

*République Algérienne Démocratique et Populaire*  
*Ministère de l'Enseignement Supérieur et de la Recherche Scientifique*

*UNIVERSITÉ de JIJEL*



*Département de Génie Civil et Hydraulique*

Envue de l'obtention du Diplôme de

**MASTER ACADEMIQUE**

**Filière : Hydraulique**

**Option: Hydraulique urbaine**

**Modélisation des apports du barrage Boukerdane  
par la méthode de Box-Jenkins**

**Présenté par:**

➤ **Ben Salah Mohammed**

**Invitation : Dr .F boumizez**

**Promotion 2021**

**Encadre par :**

**Mr : Boudjerda M**

# Résumé

L'objectif de ce travail est d'étudier les techniques d'analyse et de modélisation des séries chronologiques. On s'intéresse essentiellement aux modèles *ARIMA*; en utilisant les techniques de Box et Jenkins. Cette méthode vise à formuler un modèle permettant de représenter une chronique avec comme finalité de prévoir des valeurs futures. On termine ce travail par le principe de simulation sur les séries chronologiques avec *R*; en attachant quelques exemples. A la fin on traite une application sur des données réelles.

**Mots clés:** Série chronologique ; Box-Jenkins; Modèles *ARIMA*; Simulation et Prévision.

## *Abstract*

The objective of this work is to study the techniques of analysis and modeling of timeseries. We are mainly interested in *ARIMA* models; using Box and Jenkins techniques. This method is intended to formulate a model for representing a chronicle with the purpose of predicting future values. This work is completed by the simulation principle on the timeseries with *R*; we attach a few examples. At the end we have to treat an application on real data.

**Keywords:** Timeseries; Box-Jenkins; *ARIMA* models; Simulation and Forecast.

## الملخص

الهدف من العمل هو دراسة تقنيات تحليل ونمذجة السلاسل الزمنية. نهتم من خلال هذه الدراسة بدراسة نماذج أريما باستخدام تقنيات بوكس-جنكيز، التي تهدف إلى صياغة وقائع لنمذجة ظواهر بهدف التنبؤ بالقيم المستقبلية. نقوم بتطبيق هذا النموذج من خلال مبدأ المحاكاة في السلسلة الزمنية في برنامج *R*، كما نقوم بتطبيق هذه التقنية على معطيات حقيقية تتمثل في دراسة سعر البرميل الخام بترول برنت.

**كلمات مفتاحية:** سلسلة زمنية، بوكس=جنكيز، نماذج أريما، المحاكاة والتنبؤ.



# *Remerciements*

Je dédie ce modeste travail à ceux qui sont toujours.

Présents dans mon cœur.



A ma très **chère mère**, A mon très **cher père**

A mes amis toute ma famille.

A tous mes professeurs.

A tous mes collègues et mes amis.

A ma promotion **(2021)**.





# *Dédicace*

Avant tout, Je remercie Dieu le tout puissant, pour m’ avoir donné la santé, le courage et la volonté d’étudier et pour m’ avoir permis de réaliser ce modeste travail dans les meilleures conditions.

Je remercie fortement mon promoteur : **Mr.Boudjerda Merouane** et **Dr . F. Boumizez**de m’ avoir orienté par ses conseils judicieux dans le but de mener à bien ce travail.

Mon derniers remerciements, mais non les moindres s’ adressent à tous mes amies de la promotion 2021, pour avoir beaucoup de sollicitudes et de chaleur humaine dont nous avions tant besoin.

Et en fin, je remercie tous ceux qui ont contribué de loin ou de près à la réalisation de ce De travail.

# Sommaire

Liste des abréviations.....	vii
List des figures .....	viii
List des tableaux .....	ix
Introduction générale .....	01

## Chapitre I :Modélisation Hydrologique

I. Généralités sur la modélisation hydrologique.....	03.
Introduction .....	03
I.2. Les modèles hydrologiques.....	03
I.2.2. Principes des modèles hydrologiques.....	03.
I.3. Utilisation des modèles hydrologiques.....	04
I.4. Pourquoi du modèle hydrologique.....	04
I.5. Critères de qualité d'un modèle hydrologique.....	05
I.6. Caractéristiques intrinsèques d'un modèle hydrologique.....	05
I.7. Classification des modèles hydrologiques.....	06
I.7.1. Modèles stochastique.....	07
I.7.2. Modèles déterministes.....	07
I.7.3. Modèles boites noirs.....	08
I.7.4. Modèles à base physique.....	08
I.7.5. Modèles par a métriques.....	09
I.7.6. Modèles empiriques .....	09
I.7.7. Modèles analytiques.....	09
I.7.8. Modèles conceptuels.....	09
I.7.9. Modèles globaux.....	09
I.7.10. Modèles Spatialisés.....	10
I.7.11. Modèles conceptuels spatialisés ou semi-spatialisés .....	10
I.7.12 Modèles physiques spatialisés.....	10
I.7.13. Modèles physiques conceptuels semi spatialisés.....	11
I.7.14. Modèle pluie-débit.....	13

I.7.15. Modèle débit-débit.....	14
I.7.16. Le modèle des réseaux de neurones .....	15
I.8. La nature des unités élémentaires.....	16
I.9. Calage et performance d'un modèle hydrologique.....	16
I.9.1. Généralitésurlecalageetlaperformanced'unmodèle.....	16
I.9.2. La nécessité de cette phase de calage des paramètres a plusieurs origins.....	17
I.9.3. Les méthodes automatiques de calage des modèles.....	18
I.10.1. Généralité sur le calage et la performance d'un modèle.....	18
I.10. Calage et performance d'un modèle hydrologique.....	18

## **Chapitre II :Les Séries Chronologiques**

1.Introduction sur les séries chronologiques.....	21
I.1 Définition :.....	21
I.2. Types importants de séries chronologiques.....	21
I.2.1 Objectifs et problèmes de l'analyse des séries chronologiques .....	23
I. 2.2. Problèmes d'analyse des séries chronologiques.....	23
I.2. 3.Lacomposantesaisonnaireouvariationssaisonnieres .....	26
3.La composante résiduelle.....	26
I.3.11. Série désaisonnalisée ou série CVS.....	33
I.3.13.2. Conditions de stationnarité et d'invisibilité pour AR(p) et MA(q).....	37
I.5.14. Les processus aléatoires non stationnaires.....	37
I.5.15.....	39
I.5.16 modele ARCH ET GARCH.....	40
Conclusion.....	41

## **Chapitre III :La Zone D'étude**

Barrage de Boukerdane.....	42
Caractéristiques climatologiques et hydrométriques du barrage de Foug el Khorza .....	44
Gestion actuelle du barrage .....	48
Ajustement des précipitations maximales journalières a la loi deGumbel.....	49

## Chapitre IV : Méthodologie de Box et Jenkins

IV. 1. Introduction.....	60
IV. 2. Historique .....	60
IV. 3. La méthodologie de Box & Jenkins.....	60
III. 4. Notions sur la méthode de Box et Jenkins.....	62
IV. 4.2. Stabiliser des processus TS.....	62
IV.5. Test des composantes saisonnières et tendancielles .....	63
IV.5.1. Test de Fisher:.....	63
2. Test de Dickey-Fuller:.....	65
IV.7. Les étapes de Box et Jenkins :.....	68
IV.7.1. Identification du modèle:.....	68
IV.7.3. 2. Tests sur les résidus :.....	72
IV.7.3.3. Test de Box-Ljung :.....	72
IV.7.3. 4. Test de normalité :.....	73
IV.7.3. 5. Test de Durbin et Watson :.....	74
IV.7.4. La prévision :.....	74
Conclusion.....	77

## Chapitre V : APPLICATION DE BOX JINKINS

1-présentation du logiciel R .....	79
1-1-Installation du système de base .....	79
1-2-gestion des packages.....	79
1-3--manipulation.....	81
1.4-La précision des modèles.....	105
<b>Conclusion générale</b> .....	115
<b>Références bibliographiques</b> .....	118
<b>ANEXES</b> .....	119

# *List des abréviations*

<b>ADF</b>	Augmented Dickey Fuller
<b>AIC</b>	Akaike
<b>AR</b>	AutoRégressif
<b>ARMA</b>	Auto Régressif
<b>DF</b>	Dickey Fuller.
<b>DS</b>	Differency Stationnary
<b>DW</b>	Durbin Watson.
<b>FAC</b>	Fonction d'AutoCorrélation
<b>FAP</b>	Fonction d'Autocorrélation Partielle
<b>MA</b>	Moyens Mobile
<b>SC</b>	Schwarz
<b>TS</b>	Trend Stationnary.



# *List des figures*

<b>Figure I.1</b> :Structure d'un réseau de neurons .....	14
<b>Figure II.1</b> La démarche de la Modélisation.....	23
<b>Figure II.2</b> :Une série chronologique qui suit un modèle additif.....	27
<b>Figure III.1</b> :Localisation du barrage de Boukerdane.....	42
<b>Figure III.2</b> :barrage de Boukerdane.....	43
<b>Figure III.4</b> :Bassin versant du barrage de boukerdane .....	44
<b>Figure III.5</b> :Volumes moyens annuels évaporés .....	47
<b>Figure III.6</b> :Volumes moyens annuels des fuites.....	45
<b>Figure III.7</b> :Ajustement à la loi de Gumbel.....	53
<b>Figure III.8</b> :Ajustement à la loi de Gumbel (HydroLab).....	57
<b>Figure IV.1</b> ::procédure itérative de Box-Jenkins .....	62
<b>Figure IV.2</b> : <i>Résumé de la stratégie de test de Dickey-Fuller</i> .....	67
<b>Figure IV.4</b> : Synthèse de Box et Jenkins.....	76
<b>Figure V.1</b> :Cran .....	81
<b>Figure V.2</b> :Installes de Package .....	82
<b>Figure V.3</b> Procédure de déconstruction de la méthode Box-Jenkins.....	84
<b>Figure V.4</b> le graph de la série apport.....	86

<b>Figure V.5 :</b> Le corrélogramme(acf)de la série apport .....	86
<b>Figure V.6 :</b> Le corrélogramme partiel(pacf) de la série apport .....	87
<b>Figure V.7:</b> le graphe de la série apport diff.....	87
<b>Figure V.8 :</b> Le corrélogramme(acf) de la série apport diff.....	91
<b>Figure V.9 :</b> Le corrélogramme partiel (pacf)de la série apport diff.de.....	91
<b>Figure V.10 :</b> le graphe de la série résiduelle pour le modèle MA.....	93
<b>Figure V.11 :</b> le corrélogramme de la série résiduelle pour le modèle MA.....	94
<b>Figure V.12 :</b> le corrélogramme partiel de la série résiduelle pour le modèle MA.....	94.
<b>Figure V.13 :</b> le graphe de la série résiduelle pour le modèle ARIMA.....	95
<b>Figure V.14 :</b> le corrélogramme de la série résiduelle pour le modèle ARIMA.....	95
<b>Figure V.15 :</b> le corrélogramme partiel de la série résiduelle pour le modèle SARIMA. ....	96
<b>Figure V.16 :</b> le graphe de la série résiduelle pour le modèle ARMA(1,0,0). ....	96
<b>Figure V.17 :</b> le corrélogramme de la série résiduelle pour le modèle ARMA(1,0,0). ....	97
<b>Figure V.18 :</b> le corrélogramme partiel de la série résiduelle pour le modèle ARMA(1,0,0)....	97
<b>Figure V.19 :</b> l'histogramme des résidus pour le modèle AR(1,0,0 :). ....	98
<b>Figure V.20 :</b> Graphe de la prévision.....	100

## *List des tableaux*

<b>Tableau I.1:</b> Différents types de modèle.....	07
<b>Tableau III.1 :</b> Caractéristiques générale de barrage de Boukerdane.....	43
<b>Tableau III.2 :</b> Stations pluviométriques au niveau du bassin versant.....	44
<b>Tableau III.3 :</b> Précipitations moyennes mensuelles.....	45
<b>Tableau III.4 :</b> Températures moyennes mensuelles.....	45
<b>Tableau III.5 :</b> Apports moyennes mensuelles.....	45
<b>Tableau III.6 :</b> Moyennes mensuelles des mesures d'évaporation au barrage.....	46
<b>Tableau III.7 :</b> Volumes moyennes mensuelles des fuites.....	47
<b>Tableau III.7 :</b> Volumes moyennes mensuelles des fuites.....	48
<b>Tableau III.8 :</b> Deman des mensuelles.....	48
<b>Tableau III.9 :</b> lâchers annuelles (en Mm <sup>3</sup> ) pour l'AEP et l'irrigation.....	48
<b>Tableau 9.1.</b> Précipitations maximales journalières (P <sub>max,j</sub> ).....	49
<b>Tableau III.11 :</b> Calcul de la variable de Gumbely.....	52
<b>Tableau III.12 :</b> Calcul des intervalles de confiance.....	55
<b>Tableau III.13 :</b> Calculs de la variable réduite y de Gumbel.....	55
<b>Tableau IV.1.:</b> Tableau résumant les propriétés des AR(p), MA(q) et ARMA(p,q).....	68
<b>Tableau IV.2:</b> spécification du modèle.....	70
<b>Tableau des apport</b> .....	82
<b>Tableau V.2 :</b> estimation des paramètres des modèles SARIMA .....	90
<b>Tableau V.4 :</b> la table des critères du choix de meilleur modèle.....	95
<b>Tableau V.5 :</b> la table des prévisions.....	96

## **Introduction générale**

Depuis toujours, l'homme a voulu prédire l'avenir, que ce soit pour prendre de meilleures décisions ou simplement pour satisfaire sa curiosité. Les premières tentatives étaient

Basées sur l'astrologie ou autres superstitions. Toute fois, ces méthodes de prédiction ne peuvent pas être considérées comme rigoureuses aux yeux de la science. Heureusement, la science a énormément évolué au cours des dernières décennies : les méthodes de prévisions qui en découlent ne font pas exception.

Une méthode de prévision très populaire est basée sur l'étude rigoureuse de séries Chronologiques (temporelles). Cette méthode permet de prédire, par exemple, de nombreux phénomènes naturels et financiers. Une série chronologique est constituée de valeurs observées à des intervalles de temps réguliers. Par exemple, les débits annuels sur un cours d'eau ou encore les valeurs mensuelles de titres boursiers sont des séries chronologiques.

A` la base, l'étude formelle des séries chronologiques consiste à trouver un modèle mathématique qui explique le mieux possible les données observées. A partir de ce modèle, il est possible de faire de la prévision, ce pendant, la justesse des prévisions dépend fortement de la qualité du modèle choisi. Il est donc primordial de trouver des modèles qui reflètent le mieux possible la réalité afin de minimiser les erreurs de prévisions.

Dans ce mémoire, nous proposons la méthodologie de Box-Jenkins pour la modélisation et la prédiction en se basant sur des modèles auto-régressif (AR), les modèles moyenne mobile (MA), les modèles ARMA,...etc, ce mémoire est constitué de la présente introduction, de trois chapitres. Il est organisé comme suit :

Le premier chapitre, consacré à la présentation des outils indispensables pour l'étude de la méthodologie de Box-Jenkins et d'introduire la présentation de deux types de processus stochastique linéaire, à savoir, les modèles linéaires stationnaires dans lesquels on trouve le modèle auto-régressif (AR), le modèle moyenne mobile (MA) et le modèle auto-régressif moyenne mobile (ARMA) et celui du modèle linéaire non stationnaire ou l'on présente le modèle auto-régressif moyenne mobile intégré (ARIMA) et le modèle auto-régressif moyenne mobile

intégré saisonnier (SARIMA) cette présentation suppose qu'on définisse essentiellement des notions sur l'analyse de ces modèles, et en particulier la notion de stationnarité, la causalité et l'invisibilité, ainsi que la fonction d'auto-corrélation (ACF) et la fonction d'auto-corrélation partielle (PACF) et dans 2eme et 3eme chapitre sur les série chronologique et la zone d'étude

Dans le 5ème chapitre, on a abordé l'estimation des paramètres du modèle ARMA en utilisant deux méthodes les plus usuelles, la première étant celle de Yule-Walker (Y-W) et la seconde est celle du maximum de vraisemblance (MMV), ainsi que la validation. En outre, on a destiné à la prévision en utilisant les processus AR, MA et ARMA.

Dans le troisième chapitre, on a proposé une application numérique sur des données réelles en utilisant le logiciel R.

# **Chapitre I:**

## **MODELISATION HYDROLOGIQUE**

## CHAPITRE I: MODELISATION HYDROLOGIQUE

### Généralité sur la modélisation hydrologique<sup>1</sup>

#### I.1. Introduction :

D'après le Dictionnaire de l'Environnement, le modèle est "une représentation simplifiée, relativement abstraite, d'un processus, d'un système, en vue de le décrire, de l'expliquer ou de le prévoir". La modélisation hydrologique est donc une représentation, partielle ou totale, du cycle de l'eau. Dans ce travail, on se limite à l'étude des modèles pluie-débit, qui représentent la transformation de la pluie en écoulement sur les surfaces continentales. Les applications de ces modèles sont multiples, et permettent de simuler l'impact d'aménagements anthropiques sur l'hydrologie d'un bassin versant (construction d'un barrage, imperméabilisation d'une zone par construction), de gérer les alertes de crue sur les bassins versants ou encore de reconstituer des chroniques de débits sur des bassins sur lesquels on ne dispose que de chroniques de pluies.

#### I.2. Les modèles hydrologiques

##### I.2.1. Définition

Un modèle hydrologique, ou modèle pluie-débit, est un outil numérique de représentation de la relation pluie-débit à l'échelle d'un bassin versant. Il permet de transformer des séries temporelles dérivant du climat d'un bassin versant donné (séries de précipitations et de températures par exemple, séries qui sont les entrées du modèle hydrologique) en une série de débits (sortie du modèle hydrologique).

De nombreux modèles hydrologiques ont été développés depuis la fin des années 1960. Le choix du type de modèle à utiliser dépend généralement de l'objectif de modélisation ainsi que des données d'entrées disponibles.

##### I.2.2. Principes des modèles hydrologiques

Les modèles hydrologiques permettent donc de transformer des séries décrivant le climat d'un bassin versant (typiquement des séries de précipitations et de températures) en une série de débits. Cette transformation est souvent divisée en deux parties :

- Une première partie souvent appelée la "production", qui consiste en la détermination de bilans d'eau à l'échelle du bassin versant. Ce bilan permet notamment de répartir

la pluie brute observée (la totalité de la pluie qui est tombée sur le bassin versant et qui est mesurée par un ou plusieurs pluviomètres) en pluie "nette" (la proportion de la pluie brute qui participe au débit du bassin versant étudié), en quantité d'eau évapotranspirée et en quantité d'eau stockée par le bassin versant.

- Une deuxième partie souvent appelée le "transfert" ou "le routage", qui consiste à répartir dans le temps la quantité d'eau partic dans le débit du bassin versant étudié.

### I.3. Utilisation des modèles hydrologiques

- Un modèle hydrologique peut être utilisé dans plusieurs contextes :
- Prédiction de crues.
- Prédiction d'étiages.
- Prédétermination des débits extrêmes.
- Études d'impact anthropique sur l'hydrologie (construction d'aménagements hydrauliques (barrage par exemple), changements d'occupation du sol, etc.).
- Études d'impact de changements climatiques sur l'hydrologie.
- Simulation de débits pour combler des lacunes dans les données et reconstituer des séries de débits historiques.

### I.4. Pourquoi un modèle hydrologique

Les modèles hydrologiques, nés bien avant l'avènement de l'informatique, ils sont utilisés soit pour effectuer des prévisions. Soit pour dimensionner des ouvrages ou enfin pour délimiter des zones inondables.

On peut distinguer trois types d'utilisation de la modélisation hydrologique à savoir:

1. La modélisation comme outil de recherche dans ce cas elle peut être utilisée pour interpréter des données mesurées. Différents scénarios de fonctionnement hydrologique de bassins versants peuvent alors être confrontés aux mesures.
2. La modélisation comme outil de prévision elle sert à anticiper les évolutions futures du débit il s'agit de l'utilisation opérationnelle la plus courante des modèles hydrologiques.



3. La modélisation comme outils d'extrapolation elle permettra alors la reconstitution des débits plausibles.

### I.5. Critères de qualité d'un modèle hydrologique

Lors de l'élaboration d'un modèle, l'ambition de l'hydrologue est d'obtenir un outil le plus performant possible. Il est de coutume d'apprécier la qualité d'un modèle et ses aptitudes selon certains critères tels que:

- **Précision:** Il doit être capable de reconstituer au mieux les débits observés.
- **Robustesse:** Il doit pouvoir garder une bonne précision même dans des conditions extrêmes.
- **Généralité:** Il doit pouvoir simuler aussi bien les crues que les étiages.
- **Universalité:** Il doit pouvoir être réutilisé pour des bassins versants très différents.
- **Richesse :** Il doit pouvoir intégrer un grand nombre de caractéristiques permettant de différencier le comportement d'un bassin à un autre.
- **Adaptabilité :** Il doit pouvoir être corrigé pour tenir compte de nouvelles données sur le bassin modélisé. En réalité, il est pratiquement impossible de satisfaire tous ces critères simultanément. L'objectif de l'utilisateur du modèle définira donc ceux qu'il faudra privilégier par rapport aux autres.

### I.6. Caractéristiques intrinsèques d'un modèle hydrologique

Chaque modèle traite le problème avec un point de vue qui dépend de sa finalité. Cependant, quelques caractéristiques fondamentales concernant le découpage dans le temps et dans l'espace se retrouvent dans tous les modèles

- **Découpage dans le temps** Celui-ci est souvent fait par intervalles égaux. On différencie trois pas de temps :
  - Pas de temps de fonctionnement ou de calcul Il correspond à celui des données d'entrée. C'est donc le plus fin et celui auquel l'on se réfère pour un modèle donné.
  - Pas de temps de calcul c'est celui utilisé pour le calcul des différents critères statistiques et

pour le calage des paramètres du modèle. Il correspond généralement à pas de temps des grandeurs sorties (débits).

- **Pas de temps d'utilisation** Il peut différer des deux précédents et dépend essentiellement de l'utilisation que l'on fait du modèle.

L'utilisation de plusieurs pas de temps, mensuel, journalier et par événements, permet de mieux appréhender les phénomènes physiques. D'après Ibiza (1988), la description phénoménologique fine et complète n'est possible qu'aux petits pas de temps inférieurs à l'heure. À pas de temps journalier, les différents écoulements rapides sont déjà regroupés. On ne possède donc plus d'informations sur la forme de l'averse et l'on doit se baser sur une substitution par une forme moyenne d'averse pour un climat donné. Au pas de temps mensuel, la détermination des écoulements se base sur des considérations encore plus globales et les circonstances particulières de chaque événement sont plus apparentes. Le nombre de paramètres augmente au fur et à mesure que la description des phénomènes devient plus précise donc que le pas de temps s'affine.

- **Découpage dans l'espace:** Le modèle s'applique à un bassin versant d'une superficie donnée. Ce bassin peut être considéré comme une unité ponctuelle : le modèle est donc global, ou comme une juxtaposition de sous-bassins globaux entre lesquels s'effectuent des transferts : le modèle est alors distribué

### **I.7. Classification des modèles hydrologiques**

Il existe, dans la littérature de nombreux modèles hydrologiques de bassins versant ayant, chacun leurs spécificités et leurs domaines d'application et peuvent être fonction de plusieurs caractéristiques<sup>2</sup>.

Une fois le modèle choisi, il faut évaluer sa capacité à représenter la réalité, ceci se fait le plus souvent en comparant les résultats du modèle avec les observations, c'est une étape très délicate., le tableau suivant propose de classer les différents types des modèles selon leurs caractéristiques.

Tableau(I.1): Différents types de modèle.<sup>3</sup>

Caractéristiques	Type de modèle
Selon la nature de la variable	Modèle déterministe (variable non aléatoires) Modèle stochastique (variable aléatoires)
Traitement de l'espace (homogénéité du bassin)	Modèle globale Modèle semi-distribué (ou semi-global) Modèle spécialisé
Fondement théorique et approche de développement	Modèle empirique Modèle conceptuelle Modèle à base physique
L'évolution temporelle	Modèle cinématique (descriptifs) Modèle dynamique (explicatifs)
Fonction objectif	Modèle de simulation Modèle de gestion

## I.7.1. Modèles stochastiques

La modélisation stochastique peut être envisagée dans le cas où il y a des incertitudes sur les données ou parfois même sur les processus mis en jeu. Les modèles stochastiques utilisent des variables aléatoires pour représenter l'incertitude de processus et génèrent des résultats différents à partir d'un ensemble de données d'entrée et de valeurs de paramètres lorsqu'ils s'exécutent dans des conditions identiques « vues de l'extérieur ».<sup>4</sup>

Toutefois cette approche, qui semble par ailleurs le meilleur moyen de caractériser la variabilité des grandeurs, nécessite la connaissance des lois de probabilités plus courantes pour la variabilité considérée ou au moins de leurs premiers moments.

## I.7.2. Modèles déterministes

Un modèle est dit déterministe (par opposition à stochastique) si aucune de ses grandeurs n'est considérée comme aléatoire, c'est-à-dire résultant soit de l'observation soit de grandeurs reconstituées. La plupart des modèles hydrologiques sont déterministes. Ces modèles associés à chaque jeu de variable de forçage, de variables d'état et de paramètres, une valeur de réalisation unique des variables de sortie (il s'agit essentiellement des débits simulés à l'exutoire d'un bassin

<sup>3</sup>:A. MARRE 1987, « Le Tell Oriental Algérien de Collo à la Frontière Tunisienne

versant).

### I.7.3. Modèles boîtes noires

(«Black Box» ou «Data-Driven» en anglais)

- On ne cherche pas à identifier ou à comprendre les mécanismes.
- On ajuste des fonctions de transfert entre variables d'entrée et variables de sortie.

#### Avantages

- Réseau de mesures minimal (longues séries temporelles, mais en peu de points).
- Nombre d'outils disponibles (régression, réseaux de neurones, Fourier, ondelettes, théorie du chaos).
- Modèles très rapides (important pour le temps réel !).

#### Inconvénients

- Productivité parfois douteuse (Ex. Fourier pour la prévision des débits).
- Les paramètres de la fonction de transfert n'ont généralement pas de signification physique.
- Si la physique change, il faut tout recommencer de zéro.
- Que faut-il intervenir l'intelligence artificielle on peut citer le réseau de neurones.
- Les algorithmes génétiques, le data Manning, neuroflou, logique floue, ARIMA, SARIMA... etc.

### I.7.4. Modèles à base physique

Le modèle à base physique est basé uniquement sur des équations de la physique, et ne comportant idéalement aucun paramètre. Il n'existe pas de modèle à base physique au sens strict en hydrologie. L'importance de l'hétérogénéité spatiale dans la réponse hydrologique des bassins versants rend cependant difficile voire impossible l'utilisation de tels modèles. La précision spatiale des données disponibles en particulier concernant les types de sols et leurs profondeurs n'est pas suffisante. Dans la pratique, les profondeurs et les conductivités moyennes des sols représentatifs de sous parties du bassin versant doivent être évaluées par calage.

### I.7.5. Modèles paramétriques

Les modèles paramétriques sont les modèles incluant des paramètres dont la valeur doit être estimée par calage.

### **I.7.6. Modèles empiriques :**

Les modèles empiriques reposent sur les relations observées entre les entrées et les sorties du système considéré. Ils expriment la relation entre variables d'entrée et de sortie du système (relation pluie-débit) à l'aide d'un ensemble d'équations développées et ajustées sur la base des données obtenues sur le système. Un modèle empirique ne cherche pas à décrire les causes du phénomène hydrologique considéré ni de ni à expliquer le fonctionnement du système, le système est vu comme une boîte noire.

