

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université M'hamed Bougara Boumerdes
Faculté des Sciences
Département de Mathématiques



Mémoire Présenté
Pour L'Obtention Du Diplôme De Master
En Recherche Opérationnelle
Option : Recherche Opérationnelle et mathématique de la gestion
(ROMAGE)

Par : Hammoudi meriem
Et : Kouihal fatima

**Prévision de la production d'électricité de type
Energies Renouvelables
Cas : éolienne et photovoltaïque**

soutenu à UMBB, le 19/06/2016, devant le jurys composé de :

| | |
|-----------------------------------|------------|
| M ^{me} Benmensour Madina | Encadreur. |
| M ^{me} Zouaoui Selma | Président |
| M ^r Cheurfa Fateh | Examineur. |
| M ^{me} Berrah Yasmine | Invitée . |

Année Universitaire 2015 – 2016

Remerciement

Au terme de ce travail, nous voudrions exprimer nos profonde gratitude envers Dieu le tout puissant qui, grâce à son aide, nous avons pu finir ce travail.

Nous tenons à exprimer nos vifs remerciements et profonde gratitude à notre encadreur madame Benmenseur Madina pour l'aide et le soutien qu'elle nous a apportés, et pour ses précieux conseils qu'elle nous a fournis durant la réalisation de ce mémoire.

Que madame Berrah Yasmine, notre encadreuse au sein de la compagnie de SONEL-GAZ, trouve ici l'expression de notre reconnaissance pour son aide, sa disponibilité, et ses encouragement.

Nous tenons à remercier sincèrement madame Zouaoui Selma qui nous nous fait le grand honneur de présider notre jury de soutenance.

Nous remercions également Monsieur Fateh Cheurfa d'avoir accepter d'examiner notre travail et aussi à son aide.

En fin, nous tenons a rendre un grand hommage tous nous enseignants pour la qualité des enseignements qu'ils nous ont dispensés.

Dédicace

Je dédie ce mémoire

A mes parents pour leur amour inestimable, leur confiance, leur soutien, leurs sacrifices
et toutes les valeurs qu'ils ont su m'inculquer.

Qu'ils trouvent en moi la source de leur fierté

A qui je dois tout

A ma sœur mon âme ainsi qu'à mes deux frères pour leur tendresse, leur complicité et
leur présence dans ma vie.

A qui je souhaite un avenir radieux plein de réussit

A toute ma famille ainsi qu'à mes amis

MERIEM .

Dédicace

Je dédie ce mémoire avec plaisir

A mon père et ma mère en signe de reconnaissance pour leur soutien tout au long de mes études et leurs innombrables sacrifices en leur souhaitant une longue vie.

A mes frères et mes sœurs et tous les membres de ma famille pour qui je souhaite tout le bonheur.

A tous mes amis qui ont toujours fait preuve d'un esprit de collaboration et de serviabilité

Fatima.

Table des matières

| | |
|--|-----------|
| Dédicace | i |
| Dédicace | i |
| <i>Introduction générale</i> | 2 |
| 1 Présentation de l'entreprise | 4 |
| Introduction | 4 |
| 1.1 L'énergie électrique et ses usages | 4 |
| 1.1.1 Définition et historique de l'énergie | 4 |
| 1.1.2 Définition de l'énergie électrique (Qu'est-ce que l'électricité?) | 5 |
| 1.1.3 Les sources énergétiques | 5 |
| 1.2 le groupe Sonelgaz | 7 |
| 1.2.1 Présentation du groupe " Sonelgaz " | 7 |
| 1.2.2 Présentation L'Opérateur du Système Electrique | 8 |
| 1.3 Production de l'électricité en Algérie | 11 |
| 1.3.1 Représentation des centrales de production par type | 11 |
| 1.4 Problématique | 15 |
| 2 Aspect théorique des séries temporelles | 16 |
| Introduction | 16 |
| 2.1 Généralité sur les séries temporelles | 16 |
| 2.1.1 Définitions de séries temporelle | 16 |
| 2.1.2 Objectifs principaux | 17 |
| 2.1.3 Caractéristiques d'une série chronologique | 17 |
| 2.1.4 Les étapes de traitement d'une série chronologique | 18 |
| 2.1.5 Les composantes d'une série temporelle : | 19 |
| 2.1.6 Modèle de décomposition : | 20 |
| 2.2 Analyse d'une série chronologique | 22 |
| 2.2.1 Les opérateurs linéaires de retard, d'avance et de différenciation | 22 |
| 2.2.2 Les effets des opérateurs | 23 |
| 2.2.3 Processus aléatoires stationnaire : | 23 |
| 2.2.4 Processus aléatoires non stationnaire : | 24 |
| 2.2.5 Modélisation des séries temporelle | 26 |
| 2.2.6 Les modèles ARMAX | 28 |
| 2.2.7 Étude de la stationarité d'une serie temporelle : | 28 |
| Conclusion | 30 |

| | | |
|----------|---|-----------|
| 3 | Méthodes de prévisions | 31 |
| | Introduction | 31 |
| 3.1 | Qu'esque qu'une prévision? | 31 |
| 3.2 | Ce qui caractérise un problème de prévision | 31 |
| 3.3 | Les méthodes de prévisions | 32 |
| 3.4 | La prévision à court terme | 32 |
| 3.5 | Critères de validité de la méthode de prévision | 32 |
| 3.6 | Prévision des énergies renouvelables | 33 |
| 3.7 | Régression linéaire par la méthode des moindres carrés | 33 |
| 3.7.1 | Principe de la régression linéaire | 33 |
| 3.7.2 | Significativité de la régression | 36 |
| 3.7.3 | Comparaison de modèles et critères d'information | 38 |
| 3.7.4 | Intervalle de confiance (<i>IC</i>) | 39 |
| 3.7.5 | Prédiction | 39 |
| 3.8 | Régression non linéaire | 40 |
| 3.9 | Prévision par le modèle <i>ARMAX</i> | 40 |
| 3.9.1 | Identification des résidus par la méthode des moindres carrés ordi- naires MCO | 41 |
| 3.9.2 | Estimation du modèle <i>ARMAX</i> | 42 |
| 3.9.3 | Validation du modèle <i>ARMAX</i> | 43 |
| 3.9.4 | Prévision du modèle <i>ARMAX</i> | 43 |
| | Conclusion | 43 |
| 4 | Application | 44 |
| | Introduction | 44 |
| 4.1 | Etude prévisionnelle de la centrale éolienne de kaberten | 44 |
| 4.1.1 | Régression linéaire par la méthode des moindres carrés | 46 |
| 4.1.2 | Prévision par le modèle <i>ARMAX</i> | 54 |
| 4.1.3 | Comparaison et interprétation des résultats | 61 |
| 4.2 | Etude prévisionnelle de la centrale photovoltaïque de Adrar | 62 |
| 4.2.1 | Régression linéaire par la méthode des moindres carrés | 63 |
| 4.2.2 | Prévision par le modèle <i>ARMAX</i> | 68 |
| 4.2.3 | Comparaison et interprétation des résultats | 74 |
| | Conclusion | 74 |
| | Conclusion générale | 77 |
| | Bibliographie. | 78 |

