \infty$, quelque soit z_0 .

Notons que l'espérance mathématique utilisée dans le théorème 5.4.3' est supposée calculée pour le vrai modèle qui n'est pas nécessairement celui défini par \hat{A} , \hat{C} , \hat{D} , \hat{H} .

Remarques.

1.- Le théorème 5.4.3' est très utile en vue de faire de l'estimation des paramètres en particulier par la méthode des moindres carrés, (cf. chapitre VI).

2.- La condition (5.4.3) obtenue peut paraître difficile à exploiter, néanmoins elle donne un domaine d'inversibilité. La difficulté réside dans l'étude du comportement du produit de

matrices aléatoires $\hat{K}(t-k)$. H.KESTEN et F.SPITZEK (1978) ont étudié de tels produits, malheureusement leurs résultats ne sont pas applicables ici car les matrices ne sont pas indépendantes. Dans le cas scalaire, on se ramène à des résultats du type de ceux obtenus par T. PHAM DINH et L. TAT TRAN (1981), (cf. chapitre IV, théorème 4.4.3).

3.- Il est important de noter que la démonstration du théorème 5.4.3 repose essentiellement sur la condition $B = DH$, et donc, pour l'instant, on ne dispose de résultats que pour les modèles diagonaux.

V - APPLICATION DE LA REPRESENTATION MARKOVIENNE A L'ETUDE DE MODELES BILINEAIRES.

1. Ergodicité et ergodicité géométrique pour les représentations markoviennes.

Au chapitre III nous avons rappelé les définitions d'ergodicité et d'ergodicité géométrique qui sont importantes dans le cas de variables dépendantes afin de pouvoir utiliser des théorèmes du type de la loi des grands nombres.

H. TONG (1981) a établi l'ergodicité du processus markovien $X(t)$, $t \in \mathbb{Z}$ défini par l'équation suivante :

$$(5.5.1) \quad X(t) = aX(t-1) + bX(t-1)\epsilon(t) + \epsilon(t)$$

où $\epsilon(t)$ est un bruit blanc gaussien, à partir d'un théorème de R.L.TWEEDIE (1975), (cf. chapitre III, théorème 3.3.1). D. GUEGAN (1983) a ensuite établi le résultat suivant :

Théorème 5.5.1 (GUEGAN, 1983, p. 254)

Etant donné le processus markovien réel $Z(t)$ défini par :

$$(5.5.2) \quad Z(t) = aZ(t-1) + bZ(t-1)\epsilon(t) + \epsilon(t) + d(\epsilon^2(t)-1)$$

où $\epsilon(t)$ est un bruit blanc gaussien centré, de variance $\sigma^2 = 1$, une condition suffisante d'ergodicité est donnée par $y < 1$ où $y = E|a+b\epsilon(t)|$.

Ce résultat se généralise au cas des processus markoviens vectoriels suivants (cf. D. GUEGAN, 1983, théorème 2) :

$$(5.5.3) \quad Z(t) = AZ(t-1) + BZ(t-1)\epsilon(t) + C\epsilon(t) + D(\epsilon^2(t)-1) ,$$

où $\epsilon(t)$ est un bruit blanc gaussien, $A = (a_{ij})_{1 \leq i, j \leq n}$ et $B = (b_{ij})_{1 \leq i, j \leq n}$ sont des matrices (n, n) .

P.D. FEIGIN et R.L. TWEEDIE (1985) ont établi l'ergodicité géométrique du processus (5.5.1) à partir d'un résultat établi par R.L. TWEEDIE (1983 a), (cf. chapitre III, théorème 3.3.2), dans le cas où $\epsilon(t)$ est un bruit blanc gaussien. On peut noter que ce résultat peut se généraliser aux processus (5.5.2) et (5.5.3). D'autre part dans le cas où $\epsilon(t)$ n'est pas gaussien, une condition suffisante pour avoir la φ -irréductibilité de la chaîne (5.5.1) pour une certaine mesure φ σ -finie, est obtenue en supposant que la distribution de $\epsilon(t)$ a une densité positive sur l'intervalle $(-\eta, +\eta)$ ($\eta > 0$) : φ peut alors être une mesure uniforme sur cet intervalle et zéro ailleurs.

T. PHAM DINH (1986) a généralisé ces résultats et établi l'ergodicité géométrique pour le processus $Z(t)$, $t \in \mathbb{Z}$ vérifiant les équations (5.2.6.a). Ce processus est clairement markovien avec

une probabilité de transition $P(z, \cdot)$ égale à la distribution de $A(\epsilon(t))z + H(\epsilon(t))$. Afin d'appliquer le théorème 3.3.2 du chapitre III, on doit prouver que le processus markovien $Z(t)$, $t \in \mathbb{Z}$ est de Feller et φ -irréductible pour une mesure φ σ -finie, et les critères du théorème 3.3.2. Le premier point est facile à établir. Quant au dernier point il faut définir une bonne fonction g . Une candidate naturelle semble être $g(z) = 1 + \varphi(z)^s$ où φ est une certaine norme vectorielle et s un nombre strictement positif. Soit les notations suivantes : Si Z est un vecteur aléatoire et A une matrice aléatoire, on pose :

$$\|Z\|_{\varphi, s} \text{ la norme } L^s \text{ de } \varphi(Z) \text{ et}$$

$$\|A\|_{\varphi, s} = \sup_{z \neq 0} \frac{\|Az\|_{\varphi, s}}{\varphi(z)}$$

Alors en prenant $K = \{z, \varphi(z) < M\}$ pour un M suffisamment grand et en utilisant l'inégalité triangulaire, le théorème 3.3.2 permet d'obtenir le théorème suivant :

Théorème 5.5.2. (PHAM DINH, 1986)

Supposons que la chaîne de Markov (5.2.6.a) soit φ -irréductible avec le support de φ ayant un intérieur non vide, et que pour une norme vectorielle φ , et un $s > 0$, $\|A(\epsilon(t))\|_{\varphi, s} < 1$ et $\|H(\epsilon(t))\|_{\varphi, s} < \infty$, alors la chaîne est géométriquement ergodique.

Il est possible d'obtenir une condition suffisante permettant d'avoir l'ergodicité géométrique, et qui soit plus maniable que celle donnée à l'aide des normes. On procède de la manière suivante : On définit une condition portant sur les valeurs propres de l'espérance d'une certaine puissance tensorielle symétrique de $A(t)$. Pour cela, on définit $\{\mathbb{R}^d\}^{\otimes s}$ comme l'espace des vecteurs $u = \{u_{i_1 \dots i_s}, i_j = 1, \dots, d, j = 1, \dots, s\}$ qui sont invariants pour toute permutation portant sur les indices des composantes. Si $A = (A_{ij})$ est une matrice carrée d'ordre d , la

$s^{\text{ième}}$ puissance tensorielle symétrique $A^{\otimes s}$ est définie comme l'opérateur sur $(\mathbb{R}^D)^{\otimes s}$ qui envoie le vecteur u sur le vecteur $A^{\otimes s}u = v$ où

$$v_{i_1 \dots i_s} = \sum_{j_1 \dots j_s} A_{i_1 j_1} \dots A_{i_s j_s} u_{j_1 \dots j_s} .$$

On a alors le théorème suivant :

Théorème 5.5.3. (PHAM-DINH, 1986)

Soit A une matrice aléatoire avec des moments finis d'ordre $2m$, alors $\|A\|_{\psi, 2m} < 1$ pour une norme vectorielle ψ , si $E[A^{\otimes 2m}]$ a ses valeurs propres de module strictement plus petit que 1. Pour $m = 1$, ψ^2 peut être choisi comme une forme quadratique, alors les conditions ci-dessus sont aussi nécessaires.

Il reste à établir maintenant que la chaîne (5.2.6) est ψ -irréductible, condition pas très facile à établir car la distribution jointe de $A(\epsilon(t))$ et de $H(\epsilon(t))$ n'a pas en général de densité par rapport à la mesure de Lebesgue : en effet $A(\epsilon(t))$ et $H(\epsilon(t))$ contiennent d'une part des éléments déterministes, d'autre part dépendent du même bruit $\epsilon(t)$. Des conditions suffisantes ont cependant été obtenues pour l'irréductibilité des modèles à coefficients aléatoires généralisés par T. PHAM DINH (1986).

La propriété d'ergodicité géométrique englobe la condition la plus forte que l'on puisse espérer obtenir pour un processus de Markov. Celle-ci englobe les propriétés comme Harris Récurrente, ergodicité et Mélange. Rappelons que la propriété Harris Récurrente avait été obtenue par D. GUEGAN (1984 b) pour des processus $Z(t)$, $t \in \mathbb{Z}$ vérifiant l'équation (5.2.6 a).

Ce qui vient d'être démontré précédemment concerne le processus d'état $Z(t)$, or on s'intéresse au processus d'observation

$X(t)$.

L'ergodicité géométrique étant une notion propre aux processus markoviens, ce qui n'est pas le cas du processus $X(t)$, on introduit donc une autre notion appelée absolue régularité définie de la façon suivante. Soit $Y(t)$, $t \in \mathbb{Z}$ un processus vectoriel stationnaire, on pose :

$$\Delta_n = \sup_{h \in S_n} E[h|Y(s), s < 1] - E[h]$$

où S_n représente l'ensemble des variables aléatoires mesurables pour la tribu engendrée par $Y(t)$, $t \geq n$, et bornées par 1 en valeur absolue. Le processus est dit absolument régulier si $E[\Delta_n]$ tend vers 0 quand n tend vers l'infini. Dans le cas d'un processus de Markov, l'ergodicité géométrique avec une vitesse de convergence exponentielle vers zéro de $E[\Delta_n]$, est équivalente à l'absolue régularité.

Le processus $Z(t)$, $t \in \mathbb{Z}$, est donc absolument régulier sous les conditions du théorème 5.5.1. D'autre part le processus $X(t)$, défini par les équations (5.2.6) est subordonné au processus $Z(t)$ dans le sens où la distribution conditionnelle de $(X(t), Z(t))$, étant donné $(X(s), Z(s))$, $s < t$ dépend uniquement de $Z(t-1)$, et l'on montre que cela implique que le processus $X(t)$, $t \in \mathbb{Z}$, est aussi absolument régulier avec une vitesse de convergence vers 0 de $E[\Delta_n]$, plus petite que celle du processus $Z(t)$.

2. Calcul des moments des représentations markoviennes

Grâce à la représentation markovienne, le calcul des moments peut se faire plus facilement que quand on utilise l'approche directe.

a. Modèles superdiagonaux

Pour les modèles superdiagonaux qui admettent la représentation (5.2.1), T. PHAM DINH (1985) a fourni l'expression des moments.

Soit $X(t)$, $t \in \mathbb{Z}$ un processus stationnaire défini par l'équation (5.2.1) et posons :

$$\Sigma = CC'\sigma^2 + DD'\text{var}[\epsilon^2(t)] + [CD'+D'C]E[\epsilon^3(t)].$$

Sa fonction de covariance est donnée par :

$$R(0) = E[X^2(t)] = HQH' + \sigma^2$$

où Q est la fonction de covariance de $Z(t)$, et

$$\begin{aligned} R(t) &= E[X(t+k)X(t)] = HA^{k-1} E[Z(t)X(t)] \\ &= HA^kQH' + HA^{k-1}C \sigma^2 + HA^{k-1}D E[\epsilon^3(t)], \quad k \geq 1 \end{aligned}$$

avec Q solution positive de

$$Q = AQA' + BQB' \sigma^2 + \Sigma.$$

Il est alors facile de voir que la fonction de covariance d'un processus admettant une représentation bilinéaire markovienne a la même forme que celle d'un processus ARMA, d'où le théorème suivant :

Théorème 5.5.4. (PHAM DINH, 1985)

La fonction de covariance du processus bilinéaire (5.2.3) est la même que celle d'un processus ARMA ayant pour coefficients

autorégressifs a_1, \dots, a_p , et pour coefficients moyenne mobile des fonctions de a_i, c_j et b_{ij} dont le nombre est égal à $\max(q, Q)$.

Il est aussi possible de calculer à partir de la représentation (5.2.1) les moments d'ordre supérieur pour ces modèles (cf. T. PHAM DINH, 1985).

H. TONG (1981) a donné pour le modèle markovien (5.5.1) l'expression des moments jusqu'à l'ordre 6. P.D. FEIGIN et R.L. TWEEDIE (1985) ont généralisé ce résultat à tous les moments d'ordre pair, à condition que :

$$(5.5.4) \quad \sum_{n=0}^k \binom{k}{n} a^{k-n} E[\varepsilon(t)^n] < 1, \text{ pour tout } k \text{ pair.}$$

Ce résultat est obtenu à partir du théorème suivant démontré par R.L. TWEEDIE (1983, b).

Théorème 5.5.5. (TWEEDIE, 1983 b)

Soit $X(t)$, $t \in \mathbb{Z}$ une chaîne de markov Harris récurrente, ergodique avec $\pi(x) = 1$ (où π est une mesure de probabilité sur l'espace d'états \mathcal{E}), et f une fonction non négative \mathcal{E} -mesurable. Supposons que pour un ensemble petit A , (on prend ici $\varphi = \pi$ dans la définition des petits ensembles, cf. chapitre IV) :

- (i) f soit bornée de zéro à l'infini sur A
- (ii) $\sup E[f(X(t)) | X(t-1) = x] < \infty \quad x \in A$
- (iii) il existe $\delta > 0$ tel que :
 $E[f(X(t)) | X(t-1) = x] \leq (1-\delta)f(x), \quad x \in A^c$
 (où δ est un nombre positif).

Alors :

$$\int_E f(x) \pi(dx) < \infty.$$

D'autre part, pour $\rho < 1$

$$\int_E \pi(dx) \sup_{|h| \leq f} |D[h(X(t)) | X(0)=x] - \int_E \pi(dy) h(y)| = O(\rho^t), \quad t \rightarrow \infty.$$

b. Modèles polynomiaux en les entrées affines en l'état

D. GUEGAN (1987) obtient l'expression des moments d'ordre $2k$ pour les modèles polynomiaux en les entrées affines en l'état.

On donne auparavant les conditions de stationnarité à l'ordre 1 et 2, les notations sont les mêmes que celles introduites au paragraphe précédent. M étant une matrice (n,n) , on désigne par $\rho(M)$ la plus grande valeur propre de la matrice M .

Théorème 5.5.6. (GUEGAN, 1987, p. 294)

Soit un processus $X(t)$, $t \in \mathbb{Z}$ défini par l'équation (5.2.6).

1.- Si $\rho(\tilde{A}) < 1$ alors le processus $X(t)$ est stationnaire au premier ordre et $m = \tilde{B}\mu$ avec $\mu = (I - \tilde{A})^{-1}\tilde{H}$.

2.- Si $W = E[Z(t)Z'(t)]$ satisfait l'équation (5.3.3) et si $\rho(\tilde{A} \otimes A) < 1$, alors le processus $X(t)$ est stationnaire au second ordre, et $V = E[X(t)X(t)']$ est donné par

$$\text{vec } V = \tilde{B} \otimes B \text{ vec } W + \Delta \quad \text{où } W = (I - \tilde{A} \otimes A)^{-1} \Delta' \quad \text{avec}$$

Δ et Δ' donnés respectivement par les équations (5.3.5) et (5.3.4).

On peut alors calculer la fonction de covariance. Posons :

$$C(s) = E[(Z(t+s) - \mu)(Z(t) - \mu)'] \quad \text{et}$$

$$\Gamma(s) = E[(X(t+s) - m)(X(t) - m)'] .$$

Comme

$$E[Z(t+s)Z'(t)] = \tilde{A}^s E[Z(t)Z'(t)] + \sum_{k=0}^{s-1} \tilde{A}^k \tilde{H} \mu$$

on a :

$$C(s) = \tilde{A}^s W + \sum_{k=1}^{s-1} \tilde{A}^k \tilde{H} \mu - \mu^2, \quad s > 0$$

$$C(0) = W - \mu^2$$

et

$$\Gamma(s) = \tilde{B} C(s) \tilde{B}' - m^2 \quad s \geq 0.$$

On peut remarquer que dans le cas où le processus est centré :

$$C(s) = \tilde{A}^s W, \quad s \geq 0 \quad \text{et} \quad \Gamma(s) = \tilde{B} \tilde{A}^s W \tilde{B}' \quad s \geq 0,$$

donc la fonction de covariance du processus centré $X(t)$, $t \in \mathbb{Z}$ qui a pour représentation markovienne (5.2.6), a la même forme que celle d'un processus ARMA. On retrouve ici un résultat déjà annoncé dans le cas des processus superdiagonaux admettant la représentation markovienne (5.2.1).

Afin de simplifier l'expression des moments d'ordre supérieur, on suppose le processus $X(t)$, $t \in \mathbb{Z}$ centré. Soit

$$\Omega_{k_1, \dots, k_j, k} = E[A(\epsilon)^{\otimes k_1 - 1} \otimes H(\epsilon) \otimes A(\epsilon)^{\otimes (k_2 - k_1 - 1)} \otimes \dots \otimes H(\epsilon) \otimes A(\epsilon)^{\otimes (k - 1)}]$$

où $1 \leq k_1 < k_2 < \dots < k_j \leq k$, $A^0 =$ matrice unité par convention et où pour des raisons de simplicité on a supprimé la variable t dans $\epsilon(t)$. Soit alors :

$$\Omega_{k-j, j} = \sum_{1 \leq k_1 < \dots < k_j \leq k} \Omega_{k_1, \dots, k_j, k} \quad \text{et}$$

$$\Lambda_{k_1, \dots, k_j, k} = E[B(\epsilon)^{\otimes (k_1 - 1)} \otimes B(\epsilon)^{\otimes (k_2 - k_1 - 1)} \otimes \dots \otimes B(\epsilon)^{\otimes (k - k_j)} C(\epsilon)^j]$$

$M(\epsilon)^{\otimes n} = M(\epsilon) \otimes \dots \otimes M(\epsilon)$ où le produit est fait n fois.

On a le théorème suivant :

Théorème 5.5.7. (GUEGAN, 1987, p. 295)

Le processus $X(t)$, $t \in \mathbb{Z}$ défini par (5.2.6) est stationnaire d'ordre k , avec $k = 2n$, si $\rho(\Omega_{2n,0}) < 1$. Alors en posant :

$$m_k = E[X(t)^k] \text{ et } \mu_k = E[Z(t)^{\otimes k}], \text{ on a :}$$

$$(5.5.5) \quad \text{vec } m_k = B^{\otimes k} \text{vec } \mu_k + \sum_{j=1}^k \Lambda_{k-j,j} \text{vec } \mu_{k-j}$$

$$(5.5.6) \quad \text{vec } \mu_k = \Omega_{k,0} \text{vec } \mu_k + \sum_{j=1}^k \Omega_{k-j,j} \text{vec } \mu_{k-j}$$

avec $\mu_0 = 1$.

Dans le cas de modèles simples, les moments peuvent être calculés explicitement.

3. Application au modèle ARCH(p)

Les modèles ARCH (en anglais Autoregressive Conditional Heteroscedastic) ont été construits par R.F. ENGLE (1982) dans le but d'étendre le concept de variance constante dans les modèles économiques. L'idée de base était d'introduire une variable qui prédisait la variance.

Le processus $X(t)$, $t \in \mathbb{Z}$, défini par R.F. ENGLE et appelé modèle ARCH(p), vérifie les équations suivantes :

$$(5.5.7) \quad \begin{cases} X(t) = \eta(t)V(t)^{1/2} \\ V(t) = \alpha_0 + \alpha_1 X^2(t-1) + \dots + \alpha_p X^2(t-p) \end{cases}$$

où $\eta(t)$ est un bruit blanc gaussien centré, de variance 1, et où les coefficients α_j , $j = 1, \dots, p$ sont positifs ou nuls. La distribution de $X(t)$ sachant les valeurs passées $X(t-i)$, $i \geq 1$, suit une loi normale centrée, de variance $V(t)$.

Dans son article, R.F. ENGLE donne une condition nécessaire et suffisante d'existence pour les moments d'ordre $2r$ du modèle ARCH(1) :

$$\alpha_1^2 \sum_{j=1}^r (2j-1) < 1 .$$

Il donne une condition nécessaire et suffisante de stationnarité au second ordre du modèle ARCH(p) et fournit l'expression de la variance du processus $X(t)$:

$$E[X^2(t)] = \alpha_0 / (1 - \sum_{j=1}^p \alpha_j) .$$

L'estimation des paramètres par la méthode du maximum de vraisemblance est aussi considérée dans son article ainsi que dans l'article de A.A. WEISS (1984).

A. MILHØJ (1985) fournit des conditions d'existence de moments d'ordre supérieur pour le modèle ARCH(p), ainsi que la fonction de covariance du processus $X^2(t)$, $t \in \mathbb{Z}$, $X(t)$ vérifiant les équations (5.5.7).

Nous allons montrer qu'à partir d'une réécriture du processus $X^2(t)$, on peut obtenir une représentation markovienne de ce processus. Il est alors possible de l'étudier à partir des

résultats obtenus dans les paragraphes précédents.

Avec un changement de paramètres, le modèle (5.5.7) peut s'écrire sous la forme :

$$\begin{cases} X(t) = \eta(t)V^{1/2}(t) \\ V(t) = \alpha_0 + \alpha_1(X^2(t-1) - \alpha_0) + \dots + \alpha_p(X^2(t-p) - \alpha_0) \end{cases}$$

Alors

$$X^2(t) = \eta^2(t) \left[\alpha_0 + \sum_{i=1}^p \alpha_i (X^2(t-i) - \alpha_0) \right].$$

Posons :

$$\begin{aligned} \eta^2(t) &= \epsilon(t) + 1 \\ X^2(t) &= Y(t) + \alpha_0. \end{aligned}$$

Alors :

$$Y(t) + \alpha_0 = (\epsilon(t) + 1) \left[\alpha_0 + \sum_{i=1}^p \alpha_i Y(t-i) \right].$$

Et l'on obtient la représentation markovienne suivante :

$$(5.5.8) \quad \begin{cases} X^2(t) = Y(t) + \alpha_0 \\ Y(t) = \sum_{i=1}^p \alpha_i Y(t-i) + \sum_{i=1}^p \alpha_i Y(t-i) \epsilon(t) + \alpha_0 \epsilon(t) \end{cases}$$

où $\epsilon(t)$ est un bruit blanc centré qui, à une constante près suit la loi du χ^2 à un degré de liberté. La représentation (5.5.8) peut s'écrire sous forme vectorielle :

$$(5.5.9) \quad \begin{cases} X^2(t) = HZ(t) + \alpha_0 \\ Z(t) = AZ(t-1) + BZ(t-1)\epsilon(t) + C\epsilon(t) \end{cases}$$

$$\text{où } Z(t) = [X(t), \dots, X(t-p)]'$$

$$H = [1, \dots, 0]$$

$$A = \begin{bmatrix} \alpha_1 & 1 & 0 & \cdot & \cdot & \cdot & 0 \\ \cdot & 0 & 1 & \cdot & \cdot & \cdot & 0 \\ \cdot & & & & & & \cdot \\ \cdot & & & & & & \cdot \\ \cdot & & & & & & \cdot \\ \cdot & & & & & & 1 \\ \alpha_p & \cdot & \cdot & & & & 0 \end{bmatrix} \quad B = \begin{bmatrix} \alpha_1 & 0 & \cdot & \cdot & \cdot & 0 \\ \cdot & \alpha_2 & \cdot & \cdot & \cdot & 0 \\ \cdot & & & & & \cdot \\ 0 & \cdot & \cdot & & & \alpha_p \end{bmatrix}$$

$$C = [\alpha_0 \ 0 \ \dots \ 0]'$$

On peut remarquer que $\epsilon(t)$ possède des moments finis d'ordre 4 et que Σ la matrice de covariance de $C\epsilon(t)$ est égale à :

$$\Sigma = 2 \begin{bmatrix} \alpha_0^2 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & & & & & \\ \cdot & & & & & \\ \cdot & & & & & \\ \cdot & & & & & \\ 0 & & & & & 0 \end{bmatrix}$$

Alors en appliquant le théorème 5.3.1, on dispose d'une condition nécessaire et suffisante d'existence d'une solution stationnaire au second ordre de (5.5.9), à savoir :

(C₁) L'équation matricielle $Q = AQA' + 2BQB' + \Sigma$ admet une solution positive en Q .

Cette condition conduit à un système d'équations que doivent vérifier les coefficients α_i , et qui pour des valeurs numériques données se résolvent assez facilement vue la forme des matrices A, B et C.

Le théorème 5.3.2 fournit une condition suffisante, plus simple à obtenir :

(C₂) La matrice $A^{\otimes 2} + 2B^{\otimes 2}$ a toutes ses valeurs propres de module strictement inférieur à l'unité.

On peut par exemple expliciter ces conditions dans le cas d'un modèle ARCH(2).

Posons :

$$\Delta = (1-3\alpha_1^2)(1-2\alpha_1^2)(1-\alpha_2-2\alpha_1\alpha_2)-\alpha_2^2(1-\alpha_2-2\alpha_1\alpha_2)-2\alpha_1^2\alpha_2(1-2\alpha_2^2)$$

On obtient la condition (C₁) si :

$$(1-\alpha_2-2\alpha_1\alpha_2)/\Delta > 0 \quad \text{et} \quad [\alpha_1^2/(1-\alpha_2-2\alpha_1\alpha_2)^2] < [1/(1-2\alpha_2^2)]$$

La condition (C₂) fournit le polynôme caractéristique suivant :

$$\begin{aligned} P(\lambda) = & -\lambda^3 + \lambda^2(3\alpha_1^2 + 2\alpha_2^2 + 2\alpha_1\alpha_2 + \alpha_2) \\ & + \lambda[2\alpha_1\alpha_2(-2\alpha_2^2 - 3\alpha_1^2 - 3\alpha_1\alpha_2 - \alpha_1) + \alpha_2^2(-2\alpha_2^2 + 1)] \\ & + \alpha_2^3(12\alpha_1^3 + 2\alpha_1^2 - 2\alpha_1 - 1) \end{aligned}$$

dont les racines doivent être de module strictement inférieur à 1.

Etant donné un modèle ARCH(p), l'une de conditions (C₁) ou (C₂) étant vérifiées, la solution stationnaire est donnée de manière unique par

$$(5.5.10) \quad \begin{cases} X^2(t) = HZ(t) + \alpha_0 \\ Z(t) = \sum_{j=0}^{\infty} \prod_{k=1}^j [A + B\epsilon(t-k+1)] C\epsilon(t-j) \end{cases}$$

D'autre part la représentation (5.5.9) est minimale dès

que les conditions du théorème 5.4.1 sont vérifiées. On obtient à nouveau un ensemble d'équations dépendant des coefficients α_i . Celles-ci, il est vrai, sont d'autant plus compliquées que la dimension des matrices est grande.

La fonction de covariance du processus $X^2(t)$, $t \in \mathbb{Z}$ s'obtient facilement à partir des équations suivantes :

$$\begin{cases} R(0) = HQH' \\ R(k) = HA^kQH' \end{cases}$$

où les matrices sont celles définies dans l'équation (5.5.9) et où Q vérifie $Q = AQA' + 2BQB' + \Sigma$.

Notons d'autre part que le processus $Y(t) = X^2(t) - \alpha_0$ peut s'écrire sous la forme :

$$Y(t) = \sum_{i=1}^p \alpha_i Y(t-i) + \xi(t) \quad \text{où} \quad \xi(t) = (\alpha_0 + \sum_{i=1}^p \alpha_i Y(t-i))\epsilon(t).$$

C'est donc un processus autorégressif.

La fonction de covariance du processus $X^2(t)$ est donc celle d'un processus autorégressif d'ordre p . Résultat déjà annoncé dans A. MILHØJ, (1985).

Pour obtenir les moments d'ordre supérieur du processus $X^2(t)$ on applique le théorème 5.5.7. Pour cela on utilise l'équation (5.2.6) avec $A(\epsilon(t)) = A + B\epsilon(t)$ et $H(\epsilon(t)) = C\epsilon(t)$, A, B et C étant définis ci-dessus. Le théorème permet de calculer les moments d'ordre supérieur de $Z(t)$, et on en déduit ceux de $X^2(t) = HZ(t)$.

4. Prédiction

La possibilité de construire des prédicteurs optimaux au

sens des moindres carrés est fondamentale et permet de faire des prévisions et l'on sait que souvent la performance des modèles de séries chronologiques est jugée sur la base de la performance de leurs prédicteurs.

Le caractère markovien permet de faire la prédiction des modèles bilinéaires de façon très simple, à condition d'avoir l'inversibilité.

Considérons le processus $Z(t)$, $t \in \mathbb{Z}$ défini par l'équation (5.2.6 b) tel que $E[H(\epsilon(t))] = 0$. Alors comme dans le cas linéaire on prend comme prédicteur au temps $t + 1$ l'espérance conditionnelle de $Z(t)$ par rapport aux valeurs passées du processus :

$$E[Z(t+1) | s] = E[A(\epsilon(t))]E[Z(t) | s] \quad s > t$$

Donc avec les notations des paragraphes II et III on a :

$$\begin{aligned} \hat{Z}(t+1 | s) &= \tilde{A} \hat{Z}(t | s) \\ \hat{Z}(s+k | s) &= \tilde{A}^k \hat{Z}(s | s) = \tilde{A}^k Z(s) . \end{aligned}$$

Et donc pour calculer le prédicteur $\hat{Z}(s+k | s)$, il faut pouvoir calculer $Z(s)$, or

$$Z(s) = A(\epsilon(s))Z(s-1) + H(\epsilon(s))$$

donc $Z(s)$ dépend de $Z(s-1)$, de $A(\epsilon(s))$ et de $H(\epsilon(s))$.

Il est alors nécessaire de savoir calculer $\epsilon(s)$ en fonction des valeurs passées du processus (d'où la nécessité de l'inversibilité du modèle), $Z(s)$ sera alors calculée en fonction des valeurs passées du processus soit $X(t)$, $X(t-1)$, etc ... On pourra alors en déduire l'expression de $\hat{X}(s+k | s)$ à partir des équations (5.2.6). Pour pouvoir se servir en pratique des

prédicteurs $\hat{X}(s+k|s)$, $k = 0, 1, \dots$, il faut savoir estimer les paramètres du modèle.

VI - CONCLUSION

Il est important de remarquer que les deux représentations markoviennes considérées ont permis en partie de traiter des problèmes que l'approche directe avait à peine permis d'aborder à savoir les propriétés d'ergodicité, d'inversibilité, le calcul des moments. On peut penser que cette approche permettra d'aller encore plus loin dans l'étude probabiliste de ces modèles.

On peut aussi envisager de généraliser à des processus n-linéaires la représentation markovienne ainsi trouvée et donc montrer que des systèmes multilinéaires peuvent admettre des représentations markoviennes.

Nous n'avons pas abordé dans ce chapitre le modèle "S.D.M." introduit par M.B. PRIESTLEY (cf. chapitre I) qui admet une représentation markovienne et qui englobe les modèles bilinéaires. Ce modèle est en fait trop général pour pouvoir être utile. En particulier il n'est pas affine en l'état et l'unicité de la représentation n'est pas garantie.

CHAPITRE VI

INFERENCE STATISTIQUE

I - INTRODUCTION

Dans les chapitres précédents nous avons traité de la présentation et de l'analyse probabiliste des modèles bilinéaires, soit considérés à partir d'une approche directe, soit à partir d'une représentation markovienne. Tout au long de cette démarche, nous avons donc supposé que le modèle considéré était celui qui collait le mieux à la réalité des données. Nous allons maintenant étudier les méthodes statistiques qui aident à l'interprétation des données. Les problèmes qui se posent sont de quatre types :

- (i) le tests de linéarité,
- (ii) le choix des modèles, comprenant le choix de la classe des modèles postulés (modèles autorégressifs à seuil, bilinéaires, etc, ...) et la détermination de l'ordre,
- (iii) l'estimation des paramètres du modèle,
- (iv) les tests de validité des modèles ajustés.

Pour les tests de linéarité, quelques procédures ont été proposées, mais on connaît encore très mal leur performance.

En ce qui concerne le choix des modèles il n'y a pratiquement pas de résultats. Les modèles non linéaires que l'on pourrait envisager sont trop nombreux, et il est à craindre qu'il n'y ait pas de procédures statistiques permettant de discriminer entre une classe et une autre. Toutefois si on se restreint à une classe particulière, bilinéaire par exemple, on peut poser la

question du choix de son ordre. Ce problème là n'a jamais reçu de solutions vue la complexité et la richesse des modèles bilinéaires.

Les problèmes d'estimation sont les problèmes les plus étudiés et on dispose de procédés d'estimation consistants pour certains modèles bilinéaires. Toutefois il y a des difficultés dues à la non existence des moments d'ordre élevé pour ce type de modèles.

Le dernier point n'a pas encore été développé.

II - ESTIMATION DES PARAMETRES DES MODELES BILINEAIRES

1. Méthode générale d'estimation

L'estimation des paramètres des modèles bilinéaires par la méthode du maximum de vraisemblance a été entreprise à partir de la représentation suivante des modèles :

$$(6.2.1) \quad X(t) = \sum_{i=1}^p a_i X(t-i) + \epsilon(t) + \sum_{j=1}^q c_j \epsilon(t-j) + \sum_{j=1}^Q \sum_{k=1}^P b_{jk} X(t-k) \epsilon(t-j).$$

Elle fut tout d'abord abordée par T. SUBBA RAO (1981) pour les modèles superdiagonaux $BL(p,0,p,p)$, parallèlement par D. GUEGAN (1981 a) pour le modèle superdiagonal à un terme $BL(0,0,2.1)$, enfin ces résultats ont été généralisés aux modèles superdiagonaux $BL(p,0,P,Q)$ par T.SUBBA RAO et M.M. GABR (1984). Nous allons donner le principe de la méthode dans le cas où le modèle (6.2.1) est inversible et où les variables aléatoires $\epsilon(t)$, $t \in \mathbb{Z}$ sont supposées gaussiennes. L'hypothèse d'inversibilité est fondamentale pour l'estimation des paramètres dans cette méthode. On sait alors que :

$$\epsilon(t) = f(X(t), X(t-1), \dots)$$

pour une certaine fonction f .

Soit $\hat{\epsilon}(t)$ défini de la même façon en prenant $X(s) = 0$ pour $s \leq 0$. Généralement, on a :

$$\hat{\epsilon}(t) - \epsilon(t) \rightarrow 0 \text{ quand } t \rightarrow \infty.$$

Pour un échantillon $X(1), \dots, X(n)$ donné, les variables aléatoires $\hat{\epsilon}(1), \dots, \hat{\epsilon}(n)$ sont l'image de $X(1), \dots, X(n)$ par une application bijective de jacobien 1 en général, et donc la fonction de vraisemblance peut s'écrire, en supposant que $\hat{\epsilon}(1), \dots, \hat{\epsilon}(n)$ sont de loi conjointe gaussienne :

$$-\frac{1}{2} \log \det 2\pi \Sigma - \frac{1}{2} [\hat{\epsilon}(1), \dots, \hat{\epsilon}(n)]' \Sigma^{-1} [\hat{\epsilon}(1), \dots, \hat{\epsilon}(n)]'$$

où Σ est la matrice de variance covariance du vecteur $(\hat{\epsilon}(1), \dots, \hat{\epsilon}(n))$.

Quand n est grand, les $\hat{\epsilon}(t)$, $t = k, \dots, n$, pour k assez grand sont proches de $\epsilon(t)$, $t = k, \dots, n$. Si k est petit devant n , la contribution de la fonction log vraisemblance de $\hat{\epsilon}(1), \dots, \hat{\epsilon}(k)$ sera négligeable par rapport aux autres variables. Dans ces conditions, on pourra espérer approcher la fonction de vraisemblance en remplaçant Σ par σ^2 fois la matrice unité, cela conduit à l'approximation de la log vraisemblance :

$$-\frac{1}{2} \log \det(2\pi \sigma)^n - \frac{1}{2\sigma^2} \sum_{t=1}^n \hat{\epsilon}^2(t).$$

L'estimation par maximum de vraisemblance consiste donc à minimiser $\sum_{t=1}^n \hat{\epsilon}^2(t)$, et à estimer σ^2 par $\frac{1}{n}$ fois le minimum ainsi obtenu. Cette méthode revient aussi à faire l'estimation par les moindres carrés. Par conséquent la méthode du maximum de vraisemblance modulo l'approximation relative aux conditions de

bord, se réduit à la méthode des moindres carrés.

Notons $\underline{\theta}$ le vecteur des paramètres à estimer. Soit $\underline{\theta} = (\theta_1, \dots, \theta_N) = (a_1, a_2, \dots, a_p, c_1, c_2, \dots, c_p, b_{11}, \dots, b_{pp})$.

$$\text{Posons } Q(\underline{\theta}) = \sum_{t=1}^n \hat{\epsilon}^2(t) .$$

Pour obtenir les estimateurs du maximum de vraisemblance on peut par exemple utiliser la méthode de Newton-Raphson. Si $\underline{\theta}^{(i)}$ représente l'ensemble des estimateurs obtenus à la $i^{\text{ème}}$ étape de l'itération, on a la relation suivante :

$$(6.2.2) \quad \underline{\theta}^{(i+1)} = \underline{\theta}^{(i)} - H^{-1}(\underline{\theta}^{(i)}) G(\underline{\theta}^{(i)})$$

avec

$$(6.2.3) \quad G(\underline{\theta}) = \left[\frac{\partial Q(\underline{\theta})}{\partial \theta_1} \quad \dots \quad \frac{\partial Q(\underline{\theta})}{\partial \theta_N} \right]$$

et

$$(6.2.4) \quad H(\underline{\theta}) = \left[\frac{\partial^2 Q(\underline{\theta})}{\partial \theta_i \partial \theta_j} \right]_{N \times N}$$

où les dérivées de $Q(\underline{\theta})$ par rapport à $\underline{\theta}$ sont données par :

$$(6.2.5) \quad \left\{ \begin{array}{l} \frac{\partial Q(\underline{\theta})}{\partial \theta_i} = 2 \sum_{t=1}^n \hat{\epsilon}(t) \frac{\partial \hat{\epsilon}(t)}{\partial \theta_i} \quad i = 1, \dots, N \\ \frac{\partial^2 Q(\underline{\theta})}{\partial \theta_i \partial \theta_j} = 2 \sum_{t=1}^n \frac{\partial \hat{\epsilon}(t)}{\partial \theta_i} \frac{\partial \hat{\epsilon}(t)}{\partial \theta_j} + 2 \sum_{t=1}^n \hat{\epsilon}(t) \frac{\partial^2 \hat{\epsilon}(t)}{\partial \theta_i \partial \theta_j} \quad i, j = 1, \dots, N \end{array} \right.$$

Les dérivées partielles de $\hat{\epsilon}(t)$ peuvent être exprimées à partir de celle de $\hat{\epsilon}(t-1), \dots, \hat{\epsilon}(t-\max(q, Q))$ et de $X(t-1), \dots, X(t-\max(p, P))$.

Par exemple dans le cas où $b_{ij} = 0$ si $i \neq j$, et $c_j = 0 \forall j$, (ce qui correspond à un modèle diagonal sans partie moyenne mobile), on a les relations suivantes :

$$(6.2.6) \left\{ \begin{array}{l} \frac{\partial \hat{\epsilon}(t)}{\partial a_i} + \sum_{j=1}^p b_{jj}(t-j) \frac{\partial \hat{\epsilon}(t-j)}{\partial a_i} = -X(t-i) \quad i = 1, \dots, p \\ \frac{\partial \hat{\epsilon}(t)}{\partial b_{ii}} + \sum_{j=1}^p b_{jj}X(t-j) \frac{\partial \hat{\epsilon}(t-j)}{\partial b_{ii}} = -X(t-i)\hat{\epsilon}(t-i) \quad i=1, \dots, p \end{array} \right.$$

et

$$(6.2.7) \left\{ \begin{array}{l} \frac{\partial^2 \hat{\epsilon}(t)}{\partial a_i \partial a_j} = 0 \quad i = 1, \dots, j = 1, \dots, p \\ \frac{\partial^2 \hat{\epsilon}(t)}{\partial a_i \partial b_{jj}} + \sum_{k=1}^p b_{kk}X(t-k) \frac{\partial^2 \hat{\epsilon}(t-k)}{\partial a_i \partial b_{jj}} = X(t-j) \frac{\partial \hat{\epsilon}(t-j)}{\partial a_i} \\ \quad i = 1, \dots, p, j = 1, \dots, p \\ \frac{\partial^2 \hat{\epsilon}(t)}{\partial a_{ii} \partial b_{jj}} + \sum_{k=1}^p b_{kk}X(t-k) \frac{\partial^2 \hat{\epsilon}(t-k)}{\partial b_{ii} \partial b_{jj}} = \\ \quad -X(t-i) \frac{\partial \hat{\epsilon}(t-i)}{\partial b_{ii}} - X(t-j) \frac{\partial \hat{\epsilon}(t-j)}{\partial b_{jj}} \\ \quad i = 1, \dots, p, j = 1, \dots, p \end{array} \right.$$

Ainsi étant donné un ensemble de valeurs $\{a_i\}$ et $\{b_{jj}\}$ on peut évaluer les premières et les secondes dérivées partielles de $\hat{\epsilon}(t)$ en utilisant les équations récursives (6.2.5), (6.2.6) et (6.2.7) ; puis évaluer le vecteur quotient G et la matrice Hessienne H à l'aide des équations (6.2.3) et (6.2.4), et terminer en utilisant l'équation de récurrence (6.2.2). L'itération s'arrête dès que la précision requise est atteinte. L'itération précédente converge si elle est bien initialisée. Il n'y a pas de méthode pour obtenir des estimateurs initiaux. L'approche de T. SUBBA RAO consiste à ajuster un modèle ARMA et à utiliser les paramètres de

ce modèle comme estimateurs initiaux. Toutefois on n'a pas de résultats de convergence pour les estimateurs du maximum de vraisemblance, ni de propriétés asymptotiques contrairement à ce qui se passe dans le cas indépendant.

