

وزارة التعليم و البحث العلمي  
MINISTERE DE L'ENSEIGNEMENT ET DE LA RECHERCHE SCIENTIFIQUE

ECOLE NATIONALE POLYTECHNIQUE

DEPARTEMENT : ELECTRONIQUE

المدرسة الوطنية المتعددة التقنيات  
BIBLIOTHEQUE — المكتبة  
Ecole Nationale Polytechnique

PROJET DE FIN D'ETUDES

SUJET

ANALYSE DES DONNEES  
PAR LA METHODE DE  
BOX ET JENKINS

Proposé par :  
A. MAAFI

Etudié par :  
D. BELGROUNE  
N. BENSALAM

Dirigé par :  
A. MAAFI

PROMOTION :  
JANVIER 1986

DEDICACES

المدرسة الوطنية المتعددة التقنيات  
المكتبة — BIBLIOTHEQUE  
Ecole Nationale Polytechnique

A mon père

A ma mère

A mes frères et soeurs

A mes amis (es)

D . BELGROUNE

A mon père

A ma mère

A mes frères et soeurs

A mes amis (es)

N . BENSALAM .

REMERCIEMENTS

المدرسة الوطنية المتعددة التقنيات  
المكتبة — BIBLIOTHEQUE  
Ecole Nationale Polytechnique

Nous adressons nos plus vifs remerciements à notre promoteur ;  
Monsieur A. MAAFI , pour toute l'aide qu'il nous a apportée  
pour l'élaboration de ce travail et nous avoir guidé inlassable-  
ment durant toute sa durée .

Nous voulons que Monsieur K. BACHAR professeur au C.E.R.I  
trouve ici l'expression de nos remerciements les plus chaleureux  
et notre sincère reconnaissance pour ses conseils et sa collabo-  
ration pendant la réalisation de ce projet .

Nous tenons à remercier également , Monsieur Louni , secrétaire  
général du C.E.R.I , pour nous avoir bien accueilli dans son  
centre .

Que tous les professeurs qui ont contribué à notre formation,  
veuillent bien trouver ici l'expression de notre profonde gra-  
titude .

D . BELGROUNE

N. BENSALÉM

TABLE DES MATIERES

INTRODUCTION .

CHAPITRE I- MODELISATION DES SERIES TEMPORELLES

I - INTRODUCTION

II- LES DIFFERENTES CLASSES DE MODELES

II.1- Les modèles explicatifs

II.1.1 - Les modèles explicatifs ~~statiques~~ statiques

II.1.2- Les modèles explicatifs dynamiques

II.2 - Les modèles non explicatifs

II.2.1 - Les modèles d'ajustements

2.1.1 - L'ajustement global

2.1.2 - L'ajustement local

2.1.3 - L'ajustement additifs

2.1.4 - L'ajustement multiplicatif

2.1.5 - L'ajustement déterministe et L'ajustement aléatoire.

II.2.2 - Les modèles auto-projectifs

III - TECHNIQUE DE CONCEPTION D'UN MODELE DANS LE CAS  
DES SERIES TEMPORELLES .

III.1. - Identification

III.2. - Estimation

III.3. - Test

III.4. - Utilisation.

IV - CRITERE DE SELECTION D'UN MODELE

V - CONCLUSION

CHAPITRE II - LES MODELES UTILISES DANS L'APPROCHE  
DE BOX ET JENKINS .

I - INTRODUCTION

II - LES EQUATIONS LINEAIRES AUX DIFFERENCES

### III - LES PROCESSUS STOCHASTIQUES

III.1 - L'autocovariance

III.2 - L'autocorrélation

III.3 - Les processus stationnaires du second ordre

### IV - LES MODELES UTILISES DANS L'APPROCHE DE BOX ET JENKINS

IV.1. - Les modèles autorégressif ou AR(P)

IV.1.1 - La fonction d'autocorrélation

IV.1.2 - La fonction d'autocorrélation partielle .

IV.2. - Les modèles à moyenne mobile ou MA(q).

IV.2.1 - La fonction d'autocorrélation

IV.3 - Les modèles mixtes ou ARMA (P,q ).

IV.3.1 - La fonction d'autocorrélation .

V - LES MODELES ARIMA ( p,d,q)

VI - DISCUSSION .

### CHAPITRE III - LA PROCEDURE DE BOX ET JENKINS

I - INTRODUCTION

II - LE PRINCIPE DE LA METHODE

III - ETUDE DES DIFFERENTES PHASES

III.1. - Le traitement préalable des données .

III.2. - La phase d'identification

III.2.1. - détermination de l'ordre d de l'opérateur de différence

III.2.2 - Détermination des degrés p et q

2.2.1. - Identification d'un modèle ARMA(p,q)

2.2.2 - Identification d'un modèle SARIMA(p,d,q)(P.D.Q)

III.2.3. - Détermination des valeurs initiales des paramètres  $\theta$  et  $\phi$  intervenant dans le modèle .

III.3. - La phase d'estimation .

III.3.1 - L'estimation des paramètres par le maximum de vraisemblance .

- 3.1.1. - La fonction de vraisemblance pour un processus MA( $q$ ).
- 3.1.2. - Application de la méthode du maximum de vraisemblance à l'estimation des paramètres d'un processus ARMA( $p, q$ ) .

III.3.2. - L'estimation des paramètres par la méthode de Back forecasting .

III.4. - La phase de vérification

III.4.1. - test sur les paramètres estimés

III.4.2. - l'Analyse des résidus

III.4.3. - le choix d'un modèle .

III.5.. - utilisation .

#### CHAPITRE . IV - *organigramme des traitements .*

I - INTRODUCTION

II - PRESENTATION DES DIFFERENTES ETAPES

II.1. - l'étape d'identification

II.1.1 - Les entrées

II.1.2. - Le traitement

II.1.3.- Les sorties

II.2.- l'étape d'estimation

II.2.1 - Les entrées

II.2.2 - le traitement

II.2.3.- les sorties

II.3 - l'étape de verification

II.3.1 - les entrées

II.3.2 - le traitement

II.3.3. Les sorties

II.4. - Organigramme de l'étape (algorithmique) d'identification de la méthodologie .

## CHAPITRE. V - APPLICATION

### I - INTRODUCTION

### II - les processus autoregressif d'ordre

II.1 - les processus autoregressif d'ordre 1

II.2 - les processus autogressif d'ordre 2

II.3 - les processus moyenne mobiles d'ordre 1

II.4 - les processus moyennes mobiles d'ordre 2

II.5 . - les processus mixtes d'ordre 1

### III. - PRESENTATION D'UNE DISCUSSION D'UN MODELE STATIQUE EN METEOROLOGIE SOLAIRE

III.1 - Description du modèle utilisé

III.1.1 - Nature des données

III.1.2 - Recherche du modèle

III.1.3 - Determination des paramètres

III.2. - utilisation du modèle .

Résumé : présentation sommaire de la méthodologie de MM. BOX et JENKINS dans l'étude des séries chronologiques scalaires discrètes.

Mots-clés: Processus stochastiques ; processus stationnaires ; opérateurs avance et retard ; séries chronologiques ; moyenne mobile ; identification de processus ; vérification ; fonction d'auto-corrélation, etc...

---

AVANT - PROPOS  
=====

المدرسة الوطنية المتعددة التقنيات  
المكتبة — BIBLIOTHEQUE  
Ecole Nationale Polytechnique

L'objectif de ce travail consiste dans l'étude d'une méthode, relativement récente, traitant la modélisation, l'estimation et la prédiction d'un signal (ou autre phénomène) pouvant être chiffré.

Le formalisme mathématique utilisé dans notre document ne doit pas donner à penser qu'il s'agit d'un traité purement mathématique. Les formules relativement simple utilisées, sont indispensables pour fournir au PRATICIEN les idées et les techniques nécessaires à son élaboration concrète.

Les signaux (chiffrés) sont supposés représentables par une suite d'observations numériques ordonnées dans l'espace des temps. Nous nous limiterons dans notre travail, aux suites (ou séries) scalaires discrètes (il est intuitivement satisfaisant de penser à "discrétiser" sur un phénomène continu).

La méthode que nous tentons de présenter (méthode de BOX-JENKINS [8]) est basée sur l'idée que la plupart des séries peuvent être considérées comme des trajectoires de processus stochastiques ; les modèles associés, qui ne sont plus des modèles explicatifs (au sens de la causalité), sont dits des modèles AUTO-PROJECTIFS, c'est-à-dire des modèles ALEATOIRES NON EXPLICATIFS, où l'on suppose que le future se déduit tout naturellement (‡) du passé du phénomène à projeter éventuellement.

Ces modèles sont connus sous le nom "AUTO-REGRESSIFS MOVING AVERAGE" (ou ARMA, dans la littérature anglo-saxonne). Ils correspondent aux processus stochastiques STATIONNAIRES (Croissance différentielle du processus).

Ces modèles "ARMA" font dépendre la valeur de la variable (série) d'une d'une combinaison linéaire d'observations passées et d'une combinaison linéaire d'écart - résiduels passées (qui doivent s'interpréter comme un processus type "bruit blanc").

Notre plan sera le suivant.

après une étude relativement détaillée des modèles "ARMA" et les problèmes statistiques qui en découlent, nous exposerons l'application à des cas réels.

Nous compléterons notre document par une conclusion portant sur les limites et les extensions possibles de la méthode et nous donnons enfin, une bibliographie non exhaustive.

# CHAPITRE I

## MODELISATION DES SERIES TEMPORELLES

## I INTRODUCTION

Les séries temporelles ou séries chronologiques ou encore chroniques correspondent à des séries ponctuelles, datées et ordonnées dans le temps / 1 / .

Ces séries peuvent représenter des mesures de grandeur physiques telle que le rayonnement solaire, la température ambiante etc... ou de manière générale des signaux discrétisés évoluant en fonction du temps.

A l'état brute, elles ne donnent pas une information précise sur le phénomène qui les a générés. Cependant, elles constituent un matériau de base pour décrire ce phénomène, et permettent de déduire les lois qui le régissent.

La représentation des données relatives à un phénomène physique sous forme d'équations mathématiques s'appelle modélisation.

La question que nous voulons aborder dans ce chapitre est celle concernant la modélisation des séries temporelles .

### II- Les différentes classes de modèles :

Il existe une grande variété de modèles permettant de modéliser les séries temporelles. Pour chaque phénomène et pour chaque type de données (longueur de la série ...etc), et surtout suivant l'usage que l'on veut faire du modèle, on sera amené à fixer notre choix sur le modèle qu'il convient d'utiliser.

Il convient de distinguer deux grandes classes de modèle permettant de modéliser les séries temporelles:

.../...

la classe des modèles explicatifs et la classe des modèles non explicatifs / 2 / .

La classe des modèles explicatifs contient les modèles explicatifs statiques et les modèles explicatifs dynamiques .

La classe des modèles non explicatifs est divisée en deux sous-classes :

les modèles d'ajustements et les modèles autoprojectifs fig(1).

De façon générale, ces modèles sont employés dans des domaines très divers tels que la physique, l'automatique, l'électronique, la météorologie, l'économétrie ...etc / 3,4,5,6 / .

## II.1- Les modèles explicatifs .

Dans cette catégories de modèles la variable  $Z_i(t)$  est exprimée en fonction d'un vecteur de variables observables, dites exogènes  $Z_j(t)$  et d'une variable aléatoire centrée  $a_t$  .

Les relations qui lient les variables  $Z_i(t)$  aux variables  $Z_j(t)$  s'écrivent sous la forme .

$$Z_i(t) = F (Z_j(t) / j \in I, a_t) . \quad (1.I)$$

où  $F$  est une fonction des variables  $Z_j(t)$  et de  $a_t$  et  $Z_i(t)$  représente un ensemble de variables aléatoires dépendant du temps.

Notons que  $i, j$  appartiennent à l'ensemble  $[1, 2, \dots, k]$  et  $t$  appartient à l'ensemble  $[1, \dots, T]$ .

Les variables aléatoires  $Z_j(t)$  peuvent être des variables exogènes ou des variables endogènes.

Les modèles explicatifs sont de deux types :  
Les modèles explicatifs statiques et dynamiques. / 2 / .

.../...

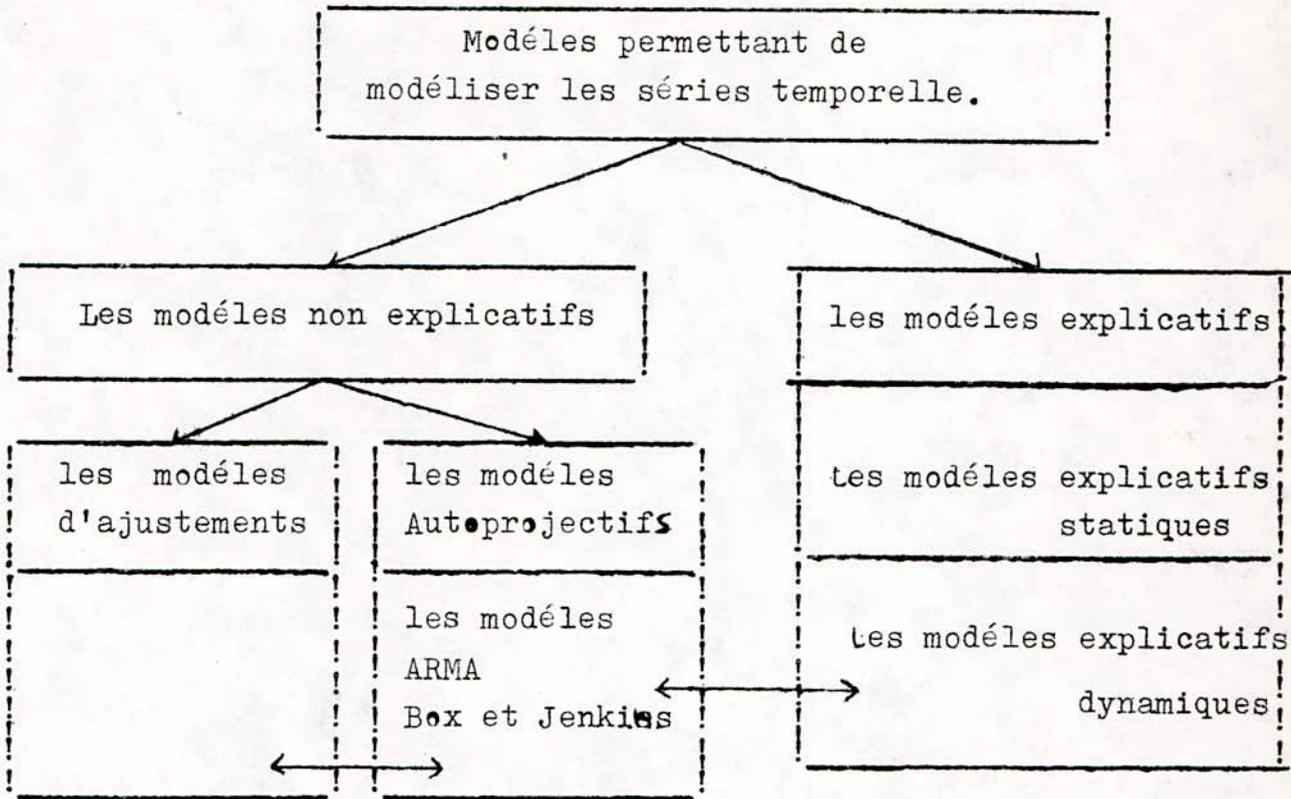


fig ( 1 ) .