### **I.7.7. Modèles analytiques:**

Ce sont des modèles pour lesquels les relations entre les variables de sortie et les variables d'entrée ont été établies par analyse de séries de données mesurées. L'exemple type est celui des modèles linéaires : les paramètres de ces modèles sont liés aux coefficients de corrélation entre les variables. Notons que l'analyse des données peut conduire au choix de relations non linéaires entre les variables.

### **I.7.8. Modèles conceptuels:**

Les modèles conceptuels considèrent en général le bassin versant, après quelques simplifications du cycle de l'eau, comme un ensemble de réservoirs interconnectés.

Ce type de modèle reproduit donc au mieux le comportement d'un système, plutôt qu'il n'apporte d'explications causales sur son comportement.

### **I.7.9. Modèles globaux**

Dans un modèle global le bassin est considéré comme une entité unique. Des relations empiriques (issues de l'expérience) relient les entrées et les sorties sans chercher à s'approcher d'une loi physique. Les modèles globaux offrent à l'utilisateur un choix très attractif, car ils présentent une structure très simplifiée, ils ne demandent pas trop de données, faciles à utiliser et à calibrer. La représentation du processus hydrologique est très simplifiée. Ils peuvent souvent mener à des résultats satisfaisants, et spécialement si l'objectif majeur est la prévision d'une crue.

### I.7.10. Modèles Spatialisés

Actuellement plusieurs modèles spatialisés correspondent des différentes écoles hydrologiques sont en phase avancée de développement. En principe, Les modèles spatialisés sont des modèles qui utilisent des entrées et des sorties où les caractéristiques des bassins versants sont distribuées dans l'espace. La spatialisation peut être arbitraire ou basée sur des divisions morphologiques naturelles (découpage en sous bassins) ou hydrologiques (aires contributives). Nous pouvons classer les modèles spatialisés en trois grands styles :

1. Modèles conceptuels spatialisés ou semi-spatialisés
2. Modèles physiques spatialisés
3. Modèles physiques conceptuels semi-spatialisés.

### I.7.11. Modèles conceptuels spatialisés ou semi-spatialisés

Les modèles conceptuels spatialisés ou semi-spatialisés représentent un grand progrès sur les modèles globaux quand il s'agit d'analyser le fonctionnement interne d'un bassin. Le bassin versant est discrétisé en nudités spatiales (mailles sous bassins) considérées comme homogènes, qui se vident les unes dans les autres de l'amont en aval. Ainsi, on a la possibilité de tenir compte de la répartition spatiale des facteurs et de suivre la genèse et la propagation des débits à l'intérieur du bassin. C'est le cas du modèle CEQUEAU et du modèle HEC-HMS.

### I.7.12. Modèles physiques spatialisés

Les modèles à base physique spatialisés sont des modèles qui décrivent les mécanismes internes d'un système (bassin versant) ayant comme base les lois de la mécanique, de la physique, de la thermodynamique, etc. Depuis un temps, ces modèles sont indépendants de tout calage parce que leurs paramètres sont mesurables. Ils permettent une description théorique unifiée de la plupart des flux observés dans un bassin versant et servent à modéliser les principaux processus hydrologiques comme:

- L'écoulement de surface à partir d'équations de Saint-Venant
- L'écoulement en milieu saturé à partir d'équations de Darcy
- L'écoulement en milieu non saturé à partir d'équations de Richards
- L'évapotranspiration à partir d'équations de conservation ou de relations entre flux.

A partir d'un découpage de l'espace en mailles de taille adaptée au problème à traiter, ils simulent les diverses composantes du cycle de l'eau sur chaque maille (Ruissellement, infiltration, évapotranspiration) et convolent les transferts de maille à maille, jusqu'à un réseau hydrographique constitué. Ces modèles sont complexes à mettre en œuvre et exigent d'importantes quantités de données. Ils sont bien adaptés à la simulation de la diversité d'un bassin.

Cependant, ces modèles ne peuvent traiter les grands bassins en raison de leur grande hétérogénéité morphologique et météorologique. Parmi ces modèles on peut citer le modèle SHE.

### I.7.13. Modèles physiques conceptuels semispatialisés

Pour dépasser les limites de chacune des approches précédentes (modèles conceptuels trop peu réalistes, modèles à base physique trop complexes), il est intéressant d'essayer une modélisation hydrologique qui peut être:

- Abas physique, fondée sur les processus réels mais simplifiés
- Semi-spatialisée, fondée sur une discrétisation en unités relativement homogènes, qui permettent de tenir compte des variabilités partiales de la structure du bassin versant.

La méthode de discrétisation spatiale varie d'un modèle à l'autre : mailles carrées, sous bassins versants, éléments de versant, plans versant et canal, unités Hydrologiques unitaires contributives (modèle TOPMODEL).

### I.7.14. Modèle pluie-débit

Il existe un grand nombre de modèles pluie-

débit, qui vont ainsi simuler un débit à l'exutoire d'un bassin versant, en utilisant la précipitation et l'évapotranspiration potentielle (ETP) comme données d'entrées. Plusieurs auteurs ont classé les modèles en trois catégories<sup>5</sup>:

- Les modèles conceptuels considèrent le bassin versant, comme un ensemble de réservoirs interconnectés.
- Les modèles empiriques de type boîte noire, cherchent à reproduire la dynamique des variables de sortie sans tenter de décrire les processus élémentaires.
- Les modèles distribués à base physique cherchent à décrire l'ensemble des processus élémentaires

sen tout point du bassin versant.

La notion de modèles physiques se oppose à celle de modèles conceptuels, ainsi que l'on oppose modèles distribués à modèles globaux au niveau de la description spatiale du bassin versant.

Il n'y a pas de conflit entre ces différents modèles qui ont été développés à des fins différentes : les modèles physiques et distribués pour une étude plus fine des processus hydrologiques, et les modèles conceptuels pour une détermination des paramètres moyens sur le bassin étudié et des débits à l'exutoire uniquement.

### ❖ Classification des modèles pluie-débit

Les modèles diffèrent suivant la nature du modèle mathématique sous-jacent, le degré d'abstraction du modèle, la nature des unités élémentaires ou encore la nature du processus représenté.

- **La nature du modèle mathématique :**

- sous-jacent. Il peut être déterministe ou probabiliste (stochastique). Les modèles statistiques ont une relation qui lie les variables d'entrée du modèle aux variables de sortie est phénoménologique ou probabiliste (stochastique). Les modèles statistiques ont un passage de la dynamique de la crue en cours.

- **Le degré d'abstraction du modèle :**

On parle de **modèles à base physique** lorsque les phénomènes sont décrits à l'aide des lois de la mécanique des fluides ou de la physique des écoulements dans les sols : loi de Richards pour le transfert en milieu non saturé, loi de Darcy pour le transfert en milieu saturé, loi de Barré-Saint-Venant pour le ruissellement de surface. Ces modèles ne nécessitent pas a priori de phase de calage de leurs paramètres puisqu'ils utilisent des paramètres mesurables sur le terrain.

Leur application en hydrologie opérationnelle sur des bassins versants d'une certaine taille est limitée par la méconnaissance de la variabilité spatiale des étapes de calage de leurs paramètres s'avère souvent nécessaire. Parmi les plus utilisés on trouve SHE<sup>6</sup> ou encore HEC-HMS<sup>7</sup>. Les modèles à base physique procèdent souvent d'une approche ascendante : on part des équations valables à l'échelle du volume élémentaire (équations issues de la mécanique des milieux



Continus le plus souvent) et on intègre ces équations à l'échelle du versant ou à l'échelle du bassin versant.

On parle de modèles empiriques lorsqu'on cherche à reproduire le comportement global du bassin à partir d'une étude systémique de la relation pluie/débit à l'échelle du bassin, sans faire d'hypothèse a priori sur les processus qui rentrent en jeu. Dans cette catégorie, on peut ranger les modèles GR<sup>8</sup>.

Le modèle SCS,

ou encore tous les modèles régressifs. L'utilisation de ces modèles empiriques nécessite une phase de calage des paramètres, puis que ceux-ci ont une signification qui n'est pas reliée a priori à des grandeurs mesurables du bassin versant.

Bien identifiés de leurs paramètres. Des relations empiriques peuvent ainsi souvent être établies entre les valeurs scalées des paramètres et un descripteur du bassin versant.

Entre les modèles empiriques et les modèles à base physique, on trouve les modèles conceptuels qui s'étudient comme les modèles empiriques (approche systémique) le modèle est le modèle TOPMODEL, qui s'appuie sur un fonctionnement du bassin mais dont l'approche est à base physique. Une bonne illustration de cette catégorie de modèles est le modèle TOPMODEL, qui s'appuie sur un fonctionnement du bassin basé sur la réaction de nappes préexistantes qui réagiront plus ou moins suivant leur position sur le bassin, l'état initial du bassin et

Les caractéristiques locales de pente et de perméabilité.

### I.7.15. Modèle débit-débit

En plus de ces modèles pluie-débit il existe des modèles débit-débit, ces modèles vont simuler un débit sortant à un point aval en fonction d'un débit entrant à un point amont. On peut classer en deux grands domaines ces modèles<sup>9</sup>

- Les modèles globaux réalisent le transfert de l'écoulement d'un point amont à un point aval. Ils ne permettent l'estimation du débit qu'au point aval.
- Les modèles distribués prennent explicitement en compte la variabilité spatiale des processus, des variables d'entrée, des conditions aux limites et des caractéristiques du bassin versant. Ils permettent donc l'estimation du débit (et de la hauteur d'eau) en tout point du réseau hydrographique.

Ces modèles débit-débit négligent en général la modélisation des apports latéraux, en le traitant principalement ou deux formes: ponctuels ou uniformément réparties. Nous proposerons ici une approche permettant de traiter la quantification des apports latéraux via une modélisation pluie-débit du bassin intermédiaire alimentant le bief dérivé considéré.

Dans notre cas, cette modélisation simplifiée des écoulements présente deux avantages:

La possibilité d'être utilisée sur des biefs sans données topographiques et la possibilité d'être appliquée sur des biefs de longueur importante.

### I.5.15 Les processus ARCH<sup>10</sup> :

- **Les processus ARCH linéaires :**

*L'hypothèse fondamentale sous-tendant les ARCH linéaires est la symétrie des spécifications quadratiques de la variance conditionnelles des erreurs.*

#### **Modèle ARCH(q):**

Rappelons qu'un modèle de type ARCH(q) consiste à spécifier la variance des erreurs de façon autorégressive conditionnellement à son information passée.

Une telle spécification peut généralement

#### **1) Modèle ARCH Généralisé (GARCH(p,q)) :**

Ils'agit d'un modèle ARCH généralisé; car, dans ce type de modèles, l'information – plus

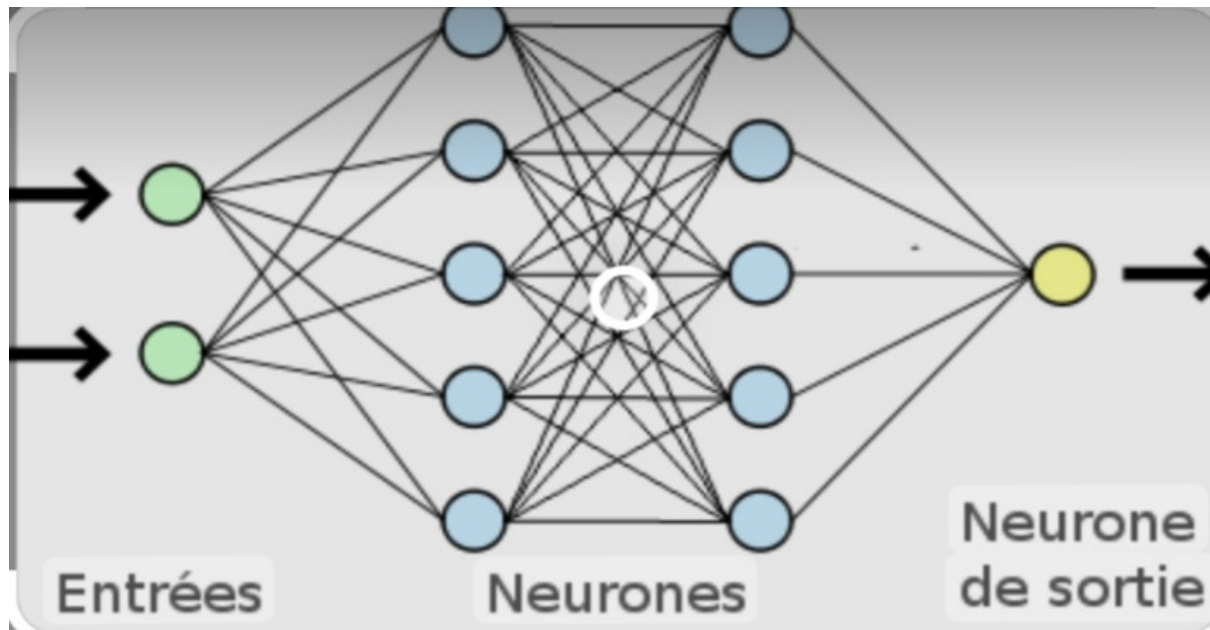
éloignée dans le passé – sur la variance conditionnelle des erreurs est prise en compte dans la spécification de celle-ci en y incluant les valeurs des variances décalées. Selon BOLLERSLEV (1986), le modèle GARCH(p,q)

Précisons aussi que nous allons estimer six modèles (ARCH, GARCH, TGARCH, EGARCH, ARCH-M et GARCH-M) parmi lesquels nous retiendrons le plus performant, et allons procéder comme suit:

- (i) Représentations graphiques des séries «EURUS» et «DEURUS»;
- (ii) Statistiques descriptives et test de normalité de la série «EURUS»;
- (iii) Etude de la volatilité de la série «EURUS» ;
- (iv) Estimation des modèles et recherche du modèle optimal
- (v) Prévision

**I.7.16. Le modèle des réseaux de neurones :**

Les réseaux de neurones sont des modèles simples' représentant le fonctionnement du système nerveux. Les unités de base sont les neurones . Ils sont généralement organisés en couche comme l'illustre la figure ci-dessous. *Figure 1. Structure d'un réseau de neurones*



*figure 2-2:* Le modèle des réseaux de neurones

Le noeud Réseau de neurones est un modèle simplifié de la manière dont le cerveau humain traite les informations. Le fonctionnement de ce modèle repose sur la simulation d'un grand nombre d'unités de traitement interconnectées, qui sont en quelque sorte des versions abstraites de nos neurones.

Ces unités de traitement sont organisées en couches. Il existe généralement trois types de couche dans un réseau de neurones : une couche d'entrée dans laquelle les unités représentent les champs

d'entrée, une ou plusieurs couches cachées, ainsi qu'une couche de sortie dans laquelle des unités représentent les champs cibles. Les unités sont reliées entre elles par des connexions de puissance (ou de pondération) différentes. Les données d'entrée sont présentées dans la première couche et les valeurs transmises entre les neurones d'une couche à l'autre. Le résultat final est obtenu à partir de la couche de sortie.

Lors de son apprentissage, le réseau procède à l'examen de tous les enregistrements afin de générer des prévisions et modifie les pondérations lorsque l'une de ses prévisions s'avère incorrecte. Ce processus se répète plusieurs fois et le réseau continue d'améliorer ses prévisions jusqu'à ce que l'un des critères d'arrêt soit atteint.

Au début, tous les coefficients de pondération sont aléatoires et les réponses en provenance du réseau risquent de ne pas avoir de sens. Le réseau apprend à travers l'apprentissage. Les exemples dont le résultat est connu sont présents à plusieurs reprises au réseau et les réponses qu'il donne sont comparées aux résultats connus. Les informations de cette comparaison sont réacheminées via le réseau, modifiant progressivement les coefficients de pondération. Au fur et à mesure de l'apprentissage, les résultats connus répliqués par le réseau sont à chaque fois plus précis. Lorsque l'apprentissage est terminé, le réseau peut être appliqué à d'autres observations pour lesquelles le résultat est inconnu.

### Conclusion

D'après le Dictionnaire de l'Environnement, le modèle est "une représentation simplifiée, relativement abstraite, d'un processus, d'un système, en vue de le décrire, de l'expliquer ou de le prévoir". La modélisation hydrologique est donc une représentation, partielle ou totale, du cycle de l'eau. Dans ce travail, on se limite à l'étude des modèles pluie-débit, qui représentent la transformation de la pluie en écoulement sur les surfaces continentales.

Les applications de ces modèles sont multiples, et permettent de simuler l'impact d'aménagement anthropique sur l'hydrologie d'un bassin versant (construction d'un barrage, imperméabilisation d'une zone par construction), de gérer les alertes de crue sur les bassins versants ou encore de reconstituer des chroniques de débits sur des bassins sur lesquels on ne dispose que de chroniques de pluies.

# **Chapitre II**

## **LES SERIESCHRONOLOGIQUES**

## Chapitre II Les séries chronologiques

### 1. Introduction sur les séries chronologiques:<sup>11</sup>

En hydrologie, les données constituent souvent des séries d'observations sur une ou plusieurs variables faites à différents pas de temps réguliers, et présentent généralement des liaisons internes.

#### I.1 Définition :

On appelle série chronologique (série temporelle ou encore chronique) une série statistique de variables  $(t, y_t)$  avec  $t \in T$ ,  $T = \{t_1, t_2, \dots\}$  où la première composante du couplet est le temps et la deuxième composante est une variable numérique  $y_t$  prenant ses valeurs aux instants. Suivant la nature du problème étudié la chronique peut être journalière (cours d'une action en bourse), mensuelle (consommation mensuelle de gaz), trimestrielle (nombre trimestriel de chômeurs), annuelle (chiffre annuel des bénéfices des exportations etc). L'étude des séries chronologiques est utile lorsque l'on cherche à analyser, comprendre ou encore prévoir un phénomène évoluant dans le temps. [16]

#### I.2. Types importants de séries chronologiques:

Il existe deux types de séries chronologiques selon que l'intervalle de temps  $T$  est discret ou continu.

- **Série discrète:**

Dans le cas où l'intervalle de temps  $T$  est discret, c'est-à-dire  $t=1, 2, \dots$ , Test le  $(T \subseteq \mathbb{N})$  nombre d'observations équidistantes, répétées avec une certaine fréquence, par exemple, une fois par an, par mois, par jour, etc., la série chronologique est dite discrète.

- **Série continue:**

Dans le cas où l'intervalle de temps  $T$  est continu,  $(T \subseteq \mathbb{R}, \mathbb{R}^+)$  (nombres réels, réels positifs), comme c'est le cas du domaine de la physique par exemple, où la variable peut être observée de façon continue, la série chronologique est dite continue. [12]

## **Chapitre II    les séries chronologiques**

---

### **Analyse des séries chronologiques:**

C'est un problème analogue à celui de l'estimation des propriétés d'une population à partir d'un échantillon tiré de cette population, puisque le principe est d'estimer et reconstituer, à partir d'un échantillon donné (la série d'observation) des propriétés du processus générateur inconnu. Pour mieux illustrer le principe de l'analyse des séries chronologiques, on est obligé de préciser son évolution historique, qui explique précisément les éléments les plus importants dans l'analyse; après avoir défini cette notion et précisé ces objectifs.

### **Définition de l'analyse des séries chronologiques:**

C'est une présentation d'un état de l'art des différentes techniques utilisées pour l'analyse exploratoire des séries chronologiques, ceci dans le cadre d'un regroupement de données. En effet, l'analyse des séries est un domaine disposant d'une longue histoire qu'ils conviendrait de prendre en compte afin de déterminer le rythme des techniques d'analyse, voir aussi de nouvelles modes de représentation.

Ces dernières décennies, l'augmentation des capacités de calcul a ouvert de nouveaux horizons, il ne s'agit plus seulement de présenter les données, mais de découvrir des structures internes. L'analyse exploratoire des séries chronologiques a pour but de permettre des données présenter en trop grand nombre pour être appréhendées individuellement. Les statistiques ont longtemps été freinées par le fait qu'elles reposaient sur le calcul, alors qu'aucune technologie vraiment efficace pour effectuer ces calculs n'a été disponible avant longtemps.

Récemment, l'ordinateur a permis une révolution des techniques statistiques. Assister par ordinateur, l'analyse statistiques permet de découvrir des informations en essayant au maximum des hypothèses techniques. Des algorithmes d'optimisation permettant de rechercher un maximum de vraisemblance pour trouver les modèles les plus adaptés aux données.

On commencera par donner un aperçu historique sur l'évolution de l'analyse des séries chronologiques, pour ensuite expliquer la problématique de cette analyse, les particularités des séries numériques en tant qu'objet d'analyse, et les formes sous lesquelles sont disponibles les données avant de pouvoir conditionner aussi fortement les techniques qui seront applicables.

## Chapitre II les séries chronologiques

### I.2.1 Objectifs et problèmes de l'analyse des séries chronologiques :

#### Modélisation :

On peut illustrer les étapes de modélisation par le schéma suivant:

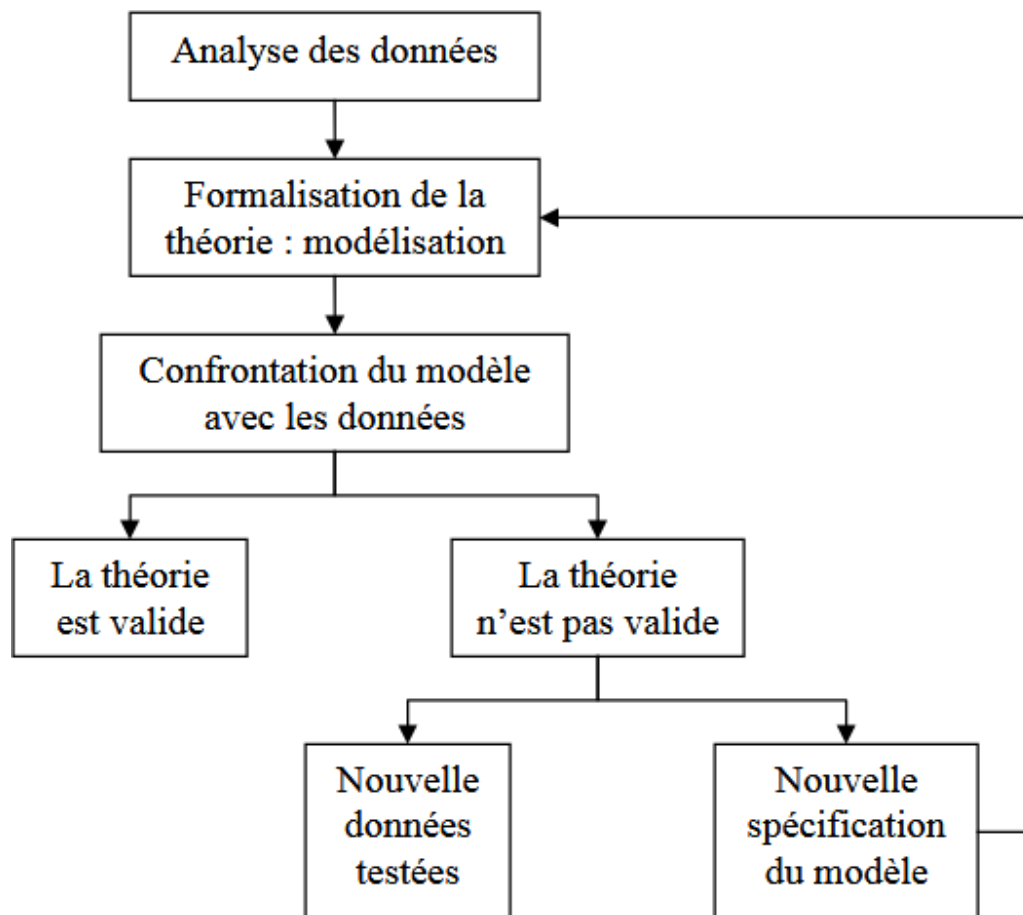


Figure II. 1 : La démarche de la modélisation.

### I.2.2. Problèmes d'analyse des séries chronologiques:

#### a) Prédiction :

Étant donné des observations  $Y_1, Y_2, \dots, Y_T$  on désire évaluer une valeur non observée  $Y_{T+1}$ .

La prédiction peut être :

- ponctuelle :  $\hat{Y}_T(h)$ .
- sous forme d'un intervalle de prédiction :  $[\hat{Y}_{T,1}(h), \hat{Y}_{T,2}(h)]$ .



## **Chapitre II      les séries chronologiques**

---

La qualité de la prévision dépend de la façon dont évolue la série. Plus la série est fonction « régulière » du temps, plus il sera facile de prévoir. Les prévisions seront par exemple bonnes pour beaucoup de variables économiques en période de décroissance, lorsque l'allure générale de la série est linéaire ou exponentielle.

En revanche, les diverses méthodes de prévision ne permettent pas de prévoir un changement d'évolution dû à une modification des structures économiques, que rien dans le passé ne permettait de supposer.

Les méthodes servant à la prévision peuvent également être utilisées pour évaluer une valeur passée de la variable, appeler valeur ajustée au lieu de prévision qui peut par exemple être utile dans le cas de données manquantes et donc permettre de compléter une série chronologique, ainsi ils peuvent aussi servir pour mesurer l'effet d'un phénomène accidentel (grève, phénomène climatique exceptionnel) ; la valeur ajustée donne une idée de la valeur qu'aurait dû prendre la variable, si le phénomène n'avait pas eu lieu.

### **b) Décomposition de la série chronologique:**

Les problèmes de décomposition les plus fréquents sont les suivants : estimer la tendance, enlever la tendance, estimer les variations saisonnières et enlever les variations saisonnières. Dans une période de décroissance, beaucoup de variables économiques ont des évolutions à moyen terme (ou tendance) analogues. Ces variables sont donc toutes fortement corrélées entre elles, sans que ceci exprime une quelconque liaison à caractère explicatif entre ces variables. Pour voir si de telles liaisons existent, il peut être utile d'enlever cette tendance.

### **c) Détection et modélisation des ruptures :(changement structurel):**

Suite à des changements politiques économiques ou à des modifications profondes des relations structurelles entre variables, les séries peuvent dans certains cas présenter des ruptures soit de niveau, soit de pente. Il est évidemment important d'essayer de prévoir ces dates de rupture ou, si ceci se révèle impossible, de repérer leur existence le plus rapidement possible.

## **Chapitre II    les séries chronologiques**

---

### **d) Étude de la dynamique entre plusieurs variables:**

On distingue deux points :

- **Causalité:**

C'est une approche qui consiste à analyser l'évolution jointe des divers variables d'intérêt est à examiner si parmi celle-ci certaines sont fixées. Une telle approche ne peut évidemment être développée qu'un modèle joint de l'ensemble des variables peut être écrit.

- **Décalage temporels :**

Cela consiste à mettre en évidence un retard temporel entre deux séries chronologiques, cela en mesurant l'importance du retard par une pente positive ou négative du diagramme représentatif de chaque série.

### **e) Séparation entre relations de court terme et relations de long terme :**

Ce point peut être résumé par le « concept de cointégration » : Les influences entre variables prennent plus ou moins de temps, sont plus ou moins persistantes. L'un des problèmes importants de la macro économétrie est de séparer ces relations persistantes (dites de long terme) de celles qui ne le sont pas. Ces dernières s'interprètent souvent en termes d'ajustement.

