
الجمهورية الجزائرية الديمقراطية الشعبية

Ministère de L'enseignement Supérieur et de la Recherche
Scientifique

UNIVERSITÉ SAAD DAHLEB DE BLIDA

Faculté des sciences

Département de Mathématiques



MEMOIRE DE MASTER

En Mathématiques

Option : Modèles Stochastiques et Statistique

THÈME :

Prévision de la consommation de l'énergie par les modèles ARIMA,
application sur les données de SONELGAZ de Blida

Réalisé par

BENSARI Sarah & HAMIDI Nourimane

Soutenu devant le Jury :

TAMI Omar	Université Blida 1	Président
FRIHI Redhouane	Université Blida 1	Examineur
RASSOUL Abdelaziz	ENSH de Blida	Promoteur

Juillet 2022

Avec l'expression de ma reconnaissance, je dédie ce modeste travail à ceux qui, quels que soient les termes embrassés, je n'arriverais jamais à leur exprimer mon amour sincère.

Quoi que je fasse ou que je dise, je ne saurai point te remercier comme il se doit. Ton affection me couvre, ta bienveillance me guide et ta présence à mes côtés a toujours été ma source de force pour affronter les différents obstacles "mon adorable mère".

Mon précieux offre du dieu, qui doit ma vie, ma réussite et tout mon respect "mon cher père".

A mon adorable soeur "Hind" qui m'a accompagné dans tous les moments de joie et de tristesse.

A toi mon grand-père , ceci est ma profonde gratitude pour ton éternel amour, que ce rapport soit le meilleur cadeau que je puisse t'offrir.

A ma grand-mère, qui je souhaite une bonne santé.

A mes frères, mes soeurs et ceux qui ont partagé avec moi tous les moments d'émotion lors de la réalisation de ce travail. Ils m'ont chaleureusement supporté et encouragé tout au long de mon parcours.

A mes oncles et mes tantes. Que Dieu leur donne une longue et joyeuse vie.

A ma famille, mes proches et à ceux qui me donnent de l'amour et de la vivacité. A tous les cousins, et les amis que j'ai connu jusqu'à maintenant qui m'ont toujours encouragé, et à qui je souhaite plus de succès. Sans oublier mon binôme "Sarah" pour son soutien moral, sa patience et sa compréhension tout au long de ce projet. Merci pour leurs amours et leurs encouragements.

"H.Nourimane"

Tout D'abord , je tiens à remercier DIEU de m'avoir donné la force et le courage de mener à bien ce modeste travail.

C'est avec un très grand plaisir que je dédie ce modeste travail aux personnes les plus chères au monde :

A mes très chers parents qui ont toujours été là pour moi, et qui m'ont donné un magnifique modèle de labeur et de persévérance.

A mes très chères sœurs; Mes frères et mon fiancé : "Mohammed" qui m'avez toujours soutenu et encouragé durant ces années d'étude .

Je tiens à remercier tout particulièrement mon promoteur, RASSOUL Abdelaziz , pour sa patience, sa disponibilité et surtout ses judicieux conseils, qui ont contribué à alimenter ma réflexion.

Et enfin , je tiens à remercier mon binôme "Nour" pour son soutien et sa Compréhension dans les moments difficiles , je vous souhaite de réussir et de réaliser tous vos rêves.

"B. Sarah"

REMERCIEMENTS

Ce travail est le résultat d'un dur labeur et de beaucoup de sacrifices, nos remerciements vont d'abord à ALLAH, créateur de l'univers qui nous a doté d'intelligence, le courage la volonté et nous a maintenu en santé mentale et physique pour mener à bien réaliser ce mémoire.

Nous aimerions exprimer nos profonde gratitude à notre promoteur RASSOUL Abdelaziz avec qui nous avons eu le plaisir de travailler sous sa direction et qui a guidé notre mémoire pour avoir accepté de nous guider par sa patience, sa disponibilité et surtout ses judicieux conseils, qui ont contribué à alimenter notre réflexion, et ses critiques constructives.

Nous remercions également tous les membres du jury pour nous avoir honorés par leur présence et pour avoir accepté d'évaluer notre travail.

Nous désirons exprimer notre reconnaissance envers Mm HAMLATI Nabila, qui nous a apporté son support moral et intellectuel tout au long de notre période de stage au niveau de la société de distribution d'électricité et du gaz -BLIDA- et pour avoir eu la patience de répondre à nos innombrables questions.

Nous tenons à exprimer notre gratitude à l'ensemble des employés et responsables de la direction de distribution d'électricité et du gaz -BLIDA-pour leur contribution et leur disponibilité pour l'apport technique concernant le domaine de travail.

Nous tenons à remercier aussi le directeur de la résidence universitaire SOUMAA-5- Mr HAMLAOUI Amar pour leur soutien et qui nous a accompagné tout au long de notre séjour, et sans oublier aussi de remercier les enseignants, les employés et les responsables de département mathématiques de l'université SAAD DAHLEB -Blida- qui nous ont toujours soutenus et encouragés pendant nos études.

Enfin, nous ne saurons terminer cette partie sans exprimer notre gratitude à nos parents, nos amis, nos famille, nos proches qui nous ont soutenus durant ce travail, et spécialement à tous ce qui nous ont encouragé à finir ce travail et qui nous ont accompagnés dans tous les moments de joie et de tristesse ,merci énormément.

"B. Sarah"

"H. Nourimane"

TABLE DES MATIÈRES

Introduction Générale	1
1 Généralités sur les séries chronologiques	3
1.1 Définition d'une série chronologique	3
1.2 Types des séries chronologiques	4
1.2.1 Séries continues	4
1.2.2 Séries discrètes	4
1.3 Les composantes d'une série chronologique	5
1.4 L'objectif d'une série chronologique	5
1.5 Domaines d'applications des séries chronologiques	6
1.6 Modèles de décomposition d'une série chronologique	8
1.6.1 Modèle additif	8
1.6.2 Modèle multiplicatif	9
1.6.3 Modèle mixte	9
1.7 Choix de modèle	10
1.8 Processus stochastique	10
1.8.1 Bruit blanc	11
1.8.2 Série linéaire	11
1.8.3 Processus stationnaire	12
1.9 Les opérateurs définis sur une série chronologique	13
1.9.1 Opérateur de retard	13

1.9.2	Opérateur de différentiation	13
1.9.2.1	L'effet de l'opérateur de différence	14
1.10	Caractéristiques et propriétés des séries chronologique	14
1.10.1	La moyenne	14
1.10.2	La variance	14
1.10.3	La fonction d'auto covariance	15
1.10.4	Fonction d'auto corrélation (ACF)	15
1.10.5	Fonction d'auto corrélation partielle (PACF)	16
1.11	Test de Dickey Fuller Augmenté	17
1.11.1	Test de Dickey et Fuller augmenté	17
1.11.2	Test de Phillips-Perron	18
1.12	Théorème de décomposition de Wold	18
1.13	Processus aléatoire stationnaire	18
1.13.1	Processus auto régressif d'ordre p : AR(p)	18
1.13.1.1	Les caractéristiques d'un processus auto régressif AR(p)	20
1.13.2	Processus de moyenne mobile d'ordre q : MA(q)	22
1.13.2.1	Les caractéristiques d'un processus moyenne mobile MA(q)	22
1.13.3	Processus auto régressif moyenne mobile d'ordre (p,q) (ARMA)	23
1.13.3.1	Les caractéristiques d'un processus ARMA	24
1.13.4	Processus auto régressif moyenne mobile saisonnier : SARMA	24
1.14	Processus linéaire non stationnaire	25
1.14.1	Processus TS	25
1.14.2	Processus DS	25
1.15	Processus auto-régressif moyenne mobile intégré ARIMA	25
1.16	Processus SARIMA	26
1.17	Processus stochastiques non linéaires	26

1.17.1	Modèles auto régressifs conditionnellement hétéroscédastiques : ARCH	27
1.17.2	Modèles auto régressifs conditionnellement hétéroscédastiques généralisés : GARCH	27
1.18	La méthode de Box&Jenkins	28
1.18.1	Identification du modèle	28
1.18.2	Estimation des paramètres du modèle	29
1.18.3	Validation du modèle	30
1.18.4	Prévision	33
2	Présentation de l'entreprise Sonelgaz	36
2.1	Introduction	36
2.2	Distribution d'électricité et du gaz en Algérie	37
2.3	Présentation de l'entreprise Sonelgaz en ALGERIE	38
2.3.1	Présentation de la société	38
2.3.2	Historique de l'évolution de la société	39
2.3.3	Organisation de Sonelgaz	41
2.4	Activités de Sonelgaz	42
2.4.1	Activités non-réglementées	42
2.4.2	Activités réglementées	42
2.4.3	Énergies renouvelables	42
2.5	Organigramme de l'entreprise Sonelgaz	43
2.6	Société de distribution d'électricité et du gaz -BLIDA-	44
2.6.1	Présentation de la wilaya de Blida	44
2.6.2	Direction de distribution d'électricité et du gaz -BLIDA-	45
2.6.3	Informations sur Sonelgaz de Blida	45
3	Application sur la consommation d'électricité et du gaz à la wilaya de Blida	47
3.1	Présentation des données	47
3.2	Méthodologie de Box & Jenkins	47

3.2.1	Création d'une série chronologique	48
3.2.2	Transformation	48
3.2.3	Identification	49
3.2.4	Estimation des paramètres	49
3.2.5	Tests d'adéquation et choix du modèle (validaion)	50
3.3	Application	50
3.3.1	Application sur les données d'électricité :	50
3.3.1.1	La consommation d'électricité pour les clients ordinaires :	50
3.3.1.2	La consommation d'électricité pour les clients administrations	58
3.3.1.3	La consommation totale d'électricité	65
3.3.2	Application sur les données du gaz	72
3.3.2.1	La consommation du gaz pour les clients ordinaires :	72
3.3.2.2	La consommation du gaz pour les clients administrations :	80
3.3.2.3	La consommation totale du gaz	88
3.4	Conclusion	97

TABLE DES FIGURES

1.1	Consommation d'électricité et du gaz au niveau mondial .	6
1.2	Décès mensuelles dus à des maladies pulmonaires au niveau mondial	7
1.3	Cours de bourse quotidiens	7
1.4	Évolution du rendement de blé tendre d'hiver	7
1.5	Température annuelle moyenne	8
1.6	Évolution de population du monde	8
1.7	Série chronologique(modèle additif)	9
1.8	Série chronologique(modèle multiplication)	9
1.9	Série chronologique(modèle mixte)	10
2.1	Station de distribution électrique	37
2.2	Station de distribution du gaz	37
2.3	Logo de l'entreprise SONELGAZ	38
2.4	Siège de la direction centrale de Sonelgaz	39
2.5	Organigramme de l'entreprise Sonelgaz	43
2.6	Organigramme de l'entreprise Sonelgaz	44
2.7	Carte administrative de la wilaya de Blida	45
2.8	Direction de distribution d'électricité et du gaz -BLIDA- .	46
3.1	La consommation d'électricité pour les clients ordinaires	51

3.2	Corrélogramme simple de la consommation d'électricité pour les clients ordinaires	51
3.3	Corrélogramme partiel de la consommation d'électricité pour les clients ordinaires	52
3.4	La série différencier de la consommation d'électricité (clients ordinaires)	53
3.5	Le corrélogramme simple de la série différencier de la consommation d'électricité (clients ordinaires)	54
3.6	Le corrélogramme partiel de la série différencier de la consommation d'électricité (clients ordinaires)	55
3.7	Le graphe qqnorm des résidus du modèle ARIMA(0,1,0) de la consommation d'électricité (clients ordinaires)	56
3.8	Les valeurs prédictives de la consommation d'électricité (clients ordinaires)	57
3.9	La consommation d'électricité pour les clients administrations	58
3.10	Corrélogramme simple de la consommation d'électricité pour les clients administrations	59
3.11	Corrélogramme partiel de la consommation d'électricité pour les clients administrations	59
3.12	La série différencier de la consommation d'électricité (clients administrations)	60
3.13	Le corrélogramme simple de la série différencier de la consommation d'électricité (clients administrations)	61
3.14	Le corrélogramme partiel de la série différencier de la consommation d'électricité (clients administrations)	62
3.15	Le graphe qqnorm des résidus du modèle ARIMA(5,1,2) pour la consommation d'électricité (clients administrations)	63
3.16	Les valeurs prédictives de la consommation d'électricité (clients administrations)	65
3.17	La consommation totale d'électricité	66

3.18	Corrélogramme simple de la consommation totale d'électricité	66
3.19	Corrélogramme partiel de la consommation totale d'électricité	67
3.20	La série différencier de la consommation totale d'électricité	68
3.21	Le corrélogramme simple de la série différencier de la consommation totale d'électricité	69
3.22	Le corrélogramme partiel de la série différencier de la consommation totale d'électricité	69
3.23	Le graphe qqnorm des résidus du modèle ARIMA (0,1,0) pour la consommation totale d'électricité	71
3.24	Les valeurs prédictives par le modèle ARIMA(0,1,0) pour la consommation totale d'électricité	72
3.25	La consommation du gaz pour les clients ordinaires . . .	73
3.26	Le corrélogramme simple de la consommation du gaz pour les clients ordinaires	73
3.27	Le corrélogramme partiel de la consommation du gaz pour les clients ordinaires	74
3.28	La série différencier de la consommation du gaz pour les clients ordinaires	75
3.29	Corrélogramme simple de la série différencier de la consommation du gaz pour les clients ordinaires	76
3.30	Corrélogramme partiel de la série différencier de la consommation du gaz pour les clients ordinaires	76
3.31	Le graphe qqnorm des résidus du modèle ARIMA de la consommation du gaz (clients ordinaires)	78
3.32	Représentation graphique des valeurs prédictives pour la consommation du gaz (clients ordinaires)	80
3.33	La consommation du gaz pour les clients administrations	81
3.34	Corrélogramme simple de la consommation du gaz pour les clients administrations	81

3.35	Corrélogramme partiel de la consommation du gaz pour les clients administrations	82
3.36	La série différencier de la consommation du gaz (clients administrations)	83
3.37	Corrélogramme simple de la série différencier de la consommation du gaz (clients administrations)	84
3.38	Corrélogramme partiel de la série différencier de la consommation du gaz (clients administrations)	84
3.39	Le graphe qqnorm du modèle ARIMA(2,1,5) pour la consommation du gaz (clients administrations)	86
3.40	Représentation graphique des valeurs prédictives pour la consommation du gaz (clients administrations)	88
3.41	La consommation totale du gaz	89
3.42	Corrélogramme simple de la consommation totale du gaz	89
3.43	Le corrélogramme partiel de la consommation totale du gaz	90
3.44	La série différencier de la consommation totale du gaz	91
3.45	Corrélogramme simple de la série différencier de la consommation totale du gaz	92
3.46	Corrélogramme partiel de la série différencier de la consommation totale du gaz	92
3.47	Le graphe qqnorm des résidus du modèle ARIMA(0,1,1) pour la consommation totale du gaz	94
3.48	Représentation graphique des valeurs prédictives de la consommation totale du gaz	95

LISTE DES TABLEAUX

1.1	Tableau comparatif des corrélogrammes et corrélogrammes partiels des processus	24
3.1	Les paramètres d'ajustement du modèle ARIMA(0,1,0) pour la consommation d'électricité (clients ordinaires)	55
3.2	Les paramètres des erreurs d'ajustement du modèle ARIMA(0,1,0) pour la consommation d'électricité (clients ordinaires)	56
3.3	Les différents modèles ajustés pour la consommation d'électricité (clients ordinaires)	57
3.4	Les valeurs prédictives par le modèle ARIMA(0,1,0) pour la consommation d'électricité (clients ordinaires)	58
3.5	Les coefficients du modèle ARIMA(5,1,2) pour la consommation d'électricité (clients administrations)	62
3.6	Les paramètres d'ajustement du modèle ARIMA(5,1,2) pour la consommation d'électricité (clients administrations)	62
3.7	Les paramètres des erreurs d'ajustement du modèle ARIMA(5,1,2) pour la consommation d'électricité (clients administrations)	63
3.8	Les différents modèles ajustés pour la consommation d'électricité (clients administrations)	64
3.9	Les valeurs de prédictives par le modèle (5,1,2) pour la consommation d'électricité (clients administrations)	65

3.10 Les paramètres d'ajustement du modèle ARIMA(0,1,0) pour la consommation totale d'électricité	70
3.11 Les paramètres des erreurs d'ajustement du modèle ARIMA(0,1,0) pour la consommation totale d'électricité	70
3.12 Les différents modèles ajustés pour la consommation totale d'électricité	71
3.13 Les valeurs prédictives de la consommation totale d'électricité	72
3.14 Les coefficients du modèle ARIMA(2,1,2) pour la consommation du gaz (clients ordinaires)	77
3.15 Paramètre d'ajustement du modèle ARIMA(2,1,2) pour la consommation du gaz (clients ordinaires)	77
3.16 Paramètre des erreurs d'ajustement du modèle ARIMA(2,1,2) pour la consommation du gaz (clients ordinaires)	77
3.17 Les différents modèles ajustés pour la consommation du gaz (clients ordinaires)	79
3.18 Les valeurs de prédiction par le modèle ARIMA(2,1,2) pour la consommation du gaz (clients ordinaires)	80
3.19 Les coefficients du modèle ARIMA(2,1,5) pour la consommation du gaz (clients administrations)	85
3.20 Les paramètres d'ajustement du modèle ARIMA(2,1,5) pour la consommation du gaz (clients administrations)	85
3.21 Les paramètre des erreurs d'ajustement du modèle ARIMA(2,1,5) pour la consommation du gaz (clients administrations)	85
3.22 Les différents modèles ajustés pour la consommation du gaz (clients administrations)	87
3.23 les valeurs de prédiction par le modèle ARIMA(2,1,5) pour la consommation du gaz (clients administrations)	88
3.24 Coefficients du modèle ARIMA(0,1,1)	93
3.25 Paramètre d'ajustement du modèle ARIMA(0,1,1)	93
3.26 paramètre des erreurs d'ajustement du modèle ARIMA(0,1,1)	93

3.27 les valeurs de la prédiction par le modèle ARIMA(0,1,1) pour la consommation totale du gaz	94
3.28 les valeurs de prédiction par le modèle ARIMA(0,1,1) pour la consommation totale du gaz	95
3.29 Les valeurs prédictives et réelles de la consommation d'électricité (clients ordinaires)	96
3.30 Les valeurs prédictives et réelles de la consommation d'électricité (clients administrations)	96
3.31 Les valeurs prédictives et réelles de la consommation totale d'électricité	96
3.32 Les valeurs prédictives et réelles de la consommation du gaz (clients ordinaires)	97
3.33 Les valeurs prédictives et réelles de la consommation du gaz (clients administrations)	97
3.34 Les valeurs prédictives et réelles de la consommation totale du gaz	97

ملخص

الهدف من هذا العمل هو دراسة و تحليل البيانات بنماذج السلاسل الزمنية. نحن مهتمون بشكل اساسي بنماذج ARIMA باستخدام تقنيات BOX & JENKINS تهدف هذه الطريقة الى صياغة نموذج يسمح بتحليل التاريخ مع نماذج و التنبؤ بالقيم المستقبلية ننهي هذا العمل من خلال النمذجة و التنبؤ بالقيم المستقبلية لاستهلاك الكهرباء و الغاز على مستوى سونلغاز ولاية البليدة باستخدام البرنامج الاحصائي .

Résumé :

L'objectif de ce travail est d'étudier et analyser des données chroniques par des modèles des séries chronologiques. On s'intéresse essentiellement aux modèles ARIMA, en utilisant les techniques de Box & Jenkins. Cette méthode vise à formuler un modèle permettant de représenter une chronique avec des modèles ARIMA et de prévoir les valeurs futures. On termine ce travail par modéliser et prédire la consommation d'électricité et du gaz au niveau de la direction de Sonelgaz de la wilaya de Blida en utilisant le logiciel statistique R et R-studio.

Abstract :

The objective of this work is to study and analyze chronic data by time

series models. We are mainly interested in ARIMA models, using the techniques of Box & Jenkins. This method aims to formulate a model to represent a chronicle with ARIMA models and to predict future values. We end this work by modeling and predicting the consumption of electricity and gas at the level of the management of SONELGAZ wilaya of blida using the statistical software R and R-studio.

INTRODUCTION GÉNÉRALE

L'augmentation de la population et le développement industriel dans la vie moderne, la nécessité des appareils électroménager, ce qui suscite une demande croissante de l'énergie afin de satisfaire les différents services.

L'augmentation de la demande de l'énergie électrique et l'adaptation avec la production a fait augmenter les risques sur les centrales de production et restent toujours instables, les conséquences inadéquates de la production avec la demande sont multiples et variées.

L'Algérie est un pays membre de l'Organisation des pays exportateurs d'électricité (OPEQ) et du Forum des pays exportateurs de gaz (FPEG). Le groupe spécialisé dans la production, le transport et de la distribution de l'électricité et de gaz est Sonelgaz.

Sonelgaz est le seul producteur et fournisseur d'électricité et le seul distributeur de gaz en Algérie. Il fournit l'électricité à 10 millions de clients et le gaz à 6 millions. L'électricité et le gaz est d'une grande importance dans le monde et sans l'électricité, les écoles et les hôpitaux ne peuvent fonctionner correctement et Le gaz naturel est utilisé comme matière première dans l'industrie chimique, notamment pour la pétrochimie et le raffinage. Le gaz naturel offre un bilan environnemental très favorable dans la production d'électricité. Et bien sûr, c'est aussi important en Algérie qui est un pays riche de l'énergie électrique et gazière.

L'importance des réserves du gaz naturel en Algérie en a fait la plus importante source de devises fortes, sur laquelle repose la politique énergétique algérienne actuelle et future, outre l'expansion de ses réseaux nationaux de consommation d'énergie primaire, et l'électricité est l'un des outils spécifiques du développement durable et un moteur important pour tous les secteurs, et est directement liée aux problèmes économiques, sociaux et environnementaux prioritaires tels que : la pauvreté, les opportunités d'emploi, les niveaux de revenus, la dégradation de l'environnement, le changement climatique,.. etc.

Notre travail est structuré de la façon suivante :

Chapitre I : nous feront des généralités sur les séries chronologiques, quelques modèles de prévision et la méthodologie de Box & Jenkins.

Chapitre II : nous présenteront l'entreprise SONELGAZ.

Chapitre III : dans ce chapitre on parlera de l'application de la méthodologie de Box & Jenkins sur la consommation d'électricité et du gaz au niveau de la wilaya de Blida.

Enfin, on terminera notre travail par une conclusion.

CHAPITRE 1

GÉNÉRALITÉS SUR LES SÉRIES CHRONOLOGIQUES

Introduction

Les séries chronologiques appelées Aussi séries temporelles ou bien séries chroniques occupent une place très importante dans tous les domaines de l'observation.

On trouve l'application de ces séries dans des domaines aussi variés que la médecine, finance, énergie, climatologie...etc. La série est dite temporelle par ce qu'elle indexe l'objet enregistré par le temps .

Dans ce chapitre, plusieurs concepts importantes liés à l'analyse des séries chronologiques seront abordés (série linéaire, stationnarité, bruit blanc, les opérateurs, les processus...etc).

Nous avons également consacré une partie pour les modèles de prévisions (ARIMA , ARMA , SARMA ...) ainsi que la méthodologie de Box & Jenkins.

1.1 Définition d'une série chronologique

Définition 1.1 *Série chronologique est une suite des valeurs numériques représentant l'évolution d'une quantité spécifique au cours de temps.*

Définition 1.2 *Une série chronologique est la réalisation d'un processus aléa-*

toire indicé par le temps, noté (Y_t) . Pour chaque t , (Y_t) est une variable aléatoire dont on a une réalisation (y_t) .

Considérons une série chronologique de longueur n , si : (voir le référence[1])

- ◇ t_1, t_2, \dots, t_n sont les n instants successifs d'observations.
- ◇ Et Y_{t_j} est la valeur mesurée à l'instant t_j . On notera la série chronologique $Y_t, t \in T$ ou T est l'ensemble des instants d'observations, ie $T = (t_1, t_2, \dots, t_n)$.

Exemple 1.1

- *Le volume de commerce extérieur d'un pays évalué mois par mois pendant une décennie.*
- *Le nombre de voyageurs transportés mensuellement par un organisme de transport.*

Remarque 1.1 ◇ Une série chronologique est aussi appelée série temporelle ou série chronique.

- ◇ La série chronologique (Y_t) n'est rien d'autre que la série statistique double (t_j, Y_{t_j}) .

1.2 Types des séries chronologiques

On distingue deux types de série chronologique :

1.2.1 Séries continues

C'est une série où l'observation est fait d'une manière continue dans le temps.