2. Mise en oeuvre de la méthode d'estimation pour certains modèles bilinéaires

Jusqu'à présent c'est la méthode des moindres carrés qui a été utilisée pour estimer les paramètres des modèles bilinéaires.

L'estimation des paramètres par la méthode des moindres carrés a été abordée tout d'abord par T. PHAM DINH et L. TAT TRAN (1981) dans le cas du modèle bilinéaire d'ordre 1 suivant :

$$(5.2.8) \quad X(t) = aX(t-1) + \epsilon(t) + b\epsilon(t-1)X(t-1).$$

D. GUEGAN et T. PHAM DINH (1987 b) ont généralisé la méthode pour traiter les modèles bilinéaires diagonaux en se servant de la représentation markovienne obtenue par T. PHAM DINH (1985). En fait la méthode est applicable pour la classe plus générale des processus admettant une représentation markovienne (5.2.1) inversible satisfaisant à certaines contraintes. On considère le modèle $X(t)$, $t \in \mathbb{Z}$ défini par les équations suivantes :

$$(6.2.9) \quad \begin{cases} X(t) = HZ(t-1) + \epsilon(t) \\ Z(t) = [A+B\epsilon(t)]Z(t-1) + C\epsilon(t) + D\epsilon^2(t) \end{cases}$$

où A et B sont des matrices carrées, H une matrice ligne, C et D des matrices colonnes, et l'on suppose que :

$$(H) : "B = DH" .$$

Ce modèle contient comme cas particulier les modèles bilinéaires diagonaux. On suppose vérifiées les conditions d'inversibilité du théorème 5.4.3'.

Soit alors la suite $\hat{\varepsilon}(t)$ construite à partir de valeurs hypothétiques $\hat{A}, \hat{C}, \hat{D}, \hat{H}$ et d'une valeur arbitraire z_0 de $Z(0)$, par :

$$(6.2.10) \quad \begin{cases} \hat{\varepsilon}(t|z_0) = X(t) - \hat{H}\hat{Z}(t-1|z_0) \\ \hat{Z}(t|z_0) = [\hat{A} + \hat{B}\hat{\varepsilon}(t|z_0)]\hat{Z}(t-1|z_0) + \hat{C}\hat{\varepsilon}(t|z_0) + \hat{D}\hat{\varepsilon}(t|z_0) \end{cases}$$

où $Z(0|z_0) = z_0$.

D'après le théorème 5.4.3', on montre que la série

$$(6.2.11) \quad \hat{Z}(t) = \sum_{j=0}^{\infty} \prod_{k=0}^{j-1} \hat{K}(t-k) [\hat{C}X(t-j) + \hat{D}X^2(t-j)]$$

où $\hat{K}(t) = \hat{A} - \hat{C}\hat{H} - \hat{D}\hat{H} X(t)$, est convergente presque sûrement, et que sa différence avec $\hat{Z}(t|z_0)$ tend vers zéro quand $t \rightarrow \infty$, quelque soit z_0 .

Il est important de constater que la représentation markovienne (6.2.9) n'est unique qu'à une transformation linéaire près du vecteur d'état car les matrices $PAP^{-1}, PC, PD, HP^{-1}$ correspondent au même modèle que les matrices A, C, D, H, P étant une matrice inversible donnée. Pour contourner cette indétermination de la représentation markovienne, on suppose que les matrices A, C, D, H dépendent d'un paramètre vectoriel θ tel que pour 2 valeurs distinctes de θ , les matrices correspondantes ne soient pas déduites l'une de l'autre par la transformation ci-dessus. Dans le cas des modèles diagonaux définis à partir de l'équation (6.2.1), on peut prendre pour θ l'ensemble des paramètres $\{a_i\}, \{c_i\}$, et $\{b_{ij}\}$.

On supposera donc que la représentation (6.2.9) est quasiminimale dans le sens où on l'a défini au chapitre V, définition 5.4.2.

Rappelons maintenant en quoi consiste la méthode standard des moindres carrés. Elle consiste à minimiser la somme des carrés des erreurs sur un certain domaine de l'espace des paramètres Θ contenant la vraie valeur de θ . On suppose que la condition de stationnarité du théorème 5.3.2, et que les conditions d'inversibilité du théorème 5.4.3 sont vérifiées, ce qui permet de délimiter le domaine des paramètres admissibles.

Cependant il n'est pas facile de vérifier la condition d'inversibilité du théorème 5.4.3', et la condition de stationnarité prend en compte le paramètre σ^2 que l'on n'estime pas pour l'instant. Pour ces raisons, nous allons considérer un espace aléatoire $\Theta_{n,\delta}$ fonction des observations défini de la manière suivante :

$$\Theta_{n,\delta} = \{ \hat{\theta} \in \Theta_0 : \prod_{t=1}^n \| \hat{A} - \hat{C}\hat{H} - \hat{D}\hat{H} X(t) \|_{\varphi} \leq (1-\delta)^n \}$$

où δ est une constante positive, et où Θ_0 est l'ensemble des θ tels que $A^{\otimes 2}$ ait ses valeurs propres de module inférieur ou égale à 1. La condition $\theta \in \Theta_0$ est moins forte que la condition du théorème 5.3.2, c'est donc une condition nécessaire mais pas suffisante pour la stationnarité du modèle. L'estimateur obtenu ne correspond pas forcément à un modèle stationnaire, mais comme on le verra plus tard, il est presque sûrement consistant, et donc la stationnarité est encore satisfaite quand la taille de l'échantillon est assez grande.

Notons que l'introduction de δ permet l'utilisation du signe \leq au lieu de $<$, qui généralement conduit à un espace compact.

L'espace $\Theta_{n,\delta}$ doit être tel que $\liminf_{n \rightarrow \infty} \Theta_{n,\delta}$ contienne la vraie valeur du paramètre. Pour cela on remarque que d'après le théorème ergodique :

$$\limsup_{n \rightarrow \infty} \Theta_{n,\delta} \subset \Theta_{\delta/2} \quad \text{où}$$

$$\Theta_{\delta/2} = \{ \hat{\theta} \in \Theta_0, E[\log \|\hat{A} - \hat{C}\hat{H} - \hat{D}\hat{H} X(t)\|_{\varphi}] \leq \log(1-\delta/2) \} .$$

Et par conséquent la condition précédente est satisfaite si

$$E[\log \|A - CH - DH X(t)\|_{\varphi}] < \log(1-\delta) .$$

La consistance forte de l'estimateur des moindres carrés $\hat{\theta}_n$ est basé que le résultat général suivant :

Théorème 6.2.1 (T.PHAM DINH, L.TAT TRAN, 1981)

Considérons $Q_n(\tilde{\theta})$ une suite de variables aléatoires continues définies sur un sous espace compact $\tilde{\Theta}$ de l'espace euclidien de dimension k , \mathbb{R}^k . Soit θ un point de $\tilde{\Theta}$, et Θ_n une suite de sous espaces compacts aléatoires de \mathbb{R}^k , tels que :

- (i) Avec la probabilité 1, $\theta \in \Theta_n \subset \tilde{\Theta}$ pour un n suffisamment grand.
- (ii) Pour $\hat{\theta} \in \tilde{\Theta}$ tel que $\hat{\theta} \neq \theta$, il existe un voisinage $U(\hat{\theta})$ de θ tel que :

$$\liminf_{n \rightarrow \infty} \{ \inf_{\tilde{\theta} \in U(\hat{\theta})} Q_n(\tilde{\theta}) - Q_n(\theta) \} > 0 \quad \text{p.s.}$$

Alors $\hat{\theta}_n$, solution de la minimisation de $Q_n(\theta)$ sur Θ_n converge sûrement vers θ .

Pour utiliser ce théorème, on pose ici :

$$Q_n(\tilde{\theta}) = \frac{1}{n} \sum_{t=-1}^n \epsilon_{\tilde{\theta}}^2(t|z_0)$$

où $\epsilon_{\tilde{\theta}}(t|z_0)$ est donné par (6.2.10) avec $\tilde{\theta}$ comme paramètre. La condition (i) du théorème est clairement satisfaite en prenant l'ensemble $\Theta_n = \Theta_{n,\delta}$. D'autre part on peut montrer que l'application $\hat{\theta} \rightarrow Z_{\hat{\theta}}(t)$, où $Z_{\hat{\theta}}(t)$ est défini par les équations (6.2.11), est continue, et la convergence de $\hat{Z}(t|z_0) - Z_{\hat{\theta}}(t)$ vers 0 où $\hat{Z}(t|z_0)$ est donnée par les équations (6.2.10) avec les paramètres $\hat{A}, \hat{C}, \hat{D}, \hat{H}$ à la place de A, C, D, H est uniforme sur $\Theta_{\delta/2}$ (cf. Théorème (5.4.3', p.181, chapitre V). La condition (ii) est vérifiée en prenant $\tilde{\theta} = \Theta_{\delta/2}$ pourvu que $E[\epsilon_{\tilde{\theta}}^2(t)] > \sigma^2$ pour tout $\hat{\theta} \neq \theta$. Comme

$$\epsilon_{\hat{\theta}}(t) = \epsilon(t) + H[Z(t-1) - Z_{\hat{\theta}}(t-1)],$$

cette dernière condition est équivalente à $HZ_{\hat{\theta}}(t) \neq HZ(t)$ pour tout $\hat{\theta} \neq \theta$. Pour obtenir cette dernière condition on se servira de la quasiminimalité du processus $X(t)$ telle qu'elle a été définie dans le chapitre V.

Théorème 6.2.2 (D.GUEGAN, T.PHAM DINH, 1987 b, p. 311)

Si le processus $X(t)$, $t \in \mathbb{Z}$ défini par les équations (6.2.9) avec la condition "B = DH" est quasiminimal et si $\epsilon(t)$ a une distribution continue, alors $HZ_{\hat{\theta}}(t) = HZ(t)$ si et seulement si

il existe une matrice inversible P telle que $\hat{A} = PAP^{-1}$,
 $\hat{C} = PC$, $\hat{D} = PD$, $\hat{H} = P^{-1}H$.

On obtient ainsi la consistance forte des estimateurs des moindres carrés pour les modèles bilinéaires diagonaux, généralisant ainsi le théorème de T.PHAM DINH et L.TAT TRAN (1981)

établi pour le modèle (6.2.8).

Remarques

1. En utilisant les résultats de T. PHAM DINH (1985), on peut voir que la représentation (6.2.9) est quasiminimale si la représentation linéaire associée :

$$\begin{cases} X(t) = HZ(t-1) + \epsilon(t) \\ Z(t) = AZ(t-1) + C\epsilon(t) \end{cases}$$

l'est.

Dans le cas particulier du modèle (6.2.1) avec $P = 0$ et $Q \leq \max(p, q)$, la minimalité de la représentation linéaire est satisfaite si les polynômes $1 - \sum_j a_j z^j$ et $1 - \sum_j c_j z^j$ n'ont pas de racines dans le cercle unité, ni de racines communes. Ceci fournit une condition suffisante pour la minimalité de la représentation (6.2.1) pour cette sous classe de modèles bilinéaires.

2. Le problème de la normalité asymptotique de tels estimateurs reste posé.

Exemples : On se place dans le cas des modèles diagonaux avec $p = q = Q = 2$, alors les matrices A, B, C, D, H de la représentation (6.2.9) ont pour expression :

$$A = \begin{bmatrix} a_1 & 1 \\ a_2 & 0 \end{bmatrix} \quad B = \begin{bmatrix} b_{01} & 0 \\ b_{02} & 0 \end{bmatrix} \quad C = \begin{bmatrix} a_1 + c_1 \\ a_2 + c_2 \end{bmatrix} \quad D = \begin{bmatrix} b_{01} \\ b_{02} \end{bmatrix}$$

$$H = [1 \quad 0] .$$

On a tout d'abord besoin de l'inversibilité du processus $X(t)$, $t \in \mathbb{Z}$. En utilisant le théorème 5.4.3, une condition

suffisante est :

$$E[\log|c_1 + b_{01} X(t)|] < 0$$

$$\text{soit } \frac{1}{2} E[\log|c_1 + b_{01} X(t)|^2] < \frac{1}{2} \log E[c_1 + b_{01} X(t)]^2 < 0$$

$$\text{donc } \log E[c_1 + b_{01} X(t)]^2 < 0 \iff E[c_1 + b_{01} X(t)]^2 < 1$$

donc la condition d'inversibilité est vérifiée pour l'ensemble :

$$\Theta_I = \{(c_1, b_{01}) \mid c_1 + b_{01} E[X^2(t)] + 2c_1 b_{01} E[X(t)] < 1\} .$$

D'autre part, l'espace Θ_0 est défini par la condition que A^{Θ_2} ait ses propres valeurs propres de module inférieur ou égal à 1. Ici :

$$A^{\Theta_2} = \begin{bmatrix} a_1^2 & 1 & 2a_1 \\ a_2^2 & 0 & 0 \\ a_1 a_2 & 0 & a_2 \end{bmatrix}$$

Le polynôme caractéristique est donné par :

$$P(\lambda) = -\lambda^3 + \lambda^2(a_1^2 + a_2) + \lambda a_2(a_1^2 + a_2) - a_2^3 .$$

Par suite :

$$\Theta_0 = \{(a_1, a_2) \mid \text{les racines de } P(\lambda) \text{ sont de module inférieur ou égal à } 1\} .$$

L'espace des paramètres est donné par la réunion de Θ_0 et de Θ_I . On définit ainsi une région dans laquelle on a la consistance des estimateurs des moindres carrés.

3. Estimation par la méthode des moments.

Nous avons présenté au chapitre III l'état actuel des recherches en ce qui concerne le calcul des moments des modèles bilinéaires régis par l'équation (6.2.1). De part les difficultés calculatoires liées à la forme générale du modèle, les résultats ne peuvent être obtenus explicitement que pour quelques cas particuliers. A partir de ces connaissances il est possible d'estimer les paramètres des modèles. D. GUEGAN (1984 a, p. 257) a appliqué la méthode des moments dans le cas des deux modèles suivants :

$$(6.2.11) \quad X(t) = \epsilon(t) + \delta\epsilon(t-1)\epsilon(t-2)$$

et

$$(6.2.12) \quad X(t) = \epsilon(t) + \gamma\epsilon(t-1)X(t-2)$$

où $\epsilon(t)$ est une suite de variables aléatoires indépendantes équidistribuées gaussiennes centrées, de variance σ^2 .

Considérons tout d'abord le modèle (6.2.11). Etant donné une suite d'observations $X(1), X(2), \dots, X(N)$, le processus étant centré et non corrélé, on considère les moments d'ordre trois et plus. Notons :

$$C(i, j) = E[X(t)X(t-i)X(t-j)]$$

$$C(i, j, k) = E[X(t)X(t-i)X(t-j)X(t-k)]$$

Ces moments sont respectivement estimés par :

$$\hat{C}(i, j) = \frac{1}{N} \sum_{i, j=1}^N X(t)X(t-i)X(t-j)$$

$$\text{et} \quad \hat{C}(i, j, k) = \frac{1}{N} \sum_{i, j, k=1}^N X(t)X(t-i)X(t-j)X(t-k);$$

On peut alors en déduire les paramètres du modèle. Notons que ceci peut se faire de différentes façons. Dans l'exemple considéré, on a choisi les expressions suivantes :

$$\hat{\delta} = \frac{\hat{C}(1,1,3)^2}{4\hat{C}^3(1,2)} \quad \text{et} \quad \frac{2\hat{C}^2(1,2)}{\hat{C}(1,1,3)} .$$

Les moments estimés convergent presque sûrement vers les vraies valeurs $C(1,2)$ et $C(1,1,3)$ et donc $\hat{\delta}$ et $\hat{\sigma}^2$ convergent fortement vers leurs vraies valeurs δ et σ^2 (cf. D. GUEGAN, 1984, a, p. 260).

De plus les estimateurs $\hat{C}(1,2)$ et $\hat{C}(1,1,3)$ sont des estimateurs asymptotiquement de loi conjointe normale, il est possible, d'en calculer la variance et la covariance asymptotiques et on peut en déduire la loi des estimateurs de $\hat{\delta}$ et $\hat{\sigma}^2$.

L'estimation des paramètres de ce modèle avait déjà donné lieu à une étude de P.M. ROBINSON (1977) sans que celui-ci donne l'expression explicite des estimateurs.

Une démarche identique a été suivie dans le cas du modèle (6.2.12), néanmoins cette fois-ci quelques précautions sont à prendre pour assurer que le processus admette des moments d'ordre suffisamment élevé.

Comme dans le cas précédent le processus est centré et non corrélé. Par suite on se servira des moments d'ordre 3 et 4 ainsi que de la variance du processus pour l'estimation des paramètres. Les équations choisies sont les suivantes :

$$\hat{\gamma} = \frac{\hat{C}(0)^2 \hat{C}(0,1,4)}{2\hat{C}(1,2)^3} \quad \text{et} \quad \hat{\sigma}^2 = \frac{\hat{C}(1,2,3)^2}{2\hat{C}(0)\hat{C}(0,1,4)} .$$

Pour assurer la convergence des moments estimés d'ordre 4, il suffit que le processus $X(t)$ admette des moments d'ordre 4 et que $u^2 < \frac{1}{3}$, où $u = \gamma^2 \sigma^2$. Toutefois pour pouvoir utiliser le théorème central limite, il est nécessaire que le processus admette des moments d'ordre 8 et satisfasse des conditions de mélange adéquates. L'existence des moments d'ordre 8 est assurée si $u^4 < \frac{1}{105}$.

L'intérêt de la méthode des moments est qu'elle permet d'obtenir des estimateurs initiaux. Son inconvénient vient de la difficulté calculatoire dans des modèles plus complexes. Dans ce contexte il serait intéressant de disposer de procédures de calculs récursifs pour obtenir les estimateurs.

4. Estimation des paramètres d'un modèle ARCH(p)

Le modèle ARCH(p) a été introduit au chapitre précédent (cf. équation 5.5.8). On peut estimer les paramètres d'un tel modèle.

A partir de l'équation 5.5.8, on a :

$$\varepsilon(t) = \frac{Y(t) - \sum_{i=1}^p \alpha_i Y(t-i)}{\alpha_0 + \sum_{i=1}^p \alpha_i Y(t-i)}$$

et $Y(t) = X^2(t) - \alpha_0$.

Si $\alpha_0 + \sum_{i=1}^p \alpha_i Y(t-i)$ est non nul, le processus $X^2(t)$ est

inversible.

Différentes approches sont envisageables pour l'estimation des paramètres du modèle.

1ère approche : On peut considérer le modèle comme un processus autorégressif d'ordre p , alors on dispose d'une estimation des paramètres par les équations de Yule-Walker, comme dans le cas des AR(P).

2ème approche : On peut utiliser la méthode des moindres carrés. Pour cela on considère :

$$M = \min_{\alpha_0, \alpha_i} \sum_{t=n}^p \left| \frac{Y(t) - \sum_{i=1}^p \alpha_i Y(t-i)}{\alpha_0 + \sum_{i=1}^p \alpha_i Y(t-i)} \right|^2 .$$

Posons :

$$\epsilon_t(\alpha_0, \underline{\alpha}) = \frac{Y(t) - \sum_{i=1}^p \alpha_i Y(t-i)}{\alpha_0 + \sum_{i=1}^p \alpha_i Y(t-i)} .$$

Alors :

$$\frac{\partial M}{\partial \alpha_0} = 2 \sum_{t=n+1}^p \frac{-Y(t-i) \left[Y(t) - \sum_{i=1}^p \alpha_i Y(t-i) \right] + Y(t-i)}{\left[\alpha_0 + \sum_{i=1}^p \alpha_i Y(t-i) \right]^2} \epsilon_t(\alpha_0, \underline{\alpha})$$

$\frac{\partial M}{\partial \alpha_0} = 0$ implique les relations suivantes :

$$\sum_{t=n+1}^p \frac{Y(t-i) - Y(t-i)Y(t) + \sum_{k=1}^p \alpha_i Y(t-k)Y(t-i)}{\left[\alpha_0 + \sum_{i=1}^p \alpha_i Y(t-i) \right]^2} \epsilon_t(\alpha_0, \underline{\alpha}) = 0$$

et

$$\sum_{t=n+1}^p \frac{Y(t-i)}{\alpha_0 + \sum_{i=1}^p \alpha_i Y(t-i)} \epsilon_t(\alpha_0, \underline{\alpha}) = 0 .$$

On dispose donc d'un système de deux équations tel que les solutions en α_i et α_0 donnent les estimateurs des moindres carrés des paramètres considérés.

3ème approche : On peut envisager l'utilisation du maximum de vraisemblance puisque par hypothèse on connaît la distribution de $X(t)$ conditionnellement aux valeurs passées. Cette approche a été considérée par R.F. ENGLE (1982).

III - TESTS

Jusqu'à présent nous avons admis que le modèle étudié était choisi, son ordre connu, mais en pratique ajuster des modèles à partir d'observations signifie que l'on dispose d'un critère pour déterminer son ordre. Dans un premier temps, on va déterminer si le modèle est un modèle linéaire ou un modèle non linéaire, d'où la nécessité de disposer de tests de linéarité.

1. Tests de linéarité

En 1980, T. SUBBA RAO et M.M. GABR ont proposé un test de linéarité basé sur un résultat concernant le spectre d'ordre supérieur établi par D.R. BRILLINGER (1965). Supposons que le processus $X(t)$, $t \in \mathbb{Z}$ ait pour représentation de type Wold

$$X(t) = \sum_{u=-\infty}^{+\infty} a(u)\epsilon(t-u)$$

où $\epsilon(t)$, $t \in \mathbb{Z}$ est un processus bruit blanc centré, de variance σ^2 tel que $\mu_3 = E[\epsilon^3(t)]$. La densité spectrale et la densité bispectrale du processus ont respectivement pour expression :

$$f(w) = \frac{\sigma^2}{2\pi} |H(w)|^2$$

et

$$f(w, w') = \frac{\mu_3}{(2\pi)^3} H(-w-w')H(w)H(w')$$

où $H(w) = \sum a(u)e^{-iuw}$.

$$\text{Posons } \varphi(w, w') = \frac{|f(w, w')|^2}{f(w)f(w')f(w, w')}$$

alors en utilisant (6.3.1) :

$$\varphi(w, w') = \frac{\mu_3^2}{2\pi\sigma^2} \quad \forall w, \quad \forall w' .$$

Donc la nullité de la densité spectrale équivaut à $\mu_3 = 0$ et ceci peut être utilisé comme indicateur du caractère gaussien du processus $X(t)$.

De la même façon $\varphi(w, w') = \text{constante}$ est un indicateur de

la linéarité du processus $X(t)$.

Dans la pratique on discrétise les fonctions f et φ , c'est-à-dire qu'on les estime sur une grille de points w_1, \dots, w_M où $w_i = \frac{i\pi}{M}$.

Le test de l'hypothèse gaussienne consiste à tester

$$H_0 : f(w_i, w_j) = 0 \quad \forall i, \quad \forall j$$

et le test de la linéarité du processus consiste à tester :

$$H_0 : \varphi(w_i, w_j) = \text{constante} \quad \forall i, \quad \forall j .$$

La méthode précédente nécessite l'estimation du bispectre. D'autre part si la distribution du processus $\epsilon(t)$, $t \in \mathbb{Z}$ est symétrique, le bispectre est identiquement nul et par suite le test précédent ne permet pas de tester la non linéarité du processus.

D.M. KEENAN (1985), à partir de l'idée que dans le cas des modèles bilinéaires la non linéarité provient de l'existence de termes produits, propose un test inspiré du test de non additivité de J.W. TUKEY pour les modèles d'analyse de la variance. Il teste en fait l'absence de termes quadratiques dans le développement de Volterra du modèle, cf. équation 4.2.1.

Le test consiste d'abord à ajuster un modèle autorégressif d'ordre M qui fournit le prédicteur $\hat{X}(t)$ de $X(t)$ et les résidus $\hat{\epsilon}(t)$. Le test est alors basé sur la corrélation entre $\hat{\epsilon}(t)$ et les résidus de la régression de $\hat{X}^2(t)$ sur $1, X(t-1), \dots, X(t-M)$.

Il applique ce test à un certain nombre de modèles linéaires ou bilinéaires simulés. Les résultats obtenus permettent d'obtenir une différenciation entre les modèles linéaires et non

linéaires.

R.S. TSAY (1986) remarque que le test précédent peut être amélioré en considérant tous les produits croisés $X(t-i)X(t-j)$ au lieu de $\hat{X}^2(t)$. Le test est alors basé sur la corrélation multiple entre $\hat{\varepsilon}(t)$, et les résidus de la régression des produits précédents sur $1, X(t-1), \dots, X(t-M)$.

Les tests de D.M. KEENAN généralisés par R.S. TSAY ont été utilisés pour des modèles de Volterra présentant des termes produits.

Une deuxième classe de test introduits par J.D. PETRUCELLI et M. DAVIES (1986) est plutôt orientée vers les non linéarités du type à seuil. Ce dernier test utilise des techniques de sommes cumulées qui ont été employées dans des tests de rupture de régression. On cherche en fait à détecter des ruptures dans la fonction d'autorégression du processus observé, (cf. D.V. HINKLEY, 1970).

L'hypothèse de non linéarité étant retenue, il s'agit maintenant de savoir avec quel type de modèles il faut travailler. En fait on ne dispose pas de tels test mais on peut s'orienter vers des tests de validité permettant de confirmer le choix d'un modèle.

2. Tests d'adéquation de modèles

Dans le cas des modèles linéaires, il existe déjà un grand nombre de tests permettant de vérifier si le modèle choisi est adéquate, (pour une revue on peut se rapporter à D. GUEGAN, 1979).

D. GUEGAN (1984 a, p. 286) a proposé des tests d'adéquation de certains modèles bilinéaires particuliers basés sur l'étude des fonctions d'autocorrélations. Il s'agit des modèles (6.2.11) et

(6.2.12).

Pour le premier modèle, le test statistique utilisé est $s = \hat{C}(0,1,4)$ qui a une espérance mathématique asymptotique nulle sous H_0 et strictement positive sous l'alternative. Cette statistique est asymptotiquement normale avec une variance asymptotique \hat{V} calculable, (cf. D. GUEGAN, 1984 a, p. 265). On trouvera des détails et des simulations sur ce test dans l'article cité.

On procède de la même manière pour le deuxième modèle en considérant le test statistique $T = \hat{C}(0,0,0) - 3\hat{C}(0,4,4)$ qui a une espérance asymptotique nulle sous H_0 et strictement positive sous l'alternative (se reporter à l'article cité ci-dessus).

Ces deux tests sont fondés sur l'étude des moments d'ordre 4 car les moments d'ordre inférieur sont insuffisants pour conclure. Si les calculs sont un peu longs pour le modèle (6.2.12), ils ne présentent pas de difficultés théoriques et peuvent être généralisés à d'autres modèles bilinéaires.

En contrepartie ces tests sont d'emploi extrêmement faciles et demandent un minimum de calculs à l'utilisateur, (l'estimation des moments d'ordre 4 nécessite environ $3N$ multiplications pour le premier test et $6N$ pour le second).

Un autre aspect positif de cette démarche est qu'elle permet de mener de front estimation des paramètres et tests. A défaut d'un calcul exact de la vraisemblance, on est nécessairement amené à employer des estimateurs empiriques. Les résultats numériques fournis dans l'article cité ci-dessus montre qu'ils sont assez bons. Toutefois pour avoir une bonne précision des estimateurs, il faut disposer de longues séries. Cela s'explique par le fait que dans un modèle bilinéaire il existe de longues périodes où $X(t)$ varie peu, interrompues de façon aléatoire par de

courtes périodes de forte excitation.

Remarque :

Le choix de l'ordre d'un modèle n'a jusqu'à présent pas été discuté. Dans le cas des modèles ARMA, le critère d'AKAIKE (1974 b) fournit un outil satisfaisant. T. SUBBA RAO et M.M. GABR (1984) l'ont utilisé pour déterminer pas à pas l'ordre de la partie bilinéaire d'un modèle.

CHAPITRE VII

PROBLEMES OUVERTS

I - INTRODUCTION

Dans le présent travail nous avons cherché à présenter l'état actuel des connaissances concernant les modèles bilinéaires. Cette mise au point doit permettre d'orienter les recherches futures dans ce domaine.

L'étude directe des modèles bilinéaires à partir de l'équation (4.1.1) ne permet d'atteindre que des résultats trop souvent limités, alors que l'approche markovienne représentée par les équations (5.2.1) (pour les modèles bilinéaires superdiagonaux), et (5.2.6) (pour l'ensemble des modèles bilinéaires), est beaucoup plus prometteuse. Il semble donc que c'est essentiellement ce dernier point de vue qui doit être considéré pour une étude plus approfondie des modèles bilinéaires.

Quelle que soit l'approche considérée, nous pouvons distinguer trois axes de recherche: le développement des propriétés probabilistes, une théorie de l'identification et des tests, et l'estimation des paramètres des modèles.

L'existence d'une solution stationnaire unique vérifiant l'équation (4.1.1) ou les équations (5.2.6) est en fait réglée; l'ergodicité et le mélange ont pu être obtenus pour ces modèles en utilisant des résultats de chaîne de Markov, (cf. R.L. TWEEDIE, 1975, 1983). Néanmoins ces résultats ne sont acquis que si la

chaîne de Markov est irréductible, condition qui n'est pas automatiquement satisfaite.

II - PROPRIETES PROBABILISTES DES MODELES BILINEAIRES.

Nous avons vu au chapitre III qu'un certain nombre de concepts sont utiles et indispensables pour la connaissance de la structure probabiliste d'un modèle. Deux notions demandent particulièrement à être étendues.: l'inversibilité et la connaissance des moments.

La notion d'inversibilité est indispensable pour espérer développer certaines propriétés statistiques des modèles.

Dans le cas des modèles diagonaux, une condition suffisante pour l'inversibilité a été établie. C'est seulement dans le cas des modèles d'ordre 1 qu'une condition nécessaire a pu être trouvée.

Le problème reste ouvert en ce qui concerne la recherche de conditions suffisantes plus faibles et leur généralisation aux modèles non diagonaux. La difficulté rencontrée dans ces modèles est due à l'existence des termes produit entre l'observation et le bruit en des instants différents. Ces termes peuvent être supprimés par l'utilisation de la représentation polynomiale en les entrées, affine en l'état, mais cela fait apparaître des termes non linéaires (car polynomiaux) en le bruit $\epsilon(t)$.

L'étude des moments d'ordre supérieur a été partiellement traitée à partir de l'équation (4.1.1) par S.A.O. SESAY et T. SUBBA RAO (1985). T. PHAM DINH (1986) et D. GUEGAN (1987) ont obtenu l'expression des moments d'ordre pair respectivement à partir des équations (5.2.1) et (5.2.6).

Il serait utile de disposer d'expressions relativement

simples des différents moments et/ou d'algorithmes récursifs permettant de calculer ceux-ci.

III - IDENTIFICATION DES MODELES BILINEAIRES.

Face à une réalisation $X(1), \dots, X(N)$ d'un processus se pose le double problème: identifier un modèle adéquate, et estimer les paramètres de ce modèle. Ces deux démarches étant à compléter par le test d'adéquation du modèle trouvé. Ceci est un schéma presque idéal pour la modélisation de données, schéma maintenant bien rodé dans le cas des modèles ARMA, et qui n'en est qu'à ses débuts pour les modèles bilinéaires.

La représentation des données peut parfois fournir une première indication sur les modèles à étudier. Certains modèles non linéaires se distinguent par l'aspect de leur évolution temporelle. Par exemple, les systèmes dynamiques bruités à cycle limite présentent des oscillations qui peuvent être mises en évidence par le tracé des trajectoires dans le plan des phases. Les processus bilinéaires présentent souvent des périodes de faible activité suivies d'une explosion.

Dans le cas des processus linéaires, l'identification est basée uniquement sur l'examen de la fonction de covariance. Dans le cas des processus bilinéaires la structure probabiliste est reflétée par les moments de tous ordres. C'est cela la difficulté de l'identification de ces modèles. Il faut en fait chercher quels moments on doit examiner pour avoir la meilleure information sur ces modèles. Pour l'instant on n'a pas encore pu déterminer les moments qui permettent de spécifier un modèle bilinéaire donné de manière unique.

Les tests de linéarité peuvent être utiles pour déceler la non linéarité d'un modèle, mais ils ne permettent pas d'identifier

la classe de celui-ci. Ils peuvent néanmoins en donner une indication dans la mesure où ils sont construits pour détecter un type donné de non linéarité.

Dans le cas des modèles bilinéaires un problème peu abordé est le choix de l'ordre du modèle. Le critère d'AKAIKE peut être utilisé à cette fin. Néanmoins celui-ci demande l'ajustement de différents modèles ce qui est très lourd. La performance statistique de cette méthode n'a pas été étudiée. La recherche de critères de choix de l'ordre reste un problème intéressant à étudier.

IV - ESTIMATION.

Pour l'estimation des paramètres on dispose pour l'instant de deux types de méthodes: la méthode du maximum de vraisemblance et la méthode des moments.

La méthode du maximum de vraisemblance (qui est équivalente à la méthode des moindres carrés), n'est applicable que si le modèle est inversible, ce qui n'est pas toujours le cas. D'autre part la maximisation de la vraisemblance, (minimisation de la somme des carrés des résidus), est un problème numérique difficile. Enfin on n'a pas pu obtenir la normalité asymptotique des estimateurs.

La méthode des moments nécessite des expressions explicites des moments en fonction des paramètres du modèle. Ces expressions peuvent être redondantes et il se pose le problème du choix d'un certain nombre d'entre elles, établissant une correspondance injective entre les paramètres du modèle et certains de ces moments. L'expression de cette correspondance d'ailleurs n'est pas toujours simple. L'estimateur des moments sous des hypothèses générales est asymptotiquement normal, mais il est difficile d'obtenir les expressions de la matrice de covariance asymptotique

des estimateurs. La bonne performance de ces estimateurs reste à prouver.

V - TESTS DE VALIDITE.

Un certain nombre de tests de validité de modèles a été proposé. Il convient néanmoins d'en étudier les performances de façon plus exhaustive pour les différents modèles alternatifs. D'autre part il est utile de disposer d'un plus grand nombre de tests, chacun sera plus sensible à un certain type de déviation, ce qui fournirait l'indication de la classe de modèles non linéaires qu'il conviendrait d'examiner.

L'analyse des résidus des modèles ajustés pourrait être utile pour déceler la mauvaise adéquation du modèle. Comment effectuer cette analyse de façon efficace est un sujet de recherche important.

BIBLIOGRAPHIE GENERALE

BIBLIOGRAPHIE

- AASE K.K. (1983) : Recursive estimation in non linear time series models of Autoregressive type. J.R.S.S. Serie B. 45. n° 2. 228-237.
- AKAIKE H. (1974 a) : Markovian representation of stochastic processes and its application to the analysis of autoregressive moving average processes. Ann. Inst. Stat. Math. 26. 363-387.
- AKAIKE H. (1974 b) : A new look at the statistical model identification. I.E.E.E. Trans. on Auto. Control. Vol. AC.19. n°6.
- AKAIKE H. (1974 c) : Stochastic theory of Minimal Realization I.E.E.E. Trans. on Auto. Control. Vol. AC.19. n° 6. 667-674.
- AKAIKE H. (1975) : Markovian representation of stochastic processes by canonical variables. SIAM J. Control. Vol.13. n° 1. 162-173.
- AKAMANAM S.I.(1983) : Some contributions to the study of Bilinear time series models. PH.D non publié. Université de Sheffield.
- AKAMANAM S.I., BHASKARA RAO M., SUBRAMANYAM S. (1986) : On the ergodicity of bilinear time series models. J.T.S.A. vol.7. n° 3. 157-163.

- BARTLETT M.S. (1946) : On the theoretical specification and sampling properties of autocorrelated time series. J.R.S.S. Serie B. 8. 27-41.
- BHASKARA RAO M., SUBBA RAO T., WALKER A.M. (1983) : On the existence of some bilinear time series models. J.T.S.A. Vol.4. n° 2. 95-110.
- BLANC LAPIERRE A., FORTET R. (1953) : Théorie des fonctions aléatoires. MASSON.
- BOURDON P. (1982): An approximation of discrete-time non linear systems using stochastic inputs. Systems and control letters . 1. 382-384.
- BOX G.E.P., JENKINS G.M. (1970) : Time series analysis, forecasting and control. Holden Day.
- BREIMAN L. (1969) : Probability and Stochastic processes with a view toward applications. Houghton Mifflin Company - BOSTON:
- BRILLINGER D.R. (1965) : An introduction to polyspectra. Ann. Math. Stat. 36. 1351-1374.
- BRILLINGER D.R. (1970) : The identification of polynomial system by means of higher order spectra. J. Sound. Vibr. 12, 3. 301-304.
- BRILLINGER D.R. (1975) : Time series Data analysis and theory. Holt, Rinehart and Wiston.
- BRILLINGER D.R., ROSENBLATT M. (1966 a): Asymptotic theory of estimates of k^{th} order spectra, in Advanced seminar in spectral analysis of time series. Ed. B. HARRIS.

John Wiley. 153-188.

BRILLINGER D.R., ROSENBLATT M. (1966 b) : Computation and interpretation of the k^{th} order spectra, in Advanced seminar in spectral analysis of time series, ed. B. HARRIS. John Wiley. 189-232.

BROCKETT R.W. (1976) : Volterra series and geometric control theory. Automatica. 12. 167-176.

BRUNI C., Di PILLO G., KOCH G. (1974) : Bilinear systems: an appealing class of "nearly linear" systems in theory and applications. I.E.E.E. Trans. Auto Control. 19. 334-348.

CARTWRIGHT P.A. (1985) : Forecasting time series : a comparative analysis of alternative classes of time series models. J.T.S.A. Vol. 6. n° 4. 203-211.

CRAMER H., LEADBETTER M. (1955) : Stationary and related stochastic processes. John Wiley.

DACUNHA CASTELLE D., DUFLO M. (1982) : Probabilités et Statistiques. Masson.

DOOB J.L. (1953) : Stochastic Processes. John Wiley.

DOUKHAN P., GHINDES M. (1980) : Etude des processus $X_n = f(X_{n-1}) + \epsilon_n$. Thèse de 3ème cycle. ORSAY. FRANCE.

ENGLE R.F. (1982) : Autoregressive Conditional heteroscedasticity with estimates of the variance of United Kingdom Inflation. Econometrica. Vol. 50. n° 4. 987-1007.

ERTEL J.E., FOWLKES E.B. (1976) : Some algorithmes for linear spline and piecewise multiple linear regression. J.A.S.A.

71. 640-648.

- FAURRE P. (1970) : Identification par minimisation d'une Representation Markovienne de processus aléatoires. Symposium on Optimization. Lectures Notes in Mathematics. 132. Springer Berlin. 87-107.
- FAURRE P., CLERGET M., GERMAIN F. (1979) : Operateurs Rationnels positifs. DUNOD - FRANCE.
- FEIGIN P.D., TWEEDIE T.W. (1985) : Random coefficient autoregressive processes : a markov chain analysis of stationary and finiteness of moments. J.T.S.A. vol. 6. n° 1. 1-14.
- FLIESS M., NORMAND-CYROT D. (1980) : Vers une approche algébrique des systèmes non linéaires en temps discret, in "Analysis and optimisation of systems". A. Bensoussan et J.L. LIONS Ed. Lect. Not. Cont. Info. Sci. 28. Springer Verlag. Berlin. 594-603.
- FUNAHASHI Y. (1979) : An observable canonical form of discrete time bilinear systems. I.E.E.E. Trans. Auto. Control. Vol. AC. 24/5. 802-803.
- GABR M.M., SUBBA RAO T. (1981) : The estimation and prediction of subset bilinear time series models with applications. J.T.S.A. Vol. 2. n° 3. 155-171.
- GODFREY M.D. (1965) : An explanatory study of the bispectrum of economic time series. J.R.S.S. Serie C. 14. 48-69.
- GRANGER C.W.J., ANDERSEN A.P. (1978 a) : An introduction to bilinear time series analysis. Vandenhoeck and Ruprecht. Gottingen.