### **f) Étude des anticipations et contrôle:**

L'analyse des séries temporelles permet aussi d'étudier comment les agents réagissent vis-à-vis du temps. Ainsi on dispose souvent de données jointes sur les valeurs prises par certaines variables économiques et sur les prévisions de ces valeurs faites avant. On peut se demander s'ils anticipent bien le niveau de la variable, son évolution ; on peut essayer de comprendre comment ils calculent implicitement leurs anticipations.  
[18]

- ✓ **Les composantes d'une série chronologique :**

Dans un premier temps, l'examen graphique de la série étudiée permet de dégager, un certain nombre de composantes fondamentales de l'évolution de la grandeur étudiée. Il faut

## Chapitre II Les séries chronologiques

alors analyser ces composantes, en les dissociant les unes des autres, c'est-à-dire en considérant une série qui résulte de la combinaison de différentes composantes.

### La tendance ou « trend »:

Notée  $f_t$ , censée décrire le mouvement de long terme, de fond ou encore la structure de long terme du phénomène. Ce mouvement est traditionnellement représenté par des formes analytiques simples.

### La composante cyclique:

Notée  $C_t$  qui regroupe des variations à période moins précise autour de la tendance. Ces phases durent généralement plusieurs années, mais n'ont pas de durée fixe. Sans informations spécifiques, il est généralement très difficile de dissocier la tendance du cycle. Dans la plupart des travaux sur les séries temporelles la tendance regroupe aussi la composante cyclique.

### 1.2.3. La composante saisonnière ou variation saisonnière

Notée  $S_t$  sont des variations qui se reproduisent périodiquement à des moments bien déterminés et qui sont liées au rythme imposé par les variations météorologiques des saisons (production agricole, consommation de gaz, . . .), ou encore par des activités économiques et sociales (fêtes, vacances, solde, le ramadhan, etc.)

### 3. La composante résiduelle

Notée  $\varepsilon_t$ . Elle rassemble tout ce que les autres composantes n'ont pu expliquer du phénomène observé. Elle contient donc de nombreuses fluctuations, en particulier accidentelles, dont le caractère est exceptionnel et imprévisible, (catastrophes naturelles, grèves, guerre, etc.).

Comme par hypothèse de type d'événement censé être corrigé, le résidu présente en général une allure aléatoire plus ou moins stable autour d'une moyenne. [7]

### Schémas de décomposition d'une série chronologique

La technique de décomposition d'une série chronologique, repose sur un modèle qui l'autorise. Ce modèle porte le nom de schéma de décomposition. Il en existe essentiellement deux grands types:

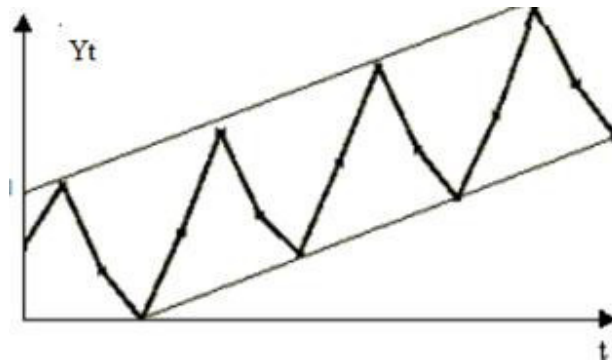
## Chapitre II les séries chronologiques

---

- **Schéma additif**

Dans un modèle additif, on suppose que les 3 composantes : tendance, variations saisonnières et variations accidentelles sont indépendantes les unes des autres. On considère que la série  $Y_t$  s'écrit comme la somme de ces 3 composantes :

$$Y_t = C_t + S_t + \varepsilon_t \quad (I.1)$$



**Figure I.2 : Une série chronologique qui suit un modèle additif.**

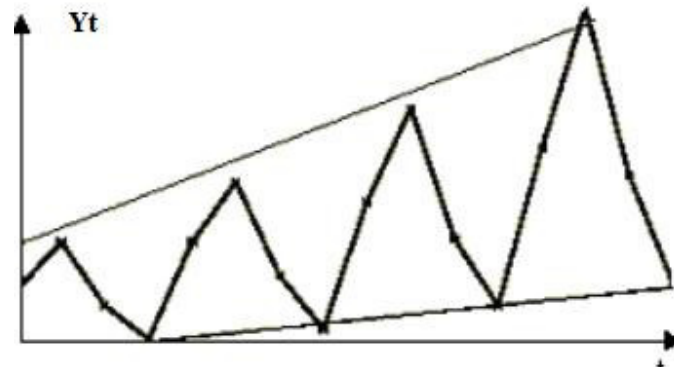
Graphiquement, l'amplitude des variations est constante autour de la tendance.

### **Schéma multiplicatif**

Graphiquement, l'amplitude des variations (saisonnières) varie. On suppose que les variations saisonnières et les variations accidentelles dépendent de la tendance et on considère que  $Y_t$  s'écrit de la manière suivante :

$$Y_t = C_t \times S_t \times \varepsilon_t \quad (I.2)$$

## Chapitre II les séries chronologiques



**Figure I. 3: Une série chronologique qui suit un modèle multiplicatif.**

Ce modèle multiplicatif se ramène à un modèle additif en considérant la série  $\ln(Y_t)$ :

$$\ln Y_t = \ln C_t + \ln S_t + \ln \varepsilon_t; \quad (I.3)$$

### Caractéristiques d'une série temporelle

#### Moyenne et variance

Soit une série temporelle stationnaire  $Y_t$ ,  $t=1 \dots T$ .

Les expressions de la moyenne et de la variance sont:

- Moyenne :  $E(Y_t) = \frac{1}{T} \sum_{t=1}^T Y_t \quad (I.4)$

- Variance :  $V(Y_t) = \frac{1}{T} \sum [Y - E(Y)]^2 \quad (I.5)$

#### Fonction d'auto covariance

Soit  $Y_t$  un processus aléatoire de variance finie. On appelle fonction d'autocovariance  $\gamma_h$  de  $Y_t$  la fonction :

$$\gamma_h = \text{cov}(Y_t, Y_{t+h}) = E \left[ [Y_t - E(Y_t)](Y_{t+h} - E(Y_{t+h})) \right] \quad (I.6)$$

La fonction d'autocovariance fournit des informations sur la variabilité de la série et sur les liaisons temporelles entre les composantes de la série  $Y_t$ .

La fonction d'auto covariance d'un processus  $X_t$  stationnaire vérifié les propriétés suivantes

:

## Chapitre II les séries chronologiques

$$\gamma_0 = \text{cov}(Y_t, Y_t) = [Y_t - E(Y_t)]^2 = V(Y_t) = \sigma_x^2 \quad (I.7)$$

$$|\gamma_h| \leq \gamma_0 \quad (I.8)$$

$$\gamma_h = \gamma_{-h}: \text{Fonction paire} \quad (I.9)$$

### La fonction d'autocorrélation

Soit  $Y_t$  un processus stationnaire. On appelle fonction d'autocorrélation la  $\rho_h$  fonction :

$$\rho_h = \frac{\gamma_h}{\gamma_0}, h \in Z \quad (I.10)$$

La fonction d'autocorrélation permet de mesurer les liaisons temporelles entre les différentes composantes de  $Y_t$ . En effet:

$$\rho_h = \frac{\text{cov}(X_t, X_{t+h})}{\sigma_{X_t} \sigma_{X_{t+h}}} = \frac{\gamma_h}{\sqrt{\gamma_0} \sqrt{\gamma_0}} = \frac{\gamma_h}{\gamma_0} \quad (I.11)$$

La fonction d'autocorrélation d'un processus  $Y_t$  stationnaire vérifie les propriétés suivantes:

$$\rho_0 = 1 \quad (I.12)$$

$$|\rho_h| \leq \rho_0 \quad (I.13)$$

$$\rho_h = \rho_{-h}: \text{Fonction paire} \quad (I.14)$$

### L'autocorrélation partielle

L'autocorrélation partielle est définie comme la corrélation entre  $Y_t$  et  $Y_{t-h}$  corrigées de l'influence des autres variables  $Y_{t-h+i}$  (pour  $i < h$ ). Plus précisément, le coefficient de corrélation partiel entre  $Y_t$  et  $Y_{t-h}$  peut être défini comme le coefficient de corrélation entre le résidu de la Régression  $Y_t$  sur les variables  $Y_{t-h+i}$  (pour  $i < h$ ), et celui de  $Y_{t-h}$  sur ces mêmes variables.

### Objectifs de l'étude des séries chronologiques

Les séries chronologiques sont le plus simple exemple d'une thématique plus large : l'estimation et la prédiction des processus stochastiques, c.à.d. des familles des variables aléatoires  $Y_t$ . Leur particularité, par rapport aux autres types de données statistiques, tient à la présence d'une relation d'antériorité qui ordonne l'ensemble des observations.

## Chapitre II les séries chronologiques

On se propose d'estimer la valeur de la variable  $Y_t$  en un point quelconque connaissant les valeurs  $Y_t$  aux points  $t_i$ , pour  $i=1, \dots, N$ . Le but principal est le choix d'un modèle raisonnable, qui permettra à partir des valeurs connues la prédiction des valeurs inobservables (comme les valeurs futures des séries chronologiques, ou moins accessibles physiquement, coûteuses, etc.) avec une élimination du bruit d'observation éventuel.

- **Description et modélisation :**

Le but est ici de déterminer les différentes composantes d'une série chronologique  $Y_t$ , en particulier, obtenir la série corrigée des variations saisonnières (dessaïsonalisation). Pour les séries stationnaires, on peut aussi chercher à modéliser la série à l'aide d'un modèle ARMA, par exemple dans le but de faire de la prédiction ou de la simulation.

- **Prediction :**

Sur la base des observations  $Y_1, \dots, Y_T$  le but est de faire une prédiction, à la date  $T$ , de la réalisation  $T+h$ , notée  $\hat{Y}(T+h)$ , de manière à ce que la somme des carrés des erreurs de prédiction soit minimale. Dans le cas des modèles ARMA, de nombreuses relations existent à fin de faire de la prédiction avec un intervalle de confiance.

- **Filtrage:**

C'est un lissage qui consiste à transformer une série chronologique de façon à détecter (pour éliminer ou au contraire conserver) certaines caractéristiques (composante saisonnière, observations aberrantes, ... etc.). Cette méthode permet aussi de détecter les ruptures au sein d'une série chronologique.

### Types de modèles ajustés aux séries chronologiques

Il existe deux catégories de modèles pour rendre compte d'une série temporelle. Les premiers considèrent que les données sont une fonction du temps ( $y = f(t)$ ). Cette catégorie de modèle peut être ajustée par la méthode des moindres carrés, ou d'autres méthodes itératives. L'analyse des modèles par transformée de Fourier est une version sophistiquée de ce type de modèle. Une seconde catégorie de modèles cherche à déterminer chaque valeur de la série en fonction des valeurs qui la précèdent ( $Y_t = f(Y_{t-1}, Y_{t-2}, \dots)$ ). C'est le cas des modèles ARIMA (Auto Regressive Integrated Moving Average). Cette catégorie de modèles a été popularisée

## Chapitre II Les séries chronologiques

---

et formalisée par Box & Jenkins (1970).

Autre que le choix de l'un ou l'autre type de modèle est surtout théorique : est-il raisonnable de penser que dans un phénomène donné, les points sont fondamentalement fonction des points précédents et de leurs erreurs, plutôt qu'un signal, périodique ou non, entaché de bruit.

On peut noter cependant que souvent, on a recours à l'analyse de variance pour traiter les séries temporelles. Or une des assumptions majeures de l'ANOVA est que les résidus des différentes mesures ne sont pas auto-corrélés. Ce n'est évidemment pas le cas si la performance à l'essai  $t$  est liée à la performance réalisée à l'essai  $t-1$ .

### Modèles déterministes :

C'est des modèles où la théorie des probabilités n'intervient pas. La série chronologique peut être considérée comme :

- Une fonction déterministe du temps :  $Y_t = f(t)$  (I.15)
- Ou une fonction de récurrence :  $Y_t = f(t, Y_{t-1}, Y_{t-2}, \dots)$  ..... (I.16)

Pourvue que la fonction  $f(\cdot)$  est si nécessaire, les valeurs passées de  $Y_t$  soient connues, il faut noter qu'un modèle déterministe permet une prédiction par faite

### Modèles stochastiques :

C'est des modèles où les éléments de la série chronologique sont considérés comme des variables aléatoires. Lorsqu'on considère une série  $(Y_t, t \in T)$  de variables aléatoires, on parle de processus stochastique (ou fonction aléatoire). La théorie des processus stochastiques est la base théorique de l'étude des modèles stochastiques. Il faut noter que c'est cette classe de modèles qui va être préconisée pour la modélisation des séries chronologiques abordées dans le contexte du présent travail.

### Description de la tendance :



## Chapitre II les séries chronologiques

La description initiale de la tendance repose sur l'interprétation de la représentation graphique de la série. On appelle tendance  $l$  (ou variation à long terme ou trend) de la série  $Y_t$  la série  $C_t$  résultant de la totalité des effets permanents auxquels est soumise la série  $Y_t$ .

Pour faire apparaître plus clairement la tendance, il faut atténuer la composante accidentelle. Nous utilisons pour cela les moyennes mobiles définies de la façon suivante:

- On appelle moyenne mobile centrée de longueur impaire  $l_i = 2k + 1$  à l'instant  $t$  la valeur moyenne  $mm_t$  des observations  $Y_{t-k}, Y_{t-k+1}, \dots, Y_t, \dots, Y_{t+1}, Y_{t+k}$

$$mm_t = (Y_{t-k}, Y_{t-k+1}, \dots, Y_t, \dots, Y_{t+1}, Y_{t+k}) / l_i \quad (I.17)$$

- On appelle moyenne mobile centrée de longueur paire  $l_p = 2k$  à l'instant  $t$  la valeur moyenne  $mm_t$  des observations  $Y_{t-k}, Y_{t-k+1}, \dots, Y_t, \dots, Y_{t+1}, Y_{t+k}$ : la première et la dernière étant pondérées par 0.5:

$$mm_t = (0.5Y_{t-k}, Y_{t-k+1}, \dots, Y_t, \dots, Y_{t+1}, 0.5Y_{t+k}) / l_p \quad (I.18)$$

Dans la première formule, le nombre de termes de la somme est égal à  $2k + 1$ : il s'agit bien d'une moyenne. Dans la seconde, la somme des coefficients est égale à  $2k$ , puisque le premier et le dernier terme sont égaux à 0.5: ils'agit d'une moyenne pondérée. Dans les deux cas, le nombre d'observations prise en compte avant l'instant  $t$  est égal au nombre d'observations prises en compte après l'instant  $t$ : c'est pour cela que les moyennes sont dites centrées.

### Définition des coefficients saisonniers

On sait que l'influence des variations saisonnières doit être neutre sur l'année et que ces variations ( $S_t$ ) se répètent théoriquement à l'identique de période en période. Dans toute série chronologique observée sur un cas réel, les variations saisonnières ne sont jamais identiques.

Donc, pour satisfaire aux exigences du modèle théorique, et pour pouvoir dire de la série réelle, à la place des ( $S_t$ ) observées, des variations périodiques identiques chaque année (mois par mois, ou trimestre par trimestre) qu'on appelle coefficients saisonniers. On les note  $S_j / j = 1$  à 12 pour des données mensuelles,  $j = 1$  à 4 pour des données trimestrielles.

### Méthode de calcul des coefficients saisonniers:

## Chapitre II Les séries chronologiques

La série  $S_t$  est observée sur années par période «p».  $p=12$  mois ( $j=1,2,\dots,12$ ) ou 4 trimestres ( $j=1,2,3$  ou 4). Les variations saisonnières  $S_j$  sont égales, par hypothèse du modèle additif à :

$$S_t = y_t - f_t \quad (I.19)$$

Nous obtenons donc  $n \times j$  valeurs de  $S_t$ , que nous pouvons s'écrire  $S_{ij}$ . On retiendra 12 valeurs de  $S_j$  (mois) ou 4 valeurs de  $S_j$  (trimestres) comme coefficients saisonniers, en calculant, mois par mois, ou trimestre par trimestre, la moyenne arithmétique des  $S_{ij}$ , sur l'ensemble des années, on obtient :

$$S_j = \frac{1}{n} \sum_{i=1}^n S_{ij} \quad (I.20)$$

La somme sur l'année de ces coefficients saisonniers  $S_j$  devrait tout logiquement être égale à 0. En fait, bien souvent, les approximations des calculs conduisent à un résultat légèrement différent. Dès lors, dans le cas où la somme des  $S_j$  est différente de 0, on calcule un coefficient correcteur «p» qui est la moyenne des  $S_j$  sur l'année.

$$P = \frac{1}{12} \sum_{j=1}^{12} S_j \quad (I.21)$$

Et l'on retient en définitive, comme coefficient saisonnier corrigé la valeur :

$$S_j^* = S_j - P \quad (I.22)$$

Le principe théorique selon lequel la moyenne (ou la somme) des coefficients saisonniers est égale à zéro est respecté par les  $S_j^*$  (coefficients saisonniers corrigés). Pour le modèle multiplicatif la moyenne des coefficients saisonniers doit être égale à 1.

### I.3.11. Série désaisonnalisée ou série CVS

Nous appelons série désaisonnalisée ou série corrigée des variations saisonnières notée série CVS, la série chronologique  $y_t^*$  à laquelle on a enlevé les variations saisonnières.

Dans le cas du modèle additif :

$$\text{La série désaisonnalisée est : } y_t^* = y_t - S_t \quad (I.23)$$

## Chapitre II les séries chronologiques

Dans le cas du modèle multiplicatif:

$$\text{La série désaisonnalisée est : } y_t^* = \frac{y_t}{s_t} \quad (I.24)$$

La particularité de la série CVS est que les données de  $y_j^*$  sont directement comparables : on a enlevé l'effet des saisons et donc le caractère propre de chaque mois on peut par exemple comparer les données d'un mois de janvier et celle d'un mois de juillet. A partir de la série CVS, on peut réévaluer la tendance par ajustement ou lissage (moindres carrés ou moyennes mobiles sur  $y_{j^*}$ ), afin d'avoir une meilleure estimation de la tendance.

### Opérateurs sur les séries chronologiques

Opérateur retard et avance Pour formaliser le déplacement dans le temps de la série chronologique, on définit une application, qui, à partir d'une observation prise à une date donnée nous permet d'exprimer les observations passées ou les valeurs futures.

Ainsi on introduit l'opérateur retard noté B (comme Backward) comme l'application

$$Y_t \rightarrow BY_t = Y_{t-1} \quad (I.25)$$

On peut alors établir une formule de récurrence selon

$$BY_t = Y_{t-1}, B^2Y_t = Y_{t-2}, \dots, B^nY_t = Y_{t-n} \quad (I.26)$$

De façon analogue on définit l'opérateur avance noté F (comme Forward) tel que:

Propriétés de l'opérateur retard

$$B^0Y_t = Y_t \quad (I.27)$$

Si:

$$Y_t = c \quad \forall t \in T (c \in \mathbb{R}) \text{ alors } B^jY_t = B^j c = c \quad \forall j \in T \quad (I.28)$$

$$B^i B^j Y_t = B^{i+j} Y_t = Y_{t-(i+j)} \quad \forall (i, j) \in T^2 \quad (I.29)$$

$$(B^i + B^j)Y_t = B^i Y_t + B^j Y_t = Y_{t-i} + Y_{t-j} \quad \forall (i, j) \in T^2 \quad (I.30)$$

### Autre opérateurs

- Opérateur de différence première (ordinaire)  $\nabla = (1 - B)$  (I.31)
- Opérateur de différences saisonnières  $\nabla^2 = (1 - B^2)$  (I.32)

## Chapitre II Les séries chronologiques

---

### III-Prévision d'une série chronologique :

#### 1-Processus stochastique

D'un point de vue sémantique, un processus est un ensemble de phénomènes, conçus comme une chaîne causale progressive à travers le temps, du point de vue mathématique, c'est la solution d'une équation composée de variables indicées par le temps ; si l'évolution temporelle d'une ou plusieurs variables n'est pas totalement déterministe (présence d'une ou plusieurs variables aléatoires) nous parlons alors de processus stochastique.

#### 2-Processus stationnaire

Une classe très spéciale des processus stochastiques dite processus stationnaire démarre de l'hypothèse que le processus repose sur un équilibre statistique. Ceci implique que la série ne comporte ni de tendance ni de saisonnalité. Le plus simple processus stationnaire en analyse de séries temporelles est appelé : processus bruit blanc.

#### 3-Processus bruit blanc

Un processus de bruit blanc est une suite de variables aléatoires  $(\varepsilon_t)$  de même distribution (mêmes moyenne et variance) et mutuellement indépendantes. C'est un terme qui a été emprunté aux physiciens dans le domaine de la lumière. Il fait illusion au spectre de la lumière blanche.

Dans le cas de l'analyse de séries chronologiques un bruit blanc  $\{\varepsilon_t, t \in T\}$  est tel que :

$$E(\varepsilon_t) = 0, \quad \forall t \in T \quad (I.33)$$

$$Var(\varepsilon_t) = \sigma^2 \quad \forall t \in T \quad (I.34)$$

$$Cov(\varepsilon_t, \varepsilon_{t+h}) = \begin{cases} \sigma^2 & \text{si } h = 0 \\ 0 & \text{sinon} \end{cases} \quad (I.35)$$

#### 4-Processus AutoRégressif d'ordre AR(p) :

Dans le processus autorégressif d'ordre  $p$ , l'observation présente  $Y_t$  est générée par une

## Chapitre II les séries chronologiques

moyenne pondérée des observations passées jusqu'à la p-ième période sous la forme suivante :

$$Y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t \quad (I.36)$$

L'équation (I.36) peut aussi s'écrire:

Avec:

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \quad (I.37)$$

Sont des paramètres réels à estimer  $\varepsilon_t \in Z$  est un bruit blanc. Le degré est appelé le degré d'autocorrélation, il indique la profondeur de la mémoire

### • Processus Moyenne mobile d'ordre q, MA(q)

Dans le processus moyenne mobile d'ordre q, chaque observation  $Y_t$  est générée par une moyenne pondérée d'aléas jusqu'à la q-ième période.

$$Y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} \dots - \theta_q \varepsilon_{t-q} \quad (I.38)$$

L'équation (I.38) peut aussi s'écrire:

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q \quad (I.39)$$

Où  $\theta_1, \theta_2, \dots, \theta_q$  sont des paramètres réels à estimer et  $\varepsilon_t \in Z$  est un bruit blanc. Le degré est nommé le degré de la moyenne mobile.

Processus ARMA(p,q)

Les processus ARMA sont donc représentatifs d'un processus généré par une combinaison des valeurs passées et des erreurs passées.

Un processus stationnaire  $Y_t$  suit un ARMA(p, q) s'il vérifie l'équation stochastique suivante:

$$\phi(B)Y_t = \theta(B)\varepsilon_t \quad (I.40)$$

Avec:

## Chapitre II Les séries chronologiques

$$- \phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \quad (I.41) \text{ Le polynôme autorégressif d'ordre } p.$$

$$- \theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q \quad (I.42) \text{ Le polynôme moyen mobile d'ordre } q.$$

$\varepsilon_t, t \in \mathbb{Z}$  Processus bruit blanc  $(0, \sigma_\varepsilon^2)$ .

$$\text{Nous avons: } ARMA(p,0) = AR(p) \text{ et } ARMA(0,q) = MA(q) \quad (I.43)$$

### Les processus mixtes intégrés ARIMA(p,d,q)

Les séries chronologiques en hydrologie ne sont pas toujours stationnaires. La condition de stationnarité des processus présentés jusque-là n'est évidemment pas toujours convenable.

On peut intégrer certains types de non-stationnarité en élargissant le processus ARMA.

Ces élargissements consistent en une série d'opérations préalables visant à éliminer les facteurs de non-stationnarité. Le principe de ces modèles est de se ramener à un modèle ARMA (p, q) en opérant des différenciations sur la série.

On appelle processus ARIMA(p,d,q) un processus pour lequel le processus différencié  $D^d$  d'ordre d.

$$Y_t = (1 - B)^d Y_t, \quad t \in \mathbb{Z} \quad (I.44)$$

vérifier une relation de récurrence ARMA (p, q):

$$Y_t = \sum_{i=1}^p \phi_i Y_{t-i} - \sum_{i=0}^q \theta_i \varepsilon_{t-i} \quad \forall t \in \mathbb{Z} \quad (I.45)$$

Où:  $\phi_i$  et  $\theta_i$  sont des réels et  $\varepsilon_t$  est un bruit blanc

Le processus ARIMA appliqué à est équivalent au processus ARMA appliqué sur le processus différencié  $Y_t$ :

$$Y_t = ARMA(p, q) \leftrightarrow X_t = ARIMA(p, d, q) \quad (I.46)$$

Le «I» d'ARIMA signifie «integrated» comme réciproque de la différenciation.

Evidemment, le degré  $d$  n'est généralement pas connu. Pour le déterminer on peut agir par tâtonnements ou avoir recours à des tests de stationnarité: puisque un processus ARMA(p,q) est stationnaire, on cherche  $d$  tel qu'on puisse accepter l'hypothèse de stationnarité pour le processus

$$X_t = \nabla^d Y_t \quad (I.47)$$

## Chapitre II Les séries chronologiques

Dans la pratique  $d=1$ , et très rarement  $d=2$ .

### I.3.13.2. Conditions de stationnarité et d'invisibilité pour AR(p) et MA(q)

1. Un processus AR est toujours inversible. Il est stationnaire lorsque les racines de (L) en module sont strictement supérieures à 1.
2. Un processus MA est toujours stationnaire. Il est inversible si les racines de (L) sont en module strictement supérieures à 1.

Nous pouvons maintenant combiner les deux modèles, moyennement mobile et autorégressif.

### I.5.14. Les processus aléatoires non stationnaires

Les processus stochastiques non stationnaires sont caractérisés par des propriétés stochastiques qui évoluent en fonction du temps. On distingue deux types de processus stochastiques non stationnaires : une non stationnarité de nature déterministe (TS) et une non stationnarité de nature stochastique (DS)

#### Description des processus TS

Un processus  $(y_t, t \in \mathbb{Z})$  présente une non stationnarité de type déterministe TS (Trend Stationary), s'il peut se décomposer en une somme de deux fonctions :

$$Y_t = \varepsilon_t + f_t \quad (I.48)$$

Tel que  $\varepsilon_t$  est un processus stationnaire de type ARMA.

$f_t$  est une fonction polynomiale du temps. Polynomiale de degré 1, ils écrivent :

$$Y_t = \alpha_0 + \alpha_1 t + \varepsilon_t \quad (I.49)$$

Où  $\alpha_0, \alpha_1 \in \mathbb{R}$

$\varepsilon_t$  : Bruit blanc

Les caractéristiques de ce processus sont :

$$E(y_t) = \alpha_0 + \alpha_1 t + \varepsilon_t \quad (I.50)$$

$$V(y_t) = \delta^2 \quad (I.51)$$

$$Cov(y_t, y_{t-h}) = 0 \quad \forall h \neq 0 \quad (I.52)$$

La non stationnarité de ce processus est dû au fait que son espérance dépend du temps.

## Chapitre II les séries chronologiques

La méthode pour stationnariser un processus TS est d'estimer les coefficients  $\alpha_0, \alpha_1$  par MCO (Moindre Carrés Ordinaires) et de retrancher de la valeur de  $y_t$  la valeur estimée de sa moyenne  $\hat{\alpha}_0 + \hat{\alpha}_1 t$ .