Exemple : signal radio, résultat d'un électrocardiogramme...etc.

1.2.2 Séries discrètes

C'est une série où l'observation est fait sur des intervalles de temps fixés a priori. Dans ce cas $t \in \mathbb{Z}$, et les dates d'observations sont le plus

souvent équidistantes.

Exemple : cas de nombre de voyageurs d'un vol d'avion.

1.3 Les composantes d'une série chronologique

Une série chronologique est constituée de quatre composantes : la tendance, le cycle, la saisonnalité et la composante accidentel ou résiduel. La tendance reste la composante la plus importante. (voir de référence [2])

- ▶ **La tendance (F_t) :** (trend) observé sur une longue durée, elle traduit l'orientation (baisse ou hausse) générale de la série étudiée.
- ▶ **La cycle (C_t) :** tout comme le mouvement saisonniers, il s'agit d'un phénomène qui se répète dans le temps mais sur les durées plus longues et irrégulières.

L'existence de ce cycle apparaît rarement évident et est donc souvent associés à la notion de tendance.

- ▶ **La saisonnalité (S_t) :** (la composante saisonnière) correspond à un phénomène qui se répète à intervalles de temps réguliers.

En général c'est un phénomène saisonnier d'où le terme de variations saisonnières (toutes les 12 période pour les données mensuelles, toutes les 7 périodes pour les données quotidiennes...).

- ▶ **Composante accidentel ou résiduel (ϵ_t) :** est de faible variabilité par rapport aux autres composantes.

Elles proviennent de circonstance non prévisible : catastrophes naturelles, crise boursière ...etc.

1.4 L'objectif d'une série chronologique

L'étude d'une série chronologique permet d'analyser, de décrire et d'expliquer un phénomène au cours de temps et d'en tirer des conséquences pour des prises de décision.

L'un des objectifs principaux de l'étude d'une série chronologique est la prévision qui consiste à prévoir les valeurs futures de la série chrono-

gique à partir de ces valeurs observées.

Les objectifs de l'analyse d'une série temporelle sont :

◆ **Modélisation** : elle consiste à :

- ▶ Développer des modèles permettant de décrire le comportement d'une ou plusieurs séries temporelles.
- ▶ Mettre au point une méthodologie pour :
 - ◇ Spécifier.
 - ◇ Estimer.
 - ◇ Valider(juger).

◆ **Prédiction** : prédire les observation dans le future.

La description est généralement réalisé à l'aide d'une présentation graphique.

1.5 Domaines d'applications des séries chronologiques

Nous trouvons l'application des séries chronologiques dans divers domaines, la liste suivante présente qu'un échantillon :

▶ **Énergie** : consommation d'électricité et du gaz au niveau.

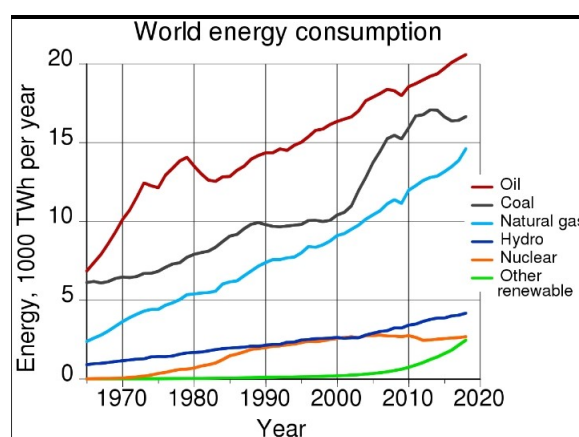


FIGURE 1.1 – Consommation d'électricité et du gaz au niveau mondial

▶ **Médecine** : décès mensuels dus à des maladies pulmonaires au niveau mondial.

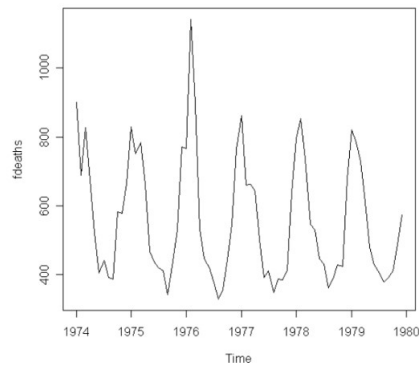


FIGURE 1.2 – Décès mensuelles dus à des maladies pulmonaires au niveau mondial

► **Finance** : cours de bourse quotidiens.

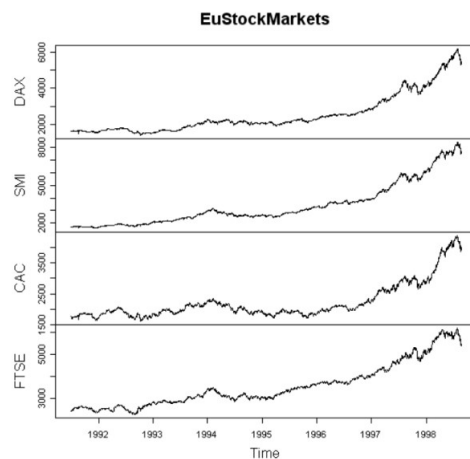


FIGURE 1.3 – Cours de bourse quotidiens

► **Agronomie** : évolution du rendement de blé tendre d'hiver.

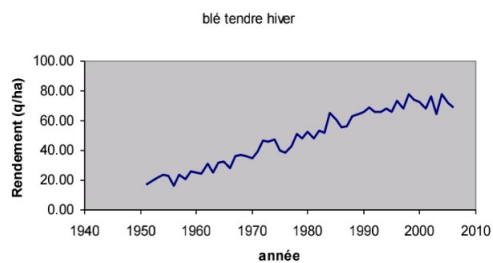


FIGURE 1.4 – Évolution du rendement de blé tendre d'hiver

► **Climatologie** : analyse de données climatiques. température annuelle moyenne.

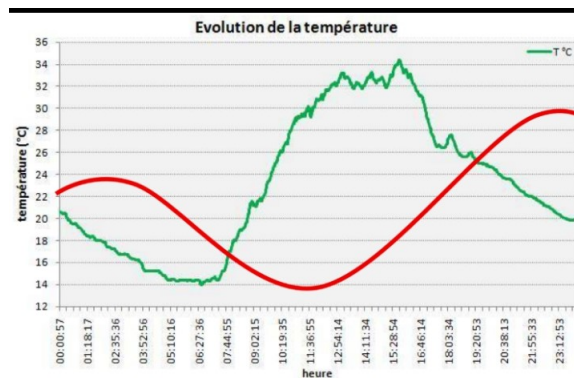


FIGURE 1.5 – Température annuelle moyenne

► **Démographie** : l'évolution d'une population.

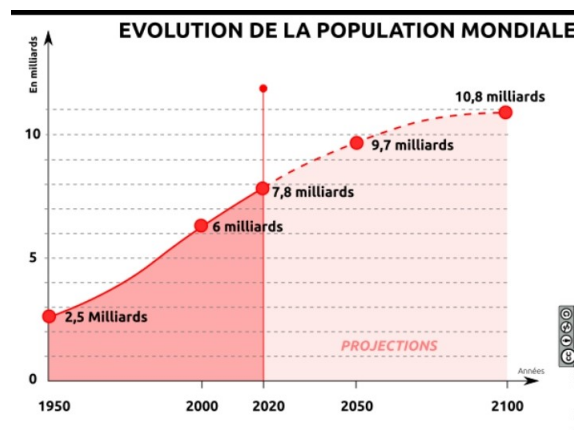


FIGURE 1.6 – Évolution de population du monde

1.6 Modèles de décomposition d'une série chronologique

La décomposition d'une série chronologique est une technique qui divise la série en plusieurs composantes, elle repose sur un modèle qui l'autorise.

La plupart des séries chronologiques se décompose selon les modèles suivants :(voir de référence[3])

1.6.1 Modèle additif

Représente la série chronologique comme une addition des composantes.

$$Y_t = F_t + S_t + C_t + \epsilon_t.$$

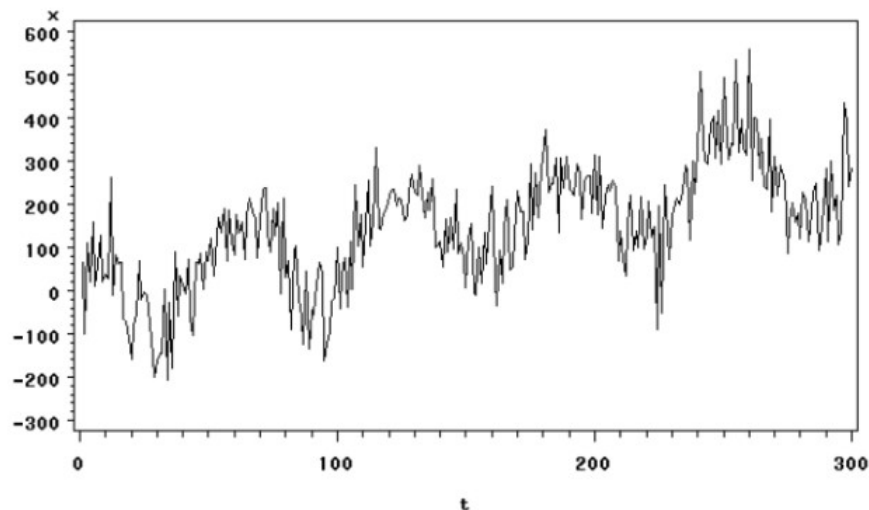


FIGURE 1.7 – Série chronologique(modèle additif)

1.6.2 Modèle multiplicatif

Représente la série chronologique comme une multiplicatif.des composantes.

$$Y_t = F_t * S_t * C_t * \epsilon_t.$$

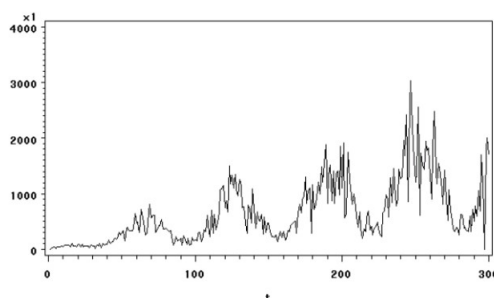


FIGURE 1.8 – Série chronologique(modèle multiplication)

1.6.3 Modèle mixte

Il s'agit d'une combinaison d'addition et de multiplication des composantes.

$$Y_t = F_t * S_t * C_t + \epsilon_t.$$

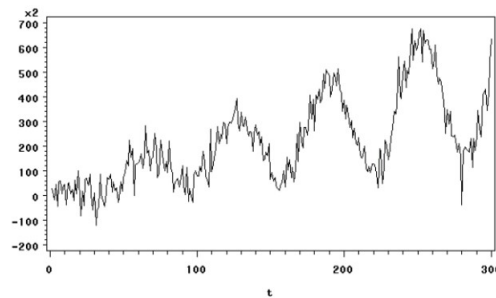


FIGURE 1.9 – Série chronologique(modèle mixte)

1.7 Choix de modèle

- ◆ **Graphiquement** : constatation graphique du parallélisme ou non des droites des minima et de maxima :
 - ▶ **Droites parallèles** : modèle additif.
 - ▶ **Droites non parallèles** : modèle multiplicatif.
- ◆ **Algébriquement** : on recherche le coefficient directeur de ces deux droites par une méthode analytique (deux points).
 - ▶ **Coefficients directeurs très proches** : le modèle choisi sera additif .
 - ▶ **Coefficients directeurs très différents** : le modèle choisi sera multiplicatif.

1.8 Processus stochastique

Définition 1.3 *Un processus stochastique est une suite de variables aléatoires réelles qui sont indexées par le temps (X_t) , $t \in \mathbb{Z}$.*

Un processus stochastique est donc une famille de variables aléatoires X dont on va observer des valeurs réelles issues de l'espace S des échantillons selon une certaine loi de probabilité. Pour chaque point de l'espace des échantillons S , la fonction $X_t(s)$ est appelée le trajectoire du processus.

Les observations successives forment l'histoire du processus. On peut les noter (x_0^t) pour designer l'histoire du processus entre 0 et t .

1.8.1 Bruit blanc

Définition 1.4 *Un processus de bruit blanc est une suite de variables aléatoires (X_t) t indépendantes. Si l'espérance est nulle, le bruit blanc est centré, et si les variables aléatoires sont gaussiennes, le bruit blanc est gaussien. (voir le référence[4]).*

On dit que la suite des variables aléatoires (ϵ_t) constitue un bruit blanc faible si elle possède les propriétés suivantes :

- ◇ $E(\epsilon_t) = 0, \forall t.$
- ◇ $E(\epsilon_t^2) = (\sigma_\epsilon)^2.$
- ◇ $cov(\epsilon_s; \epsilon_t) = 0, sit \neq s.$

c'est à dire :

- ◇ *Les espérances sont égale à 0.*
- ◇ *Les variances sont égale à σ^2 .*
- ◇ *les variables sont non corrélés.*

Remarque 1.2 *On dit que (ϵ_t) est un bruit blanc fort s'il est un bruit blanc faible et que les variables aléatoires ϵ_t sont i.i.d (indépendantes et identiquement distribuées)*

Un bruit blanc Gaussien (ϵ_t) est une suite des v.a.i.i.d $\hookrightarrow \mathcal{N}(0, \sigma_\epsilon^2)$, on note : $\epsilon_t \hookrightarrow BBN(0, \sigma_\epsilon^2)$.

1.8.2 Série linéaire

Définition 1.5 *Une série (Y_t) est dite linéaire si elle peut écrire sous la forme : [5]*

$$Y_t = m + \sum_{i=-\infty}^{+\infty} (Y_i \epsilon_{t-i}) \quad (1.1)$$

où

- ◇ *(ϵ_t) est un bruit blanc $(0, (\Sigma_\epsilon)^2)$.*
- ◇ $Y_0 = 1$.

La suite (Y_i) est absolument sommable c'est à dire $\sum |Y_i| < \infty$.

1.8.3 Processus stationnaire

Définition 1.6 *La stationnarité est une caractéristique d'une série chronologies qui implique que le comportement de la série ne dépend pas de temps. En particulier, on dit qu'une série X_t est stable si elle ne comporte pas la tendance à la hausse ou à la baisse.*

Plus formellement, on distingue deux types de stationnarité, à savoir stricte et faible (voir les références .[6][7])

Stationnarité Stricte :

Définition 1.7 *Le processus aléatoire $x_t, t \in T$ est dite strictement stationnaire si :*

La loi de $(x_{t_1}, x_{t_2}, \dots, x_{t_n})$ est la même que la loi $(x_{t_1+h}, x_{t_2+h}, \dots, x_{t_n+h})$ pour tout (t_1, t_2, \dots, t_n) , avec $t_i \in T$, pour $i = 1, \dots, n$ et pour tout $h \in T$ avec $t_{(i+h)} \in T$.

Ainsi un processus aléatoire est strictement stationnaire si toutes ces caractéristiques c'est-à-dire tous ces moments sont invariants pour tout changement de l'origine du temps.

Stationnarité Faible (stationnarité de second ordre) :

Définition 1.8 *Un processus aléatoire $X_t, t \in T$ est dit stationnaire au sens faible c'est-à-dire :*

$$E(X_t) = \mu < \infty, t \in T \text{ et } var(x_t) = \sigma^2, \forall t \in T$$

$$cov(X_t, X_{(t+h)}) = \gamma(t), \forall t, h \in T.$$

En résumé, un processus X_t est dite stationnaire du second ordre si sa moyenne, sa variance et sa covariance sont indépendantes du temps et sa variance est finie.

Relation entre stationnarité faible et stricte

Un processus strictement stationnaire du second ordre est faiblement stationnaire.

La réciproque n'est pas vraie en général.

1.9 Les opérateurs définis sur une série chronologique

1.9.1 Opérateur de retard

Définition 1.9 *L'opérateur de retard noté B (ou L quelquefois), est l'opérateur qui a tout élément d'une série chronologique associe l'observation précédente. (voir le référence[8])*

$$BY_t = Y_{t-1}, \forall t > 1. \quad (1.2)$$

Généralisation : pour un décalage de plusieurs unités, on utilise plusieurs fois de suite cet opérateur, ce qui l'on note L élevé à une certaine puissance.

$$B^k Y_t = Y_{t-k}. \quad (1.3)$$

Remarque 1.3 *Une généralisation est de décaler non-plus dans le passé mais dans le futur par un exposant négatif.*

$$B^{-1} Y_t = Y_{t+1} \quad (1.4)$$

1.9.2 Opérateur de différentiation

L'opérateur Δ fait la différence entre le processus et sa version décalée d'une unité de temps .[9]

Opérateur de différence première : L'opérateur de différence première est un polynôme retard spécial :

$$\Delta Y_t = Y_t - Y_{t-1}. \quad (1.5)$$

$$\Delta Y_t = (1 - B)Y_t. \quad (1.6)$$

En effet, par écriture purement formelle on peut écrire :

$$\Delta Y_t = Y_t - BY_t = (1 - B)Y_t \quad (1.7)$$

On peut écrire Δ sous la forme d'un polynôme en B avec :

$$\Delta = 1 - B$$

Opérateur de différence seconde :

$$\Delta^2(Y_t) = (1 - B)^2 Y_t$$

i-ème opérateur de différence :

$$\Delta^i(Y_t) = (1 - B)^i Y_t$$

1.9.2.1 L'effet de l'opérateur de différence

- ◇ Permet d'éliminer la tendance de la série.
- ◇ Peut-être répété plusieurs fois si la tendance n'est pas linéaire.

1.10 Caractéristiques et propriétés des séries chronologique

1.10.1 La moyenne

La formule générale de la moyenne d'une série chronologique de terme générale X_t , pour laquelle nous disposons de T observations est :[10]

$$\bar{X} = \frac{1}{T} \sum X_t, \quad t = 1 \dots T \quad (1.8)$$

1.10.2 La variance

La variance d'une série chronologique permet d'évaluer la dispersion autour de la moyenne.(voir le référence[10])

La formule de la variance (var) dans le cas d'une population connue totalement est :

$$\text{var}(X) = \frac{1}{T} \sum (X_t - \bar{X})^2. \quad (1.9)$$

1.10.3 La fonction d'auto covariance

Soit (Y_t) un processus stochastique.(voir le référence[11])

Définition 1.10 On appelle fonction d'auto covariance du processus (X_t) la fonction suivante :

$$\forall h, t \in \mathbb{Z} : \gamma(h) = \text{cov}(X_t, X_{t+h}) \quad (1.10)$$

Proposition 1.1 La fonction d'auto covariance vérifie les propriétés suivantes :

- ◇ $\gamma(0) = \text{var}(X_t) \geq 0$.
- ◇ $|\gamma(h)| \leq \gamma(0)$, $\forall h$; la fonction auto covariance est donc bornée .
- ◇ $\gamma(h) = \gamma(-h)$, $\forall h > 0$; autrement dit l'auto covariance est une fonction paire.

1.10.4 Fonction d'auto corrélation (ACF)

Définition 1.11 La fonction d'auto corrélation est la fonction noté $\rho(h)$ qui mesure la corrélation de la série elle-même décalée d'un pas h :(voir le référence[11])

$\rho(h)$ mesure donc l'intensité de la liaison entre X_t et X_{t+h} .

Le graphe de cette fonction est appelé corrélogramme simple.

$$\rho(h) = \rho(X_t, X_{t+h}) = \frac{\gamma(h)}{\gamma(0)}. \quad (1.11)$$

Proposition 1.2 La fonction d'auto corrélation d'un processus stationnaire vérifie les conditions suivantes :

- ◇ $\rho(0) = 1$.
- ◇ $|\rho(h)| \leq 1$, $\forall h > 0$.
- ◇ $\forall h \in \mathbb{Z}, \rho(-h) = \rho(h)$; elle est paire.

1.10.5 Fonction d'auto corrélation partielle (PACF)

Elle mesure la liaison (linéaire) entre :

$$X_t \text{ et } X_{t-h}$$

Une fois retiré les liens transitant par les variables intermédiaires $X_{t-1}, \dots, X_{t-h+1}$.

Définition 1.12 Le coefficient d'auto corrélation partielle d'ordre h , noté $r(h)$, est définie par :

$$r(h) = \text{cor}(X_t, X_{t-h} / X_{t-1}, \dots, X_{t-h+1}) \quad (1.12)$$

Le coefficient d'auto corrélation partielle d'ordre h d'un processus stationnaire se calcule de la manière suivante :

$$r(h) = \frac{|R(h)^*|}{|R(h)|}$$

avec :

$$R(h)^* = \begin{pmatrix} 1 & \rho(1) & \dots & \rho(h-1) \\ \rho(1) & 1 & \dots & \rho(h-2) \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \rho(h-1) & \rho(h-2) & \dots & 1 \end{pmatrix}$$

et $R(h)^*$ la matrice $R(h)$ dans laquelle on remplacé la colonne h par :

$$(\rho(1) \quad \rho(2) \quad \dots \rho(h) \quad)'$$

soit :

$$R(h) = \begin{pmatrix} 1 & \rho(1) & \dots & \rho(1) \\ \rho(1) & 1 & \dots & \rho(2) \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \rho(h-1) & \rho(h-2) & \dots & \rho(h) \end{pmatrix}$$

ainsi,

$$r(1) = \rho(1), \rho(2) = \frac{\rho(2) - \rho(1)^2}{1 - \rho(1)^2}, \dots$$

De manière empirique, les auto corrélations partielle s'estiment par l'estimation des auto corrélations simples et en calculant $r(\hat{h})$ à partir de la formule ci dessus.

1.11 Test de Dickey Fuller Augmenté

Les tests de stationnarité permettent de vérifier si une série est stationnaire ou non.

Le test de Dickey-Fuller, le test augmenté de Dickey-Fuller (ADF), ou encore le test de Phillips-Perron (PP) sont des tests de racine unitaire, pour lesquels l'hypothèse nulle est que la série a été générée par un processus présentant une racine unitaire, et donc qu'elle n'est pas stationnaire. On donne ici le test ADF.

Les hypothèses du test de ADF se définissent de la manière suivante :

Un processus non stationnaire, il correspond à une de ces formes de non stationnarité :

Modèle[1] : $\Delta X_t = \phi X_{t-1} + \sum_{i=1}^p \alpha_i \Delta X_{t-i} + \epsilon_t$ avec $\phi = \rho - 1$.

Modèle[2] : $\Delta X_t = \alpha + \phi X_{t-1} + \sum_{i=1}^p \alpha_i \Delta X_{t-i} + \epsilon_t$.

Modèle[3] : $\Delta X_t = \alpha + \beta t + \sum_{i=1}^p \alpha_i \Delta X_{t-i} + \epsilon_t$.

1.11.1 Test de Dickey et Fuller augmenté

Robuste à l'auto corrélation par rapport au test de Dickey-Fuller.

$$\begin{cases} H_0 : X_t \implies I(1) \\ H_1 : X_t \text{ n'est pas } I(1) \end{cases}$$

tel que : $I(1)$ est une racine unitaire, et

$$ADF = \frac{T \hat{\phi}}{1 - \hat{\alpha}_1 - \dots - \hat{\alpha}_p} \tag{1.13}$$

1.11.2 Test de Phillips-Perron

Robuste à l'hétéroscédasticité.

$$PP = T\hat{\phi} - \frac{1}{2}T^2 \frac{SE\hat{\phi}}{\hat{\sigma}^2} (\hat{\lambda}^2 - \hat{\sigma}^2) \quad (1.14)$$

$$\hat{\sigma}^2 = \frac{1}{T-k} \sum_{t=1}^T (\hat{\epsilon}_t^2 - \hat{\epsilon}_t)$$

1.12 Théorème de décomposition de Wold

Le Théorème de Wold est fondamental pour l'analyse de séries temporelles stationnaires. (voir le référence[4])

Théorème 1.1 *Tout processus $(X_t)_t \in \mathbb{Z}$ faiblement stationnaire peut s'écrire sous la forme :*

$$X_t = \sum_{j=0}^{+\infty} \psi_j \epsilon_{t-j} + k_t \quad (1.15)$$

Où les paramètres $\psi_1, \psi_2 \dots$ sont des réels tels que :

$$\psi_0 = 1 \quad \text{et} \quad \sum_{j=0}^{\infty} \psi_j^2 < +\infty$$

$(\epsilon_t)_{t \in \mathbb{Z}}$: est un bruit blanc et k_t est une composante linéaire telle que :
 $Cov(k_t, \epsilon_t) = 0$ pour tout $j \in \mathbb{Z}$.

Selon le Théorème de Wold, tout processus stationnaire d'ordre 2 peut être représenté comme une somme pondérée infinie de chocs passés, caractérisés par un bruit blanc.