Introduction générale

De nos jours, la production de l'énergie en général est considérée comme un indice représentatif du développement économique d'un pays, et plus particulièrement l'énergie électrique. Celle-ci est à la base de toutes activités industrielles, elle constitue par conséquent un facteur très important à l'amélioration des conditions de vie des citoyens, à l'accession et à la modernité.

La croissance rapide de la consommation de cette énergie est due essentiellement au développement industriel, à la croissance démographique ainsi qu'à l'urbanisation de plus en plus des villes et des campagnes et l'utilisation des biens électroménagers, ce qui nécessite un réseau performant capable de répondre en permanence à cette demande sans cesse grandissante.

Face à cette demande croissante d'énergie et à l'épuisement à plus ou moins long terme des énergies fossiles ainsi que les problèmes rencontrés en termes de pollution des déchets radioactifs, du démantèlement prochain des vieilles centrales et du risque industriel, différentes solutions de substitution ont été envisagées. Certains pays ont mené une politique orientée vers le nucléaire alors que d'autres ont massivement utilisé des énergies renouvelables.

L'Algérie comme tout autre pays s'engage avec détermination sur la voie des énergies renouvelables afin d'apporter des solutions globales et durables aux défis environnementaux et aux problématiques de préservation des ressources énergétiques d'origine fossile.

Ce choix stratégique est motivé par l'immense potentiel en énergie solaire. Cette énergie constitue l'axe majeur du programme qui consacre au solaire thermique et au solaire photovoltaïque une part essentielle. Le solaire devrait atteindre d'ici 2030 plus de 37% de la production nationale d'électricité.

Malgré un potentiel assez faible, le programme n'exclut pas l'éolienne qui constitue le second axe de développement et dont la part devrait avoisiner les 3% de la production d'électricité en 2030.

L'Algérie prévoit également l'installation de quelques unités de taille expérimentale afin de tester les différentes technologies en matière de biomasse, de géothermie et de dessalement des eaux saumâtres par les différentes filières d'énergie renouvelable.

Ainsi le secteur de l'énergie électrique et gazier représenté par la Société Nationale de l'Electricité et du Gaz Sonelgaz est considéré comme l'un des piliers de l'économie. Il a

joué dans le passé un rôle primordial dans l'industrialisation du pays et il continue encore d'assurer les besoins nationaux dans les meilleures conditions de qualité de service, de sécurité et de coût.

Comme L'énergie électrique est un produit non stockable, sa production, son transport et son utilisation nécessitent d'énormes investissements et aussi une planification et une gestion rigoureuse ainsi que des prévisions sur la consommation et la production d'électricité à partir d'énergie fossile ou d'énergie renouvelable. Et comme la production d'électricité dépend du changement climatique (instabilité) le besoin de prévoir devient une nécessité pour la Sonelgaz.

On s'intéressera dans ce modeste travail à la prévision de la production d'énergie renouvelable (photovoltaïque et éolienne). Ce qui nous mène a utilisé l'approche de séries temporelles, le modèle ARMAX et la régression linéaire des moindres carrés comme méthodes de prévision.

Pour cela on prend à titre d'exemple la série de la production de l'énergie renouvelable photovoltaïque d'Adrar ainsi que la série de la production de l'énergie éolienne de Kaberten pendant Une semaine de 08/04/2016 au 14/04/2016 pour la centrale éolienne de Kaberten.

Pour mener à bien cette étude et afin d'atteindre nos objectifs, nous allons dans un premier temps consacré le premier chapitre a la présentation de l'organisme d'accueil et à la définition des notions de base sur les types des énergies renouvelables entamé.

Le deuxième chapitre seras un chapitre introductif a la prévision en sens générale et les déffirentes approches utilisées dans le domaine de prévision des énergies renouvelables.

Le troisième chapitre, de ce mémoire fournit au lecteur une connaissance suffisante sur l'aspect théorique des séries temporelles .

Dans le quatrième chapitre, nous exposons toutes les notions de la régression par la Méthode des moindres carrés, dans le but de modélisation et de prévision on vue d'en faire une application sur une série de production pour la centrale éolienne dans le dernier chapitre.

Nous terminons notre travail par une conclusion générale.

1

Présentation de l'entreprise

Introduction

Le système économique Algérien est basé principalement sur le secteur énergétique qui procure 96% des recettes extérieures. Ce secteur joue un rôle très important dans la vie économique du pays par le financement du développement économique et la satisfaction des besoins du secteur productif et des ménages en matière d'énergie notamment l'électricité et le gaz. Notre chapitre est composé de trois sections la première comporte des définitions de l'énergie électrique et ses sources, la deuxième fait la présentation de groupe Sonelgaz , sa filiale Operateur système ainsi que l'organisme d'accueil et la troisième éclaire la production d'électricité en Algérie.

1.1 L'énergie électrique et ses usages

1.1.1 Définition et historique de l'énergie

Définition de l'énergie :

Le mot énergie est à la fois d'origine latine " énergia " et grecque, et qui veut dire " puissance qui permet d'agir et de réagir ". L'énergie est définie comme la force utilisée à partir des ressources naturelles (charbon, gaz, pétrole) et des industries spécialisées (électricité, nucléaire, raffinage pétrolier...etc.) pour faire fonctionner les machines utilisées aux individus et à l'industrie.

Et d'une autre manière l'énergie caractérise la capacité à produire des actions, par exemple à engendrer du mouvement, modifier la température d'un corps ou à transformer la matière. L'énergie provient de différentes sources que l'on trouve dans la nature : le bois, le charbon, le pétrole, le gaz, le vent, le rayonnement solaire, les chutes d'eau, la

chaleur interne de la terre, l'uranium. Elle peut prendre différentes formes : chaleur, énergie musculaire, énergie mécanique, énergie chimique, énergie électrique, Ses formes multiples peuvent se transformer l'une en l'autre.