- **Description d'un processus DS sans dérive ; [11]**

Soit le processus DS sans dérive (AR(1)) :

$$y_t = y_{t-1} + \varepsilon_t \quad (I.53)$$

$$(1 - B)y_t = \varepsilon_t \quad (I.54)$$

La racine du polynôme caractéristique  $(1-B)$  est égale à 1. On dit que le processus  $y_t$  a une racine unité, il est donc non stationnaire. Ce processus DS sans dérive peut se réécrire sous la forme :

$$y_t = y_{t-1} + \varepsilon_t \quad (I.55)$$

$$y_t = y_{t-h} + \varepsilon_t + \dots + \varepsilon_t \quad (I.56)$$

$$y_t = y_0 + \sum_{j=1}^t \varepsilon_j \quad (I.57)$$

Un processus DS sans dérive est un processus stationnaire en moyenne et non stationnaire en variance.

### DS avec dérive : [11]

Considérons un processus DS avec dérive :

$$y_t = \mu + y_{t-1} + \varepsilon_t = t\mu + y_0 + \sum_{j=1}^t \varepsilon_j \quad (I.58)$$

Un processus DS avec dérive est un processus non stationnaire en moyenne et en variance.

Ces moments évoluent en fonction du temps  $t$ . Un processus DS est un processus que l'on peut stationnariser par l'application du filtre aux différences :

$$x_t = x_{t-1} + \varepsilon_t \quad (I.59)$$



## Chapitre II Les séries chronologiques

---

$$(x_t - x_{t-1}) = \varepsilon_t \quad (I.60)$$

$$\nabla x_t = \varepsilon_t \quad (I.61)$$

- **PROCESSUS SARMA :**

Définition 2.7 Le processus  $(X_t)_{t \in \mathbb{Z}}$  satisfait une représentation d'un processus autorégressif moyennement bilésaisonnier, noté SARMA.

### Modèles linéaires non-stationnaires :

À la section précédente, quelques processus aléatoires stationnaires ont été présentés et étudiés. Toutefois, dans la réalité, il arrive souvent que des séries observées montrent un comportement non stationnaire. La classe des processus non stationnaire est relativement vaste, il existe différents types de non stationnarité, on présente deux classes des processus non stationnaires les processus  $TS$  et  $DS$  :

### 5- Mise en oeuvre sous R

Pour utiliser les fonctions spécifiques à l'étude des modèles ARCH et GRACH, il faut avant tout charger le package

tseries à l'aide de la commande `library(tseries)`.

La fonction `garch` permet d'estimer un GARCH $p,q$  : `serie <- garch(data, order=c(q,p))`

Parmi les sorties de cette fonction : `coef`, `residuals`, `fitted.values`.

La prédiction se fait de la même façon que pour les modèles de type ARIMA

### Conclusion :

L'étude des séries chronologiques, correspond à l'analyse d'observations régulièrement espacées dans le temps. Cette analyse est un outil statistique qu'utilise l'administrateur afin d'étudier les situations aléatoires qui évoluent dans le temps, ce qui aide les entreprises à représenter et à faire face afin d'obtenir un modèle et l'utiliser selon les objectifs recherchés ; comme la prévision ou le contrôle.

# **Chapitre III**

## **Présentation de La zone d'étude**

### Chapitre III:

#### Introduction :

Un barrage est un ouvrage d'Art placé en travers d'un cours d'eau, destiné à retenir et stocker de l'eau ou à la dériver.

Les barrages sont construits à d'autres fins : protection contre les crues, lutte contre l'incendie, navigation, création de polders, protection des estuaires contre les remontées d'eau salée depuis la mer, alimentation en eau potable, usages industriels, tourisme et loisirs. [10]

#### ✓ **Barrage de Boukerdane:**

Le barrage de Boukerdane est un barrage en terre, qui est située à l'ouest d'Alger dans la région de wilaya de Tipaza, il se trouve à 11 Km du bord de la méditerranée. Ce barrage a été construit en 1992 pour une capacité initiale de 105 Hm<sup>3</sup>.



Figure III.1 : Localisation du barrage de Boukerdane.



**Figure III.2 :** barrage de boukerdane.

Les caractéristiques du barrage de Boukerdane qui est de type terre zoné, sont récapitulées dans le tableau.

Les courbes caractéristiques du barrage sont données par les relations suivantes :

$$H = 75.65 + 1.93 * V - 0.05 * V^2 + 3 * 10^{-6} * V^4 \quad (1)$$

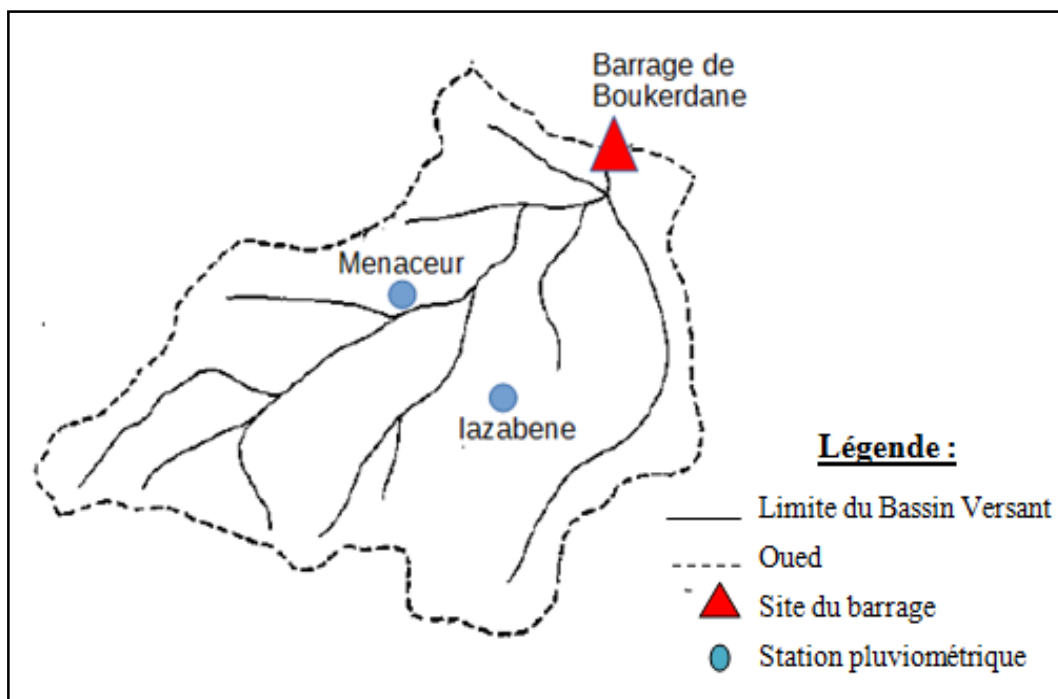
$$S = 0.28 + 0.104 * V - 0.001 * V^2 + 5 * 10^{-6} * V^3 \quad (2)$$

**Tableau III.1 :** Caractéristiques générale de barrage de Boukerdane.

Côte du niveau normale de la retenue	119,50 m.
Cote du niveau de plus hautes eaux	123,00 m.
Hauteur du barrage (Fondation du barrage)	15m
Longueur du barrage	68m
Epaisseur du barrage (au sommet)	3,25 Mm3
Epaisseur du barrage (à la base)	7,5 Mm3
Aire de la retenue au niveau normale	536ha
Capacité totale de la retenue	105 Mm3
Volume régularisable	49m

**Caractéristiques climatologiques et hydrométriques du barrage de Foum el Khorza :**

Le barrage de la Cheffia est implanté à l'exutoire est construit sur le lit de l'oued El-Hachem au confluent des deux principaux oueds : Menacer boukadir et fedjana, il draine un bassin versant d'environ 156 Km<sup>2</sup>de surface.



**Figure III.3 :** Bassin versant du barrage de Boukerdane.

Le bassin versant du barrage de boukerdane contient deux stations pluviométriques qui sont : la station de Menaceur et Iazabane, dont ces caractéristiques sont récapitulées dans le tableau.

**Tableau III.2 :** Stations pluviométriques au niveau du bassin versant.

Nom de la station	Code	X(km)	Y(km)	Z(m)
<b>Menaceur</b>	020303	458.25	354.45	250
<b>Iazabane</b>	020304	461.15	371.3	270

### ✓ Précipitations :

Les précipitations moyennes mensuelles sur le site du barrage pour une période allant de 1968 à 2009, sont récapitulées dans le tableau (III.3).

**Tableau III.3 :** Précipitations moyennes mensuelles.

Mois	sept	oct	nov	déc	janv	févr	mars	avr	mai	juin	juil	août	Total annuel
<b>Précipitation moyenne (mm)</b>	75,88	72,73	69,66	53,23	37,68	7,78	2,275	4,88	27,45	54,87	81,1	84,47	572,005

### ✓ La température

Le barrage de Boukerdane est influencé par un climat Méditerranéen, avec des valeurs minimales pendant les mois de Décembre et Janvier, correspond jusqu'à 17°C°. Tandis que les valeurs maximales coïncident avec les mois de Juillet et Août, correspond jusqu'à 35°C°.

**Tableau III.4 :** Températures moyennes mensuelles.

Mois	sept	oct	nov	déc	janv	févr	mars	avr	mai	juin	Juil	août	Totl annuel
<b>Température moyen (C°)</b>	22.5	17.1	14.2	11.5	10.8	11.2	12.3	14.4	17.3	21.4	24.4	24.5	201.6

### ✓ Apport :

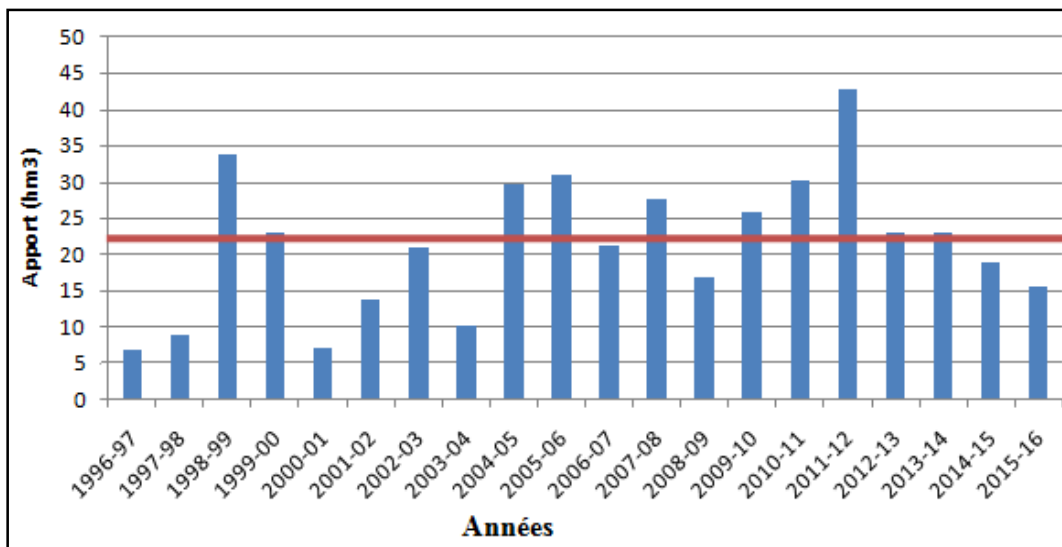
Les apports mensuels sont reconstitués à partir de l'équation du bilan hydrique pour une période de 1996 à 2016. Les apports moyens mensuels sont récapitulés dans le tableau.

**Tableau III.5 : Apports moyennes mensuelles.**

Mois	sept	oct	nov	déc	janv	févr	mars	avr	mai	juin	Juil	août	Total annuel
Apport moyen (Hm <sup>3</sup> )	0,22	0,34	1,53	2,38	2,57	4,34	5,36	2,33	1,93	0,49	0,17	0,12	19.94

Les apports moyens annuels sont représentés dans la figure (III.3).

**Figure III.4 : Apport moyen annuel.**



**Figure III.4 : Apport moyen annuel.**

D’après la figure, nous constatons que le nombre d’années sèches et humides sont égaux. Ainsi on peut remarquer deux périodes. Dans la première qui s’étale de 1996 à 2004, on dénombre mois d’années humides (2/8) que d’années sèches (6/8), ces dernières ont des valeurs plus inférieurs que celle de l’apport moyen annuel qui est de 21,61Mm<sup>3</sup>, où on enregistre que l’année 1996-1997 est la plus faible avec un apport de 7 Mm<sup>3</sup>.

Dans la deuxième période (de 2004 à 2016), on dénombre plus d’années humides (8/12) que d’années sèches (4/12). Les valeurs de l’apport sont proches de la moyenne excepte pour celle de l’année 2011-2012 où on enregistre un apport maximal de 43Mm<sup>3</sup>.

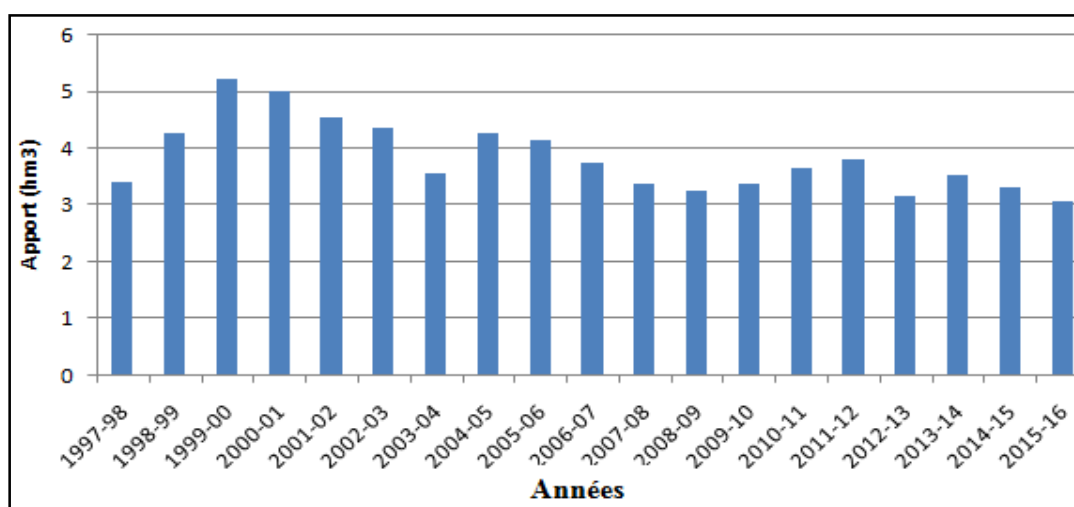
**Evaporation**

L’évaporation est mesurée à la station météorologique située au niveau du barrage. Le tableau ci-dessous décrit les moyennes mensuelles de l’évaporation observée depuis la mise en service du barrage (sur la période 1996-2010).

**Tableau III.6 :** Moyennes mensuelles des mesures d'évaporation au barrage.

Mois	sept	oct	Nov	déc	janv	févr	mars	avr	mai	juin	juil	août
<b>Evaporation (mm/jour)</b>	4.72	3.41	1.9	1.47	1.53	1.82	2.74	3.39	4.37	6.14	6.65	6.17

Le volume net moyen annuel perdu par évaporation est estimé à 3,9 hm<sup>3</sup>, qui représente environ de 18% de l'apport moyen annuel.



**Figure III.5 :** Volumes moyens annuels évaporés.

✓ **Infiltration :**

Depuis son entrée en exploitation, le barrage de Boukerdane a un problème des infiltrations où des résurgences sont apparues en aval, les volumes des infiltrations sont présentés dans le tableau.

**Tableau III.7 :** Volumes moyennes mensuelles des fuites.

Mois	sept	oct	Nov	déc	janv	févr	mars	avr	mai	juin	juil	août
<b>Volume des fuites (Hm<sup>3</sup>)</b>	0,39	0,4	0,41	0,4	0,5	0,44	0,39	0,38	0,35	0,35	0,38	0,38

Le volume moyen annuel perdu par les infiltrations est estimé à 4.45 hm<sup>3</sup>, qui représente environ de 20,6% de l'apport moyen annuel.



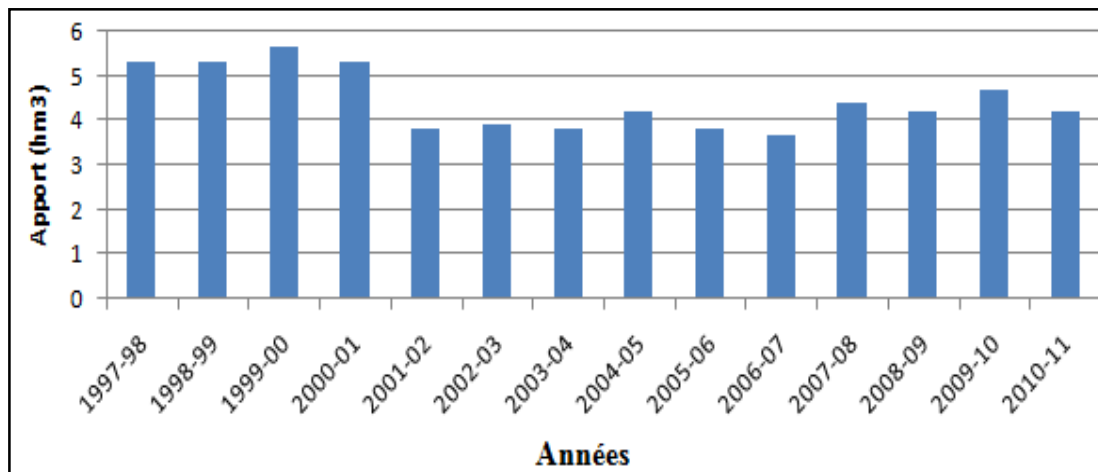


Figure III.6 : Volumes moyens annuels des fuites.

**Gestion actuelle du barrage :**

Le barrage de Boukerdane est destiné à l'alimentation en eau de trois secteurs :

- 1- L'Alimentation en Eau Potable de la ville de Tipaza et ces environs
- 2- L'irrigation de la vallée de l'oued EL Hachem et une partie de l'ouest de la Mitidja

Les demandes mensuelles (en hm3) pour l'AEP en 2002 ainsi que Les besoins en eau pour l'irrigation du futur périmètre du sahel Algérois ouest sont récapitulés dans le tableau (III.8).

Tableau III.8 :Demandes mensuelles.

Mois	sept	oct	Nov	déc	janv	févr	mars	Avr	mai	juin	juil	août	Totale annuelle
AEP	0,74	0,77	0,74	0,77	0,77	0,74	0,77	0,74	0,77	0,74	0,77	0,77	9
Irrig	1,7	0,3	0,07	0,02	0,04	0,12	0,77	1,7	1,7	2,97	4,53	4,12	18

À partir du tableau (III.9) qui mentionne les lâchers annuels pour l'AEP et l'irrigation, on peut remarquer une nette augmentation des lâchers pour l'AEP pour cause de l'accroissance de la population. Ainsi une variation des lâchers pour l'irrigation, qui sont directement liés aux volumes des apports annuels au barrage.

**Tableau III.9 :** lâchers annuelles (en Mm<sup>3</sup>) pour l'AEP et l'irrigation.

Années	Lâchers pour AEP	Lâchers pour Irrig	Volume demandée	Taux de satisfaction
2004-05	8,46	2,82	5	56,4
2005-06	7,89	5,43	7,15	76
2006-07	9,68	6,72	6,5	103,46
2007-08	9,70	5,48	4	137,12
2008-09	11,23	1,50	5	30,14
2009-10	12,26	2,07	3	69
2010-11	13,04	4,33	7	61,86
2011-12	12,91	8,09	7,5	107,87
2012-13	12,79	2,95	8	36,87
2013-14	12,98	3,9	8	48,75
2014-15	12,8	3,87	5,5	70,36
2015-16	13,24	5,08	9,5	53,47

Le taux de satisfaction globale est de 68,63%, où on enregistre des lâchers pour l'irrigation supérieurs à la demande, pour les années : 2006-2007, 2007-2008 et 2011-2012, cette dernière correspond à l'année d'apport maximal avec 43 Mm<sup>3</sup>, Alors que le taux de satisfaction pour l'année suivante atteint 36,87% malgré un apport de 23Mm<sup>3</sup>.

Le taux de satisfaction minimal (30,14%) est enregistré pour l'année 2008-2009 qui correspond à l'année de plus faible apport 17Mm<sup>3</sup>.

**✓ Ajustement des précipitations maximales journalières alaloide Gumbel :**

Soient les précipitations maximales journalière  $P_{max,j}$  enregistrées à la station de Ouled Mimoun dans la wilaya de Tlemcen, observée sur une période de 64 ans, on demande:

1. Faire l'ajustement des précipitations maximales journalières à l'aloide Gumbel (loi doublement exponentielle) ;
2. Calculer les quantiles de période de retour 10, 50 et 100 ans
3. Déterminez leur intervalle de confiance pour  $1-\alpha=95\%$

**Données :**

**Tableau9.1.**Précipitations maximales journalières ( $P_{max,j}$ )

Année	$P_{max,j}$ (mm)	Année	$P_{max,j}$ (mm)	Année	$P_{max,j}$ (mm)
1997	75,5	1946/47	56,5	1985/86	35,5
1998	30,2	1947/48	42,7	1986/87	31,8
1999	25	1948/49	66,9	1987/88	16,4
2000	52	1949/50	45,1	1988/89	45,4
2001	60	1950/51	59,5	1989/90	18,7
2002	35,6	1951/52	42,4	1990/91	51,8
2003	32,5	1952/53	24,5	1992/93	48,4
2004	28,5	1953/54	52,5	1994/95	32,3
2005	36	1954/55	51,9	1995/96	25,4
2006	35,8	1955/56	21	1997/98	18,6
2007	43	1956/57	57	1998/99	29,5
2008	48	1957/58	39	1999/00	35,8
2009	55	1958/59	54,3	2001/02	49,7
2010	35	1959/60	20	2002/03	29,1
2011	52,9	1960/61	30,4	2003/04	32,5
2012	39,3	1961/62	45,3	2005/06	22,8
2013	34,9	1979/80	74	2006/07	28,2
2014	36,3	1980/81	41,8	2007/08	33,7
2015	32,5	1982/83	25,8	2008/09	41,4
2016	54,2	1983/84	29,7	2009/10	37,5
2017	43,3	1984/85	33,8	2010/11	34
2018	31,4				

**Rapporte:**

La fonction de répartition de la loi de Gumbel ou loi doublement exponentiel est :

$$F(x) = e^{-e^{-y}} \text{ et } y = \alpha(x - x_0) \dots \dots \dots (III.1)$$

$F(x)$ : fréquence au dépassement de la valeur de  $x$ .

$\alpha, x_0$ : coefficients d'ajustement.

$x_0$ : paramètre de position (mode).

$\alpha$ : paramètre d'échelle différent de zéro.

Par un changement de variable  $y = \alpha(x - x_0)$ , la loi de Gumbel s'écrit:

$$F(x) = e^{-e^{-y}} \dots \dots \dots (III.2)$$

Avec,  $y = \alpha(x - x_0)$ , est la variable réduite de Gumbel

L'équation  $y = \alpha(x - x_0)$ , présentée sous la forme  $x = \frac{1}{\alpha} y + x_0$

est l'équation d'une droite qui représente la loi de Gumbel sur une échelle Gumbelienne. Ce papier

porte en abscisse une échelle arithmétique de la variable réduite  $y$ . L'ordonnée représente une échelle arithmétique de la variable étudiée soit  $P_{\max,j}$  et l'abscisse est une échelle arithmétique de la variable de Gumbel  $y$ .

• **1-Procédé d'ajustement**

- Classement des valeurs par ordre croissant en leur affectant un numéro d'ordre;
- Calculer la fréquence expérimentale par la formule de Hazen :

$$F(x) = (m - 0.5) / n \dots \dots \dots (III.3)$$

- Calculer les caractéristiques empiriques de la série.
- Calculer la variable de Gumbel pour chaque valeur observée :

$$y = -[\ln - \ln F(x)] \dots \dots \dots (III.4)$$

- Reporter les valeurs observées sur papier à échelle Gumbelienne.
- Calculer le coefficient de corrélation entre les valeurs observées et la variable de Gumbel.
- Si la corrélation est satisfaisante, calculer les paramètres d'ajustement de la droite de Gumbel. La droite de régression ou droite de Gumbel est:

$$P_{\max,j} = \frac{1}{\alpha} y + x_0 \dots \dots \dots (III.5)$$

Où :

$\frac{1}{\alpha}$  est la pente de la droite (Gradex) et  $x_0$  est l'ordonnée à l'origine (mode).

$Y$  : est la variable de Gumbel pour une probabilité donnée. Les paramètres

$\frac{1}{\alpha}$  et  $x_0$  sont déterminés par la méthode des moindres carrés.

**NB:** Il y a lieu de noter que plusieurs méthodes d'ajustement sont possibles et par miles plus classiques, on peut citer celle des moments, celle du maximum de vraisemblance, celle des moyennes pondérées ..... Ici, nous avons utilisé une méthode des moins classiques et celle des calculs des paramètres d'ajustement par la méthode des moindres carrés. Dans la pratique, les estimations sont très proches de celle des moments et les écarts sont quasi-négligeables en regard de sources d'incertitude.

- Tracer la droite de régression sur papier Gumbel. Pour cela, il suffit de prendre 2 valeurs

de probabilité pour lesquelles, il faut déterminer la variable  $y$  qu'il suffira de remplacer dans la droite de régression.

*Ex :*

$$F(x)=0,50 \quad \text{on a} \quad y=-[\ln -\ln (0,50)] = + 0,36\dots(III.6)$$

$$F(x)=0,80 \quad \text{on a} \quad y=-[\ln -\ln (0,80)] = + 1,49\dots(III.7)$$

En remplaçant, ces valeurs dans l'équation, nous retrouvons respectivement les précipitations de probabilité 50% et 80%.

- **2-La valeur extrême ou quantile peut être déterminée graphiquement :**

sur papier Gumbel ou analytiquement en utilisant simplement la droite de régression ou la droite de Gumbel.

- **3-Ajustement des précipitations annuelles à la loi de Gumbel :**

Les caractéristiques de la série pluviométrique initiale sont :

- Moyenne =  $P_{\max, j} = 39,52$  mm
- Ecart type =  $\sigma = 13,24$  mm.

Le tableau (III.11) résume les résultats et la représentation graphique est en figure (III.6).