1.13 Processus aléatoire stationnaire

1.13.1 Processus auto régressif d'ordre p : AR(p)

Définition 1.13 *Dans un modèle auto-régressif d'ordre p noté AR(p), la valeur future d'une variable est supposée être une combinaison linéaire de p*

observations passées et d'une erreur aléatoire associée à un terme constant, le modèle $AR(p)$ peut être exprimé par :

$$X_t = c + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \epsilon_t \quad (1.16)$$

Parfois le terme constant commis pour des raisons de simplicité d'où :

$$X_t = \sum_{i=1}^p \phi_i X_{t-i} + \epsilon_t \quad (1.17)$$

alors :

$$X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \dots - \phi_p X_{t-p} = \epsilon_t \quad (1.18)$$

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) X_t = \epsilon_t \quad (1.19)$$

$$\phi(B) X_t = \epsilon_t \quad (1.20)$$

où :

X_t et ϵ_t sont respectivement la valeur réelle et l'erreur aléatoire à la période t .

ϕ_i des coefficients réels tel que $i = 1, \dots, p$.

$\phi(B)$ est l'équation caractéristique de B (l'opérateur de retard).

Définition 1.14 $(X_t), t \in Z$ est un processus $AR(p)$ si :

- ◇ (X_t) est stationnaire .
- ◇ (X_t) Vérifie une équation :

$$X_t = c + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \epsilon_t \quad (1.21)$$

avec : $\epsilon_t \neq 0$ et ϵ_t est un $BB(0, \sigma^2)$

Notation :

$$\phi(B) X_t = c + \epsilon_t$$

où

$$\phi(B) = 1 - (\phi_1 B + \dots + \phi_p B^p)$$

Proposition 1.3 \diamond Les racines de ϕ sont de module strictement supérieur à 1.

\diamond ϵ_t suit un bruit blanc.

On peut se ramener ensuite à $c = 0$ par centrage car :

$$\phi(B)(X_{t-m}) = \epsilon_t$$

où

$$m = c/\phi_1 = E(X_t)$$

On considère donc le cas où :

$$\phi(B)X_t = \epsilon_t \text{ (et } E(X_t) = 0).$$

1.13.1.1 Les caractéristiques d'un processus auto régressif AR(p)

L'auto covariance :

La fonction d'auto-covariance d'un processus auto régressif AR(p) est donnée par l'équation de récurrence suivante :

$$\text{pour } h > 0 : \gamma(h) = \text{cov}(X_t, X_{t-h}) = E(X_t X_{t-h}) \quad (1.22)$$

$$X_t X_{t-h} = \phi_1 X_{t-1} X_{t-h} + \dots + \phi_p X_{t-p} X_{t-h} + \epsilon_t X_{t-h}$$

d'où

$$\gamma(h) = \phi_1 \gamma_{h-1} + \dots + \phi_p \gamma_{h-p} + E(\epsilon_t X_{t-h})$$

avec

$$E(\epsilon_t X_{t-h}) = 0 \quad \text{car} \quad \epsilon_t X_{t-h}$$

$$\gamma(h) = \sum_{i=1}^p \phi_i \gamma(h-i) \quad (1.23)$$

L'équation $\gamma(h)$ est une fonction de Yule-Walker.

Les équation de Yule-Walker établissent une correspondance directe entre les paramètres du modèle et ses auto-covariances.

Elles sont utiles pour déterminer la fonction d'auto-corrélation ou estimer les paramètres.

Elles établissent que :

$$\gamma(j) = \sum_{k=1}^p \phi_k \gamma_{j-k} \quad \forall j = 1 \dots p \quad (1.24)$$

Les coefficient $\gamma(h)$ représentent la fonction d'auto-covariance de X d'ordre j.

Lorsque l'on inclut également l'auto-covariance d'ordre 0 (en fait la variance), il faut également rajouter la variance des résidus pour la première équation. Ce terme supplémentaire ne se retrouve que dans la première équation car on a fait l'hypothèse d'indépendance des résidus (et donc $cov(\epsilon) = 0$)

$$\gamma(h) = \sum_{i=1}^p \phi_i \gamma(h-i) + \sigma_\epsilon^2 \quad (1.25)$$

En divisant l'équation $\gamma(h)$ par $\gamma(0)$:

d'où

$$\gamma(0) = \phi_1 \gamma(1) + \phi(2) \gamma(2) + \dots + \phi(p) \gamma(p) + \sigma_\epsilon$$

On obtient la fonction d'auto-corrélation satisfaisant à l'équation de récurrence suivante :

$$\rho(h) = \frac{\rho(h)}{\rho(0)} = \phi_1 \rho(h-1) + \dots + \phi_p \rho(h-p) \quad \forall h > 0 \quad (1.26)$$

$$\rho(h) = \sum_{i=1}^p \phi_i \rho(h-i).$$

Proposition 1.4 Si X_t AR(p) alors les $|\rho(h)|$ et les $\gamma(h)$ décroissent vers 0 exponentiellement avec h.

1.13.2 Processus de moyenne mobile d'ordre q : MA(q)

Définition 1.15 On appelle moyenne mobile (Moving Average) d'ordre q un processus de la forme :

$$X_t = \epsilon_t - \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q} \quad (1.27)$$

$$X_t = \theta(L)\epsilon_t$$

où ϵ_t est un processus bruit blanc et θ est un polynôme de degré q dont les coefficients sont $1, \theta_1, \dots, \theta_q$.

On notera parfois

$$X_t = \sum_{j=0}^q b_j \epsilon_{t-j} \quad (1.28)$$

En imposant $b_0 = 1$.

Proposition 1.5 Un processus auto régressif est un processus moyenne mobile infini, et réciproquement un processus moyenne mobile est un processus d'auto régressif d'ordre infini.

1.13.2.1 Les caractéristiques d'un processus moyenne mobile MA(q)

La fonction d'auto covariance :

La fonction d'auto covariance d'un processus moyenne mobile MA(q) donnée par :

$$\gamma(h) = \begin{cases} \sigma^2 \sum_{k=0}^{q-h} \theta_k \theta_{q+h-k} & \forall h \leq q \\ 0 & \forall h > q \end{cases}$$

La fonction d'auto corrélation :

La fonction d'auto-corrélation d'un processus moyenne mobile MA(q)

donnée par le système d'équations suivant :

$$\rho(h) = \begin{cases} 1, & \text{si } h = 0 \\ \frac{-\theta}{1+\theta^2}, & \text{si } h = 1 \\ 0, & \text{si } h > 1 \end{cases}$$

La fonction d'auto corrélation partielle :

$$\Gamma(h) = \frac{-\theta^h}{\sum_{i=0}^p \theta^{2i}} \quad (1.29)$$

Remarque 1.4 *Un processus MA est toujours stationnaire.*

Il est inversible si les racines de $\theta(B)$ sont en module strictement supérieures à 1.

1.13.3 Processus auto régressif moyenne mobile d'ordre (p,q)(ARMA)

Les modèles ARMA permettent de présenter un grand nombre de processus aléatoires stationnaires.

Définition 1.16 *Un processus stationnaire Y_t admet une représentation ARMA(p,q) minimale h.*

$$Y_t - \sum_{i=1}^p \phi_i Y_{t-i} = \epsilon_t - \sum_{i=1}^q \theta_i \epsilon_{t-i} \quad (1.30)$$

Qui peut s'écrire aussi :

$$\phi(L)Y_t = \theta(L)\epsilon_t \quad (1.31)$$

S'il vérifie les conditions suivantes :

- ▶ $\phi_p \neq 0$ et $\theta_q \neq 0$.
- ▶ Les polynômes ϕ et θ ont tous leurs racines de module strictement supérieur à 1.
- ▶ ϕ et θ n'ont pas de racine commune.
- ▶ ϵ_t est un bruit blanc de variance $\sigma^2 \neq 0$.

Remarque 1.5 Les processus ARMA sont des mélanges des processus AR et MA.

1.13.3.1 Les caractéristiques d'un processus ARMA

L'auto covariance et l'autocorrélation :

La fonction d'auto covariance et d'auto corrélation d'un processus ARMA données par :(voir le référence[12])

$$\begin{cases} \gamma(h) = \sum_{i=1}^p \phi_i \\ \rho(h) = \gamma(h)(h-i) \sum_{i=1}^p \phi_i \rho(h-i) \quad \text{si } h > q \end{cases}$$

Ce qui est un système d'équation de type Yule-Walker.

La fonction d'auto-corrélation partielle :

Pour $h \leq p$, il n'y a pas de schéma bien précis.

Pour $h > p$, on retrouve le même comportement que pour un MA(q).

Proposition 1.6 X_t est un processus ARMA(p, q), alors la fonction d'auto-corrélation $\rho(h)$ décroît exponentiellement vers 0.

On peut ainsi dresser un tableau comparatif des corrélogrammes et corrélogrammes partiels des processus :

TABLE 1.1 – Tableau comparatif des corrélogrammes et corrélogrammes partiels des processus

processus	auto corrélation $\rho(h)$	auto corrélation partielle $\Gamma(h)$
AR(p)	amortie	nulle pour $h > p$
MA(q)	nulle pour $h > q$	amortie
ARMA(p, q)	amortie	amortie

1.13.4 Processus auto régressif moyenne mobile saisonnier : SARMA

Définition 1.17 , noté SARMA si et seulement si :

$$\sum_{j=0}^p \phi_{js} X_{t-js} = c + \sum_{j=0}^q \theta_{js'} \epsilon_{t-js'j} \quad (1.32)$$

Où c est constant et $(\epsilon_t)_{t \in \mathbb{Z}}$ sont iid $\mathcal{N}(0, \sigma^2)$.

Ici, s désigne la période de saisonnalité de la composante AR, alors que s' représente la période de saisonnalité de MA.

1.14 Processus linéaire non stationnaire

1.14.1 Processus TS

Définition 1.18 Un processus $(X_t)_{t \in \mathbb{Z}}$ fait partie de la famille des modèles à tendance saisonnière (TS) s'il peut s'écrire sous la forme :

$$X_t = f(t) + \epsilon_t \quad (1.33)$$

avec :

f : est une fonction de temps.

ϵ_t : est un bruit blanc.

Remarque 1.6 Pour stationnariser un processus de type TS on estime $f(t)$ par les moindres carrés ordinaire.

1.14.2 Processus DS

Définition 1.19 Un processus non stationnaire fait partie de la famille des modèles à différence stationnaire DS d'ordre d si le processus filtré comme cas particuliers des modèles DS, $(0,1,1)$ retrouve les modèles ARIMA et SARIMA

On représente les modèles dans la suite.

1.15 Processus auto-régressif moyenne mobile intégré ARIMA

L'objectif essentiel des modèles ARIMA est de permettre une prédiction de l'évolution future d'un phénomène.

Définition 1.20 Un modèle ARIMA est noté (p,d,q) .

◇ p : correspondant au nombre de termes auto-régressifs.

◇ d : correspondant au nombre de différenciations.

◇ q : correspondant au nombre de termes de moyenne mobile.

De nombreuses combinaisons sont envisageables.

Le processus stationnaire satisfait une représentation ARIMA(p, d, q) si et seulement si :

$$\phi(B^s)(1 - B^s)^d X_t = c + \theta(B^s)\epsilon_t \quad (1.34)$$

avec $c \in \mathbb{R}$ et

$$\phi(B) = \sum_{j=0}^p \phi_j B^j \quad \text{et} \quad \theta(B) = \sum_{j=0}^q \theta_j B^j.$$

1.16 Processus SARIMA

Définition 1.21 Le processus SARIMA est un modèle de ARIMA dans laquelle une composante saisonnière est ajoutée.

Ainsi $(X_t), t \geq 0$, est un processus SARIMA(p, d, q) de période s et s' si et seulement si :

$$\phi(B^s)(1 - B^s)^d X_t = c + \theta(B^{s'})\epsilon_t \quad (1.35)$$

avec $c \in \mathbb{R}$ et

$$\phi(B) = \sum_{j=0}^p \phi_j B^j \quad \text{et} \quad \theta(B) = \sum_{j=0}^q \theta_j B^j.$$

La période de saisonnalité s est associée à la composante AR alors que la période de saisonnalité s' est associée à la composante MA.

On suppose que toutes les racines polynômes $\phi(B)$ et $\theta(B)$ sont situées à l'extérieur du cercle unité.

1.17 Processus stochastiques non linéaires

Dans le cas où les résidus ne suivent pas un bruit blanc, il est possible de le modéliser avec un modèle ARCH / GARCH (auto régressive conditional heteroscedastik/(généralise ARCH))

1.17.1 Modèles auto régressifs conditionnellement hétéroscédastiques : ARCH

Définition 1.22 Le processus $(X_t)_{t \geq 0}$ satisfait une représentation d'un processus auto régressif conditionnellement hétéroscédastique d'ordre $p \geq 1$, noté ARCH(p), si et seulement si :

$$X_t = \sigma_t \epsilon_t \quad \text{et} \quad \sigma_t^2 = \alpha_0 + \sum_{j=0}^p \alpha_j X_{t-j}^2 \quad (1.36)$$

Ici, $\alpha_0, \alpha_1, \dots, \alpha_p$ sont des paramètres à valeurs dans \mathbb{R}^+ et le processus $(\epsilon_t)_{t \geq 0}$ est un bruit blanc tel que ϵ_t est indépendant de

$$X_{t-l}, l \geq 1$$

pour tout t .

Contrairement aux modèles ARMA, les modèles ARCH permettent à la volatilité, c'est à dire la variance de varier en fonction du temps.

1.17.2 Modèles auto régressifs conditionnellement hétéroscédastiques généralisés : GARCH

Définition 1.23 Le modèle GARCH(p, q) pour un processus $(X_t)_{t \geq 0}$ s'écrit sous la forme

$$X_t = \sigma_t \epsilon_t \quad \text{et} \quad \sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i X_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2 \quad (1.37)$$

Ici, $\alpha_0, \alpha_1, \dots, \alpha_p$ et β_1, \dots, β_q sont des paramètres à valeurs dans \mathbb{R}^+ et le processus $(\epsilon_t)_{t \geq 0}$ est un bruit blanc tel que ϵ_t est indépendant de $X_{t-l}, l \geq 1$ pour tout t .

Les modèles GARCH généralisent les modèles ARCH en permettant à la variance conditionnelle de suivre un processus ARMA(p, q) via l'introduction d'une composante auto régressive.

1.18 La méthode de Box&Jenkins

Introduction

Box & Jenkins (1970)(voir le référence[13]) ont élaboré une méthodologie pour identifier un modèle adéquat pour une série chronologique

L'approche de Box & Jenkins consiste en une méthodologie rigoureuse d'étude systématique des série chronologique à partir de leur caractéristique.

L'objectif est de déterminer le modèle le plus adapté à représenter le phénomène étudié. La méthodologie de Box&Jenkins peut se décomposer en quatre étapes :

1. Identification du modèle
2. Estimation des paramètres du modèle
3. Validation du modèle
4. Prédiction

1.18.1 Identification du modèle

L'étape d'identification d'un processus ARMA (choix entre AR ,MA et ARMA et choix de p et q) est basé sur la comparaison des caractéristiques théoriques à leur équivalent empirique (c.à.d calculer sur la série observée), les caractéristiques utilisées sont les AC simples et partielles

- ◆ Si le corrélogramme simple n'a que ses q premiers termes différents de zéro et que les termes du corrélogramme partiel diminuent lentement , nous pouvons pronostiquer un MA(q).
- ◆ Si le corrélogramme partiel n'a que ses p premiers termes différents de zéro et que les termes du corrélogramme simple diminuent lentement, cela caractérise un AR(p).
- ◆ Si les fonctions d'AC simple et partiel ne paraissent pas tronquées, il s'agit d'un processus ARMA.

1.18.2 Estimation des paramètres du modèle

L'estimation des paramètres d'un modèle lorsque p, d, q sont supposés connus peut se réaliser par différentes méthodes dans le domaine temporel : (voir la référence [12])

1-Maximum de vraisemblance :

Une méthode populaire pour estimer les paramètres d'un modèle est le maximum de vraisemblance.

La fonction de vraisemblance associée à un échantillon X_1, \dots, X_T iid d'une loi dont la densité est $f(x/\theta)$ avec $\theta = (\theta_1, \dots, \theta_k) \in R^+$ est défini par :

$$L(\theta) = \prod_{h=1}^T f(x/\theta) \quad (1.38)$$

L'estimateur du maximum de vraisemblance (EMV) est la valeur $\widehat{\theta}_{EMV}$ et de résoudre le système à k équations .

$$\frac{dL(\theta)}{d\theta_j} = 0 \text{ ou } j = 1, \dots, k \quad (1.39)$$

$$L(x, \phi, \theta, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{T/2}} \frac{1}{\sqrt{|\Omega(\theta, \phi)|}} \exp\left(-\frac{1}{2\sigma^2} X^T \Omega^{-1}(\theta, \phi) X\right) \quad (1.40)$$

$L(X, \phi, \theta, \sigma^2)$: fonction de vraisemblance.

où $X = (X_1, \dots, X_d)^T$.

σ^2 : est la variance.

$\Omega(\theta, \phi) \in R^{T+T}$: est la matrice de variance covariance de X .

$\theta = (\theta_1, \dots, \theta_q), \phi = (\phi_1, \dots, \phi_p)$: les paramètres des processus .

2-Méthode des moments :

Pour le modèle donné, on sait que les auto-corrélations ρ_h et γ_h , des processus AR et MA respectivement, dépendent des paramètres $\xi = (\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q)$ selon des équations théoriques connus.

$$(\rho_1, \dots, \rho_p, \Gamma_1, \dots, \Gamma_q) = F(\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q).$$

On sait facilement estimer les auto-corrélations ρ_h et ω_h , il suffit donc

d'inverser les équations pour estimer les paramètres.

$$(\hat{\phi}_1, \dots, \hat{\phi}_p, \hat{\theta}_1, \dots, \hat{\theta}_q) = F^{-1}(\hat{\rho}_1, \dots, \hat{\rho}_p, \hat{\Gamma}_1, \dots, \hat{\Gamma}_q).$$

3-Méthode des moindres carrés :

► Modèle AR(p) :

Soit

$$X_t = \delta + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \epsilon_t \quad o \quad \epsilon_t \hookrightarrow BB(0, \sigma^2).$$

En donnant N observations x_1, \dots, x_N , les paramètres $\delta, \phi_1, \dots, \phi_p$ peuvent être estimés par la méthode des moindres carrés en minimisant :

$$S = \sum_{t=p+1}^N (x_t - \delta + \phi_1 x_{t-1} + \dots + \phi_p x_{t-p})^2 \quad (1.41)$$

par rapport à δ, ϕ_1, ϕ_p .

Si $\epsilon_t \rightarrow \mathcal{N}(0, 1)$ les estimateurs de MC sont aussi les estimateurs du MV conditionnel sur les p premières valeurs.

► Modèle MA : Le problème d'estimation est plus difficile pour le modèle MA, on ne peut pas trouver des estimateurs explicitement on obtient une forme itérative numérique.

1.18.3 Validation du modèle

Cette étape consiste à s'assurer que le retenu décrit bien la série d'observations. Par exemple, on peut vérifier si la présence ou non d'une tendance dans le modèle est justifiée considérant la représentation graphique de la série.

Aussi, il convient de s'assurer que les résidus ne possèdent pas d'information permettant de prétendre à un meilleur modèle. Pour cela, il s'agit de vérifier les trois conditions d'un processus bruit blanc à l'aide de tests appropriés.

Dans cette optique, regardons maintenant quelques tests permettant

de vérifier certains de ces conditions.

Test sur les résidus :

Dans l'analyse des séries chronologiques par processus, le bruit blanc joue un rôle particulier puisque c'est un processus sans aucune structure.

Le test préliminaire concerne ce processus.

la première hypothèse testée :

$$H_0 : \rho_1 = \rho_2 = \dots = \rho_h = 0.$$

Contre :

H_1 : il existe au moins ρ_i significativement différent de 0 pour effectuer ces tests, on recourt à la statistique Q de Box-Pierre ou Q' de Ljung-Box qui sont données par :

$$Q = \sum_{h=1}^h \hat{\rho}_h^2, \tag{1.42}$$

telle que :

T : le nombre d'observations.

h : le nombre de retards.

$\hat{\rho}_h$: auto corrélation empirique d'ordre h.

Le test de Student sur les coefficients :

On vérifie d'abord que les racines des polynômes AR et MA ne sont pas égale à 1.

On teste la significative des coefficients par des tests de student.

On calcule donc la statistique de Student du coefficients $t_{\hat{\phi}_j} = \frac{\hat{\phi}_j}{\sqrt{\hat{V}(\hat{\phi}_j)}}$

Si $|t_{\hat{\phi}_j}| < t_{1-\frac{\alpha}{2}}$, on accepte l'hypothèse nulle $H_0 : \phi_j = 0$.

Si $|t_{\hat{\phi}_j}| > t_{1-\frac{\alpha}{2}}$, l'hypothèse nulle est rejetée : $\phi_j \neq 0$.

Ou $t_{1-\frac{\alpha}{2}}$ est le quartile d'ordre $(1-\frac{\alpha}{2})$ de loi de student à $(T - h)$ degrés de liberté, h étant le nombre de paramètres estimés.

choix de modèle :

Après examen des coefficients et des résidus, certains modèles sont écartés. Pour départager les modèles restants, on fait appel aux critères standards et critères d'information .

■ **Critères standards :**

L'erreur absolue moyenne (MAE) :

$$MAE = \frac{1}{T} \sum_t |e_t| \quad (1.43)$$

Racine de l'erreur quadratique moyenne (RMSE) :

$$RMSE = \sqrt{\frac{1}{T} \sum_t e_t^2} \quad (1.44)$$

Ecart absolu moyen en pourcentage (MAPE) :

$$MAPE = 100 \frac{1}{T} \sum_t \left| \frac{e_t}{X_t} \right| \quad (1.45)$$

Plus la valeur de ces critères est faible, plus le modèle estimé est proche des observations.

■ **Critère d'information :**

► **AIC (Akaike) :**

$$AIC(p, q) = \log(\hat{\sigma}^2) + \frac{2(p+q)}{T} \quad (1.46)$$

► **Le critère de Schwarz (1978) :**

$$SC(p, q) = (\log \sigma_{\hat{\epsilon}_t}^2) + (p+q) \frac{\log T}{T} \quad (1.47)$$

On choisit le modèle qui minimise les critères standards et les critères d'information. Le modèle sélectionné sera alors utilisé pour la prévision. L'AICc est une correction de l'AIC pour le cas d'échantillons de petite taille :

$$AICc = AIC + \frac{2k(k+1)}{n-k-1} \quad (1.48)$$

Le critère d'information bayésien (BIC) est un critère d'information dé-

rivé du critère d'information d'Akaike :

$$BIC = -2\ln(L) + k.\ln(N) \quad (1.49)$$

1.18.4 Prévision

Après validation du modèle on procède à des prévisions sur un horizon limité car la variance de l'erreur de prévision croit rapidement avec l'horizon.

Si l'on désire avoir la valeur de X_{t+h} , tout en supposant qu'on est à l'instant t , on utilise l'estimateur :

$$\widehat{X}_t(h) = E(X_{t+h}|X_t, X_{t-h}, \dots, X_1)$$

avec t l'origine de prévision de h l'horizon de prévision .

Dans le cas d'un modèle ARMA , on aura la forme :

$$\begin{aligned} \widehat{X}_t(h) &= \phi_1 E(X_{t+h-1}|X_t, X_{t-1}, \dots, X_1) + \dots + \phi_p E(X_{t+h-p}|X_t, X_{t-1}, \dots, X_1) \\ &- \theta_1 E(\epsilon_{t+h-1}|X_t, X_{t-1}, \dots, X_1) - \dots - \theta_q E(\epsilon_{t+h-q}|X_t, X_{t-1}, \dots, X_1) + E(\epsilon_{t+h}|X_t) \end{aligned}$$

avec :

$$\begin{aligned} E(X_{t-j}/X_t, X_{t-1}, \dots, X_1) &= X_{t-j} \\ E(X_{t+j}/X_t, X_{t-1}, \dots, X_1) &= \widehat{X}_t(j) \\ E(\epsilon_{t-j}/X_t, X_{t-1}, \dots, X_1) &= \epsilon_{t-j} \\ E(\epsilon_{t+j}/X_t, X_{t-1}, \dots, X_1) &= 0 \end{aligned}$$

Erreur de prévision :

On considère un processus ARMA telle que :

$$X_t = \delta + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \dots - \theta_q \epsilon_{t-q}$$

avec $(\phi_p, \theta_q) \in \mathbb{R}^{k*}$ et

$$\epsilon_t \implies IID(0, \sigma_\epsilon)^2$$

Appliquons le Théorème de Wold au processus $(X_t, t \in \mathbb{Z})$ et considérons la forme $MA(\infty)$ correspondante : (avec $\delta = 0$)

$$X_t = \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i}; \psi_0 = 1 \quad (1.50)$$

Il s'ensuit que la meilleure prévision que l'on peut faire pour X_{t+1} compte tenu de toute l'information disponible jusqu'à la date t notée $\widehat{X}_t(1)$ est donnée par :

$$\widehat{X}_t(1) = E(X_{t+1} | X_t, X_{t-1}, \dots, X_1) = E(X_{t+1} | \epsilon_t, \epsilon_{t-1}, \dots, \epsilon_1) = \sum_{i=1}^{\infty} \psi_i \epsilon_{t+1-i}$$

L'erreur de prévision ϵ_t est appelé aussi l'innovation :

$$X_{t+1} - \widehat{X}_t(1) = \epsilon_{t+1}$$

Plus généralement pour une prévision à un horizon k on a :

$$\widehat{X}_t(k) = \sum_{i=0}^{\infty} \psi_i \epsilon_{t+k-i} \quad (1.51)$$

$$X_{t+k} - \widehat{X}_t(k) = \sum_{i=0}^{k-1} \psi_i \epsilon_{t+k-i} \quad (1.52)$$

Déterminons un intervalle de confiance sur la prévision $\widehat{X}_t(k)$, sous l'hypothèse de normalité des résidus ϵ_t .