L'énergie primaire : C'est la première forme de l'énergie directement disponible dans la nature : bois, charbon, gaz naturel, pétrole, vent, rayonnement solaire, eau, chaleur du sous-sol...etc.

L'énergie secondaire : L'énergie secondaire est une énergie obtenue par la transformation d'une énergie primaire au moyen d'un système de conversion. Par exemple, une centrale thermique produit de l'électricité (énergie secondaire) à partir de charbon (énergie primaire), une énergie secondaire peut aussi résulter de la transformation d'une autre énergie secondaire, c'est le cas d'une centrale thermique alimentée en gaz de haut fourneau.

L'électricité est donc une énergie secondaire qui peut être produite de différente manière.

1.1.2 Définition de l'énergie électrique (Qu'est-ce que l'électricité?)

L'électricité est ce qui fait marcher les multitudes des machines que nous utilisons dans notre vie quotidienne pour faire le travail pénible à notre place, c'est une source d'énergie qui sert à des usages domestiques et industriels.

L'énergie électrique correspond à la charge en électricité que possède un corps. Elle se manifeste lorsqu'on relie par un fil conducteur deux corps dont la charge électrique est différente.

L'électricité n'est pas proprement parler une source d'énergie, mais un vecteur énergétique, c'est-à-dire un mode de transport de l'énergie. Elle est invisible contrairement au pétrole ou au charbon.

1.1.3 Les sources énergétiques

- **Les énergies renouvelables :**

Une énergie renouvelable est une énergie renouvelée ou régénérée naturellement à l'échelle d'une vie humaine. Les énergies renouvelables sont issues de phénomènes naturels, réguliers ou constants, provoqués par les astres.

L'énergie solaire : L'énergie solaire est l'énergie que dispense le soleil par son rayonnement (le rayonnement est un transfert d'énergie sous forme d'ondes ou de particules, qui peut se produire par rayonnement...) directement ou de manière diffuse à travers l'atmosphère.

L'énergie éolienne : L'énergie éolienne est l'énergie du vent et plus spécifiquement l'énergie tirée du vent au moyen d'un dispositif aérogénérateur comme une éolienne ou un moulin à vent. L'énergie éolienne peut être utilisée de deux manières :

- Conservation de l'énergie mécanique.
 - Transformation en énergie électrique.
-

L'énergie hydraulique : L'énergie hydraulique est l'énergie fournie par le mouvement de l'eau, sous toutes ses formes, chutes, cours d'eau, marée. Ce mouvement peut être utilisé directement, par exemple avec un moulin à eau, ou plus couramment être converti, par exemple en énergie électrique dans une centrale hydroélectrique. L'énergie hydraulique est en fait cinétique dans le cas des marées et cours d'eau, et potentielle dans le cas des chutes.

La biomasse : Il s'agit d'énergie solaire stockée sous forme organique grâce à la photosynthèse. Cette énergie est exploitée par combustion. Elle est considérée comme renouvelable si on admet que les quantités brûlées n'excèdent pas les quantités produites. On peut citer notamment le bois et les biocarburants.

La géothermie : La géothermie vise à exploiter la chaleur naturelle du globe terrestre. Les gisements de chaleur sont de deux sortes : les sources géothermale haute énergie, où la vapeur sèche ou humide peut être utilisée pour la production d'électricité et les sources géothermales basse énergie, telles les sources d'eau chaude.

- **Les énergies non renouvelables :**

Une énergie non renouvelable est une source d'énergie qui ne se renouvelle pas assez rapidement pour être considérée comme inépuisable à l'échelle de l'homme, ou même qui ne se renouvelle pas du tout, par opposition aux énergies renouvelables.

Le charbon : Le charbon "la houille" (mot wallon francisé), est une roche carbonée. C'est également une roche combustible (fossile solide provenant de la décomposition d'organismes du carbonifère). Utilisé depuis le *XI^e* siècle, l'extraction de ce combustible dans les mines a rendu possible la révolution industrielle au *XIX^e* siècle. Pendant des décennies, la houille constitua la source principale d'énergie aussi son nom fut décliné pour désigner des énergies naturelles potentiellement utilisables pour les besoins industriels.

Le pétrole : Le pétrole a été découvert à la fin de l'avant dernier siècle aux états-Unis, il s'est généralisé par la suite dans tous les pays industriels. Son véritable problème c'est qu'il est disponible en quantités limitées. Il faut en effet noter que ce vecteur cumule des points forts en ce qui concerne sa transportabilité, son stockage et sa facilité d'utilisation. Tous ces facteurs ont fait de lui le produit le plus échangé internationalement.

Le pétrole est considéré comme énergie indispensable dans le fonctionnement de la machine économique, il est utilisé dans différents domaines comme l'essence et le gasoil dans les véhicules et le kérosène dans le trafic aérien, le fuel dans le trafic maritime et il est utilisé pour plein d'autres usages.

Le gaz : Le gaz est un carburant ou combustible domestique ou industriel, c'est une source d'énergie volatile et invisible recueillie dans des gisements souterrains pouvant être conditionné sous forme liquide.

Le gaz naturel est utilisé comme charge d'alimentation, comme source d'hydrogène et comme carburant dans la production d'électricité.

C'est aussi le carburant idéal pour les fours industriels et les hauts fourneaux à cause de la chaleur constante, de l'absence d'huiles et de cires et de sa très faible production de soufre et de poussière. Comme il brûle proprement, le gaz

utilisé à feu direct peut éliminer le recours à des accessoires tels des échangeurs de chaleur.

Le gaz est aussi l'énergie idéale de la maison il a trois usages principaux : chauffage, eau chaude sanitaire et cuisson.

L'électricité : L'électricité est un phénomène énergétique associé à la mobilité ou au repos d'électrons, d'ions ou d'autres particules chargées positivement ou négativement.

L'électricité est en effet une énergie secondaire, elle résulte de la transformation de la chaleur en énergie mécanique, laquelle est ensuite transformée en énergie électrique.

La chaleur peut provenir soit de la combustion d'un combustible " traditionnel" (charbon, fuel, gaz, bois), soit de la radioactivité d'un combustible nucléaire, soit du soleil.