**Tableau III.11 : Calcul de la variable de Gumbely.**

<b>M</b>	<b>P<sub>max,j</sub></b>	<b>F(x)</b>	<b>y= -[ln-lnF(x)]</b>	<b>m</b>	<b>P<sub>max,j</sub></b>	<b>F(x)</b>	<b>y= -[ln-lnF(x)]</b>
1	16,4	0,0078	-1,57940	33	36	0,5078	0,38913
2	18,6	0,0234	-1,32267	34	36,3	0,5234	0,43489
3	18,7	0,0391	-1,17637	35	37,5	0,5391	0,48139
4	20	0,0547	-1,06682	36	39	0,5547	0,52873
5	21	0,0703	-0,97637	37	39,3	0,5703	0,57702
6	22,8	0,0859	-0,89777	38	41,4	0,5859	0,62634
7	24,5	0,1016	-0,82728	39	41,8	0,6016	0,67683
8	25	0,1172	-0,76266	40	42,4	0,6172	0,72860
9	25,4	0,1328	-0,70251	41	42,7	0,6328	0,78180
10	25,8	0,1484	-0,64584	42	43	0,6484	0,83658
11	28,2	0,1641	-0,59195	43	43,3	0,6641	0,89311
12	28,5	0,1797	-0,54031	44	45,1	0,6797	0,95160
13	29,1	0,1953	-0,49051	45	45,3	0,6953	1,01227
14	29,5	0,2109	-0,44224	46	45,4	0,7109	1,07537
15	29,7	0,2266	-0,39524	47	48	0,7266	1,14121
16	30,2	0,2422	-0,34928	48	48,4	0,7422	1,21015
17	30,4	0,2578	-0,30419	49	49,7	0,7578	1,28259
18	31,4	0,2734	-0,25981	50	51,8	0,7734	1,35903
19	31,8	0,2891	-0,21601	51	51,9	0,7891	1,44008
20	32,3	0,3047	-0,17267	52	52	0,8047	1,52647
21	32,5	0,3203	-0,12967	53	52,5	0,8203	1,61914
22	32,5	0,3359	-0,08694	54	52,9	0,8359	1,71924
23	32,5	0,3516	-0,04437	55	54,2	0,8516	1,82833
24	33,7	0,3672	-0,00188	56	54,3	0,8672	1,94841
25	33,8	0,3828	0,04060	57	55	0,8828	2,08231
26	34	0,3984	0,08316	58	56,5	0,8984	2,23401
27	34,9	0,4141	0,12586	59	57	0,9141	2,40954
28	35	0,4297	0,16878	60	59,5	0,9297	2,61857
29	35,5	0,4453	0,21198	61	60	0,9453	2,87813
30	35,6	0,4609	0,25555	62	66,9	0,9609	3,22274
31	35,8	0,4766	0,29954	63	74	0,9766	3,74158
32	35,8	0,4922	0,34405	64	75,5	0,9922	4,84811

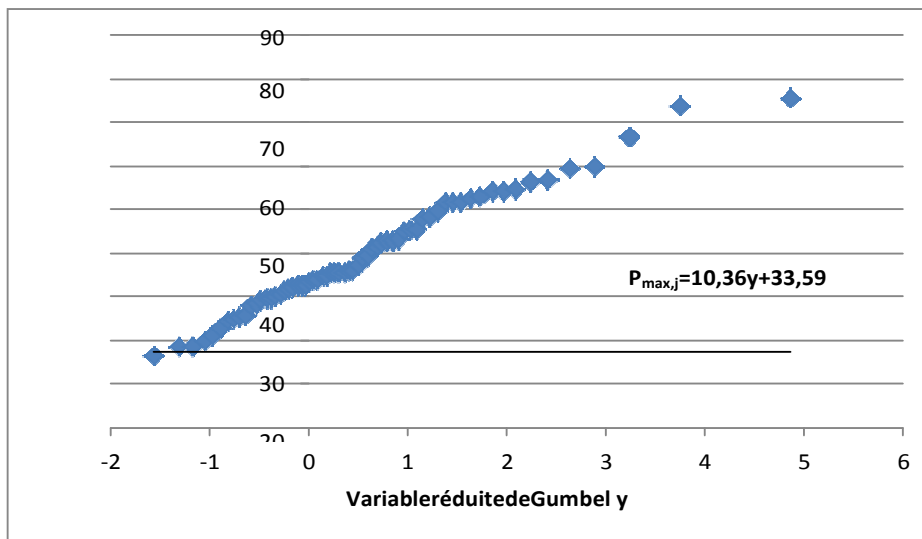


Fig III.7 :Ajustement à la loi de Gumbel.

• **4- Calcul des quantiles**

Une fois l'adéquation vérifiée, il faut procéder au calcul des quantiles en appliquant l'équation de la droite de Gumbel, soit :

$$P_{max,j,p\%} = 10,36y + 33,59 \dots (9.8)$$

Pour T= 10 ans;  $F(x)=0,10=10\%$  d'où  $y = -[\ln(-\ln 0,90)] = 2,25$  Soit:  $P_{max,j,10\%} = (10.36 * 2,25) + 33.59 = 57\text{mm}$

Pour T= 50 ans;  $F(x)=0,02=2\%$  d'où  $y = -[\ln(-\ln 0,98)] = 3,902$  Soit:  $P_{max,j,2\%} = (10.36 * 3,902) + 33.59 = 74\text{ mm}$

Pour T= 100ans;  $F(x)=0,01=1\%$  d'où  $y = -[\ln(-\ln 0,99)] = 4,60$  Soit:  $P_{max,j,1\%} = (10.36 * 4,60) + 33.59 = 81\text{mm}$

• **5- Intervalles de confiance**

BERNIER et VERON ont étudié les intervalles de confiance de la loi de Gumbel. En outre, pour ces intervalles de confiance, on suppose un ajustement par la méthode des moments (pour le calcul du mode et du gradex). Pour un seuil  $\alpha$ , l'intervalle de confiance sur un quantile  $X_{p\%}$  s'exprime en fonction de l'écart-type  $\sigma_x$  par l'expression 9.9

Où  $h_1$  et  $h_2$  sont des paramètres dépendant de la taille de l'échantillon et de la fréquence (Probabilité) et de la valeur de  $\alpha$ .

- $h_1$  et  $h_2$  seront évalués par la formule suivante (avec le signe + pour  $h_2$  et le signe - pour  $h_1$ ):

$$h_{1,2} = \frac{\sqrt{n} \sqrt{1 + 1.13 y_{P\%} + 1.1 y_{P\%}^2} \pm \frac{u_\alpha^2}{n} (1.1 y_{P\%} + 0.57)}{1 - 1.1 \frac{n \alpha^2}{n}} \dots \dots (III.10)$$

- $u_\alpha$  est la variable réduite de Gauss correspondant à la fréquence  $1 - \frac{1-\alpha}{2}$
- $y_{P\%}$  est la variable réduite de Gumbel correspondant à la fréquence  $1 - P\%$ , ramenée à sa moyenne et à son écart-type:

$$y_{P\%} = \frac{-\text{Ln}[-\text{Ln}(F(x))] - 0.577}{1.28} \dots \dots (III.11)$$

Pour  $\alpha = 5\%$       $u_\alpha = 1.96$

$$\hat{x}_{P\%} - h_1 \sigma_x \leq P_{\text{max},j,P\%} < \hat{x}_{P\%} + h_2 \sigma_x \dots \dots (III.12)$$

$$P_{\text{max},j,1,P\%} \leq P_{\text{max},j,P\%} < P_{\text{max},j,2,P\%}$$

$\sigma_x$  : Ecart type de la variable à étudier, soit  $\sigma_{P_{\text{max},j}} = 13.24$

Avec :

$$P_{\text{max},j,1,P\%} = P_{\text{max},j,P\%} - h_1 \sigma_{P_{\text{max},j}}$$

$$P_{\text{max},j,2,P\%} = P_{\text{max},j,P\%} + h_2 \sigma_{P_{\text{max},j}}$$

Le tableau (III.12) résume le calcul des intervalles de confiance.

**Tableau III.12 : Calcul des intervalles de confiance**

Période de retour (an)	1-F(x)	$y_{P\%}$	$h_1$	$h_2$	$P_{\text{max},j,1,P\%}$ % mm	$P_{\text{max},j,P\%}$ mm	$P_{\text{max},j,2,P\%}$ % mm
10	0,90	1,3073	0,418	0,677	51	57	71
50	0,98	2,5976	0,664	1,104	66	74	88
100	0,99	3,1430	0,810	1,250	71	81	98

**On écrit :**

$$\text{Prob} [P_{\text{max},j,1,P\%} \leq P_{\text{max},j,P\%} < P_{\text{max},j,2,P\%}] = 1 - 0,05 = 0,95 = 95\%$$

$$F(x)=10\%: \quad \text{Prob}[51 \leq 57 < 71] = 95\%$$

$$F(x)=2\%: \quad \text{Prob}[66 \leq 74 < 88] = 95\%$$



$F(x)=1\%: \text{Prob}[71 \leq 81 < 98] = 95\%$

Le même résultat peut être trouvé en utilisant Hydrolab donnant aussi les intervalles de confiance. Les calculs très utiles dans le tableau (III.13) et l'ajustement est donné en figure (III.7).

**Tableau(III.13) :** Calculs de la variable réduite de Gumbel.

<b>Mode=</b>	<b>33,56</b>				<b>I.C.(%)=95</b>	
<b>Gradex=</b>	<b>10,33</b>	<b>F(x)</b>	<b>y</b>	<b>(64)</b>	<b>uGauss=1,96</b>	
<b>Pmax,j classées</b>	<b>Ordre de classement</b>	<b>Fréquence expérimentale</b>	<b>Variable réduite</b>	<b>Valeur théorique</b>	<b>Borne inférieure</b>	<b>Borne supérieure</b>
16,4	1	0,0078	-1,579	17,2551	10,99250856	21,3349465
18,6	2	0,0234	-1,323	19,90626	14,41488904	23,5901624
18,7	3	0,0391	-1,176	21,41698	16,34485774	24,895483
20	4	0,0547	-1,067	22,54831	17,77790325	25,8852311
21	5	0,0703	-0,976	23,48233	18,95168411	26,7116969
22,8	6	0,0859	-0,898	24,29397	19,9638359	27,4377217
24,5	7	0,1016	-0,827	25,02198	20,86473626	28,0958873
25	8	0,1172	-0,763	25,68921	21,68403993	28,7054869
25,4	9	0,1328	-0,703	26,31038	22,44079189	29,2790039
25,8	10	0,1484	-0,646	26,89559	23,1480097	29,8250512
28,2	11	0,1641	-0,592	27,45212	23,81501968	30,3498712
28,5	12	0,1797	-0,540	27,98539	24,44875552	30,8581697
29,1	13	0,1953	-0,491	28,49961	25,05452969	31,3536118
29,5	14	0,2109	-0,442	28,99808	25,63651649	31,8391318
29,7	15	0,2266	-0,395	29,48351	26,19806803	32,3171362
30,2	16	0,2422	-0,349	29,9581	26,74192839	32,7896408
30,4	17	0,2578	-0,304	30,42373	27,27038347	33,2583664
31,4	18	0,2734	-0,260	30,88201	27,78536859	33,7248079
31,8	19	0,2891	-0,216	31,33433	28,2885476	34,1902848
32,3	20	0,3047	-0,173	31,78191	28,78137229	34,6559788
32,5	21	0,3203	-0,130	32,22586	29,26512788	35,1229636
32,5	22	0,3359	-0,087	32,66718	29,74096831	35,5922279
32,5	23	0,3516	-0,044	33,10678	30,20994426	36,0646938
33,7	24	0,3672	-0,002	33,54554	30,67302551	36,5412326
33,8	25	0,3828	0,041	33,98426	31,13111922	37,0226777
34	26	0,3984	0,083	34,42372	31,58508499	37,5098365
34,9	27	0,4141	0,126	34,86468	32,0357475	38,0035011
35	28	0,4297	0,169	35,30787	32,48390738	38,5044576
35,5	29	0,4453	0,212	35,75403	32,9303506	39,0134965
35,6	30	0,4609	0,256	36,20391	33,37585701	39,5314212
35,8	31	0,4766	0,300	36,65824	33,8212081	40,0590584
35,8	32	0,4922	0,344	37,11782	34,2671945	40,5972676
36	33	0,5078	0,389	37,58342	34,71462332	41,1469516
36,3	34	0,5234	0,435	38,05589	35,16432564	41,7090682

### Chapitre III

### présentation de La zone d'étude

37,5	35	0,5391	0,481	38,53611	35,61716446	42,2846427
39	36	0,5547	0,529	39,02502	36,07404329	42,8747815
39,3	37	0,5703	0,577	39,52362	36,53591573	43,4806884
41,4	38	0,5859	0,626	40,033	37,00379637	44,1036822
41,8	39	0,6016	0,677	40,55436	37,47877342	44,745218
42,4	40	0,6172	0,729	41,08899	37,96202353	45,406912
42,7	41	0,6328	0,782	41,63835	38,45482942	46,0905709
43	42	0,6484	0,837	42,20402	38,95860115	46,7982272
43,3	43	0,6641	0,893	42,78783	39,47490202	47,5321821
45,1	44	0,6797	0,952	43,39181	40,00548037	48,2950589
45,3	45	0,6953	1,012	44,01829	40,55230928	49,0898687
45,4	46	0,7109	1,075	44,66995	41,11763646	49,9200931
48	47	0,7266	1,141	45,34988	41,70404789	50,7897906
48,4	48	0,7422	1,210	46,06172	42,31455003	51,7037329
49,7	49	0,7578	1,283	46,80977	42,95267756	52,6675848
51,8	50	0,7734	1,359	47,59916	43,62263694	53,688143
51,9	51	0,7891	1,440	48,43611	44,32950116	54,7736609
52	52	0,8047	1,526	49,32828	45,07947931	55,9342986
52,5	53	0,8203	1,619	50,28519	45,88029865	57,1827606
52,9	54	0,8359	1,719	51,31899	46,74176054	58,5352241
54,2	55	0,8516	1,828	52,44543	47,67657514	60,0127345
54,3	56	0,8672	1,948	53,68553	48,70166178	61,6433783
55	57	0,8828	2,082	55,06819	49,8402661	63,4658255
56,5	58	0,8984	2,234	56,63478	51,12560038	65,5354198
57	59	0,9141	2,410	58,44745	52,60754644	67,9354007
59,5	60	0,9297	2,619	60,60604	54,36615249	70,7995089
60	61	0,9453	2,878	63,2864	56,5423446	74,3634528
66,9	62	0,9609	3,223	66,845	59,42156028	79,1051502
74	63	0,9766	3,742	72,20295	63,74088397	86,2601507
75,5	64	0,9922	4,848	83,62967	72,91324066	101,558683

**NB:**

Ilestànoterquelemode (paramètredeposition)etlegradex (paramètre dedispersion)  
sontégaux à :

$$\text{Gradex } G = 0,78 \text{ } \sigma \text{ et}$$

$$x_0 = \bar{X} - 0.577 G \quad \text{et} \quad \bar{X} = \bar{P}_{\max,j}$$

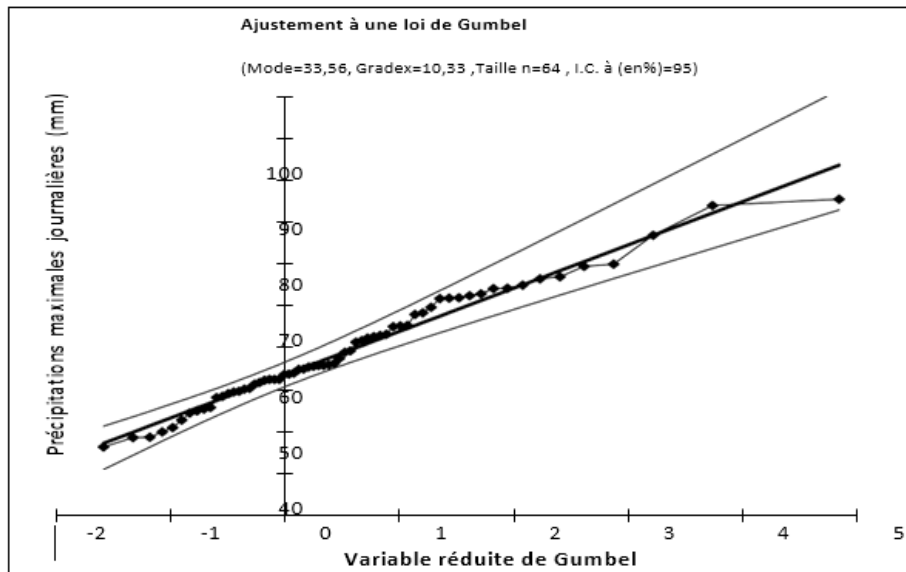


Fig III.8 : Ajustement à la loi de Gumbel (Hydrolab).

La droite de Gumbel est:  $P_{\max,j,p\%} = 10,33y + 33,56$

Le mode et le gradex sont calculés par la méthode des moments. Les quantiles sont résumés dans le tableau 9.4.

Tableau III.14. Quantiles (Hydrolab).

Période de retour an	Fréquence 1-F(x)	Variable réduite	$P_{\max,j,p\%}$ Mm	Borne inférieure	Borne supérieure
10	0,9	2,250	<b>56,8=57</b>	51,3	65,8
50	0,98	3,902	<b>73,9=74</b>	65,1	88,5
100	0,99	4,600	<b>81,1=81</b>	70,9	98,1

**On écrira:**

$$\text{Prob}[P_{\max,j,1,10\%} < P_{\max,j,10\%} < P_{\max,j,2,10\%}] = 1 - \alpha = 95\%$$

Soit: **Prob[51 < 57 < 66] = 1 - 5% = 95%**

$$\text{Prob}[P_{\max,j,2,2\%} < P_{\max,j,2\%} < P_{\max,j,2,2\%}] = 1 - \alpha = 95\%$$

Soit: **Prob[65 < 74 < 89] = 1 - 5% = 95%**

$$\text{Prob}[P_{\max,j,1,1\%} < P_{\max,j,1\%} < P_{\max,j,2,1\%}] = 1 - \alpha = 95\%$$

Soit: **Prob[71 < 81 < 98] = 1 - 5% = 95%**

### **Conclusion:**

Ce chapitre, nous a permis de récolter les caractéristiques générales de notre zone d'étude ainsi que les données relatives à la situation hydraulique actuelle qui vont nous servir pour la modélisation dans les chapitres suivants.

# *Chapitre IV*

## **Chapitre IV Méthodologie de box-jenkins**

---

### **CHAPITRE 4: METHODE DE BOX ET JENKINS**

#### **IV. 1. Introduction:**

L'étude des séries chronologiques ou temporelles correspond à l'analyse statistique d'observations régulièrement espacées dans le temps, dans le but de présenter des phénomènes aléatoires qui évoluent dans le temps, le modèle obtenu sera par la suite utilisé selon les objectifs désirés, tels que la prévision ou le contrôle.

Une démarche générale de prévision a été popularisée par Box et Jenkins (1976), c'est une méthode extrapolative du fait que seule la variable étudiée est utilisée.

Cette méthode est fondée sur la notion des modèles ARMA, et utilise des concepts et résultats de la théorie des processus aléatoires. Pour cela nous allons donner les définitions et les concepts de base. [13]

#### **IV. 2. Historique:**

La fonction de Box et Cox (1964) est une généralisation astucieuse de la transformation logarithmique qui permet, dans des situations favorables, de justifier objectivement de l'utilisation d'icelle. Dans des situations moins favorables elle permet au minima de proposer une transformation qui permet de mieux appréhender la variabilité des données. [8]

Box et Jenkins ont proposé une procédure unifiée de traitement des processus stochastiques stationnaires et ergodiques qui sert toujours de référence. La méthodologie qu'ils ont promue consiste à modéliser les séries temporelles univariées au moyen des processus ARMA. Ces processus sont parcimonieux et constituent une bonne approximation de processus plus généraux pourvu que l'on s'en tienne au cadre linéaire.

Cette méthode de Box-Jenkins vise à formuler un modèle permettant de représenter une série chronologique.

Son idée principale est le concept de parcimonie ou de la minimisation du nombre de paramètres. En pratique, ces derniers étant inconnus, ils sont remplacés par leurs valeurs initiales sachant que plus il y a de paramètres plus nombreuses sont les chances de se tromper. Dans cette approche une hypothèse et quatre étapes sont considérées. [8]

#### **IV. 3. La méthodologie de Box & Jenkins:**

La méthode de Box et Jenkins permet en plusieurs étapes de trouver un modèle.

## Chapitre IV Méthodologie de box-jenkins

---

ARMA susceptible de représenter une série chronologique (LS agit d'une méthode de prévision extrapolative puisque seul le passé de la variable est utilisé à cette fin, sans apport d'information ex tertiaire).

La méthode de prévision de Box-Jenkins est particulièrement bien adaptée au traitement de séries chronologiques complexes et en général aux situations dans lesquelles la loi de base n'est pas immédiatement apparente. Comme elle traite des situations beaucoup plus compliquées, il est bien plus difficile de saisir les principes de cette technique, d'autant plus que son application présente certaines limites.

En décrivant leur méthode de prévision George Box et Gwilym Jenkins ont établi le diagramme représenté par la figure III.1. [9]

Cette démarche divise le problème de prévision en 5 étapes pour pouvoir trouver le modèle ARMA susceptible de représenter la série chronologique, et la dernière étape c'est l'étape de prévision:

1. Analyses préliminaires (Stationnarisation).
2. Identification.
3. Estimation.
4. Tests diagnostiques (validation du modèle).
5. Prévision.

## Chapitre IV Méthodologie de box-jenkins

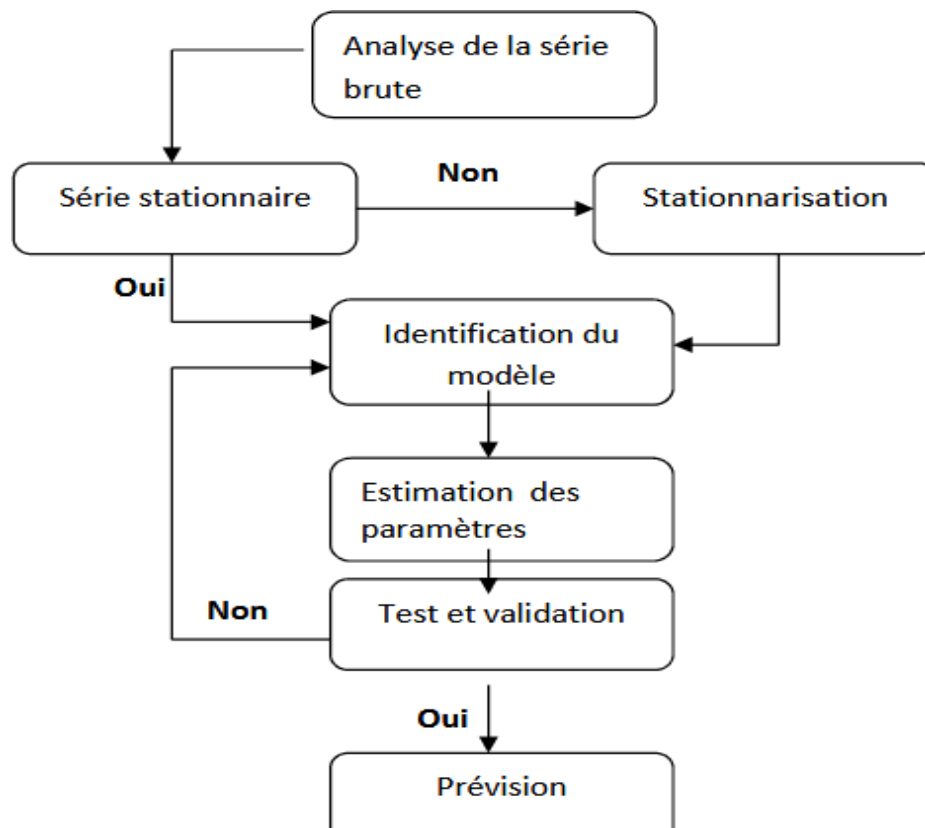


Figure IV .1: procédure itérative de Box-Jenkins.[9]

### IV. 4. Notions sur la méthode de Box et Jenkins:

C'est une approche itérative qui consiste à identifier le modèle susceptible de représenter le phénomène étudié, elle se réfère à deux types de modèles autorégressifs et une moyenne mobile ou à une combinaison des deux, cependant ; la modélisation d'une chronique nécessite au préalable une stationnarisation.

La méthode de prévision de Box-Jenkins est particulièrement bien adaptée au traitement de séries chronologiques complexes et à d'autres situations où la loi de base est difficile à appréhender. Cependant, il est bien plus difficile de cerner les principes de la technique, ainsi que les limites de son application.

Dans la méthode de Box-Jenkins, il n'y a pas besoin de faire dès le départ l'hypothèse d'une loi rigide mais au lieu de cela, on commence par une loi expérimentale adaptée au don de vue de minimiser l'erreur. Ensuite, la méthode prend un aspect subjectif puisqu'elle fournit aux personnes qui appliquent cette méthode des informations explicites qui leur donnent la possibilité de juger si la loi adoptée est correcte ou non pour la situation considérée.



## Chapitre IV Méthodologie de box-jenkins

### IV. 4.1. Stabiliser les processus aléatoires non stationnaires:

Les processus stochastiques non stationnaires sont caractérisés par des propriétés stochastiques qui évoluent en fonction du temps. On distingue deux types de processus stochastiques non stationnaires : un non stationnarité de nature déterministe (TS) et une non stationnarité de nature stochastique (DS).

### IV. 4.2: Stabiliser des processus TS:

Un processus présente une non stationnarité de type déterministe TS (Trend Stationary), s'il peut se décomposer en une somme de deux fonctions:

$$X_t = f(t) + a_t \dots \dots \dots (III. 1)$$

Tel  $a_t$  Est un processus stationnaire de type ARMA.

$f(t)$ : Est une fonction polynomiale du temps.

La non stationnarité de ce processus est dû au fait que son espérance dépend du temps.

La méthode pour stationnariser un processus TS se traduit d'estimer les coefficients  $a_0, a_1, \dots$

par MCO (Moindre Carrés Ordinaires) et de retrancher de la valeur de

$X_t$  la valeur

estimée de sa moyenne.

Il faut transformer la série observée de manière à :

- Enlever la tendance,
- Enlever la saisonnalité,
- Stabiliser la variance.

### IV.5. Test des composantes saisonnières et tendancielles :

#### IV.5.1. Test de Fisher:

Avant toute étude d'une chronique, il convient d'élaborer un test permettant de détecter l'existence d'une saisonnalité. Le test le plus communément employé est celui de Fisher par analyse de la variance du facteur période (mensuel, trimestriel...).

On considère :

n: Le nombre d'années,

P : Le nombre d'observations dans l'année

$X_{ij}$ : La valeur de la série pour la  $i^{\text{ème}}$  année et la  $j^{\text{ème}}$  période

La moyenne générale  $\bar{X}$  la moyenne de l'année  $\bar{X}_i$ , la moyenne de la période  $\bar{X}_j$

## Chapitre IV Méthodologie de box-jenkins

La variance année et la variance période sont définies respectivement par :

$$var_A = \frac{p \sum_{i=1}^n (\bar{X}_i - \bar{X} \dots)^2}{n-1} \dots \dots \dots (III.2)$$

$$var_p = \frac{n \sum_{j=1}^p (\bar{X}_j - \bar{X} \dots)^2}{j-1} \dots \dots \dots (III.3)$$

La variance résiduelle :  $var_R = \frac{\sum_{i=1}^n \sum_{j=1}^p (X_{ij} - \bar{X}_i - \bar{X}_j - \bar{X} \dots)^2}{(n-1)(j-1)} \dots \dots \dots (III.4)$

L'équation de la variance :

$$var_T = var_A + var_p + var_R = \frac{\sum_{i=1}^n \sum_{j=1}^p (X_{ij} - X \dots)^2}{n-1} \dots \dots \dots (III.5)$$

L'hypothèse est :

H0 : « pas de saisonnalité »

H1 : « il existe une saisonnalité »

La valeur calculée  $F = \frac{var_p}{var_R} \dots \dots \dots (III.6)$

Quel'on compare à la valeur tabulée  $F_{v_1, v_2}^\alpha$

Avec :  $v_1 = (p - 1), v_2 = (n - 1)(p - 1) \dots \dots \dots (III.7)$  degré de liberté

Si  $F_0 > F_{v_1, v_2}^\alpha$  on rejette  $H_0$  : la série est saisonnière.

Soient les hypothèses :

H0 : « La série n'est pas affectée d'une tendance »

H1 : « La série est affectée d'une tendance »

On calcule  $F_1 = \frac{var_A}{var_R} \dots \dots \dots (III.8)$

quel'on compare avec  $F_{v_1, v_2}^\alpha$

Avec  $v_3 = (p - 1), v_2 = (n - 1)(p - 1) \dots \dots \dots (III.8)$  degré de liberté

❖ Si  $F_0 > F_{v_1, v_2}^\alpha$  on rejette l'hypothèse, la série est affectée d'une tendance.

## Chapitre IV Méthodologie de box-jenkins

Concernant l'existence de la tendance, le test de Fisher's a été rendu faible, il convient d'effectuer un autre test. [15]

### 2. Test de Dickey-Fuller:

Il permet de mettre en évidence le caractère stationnaire ou non d'une chronique par la détermination d'une tendance déterministe ou stochastique.