On montre que :

$$E([X_{t+k} - \widehat{X}_t(k)]^2) = E([\sum_{i=0}^{k-1} \psi_i \epsilon_{t+k-i}]^2) = \sum_{i=0}^{k-1} \psi_i^2 \sigma_\epsilon^2$$

d'où :

$$\frac{X_{t+k} - \widehat{X}_t(k)}{\sigma_\epsilon [\sum_{i=0}^{k-1} \psi_i^2]^{1/2}} \text{-----} Nt(0, 1)$$

On peut donc construire l'intervalle de confiance sous la forme :

$$IC = [\widehat{X}_t(k) + -t_{\frac{\alpha}{2}} \hat{\sigma} [\sum_{i=0}^{k-1} \psi_j^2]^{1/2}].$$

CHAPITRE 2

PRÉSENTATION DE L'ENTREPRISE SONELGAZ

2.1 Introduction

Dans sa politique énergétique, l'Algérie a opté dès son indépendance, pour le développement des infrastructures électriques et gazières, et l'accès de la population à l'électricité et au gaz naturel, ces deux vecteurs constituant une priorité qui vise l'amélioration de la qualité de vie du citoyen d'une part, et de la situation économique du pays, d'autre part.

Conscient de cet enjeu, le secteur s'est fixé comme priorité de développer tous les axes permettant de garantir la couverture à long terme, des besoins en électricité et en gaz du pays, notamment par la diversification des sources d'énergie, le développement du parc de production électrique et des infrastructures de transport et de distribution de l'électricité et du gaz. Sonelgaz en tant qu'entreprise publique et outil de l'État, assume la mission de service public dans le domaine de la distribution de l'électricité et du gaz.

La promulgation de la loi n° 02-01 du 05 février 2002 relative à l'électricité et à la distribution du gaz par canalisations a permis la réorganisation de Sonelgaz, avec la création de la société holding « Sonelgaz » ainsi qu'une quarantaine de filiales (métiers, travaux et services) et de sociétés en participation. Cette loi a également créé la commission

de régulation de l'électricité et du gaz « CREG » installée en 2005, qui encadre et régule les activités du secteur de l'électricité et gaz.

2.2 Distribution d'électricité et du gaz en Algérie

Les producteurs d'électricité et du gaz sont des organisations spécialisés dans la production de ces deux derniers. Le secteur d'électricité et du gaz accélère sa révolution technologique pour maintenir le relationnel avec ses clients vu que leurs consommations accélèrent et se développent chaque jour. Ils ont connu un grand élargissement et développement de la technologie moderne car elles sont toujours produises et consommés que ce soit pour l'utilisation quotidienne ou des utilisations industrielles.



FIGURE 2.1 – Station de distribution électrique



FIGURE 2.2 – Station de distribution du gaz

2.3 Présentation de l'entreprise Sonelgaz en ALGERIE



FIGURE 2.3 – Logo de l'entreprise SONELGAZ

2.3.1 Présentation de la société

Sonelgaz est l'opérateur historique dans le domaine de la fourniture des énergies électrique et gazière en Algérie, crée en **1969**. **Sonelgaz**, œuvre depuis un demi-siècle au service du citoyen algérien en lui apportant cette source énergétique essentielle à la vie quotidienne.

A la faveur de la promulgation de la loi sur l'électricité et la distribution du gaz par canalisations, **Sonelgaz** est passée d'une entreprise verticalement intégrée à une holding pilotant un groupe industriel multi-sociétés et multi-métiers.

Sonelgaz a toujours joué un rôle majeur dans le développement économique et social du pays. Sa contribution dans la concrétisation de la politique énergétique nationale est à la mesure des importants programmes réalisés, en matière d'électrification rurale et de distribution publique gaz, ce qui a permis de hisser le taux de couverture en électricité à 98% pour **10 494 000** clients et un taux de pénétration du gaz à 65% pour **6 451 000** clients.

Aujourd'hui, le groupe **Sonelgaz** est composé de 21 sociétés dont 02 Holding, dont 14 sociétés directement pilotées par la Holding, 02 sociétés contrôlées à hauteur de 50% et 51% de 05 sociétés en participations

avec des tiers.



FIGURE 2.4 – Siège de la direction centrale de Sonelgaz

2.3.2 Historique de l'évolution de la société

De **EGA** à **SONELGAZ** d'aujourd'hui : Plus d'un demi siècle d'existence...

En 1947 : est créé l'établissement public « **électricité et gaz d'Algérie** » par abréviation **EGA**, auquel est confié le monopole de la production, du transport et de la distribution de l'électricité et du gaz.

EGA regroupe les anciennes entreprises de production et de distribution de statut privé, notamment Le bon et **CIE** et **SAE** (Société Algérienne de l'électricité et du gaz), tombant sous le coup de la loi de nationalisation de **1946**, promulguée par l'état français.

En 1962 : **EGA** est pris en charge par l'état algérien indépendant, en quelques années grâce à un formidable effort de formation, l'encadrement et le personnel algérien assurent effectivement le fonctionnement de l'établissement.

En 1969 : **EGA** devient **Sonelgaz**, devient société nationale de l'électricité et du gaz, à ce moment c'est déjà une entreprise de taille importante dont le personnel est de quelques **6000 agents**.

La transformation de la société avait pour objectif de conférer à l'entreprise les capacités organisationnelles et gestionnaires pour accompa-

gner et soutenir le développement économique du pays. Il s'agit notamment du développement industriel, et de l'accès du plus grand nombre à l'énergie électrique. (L'électrification rurale), projet inscrits dans le plan de développement élaboré par les autorités publiques.

En 1983, l'entreprise se dote de cinq (05) filiales travaux spécialisées : **KAHRIF** de l'électrification.

KAHRAKIB : Infrastructures et installation électrique.

KANAGAZ : Réalisation des réseaux gaz.

INERGA : Génie Civil.

ETTERKIB : Montage industriel.

Et l'entreprise **AMC** : fabrication des compteurs et appareils de mesure et de contrôle.

C'est grâce à ces filiales que Sonelgaz dispose actuellement d'infrastructures électriques et gazières répondant aux besoins du développement économique et sociale du pays.

En 1991 : **SONELGAZ** devient établissement public à caractère industriel et commercial (**EPIC**), la reprise de statut, tout en confirmant la mission de service public pose la nécessité de la gestion économique et de la prise en compte de la commercialise.

En 2002 : Dans ce même objectif l'établissement devient, une société par action (**SPA**).

Cette promotion donne **Sonelgaz** la possibilité d'élargir ses activités à d'autres domaines relevant du secteur de l'énergie et aussi d'intervenir hors des frontières de l'Algérie.

En tant que **SPA** elle doit détenir un porte feuille d'actions et autres valeurs mobilières et à la possibilité de prendre des participations dans d'autres sociétés.

En 2004 : L'annonce d'évolution de la **Sonelgaz** qui devient un groupe ou holding. Durant les années **2004** à **2006**, devenant une holding ou groupe d'entreprises, **Sonelgaz** se restructure en filiales chargées de ses activités de base :

- ▶ Sonelgaz Production Électricité (SPE).
- ▶ Gestionnaire Réseau Transport Électricité (GRTE).
- ▶ Gestionnaire Réseau Transport Gaz (GRTG).

En 2006 la fonction distribution est structurée en quatre filiales :

- Alger.
- Région Centre.
- Région Est.
- Région Ouest.

2.3.3 Organisation de Sonelgaz

La **Sonelgaz** est organisée en groupe industriel constitué de 39 filiales et cinq sociétés en participation exerçant des métiers de bases, travaux, périphériques.

Parmi ces filiales :

- ▶ La société de production de l'électricité .
- ▶ Shariket kahraba wa takat moutadjadida (SKTM).
- ▶ La société de l'engineering de l'électricité et du gaz (CEEG).
- ▶ La société de gestion du réseau de transport de l'électricité (GRTE).
- ▶ La société de gestion du réseau de transport gaz (GRTG).
- ▶ L'opérateur système électrique (os), chargée de la conduite du système de production et de transport de l'électricité.
- ▶ La société algérienne de distribution de l'électricité et du gaz. (**SAD-DEG**), créée en 2017 par fusion des sociétés **SDC, SDA, SDE et SDO**

Il y a aussi Kahrif, Kahrakib, Etterkib, Ingra et Kanaghaz, qui sont spécialisées dans le domaine de la réalisation des infrastructures énergétiques (engineering, montage industriel, réalisation de réseaux ...).

Début **octobre 2020**, les sociétés filiales SKT (Shariket Kahraba Terga), SKD (Shariket Kahraba Koudiat Eddraouch) et SKB (Shariket Kahraba Berrouaghia) sont absorbées par la société SKS (Shariket Kahraba Skikda), en vue de permettre une meilleure optimisation de la production d'élec-

tricité .

SKS créée en mai 2003, passerait ainsi d'une capacité de production de **825 Méga-Wattes (MW)** à une nouvelle capacité de production de **3582 MW**.

En 2020, **Sharikat Kahraba El Djazair (SKE)**, chargée principalement de la production électrique d'origine renouvelable, détenue par **SONELGAZ** à hauteur de 51% du capital contre 49% pour **Sonatrach**, est le résultat de la fusion/absorption des sociétés filiales **SKD, SKT, SKB** et **SKS** mentionnées plus haut.

2.4 Activités de Sonelgaz

2.4.1 Activités non-réglementées

Sonelgaz exerce ses activités dans les secteurs dé réglementés de la production d'électricité, de la fourniture de gaz naturel et d'électricité. La société dispose d'installations qui ont une puissance installée totale de **20963 MW** à la fin de **2016**.

Elle gère des centrales hydroélectriques, thermiques classiques à cycle combiné de cogénération. La production du groupe en Algérie s'est chiffrée à **70904 GWh** en **2017**.

2.4.2 Activités réglementées

Sonelgaz distribue de l'énergie à près de **10 millions** de clients en Algérie.

En outre, elle compte près de **10 millions** d'abonnés en électricité et **6 millions** en gaz naturel en **2017**.

2.4.3 Énergies renouvelables

Sonelgaz inscrit dans son plan stratégique, adopté en **2009** par le conseil d'administration, un axe de diversification du mix énergétique en adoptant les ENR (solaire et éolien).

En 2018, la capacité installée des ENR est de 354 mégawatts dont 344.1 MW d'énergie photovoltaïque, produite par 24 centrales et une infime partie 10.2 MW d'énergie éolienne.

Afin de répondre aux besoins de plus en plus croissants de la consommation national, Sonelgaz investit de plus en plus dans les énergies renouvelables avec de nombreux projets et plans à travers le pays.

2.5 Organigramme de l'entreprise Sonelgaz

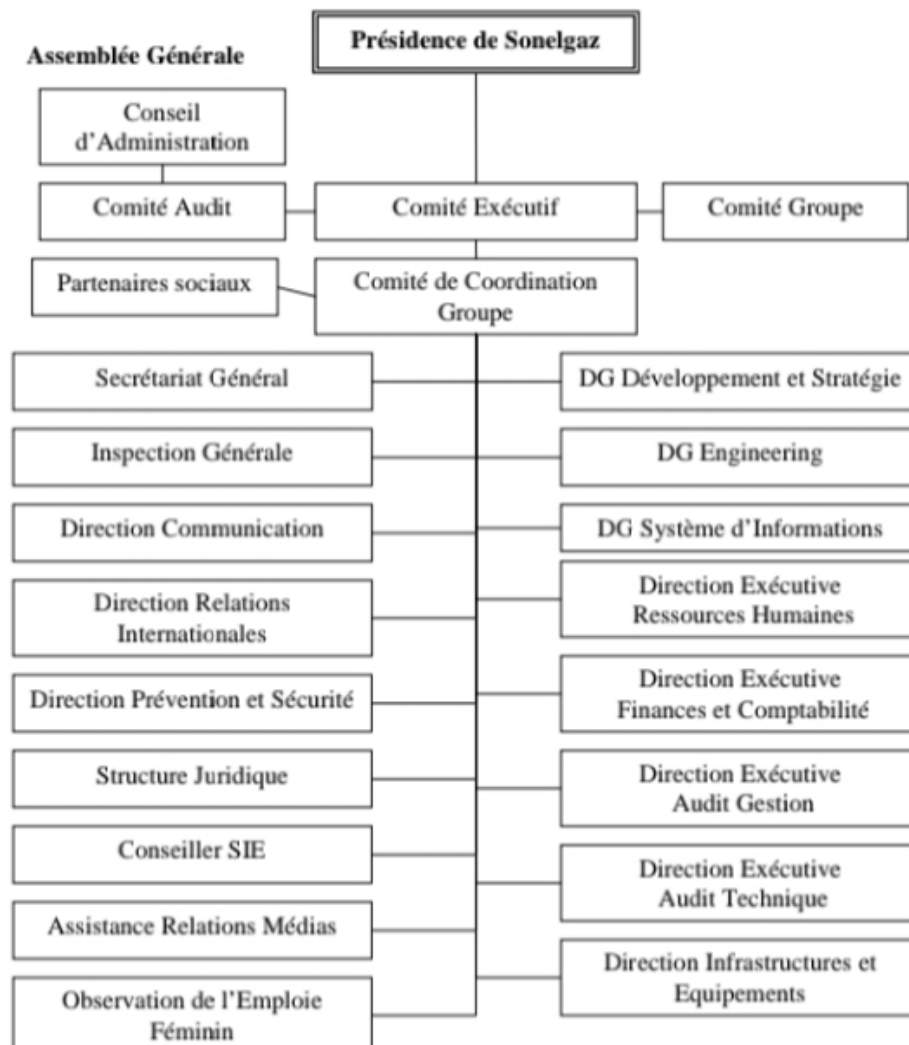


FIGURE 2.5 – Organigramme de l'entreprise Sonelgaz

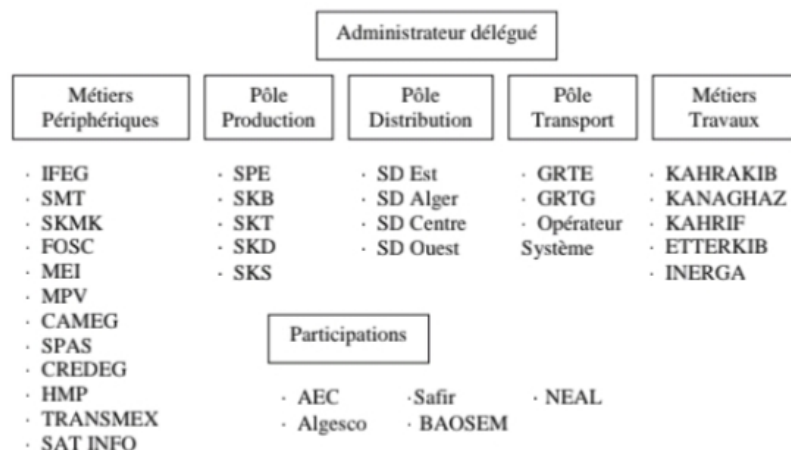


FIGURE 2.6 – Organigramme de l'entreprise Sonelgaz

2.6 Société de distribution d'électricité et du gaz -BLIDA-

2.6.1 Présentation de la wilaya de Blida

La wilaya de Blida se situe dans la partie nord du pays, dans la zone géographique du Tell central. Elle est limitée au nord par la wilaya de Tipaza et la wilaya d'Alger, à l'ouest par la wilaya de Ain Defla, au sud par la wilaya de Médéa et à l'Est par les wilayas de Boumerdes et de Bouira. Faisant partie de la wilaya d'Alger à l'époque, Blida fut promue au rang de la wilaya à partir du découpage administratif de 1974, en application des dispositions de l'ordonnance n° 74-69 du 02 Juillet 1974 portant refonte de l'organisation territoriale des wilayas.

A partir de l'année 1984 et conformément à la loi n° 84-09 du 04 Février 1984, la wilaya de Blida a fait l'objet d'une nouvelle délimitation de son territoire, avec 29 communes qui se repartissent en 12 daïras.

Mais à partir de Juillet 1997, quatre communes (Sidi Moussa, Ouled Chebel, Birtouta et Tessala El Merdja) sont intégrées au gouvernorat du Grand Alger. Ainsi la wilaya de Blida passe de 29 à 25 communes se répartissant sur 10 daïras.

La wilaya de Blida s'étend sur une superficie de 1478,62 Km².

Sa population en fin de l'année 2014, est estimée à 1 178 205 habitants.



FIGURE 2.7 – Carte administrative de la wilaya de Blida

2.6.2 Direction de distribution d'électricité et du gaz -BLIDA-

La direction de Distribution d'électricité et de gaz du Centre (SDC) est en charge de la distribution et de la commercialisation d'électricité en tant que filiale de la Sonelgaz à Blida et dans toute la wilaya de Blida. Dans la wilaya de Blida, la société de distribution d'électricité et de gaz du centre s'occupe de tout ce qui a trait à votre compteur électrique et gazière, aux lignes et à la facturation.

- ▶ Raccordement à l'électricité d'un nouveau compteur à Blida.
- ▶ Coupure d'électricité et du gaz.
- ▶ Problème ou question de compteur électrique et gazière.
- ▶ Factures et tarifs de l'électricité et du gaz à Blida.

2.6.3 Informations sur Sonelgaz de Blida

- ▶ Localisation : Boulevard Mohamed BOUDIAF-Blida-
- ▶ Nombre de communes :25.
- ▶ Nombre d'agences commerciales : 08(Blida Est, Blida Ouest, Boufarik , Mouzaia, Larbaa, Meftah, Oued Alleug, Ouled Yaich)
- ▶ Nombre de clients pour l'électricité :
 - ◇ Nombre de clients total : 347 818.

- ◇ Nombre de clients ordinaires : 336 465.
 - ◇ Nombre de clients administrations : 8 544.
 - ◇ Nombre de clients (MT) : 2 809.
 - ◇ Taux de pénétration : 99.5%.
- Nombre de clients pour le gaz :
- ◇ Nombre de clients total : 216 334.
 - ◇ Nombre de clients ordinaires : 213 863.
 - ◇ Nombre de clients administrations : 2 094.
 - ◇ Nombre de clients (MP) : 377.
 - ◇ Taux de pénétration : 68%.



FIGURE 2.8 – Direction de distribution d'électricité et du gaz -BLIDA-

CHAPITRE 3

APPLICATION SUR LA CONSOMMATION D'ÉLECTRICITÉ ET DU GAZ À LA WILAYA DE BLIDA

Introduction

Le présent chapitre est consacré pour une application de la méthodologie de Box&Jenkins, pour étudier, modéliser et prédire la consommation d'électricité et de gaz au niveau de la direction de Sonelgaz la Wilaya de Blida. Cette partie est élaboré par l'utilisation du logiciel statistique R et R-Studio, et à l'aide d'un ensemble des paquets des programme, tseries, forecast,....

3.1 Présentation des données

L'ensemble des données représente la consommation d'électricité et du gaz au niveau de la direction de Sonelgaz de la wilaya de Blida.

3.2 Méthodologie de Box & Jenkins

L'approche de Box& Jenkins (1976) est une méthodologie d'étude systématique des séries temporelles en fonction de leurs caractéristiques pour déterminer, dans la famille des modèles ARIMA, le plus adéquate à représenter le phénomène étudié. Le seul problème avec la modéli-

sation ARIMA est mathématiquement difficile et nécessite une analyse approfondie connaissance de la méthode. Il n'est pas facile de construire un modèle ARIMA sans une formation en analyse statistique et une connaissance de la méthodologie. Un modèle ARIMA combine trois types de processus aléatoires, la contribution de chacun étant spécifiée par l'ARIMA (p, d, q), où est l'ordre du processus auto régressif $AR(p)$, d le degré d'intégration d'un processus $I(d)$, et q l'ordre de la moyenne mobile $MA(q)$.(voir le référence[14])

3.2.1 Création d'une série chronologique

Créer un objet de type série temporelle :

serie = ts (data,start,end,frequency).

data : contient le vecteur des données.

start et end : mentionne les dates de début et de fin de la série.

frequency : mentionne le nombre de données par unité de temps.

Pour représenter la série temporelle graphiquement on utilise la commande : **plot.ts(nom de la série).**

3.2.2 Transformation

La fonction `log` permet de stabiliser la variance. La fonction `diff` permet d'appliquer l'opérateur de différenciation Δ_T^k en utilisant la fonction :

diff(serie,lag = T, difference = k).

Δ_T^k élimine une tendance et une saisonnalité d'ordre k . On peut aussi utiliser la fonction **decompose** pour extraire les composantes d'une série temporelle (via la méthode de la moyenne mobile) :

`decomp <- decompose(serie,type=c('additive','multiplicative')).`

la composante saisonnière: par `decompseasonal`, que

l'on suppose additive ou multiplicative dans l'option `type`.

la tendance : par `decomptrend`.

la partie aléatoire stationnaire de la série: par `decomprandom`.

Après avoir éliminer la tendance et la saisonnalité, l'étape suivante consiste à modéliser la série résiduelle obtenue.

La première chose à faire est donc de tester s'il y a une dépendance entre les termes de cette série. Si cette série est un bruit blanc, il n'est donc pas utile d'aller plus loin dans la modélisation Pour cela on utilise les fonctions suivantes :

1. La fonction **Box.test(série)** teste la bruité de la série. Si p-value > à 0.05 la série est un bruit blanc.
2. **qqplot**, **qqnorm** et **jarque.bera.test** ce sont des fonctions pour tester la normalité des résidus.
3. La fonction **adf.test(série)** teste la stationnarité de la série. si p-value < à 0.05 la série est stationnaire.

On peut également utiliser la fonction **KPSS.test(série)**, Si p-value > à 0.05 la série est stationnaire.

On peut aussi utiliser la fonction **pp.test(série)**, Si p-value < à 0.05 la série est stationnaire et admet une racine unitaire.

3.2.3 Identification

La fonction $acf(x)$ calcule (et trace si l'option plot est à TRUE) les lag.max premières auto-corrélations empiriques et auto-covariances.

La fonction $pacf(x)$ calcule et trace les auto-corrélations partielles empiriques.

3.2.4 Estimation des paramètres

Estimation des paramètres du modèle :

Dans le cas général la fonction arima permet d'estimer les paramètres :

- D'un ARMA(p,q) : `out=arima(serie,order=c(p,0,q))`.
- D'un ARIMA(p,d,q) : `out=arima(serie,order=c(p,d,q))`.
- D'un SARIMA(p,d,q,T) :
`out=arima(serie,order=c(p,d,q),seasonal=list(order=c(P,D,Q)))`.

3.2.5 Tests d'adéquation et choix du modèle (validaion)

Après avoir estimer le modèle, il faut s'assurer encore une fois que les résidus vérifient les propriétés d'un bruit blanc sans oublier de tester aussi la normalité, on utilise alors :

- ♣ `Box.test(outresid)` pour tester la bruité.
- ♣ `adf.test(outresid)` pour tester la stationnarité.
- ♣ `qqnorm(outresid)` pour tester la stationnarité graphiquement.
- ♣ `jarque.bera.test(outresid)` pour tester la normalité la normalité des résidus.

Prévision :

Si à la fin on a plusieurs modèles valides on choisit parmi eux celui qui donne les meilleurs critères d'ajustement et les meilleurs performances en prévision.

La fonction `predict` permet de faire les prévisions à l'horizon h :

```
prevision1=predict(out,n.ahead=h)
```

3.3 Application

Les données considérées sont mensuelles et la période retenue pour l'étude est de premier mois de 2006 à 12^{ime} mois de 2021.