- GRANGER C.W.J., ANDERSEN A.P. (1978 b) : On the invertibility of time series models. Stoch. Proc. and their appl. 8. 27-92.
- GRANGER C.W.J., ANDERSEN A.P. (1978 c) : Non linear time series analysis, in Applied time series analysis, ed. F. FINDLEY. Academic Press. 25-38
- GUEGAN D. (1979) : Tests d'adéquation pour des processus stationnaires. Prépublications de l'Université Paris-Nord. n° 10.
- GUEGAN D. (1980) : Tests of adequacy for ARMA models and tests of separated hypothesis, in Stochastic Systems : the Mathematics of filtering and identification and Applications. Eds. M. HAZEWINKEL et J.C. WILLEMS. D. REIDEL publ. Cie. 291-297.
- GUEGAN D. (1981 a) : Etude d'un modèle non linéaire, le modèle superdiagonal d'ordre 1. C.R.A.S. t. 293. Serie I. 95-98.
- GUEGAN D. (1981 b) : Etude de modèles bilinéaires. Existence, stationnarité, inversion, moments. Document de Travail non publié, (17 pages).
- GUEGAN D. (1981 c) : Processus ARMA multidimensionnels. Formalisme. Representation markovienne. Problèmes d'identification. Prépublication U.P.N. n° 29.
- GUEGAN D. (1982 a) : Modèles non linéaires dans les séries chronologiques. Document de travail non publié (28 pages).
- GUEGAN D. (1982 b) : Frame of study for non linear models. Document de travail non publié (22 pages).
- GUEGAN D. (1983 a) : Cadre d'étude pour des modèles non linéaires.

C.R.A.S. t. 296. Serie I. 167-170.

GUEGAN D. (1983 b) : Une condition d'ergodicité pour des modèles bilinéaires à temps discret. C.R.A.S. t. 297. Serie I. 537-540.

GUEGAN D. (1984 a) : Tests de modèles non linéaires, in Proceedings of the 3rd Franco Belgian meeting of statisticians. Nov. 82. BRUXELLES. 45-65.

GUEGAN D. (1984 b) : Propriétés de mélange et Théorème central limite. Quelques résultats dans le cas markovien. Applications aux modèles bilinéaires discrets. Prepublications U.P.N. n° 53 (30 pages).

GUEGAN D. (1986) : Représentation ℓ -markovienne et existence d'une représentation affine en l'état des modèles bilinéaires. C.R.A.S. t. 302. Serie I. n° 7. 289-282.

GUEGAN D. (1987) : Different representations of bilinear models. J.T.S.A. Vol. 8. n° 4. 389-408.

GUEGAN D., PHAM DINH T. (1987 a) : Minimalité et inversibilité des modèles bilinéaires à temps discret. C.R.A.S. Serie I. t. 448. 159-162.

GUEGAN D., PHAM DINH T. (1987 b) : A note on the estimation of the parameters of the diagonal bilinear models by the least squares method. A paraître dans Scand. Journ. of Stat. Theory and Appl.

GUEGAN D., PHAM DINH T. (1987 c) : Tests de scores pour des modèles bilinéaires. (en préparation)

HAGGAN V., HERAVI S.M., PRIESTLEY M.B. (1984) : A study of

application of state - dependent models in non linear time series analysis. J.T.S.A. Vol. 5. 69-102.

- HAGGAN V., OZAKI T. (1981) : Modelling non linear random vibrations using an amplitude - dependent autoregressive time series models. Biometrika. 68.1. 189-196.
- HANNAN E.J. (1969) : The identification of vector mixed autoregressive moving average systems. Biometrika. 56. 223-225.
- HANNAN E.J. (1979) : The statistical theory of linear systems, in Developments in Statistics. Vol. 2, ed. by KRISNAYAH. 83-121. Academic Press. New-York.
- HANNAN E.J. (1982) : A note on bilinear time series models. Stoch. Proc. and their Appl. 12. 221-224.
- HANNAN E.J., HEYDE C.C. (1972) : On limit theorems for quadratic functions of discrete time series. Ann. Math. Stat. Vol. 4. n°.6. 2058.2066.
- HIDA T. (1980) : Brownian Motion. Springer Verlag. New York.
- HINNICH M.J. (1982) : Testing for gaussianity and linearity of a stationary time series. J.T.S.A. 3. 169-176.
- HO, B.L., KALMAN R.E. (1966) : Effective construction of linear state - variables models from input/output functions. Regelungstechnik. 14. 545-548.
- HUBER P.J., KLEINER B., GASSER T. (1971) : Statistical methods for investigating phase relations in stationary stochastic processes. I.E.E.E. Trans. on Audio and Electroacoustics. Vol. AU. 19-20. 78-86

- IBRAJIMOV I.A., ROZANOV Y. (1974) : Processus Aléatoires Gaussiens. MIR. MOSCOU.
- IRVING M., CROUCH C.C. (1984) : Hamiltonian realization of finite Volterra series. Technical Report of Control Center. Coventry.
- JONES D.A. (1978) : Non linear autoregressive processes. Proc. Roy. Soc. London. A-360. 71-95.
- KALMAN R.E., FALB P.L., ARBIB M.A. (1969) : Topics in Mathematical System Theory. Mc Graw Hill. New York.
- KEENAN D.M. (1985) : A tukey non additivity type test for time series non linearity. Biometrika. 72,1. 39-44.
- KESTEN H., SPITZER F. (1978) : Convergence in distribution of products of random variables. Z.W. 67. 363-386.
- KLIMKO L.A., NELSON P.I. (1978) : On conditional least squares estimation for stochastic process. Annals. of Stat. Vol. 6. n° 3. 629-642.
- LEONOV V.P., SHIRYAEV A.W. (1959) : On a method of calculations of semi invariants. Theor. Prob. Appl. 4. 319-329.
- LI K. (1984) : On the autocorrelation structure and identification of some bilinear time series. J.T.S.A. Vol. 5. n° 3. 173-181.
- LIM K.S., TONG H. (1983) : A statistical approach to difference - delay equation modelling in ecology. Two cases studies. J.T.S.A. Vol. 4. n° 4. 239-267.

- MARAVALL A. (1983) : An application of non linear time series forecasting. Journ. Bus. Econ. Stat. Vol. 1. n° 1. 66-74.
- MASANI P., WIENER W. (1959) : Non linear prediction, in Probability and statistics. Ed. U. GRENANDER. WILEY.
- Mc LEOD, LI W.K. (1983) : Diagnostic cheking ARMA time series models using squared residual autocorrelations. J.T.S.A. Vol. 4. n° 4. 269-273.
- MILHØJ A. (1985) : The moment structure of ARCH processes. Scand. J. Stat. 12. 281-292.
- MOLHER R.R. (1970) : Natural Bilinear control processes. I.E.E.E. Trans. Syst. Sci. Cybern. Vol. S.C.C.G. 192-197.
- MOLHER R.R. (1973) : Bilinear Control processes with applications to engineering, ecology and medecine. Academic Press. New York.
- MOLHER R.R., RINK R.E. (1969) : Control with a multiplicative model. TRANS. A.S.M.E. Vol. 91.
- MURPHY D.N.P. (1979) : Controllability of a discrete time bilinear system. I.E.E.E. Trans. Auto. Control. AC 24/6. 974-975.
- NELSON J.Z., VAN NESS J.W. (1973) : Formulation of a non linear predictor. Technometrics. V. 15. 11-12.
- NEUDECKER H. (1969) : Some theorems on matrix differentiation with special reference to Kronecker matrix products. J. Amer. Stat. Ass. 64. 953-963.
- NEVEU J. (1968) : Processus Aléatoires Gaussiens. Cours de Montréal. P.U.F.

- NICHOLLS D.F., QUINN B.G. (1982) : Random coefficient autoregressive models. An introduction. Lectures notes in Statistics n° 11. Springer Verlag.
- NORMAND CYROT D. (1983) : Théorie et pratique des systèmes non linéaires en temps discret. Thèse d'état de Physique. ORSAY n°2687.
- NUMMELIN E., TUOMINEN P. (1982) : Geometric ergodicity of Harris recurrent Markov chains with application to renewal theory. Stoch. Proc. and Appl. 12. 187-202.
- OZAKI T. (1980) : Non linear time series models for non linear random vibrations. J. Appl. Prob. 17. 84-93.
- OZAKI T. (1982) : The statistical analysis of perturbed limit cycle processes using non linear time series models. J.T.S.A. Vol. 3. n° 1. 29-41.
- OZAKI T. (1985 a) : Statistical identification of storage models with application to stochastic hydrology. Water Ressources Bull. 21. n° 4. 663-675.
- OZAKI T. (1985 b) : Non linear time series models and dynamical system, in Handbook of Statistics, vol. 5, publié par E.J. HANNAN, P.R. KRISNAIAH et M.M. RAO. 25-83.
- PEMBERTON J., TONG H. (1981) : A note on the distributions of non linear autoregressive stochastic models. Vol. 2. n° 1. 49-52.
- PETRUCELLI J.D., DAVIES N. (1986) : A portmanteau test for self exciting threshold autoregressive type non linearity in time series. Biometrika 73.3. 687-694.

- PHAM DINH T. (1985) : Bilinear markovian representation and bilinear models. Stoch. Processes and their appl. 20. 295-306.
- PHAM DINH T. (1986) : The mixing property of bilinear and generalised random coefficient autoregressive models. Stoch. Proc. and their appl. 23. 291-300.
- PHAM DINH T., TAT TRAN L. (1981) : On the first order bilinear time series models. J. Appl. Prob. 617-627.
- PRIESTLEY M.B. (1980) : State dependent models : a general approach to non linear time series analysis. J.T.S.A. Vol. 1. 47-71.
- QUENOUILLE M.H. (1949) : Approximate tests of correlation in time series, J.R.S.S. Serie B. 11. 68-84.
- QUINN B.G. (1982 a) : Stationarity and invertibility of simple bilinear models. Stoch. Proc. and their appl. 12. 225-230.
- QUINN B.G. (1982 b) : A note on the existence of strictly stationary solutions to bilinear equations. J.T.S.A. Vol.3. n° 4. 249-252.
- RISSANEN J. (1973) : Algorithms for triangular decomposition of block Hankel and Toeplitz matrices with application to factorizing positive matrix polynomials. Mathematics of Computations. 27. 147-154.
- ROBINSON P.M. (1977) : The estimation of a non linear moving average models. Stoch. Proc. and their appl. 1. 81-90.
- ROBINSON P.M. (1983) : Non parametric estimators for time series. J.T.S.A. Vol. 4. n° 3. 185-207.

- ROSENBLATT M. (1971) : Markov processes. Structure and asymptotic behaviour. Springer Verlag. New York.
- ROSENBLATT M. (1979) : Linearity and non linearity in time series prediction. Bull. Inst. Stat. Inst. 42.
- ROSENBLATT M. (1980) : Linear processes and bispectra. J. Appl. Prob. 17. 265-270.
- ROSENBLATT M. (1983) : Cumulants and cumulant spectra, in Handbook of statistics, Vol. 3, ed. D.R. BRILLINGER and P.R. KRISNAIAH. 369-382.
- ROSENBLATT M. (1985) : Stationary sequences and random fields. BIRKHAUSER. Boston.
- ROSENBLATT M., VAN NESS J.W. (1965) : Estimation of the bispectrum. Ann. Math. Stat. 36. 1120-1136.
- RUBERTI A., ISIDORI A., d'ALESSANDRO P. (1972) : Theory of bilinear dynamical models. I.S.I. Meeting. New-Delhi.
- SCHWARTZ L. (1965) : Méthodes mathématiques pour les Sciences Physiques. Hermann. Paris.
- SESAY S.A.O., SUBBA RAO T. (1986) : Yule walker Type Difference Equations for Higher Order Moments (and cumulants) for bilinear time series models. Technical report n'179. U.M.I.S.T.
- SHEPP L.A., SPLEPIAN D., WYNER A.D. (1980) : On prediction of moving average processes. The Bell System Technical Journal. 367-415.
- SHIRYAEV A.N. (1960) : Some problems in the spectral theory of

higher order moment. I. Theory Prob. and its appl.,
Vol. 5, 3. 205-284.

SONTAG E.D. (1970) : Polynomial response maps. Lect. Notes
Control Info. Sci. 13. Springer Verlag. BERLIN.

SONTAG E.D. (1979) : Realization theory of discrete time non
linear systems. I. The bounded case. I.E.E.E. Trans.
Circuits. Syst. 26. 342-356.

STENSHOLT B. K., SUBBA RAO T. (1987) : On the theory of
multivariate time series Models. Technical Report n°183.
U.M.I.S.T.

STENSHOLT B.K., TJOSTHEIM D. (1987) : Multiple bilinear time series
models. J.T.S.A. Vol. 8. n° 2. 221-233.

SUBBA RAO T. (1981) : On the theory of bilinear time series models.
J.R.S.S. Serie B. Vol. 43. 2. 224-255.

SUBBA RAO T. (1983) : The bispectral analysis of non linear sta-
tionary time series with reference to bilinear time series
models, in Handbook of statistics. Vol. 3, published by
D.R. BRILLINGER et P.R. KRISNAIAH. 293-319.

SUBBA RAO T., GABR M.M. (1980) : A test for linearity of statio-
nary time series. J.T.S.A. Vol. 1. n°1. 145-158.

SUBBA RAO T., GABR M.M. (1984) : An introduction to bispectral
analysis and bilinear time series models. Lectures Notes
in statistics. 24. Springer Verlag.

SUBBA RAO T., NUNES A.M.D. (1985) : Identification of non linear
(quadratic) systems using higher order spectra. Technical
report. n°174. U.M.I.S.T.

- SUSSMAN H.J. (1976) : Existence and uniqueness of minimal realisations of non linear systems. I. Initialised system. J. Math. Syst. Theory.
- TJOSTHEIM D. (1986) : Estimation in non linear time series models. Stoch. Proc. and their appl. 21. 251-273.
- TONG H. (1981) : A note on a markov bilinear stochastic process in discrete time. J.T.S.A. Vol. 2. n°4. 279-284.
- TONG H. (1982) : An index of linearity in time series analysis. Technical. Report U.M.I.S.T. n° 155.
- TONG H. (1983) : Threshold models in non linear time series Analysis. Lectures notes in statistics. 21. Springer Verlag. NEW-YORK.
- TONG H., LIM K.S. (1980) : Threshold Autoregression, limit cycles and cyclical data. J.R.S.S. Serie B. 42. n°3. 245-292.
- TONG H., THANOON B., GUDMUNDSSON G. (1985) : Threshold time series modelling of two icelandic riverflow systems. Water Resources Bull. Vol. 21. n°4. 651.661.
- TSAY R.S. (1986) : Non linearity tests for time series. Biometrika 73. 2. 461-466.
- TUKEY J.W. (1949) : One degree of freedom for non additivity. Biometrics. 5. 232-242.
- TWEEDIE R.L. (1975) : Sufficient conditions for ergodicity and ergodicity of markov chains on a general state space. Stoch. Proc. and their Appl. 3. 385-403.

- TWEEDIE R.L. (1983 a) : The existence of moments for stationary markov chains. J.A.P. 191-196.
- TWEEDIE R.L. (1983 b) : Criteria for rate of convergence of Markov chains with applications to queueing and storage theory, in J.F.C. Kingman and G.E.H. Reuter Eds. Probability, Statistics and Analysis. Cambrigde University Press.
- VAN NESS W. (1966) : Asymptotic normality of bispectral estimates. Ann. Math. Stat. 37. 1257-1272.
- VILLEGAS C. (1976) : On a multivariate central limit theorem for stationary bilinear processes. Stoch. Proc. and their Appl. 4. 121.133.
- WANG S.R., AN H.Z., TONG H. (1983) : On the distribution of a simple stationary bilinear process. J.T.S.A. Vol. 4. n'3. 209-216.
- WEISS A.A. (1984) : ARMA models with ARCH. errors. J.T.S.A. Vol.5. n' 2, 129-143.
- WHITTLE H. (1969) : A view of stochastic control theory. J.R.S.S. A. 132. 320-334.
- WIENER N. (1958) : Non linear problems in random theory. M.I.T. Press.

APPENDICE

ARTICLES PUBLIES

THÉOREME 3. - *Étant donné un processus (Y_t) vérifiant (1), il est stationnaire en loi si (ε_t) est stationnaire en loi.*

$Y_t(a)$ est alors stationnaire en loi, et l'on passe à la limite. ■

PROPRIÉTÉS DES PROCESSUS (Y_t) DÉFINIS PAR (1). - *Moments d'ordre k de (Y_t) .* - Pour k impair, $EY_t^k = 0$ et pour k pair, $EY_t^k = b^k E[\varepsilon_t^k] E[Y_{t-2}^k] + w_t$, où w_t contient les termes des moments de ε_t et de Y_t d'ordre inférieur à k . Supposons les ε_t i.i.d. et bornées : si $S = \text{ess sup } |\varepsilon_t|$, $|Y_t| \leq |b|S|Y_{t-1}| + S$ et si $|b|S \leq 1$, $|Y_t| \leq S/(1 - |b|S)$.

THÉOREME 4. - *Si S est fini, et $|b|S < 1$, (Y_t) est borné. Si S est infini, (ce qui se produit si les ε_t sont gaussiennes), alors (Y_t) n'a pas tous ses moments finis.*

Bruit blanc. Définition 1. - On appelle bruit blanc une suite de variables aléatoires, de même variance, centrées et décorrélées.

Régularité et singularité. Définition 2. - Si $H_{-}(Y) = \{0\}$, le processus (Y_t) est dit régulier linéaire, et si $H_{-}(Y) = H_{+}(Y) = H$, le processus (Y_t) est singulier linéaire. Si $\mathcal{F}_{-}(Y) = \{\emptyset, \Omega\}$, alors le processus (Y_t) est dit régulier fort, et si $\mathcal{F}_{-}(Y) = \mathcal{F}_{+}(Y) = \mathcal{F}$, le processus (Y_t) est singulier fort.

Par le théorème de décomposition de Wold (cf. [2]), on sait que tout processus (Y_t) régulier linéaire et stationnaire au second ordre s'écrit sous la forme :

$$(3) \quad Y_t = \sum_{n=0}^{\infty} \theta_n \varepsilon_{t-n},$$

où les coefficients θ_n sont non aléatoires, et où (ε_t) est un bruit blanc, appelé processus d'innovation du processus (Y_t) au sens de Wold.

Pour le processus (Y_t) défini par (1), $\forall k \neq 0$, $\text{cov}(Y_t, Y_{t-k}) = 0$; les processus décorrélés sont réguliers linéaire, donc :

PROPOSITION 2. - *Le processus (Y_t) défini par (1) est un bruit blanc, il est régulier linéaire, c'est son propre processus d'innovation au sens de Wold.*

PROPOSITION 3. - *Le processus (Y_t) défini par (1) est un processus régulier fort. En effet $\mathcal{F}_{-}(Y) \subseteq \mathcal{F}_{-}(\varepsilon) = \{\emptyset, \Omega\}$.* ■

Il existe des relations entre les processus « linéaires » et les processus « forts » (sans condition de stationnarité).

PROPOSITION 4. - *Tout processus (Y_t) régulier fort est régulier linéaire :*

$$\sigma(H_t(Y)) = \mathcal{F}_t(Y) \quad \text{donc} \quad \forall t, \quad \sigma(H_{-}(Y)) \subseteq \sigma(H_t(Y)) = \mathcal{F}_t(Y)$$

et :

$$\sigma(H_{-}(Y)) \subseteq \bigwedge_t \mathcal{F}_t(Y) = \mathcal{F}_{-}(Y) = \{\emptyset, \Omega\}.$$

PROPOSITION 5. - *Tout processus (Y_t) singulier linéaire est singulier fort :*

$$\forall t, \quad H_t(Y) = H, \quad \text{or} \quad \mathcal{F}_t(Y) = \sigma(H_t(Y)) = \mathcal{F},$$

donc :

$$\bigwedge_t \mathcal{F}_t(Y) = \mathcal{F} \quad \text{et} \quad \forall \mathcal{F}_t(Y) = \mathcal{F}. \quad \blacksquare$$

PROPOSITION 6. - *Il existe des processus stationnaires au second ordre :*

- (i) réguliers linéaires et singuliers forts;
- (ii) singuliers forts et non singuliers linéaires.

(a) Soit ε une variable $\mathcal{N}(0,1)$, considérons le processus (Y_t) défini par $\forall t < 0$, $Y_t = \varepsilon^{-t} - E(\varepsilon^{-t})$.

LEMME 1. - *Soit $K_t = \overline{\text{sp}} \{Y_u, t \leq u < 0\}$, alors $\dim K_t = t$.*

Ceci résulte du fait que la v. a. ε^n n'est pas combinaison linéaire de $1, \varepsilon, \varepsilon^2, \dots, \varepsilon^{n-1}$. ■

Posons $I_t = K_t \ominus K_{t-1}$, espaces complémentaires dans L^2 , alors $\dim I_t = 1$.

Soit $X_t = Y_t / \|Y_t\|$ une base de l'espace vectoriel I_t ; par construction le processus (X_t) est

non corrélé et centré donc (X_t) est régulier linéaire. Or $\forall t, \mathcal{F}_t(X) = \mathcal{F}(\varepsilon)$ donc $\mathcal{F}_{-t}(X) = \mathcal{F}(\varepsilon) = \mathcal{F}$. La σ -algèbre associée au processus (X_t) est constante, donc (X_t) est singulier fort, d'où (i).

(b) Soit φ une v. a. indépendante de ε uniformément répartie sur $[0, 2\pi]$ et ω un réel fixé ($\neq 0, \text{ mod } \pi$). On considère le processus suivant :

$$\begin{cases} i < 0, & Z_i = \varepsilon \sin(\omega t + \varphi), \\ i = 1, & Z_1 = (\varepsilon^2 - 1) \sin(\omega t + \varphi), \\ i > 1, & Z_i = \varepsilon \sin(\omega t + \varphi). \end{cases}$$

Si $A = H_t(Z)$ pour $t < 0$ et $B = H_t(Z)$ pour $t > 0$, alors $H_{-t}(Z) = A$ et $H_{+t}(Z) = B$, on a $A \neq B$. Par contre, pour tout $t, \mathcal{F}_t(Z) = \sigma(\varepsilon)$. Donc Z , stationnaire au second ordre, est singulier fort sans être singulier linéaire, d'où (ii).

Par analogie avec l'innovation au sens de Wold, posons :

DEFINITION 3. - Le processus (ε_t) défini par (3) est appelé *processus d'innovation au sens fort* si $\forall t, \mathcal{F}_t(\varepsilon) \neq \mathcal{F}_{t+1}(\varepsilon)$.

Considérons le processus (ε_t) suivant : $i \leq 0, \varepsilon_i \sim N(0, 1)$ et $i > 0, \varepsilon_i = 1/\sqrt{2}(\varepsilon_{i-1}^2 - 1)$. Ce processus est stationnaire au second ordre, mais il n'est pas stationnaire en loi. C'est un processus régulier linéaire car $H_{-t}(\varepsilon) = \{0\}$, et régulier fort car $\mathcal{F}_{-t}(\varepsilon) = \{\emptyset, \Omega\}$. C'est un processus d'innovation au sens de Wold, en effet $H_{t+1}(\varepsilon) = H_t(\varepsilon) \oplus I_{t+1}(\varepsilon)$, mais ce n'est pas un processus d'innovation au sens fort $\forall t > 0, \mathcal{F}_t(\varepsilon) = \mathcal{F}_0(\varepsilon)$.

On constate donc que parmi les processus réguliers, il convient de distinguer les processus stationnaires au second ordre des processus stationnaires au second ordre et non stationnaires en loi et parmi ces deux catégories, ceux qui sont des processus d'innovation au sens de Wold et forts, de ceux qui sont processus d'innovation au sens de Wold et non forts.

CONDITION D'INVERTIBILITÉ DU PROCESSUS (Y_t) DÉFINI PAR (1).

DEFINITION 4. - Le modèle (1) est dit *inversible* si et seulement si il existe une suite d'estimateurs $(\hat{\varepsilon}_t)$, consistants dans L^2 quand $t \rightarrow +\infty, \varepsilon_t$ étant fonction des seuls $Y_n, n \leq t$.

THEOREME 5. - Si $2b^2\sigma^2 < 1$, alors le processus (Y_t) défini par (1) est inversible.

Définissons $\hat{\varepsilon}_t$ par la relation de récurrence, $\hat{\varepsilon}_0 = 0$ et $\hat{\varepsilon}_t = -bY_{t-2}\hat{\varepsilon}_{t-1} + Y_{t-1}$, (4). Soit $e_t = \hat{\varepsilon}_t - \varepsilon_t = -bY_{t-2}e_{t-1} + \varepsilon_t$ et $\hat{\varepsilon}_t$ sont centrées donc $E(e_t) = 0, \hat{\varepsilon}_t$ est consistant si et seulement si $E(e_t^2) \rightarrow 0$. On montre que ceci est réalisé à l'aide du théorème d'ergodicité classique (cf. [3]) dès que $2b^2\sigma^2 < 1$. On a alors $E(e_t^2) < C\rho$, avec $\rho = b^2A, C$ constante. Ω étant un borélien de \mathbb{R}^n , on a alors le lemme suivant :

LEMME 2. - $\forall \alpha > 0, \forall k_0, \exists C', \exists \Omega', \mathbb{P}(\Omega') \geq 1 - \alpha, \forall t \geq k_0 :$

$$|\hat{\varepsilon}_t - \varepsilon_t| \leq C' \rho^t \text{ sur } \Omega'.$$

A t fixé, soit $\alpha_t = \alpha/2^t$, alors il existe Ω_t tel que $\mathbb{P}(\Omega_t) \geq 1 - \alpha_t$ et $|\hat{\varepsilon}_t - \varepsilon_t| < C' \rho^t$. Il suffit de prendre $\Omega' = \bigcap_t \Omega_t$. ■

THEOREME 6. - Si le processus (Y_t) défini par (1) est inversible, alors le processus (ε_t) a pour représentation :

$$(5) \quad \varepsilon_t = Y_t + \sum_{n=1}^t (-b)^n Y_{t-n} \prod_{j=1}^n Y_{t-j-1}.$$

On pose :

$$\varepsilon_t(a) = Y_t + \sum_{n=1}^{t-1-a} (-b)^n Y_{t-n} \prod_{j=1}^n Y_{t-j-1} + (-b)^{t-a} \prod_{j=1}^{t-1} Y_{t-j-1} \varepsilon_a$$

et on montre que la série $\eta_t = \varepsilon_t - \varepsilon_t(a)$ tend p. s. vers zéro quand a tend vers $-\infty$ et que la

série :

$$Y_t + \sum_{n=1}^{t-1-u} (-b)^n Y_{t-n} \prod_{j=1}^n Y_{t-j-1},$$

converge p. s. et a pour somme (5). ■

Notons que le processus (ε_t) défini par (5) est $\mathcal{F}_t(Y)$ -mesurable, on a alors $\mathcal{F}_t(\varepsilon) = \mathcal{F}_t(Y)$.

THÉOREME 7. — Si le processus (Y_t) défini par (1) est inversible alors il est son propre processus d'innovation au sens fort.

En effet si (Y_t) n'est pas processus d'innovation au sens fort, alors $\exists t$, $\varepsilon_{t+1} \in \mathcal{F}_{t+1}(Y) = \mathcal{F}_t(Y) = \mathcal{F}_t(\varepsilon)$ ce qui est impossible. ■

FONCTION DE VRAISEMBLANCE. — On veut calculer la fonction de vraisemblance du processus (Y_t) défini par (1) dans le cas où les v. a. ε_t sont gaussiennes. On suppose le modèle inversible et les $\hat{\varepsilon}_t$ construits à l'aide de la relation (4). Étant donné $(y_1, \dots, y_n) \in \mathbb{R}^n$, on considère l'application φ de \mathbb{R}^n dans \mathbb{R}^n telle que $\varphi(y_1, \dots, y_n) = (x_1, \dots, x_n)$ où $y_i = by_{i-2}x_{i-1} + x_i$ avec $x_0 = 0$.

Alors si on note $L_Y(y_1, \dots, y_n)$ la vraisemblance de (Y_t) calculée en (y_1, \dots, y_n) , on a : $L_Y(y_1, \dots, y_n | \varepsilon_0 = 0) = L_\varepsilon(x_1, \dots, x_n | \varepsilon_0 = 0)$, le jacobien de φ étant égal à 1. \mathcal{B} étant la σ -algèbre des boréliens de \mathbb{R}^n , on considère $B \in \mathcal{B}$ t. q. $B = \prod_{i=k}^n B_i$, $B_i = [x_i - \delta/2, x_i + \delta/2]$.

Posons $\varepsilon = (\varepsilon_k, \dots, \varepsilon_n)$ et $\hat{\varepsilon} = (\hat{\varepsilon}_k, \dots, \hat{\varepsilon}_n)$.

LEMME 3. — $\forall B \in \mathcal{B}$, $\exists \eta > 0$, $\exists k_0$, $\forall k \geq k_0$, $\forall n \geq n_0$:

$$|P(\hat{\varepsilon} \in B) - P(\varepsilon \in B)| < \eta.$$

On obtient ce lemme en appliquant le lemme 2 et en écrivant :

$$|P(\hat{\varepsilon} \in B) - P(\varepsilon \in B)| \leq P(|\hat{\varepsilon} - \varepsilon| \geq C \rho^t) + P(\hat{\varepsilon} \in F(B)),$$

où $F(B)$ représente la frontière du borélien B d'épaisseur $C \rho^t$ dont la mesure gaussienne est équivalente à $(\sqrt{2\pi}\sigma)^{-n} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=k}^n x_i^2\right\} \cdot (C\rho)^{n-k} \delta^n$. ■

THÉOREME 8. — Si les v. a. ε_t sont i. i. d. et gaussiennes centrées et si le modèle (1) est inversible, alors :

$$|\text{Log } L_\varepsilon(x_k, \dots, x_n | \varepsilon_0 = 0) - \text{Log } L_{\hat{\varepsilon}}(x_k, \dots, x_n | \varepsilon_0 = 0)| = O(1).$$

On remarque que à (x_1, \dots, x_n) fixé, $L_\varepsilon(x_1, \dots, x_n) = \lim_{\delta \rightarrow 0} \delta^{-n} P(\varepsilon \in B)$. On montre que $\log L_\varepsilon(x_1, \dots, x_{k-1})/n \xrightarrow{p} 0$ de même pour $L_{\hat{\varepsilon}}(x_1, \dots, x_{k-1})$ et on applique le lemme 3. ■

PROPOSITION 7. — Si les v. a. ε_t sont i. i. d. gaussiennes centrées et si le modèle (1) est inversible, sa log vraisemblance est équivalente à $\sum_{i=1}^n \hat{\varepsilon}_i^2$.

Ceci est dû à l'indépendance des v. a. ε_t et au théorème 8. ■

On obtient alors un résultat similaire à celui obtenu dans le cas des modèles ARMA. Il permet en particulier de rechercher les estimateurs du maximum de vraisemblance des paramètres du modèle.

(*) Remise le 22 juin 1981

[1] C. W. J. GRANGER et A. ANDERSEN, *An Introduction to Bilinear Time Series Models*, Wanderhoeck and Reprecht, Gottingen, 1978

[2] YU. A. ROZANOV, *Stationary Random Process*, Holden day, 1967.

[3] I. BREITMAN, *Probability and Stochastic Processes*, 1969.

PROBABILITÉS. — *Cadre d'étude pour des modèles non linéaires.* Note (*) de Dominique Guégan, présentée par Robert Fortet.

A partir de la présentation d'un certain nombre de propriétés probabilistes du processus bilinéaire $X(t)$ régi par l'équation $X(t) = \alpha \varepsilon(t-2)\varepsilon(t-1) + \varepsilon(t)$ où $\varepsilon(t)$ est une suite de v. a. i. gaussiennes centrées, nous montrons comment les techniques hilbertiennes fondées sur la seule covariance sont insuffisantes pour l'étude de ce type de modèle. On peut alors étudier ces modèles dans le cadre plus large des chaos de Wiener. Dans ce contexte-là nous proposons un essai de classification des processus englobant les processus linéaires et non linéaires. Nous nous intéressons principalement aux processus réguliers, singuliers, processus d'innovations.

PROBABILITY THEORY. — Frame for the Study of non Linear Processes.

From the study of probabilistic properties of the bilinear process $X(t)$ defined by the equation $X(t) = \alpha \varepsilon(t-2)\varepsilon(t-1) + \varepsilon(t)$, where $\varepsilon(t)$ are i. i. d. gaussian centered random variables, we show why the hilbertian techniques based only on the covariance are non sufficient for the study of this kind of models. So we can consider a new approach for these models, in the frame of Wiener's chaos. In this context we suggest a classification of linear and non linear processes. We are particularly interested in the regular, singular and innovation processes.

INTRODUCTION. — La théorie classique du modèle linéaire est une théorie hilbertienne : les processus étudiés sont stationnaires au second ordre, réguliers (cf. [1] et [2]), et possèdent la décomposition de Wold suivant des v. a. non corrélées. Toute procédure statistique sur ces modèles ne se fait qu'à partir des covariances et des quantités déduites de celles-ci. Cette attitude, dans le cas de modèles non linéaires, (cf. [4], [5], [6]) amène à oublier des structures : le processus $X(t) = \varepsilon(t) + \alpha \varepsilon(t-1)\varepsilon(t-2)$ où $\varepsilon(t)$ est un bruit blanc gaussien (b. b. g.), en tant que processus au second ordre est un b. b. alors que deux observations de ce processus sont fortement dépendantes. La covariance apparaît insuffisante pour caractériser l'ensemble des structures de ce processus. Nous allons donc nous attaquer aux points de vue contradictoires suivants : sortir du cadre second ordre tout en gardant un cadre hilbertien. Pour cela nous allons étendre deux outils fondamentaux dans la théorie du modèle linéaire : l'espace linéaire engendré par un b. b. $\varepsilon(t)$ sera remplacé par un espace plus gros isomorphe à un chaos de Wiener et un opérateur \mathcal{B} analogue à l'opérateur de translation B sera défini sur l'espace linéaire.

RAPPELS ET NOTATIONS. — Nous nous plaçons dans le cadre restrictif suivant : $\varepsilon(t)$, $t \in \mathbb{Z}$ est un b. b. g.. On note $\mathcal{F} = \mathcal{F}(\varepsilon)$ la σ -algèbre engendrée par tous les $\varepsilon(t)$, \mathcal{F}_t celle engendrée par les $\varepsilon(u)$, $u \leq t$, $\mathcal{F}_{-x} = \bigwedge_t \mathcal{F}_t$, $\mathcal{F}_{+x} = \bigvee_t \mathcal{F}_t$. Le chaos de Wiener \mathcal{H} construit sur les $\varepsilon(t)$, $t \in \mathbb{Z}$ est défini ainsi : $\xi = \bigoplus_{n \geq 0} H^{\otimes n}$ où $H^{\otimes 0}$ est l'ensemble des constantes et où $\forall n \geq 1$, $H^{\otimes n} = \overline{\text{sp}} \{ \varepsilon(t_1) \otimes \dots \otimes \varepsilon(t_n) \}$ est le sous-espace fermé engendré par les combinaisons linéaires des produits symétriques formels $\varepsilon(t_1) \otimes \dots \otimes \varepsilon(t_n)$, $\forall i, i = 1, \dots, n$, $t_i \in \mathbb{Z}$, muni de la norme

$$\langle \varepsilon(t_1) \otimes \dots \otimes \varepsilon(t_n) \rangle_{H^{\otimes n}} \leq \sum_{\sigma} \prod_{k=1, n} \text{cov}(\varepsilon(t_k), \varepsilon(\sigma(t_k)))$$

où σ est une permutation parcourant le n -ième groupe symétrique. On note \mathcal{H} l'espace de Hilbert, complété de ξ par cette norme. On munit l'espace $L^2(\mathcal{F})$ d'une structure de chaos de Wiener à l'aide de l'isomorphisme $\Psi^{-1} : \mathcal{H} \rightarrow L^2(\mathcal{F})$ tel que si on choisit $X \in \mathcal{H}$ de la forme $X = \exp \odot h$, $h \in H^{\otimes 1}$, alors $\Psi^{-1}(X) = \exp(X - 0,5 E(X^2))$ cf. [3]. On identifiera \mathcal{H} et $L^2(\mathcal{F})$. Notons \mathcal{H}_t le chaos de Wiener construit sur les $\varepsilon(u)$, $u \leq t$. La norme définie sur ces espaces

est identique à la norme L^2 , et un processus $X(t)$ qui appartient à \mathcal{H}_t est \mathcal{F}_t -mesurable. L'image par Ψ^{-1} du vecteur de base $\varepsilon(t_1) \otimes \dots \otimes \varepsilon(t_n)$ vaut $h_n(\varepsilon(t_1), \dots, \varepsilon(t_n))$ où h_n désigne le polynôme d'Hermite de degré n . Nous utiliserons cette notation malgré l'abus qui consiste à noter de la même façon le polynôme d'Hermite dont les arguments sont égaux ou différents. On obtient donc une base de $H^{\otimes n}$ constituée de l'ensemble des polynômes d'Hermite en $\varepsilon(t)$ de degré n , en nombre dénombrable, numérotés par p , soit $\{h_{np}\}$. Enfin, f étant une fonction de \mathbb{R}^n dans \mathbb{R} , on note $f(\underline{\varepsilon}(t)) = f(\varepsilon(t), \varepsilon(t-1), \dots, \varepsilon(t-n))$. Nous considérerons par la suite des processus $X(t)$ appartenant à \mathcal{H}_t qui admettent la décomposition :

$$X(t) = \sum_{n \geq 0} \sum_{p=1, \dots, x} \alpha_{np} h_{np}(\underline{\varepsilon}(t)), \quad \alpha_{np} \in \mathbb{R}.$$

LES MODÈLES BILINÉAIRES COMME ÉLÉMENTS DE \mathcal{H}_t . — Le cadre qui vient d'être défini est bien celui dans lequel se glissent les modèles bilinéaires :

(a) Les modèles linéaires. Ils appartiennent à $H_t^{\otimes 1}$.
 (b) les modèles non linéaires, type moyenne mobile : processus qui sont des polynômes en $\varepsilon(t), \varepsilon(t-1), \dots$, de degré fini, s'écrivant sous la forme $X(t) = \sum_{i \in I} \prod_{k=1, \dots, d} \alpha_{ik} \varepsilon_{i-k}$, $I \subset \mathbb{N}$. Ils appartiennent à un nombre fini de composantes du chaos, $X(t) \in H_t^{\otimes 0} + H_t^{\otimes 1} + \dots + H_t^{\otimes d}$.

(c) Les modèles non linéaires à partie autorégressive : processus possédant des contributions du type $\alpha_{kl} X(t-k) \varepsilon(t-l)$, $k, l \in I \subset \mathbb{N}$. Aucune somme finie $H_t^{\otimes 0} + H_t^{\otimes 1} + \dots + H_t^{\otimes n}$ ne va suffire à les représenter. Ils appartiennent à l'espace \mathcal{H}_t , tout entier. Considérons les représentations de certains de ces modèles :

(i) $X(t) = \varepsilon(t) + aX(t-1)\varepsilon(t-1)$, $|a| < 1$, $t \in \mathbb{N}$, $X(0) = 1$: $X_1(t) = \varepsilon(t) + a\varepsilon(t-1)$, $\forall n > 1$, $X_n(t) = a[\varepsilon(t-1)\varepsilon(t-2) \dots \varepsilon^2(t-n+1) + \varepsilon(t-1) \dots \varepsilon(t-n)]$.

(ii) $X(t) = \varepsilon(t) + bX(t-2)\varepsilon(t-1)$, $|b| < 1$, $t \in \mathbb{N}$, $X(0) = 1$: $X_1(t) = \varepsilon(t)$, $\forall n > 1$, $X_n(t) = b\varepsilon(t-1)\varepsilon(t-2) \dots \varepsilon(t-2n-1)\varepsilon(t-2n)$.