L'électricité en Algérie est produite essentiellement par cinq types de centrales, les turbines à gaz, les turbines à vapeur, les turbines Diesel, les turbines Hydraulique et les cycles combinés.

Le nucléaire : L'énergie nucléaire (on dit aussi énergie atomique) est dégagée par la désintégration du noyau de certains atomes, des atomes instables. Pour acquérir une meilleure stabilité l'atome instable va se transformer en un autre type d'atome en explosant de l'énergie sous forme de rayonnements : c'est le phénomène de la radioactivité.

Les applications de l'énergie nucléaire concernent pour l'essentiel deux domaines :

- La production d'électricité dans des centrales nucléaires.
- La propulsion (La propulsion est le principe qui permet à un corps de se mouvoir dans son espace environnant). Elle est utilisée dans le domaine naval (principalement pour les flottes militaires, dans les sous-marins et les porte-avions).

Autre application est la production d'isotopes radioactifs utilisés dans l'industrie (radiographie de soudure par exemple) et en médecine (médecine nucléaire et radiothérapie).

1.2 le groupe Sonelgaz

1.2.1 Présentation du groupe " Sonelgaz "

La société nationale de l'électricité et du gaz par convention Sonelgaz a été créé par ordonnance N°69 – 59 du 28 juillet 1969, publiée dans le journal officiel du 1 Août 1969, portant dissolution de l'établissement public électricité et gaz d'Algérie (*E.G.A*) et l'ensemble des biens, droits et obligations de l'ex *E.G.A*. a été transféré à Sonelgaz et lui définit ses missions et ses attributions entre autres.

Il est attribué à Sonelgaz le monopole de la production, du transport, de la distribution de l'énergie électrique, et de l'importation et l'exportation de gaz.

La Sonelgaz est chargée du transport du gaz naturel par canalisation sur l'ensemble du territoire national.

La Sonelgaz est chargée de la distribution du gaz naturel tant à usage domestique qu'industriel, sur le territoire national.

Suite à l'avènement de la loi N°01.02 du 05 février 2002, la restructuration de Sonelgaz, s'est achevée avec la création de l'ensemble des filiales.

La Sonelgaz est aujourd'hui érigé en Groupe industriel composé de 39 filiales et 5 sociétés en participation. Ainsi, ses filiales métiers de base assurent la production, le transport et la distribution de l'électricité ainsi que le transport et la distribution du gaz par canalisations.

On compte :

- La Société de Production de l'Electricité (*SPE*).
- Sharikat 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'Electricité (*GRTE*).
- La Société de Gestion du Réseau de Transport Gaz (*GRTG*).
- L'Opérateur Système électrique (*OS*), chargée de la conduite du système Production / Transport de l'électricité.
- La Société de Distribution de l'électricité et du gaz d'Alger (*SDA*).
- La Société de Distribution de l'électricité et du gaz du Centre (*SDC*).
- La Société Distribution de l'électricité et du gaz de l'Est (*SDE*).
- La Société de Distribution de l'électricité et du gaz de l'Ouest (*SDO*).

Les sociétés travaux du Groupe Sonelgaz, Kahrif, Kahrakib, Etterkib, Inerga et Kana-ggaz, sont spécialisées dans le domaine de la réalisation des infrastructures énergétiques (engineering, montage industriel, réalisation de réseaux...) et c'est grâce à ces sociétés que l'Algérie dispose aujourd'hui d'infrastructures électriques et gazières répondant aux besoins du développement économique et social du pays.

Par ailleurs, les filiales métiers périphériques sont en charge d'activités annexes, telles que la maintenance d'équipements énergétiques, la distribution de matériel électrique et gazier, le transport et la manutention exceptionnels, ... etc.

De plus, Sonelgaz détient également des participations dans des sociétés, dont le métier est en rapport avec le domaine de l'électricité et du gaz.

1.2.2 Présentation L'Opérateur du Système Electrique

Il a été érigé en filiale du groupe SONELGAZ en 2006, il est le gestionnaire du Système Production - Transport de l'électricité et doit assurer la coordination du système de Production - Transport de l'électricité, conformément à la loi N°02 – 01 du 5 février 2002.

L'OS.Spa veille en particulier à l'équilibre permanent entre consommation et production, à la sécurité, à la fiabilité et à l'efficacité de l'alimentation électrique.

Organisation de la Filiale L'Opérateur du Système Electrique

La filiale OS. Spa est administrée par un conseil d'Administration composé de trois membres au moins et de douze au plus, dont un membre représentant le Ministère de l'Energie et des Mines. La gestion de la Société est assurée par un Président Directeur Général.

L'adaptation de l'organisation de l'OS.Spa, obéit au souci de performance et de maîtrise des activités de l'OS.Spa afin d'atteindre ses objectifs et accomplir les missions qui lui sont dédiées par la loi n°02 – 01 du 05 Février 2002 relative à l'Electricité et à la distribution du gaz par canalisation notamment les articles n°35 et 36, et des statuts de l'Opérateur du système électrique.

Pour assurer l'ensemble de ses activités et missions, le schéma organisationnel de l'Opérateur du Système Electrique est constitué des structures et fonctions suivantes :

- Une Direction Générale avec des assistants dans les domaines communication, Juridique, Audit et Sûreté Interne de l'Etablissement .
 - Une Direction Conduite du Système Electrique.
 - Une Direction Etudes et Analyse Fonctionnement du SPTE .
 - Une Direction planification Réseau de Transport et Raccordement.
 - Un Département Système d'information et Statistiques.
 - Une Division Finances Comptabilité et contrôle de Gestion.
 - Une Division Ressources Humaines.
 - Un Service Administration Générale.
 - Un Service Hygiène et sécurité.
-