3.3.1 Application sur les données d'électricité :

3.3.1.1 La consommation d'électricité pour les clients ordinaires :

Analyse graphique de la série "consommation d'électricité pour les clients ordinaires" :

A partir des données de la série et avec l'ensemble des commandes suivantes :

```
ts=ts(x,frequency=12)
plot(ts)
```

Nous obtenons le graphe suivant :

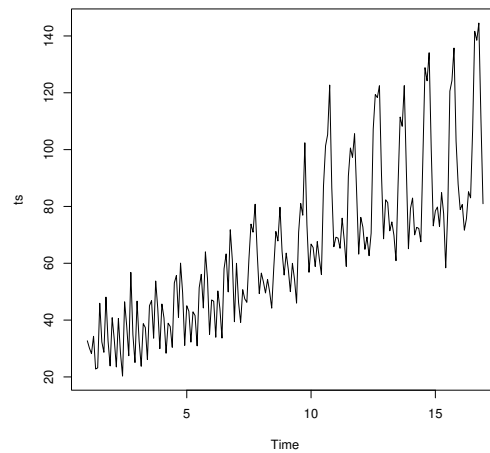


FIGURE 3.1 – La consommation d'électricité pour les clients ordinaires

D'après la figure 3.1 on remarque que la tendance de la consommation d'électricité pour les clients ordinaires augmente.

Le corrélogramme simple :

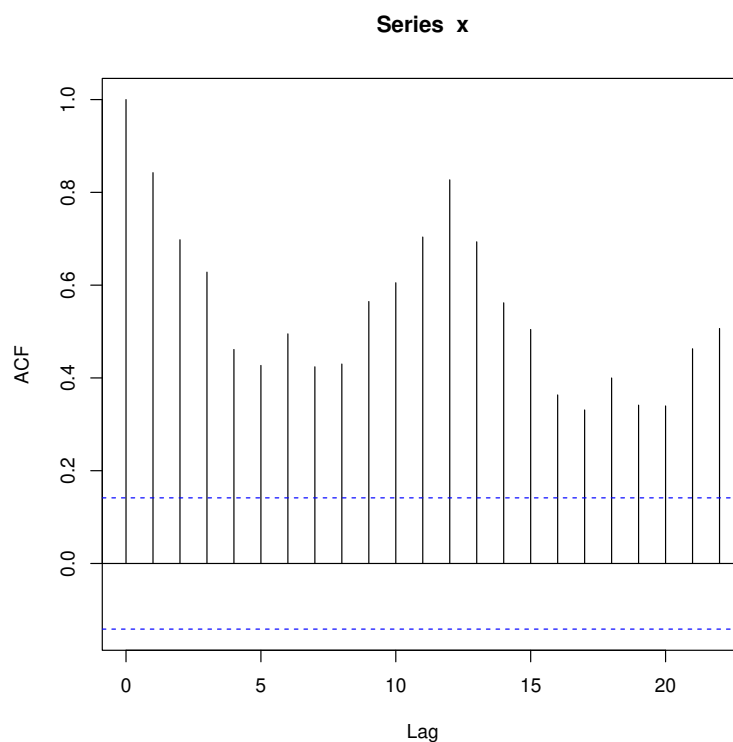


FIGURE 3.2 – Corrélogramme simple de la consommation d'électricité pour les clients ordinaires

Le corrélogramme partiel :

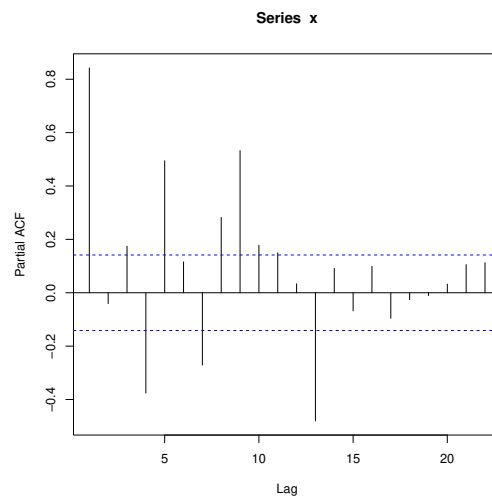


FIGURE 3.3 – Corrélogramme partiel de la consommation d'électricité pour les clients ordinaires

La visualisation des corrélogrammes simple et partiel montre une non stationnarité propre, l'ACF ne coupe pas.

On utilise aussi le test de racine unitaire (test ADF) pour confirmer la non stationnarité :

```
library(tseries)
```

```
adf.test(x)
```

Augmented Dickey-Fuller Test

```
data: x
```

```
Dickey-Fuller = -7.0793, Lag order = 5, p-value = 0.1
```

p-value = 0.1 > 0.05 donc la série n'est pas stationnaire.

(Admet une racine unitaire).

On peut aussi utiliser la fonction **decompose** pour extraire les composantes d'une série temporelle.

```
dec.x=decompose(x)
```

```
plot(dec.x)
```

Transformation de la série :

On utilise la commande **ndiffs** pour savoir l'ordre de différenciation qui éliminé la tendance et la saisonalité.

```
library(forecast)
```

```
ndiffs(x)
```

```
[1] 1
```

Donc :

```
d=1
```

```
x5=diff(x,df=5)
```

La représentation graphique est :

```
plot.ts(x5,col='red')
```

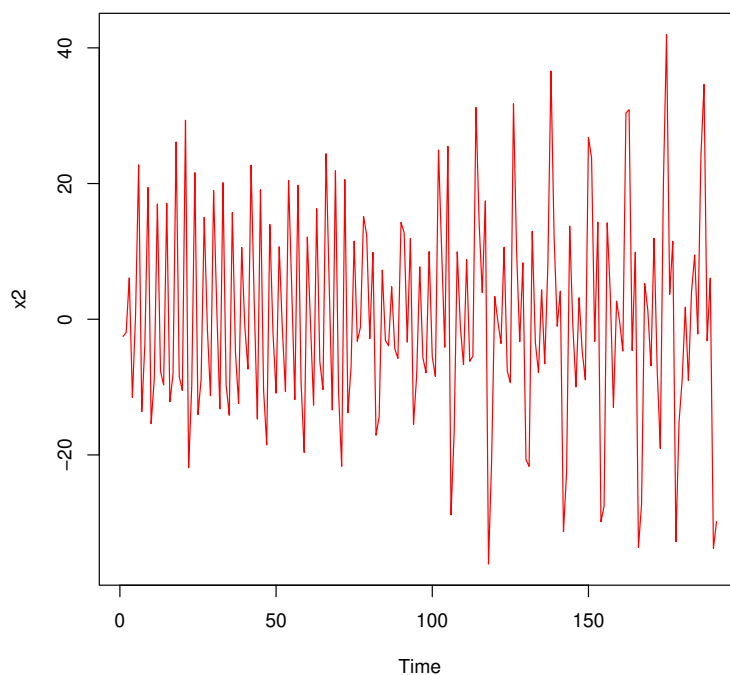


FIGURE 3.4 – La série différencier de la consommation d'électricité(clients ordinaires)

On applique l'ensemble des tests de la stationnarité on obtient :

```
library(tseries)
```

```
kpsstest(x5)
```

KPSS Test for Level Stationarity

```
data: x5
```

```
KPSS Level = 0.013051, Truncation lag parameter = 4, p-value = 0.1
```

```
library(tseries)
```

```
pp.test(x5)
```

```
Phillips-Perron Unit Root Test
```

```
data: x5
```

```
Dickey-Fuller Z(alpha) = -174.68, Truncation lag parameter = 4, p-value = 0.01
```

```
alternative hypothesis: stationary
```

```
library(tseries)
```

```
adf.test(x5)
```

```
Augmented Dickey-Fuller Test
```

```
data: x5
```

```
Dickey-Fuller = -6.2294, Lag order = 5, p-value = 0.01
```

D'après les résultats des trois test on résume que la série différencier est stationnaire.

Identification :

Les graphes d'auto-corrélation simple et partiel de la série différencier sont :

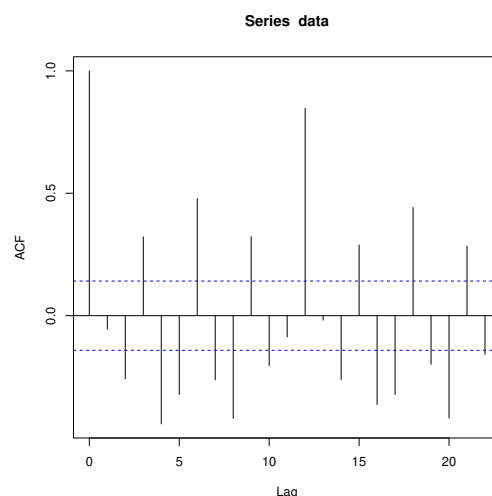


FIGURE 3.5 – Le corrélogramme simple de la série différencier de la consommation d'électricité(clients ordinaires)

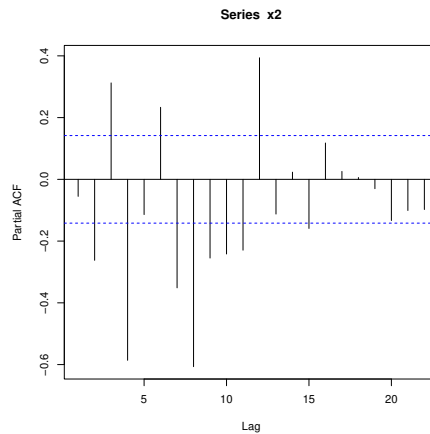


FIGURE 3.6 – Le corrélogramme partiel de la série différencier de la consommation d'électricité(clients ordinaires)

Estimation des paramètres :

On utilise la commande `auto.arima` pour construire le meilleur modèle qui représente la série.

```
library(forecast)
```

```
Modèle=auto.arima(x,trace=TRUE,test="kpss",ic="aic")
```

Qui donne que le meilleur modèle (0,1,0) car il minimise le critère AIC.

Aussi on a la commande `summary(Modèle)` qui donne les coefficients du modèle **ARIMA(0,1,0)**

```
summary(Modèle)
```

```
Series: x
```

```
ARIMA(0,1,0)
```

TABLE 3.1 – Les paramètres d'ajustement du modèle ARIMA(0,1,0) pour la consommation d'électricité(clients ordinaires)

σ^2	235.6
log likelihood	-792.64
AIC	1587.29
AICc	1587.31
BIC	1590.54

Training set error measures : TSEM

Training set : TS

TABLE 3.2 – Les paramètres des erreurs d'ajustement du modèle ARIMA(0,1,0) pour la consommation d'électricité (clients ordinaires)

TSEM	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
TS	0.2514204	15.30858	12.36736	-3.023391	21.36937	0.9948054	-0.05511583

Validation :

Nous avons diagnostiqué notre modèle à partir des tests suivants pour montrer que le modèle choisi est valide.

```
library(tseries)
```

```
Test ADF:
```

```
adf.test(Modèle$residuals)
```

```
Augmented Dickey-Fuller Test
```

```
data: Modèle$residuals
```

```
Dickey-Fuller = -5.9886, Lag order = 5, p-value = 0.01
```

```
alternative hypothesis: stationary
```

```
Test de Box-pierce :
```

```
Box.test(Modèle$residuals)
```

```
Box-Pierce test
```

```
data: Modèle$residuals
```

```
X-squared = 1.2931, df = 1, p-value = 0.2555
```

Tester de la stationnarité graphiquement par la commande : **qqnorm(x5)**.

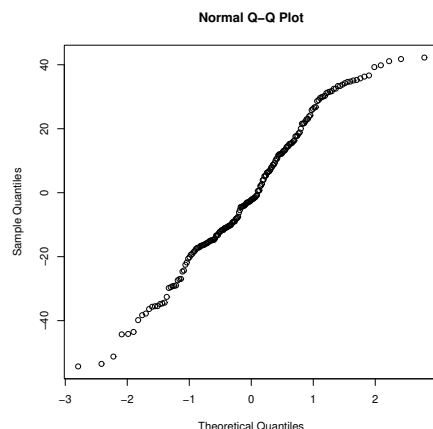


FIGURE 3.7 – Le graphe qqnorm des résidus du modèle ARIMA(0,1,0) de la consommation d'électricité (clients ordinaires)

TABLE 3.3 – Les différents modèles ajustés pour la consommation d'électricité (clients ordinaires)

Modèle	la valeur de AIC
ARIMA(2,1,2) with drift	Inf
ARIMA(0,1,0) with drift	1583.772
ARIMA(1,1,0) with drift	1586.148
ARIMA(0,1,1) with drift	1584.63
ARIMA(0,1,0)	1581.824
ARIMA(1,1,1) with drift	Inf

Meilleur modèle est : ARIMA(0,1,0) avec AIC=1581.824.

Conclusion : Le Modèle ARIMA(0,1,0) est valable.

Prévision :

D'après le modèle choisis précédents on peut prédire des résultats pour l'année prochaine.

On utilise l'ensemble des commandes suivantes :

```
library(forecast)
m_aa=auto.arima(x)
f_aa=forecast(m_aa,h=12)
f_aa
```

La représentation graphique des valeurs prédictives est : `plot(f_aa,col='red')`.

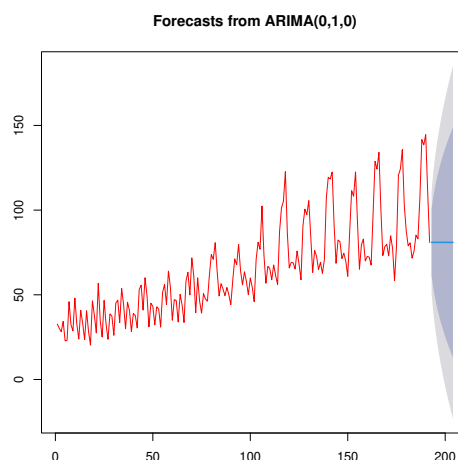


FIGURE 3.8 – Les valeurs prédictives de la consommation d'électricité (clients ordinaires)

TABLE 3.4 – Les valeurs prédictives par le modèle ARIMA(0,1,0) pour la consommation d'électricité (clients ordinaires)

Point	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
193	80.95	61.2799720	100.6200	50.867288	111.0327
194	80.95	53.1323796	108.7676	38.406621	123.4934
195	80.95	46.8805121	115.0195	28.845215	133.0548
196	80.95	41.6099439	120.2901	20.784576	141.115
197	80.95	36.9664802	124.9335	13.683011	148.2170
198	80.95	32.7684681	129.1315	7.262706	154.6373
199	80.95	28.9079975	132.9920	1.358626	160.5414
200	80.95	25.3147592	136.5852	-4.136758	166.0368
201	80.95	21.9399159	139.9601	-9.298136	171.1981
202	80.95	18.7479098	143.1521	-14.179888	176.0799
203	80.95	15.7118974	146.1881	-18.823068	180.7231
204	80.95	12.8110241	149.0890	-23.259571	185.1596

3.3.1.2 La consommation d'électricité pour les clients administrations

Analyse graphique de la série "consommation d'électricité pour les clients administrations" :

A partir des données de la série et avec l'ensemble des commandes suivantes :

```
ts1=ts(x, frequency=12)
plot(ts1)
```

On obtient le graphe suivant :

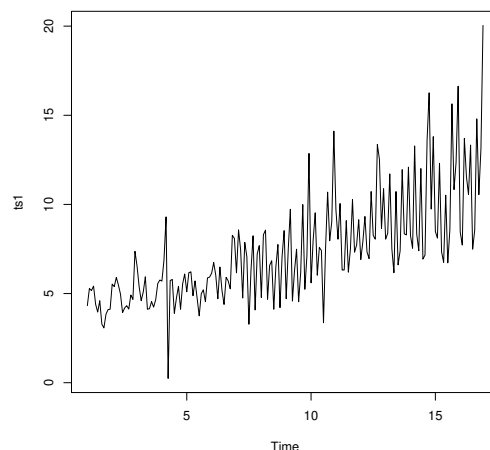


FIGURE 3.9 – La consommation d'électricité pour les clients administrations

D'après la figure 3.9 on remarque que la tendance de la consommation d'électricité pour les clients administrations augmente.

Le corrélogramme simple :

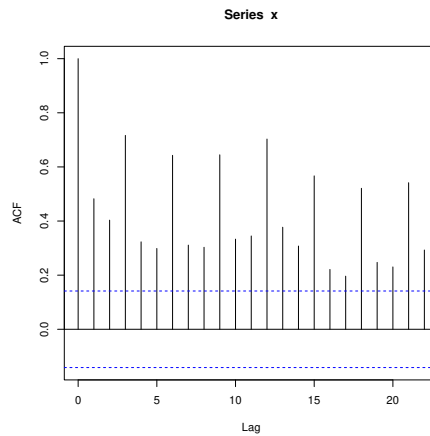


FIGURE 3.10 – Corrélogramme simple de la consommation d'électricité pour les clients administrations

le corrélogramme partiel :

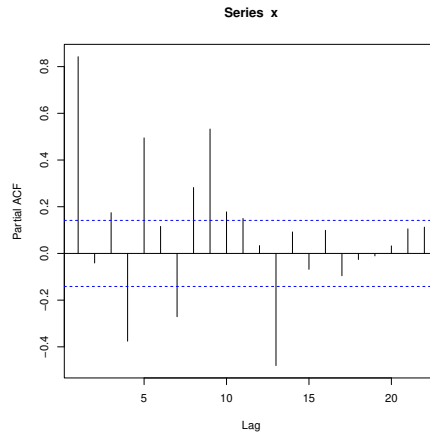


FIGURE 3.11 – Corrélogramme partiel de la consommation d'électricité pour les clients administrations

La visualisation des corrélogrammes simple et partiel montre une non stationnarité propre, l'ACF ne coupe pas.

On utilise aussi le test de racine unitaire (**test ADF**) pour confirmer la non stationnarité :

```
library(tseries)
```

```
adf.test(x)
```

Augmented Dickey-Fuller Test

```
data: x
```

```
Dickey-Fuller = -5.2984, Lag order = 5, p-value = 0.1
```

```
p-value=0.1>0.05 Donc la série est n'est pas stationnaire.\\  
(Admet une racine unitaire).
```

On peut aussi utiliser la fonction **decompose** pour extraire les composantes d'une série temporelle.

```
dec.x=decompose(x)  
plot(dec.x)
```

Transformation de la série :

On utilise la commande **ndiffs** pour savoir l'ordre de différenciation qui élimine la tendance et la saisonnalité .

```
library(forecast)
```

```
ndiffs(x)
```

```
[1] 1
```

```
Donc : df= 1
```

```
x5=diff(x,df=1)
```

La représentation graphique de la série différencier est : **plot.ts(x5,col='red')**

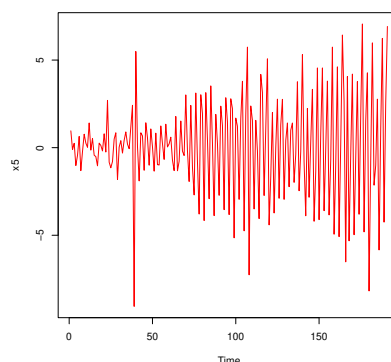


FIGURE 3.12 – La série différencier de la consommation d'électricité(clients administrations)

On applique l'ensemble des tests de la stationnarité on obtient :

```
library(tseries)
```

```
kpss.test(x5)
```

KPSS Test for Level Stationnarité

```
data: x5
```

KPSS Level = 0.069351, Truncation lag paramètre = 4, p-value = 0.1

```
pp.test(x5)
```

Phillips-Perron Unit Root Test

Dickey-Fuller Z(alpha) = -223.16, Truncation lag parameter = 4, p-value = 0.01

alternative hypothèses: stationary

```
adf.test(x5)
```

Augmented Dickey-Fuller Test

Dickey-Fuller = -7.2449, Lag order = 5, p-value = 0.01

alternative hypothèses: stationary

D'après les résultats des trois test on résume que la série différencier est stationnaire.

Identification :

Les graphes d'auto corrélation simple et partiel de la série différencier sont :

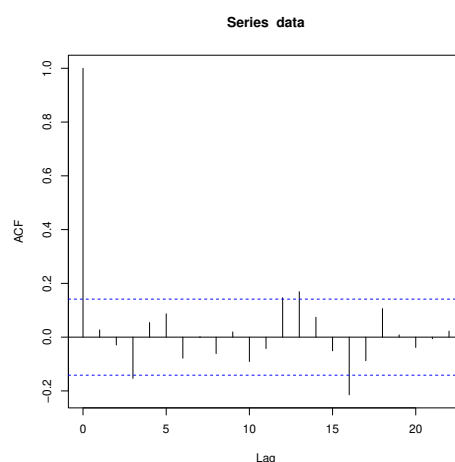


FIGURE 3.13 – Le corrélogramme simple de la série différencier de la consommation d'électricité(clients administrations)

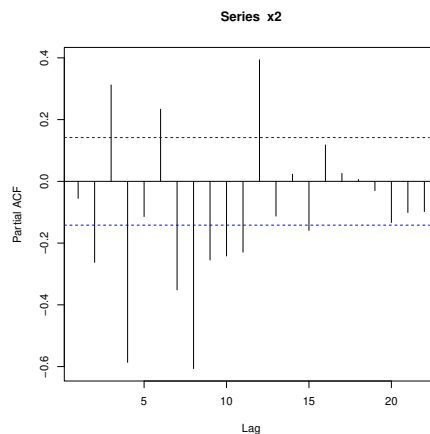


FIGURE 3.14 – Le corrélogramme partiel de la série différencier de la consommation d'électricité (clients administrations)

Estimation des paramètres :

On utilise la commande `auto.arima` pour construire le meilleur modèle qui représente la série .

```
library(forcast)
```

```
Modèle=auto.arima(x, trace=TRUE, test="kpss", ic="aic")
```

Qui donne le meilleur modèle(5,1,2) car il minimise le critère AIC.

Aussi on a la commande `summary(Modèle)` qui donne les coefficients du modèle **ARIMA(5,1,2)**

```
summary(Modèle)
```

```
séries=x
```

TABLE 3.5 – Les coefficients du modèle ARIMA(5,1,2) pour la consommation d'électricité (clients administrations)

	ar1	ar2	ar3	ar4	ar5	ma1	ma2	drift
coefficients	-0.3348	0.0096	0.3430	-0.1772	-0.4879	-0.3716	-0.5487	0.0396
s.e	0.1369	0.0789	0.0748	0.0974	0.0808	0.1379	0.1350	0.0059

TABLE 3.6 – Les paramètres d'ajustement du modèle ARIMA(5,1,2) pour la consommation d'électricité (clients administrations)

σ^2	2.264
log likelihood	-347.47
AIC	712.94
AICc	713.93
BIC	742.21

Training set erreur mesures : TSEM

Training set : TS

TABLE 3.7 – Les paramètres des erreurs d'ajustement du modèle ARIMA(5,1,2) pour la consommation d'électricité (clients administrations)

TSEM	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
TS	-0.02205565	1.468887	1.018467	-16.65264	26.97151	0.4623123	0.02694084

Validation :

Nous avons diagnostiqué notre modèle à partir des tests suivants pour montrer que le modèle choisis est valide.

```
library(tséries)
```

```
test adf:
```

```
Augmented Dickey-Fuller Test
```

```
data: Modèle$residuals
```

```
Dikey-Fuller = -5.9133, Lag order = 5, p-value = 0.01
```

```
alternative hypothèses: stationary
```

```
Test de Box-pierce:
```

```
Box-Pierce test
```

```
X-squared = 0.13936, df = 1, p-value = 0.7089
```

Tester de la stationnarité graphiquement la commande :**qqnorm(x5)**.