(iii) $X(t) = \varepsilon(t) + cX(t-1)\varepsilon(t-2)$,
 $t \in \mathbb{N}$, $X(0) = 1$, $|c| < 1$: $X_1(t) = \varepsilon(t) + c\varepsilon(t-3)$,
 $\forall n > 1$, $X_n(t) = c^{n-1}\varepsilon(t-2)\varepsilon(t-3) \dots \varepsilon(t-n+1)\varepsilon(t-n) + c\varepsilon(t-2)\varepsilon(t-3) \dots \varepsilon(t-n+1)$.

Tout processus bilinéaire $X(t)$ appartient à \mathcal{H}_t et a une décomposition sur le chaos de Wiener. Nous allons considérer le problème réciproque suivant. Posons $\mathcal{F}_t(X) = \sigma(X(u),$

$u \leq t)$, $\mathcal{F}_{-x}(X) = \bigwedge \mathcal{F}_t(X)$, $\mathcal{F}_{+x}(X) = \bigvee \mathcal{F}_t(X)$, on a, cf. [2] :

DÉFINITIONS. — Soit $X(t)$ un processus stationnaire en loi. $X(t)$ est régulier fort si et seulement si $\mathcal{F}_{-x}(X) = \{\emptyset, \Omega\}$. $X(t)$ est singulier fort si et seulement si $\mathcal{F}_{-x}(X) = \mathcal{F}_{+x}(X) = \mathcal{F}$.

Pour un processus régulier fort $X(t)$ de carré sommable, existe-t-il un b. b. g. $\varepsilon(t)$ tel que $X(t)$ soit \mathcal{F}_t -mesurable? Le problème ainsi posé n'a pas encore de solution : étant donné une suite de v. a. $X(t)$, on peut considérer à chaque temps t , une σ -algèbre $\mathcal{F}_t(\varepsilon)$ (où ε est une v. a. gaussienne centrée) qui ne dépend ni de ce qui se passe avant t , ni après, donc qui ne permet pas de rendre compte de l'évolution globale du processus. Par contre, disposant d'un b. b. g. $\varepsilon(t)$ et de l'espace \mathcal{H} associé, nous allons construire sur cet espace des processus $Y(t)$ stationnaires en loi $\varepsilon(t)$ -mesurables. En cela nous nous plaçons dans un contexte un peu différent du cas linéaire pour lequel : soit on se donne le processus Y et on obtient la

représentation de Wold, soit on se donne un b. b. $\varepsilon(t)$ et on se trouve dans l'une des situations suivantes :

- si $\varepsilon(t)$ est un b. b. g., on construit un processus stationnaire au second ordre et gaussien donc stationnaire en loi;
- si $\varepsilon(t)$ est une suite de v. a. non corrélées, on construit un processus stationnaire au second ordre, mais pas stationnaire en loi en général;
- si $\varepsilon(t)$ est un b. b. stationnaire en loi, alors on construit un processus stationnaire en loi.

OPÉRATEUR \mathcal{B} . — Nous disposons de deux approches pour le définir :

DÉFINITION. — 1. $\forall U \in \mathcal{H}_t, U = \sum_{n=0, \infty} \sum_{p=1, \infty} \lambda_{np} \varepsilon(t_1) \odot \dots \odot \varepsilon(t_n)$. On appelle \mathcal{B} , l'opérateur défini de \mathcal{H}_t dans \mathcal{H}_{t-1} tel que (1)

$$U = \sum_{n=0, \infty} \sum_{p=1, \infty} \lambda_{np} \varepsilon(t_1 - 1) \odot \dots \odot \varepsilon(t_n - 1).$$

2. $\forall X \in L^2(\mathcal{F}_t), X = \lim P_n(\underline{\varepsilon}(t))$ [limite de polynômes en $\varepsilon(t)$]. On appelle \mathcal{B} , l'opérateur défini de $L^2(\mathcal{F}_t)$ dans $L^2(\mathcal{F}_{t-1})$ tel que (2) $X = \lim P_n(\underline{\varepsilon}(t-1))$.

- PROPRIÉTÉS. — 1. Les écritures (1) et (2) sont compatibles car $\mathcal{B} \circ \Psi = \Psi \circ \mathcal{B}$.
- 2. \mathcal{B} est un opérateur unitaire.
 - 3. $\forall X \in L^2(\mathcal{F}_t)$, X et $\mathcal{B}X$ ont même loi.
 - 4. Si $\forall i, i=1, \dots, n, X(t_i) \in L^2(\mathcal{F}_t)$, les suites de v. a. $X(t_1), \dots, X(t_n)$ et $\mathcal{B}X(t_1), \dots, \mathcal{B}X(t_n)$, ont même loi.

CONSTRUCTION DU PROCESSUS $Y(t)$. — (a) Approche spectrale. L'opérateur \mathcal{B} étant unitaire, il a une représentation spectrale (théorème de Böchner). Notons : $\mathcal{B}^t = \int_{-\pi}^{\pi} e^{-i\lambda t} \varphi(d\lambda)$, où $\varphi(d\lambda)$ est la famille spectrale d'opérateur de projection sur \mathcal{H} . Soit $Y(0) \in \mathcal{H}_0$, alors $Y(t) = \mathcal{B}^{-t} Y(0)$ et :

$$(3) \quad Y(t) = \int_{-\pi}^{\pi} e^{i\lambda t} \varphi(d\lambda) Y(0) = \int_{-\pi}^{\pi} e^{i\lambda t} \mathcal{F}(d\lambda),$$

où $\mathcal{F}(d\lambda) = \varphi(d\lambda) Y(0)$ est une mesure aléatoire σ -additive sur $[-\pi, \pi]$ à valeurs dans \mathcal{H} telle que $E|\mathcal{F}(\Delta)|^2 \leq E|Y(0)|^2$ pour tout Δ borélien de $[-\pi, \pi]$ et $E(\mathcal{F}(\Delta)\mathcal{F}(\Delta')) = 0$ pour tous Δ et Δ' boréliens disjoints de $[-\pi, \pi]$.

PROPOSITION. — Le processus $Y(t)$ défini par (3) est stationnaire en loi.

Posons $\mathcal{F}(\Delta) = E(\mathcal{F}(\Delta))^2$, Δ borélien de $[-\pi, \pi]$.

DÉFINITION. — $\mathcal{F}(\Delta)$ est la mesure spectrale du processus $Y(t)$ sur \mathcal{H}_t .

$\mathcal{F}(\Delta)$ permet de calculer les covariances du processus $Y(t) : C(s) = E(Y(t)Y(t+s))$, $\forall s \geq 0$, qui ne dépendent pas de t . $C(s)$ est le s -ième coefficient de Fourier de $\mathcal{F}(\cdot)$. Mais $\mathcal{F}(\Delta)$ ne donne qu'une approche du processus $Y(t)$ au second ordre et ne suffit pas à la caractériser. Posons $\mathcal{G}(\Delta, \Delta') = E(\mathcal{F}(\Delta \setminus \Delta')\mathcal{F}(\Delta)\mathcal{F}(\Delta'))$, Δ, Δ' boréliens de $[-\pi, \pi]^2$.

DÉFINITION. — $\mathcal{G}(\Delta, \Delta')$ est la mesure spectrale au troisième ordre du processus $Y(t)$ sur \mathcal{H}_t . $\mathcal{G}(\Delta, \Delta')$ permet alors d'étudier les moments d'ordre 3 du processus $Y(t)$:

$$C(r, s) = E(Y(t)Y(t+r)Y(t+s)) = \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} e^{-i(\lambda r + \mu s)} \mathcal{G}(d\lambda, d\mu), \text{ ils ne dépendent pas de } t \text{ et } C(r, s) \text{ est le } (r+s)\text{-ième coefficient de Fourier de } \mathcal{G}(\cdot, \cdot).$$

(b) *Espaces d'innovations du processus Y(t).* Considérons la projection orthogonale $E(\cdot/\mathcal{F}_t) : \mathcal{H}_{t+1} \rightarrow \mathcal{H}_t$. Soit \mathcal{N}_{t+1} son noyau, alors $\mathcal{N}_{t+1} = \mathcal{H}_{t+1} - \mathcal{H}_t$, et l'on a :

THÉORÈME. — $\forall Y \in \mathcal{N}_{t+1} : 1, (4) Y = \sum_{k=0, \infty} Q_k \cdot \varepsilon^k(t+1)$ avec Q_k, \mathcal{F}_t -mesurable et $\sum_{k=0, \infty} m_k \cdot Q_k = 0, m_k$ étant le moment d'ordre k de $\varepsilon(t)$.

2) (5) $Y = \sum_{k=0, \infty} R_k \cdot \varepsilon^k(t+1), R_k \in \mathcal{H}_t$ et $\sum_{k=0, \infty} \lambda_k \cdot R_k = 0, \lambda_k$ dépendant des moments de $\varepsilon(t)$ jusqu'à l'ordre k .

PROPRIÉTÉS. — 1. Les écritures (4) et (5) se correspondent par l'isomorphisme Ψ .

2. $\forall t, \forall s, \mathcal{H}^{-s} \cdot \mathcal{N}_t = \mathcal{N}_{t+s}$.

3. $\mathcal{H}_0 = \bigoplus_{t \leq 0} \mathcal{N}_t$.

4. $\forall Y(t) \in \mathcal{H}_t, Y(t)$ a une représentation type moyenne mobile : (6) $Y(t) = \sum_{s \leq 0} \sum_{i, k} \lambda_{iks} \mathcal{S}_{ik}(t-s)$ où \mathcal{S}_{ik} est une base de \mathcal{N}_0 . Les écritures (3) et (6) sont deux représentations du processus $Y(t) \in \mathcal{H}_t$. Cet espace ne dépend que de $\varepsilon(t)$ et est de dimension infinie.

Si on veut disposer comme dans le cas linéaire d'un espace d'innovation de dimension finie, on procède de la manière suivante : on pose $V_{t, k} = E(X(t+k)/\mathcal{F}_t) - E(X(t+k)/\mathcal{F}_{t-1})$. Pour $k > 0, V_{t, k} \in \mathcal{N}_t$.

ESPACE D'INNOVATION. — L'espace $\mathcal{I}_t = \text{sp}\{V_{t, k}, k=1, \dots, n\}$ est appelé espace d'innovation en t .

Exemples. — 1. $X(t) = \sum_{k=0, p} a_k \varepsilon(t-k), \dim \mathcal{I}_t = 1$.

2. $X(t) = \sum_{k=0, p} a_k \varepsilon(t-k) + \sum_{l=1, q} b_l \varepsilon^2(t-l), \dim \mathcal{I}_t = 2$ si et seulement si les vecteurs $(a_0, \dots, a_p)'$ et $(b_1, \dots, b_q)'$ ne sont pas colinéaires.

3. $X(t) = \sum_{k=0, p} a_k \varepsilon(t_k-k) + \sum_{p=1, q} b_p \varepsilon^2(t_p-l) + \sum_{i, j} c_{ij} \varepsilon(t_i-i) \varepsilon(t_j-j), i, j \in \mathbb{I} \subset \mathbb{N}, \dim \mathcal{I}_t = 3$ ssi les vecteurs $(a_0, \dots, a_p)'$ et $(b_1, \dots, b_q)'$ étant choisis comme précédemment, les coefficients c_{ij} sont déterminés de telle manière que $t_i \neq t_k, t_l \forall k, \forall l$ et $t_j \neq t_k, t_l \forall k \forall l$.

4. $X(t) = \varepsilon(t) + \alpha X(t-2) \varepsilon(t-1) \dim \mathcal{I}_t = 2$.

(*) Remise le 20 décembre 1982.

[1] Y. A. ROZANOV, *Stationary Random Process*, Holden day, 1967.

[2] D. GUEGAN, *Comptes rendus*, 293, série I, 1981, p. 95.

[3] J. NEVEU, *Processus aléatoires gaussiens*, Cours de Montréal, 1968.

[4] M. B. PRIESTLEY, *J.T.S.A.*, 1, n° 1, 1980.

[5] T. SUBBA RAO, *J.R.S.S.*, 43, n° 2, 1981.

[6] T. SUBBA RAO et M. M. GABR, *J.T.S.A.*, 1, n° 1, 1980.

Département de mathématiques, C.S.P. U.P.N.,
avenue Jean-Baptiste-Clément, 93470 Villetaneuse.

PROBABILITÉS. — Une condition d'ergodicité pour des modèles bilinéaires à temps discret. Note (*) de Dominique Guégan, présentée par Robert Fortet.

A partir de la représentation markovienne des modèles bilinéaires :

$$X(t) = \sum_{i=1}^p \alpha_i X(t-i) + \sum_{j=1}^q \beta_j \varepsilon(t-j) + \varepsilon(t) + \sum_{k=0}^p \sum_{l=1}^q \delta_{kl} X(t-k-l) \varepsilon(t-l),$$

on donne une condition suffisante d'ergodicité pour ces modèles. On étudie d'abord le cas scalaire puis le cas vectoriel.

PROBABILITY THEORY. — A Condition of Ergodicity for Discrete Time Bilinear Models.

From the markovian representation of the bilinear models:

$$X(t) = \sum_{i=1}^p \alpha_i X(t-i) + \sum_{j=1}^q \beta_j \varepsilon(t-j) + \varepsilon(t) + \sum_{k=0}^p \sum_{l=1}^q \delta_{kl} X(t-k-l) \varepsilon(t-l),$$

we give a sufficient condition for ergodicity. First we study the scalar case, then the vectorial case.

I. Considérons le processus réel $Z(t)$ régi par l'équation :

$$Z(t) = aZ(t-1) + bZ(t-1)\varepsilon(t) + c\varepsilon(t) + d(\varepsilon^2(t) - 1),$$

où $\varepsilon(t)$ est un bruit blanc gaussien centré de variance 1 (b.b.g.). Ce processus est markovien et sa fonction de transition est donnée par :

$$p(r, u) = \mathbb{P}(Z(t) \leq u/a \mid Z(t-1) = r) \quad \text{où } u \in \mathbb{R}, \quad r \in \mathbb{R}.$$

Soit $f(v)$ la fonction de densité suivante : si $d \neq 0$:

$$f(v) = \frac{1}{\sqrt{2\pi(s^2 + 4d(d+v))}} \left[\exp\left(-\frac{1}{8d^2}\right) \left(\exp\{-s + \sqrt{s^2 + 4d(d+v)}\}^2 + \exp\{-s - \sqrt{s^2 + 4d(d+v)}\}^2 \right) \right],$$

si $d=0$:

$$f_0(v) = \frac{1}{\sqrt{2\pi s}} \exp\left\{-\frac{v^2}{2s^2}\right\}.$$

Si $v = u - r$ et $s = bZ(t-1) + c$, alors la densité de transition du processus $Z(t)$ est telle que :

* $p(r, u) = f(v)$ si $d > 0$ et si l'on se trouve dans l'un des cas suivants :

- (1) $u \geq r$;
- (2) $u < r$ et $|d| > |v|$;
- (3) $u < r$ et $|d| < |v|$ et $s > \pm \sqrt{|4d(d+v)|}$.

* $p(r, u) = -f(v)$ si $d < 0$ et si l'on se trouve dans l'un des cas suivants :

- (1) $u < r$;
- (2) $u \geq r$ et $|d| > v$;
- (3) $u \geq r$ et $|d| < v$ et $\beta > \pm \sqrt{|4d(d+v)|}$.

* $p(r, u) = f_0(v)$ si $d=0$ et $s > 0$.

* $p(r, u) = -f_0(v)$ si $d=0$ et $s < 0$.

* le cas $d=0$ et $s=0$ correspond au modèle AR(1).

La densité $p(r, u)$ du processus $Z(t)$ est absolument continue. On a alors le résultat suivant :

THÉORÈME 1. — *Étant donné le processus markovien réel $Z(t)$ défini par $Z(t) = aZ(t-1) + bZ(t-1)\varepsilon(t) + c\varepsilon(t) + d(\varepsilon^2(t)-1)$ où $\varepsilon(t)$ est un b.b.g., une condition suffisante d'ergodicité est donnée par $y < 1$ où $y = E|a + b\varepsilon(t)|$.*

Soit la norme suivante :

$$\gamma_x = E[|\|Z(t)\| - \|Z(t-1)\|| / |Z(t-1)| = |x|],$$

$$\gamma_x \leq E[|a + b\varepsilon(t)| |x| + |c\varepsilon(t) + d(\varepsilon^2(t)-1)|] |x|.$$

Posons $w = E|c\varepsilon(t) + d(\varepsilon^2(t)-1)|$, soit λ une constante positive plus grande que $-w$, si $y < 1$, on considère le compact :

$$K = \left\{ x \in \mathbb{R}; |x| < \frac{\lambda w}{1-y} \right\},$$

alors $\gamma_x < -\lambda$ pour $x \notin K$; d'autre part $\forall x \in \mathbb{R}$, $\gamma_x \leq w < \infty$. Il suffit alors d'appliquer le théorème 5.1 de R. L. Tweedie [4] pour conclure.

Notons que les processus considérés ici peuvent être généralisés à des processus régis par l'équation suivante :

$$Z(t) = aZ(t-1) + bZ(t-1)\varepsilon(t) + c\varepsilon(t) + d\varepsilon^2(t) + f.$$

Les densités de transition de ces processus ont la même forme que précédemment, le théorème 1 est encore vrai. Seule la valeur de w change et devient : $w = E|c\varepsilon(t) + d\varepsilon^2(t) + f|$.

II. On considère maintenant le processus $Z(t)$ n -vectoriel régi par l'équation :

$$Z(t) = AZ(t-1) + BZ(t-1)\varepsilon(t) + C\varepsilon(t) + D(\varepsilon^2(t)-1),$$

A et B étant des matrices (n, n) , C et D des vecteurs colonnes $(1, n)$.

On pose $Z(t) = (Z_i(t))_{1 \leq i \leq n}$, $AZ(t) = (\gamma_i Z_i(t))_{1 \leq i \leq n}$.

La fonction de transition du processus $Z(t)$ est donnée par :

$$p(r, u) = (p_i(r_i, u_i))_{1 \leq i \leq n}$$

où :

$$p_i(r_i, u_i) = \mathbb{P}(Z_i(t) \leq u_i / \gamma_i Z_i(t-1) = r_i, \forall i)$$

avec :

$$u = (u_i)_{1 \leq i \leq n} \in \mathbb{R}^n \quad \text{et} \quad r = (r_i)_{1 \leq i \leq n} \in \mathbb{R}^n.$$

La densité de transition du processus $Z(t)$ est alors n -vectorielle, chacune de ses composantes se calcule comme précédemment suivant le signe des composantes du vecteur D . Elles ont pour expression $f(v_i)$, $-f(v_i)$, $f_0(v_i)$, $-f_0(v_i)$ suivant le signe de d_i , $1 \leq i \leq n$, $v = (v_i)_{1 \leq i \leq n}$. La densité $p(r, u)$ est absolument continue, d'où le théorème :

THÉORÈME 2. — *Étant donné le processus $Z(t)$ n -vectoriel défini par l'équation :*

$$Z(t) = AZ(t-1) + BZ(t-1)\varepsilon(t) + C\varepsilon(t) + D(\varepsilon^2(t)-1),$$

où $\varepsilon(t)$ est un b.b.g., $A=(a_{ij})_{1 \leq i, j \leq n}$, $B=(b_{ij})_{1 \leq i, j \leq n}$ des matrices (n, n) , une condition suffisante d'ergodicité de ce processus est donnée par $z < 1$ avec :

$$z = \xi \left(\sum_{i,j} |a_{ij}| + \sqrt{\frac{2}{\pi}} \sum_{i,j} |b_{ij}| \right) \quad \text{où } \xi = \sup_i |x_i| \quad \text{et} \quad x_i = Z_i(t-1), \quad \forall i.$$

On obtient un résultat du même type que précédemment. Il suffit d'écrire :

$$\gamma_x < E \left[\sum_i |Z_i(t)| / |Z(t-1)| = |x| \right] - \sum_i |x_i| \quad \text{où } x = (x_i)_{1 \leq i \leq n} \in \mathbb{R}^n,$$

soit :

$$\gamma_x < \xi \left[\sum_{i,j} |a_{ij}| + \sum_{i,j} |b_{ij}| E|\varepsilon(t)| + \sum_i E|c_i \varepsilon(t) + d_i(\varepsilon^2(t) - 1)| - \sum_i |x_i| \right].$$

On pose $w = \sum_i E|c_i \varepsilon(t) + d_i(\varepsilon^2(t) - 1)|$ et par le même raisonnement que précédemment, on obtient le théorème 2.

La remarque précédente pour des modèles généralisés est encore valable ici.

III. Les processus bilinéaires définis par l'équation :

$$(1) \quad X(t) = \sum_{i=1}^p \alpha_i X(t-i) + \sum_{j=1}^q \beta_j \varepsilon(t-j) + \varepsilon(t) + \sum_{k=0}^P \sum_{l=1}^Q \gamma_{kl} X(t-k-l) \varepsilon(t-l),$$

où $\varepsilon(t)$ est un b.b.g., admettent la représentation markovienne suivante :

$$\begin{cases} \text{(i)} & X(t) = HZ(t-1) + \varepsilon(t), \\ \text{(ii)} & Z(t) = AZ(t-1) + BZ(t-1)\varepsilon(t) + C\varepsilon(t) + D(\varepsilon^2(t) - 1) \end{cases}$$

(cf. T. Pham Dinh [2]). A l'aide du théorème 2, on obtient alors le résultat suivant :

THÉORÈME 3. — Sous les conditions du théorème 2, le processus bilinéaire $X(t)$ régi par l'équation (1) est ergodique.

Si les processus $X(t)$ régis par une équation du type (1) ne sont pas centrés, ils admettent alors une représentation markovienne où l'équation (ii) est remplacée par l'équation (iii).

$$\text{(iii)} \quad Z(t) = AZ(t-1) + BZ(t-1)\varepsilon(t) + C\varepsilon(t) + D(\varepsilon^2(t) + F),$$

le théorème 3 est alors vérifié grâce aux remarques faites dans les paragraphes I et II.

IV. Nous obtenons donc pour un ensemble de modèles bilinéaires une condition suffisante d'ergodicité. Celle-ci est en particulier vérifiée pour le modèle $X(t) = \varepsilon(t) + \delta\varepsilon(t-1)X(t-2)$ étudié dans [1]. Il suffit de remarquer qu'une représentation markovienne de ce processus est obtenue pour :

$$A = \begin{pmatrix} 0 & \delta \\ 0 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \quad C = (1, 0)', \quad D = 0,$$

et par exemple $H = \begin{pmatrix} 0 & 0 \\ 0 & \delta \end{pmatrix}$, sa densité étant égale à $f_0(v)$.

Notons enfin que le résultat énoncé dans le cas scalaire (paragraphe III) généralise le résultat de H. Tong donné dans [3] uniquement pour $d=0$.

(*) Remise le 24 octobre 1983.

-
- [1] D. GUÉGAN, *Test de modèles non linéaires*, à paraître chez Springer-Verlag.
[2] T. PHAM DINH, *Modèles bilinéaires et représentation markovienne* (Communication au Séminaire d'Orsay, mai 1983).
[3] H. TONG, *J.T.S.A.*, 2, n° 4, 1981.
[4] R. L. TWEEDIE, *Stoch. Proc. and its appl.*, 3, p. 385-403.

*Département de Mathématiques, Université Paris-Nord,
avenue J.-B. Clément, 93430 Villetaneuse.*

STATISTIQUE. - Étude d'un modèle non linéaire, le modèle superdiagonal d'ordre 1.
 Note (*) de **Dominique Guegan**, transmise par Robert Fortet.

Nous présentons ici quelques résultats relatifs au modèle superdiagonal $X_t = bX_{t-2}\varepsilon_{t-1} + \varepsilon_t$: stationnarité, régularité, processus d'innovation, inversion, fonction de vraisemblance. Les techniques habituelles relatives à l'espace L^2 , utilisées pour les modèles linéaires faisant défaut ici, nous donnons une extension des notions de régularité et de singularité en vue d'une classification plus générale des processus.

We present here some results on the superdiagonal model $X_t = bX_{t-2}\varepsilon_{t-1} + \varepsilon_t$, stationarity, regularity, innovation process, invertibility, likelihood. The usual techniques on the L^2 space, used for linear models are not adequate here, so we give an extension of the notions of regularity and singularity in view of a more general classification of the processes.

INTRODUCTION ET NOTATIONS. - Soit (Ω, \mathcal{A}, P) un espace de probabilité et on note $\| \cdot \|$ la norme dans $L^2 = L^2(\Omega, \mathcal{A}, P)$. On se donne dans cet espace une suite (ε_t) de variables aléatoires indépendantes, centrées et de même variance σ^2 . On note $\mathcal{F}_t(\varepsilon)$ la tribu $\sigma(\varepsilon_n, n \leq t)$ et on considère l'équation : (1) $Y_t = bY_{t-2}\varepsilon_{t-1} + \varepsilon_t$. On cherche à étudier les processus (Y_t) solutions de cette équation et non anticipatifs, i. e., tels que $\forall t, Y_t$ soit $\mathcal{F}_t(\varepsilon)$ -mesurable. Ce modèle fait partie des modèles bilinéaires dont on peut trouver une présentation générale dans [1]. On utilise les notations suivantes : $H_t(Y) = \overline{\text{sp}} \{ X_u, u \leq t \}$, sous espace de Hilbert de L^2 , $H_{-}(Y) = \bigcap H_t(Y)$ et $H_{+}(Y) = \bigcup H_t(Y)$, et $\mathcal{F}_{-}(Y) = \bigwedge \mathcal{F}_t(Y)$, $\mathcal{F}_{+}(Y) = \bigvee \mathcal{F}_t(Y)$. On désigne par A la quantité $\sigma^2 / (1 - b^2 \sigma^2)$.

EXISTENCE D'UN PROCESSUS STATIONNAIRE UNIQUE VÉRIFIANT (1).

THÉORÈME 1. - Pour que (1) ait dans L^2 une solution stationnaire non anticipative, il est nécessaire que $b^2 \sigma^2 < 1$.

Appelons Y_t cette solution. Y_t étant $\mathcal{F}_t(\varepsilon)$ -mesurable, un simple maniement d'espérances conditionnelles sur (1) montre que $E(Y_t) = 0$, et que $\text{var}(Y_t) = A$. ■

THÉORÈME 2. - Si $b^2 \sigma^2 < 1$, alors (1) a une solution stationnaire unique, donnée par :

$$(2) \quad \dot{Y}_t = \sum_{j=1}^{\infty} b^j \varepsilon_{t-2j} \prod_{k=1}^j \varepsilon_{t-2k+1} + \varepsilon_t.$$

Supposons qu'il existe deux solutions Y_t et Z_t de (1). Posons $\gamma_t = EZ_t Y_t = b^2 \sigma^2 \gamma_{t-2} + \sigma^2$. Soit t_0 tel que $\gamma_{t_0} \neq A$, alors $\gamma_{t-2a} - A = (\gamma_{t_0} - A) / (b^2 \sigma^2)^a$ donc $|\gamma_{t-2a}| \rightarrow +\infty$ quand $a \rightarrow +\infty$. Or $\text{var} Y_t = \text{var} Z_t = A$, donc la covariance $|\gamma_{t-2a}|$ devient plus grande que $\text{var} X_t$, d'où la contradiction, donc $\gamma_t \equiv A$ et $Y_t \equiv Z_t$.

Notons :

$$Y_t(a) = \sum_{j=1}^a b^j \varepsilon_{t-2j} \prod_{k=1}^j \varepsilon_{t-2k+1} + \varepsilon_t,$$

alors pour $n < m$:

$$\| Y_t(m) - Y_t(n) \| < \sum_{a=n}^m \sigma^2 (b^2 \sigma^2)^a = A (b^2 \sigma^2)^n,$$

et sous l'hypothèse $b^2 \sigma^2 < 1$, cette quantité tend vers zéro quand $n \rightarrow +\infty$, donc le processus (Y_t) défini par (2) existe. Posons $Y_t = \lim_{a \rightarrow +\infty} Y_t(a)$, Y_t vérifiant (2), alors

$\| Y_t(a) - bY_{t-2}(a)\varepsilon_{t-1} - \varepsilon_t \| = \sigma^2 (b^2 \sigma^2)^{a+1}$, qui tend vers zéro, donc le processus (Y_t) défini par (2), vérifie (1). ■

PROPOSITION 1. - Une condition nécessaire et suffisante pour que (1) ait une solution stationnaire unique non anticipative définie par (2) est $b^2 \sigma^2 < 1$.

TESTS DE MODELES NON LINEAIRES

D. GUEGAN

Département de Mathématiques, Université Paris XIII, France

*(paru dans Proceeding of the 3rd franco belgium meeting)*Abstract

From the example of two non-linear models, $X(t) = \varepsilon(t) + \delta\varepsilon(t-1)\varepsilon(t-2)$ and $X(t) = \varepsilon(t) + \gamma\varepsilon(t-1)X(t-2)$, $\varepsilon(t)$ being a Gaussian white noise, we show how we can extend Bartlett's and Quenouille's tests for choosing between these two models. We estimate consistently the parameters of the two models and with the help of two statistics whose laws are asymptotically known under the null and alternatives hypothesis, we construct two tests, whose power is calculated for some values of the parameters.

Keywords : Bilinear models, Estimation, Tests de modèles.

1. INTRODUCTION

Depuis quelques années on s'est beaucoup intéressé à l'étude des modèles bilinéaires : à côté des études faites par les contrôleurs, cfr Bruni *et al.* (1974), on trouve principalement les travaux de l'équipe de Manchester (Priestley, 1980; Subba Rao, 1980, 1981), qui a choisi une approche essentiellement spectrale, approche dont on connaît déjà les difficultés d'utilisation dans le cas linéaire. Dans de précédents articles (Guégan, 1981, 1983), nous avons étudié ces modèles d'un point de vue probabiliste et montré que les chaos de Wiener en étaient le bon cadre. Nous présentons ici à titre d'exemples une analyse statistique complète de deux modèles particuliers, basée sur une généralisation des tests de Bartlett (1946) Quenouille (1949). Les calculs sont particulièrement fastidieux, mais la généralisation à des modèles plus compliqués ne pose pas de problèmes théoriques.

$\varepsilon(t)$ étant un bruit blanc gaussien, de variance σ^2 , les deux modèles considérés tout au long de l'étude sont les suivants :

$$(1) \quad X(t) = \varepsilon(t) + \delta \varepsilon(t-1) \varepsilon(t-2)$$

$$(2) \quad X(t) = \varepsilon(t) + \gamma \varepsilon(t-1) X(t-2).$$

On posera

$$c(h) = E[X(t) X(t-h)]$$

$$c(i,j) = E[X(t) X(t-i) X(t-j)]$$

$$c(i,j,k) = E[X(t) X(t-i) X(t-j) X(t-k)].$$

Le Tableau 1 donne les résultats relatifs à l'étude des premiers moments (d'ordre 1, 2, 3).

Au vu du Tableau 1, on constate que la procédure statistique habituelle basée sur des quantités déduites des seules covariances ne suffit pas à différencier ces deux modèles, qui, on le voit, se comportent tous les deux comme des bruits blancs (non gaussiens). L'étude des moments d'ordre 3 ne le permet pas non plus : un seul moment est non nul, $c(1,2)$. (Ce résultat a lieu d'ailleurs pour tous les modèles (3) et (4).) Il est nécessaire d'en étudier les moments d'ordre 4 dont le Tableau 2 nous donne les valeurs.

(1) $X(t) = \epsilon(t) + \delta\epsilon(t-1)\epsilon(t-2)$ $v = \delta^2\sigma^2$	(2) $X(t) = \epsilon(t) + \gamma\epsilon(t-1)X(t-2)$ $u = \gamma^2\sigma^2$
Moments d'ordre 4 $c(i,j,k) = EX(t)X(t-i)X(t-j)X(t-k)$	
$c(0,0,0) = 3\sigma^4(1 + 2v + 3v^2)$	$c(0,0,0) = \frac{3\sigma^4(1+u)}{(1-u)(1-3u^2)}$
$c(0,1,1) = \sigma^4(1 + 4v + 3v^2)$	$c(0,1,1) = \frac{\sigma^4(1+2u)}{(1-u)^2}$
$c(0,2,2) = \sigma^4(1 + 4v + v^2)$	$c(0,2,2) = \frac{\sigma^4(1+3u)}{(1-u)(1-3u^2)}$
$c(0,k,k) = \sigma^4(1+v)^2 \quad \forall k \geq 3$	$c(0,2k-1,2k-1) = \frac{\sigma^4(1+2u^k)}{(1-u)^2}$
$c(1,1,3) = 2v\sigma^4$	$c(0,2k,2k) = \frac{\sigma^4}{(1-u)^2} + \frac{2\sigma^4 u^{2k}}{(1-u)^2(1-3u^2)}$
$c(0,1,4) = 0$	$c(1,1,3) = \frac{2u\sigma^4}{1-u}$
$c(0,2k+1,2k+3) = 0$	$c(0,1,4) = \frac{2u^2\sigma^4}{1-u}$
$c(1,3,4) = 0$	$c(0,2k+1,2k+4) = \frac{2\sigma^4 u^{k+2}}{1-u}$
	$c(1,3,4) = \frac{u\sigma^4}{1-u}$
Tous les autres moments d'ordre 4 sont nuls	

Tableau 2

On constate que certains moments d'ordre 4 sont nuls pour le modèle (1) et non nuls pour le modèle (2), en particulier $c(0,1,4)$. On va donc pouvoir construire un test basé sur cette statistique qui permette de différencier les deux modèles. Inversement, il est facile de trouver une combinaison linéaire des moments d'ordre 4,

nulle pour le modèle (2), et non nulle pour le modèle (1), par exemple celle faisant intervenir les plus petits écarts, soit $T = c(0,0,0) - 3c(0,4,4)$.

Notons que tous les moments dépendent de v (soit de δ et de σ^2) pour le modèle (1), et de u (soit de γ et de σ^2) pour le modèle (2), qui sont des quantités inconnues lors de l'étude statistique d'une série.

2. ESTIMATION DES PARAMETRES DU MODELE (1)

Soit $X(t)$ le processus régi par l'équation (1); $\forall p, X(t) \in L^p$.

Quand on observe $X_1, X_2 \dots X_N$, on peut donc estimer de façon consistante tous les moments du modèle (1) par les moyennes empiriques correspondantes (en posant $X(u) = 0$ si $u < 0$)

$$c(i,j) \text{ par } \hat{c}(i,j) = \frac{1}{N} \sum_{i=1}^N X(t) X(t-i) X(t-j)$$

$$c(i,j,k) \text{ par } \hat{c}(i,j,k) = \frac{1}{N} \sum_{i=1}^N X(t) X(t-i) X(t-j) X(t-k).$$

Leurs variances sont parfaitement connues et les estimateurs $\hat{c}(i,j)$ et $\hat{c}(i,j,k)$ sont asymptotiquement normaux. Il y a convergence presque sûre et dans L^2 vers la vraie valeur des paramètres.

En particulier pour le moment $c(0,1,4)$, on a :

$$E \hat{c}(0,1,4) = 0$$

et

$$(5) \quad \text{Var } \hat{c}(0,1,4) = \hat{v}^2 = \frac{\sigma^8}{N} [3 + 24v + 88v^2 + 120v^3 + 45v^4].$$

Les paramètres du modèle vont être calculés à partir de ces moments. Nous avons retenu les expressions suivantes :

$$\delta = \frac{\hat{c}(1,1,3)^2}{4\hat{c}^3(1,2)} \quad \text{et} \quad \sigma^2 = \frac{2\hat{c}^2(1,2)}{\hat{c}(1,1,3)}$$

Au vu de simulations de longueur 1000 et 10000 faites pour δ variant de -1 à +1, avec un pas de 0,1, on constate que ces estimateurs "marchent bien", sauf en $\delta = 0$ où le processus $X(t)$ est identique au bruit $\varepsilon(t)$, cfr. Tableaux 3.

Les estimateurs $\hat{c}(1,2)$ et $\hat{c}(1,1,3)$ sont des estimateurs asymptotiquement normaux non indépendants. On peut calculer la loi du couple :

δ théorique	100 simulations de longueur 1000 $\sigma^2 = 1$			10 simulations de longueur 10000 $\sigma^2 = 1$		
	δ observé (en moyenne)	Ecart type théorique	Ecart type observé	δ observé (en moyenne)	Ecart type théorique	Ecart type observé
-1	-1,0026	0,4450	0,807	-0,9485	0,1407	0,147
-0,9	-0,9555	0,3505	0,682	-0,8637	0,1109	0,511
-0,8	-0,7876	0,2817	0,383	-0,7752	0,0891	0,091
-0,7	-0,7190	0,2380	0,377	-0,6889	0,0753	0,079
-0,6	-0,6250	0,2178	0,265	-0,6292	0,0689	0,070
-0,5	-0,5142	0,2167	0,269	-0,5121	0,0685	0,070
-0,4	-0,4099	0,2311	0,215	-0,4093	0,0731	0,093
-0,3	-0,3959	0,2631	0,289	-0,3736	0,0832	0,106
-0,2	-0,3109	0,3353	0,324	-0,2101	0,1060	0,069
-0,1	-0,2634	0,5819	15,262	-0,1783	0,1840	0,211
0	-	-	-	-	-	-
0,1	49,2714	0,5819	440,644	0,2543	0,1840	0,298
0,2	0,3133	0,3353	0,437	0,2010	0,1060	0,078
0,3	0,3407	0,2631	0,237	0,3016	0,0832	0,077
0,4	0,4631	0,2311	0,264	0,4518	0,0731	0,075
0,5	0,5344	0,2167	0,269	0,5366	0,0685	0,082
0,6	0,6119	0,2178	0,379	0,6007	0,0689	0,124
0,7	0,7803	0,2380	0,370	0,7431	0,0753	0,130
0,8	0,7637	0,2817	0,382	0,7506	0,0891	0,087
0,9	1,0370	0,3505	0,998	0,9396	0,1109	0,148
1	1,0670	0,4450	0,712	0,9529	0,1407	0,234

Tableau 3.1
Estimation de δ .

δ théorique	100 simulations de longueur 1000			10 simulations de longueur 10000		
	σ^2 observé (en moyenne)	Ecart type théorique	Ecart type observé	σ^2 observé (en moyenne)	Ecart type théorique	Ecart type observé
-1	1,1615	0,2449	0,364	1,0682	0,0775	0,108
-0,9	1,1280	0,2052	0,434	1,0306	0,0649	0,097
-0,8	0,9436	0,1747	1,294	1,0225	0,0533	0,067
-0,7	1,0681	0,1586	0,265	1,0147	0,0501	0,069
-0,6	1,0497	0,1642	0,217	0,9966	0,0519	0,060
-0,5	1,0776	0,1993	0,284	0,9913	0,0630	0,074
-0,4	1,6458	0,2751	4,562	1,0021	0,0870	0,119
-0,3	1,1039	0,4303	0,752	0,9364	0,1361	0,117
-0,2	0,8862	0,8359	2,336	1,0383	0,2643	0,277
-0,1	-0,3480	2,9119	5,636	0,8676	0,9208	2,896
0	0,0901		1,021	-0,224		0,064
0,1	-0,0127	2,9119	1,792	2,0196	0,9208	2,814
0,2	2,8683	0,8359	13,729	1,0679	0,2643	0,210
0,3	-1,1352	0,4303	21,646	1,0216	0,1361	0,123
0,4	1,1008	0,2751	0,511	0,9525	0,0870	0,080
0,5	1,0729	0,1993	0,360	0,9773	0,0630	0,075
0,6	1,0934	0,1642	0,275	1,0166	0,0519	0,092
0,7	1,0142	0,1586	0,272	0,9720	0,0501	0,100
0,8	1,1328	0,1747	0,301	1,0600	0,0533	0,077
0,9	1,0511	0,2052	0,316	0,9936	0,0649	0,098
1	1,1413	0,2449	1,682	1,0411	0,0775	0,128

Tableau 3.2.
Estimation de σ^2
(valeur théorique \hat{i})

$$\begin{pmatrix} \hat{c}(1,2) \\ \hat{c}(1,1,3) \end{pmatrix} \cong N \left(\begin{pmatrix} \partial \sigma^4 \\ 2u\sigma^4 \end{pmatrix}, \frac{\sigma^6}{N} \begin{bmatrix} 1+14v+26v^2+9v^3 & \delta\sigma^2(3+70v+81v^2) \\ \delta\sigma^2(3+70v+81v^2) & 3\sigma^2(1+16v+84v^2+108v^3+45v^4) \end{bmatrix} \right)$$

et en déduire la loi des estimateurs $\hat{\delta}$ et $\hat{\sigma}^2$.

Les moments estimés $\hat{c}(1,2)$ et $\hat{c}(1,1,3)$ convergent respectivement p.s. vers $c(1,2)$ et $c(1,1,3)$, donc $\hat{\sigma}^2$ et $\hat{\delta}$ convergent fortement respectivement vers σ^2 et δ , les vraies valeurs des paramètres.