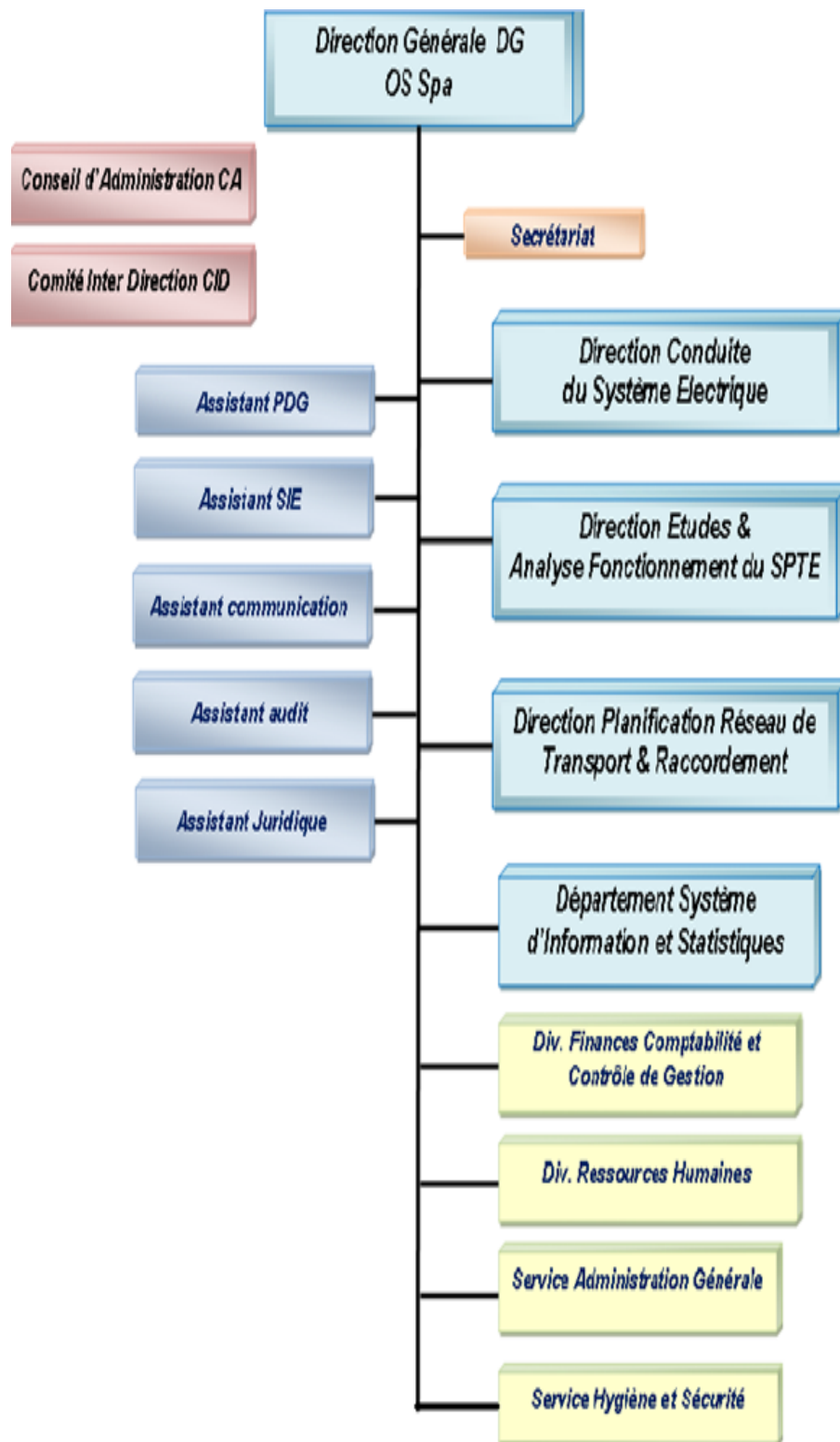


FIGURE 1.1 – Organigramme de la direction des opérateurs systèmes

Présentation de l'organisme d'accueil

Le Département Système Informatique et Statistiques (*SIS*) est indépendant des autres directions, il est relié directement au *PDG*. Il se compose de deux pôles :

Pôle statistique système électrique : – Collecte des données statistiques journalières.

- Traitement des données et analyse l'évolution de la demande d'Energie Electrique.
- Prévision de la demande électrique à court terme Nationale, Régionale et pole d'Adrar.
- Etablir des rapports de synthèse et des bilans énergétiques mensuels, semestriels et annuels.

Pole système d'information, gestion bases de données et site web : – Gérer et actualiser les bases de données statistiques de l'Opérateur Système.

- Concevoir et suivre le site Web de l'Opérateur Système.

1.3 Production de l'électricité en Algérie

1.3.1 Représentation des centrales de production par type

Les usines génératrices de l'énergie électrique peuvent se subdiviser en quatre catégories :

- **Les centrales hydrauliques :**

Dans ces centrales l'énergie est fournie par une chute d'eau provoquée par les vannes d'un barrage. Selon ce cas, l'usine est située au pied des barrages. L'eau arrive alors jusqu'à l'usine par une conduite forcée, dans laquelle elle circule sous forte pression qui fait tourner la turbine et sa rotation produit de l'énergie mécanique qui fait tourner à son tour l'alternateur qui produit du courant électrique.

- **Les centrales thermiques :**

La production de l'électricité dispose à quatre types de centrales thermiques à savoir : turbine vapeur (*TV*), turbine gaz (*TG*), cycle combiné (*CC*) et centrale diesel (*CD*).

Centrale turbine à vapeur (*TV*) : Les turbines peuvent fonctionnées en utilisant plusieurs types de combustibles : gaz naturel, charbon, fuel, bois, gasoil...etc. L'énergie chimique du combustible se transforme en énergie calorifique dans une grande chaudière, cette énergie chauffe l'eau qui se transforme alors en vapeur, la vapeur fait tourner la turbine et sa rotation produit de l'énergie mécanique, laquelle fait tourner à son tour l'alternateur ($3000\text{tour}/\text{min}$) qui produit à son tour de l'énergie électrique.

Centrale turbine à gaz (*TG*) : Dans les centrales turbines gaz, les turbines fonctionnent en utilisant un seul combustible, le gaz. Le rendement faible de la turbine à gaz (25 à 35

Centrale à cycle combiné (*CC*) : Les centraux cycles combinés sont une combinaison entre les turbines vapeurs et les centrales turbines gaz. Les cycles combinés sont des candidats techniquement et économiquement intéressants pour

un fonctionnement en base, en raison de leur très bon rendement et de leurs excellentes prestations environnementales lorsqu'on les compare, à puissances égales, aux autres types d'unités utilisant des combustibles fossiles.

Centrale diesel (CD) : Les turbines à diesel sont nécessaires dans les zones où l'acheminement de l'électricité est impossible ou s'il y a une panne d'électricité dans les infrastructures importantes. Les moteurs diesel utilisant des gasoils, entraînent directement un alternateur qui produit l'électricité, mais ils ont un faible rendement et un caractère polluant.