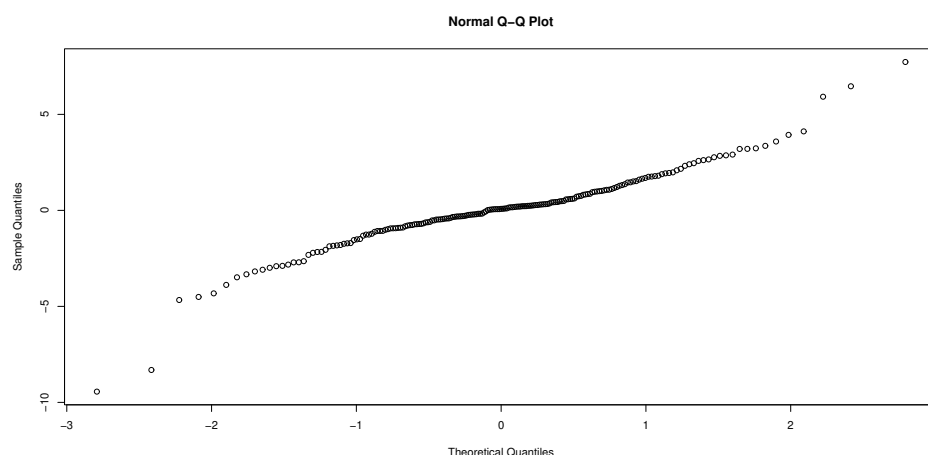


FIGURE 3.15 – Le graphe qqnorm des résidus du modèle ARIMA(5,1,2) pour la consommation d'électricité(clients administrations)

TABLE 3.8 – Les différents modèles ajustés pour la consommation d'électricité (clients administrations)

Modèle	Valeurs de AIC
ARIMA(2,1,2) with drift	788.4429
ARIMA(0,1,0) with drift	951.0428
ARIMA(1,1,0) with drift	917.07 1
ARIMA(0,1,1) with drift	828.6382
ARIMA(0,1,0)	949.1962
ARIMA(1,1,2) with drift	813.529
ARIMA(2,1,1) with drift	762.0698
ARIMA(1,1,1) with drift	834.9924
ARIMA(2,1,0) with drift	779.3275
ARIMA(3,1,1) with drift	760.4677
ARIMA(3,1,0) with drift	759.8818
ARIMA(4,1,0) with drift	757.9319
ARIMA(5,1,0) with drift	734.5987
ARIMA(5,1,1) with drift	713.5755
ARIMA(4,1,1) with drift	Inf
ARIMA(5,1,2) with drift	709.2565
ARIMA(4,1,2) with drift	Inf
ARIMA(5,1,3) with drift	Inf
ARIMA(4,1,3) with drift	Inf
ARIMA(5,1,2)	722.173

Meilleur modèle est ARIMA(5,1,2) avec AIC=709.2565.

Conclusion :

Le modèle ARIMA(5,1,2) est valable .

Prévision :

D'après le modèle choisis précédent on peut prédire des résultats pour l'année prochaine.

On utilise l'ensemble des commandes suivant :

```
library(forecast)
m_aa=auto.arima(x)
f_aa=forecast(m_aa,h=12)
f_aa
```

La représentation graphique des valeurs prédictives est :
plot(f_aa,col='red').

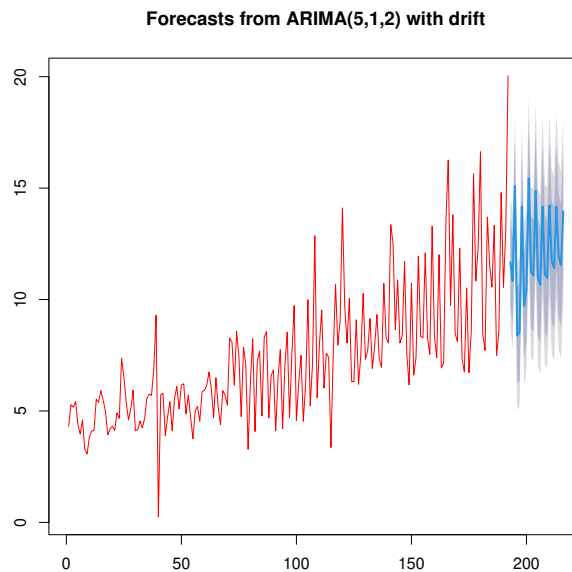


FIGURE 3.16 – Les valeurs prédictives de la consommation d'électricité (clients administrations)

TABLE 3.9 – Les valeurs de prédictives par le modèle (5,1,2) pour la consommation d'électricité (clients administrations)

Point	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
193	11.680601	9.752412	13.60879	8.731691	14.62951
194	10.832379	8.822825	12.84193	7.759032	13.90573
195	15.093154	13.083526	17.10278	12.019693	18.16662
196	8.376527	6.203665	10.54939	5.053422	11.69963
197	8.545371	6.355726	10.73502	5.196598	11.89414
198	14.180037	11.850614	16.50946	10.617492	17.74258
199	9.715563	7.347235	12.08389	6.093517	13.33761
200	10.498787	8.119526	12.87805	6.860021	14.13755
201	15.438818	13.027205	17.85043	11.750574	19.12706
202	11.245600	8.733659	13.75754	7.403918	15.08728
203	11.072660	8.556332	13.58899	7.224268	14.92105
204	14.889472	12.365791	17.41315	11.029834	18.74911

3.3.1.3 La consommation totale d'électricité

Analyse graphique de la série " consommation totale d'électricité " :

A partir des données de la série , et avec l'ensemble des commandes suivantes :

```
ts3=ts(x, frequency=12)}
```

```
plot(ts3)
```

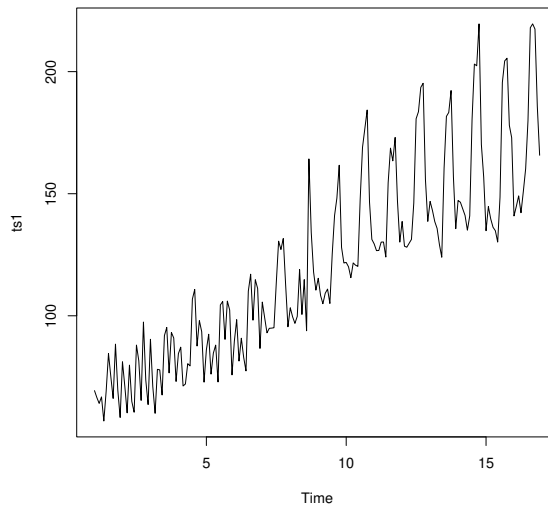


FIGURE 3.17 – La consommation totale d'électricité

D'après la figure 3.17 on remarque que la tendance de la consommation totale d'électricité augmente.

Le corrélogramme simple :

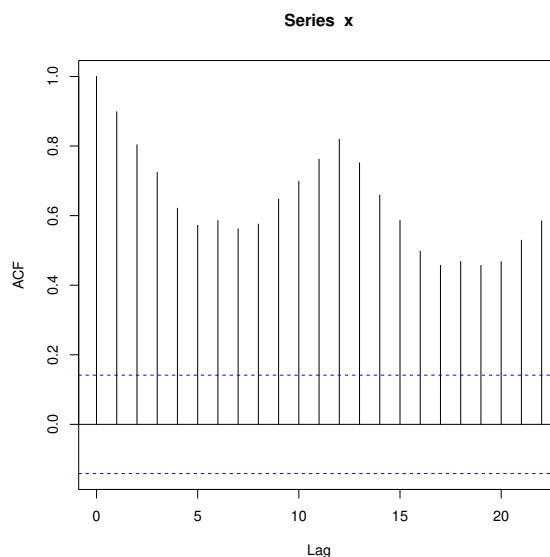


FIGURE 3.18 – Corrélogramme simple de la consommation totale d'électricité

Le corrélogramme partiel :

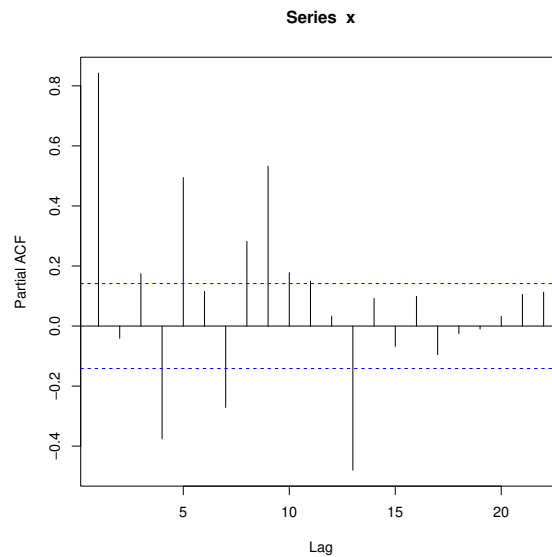


FIGURE 3.19 – Corrélogramme partiel de la consommation totale d'électricité

La visualisation des corrélogrammes simple et partiel montre une non stationnarité propre, l'ACF ne coupe pas.

On utilise aussi le test de racine unitaire (test ADF) pour confirmer la non stationnarité :

```
library(tseries)
adf.test(x)
Augmented Dickey-Fuller Test
data: x
Dickey-Fuller = -6.9333, Lag order = 5, p-value = 0.1
p-value=0.1>0.05 donc la série n'est pas stationnaire .\
(Admet une racine unitaire).
```

On peut aussi utiliser la fonction **décompose** pour extraire les composantes d'une série temporelle.

Transformation de la série :

On utilise la commande **ndiffs** pour savoir l'ordre de différenciation qui élimine la tendance et la saisonnalité.

```
library(forecast)
```

```
ndiffs(x)
```

```
x5=diff(x,df=5)
```

La représentation graphique de la série différencier est :

```
plot.ts(x5,col='red')
```

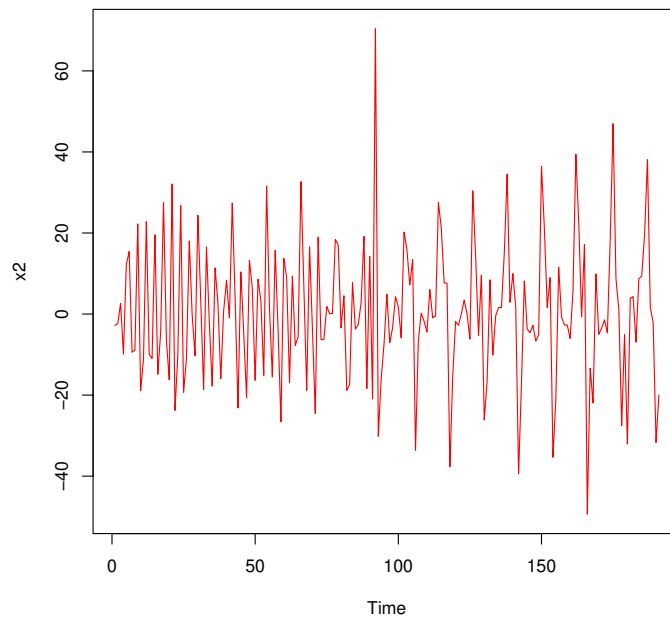


FIGURE 3.20 – La série différencier de la consommation totale d'électricité

On applique l'ensemble des test de la stationnarité on obtient :

```
library(tseries)
```

```
kps.test(x5)
```

```
KPSS Test for Level Stationarity
```

```
data: x5
```

```
KPSS Level = 0.0099144, Truncation lag parameter = 4, p-value = 0.1
```

```
Phillips-Perron Unit Root Test
```

```
Dickey-Fuller Z(alpha) = -183.34, Truncation lag parameter = 4, p-value = 0.01
```

```
alternative hypothesis: stationary
```

```
adf.test(x5)
```

```
Augmented Dickey-Fuller Test
```

Dickey-Fuller = -7.2331, Lag order = 5, p-value = 0.01

alternative hypothesis: stationary

D'après les résultats des trois test on résume que la série différencier est stationnaire.

Identification :

Les graphes d'autocorrélations simple et partiel de la série différencier :

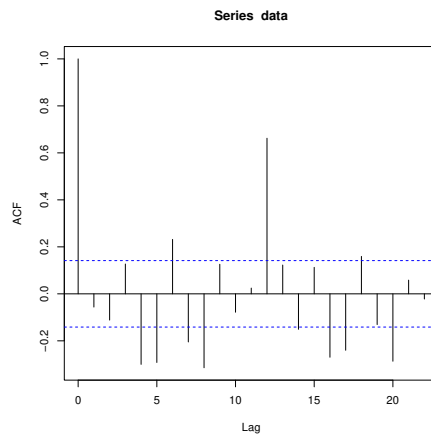


FIGURE 3.21 – Le corrélogramme simple de la série différencier de la consommation totale d'électricité

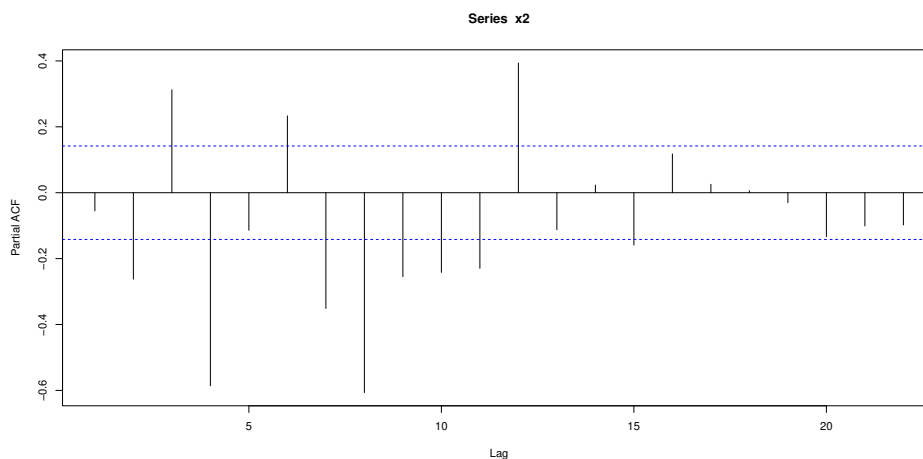


FIGURE 3.22 – Le corrélogramme partiel de la série différencier de la consommation totale d'électricité

Estimation des paramètres :

On utilise la commande `auto.arima` pour construire le meilleur modèle qui représente la série.

```
library(forecast)
```

```
Modèle=auto.arima(x, trace=TRUE, test="kpss", ic="aic")
```

Qui donne que le meilleur modèle est ARIMA(0,1,0) car il minimise le critère AIC.

Aussi on a la commande **summary(Modèle)** qui donne :

TABLE 3.10 – Les paramètres d’ajustement du modèle ARIMA(0,1,0) pour la consommation totale d’électricité

σ^2	log likelihood	AIC	AICc	BIC
289.6	-812.37	1626.74	1626.76	1629.99

Training set error measures : TSEM

Training set : TS

TABLE 3.11 – Les paramètres des erreurs d’ajustement du modèle ARIMA(0,1,0) pour la consommation totale d’électricité

TSEM	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
TS	0.5024966	16.97432	12.69989	-0.6537077	11.29615	0.99482	-0.05630178

Validation :

Nous avons diagnostiqué notre modèle à partir des tests suivants pour :

Montrer que le modèle choisis est valide.

```
library(tseries)
```

```
Test adf.test(Modèle$residuals)
```

Augmented Dickey-Fuller Test

```
Data: Modèle$residuals
```

```
Dickey-Fuller = -7.237, Lag order = 5, p-value = 0.01
```

```
Alternative hypothesis: stationary
```

```
Test de box-pierce:
```

```
Box-Pierce test
```

```
Data: Modèle$residuals
```

```
X-squared = 0.60862, df = 1, p-value = 0.4353
```

Tester les stationnarité graphiquement la commande : **qqnorm(x5)**.

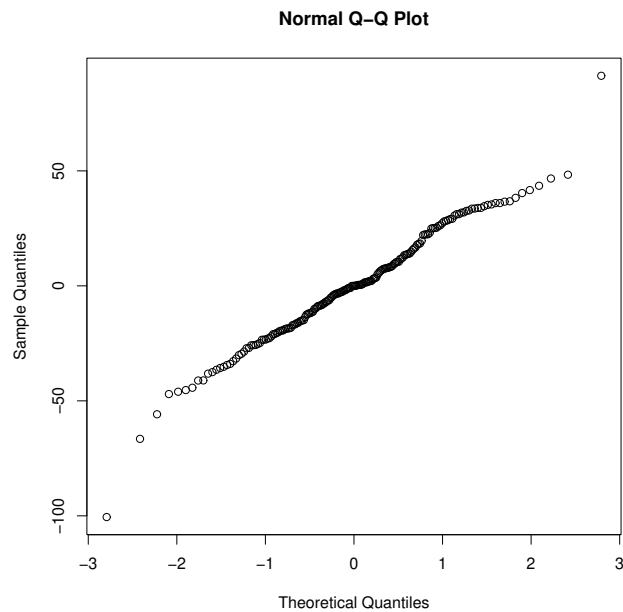


FIGURE 3.23 – Le graphe qqnorm des résidus du modèle ARIMA (0,1,0) pour la consommation totale d'électricité

TABLE 3.12 – Les différents modèles ajustés pour la consommation totale d'électricité

Modèle	la valeur de AIC
ARIMA(2,1,2) with drift	Inf
hline ARIMA(0,1,0) with drift	1622.905
hline ARIMA(1,1,0) with drift	1625.258
hline ARIMA(0,1,1) with drift	1624.125
hline ARIMA(0,1,0)	1621.073
hline ARIMA(1,1,1) with drift	Inf

Meilleur modèle est :ARIMA(0,1,0) avec AIC=1621.073.

Conclusion : le modèle ARIMA(0,1,0) est valable.

Prévision :

D'après le modèle choisis précédent on peut prédire des résultats pour les deux prochaines années.

on utilisé l'ensemble des commandes suivant :

```
library(forecast)
m_aa=auto.arima(x)
f_aa=forecast(m_aa,h=12)
f_aa
```

La représentation graphique des valeurs prédictives est :

`plot(f_aa,col='red')`.

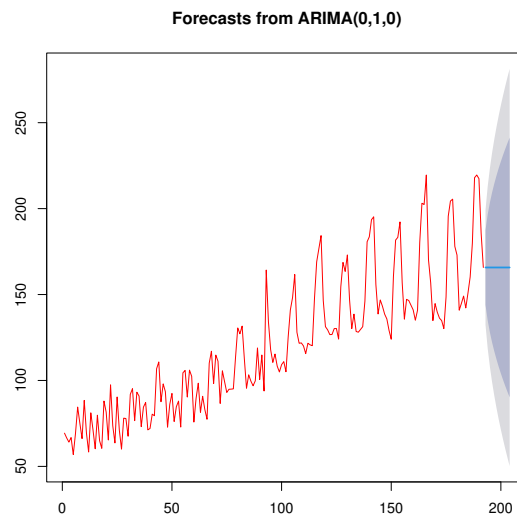


FIGURE 3.24 – Les valeurs prédictives par le modèle ARIMA(0,1,0) pour la consommation totale d'électricité

TABLE 3.13 – Les valeurs prédictives de la consommation totale d'électricité

Point	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
193	165.75	143.93966	187.5603	132.39397	199.1060
194	165.75	134.90552	196.5945	118.57745	212.9226
195	165.75	127.97339	203.5266	107.97566	223.5243
196	165.75	122.12932	209.3707	99.03793	232.4621
197	165.75	116.98060	214.5194	91.16364	240.3364
198	165.75	112.32580	219.1742	84.04474	247.4553
199	165.75	108.04527	223.4547	77.49823	254.0018
200	165.75	104.06105	227.4390	71.40489	260.0951
201	165.75	100.31899	231.1810	65.68190	265.8181
202	165.75	96.77966	234.7203	60.26896	271.2310
203	165.75	93.41329	238.0867	55.12055	276.3794
204	165.75	90.19677	241.3032	50.20131	281.2987

3.3.2 Application sur les données du gaz

3.3.2.1 La consommation du gaz pour les clients ordinaires :

Analyse graphique de la série "consommation du gaz pour les clients ordinaires" :

A partir des données de la série, et avec l'ensemble des commandes suivantes :

```
ts1=ts(x, frequency=12)  
plot(ts1)
```

Nous obtenons le graphe suivant :

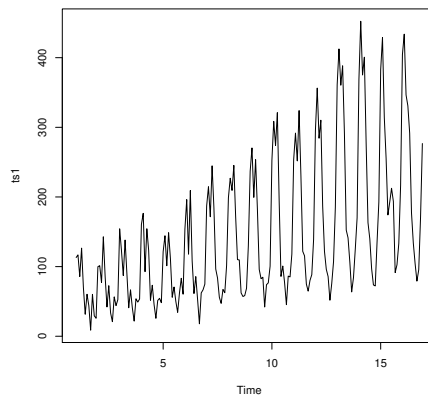


FIGURE 3.25 – La consommation du gaz pour les clients ordinaires

D'après la figure 3.25 on remarque que la tendance de la consommation du gaz pour les clients ordinaires augmente .

Le corrélogramme simple :

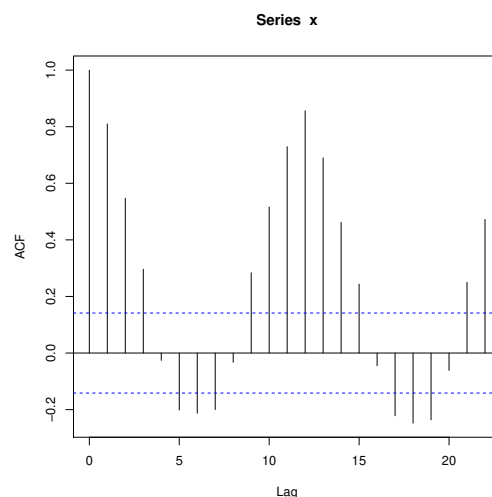


FIGURE 3.26 – Le corrolégramme simple de la consommation du gaz pour les clients ordinaires

Le corrélogramme partiel :

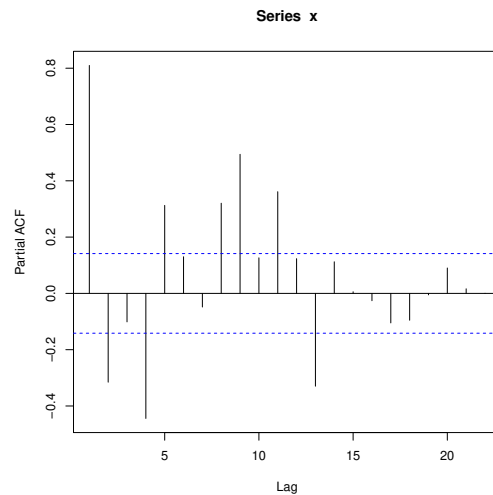


FIGURE 3.27 – Le corrolégramme partiel de la consommation du gaz pour les clients ordinaires

La visualisation des corrélogrammes simple et partiel montre une non stationnarité propre, l'ACF ne coupe pas.

On utilise aussi le test de racine unitaire (test ADF) pour confirmer la non stationnarité :

```
library(tseries)
```

```
adf.test(x)
```

Augmented Dickey-Fuller Test

Dickey-Fuller = -10.25, Lag order = 5, p-value = 0.1

p-value = 0.1 > 0.05 donc la série est n'est pas stationnaire.

(Admet une racine unitaire).

On peut aussi utiliser la fonction **decompose** pour extraire les composantes d'une série temporelle.

```
dec.x=decompose(datx)
```

```
plot(dec.x)}
```

Transformation de la série :

On utilise la commande **ndiffs** pour savoir l'ordre de différenciation qui élimine la tendance et la saisonnalité.

```
ndiffs(x)
```

```
data1=diff(x,df=5)
```

La représentation graphique de la série différencier est : `plot.ts(data1,col='red')`

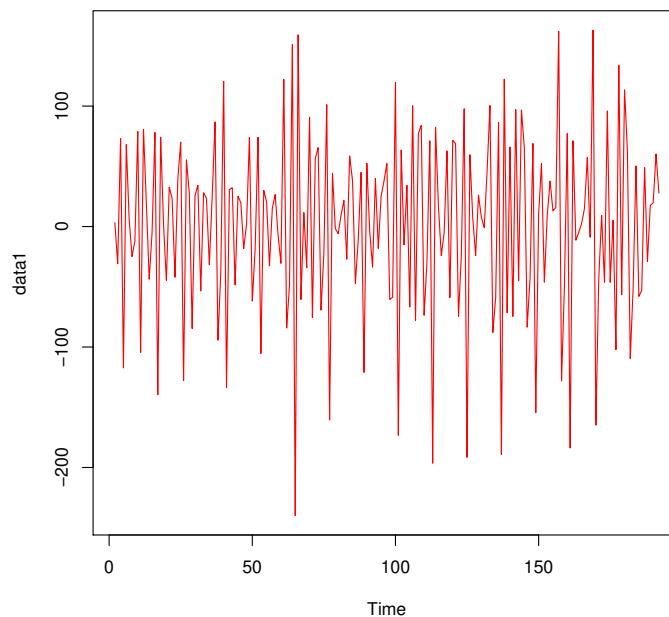


FIGURE 3.28 – La série différencier de la consommation du gaz pour les clients ordinaires

On applique l'ensemble des tests de la stationnarité on obtient :

```
library(tseries)
kpss.test(data1)
KPSS Test for Level Stationnarity
data: data1
KPSS Level = 0.010331, Truncation lag paramètre = 4, p-value = 0.1
library(tseries)
pp.test(data1)
Phillips-Perron Unit Root Test
data: data1
Dickey-Fuller Z(alpha) = -147.39, Truncation lag paramètre = 4, p-
value = 0.01
alternative hypothèses: stationary
library(tseries)
adiseseriesf.test(data1)
```

Augmentes Dickey-Fuller Test

data: data1

Dickey-Fuller = -8.432, Lag order = 5, p-value = 0.01

alternative hypothèses: stationary

D'après les résultats des trois tests on résume que la série différencier est stationnaire.