Remarquons que si on note θ le paramètre étudié, $\theta = (\delta, \sigma^2)$, $X = \hat{c}(1,2)$, $Y = \hat{c}(1,1,3)$, on a $\theta = f(X, Y)$; on peut prendre comme approximation de la variance de $\hat{\theta}$:

$$(6) \quad \text{var } \hat{\theta} \cong \left(\frac{\partial f(X, Y)}{\partial X} \right)^2 \text{var } X + \left(\frac{\partial f(X, Y)}{\partial Y} \right)^2 \text{var } Y + 2 \left(\frac{\partial f(X, Y)}{\partial X} \cdot \frac{\partial f(X, Y)}{\partial Y} \right) \text{cov}(X, Y)$$

On obtient ainsi :

$$\begin{aligned} \text{var } \hat{\delta} &\cong \frac{3}{Nv\sigma^2} [1 + 13v - 14v^2 + 24v^3 + 42v^4] \\ \text{var } \hat{\sigma}^2 &\cong \frac{\sigma^4}{4v^2N} [3 + 40v - 84v^2 + 92v^3 + 189v^4] \end{aligned}$$

A titre de comparaison on donne dans le Tableau 4 les intervalles de confiance obtenus par $\hat{\delta}$ et $\hat{\sigma}^2$.

3. ESTIMATION DES PARAMETRES DU MODELE (2)

Soit $X(t)$ le processus régi par l'équation (2) :

1) Si $u < 1$ alors $X(t)$ existe et appartient à L^2 (cfr. Guégan, 1981). Sous cette condition, observant X_1, X_2, \dots, X_N , la moyenne empirique $\bar{X} \in L^2$ et converge vers la vraie moyenne (soit zéro); les moments d'ordre 2 existent, mais pour estimer ces moments, on a besoin que $X(t) \in L^4$, soit $u < 3^{-1/2}$.

Pour simplifier un peu les formules, nous nous limiterons au cas où N est pair, et l'on a bien sûr des résultats analogues pour N impair. On posera $N = 2M$.

2) Si $u^2 < \frac{1}{3}$, alors $X(t) \in L^4$, on en déduit que $E(X^4(t))$ est fini, donc aussi $c(i, j, k) \forall i, \forall j, \forall k$, que l'estimation de $c(h)$ par $\hat{c}(h) = \frac{1}{2M} \sum_{t=1}^{2M} X(t)X(t-h)$ converge dans L^2 vers $c(h)$. En particulier :

δ	v	Estimation de δ		Estimation de σ^2	
		Ecart type pour N = 1000	Ecart type pour N = 10000	Ecart type pour N = 1000	Ecart type pour N = 10000
0,1	0,01	0,5818	0,1840	2,9119	0,9208
0,2	0,04	0,3353	0,1060	0,8359	0,2643
0,3	0,09	0,2631	0,0832	0,4302	0,1360
0,4	0,16	0,2310	0,0730	0,2751	0,0869
0,5	0,25	0,2167	0,0685	0,1992	0,0630
0,6	0,36	0,2177	0,0688	0,1642	0,0519
0,7	0,49	0,2380	0,0752	0,1585	0,0501
0,8	0,64	0,2816	0,0890	0,1747	0,0552
0,9	0,81	0,3505	0,1108	0,2052	0,0648
1	1	0,4449	0,1407	0,2449	0,0774

Tableau 4

Ecart type des estimateurs de δ et σ^2

$$\hat{c}(0) \xrightarrow{L^2} c(0) = \frac{\sigma^2}{1-u}$$

et

$$\frac{1}{\sqrt{2M}} [\hat{c}(0) - c(0)] \longrightarrow N(0, \tilde{v}_0^2)$$

où

$$(7) \quad \tilde{v}_0^2 = \frac{\sigma^4(1-3u-u^2-7u^3)}{M(1-u)^3(1-3u^2)}$$

3) Pour que chacun des estimateurs $\hat{c}(h,k,l)$ converge vers le terme $c(h,k,l)$ correspondant, il faut de même que $X(t) \in L^8$ (c'est en particulier nécessaire pour que $c(0,0,0) \rightarrow E[X^4(t)]$)

$$X(t) \in L^8 \iff u^4 < \frac{1}{105}$$

Il se trouve que, pour que les estimateurs de $c(1,1,3)$ et $c(0,1,4)$ qui interviennent dans notre procédure convergent dans L^2 , il suffit que $X(t) \in L^6$, ce qui est réalisé pour $u^3 < \frac{1}{15}$.

Leurs variances sont calculées ci-dessus, ces estimateurs sont asymptotiquement normaux et convergent p.s. vers la vraie valeur des paramètres :

$$a) \quad \frac{1}{\sqrt{2M}} [\hat{c}(1,2) - c(1,2)] \sim N(0, \hat{v}_1^2)$$

où

$$(8) \quad \hat{v}_1^2 = \frac{\sigma^6 (1 + 22u + 60u^2 + 24u^3 - 267u^4 - 99u^5 - 126u^6 + 36u^7 + 216u^8 + 108u^{10})}{2M(1-u)^2(1-3u^2)(1-3u^3)}$$

$$b) \quad \frac{1}{\sqrt{2M}} [\hat{c}(1,1,3) - c(1,1,3)] \sim N(0, \hat{v}_2^2)$$

où

$$(9) \quad \hat{v}_2^2 = \frac{\sigma^8}{2M(1-u)^3(1-3u^2)(1-3u^3)(1-15u^3)(1-15u^4)} \times \\ (1 + 15u + 92u^2 + 118u^3 - 249u^4 - 2197u^5 - 1175u^6 + 1226u^7 - 16473u^8 \\ + 30051u^9 + 36204u^{10} + 25971u^{11} - 240988u^{12} - 43335u^{13} - 270570u^{14} \\ - 26865u^{15} + 55620u^{16} + 32400u^{17} + 32400u^{18} + 97200u^{19})$$

$$c) \quad \frac{1}{\sqrt{2M}} (\hat{c}(0,1,4) - c(0,1,4)) \sim N(0, \hat{v}_3^2)$$

où

$$(10) \quad \hat{v}_3^2 = \frac{\sigma^8}{4M(1-u)^3(1-3u^2)(1-3u^3)(1-15u^3)(1-15u^4)} \times \\ (6 + 114u + 360u^2 - 66u^3 + 732u^4 - 1960u^5 - 14808u^6 - 30564u^7 \\ + 9306u^8 + 188910u^9 - 383802u^{10} + 840696u^{11} + 105600u^{12} + 705120u^{13} \\ + 22930u^{14} - 387570u^{15} + 5164700u^{16} + 271200u^{17} - 2223600u^{18})$$

On peut alors comme précédemment calculer les paramètres du modèle à l'aide des moments estimés. On retient les expressions suivantes :

$$\hat{\gamma} = \frac{\hat{c}(0)^2 \hat{c}(0,1,4)}{2\hat{c}(1,2)^3} \quad \hat{\beta}^2 = \frac{\hat{c}(1,1,3)^2}{2\hat{c}(0) \hat{c}(0,1,4)}$$

Les estimateurs $\hat{c}(0)$, $\hat{c}(1,2)$, $\hat{c}(0,1,4)$ et $\hat{c}(1,1,3)$ sont asymptotiquement normaux non indépendants, on peut calculer la loi de chaque triplet $(c(0), c(1,2), c(0,1,4))$ et $(c(0), c(0,1,4), c(1,1,3))$ et en déduire la loi des estimateurs de \bar{y} et de σ^2 .

Ainsi :

$$\begin{pmatrix} \hat{c}(0) \\ \hat{c}(1,2) \\ \hat{c}(0,1,4) \end{pmatrix} \sim N \left(\begin{pmatrix} c(0) \\ c(1,2) \\ c(0,1,4) \end{pmatrix}, \begin{bmatrix} \hat{v}_0^2 & \hat{c}_{01} & \hat{c}_{03} \\ \hat{c}_{01} & \hat{v}_1^2 & \hat{c}_{13} \\ \hat{c}_{03} & \hat{c}_{13} & \hat{v}_3^2 \end{bmatrix} \right)$$

où \hat{v}_0^2 , \hat{v}_1^2 et \hat{v}_3^2 sont données par les formules (7), (8) et (10),

\hat{c}_{01} , \hat{c}_{03} et \hat{c}_{13} étant données par les formules (11), (12), (13) :

$$(11) \quad \hat{c}_{01} = \text{cov}(\hat{c}(0), \hat{c}(1,2)) = \frac{2\gamma\sigma^6(1+8u+8u^2-27u^3-6u^4+18u^5)}{N(1-u)^3(1-3u^2)}$$

$$(12) \quad \hat{c}_{03} = \text{cov}(\hat{c}(0), \hat{c}(0,1,4)) = \frac{2u^2\sigma^6(5+43u+31u^2-21u^3-102u^4-36u^5+108u^6)}{N(1-u)^3(1-3u^2)}$$

$$(13) \quad \hat{c}_{13} = \text{cov}(\hat{c}(1,2), \hat{c}(0,1,4)) = \frac{2u\gamma\sigma^8(6+193u-311u^2+83u^3-9u^4+162u^5+108u^6)}{N(1-u)^3(1-3u^2)}$$

et

$$\begin{pmatrix} \hat{c}(0) \\ \hat{c}(1,1,3) \\ \hat{c}(0,1,4) \end{pmatrix} \sim N \left(\begin{pmatrix} c(0) \\ c(1,1,3) \\ c(0,1,4) \end{pmatrix}, \begin{bmatrix} \hat{v}_0^2 & \hat{c}_{02} & \hat{c}_{03} \\ \hat{c}_{02} & \hat{v}_2^2 & \hat{c}_{23} \\ \hat{c}_{03} & \hat{c}_{23} & \hat{v}_3^2 \end{bmatrix} \right)$$

où \hat{v}_0^2 , \hat{v}_2 et \hat{v}_3 sont données par les formules (7), (9) et (10)

\hat{c}_{02} , \hat{c}_{03} et \hat{c}_{23} étant données par les formules (14), (12) et (15) :

$$(14) \quad \hat{c}_{02} = \text{cov}(\hat{c}(0), \hat{c}(1,1,3)) = \frac{4u\sigma^6(1+11u+20u^2-21u^3-65u^4-69u^5+54u^6)}{N(1-u)^3(1-3u^2)}$$

$$(15) \quad \hat{c}_{23} = \text{cov}(\hat{c}(0,1,4), \hat{c}(1,1,3)) = \frac{u\sigma^8}{N(1-u)^3(1-3u^2)} \times \\ (9+198u+722u^2+290u^3-3391u^4+3033u^5-1308u^6+1728u^7-432u^9).$$

Les moments estimés convergent p.s. vers leurs vraies valeurs, les estimateurs

$\hat{\sigma}^2$ et $\hat{\gamma}$ convergent donc fortement vers σ^2 et γ , les vraies valeurs des paramètres.

L'analogie de la formule (6) au cas de 3 variables permet d'obtenir une approximation de la variance de chacun des estimateurs $\hat{\gamma}$ et $\hat{\sigma}^2$:

$$\text{var } \hat{\gamma} \approx \frac{1}{N^2 u^3 (1-u)(1-3u^2)(1-3u^3)(1-15u^3)(1-15u^4)} \times$$

$$(3 + 57u + 180u^2 - 60u^3 - 567u^4 + 1516u^5 - 7585u^6 + 19u^7$$

$$- 25602u^8 + 65856u^9 - 200921u^{10} + 296796u^{11} + 719457u^{12}$$

$$+ 193305u^{13} + 87148u^{14} - 1804875u^{15} + 3019845u^{16} + 592980u^{17}$$

$$- 197300u^{18} + 340200u^{19} - 364500u^{20} - 218700u^{21})$$

$$\text{var } \hat{\sigma}^2 \approx \frac{\sigma^4}{8Nu^4 (1-u)(1-3u^2)(1-3u^3)(1-15u^3)(1-15u^4)} \times$$

$$(6 + 114u + 312u^2 - 1290u^3 - 2804u^4 - 92u^5 + 35104u^6 - 138916u^7$$

$$+ 117338u^8 - 321186u^9 - 1391059u^{10} + 2428536u^{11} + 1607186u^{12}$$

$$+ 8422944u^{13} - 8796382u^{14} + 2407830u^{15} - 21094140u^{16} + 1799184u^{17}$$

$$- 4039920u^{18} + 13478400u^{19} + 4276800u^{20} + 2332800u^{21}).$$

On donne dans le tableau 5 les intervalles de confiance obtenues pour $\hat{\gamma}$ et $\hat{\sigma}^2$

δ	u	Estimation de $\hat{\gamma}$		Estimation de $\hat{\sigma}^2$	
		Ecart type pour N = 1000	Ecart type pour N = 10000	Ecart type pour N = 1000	Ecart type pour N = 10000
0,1	0,01	60,21	19,04	300,93	95,16
0,2	0,04	4,0465	1,2796	23,69	7,49
0,3	0,09	1,2518	0,3958	6,2021	1,9613
0,4	0,16	0,5256	0,1852	2,6052	0,8238

Tableau 5

Ecart type des estimateurs de $\hat{\gamma}$ et $\hat{\sigma}^2$

4 TESTS DE MODELES

4.1 Test du modèle (1) contre le modèle (2)

Pour tester le modèle (1) contre le modèle (2), on construit un test asymptotique à partir de la statistique $S = \hat{c}(0,1,4)$ sachant que :

Sous H_0 : $E\hat{c}(0,1,4) = 0$ et $S \sim N(0, \hat{V}^2)$ où \hat{V}^2 est donnée par l'équation (5)

Sous H_1 : $E\hat{c}(0,1,4) > 0$.

Pour ce test, on obtient au niveau 5% les seuils unilatéraux donnés par le Tableau 6.

δ	N = 1000		N = 10000	
	V	1,645 V	V	1,645 V
0,1	0,0565	0,0929	0,0178	0,0292
0,3	0,0768	0,1263	0,0242	0,0398
0,5	0,1296	0,2131	0,0409	0,0672
0,7	0,2290	0,3767	0,0721	0,1186
0,9	0,4041	0,6647	0,1276	0,2099

Tableau 6

Dans les simulations de longueur 1000 et 10000 on constate que la variance observée pour S varie entre $\frac{1}{2}V$ et $2V$, ce qui est un bon résultat (Tableau 7).

Sachant qu'asymptotiquement la statistique S suit sous H_1 une loi normale $\left(\frac{2u^2\sigma^4}{1-u}, \hat{V}_3^2\right)$ où \hat{V}_3^2 est donnée par la formule (10), on peut obtenir la puissance du test basé sur cette statistique pour différentes valeurs de u et de v (pour N = 1000 et N = 10000, Tableau 8), les calculs étant faits pour la valeur théorique $\sigma^2 = 1$. La puissance du test sera égale à

$$1 - \Phi\left\{\frac{t(\alpha)\sqrt{\hat{V}_3} - E_1}{s_1}\right\}$$

où Φ est la fonction de répartition de la gaussienne centrée réduite, E_1 et s_1 res-

δ théorique	N = 1000			N = 10000		
	Variance théorique V_T	Variance observée V_0	$\frac{V_0}{V_T}$	Variance théorique V_T	Variance observée V_0	$\frac{V_0}{V_T}$
- 1	0,280	0,510	1,82	0,028	0,061	2,17
- 0,9	0,163	0,170	1,04	0,016	0,023	1,43
- 0,8	0,093	0,081	0,87	0,0093	0,002	0,21
- 0,7	0,052	0,052	1	0,005	0,004	0,8
- 0,6	0,029	0,031	1,06	0,0029	0,003	1,03
- 0,5	0,016	0,012	0,75	0,0016	0,001	0,62
- 0,4	0,009	0,009	1	0,0009	0,001	1,11
- 0,3	0,005	0,010	2	0,0005	0,001	2
- 0,2	0,004	0,004	1	0,0004	0,0004	1
- 0,1	0,003	0,003	1	0,0003	0,0003	1
0	0,003	0,003	1	0,0003	0,0003	1
0,1	0,003	0,003	1	0,0003	0,0003	1
0,2	0,004	0,004	1	0,0004	0,0004	1
0,3	0,005	0,006	1,2	0,0005	0,001	2
0,4	0,009	0,010	1,11	0,0009	0,0003	0,33
0,5	0,016	0,023	1,43	0,0016	0,003	1,875
0,6	0,029	0,056	0,51	0,0029	0,003	1,03
0,7	0,052	0,042	0,80	0,005	0,006	1,2
0,8	0,093	0,132	1,4	0,0093	0,010	1,07
0,9	0,163	0,154	0,94	0,016	0,008	0,5
1	0,280	0,212	1,32	0,028	0,025	0,89

Tableau 7 : Etude de la variance de la statistique S

N = 1000								
v \ u	0,0	0,1	0,15	0,20	0,25	0,30	0,35	0,40
0	16,2	29,0	41,4	51,7	59,1	63,3	63,6	55,2
0,1	7,1	18,7	32,2	44,7	54,2	60,1	61,7	54,7
0,2	2,0	9,6	22,3	36,3	48,1	56,0	59,3	54,1
0,3	0,3	3,7	13,3	27,3	41,0	51,2	56,5	53,4
0,4	0	1,0	6,6	18,6	33,3	45,7	53,2	52,6
0,5	0	0,2	2,6	11,3	25,6	39,7	50,4	51,7
0,6	0	0	0,7	5,9	18,3	32	45,6	50,8
0,7	0	0	0,2	2,6	12	27,1	41,3	49,7
0,8	0	0	0	0,9	7,1	21	36,8	48,6
0,9	0	0	0	0,3	3,8	15,4	32,2	47,4
1	0	0	0	0	1,7	10,6	27,3	46,1

N = 10000								
0	19,7	43,6	67,5	83,1	90,5	92,9	90,9	68,2
0,1	9,1	30,9	58,3	78,3	88,3	91,7	90,1	67,5
0,2	2,7	18	46,4	71,4	84,8	90,0	88,9	67,0
0,3	0,4	8,1	32,9	62,3	80,4	87,7	87,5	66,4
0,4	0	2,6	20,1	51,1	74,3	84,7	85,8	65,6
0,5	0	0,5	10,1	38,5	66,5	80,8	83,6	64,8
0,6	0	0	3,9	26	57,1	75,8	81,0	63,9
0,7	0	0	1,1	15,3	46,4	69,8	77,9	62,9
0,8	0	0	0,2	7,5	35,1	62,6	74,3	61,9
0,9	0	0	0	3,0	24,3	54,4	70,0	60,7
1	0	0	0	0,9	15,1	45,3	65,2	59,4

Tableau 8

Puissance du test à partir de la statistique S

pectivement l'espérance et l'écart type de S sous H_1 dont les valeurs sont données dans le Tableau 9. Le test ainsi construit n'est évidemment pas le plus puissant, mais à u et v fixés on acceptera H_0 si $S < t(\alpha) \sqrt{V_3}$ où $P(N(0,1) > t(\alpha)) = \alpha$, α niveau choisi.

u	E_1	$N = 1000$ s_1	$N = 10000$ s_1
0	0	0,3950	0,1249
0,1	0,27	0,5677	0,1795
0,2	0,48	0,8153	0,2578
0,3	0,63	1,1361	0,3593
0,4	0,72	1,5294	0,4837
0,5	0,75	1,9950	0,6309
0,6	0,72	2,5325	0,8009
0,7	0,63	3,1421	0,9936
0,8	0,48	3,8237	1,2091
0,9	0,27	4,5772	1,4474
1	0	5,4028	1,7085

Tableau 9

La lecture du Tableau 8 permet de distinguer 3 zones spécifiques :

- Une zone où la puissance est mauvaise : ce qui correspond aux petites valeurs de $u = s^2$. Le processus est alors presque un bruit blanc.
- Une zone où la puissance est moyenne (environ 30%), puis,
- Une zone où la puissance est bonne et même très bonne.

4.2 Test du modèle (2) contre le modèle (1)

La démarche est analogue : on teste :

$$H_0 : X(t) \text{ suit le modèle (2) : } X(t) = \varepsilon(t) + v\varepsilon(t-1) + \dot{\varepsilon}(t-2)$$

contre

$$H_1 : X(t) \text{ suit le modèle (1) : } X(t) = \varepsilon(t) + \delta\varepsilon(t-1) + \varepsilon(t-2).$$

Le Tableau 2 permet de trouver un grand nombre de statistiques, combinaisons linéaires de $c(h,k,l)$, d'espérance nulle sous H_0 et non nulle sous H_1 . Par exemple :

$$T = \bar{c}(0,0,0) - 3\bar{c}(0,4,4) \text{ vérifie}$$

$$\text{sous } H_0 : E(T) = 0$$

$$\text{sous } H_1 : E(T) = 6\sigma^4 v^2 > 0.$$

Cette statistique semble préférable aux autres que l'on pourrait utiliser, du fait que, faisant intervenir les plus petites valeurs de (h,k,l) , ce sera celles que l'on estimera le mieux à partir d'un échantillon.

On construit alors un test asymptotique sur cette statistique, et à partir du calcul de $\Sigma = \text{var } T$ sous H_0 , on obtient au niveau de 5%, les seuils unilatères correspondant à ce test, pour quelques valeurs de γ (Tableau 10).

$$\Sigma = \frac{3\sigma^8}{4M(1-u)^3(1-3u)^2(1-3u^2)^3(1-3u^3)(1-15u^3)(1-105u^4)} \times$$

$$(76 + 90u + 596u^2 - 9708u^3 - 81496u^4 - 337338u^5 + 453268u^6$$

$$- 2275944u^7 - 2041407u^8 + 4013991u^9 + 14395209u^{10} - 10519470u^{11}$$

$$+ 8666355u^{12} - 641889285u^{13} + 2399960262u^{14} + 976505757u^{15}$$

$$- 11686122760u^{16} - 1551163101u^{17} + 21915666220u^{18} + 779288904u^{19}$$

$$- 7894194897u^{20} - 74072973820u^{21} + 56678733770u^{22} + 12234819790u^{23}$$

$$+ 28286918330u^{24} + 2573634060u^{25} - 106371270u^{26} + 18874020060u^{27}$$

$$- 247926420u^{28} + 675126900u^{29} - 3152429280u^{30} + 2758375620u^{31})$$

γ	$M = 500$		$M = 5000$	
	Σ	$1,65 \Sigma$	Σ	$1,65 \Sigma$
0,1	0,3397	0,5588	0,1074	0,1767
0,2	0,3392	0,5580	0,1072	0,1764
0,3	0,3357	0,5523	0,1061	0,1746
0,4	0,1876	0,3087	0,0593	0,0976

Tableau 10

La statistique T suit asymptotiquement une loi normale sous H_1 $N(6\sigma^4 v^2, \hat{v}_4^2)$
où

$$\hat{v}_4^2 = \frac{30^8}{N} (38 + 206v + 1023v^2 + 2532v^3 + 4008v^4).$$

Le tableau suivant (Tableau 11) donne l'espérance E_2 et l'écart type s_2 de T sous H_1 .

v	E_2	M = 500	M = 5000
		s_2	s_2
0	0	0,114	0,0114
0,1	0,06	0,215	0,0215
0,2	0,24	0,440	0,0440
0,3	0,54	0,878	0,0878
0,4	0,96	1,646	0,1646
0,5	1,50	2,891	0,2891
0,6	2,16	4,789	0,4789
0,7	2,94	7,543	0,7543
0,8	3,84	11,387	1,1387
0,9	4,86	16,583	1,6583
1	6	23,421	2,3421

Tableau 11

Bien que ce test ne soit pas non plus le plus puissant, c'est celui-là que l'on choisit car $E_{H_1}(T) > E_{H_0}(T)$. On peut comme précédemment en calculer la puissance en évaluant

$$1 - \beta = \left\{ \frac{t(x) \sqrt{\hat{v}_4^2} - E_2}{s_2} \right\}$$

où $s_2 = \sqrt{s^2}$ (Tableau 12).

Les résultats sont donnés pour différentes valeurs de u et de v, pour M = 500, et M = 5000.

A la différence du Tableau 8, on ne retrouve pas ici (Tableau 11) les 3 zones

N = 2M = 1000				
v \ u	0,1	0,2	0,3	0,4
0	0	0	0	4,1
0,1	0	0	0,5	12,8
0,2	0	0,0	6,4	30,1
0,3	0	0,0	22,4	48,1
0,4	0	0,8	41,1	61,5
0,5	0	6,7	55,8	70,5
0,6	0	19,4	66,1	76,4
0,7	0	34,3	73,1	80,4
0,8	0	47,5	77,9	83,3
0,9	0	57,9	81,2	85,3
1	0	65,7	83,7	86,8
N = 2M = 10000				
0	0	0	0	0
0,1	0	0	0	0
0,2	0	0	0	5,0
0,3	0	0	0,8	43,9
0,4	0	0	23,7	82,2
0,5	0	0	67,9	95,5
0,6	0	0,3	90,6	98,9
0,7	0	10,1	97,4	99,7
0,8	0	42,2	99,2	99,9
0,9	0	73,5	99,7	100
1	0	89,9	99,9	100

Tableau 12

Puissance du test à partir de la statistique T

distinctes. En effet, soit la puissance de T est très mauvaise (pour les petites valeurs de u et de v) : on est proche des bruits blancs, soit elle est bonne ou très bonne (ce qui correspond au trait renforcé).

On constate d'autre part que ces 2 puissances sont en quelque sorte "distribuées de manière symétrique".

Il est en fait difficile de comparer ces 2 tests au vu du calcul de ces puissances. On constate simplement que à u et v fixés, on atteint de meilleures puissances dans le cas du 2ème test.

5. CONCLUSION

Les tests que nous avons construits ici sont fondés sur l'étude des moments d'ordre 4 des modèles (1) et (2), et l'on a vu que les moments d'ordre inférieurs étaient insuffisants pour conclure. Ils permettent de décider entre un modèle appartenant à l'une des composantes du chaos de Wiener construit à partir des $\varepsilon(t)$, et un modèle appartenant au chaos tout entier. Si les calculs sont extrêmement longs et fastidieux, surtout pour le modèle (2), ils ne présentent pas de difficultés théoriques, et peuvent être généralisés à d'autres modèles bilinéaires ou même à des degrés plus élevés : par exemple, sachant que tout modèle "moyenne mobile" appartient à une somme finie de composantes du chaos (cfr. Quenouille, 1949), on peut se servir de ces tests pour différencier deux types de modèles, les modèles "moyenne mobile" contre les modèles "à partie autorégressive".

Mais en contrepartie, ces tests sont d'emploi extrêmement facile et demandent un minimum de calculs à l'utilisateur : l'estimation de moments d'ordre 4 (soit environ $3N$ multiplications pour le premier test et $6N$ pour le deuxième), évite le passage en spectral, et toute transformation de Fourier sur les données, transformation qui demande un temps de calcul d'autant plus long que pour estimer correctement le spectre, il faut moyennner sur une "fenêtre".

Un autre aspect positif de la démarche que nous proposons est qu'elle permet de mener de front estimation des paramètres et tests. A défaut d'un calcul exact de la vraisemblance, on est nécessairement amené à employer des estimateurs empiriques. Les résultats numériques que nous fournissons nous montrent qu'ils sont assez bons. Dans une étape ultérieure, quand on aura la forme de la vraisemblance, ils serviront efficacement d'estimateurs initiaux dans l'algorithme de maximisation de la vraisemblance. Il conviendra aussi de déterminer dans une étude ultérieure dans quelle mesure le remplacement dans les tests de "vraies" valeurs σ^2, δ, γ par leurs estimations $\hat{\sigma}^2, \hat{\gamma}, \hat{\delta}$ perturbe les tests. On rejoint là la théorie maintenant classique des tests avec estimation (cfr. Akaike, Atkinson, etc...).

Nous concluons en comparant nos résultats aux résultats analogues obtenus dans les modèles linéaires. Il faut constater que pour obtenir une précision analogue, on doit observer des séries environ 10 fois plus longues. Sachant que les modèles bilinéaires se caractérisent par de longues périodes où $X(t)$ varie peu, interrompues de façon aléatoire par de courtes périodes de forte excitation, il fallait s'attendre à un tel résultat : une observation de longueur 100 tombera souvent dans une plage d'inactivité relative, ne permettant aucune identification, quelquefois dans une période mouvementée laissant croire à un fort bruit. Il est nécessaire de compenser cette irrégularité de l'observation en x donnant une période assez longue pour espérer observer les deux types de comportement, et ceci demeurera essentiellement vrai quelle que soit la méthode que l'on pourra proposer.

REFERENCES

- Bartlett, M.S. (1946), On the Theoretical Specification and Sampling Properties of Autocorrelated Time Series, *Journal of the Royal Statistical Society*, Series B, 8, 27-41.
- Bruni, C., Dipillo, G., and Koch, G. (1974), Bilinear Systems : An Appealing Class of "Nearly Linear Systems", *I.E.E. Transactions on Automatic Control*, A.C. 19(4), 334-348.
- Guegan, D. (1981), Etude d'un modèle non linéaire, le modèle superdiagonal d'ordre 1, C.R.A.S., Série A, t. 293.
- Guegan, D. (1983), Cadre d'étude pour des modèles non linéaires, C.R.A.S., Série I, t. 296.
- Priestley, M.B. (1980), State Dependent Models : A General Approach to Non-Linear Time Series Analysis, *Journal of Time Series Analysis*, 1(1), 47-71.
- Quenouille, M.H. (1949), Approximate Tests of Correlation in Time Series, *Journal of the Royal Statistical Society*, Series B, 11, 68-84.
- Subba Rao, T. (1981), On the Theory of Bilinear Time Series Models, *Journal of the Royal Statistical Society*, Series B, 43(2), 244-255.
- Subba Rao, T. et Gabr, M. (1980), A Test for Linearity of Stationary Time Series, *Journal of Time Series Analysis*, 1(1), 145-158.

STATISTIQUE. — *Représentation l -markovienne et existence d'une représentation affine en l'état des modèles bilinéaires.* Note de Dominique Guegan, présentée par Robert Fortet.

Les modèles étudiés sont les modèles bilinéaires (S) :

$$(S) \quad X(t) = \sum_{i=1}^p a_i X(t-i) + \varepsilon(t) + \sum_{h=1}^q c_h \varepsilon(t-h) + \sum_{j=1}^Q \sum_{m=1}^P b_{jm} X(t-m) \varepsilon(t-j),$$

qui sont une généralisation non linéaire des modèles ARMA. Dans un premier temps, on généralise à tous ces modèles un résultat obtenu par P. Dinh dans [5] pour une sous-classe. On obtient ainsi une représentation l -markovienne des modèles bilinéaires. Dans une deuxième partie, on considère une représentation affine en l'état de ces modèles et on en donne une condition suffisante d'existence.

STATISTICS. — *l -markovian representation and existency of a state affine representation of the bilinear models.*

We investigate the bilinear models (S) which appear as a natural non linear generalization of the ARMA models. First we generalize to all these models a result obtained by P. Dinh in [5] for a subclass. Thus we obtain a l -markovian representation of these models. Secondly we consider a state affine representation of these models and we give for these models a sufficient condition of existency.

Les modèles bilinéaires (S) introduits par C. W. J. Granger et A. P. Anderson en 1978 apparaissent comme une généralisation non linéaire des modèles ARMA. Les développements probabilistes et statistiques de ces modèles n'en sont à l'heure actuelle qu'à leurs balbutiements à cause de la grande complexité de leur structure due au retard existant entre les observations et les bruits dans les termes produits. Avoir une représentation markovienne de ces modèles est un atout important car elle permet une meilleure connaissance de la structure probabiliste profonde de ces modèles. Cette Note est composée de deux parties. Dans un premier temps nous généralisons un résultat obtenu par P. Dinh dans [5] pour $l < 0$ [modèles (B₁)] : Nous donnons une représentation markovienne des modèles (B) dans le cas où $l > 0$; ces modèles appelés super-diagonaux sont notés (B₂). La situation des modèles (B₂) est beaucoup plus complexe que celle des modèles (B₁) car le bruit opère avec un certain retard sur les observations précédentes, situation qui intervient en particulier dans les phénomènes de stockage. Les résultats énoncés ne concernent que les modèles (B₂), une présentation générale pouvant être trouvée dans [2]. Dans la deuxième partie, on associe aux modèles (B) une représentation affine en l'état (A) pour laquelle on donne une condition suffisante d'existence.

I. REPRÉSENTATION l -MARKOVIENNE DES MODÈLES BILINÉAIRES. — Soit $X(t)$ l'observation d'un processus stochastique à temps discret défini sur un certain espace de probabilité (Ω, \mathcal{A}, P) , régie par :

$$(B) \quad X(t) = \sum_{i=1}^p a_i X(t-i) + \varepsilon(t) + \sum_{h=1}^q c_h \varepsilon(t-h) + \sum_{j=1}^Q \sum_{m=1}^P b_{jm} X(t-j+l) \varepsilon(t-j),$$

où $\varepsilon(t)$ est une suite de variables aléatoires indépendantes gaussiennes centrées de variance σ^2 définies aussi sur (Ω, \mathcal{A}, P) , (a_i) , (b_{jm}) et (c_h) étant des suites de nombres réels. On cherche une « représentation markovienne » pour une sous-classe des modèles (B) quand $l > 0$; nous dénommons (B₂) ces modèles.

Pour cela on considère comme espace des prédicteurs :

$$\mathcal{P}_t = \mathcal{P}_t \cup \{X_p, \dots, X_{t-p+\max(p, q, Q)+1}\} \cup \{\text{constantes}\},$$

où \mathcal{P}_t est le sous-espace fermé engendré par les éléments $\hat{X}_{t+k|t} = E[X(t+k)/\mathcal{F}_t]$, $k=1, 2, \dots$, avec \mathcal{F}_p la σ -algèbre engendrée par tous les ε_s , $s < t$. Les éléments de \mathcal{P}'_t ont alors pour expression :

$$\forall k > \max(p, q, Q), \hat{X}_{t+k|t} = \sum_{i=1}^p a_i \hat{X}_{t+k-i|t} + c,$$

où c est calculé à partir des coefficients du modèle. Il s'ensuit que pour $k > \max(p, q, Q)$, $\hat{X}_{t+k|t}$ peut être exprimé comme une combinaison linéaire de $\hat{X}_{t+1|t}, \dots, \hat{X}_{t+\max(p, q, Q)|t}$ plus une constante. L'espace des prédicteurs \mathcal{P}'_t est donc de dimension finie. Elle est au plus $\max(p, q+p, Q+p) + 1$.

Notons \mathcal{I}_t l'espace d'innovation, sous-espace fermé engendré par :

$$\hat{\eta}_t = \hat{X}_{t+k|t} - \hat{X}_{t+k|t-1}, \quad k=1, 2, \dots, \quad \mathcal{I}_t \subset \mathcal{P}'_t.$$

En remplaçant X_{t+k} par son expression (B₂) dans $\hat{\eta}_t$, et en calculant la contribution de chaque terme en fonction de k , les indices étant fixés, on obtient les éléments de l'espace d'innovation :

$$\hat{\eta}_t = \sum_{i=1}^{\max(p, q, Q)} a_i \hat{\eta}_{t-i} + c_k \varepsilon(t) + S_k - c,$$

où, à k fixé, $S_k = \sum_{m=1}^p b_{km} R_{t+k-m}$ avec $R_{t+k-m} = E[X(t+k-m)\varepsilon(t)/\mathcal{F}_t]$.

On exprime R_{t+k-m} à l'aide de variables \mathcal{F}_t -mesurables, en remplaçant $X(t+k-m)$ par son expression (B₂) autant de fois que nécessaire. Le calcul s'effectue en fait en prenant $X(t+l)$ car on travaille avec $k=j$ dans l'expression (B₂). On trouvera le détail des calculs dans [2].

Enfin pour obtenir la représentation markovienne, on choisit comme variables d'état, les éléments d'une base de \mathcal{P}'_t . On pose $n=r+s$ où r représente le nombre de variables $X(t)$ que l'on a besoin de mettre en mémoire et s le nombre de variables $X(t)$ que l'on peut prédire en utilisant les éléments de l'espace d'innovation. Alors :

$$r = \max(p, q - \max(q, p)) \quad \text{et} \quad s = n - r.$$

On pose :

$$Z_k(t) = X(t+k-n), \quad k=1, \dots, r$$

et

$$Z_{r+k}(t) = \sum_{i=k}^n a_i X(t+k-i) + \sum_{j=k}^s [c_j \varepsilon(t+k-j) + S_j], \quad k=1, \dots, s.$$

En calculant Z_k pas à pas, pour $k=1, \dots, r+s$, on obtient le théorème 1.

THÉORÈME 1. — Soit $X(t)$ le processus défini par les équations (B₂); $X(t)$ admet une représentation l-markovienne :

$$(M) \left\{ \begin{aligned} X(t) &= Z_r(t-1) + \varepsilon(t), \\ Z_k(t) &= Z_{k+1}(t-1), \quad 1 \leq k < r, \\ Z_r(t) &= Z_{r+1}(t-1) + \varepsilon_r, \\ Z_{r+k}(t) &= a_k Z_{r+1}(t-1) + Z_{r+k+1}(t-1) + (a_k + c_k) \varepsilon(t) \\ &\quad + \sum_{m=1}^p b_{km} V_m(t) + c_1 \varepsilon^2(t), \quad r < k \leq s-1, \\ Z_{r+s}(t) &= \sum_{i=s}^n a_i Z_{r+s+1-i}(t-1) + (a_s + c_s) \varepsilon(t) + \sum_{m=1}^p b_{sm} V_m(t) + c_2 \varepsilon^2(t), \end{aligned} \right.$$

où c_1 et c_2 sont calculés à partir de c et où :

$$\begin{aligned}
 V_m(l) = & \sum_{\lambda=1}^l \sum_{i_1 \dots i_\lambda=1}^p a_{i_1} \dots a_{i_\lambda} Z_{r+1-m}(t-1) \varepsilon(t) \\
 & + \sum_{\lambda=0}^l \sum_{i_1 \dots i_{\lambda_1}=1}^p \sum_{j_1 \dots j_{\lambda_2}=1}^Q \sum_{m_1 \dots m_{\lambda_2}=1}^P a_{i_1} \dots a_{i_{\lambda_1}} b_{j_1 m_1} \dots b_{j_{\lambda_2} m_{\lambda_2}} \\
 & \quad i_1 + \dots + i_{\lambda_1} + m_1 + \dots + m_{\lambda_2} \geq \lambda \quad \text{avec } \lambda_1 + \lambda_2 = \lambda \\
 & \times [Z_{r+1-m}(t-1) \varepsilon(t) + \varepsilon^2(t)] \prod_{\mu=1}^l \varepsilon(t+l-\mu-(j_1 + \dots + j_{\lambda_2})) \\
 & + \sum_{\lambda=1}^l \sum_{j_1 \dots j_\lambda=1}^Q \sum_{m_1 \dots m_\lambda=1}^P b_{j_1 m_1} \dots b_{j_\lambda m_\lambda} Z_{r+1-m}(t-1) \varepsilon(t) \\
 & \quad \times \prod_{\mu=1}^l \varepsilon(t+l-\mu-(j_1 + \dots + j_\lambda)).
 \end{aligned}$$

Remarque. — Dans l'expression de $Z_{r+s}(t)$ apparaît le produit $P = \sum_{v=0}^l \varepsilon(t-v) \varepsilon(t)$, on a au plus $(l+1)$ termes. On peut donc envisager une classification des modèles (B_2) en fonction du retard entre $X(t)$ et $\varepsilon(t)$, soit l . Plus cette distance est grande entre X et ε , plus il y a de termes dans le produit P , et moins la représentation (M) obtenue est markovienne, d'où l'expression choisie : représentation l -markovienne. On dira que les modèles (B_1) ont une représentation 1-markovienne.

II. REPRÉSENTATION AFFINE EN L'ÉTAT. — La représentation (M) ainsi obtenue des modèles bilinéaires apporte une connaissance très importante quant à la structure probabiliste des modèles (B). Un résultat de E. D. Sontag en 1979, [6], montre que l'on peut associer aux représentations (M) une représentation affine en l'état réalisable de dimension finie régie par les équations suivantes :

$$(A) \quad Z(t+1) = AZ(t) + P(\varepsilon(t))Z(t) + H\varepsilon(t+1), \quad X(t) = BZ(t),$$

où $Z(t) \in \mathcal{P} \subset \mathbb{R}^l$, $P(\varepsilon(t)) = \sum_{i=1}^n A_i \varepsilon^i(t)$, A et A_i étant des (l, l) matrices, B et H des vecteurs colonne. Le degré du polynôme P est égal à 2 pour les modèles (B_1) et est fonction de l pour les modèles (B_2). Il est alors possible pour ces modèles, comme pour les modèles (B) de donner leur développement de Volterra, cf. [3], et on peut obtenir une condition suffisante d'existence. Pour cela $\varepsilon(t)$ étant un bruit blanc, nous posons :

\mathcal{H}_0 : $\varepsilon(t)$ est strictement stationnaire;

\mathcal{H}_1 : $E[|\varepsilon^j(t)|] < \infty$, $1 \leq j \leq m$ pour un certain $m \in \mathbb{N}$;

\mathcal{H}_2 : $E[|\varepsilon^j(t)|^2] < \infty$, $1 \leq j \leq m$;

\mathcal{H}_3 : $\rho [I + 2A + A \otimes A + \sigma_m [2P + A \otimes P + P \otimes A] + \sigma_{2m} P \otimes P] = \mu < 1$,

où \otimes représente le produit matriciel de Kronecker défini par Neudecker [4] et $\rho(M)$ représente le rayon de la matrice M , soit $\rho(M) = \max[\lambda_i(M)]$, où $\lambda_i(M)$ est la i -ième valeur propre de M , et σ_k est le moment d'ordre k de $\varepsilon(t)$.