.1.1- Les modèles explicatifs statiques :

Dans ces modèles, les variables  $Z_j(t)$  dites exogènes ne contiennent pas des valeurs prises dans le passé de  $Z_i(t)$  et les  $a_t$  sont indépendants entre eux. / 2 / .

1.2. Les modèles explicatifs dynamiques :

Un modèle explicatif peut être dynamique soit parce que les  $a_t$  sont autocorrélés, soit parce que l'ensemble de  $Z_j(t)$  contient des valeurs prises dans le passé de  $Z_i(t)$ , c'est à dire les variables sont endogènes / 2 / .

II-2- Les modèles non-explicatifs :

II.2.1- Les modèles d'ajustements :

Ces modèles sont universellement employés. Dans le cas général, ils sont représentés par la relation suivante :

$$Z_i(t) = f(t, a_t) \quad (2, I)$$

où  $Z_i(t)$  est une fonction du temps  $t$  et d'une variables aléatoire centrée  $a_t$ . Elle est indexée par un nombre fini de paramètres inconnus. / 2 /

2.1.1. L'ajustement global .

Dans ce cas la fonction  $f$  définie par la relation(2.I) est déterminée à partir de l'ensemble des données disponibles.

De façon générale  $f(t, a_t)$  est représentée par une forme polynomiale en  $t$  perturbée par un bruit blanc  $a_t$  .

2.1.2- L'ajustement local .

.../...

Dans ce type d'ajustement,  $f$  sera estimée uniquement au voisinage du présent, cela veut dire au voisinage de  $T$ , si l'ensemble des observations est connu dans l'intervalle de temps  $(1, T)$ .

.Donc, on privilège les informations les plus récentes au détriment des plus anciennes / 7 /  
L'exemple le plus connu de ce type d'ajustement est celui du lissage exponentiel simple .

### 2.1.3- L'ajustement additif .

. La fonction d'ajustement  $f(t, a_t)$  est additive lorsqu'elle peut se décomposer de la façon suivante :

$$f(t, a_t) = g(t) + a_t \quad (3.I)$$

où  $g(t)$  est une fonction du temps inobservable mais déterministe / 2 /

### 2.1.4- L'ajustement multiplicatif .

On dit que le modèle d'ajustement est multiplicatif si la fonction  $f(t, a_t)$  peut s'écrire sous la forme du produit :

$$f(t, a_t) = g(t) \cdot a_t \quad (4.I)$$

notons que lorsque les diverses variables sont positives, le passage d'un modèle multiplicatif à un modèle additif se fait à l'aide de la fonction logarithmique / 2 /

### 2.1.5 L'ajustement déterministe et l'ajustement aléatoire

Le modèle d'ajustement additif s'écrit :

$$Z_i(t) = g(t) + a_t .$$

où  $g(t)$  est soit une fonction du temps inobservable mais déterministe, soit un processus indépendant du processus  $a_t$  . (2)

.../...

### II.2.2. Les modèles autoprojectifs .

Dans ces modèles, on suppose que  $Z_i(t)$  est une fonction de ses valeurs prises dans le passé et d'une perturbation aléatoire  $a_t$  :

$$Z_i(t) = f(Z_i(t-1), Z_i(t-2), \dots ; a_t) \quad (5,1)$$

Dans cette classe de modèles on peut trouver par exemple les modèles de "BOX et Jenkins " . / 8 /

L'avantage essentiel de ces modèles réside dans le choix de celui qui semble le mieux adopté aux données dont on dispose et dans la sélection de la méthode optimale dans un vaste ensemble de possibilité , tandis-que dans les modèles classiques (les modèles d'ajustements) le choix de la fonction utilisée pour l'ajustement est arbitraire .

### III- Technique de conception d'un modèle dans le cas des séries temporelles.

La première étapes de la conception d'un modèle pour d'écrire une série temporelle donnée et le choix du type de modèle qui convient au mieux aux données, cette opération s'appelle identification .

Afin de déterminer complètement le modèle on doit estimer les paramètres du modèle considérée à partir des données . Enfin on effectuera un test de crédibilité pour juger de la qualité du modèle .

Ces étapes sont résumées dans l'organigramme de la figure (2) .

#### III-1. Identification .

Par identification on entend le choix d'un modèle bien précis parmi tous les autres dans le cas des modèles autoprojectifs, et la précision de la nature de la fonction F dans le cas

.../...

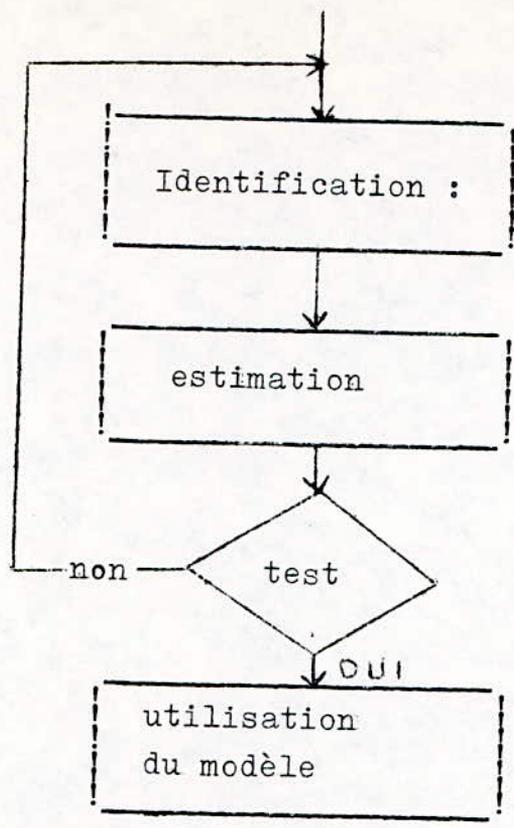


fig ( 2 ) .

des modèles traditionnels (ajustements).

### III.2- estimation .

L'estimation consiste à évaluer les paramètres du modèle identifié ainsi que les caractéristiques de la distribution du bruit  $a_t$  .

Dans le cas d'un modèle linéaire, on estimera les paramètres et les premiers moments de la variable aléatoire  $a_t$  .

Dans ce cas bien précis, les techniques du type moindres carrés permettent d'estimer les paramètres sous certaines conditions: les moments d'ordre un et deux de  $a_t$  doivent être finis et le processus  $a_t$  non autocorrélé.

### III-3- Test .

Le test consiste à vérifier si l'ajustement est correct. Pour ce faire le modèle sera confronté aux faits réels afin de détecter les contradictions éventuelles .

### III-4- utilisation .

L'utilisation du modèle se fera, une fois que toutes les étapes précédentes soient terminées et que les résultats du test de crédibilité du modèle soient bons .

### IV- Critère de sélection d'un modèle .

Ce choix, en termes économiques, doit s'opérer en recherchant à minimiser un coût .

Mais il faut tenir compte également du type de série temporelle et de la finalité du traitement qui jouent le rôle de contraintes dans cette recherche d'optimum .

.../...

Le coût de traitement dépend de deux facteurs qui sont :

- L'algorithme de calcul inhérent à la technique choisie et de ce point de vue certaines techniques impliquent l'usage de gros ordinateurs .

- longueur de l'historique qu'il faut conserver ce qui se traduit par un coût de stockage de données et de traitement plus ou moins important .

En général, ces deux facteurs cités ne sont pas indépendants; les techniques qui utilisent le maximum d'information sont celles qui utilisent les algorithmes les plus complexes .

Cependant, un compromis doit toujours être effectué entre coût de constitution et de traitement d'une information et les avantages que l'on retire .

#### V- Conclusion :

Les différents types de modèles que l'on vient de présenter ne sont pas indépendants, en effet, on peut montrer, par exemple, que certains modèles d'ajustements (lissage exponentiel simple ou double) sont des modèles très particuliers des modèles autoprojectifs du type ARMA. Ce sont ces modèles qui feront l'objet du chapitre suivant .

# CHAPITRE. II

LES MODELES UTILISES

DANS L'APPROCHE

DE BOX ET JENKINS

## I. INTRODUCTION

Deux approches sont possibles pour étudier les séries temporelles . L'une est descriptive et se base souvent sur des modèles simples(modèles d'ajustements), l'autre adopte un point de vue dynamique et explique le passage d'une observation à l'autre(modèles stochastiques ou processus).

La première approche est très traditionnelle, la seconde est plus récente et elle se base sur les travaux de BOX et JENKINS / 8 /. C'est cette approche que nous développons dans ce chapitre .

Au cours de cette étude, nous utiliserons des notions sur les équations linéaires aux différences et sur les processus stochastiques qu'il convient de rappeler .

## II. Les équations linéaires aux différences :

Les équations linéaires aux différences sont les équivalents discrets des équations différentielles des systèmes continus.Elles utilisent des opérateurs discrets / 9 / .

Nous considérons l'ensemble des suites numériques  $\{Z_t\}$  sur lequel nous définissons :

i) l'opérateur de retard B ou (Backward)

$$B Z_t = Z_{t-1} \quad (1.II)$$

ii) l'opérateur d'avance F ou ( Forward)

$$F.Z_t = Z_{t+1} \quad (2.II)$$

.../...

iii) l'opérateur de différence  $\nabla$

$$\nabla Z_t = Z_t - Z_{t-1} \quad (3.II)$$

avec  $\nabla = 1 - B$  (4.II)

Les modèles utilisés dans l'approche de BOX et JENKINS sont basés dans le cas le plus général sur les équations linéaires aux différences de la forme suivante :

$$Z_t - \phi_1 Z_{t-1} - \phi_2 Z_{t-2} \dots - \phi_p Z_{t-p} = a_t - \theta_1 a_{t-1} \dots - \theta_q a_{t-q} \quad (5.II)$$

où  $p$  et  $q$  sont des entiers naturels et  $\phi$  et  $\theta$  sont des réels. Cette équation peut s'écrire de façon plus condensée :

$$\phi(B). Z_t = \theta(B). a_t \quad (6.II)$$

### III. Les processus stochastiques

Un processus stochastique est une famille de variables aléatoires indicées par un ensemble ordonné .

On définit la loi temporelle du processus comme la loi de tout vecteur  $(z_{t_1}, \dots, z_{t_k})$  .

Un tel processus est stationnaire si sa loi temporelle est invariante par translation sur l'axe des temps / 9 / .

#### III-1. L'autocovariance .

Si  $E(Z_t)$  est l'espérance mathématique de  $Z_t$  et  $\text{var}(Z_t)$  est sa variance tel que :

$$m_t = E ( Z_t ) \quad (7.II)$$

et  $\text{Var} (Z_t) = E(Z_t - m_t)^2$  (8.II)

.../...

L'autocovariance s'écrit dans le cas scalaire .

$$\text{cov} (Z_{S_1}, Z_t) = \gamma (S_1, t) = E \left[ (Z_{S_1 - m_{S_1}})(Z_{t - m_t}) \right] \quad (9.II)$$

### III.2- L'autocorrélation .

Elle est définie par :

$$\rho_{ZZ}(S, t) = \frac{\text{cov} ( Z_S, Z_t )}{\sqrt{\text{Var } Z_S \cdot \text{Var } Z_t}} = \frac{\gamma (S, t)}{\sqrt{\gamma(s;s) \cdot \gamma(t, t)}} \quad (10.II)$$

### III.3- Les processus stationnaires du second ordre .

Ce sont ceux qui ont une moyenne et une variance indépendante du temps et dont la fonction de covariance ne dépend que de la différence des temps / 2 / tel que

$$\text{cov} (Z_s, Z_t) = \gamma / t - s / \quad (11.II)$$

Les notions sur les équations aux différences et les processus stochastiques que nous venons de présenter sont utiles pour l'étude et la compréhension des modèles utilisés dans l'approche de BOX et JENKINS .

## IV Les modèles utilisés dans l'approche de BOX et JENKINS .

### III.1- Les modèles autoregressifs ou AR(p).

On considère  $\bar{Z}_t = Z_t - \mu$  un processus centré où  $\mu$  est la moyenne de  $Z_t$  .

$\bar{Z}_t$  est un processus autoregressif d'ordre p s'il vérifie l'équation aux différences suivantes :

$$\bar{Z}_t = \phi_1 \bar{Z}_{t-1} + \phi_2 \bar{Z}_{t-2} + \dots + \phi_p \bar{Z}_{t-p} + a_t \quad (12.II)$$

.../...

en utilisant l'opérateur B , celle-ci peut s'écrire

$$\bar{z}_t = \phi^{-1}(B) a_t = \Psi(B) \cdot a_t \quad (13.II)$$

#### IV.1.1 La fonction d'autocorrélation .

La fonction d'autocorrélation du processus AR(p) est définie par :

$$\rho_k = \frac{\gamma_k}{\gamma_0} \quad (14.II)$$

où  $\gamma_k$  est la fonction d'autocovariance .  
celle-ci vaut .

$$\gamma_k = E(\bar{z}_t, \bar{z}_{t-k}) = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} + \dots + \phi_p \gamma_{k-p} \quad (15.II)$$

La valeur initiale de  $\gamma_k$  est

$$\gamma_0 = \phi_1 \gamma_{-1} + \phi_2 \gamma_{-2} + \dots + \phi_p \gamma_{-p} + \sigma_a^2 \quad (16.II)$$

D'où

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \dots + \phi_p \rho_{k-p} \quad (17.II)$$

• où 
$$\rho_k - \sum_{i=1}^p \phi_i \rho_{k-i} = 0 \quad (18.II)$$

Notons que cette équation aux différences permet de définir le système de YULE WALKER qui permet de calculer les valeurs  $\phi_i$  à partir des  $\rho_i$  .

Sachant que  $\rho_0 = 1$  , on aura en développant l'équation (17.II)

$$\begin{aligned} \rho_1 &= \phi_1 + \phi_2 \rho_1 + \dots + \phi_p \rho_{p-1} \\ \rho_2 &= \phi_1 \rho_1 + \phi_2 \rho_2 + \dots + \phi_p \rho_{p-2} \\ &\vdots \\ \rho_p &= \phi_1 \rho_{p-1} + \phi_2 \rho_{p-2} + \dots + \phi_p \end{aligned}$$

d'où le système de YULE-WALKER

$$\begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_p \end{bmatrix} = \begin{bmatrix} 1 & \rho_1 & \dots & \rho_{p-1} \\ \rho_1 & 1 & & \\ \vdots & & \ddots & \\ \rho_{p-1} & & & 1 \end{bmatrix} \cdot \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix} \quad (19.II)$$

De l'équation (16.II) et compte - tenu de la propriété de symétrie de la fonction d'autocovariance on obtient l'expression de la variance du processus .

$$\sigma_z^2 = \frac{\sigma_a^2}{1 - \phi_1 \rho_1 - \phi_2 \rho_2 - \dots - \phi_p \rho_p} \quad (20.II)$$

#### IV.1.2 La fonction d'autocorrelation partielle .

En remplaçant dans l'équation (16.II) k par j, on obtient le système de YULE-WALKER :

$$\rho_j = \phi_{1j} \rho_{j-1} + \phi_{1_2} \rho_{j-2} + \dots + \phi_{pp} \rho_{j-1} \quad (21.II)$$

$$j = 1, \dots, l$$

où  $\phi_{pp}$  est la fonction d'autocorrélation partielle .