Pour la construction de ce test, Dickey et Fuller ont proposé trois modèles de base

$(1 - \phi_1 B)X_t = \varepsilon_t$ : Modèle autorégressif d'ordre 1.....Modèle (1)...

$(1 - \phi_1 B)(X_t - c) = \varepsilon_t$ : Modèle autorégressif d'ordre 1 avec constante.....Modèle (2)

$(1 - \phi_1 B)(X_t - c - bt) = \varepsilon_t$ : Modèle autorégressif d'ordre 1 avec tendance.....Modèle (3)

Les hypothèses à tester sont:  $H_0: \phi_1 = 1$

$H_1: \phi_1 < 1$

C'est un test itératif qui se déroule comme suit:

- Si dans l'un des modèles la statistique relative à  $\phi_1$  est supérieure aux valeurs tabulées  $t_{0,05}$ , on déduit qu'il existe une racine unitaire. On estime en premier le modèle (3); on commence par tester la significativité de la tendance:
  - Si la tendance n'est pas significative alors on teste la significativité de la constante
  - Si elle est significative le processus est de type DS avec dérive, pour stationnariser la chronique on applique le filtre aux différences et recommence la procédure.
  - Si elle n'est pas significative, le processus est donc une marche aléatoire dit DS sans dérive, on différencie la série une fois et on recommence jusqu'à ce que la série soit stationnaire.
- Sinon, le processus est déterministe; pour le stationnariser on e- Sinon Si dans le modèle (3), le coefficient de la tendance est significativement différent de 0 alors le processus est TS.
- Sinon la série est stationnaire Effectue une régression sur le temps et on recommence le test.
- Sinon Si dans le modèle (3), le coefficient de la tendance est significativement différent de 0 alors le processus est TS;

## Chapitre IV Méthodologie de box-jenkins

- Sinon la série est stationnaire. [15]

- **La stratégie de test de Dickey Fuller:**

Le test de Dickey Fuller (1979) est un test de racine unitaire (ou de non stationnarité) dont l'hypothèse nulle est la non stationnarité d'un processus autorégressif d'ordre 1.

Considérons un processus  $(X_t, t \in \mathbb{Z})$  satisfaisant la représentation AR(1) suivante:

$$X_t = \rho X_{t-1} + \varepsilon_t \quad \text{Avec } \varepsilon_t \text{ i.i.d. } (0, \sigma_\varepsilon^2) \text{ et } \rho \in \mathbb{R} \dots \dots \dots \text{(III.10)}$$

Le principe général du test de Dickey Fuller consiste à tester l'hypothèse nulle aléatoire (Random Walk Process). L'hypothèse nulle testée correspond ainsi à une hypothèse de non stationnarité stochastique. Ce test, comme tout test non symétrique, peut être réalisé de différentes façons. La plus simple consiste à utiliser une statistique de Student associée à l'hypothèse  $H_0$ . En cela, le test de Dickey Fuller ne se distingue pas d'un test quelconque d'une hypothèse non symétrique. Là où l'application du test de Dickey Fuller diffère de celle d'un test standard, c'est dans la distribution asymptotique de la statistique de Student associée au test  $H_0$ .


En effet, la distribution asymptotique de l'estimateur des moindres carrés du paramètre  $\rho$ , sous l'hypothèse de non stationnarité, est non standard. On n'a plus dans ce cas une distribution asymptotique normale, comme en économétrie de base. De la même façon, la statistique de

Student associée au test  $\rho=1$ , n'a pas une distribution asymptotique standard (distribution de Student approximée par une distribution normale). C'est pourquoi, l'application du test de Dickey Fuller nécessite que l'on utilise des seuils différents de ceux que l'on utilise traditionnellement pour des statistiques de Student.

Aussi la distribution asymptotique de la statistique de Student associée au test  $H_0$  n'est pas la même suivant que dans le modèle on inclut ou non, une constante et un trend déterministe.

Dès lors, puisque a priori, on ne sait pas si l'on doit inclure cette constante et centred, il convient d'appliquer non pas un test simple, mais une stratégie de tests de Dickey Fuller.

On considère les trois modèles définis comme suit:

 Modèle 1:  $\Delta X_t = \rho X_{t-1} + \varepsilon_t \dots \dots \dots \text{(III.11)}$

 Modèle 2:  $\Delta X_t = \rho X_{t-1} + c + \varepsilon_t \dots \dots \dots \text{(III.12)}$

## Chapitre IV Méthodologie de box-jenkins

---

 Modèle 3:  $\Delta X_t = \phi X_{t-1} + \beta t + c + \varepsilon_t \dots \dots \dots (III.13)$

Avec  $\varepsilon_t \text{ i.i.d. } (0, \sigma_\varepsilon^2)$ . On cherche à tester l'hypothèse de racine unitaire :

$$H_0: \phi = 0 \text{ contre } H_1: \phi < 0$$

### 1. Procédure:

Le principe général de la stratégie de test est illustré par la figure : III.3

Il s'agit de partir du modèle le plus général, d'appliquer le test de racine unitaire en utilisant les seuils correspondant à ce modèle, puis de vérifier par un test approprié que le modèle retenu était le bon. En effet, si le modèle n'était pas le bon, les seuils utilisés pour le test de racine unitaire ne sont pas valables. On risque alors de commettre une erreur de diagnostic quant à la stationnarité de la série.

Il convient dans ce cas, de recommencer le test de racine unitaire dans un autre modèle, plus contraint. Et ainsi de suite, jusqu'à trouver le bon modèle, les bons seuils et bien entendu les bons résultats. [3]

## Chapitre IV Méthodologie de box-jenkins

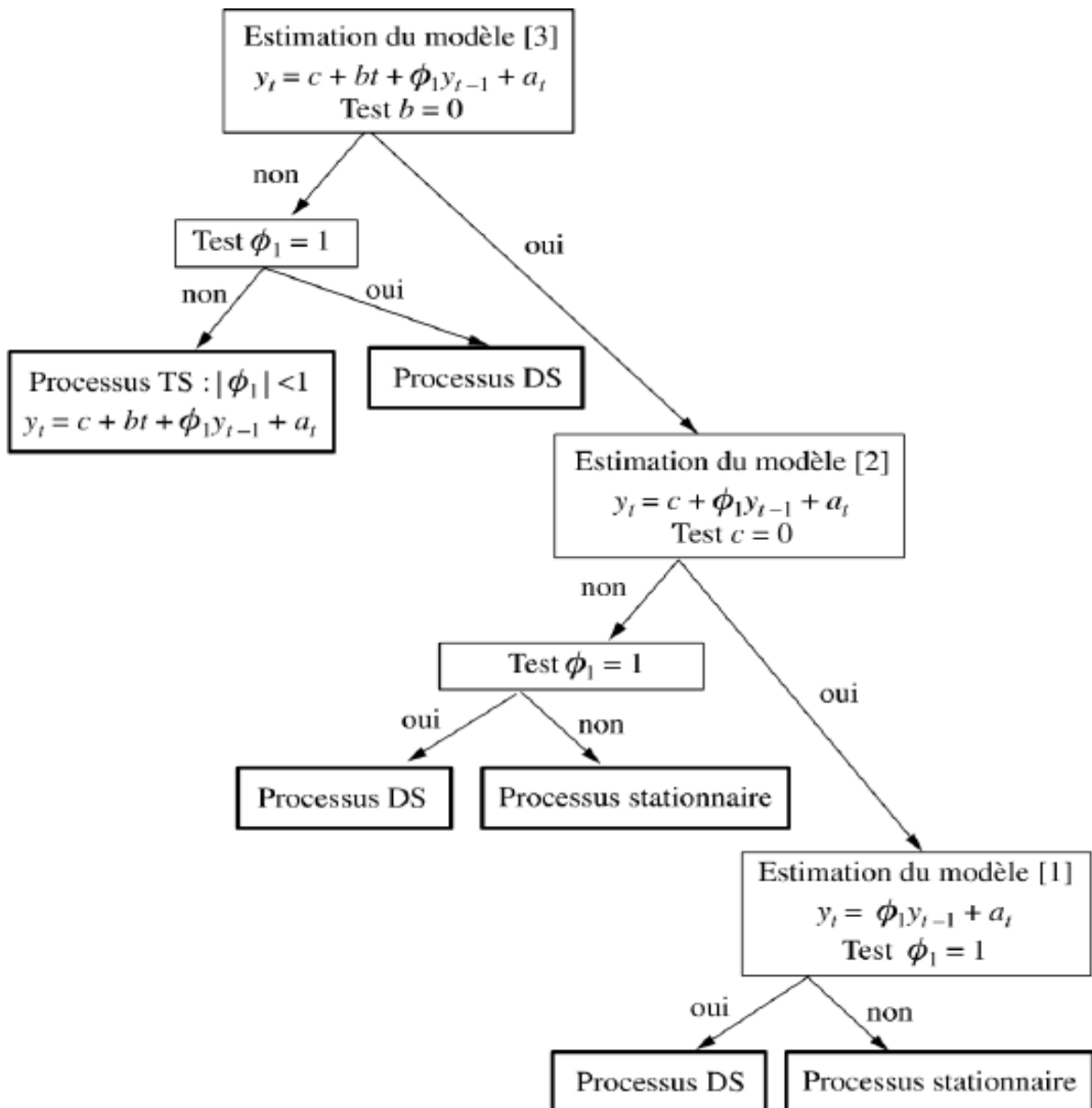


Figure IV.2: Résumé de la stratégie de test de Dickey-Fuller. [15]

### IV.7. Les étapes de Box et Jenkins :

#### IV.7.1 Identification du modèle :

Cette étape consiste à identifier le modèle ARMA susceptible de représenter la série, c'est pour cela qu'il est important de se familiariser avec les données en examinant le graphique de la série chronologique (présence de saisonnalité, stationnarité,...) qui permet de faire

## Chapitre IV Méthodologie de box-jenkins

une analyse préliminaire qui consiste par exemple à corriger les données aberrantes, transformer les données (transformation logarithmique, inverse, racine carrée,...) puisqu'il faut se ramener à une série stationnaire, le recours aux différences premières ordinaires, différences premières saisonnières, différences ordinaires et saisonnières. Le choix de la transformation est dicté par l'allure graphique de la série.

D'ailleurs le choix de la transformation des données est plus facile après avoir appliqué

les opérateurs de différence adéquats. Il est conseillé de comparer les variances des différentes séries. Comme l'inspection des autocorrélations partielles (PAC) donne une idée sur l'ordre du modèle autoregressif et celle des autocorrélations simples (AC) donne une idée sur l'ordre du modèle moyenne mobile. Plus précisément si la fonction d'autocorrélation simple décroît rapidement vers zéro et la fonction d'autocorrélations partielles présente un cut-off après  $p$  retard, on peut conclure que la série provient d'un processus autorégressif d'ordre  $p$  (AR( $p$ )).

Par contre si la fonction d'autocorrélation simple présente un cut-off après  $q$  retards et que la fonction d'autocorrélation partielles décroît rapidement vers zéro, alors on peut déduire que la série est générée à partir d'un modèle moyenne mobile d'ordre  $q$  (MA( $q$ )). Evidemment il se peut aussi que les fonctions d'autocorrélation simple et d'autocorrélations partielles présentent une forme exponentielle ou sinusoidale, on constate dans ce cas qu'on est en présence d'un processus autorégressif moyenne mobile ARMA ( $p ; q$ ).

Le tableau qui suit synthétise les caractéristiques en termes de corrélogramme des processus AR, MA et ARMA.

**Tableau IV.1.: Tableau résumant les propriétés des AR( $p$ ), MA( $q$ ) et ARMA( $p,q$ ).**

Processus	FAC	FAP
AR( $p$ )	Décroissance exponentielle et /ou sinusoidale	Pic significatifs pour les $p$ premiers retards, les autres coefficients sont nuls pour des retards $> p$
MA( $q$ )	Pic significatifs pour les $q$ premiers retards, les autres coefficients sont nuls pour des retards $> q$	Décroissance exponentielle et /ou Sinusoidale
ARMA( $p,q$ )	Décroissance exponentielle ou sinusoidale amortie tronquée après $(q-p)$ retards	Décroissance exponentielle ou sinusoidale amortie tronquée après $(q-p)$ retards

## Chapitre IV Méthodologie de box-jenkins

---

A priori, n'est pas une tâche aisée tant et demande beaucoup d'expertise ce qui révèle une part de subjectives.

Il existe cependant des méthodes d'identification automatiques, basées sur le critère d'information.

- **Familiarisation avec les données:**

Cette étape consiste à s'informer sur le domaine dont relèvent les données, les objectifs poursuivis (prévision ponctuelle ou par intervalle, et la détection d'un changement de comportement etc.), examiner les représentations graphiques des données.

- **Analyse préliminaire:**

Dans cette étape, on est amené à faire les changements nécessaires sur notre série, par exemple : abandonner une partie des données au début de cette dernière, corriger les données aberrantes, compléter les données manquantes.

On doit aussi choisir de travailler:

- En différences premières ordinaires (opérateur  $\nabla$ ) en présence de latence de variation de niveau.
- En différences premières saisonnières (opérateur  $\nabla\{s\}$ ), en présence d'un comportement périodique de périodes.
- En différences premières ordinaires et saisonnières (opérateur  $\nabla\nabla\{s\}$ ), voire en différences secondes (opérateur  $\nabla^2$ ).

Pour s'amener à un modèle ARMA. D'ailleurs le choix de la transformation des données est plus facile après avoir appliqué les opérateurs de différences adéquates.

- **Spécification du modèle:**

Cette étape consiste à déterminer les paramètres  $p, d, q$  du modèle ARIMA en se basant sur les représentations graphiques de l'autocorrélation simple et l'autocorrélation partielle de la série stationnaire étudiée.



## Chapitre IV Méthodologie de box-jenkins

L'idées ou sjacente est que chaque modèle ARMA possède des fonctions d'autocorrélation théoriques et estimées. On peut alors choisir un ou plusieurs modèles théoriques. Cette étape est la partie la plus difficile car dans la pratique les corrélogrammes observés n'engendrent pas toujours un choix évident.

Le tableau suivant résume les différents cas possibles :

**Tableau IV.2: spécification du modèle.**

Modèles	Autocorrélation simple	Autocorrélation partielle
AR(p)	Décroit vers zéro	Tronqué après un retard p
MA(q)	Tronqué après un retard q	Décroit vers zéro
ARMA(p;q)	Décroit vers zéro	Décroit vers zéro

### 2. Estimation :

Les paramètres sont les coefficients des polynômes AR et MA ainsi que des polynômes saisonniers éventuels SAR et SMA.

La méthode d'estimation des paramètres du modèle identifié dépend du type de processus diagnostiqué :

- Dans le cas d'un modèle AR(p), on peut appliquer la méthode des moindres carrés ordinaires MCO, pourvu que les résidus résultant soient non corrélés, on peut aussi exploiter les équations Yule-Walker. [6]
- Dans le cas d'un MA(q), l'estimation s'avère plus complexe, plusieurs méthodes itératives ont été mises au point, nous citons, à titre d'exemple : la procédure itérative de balayage, proposée par Box et Jenkins. Mais cette méthode n'est applicable que si le nombre de paramètres à estimer n'est pas trop important, en plus la plage de variation de ces paramètres est difficile à déterminer.

Cependant, il existe une procédure itérative de deux étapes pour l'estimation des (p+q) paramètres du modèle ARMA (p; q), ces étapes sont :

**Etape 1 :** Pour obtenir les estimateurs des paramètres autorégressifs, on résout le système de Yule-Walker suivant :

$$\hat{\rho}(h) - \sum_{i=1}^p \phi_i \rho(h-i) = 1 \text{ pour } h = 1, \dots, p, \hat{\rho}(0) = 1 \dots \dots \dots \text{(III. 14)}$$

## Chapitre IV Méthodologie de box-jenkins

On obtient alors la solution  $\boldsymbol{\varphi} = (\boldsymbol{\varphi}_1, \boldsymbol{\varphi}_2, \dots, \boldsymbol{\varphi}_p) \dots \dots \dots$  (III. 15)

Pour estimer les paramètres moyens mobiles, on utilise l'algorithme de Newton-Raphson:

**Etape 2:** Une fois l'estimation initiale obtenue, on utilise l'algorithme de Marquardt pour les  $(p+q)$  paramètres définitifs.

### IV.7.3. Validation du modèle (Adéquation) :

Après la phase précédente, si on dispose de plusieurs modèles candidats, il est nécessaire d'effectuer des tests sur les paramètres et sur les résidus afin de les départager.

#### IV.7.3.1. Test sur les paramètres :

C'est un test classique de Student qui permet de tester la significativité de chaque paramètre du processus en utilisant les hypothèses:

$H_0$ : « Le coefficient = 0 » contre

$H_1$ : « Le coefficient est significativement différent de zéro »

La statistique de Student représenterait du coefficient à estimer sur son écart type:

$$\frac{\hat{B}_p}{\hat{\sigma}_{\hat{B}_p}}$$

$\frac{\hat{B}_p}{\hat{\sigma}_{\hat{B}_p}} > \sigma$  On accepte  $H_1$ , dans le cas contraire on rejette le modèle et on envisage une nouvelle

spécification (retour à l'étape d'identification). Après avoir validé plusieurs modèles, il convient de choisir le modèle optimal en se basant sur une comparaison de leur qualité en utilisant les critères d'information et / ou le principe de parcimonie.

#### ✓ Critères d'informations:

Ce sont des statistiques aidant à décider de l'ordre d'un modèle dans le cas où il en existerait plusieurs. Entre autres ; on peut citer:

#### ✓ Critère d'information d'Akaike (AIC):

En anglais Akaike information criterion est une mesure de la qualité d'un modèle statistique proposée par Hirotugu Akaike en 1973.

Lorsque l'on estime un modèle statistique, il est possible d'augmenter la

## Chapitre IV Méthodologie de box-jenkins

vraisemblance du modèle en ajoutant un paramètre. Le critère d'information d'Akaike, tout comme le critère d'information bayésien, permet de pénaliser les modèles en fonction du nombre de

paramètres afin de satisfaire le critère de parcimonie. On choisit alors le modèle avec le critère d'information d'Akaike le plus faible.

Il tient compte de la qualité d'ajustement du modèle à la série observée et du nombre de paramètres utilisés dans l'ajustement.)

$$AIC(p, q) = n \log \hat{\sigma}_\varepsilon^2 + 2(p + q) \dots \dots \dots (III. 16)$$

✓ Critère de Schwartz (1978) :

$$SC(p, q) = n \log \hat{\sigma}_\varepsilon^2 + (p + q) \log n \dots \dots \dots (III. 17)$$

### IV.7.3. 2. Tests sur les résidus :

Lorsque le modèle est bien estimé, les résidus entre les valeurs observées et les valeurs estimées doivent se comporter comme un bruit blanc (gaussien ou pas). Les résidus (ou erreurs de prévision) sont notés  $\varepsilon_t, t \in \mathbb{Z}$ .

### IV.7.3.3. Test de Box-Ljung :

Appelé aussi test de « porte manteau », il permet de vérifier l'hypothèse de bruit blanc des résidus, ce qui signifie qu'il n'existe aucune autocorrélation significativement non nulle, il se base sur la fonction d'autocorrélation en utilisant deux hypothèses :

$$H_0: \rho_1 = 0, \rho_2 = 0, \dots, \rho_k = 0$$

$H_1$ : Il existe au moins un  $\rho$  significativement différent de zéro.

$$\text{Statistique est } Q = n(n+1) \sum_{h=1}^K \frac{r_h^2(\varepsilon_t)}{n-h} \dots \dots \dots (III. 18)$$

Avec:

$K$  = nombre de retards choisis et  $n$  = nombre d'observations

En l'absence d'autocorrélation la statistique  $Q$  obéit à une  $\chi^2(K - p - q - P - Q)$  de degrés de liberté.

## Chapitre IV Méthodologie de box-jenkins

L'hypothèse  $H_0$  est rejetée à un seuil de 5% si  $Q$  est supérieure à la quantile 0.95 de la loi de  $\chi^2$ .

### IV.7.3.4. Test de normalité :

Le test le plus fréquent qui permet de vérifier la normalité d'une distribution statistique est celui de Jarqueet Bera (1984), ce dernier est fondé sur la notion de Skewness (asymétrie) et de Kurtosis (queue de distribution), par ailleurs il existe un autre test celui de Kolmogorov-Smirnov.

Ces deux tests permettent par la suite de calculer les intervalles de prévisions

Soit :  $u_k = \frac{1}{n} \sum_{t=1}^n (\varepsilon_t - \bar{\varepsilon})^k$  le moment centré d'ordre  $k$  du processus  $\{\varepsilon_t, t \in \mathbb{Z}\}$

- ❖ Le coefficient de Skewness est défini par :

$$B_1^{1/2} = \frac{u_3}{u_2^{3/2}} \dots \dots \dots \text{(III. 19)}$$

- ❖ Le coefficient de Kurtosis est défini par :

$$B_2 = \frac{u_4}{u_2^2} \dots \dots \dots \text{(III. 20)}$$

Si le nombre d'observation grand ( $n > 30$ ) on suit alors la loi normale

$$B_1^{1/2} \rightarrow N\left(0; \sqrt{\frac{6}{n}}\right) \text{ et } B_2 \rightarrow N\left(3; \sqrt{\frac{24}{n}}\right) \dots \dots \dots \text{(III. 21)}$$

On construit alors les statistiques centrées réduites correspondantes à  $B_1^{1/2}$  et  $B_2$

$$v_1 = \frac{|B_1^{1/2} - 0|}{\sqrt{\frac{6}{n}}} \text{ et } v_2 = \frac{|B_2 - 3|}{\sqrt{\frac{24}{n}}} \dots \dots \dots \text{(III. 22)}$$

Quel'on compare à 1.96 (valeur de la loi normale à un seuil de 5%) Soient les hypothèses :

$H_0: v_1 = 0$  (symétrie) et  $v_2 = 0$  (aplatissement normal) On acceptera l'hypothèse  $H_0$  si  $v_1 \leq 1.96$  et  $v_2 \leq 1.96$ ; dans le cas contraire, l'hypothèse de normalité est rejetée.

Le test de Jarqueet Bera synthétise les résultats précédents

- Si  $B_1^{1/2}$  et  $B_2$  obéissent à des lois normales alors la statistique :

$$JB = \frac{n}{6} B_1 + \frac{n}{24} (B_2 - 3)^2 \dots \dots \dots \text{(III. 23)}$$

## Chapitre IV Méthodologie de box-jenkins

Suite  $\chi^2$  à deux degrés de liberté.

- Si  $J/B > \chi^2_{1-\alpha}$  on rejette l'hypothèse de normalité des résidus au seuil  $\alpha$ .

### IV.7.3. 5. Test de Durbin et Watson :

Les modèles ajustés à des séries chronologiques manifestent parfois un certain degré de corrélation entre les valeurs successives des erreurs. En terme probabiliste, cela signifie que les erreurs sont autocorrélées, ou encore qu'une erreur produite en  $t-1$  a une influence sur l'erreur en  $t$ . Le test de Durbin et Watson (1951) permet de détecter l'autocorrélation des résidus pour un ordre  $n$  (corrélation entre  $\varepsilon_t$  et  $\varepsilon_{t-1}$ ) sous la forme :

$$\varepsilon_t = \rho \varepsilon_{t-1} + v_t \text{ ou } v_t \rightarrow N(0; \sigma^2 V)$$

On teste  $H_0: \rho = 0$  (absence d'autocorrélation à l'ordre 1 des résidus).

Contre :  $H_1: \rho \neq 0$  (présence d'autocorrélation à l'ordre 1 des résidus).

La statistique de Durbin et Watson, notée DW, est donnée par :

$$DW = \frac{\sum_{t=2}^n (\varepsilon_t - \varepsilon_{t-1})^2}{\sum_{t=1}^n \varepsilon_t^2} \dots \dots \dots \text{(III. 24)}$$

Avec :

$\varepsilon_t$  : sont les résidus de l'estimation du modèle.

Depuis sa construction, cette statistique varie entre 0 et 4 et nous avons  $DW=2$  lorsque  $\hat{\rho}$  est l'estimateur de  $\rho$  : on a  $DW=0$ , il existe une autocorrélation positive ;  $DW=4$ , il existe une autocorrélation négative ;  $DW=2$ , indique l'absence d'autocorrélation. [15]

### IV.7.4. la prévision :

Lorsque pour identifier le processus étudié à un processus ARMA, on applique différentes transformations (exemple différenciation dans le cas d'une série I(1)), il est nécessaire lors de la phase de prévision de prendre en compte la transformation retenue et de recolorer la prévision. Plusieurs cas sont possibles :

- Si le processus contient une tendance déterministe, on extrait cette dernière par régression afin d'obtenir une série stationnaire lors de la phase d'estimation. Ensuite, lors de la phase de prévision, on adjoint aux prévisions réalisées sur la composante ARMA stationnaire, la projection de la tendance.
- Si la transformation résulte de l'application d'un filtre linéaire (de type par

## **Chapitre IV      Méthodologie de box-jenkins**

---

exemple de différences premières), on réalise les prévisions sur les séries filtrées stationnaires et l'on reconstruit ensuite par inversion du filtre les prévisions sur la série initiale.

## Chapitre IV Méthodologie de box-jenkins

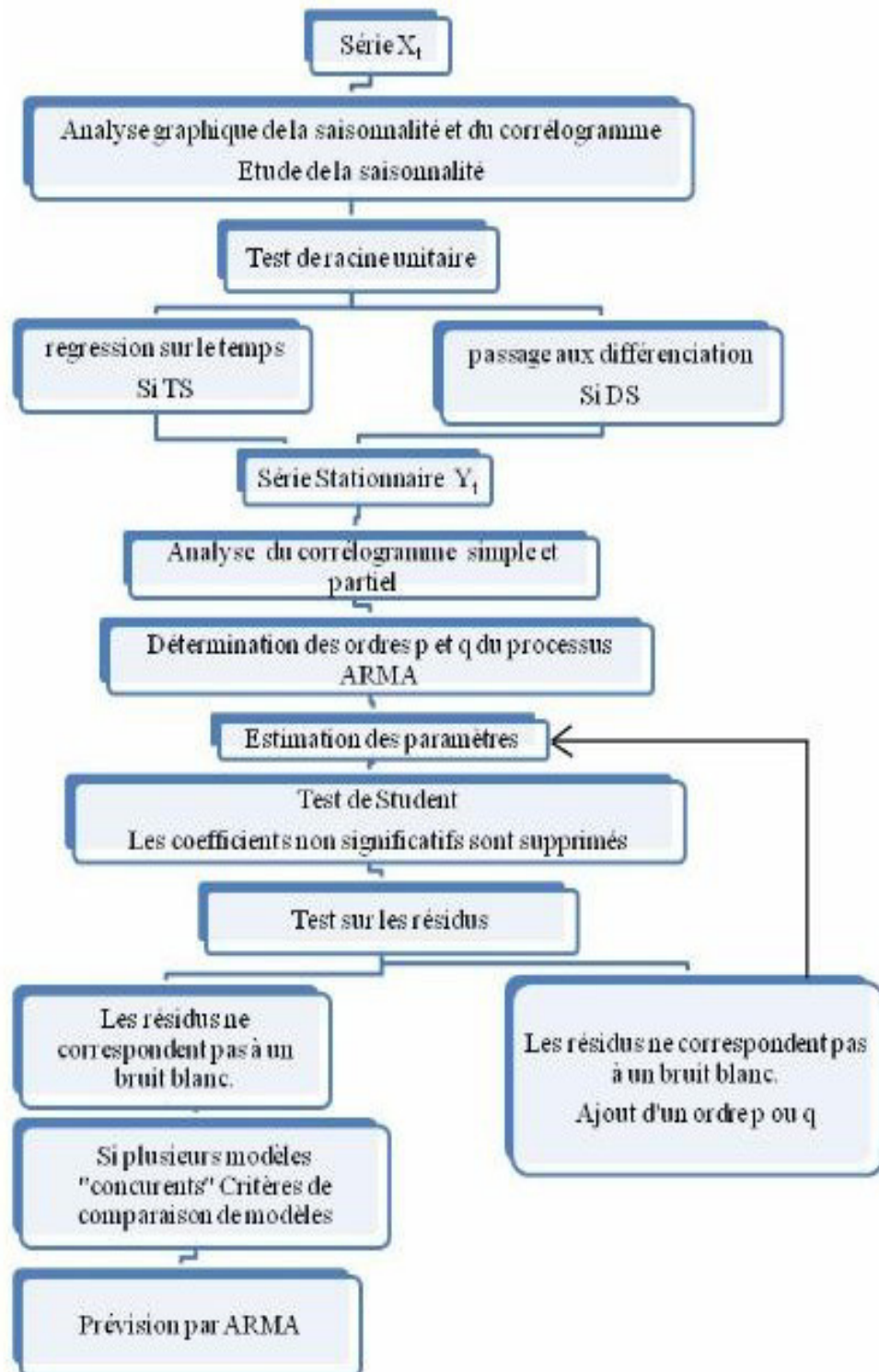


Figure IV. 4: Synthèse de Box et Jenkins. [5]

## **Chapitre IV      Méthodologie de box-jenkins**

---

- **L'interprétation des résultats :**

Elle n'est pas toujours aisée. Les éléments les plus importants, pour lesquels il faudrait trouver une explication, sont les opérations de différence utilisées et éventuellement les constantes.