THÉORÈME 2. — Étant donné les hypothèses \mathcal{H}_0 - \mathcal{H}_3 , soit $X(t)$ tel que :

$$(2.1) \quad X(t) = BH\varepsilon(t) + B \sum_{k=1}^{t-1} \left[\sum_{j=k}^{t-1} [A + P(\varepsilon(j))] H\varepsilon(k) \right],$$

Alors $X(t)$ est un processus strictement stationnaire satisfaisant (A), $\forall t \in \mathbb{Z}$, et inversement, si $X(t)$, $t \in \mathbb{Z}$ est un processus strictement stationnaire satisfaisant (A) sous les hypothèses \mathcal{H}_0 - \mathcal{H}_3 , alors $X(t)$ est défini par (2.1).

La démonstration de ce théorème repose sur un lemme. Posons :

$$W_k = \sum_{j=k}^t [A + P(\varepsilon(j))] H \varepsilon(k);$$

$(W_k)_{k>1}$ est une suite de vecteurs aléatoires :

LEMME. — Sous les hypothèses \mathcal{H}_1 - \mathcal{H}_3 , la série $\sum_{k>1} W_k$ converge absolument presque sûrement.

On montre alors que sous \mathcal{H}_0 , $X(t)$ défini par (2.1) est strictement stationnaire et que par construction ce processus est solution de (A). Inversement on se sert du développement formel suivant :

$$(2.2) \quad X(t) = B H \varepsilon(t) + B \prod_{j=0}^n [A + P(\varepsilon(j))] X(t-n) + B \sum_{k=1}^{n-1} \left[\prod_{j=k}^{n-1} [A + P(\varepsilon(j))] H \varepsilon(k) \right],$$

$X(t)$ étant strictement stationnaire, $\{X(t-n), n > 1\}$ converge vers $H z(0)$ en distribution et $\prod_{i=0}^n [A + P(\varepsilon(j))] X(t-n)$ converge encore vers 0 en distribution; on peut donc trouver une sous-suite convergeant vers 0 p. s. En prenant la limite dans (2.2), on obtient (2.1) pour tout $t \in \mathbb{Z}$.

Remise le 18 novembre 1985.

RÉFÉRENCES BIBLIOGRAPHIQUES

- [1] C. W. J. CRANGER et A. P. ANDERSON, *An introduction to bilinear time series analysis*, Vanderhoeck et Ruprecht, Gottingen, 1978.
- [2] D. GUEGAN, *l-markovian representation of subdiagonal bilinear models*, Prepublication, Paris-Nord, n° 57, 1985.
- [3] D. CUEGAN, *Existence and stationarity for a polynomial state affine process*, Prepublication, Paris-Nord, n° 61, 1985.
- [4] H. NEUDECKER, Some theorem on matrix differentiation with special reference to Kronecker matrix product, *J. Amer. Stat. Assoc.*, 69, 1969, p. 153-159.
- [5] T. PHAM DINH, *Bilinear markovian representation and bilinear models*, Technical Reports, U.M.I.S.T., n° 169, 1983.
- [6] E. D. SONTAG, Realization theory of discrete time non linear systems, I, the bounded case, *I.E.E.E. Trans. Circuits*, 26, 1979, p. 342-356.

Équipe C.N.R.S.-U.A. n° 743, Département de Mathématiques, C.S.P., Université Paris-Nord, avenue Jean-Baptiste-Clément, 93430 Villetaneuse.

DIFFERENT REPRESENTATIONS FOR BILINEAR MODELS

BY DOMINIQUE GUEGAN

C.N.R.S., Orsay, France

Abstract. In this paper, we investigate different representations of bilinear models introduced by Granger and Andersen (1978). We consider first the decomposition of these models on the chaos of Wiener, then we give for the subdiagonal models a l -Markovian representation, generalizing a result obtained by Pham Dinh (1985) in the case of the superdiagonal models. Thus we consider the polynomial state affine representation. For these representations, we are interested in the existence and the uniqueness of the decomposition and the expression of the different moments of the models.

Keywords. Bilinear models; Wiener's decomposition; Markovian representation; state affine model; existence and stationarity of solutions; k -order moments.

INTRODUCTION

The general bilinear process of order p, q, P, Q defined by:

$$(A) \quad X_t = \sum_{i=1}^p a_i X_{t-i} + \varepsilon_t + \sum_{h=1}^q c_h \varepsilon_{t-h} + \sum_{j=1}^Q \sum_{k=1}^P b_{jk} X_{t-k} \varepsilon_{t-j},$$

where the observation X_t is a real stochastic process defined on some probability space $(\Omega, \mathfrak{A}, P)$ for every $t \in \mathbb{Z}$, for some constants $\{a_i, 1 \leq i \leq p\}$, $\{c_j, 1 \leq j \leq q\}$ and $\{b_{jk}, 1 \leq j \leq Q, 1 \leq k \leq P\}$, and some sequence $\{\varepsilon_t, t \in \mathbb{Z}\}$ of independent Gaussian random variables defined on $(\Omega, \mathfrak{A}, P)$ with common mean and common variance $\sigma^2 < \infty$, was introduced in 1978 by Granger and Andersen. We can distinguish two subclasses of these models: the (A_1) models called diagonal and superdiagonal models, for which $b_{jk} = 0$ if $k \leq j$; and the (A_2) models called subdiagonal models, for which $b_{jk} = 0$ if $k > j$. From the literature it appears that the second kind of model has never been investigated seriously. The (A_1) models have been investigated in detail because of the simplicity of their expansions; from a physical point of view, it seems more interesting to investigate the (A_2) class of models. In these models the noise operates with a delay, and this situation can arise particularly in stocking phenomena. Knowledge of the probabilistic structure of these models is very incomplete, because of the complexity of their structure. In this article, three different approaches are proposed for the study of the (A) models. For each of them, we analyse its limits and the future extensions. A first approach in terms of Volterra series and Wiener's chaos respectively has been made by Subba Rao and Gabr (1984) and Guegan (1983) for some models. In Section 1 we generalize the expression of the Volterra representation for all models (A) . Such an expansion, equation (B), based on the refor-

mulation of the model in vector form, is essentially formal. It can lead to sufficient conditions for the existence of a stationary solution of (A) in some particular cases, but it does not allow us to determine the all-important problem of the uniqueness of the solution, or the explicit expression of the moments, because of the tedious algebra of the Wiener's chaos.

A result which produces better understanding of the deep probabilistic structure of these models, and which is interesting in itself, is the Markovian representation of the models (A). Pham Dinh (1985) has established it for the (A_1) models. In the second part of this paper, we give a representation of the (A_2) models (equation (C)), generalizing the result obtained by Pham Dinh for the (A_1) models. The representation obtained in this way is not exactly Markovian, we call it an l -Markovian representation; nevertheless, it permits us to show the complexity of the probabilistic structure of these models. In fact, with this result we close the problem of the Markovian representation for all the nonlinear models thanks to a theorem proved by Bourdon (1982) which shows that the class of the bilinear models is dense in the class of nonlinear models.

Though the two previous representations (B) and (C) are natural because of the expression of the representation (A), they do not permit a real statistical study of these models. A point of view which seems interesting is that offered by the state affine representation as introduced by the automaticians and first investigated in discrete time by Sontag (1979) with the help of algebraic geometry. From a theorem of Sontag which says that bilinear (input-output) systems, among others (and whose models (A) are a particular case), can be represented by a polynomial state affine system (equation (D) in the third part of this paper), we are interested in these later models when the inputs are sequences of independent identically distributed Gaussian random variables. We establish under what conditions these models exist and are stationary and then give the expression for the different moments when they exist.

Throughout, we use the following notation: let B denote the closed subspace generated by ε_t ; B_t the one spanned by ε_s , $s \leq t$, $\mathfrak{F} = \mathfrak{F}(t)$ the σ -algebra generated by all the ε_t , and $\mathfrak{B} = \bigoplus_{n \geq 0} B^{\odot n}$ the direct Hilbertian sum of the symmetric tensorial powers of the space B .

1. FORMAL DECOMPOSITION ON THE COMPONENTS OF WIENER'S CHAOS

As we know that the Hilbertian frame of $L^2(\mathfrak{F})$ is not sufficient for investigation of the class of models (A), we can use $L^2(\mathfrak{F})$ with its Wiener's chaos structure obtained by means of an isomorphism which maps \mathfrak{B} on to $L^2(\mathfrak{F})$, and $B_t^{\odot n}$ on to the space of polynomials of degree n in $\varepsilon(s)$, $s \leq t$, and their limit in mean square (cf. Neveu, 1968), to investigate them. To obtain the decomposition on the chaos, we need to write the models (A) in a matrix form. Without losing generality, in this section, we let $n = \max(p, q, P, Q)$; also, we add as many zero as necessary in the matrix we use.

The equation (A) may be written in the following form:

$$(A') \quad \begin{aligned} \vec{X}_t &= A\vec{X}_{t-1} + C\varepsilon_t + \sum_{j=1}^n C_j \varepsilon_{t-j} + \sum_{j=1}^n B_j \vec{X}_{t-1} \varepsilon_{t-j}, \\ X_t &= H\vec{X}_t, \end{aligned}$$

where $\vec{X}_t = [X_t, \dots, X_{t-n}]$, $H = [1 \ 0 \ \dots \ 0]$, $C_j = [c_j \ 0 \ \dots \ 0]$ ($0 \leq j \leq n$), $C_0 = C = I$ with I the identity matrix.

$$A = \begin{pmatrix} a_1 & a_2 & \dots & a_n \\ 1 & 0 & & 0 \\ 0 & 1 & \dots & \dots \\ \vdots & \vdots & & \vdots \\ 0 & & & 0 \end{pmatrix}, \quad B_j = \begin{pmatrix} b_{j1} & b_{j2} & \dots & b_{jj} & \dots & b_{jn} \\ 0 & 0 & & & & 0 \\ \vdots & \vdots & & \vdots & & \vdots \\ 0 & \dots & \dots & \dots & \dots & 0 \end{pmatrix}$$

In the case of the (A₁) models, $b_{ji} = 0 \ \forall i \leq j$.
 In the case of the (A₂) models, $b_{ji} = 0 \ \forall i > j$.

THEOREM 1. *The formal decomposition of the models (A) on the Wiener's chaos is given by*

$$X_t = H \sum_{i=1}^{\infty} x_i(t),$$

where

$$(B) \quad \begin{aligned} x_i(t) &= \sum_{u_1=0}^i \dots \sum_{u_i=0}^i \sum_{j=1}^n \sum_{k=0}^n A^{u_1-1} B_j \delta(u_1 - l) \\ &\quad \times \prod_{l=2}^{i-1} A^{u_l - u_{l-1} - 1} B_j \delta(u_l - u_{l-1} - l) A^{u_i - u_{i-1} - 1} C_k \delta(u_i - u_{i-1}) \prod_{l=1}^{i-1} \varepsilon_{t-u_l-j+1} \varepsilon_{t-u_l-k} \end{aligned}$$

with $\delta(u)$ a step function defined by $\delta(u) = 1$ if $u \geq 0$ and $\delta(u) = 0$ otherwise, and H, A, B_j and C_k respectively square, square, row and column matrices, obtained from the vectorial representation (A') of the models (A).

To establish this result, we seek a solution of the equation (A') in the form of a power series expansion. For that, we use a method developed by Subba Rao and Gabr (1984) in the case $j = 1$ and $k = 0$. If λ is a numerical parameter, we consider the equation:

$$\vec{X}_t = A\vec{X}_{t-1} + \lambda \sum_{k=0}^n C_k \varepsilon_{t-k} + \lambda \sum_{j=1}^n B_j \vec{X}_{t-1} \varepsilon_{t-j}, \tag{1.1}$$

and we want a solution in the form

$$\vec{X}_t = \sum_{i=1}^{\infty} \lambda^i x_i(t) \tag{1.2}$$

(ultimately λ becomes equal to unity). We substitute (1.2) into (1.1) and equate the powers of λ on both sides to get:

$$x_1(t) = Ax_1(t-1) + \sum_{k=0}^n C_k \varepsilon_{t-k},$$

$$x_i(t) = Ax_i(t-1) + \sum_{j=1}^n B_j x_{i-j}(t-1) \varepsilon_{t-j}, \quad i > 1.$$

By recurrence, with the following initializations:

$$x_i(t) = 0 \quad \text{if } t \leq 0, \quad \text{and } \varepsilon_t = 0 \quad \text{if } t < 0,$$

we show that

$$x_1(t) = \sum_{k=0}^n \sum_{j=0}^t A^j C_k \varepsilon_{t-j-k},$$

$$x_i(t) = \sum_{l=0}^n \sum_{j=0}^{t-1} A^j B_l x_{i-1}(t-j-1) \varepsilon_{t-j-l}, \quad i > 1,$$

and obtain equation (B), and so prove Theorem 1.

As we know there exists an isomorphism between $L^2(\mathfrak{F})$ and \mathfrak{B} , we can use the result obtained by Pham Dinh (1985) to give a sufficient condition of existence for the formal decomposition (B) only for the (A_1) models. The condition is that 'the matrix $A^{\odot 2} + B^{\odot 2} \sigma^2$ has no eigenvalues of modulus greater than or equal to 1', where A and B are the matrices defined in the definition of the (A_1) models, and $A^{\odot 2}$ denotes the matrix associated with the linear transformation $W = AW A'$, where W represents the covariance matrix of the process $X(t)$ defined by equation (A). This condition depends of the covariance structure of the models and this one is always not known for the (A_2) models, even with the representation (B), so that the problem is always open for these models.

2. THE 'I-MARKOVIAN' REPRESENTATION OF THE (A_2) MODELS

In this section, for convenience, we consider the two following classes of models:

$$(A_1) \quad X_t = \sum_{i=1}^p a_i X_{t-i} + \varepsilon_t + \sum_{j=1}^q c_j \varepsilon_{t-j} + \sum_{j=1}^Q \sum_{m=0}^P b_{jm} X_{t-j-m} \varepsilon_{t-j},$$

where $b_{jm} = 0$ for all couples (j, m) for which $m < j$;

$$(A_2) \quad X_t = \sum_{i=1}^p a_i X_{t-i} + \varepsilon_t + \sum_{j=1}^q c_j \varepsilon_{t-j} + \sum_{j=1}^Q \sum_{m=1}^P b_{jm} X_{t-m} \varepsilon_{t-j},$$

where $b_{jm} = 0$ for all couples (j, m) for which $m > j$.

Note that only the definition of b_{jm} in (A_1) has been changed; (A_2) remains as (A) with the condition given. We generalize to the bilinear models (A_2) the result obtained in the linear case by Akaike (1975) (see also Guegan, 1981), and for the (A_1) models by Pham Dinh (1985). For this we consider the notion of predictor

space in a similar manner to Akaike. By analogy with the linear case, we define the nonlinear predictor space \mathbb{P}_t as the closed space spanned by the predictors $\hat{X}_{t+k/t} = E[X_{t+k}/\mathfrak{F}_t]$, $k = 1, 2, \dots$ and the innovation space \mathbb{M}_t as the closed subspace spanned by $\hat{\eta}_{t,k} = \hat{X}_{t+k/t} - \hat{X}_{t+k/t-1}$, $t = 1, 2, \dots$. Note that $\mathbb{M}_t \subset \mathbb{P}_t$, and all the elements of \mathbb{M}_t are noncorrelated with those of \mathbb{B}_{t-1} . For simplicity, we suppose first that the distance between X and ε is constant in the product terms in the equation (A₂). We denote this distance by l . We then let $b_{jj+m} = 0$ for all couples (j, m) for which $j - m \neq l$, so the equations (A₂) become:

$$(A'_2) \quad X_t = \sum_{i=1}^p a_i X_{t-i} + \varepsilon_t + \sum_{k=1}^q c_k \varepsilon_{t-k} + \sum_{j=1}^Q b_{jj+l} X_{t-j+l} \varepsilon_{t-j}.$$

To obtain the " l -Markovian representation" of the models (A'₂), we consider the predictor space \mathbb{P}'_t spanned by $\mathbb{P}_t, X_t, \dots, X_{t-\max(p, Q)+1}$ and products of ε_t :

$$\mathbb{P}'_t = \mathbb{P}_t \cup \{X_{t-m}, m = 0, \dots, p - \max(p, Q) + 1\} \\ \cup \left\{ \varepsilon_t, \sum_{m=1}^l \varepsilon_t \varepsilon_{t-1} \dots \varepsilon_{t-m}, \text{const} \right\}.$$

We can see that X_{t-m} , $m = 0, \dots, p - \max(p, Q) + 1$ belong to \mathbb{P}'_t , taking $p > \max(q, Q)$, and $k = \max(q, Q + 1), \dots, p$, or $k = p - 1$. The dimension of \mathbb{P}'_t is at most $\max(p, q, Q + p) + 2^{l+1}$ (cf. Lemma 1). We call this dimension n . We take as the state variables $Z(t) = (Z_1(t), \dots, Z_n(t))$, the elements of a basis of \mathbb{P}'_t . Let $n = r + s$, where r represents the number of the variables X_t we need to put in memory, and s the number of variables X_t we are going to predict using the elements of the innovation space (cf. Lemma 2). So $r = \max(Q, p - \max(q, Q))$ and $s = n - r$. Then we put

$$Z_k(t) = X_{t+k-r} \quad (k = 1, \dots, r), \\ Z_{r+k}(t) = \sum_{i=k}^n a_i X_{t+k-i} + \sum_{j=k}^s [c_j \varepsilon_{t+k-j} + T_{t+k,j}] \quad (k = 1, \dots, s), \quad (2.1)$$

where

$$T_{t+k,j} = \sum_{m=1}^p b_{jm} X_{t+k-m} \varepsilon_{t+k-j}$$

for the (A₁) models and

$$T_{t+k,j} = b_{j,j+l} E[X_{t+k-j+l} \varepsilon_{t+k-j} / \mathfrak{F}_t].$$

Then computing $Z(t)$ step by steps for every $k = 1, \dots, r + s$, and using lemmas 1 and 2, we get the following theorem.

THEOREM 2. *Let (X_t) be the process defined by equation (A₂). Then:*

- (i) *The dimension of the predictor space \mathbb{P}'_t is finite, at most equal to $\max(p, q, Q + p) + 2^{l+1}$.*
- (ii) *The expressions for $\hat{\eta}_{t,k}$ are given by (2.5) and (2.6).*

(iii) X_t admits the following 'l-Markovian representation':

$$\begin{aligned}
 X_t &= Z_r(t-1) + \varepsilon_t, \\
 Z_k(t) &= Z_{k+1}(t-1), \quad 1 \leq k < r, \\
 \text{(C)} \quad Z_r(t) &= Z_{r+1}(t-1) + \varepsilon_t, \\
 Z_{r+k}(t) &= a_k Z_{r+k}(t-1) + Z_{r+k-1}(t-1) + (a_k + c_k)\varepsilon_t \\
 &\quad + b_{kk+l} W_t(l), \quad k = r+1, \dots, s-1, \\
 Z_{r+s}(t) &= \sum_{i=s}^n a_i Z_{r+s+1-i}(t-1) + (a_s + c_s)\varepsilon_t + b_{ss+l} W_t(l)
 \end{aligned}$$

where

$$\begin{aligned}
 W_t(l) &= \sum_{\lambda=1}^l \sum_{n=0}^{\lambda} \sum_{i_1 \dots i_{\lambda_1}=0}^p \sum_{j_1 \dots j_{\lambda_2}=0}^Q \\
 &\quad \times a_{i_1} \dots a_{i_{\lambda_1}} b_{j_1 j_1+l} \dots b_{j_{\lambda_2} j_{\lambda_2}+l} Z_{t+nl+1}(t-1) \\
 &\quad \times \varepsilon_t \prod_{\mu=0}^m \varepsilon_{t+nl-\mu} + P(\varepsilon_t \dots \varepsilon_{t-l}),
 \end{aligned}$$

where $m \in \mathbb{N}^*$ and $m \geq l$, with

$$\begin{aligned}
 P_t(\varepsilon_t \dots \varepsilon_{t-l}) &= \sum_{\lambda=1}^l \sum_{n=0}^{\lambda} \sum_{i_1 \dots i_{\lambda_1}=0}^p \sum_{j_1 \dots j_{\lambda_2}=0}^Q a_{i_1} \dots a_{i_{\lambda_1}} b_{j_1 j_1+l} \dots b_{j_{\lambda_2} j_{\lambda_2}+l} \\
 &\quad \times \left[\sum_{k \geq nl-m} c_k \prod_{\mu=0}^m \varepsilon_{t+nl-\mu-k} \varepsilon_t - c_{nl-m} \sigma^2 \prod_{\mu=0}^m \varepsilon_{t+nl-\mu} - c_{nl-m} \sigma^2 \right].
 \end{aligned}$$

In the theorem, the summations are made for $\lambda_1 + \lambda_2 = \lambda$ and $i_1 + \dots + i_{\lambda_1} + j_1 + \dots + j_{\lambda_2} = m$ with the conventions $i_0 = 1$ and $b_{j0} = 1$.

REMARKS

(1) In the case of the (A_1) models, we use the predictor space $P'_t = \mathbb{P}_t \cup \{X_{t-m}, m = 0, \dots, p - \max(q + Q) + 1\} \cup \{\text{constants}\}$, and the relations (2.1) on $Z_{r+k}(t)$ for $r \leq k \leq r + s$ are easier: so we have

$$\begin{aligned}
 Z_{r+k}(t) &= a_k Z_{r+1}(t-1) + Z_{r+k+1}(t-1) + (a_k + c_k)\varepsilon_t \\
 &\quad + \sum_{m=0}^r b_{km} Z_{r+1-m}(t-1)\varepsilon_t + b_{k0} \varepsilon_t^2, \quad 1 \leq k \leq s-1 \\
 Z_{r+s}(t) &= \sum_{i=s}^n a_i Z_{r+s+1-i}(t-1) + (a_s + c_s)\varepsilon_t \\
 &\quad + \sum_{m=0}^r Z_{r+1-m}(t-1)\varepsilon_t + b_{(r+s)0} \varepsilon_t^2,
 \end{aligned}$$

expressions already found by Pham Dinh (1985).

(2) In the expression of $W_l(l)$ appears the product

$$\prod_{\mu=1}^l \varepsilon(t + nl - \mu).$$

This product has at most l terms where l represents the delay between the observation X and the noise ε , so we can think of ordering the subdiagonal model as a function of l . The greater the delay between X and ε , the more terms we have in the product

$$\prod_{\mu=1}^l \varepsilon(t + nl - \mu).$$

We call such a representation, a l -Markovian representation: in fact we suppose that independence is assumed between $\varepsilon(t)$ and $Z(s)$, $s < t$, and we note that the representation (B) depends at time t on a product of l terms in $\varepsilon(t - j)$.

(3) The proof of the theorem relies on two lemmas; we give them briefly here and we shall find their proofs in the Appendix. The conventions are the same as those used in Theorem 2.

LEMMA 1. *The elements of the predictor space are given by expression (2.2) for the (A_1) model, and expression (2.3) for the (A_2) model:*

$$\forall k > \max(p, q, Q) \quad \hat{X}_{t+k/t} = \sum_{i=1}^p a_i \hat{X}_{t+k-i/t} + \sum_{j=1}^q b_{j0} \sigma^2, \quad (2.2)$$

$$\forall k > \max(p, q, Q) \quad \hat{X}_{t+k/t} = \sum_{i=1}^p a_i \hat{X}_{t+k-i/t} + \sum_{j=1}^Q b_{jj+1} P_j(l) + C, \quad (2.3)$$

where $P_j(l)$ is a polynomial of bounded degree in ε , and C is some constant which depends on the parameters a_i , b_{jj+1} , c_k and σ^2 .

It follows from this lemma that for $k > \max(p, q, Q)$, $\hat{X}_{t+k/t}$ can be expressed as a linear combination of $\hat{X}_{t+k/t}, \dots, \hat{X}_{t+\max(p, q, Q)/t}$, ε_t , and products $\varepsilon_t \varepsilon_{t-1} \dots \varepsilon_{t-j}$, $j = 1, \dots, l$, plus a constant. So the prediction space has a finite dimension.

LEMMA 2. *The expressions for $\hat{\eta}_{t,k}$, k fixed, are given by (2.4) for the (A_1) models, and by (2.5)–(2.7) for the (A_2) models.*

$$\hat{\eta}_{t,k} = \sum_{i=1}^{\max(p, q, Q)} a_i \hat{\eta}_{t-i,k} + c_k \varepsilon_t + \sum_{m=1}^p b_{km} X_{t-m} \varepsilon_t - b_{k0} \sigma^2, \quad (2.4)$$

$$\hat{\eta}_{t,k} = \sum_{i=1}^{\max(p, q, Q)} a_i \hat{\eta}_{t-i,k} + c_k \varepsilon_t + b_{kk+1} R_t(l) - C, \quad (2.5)$$

where

$$\begin{aligned}
 R_t(l) &= \sum_{\lambda=1}^l \sum_{n=0}^{\lambda} \sum_{i_1 \dots i_{\lambda 1}=0}^p \sum_{j_1 \dots j_{\lambda 2}=0}^Q a_{i_1} \dots a_{i_{\lambda 1}} b_{j_1 j_1 + 1} \dots \\
 &\quad \times b_{j_{\lambda 2} j_{\lambda 2} + 1} X_{t+nl-m} \prod_{\mu=0}^m \varepsilon_{t+nl-\mu} \varepsilon_t \\
 &+ \sum_{\lambda=1}^l \sum_{n=0}^{\lambda} \sum_{i_1 \dots i_{\lambda 1}=0}^p \sum_{j_1 \dots j_{\lambda 2}=0}^Q a_{i_1} \dots a_{i_{\lambda 1}} b_{j_1 j_1 + 1} \dots b_{j_{\lambda 2} j_{\lambda 2} + 1} \\
 &\quad \times \left[\sum_{k > nl-m} c_k \prod_{\mu=0}^m \varepsilon_{t+nl-k-\mu} \varepsilon_t - c_{nl-m} \sigma^2 \prod_{\mu=0}^m \varepsilon_{t+nl-\mu} \right] \tag{2.6}
 \end{aligned}$$

and where

$$\begin{aligned}
 C &= \sum_{\lambda=1}^l \sum_{n=0}^{\lambda} \sum_{i_1 \dots i_{\lambda 1}=0}^p \sum_{j_1 \dots j_{\lambda 2}=0}^Q \\
 &\quad \times a_{i_1} \dots a_{i_{\lambda 1}} b_{j_1 j_1 + 1} \dots b_{j_{\lambda 2} j_{\lambda 2} + 1} c_{nl-m} \sigma^2 \tag{2.7}
 \end{aligned}$$

REMARK. The expression of $R_t(l)$ is the key of the representation we have obtained. In fact, what is important to note is that, in the product in ε , there are at most l terms in ε_{t-j} ; we can see that in the previous terms of the expression (2.6), and in the preceding terms, we have $(l-1)$ terms, $(l-2)$ terms, etc. The number of terms in the products depends on the maximal iterations that we need to obtain $R_t(l)$. If we suppose that l is nonconstant in the (A_2) expression, then we have some terms $X\varepsilon$, where the distance between X_{t-k} and ε_{t-j} is not always the same, and the state variables $Z(t)$ become

$$\begin{aligned}
 Z_k(t) &= X_{t+k-r} \quad (k = 1, \dots, r), \\
 Z_{r+k}(t) &= \sum_{i=k}^n a_i X_{t+k-i} + \sum_{j=k}^s [c_j \varepsilon_{t+k-j} + T_{t,j}] \quad (k = 1, \dots, s),
 \end{aligned}$$

where

$$T_{t,j} = \sum_l \sum_{m < j} b_{jm} R_t(j-m).$$

The (C) representation in this case is always the one given in Theorem 2, where we let $l = \sup_{\gamma} l_{\gamma}$ with $l_{\gamma} = j_{\gamma} - m_{\gamma}$ and $1 < \gamma < \max(P, Q)$. The expression (C), in spite of its intrinsic interest, again does not lead to general expressions for the covariance function or obtain theorems of existence such as those obtained by Pham Dinh for the (A_1) models.

3. POLYNOMIAL STATE AFFINE REPRESENTATION

We recall the definition of a state affine model and the result of Sontag on which we have based this work. The problem is the following. We consider a black box

in which some information, which we formalize by a control $u(t)$, enters; we denote by f the response we obtain. For investigation of this response, corresponding to observations X on the system, we consider two points of view. We can consider the system as an input-output system, or as a state affine system. Sontag's result is interesting because it establishes a link between two points of view.

(1) An output affine difference equation of order r for the response f is an equation such as:

$$\sum_{i=1}^r p_i(u(t), \dots, u(t-r))X(t-i) + p_{r+1}(u(t), \dots, u(t-r)) = 0$$

with p_i a polynomial in $u(t), \dots, u(t-r)$, $p_0 \neq 0$, which is understood to hold for all input-output pairs $(u(\cdot), x(\cdot))$ with $x(t) = f(u(1), \dots, u(t))$, $t > r$.

(2) A polynomial state affine system corresponds to the set of difference equations

$$\begin{aligned} \text{(S.A.)} \quad z(t+1) &= F(u(t))z(t) + H(u(t)), \\ X(t) &= G(u(t))z(t) + I(u(t)), \end{aligned}$$

where $F(u)$, $G(u)$ are matrices and $H(u)$, $I(u)$ are vectors, of polynomials in u . The dimension of the (S.A.) system is the dimension of the state space. A response f is realizable for a finite dimensional (S.A.) system if there exists some finite dimensional system realizing f . That a system realizes f means that if $u(\cdot)$ is the input, then the output of the system is $f(u, \cdot)$. Sontag's theorem is then: 'A polynomial response satisfies an output affine difference equation if and only if it is a polynomial state affine finitely realizable system'. If we consider equation (A) of the bilinear models, we have an input-output difference equation (E) of order $r = \max(p, q, Q)$, where the controls $u(t)$ are stochastic noises $\varepsilon(t)$. Thus, thanks to Sontag's theorem, with all bilinear models (A) there is associated a polynomial state affine model (S.A.) finitely realizable. The (S.A.) representation also can be obtained from the representation (C), if we consider a bigger predictor space, cf. Pham Dinh (1986).

In this section, we consider a discrete-time stochastic process $X(t)$ governed by a discrete-time polynomial state affine model, defined by the equations:

$$\begin{aligned} \text{(D)} \quad \text{(a)} \quad z(t+1) &= A(\varepsilon(t))z(t) + H(\varepsilon(t)), \\ \text{(b)} \quad X(t) &= B(\varepsilon(t))z(t) + C(\varepsilon(t)), \end{aligned}$$

where $X(t)$ represents the observation of the process at time t , and where we make the following hypothesis:

- (i) $z(t)$ is the state of the system, it belongs to the state space $Z \subset \mathbb{R}^n$: $z(t) = [z_1(t), \dots, z_n(t)]'$. $z(0)$ is given and is independent of $A(\varepsilon(t))$, $H(\varepsilon(t))$, for $t = 1, 2, \dots$.
- (ii) $\{\varepsilon(t)\}$, $t \in \mathbb{Z}$, is a sequence of independent identically distributed random variables, centred Gaussian with a finite variance σ^2 .
- (iii) A, B, H are respectively square, row and column matrices of finite degree polynomials in $\varepsilon(t)$. Their dimension are respectively $n \times n$, $n \times 1$ and $1 \times n$.

(iv) $C(\varepsilon(t))$ is a finite degree polynomial in $\varepsilon(t)$, and we suppose that $E[C(\varepsilon(t))] = 0$.

In all this part, we use the following notation: if M and N are two matrices: $E[\widetilde{M(\varepsilon(t))} \otimes \widetilde{N(\varepsilon(t))}] = \widetilde{M} \otimes \widetilde{N}$, $E[M(\varepsilon(t))N(\varepsilon(t))] = \widetilde{MN}$, $E[M(\varepsilon(t))] = \widetilde{M} \forall t$, where $\widetilde{M} \otimes \widetilde{N}$, \widetilde{MN} and \widetilde{M} are some functions of σ . To simplify the notations, we omit σ . \otimes represents the Kronecker product of matrices.

The main result of this section consists in finding a unique stationary solution $\{X(t)\}$ for which $X(t)$ is measurable with respect to \mathfrak{F}_t .

DEFINITION. The process $\{X(t)\}$, $t > 0$ is second order stationary if and only if $m_t = E[X(t)]$, $t > 0$ is constant and $V_{t,s} = E[(X(t) - m)(X(s) - m)']$, $t, s > 0$ depends only on the value of $(t - s)$.

In order to establish stationarity conditions for process $\{X(t)\}$ satisfying (D), we examine first the second order stationarity of the process $\{z(t)\}$ defined by (D(a)), and we show that second order properties of $\{X(t)\}$ may be found from those of the sequence $\{z(t)\}$. We let $\mu_t = E[z(t)]$ and $W_{t,s} = E[z(t)z'(s)]$.

THEOREM 3.1. *The process $\{z(t)\}$, $t > 0$ generated by (D(a)) from $t = 1$ is second order stationary if and only if $\mu_1 = \mu_0$ and $W_{1,1} = W_{0,0}$.*

The necessity of the conditions is obvious from the definition of stationarity. To prove their sufficiency, we use induction. Suppose $\mu_t = \mu$, $t = 0, 1, \dots, k$, and $W_{t,t-s} = W_{t-u,t-s-u} = W_s$, $t = s + 1, \dots, k$, $u = 1, \dots, t - s$, $s = 0, 1, \dots, k$. In the case $k = 1$, these conditions reduce to $\mu_1 = \mu_0 = \mu$ and $W_{1,1} = W_{0,0} = W$. From (D(a)), we have, under the induction hypothesis:

$$\begin{aligned} \mu_{k+1} &= E[z(k + 1)] = E[A(\varepsilon(k))z(k) + H(\varepsilon(k))] \\ &= E[A(\varepsilon(k))]E[z(k)] + E[H(\varepsilon(k))] \\ &= E[A(\varepsilon(k - 1))]E[z(k - 1)] + E[H(\varepsilon(k - 1))] = \mu_k, \end{aligned}$$

since $z(k)$ is \mathfrak{F}_{k-1} measurable and $A(\varepsilon(k))$ and $H(\varepsilon(k))$ are independent of $\{A(\varepsilon(1)), \dots, A(\varepsilon(k - 1))\}$, $\{H(\varepsilon(1)), \dots, H(\varepsilon(k - 1))\}$ and $z(0)$. For $1 \leq s \leq k$,

$$\begin{aligned} W_{k+1, k+1-s} &= E[z(k + 1)z'(k + 1 - s)] = E[(A(\varepsilon(k))z(k) \\ &\quad + H(\varepsilon(k)))z'(k + 1 - s)] \\ &= E[A(\varepsilon(k))]E[z(k)z'(k + 1 - s)] \\ &\quad + E[H(\varepsilon(k))]E[z'(k + 1 - s)] \\ &= E[A(\varepsilon(k - 1))]E[z(k - 1)z'(k - s)] \\ &\quad + E[H(\varepsilon(k - 1))]E[z'(k - s)] = W_{k, k-s}, \end{aligned}$$

while,

$$\begin{aligned} W_{k+1, k+1} &= E[z(k+1)z'(k+1)] \\ &= E[\{A(\varepsilon(k))z(k) + H(\varepsilon(k))\}\{A(\varepsilon(k))z(k) + H(\varepsilon(k))\}']. \end{aligned}$$

We compute $\text{vec } W_{k+1, k+1}$ (vec denotes the vector formed by stacking the columns of the corresponding matrix one on top of the other; see Neudecker, 1969) and, noting that the algebra of the matrix with the Kronecker product is a vectorial space, we have:

$$\begin{aligned} \text{vec } W_{k+1, k+1} &= \widetilde{A \otimes A} \text{vec } W_{k, k} + (\widetilde{A \otimes H} + \widetilde{H \otimes A})\mu + \widetilde{H \otimes H} \\ &= \widetilde{A \otimes A} \text{vec } W_{k-1, k-1} + (\widetilde{A \otimes H} + \widetilde{H \otimes A})\mu \\ &\quad + \widetilde{H \otimes H} = \text{vec } W_{k, k}. \end{aligned}$$

COROLLARY 3.1

(a) The process $\{z(t)\}$, $t \in \mathbb{Z}$ is second order stationary if and only if $\mu = E[z(t)]$ and $W = E[z(t)z'(t)]$ satisfy

$$\mu = \tilde{A}\mu + \tilde{H}, \quad (3.1)$$

with

$$\text{vec } W = \widetilde{A \otimes A} \text{vec } W + \Delta', \quad (3.2)$$

$$\Delta' = (\widetilde{A \otimes H} + \widetilde{H \otimes A})\mu + \widetilde{H \otimes H}. \quad (3.3)$$

(b) The process $\{X(t)\}$, $t \in \mathbb{Z}$ is second order stationary if and only if $m = E[X(t)]$ and $V = E[X(t)X(t)]$ satisfy $m = \tilde{B}\mu$ where μ is given by (3.1), and

$$\text{vec } V = \widetilde{B \otimes B} \text{vec } W + \Delta, \quad (3.4)$$

where W is given by (3.2), and

$$\Delta = \widetilde{2B \otimes C}\mu + \widetilde{C \otimes C}. \quad (3.5)$$

From the proof of Theorem 3.1, $\mu_1 = E[z(1)] = \mu_0$ and

$$\text{vec } W_{1, 1} = \text{vec } E[z(1)z'(1)] = \widetilde{A \otimes A} \text{vec } W_{0, 0} + (\widetilde{A \otimes H} + \widetilde{H \otimes A})\mu + \widetilde{H \otimes H}.$$

Thus we obtain (a). To get (b), we develop, $\forall t$, $E[X(t)]$ and $\text{vec}[X(t)X(t)]$ and use (3.2) and (3.3). Note that obviously the stationarity of $z(t)$ implies that of $X(t)$.

To obtain the existence of a stationary solution of (D), we prove the existence of two stationary processes $\{X(t)\}$ and $\{z(t)\}$ satisfying (D) for a given sequence $\{\varepsilon(t)\}$, which verifies (ii). In an attempt to find a \mathfrak{F}_t -measurable solution to (D), it is advantageous to obtain an expansion for $X(t)$ in terms of measurable function on \mathfrak{F}_t , by iterating the equation (D). For that we define the matrix product:

$$\prod_{k=1}^j M_k = \begin{cases} M_i M_{i+1} \cdots M_j & i \leq j, \\ I & j = i - 1, \end{cases}$$

and let $P_{t,r} = \prod_{k=1}^r A(\varepsilon(t-k))$, $r > 0$, and

$$P_{t,0} = I \quad \text{and} \quad Q_{t,r} = \sum_{j=0}^r P_{t,j-1} H(\varepsilon(t-j-1)).$$

THEOREM 3.2. *Suppose that the sequence $\{\varepsilon(t)\}$ verifies the hypothesis (ii) and that $E[H(\varepsilon(t))] = 0$, then a necessary and sufficient condition for the existence of a stationary solution of (D) is that either one of the following conditions hold:*

(a) *The matrix equation $\text{vec } W = \overline{A \otimes A} \text{vec } W + \overline{H \otimes H}$ admits a positive solution in W .*

(b) *The sequence W_k defined by $W_0 = W$, $W_k = E[A(\varepsilon(t))W_{k-1}A'(\varepsilon(t))] + \Delta_1$, where*

$$\Delta_1 = E[A(\varepsilon(t))z(t)H'(\varepsilon(t)) + H(\varepsilon(t))z'(t)A'(\varepsilon(t)) + H(\varepsilon(t))H'(\varepsilon(t))]$$

is convergent. Under the above conditions, a stationary solution of (D) is given by

$$\begin{aligned} X(t) &= B(\varepsilon(t))z(t) + C(\varepsilon(t)), \\ z(t) &= \sum_{j \geq 0} P_{t,j-1} H(\varepsilon(t-j-1)). \end{aligned} \tag{3.6}$$

The solution is unique if the equation $W = E[A(\varepsilon(t))WA'(\varepsilon(t))]$ admits no positive nontrivial solution.