Identification :

Les graphes d'auto corrélation simple et partiel de la série diff sont :

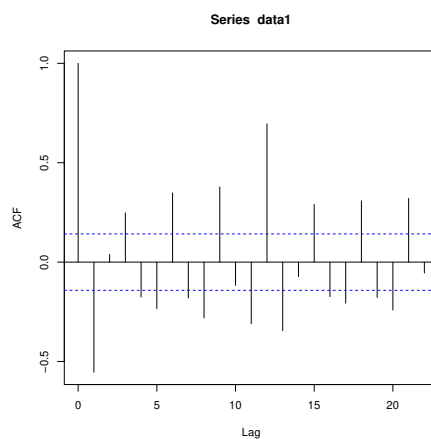


FIGURE 3.29 – Corrélogramme simple de la série différencier de la consommation du gaz pour les clients ordinaires

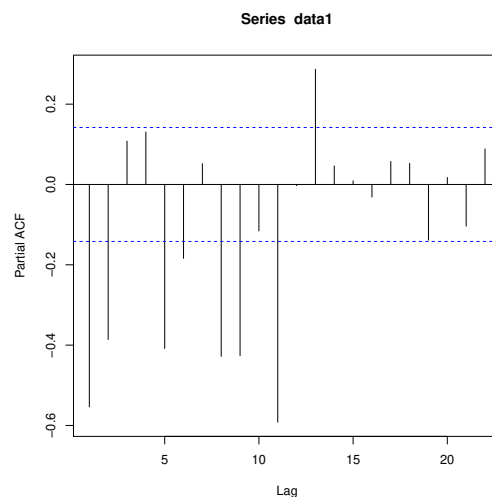


FIGURE 3.30 – Corrélogramme partiel de la série différencier de la consommation du gaz pour les clients ordinaires

Estimation des paramètres :

On utilise la commande (auto :arima) pour construire le meilleur modèle qui représente la série.

library(forecast)

Modèle=auto.arima(x, trace=TRUE, test="kpss", ic="aic")

Qui donne que le meilleur modèle est ARIMA(2,1,2) car il est minimise le critère AIC.

Aussi on a la commande **summary(Modèle)** qui donne les coefficients du modèle **ARIMA(2,1,2)**.

TABLE 3.14 – Les coefficients du modèle ARIMA(2,1,2) pour la consommation du gaz (clients ordinaires)

	ar1	ar2	ma1	ma2
Coefficients	0.3732	0.1662	-0.2073	-0.7355
s.e.	0.1003	0.0964	0.0647	0.0631

TABLE 3.15 – Paramètre d'ajustement du modèle ARIMA(2,1,2) pour la consommation du gaz (clients ordinaires)

σ^2	2963
log likelihood	-1033.26
AIC	2076.52
AICc	2076.85
BIC	2092.78

Training set error measures :TSEM

Training set :TS

TABLE 3.16 – Paramètre des erreurs d'ajustement du modèle ARIMA(2,1,2) pour la consommation du gaz (clients ordinaires)

TSEM	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
TS	5.39919	53.7206	40.3457	-10.4579	36.3987	0.83202	0.0084

Validation :

Nous avons diagnostiqué notre modèle à partir des tests suivants pour montrer que le modèle choisis est valide.

Test ADF `adf.test(Modèle$residuals)`

Augmented Dickey-Fuller Test

```
data: Modèle$residuals
```

```
Dickey-Fuller = -9.4628, Lag order = 5, p-value = 0.01
```

```
alternative hypothesis: stationary
```

```
Box-Pierce test
```

```
data: Modèle$residuals
```

```
X-squared = 0.013601, df = 1, p-value = 0.9072
```

```
acf(data)
```

```
data1=diff(x,df=5)
```

Tester de la stationnarité graphiquement par la commande :

qqnorm(data1)

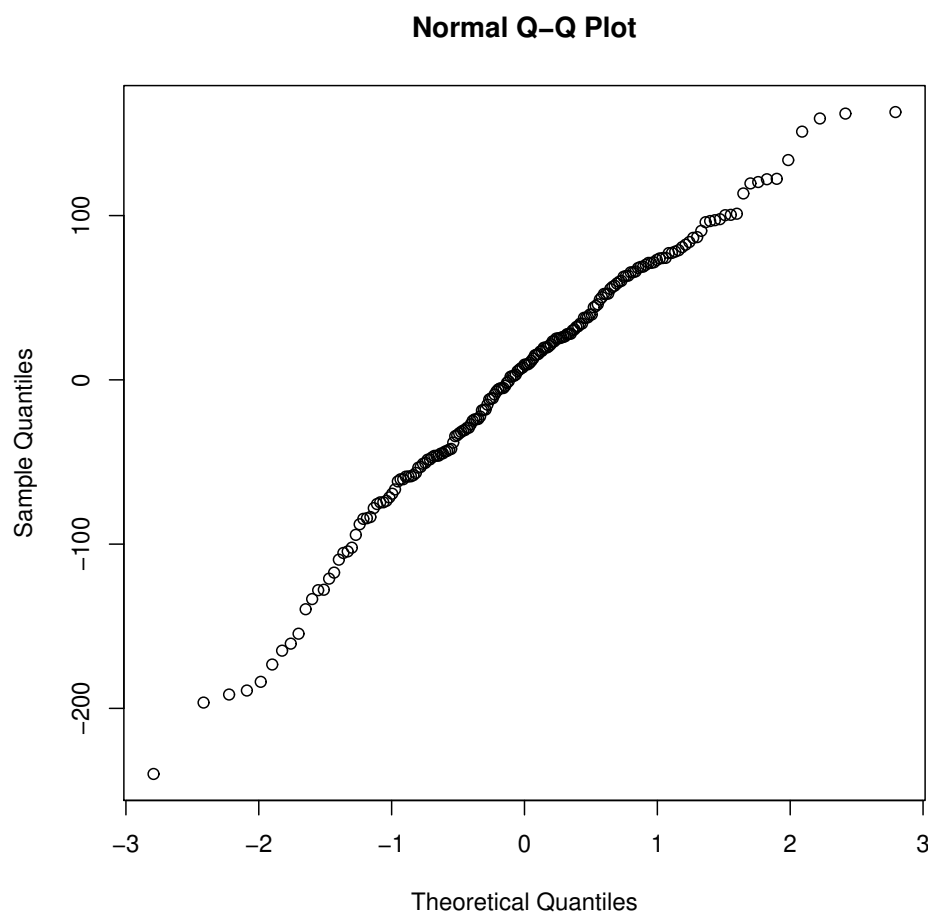


FIGURE 3.31 – Le graphe qqnorm des résidus du modèle ARIMA de la consommation du gaz (clients ordinaires)

TABLE 3.17 – Les différents modèles ajustés pour la consommation du gaz (clients ordinaires)

Modèles	La valeur AIC
ARIMA(2,1,2) with drift	Inf
ARIMA(0,1,0) with drift	2117.431
ARIMA(1,1,0) with drift	2111.348
ARIMA(0,1,1) with drift	2106.549
ARIMA(0,1,0)	2115.467
ARIMA(1,1,1) with drift	2094.282
ARIMA(2,1,1) with drift	2094.814
ARIMA(1,1,2) with drift	2096.027
ARIMA(0,1,2) with drift	2096.303
ARIMA(2,1,0) with drift	2113.217
ARIMA(1,1,1)	2092.283
ARIMA(0,1,1)	2104.581
ARIMA(1,1,0)	2109.379
ARIMA(2,1,1)	2092.893
ARIMA(1,1,2)	2094.03
ARIMA(0,1,2)	2094.347
ARIMA(2,1,0)	2111.262
ARIMA(2,1,2)	2068.188
ARIMA(3,1,2)	Inf
ARIMA(2,1,3)	Inf
ARIMA(1,1,3)	2080.19
ARIMA(3,1,1)	2081.91
ARIMA(3,1,3)	Inf

Meilleur modèle est : $ARIMA(2,1,2) = 2068.188$

Conclusion : Le modèle $ARIMA(2,1,2)$ est valable.

Prévision :

D'après le modèle choisis précédent on peut prédire des résultats pour l'année prochaine.

On utilise l'ensemble des commandes suivantes :

```
library(forecast)
m_aa=auto.arima(x)
f_aa=forecast(m_aa,h=12)
f_aa
```

La représentation graphique des valeurs prédictives est :

`plot(f_aa,col='red')`.

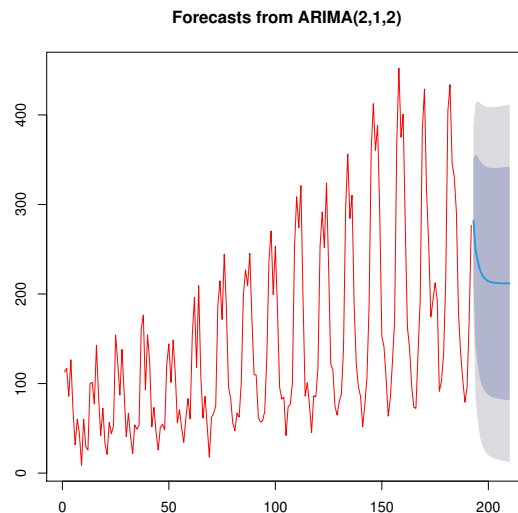


FIGURE 3.32 – Représentation graphique des valeurs prédictives pour la consommation du gaz (clients ordinaires)

TABLE 3.18 – Les valeurs de prédiction par le modèle ARIMA(2,1,2) pour la consommation du gaz (clients ordinaires)

Point	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
193	281.8282	212.06810	351.5882	175.13934	388.5170
194	248.7264	141.57340	355.8794	84.85004	412.6028
195	237.2166	120.63234	353.8008	58.91638	415.5167
196	227.4207	105.79668	349.0447	41.41282	413.4286
197	221.8523	97.77453	345.9301	32.09172	411.6129
198	218.1465	92.63579	343.6571	26.19446	410.0984
199	215.8381	89.41909	342.2572	22.49689	409.1794
200	214.3609	87.29769	341.4241	20.03449	408.6873
201	213.4260	85.86375	340.9883	18.33637	408.5157
202	212.8316	84.85259	340.8107	17.10457	408.5587
203	212.4545	84.10723	340.8017	16.16430	408.7447
204	212.2150	83.52901	340.9009	15.40677	409.0232

3.3.2.2 La consommation du gaz pour les clients administrations :

Analyse graphique de la série " consommation du gaz pour les clients administrations ".

A partir des données de la série, et avec l'ensemble des commandes suivantes :

$ts2=ts(x, \text{fréquence}=12)$

plot(ts2)

Nous obtenons le graphe suivant :

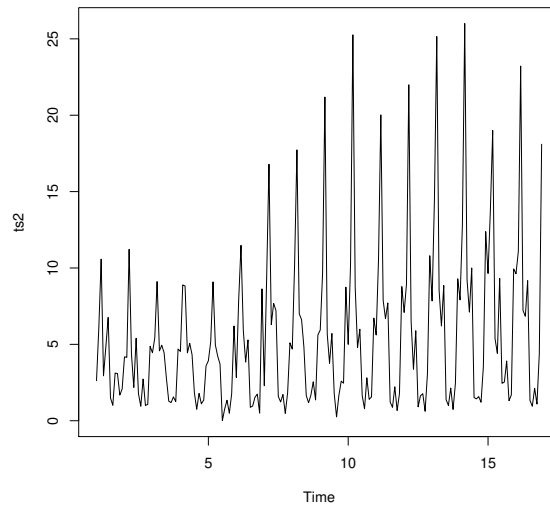


FIGURE 3.33 – La consommation du gaz pour les clients administrations

D'après la figure 3.33 on remarque que la tendance de la consommation du gaz pour les clients administrations augmente.

Le corrélogramme simple :

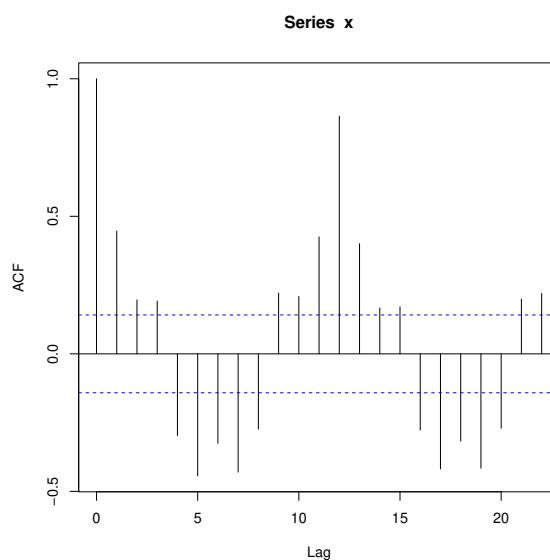


FIGURE 3.34 – Corrélogramme simple de la consommation du gaz pour les clients administrations

Le corrélogramme partiel :

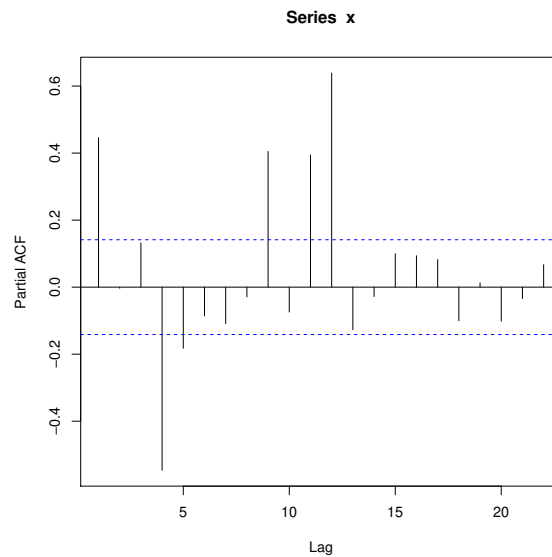


FIGURE 3.35 – Corrélogramme partiel de la consommation du gaz pour les clients administrations

La visualisation des corrélogramme simple et partiel montre une non stationnarité propre, l'ACF ne coupe pas, On utilise aussi le test de racine unitaire (**test ADF**) pour confirmer la non stationnarité :

```
library(tseries)
```

```
adf.test(x):
```

Augmented Dickey-Fuller Test

```
data: x
```

```
Dickey-Fuller = -11.214, Lag order = 5, p-value = 0.1
```

```
p-value = 0.1 > 0.05 donc la série est n'est pas stationnaire .
```

(Admet une racine unitaire)

On peut utiliser la fonction **décompose** pour extraire les composantes d'une série temporelle .

```
dec.x=décompose(x)
```

```
plot(dec.x)
```

Transformation de la série :

On utilise la commande **ndiffs** pour avoir l'ordre de différenciation qui élimine la tendance et la stationnaire.

```
library(forecast)
ndiffs(x)
data1=diff(x,df=5)
```

La représentation graphique de la série différencier est : `plot.ts(data1,col='red')`

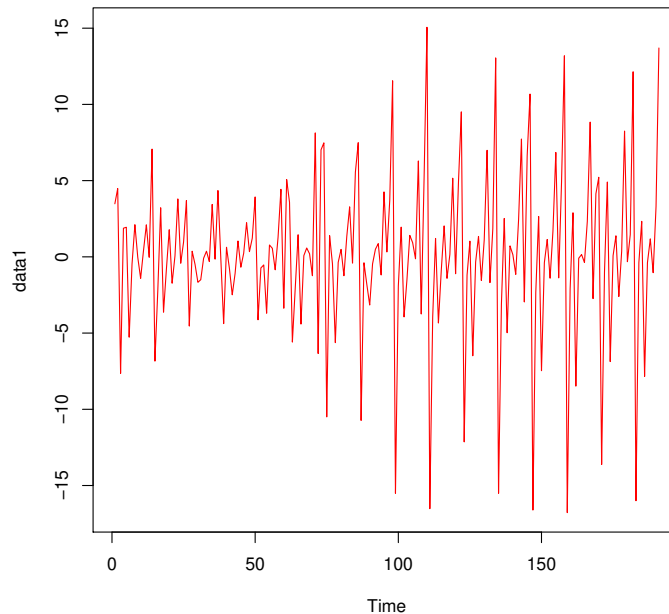


FIGURE 3.36 – La série différencier de la consommation du gaz(clients administrations)

On applique l'ensemble des test de la stationnarité on obtient :

```
library(tseries)
kpss.test(data1)
KPSS Test for Level Stationarity
data: data1
KPSS Level = 0.013051, Truncation lag paramètre = 4, p-value = 0.1
Phillips-Perron Unit Root Test
pp.test(data1)
data: data1
Dickey-Fuller Z(alpha) = -174.68, Truncation lag paramètre = 4, p-
value = 0.01
alternative hypothèses: stationary
```

```
adf.test(data1)
```

Augmentes Dickey-Fuller Test

```
data: data1
```

```
Dickey-Fuller = -11.214, Lag order = 5, p-value = 0.01
```

```
alternative hypothèses: stationary
```

D'après les résultats de trois tests on résume que la série différencier est stationnaire.

Identification :

Les graphes d'auto corrélation simple et partiel de la série diff sont :

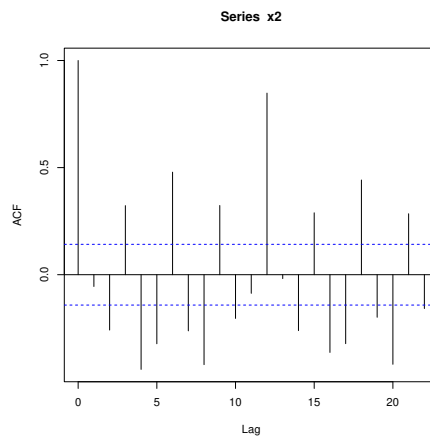


FIGURE 3.37 – Corrélogramme simple de la série différencier de la consommation du gaz (clients administrations)

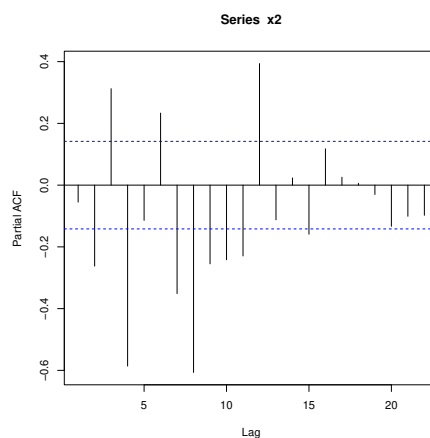


FIGURE 3.38 – Corrélogramme partiel de la série différencier de la consommation du gaz (clients administrations)

Estimation des paramètres

On utilise la commande **auto.arima** pour construire le meilleur modèle qui représente la série .

```
library(forecast)
```

```
Modèle=auto.arima(x,trace="TRUE",test="kpss",ic="aic")
```

Qui donne que le meilleur modèle est (2,1,5) car il minimise le critère AIC .

Aussi on a la commande **summary(Modèle)** qui donne les coefficients du modèle ARIMA(2,1,2)

TABLE 3.19 – Les coefficients du modèle ARIMA(2,1,5) pour la consommation du gaz (clients administrations)

	ar1	ar2	ma1	ma2	ma3	ma4	ma5
Coefficients	-0.6408	-0.0535	0.2496	-0.4535	0.3076	-0.3615	-0.6684
s.e.	0.1335	0.1080	0.1139	0.0799	0.1067	0.0573	0.0828

TABLE 3.20 – Les paramètres d'ajustement du modèle ARIMA(2,1,5) pour la consommation du gaz (clients administrations)

σ^2	15.39
log likelihood	-530.77
AIC	1077.54
AICc	1078.33
BIC	1103.55

Training set erreur mesures : TSEM

Training set : TS

TABLE 3.21 – Les paramètre des erreurs d'ajustement du modèle ARIMA(2,1,5) pour la consommation du gaz (clients administrations)

TSEM	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
TS	0.31324	3.840927	2.833234	-393.2523	442.9714	0.79118	-0.00951

Validation :

Nous avons diagnostiqué notre modèle à partir des tests suivants pour montrer que le modèle choisis est valide

```
adf.test(Modèle$residuals)
```

Augmentes Dickey-Fuller Test

```
data: Modèle$residuals
```

```
Dickey-Fuller = -8.1881, Lag order = 5, p-value = 0.01
```

```
alternative hypothèses: stationary
```

```
Box.test(Modèle$residuals)
```

Box-Pierce test

```
X-squares = 0.017355, df = 1, p-value = 0.895
```

Tester de la stationnarité graphiquement par la commande :

```
qqnorm(data1)
```

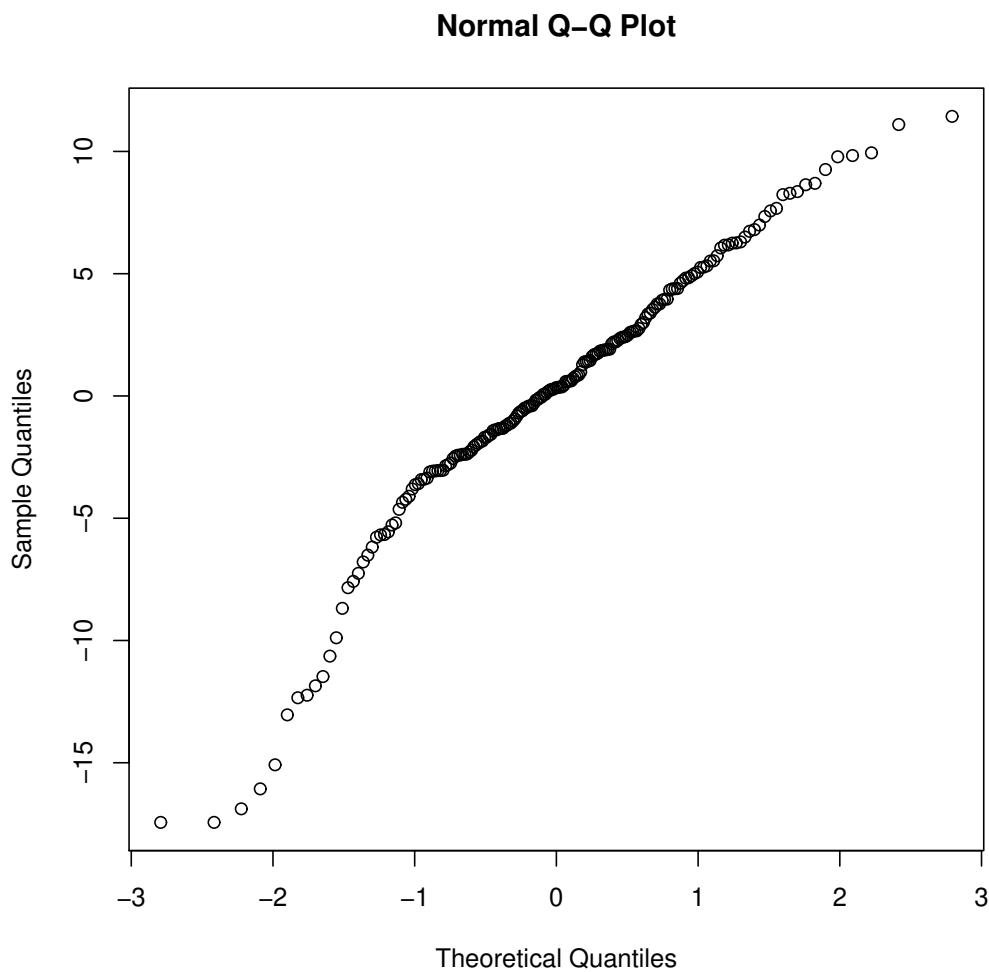


FIGURE 3.39 – Le graphe qqnorm du modèle ARIMA(2,1,5) pour la consommation du gaz (clients administrations)

TABLE 3.22 – Les différents modèles ajustés pour la consommation du gaz (clients administrations)

Modèles	la valeur de AIC
ARIMA(2,1,2) with drift	1124.773
ARIMA(0,1,0) with drift	1181.091
ARIMA(1,1,0) with drift	1170.963
ARIMA(0,1,1) with drift	1163.882
ARIMA(0,1,0)	1179.135
ARIMA(1,1,2) with drift	Inf
ARIMA(2,1,1) with drift	1129.336
ARIMA(3,1,2) with drift	Inf
ARIMA(2,1,3) with drift	Inf
ARIMA(1,1,1) with drift	1134.418
ARIMA(1,1,3) with drift	1139.631
ARIMA(3,1,1) with drift	1126.825
ARIMA(3,1,3) with drift	1080.3
ARIMA(4,1,3) with drift	Inf
ARIMA(3,1,4) with drift	Inf
ARIMA(2,1,4) with drift	1080.092
ARIMA(1,1,4) with drift	Inf
ARIMA(2,1,5) with drift	1078.746
ARIMA(1,1,5) with drift	Inf
ARIMA(3,1,5) with drift	Inf
ARIMA(2,1,5) with drift	1076.84
ARIMA(1,1,5) with drift	Inf
ARIMA(2,1,4) with drift	1078.145
ARIMA(3,1,5) with drift	Inf
ARIMA(1,1,4) with drift	Inf
ARIMA(3,1,4) with drift	Inf

Meilleur modèle est : ARIMA(2,1,5) avec $AIC = 1076.84$.