Ceux-ci déterminent en effet le comportement de la prévision à long terme. Le polynôme autorégressif joue également un rôle essentiel sur la fonction de prévision. En effet, pour celui-ci les racines réelles induisent une composante amortie dans la fonction de prévision. Mais, les racines complexes donnent lieu à une composante pseudopériodique amortie. [5]

### **Conclusion:**

Ce chapitre constitue un arrière-plan théorique de la méthode de modélisation stochastique de Box-Jenkins et sachant que la procédure est valable pour tout autre domaine.

Et nous pouvons dire pour cette méthode, qu'en dépit de sa simplicité dans son principe de base, elle semble complexe dans sa mise en œuvre, d'autant plus qu'elle ne prend pas en compte les informations externes. Néanmoins, elle permet non seulement de traiter et d'analyser les processus stationnaires mais également les processus qui ne le sont pas.



# *Chapitre V*

**Chapitre V : application de méthode Box Jenkins par R :****1-présentation du logiciel R**

Le système R connaît depuis plus d'une décennie une progression remarquable dans ses fonctionnalités, dans la variété de ses domaines d'application ou, plus simplement, dans le nombre de ses utilisateurs. La documentation disponible suit la même tangente : plusieurs maisons d'édition proposent dans leur catalogue des ouvrages — voire des collections complètes — dédiés spécifiquement aux utilisations que l'on fait de R en sciences naturelles, en sciences sociales, en finance, etc. Néanmoins, peu d'ouvrages se concentrent sur l'apprentissage de R en tant que langage de programmation sous-jacent aux fonctionnalités statistiques. C'est là une niche que nous tâchons d'occuper.

R est un environnement permettant de faire des analyses statistiques et de produire des graphiques. C'est un logiciel libre, clone d'un autre logiciel très célèbre dans la communauté statisticienne S+ (notez le clin d'œil). Il peut être téléchargé gratuitement sur [www.r-project.org](http://www.r-project.org). Sur ce site vous trouverez également des documentations très complètes, notamment le manuel d'Emmanuel Paradis, chercheur en biologie à... Montpellier. Comme tous les logiciels libres, le développement et l'amélioration de R peuvent être effectués partout un chacun. Actuellement, R propose bien plus de fonctions statistiques que vous ne pourrez en utiliser pendant ce cours. Les développeurs insistent sur le fait que R permet un calcul vectoriel lui ouvrant des applications dans d'autres domaines que les statistiques. R est plus précisément un langage orienté-objet interprété. Outre les diverses procédures intégrées, R est très connu pour son interface graphique souple et qui permet d'exporter des graphiques de très bonne qualité à des formats variés très simplement. Enfin, contrairement à d'autres logiciels « boîtes noires », R oblige son utilisateur à une programmation minimale, ce qui le rend également plus souple que SAS ou SPSS.

**1-1-****Installation du système  
de base**

Le site principal du logiciel R est [www.r-project.org](http://www.r-project.org). Le téléchargement de R se fait à partir d'un des sites du "Comprehensive R archive Network" (CRAN), par exemple [cran.cict.fr](http://cran.cict.fr)

L'interface R:

**1-2-gestion des packages**

Un « package », ou bibliothèque, est généralement, un ensemble de fonctions R qui sont

mises à la disposition des utilisateurs pour effectuer certains traitements qui n'existaient pas dans R

Les packages présentés sont de deux types :

— packages de « base » au nombre de 14 : base, compiler, datasets, graphics, grDevices, grid, methods, parallel, splines, stats, stats4, tcltk, tools, utils. Ils constituent en quelque sorte, le cœur du logiciel. Ils définissent le langage, permettent de faire les analyses courantes, donnent accès aux fonctions de base (mathématique, etc.) et aux fonctions permettant de faire les graphiques.

— packages « recommended » au nombre de 15 : boot, class, cluster, codetools, foreign, KernSmooth, lattice, MASS, Matrix, mgcv, nlme, nnet, rpart, spatial, survival. Ces packages complètent les fonctions de base. Les packages installés contiennent de très nombreuses fonctions mais évidemment pas toutes les fonctions dont on pourrait avoir besoin. Cela n'est pas un problème puisque, R étant un langage de programmation, on peut écrire ses propres fonctions et les utiliser de la même façon que les fonctions déjà présentes.

Il existe plusieurs manières d'installer un nouveau package. L'installation peut se faire :

— à partir de l'interface graphique de R : menu Packages de la fenêtre R Console ;

Dans ce menu, vous sélectionnez l'item Installer le(s) package(s) . . . R vous demande alors de choisir le site à partir duquel il va effectuer le téléchargement.

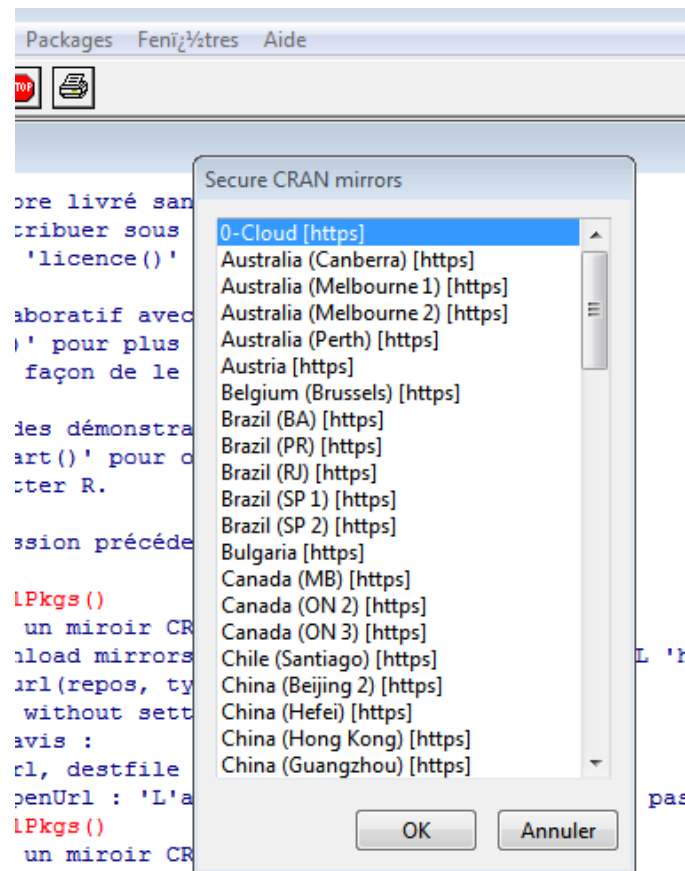
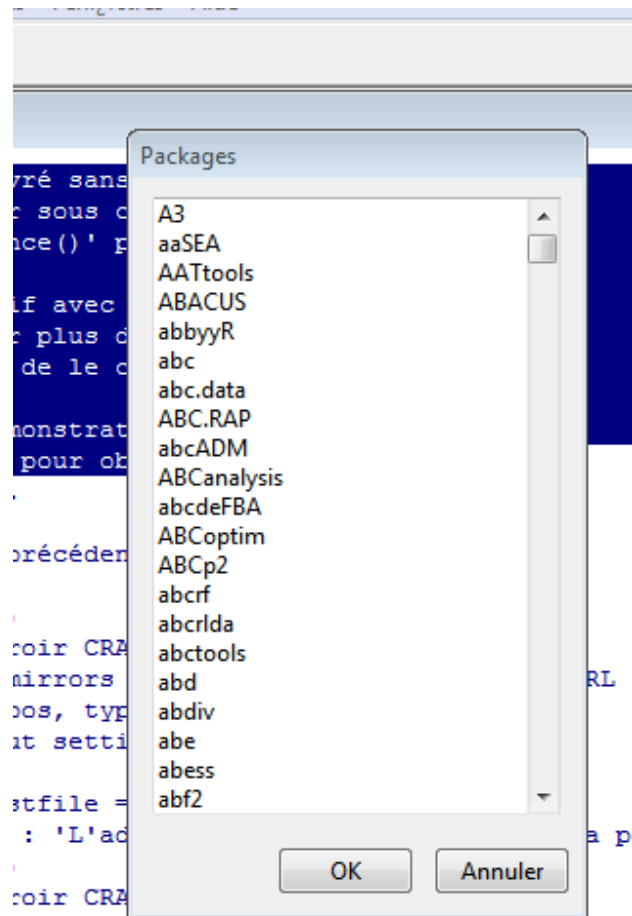


Figure : v-1 Cran de programme R



**Figure : V-2 Installes de package de programme R**

Demême:

— en utilisant la fonction `install.package()`.

Importer/exporter des données

1. Importer une suite : `x=scan("data.dat")` : pour créer un vecteur `a partir de données stockées dans un fichier, icidata.dat.
2. Importer un tableau : `x=read.table("data.dat")` ou `x=read.table("data.dat", header=TRUE)` L'instruction `header=TRUE` permet de préciser que la première ligne du fichier contient le nom des colonnes du tableau.
3. Exporter: `write`, `write.table`

### 1-3-manipulation(Graphiquesetfigures)

`windows()` ouvre une nouvelle fenêtre graphique  
`pdf()`, `png()`, `jpeg()`, `bitmap()`, `xfig()`, `pictex()`, `postscript()` : pilote graphique produisant des sorties dans des fichiers plutôt qu'à l'écran  
`dev.off()` ferme le pilote de sortie graphique pour clore le fichier de sortie  
`(mfrow=c(n,m))`, on obtient alors  $n \times m$  graphiques sur une même page répartis sur  $n$  lignes et  $m$  colonnes.

### Commandes graphiques haut niveau

`plot(x)` trace les valeurs contenues dans  $x$  sur l'axe des  $y$ ; s'adapte à la classe de l'objet  $x$   
`plot(x,y)` graphe bivarié ( $x$  sur l'axe des  $x$ ,  $y$  sur l'axe des  $y$ )

`hist(x)` histogramme des fréquences de  $x$

`curve(expr)` trace la fonction définie par l'expression  $exp$ .

- **Application**

Nous intéressons à appliquer la méthode de Box-Jenkins (voir la figure 3.1) sur des données réelles notée "apport" qui représente les apports de barrage Boukerdan wilaya de Tibaza, la période retenue pour l'étude entre janvier 1995 et août 2016.

## Chapitre V

## Application methode de BOX- JENKINS

Tableau V-1 : les Apport

An / Mois	sept	oct	nov	déc	janv	févr	mars	avr	mai	juin	juil	AU
1995-96	0	0	0	0	2,268	2,89	2,914	11,069	3,66	0,966	0,39	0
1996-97	0,471	1,1	0,379	0,353	0,814	1,405	0,357	1,215	0,45	0,137	0,126	0
1997-98	0,154	0,537	1,805	1,406	0,545	0,583	0,685	0,789	1,838	0,285	0,124	0
1998-99	0,293	0,269	0,81	0,84	6,035	8,212	14,521	1,621	0,525	0,224	0,247	0
1999-00	0,26	0,282	0,959	12,327	4,306	2,033	1,07	0,569	0,658	0,16	0,186	0
2000-01	0,148	0,413	1,019	0,435	1,948	1,05	0,665	0,75	0,464	0,08	0,087	0
2001-02	0,004	0,254	6,907	1,086	0,936	0,977	2,527	0,573	0,278	0,104	0,233	0
2002-03	0,099	0,157	1,152	1,404	3,503	6,802	3,499	2,538	1,258	0,414	0,136	0
2003-04		0,051	0,781	0,776	1,63	0,88	1,83	2,3	1,59	0,52	0	0
2004-05	0,06	0,03	0,8	1,52	2,45	7,98	13,31	2,57	0,76	0,19	0,170	0
2005-06		0,17	0,38	3,77	3,91	8,99	4,29	1,07	7,26	0,8	0,38	
2006-07	0,7	0,17	0,17	3,74	0,99	1,01	5,1	7,34	1,66	0,36	0,05	0
2007-08	0,46	0,96	9,13	6,94	5,38	1,48	2,07	0,7	0,37	0,33	0,04	0
2008-09	0,18	0,38	0,92	3,13	5,54	2,2	1,92	1,82	0,52	0,07	0,08	0
2009-10	0,28	0,121	0,37	1,77	3,674	5,418	10,288	2,532	1,144	0,284	0,048	0
2010-11	0,087	0,545	0,912	1,895	1,965	7,743	5,42	2,288	6,959	2,159	0,266	0
2011-12	0,088	0,243	0,811	1,379	1,233	18,431	9,816	8,014	2,17	0,609	0,039	0
2012-13	0,078	0,31	1,323	0,745	1,555	2,409	2,626	3,607	8,076	1,698	0,44	0
2013-14	0,04	0,08	1,09	2,46	2,36	1,81	11,31	2,05	0,86	1	0,07	0
2014-15	0,02	0,35	0,5	1,4	1,87	5,95	6,01	1,87	0,37	0,14	0,44	0
2015-16	0,6	0,47	0,5	0,28	0,8	1,53	9,84	7,41	1,42	0,2	0,04	0

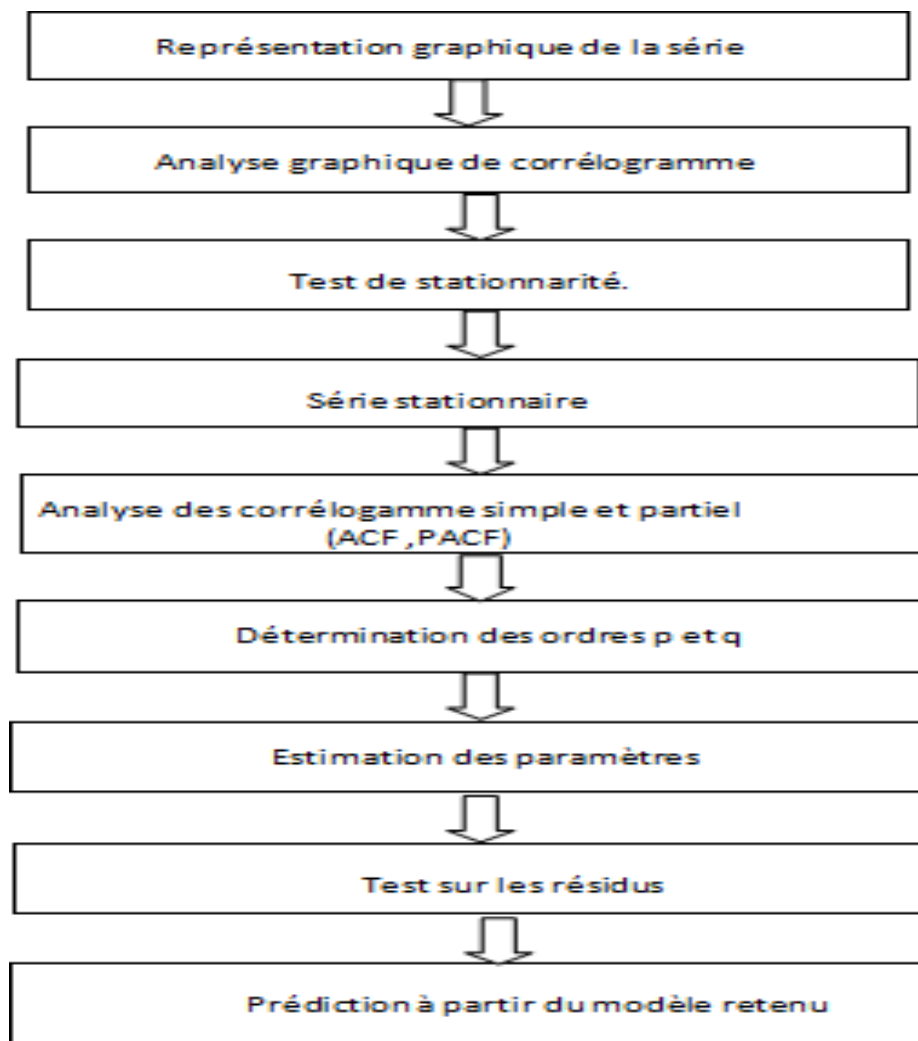


FIGURE V-1:Procédure deconstructionde laméthodeBox-Jenkins.



Le graphe de la série apport diff mon tre que la série est stationnaire. On utilise aussi le test de racine unitaire ( test ADF) pour confirmer la stationnarité.

Cette test donne un p-value= 0.01 <0.05 ( la probabilité critique affectées à la constante est inférieure à 0.05).

**adf. test (apport)**

**Augmented Dickey-Fuller Test**

**data: apport**

**Dickey-Fuller = , Lag order = 6, p-value = 0.1**

**alternative hypothèses: non stationnaire**

- **Test d'influence du facteur ligne (mois):**

H0: Pas d'influence du facteur mois: pas de saisonnalité.

H1: Il y a d'influence du facteur mois: existence de la saisonnalité.

- **Test d'influence du facteur colonne (années):**

H0 : La série n'est pas affectée d'une

tendance. H1: La série est tendancielle.

Nous avons utilisé le logiciel Excel pour l'analyse de variance

à parti

le test de l'ANOVA, les résultats résumés dans le tableau suivant

Test de stationnarité :

Tableau V-2 : Test de stationnarité

<i>RAPPORT DÉTAILLÉ</i>	<i>Nombre d'échantillons</i>	<i>Somme</i>	<i>Moyenne</i>	<i>Variance</i>
1996	12	26,681	2,22341667	9,21971717
1997	12	8,575	0,71458333	0,35685408
1998	12	7,243	0,60358333	0,21679863
1999	12	45,551	3,79591667	27,0872944
2000	12	11,172	0,931	1,42600891
2001	12	13,338	1,1115	3,65985755
2002	12	8,44	0,70333333	0,55306442
2003	12	19,785	1,64875	4,3263693
2004	12	11,16	0,93	0,6768
2005	12	31,89	2,6575	16,7040568
2006	12	31,63	2,63583333	9,12142652
2007	12	34,02	2,835	10,9982636
2008	12	14,98	1,24833333	2,57065152
2009	12	14,741	1,22841667	2,55175917
2010	12	26,999	2,24991667	9,13728281
2011	12	29,458	2,45483333	7,44697997
2012	12	42,859	3,57158333	32,3838848
2013	12	24,207	2,01725	4,98711057
2014	12	21,73	1,81083333	9,60839015
2015	12	18,5	1,54166667	4,65954242
janvier	20	52,912	2,6456	2,78694962
février	20	88,253	4,41265	19,3657102
mars	20	100,228	5,0114	19,6461286
avril	20	55,285	2,76425	7,86218072
mai	20	40,87	2,0435	6,08221184
juin	20	10,53	0,5265	0,31363237
juillet	20	3,552	0,1776	0,01994509
août	20	2,041	0,10205	0,00879994
septembre	20	4,022	0,2011	0,04322631
octobre	20	6,892	0,3446	0,07783794
novembre	20	30,718	1,5359	5,18742841
décembre	20	47,656	2,3828	7,94164533

**ANALYSE DE VARIANCE :**

**Analyse de variance: deux facteurs sans répétition d'expérience :**

**Tableau V- 3:**Analyse de variance: deux facteurs sans répétition d'expérience

<i>RAPPORT DÉTAILLÉ</i>	<i>Nombre d'échantillons</i>	<i>Somme</i>	<i>Moyenne</i>	<i>Variance</i>
			-	
1997	12	-17,402	1,45016667	8,6806307
1998	12	1,957	0,16308333	0,56602863
1999	12	25,002	2,0835	21,049625
2000	12	-11,319	-0,94325	31,325622
			-	
2001	12	-14,407	1,20058333	11,9619915
2002	12	5,67	0,4725	3,42839645
2003	12	7,334	0,61116667	6,83243342
			-	
2004	12	-10,855	0,90458333	2,94600154
2005	12	19,622	1,63516667	13,8422452
2006	12	1,19	0,09916667	12,2338629
			-	
2007	12	-9,68	0,80666667	12,4152061
2008	12	6,37	0,53083333	14,7536811
			-	
2009	12	-11,05	0,92083333	6,73408106
2010	12	9,035	0,75291667	7,29389845
2011	12	4,216	0,35133333	6,33343642
2012	12	13,682	1,14016667	16,1157369
			-	
2013	12	-20,75	1,72916667	29,959692
2014	12	0,137	0,01141667	12,2365328
			-	
2015	12	-4,21	0,35083333	4,2442447
janvier	19	0,02	0,00105263	0,0586975
février	19	0,35	0,01842105	0,17247181
mars	19	0,5	0,02631579	12,2494552
avril	19	1,4	0,07368421	17,372892
			-	
mai	19	-0,398	0,02094737	5,07566316

## Chapitre V

## Application methode de BOX- JENKINS

juin	19	3,06	0,16105263	38,2656757
juillet	19	3,096	0,16294737	48,7289433
août	19	-9,199	0,48415789	13,3789758
septembre	19	-3,29	0,17315789	13,0478157
octobre	19	-0,826	0,04347368	0,60011949
novembre	19	0,05	0,00263158	0,05464125
			-0.001163	
décembre	19	-0,221		0,08952558

**Tableau V-4** :ANALYSE DE VARIANCEavecprobabilite

<i>Source des variations</i>	<i>Somme des carrés</i>	<i>Degré de liberté</i>	<i>Moyenne des carrés</i>
Lignes	237,280802	18	13,1822668
Colonnes	6,05983988	11	0,55089453
Erreur	2446,42698	198	12,3556918
Total	2689,76762	227	
		Valeur critique	
F	Probabilité	pour F	
1,06689832	0,38826807	1,6561749	
0,0445863	0,99999861	1,8372598	

**Tableau V-5 : La série apport stationnaire âpre le test la Nova**

	janvier	février	mars	avril	mai	juin	juillet	août	septembre	octobre	novembre	décembre
1997,00	0,47	1,10	0,38	0,35	-1,45	-1,49	-2,56	-9,85	-3,21	-0,83	-0,26	-0,05
1998,00	-0,32	-0,56	1,43	1,05	-0,27	-0,82	0,33	-0,43	1,39	0,15	0,00	0,01
1999,00	0,14	-0,27	-1,00	-0,57	5,49	7,63	13,84	0,83	-1,31	-0,06	0,12	0,16
2000,00	-0,03	0,01	-0,22	11,49	-1,73	-6,18	-13,45	-1,05	0,13	-0,06	-0,06	-0,16
2001,00	-0,11	0,13	0,43	-11,89	-2,36	-0,98	-0,41	0,18	-0,19	0,64	-0,10	0,26
2002,00	-0,14	-0,16	5,89	0,65	-1,01	-0,07	1,86	-0,18	-0,19	-0,70	0,15	-0,43
2003,00	0,10	-0,10	-5,76	0,32	2,57	5,83	0,97	1,97	0,98	0,31	0,13	0,03
2004,00	-0,10	-0,11	-0,37	-0,63	-1,87	-5,92	-1,67	-0,24	0,33	0,11	-0,36	-0,03
2005,00	0,06	-0,02	0,02	0,74	0,82	7,10	11,48	0,27	-0,83	-0,33	0,17	0,14
2006,00	-0,06	0,14	-0,42	2,25	1,46	1,01	-9,02	-1,50	6,50	0,61	0,21	0,01
2007,00	0,70	0,00	-0,21	-0,03	-2,92	-7,98	0,81	6,27	-5,60	-0,44	-0,33	0,05
2008,00	-0,24	0,79	8,96	3,20	4,39	0,47	-3,03	-6,64	-1,29	-0,03	-0,01	-0,20
2009,00	-0,28	-0,58	-8,21	-3,81	0,16	0,72	-0,15	1,12	0,15	-0,26	0,04	0,05
2010,00	-0,15	-0,26	-0,55	-1,36	-1,87	3,21	8,37	0,71	0,62	0,21	-0,03	0,12
2011,00	0,06	0,42	0,54	0,13	-2,02	2,33	-4,87	-0,24	5,82	1,88	0,22	-0,04
2012,00	0,00	-0,30	-0,10	-0,52	-0,42	10,69	4,40	5,73	-4,79	-1,55	-0,23	0,77
2013,00	-0,01	0,07	0,51	-0,63	0,32	-16,02	-7,19	-4,41	5,91	1,09	0,40	-0,78
2014,00	-0,04	-0,23	-0,23	1,72	0,81	-0,60	8,68	-1,56	-7,22	-0,70	-0,37	-0,13
2015,00	-0,02	0,27	-0,59	-1,06	-0,49	4,14	-5,30	-0,18	-0,49	-0,86	0,37	0,00
2016	0,58	0,12	0,00	-1,12	-1,07	-4,42	3,83					

Analyse préliminaire de la série Apport:

Représentation graphique:

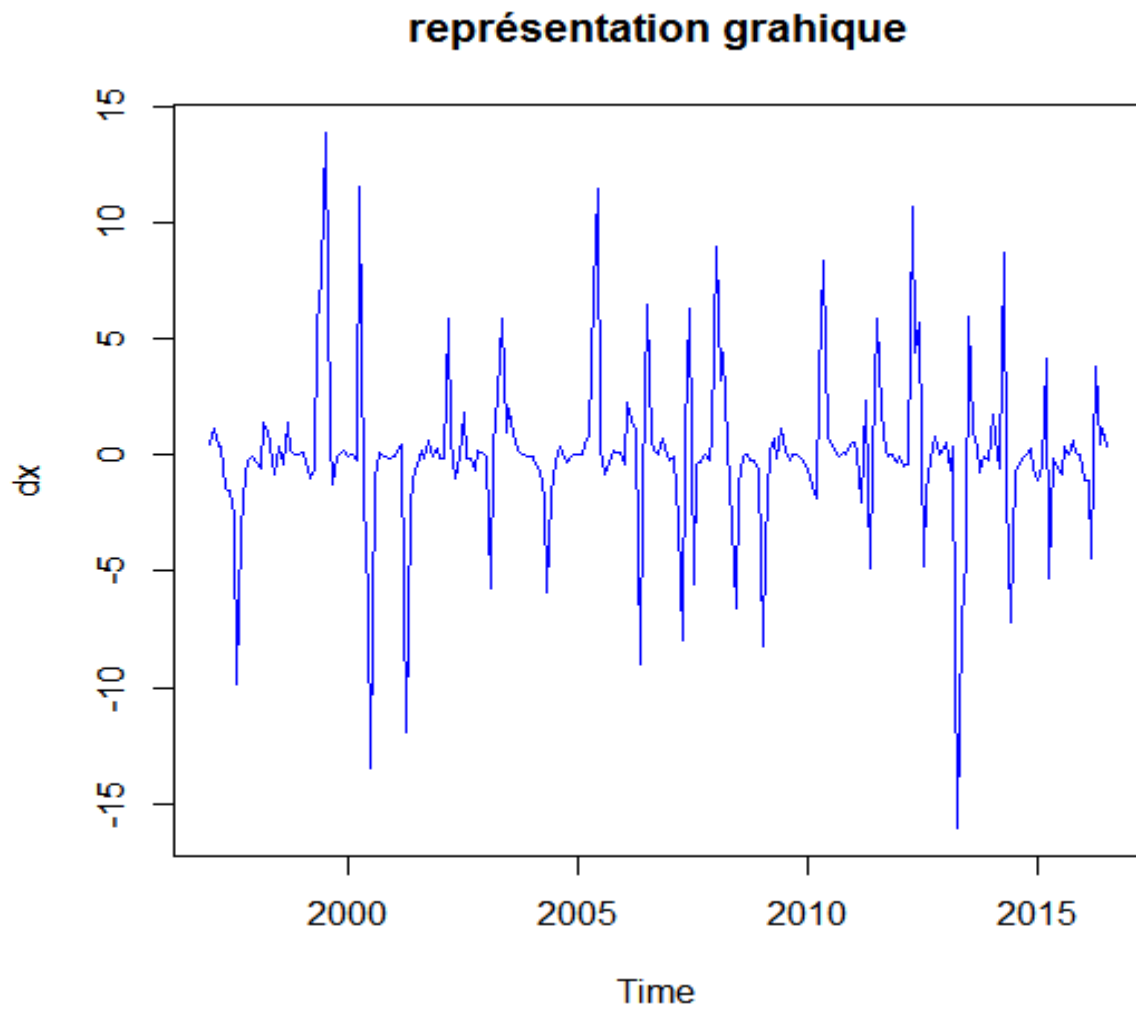
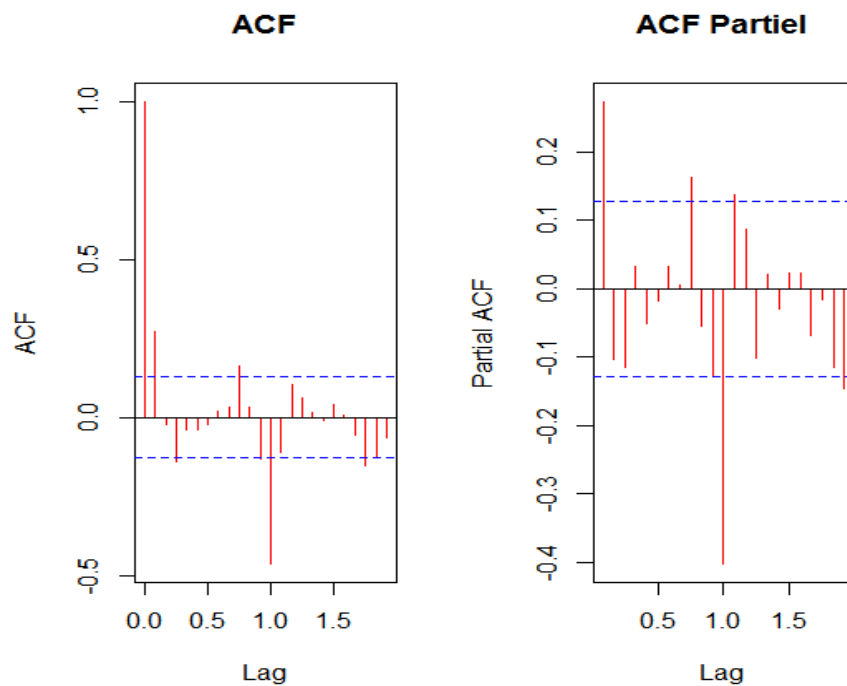


Figure V-6: representation graphique d apport

Etude du corrélogramme de la série Apport :



**Figure V-7: corrélogramme de la série Apport**

La lecture visuelle des graphes ci -dessus montre que la série dx est stationnaire , tout cela va etre vérifié par le test ADF.

**Test de Dickey fuller (stationnarité):**

Augmented Dickey-Fuller Test

data: dx

Dickey-Fuller = -5.7758, Lag order = 6, p-value = 0.01

alternative hypotesis: stationary

Message d'avis :

Dans `adf.test(dx)` : p-value smaller than printed p-value

ce test donne une p-value = 0.01 < 0.05 (la probabilité critique est inférieure 0.05), ce qui montre que la série est stationnaire.

### **Auto-selection du Meilleur modèle :**

Fitting models using approximations to speed things up...

ARIMA(2,0,2)(1,0,1)[12] with non-zero mean : 1172.14

ARIMA(0,0,0) with non-zero mean : 1247.159

ARIMA(1,0,0)(1,0,0)[12] with non-zero mean : 1176.516

ARIMA(0,0,1)(0,0,1)[12] with non-zero mean : 1169.467

ARIMA(0,0,0) with zero mean : 1245.133

ARIMA(0,0,1) with non-zero mean : 1230.172

ARIMA(0,0,1)(1,0,1)[12] with non-zero mean : 1165.699

ARIMA(0,0,1)(1,0,0)[12] with non-zero mean : 1176.308

ARIMA(0,0,1)(2,0,1)[12] with non-zero mean : 1176.713

ARIMA(0,0,1)(1,0,2)[12] with non-zero mean : 1167.798

ARIMA(0,0,1)(0,0,2)[12] with non-zero mean : 1171.493

ARIMA(0,0,1)(2,0,0)[12] with non-zero mean : 1183.819

ARIMA(0,0,1)(2,0,2)[12] with non-zero mean : 1178.001

ARIMA(0,0,0)(1,0,1)[12] with non-zero mean : 1176.033

ARIMA(1,0,1)(1,0,1)[12] with non-zero mean : 1168.773

ARIMA(0,0,2)(1,0,1)[12] with non-zero mean : 1167.64

ARIMA(1,0,0)(1,0,1)[12] with non-zero mean : 1167.259

ARIMA(1,0,2)(1,0,1)[12] with non-zero mean : 1170.357

ARIMA(0,0,1)(1,0,1)[12] with zero mean : 1163.918

ARIMA(0,0,1)(0,0,1)[12] with zero mean : 1167.401

ARIMA(0,0,1)(1,0,0)[12] with zero mean : 1174.274

ARIMA(0,0,1)(2,0,1)[12] with zero mean : 1174.906



ARIMA(0,0,1)(1,0,2)[12] with zero mean : 1165.989  
ARIMA(0,0,1) with zero mean : 1228.125  
ARIMA(0,0,1)(0,0,2)[12] with zero mean : 1169.408  
ARIMA(0,0,1)(2,0,0)[12] with zero mean : 1181.848  
ARIMA(0,0,1)(2,0,2)[12] with zero mean : 1176.245  
ARIMA(0,0,0)(1,0,1)[12] with zero mean : 1174.508  
ARIMA(1,0,1)(1,0,1)[12] with zero mean : 1166.958  
ARIMA(0,0,2)(1,0,1)[12] with zero mean : 1165.8  
ARIMA(1,0,0)(1,0,1)[12] with zero mean : 1165.439  
ARIMA(1,0,2)(1,0,1)[12] with zero mean : 1168.447

Now re-fitting the best model(s) without approximations...

ARIMA(0,0,1)(1,0,1)[12] with zero mean : 1168.181

Best model: ARIMA(0,0,1)(1,0,1)[12] with zero mean

Series: x

ARIMA(0,0,1)(1,0,1)[12] with zero mean

Coefficients:

ma1 sar1 sma1

0.2368 -0.0205 -0.5542

s.e. 0.0609 0.1391 0.1296

sigma<sup>2</sup> estimated as 8.093: log likelihood=-580

AIC=1168.01 AICc=1168.18 BIC=1181

**Estimation du meilleur modèle :**

initial value 1.229582

iter 2 value 1.113819

iter 3 value 1.048069

iter 4 value 1.046539

iter 5 value 1.042214

iter 6 value 1.041208

iter 7 value 1.040251

iter 8 value 1.040111

iter 9 value 1.040103

iter 10 value 1.040094

iter 11 value 1.040093

iter 12 value 1.040093

iter 13 value 1.040092

iter 13 value 1.040092

iter 13 value 1.040092

final value 1.040092

converged

initial value 1.049519

iter 2 value 1.049501

iter 3 value 1.049163

iter 4 value 1.049163

iter 5 value 1.049162

iter 6 value 1.049162

iter 7 value 1.049162

iter 7 value 1.049162

iter 7 value 1.049162

final value 1.049162

converged

>summary(Model)

	Length	Class	Mode
fit	14	Arima	list
degrees_of_freedom	1	-none-	numeric
ttable	12	-none-	numeric

```
AIC          1  -none- numeric
AICc         1  -none- numeric
BIC          1  -none- numeric
>Model$fit
```

Call:

```
arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D, Q), period = S),
xreg = xmean, include.mean = FALSE, transform.pars = trans, fixed = fixed,
optim.control = list(trace = trc, REPORT = 1, reltol = tol))
```

Coefficients:

```
      ma1      sar1      sma1
0.2368 -0.0205 -0.5542
s.e. 0.0609 0.1391 0.1296
```

sigma<sup>2</sup> estimated as 7.99: log likelihood = -580, aic = 1168.01

diagnostique des résidus :

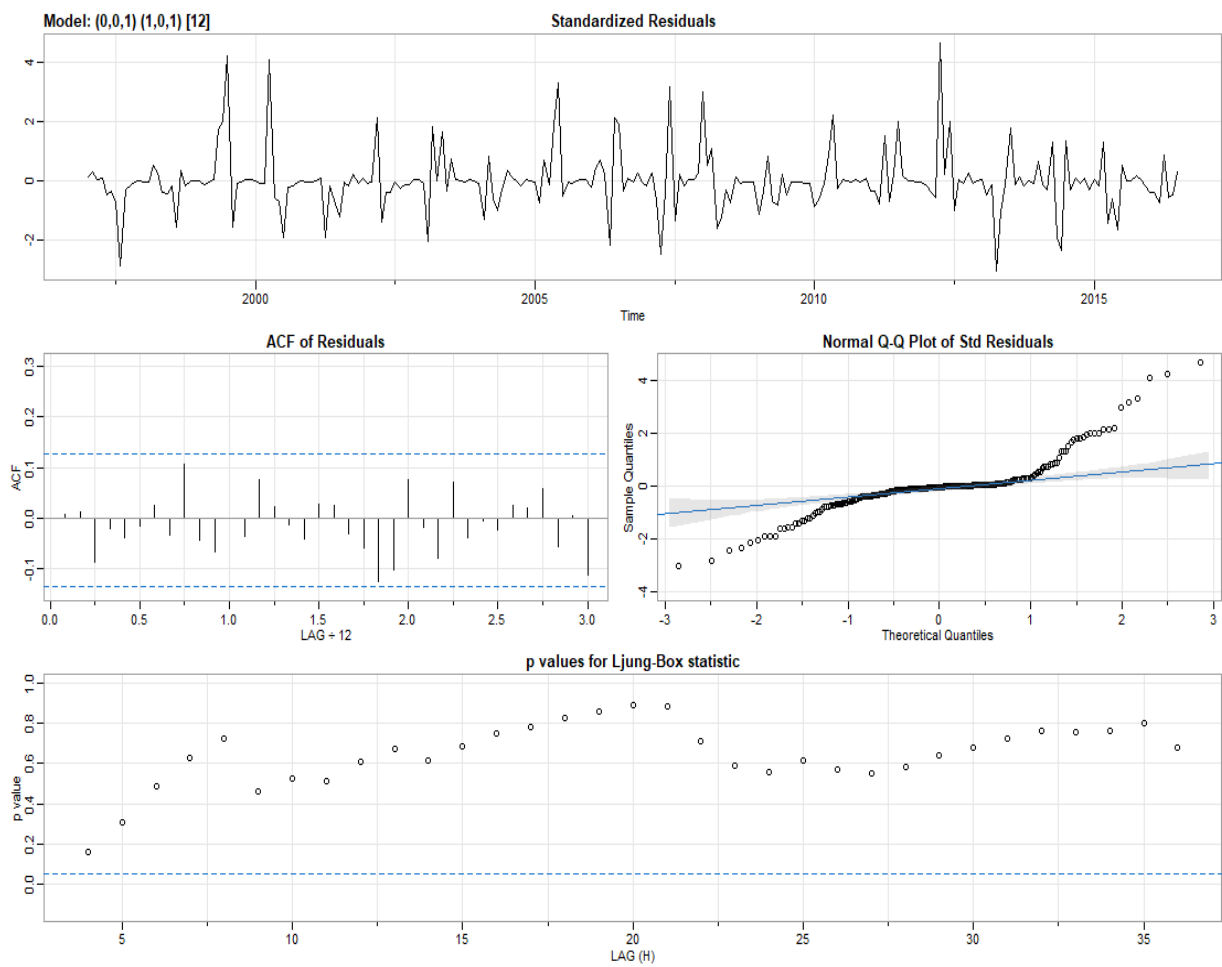
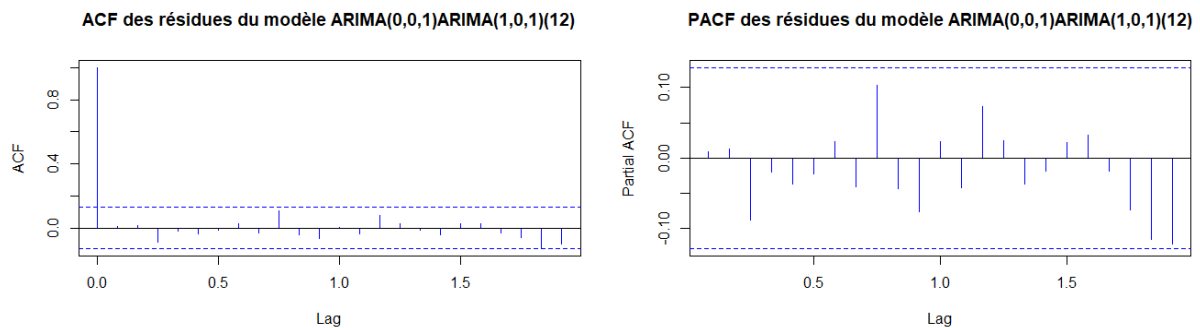


Figure V-8 :diagnostique des résidus

**ACF et PACF des residues du meilleur modèle :**



**Figure V- 9:**ACF et PACF des residues du meilleur modèle

**test de bruit blanc (test de Box - Pierce) du meilleur modèle :**

Box-Pierce test

X-squared = 6.9636, df = 12, p-value = 0.86

**test de normalité des résidus (test de JarqueBera) :**

JarqueBera Test

X-squared = 322.57, df = 2, p-value < 2.2e-16

**test ARCH :**

ARCH LM-test; Null hypothesis: no ARCH effects

Chi-squared = 32.635, df = 12, p-value = 0.001103

**Ecriture du modèle :**

$$apports_t = -0,0205apports_{t-12} + \varepsilon_t + 0,2368\varepsilon_{t-1} - 05542\varepsilon_{t-12}$$

**Prévisions mensuelles de la série “aprts”:**

Tableau 9: Prévisions mensuelles de la série “aprts”

Mois	Prévisions	Écart-type de prévisions
Aug. 2016	-0.03369241	2.826625
Sep. 2016	-0.01680538	2.904777
Oct. 2016	-0.26522761	2.904777
Nov. 2016	-0.14453379	2.904777
Dec.2016	0.23065365	2.904777
Jan. 2017	0.65748356	2.904777
Feb. 2017	0.7536302	2.904777
Mar. 2017	1.39159862	2.904777
Apr. 2017	-1.16925979	2.904777
May. 2017	0.5583954	2.904777
Jun. 2017	0.91326153	2.904777
Jul. 2017	-0.3142305	2.904777

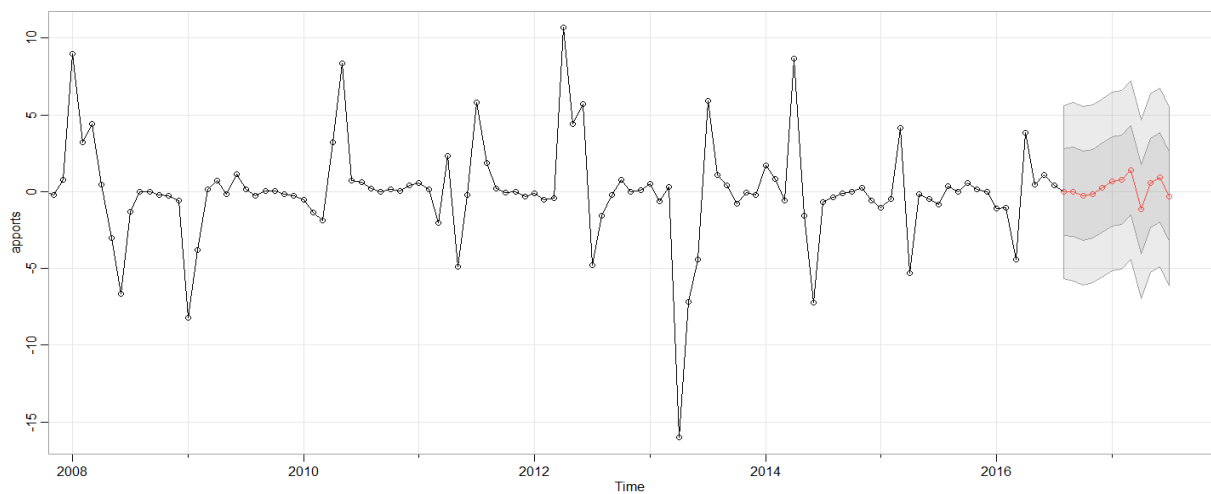


Figure V- 9 : la prévision

## Conclusion Générale:

En guise de conclusion générale, nous allons tenter d'établir une synthèse globale sur le travail qui a été réalisé dans ce mémoire. L'objectif principal de ce mémoire est de développer des modèles statistiques qui peuvent être utilisés pour prédire le nombre de sacs d'APPORT d'une série temporelle réelles, qui sera valable pour des prédictions futures.

Les résultats de cette étude confirment la capacité des modèles autorégressives dans le domaine de la prédiction des séries temporelles. Ce qui est vérifié par les critères de performances et de précision, l'utilisation de la méthode de Box-Jenkins pour construire les différents modèles : AR, MA, SARIMA qui peuvent être utilisés pour résoudre le problème de la prédiction les résultats de la prédiction ont été acceptables.

D'après l'analyse de la série additive des apports et le test de Dickey-Fuller on a trouvé que le modèle est de type SARMA (ARIMA(0,0,1)(1,0,1)[12] with zero mean : (1168.181))

$$\text{apports}_t = -0,0205 \text{apports}_{t-12} + \varepsilon_t + 0,2368 \varepsilon_{t-1} - 0,5542 \varepsilon_{t-12}$$

## REFERENCES BIBLIOGRAPHIQUES :

- [1] :A. Bermad,  
(1990)«Thèsedemagister:Elaborationd'unmodèlededestiond'unbarrageréservoir
- [2] :A. MARRE 1987,« Le Tell Oriental Algérien de Collo à la Frontière Tunisienne»,EtudeGéomorphologique, Volume1, OfficeDesPublicationsUniversitaires.
- [3] : BOURBONNAIS Régis:«Econométrie»,4èmeéditionDunod,Paris2002
- [4] : BOURBONNAIS Régis–  
TERRAZAMichel:«Analysedessériestemporelles»,éditionDunod, Paris1998, page149.
- [5] CfBOURBONNAIS R.etTERRAZAM.chapitre5,2010.
- [6] :C.Gourieroux. et A Monfort. "Séries temporelles et modèles dynamiques",Economica,Paris1995.
- [7]:E.Dor:«Econométrie»,PearsonEducationFrance,2009, Page162.
- [8] :E,Parzen,(1974)Somerecentadvancesintimeseriesmodelling.IEEETransactionsonAutomaticControl, AC-19, 723–729 .
- [9] :H,Akaike,(1974)Anewlookatthestatisticalmodelidentification.IEEETransactionsonAutomaticControl. 19(6), 716–723.
- [10] : J.M.Duran et al : Technique des petites barrages en Afrique sahélienne et équatoriale,cemagreféditions–ISBN2-85362 -511-7/1999.
- [11] : kH .Bekai,N.Hammadouche,N .Boukrif : « Etude prévisionnelle des ventes d'huile parlaméthodedeBOXetJENKINS»,MémoiredeMaster,UniversitédeBejaia,Algérie .2013
- [12] : KH.Boukharouba : «Elaboration d'unmodèle stochastique de type Box-Jenkinspour la prédiction simple et filtre de kalman pourla prédiction multi-site des apports desouedsaunordd'Algérie »,MémoiredeDoctorat,Ecolenationalepolytechnique,Algérie.2010.
- [13] :O.D,Anderson,(1976)TimeSeriesAnalysisandForecasting:The BOXJENKINSApproach.Butterworth,London.
- [14] :O. Derdous , 2016 ;Conditions d'écoulement à l'aval d'un barrage en cas de rupturede la digue-Application aux cas Algériens, du diplôme de Doctorat 3ème cycle,



Université Badji Mokhtar Annaba.

[15] : P. G. Box. & Jenkins, P. M, (1970) Time Series Analysis. Holden Day, San Francisco, USA.

[16] : R. Bourbonnais– M. Terraza: « Analyse des séries temporelles », édition Dunod, Paris 2004.

[17] : S. Affoune " Ressources en eaux, mobilisation et utilisation dans le bassin versant de lamafragh, mémoire de magister, université Mentouri de Constantine, Algérie 2006.

[18] : S. Senouci : « Essais d'application des modèles de prévision univariés sur la consommation d'énergie électrique en Algérie », Mémoire de Magister, Université d'Oran, Algérie. 2012.

[19] : Wikipédia: site internet

[https : www.edf.fr](https://www.edf.fr) (se site EDF présente les barrages et leurs fonctionnements). <https://pbil.univ-lyon1.fr/R/pdf/tdr29.pdf>.

[20] : R. Bourbonnais– M. Terraza: « Analyse des séries temporelles », édition Dunod, Paris 2004.

[21] : S. Affoune " Ressources en eaux, mobilisation et utilisation dans le bassin versant de lamafragh, mémoire de magister, université Mentouri de Constantine, Algérie 2006.

[22] : S. Senouci : « Essais d'application des modèles de prévision univariés sur la consommation d'énergie électrique en Algérie », Mémoire de Magister, Université d'Oran, Algérie. 2012.

[23] : Wikipédia: site internet

[https : www.edf.fr](https://www.edf.fr) (se site EDF présente les barrages et leurs fonctionnements). <https://pbil.univ-lyon1.fr/R/pdf/tdr29.pdf>.

# Annex

## **Leprogrammed'applicationsousR :**

```
# Bensalah mohammed Prog.
# Loading packages
library(aTSA)
library(astsa)
library(tseries)
library(e1071)
library(fGarch)
library(forecast)
library(FitARMA)
library(urca)
library(FinTS)
# Loading data...
xdata=c(0.47,1.10,0.38,0.35,-1.45,-1.49,-2.56,-9.85,-3.21,-0.83,-0.26,-0.05,-0.32,
-0.56,1.43,1.05,-0.27,-0.82,0.33,-0.43,1.39,0.15,-0.00,0.01,0.14,-0.27,
-1.00,-0.57,5.49,7.63,13.84,0.83,-1.31,-0.06,0.12,0.16,-0.03,0.01,-0.22,
11.49,-1.73,-6.18,-13.45,-1.05,0.13,-0.06,-0.06,-0.16,-0.11,0.13,0.43,
-11.89,-2.36,-0.98,-0.41,0.18,-0.19,0.64,-0.10,0.26,-0.14,-0.16,5.89,0.65,
-1.01,-0.07,1.86,-0.18,-0.19,-0.70,0.15,0.10,-0.10,-5.76,0.32,2.57,5.83,
0.97,1.97,0.98,0.31,0.13,0.03,-0.10,-0.11,-0.37,-0.63,-1.87,-5.92,-1.67,
-0.24,0.33,0.11,-0.36,-0.03,0.06,-0.02,0.02,0.74,0.82,7.10,11.48,0.27,
-0.83,-0.33,0.17,0.14-0.06,0.14,-0.42,2.25,1.46,1.01,-9.02,-1.50,6.50,
0.61,0.21,0.01,0.70,0.00,-0.21,-0.03,-2.92,-7.98,0.81,6.27,-5.60,-0.44,
-0.33,0.05,-0.24,0.79,8.96,3.20,4.39,0.47,-3.03,-6.64,-1.29,-0.03,-0.01,
-0.20,-0.28,-0.58,-8.21,-3.81,0.16,0.72,-0.15,1.12,0.15,-0.26,0.04,0.05,
-0.15,-0.26,-0.55,-1.36,-1.87,3.21,8.37,0.71,0.62,0.21,-0.03,0.12,0.06,
0.42,0.54,0.13,-2.02,2.33,-4.87,-0.24,5.82,1.88,0.22,-0.04,0.00,-0.30,
-0.10,-0.52,-0.42,10.69,4.40,5.73,-4.79,-1.55,-0.23,0.77,-0.01,0.07,0.51,
-0.63,0.32,-16.02,-7.19,-4.41,5.91,1.09,0.40,-0.78,-0.04,-0.23,1.72,0.81,
-0.60,8.68,-1.56,-7.22,-0.70,-0.37,-0.13,-0.02,0.27,-0.59,-1.06,-0.49,4.14,
-5.30,-0.18,-0.49,-0.86,0.37,0.00,0.58,0.12,0.00,-1.12,-1.07,-4.42,3.83)
```

```

# Convert data to a time serie
x=ts(xdata,start=c(1997,1),end=c(2016,07),frequency=12)

# Plot data
plot(x, col="blue")

# Correlogram of dx
par(mfrow=c(1,2))
acf(x, lag.max = 24, col="red")
pacf(x, lag.max = 24,col="red")
par(mfrow=c(1,1))

# Unit root test (ADF and PP)
adf.test(dx)

# Auto-selection of the best model
auto.arima(x,trace=TRUE)

# Fitting models
Model <- sarima(x, 0, 0, 1, P = 1, D = 0, Q = 1, S = 12, details = TRUE, xreg=NULL, Model=TRUE,
fixed=NULL, no.constant = TRUE)
summary(Model)
Model$fit

# Extract residual nserie
res<- Model$fit$residuals

# donner la série des résids d'un modèle ARMA(1,2)
par(mfcol=c(1,2))
acf(res,main="ACF des résidues du modèle ARIMA(0,0,1)ARIMA(1,0,1)(12)",col="blue")
pacf(res,main="PACF des résidues du modèle ARIMA(0,0,1)ARIMA(1,0,1)(12)",col="blue")
par(mfcol=c(1,1))

# le test de bruit blanc (test de Box - Pièce)

```

```
Box.test(res, lag = 12)
```

```
# le test de bruit blanc
```

```
# test de normalité des résidus
```

```
jarque.bera.test(res)
```

```
# test ARCH
```

```
ArchTest(res)
```