- **Les centrales éoliennes :**

Définition de l'énergie éolienne :

L'énergie éolienne est produite par la force que le vent exerce sur les pales d'une éolienne, les faisant tourner à entre 10 et 20 tours par minute. Ces pales sont reliées à un alternateur qui transforme l'énergie mécanique en énergie électrique.

Fonctionnement :

Le principe de fonctionnement de l'énergie éolienne est relativement simple ; L'énergie du vent captée sur les pales entraîne le rotor qui, couplé à une génératrice, convertit l'énergie mécanique en énergie électrique. L'électricité produite est dirigée vers le réseau électrique ou vers des batteries de stockage.

Les modes d'exploitation de l'énergie éolienne :

- Les éoliennes terrestres dites " onshore " sont installées sur la terre.
- Les éoliennes dites " offshore " sont installées en mer.

Les éoliennes exigent :

- un apport minimal de vent, généralement de $3m/s$ à $3.8m/s$, pour commencer à tourner et à produire de l'électricité.
- des vents de $13m/s$ à $17m/s$ pour produire à pleine puissance
- des vents en deçà de $25m/s$; quand le cap des $25m/s$ est dépassé, il faut interrompre la production pour éviter le bris d'équipement.

Avantages :

- L'énergie éolienne est indéfiniment durable et propre, elle ne nécessite aucun carburant, ne crée pas de gaz à effet de serre, ne produit pas de déchets toxiques ou radioactifs.
- Le terrain où les éoliennes sont installées reste toujours exploitable pour les activités industrielle et agricole. L'installation peut être démantelée relativement facilement.

Inconvénients :

- L'énergie éolienne dépend de la puissance et de la régularité du vent, c'est une source d'énergie intermittente.
 - Des effets sur le paysage (esthétique), problème de bruit (que l'on remarque essentiellement avec des éoliennes importées d'Asie).
-

- Il peut exister des conflits d'utilisation de l'espace terrestre ou marin avec les autres usagers (exemple : pêcheurs, plaisanciers).

En Algérie :

L'Algérie a un régime de vent modéré (2 à 6m/s). Ce potentiel énergétique convient parfaitement pour le pompage de l'eau particulièrement sur les Hauts Plateaux.

Le programme ENR algérien prévoit dans un premier temps, sur la période 2011 – 2013, l'installation de la première ferme éolienne d'une puissance de 10MW à Adrar. Entre 2014 et 2015, deux fermes éoliennes de 20MW chacune devraient être réalisées. Des études seront menées pour détecter les emplacements favorables afin de réaliser d'autres projets sur la période 2016 – 2030 pour une puissance d'environ 1700MW.

Paramètres influençant la production électrique :

La production d'électricité de l'éolienne va dépendre de plusieurs facteurs liés au site choisi et aux caractéristiques de l'éolienne, qui jouent un rôle important dans la quantité d'énergie transformée :

- **La vitesse du vent :** la quantité d'électricité fournie par une éolienne dépend principalement de la vitesse du vent car l'énergie du vent varie avec le cube de la vitesse moyenne .

Une éolienne capte l'énergie en freinant le vent, c'est-à-dire que les forces s'exerçant sur les pâles sont transformées en puissance de rotation.

De plus, une éolienne va se mettre en mouvement à une vitesse minimale de 2m/s, mais ne commencera à produire qu'à partir de 3.5m/s.

- **La surface balayée par le rotor :** plus le diamètre du rotor est important et plus la surface de balayage augmente. L'énergie récupérée va donc augmenter avec la surface balayée.

- **la densité de l'air :** plus l'air est dense et plus l'éolienne peut récupérer d'énergie car l'énergie cinétique est proportionnelle à la masse volumique de l'air.

Ceci dit, la vitesse du vent est le facteur principal. Si l'énergie du vent varie avec le cube de la vitesse moyenne alors un doublement de la vitesse de vent induit une augmentation de huit fois sa capacité énergétique.

- **Les centrales photovoltaïques :**

Définition de l'énergie solaire :

L'énergie solaire photovoltaïque est une énergie électrique produite à partir du solaire. L'énergie produite par une centrale photovoltaïque est dite renouvelable, car sa source est considérée comme inépuisable à l'échelle de temps humaine.

Fonctionnement :

La production de courant par des cellules photovoltaïques repose sur le principe de l'effet photoélectrique. Ces cellules produisent du courant continu à partir du rayonnement solaire. L'utilisation de ce courant continu diffère d'une installation à l'autre, selon le but de celle-ci.

On distingue principalement deux types d'utilisation, celui où l'installation photovoltaïque est connectée à un réseau de distribution d'électricité et celui où elle ne l'est pas.

Modules photovoltaïques :

Il existe plusieurs techniques de modules solaires photovoltaïques :

- **les modules solaires monocristallins** : possèdent le meilleur rendement au m^2 et sont essentiellement utilisés lorsque les espaces sont restreints. Le coût, plus élevé que celui d'autres installations de même puissance, contrarie le développement de cette technique ;
- **les modules solaires polycristallins** : ont actuellement le meilleur rapport qualité/prix, c'est pourquoi ce sont les plus utilisés. Ils ont un bon rendement et une bonne durée de vie (plus de 35 ans).
- **les modules solaires amorphes** : auront certainement un bon avenir car ils peuvent être souples et ont une meilleure production par faible lumière. Cependant, le silicium amorphe possède un rendement divisé par deux par rapport à celui du cristallin, cette solution nécessite donc une plus grande surface pour la même puissance installée. Toutefois, le prix au m^2 installé est plus faible que pour des panneaux solaires composés de cellules cristallines.

Avantages :

- Les contraintes d'installation sont peu nombreuses
- L'énergie propre et produite localement.
- Une source d'énergie non polluante.
- La centrale photovoltaïque peut être installée partout, même en ville.
- Fiabilité, longue durée de vie (environ 30 ans).

Inconvénients :

- Production intermittente, parfois difficile à prévoir selon l'heure du jour, les conditions climatiques et la saison de l'année.
- L'absence de vision à long terme sur les rendements au futur.
- le coût d'investissement des panneaux photovoltaïques est élevé.
- Le rendement réel de conversion d'un module est faible.

En Algérie :

De par sa situation géographique, l'Algérie dispose d'un des gisements solaire les plus élevés au monde. La durée d'insolation sur la quasi-totalité du territoire national dépasse les 2000 heures annuellement et peut atteindre les 3900 heures (hauts plateaux et Sahara). L'énergie reçue quotidiennement sur une surface horizontale de $1m^2$ est de l'ordre de $5KWh$ sur la majeure partie du territoire national, soit près de $1700KWh/m^2/an$ au Nord et $2263kWh/m^2/an$ au Sud du pays.