It is clear that (b) implies (a) since the limit W of W_k , if it exists, satisfies the equation in (a). Conversely, suppose that the equation admits a positive solution W ; then W_k being defined as in (b), we have $W_0 \leq W_1$ and hence we can see that $W_k \leq W \forall k$, and if we start with $W_0 \leq W_1$ then $W_k \leq W_{k+1} \forall k$, thus the sequence converges. Elsewhere W_k is the covariance matrix of $Q_{t,r}$, and since, for all t , $\varepsilon(t)$ is independent of $\varepsilon(1), \dots, \varepsilon(t-1)$, then $H(\varepsilon(t-j-1))$ is independent of $P_{t,j-1}$ and for any vector x , the sequence $xQ_{t,r}$, $r > 0$ is a Martingale sequence with second moment bounded by $x'Wx$ and hence converges in mean square and almost surely to a limit $x'Q$, x (cf. Dacunha-Castelle and Duflo (1983)). It is easy to see that $z(t)$ is a solution of (D(a)). Conversely, if (D) admits a stationary solution, then (a) holds since the covariance matrix W of $z(t)$ satisfies (3.2). To see the uniqueness of the solution, we suppose that (D(a)) admits two solutions, $z_1(t)$ and $z_2(t)$. Let $U(t) = z_1(t) - z_2(t)$, then $U(t)$ satisfies $U(t) = A(\varepsilon(t-1))U(t-1)$, and since $U(t)$ is \mathfrak{F}_t measurable,

$$E[U(t)U'(t)] = E[A(\varepsilon(t-1))U(t-1)U'(t-1)A'(\varepsilon(t-1))].$$

However since $\{U(t)\}$ is stationary, we must have

$$E[U(t)U'(t)] = E[A(\varepsilon(t))U(t)U'(t)A'(\varepsilon(t))]$$

and $E[U(t)U'(t)] = 0$, and thus $U(t)$ has covariance matrix W . A solution of $W = E[A(\varepsilon(t))WA'(\varepsilon(t))]$ must be zero by assumption.

REMARK. If the process is not centred, it can always be reduced to a centred process.

In the following we give a sufficient condition for the existence of a unique stationary solution of equation (D).

THEOREM 3.3. *If the matrices \tilde{A} and $\widetilde{A \otimes A}$ have all their eigenvalues of modulus strictly less than unity, then the series (3.6) is convergent in the mean and if the $\varepsilon(t)$ verify hypothesis (ii), the series is convergent almost surely, and defines a unique stationary solution of (D).*

We shall show that the general term $P_{t,j-1}H(\varepsilon(t-j-1))$ of the series in (3.6) satisfies:

$$\sum_{j \geq 0} E |P_{t,j-1}H(\varepsilon(t-j-1))| < \infty, \tag{3.7}$$

and by the same argument used in the proof of Theorem 3.2, this would imply that the series converges in the mean, and if the $\varepsilon(t)$ are independent, converges almost surely. To establish this result, let σ_{2m} be the moments of order $2m$ of the sequence $\{\varepsilon(t)\}$; K_0, K_1, K_2 , be constants which depend on σ_{2m} and respectively on the matrices $\widetilde{A \otimes A}, \tilde{A}, \tilde{H}$; let $\lambda_0 = \rho(\widetilde{A \otimes A})$ and $\lambda_1 = \rho(\tilde{A})$, where $\rho(M)$ represents the maximum modulus of the eigenvalues of the matrix M . Then

$$E |P_{t,j-1}H(\varepsilon(t-j-1))| = E |P_{t,j-1}| E |H(\varepsilon(t-j-1))|$$

and we have, using the definition of $P_{t,j-1}$ and the properties of the Kronecker product of the matrices:

$$E |H(\varepsilon(t))| < K_2 \forall t \tag{3.8}$$

and

$$E |P_{t,j-1}| \leq nK_0 K_1 \lambda_0^{j-1} \lambda_1^{j-1}. \tag{3.9}$$

Then if $\lambda_0 < 1$ and $\lambda_1 < 1$, using (3.8) and (3.9), we have (3.7), and the convergence of the series $Q_{t,j}$ is assumed. We can see easily that the process defined by (3.6) is stationary and is a solution of (D) by construction. Conversely, by repeated application of (D), we obtain

$$\begin{aligned} X(t) &= B(\varepsilon(t))z(t) + C(\varepsilon(t)), \\ z(t) &= P_{t,r}z(t-r) + \sum_{j=0}^r P_{t,j-1}H(\varepsilon(t-j-1)). \end{aligned} \tag{3.10}$$

We previously showed the absolute convergence of the series $Q_{t,r}$ and as before we show that $E |P_{t,r}| < nK_0 \lambda_0^{r/2}$, so $\lim_{r \rightarrow \infty} E |P_{t,r}| = 0$. As the process $\{z(t)\}$ is stationary, $\{z(t-r), r > 0\}$ converges to $A(\varepsilon(0))z(0)$ in distribution and $P_{t,r}z(t-r)$ converges to 0 in distribution, hence in probability. So we can find a subsequence of this sequence which converges to 0 a.e. Taking limits all along this subsequence in (3.10), we obtain (3.6), and for $t > 0$ the uniqueness of the solution follows. Note that for the uniqueness of the solution, the condition given in the theorem is equivalent to that given in Theorem 3.2. In effect

$W = E[A(e(t))WA'(e(t))]$ is equivalent to $\text{vec } W = \overline{A \otimes A} \text{vec } W$. Otherwise the result we obtain here is similar to that of Bhaskara Rao *et al.* (1982), and Pham Dinh (1985) for some subclasses of the bilinear model (A).

We can also find an explicit condition for the L^2 convergence of the series $Q_{t,j}$. For that we must compute $E[Q_{t,j}Q_{t,j}]$, particularly $\text{vec } E[Q_{t,j}Q_{t,j}]$. Then using the properties of the Kronecker products of matrices, we obtain:

$$\text{vec } E[Q_{t,j}Q_{t,j}] = \sum_{j \geq 0} (\overline{A \otimes A})^{j-1} \text{vec } \overline{HH'} + \sum_{j \neq 1} (\tilde{A})^{j-1} \times (\tilde{A})^{j-1} \text{vec } \overline{HH'}$$

and a sufficient condition for the convergence of the series is given by $\rho(\overline{A \otimes A}) < 1$ and $\rho(\tilde{A}) < 1$.

We shall study the moments of processes satisfying equation (D). At the same time, we give sufficient conditions for the k -order stationarity of such processes.

THEOREM 3.4. *If $\rho(\tilde{A}) < 1$ then the process $\{X(t)\}$ is first order stationary and $m = \tilde{B}\mu$ with $\mu = (I - \tilde{A})^{-1}\tilde{H}$.*

The assumption $\rho(\tilde{A}) < 1$ implies that the matrix $I - \tilde{A}$ is invertible. It implies too the convergence of the series $\sum \tilde{A}^k$ (with limit $(I - \tilde{A})^{-1}$), and hence that of $\tilde{A}^1\mu_0 + \sum_{k=0}^{t-1} \tilde{A}^k\tilde{H}$ ($t \rightarrow \infty$), which proves the theorem.

THEOREM 3.5. *If $W_t = E[z(t)z'(t)]$ satisfies equation (3.2) and if $\rho(\overline{A \otimes A}) < 1$, then the process is second order stationary; then $V = E[X(t)X(t)]$ is given by $\text{vec } V = \overline{B \otimes B} \text{vec } W + \Delta$, where $W = (I - \overline{A \otimes A})^{-1}\Delta'$ with Δ and Δ' given by (3.5) and (3.3).*

The second order stationarity of the process $\{X(t)\}$ comes from the fact that $\text{vec } W_t = \overline{A \otimes A} \text{vec } W_{t-1} + \Delta'$. If $\rho(\overline{A \otimes A}) < 1$, the matrix $(I - \overline{A \otimes A})$ is invertible, and the solution of equation (3.2) converges. Now $V = E[X(t)X(t)]$ satisfies equation (3.4) and the result follows.

We can now compute the covariance function. Let

$$C(s) = E[(z(t+s) - \mu)(z(t) - \mu)'] \quad \text{and} \quad \Gamma(s) = E[(X(t+s) - m)(X(t) - m)']$$

As

$$E[z(t+s)z'(t)] = \tilde{A}^s E[z(t)z'(t)] + \sum_{k=0}^{s-1} \tilde{A}^k \tilde{H}\mu,$$

$$C(s) = \tilde{A}^s W + \sum_{k=1}^{s-1} \tilde{A}^k \tilde{H}\mu - \mu^2, \quad s > 0,$$

$C(0) = W - \mu^2$, and $\Gamma(s) = \tilde{B}C(s)\tilde{B}' - m^2$, $s > 0$ and $\Gamma(0) = \tilde{B}C(0)\tilde{B}' - m^2$. If the process is centred, the previous results reduce to $C(s) = \tilde{A}^s W$, $s > 0$, and $\Gamma(s) = \tilde{B}\tilde{A}^s W \tilde{B}'$, $s > 0$, so the covariance function of a centred process admitting the representation (D) has the same form as that of a certain autoregressive moving average process.

Computation of higher order moments is in principle possible. We generalize the preceding results. We suppose the process centred. Let

$$\Omega_{k_1, \dots, k_j, k} = E[A(\varepsilon)^{\otimes(k_1-1)} \otimes H(\varepsilon) \otimes A(\varepsilon)^{\otimes(k_2-k_1-1)} \otimes \dots \otimes H(\varepsilon) \otimes A(\varepsilon)^{\otimes(k-k_j)}],$$

where $1 \leq k_1 < k_2 < \dots < k_j \leq k$, and we have dropped the subscript t in $\varepsilon(t)$ (by convention A^0 is the unit matrix). Then one may put

$$\Omega_{k-j, j} = \sum_{1 \leq k_1 < \dots < k_j \leq k} \Omega_{k_1, \dots, k_j, k}.$$

We use the same convention for:

$$\Lambda_{k_1, \dots, k_j, k} = E[B(\varepsilon)^{\otimes(k_1-1)} \otimes B(\varepsilon)^{\otimes(k_2-k_1-1)} \otimes \dots \otimes B(\varepsilon)^{\otimes(k-k_j)} C(\varepsilon)^T]$$

and

$$M(\varepsilon)^{\otimes n} = M(\varepsilon) \otimes \dots \otimes M(\varepsilon),$$

where we take the products n times.

THEOREM 3.6. *The process $\{X(t)\}$ defined by (D) is k -order stationary, with $k = 2n$ if $\rho(\Omega_{2n, 0}) < 1$. Then if $m_k = E[X(t)^k]$ and $\mu_k = E[z(t)^{\otimes k}]$ we have the relations:*

$$\text{vec } m_k = B^{\otimes k} \text{vec } \mu_k + \sum_{j=1}^k \Lambda_{k-j, j} \text{vec } \mu_{k-j}, \quad \text{for } k = 2n, \quad (3.10)$$

$$\text{vec } \mu_k = \Omega_{k, 0} \text{vec } \mu_k + \sum_{j=1}^k \Omega_{k-j, j} \text{vec } \mu_{k-j}, \quad \text{with } \mu_0 = I. \quad (3.11)$$

Developing the expression in z , we have $z(t) = \sum_{j=0}^{\infty} P_{t, j-1} H(\varepsilon(t-j-1))$. The general term of this series, if $\rho(\Omega_{2n, 0}) < 1$, has moment of order $2n$ equal to $\Omega_{2n, 0}^j \Omega_{0, 2n}$, and the process in z is stationary. The moments μ_2, \dots, μ_{2n} must satisfy equation (3.12) and can be computed recursively. We can obtain in the same way the expressions of the moments of $\{X(t)\}$. Note that we proceed by induction to obtain the expressions (3.11) and (3.12). From the expression of $\text{vec}(z(t)z'(t))$, we compute $\text{vec}(\text{vec}(z(t)z'(t))z'(t)) = \text{vec}(z(t)z'(t)z'(t))$, and so on.

5. CONCLUSION

The representation (B) gives the explicit representation of the models (A) on Wiener's chaos. Algebra on Wiener's chaos is very tedious, so, for the moment, this approach does not permit generalization for all the models of the techniques of identification, particularly that based on the moments of the processes, as developed by Guegan (1982) to identify easy models. The (C) representation allows a better understanding of the deep probability structure of the models (A). It leads to the introduction of the representation (D), which offers more possibilities. This first approach of the polynomial state affine model deals with the problem of stationarity, existence and uniqueness of the solution of the equation

(D). We obtain, in fact, conditions for second order stationarity and existence for a process $X(t)$ defined by (D) for every t belonging to \mathbb{Z} . If we start with a certain $z(0)$ and if we consider the process defined by (D) with t belonging to \mathbb{N} , we need for $z(0)$ a suitable chosen distribution; in particular we need $z(0)$ independent of all functions of $\varepsilon(t)$ for $t > 1$ and admitting second order moments. Some problems occur when we begin with an inadequate initial value for $z(0)$. We then need to study the stability of the system generated by a such $z(0)$. This will be important when we consider procedures to identify such systems. Otherwise, from the expressions of the different moments we can hope to estimate the parameters of the models and establish tests to decide between different kind of models. Note that throughout this paper we suppose $\varepsilon(t)$ is Gaussian; in fact this restriction can be relaxed and it suffices to suppose the $\varepsilon(t)$ independent with finite moments for some order $m \geq 2$.

APPENDIX. PROOFS OF THEOREM 2 AND LEMMA 1.2

LEMMA 1. To compute the elements of the predictor space, for the (A_2) models, we use the (A_2) expression. We have:

For the (A_1) models:

$$(L.1) \quad \hat{X}_{t+k|t} = \sum_{i=1}^p a_i \hat{X}_{t+k-i|t} + \sum_{j=1}^q c_j \hat{\varepsilon}_{t+k-j|t} + \sum_{j=1}^q \sum_{m=0}^P b_{jm} E[X_{t+k-j-m} \varepsilon_{t+k-j} / \mathfrak{F}_t].$$

For the (A_2) models:

$$(L.2) \quad \hat{X}_{t+k|t} = \sum_{i=1}^p a_i \hat{X}_{t+k-i|t} + \sum_{j=1}^q c_j \hat{\varepsilon}_{t+k-j|t} + \sum_{j=1}^q b_{jj+l} E[X_{t+k-j+l} \varepsilon_{t+k-j} / \mathfrak{F}_t].$$

We investigate successively the contribution of each term of these expressions. For the two first terms on the right-hand side of these expressions, we observe that: at i fixed $\hat{X}_{t+k-i|t} \in \mathbb{P}_i$ if $k > i$, and $\hat{X}_{t+k-i|t} = X_{t+k-i}$, if $k \leq i$; at j fixed, $\hat{\varepsilon}_{t+k-j|t} = 0$ if $k > j$ and $\hat{\varepsilon}_{t+k-j|t} = \varepsilon_{t+k-j}$ if $k \leq j$. If we consider the third term, using the fact that ε_t is independent of X_t, X_{t-1}, \dots , we obtain, at j fixed:

For the (A_1) models:

$$E[X_{t+k-j-m} \varepsilon_{t+k-j} / \mathfrak{F}_t] = \begin{cases} \sigma^2 & \text{if } m = 0 \text{ and } k > j, \\ 0 & \text{if } m > 0 \text{ and } k > j, \end{cases}$$

For the (A_2) models: in the expectation of the expression (L.2) we replace $X_{t+k-j+l}$ by its corresponding (A_2) expression. For each ε which appears, we compare its index with that of ε_{t+k-j} ; when two indices are identical, we get a contribution. Finally we have $\forall k > j: E[X_{t+k-j+l} \varepsilon_{t+k-j} / \mathfrak{F}_t] = C + P_l(l)$, where C is a constant which depends on σ^2 , and on the parameters a_i, b_{jj+l} and c_j , and where $P_l(l) = \sum_{m=1}^l \alpha_{i_1} \dots \alpha_{i_m} \varepsilon_{t-j_1} \dots \varepsilon_{t-j_m}$, with $\alpha_{i_1} \dots \alpha_{i_m}$ being coefficients which depend on the parameters a_i, b_{jj+l} and c_j , and $1 \leq j_m \leq l-1$.

LEMMA 2. To find the elements of the innovation space, we need to evaluate:

For the (A_1) models:

$$(L.4) \quad \hat{X}_{t+k|t} - \hat{X}_{t+k|t-1} = \sum_{i=1}^p a_i [\hat{X}_{t+k-i|t} - \hat{X}_{t+k-i|t-1}] + \sum_{j=1}^q c_j [\hat{\varepsilon}_{t+k-j|t} - \hat{\varepsilon}_{t+k-j|t-1}] + S_{t+k}^1,$$

where

$$(L.5) \quad S_{t+k}^1 = \sum_{j=1}^q \sum_{m=0}^P b_{jm} \{E[X_{t+k-j-m} \varepsilon_{t+k-j} / \mathfrak{F}_t] - E[X_{t+k-j-m} \varepsilon_{t+k-j} / \mathfrak{F}_{t-1}]\}.$$

For the (A_2) models:

$$(L.6) \quad \hat{X}_{t+k|t} - \hat{X}_{t+k|t-1} = \sum_{i=1}^p a_i [\hat{X}_{t+k-i|t} - \hat{X}_{t+k-i|t-1}] + \sum_{j=1}^q c_j [\hat{\varepsilon}_{t+k-j|t} - \hat{\varepsilon}_{t+k-j|t-1}] + S_{t+k}^2,$$

where

$$(L.7) \quad S_{t+k}^2 = \sum_{j=1}^Q b_{j,j+k} (E[X_{t+k-j+1} \varepsilon_{t+k-j} / \mathfrak{F}_t] - E[X_{t+k-j+1} \varepsilon_{t+k-j} / \mathfrak{F}_{t-1}]).$$

In each expression (L.4) and (L.6), when $k = j$, the second summation reduces to $c_k \varepsilon_t$. In (L.5), at j fixed, we put $k = j$, then: if $m = 0$, $S_{t+k}^1 = b_{k0} X_t \varepsilon_t - b_{k0} \sigma^2$, and if $m > 0$ $S_{t+k}^2 = \sum_{m=1}^p b_{km} X_{t-m} \varepsilon_t$. In (L.6), at j fixed, for $k = j$ we must evaluate $S_{t+k}^2 = b_{kk} V_t(l) - Q_t(l)$, where $V_t(l) = E[X_{t+l} \varepsilon_t / \mathfrak{F}_t]$ and $Q_t(l) = E[X_{t+l} \varepsilon_t / \mathfrak{F}_{t-1}]$. To compute $V_t(l)$ in terms of the products of ε_{t-j} , $j \geq 0$, we replace X_t by its (A_2) expression, so we obtain:

$$(L.8) \quad V_t(l) = E \left[\sum_{i=1}^p a_{i1} X_{t+1-i} \varepsilon_t + \sum_{j=1}^q c_j \varepsilon_{t+j} \varepsilon_t + \sum_{j=1}^Q b_{j_1 j_1 + 1} X_{t+2l-j_1} \varepsilon_{t+l-j_1} \varepsilon_t / \mathfrak{F}_t \right]$$

If necessary, we begin again; for example at the third step we obtain:

$$(L.9) \quad \begin{aligned} V_t(l) = E \left[\sum_{i_1 i_2 i_3} a_{i_1} a_{i_2} a_{i_3} X_{t+l-i_1-i_2-i_3} \varepsilon_t + \sum_j c_j \varepsilon_{t+j} \varepsilon_t \right. \\ + \sum_h c_h \left[\sum_{i_1} a_{i_1} \varepsilon_{t+l-i_1-h} + \sum_{i_1 i_2} a_{i_1} a_{i_2} \varepsilon_{t+l-i_1-i_2-h} + \sum_{j_1} b_{j_1 j_1 + 1} \varepsilon_{t+2l-j_1-h} \right] \varepsilon_t \\ + \sum_h \sum_{i_1} \sum_{j_1} a_{i_1} b_{j_1 j_1 + 1} c_h (\varepsilon_{t+2l-j_1-i_1-h} \varepsilon_{t+l-j_1-i_1} + \varepsilon_{t+3l-j_1-i_1-h} \varepsilon_{t+l-j_1}) \varepsilon_t \\ + \sum_{j_1 j_2} \sum_h b_{j_1 j_1 + 1} b_{j_2 j_2 + 1} c_h \varepsilon_{t+3l-j_1-j_2-h} \varepsilon_{t+2l-j_1-j_2} \varepsilon_{t+l-j_1} \varepsilon_t \\ + \sum_{i_1 i_2} \sum_{j_1} a_{i_1} a_{i_2} b_{j_1 j_1 + 1} X_{t+2l-j_1-i_1-i_2} [\varepsilon_{t+l-i_1-i_2-j_1} + \varepsilon_{t+l-i_1-j_1}] \varepsilon_t \\ + \sum_{i_1 i_2} \sum_{j_1} a_{i_1} a_{i_2} b_{j_1 j_1 + 1} X_{t+3l-j_1-i_1-i_2} \varepsilon_{t+l-j_1} \varepsilon_t \\ + \sum_{i_1} \sum_{j_1 j_2} a_{i_1} b_{j_1 j_1 + 1} b_{j_2 j_2 + 1} X_{t+3l-j_1-j_2-i_1} \\ \times [\varepsilon_{t+2l-j_1-j_2-i_1} \varepsilon_{t+l-j_1-i_1} + \varepsilon_{t+2l-j_1-j_2} \varepsilon_{t+l-j_1}] \varepsilon_t \\ + \sum_{i_1} \sum_{j_1 j_2} a_{i_1} b_{j_1 j_1 + 1} b_{j_2 j_2 + 1} X_{t+4l-j_1-j_2-i_1} \varepsilon_{t+3l-j_1-j_2-i_1} \varepsilon_{t+l-j_1} \varepsilon_t \\ + \sum_{j_1 j_2 j_3} b_{j_1 j_1 + 1} b_{j_2 j_2 + 1} b_{j_3 j_3 + 1} X_{t+4l-j_1-j_2-j_3} \\ \times \varepsilon_{t+3l-j_1-j_2-j_3} \varepsilon_{t+2l-j_1-j_2} \varepsilon_{t+l-j_1} \varepsilon_t / \mathfrak{F}_t \left. \right] \end{aligned}$$

and so on. Then

$$(L.10) \quad \begin{aligned} V_t(l) = \sum_{\lambda=1}^l \sum_{n=0}^{\lambda} \sum_{i_1 \dots i_{\lambda-1}=0}^p \sum_{j_1 \dots j_{\lambda-2}=0}^Q a_{i_1} \dots a_{i_{\lambda-1}} b_{j_1 j_1 + 1} \dots b_{j_{\lambda-2} j_{\lambda-2} + 1} X_{t+nl-m} \prod_{\mu=0}^m \varepsilon_{t+nl-\mu} \varepsilon_t \\ + \sum_{\lambda=1}^l \sum_{n=0}^{\lambda} \sum_{i_1 \dots i_{\lambda-1}=0}^p \sum_{j_1 \dots j_{\lambda-2}=0}^Q \sum_{h \geq nl-m} a_{i_1} \dots a_{i_{\lambda-1}} b_{j_1 j_1 + 1} \dots \\ \times b_{j_{\lambda-2} j_{\lambda-2} + 1} c_h \prod_{\mu=0}^m \varepsilon_{t+nl-h-\mu} \varepsilon_t, \end{aligned}$$

where the summations are made for $\lambda_1 + \lambda_2 = \lambda$ and $i_1 + \dots + i_{\lambda_1} + j_1 + \dots + j_{\lambda_2} = m$ with the conventions $i_0 = 1$ and $b_{j0} = 1$, and with $m \geq l$. We can see in the product $\prod_{a>0} \varepsilon_{t-a}$, there are at most l terms. To obtain the expression for $Q_t(l)$, we proceed in the same way as for $V_t(l)$:

$$Q_t(l) = E \left[\sum_{i=1}^p a_{i1} X_{t+1-i} + \varepsilon_{t+1} \varepsilon_t + \sum_{h=1}^q c_h \varepsilon_{t+h} \varepsilon_t + \sum_{j_1=1}^Q b_{j_1 j_1 + 1} X_{t+2l-j_1} \varepsilon_t / \mathfrak{F}_t \right],$$

but in this case, the only terms which are non-null when we develop $Q_t(l)$ with \mathfrak{F}_{t-1} measurable functions are those in which it appears as ε_t^2 . We do the same as for expression (L.8) and (L.9) and

finally, with the same notation, we get for $m \geq l$:

$$(L.11) \quad Q_t(l) = c_l \sigma^2 + \sum_{\lambda=1}^l \sum_{n=0}^{\lambda} \sum_{i_1 \dots i_{\lambda 1}=0}^p \sum_{j_1 \dots j_{\lambda 2}=0}^q a_{i_1} \dots a_{i_{\lambda 1}} b_{j_1 j_1 + 1} \dots b_{j_{\lambda 2} j_{\lambda 2} + 1} c_{nl-m} \sigma^2 + \sum_{\lambda=1}^l \sum_{n=0}^{\lambda} \sum_{i_1 \dots i_{\lambda 1}=0}^p \sum_{j_1 \dots j_{\lambda 2}=0}^q a_{i_1} \dots a_{i_{\lambda 1}} b_{j_1 j_1 + 1} \dots b_{j_{\lambda 2} j_{\lambda 2} + 1} c_{nl-m} \sigma^2 \prod_{\mu=0}^m \varepsilon_{t+nl-\mu}.$$

Then $S_{t+k}^2 = b_{kk+l} [V_t(l) - Q_t(l)]$, where $V_t(l)$ and $Q_t(l)$ are given by (L.10) and (L.11). We can write this expression as $S_{t+k}^2 = b_{kk+l} R_t(l) - C$, where for $m \geq l$:

$$(L.12) \quad R_t(l) = \sum_{\lambda=1}^l \sum_{n=0}^{\lambda} \sum_{i_1 \dots i_{\lambda 1}=0}^p \sum_{j_1 \dots j_{\lambda 2}=0}^q a_{i_1} \dots a_{i_{\lambda 1}} b_{j_1 j_1 + 1} \dots b_{j_{\lambda 2} j_{\lambda 2} + 1} X_{t+nl-m} \prod_{\mu=0}^m \varepsilon_{t+nl-\mu} \varepsilon_t + \sum_{\lambda=1}^l \sum_{n=0}^{\lambda} \sum_{i_1 \dots i_{\lambda 1}=0}^p \sum_{j_1 \dots j_{\lambda 2}=0}^q a_{i_1} \dots a_{i_{\lambda 1}} b_{j_1 j_1 + 1} \dots b_{j_{\lambda 2} j_{\lambda 2} + 1} \times \left[\sum_{h \geq nl-m}^m c_h \prod_{\mu=0}^m \varepsilon_{t+nl-\mu-h} - c_{nl-m} \sigma^2 \prod_{\mu=0}^m \varepsilon_{t+nl-\mu} \right]$$

and

$$(L.13) \quad C = \sum_{\lambda=1}^l \sum_{n=0}^{\lambda} \sum_{i_1 \dots i_{\lambda 1}=0}^p \sum_{j_1 \dots j_{\lambda 2}=0}^q a_{i_1} \dots a_{i_{\lambda 1}} b_{j_1 j_1 + 1} \dots b_{j_{\lambda 2} j_{\lambda 2} + 1} b_{kk+l} c_{nl-m} \sigma^2$$

and so we get Lemma 2.

PROOF OF THEOREM 2. We put

$$Z_k(t) = X_{t+k-r} \quad (k = 1, \dots, r),$$

$$Z_{r+k}(t) = \sum_{i=k}^n a_i X_{t+k-i} + \sum_{j=k}^s [c_j \varepsilon_{t+k-j} + T_{t+k,j}] \quad (k = 1, \dots, s),$$

where

$$T_{t+k,j} = \sum_{m=k}^p b_{jm} X_{t+k-m} \varepsilon_{t+k-j}$$

for the (A_1) models and

$$T_{t+k,j} = b_{kk+l} E[X_{t+k-j+l} \varepsilon_{t+k-j} / \mathfrak{F}_t]$$

for the (A_2) models. It can be checked, using (A), the expressions for S_{t+k}^2 , $i = 1, 2$, and the relation:

$\hat{X}_{t+l} = \hat{X}_{t+l-1} + \varepsilon_t$, that:

For the (A_1) models:

$$Z_k(t) = Z_{k+1}(t-1) \quad (k = 1, \dots, r-1),$$

$$Z_r(t) = Z_{r-1}(t-1) + \varepsilon_t,$$

$$Z_{r+k}(t) = a_k Z_{r+k}(t-1) + Z_{r+k+1}(t-1) + (a_k + c_k) \varepsilon_t + \sum_{m=0}^p b_{km} Z_{r+1-m}(t-1) \varepsilon_t + b_{k0} \sigma^2,$$

$$(k = 1, \dots, s-1)$$

$$Z_{r+s}(t) = \sum_{i=1}^n a_i Z_{r+s+1-i}(t-1) + (a_s + c_s) \varepsilon_t + \sum_{m=0}^p b_{sm} Z_{r+1-m}(t-1) \varepsilon_t + b_{(r+s)0} \varepsilon_t^2.$$

and we find the expressions already given by Pham Dinh.

For the (A_2) models:

$$\begin{aligned} Z_k(t) &= Z_{k+1}(t-1) \quad (k = 1, \dots, r-1), \\ Z_r(t) &= Z_{r+1}(t-1) + \varepsilon_t, \\ Z_{r+k}(t) &= Z_{r+k+1}(t-1) + a_k X_t + c_k \varepsilon_t + b_{kk+i} R_i(t) \quad (k = 1, \dots, s-1), \\ Z_{r+s}(t) &= \sum_{i=s}^n a_i X_{t+s-i} + c_s \varepsilon_t + b_{ss+i} R_i(t). \end{aligned}$$

Replacing X_t by $Z_{r+1}(t-1) + \varepsilon_t$ and $R_i(t)$ by $W_i(t)$, the last two relations become:

$$\begin{aligned} Z_{r+k}(t) &= a_k Z_{r+1}(t-1) + Z_{r+k+1}(t-1) + (a_k + c_k) \varepsilon_t + b_{kk+i} W_i(t), \quad (k = 1, \dots, s-1), \\ Z_{r+s}(t) &= \sum_{i=s}^n a_i Z_{r+s+1-i}(t-1) + (a_s + c_s) \varepsilon_t + b_{ss+i} W_i(t), \end{aligned}$$

where $W_i(t)$ is obtained from $R_i(t)$ and C is given by equations (L.12) and (L.13). Using the relation for $Z_i(t)$, we have for $m \geq t$:

$$\begin{aligned} W_i(t) &= \sum_{\lambda=1}^l \sum_{n=0}^{\lambda} \sum_{i_1 \dots i_{\lambda 1}=0}^p \sum_{j_1 \dots j_{\lambda 2}=0}^q a_{i_1} \dots a_{i_{\lambda 1}} b_{j_1 j_1 + i} \dots b_{j_{\lambda 2} j_{\lambda 2} + i} Z_{t+n+1}(t-1) \\ &\quad \times \prod_{\mu=0}^m \varepsilon_{t+n-\mu} + P_i(\varepsilon_t, \varepsilon_{t-1}, \dots, \varepsilon_{t-m}), \end{aligned}$$

with

$$\begin{aligned} P_i(\varepsilon_t, \dots, \varepsilon_{t-m}) &= \sum_{\lambda=1}^l \sum_{n=0}^{\lambda} \sum_{i_1 \dots i_{\lambda 1}=0}^p \sum_{j_1 \dots j_{\lambda 2}=0}^q a_{i_1} \dots a_{i_{\lambda 1}} b_{j_1 j_1 + i} \dots b_{j_{\lambda 2} j_{\lambda 2} + i} \\ &\quad \times \left[\sum_{h > n-m} c_h \prod_{\mu=0}^m \varepsilon_{t+n-h-\mu} \varepsilon_t - c_{n-m} \sigma^2 \prod_{\mu=0}^m \varepsilon_{t+n-\mu} - c_{n-m} \sigma^2 \right] \end{aligned}$$

and so we get Theorem 2.

ACKNOWLEDGEMENT

The author wishes to thank the referee for many helpful suggestions and comments which led to the final version of this paper.

REFERENCES

- AKAIKE, H. (1975) Markovian representation of stochastic processes by canonical variables. *SIAM J. Control*, 13(1), 162-173.
- BHASKARA RAO, M., SUBBA RAO, T., WALKER, A. M. (1983) On the existence of some bilinear time series models. *J. Time Series Anal.* 4, 95-110.
- BOURDON, P. (1982) An approach of discrete time non linear systems using stochastic inputs. *Systems Control Lett.* 1(6), 382-4.
- DACUNHA-CASTELLE, D. and DUFLO, M. (1983) *Probabilités et Statistiques*, Vol. 2. Paris: Masson.
- GRANGER, C. W. J. and ANDERSEN, A. P. (1978) *An Introduction to Bilinear Time Series Analysis*. Groningen: Vanderhoeck and Ruprecht.
- GUEGAN, D. (1981) Processus ARMA multidimensionnels—Représentation markovienne. Prépublications U.P.N. No. 19.
- (1982) Tests de modèles non linéaires. In *Alternative Approaches to Time Series Analysis* (Proceeding of the 3rd Franco Belgium Meeting of Statisticians). Publ. de la Fac. St. Louis, Belgique, 45-65.
- (1983) Cadre d'étude pour des modèles non linéaires. *Compte Rend. Acad. Si. Paris* 296, Serie I, 167-70.
- NEVEU, J. (1968) *Processus Aléatoires Gaussiens*. Presses de L'Université de Montréal.
- NEUDECKER, H. (1969) Some theorems on matrix differentiation with special reference to Kronecker matrix product. *J. Amer. Statist. Assoc.* 69, 953-63.

- PHAM DINH, T. (1985) Bilinear markovian representation and bilinear models. *Stochastic Processes and their Applications*. 20, 295–306.
- (1986) The mixing property of bilinear and generalised random coefficient autoregressive models. To appear in *Stochastic Processes and their Applications*.
- PRIESTLEY, M. B. (1981) *Spectral Analysis and Time Series*. London: Academic Press.
- SONTAG, E. D. (1979) Realization theory of discrete time non linear systems. I. The bounded case. *I.E.E.E. Trans. Cir. Syst.* 26, 342–56.
- SUBBA RAO, T. (1981) On the theory of bilinear time series models. *J. Roy. Statist. Soc. B* 43, 2, 224–55.
- and GABR, M. M. (1984) An introduction to bispectral analysis and bilinear time series models. *Lecture Notes in Statistics* No. 24. Berlin: Springer Verlag.

STATISTIQUE. — *Minimalité et inversibilité des modèles bilinéaires à temps discret.*
 Note de Dominique Guegan et Tuan Pham Dinh, présentée par Robert Fortet.

On présente ici les conditions de minimalité pour des modèles polynomiaux affines en l'état, ainsi qu'une condition suffisante d'inversibilité pour des modèles bilinéaires diagonaux.

STATISTICS. — *Minimality and inversibility of discrete time bilinear models.*

We present here some conditions of minimality for polynomial state affine models, as a sufficient condition of invertibility for diagonal bilinear models.

Les modèles bilinéaires à temps discret tels qu'ils ont été introduits par C. W. J. Granger et A. P. Andersen [1], ont été étudiés depuis une dizaine d'années. Dans le cas de la représentation (1) nous généralisons aux processus non centrés le théorème d'existence obtenu par D. Guegan ([2], [3]) pour des processus centrés, puis nous donnons des conditions pour avoir la quasi-minimalité de la représentation. Pour une sous classe des modèles (3) introduits par T. Pham Dinh, [4], nous établissons une condition suffisante d'inversibilité de ces modèles, condition importante pour obtenir des algorithmes d'estimation des paramètres.

MODÈLES POLYNOMIAUX À ÉTAT AFFINE : EXISTENCE ET QUASI-MINIMALITÉ. — On considère les modèles suivants :

$$(1) \quad \begin{cases} (a) & z(t) = A(\varepsilon(t))z(t-1) + H(\varepsilon(t)) \\ (b) & X(t) = B(\varepsilon(t))z(t-1) + \varepsilon(t) \end{cases}$$

où z est le vecteur d'état, X l'observation du système et $\varepsilon(t)$ une suite de v. a. i. équidistribuées, gaussiennes, centrées, de variance σ^2 . Si M et N sont deux matrices quelconques, on pose $\widetilde{M \otimes N} = E[M(t) \otimes N(t)]$, où \otimes représente le symbole de Kronecker ($\widetilde{M \otimes N}$ est en fait une fonction de σ dans laquelle on omet la variable σ pour des raisons de simplicité).

THÉOREME 1. — *Une condition nécessaire et suffisante d'existence d'une solution stationnaire au second ordre de (1), c'est-à-dire l'existence de deux processus stationnaires au second ordre satisfaisant (1) est que*

(a) *l'équation $\mu = [EA(\varepsilon(t))]\mu + EH(\varepsilon(t))$ admette une solution en μ , et que l'une des conditions suivantes soit vraie :*

(b) *l'équation matricielle $\text{vec } W = \widetilde{A \otimes A} \text{vec } W + \Delta$ où*

$$\Delta = E[A(\varepsilon(t))\mu H'(\varepsilon(t)) + H(\varepsilon(t))\mu' A'(\varepsilon(t)) + H(\varepsilon(t))H'(\varepsilon(t))]$$

admet une solution $W \geq \mu\mu'$:

(c) *la suite W_k définie par $W_0 = \mu\mu'$, $\text{vec } W_k = \widetilde{A \otimes A} \text{vec } W_k + \Delta$ est convergente.*

Sous ces conditions, une solution stationnaire est donnée par

$$\begin{cases} X(t) = B(\varepsilon(t))z(t-1) + \varepsilon(t) \\ z(t) = \sum_{j \geq 0} \prod_{k=1}^{j-1} A(\varepsilon(t-k))H(\varepsilon(t-j-1)). \end{cases}$$

La solution est unique si $I - EA(\varepsilon(t))$ est inversible et si l'équation $W = E[A(\varepsilon(t))WA'(\varepsilon(t))]$ n'admet pas de solution positive non triviale.

Pour la preuve, on montre tout d'abord que la stationnarité implique (a) et (b), puis que sous (a), (b) implique (c), enfin que (a) et (c) impliquent la stationnarité.

On peut remarquer que μ et W sont respectivement la moyenne et la covariance du processus $z(t)$ défini par les équations (1). On suppose par la suite que $EH(\varepsilon(t))=0$, et donc on peut prendre $\mu=0$. C'est une solution triviale de (a). Si $\mu \neq 0$, on peut toujours se ramener au cas $\mu=0$ en travaillant sur le processus $\tilde{z}(t)=z(t)-\mu$. La démonstration du théorème 1 se fait alors à partir du processus $\tilde{z}(t)$ défini par

$$\tilde{z}(t) = A(\varepsilon(t))\tilde{z}(t-1) + \xi(\varepsilon(t))$$

où

$$\xi(\varepsilon(t)) = (A(\varepsilon(t)) - E[A(\varepsilon(t))])\mu + H(\varepsilon(t)) - E[H(\varepsilon(t))]$$

est un processus centré.

La notion de minimalité que nous considérons maintenant est un peu plus faible que celle habituellement utilisée dans les modèles linéaires.

DÉFINITION 1. — On dit qu'une représentation est « quasi minimale » s'il n'existe pas d'autre représentation définie à partir du même bruit et dont le vecteur d'état soit la transformation linéaire du vecteur d'état d'origine avec une dimension plus petite.

THÉORÈME 2. — Soient $\{X(t)\}$ le processus défini par les équations (1), p la dimension de l'espace d'état, Δ et W_k définies comme dans le théorème 1, P_k la suite définie par

$$P_{k+1} = E[A(\varepsilon(t))P_k A'(\varepsilon(t))] + D$$

où

$$D = E[B'(\varepsilon(t))B(\varepsilon(t))].$$

Soit enfin la décomposition de $A(\varepsilon(t))$ suivant les polynômes d'Hermite h_j :

$$A(\varepsilon(t)) = \sum_{j=0}^r A_j h_j(\varepsilon(t)).$$

Alors la représentation (1) est quasi minimale si et seulement si

l'une des conditions équivalentes de (a) et l'une des conditions équivalentes de (b) est réalisée :

- (a) (i) W_k est non singulière pour $k \leq p$.
- (ii) La matrice C_k formée par les colonnes des matrices $\Delta, A_{i_1} \dots A_{i_n} \Delta, 1 \leq n \leq k$ est de rang p , pour un $k \leq p$.
- (b) (i) P_k est non singulière pour $k \leq p$.
- (ii) La matrice L_k formée par les lignes des matrices $D, DA_{i_1} \dots A_{i_n}, 1 \leq n \leq k$, est de rang p , pour un $k \leq p$.