En résolvant ( 21.II) on constate que  $\phi_{11}$  est le dernier terme du vecteur  $\phi_i$  .

de façon générale on a à résoudre

$$\begin{bmatrix} \rho_1 \\ \vdots \\ \rho_l \end{bmatrix} = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{l-1} \\ \rho_1 & 1 & & & \\ \rho_2 & & \ddots & & \\ \rho_{l-1} & & & \ddots & \\ \rho_l & & & & 1 \end{bmatrix} \begin{bmatrix} \phi_{11} \\ \phi_{12} \\ \vdots \\ \phi_{ll} \end{bmatrix} \begin{matrix} \phi_{11} \\ \phi_{12} \\ \vdots \\ \phi_{ll} \end{matrix} \dots / \dots$$

en pratique on utilise les valeurs estimés qui sont définit par

$$\phi_{ll} = \begin{cases} r_1 & l=1 \\ \frac{r_l - \sum_{j=1}^{l-1} \phi_{l-1,j} r_{l-j}}{1 - \sum_{j=1}^{l-1} \phi_{l-1,j} r_j} & l=2,3 \dots \end{cases} \quad (22.II)$$

d'après la méthode de DURBIN / 8 /

IV.2- Les modèles à moyenne mobile ou MA(q).

$\bar{z}_t$  est un processus à moyenne mobile d'ordre q noté MA(q) s'il s'écrit :

$$\bar{z}_t = a_t - \theta_1 a_{t-1} \dots - \theta_q a_{t-q} \quad (23.II)$$

en utilisant l'opérateur B on obtient :

$$a_t = \theta^{-1} (B) \bar{z}_t \quad (24.II)$$

IV.2.1. La fonction d'autocorrelation .

La fonction d'autocorrelation du processus MA(q) est définie toujours par

$$\rho_k = \frac{\gamma_k}{\gamma_0}$$

où  $\gamma_k = E(\bar{z}_t, \bar{z}_{t-k}) = E(a_t - \theta_1 a_{t-1} \dots - \theta_q a_{t-q})(a_{t-k} - \theta_1 a_{t-k-1} \dots - \theta_q a_{t-k-q})$  (25. (25.II)

$a_t$  étant un bruit blanc il est défini par .

$$\begin{aligned} E(a_t) &= 0 \\ E(a_t - a_{t'}) &= \sigma_a^2 \delta_{tt'} = \begin{cases} \sigma_a^2 & t=t' \\ 0 & t \neq t' \end{cases} \end{aligned} \quad (26.II)$$

d'où

$$\begin{cases} \gamma_k = (-\theta_k + \theta_1 \theta_{k+1} + \dots + \theta_{q-k} \theta_q) \sigma_a^2 & k=1, \dots, q \\ \gamma_k = 0 & k > q \end{cases} \quad (27.II)$$

et par conséquent la fonction d'autocorrélation sera :

$$\begin{cases} \rho_k = \frac{\gamma_k}{\gamma_0} = \frac{-\theta_k + \theta_1 \theta_{k+1} + \dots + \theta_{q-k} \theta_q}{1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2} & \text{pour } k=1, \dots, q \\ \rho_k = 0 & \text{pour } k > q \end{cases} \quad (28.II)$$

On pourrait trouver les  $\theta_i$  en fonction des  $\rho_i$  à l'aide du système (28.II) mais on remarque que dans le cas des modèles à moyenne mobile le système est non linéaire d'où la difficulté de déterminer les  $\theta_i$ .

En remplaçant  $k$  par zéro dans l'équation (27.II) on obtient l'expression de la variance :

$$\sigma_z^2 = (1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2) \sigma_a^2 \quad (29.II)$$

#### IV.3 . Les modèles mixtes ou ARMA(p,q)

On appellera un processus mixte d'ordre (p,q) un processus stationnaire  $Z_t$  satisfaisant :

$$\bar{z}_t - \sum_{i=1}^p \phi_i \bar{z}_{t-i} = a_t - \sum_{i=1}^q \theta_i a_{t-i} \quad (30.II)$$

ou encore

$$\phi(B) \bar{z}_t = \theta(B) a_t \quad (31.II)$$

##### IV.3.1. La fonction d'autocorrélation

Après multiplication de l'équation (30,II) par  $\bar{z}_{t-k}$  et passage à l'espérance mathématique on aura la fonction d'autocovariance :

.../...

$$E(\bar{Z}_t, \bar{Z}_{t-k}) = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} + \dots + \phi_p \gamma_{k-p} + \gamma_{Z_a}^{(k)} - \theta_1 \gamma_{Z_a}^{(k-1)} - \theta_q \gamma_{Z_a}^{(k-q)}. \quad (32.II)$$

Sachant que :  $\gamma_{Z_a}^{(k)} = E(\bar{Z}_{t-k} a_t)$

$$\begin{aligned} \text{on a : } \gamma_{Z_a}^{(k)} &= 0 && \text{pour } k > 0 \\ \gamma_{Z_a}^{(k)} &\neq 0 && \text{pour } k=0 \end{aligned} \quad (33.II)$$

donc  $\gamma_k$  s'écrira :

$$\gamma_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \dots + \phi_p \rho_{k-p} \quad \text{pour } k \geq q+1 \quad - (34.II)$$

la fonction d'autocorrélation sera :

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \dots + \phi_p \rho_{k-p} \quad (35.II)$$

d'une façon condensée et en utilisant l'opérateur de retard B on aura :

$$\phi(B) \rho_k = 0 \quad \text{pour } k \geq q+1 \quad (36.II)$$

Les modèles AR(p), MA(q) et ARMA(p,q) sont des processus linéaires stationnaires .

Dans le cas où l'on considère des processus non stationnaires, BOX et JENKINS proposent les modèles ARIMA .

#### V . Les modèles ARIMA(p,d,q).

Si un processus  $\bar{Z}_t$  est non stationnaire , on lui appliquera d'une façon itérative l'opérateur de différence  $\nabla$  jusqu'à l'obtention d'un processus stationnaire  $w_t$ , le processus obtenu sera dit intégré et sera noté ARIMA, il est décrit alors par l'équation aux différences suivante :

$$\begin{aligned} \phi(B) \cdot w_t &= \theta(B) \cdot a_t \\ \text{où } w_t &= (1-B)^d z_t \end{aligned} \quad (37.II)$$

Certain processus possèdent des effets saisonniers et les modèles relatifs seront dit SARIMA.

L'équation aux différences représentant ces modèles sera :

$$\phi(B) \cdot \phi(B^S) \nabla^d \nabla_S^p z_t = \theta(B) \cdot \theta(B^S) a_t \quad (38.II)$$

où  $\phi$  et  $\theta$  sont des polynômes de degré P et Q respectivement et D le degré de différentiation .

## VI . Discussion .

Les fonctions d'autocorrélation et d'autocorrélation partielle permettent :

- le choix du modèle à partir de l'examen des données.
- la détermination de l'ordre des modèles caractérisés par les paramètres p et q .
- la mesure de la corrélation qui existe entre les variables .
- de vérifier la stationnarité de la série .

Leur rôle est dual dans le cas d'un AR(p) et d'un MA(q) c'est-à-dire :

lorsque  $\hat{\rho} = 0$  on a un modèle AR(p)  
et  $\hat{\pi} \rightarrow 0$

où  $\hat{\pi}$  est l'estimé de la fonction d'autocorrélation partielle tandis que lorsque :

$\hat{\rho} \rightarrow 0$   
et  $\hat{\pi} = 0$  on a un MA(q)

notons que : tout processus stationnaire du second ordre peut être approché par un processus à moyenne mobile(2)

Rappelons qu'un processus stationnaire est du second ordre si :

- $\forall t, E(Z_t^2) < \infty$
- $\forall t, E(Z_t) = m$  (indépendant de  $t$ )
- $\forall t, \forall h \in \mathbb{Z}, \text{cov}(Z_t, Z_{t+h}) = \gamma(h)$  (indépendant de  $t$ )

On pourrait se contenter d'un processus à moyenne mobile, mais il est clair que l'introduction des processus autoregressif et des processus mixtes permet d'économiser le nombre de paramètres à déterminer.

Nous pouvons montrer cela en étudiant, par exemple, le passage d'un AR(1) vers un MA (q)

Le processus AR(1) est :

$$\bar{Z}_t - \rho \bar{Z}_{t-1} = a_t \quad |\rho| < 1$$

dans ce processus on a un seul paramètre à estimer  $\rho$  et en le transformant en modèle à moyenne mobile on obtient :

$$\bar{Z}_t = \frac{1}{1 - \rho B} a_t = \sum_{i=0}^{\infty} \rho^i a_{t-i}$$

dans ce cas on a  $i$  paramètres à estimer donc on voit bien que le nombre de paramètres à estimer augmente d'une façon considérable. Par ailleurs, la détermination des  $\theta_i$  dans le cas d'un processus autorégressif est plus simple que la détermination des  $\theta_i$  dans le cas d'un processus à moyenne mobile.

# CHAPITRE III

LA PROCEDURE

DE BOX ET JENKINS

## I Introduction

Le chapitre précédent nous nous a fourni les différents modèles utilisés dans l'approche de BOX et JENKINS. La question, qui se pose maintenant, est comment choisir un modèle précis parmi tous les modèles existants .

Une fois qu'un modèle est retenu, on cherche à estimer ses paramètres en fonction-des données du problème, ces derniers sont : les mesures effectuées sur le phénomène, leur nature propre, les-considerations physiques ets ...

C'est par l'application de la méthode itérative de BOX et JENKINS , qu'on parvient à un modèle parfaitement déterminé

## II Le principe de la méthode.

C'est une procédure itérative qui comporte les phases suivantes .

- Le traitement préalable de données
- L'identification
- L'estimation
- La vérification
- L'utilisation

Le principe de la méthode est illustré par l'organigramme de la figure (3) .

- L'étape d'identification détermine l'ordre du processus, c'est à dire , fixe les valeurs de  $p, d$  et  $q$ , pour ce faire elle utilise comme outil les fonctions-d'autocorrélations, et les fonctions d'autocorrélation partielle .
- La phase d'estimation détermine les paramètres du modèle. Elle est basée sur les méthodes du type moindres carrés.
- La phase de vérification est basée sur l'analyse des résidus .

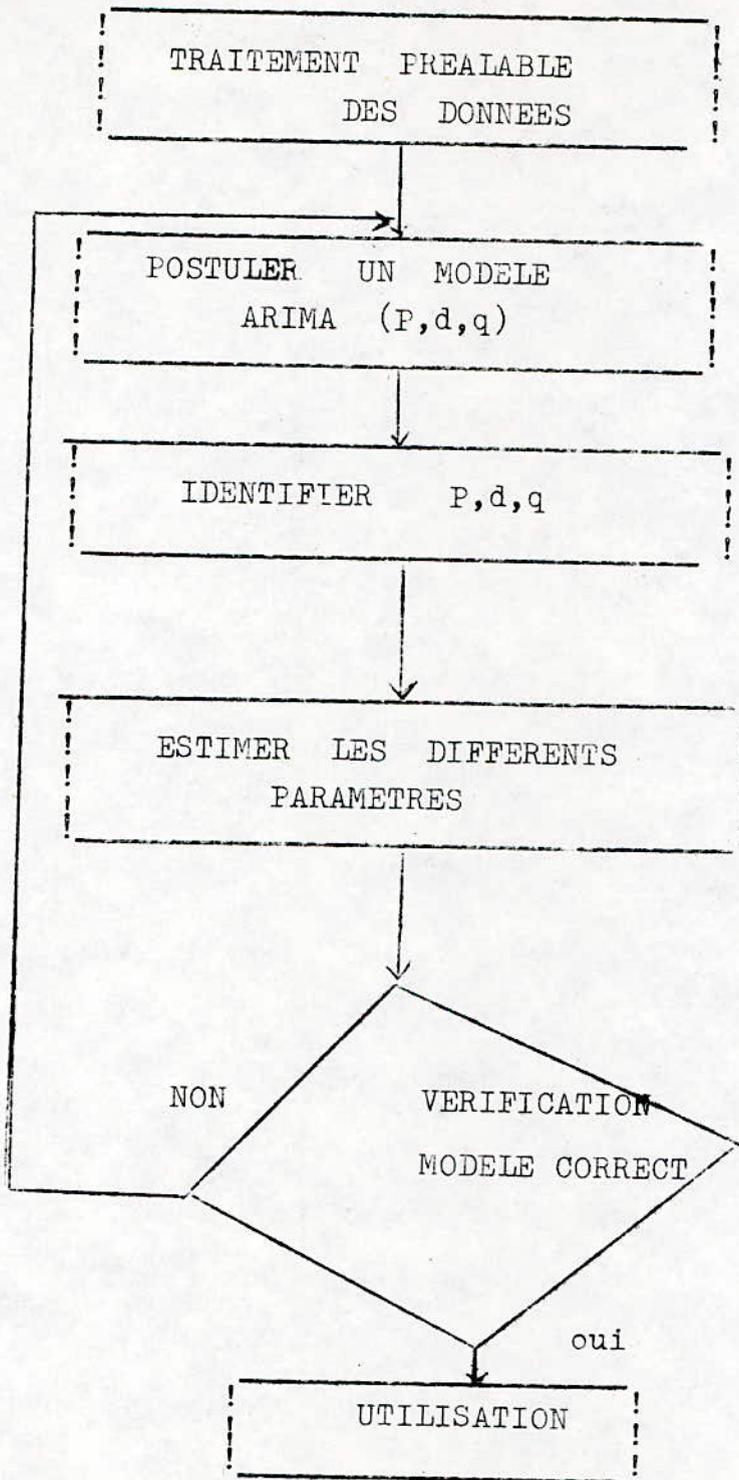


fig ( 3 ) .

Notons qu'entre la détermination de p,d,q, la phase d'identification donnera les valeurs initiales des paramètres. Ce sont ces valeurs que l'on utilisera par la suite dans la phase d'estimation.

### III Etude des différentes phases

#### III.1. Le traitement préalable des données

Il existe des séries temporelles pour lesquelles on ne pourra pas atteindre la stationnarité par l'application de l'opérateur  $\nabla^d$ .

On peut alors envisager d'appliquer à la série  $\{Z_t\}$  une transformation logarithmique avant de lui appliquer l'opérateur  $\nabla^d$ . Celle ci est très utilisée du fait de la croissance avec le temps des variations de  $Z_t$  et de  $\nabla Z_t$

D'une manière générale BOX et COX . / 8 / ont défini la classe de transformation suivante .

$$T_\lambda (Z_t) = \frac{Z_t^\lambda - 1}{\lambda} \quad \lambda \neq 0$$

$$T_0 (Z_t) = \log Z_t \quad \lambda = 0$$

(1.III)

où la paramètre  $\lambda$  sera estimé en même temps que les autres paramètres du modèle .

#### III.2 La phase d'identification

Dans la pratique , on ne connaît pas les paramètres p,d et q d'un modèle ARIMA . Cette phase vise donc à déterminer les degrés p et q et l'ordre d de différence sur la série.