Conclusion : Le modèle ARIMA(2,1,5) est valable.

Prévision :

D'après le modèle choisis précédent on peut prédire des résultats pour l'année prochaine .

On utilise l'ensemble des commandes suivantes :

```
library(forecast)
```

```
m_aa=auto.arima(data)
```

```
f_aa=forecast(m_aa,h=12)
```

```
f_aa
```

on aura :

La représentation graphique des valeurs prédictives est :
`plot(f_aa,col='red')`.

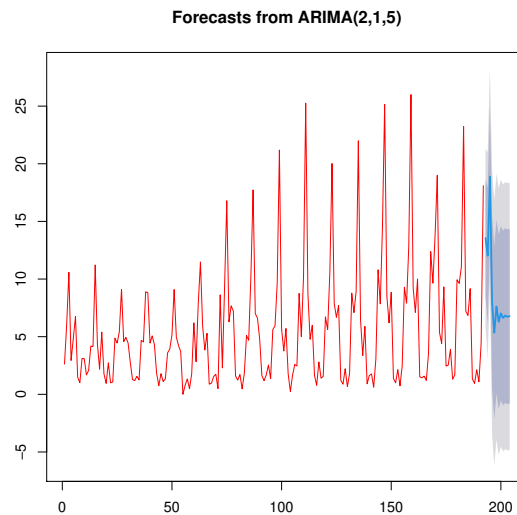


FIGURE 3.40 – Représentation graphique des valeurs prédictives pour la consommation du gaz (clients administrations)

TABLE 3.23 – les valeurs de prédiction par le modèle ARIMA(2,1,5) pour la consommation du gaz (clients administrations)

Point	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
193	13.564416	8.53603188	18.59280	5.874166	21.25467
194	12.020615	6.13331918	17.90791	3.016773	21.02446
195	18.873579	12.72513826	25.02202	9.470350	28.27681
196	7.967199	0.49218215	15.44222	-3.464853	19.39925
197	5.346629	-2.18413801	12.87740	-6.170685	16.86394
198	7.609662	0.06604218	15.15328	-3.927309	19.14663
199	6.299885	-1.26805210	13.86782	-5.274276	17.87405
200	7.017992	-0.54992155	14.58590	-4.556133	18.59212
201	6.627973	-0.94734513	14.20329	-4.957476	18.21342
202	6.839439	-0.73749382	14.41637	-4.748480	18.42736
203	6.724819	-0.85634649	14.30598	-4.869573	18.31921
204	6.786943	-0.79688274	14.37077	-4.811517	18.38540

3.3.2.3 La consommation totale du gaz

Analyse graphique de la série " consommation totale du gaz " :

A partir des données de la série, et avec l'ensemble des commandes suivantes :

```
ts3=ts(x, fréquence=12)}  
plot(ts3)
```

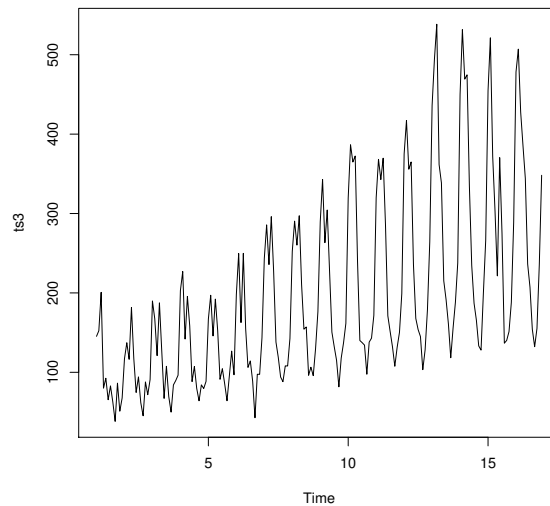


FIGURE 3.41 – La consommation totale du gaz

D'après la figure 3.41 on remarque que la tendance de la consommation totale du gaz augmente.

Le corrélogramme simple :

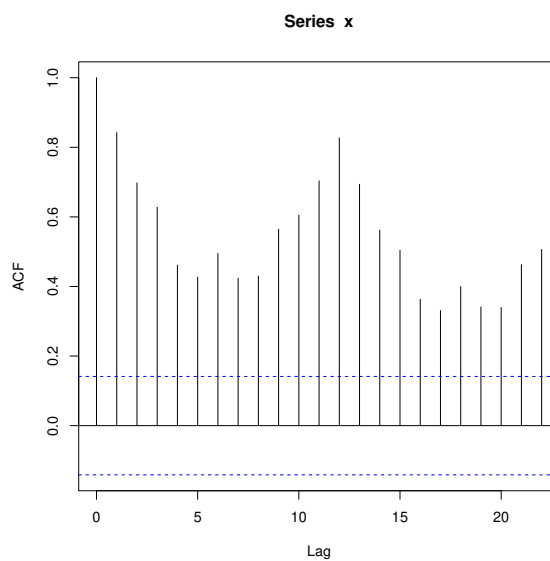


FIGURE 3.42 – Corrélogramme simple de la consommation totale du gaz

Le corrélogramme partiel :

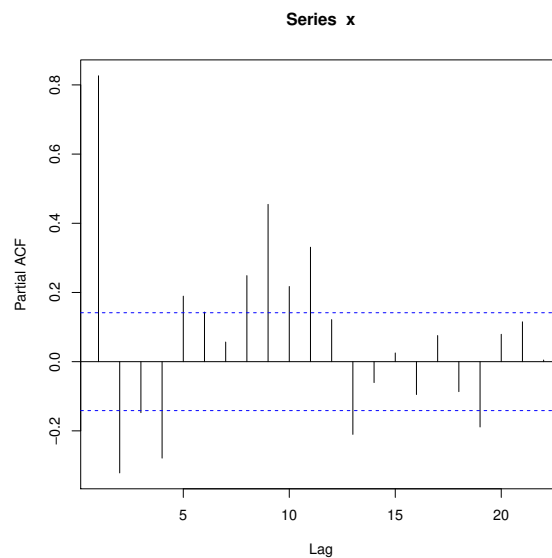


FIGURE 3.43 – Le corrélogramme partiel de la consommation totale du gaz

La visualisation des corrélogramme simple et partiel montre une non stationnarité propre, l'ACF ne coupe pas.

On utilise aussi le test de racine unitaire (**test ADF**) pour confirmer la non stationnarité :

```
adf.test(x)
```

Augmented Dickey-Fuller Test

Dickey-Fuller = -10.503, Lag order = 5, p-value = 0.1

p-value=0.1 > 0.05 donc la série est n'est pas stationnaire .
(Admet une racine unitaire).

On peut aussi utiliser la fonction **décompose** pour extraire les composantes d'une série temporelle.

Transformation de la série

On utilise la commande **ndiffs** pour savoir l'ordre de différenciation qui éliminé la tendance et la saisonnalité.

```
library(forecast)
```

```
ndiffs(x)
```

```
data1=diff(x,df=5)
```

La représentation graphique est :

`plot.ts(data1,col='red')`

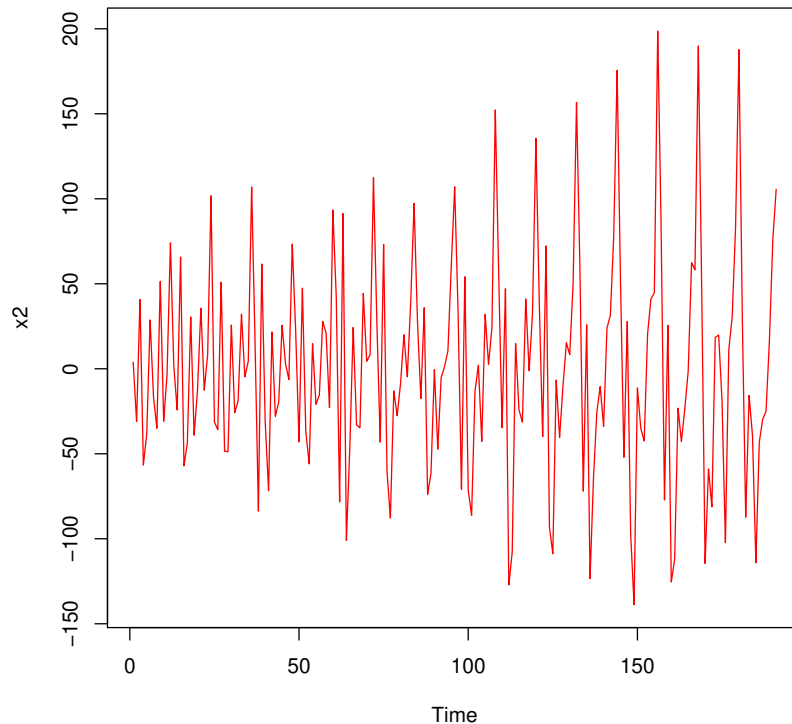


FIGURE 3.44 – La série différencier de la consommation totale du gaz

On applique l'ensemble des test de la stationnarité on obtient :

```
library(tseries)
kpss.test(data1)
KPSS Test for Level Stationarity
data: data1
KPSS Level = 0.013106, Truncation lag paramètre = 4, p-value = 0.1
pp.test(data1)
Phillips-Perron Unit Root Test
Dickey-Fuller Z(alpha) = -269.46, Truncation lag paramètre = 4, p-
value = 0.01
alternative hypothèses: stationary
adf.test(data1)
```

Augmented Dickey-Fuller Test

Dickey-Fuller = -9.2067, Lag order = 5, p-value = 0.01

alternative hypothèses: stationary

D'après les résultats des trois tests on résume que la série différencier est stationnaire. **Identification :**

Les graphes d'auto corrélation simple et partiel de la série diff sont :

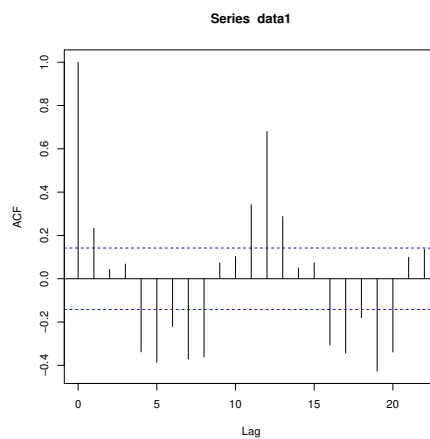


FIGURE 3.45 – Corrélogramme simple de la série différencier de la consommation totale du gaz

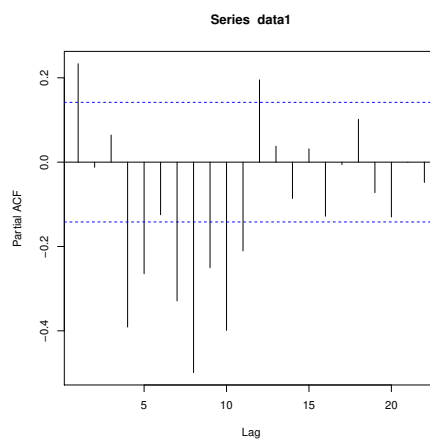


FIGURE 3.46 – Corrélogramme partiel de la série différencier de la consommation totale du gaz

Estimation des paramètres :

On utilise la commande (**auto.arima**) pour construire le meilleur modèle qui représente la série.

library(forecast)

Modèle=auto.arima(x, trace=TRUE, test="kpss", ic="aic")

Qui donne que le meilleur modèle est ARIMA(0,1,1) car il est minimisé le critère AIC.

Aussi on a la commande **summary(Modèle)** qui donne les coefficients du modèle **ARIMA(0,1,1)**

TABLE 3.24 – Coefficients du modèle ARIMA(0,1,1)

	ma1
Coefficient	0.2558
se	0.0775

TABLE 3.25 – Paramètre d'ajustement du modèle ARIMA(0,1,1)

σ^2	4282
log likelihood	-1069.15
AIC	2142.29
AICc	2142.35
BIC	2148.8

Training set erreur mesures : TSEM

Training set : TS

TABLE 3.26 – paramètre des erreurs d'ajustement du modèle ARIMA(0,1,1)

TSEM	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
TS	0.94005	65.09792	49.44135	-4.153499	28.83867	0.94742	-0.00809

Validation :

Nous avons diagnostiqué notre modèle à partir des tests suivants pour montrer que le modèle choisis est valide.

Test ADF `adf.test(Modèle$residuals)`

Augmentes Dickey-Fuller Test

data: `Modèle$residuals`

Dickey-Fuller = -9.0318, Lag order = 5, p-value = 0.01

alternative hypothèses: stationary

Box-Pierce test

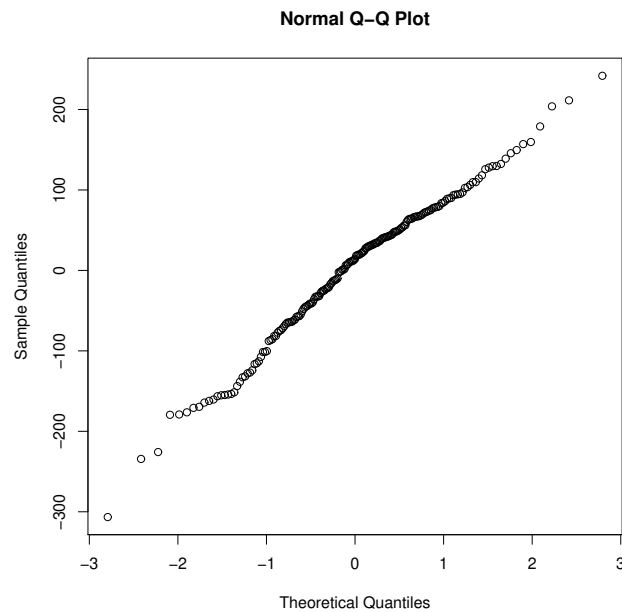


FIGURE 3.47 – Le graphe qqnorm des résidus du modèle ARIMA(0,1,1) pour la consommation totale du gaz

data: Modèle\$residuals

X-squares = 0.012557, df = 1, p-value = 0.9108

Tester de la stationnarité graphiquement avec la commande :**qqnorm(data1)**

TABLE 3.27 – les valeurs de la prédiction par le modèle ARIMA(0,1,1) pour la consommation totale du gaz

Modèle	AIC
ARIMA(2,1,2) with drift	2136.404
ARIMA(0,1,0) with drift	2145.011
ARIMA(1,1,0) with drift	2137.149
ARIMA(0,1,1) with drift	2135.771
ARIMA(0,1,0)	2143.059
ARIMA(1,1,1) with drift	2136.807
ARIMA(0,1,2) with drift	2137.427
ARIMA(1,1,2) with drift	2137.918
ARIMA(0,1,1)	2133.81
ARIMA(1,1,1)	2134.934
ARIMA(0,1,2)	2135.463
ARIMA(1,1,0)	2135.186
ARIMA(1,1,2)	2136.168

Meilleur modèle est : ARIMA(0,1,1) avec $AIC = 2142.291$.

Conclusion : le modèle ARIMA(0,1,1) est valable.

Prévision :

D'après le modèle choisis précédent on peut prédire des résultats pour l'année prochaine.

On utilise l'ensemble des commandes suivantes :

```
m_aa=auto.arima(x)
f_aa=forecast(m_aa,h=12)
f_aa
```

La représentation graphiques des valeurs prédictives est : `plot(f_aa,col='red')`

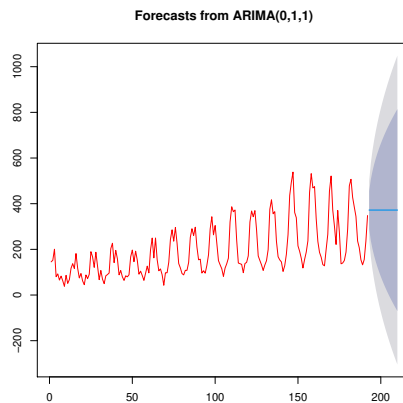


FIGURE 3.48 – Représentation graphique des valeurs prédictives de la consommation totale du gaz

TABLE 3.28 – les valeurs de prédiction par le modèle ARIMA(0,1,1) pour la consommation totale du gaz

Point	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
193	372.012	288.14775	455.8763	243.75268	500.2714
194	372.012	237.38226	506.6418	166.11356	577.9105
195	372.012	201.08183	542.9422	110.59684	633.4272
196	372.012	171.24079	572.7833	64.95891	679.0652
197	372.012	145.29403	598.7300	25.27677	718.7473
198	372.012	122.02	621.9981	-10.30859	754.3327
199	372.012	100.74654	643.2775	-42.85274	786.8768
200	372.012	81.01903	663.0050	-73.02336	817.0474
201	372.012	62.54655	681.4775	-101.27458	845.2986
202	372.012	45.11626	698.9078	-127.93190	871.9560
203	372.012	28.56946	715.4546	-153.23806	897.2621
204	372.012	12.78403	731.2400	-177.37979	921.4039

Comparaison des valeurs prédictives avec des valeurs réelles

Voici les tableaux comparatifs entre les prévisions et les résultats réels des 4 premiers mois de 2022 :

La consommation d'électricité

La consommation d'électricité pour les clients ordinaires

TABLE 3.29 – Les valeurs prédictives et réelles de la consommation d'électricité (clients ordinaires)

	valeur prédictive	valeur réelle	erreur
janvier	80.95	82,47	1.52
février	80.95	86,33	5.38
mars	80.95	76,87	4.08
avril	80.95	79,81	1.14

La consommation d'électricité pour les clients administrations

TABLE 3.30 – Les valeurs prédictives et réelles de la consommation d'électricité (clients administrations)

	valeur prédictive	valeur réelle	erreur
janvier	11.680601	9,22	2.460601
février	10.832379	8,58	2.252379
mars	15.093154	14.53	0.563154
avril	8.376527	9.55	1.173473

La consommation totale d'électricité :

TABLE 3.31 – Les valeurs prédictives et réelles de la consommation totale d'électricité

	valeur prédictive	valeur réelle	erreur
janvier	165.75	153,01	12.74
février	165.75	154,38	11.37
mars	165.75	91,40	74.35
avril	165.75	148,70	17.05

La consommation du gaz

La consommation du gaz pour les clients ordinaires

TABLE 3.32 – Les valeurs prédictives et réelles de la consommation du gaz (clients ordinaires)

	valeur prédictive	valeur réelle	erreur
janvier	281.8282	486.57	204.7418
février	248.7264	532.83	284.1036
mars	237.2166	458.25	221.0334
avril	227.4207	466.88	239.4593

La consommation du gaz pour les clients administrations

TABLE 3.33 – Les valeurs prédictives et réelles de la consommation du gaz (clients administrations)

	valeur prédictive	valeur réelle	erreur
janvier	13.564416	10.92	2.644416
février	12.020615	14.02	1.999385
mars	18.873579	29.91	11.036421
avril	7.967199	9.84	1.872801

La consommation totale du gaz

TABLE 3.34 – Les valeurs prédictives et réelles de la consommation totale du gaz

	valeur prédictive	valeur réelle	erreur
janvier	372.012	543,79	171.778
février	372.012	615,75	243.738
mars	372.012	551,16	179.148
avril	372.012	525,85	153.383

3.4 Conclusion

L'analyse et la comparaison des valeurs réelles et les valeurs de prévision par les modèles ARIMA, nous constatons que quelques valeurs sont proches et d'autres valeurs sont loin, nous pouvons conclure qu'il y'a des doutes sur les valeurs enregistrés par la société de Sonelgaz, parfois les enregistrements ne sont pas logique selon la période de la consommation, pour exemple, on trouve la consommation d'électricité en hiver est grande par rapport à la consommation en été. La justification du staffe de Sonelgaz basé sur des explications commerciales, donc la qualité des données influencent sur la crédibilité du modèle de prédiction.

CONCLUSION GÉNÉRALE

Les problèmes liés à la consommation et la production d'électricité et du gaz sont des sujets qui actuellement attire l'attention des producteurs de cette énergie et leurs bureaux de développement, c'est d'ailleurs la cause qui nous a orienté vers le choix de ce sujet.

Pour traiter les problèmes de la consommation et de la demande d'électricité et du gaz par les clients tels que la consommation quotidienne et celle de l'industrie, ainsi que pour prévoir la demande afin d'adapter la production, Sonelgaz entreprise spécialisée dans la production d'électricité et du gaz nous a proposé d'étudier la consommation afin de prévoir la future demande pour régler ce problème

La prévision de la consommation de cette énergie est un processus qui peut se faire à l'aide des méthodes mathématiques telles que l'étude des séries chronologiques. Pour réaliser ce travail, on a étudié la consommation avec le modèle ARIMA pour atteindre le niveau de précision souhaité.

Durant la période de stage de fin d'étude au sein de l'entreprise Sonelgaz, nous avons eu l'occasion de découvrir le milieu de recherche industriel, notamment celui de l'étude et de la production d'énergie électrique. Nous avons eu la possibilité d'être proche et d'être en relation avec notre domaine d'application et de découvrir les différents outils informatiques du travail.

Cette expérience nous a permis non seulement de mettre en pratique ce qu'on a acquis comme connaissances théoriques durant notre cursus universitaire mais aussi nous nous sommes familiarisés avec le milieu du travail et nous avons pu avoir un aperçu de la réalité du terrain et des problèmes imprévus auxquels il faut trouver des solutions.

L'objectif principal de notre projet était de réaliser la prévision de la consommation de l'énergie électrique et gazière par la méthode de Box-Jenkins et le modèle ARIMA. Premièrement, nous avons introduits dans le monde des séries chronologiques et de ses généralités, et on a consacré une partie pour les modèles de prévision mathématiques et leurs caractéristiques. Ensuite, nous avons présenté l'entreprise Sonelgaz, ainsi que d'application de la consommation d'électricité et du gaz au niveau de la wilaya de Blida avec la méthodologie de Box & Jenkins .

Enfin, nous espérons que notre travail servira de base pour des étudiants qui choisiront de travailler sous la technologie de prévision à l'aide des séries chronologiques.

- [1] Christian Gouriéroux and Alain Monfort. *Séries temporelles et modèles dynamiques*. Economica, 1995.
- [2] A Müller. *Comportement asymptotique de la distribution des pluies extrêmes en France*. PhD thesis, Thèse de doctorat, Université Montpellier II, 2006.
- [3] Cours Serge Dégerine. Cours de séries chronologiques. *Université Joseph Fourier Grenoble*, 1, 2005.
- [4] Yan Girard. *Séries chronologiques à une et plusieurs variables : synthèse des méthodes classiques et modèles à base de copules*. PhD thesis, Université du Québec à Trois-Rivières, 2011.
- [5] Xavier De Luna. Une approche non-paramétrique et parcimonieuse de la prévision de séries temporelles. Technical report, EPFL, 1996.
- [6] Frédéric Gugelot. *Archives de sciences sociales des religions*, (140) :119–131, 2007.
- [7] RODOLPHE RAMIHARIJAFY, JEAN MARC RABEHARISOA, ADOLPHE RATIARISON, and Jérôme VELO. Modélisation et études prévisionnelles de pluies : cas du site hydroélectrique d’andekaleka à madagascar. *Afrique Science*, 13(2) :233–249, 2017.
- [8] Samah DENDOUGA. *Séries temporelles : Théorie et application*. 2020.

- [9] Cédric Hugrée. Les sciences sociales face à la mobilité sociale. *Politix*, (2) :47–72, 2016.
- [10] Malika Chekroune. *Etude comparative de deux techniques de séchage (convection et micro-onde) par application des plans d'expérience*. PhD thesis, Boumerdes, Université M'hamed Bougara. Faculté des sciences de l'ingénieur, 2009.
- [11] KHERCHI Ilyas. *MODELISATION DES SERIES TEMPORELLES PAR LES PROCESSUS AUTOREGRESSIFS*. PhD thesis, UNIVERSITE DE MOHAMED BOUDIAF M'SILA FACULTE DE TECHNOLOGIE, 2016.
- [12] Aragon Yves. *Séries temporelles avec r. méthodes et cas*, 2011.
- [13] Application à la destination du Vietnam. Prévission de la demande touristique par méthodes delphi et box-jenkins. *Téoros. Revue de recherche en tourisme*, 29(1) :100–108, 2010.
- [14] Wayne A Woodward and Henry L Gray. On the relationship between the s array and the box-jenkins method of arma model identification. *Journal of the American Statistical Association*, 76(375) :579–587, 1981.