Remarque. — La décomposition de $A(\cdot)$ suivant les polynômes d'Hermite est utilisée pour faciliter la démonstration, mais les résultats restent vrais pour toute décomposition de $A(\cdot)$ suivant une autre base de polynômes.

1. On démontre l'équivalence des conditions (a). La démonstration est la même pour (b). En utilisant la décomposition de $A(\cdot)$, et (c) du théorème 1 :

$$W_k = \left\{ \sum_{n=0}^k \sum_{i_1 \dots i_n} \sum_{j_1 \dots j_n} A_{i_1} \dots A_{i_n} \Delta A'_{j_1} \dots A'_{j_n} \right. \\ \left. \times E[h_{i_1}(\varepsilon(t)) \dots h_{i_n}(\varepsilon(t)) h_{j_1}(\varepsilon(t)) \dots h_{j_n}(\varepsilon(t))] \right\} \\ = \sum_{n=0}^k \sum_{i_1 \dots i_n} A_{i_1} \dots A_{i_n} \Delta A'_{i_1} \dots A'_{i_n} \sigma^{2(i_1 + \dots + i_n)}.$$

La singularité de W_k équivaut à l'existence de $v \in \mathbb{R}^p$, $v \neq 0$ tel que $v' W_k v = 0$, ce qui équivaut à $v' \Delta v = 0$, $v' A_{i_1} \dots A_{i_n} \Delta A_{i_1}' \dots A_{i_n}' v = 0$, $1 \leq n \leq k$. Les dernières relations sont équivalentes à $v' \Delta = 0$, $v' A_{i_1} \dots A_{i_n} \Delta = 0$, $1 \leq n \leq k$, i.e. l'espace vectoriel engendré par les vecteurs colonnes de Δ , $A_{i_1} \dots A_{i_n} \Delta$, $1 \leq n \leq k$, est de rang $< p$.

2. On montre maintenant que si la représentation (1) n'est pas quasi minimale, alors (a) (i) ou (b) (ii) ne sont pas vérifiées. Par hypothèse il existe une matrice S et des matrices de polynômes $\tilde{A}(\cdot)$, $\tilde{B}(\cdot)$, $\tilde{H}(\cdot)$ telles que :

$$(2) \quad \begin{cases} Sz(t) = \tilde{A}(\varepsilon(t)) Sz(t-1) + \tilde{H}(\varepsilon(t)) \\ X(t) = \tilde{B}(\varepsilon(t)) Sz(t-1) + \varepsilon(t). \end{cases}$$

Si (a) (i) n'est pas vérifiée, il n'y a rien à démontrer. On suppose que (a) (i) est vérifiée, ce qui implique que la matrice de covariance W de $z(t)$ est non singulière car $W \geq W_k$. En comparant (1) et (2), on obtient $B(\cdot) = \tilde{B}(\cdot) S$, $SA(\cdot) = \tilde{A}(\cdot) S$, $SH(\cdot) = \tilde{H}(\cdot)$, ce qui implique $SA_i = \tilde{A}_i S$. Comme S a moins de lignes que de colonnes, il existe $x \in \mathbb{R}^p$, $x \neq 0$ tel que $Sx = 0$. Par suite

$$Dx = S' \tilde{D} S x = 0, \quad DA_{i_1} \dots A_{i_n} x = S \tilde{D} \tilde{A}_{i_1} \dots \tilde{A}_{i_n} S x = 0,$$

$1 \leq n \leq k$, où $\tilde{D} = E[\tilde{B}'(\varepsilon(t)) \tilde{B}(\varepsilon(t))]$, et l'espace vectoriel engendré par les lignes de D , $DA_{i_1} \dots A_{i_n}$, $1 \leq n \leq k$ est de rang p .

3. Inversement si (a) (i) n'est pas vérifiée, alors en notant $N(W_k)$ le noyau de W_k , on a $N(W_1) \supset \dots \supset N(W_p) \neq \{0\}$. Par suite $N(W_k) = N(W_{k+1})$ pour un $k \in \{1, \dots, p-1\}$. Pour $x \in N(W_{k+1})$,

$$0 = x' \left(\sum_{i=0}^r A_i W_k A_i' + \Delta \right) x,$$

ce qui implique $x' \Delta x = 0$ et $A_i' x \in N(W_k) = N(W_{k+1})$, donc

$$x' W_{k+1} x = x' \sum_{i=0}^r (A_i W_{k+1} A_i' + \Delta) x = 0.$$

En répétant cet argument, on a $N(W_j) = N(W_{j+1})$, $\forall j \geq k$. Par suite la matrice de covariance W de $z(t)$ est singulière, et donc la représentation (1) n'est pas quasi minimale. Si (b) (i) n'est pas vérifiée, les mêmes arguments ci-dessus montrent que la matrice $P = \lim P_k$ est singulière, et que l'espace vectoriel engendré par les vecteurs colonnes de D , $A_{i_1} \dots A_{i_n}$, $n = 1, 2, \dots$, est de dimension $< p$. Soit S la matrice dont les lignes forment une base de cet espace: alors comme les vecteurs lignes de $B(\varepsilon(t))$ sont dans le sous-espace vectoriel engendré par ceux de D presque sûrement, il existe $\tilde{B}(\cdot)$, tel que $B(\cdot) = \tilde{B}(\cdot) S$. Chaque ligne de S étant une combinaison linéaire de celles de $DA_{i_1} \dots A_{i_n}$, il en est de même pour SA_i , et par suite SA_i est de la forme $\tilde{A}_i S$, ce qui termine la démonstration.

MODÈLES BILINÉAIRES DIAGONAUX : INVERSIBILITÉ. — On suppose maintenant que le processus $\{X(t)\}$ est régi par les équations :

$$(3) \quad \begin{cases} (a) & X(t) = HZ(t-1) + \varepsilon(t) \\ (b) & Z(t) = [A + B\varepsilon(t)]Z(t-1) + C\varepsilon(t) + D[\varepsilon^2(t) - \sigma^2] \end{cases}$$

où $Z(t)$ est l'état du système, $\varepsilon(t)$ est une suite de v. a. i. équidistribuées gaussiennes centrées de variance σ^2 , H , A , B , C et D respectivement des matrices ligne, carrée, carrée, colonne et colonne, avec la condition (H_1) : « $B = DH$ ». Cette condition est en particulier vérifiée pour les modèles diagonaux.

DÉFINITION 2. — Le processus $\{X(t)\}$ défini par les équations (3) est inversible p. s. (P), s'il existe une suite $(\hat{\varepsilon}(t))$ fonction des $X(s)$, $s \leq t$, telle que $\hat{\varepsilon}(t) - \varepsilon(t) \rightarrow 0$, $t \rightarrow \infty$, p. s. (P).

On utilisera dans la suite la norme opérateur définie par $\|K\|_{\varphi} = \sup_{\varphi(x) \leq 1} \varphi(Kx)$, où K est une matrice carrée, et φ une norme vecteur.

THÉORÈME 3. — *Le processus $\{X(t)\}$ défini par les équations (3) avec la condition (H_1) est inversible p. s., s'il existe φ telle que $E[\log \|K(t)\|_{\varphi}] < 0$ où $K(t) = A - CH - BX(t)$.*

Pour la démonstration on définit la suite $\hat{\varepsilon}(t)$ par $\hat{\varepsilon}(t) = X(t) - H\hat{Z}(t-1)$, $\hat{Z}(t)$ étant défini par l'équation (3) (b) où on remplace $\varepsilon(t)$ par $\hat{\varepsilon}(t)$. Alors

$$\hat{Z}(t) = K(t)\hat{Z}(t-1) + CX(t) + DX^2(t) - D\sigma^2$$

où $K(t) = A - CH - BX(t)$. Il suffit alors de montrer la convergence vers 0 de

$$\hat{Z}(t) - Z(t) = \left[\prod_{k=1}^t K(k) \right] [\hat{Z}(0) - Z(0)].$$

Or

$$\frac{1}{t} \log \prod_{k=1}^t \|K(k)\|_{\varphi} \leq \frac{1}{t} \sum_{k=1}^t \log \|K(k)\|_{\varphi},$$

ce qui converge presque sûrement par le théorème ergodique vers

$$E[\log \|K(k)\|_{\varphi}] = \alpha < 0.$$

Donc pour tout α' , $0 > \alpha' > \alpha$, il existe un nombre T assez grand tel que

$$\left\| \prod_{k=1}^t K(k) \right\|_{\varphi} \leq \exp(\alpha' t), \quad t > T,$$

avec probabilité 1, ce qui termine la démonstration.

Reçue le 8 décembre 1986.

RÉFÉRENCES BIBLIOGRAPHIQUES

- [1] C. W. J. GRANGER et A. P. ANDERSEN, *An introduction to bilinear time series analysis*, Vanderhoeck and Ruprecht, Gottingen, 1978.
- [2] D. GUEGAN, Représentation l -markovienne et existence d'une représentation affine en l'état, *Comptes rendus*, 302, série I, 1986, p. 279-282.
- [3] D. GUEGAN, Different representations of bilinear models, *J.T.S.A.* (à paraître).
- [4] T. PHAM DINH, *Bilinear markovian representation and bilinear models. Stochastic Processes and their applications*, 20, 1986.

D. G. : Équipe C.N.R.S., U.A. n° 743, Université Paris-Sud,
91405 Orsay Cedex;

T.P.D. : Laboratoire TIM 3, I.M.A.G., B.P. n° 38,
38402 Saint-Martin-d'Hères Cedex.

A NOTE ON THE ESTIMATION OF THE PARAMETERS OF THE
DIAGONAL BILINEAR MODEL BY THE LEAST SQUARES METHOD

D. GUEGAN - D.T. PHAM

(à paraître dans *Journ. Scand. of Stat. Theor. Appl.*)

I - INTRODUCTION

The bilinear model, introduced by GRANGER and ANDERSEN (1978 a) is defined by

$$(1) \quad X(t) = \sum_{i=1}^p a_i X(t-i) + \epsilon(t) + \sum_{j=1}^q c_j \epsilon(t-j) + \sum_{j=1}^Q \sum_{k=1}^P b_{jk} X(t-k) \epsilon(t-j)$$

where $\epsilon(t)$ is a sequence of i.i.d. random variables with zero mean and common variance σ^2 . This model has been investigated in many recent works. See for example, SUBBA RAO (1981), PHAM and TRAN (1981), GUEGAN (1981), HANNAN (1982), QUINN (1982), BHASKARA RAO, SUBBA RAO and WALKER (1983). Most of them treat the probabilistic aspects of the model. However, the statistical problem of the estimation of the parameters for some simple models has been considered in SUBBA RAO (1981), PHAM and TRAN (1981), GUEGAN (1981) and SUBBA RAO and GABR (1984). The purpose of this article is to extend the result of PHAM and TRAN (1981), concerning the strong consistency of the least squares estimates of the first order bilinear model, to the diagonal model of arbitrary order. For this purpose we use the Markovian representation of the diagonal and superdiagonal models ($b_{ij} = 0$ for $i > j$) established by PHAM in 1985 :

$$(2) \quad \begin{cases} X(t) = HZ(t-1) + \epsilon(t) \\ Z(t) = [A+B\epsilon(t)]Z(t-1) + C\epsilon(t) + D\epsilon^2(t) \end{cases}$$

where A and B are square matrices, H is a row matrix, C and D are column matrices. A choice of A , B , C , D and H is provided in the same paper, and in the case of the diagonal model, an explicit computation gives, with $p = \max(p, q, P, Q)$:

$$A = \begin{pmatrix} a_1 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & & & & \\ \vdots & & & & 1 \\ a_p & 0 & 0 & \dots & 0 \end{pmatrix}, \quad B = \begin{pmatrix} b_{11} & & & & \\ \vdots & & & & \\ \vdots & & & & \\ \vdots & & & & \\ b_{pp} & & & & \end{pmatrix}, \quad C = \begin{pmatrix} a_1 + c_1 \\ \vdots \\ a_p + c_p \end{pmatrix}$$

$$D = [b_{11} \dots \dots b_{pp}]', \quad H = [1 \dots \dots 0].$$

We see that $B = DH$. This condition is fundamental for our method of estimation which in fact, can be used for the model (2) as well as for the model (1). Consequently in the following, we work with the model (2) with the condition $B = DH$, which is more general than the diagonal model.

II - INVERTIBILITY

In the following, we shall restrict ourselves to the stationary processes solution of (2). A sufficient condition for the existence of such a solution can be found in PHAM (1985), namely $A^{\otimes 2} + \sigma^2 B^{\otimes 2}$ has all its eigenvalues of modulus strictly less than 1. Here $A^{\otimes 2}$ denotes the symmetric tensor product of A with itself, which can be defined as the matrix representation of the linear transformation on the space of symmetric matrices $Q \rightarrow AQA'$. Note

that the Kronecker product $A \otimes A$ corresponds also to the transformation $Q \rightarrow AQA'$, but defined on the space of arbitrary matrices; hence $A^{\otimes 2} + \sigma^2 B^{\otimes 2}$ can be viewed as the restriction of $A \otimes A + \sigma^2 B \otimes B$ on the space of symmetric matrices, and therefore the eigenvalues of the former is a subset of those of the latter.

The notion of invertibility is very useful for the estimation of the parameters of the model. Several different definitions of this notion have been proposed in the literature (see GRANGER and ANDERSEN (1978 b), PHAM and TRAN (1981), GUEGAN (1981), QUINN (1982)). The weakest form of this notion, common to all the above definitions, is the existence of a sequence $\hat{\epsilon}_n$ function of $X(1), \dots, X(n)$ only, such that $\hat{\epsilon}_n - \epsilon(n)$ converges to 0 in probability as $n \rightarrow \infty$. This is equivalent to saying that $\epsilon(t)$ is measurable with respect to the σ -algebra generated by $X(s), s \leq t$. Let B be the usual backshift operator defined by $BX(t) = X(t-1)$ and for any measurable function f of $X(t), X(t-1), \dots$,

$$Bf(X(t), X(t-1), \dots) = f(X(t-1), X(t-2), \dots).$$

(Note that $B\epsilon(t) = \epsilon(t-1)$, but $B\hat{\epsilon}_n \neq \hat{\epsilon}_{n-1}$). Then for every fixed $t > 1$ and every $n \geq t$, one has $B^{n-t}[\hat{\epsilon}_n - \epsilon(n)] = B^{n-t} \hat{\epsilon}_n - \epsilon(t)$ and since the distribution of $B^k[\hat{\epsilon}_n - \epsilon(n)]$ is the same as that of $\hat{\epsilon}_n - \epsilon(n)$ for all $k > 0$ by stationarity, the convergence in probability to zero of $\hat{\epsilon}_n - \epsilon(n)$ implies that of $B^{n-t} \hat{\epsilon}_n - \epsilon(t)$ and hence $\epsilon(t)$ being the limit of the sequence $B^{t-n} \hat{\epsilon}_n$, is $\sigma(X(s), s \leq t)$ -measurable. Conversely, if $\epsilon(t)$ is $\sigma(X(s), s \leq t)$ -measurable, then $\epsilon(0)$ is the limit of a sequence $\hat{\epsilon}_n(0)$, with $\hat{\epsilon}_n(0)$ function of $X(1-n), \dots, X(0)$ only. Define $\hat{\epsilon}_n = B^{-n} \hat{\epsilon}_n(0)$, then $\hat{\epsilon}_n$ satisfies the conditions of the first definition of the invertibility.

The above definition does not specify how one can construct the sequence $\hat{\epsilon}_n$. We shall use a stronger definition defined as follow. Considering the model (2), then if $Z(0)$ is known to have the value z_0 , the corresponding $\epsilon(t)$, denoted by $\epsilon(t|z_0)$, is the

solution of the recurrence equations :

$$(3) \quad \begin{cases} \epsilon(t|z_0) = X(t) - HZ(t-1|z_0) \\ Z(t|z_0) = [A+B\epsilon(t|z_0)]Z(t-1|z_0) + C\epsilon(t|z_0) + D\epsilon^2(t|z_0), \quad t \geq 1, \end{cases}$$

with $Z(0|z_0) = z_0$. In practice $Z(0)$ is not known, but one may still use $\epsilon(t|z_0)$ for arbitrary z_0 as an estimate of $\epsilon(t)$. We say, similarly to GRANGER and ANDERSEN (1978 b) that the model is strongly invertible if $\epsilon(t|z_0) - \epsilon(t)$ converges to 0 in probability as $t \rightarrow \infty$, for all z_0 .

For a vector norm φ , we introduce the associated operator norm defined by $\|K\|_\varphi = \sup_{\varphi(x) \leq 1} \varphi(Kx)$ for a square matrix K . We have the following theorem, the proof of which can be found in GUEGAN and PHAM (1987).

Theorem 1

Suppose that there exists a vector norm φ such that $E[\log\|K(t)\|_\varphi] < 0$ where $K(t) = A - CH - DHX(t)$, then $\hat{\epsilon}(t|z_0) - \epsilon(t) \rightarrow 0$ a.s., when $t \rightarrow \infty$, for every z_0 .

In practice the parameters A, C, D, H are unknown, so the computations in the recursion scheme (3) must be done with the aid of some hypothetical values $\tilde{A}, \tilde{C}, \tilde{D}, \tilde{H}$. The following theorem strengthens the theorem 1 and can be proved in the same manner.

Theorem 1'

Let $\tilde{A}, \tilde{C}, \tilde{D}, \tilde{H}$ be such that, for a vector norm φ :

$E[\log\|\tilde{K}(t)\|_\varphi] < 0$ where $\tilde{K}(t) = \tilde{A} - \tilde{C}\tilde{H} - \tilde{D}\tilde{H}X(t)$. Then the series

$$(4) \quad \sum_{j=0}^{\infty} [\tilde{K}(t) \dots \tilde{K}(t-j+1)] [\tilde{C}X(t-j) + \tilde{D}X^2(t-j)]$$

converges almost surely and its difference with $\tilde{Z}(t|z_0)$ given by (3) with the parameters \tilde{A} , \tilde{C} , \tilde{D} , \tilde{H} , in place of A , C , D , H tends to 0 a.s., as $t \rightarrow \infty$, for every z_0 .

Remark

The mathematical expectation used in the above theorem is relative to the true model which is not necessarily the one defined by \tilde{A} , \tilde{C} , \tilde{D} , \tilde{H} .

III - CONSISTENCY OF THE LEAST SQUARES ESTIMATORS

It is clear that the Markovian representation (2) can be unique only up to a linear transformation of the state vector, because TAT^{-1} , TC , TD , HT^{-1} correspond to the same model as the A , C , D , H for all invertible matrices T . To get around this indeterminacy of the Markovian representation, we suppose in the following that the matrices A , C , D , H depend on a more fundamental vector parameter, so that for two distinct values of θ , the corresponding matrices cannot be deduced one from the other by the above transformation. In the case of the diagonal model (1), one can take for θ , the set of parameters a_i , c_i , b_{ii} . We also suppose that the representation (2) is quasiminimal in the sense that there is no other representation with the same noise and with the state vector being a linear transformation of the original state vector and having smaller dimension (see PHAM, 1985).

Using the results of PHAM (1985), it can be seen that the representation (2) is quasiminimal if the associated linear representation : $X(t) = HZ(t-1) + \epsilon(t)$, $Z(t) = AZ(t-1) + C\epsilon(t)$ is. In the particular case of model (1) with $P = 1$ and $Q \leq \max(p,q)$, the minimality of the linear representation is satisfied if the

polynomials $1 - \sum_j a_j z^j$ and $1 - \sum_j c_j z^j$ have no roots in the unity

circle and no common roots. This provides a sufficient condition for the minimality of the representation (2) for this subclass of bilinear models.

Recall that the standard method of least squares consists of minimizing the sum of squares of the errors on some given set of parameters containing the true value of θ . We suppose that the true model is stationary and fulfills the invertibility condition of the theorem 1'. The last condition is crucial in order that the error be correctly estimated for large t . However this condition cannot be easily verified. Therefore, we shall consider a random set $\Theta_{n,\delta}$ depending of the observations, defined by :

$$(5) \quad \Theta_{n,\delta} = \left\{ \tilde{\theta} \in \Theta_0 : \prod_{t=1}^n \|\tilde{A} - \tilde{C}\tilde{H} - \tilde{D}\tilde{H} X(t)\|_{\varphi} \leq (1-\delta)^n \right\}$$

where Θ_0 is a given compact set, n is the sample size and where δ is a (small) positive number. The introduction of δ permits the use of the sign \leq instead of $<$, leading to a closed, hence compact set. $\Theta_{n,\delta}$. We suppose further that Θ_0 contains the true value θ (this is a reasonable assumption if Θ_0 has been chosen large enough, but still satisfying (5)). Stationarity considerations may be helpful for this purpose since we know that the eigenvalues of $A^{\otimes 2} + \sigma^2 B^{\otimes 2}$ must lie inside the unit circle of the complex plane. However, the above condition involves the unknown values of σ^2 and hence one would need at least a lower bound for it.

For example in case $p = P = Q = 1, q = 0$, the stationarity condition is $a_1^2 + \sigma^2 b_{11}^2 \leq 1$. Thus one may take

$$\Theta_0 = \{(a_1, b_{11}) : a_1^2 + b_{11}^2 \sigma_{\min}^2 < 1\},$$

where σ_{\min}^2 is some lower bound of σ^2 . Alternatively, one may take

$$\Theta_0 = \{(a_1, b_{11}) : |a_1| \leq 1, |b_{11}| \leq M\}$$

where M is a large number. The condition $|b_{11}| \leq M$, however would be ineffective if M is very large since we have from the definition of $\Theta_{n,\delta} : |b_{11}| \leq \left| \prod_{t=1}^n X(t) \right|^{1/n(1-\delta)}$. Thus one may take

$$\Theta_0 = \{(a_1, b_{11}) : |a_1| \leq 1\}$$

in this special case.

For $p = q = P = Q = 1$, one may take :

$$\Theta_0 = \{(a_1, b_{11}, c_1) : a_1^2 + b_{11}^2 \sigma_{\min}^2 \leq 1, |c_1| \leq M\}$$

where M is a large number. The condition $|c_1| \leq M$ might again be dropped because of the definition of $\Theta_{n,\delta}$.

In case $p = q = P = Q = 2$, the stationarity condition is that the eigenvalues of the polynomial

$$P(\lambda) = -\lambda^3 + \lambda^2(a_1^2 + a_2^2 + \sigma^2 b_{11}^2) + \lambda(a_1^2 a_2^2 + a_2^2 + 2\sigma^2 a_1 b_{11} b_{22} - \sigma^2 a_2 b_{11}^2 + \sigma^2 b_{22}^2) - a_2(a_2^2 + \sigma^2 b_{22}^2)$$

have moduli strictly less than 1. The above conditions can be expressed by the inequalities (see for example BARNDORFF-NIELSEN and SCHOU, 1973) :

$$\begin{aligned} \alpha_1 + \alpha_2 + \alpha_3 &< 1 \\ -\alpha_1 + \alpha_2 - \alpha_3 &< 1 \\ \alpha_3(\alpha_3 - \alpha_1) - \alpha_2 &< 1 \\ |\alpha_3| &< 1 \end{aligned}$$

where

$$\begin{aligned}\alpha_1 &= a_1^2 + a_2 + \sigma^2 b_{11}^2 \\ \alpha_2 &= a_1^2 a_2 + a_2^2 + 2\sigma^2 a_1 b_{11} b_{22} - \sigma^2 a_2 b_{11}^2 + \sigma^2 b_{22}^2 \\ \alpha_3 &= -(a_2^3 + \sigma^2 a_2 b_{22}^2).\end{aligned}$$

However these expressions are so complicated that we do not know how to explicit them judiciously .

The condition $\theta \in \Theta_0$ is only necessary but not sufficient for the stationarity of the model. Therefore the estimator may fail to correspond to a stationary model, but, since it is strongly consistent, this will not happen for sample sizes large enough. From the ergodic theorem, we know that :

$$\limsup_{n \rightarrow \infty} \Theta_{n, \delta} \subset \Theta_{\delta/2} \quad \text{where}$$

$$(6) \quad \Theta_{\delta/2} = \left\{ \tilde{\theta} \in \Theta_0, E \left[\log \| \tilde{A} - \tilde{C} \tilde{H} - \tilde{D} \tilde{H} X(t) \|_{\varphi} \right] \leq \log \left(1 - \frac{\delta}{2} \right) \right\}$$

In the following, we suppose that $\Theta_{\delta/2}$ is compact. In the same way, we can show that the true value θ of the parameter is in the set $\liminf_{n \rightarrow \infty} \Theta_{n, \delta}$ if :

$$E \left[\log \| A - CH - DH X(t) \|_{\varphi} \right] < \log(1 - \delta)$$

Here the limit sup of a sequence of sets A_n is defined as :

$$\limsup_{n \rightarrow \infty} A_n = \bigcap_{p=0}^{\infty} \left[\bigcup_{n \geq p} A_n \right]$$

and similarly for the limit inf.

The strong consistency of the least squares estimate $\hat{\theta}_n$ is based on the following lemma, established by PHAM and TRAN (1981).

Lemma 1

Let $Q_n(\tilde{\theta})$ be a sequence of continuous random functions defined on a compact subset $\tilde{\Theta}$ of a k -dimensional Euclidean space \mathbb{R}^k , θ be a point of $\tilde{\Theta}$, and Θ_n be a sequence of random compact subsets of \mathbb{R}^k , such that :

- (i) with probability 1, $\theta \in \Theta_n \subset \tilde{\Theta}$ for n sufficiently large,
- (ii) for any $\theta' \in \tilde{\Theta}$, $\theta' \neq \theta$, there exists a neighborhood $U(\theta')$ of θ' such that

$$\liminf_{n \rightarrow \infty} \left\{ \inf_{\tilde{\theta} \in U(\theta')} Q_n(\tilde{\theta}) - Q_n(\theta) \right\} > 0 \quad \text{a.s.}$$

then $\hat{\theta}_n$, the minimizer of Q_n on Θ_n , converges almost surely to θ .

We now apply the above lemma to the function :

$$Q_n(\tilde{\theta}) = \frac{1}{n} \sum_{t=1}^n \epsilon_{\tilde{\theta}}^2(t|z_0)$$

where $\epsilon_{\tilde{\theta}}(t|z_0)$ are given by (3) with $\tilde{\theta}$ as parameter.

Theorem 2

Suppose that the process $\{X(t)\}$ defined by the equations (2) with the condition $B = DH$, is quasiminimal, and $\epsilon(t)$ has a continuous distribution. Assume further that A, B, C, D, H satisfy the conditions stated at the beginning of section III and that $\Theta_{\delta/2}$ is compact. Let $Q_n(\tilde{\theta})$ be defined as above. Then $\hat{\theta}$, the value of $\tilde{\theta}$ which minimizes Q_n over the set $\Theta_{n,\delta}$ defined by (5), is a strongly consistent estimate of $\tilde{\theta}$.

Proof

We need to verify the conditions of the lemma 1. This is

clear for the condition (i). On the other hand, by the same argument as in the proof of theorem 1', the mapping $\tilde{\theta} \rightarrow Z_{\tilde{\theta}}(t)$ where $Z_{\tilde{\theta}}(t)$ is defined by the series (4) is continuous, and the convergence of $\tilde{Z}(t|z_0) - Z_{\tilde{\theta}}(t)$ to 0, $\tilde{Z}(t|z_0)$ being as in the theorem 1', is uniform on $\Theta_{\delta/2}$. The condition (ii) can then be shown to be satisfied by the same arguments as un PHAM and TRAN (1981), provided that, for all $\tilde{\theta} \neq \theta$, $E\epsilon_{\tilde{\theta}}^2(t) > \sigma^2$, (we put $E\epsilon_{\tilde{\theta}}^2(t) = \infty$ if it does not exist), where $\epsilon_{\tilde{\theta}}(t) = X(t) - \tilde{H}Z_{\tilde{\theta}}(t)$. Since $\epsilon_{\tilde{\theta}}(t) = \epsilon(t) + HZ(t-1) - \tilde{H}Z_{\tilde{\theta}}(t-1)$, the last condition is satisfied if $\tilde{H}Z_{\tilde{\theta}}(t) \neq HZ(t)$ for all $\tilde{\theta} \neq \theta$. This property holds for quasiminimal process $X(t)$, defined by (2) with the parametrization satisfying the assumption at the beginning of the section 3, as is shown in the following lemma.

Lemma 2

If the process $\{X(t)\}$, defined by the equations (2) with the condition $B = DH$, is quasiminimal and if $\epsilon(t)$ has a continuous distribution, then $\tilde{H}Z_{\tilde{\theta}}(t) = HZ(t)$ if and only if there exists an invertible transformation T such that $\tilde{A} = TAT^{-1}$, $\tilde{C} = TC$, $\tilde{D} = TD$, $\tilde{H} = HT^{-1}$.

Proof

Note that $\tilde{H}Z_{\tilde{\theta}}(t) = HZ(t)$ holds for some t , then it also holds for all t by stationarity. On the other hand, we have

$$Z(t) = [A - CH - DHX(t)]Z(t-1) + CX(t) + DX^2(t),$$

and the same expression for $Z_{\tilde{\theta}}(t)$ with $\tilde{A}, \tilde{C}, \tilde{D}, \tilde{H}$ in place of A, C, D, H . Thus the condition $\tilde{H}Z_{\tilde{\theta}}(t) = HZ(t)$ implies

$$\begin{aligned} \tilde{H}\tilde{A}Z_{\tilde{\theta}}(t-1) - HAZ(t-1) - (\tilde{H}\tilde{C} - HC)HZ(t-1) - (\tilde{H}\tilde{D} - HD)X(t)HZ(t-1) \\ + (\tilde{H}\tilde{C} - HC)X(t) + (\tilde{H}\tilde{D} - HD)X^2(t) = 0, \quad \text{a.s.} \end{aligned}$$

Replacing $X(t)$ by $HZ(t-1) + \epsilon(t)$, the above expression reduces to:

$$\tilde{H}\tilde{A}\tilde{Z}_{\theta}(t-1) - HAZ(t-1) + (\tilde{H}\tilde{C} - HC)\epsilon(t) + (\tilde{H}\tilde{D} - HD)[HZ(t-1) + \epsilon(t)]\epsilon(t) = 0, \text{ a.s.}$$

The above left hand side is a polynomial of second degree in $\epsilon(t)$: $\alpha(t-1)\epsilon^2(t) + \beta(t-1)\epsilon(t) + \gamma(t-1)$ with coefficients $\alpha(t-1)$, $\beta(t-1)$, $\gamma(t-1)$ being functions of $X(s)$, $s \leq t-1$ only. Hence $\alpha(t-1)$, $\beta(t-1)$, $\gamma(t-1)$ are independent of $\epsilon(t)$. Thus:

$$P\{\alpha(t-1)\epsilon^2(t) + \beta(t-1)\epsilon(t) + \gamma(t-1) \neq 0\} = \int P\{\alpha\epsilon(t)^2 + \beta\epsilon(t) + \gamma \neq 0\} d\tilde{P}(\alpha, \beta, \gamma)$$

where \tilde{P} is the joint distribution of $\alpha(t-1)$, $\beta(t-1)$, $\gamma(t-1)$. But the left hand side of the above equality is zero by assumption, implying

$$P\{\alpha\epsilon(t)^2 + \beta\epsilon(t) + \gamma \neq 0\} = 0$$

for all α, β, γ outside a set of zero \tilde{P} -measure. Now $\epsilon(t)$ having a continuous distribution:

$$P\{\alpha\epsilon(t)^2 + \beta\epsilon(t) + \gamma \neq 0\} = 1 \text{ except when } \alpha = \beta = \gamma = 0.$$

Hence $\alpha(t-1) = \beta(t-1) = \gamma(t-1) = 0$ with probability one. Thus:

$$\begin{cases} \tilde{H}\tilde{D} = HD \\ \tilde{H}\tilde{C} = HC \\ \tilde{H}\tilde{A}\tilde{Z}_{\theta}(t) - HAZ(t) = 0 \end{cases}$$

Applying repeatedly the above argument, we get

$$(7) \quad \begin{cases} (a) & \tilde{H}\tilde{A}^k\tilde{D} = HAD & k \geq 0 \\ (b) & \tilde{H}\tilde{A}^k\tilde{C} = HAC & k \geq 0 \\ (c) & \tilde{H}\tilde{A}^k\tilde{Z}_{\theta}(t) - HA^kZ(t) = 0 \end{cases}$$

Considering now the expansion of $Z(t)$ from the expression (2), we have :

$$Z(t) = \sum_{j=0}^{\infty} \prod_{k=0}^{j-1} [A+B\epsilon(t)] \dots [A+B\epsilon(t-j+1)] [C\epsilon(t-j)+D\epsilon^2(t-j)]$$

$$X(t) = HZ(t-1) + \epsilon(t) .$$

By expanding the product in the above right hand side, we see that $X(t)$ can be expressed as a series of monomial in $\epsilon(t), \dots, \epsilon(t-j)$ or in $\epsilon(t), \dots, \epsilon(t-j+1), \epsilon^2(t-j)$ with coefficients of the form

$$HA^{n_1} B^{m_1} \dots A^{n_r} B^{m_r} C \quad \text{or} \quad HA^{n_1} B^{m_1} \dots A^{n_r} B^{m_r} D, \quad n_i \geq 0, m_i \geq 0, \text{ and}$$

$\sum_i (n_i + m_i) = j$. We begin with the first above product. Note that $m_1 > 0$, since otherwise we may group the factor A^{n_1} with A^{n_2} . Thus by (7.a), we have :

$$HA^{n_1} B^{m_1} = HA^{n_1} DHB^{m_1-1} = \tilde{H}A^{n_1} \tilde{D}HB^{m_1-1} .$$

If $m_1 > 1$, we write, by (7.a) :

$$HB^{m_1-1} = H(DH)^{m_1-1} = (HD)^{m_1-1} H = \tilde{H}(\tilde{H}D)^{m_1-1} H .$$

Repeating the same argument, we get finally

$$HA^{n_1} B^{m_1} \dots A^{n_r} B^{m_r} C = \tilde{H}A^{n_1} \tilde{B}^{m_1} \dots \tilde{A}^{n_r} \tilde{B}^{m_r} \tilde{C} ,$$

using (7.b) in the last step. In the same way,

$$HA^{n_1} B^{m_1} \dots A^{n_r} B^{m_r} D = \tilde{H}A^{n_1} \tilde{B}^{m_1} \dots \tilde{A}^{n_r} \tilde{B}^{m_r} \tilde{D} .$$

It can now be seen that, if we put

$$\tilde{Z}(t) = \sum_{j=0}^{\infty} [\tilde{A} + \tilde{B}\epsilon(t)] \dots [\tilde{A} + \tilde{B}\epsilon(t-j+1)] [\tilde{C}\epsilon(t-j) + \tilde{D}\epsilon^2(t-j)]$$

which should not be confounded with $Z_{\theta}(t)$, then we have also

$X(t) = \tilde{H}\tilde{Z}(t-1) + \epsilon(t)$. This shows that the process admits a second Markovian representation :

$$X(t) = \tilde{H}\tilde{Z}(t-1) + \epsilon(t)$$

$$\tilde{Z}(t) = [\tilde{A} + \tilde{B}\epsilon(t)]\tilde{Z}(t-1) + \tilde{C}\epsilon(t) + \tilde{D}\epsilon^2(t)$$

Since the Markovian representation of $X(t)$ is quasiminimal, it is unique up to a linear invertible transformation of the state vector, implying $\tilde{Z}(t) = TZ(t)$ for some invertible matrix T . Hence $\tilde{A} = TAT^{-1}$, $\tilde{C} = TC$, $\tilde{D} = TD$, $\tilde{H} = HT^{-1}$, giving the result.

Conversely, if $\tilde{A} = TAT^{-1}$, $\tilde{C} = TC$, $\tilde{D} = TD$, $\tilde{H} = HT^{-1}$, then

$$\begin{aligned} Z_{\theta}(t) &= \sum_{j=0}^{\infty} \tilde{K}(t) \dots \tilde{K}(t-j+1) [\tilde{C}X(t-j) + \tilde{D}X^2(t-j)] \\ &= \sum_{j=0}^{\infty} K(t) \dots K(t-j+1) [CX(t-j) + DX^2(t-j)] \end{aligned}$$

and the last series equals to $Z(t)$ as shown in the proof of the theorem 1.

NOTE

As remarked by a referee, the invertibility condition of the theorem 1 can be relaxed to the condition that the upper Lyapounov exponent of the matrix product :

$$K(n)K(n-1)\dots K(1), \quad n = 1, 2, \dots$$

is strictly negative. For a stationary ergodic sequence of random real matrices $K(1), K(2), \dots$, the upper Lyapounov exponent may be defined as :

$$\lim_{n \rightarrow \infty} \frac{1}{2n} \log[\text{largest eigenvalue of } K(n)^* \dots K(1)^* K(1) \dots K(n)]$$

(where A^* represents the transpose matrix A), which is a non random number (see for example COHEN et al. p.24, 1986). We may then modify the definition of $\Theta_{n,\delta}$ in (5) by the following :

$$\{\tilde{\theta} \in \Theta_0 : \|[\tilde{A}-\tilde{C}\tilde{H}-\tilde{D}\tilde{H}X(n)] \dots [\tilde{A}-\tilde{C}\tilde{H}-\tilde{D}\tilde{H}X(1)]\| \leq (1-\delta)^n\}$$

where $\|\cdot\|$ denotes any operator norm.

BIBLIOGRAPHY

- BARNDORFF-NIELSEN, O., SCHOU, G. (1973). On the Parametrization of Autoregressive Models by partial correlations. J. Multivariate Anal. 3. 408-419.
- BHASKARA RAO, H., SUBBA RAO, T., WALKER, A.M. (1983). On the existence of some bilinear time series models. J. Time Ser. Anal. 4. n° 2. 95-110.
- COHEN, J.E., KESTEN, H., NEWMAN, C.H. (1986). Random matrices and their applications. Contemporary Mathematics. Vol. 50. A.M.S. Providence. Rhode Island.
- GRANGER, C.W.J., ANDERSEN, A.P. (1978,a). An introduction to bilinear time series models. Vanderhoeck and Ruprecht. Gottingen.

- GRANGER, C.W.J., ANDERSEN, A.P. (1978,b). On the invertibility of time series models. *Stochastic Process. Appl.* 8. 87-92.
- GUEGAN, D. (1981). Etude d'un modèle non linéaire, le modèle super-diagonal d'ordre 1. *C.R. Acad. Sci. Paris, Serie I. Math.* 293. 95-99.
- GUEGAN, D., PHAM D.T. (1987). Minimalité et Inversibilité des modèles bilinéaires à temps discret. *C.R. Acad. Sci. Serie I. Math.* 448. 159-162.
- HANNAN, E.J. (1982). A note on bilinear time series models. *Stochastic Process. Appl.* 12. 221-224.
- PHAM, D.T. (1985). Bilinear Markovian representation and bilinear models. *Stochastic Process. Appl.* 20. 295-306.
- PHAM, D.T., TRAN T.L. (1981). On the first order bilinear time series models. *J. Appl. Prob.* 18. 617-627.
- QUINN, B.G. (1982). Stationarity and invertibility of simple bilinear models. *Stochastic Process. Appl.* 12. 225-230.
- SUBBA RAO, T. (1981). On the theory of bilinear time series models. *J. Roy. Statist. Soc. Ser. B* 43. 224-255.
- SUBBA RAO, T., GABR, M.M. (1984). An introduction to bispectral analysis an bilinear time series models. *Lecture Notes in Statistics.* 24. Springer-Verlag.

RESUME DE LA THESE :

Cette thèse présente l'étude probabiliste et statistique approfondie des modèles bilinéaires à temps discret. On étudie ces modèles à partir de différentes approches (directe, Markovienne). On trouve tout d'abord une présentation globale des modèles non linéaires, la description des outils probabilistes utiles à l'étude des modèles non linéaires, ainsi qu'une présentation des modèles bilinéaires à partir de simulations permettant de mettre en évidence leurs principales caractéristiques trajectorielles.

L'approche Markovienne s'avère beaucoup plus puissante que l'approche directe. Nous démontrons l'existence d'une représentation Markovienne sous la forme d'un modèle polynomial affine en l'état ; nous donnons des critères pour la minimalité et l'inversibilité de ces représentations.

Sur le plan statistique nous avons montré la convergence presque sûre des estimateurs des moindres carrés. D'autres estimateurs sont aussi envisagés permettant de mettre en place des tests d'adéquation de modèles.

Certains travaux de l'auteurs (huit articles) ont été publiés et sont regroupés dans l'annexe.

Mots clefs : 6.3

Ergodicité (Géométrique), Estimateur, Inversibilité, Minimalité, Modèles Bilineaires, Modèles Polynomiaux, Représentation Markovienne, Stationnarité, Test (d'adéquation, de linéarité).