##### III.2.1 Détermination de l'ordre d de l'opérateur de différence .

La fonction d'autocorrélation nous permet de déterminer l'ordre d de différence .

En effet si pour un assez grand nombre de  $h$ , cette dernière ne reste proche de 1, il en résulte que la série est non stationnaire et donc, une différentiation de la série est obligatoire. Le même critère sera appliqué à la série différenciée pour savoir s'il convient de la différencier une deuxième fois, etc...

Dans la pratique, on a toujours  $d \leq 2$ .

### III.2.2. Détermination des degrés $p$ et $q$ .

Elle repose sur le comportement type des fonctions d'autocorrélation  $\hat{\rho}_h$  et d'autocorrélation partielle  $\hat{\pi}_h$

Si la fonction d'autocorrélation  $\hat{\rho}_h$  décroît rapidement vers des valeurs proches de zéro pour  $h > q$ , et la fonction d'autocorrélation partielle  $\hat{\pi}_h$  tend vers zéro, on identifie un processus MA( $q$ ).

Par contre si la fonction d'autocorrélation partielle  $\hat{\pi}_h$  décroît rapidement vers des valeurs proches de zéro pour  $h > p$  et la fonction d'autocorrélation  $\hat{\rho}_h$  tend vers zéro, on identifie un processus AR( $p$ ).

Dans le cas où les deux fonctions considérées d'autocorrélation et d'autocorrélation partielle tendent vers zéro simultanément c'est à dire qu'il n'y a pas de coupures pour  $h = q$  ou  $h = p$  comme dans le cas précédent, le processus considéré est alors un processus mixte.

Un calcul des écarts-type des fonctions d'autocorrélation  $\hat{\rho}_h(h)$  et d'autocorrélation partielle  $\hat{\pi}_h(h)$  nous permet de voir si ces fonctions sont significativement non nulles.

On montre / 2 / que :

- pour un processus MA( $q$ )

$$\forall h > q : \text{Var } \hat{\rho}_h(h) \approx \frac{1}{N} (1 + 2 \sum_{i=1}^q \hat{\rho}^2(i)) \quad (2.III)$$

.../...

- pour un processus AR(p)

$$\forall h > p : \text{Var } \hat{\pi}(h) = \frac{1}{N} \quad (3.III)$$

où N est le nombre d'observations de la série différenciée.

### 2.2.1. Identification d'un modèle ARMA (p,q).

Pour déterminer p et q on passe de la représentation mixte à la représentation autoregressive et à moyenne mobile respectivement.

### 2.2.2 Identification d'un modèle SARIMA(p,d,q)(P,D,Q)

Dans le cas où le modèle est saisonnier, les paramètres P,D,Q sont identifiés par des méthodes analogues à celles décrites précédemment tout en gardant dans les estimés de  $\hat{\rho}(h)$  et  $\hat{\pi}(h)$  que les valeurs de h multiples de la période de saisonnalité .

### III.2.3. Détermination des valeurs initiales des paramètres $\theta$ et $\phi$ intervenant dans le modèle.

Il existe des méthodes propres à chaque type de processus , cependant la méthode la plus générale pour l'estimation de ces valeurs initiales d'un processus ARMA (p,q) est celle donnée par l'algorithme de BOX et JENKINS / 8 / .

Notons que dans la pratique, ces valeurs sont prises arbitrairement tout en respectant le critère de convergence du modèle .

### III.3 La phase d'estimation

Cette opération permet de déterminer les valeurs numériques des paramètres du modèle identifié.

.../...

Nous allons présenter la méthode du maximum de vraisemblance et les techniques du type moindres carrés, qui permettent de définir des procédures itératives d'estimation. Celles-ci s'articulent sur les valeurs initiales estimées lors de la phase d'identification .

III.3.1. L'estimation des paramètres par le maximum de vraisemblance .

Avant d'aborder l'estimation nous allons d'abord définir la fonction de vraisemblance.

3.1.1. La fonction de vraisemblance pour un processus MA(q)

Le processus MA(q) s'écrit :  $w_t = \theta(B) a_t$ . BOX et JENKINS montrent que pour la série d'observations  $w_1, \dots, w_n$  la vraisemblance s'écrit :

$$L(\theta, \sigma_a^2 / w_n) = (2\pi\sigma_a^2)^{-n/2} / |X \tilde{X}|^{-1/2} \exp \left\{ - \frac{1}{2\sigma_a^2} \sum_{t=1}^n [a_t]^2 \right\}$$

où  $|X \tilde{X}|$  désigne un déterminant considéré

comme négligeable par rapport à  $\sum_{t=1}^n [a_t]^2$  et où  $[a_t]^2$

est l'espérance conditionnelle de  $a_t$  sachant  $w_n$  et  $\theta$  .

$n$  étant le nombre d'observations de la série  $w$  différenciée .

3.1.2. Application de la méthode du maximum de vraisemblance à l'estimation des paramètres d'un processus ARMA(p,q)

Dans le chapitre précédent nous avons vu que tout processus mixte admet une représentation moyenne mobile , il est alors possible d'écrire la vraisemblance d'un processus ARMA ( p,q) sous la forme :

.../...

$$L(\phi, \theta, \sqrt{V_a} / w_n) = (2\pi \sqrt{V_a}^2)^{-n/2} / M_n^{1/2} \exp\left[-\frac{S(\phi, \theta)}{2 \sqrt{V_a}^2}\right] \quad (4.III)$$

$$\text{où } S(\phi, \theta) = \sum_{t=-\infty}^n [a_t / w_t, \phi, \theta]^2 \quad (5.III)$$

L'estimation par le maximum de vraisemblance conduit à rechercher les paramètres  $\phi$ ,  $\theta$  et  $\sqrt{V_a}$  qui maximisent l'expression (4.III) ou mieux encore son logarithme .

$$\text{Log } L(\theta, \phi, \sqrt{V_a} / w_n) = -\frac{n}{2} \log 2\pi \sqrt{V_a}^2 - \frac{1}{2} \log M_n - \frac{1}{2 \sqrt{V_a}^2} \sum_{t=-\infty}^n [a_t]^2 \quad (6.III) .$$

Pour maximiser l'expression (6.III) une solution consiste à minimiser les sommes :

$$\sum_{t=-\infty}^n [a_t / \phi, \theta, w_n]^2 = S(\phi, \theta)$$

Nous recherchons alors les valeurs de  $\phi$  et  $\theta$  qui minimisent  $S(\phi, \theta)$ .

En effet si les  $[a_t / \phi, \theta, w]$  sont linéaires en  $\phi$  et  $\theta$  la résolution est classique, il suffira alors de chercher les paramètres  $\phi$  et  $\theta$  qui annulent les dérivées de  $[a_t]$  par rapport à chacun de ces paramètres.

Si par contre les  $[a_t / \phi, \theta, w]$  ne sont pas linéaires en  $\phi$  et  $\theta$ , deux solutions sont encore possibles. La première consiste à linéariser le modèle et on aboutit au cas précédent, tandis que la deuxième solution est basée sur la méthode du Back Forecasting proposée par BOX et JENKINS / 8 /.

III.3.2 L'estimation des paramètres par la méthode de Back Forecasting .

C'est une méthode itérative, basée sur le principe de la minimisation de  $S(\theta, \phi)$  . Elle utilise, les valeurs initiales de  $\theta$  et  $\phi$  choisies lors de l'étape d'identification ainsi que les relations suivantes:

$$\phi(F) [\bar{w}_t] = \theta(F) [\bar{e}_t] \quad (7.III)$$

$$\phi(B) [\bar{w}_t] = \theta(B) [\bar{a}_t] \quad (8.III)$$

où  $e_t$  est un bruit blanc de même variance que les  $a_t$  , mais se déroulant dans le sens de  $t$  décroissant .

Moyennant certaines initialisations, la première relation (7.III) permet le calcul des valeurs de  $w_0, w_{-1}, \dots$  ainsi que  $Q$  pour lequel  $w_{-(1-Q)}$  est négligeable .

La deuxième relation (8.III) permet ensuite le calcul de  $[\bar{a}_t]$  pour  $t$  variant de  $1-Q$  à  $n$  , et par la suite on obtient

$$\sum_{t=1-Q}^n [\bar{a}_t]^2$$

Les paramètres  $\theta$  et  $\phi$  optimaux sont obtenus par itérations , en cherchant le minimum de  $S(\theta, \phi)$  pour  $\theta$  et  $\phi$  appartenant à l'intervalle  $[-1, 1]$  .

Notons que si  $n$  est assez grand le résultat est obtenu avec précision suffisante avec une seule itération .

III.4. La phase de vérification .

La vérification constitue l'étape finale de la méthode de BOX et JENKINS , elle consiste d'une part à tester

l'a justement d'une série identifiée et estimée d'autre part à r  
rechercher dans quelle direction il conviendrait de modifier  
le modèle,

Le modèle subira alors deux types de test .

- Un test sur les paramètres estimés .
- Un test sur l'indépendance (non corrélation) des résidus .

#### III.4.1 Test sur les paramètres estimés

On peut tester la crédibilité d'un modèle ARIMA  
(p,d,q) en augmentant d'une unité l'ordre  $p$  . Si la valeur de  
 $\phi_{p+1}$  est différente de zéro, on peut conclure que le modèle  
ARIMA (p,d,q) ne décrit pas les données expérimentales.

On peut effectuer la même opération en augmentant  $q$   
d'une unité et faire le même test | 7 |

Cependant on évitera d'augmenter simultanément  $p$  et  $q$   
d'une unité afin d'éviter le problème de redondance .

#### III.4.2. L'Analyse des résidus .

Dans ce type de test, on veut vérifier si la série  
des résidus obtenus est cohérente avec l'hypothèse selon  
laquelle  $\{a_t\}$  est un bruit blanc .  
on appelle résidu le processus  $\hat{a}$  défini par :

$$\hat{\phi}(B) w_t = \hat{\theta}(B) \hat{a}_t \quad (9.III)$$

Si l'ajustement est correct on peut montrer que :

$$\hat{a} \# \alpha$$

Il s'agit donc de vérifier que  $\hat{a}$  est voisin d'un bruit blanc.  
BOX et PIERCE /8/ ont proposé un test fondé sur la statistique

.../...

$$Q = n \sum_{h=1}^k \hat{\rho}_h^2 (\hat{a}) \quad (10.III)$$

où dans la pratique  $k = 20 + p + q$ .

$\hat{\rho}_h (\hat{a})$  représente la corrélation empirique des résidus  $(\hat{a})$ .

On montre que pour  $n$  assez grand  $Q \sim \chi^2$  à  $k-p-q$  degrés de liberté.

En fait si  $n$  est faible il est préférable de considérer / 8 / la statistique :

$$Q' = n(n+2) \sum_{h=1}^k (n-h)^{-1} \hat{\rho}_h^2 (\hat{a}) \quad (11.III)$$

### III.4.3 Le choix d'un modèle

Dans le cas où plusieurs modèles différents franchissent les tests de la vérification, le choix de l'un deux est difficile.

On pourra sélectionner celui qui minimisera le critère d'AKAIKE /2/.

En effet, supposons que le vrai modèle soit un ARMA(p,q) et supposons que l'on ait plusieurs modèles ayant franchis les tests de la vérification qu'on dénote par ARMA(p,q), ARMA(p<sub>2</sub>,q<sub>2</sub>),....., ARMA(p<sub>m</sub>,q<sub>m</sub>).

AKAIKE propose une distance entre L'ARMA(p,q) et les ARMA(p<sub>i</sub>,q<sub>i</sub>) qui est la suivante.

$$AIC = n \log \hat{\sigma}^2 a^2 + \frac{2n(p_i + q_i + 1 + \delta)}{n - d}$$

avec  $\delta = \begin{cases} 1 & \text{si } d = 0 \\ 0 & \text{si } d > 0 \end{cases} \quad (12.III)$

Le critère consiste à minimiser AIC pour un ARMA  $(p_i, q_i)$  -  
 $i = 1, \dots, m$  .

### III . 5. Utilisation .

Une fois que le modèle est identifié , estimé et testé il sera éventuellement utilisé dans un but bien précis tel que la prévision et le contrôle .

CHAPITRE IV  
ORGANIGRAMME  
DES  
TRAITEMENTS

## I. Introduction

Nous présentons une synoptique d'algorithmes relatifs aux différentes étapes de la méthodologie. (\*)

## II Présentation des différentes étapes

Les synoptiques d'algorithmes que nous présentons sont propres aux modèles stochastiques unidimensionnels (c'est à dire traitant des séries scalaires discrètes )

### II.1 L'étape d'identification

Elle consiste en l'introduction d'une série temporelle (d'une façon appropriée), transforme les observations, applique une opération de différentiation et puis calcule :

- 1 - la moyenne  $\bar{w}$  et la variante  $S_w^2$
- 2 - la fonction d'autocovariance  $C_k$
- 3 - la fonction d'autocorrélation  $r_k$
- 4 - la fonction d'autocorrélation partielle  $\hat{\phi}_{11}$

#### II.1.1 les entrées

Les informations nécessaires pour les calculs sont :  
les valeurs de la série temporelle :

$$Z_t, t = 1, 2, \dots, N$$

- le nombre d'observation :  $N$

(\*) cela devrait permettre, dans un travail ultérieur, de faciliter l'écriture dans un langage Scientifique évolué (de type FORTRAN ou PL/1 )

- le degré de différence maximum (\*) de la série désaisonnalisée:

$$d \geq 0$$

- le degré de différence de la série saisonnière

$$D \geq 0 \quad (*)$$

- la période de saisonnalité :

$$S \geq 1, \quad (N - d - SD - 1)$$

- le retard ( ou LAG) (\*\*) de la fonction d'autocorrélation : K

- le retard maximum de la fonction d'autocorrélation partielle:

$$L \leq K$$

- les paramètres de transformation :

$$\lambda \text{ et } m .$$

## II. 12 . Le traitement

Pour chaque série des valeurs (d,D,  $\lambda$  , m ) la série temporelle sera transformée comme suit :

$$Z'_t = \begin{cases} (Z_t + m) & \lambda \neq 0 \\ \text{Log}_e (Z_t + m) & \lambda = 0 \end{cases}$$

où si  $\lambda \neq 0$ , le paramètre m est choisi tel que  $Z_t + m$  est positif pour tout t et , si  $\lambda = 1$  , m est fixé à zéro tel que la série  $Z_t$  reste inchangée .

La série différenciée :

$$w_t = \nabla^d \nabla_S^D Z'_t \quad t = 1, 2, \dots, n$$

(\*) la pratique a montré que  $d \leq 2$  et  $D \leq 2$

(\*\*) en général, on fixe le LAG à 36 ou 48 .

où

$$\begin{aligned} \nabla w_t &= w_t - w_{t-1} \\ \nabla_S w_t &= w_t - w_{t-S} \\ n &= N - d - SD \end{aligned}$$

et les quantités suivantes sont alors calculées :

- La moyenne et la variance

$$\bar{w} = \frac{1}{n} \sum_{t=1}^n w_t$$

$$S_w^2 = C_0$$

où  $C_0$  est définie ci dessous .