De ce fait La stratégie énergétique de l'Algérie repose sur l'accélération du développement de l'énergie solaire. Le gouvernement prévoit le lancement de plusieurs projets solaires photovoltaïques d'une capacité totale d'environ $800MWc$ d'ici 2020.

D'autres projets d'une capacité de $200MWc$ par an devraient être réalisés sur la période 2021 – 2030.

Paramètres influençant la production électrique :

La production photovoltaïque dépendant directement de l'ensoleillement, ce pa-

ramètre varie de façon importante en fonction du temps. Cette variation est du à :

Alternance jour-nuit : Par une journée sans nuages, l'intensité de l'ensoleillement passe d'environ $1000W/m^2$ au maximum vers midi à près de $0W/m^2$ aussitôt la nuit tombée.

Couverture nuageuse : La présence de nuages diminue l'ensoleillement, ce qui réduit en conséquence la production d'énergie. Une diminution de production peut durer de quelques secondes (journées partiellement ensoleillées) à plusieurs jours (couverture nuageuse durable)

Variation quatre saisons : L'intensité de l'ensoleillement journalier varie beaucoup selon les saisons, car l'angle d'inclinaison est plus réduit en été et plus importante en hiver.

1.4 Problématique

La consommation d'électricité ne cesse d'augmenter d'année en année. L'équilibre offre-demande d'électricité s'avère nécessaire. Or la production de cette énergie est non stockable et nécessitent d'énormes investissements ce qui requiert une bonne connaissance de la demande électrique.

Dans ce fait la politique énergétique en Algérie avantage l'installation des centrales de types renouvelable afin d'encourager l'intégration des énergies renouvelables dans le mix énergétique nationale.

La production électrique *ENR's* (énergies renouvelables) qu'elle soit photovoltaïque ou éolienne est aléatoire donc difficile à appréhender. Des travaux de recherche se multiplient pour améliorer les prévisions de production électrique *ENR's* qui dépendent essentiellement sur les fortes incertitudes des simulations météorologiques.

Il existe un besoin de plus en plus pressant pour l'amélioration de la prévision de la production électrique type *ENR's* et l'estimation de son incertitude.

Se qui nous amène a posé la question suivante : quel est le modèle Statistique de prévision pour la production d'électricité *ENR's* qui nous permet d'avoir un résultat qui rapproche le plus de la réalité tout en prenant en considération les facteurs influençant sur l'évolution de cette production ?

2

Aspect théorique des séries temporelles

Introduction

Nous abordons dans ce chapitre l'analyse de données statistiques particulières en ce sens que les observations sont régulièrement échelonnées dans le temps. Ce genre de données est bien connu en économie et fréquente aussi en gestion.

Leur particularité vient de l'introduction du temps dans l'analyse de ces données : on étudie une suite de couples de la forme (Y_t, t) , où Y_t est l'observation de la variable à l'instant t .

Aussi ce chapitre a pour but d'étudier les différentes techniques permettant d'analyser, de modéliser et de prévoir les séries temporelles de type stationnaire ou non, saisonnière ou non.

2.1 Généralité sur les séries temporelles

2.1.1 Définitions de séries temporelle

Définition de processus aléatoire :

Le terme de processus aléatoire est utilisé pour décrire une variable dont le comportement ne peut pas être exprimé entièrement par une relation déterministe.

Un processus aléatoire est une suite de variables aléatoires indexées généralement dans le temps. Ainsi, pour chaque instant, la valeur de la quantité étudiée Y_t est appelée variable aléatoire et l'ensemble des valeurs Y_t quand t varie est appelé processus aléatoire.

Définition de série temporelle

Une série temporelle ou encore chronologique, est une succession d'observations au cours du temps, chacune d'entre elles étant une réalisation d'une variable aléatoire.

Ceci peut s'exprimer aussi en disant que la série temporelle est une réalisation d'une famille de variables aléatoires indicées par le temps. Cette famille de variables aléatoires s'appelle un processus aléatoire, le pas du temps des observations est considéré constant : l'heure, le jour, le mois, l'année, ...

- Une série temporelle est dite discrète si l'ensemble des instants d'observations est discret (dénombrable). Exemple :(nombre de voyageurs SNCF quotidien, température maximale...).
- Une série temporelle est dite continue si l'ensemble des instants d'observations est continu (non dénombrable). Exemple :(signal radio, résultat d'un électrocardiogramme...)

2.1.2 Objectifs principaux

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

Cette étude permet aussi de faire un contrôle, par exemple pour la gestion des stocks, le contrôle d'un processus chimique...

L'un des objectifs principaux de l'étude d'une série chronologique est la prévision qui consiste à prévoir les valeurs futures Y_{t+h} ($h = 1, 2, 3, \dots$) de la série chronologique à partir de ses valeurs observées jusqu'au temps T

Il existe encore bien d'autres objectifs immédiats relatifs à l'étude des séries chronologiques. Par exemple, si deux séries (notant X_t et Y_t) sont observées, on peut se demander quelle influence elles exercent l'une sur l'autre. Ici, deux questions se posent : tout d'abord, la question de la causalité i.e. quelle variable (X_t) va expliquer l'autre (Y_t), ce qui amène la deuxième question, celle du décalage temporel : si une influence de (X_t) sur (Y_t) existe, avec quel délai et pendant combien de temps la variable explicative (X_t) influence-t-elle la variable expliquée (Y_t).

2.1.3 Caractéristiques d'une série chronologique

Moyenne et Variance :

Soit une série temporelle stationnaire $Y_t, t = 1 \dots T$. Les expressions de la moyenne et de la variance $\forall t \in T$ sont :

$$\text{Moyen : } E(Y_t) = \frac{1}{T} \sum_{i=1}^T (Y_i)$$

$$\text{Variance : } Var(Y_t) = \frac{1}{T} \sum_{i=1}^T (Y_i - E(Y_t))^2$$

Fonction d'autocovariance :

Soit Y_t une série de variance finie. on définit la fonction d'autocovariance de Y_t Pour tout t , par :

$$\gamma(h) = cov(Y_t, Y_{t+h}) = E[(Y_t - E(Y_t))(Y_{t+h} - E(Y_{t+h}))] \text{ avec } (t, h) \in Z$$

Cette fonction mesure la covariance entre deux valeurs de Y_t séparées par un certain délai. Elle fournit des informations sur la variabilité de la série et sur les liaisons temporelles qui existent entre les différentes composantes de la série Y_t .