- la fonction d'autocovariance :

$$C_k = \frac{1}{n} \sum_{t=1}^{n-k} (w_t - \bar{w}) (w_{t+k} - \bar{w})$$

où  $k = 0, 1, \dots, K$

- la fonction d'autocorrélation :

$$r_k = \frac{C_k}{C_0}$$

où  $k = 0, 1, \dots, K$

- la fonction d'autocorrélation partielle :

$$\hat{\phi}_{ll} = \begin{cases} r_1 & l=1 \\ r_1 - \sum_{j=1}^{l-1} \hat{\phi}_{l-1,j} r_{l-j} & l=2, 3, \dots, L \\ 1 - \sum_{j=1}^{l-1} \hat{\phi}_{l-1,j} r_j & \end{cases}$$

..../....

où  $\hat{\phi}_{ij} = \hat{\phi}_{l-1,j} - \hat{\phi}_{ll} \hat{\phi}_{l-1, l-j} \quad j = 1, 2, \dots, l-1$

( Algorithme de DURBIN )

II.1.3. les sorties

On conviendra d'éditer toutes les informations d'entrées et les paramètres de transformation ( $\lambda$  et  $m$ ), les autres sorties peuvent être :

- les valeurs de la série différenciée et transformée :

$$\{w_t\} ; t = 1, 2, \dots, n$$

- le nombre de valeurs de  $w_t$  :  $n$
- la moyenne de la série  $w$  :  $\bar{w}$
- la variance de la série :  $S_w^2$
- la fonction d'autocovariance :

$$C_k , k = 0, 1, \dots, K$$

- la fonction d'autocorrélation :

$$r_k , k = 0, 1, \dots, K$$

- la fonction d'autocorrélation partielle :

$$\hat{\phi}_{ll} , l = 1, 2, \dots, L .$$

II.2 . l'étape d'estimation .

Le principe de cette étape consiste à introduire les valeurs des paramètres initiaux et ensuite calcule

- les paramètres estimés aux moindres carrés  $u, \phi, \bar{\phi}, \theta, \sigma$  et  $\sigma_a^2$  dans le modèle saisonnier :

.../...

$$\phi(B) \bar{\phi}(B^S) (w_t - \mu) = \theta(B) \theta^S(B^S) a_t$$

où  $w_t = \nabla^d \nabla_S^D Z_t'$

$Z_t'$  est la série transformée, définie par l'étape d'identification.

$\mu$  est la valeur moyenne de la série  $w$

- la variance empirique ( des résidus ) .

### II.2.1 les entrées

l'information minimale requise pour le calcul exige la connaissance des paramètres suivants :

-  $\{Z_t\}$ ,  $N, d, D, S, R, \lambda, m$ .

-  $p$  et  $q$

- le nombre de paramètres du modèle autorégressif saisonnier:

$$P \geq 0$$

- le nombre de paramètres du modèle à moyenne mobile saisonnier :

$$Q \geq 0 ; P + Q > 0$$

- le paramètre de contrôle pour la moyenne  $\mu$  :

$$M = 1 \quad \text{Si } \mu \text{ est inclu.}$$

$$M = 0 \quad \text{Si } \mu \text{ est non inclu.}$$

- les valeurs initiales estimées des paramètres du modèle autorégressif non saisonnier :

$$\hat{\phi} = ( \hat{\phi}_{10} ) \dots , ( \hat{\phi}_{P0} )$$

- les valeurs initiales estimées des paramètres du modèle autorégressif saisonnier :

$$\hat{\bar{\phi}} = ( \hat{\bar{\phi}}_{10} , \dots , \hat{\bar{\phi}}_{P0} )$$

- les valeurs initiales estimées des paramètres du modèle à moyenne mobile non saisonnier :

$$\hat{\theta} = ( \hat{\theta}_{10} , \dots , \hat{\theta}_{q_0} )$$

- les valeurs initiales estimées des paramètres du modèle à moyenne mobile saisonnier

$$\hat{\theta} = ( \hat{\theta}_{10} , \dots , \hat{\theta}_{q_0} )$$

- le nombre maximum d'itération : it

Dans la pratique , les valeurs initiales des paramètres sont prises comme désirées , tout en étant proches de zéro sans être nulles .

### II.2.2 le traitement .

Ayant les séries initiales  $w$  projetées en arrière, les résidus , pour une série des valeurs spécifiées des paramètres sont calculés ceci va nous permettre de calculer la somme des carrés des résidus en utilisant les deux équations suivantes :

$$\phi(F) \bar{\phi}(F^S) (w_t - \mu) = \theta(F) \theta(F^S) e_t$$

$$\phi(B) \bar{\phi}(B^S) (w_t - \mu) = \theta(B) \theta'(B^S) a_t$$

F : opérateur d'avance ( BF = 1 )

$a_t$  est la notation de  $\left[ a_t / u, \theta, \theta', \bar{\phi}, \phi, w \right]$  qui représente les estimations de  $a_t$  faites conditionnellement aux  $w_t$  connus et à  $\mu, \theta$  et  $\phi$  donnés .

$\mu$  apparait seulement Si  $M = 1$

$t = Q', Q' + 1, \dots, n$

.../...

où  $Q'$  est un temps au delà duquel les prévisions arrières sont négligeables .

Pour démarrer la récurrence une initialisation préalable est faite :

$$e_t = 0 \quad \text{pour } t = n$$

les prévisions arrières des valeurs  $w_0, w_{-1}, \dots$  sont arrêtées quand  $w_t - \mu$  devient très petit .

ainsi , pour les valeurs données des paramètres  $(\mu, \theta, \vartheta, \phi, \bar{\phi})$  , la somme des carrés des résidus est calculée à partir de :

$$S(u, \theta, \vartheta, \bar{\phi}, \phi) = \sum_{t=1-Q'}^n [a_t]^2$$

La valeur de la variance résiduelle empirique est obtenue à partir de la somme des carrés des résidus

$$\frac{Va^2}{n-p-q-P-Q-M} = \frac{1}{n-p-q-P-Q-M} S(\hat{u}, \hat{\theta}, \hat{\vartheta}, \hat{\theta}, \hat{\vartheta})$$

Finalement la valeur estimée du terme constant global  $\hat{\theta}_0$  est obtenu à partir de .

$$\hat{\theta}_0 = \hat{\mu} G$$

avec

$$G = \left(1 - \prod_{i=1}^P \hat{\phi}_i\right) \left(1 - \prod_{j=1}^P \hat{\phi}_j\right)$$

### II.2.3. les sorties

Celles-ci doivent inclure toute information d'entrée ainsi que :

$$\hat{u}, \hat{\phi}, \hat{\vartheta}, \hat{\theta}, \hat{\vartheta}, S(\hat{u}, \hat{\phi}, \hat{\vartheta}, \hat{\theta}, \hat{\vartheta})$$

à chaque itération :

.../...

nous aurons en sortie :

- les résidus correspondants aux estimés des moindres carrés :

$$\{\hat{a}_t\}, \quad t = 1, 2, \dots, n$$

- la variance résiduelle empirique :  $\overline{Va}^2$
- la valeur estimée du terme constant global :  $\hat{\theta}$

### II.3 l'étape de vérification .

En utilisant les résidus  $\hat{a}_t$  , les autocorrélations des résidus sont calculées à partir de :

$$r_{\hat{a}\hat{a}}(k) = C_{\hat{a}\hat{a}}(k) / C_{\hat{a}\hat{a}}(0)$$

et la statistique des résidus obtenue est comparée à la distribution de KHI-DEUX théorique (  $\chi^2$  )

#### II.3.1. les entrées

- les résidus estimés par les moindres carrés :

$$\hat{a}_t, \quad t = 1, 2, \dots, n .$$

- la variance empirique des résidus :

$$\overline{Va}^2$$

#### II.3.2 le traitement

Il calcule l'autocorrélation des résidus  $\{\hat{a}_t\}$  à partir de la relation

$$r_{\hat{a}\hat{a}}(k) = C_{\hat{a}\hat{a}}(k) / C_{\hat{a}\hat{a}}(0)$$

où 
$$C_{\hat{a}\hat{a}}(k) = \frac{1}{n} \sum_{t=1}^{n-k} (\hat{a}_t - \bar{a}) (\hat{a}_{t+k} - \bar{a})$$

et 
$$\bar{a} = \frac{1}{n} \sum_{t=1}^n \hat{a}_t$$

avec  $t = 0, 1, \dots, K$

.../...

La statistique de KHI - DEUX est calculée à partir de :

$$KHI = n \sum_{k=1}^K r_{\hat{a}\hat{a}}^2(k)$$

Elle est comparée à la distribution de HKI - DEUX avec  $(K-M-p-q-P-Q)$  degrés de libertés .

### II.3.3 les sorties

- les autocorrélations des résidus :

$$r_{\hat{a}\hat{a}}(k) , k = 1, 2, \dots, K$$

- la statistique de KHi - DEUX :

$$KHi$$

### II.4. Organigramme de l'étape (algorithmique) d'identification de la methodologie

Nous distinguons principalement l'étape d'identification (il convient , par la suite, de distinguer les autres étapes ).

L'organigramme correspondant est donné par la fig(4) par ailleurs nous avons constaté que l'organigramme fait appel à la technique des moindres carrés dans le cas où le modèle **est** linéaire .

C'est ainsi que nous avons en l'idée d'élaborer un logiciel qui permet d'effectuer un ajustement polynomial à un nuage de point donné .

L'organigramme correspondant a fait l'objet de la fig(5) et le programme a été réalisé sur le micro-ordinateur APPLE . II .

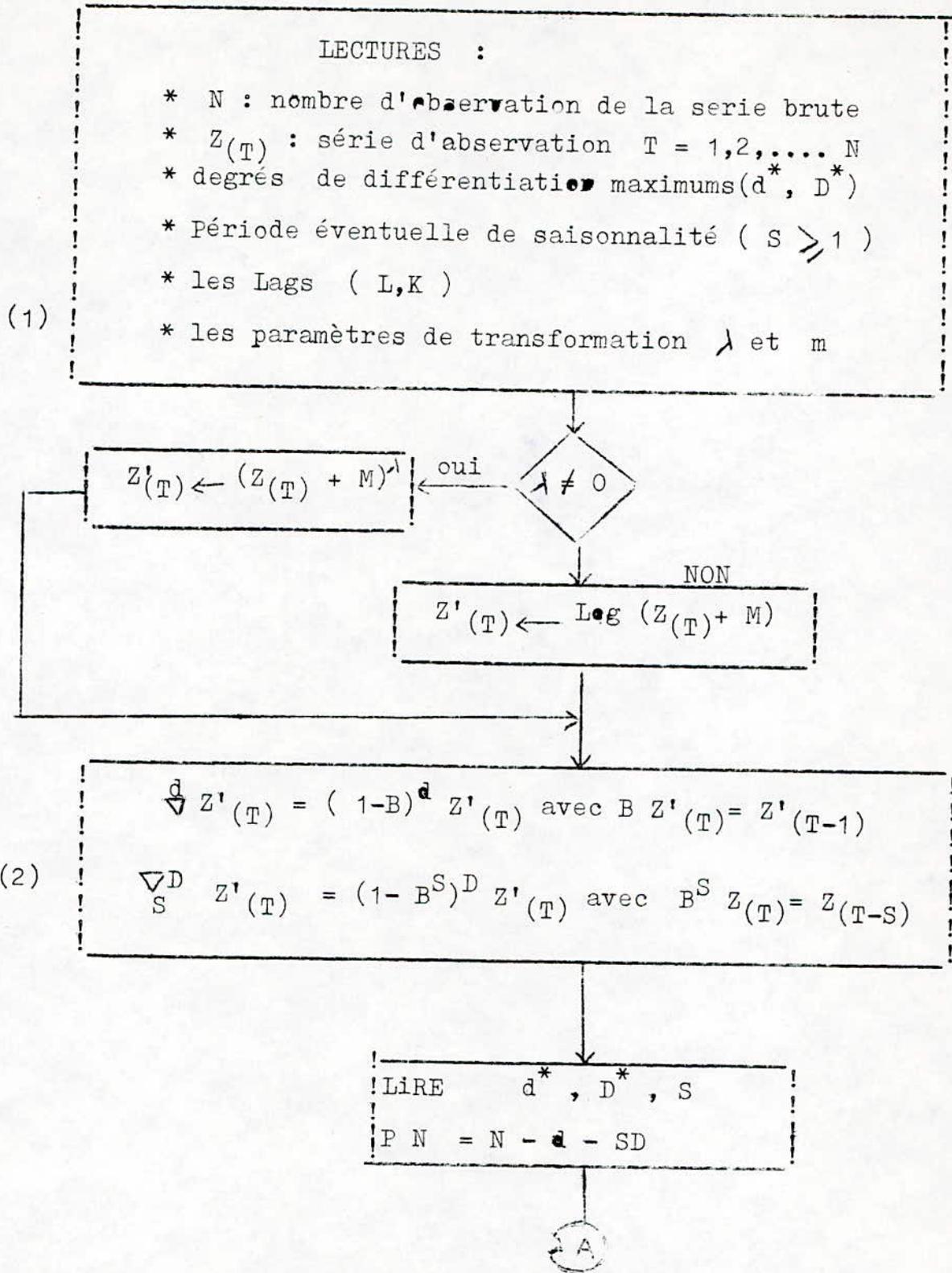
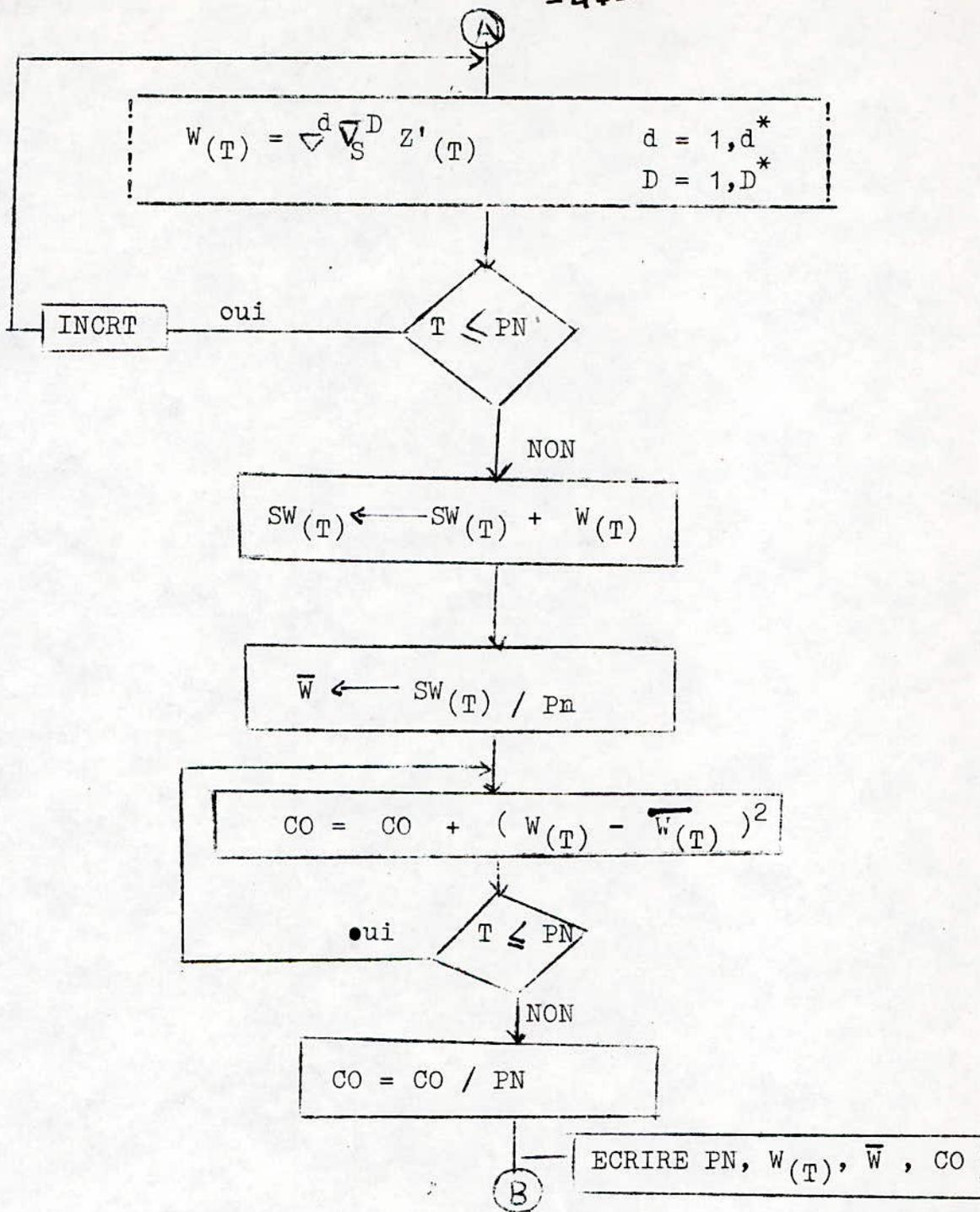
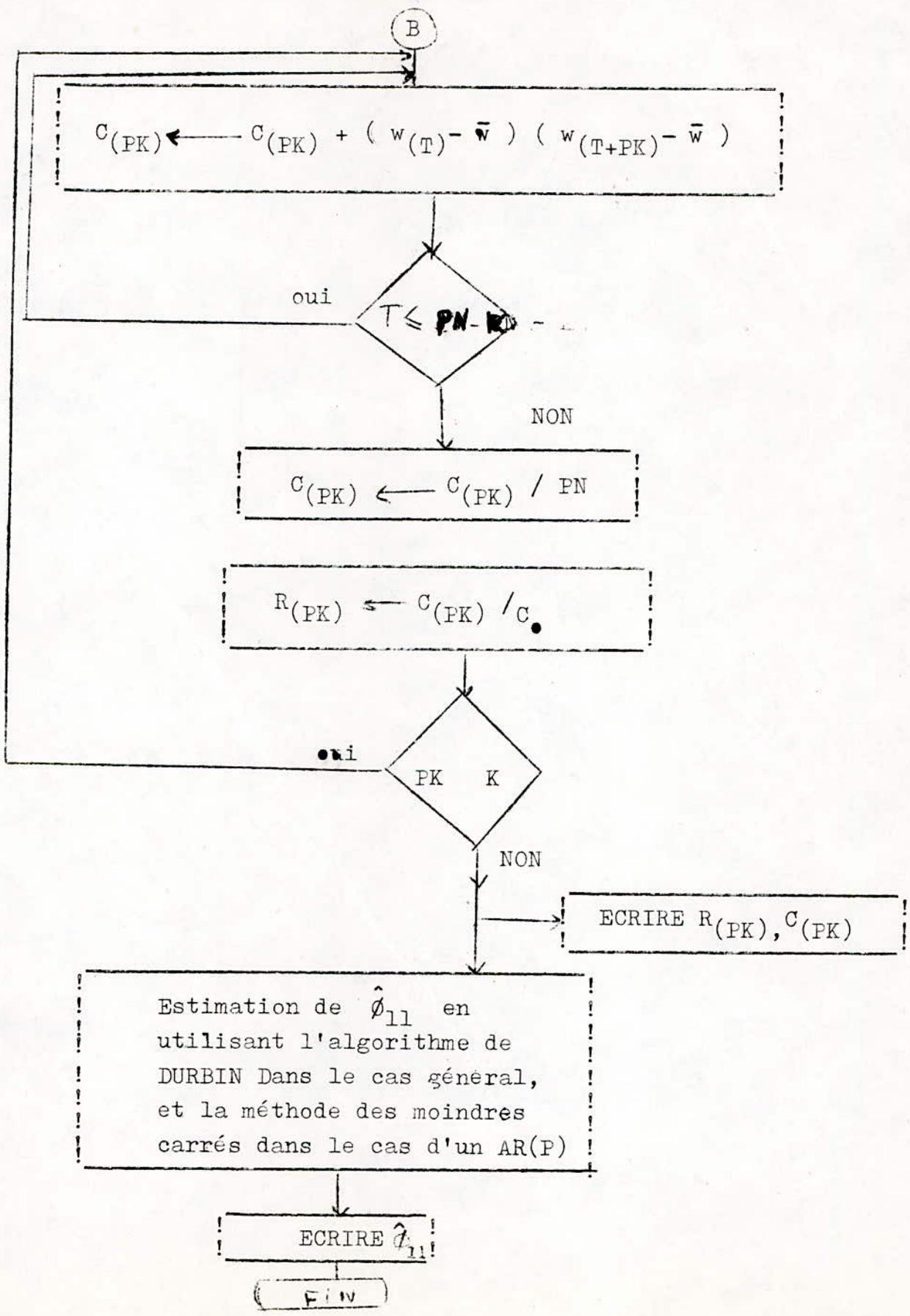


fig (4)

- (1) dans un vecteur par exemple dimension à N
- (2) se calculent par itération successives





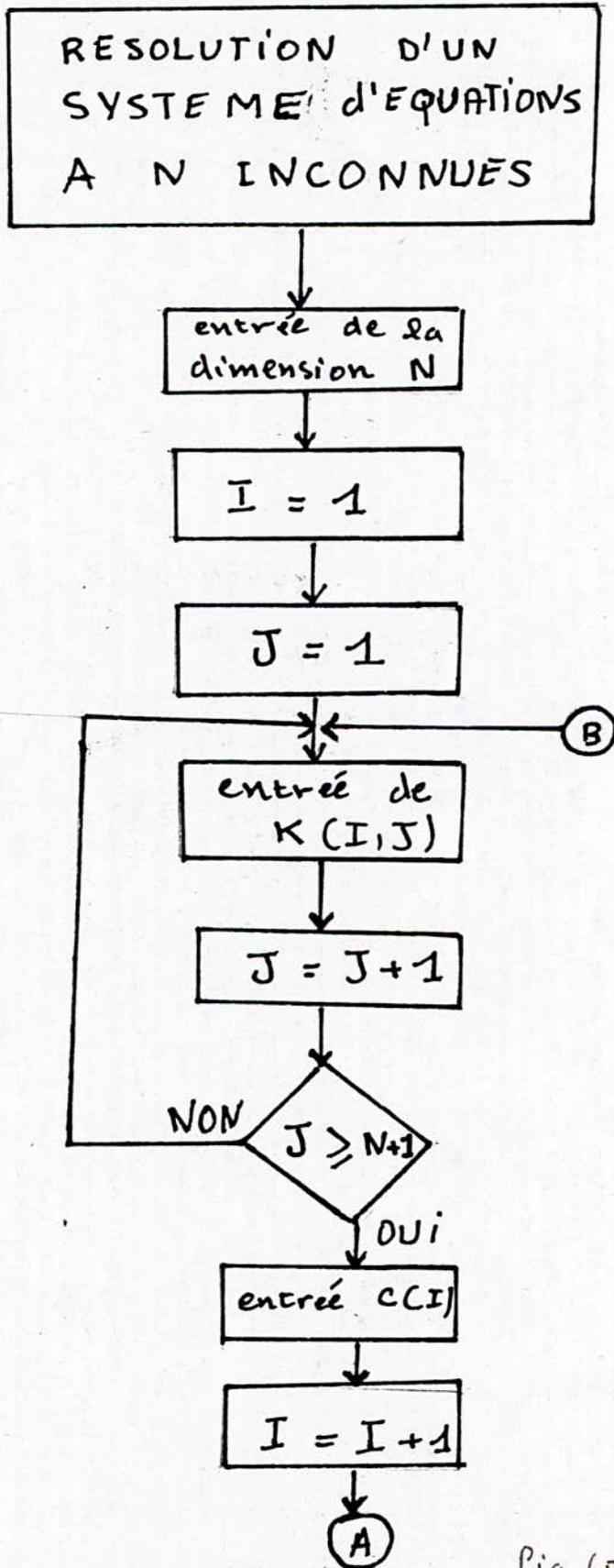
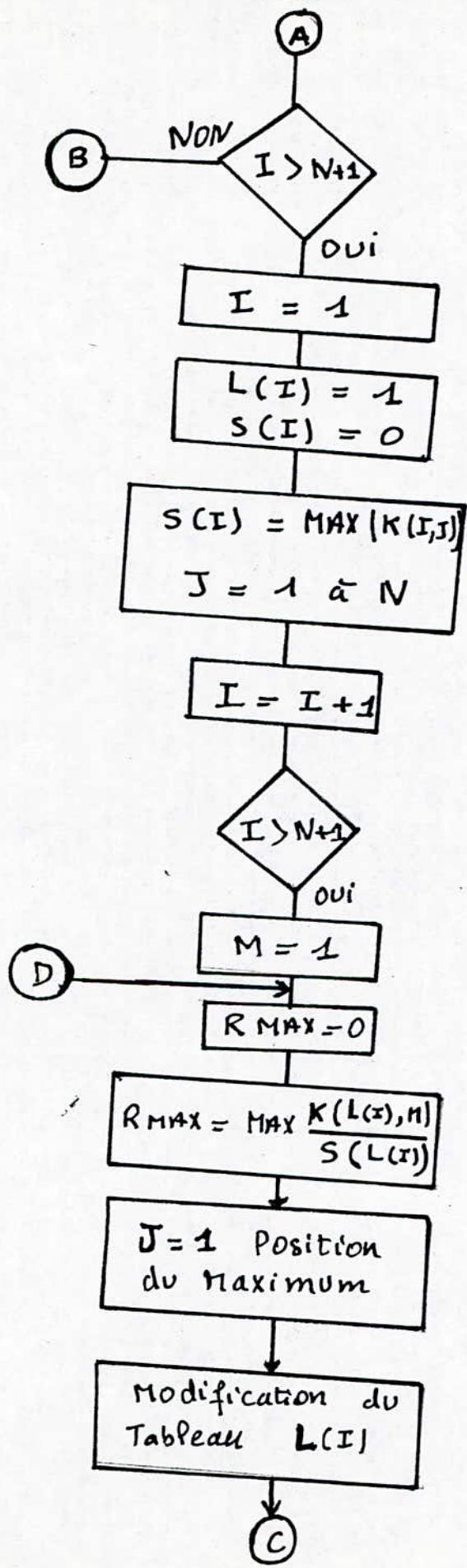
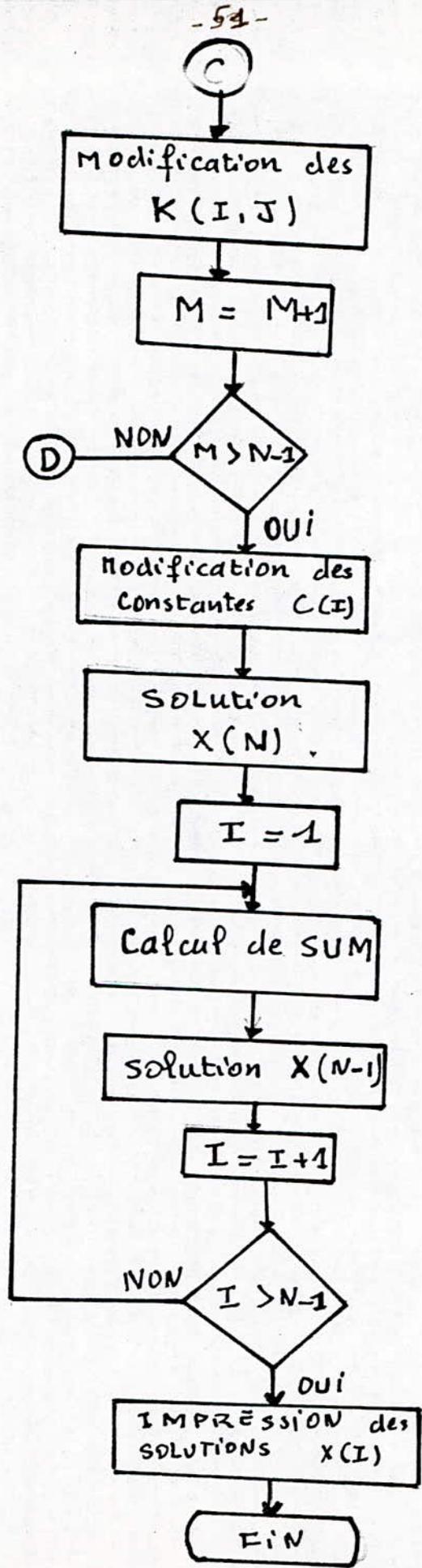


fig (5)





## LIST

```

1  INPUT "N=?";N
2  INPUT "N1=?";N1
3  N3 = 2 * N1 + 1
4  N2 = N1 + 1
5  DIM X(N),Y(N),S(N3),SU(N3),C(N3)
10 REM :LECTURE DES X(I)
15 DATA 2,3,4
20 FOR I = 1 TO N
25 READ X(I)
30 NEXT I
35 REM :LECTURE DES Y(I)
40 DATA 9,29,5
45 FOR I = 1 TO N
50 READ Y(I)
55 NEXT I
60 REM :FORMATION DES SOMMES X(I)
65 FOR LX = 0 TO 2 * N1
70 S(LX) = 0
75 FOR LK = 1 TO N
80 S(LX) = S(LX) + X(LK) ^ LX
85 NEXT LK
86 PRINT LX; TAB( 4);"S(",LX;")=";S(LX)
90 NEXT LX
95 REM :FORMATION DES SOMMES Y(I)
100 FOR LX = 0 TO N1 - 1
105 SU(LX) = 0
110 FOR LK = 1 TO N
130 C(LK) = (X(LK) ^ LX) * Y(LK)
140 SU(LX) = SU(LX) + C(LK)
145 NEXT LK
150 PRINT LX; TAB( 4);"SU(",LX;")=";SU(LX)
160 NEXT LX
190 REM :FORMATION DE LA MATRICE
200 N2 = N1 + 1
210 DIM A(N1,N2)
220 FOR I = 1 TO N1
230 FOR J = 1 TO N2
235 A(I,J) = S(I + J - 2)
240 IF J > N1 THEN A(I,J) = SU(I - 1)
245 PRINT TAB( 5 * J);A(I,J);
270 NEXT J
275 PRINT
276 PRINT
280 NEXT I
300 REM :ALGORITHME DE GAUSS
310 FOR I = 1 TO N1
320 L(I) = I
330 P(I) = 0
340 FOR J = 1 TO N1
350 D = ABS (A(I,J))
360 IF D > P(I) THEN P(I) = D
370 NEXT J
380 NEXT I
390 FOR M = 1 TO N1 - 1
405 RMAX = 0
410 FOR I = M TO N1
420 R = ABS (A(L(I),M)) / P(L(I))
430 IF R < = RMAX THEN 460
440 J = I
450 RMAX = R
460 NEXT I
470 LK = L(J)
480 L(J) = L(M)
490 L(M) = LK

```

```
500 FOR I = M + 1 TO N1
510 XM = A(L(I),M) / A(LK,M)
520 A(L(I),M) = XM
530 FOR J = M + 1 TO N2
540 A(L(I),J) = A(L(I),J) - XM * A(LK,J)
550 NEXT J
560 NEXT I
570 NEXT M
580 FOR J = 1 TO N2
590 FOR I = J + 1 TO N1
600 A(L(I),N2) = A(L(I),N2) - A(L(I),J) * A(L(I),N2)
610 NEXT I
620 NEXT J
630 Z(N1) = A(L(N1),N2) / A(L(N1),N1)
640 FOR I = 1 TO N1 - 1
650 SUM = A(L(N1) - I,N2)
660 FOR J = N1 - I + 1 TO N2
670 SUM = SUM - A(L(N1) - I,J) * Z(J)
680 NEXT J
690 Z(N1 - I) = SUM / A(L(N1 - I),N1 - I)
700 NEXT I
710 PRINT "LES SOLUTION SONT."
720 FOR I = 1 TO N1
730 PRINT "Z(" + I + ") = " + Z(I)
740 NEXT I
750 END
```

# CHAPITRE V

## APPLICATIONS

## I- INTRODUCTION .

Les modèles que nous venons d'étudier sont utilisés dans différents domaines tels que, l'automatique, l'électronique, l'économétrie, la météorologie / 3,4,5,6 / . Leur champ d'application est donc très vaste .

Au cours de ce travail nous avons remarqué que les modèles les plus utilisés se limitent à l'ordre 2 , c'est ainsi que nous avons fait l'étude des modèles AR(1), AR(2), MA(1), MA(2) et ARMA (1,1).

Afin de mieux illustrer l'application de ces modèles, dans la pratique nous présentons la (discussion d'un modèle statistique en météorologie solaire .

## II - Les processus autoregressifs .

### II.1 - Les processus autoregressifs d'ordre 1 (processus de Markov)

Les processus autoregressifs d'ordre un s'écrivent

$$\bar{z}_t = \phi_1 \bar{z}_{t-1} + a_t$$

les processus sont toujours inversibles et la condition de stationnarité est donnée par

$$|\phi_1| < 1$$

En utilisant l'expression ( 17,4) nous pouvons déterminer la fonction d'autocorrélation qui s'écrit :

$$\rho_k = \phi_1^k \quad k \geq 0.$$

Pour  $k = 1$  elle sera noté

$$\rho_1 = \phi_1$$

Les diagrammes représentant les processus

$$\bar{Z}_t = 0,8 \bar{Z}_{t-1} + a_t$$

et

$$\bar{Z}_t = - 0,8 \bar{Z}_{t-1} + a_t$$

sont données par la fig (7)

Les corrélogrammes de ces processus sont représentés par la fig (8) .

Nous constatons que la fonction d'autocorrélations décroît exponentiellement vers zero quand  $\phi_1$  est positifs, mais décroît exponentiellement vers zéro en oscillant quand  $\phi_1$  est négatif .

Rappelons que l'étude des corrélogrammes est très importante dans l'étape d'identification .

## II.2. Les processus autoregressifs d'ordre 2

Les processus autoregressifs d'ordre deux s'écrivent

$$\bar{Z}_t = \phi_1 \bar{Z}_{t-1} + \phi_2 \bar{Z}_{t-2} + a_t$$

Ces processus sont toujours inversibles . Pour satisfaire la condition de stationnarité il faut que les racines de l'équation homogène

$$\phi(B) \bar{Z}_t = 0$$

Soient situés hors du disque unité .

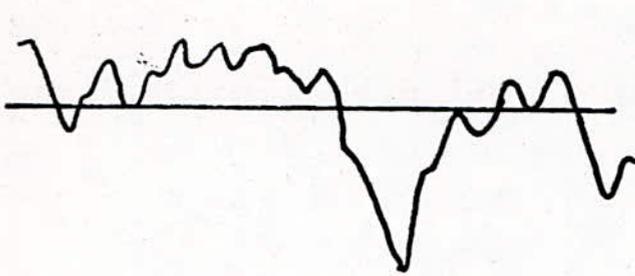
Cette condition définit la région de stationnarité par les inégalités suivantes .

$$\phi_2 + \phi_1 < 1$$

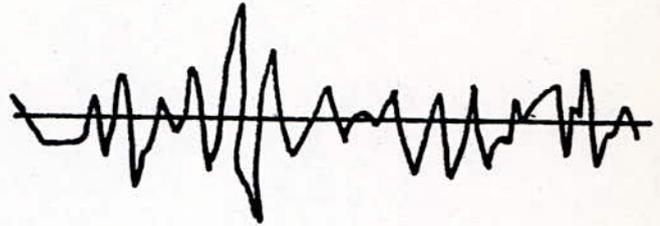
$$\phi_2 - \phi_1 < 1$$

$$-1 < \phi_2 < 1 .$$

.../...

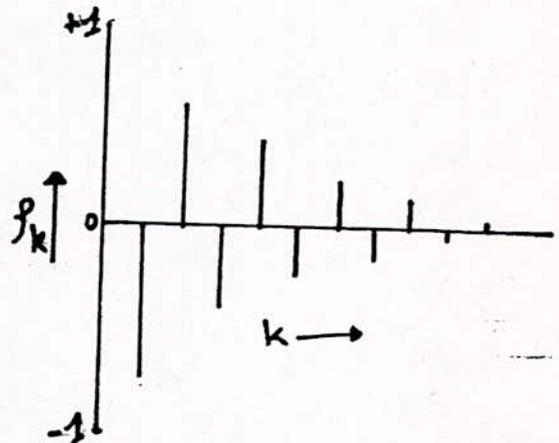
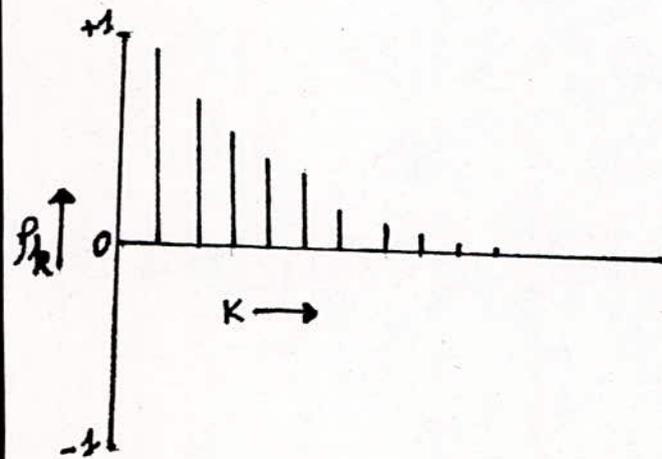


$$\bar{z}_t = 0,8 \bar{z}_{t-1} + a_t$$



$$\bar{z}_t = -0,8 \bar{z}_{t-1} + a_t$$

FIG. 7



Fonction d'autocorrélation théorique

FIG. 8

Les équations de Yule Walker permettent d'avoir avec  $\rho_0 = 1$

$$\rho_1 = \phi_1 + \rho_1 \phi_2$$

$$\rho_2 = \phi_1 \rho_1 + \phi_2$$

en utilisant les conditions de stationnarité et les expressions de  $\rho_1$  et  $\rho_2$  on obtient la région admissible de  $\rho_1$  et  $\rho_2$

telles que :

$$-1 < \rho_1 < 1$$

$$-1 < \rho_2 < 1$$

$$\rho_1^2 < \frac{1}{2} (\rho_2 + 1) .$$

les régions admissibles pour les paramètres  $\phi_1, \phi_2$  et  $\rho_1, \rho_2$  seront représentées respectivement par les fig (9) et fig (10)

### II.3 les processus moyennes mobiles d'ordre 1 ou MA(1)

Les processus moyennes mobiles d'ordre un s'écrivent

$$\bar{z}_t = a_t - \theta_1 a_{t-1}$$

ces processus sont stationnaires, quelque soit les valeurs de  $\theta_1$  .

Ils seront inversibles si

$$-1 < \theta_1 < 1$$

en utilisant l'expression (28.II) nous pouvons déterminer la fonction d'autocorrelation qui s'écrit :

$$\rho_k = \begin{cases} -\theta_1 & k = 1 \\ \frac{-\theta_1}{1 + \theta_1^2} & k \geq 2 \\ 0 & k \geq 2 \end{cases}$$

.../...

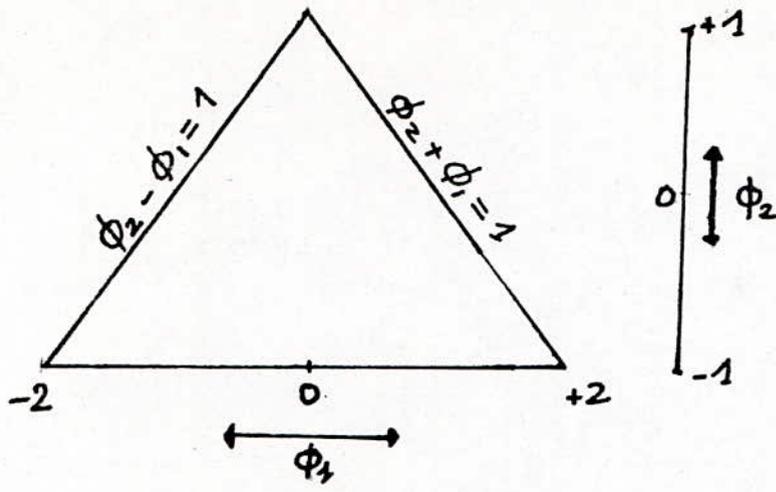


FIG. 9

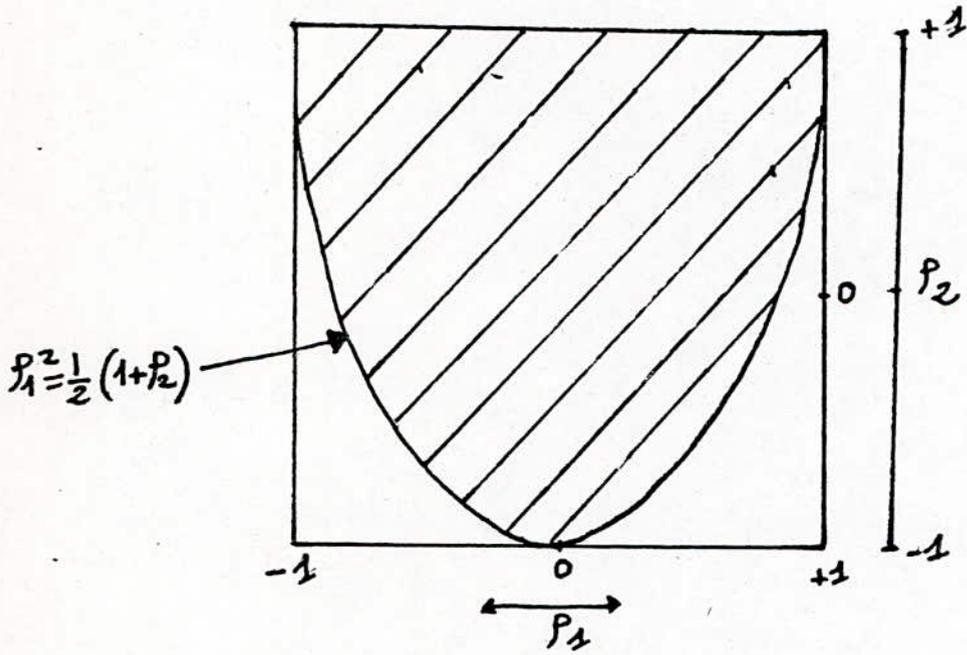


FIG. 10

## II.4 Les processus moyennes mobiles d'ordre 2 ou MA(2)

Les processus moyennes mobiles d'ordre deux s'écrivent

$$\bar{z}_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$$

Ces processus sont toujours stationnaires . Pour satisfaire la condition d'inversibilité il faut que les racines de l'équation homogène

$$\theta(B) \bar{z}_t = 0$$

Soient situées hors du disque unité. Cette condition définie :

$$\theta_1 + \theta_2 < 1$$

$$\theta_2 - \theta_1 < 1$$

$$-1 < \theta_2 < 1$$

La région admissible de  $\theta_1$  et  $\theta_2$  est représentée fig (11) en utilisant l'expression (28,II) on détermine le système de la fonction d'autocorrelation :

$$\rho_1 = \frac{-\theta_1(1-\theta_2)}{1 + \theta_1^2 + \theta_2^2}$$

$$\rho_2 = \frac{-\theta_2}{1 + \theta_1^2 + \theta_2^2}$$

$$\rho_k = 0 \quad k \geq 3$$

en utilisant les conditions d'inversibilités et les expressions de  $\rho_1$  et  $\rho_2$  , les valeurs admissibles de  $\rho_1$  et  $\rho_2$  se situe dans la région définie par .

$$\rho_1 + \rho_2 = -0,5$$

$$\rho_2 - \rho_1 = -0,5$$

$$\rho_1^2 = 4 \rho_2 (1 - 2 \rho_2)$$

la fig(12) correspond, aux régions admissibles de l'autocorrélation .

### II.5. Les Processus mixtes d'ordre 1 ou ARMA(1,1)

Les processus mixtes d'ordre un s'écrivent

$$\bar{z}_t - \phi_1 \bar{z}_{t-1} = a_t - \theta_1 a_{t-1}$$

Ces processus sont stationnaires si la condition suivante est vérifiée :

$$|\phi_1| < 1$$

de même pour la condition d'inversibilité il faut avoir  $|\theta_1| < 1$

La région admissible des paramètres  $\theta_1$  et  $\phi_1$  est le carré donné par fig (13)

les deux premières autocorrélation sont représentés par:

$$\rho_1 = \frac{(1 - \phi_1 \theta_1) (\phi_1 - \theta_1)}{1 + \theta_1^2 - 2 \phi_1 \theta_1}$$

$$\rho_2 = \phi_1 \rho_1$$

En utilisant les expressions de  $\rho_1$  et  $\rho_2$  ainsi que les conditions de stationnarité et d'inversibilité on montre que la région admissible pour  $\rho_1$  et  $\rho_2$  est définie par :

$$|\rho_2| < |\rho_1|$$

$$\rho_2 > \rho_1 (2\rho_1 + 1) \quad \rho_1 < 0$$

$$\rho_2 > \rho_1 (2\rho_1 - 1) \quad \rho_1 > 0$$

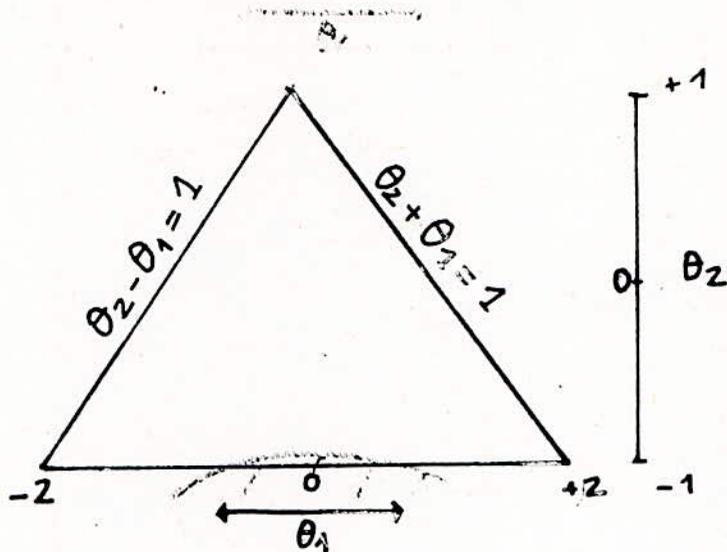


FIG. 11

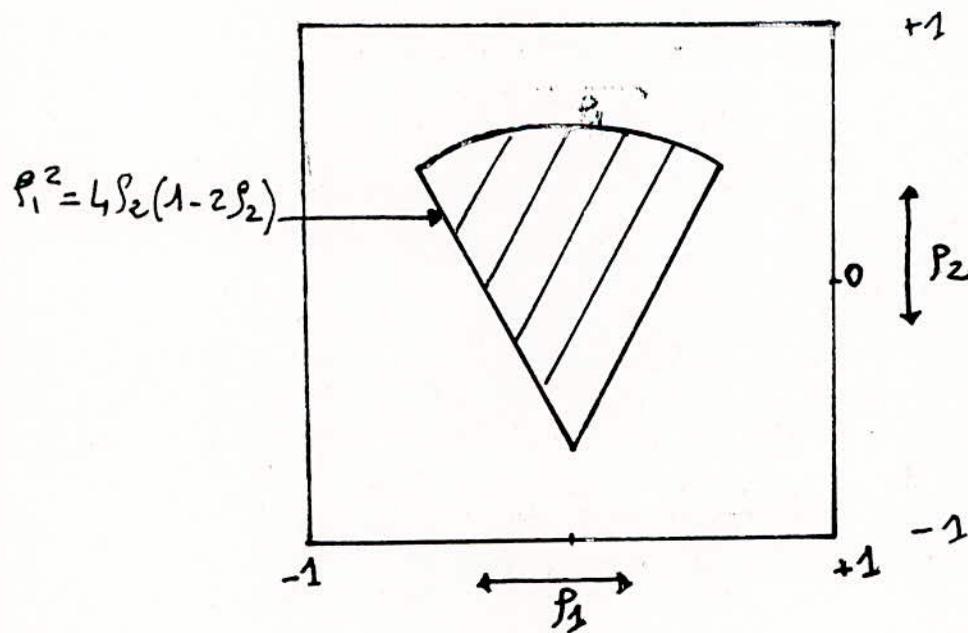


FIG 12

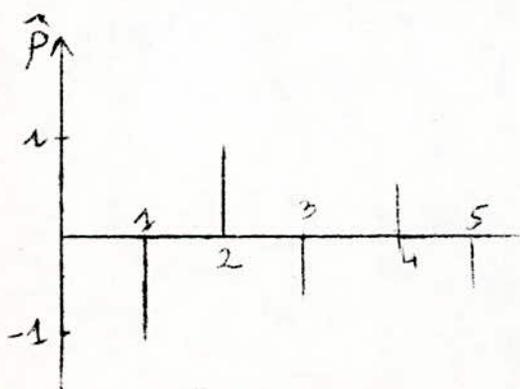
la fig(14) représente ces régions .

- à travers cette étude on constate que les processus d'ordre un et deux sont faciles à appliquer mais au delà de cet ordre une complication est apparente lors de l'estimation des différents paramètres et de la détermination des régions admissibles

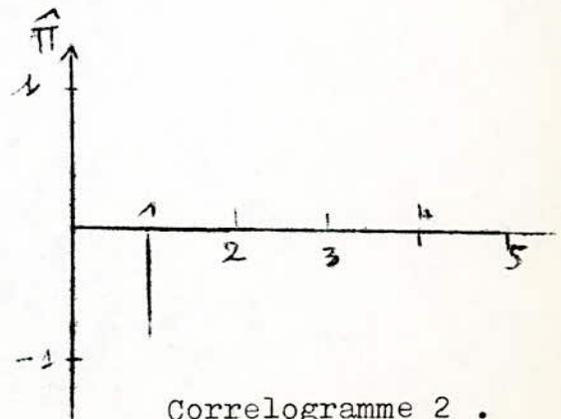
Afin de montrer comment on parvient à identifier un modèle à l'aide des fonctions d'autocorrélations et d'autocorrelation partielle. ANDERSON / 10 / a simulé 4 séries de 200 observations , qui lui ont permis de justifier à l'aide de leurs correlogrammes les modèles en question .

	A		B		C		D	
h	$\hat{\pi}$	$\hat{\rho}$	$\hat{\pi}$	$\hat{\rho}$	$\pi$	$\hat{\rho}$	$\hat{\pi}$	$\hat{\rho}$
1	- 0,800	-0,800	0,449	0,449	-0,719	-0,719	0,427	0,427
2	0,085	0,670	-0,324	-0,056	-0,375	0,337	0,358	0,475
3	0,112	-0,518	0,218	-0,023	-0,048	-0,083	-0,160	0,169
4	-0,044	0,390	-0,118	0,028	0,239	0,075	0,106	0,253
5	-0,061	-0,310	0,077	0,073	-0,173	-0,088	0,035	0,126

On trace les correlogrammes correspondant à la série A .



Correlogramme 1



Correlogramme 2 .

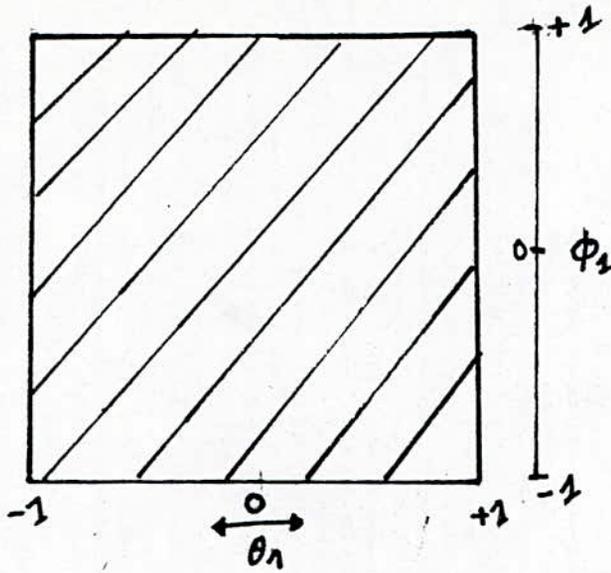


FIG 13

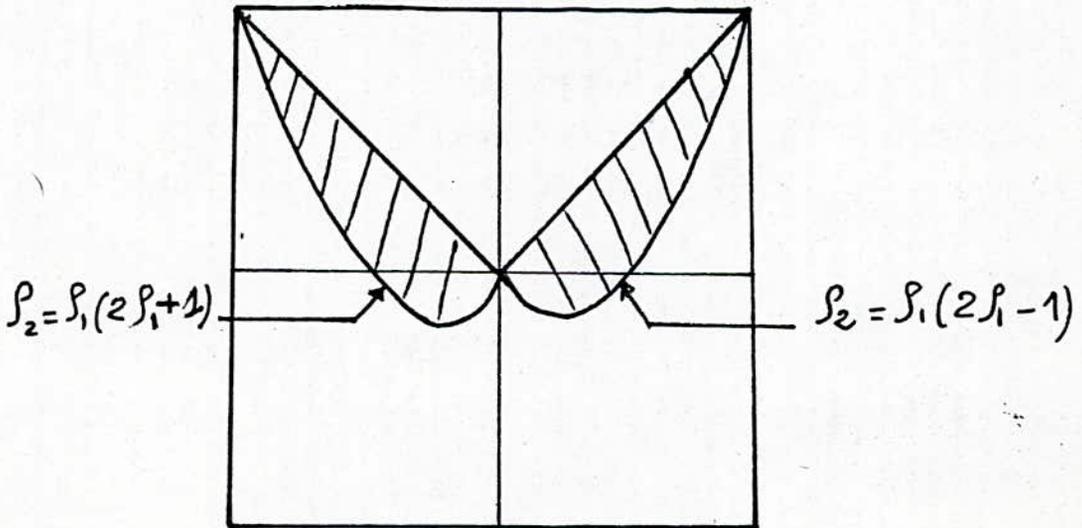


FIG 14

Ces corrélogrammes montrent qu'on est en présence d'un processus autoregressif d'ordre 1 .

Le même raisonnement est fait pour l'analyse des séries B, C et D on obtient pour :

$$B \sim MA(1) \quad ; \quad C \sim MA(2) \quad ; \quad D \sim AR(2).$$

### III- présentation d'une discussion d'un modèle statistique en météorologie solaire .

Pour montrer l'application des modèles de BOX et JENKINS , nous présentons la discussion d'un modèle statistique en météorologie solaire/5/ l'objet de cette étude est de rechercher à partir des mesures de l'irradiation journalière relevée au PEROU (à HUALLAO) , sur une période de 11ans, le modèle susceptible de représenter les fluctuations de cette irradiation . Pour cela l'auteur a utilisé les méthodes d'écrites par BOX et JENKINS afin de fixer son choix sur un modèle à deux paramètres et qui s'écrit :

$$w_t = (1 - \theta B) (1 - \theta' B) a_t$$

où  $w_t$  représente l'irradiation solaire,  $\theta$  et  $\theta'$  sont les paramètres du modèle qu'il faudrait estimer .

B étant défini par la relation (1.II) et  $a_t$  est un bruit blanc .

#### III-1 Description du modèle utilisé

##### III.1.1 nature des données

les données utilisées représentent les mesures de l'irradiation solaire journalière  $I_j$  faite en climat équatorial au PEROU sur une période de 11 ans .

L'examen graphique de l'évolution de cette irradiation solaire journalière en fonction du temps a révélé la présence des fluctuations de type saisonnier autour d'une même courbe moyenne.

de la suite  $I_j$  ( où  $j$  varie de 1 à 365 X 11), des suites saisonnières analogues  $^m I_j$  ont été déduites en remplaçant  $m$  mesures consécutives par leur valeur moyenne .

$$m\bar{I}_j = \frac{1}{m} \sum_{k=(j-1)m+1}^{k=jm} I_k$$

La representation graphique de cette nouvelle série a montré une réduction notable de l'amplitude des fluctuations et donc il sera intéressant de modeliser ces suites  $m\bar{I}_j$

### III.1.2- recherche du modèle

Pour rechercher un modèle susceptible de représenter les suites  $m\bar{I}_j$  l'auteur a employé la méthode décrite par BOX et JENKINS / 8 / . ainsi une première analyse des données a permis d'affirmer que la série est non stationnaire et donc une transformation préliminaire pour stationnariser la série est indispensable cette transformation consiste à remplacer les suites  $m\bar{I}_j$  par  $Z_j = \text{Log. } m\bar{I}_j$

Enfin l'étude simultanée de la fonction d'autocorrélation et d'autocorrélation partielle a permis de choisir le modèle le mieux adapté à la série temporelle étudiée et l'ordre de ce modèle qui est :

$$w_t = (1-\theta B)(1-\theta B). a_t$$

### III.1.3 Détermination des paramètres

La méthode retenue pour obtenir une bonne approximation des paramètres consiste à minimiser par rapport aux paramètres la somme  $\sum [a_t]^2$  où  $[a_t]$  est l'estimé de  $a$  conditionnellement aux valeurs des paramètres et de la série connue sur une période de  $T$  jours .

Le choix - de  $T$  est déterminé par la nécessité d'une précision suffisante .

### III.2 Utilisation du modèle .

Le modèle obtenu a été utilisé pour la simulation et la prévision .

En outre son utilisation peut être étendue au dimensionnement des systèmes solaires, en particulier pour dimensionner des appareils de stockage de type passif [ 6 ]

Ce modèle a donné de Bons résultats en ce qui concerne la simulation cependant il est difficilement utilisable pour la prévision du fait que les corrélations sont négligeables au delà du deuxième jour .

( c;a;d n'apportent aucun supplément d'informations).

### CONCLUSION

A l'heure actuelle ces modèles sont de plus en plus employés comme moyen d'éliminer les inconvénients (coûts, données) de l'approche par régression multiple, par exemple :

L'Etude et l'utilisation des modèles ARMA sont très bien adaptées pour la prédiction à court-terme (ce qui s'avère très intéressant dans les traitements de signaux), cette durée dépend bien entendu, de la nature du phénomène et de son unité chronologique.

L'approche ARMA peut être donc considérée comme une méthode extrêmement riche et élégante (par sa rigueur et son fondement théorique).

Cependant, et malgré ses qualités cette méthodologie reste d'une utilisation relativement limitée.

En effet, son utilisation requiert des données importantes en nombre (problèmes asymptotiques qualités statistiques...), son identification demande une analyse poussée faite par un personnel très averti, et aussi, elle ne permet pas une adaptation dynamique aux changements systématique non inscrit dans le passé du phénomène

Il est certain que les algorithmes associés aux différentes étapes soient complexes en raison du nombre d'informations et des problèmes numériques qui peuvent en découler. L'ensemble des étapes ne peut être complètement automatisé et nécessite absolument l'intervention du praticien.

Enfin, des extensions à des modèles à paramètres eux mêmes dynamiques, ainsi qu'aux séries liées, ou vectorielles donnent lieu à des recherches importantes qui devraient déboucher sur des procédures aussi opérationnelles que celles du cas scalaire que nous venons d'exposer.

## B I B L I O G R A P H I E

- 1 - G. CALLOT  
cours de statistiques descriptive  
DUNOD
- 2 - A. MONFORT  
Methode de Prevision des séries Temporelles  
1979
- 3 - A. BENVENISTE and C. CHAVRE  
AR and ARMA identification algorithms of levinson type ; an innovations  
approach I.E.E.E trans. On automatic control  
Vol. AC - 26 , N° -6 , Decembre 1981.
- 4 - D.T.L. Lee, B. FRIEDLANDER, and M. MORF  
Recursive ladder algorithms for ARMA modeling.  
I.E.E.E trans, automatic control,  
vol. AC - 27 , N° 4, august 1982
- 5 - S. M . KAY  
Noise compensation for autoregressive spectral estimates  
I.E.E.E trans. A coustics, speech, and signal processing,  
vol. AS SP - 28, N° 3, June 1980
- 6 - Rev. Phys. App, N° 14, PP 92-60, Janvier 1979
- 7 - K. BACHAR. &  
Les séries temporelles . Recueil du séminaire international  
"journée de statistique et informatique appliquées"  
Alger Mars 1982  
- Fascicule II les processus ARMA  
ENSAE , Mai 1979
- 8 - BOX and JENKINS  
Tome séries analysis for casting and control holden day SON FRANCISCO 1971

- 9 - QUITTARD PINON  
élément de statistique  
série chronologique scalaire  
O.P.U Alger 1983.
- 10 - OD. ANDERSON "The ~~BOX~~ JENKINS"  
approch to time séries analysis  
RAIRO (Vol II N° 1 février 1977.