L'auto-covariance au décalage k est définie par :

$$\gamma(h) = \frac{1}{n} \sum_{j=1}^{n-k} ((Y_{j+h} - \bar{Y})(Y_j - \bar{Y})) \text{ avec } \bar{Y} = \frac{1}{n} \sum_{i=1}^n (Y_i)$$

Fonction d'autocorrélation :

La notion d'autocorrélation découle de la notion d'autocovariance comme la corrélation de la covariance. Par définition l'autocorrélation (ou le coefficient d'autocorrélation) de la série Y_t est :

$$\rho(k) = \rho_k = \frac{\text{cov}(Y_t, Y_{t+h})}{\sqrt{\text{var}(Y_t) \cdot \text{var}(Y_{t-k})}} \quad \forall k \in \mathbb{Z}$$

Cette fonction mesure la corrélation des séries avec elle-même décalée de k période.

La représentation graphique de $\gamma(h)$ est appelée corrélogramme

Fonction d'autocorrélation partielle :

L'autocorrélation partielle d'ordre k de la série Y_t est égale au coefficient de corrélation entre :

$$Y_t - E(Y_t / Y_{t-2}, \dots, Y_{t-k+1}) \text{ Et } Y_{t-k} - E(Y_{t-k} / Y_{t-1}, Y_{t-2}, \dots, Y_{t-k+1})$$

D'autres mesures de dépendance entre variables existent en statistique. Un outil de diagnostic intéressant est la fonction d'autocorrélation partielle, que l'on note PACF (Partial ACF).

2.1.4 Les étapes de traitement d'une série chronologique

Schématiquement, les principales étapes de traitement d'une série chronologique sont les suivantes :

correction des données : Il est souvent nécessaire de traiter, modifier les données brutes avant de se lancer dans l'étude d'une série chronologique, par exemple : évaluation de données manquantes, remplacement de données accidentelles,...

observation de la série : Une règle générale en Statistique Descriptive consiste à commencer par regarder les données avant d'effectuer le moindre calcul. Ainsi, une fois la série corrigée et prétraitée, on trace son graphique c'est -à-dire la courbe de coordonnées (t, Y_t) L'observation de ce graphique est souvent une aide 'a la modélisation de la série chronologique et permet de se faire une idée des différentes composantes de la série chronologique.

modélisation : Un modèle est une image simplifiée de la réalité qui vise à traduire les mécanismes de fonctionnement du phénomène étudié et permet de mieux les comprendre. On distingue principalement deux types de modèles :

- **Les modèles déterministes :** Les deux modèles de ce type les plus usités sont les suivants
 - le modèle additif
 - le modèle multiplicatif
- **Les modèles stochastiques :** La classe des modèles de ce type la plus fréquemment utilisée est la classe des modèles ARMA (et de ses sous-modèles ARIMA, SARIMA,...).

analyse de la série à partir de ses composantes : Une fois l'étape de modélisation effectuée, on étudie les composantes du modèle les unes après les autres : tendance, facteurs saisonniers et fluctuations irrégulières.

diagnostique du modèle - ajustement au modèle : Une fois le modèle construit et ses paramètres estimés, on vérifie que le modèle proposé est bon c'est-à-dire l'ajustement au modèle :

- en étudiant les résidus
- en faisant des tests

prédiction (prévision) : Enfin, nous sommes en mesure de faire de la prédiction.

2.1.5 Les composantes d'une série temporelle :

On considère qu'une série chronologique Y_t est la résultante de différentes composantes fondamentale. La représentation graphique de la série permet d'entrevoir ces composantes fondamentales.

- **La composante tendancielle (T_t) :**
elle est censée d'écrire le mouvement à long terme d'une série chronologique. On admet d'ordinaire que cette projection à long terme est une ligne droite à la hausse ou à la baisse qui élimine toutes les fluctuations (aléatoires) dues aux facteurs saisonniers et cycliques.
- **La composante saisonnière (S_t) :**
correspond à un phénomène qui se répète à intervalle de temps réguliers (périodique). En générale c'est un phénomène saisonnier d'où le terme de variation saisonnière.
- **La composante résiduelle (E_t) :**
correspond à des fluctuations irrégulières, en générale de faible intensité mais de nature aléatoire. On parle aussi d'aléas.

Les modèles présentés dans ce travail tiennent compte de ces trois composantes (tendance, saisonnalité et fluctuations irrégulières) il faut cependant remarquer que l'on pourrait envisager d'autres composantes.

- **Des phénomènes cycliques :**

Il s'agit d'un phénomène qui se répète mais contrairement à la saisonnalité sur des durées qui ne sont pas fixes et généralement plus longues.

- **Des phénomènes accidentels :**

(grèves, conditions météorologiques exceptionnelles, crash financier) peuvent notamment intervenir.

2.1.6 Modèle de décomposition :

a : Modèle additif :

Définition :

Un modèle additif suppose que les trois composantes : tendance, variations saisonnières et variations résiduelles sont indépendantes les unes des autres. La série s'écrit comme suit :

$$X_t = T_t + S_t + E_t$$

b : Modèle multiplicatif :

Définition :

Ce modèle suppose que les trois composants sont en interaction entre elles :

$$X_t = T_t \times S_t \times E_t$$

Ce modèle peut être ramené à un modèle additif par transformation logarithmique

c : Modèle mixte :

Définition :

Il s'agit là de modèles où addition et multiplication sont utilisées. On peut supposer par exemple que la composante saisonnière agit de façon multiplicative alors que les fluctuations irrégulières sont additives :

$$X_t = T_t \times S_t + E_t, t = 1 \dots T$$

Choix de modèle :

Avant toute modélisation et étude approfondies du modèle, on tente d'abord de déterminer si on est en présence d'une série dans laquelle pour une observation Y donnée

- la variation saisonnière S s'ajoute simplement à la tendance T ; c'est le modèle additif.
- la variation saisonnière S est proportionnelle à la tendance T ; c'est le modèle multiplicatif.

Afin de faire cette distinction, on peut se baser sur une méthode graphique ou utiliser une méthode analytique.

- **Méthode graphique (Méthode de la bande :)**

On fait un graphique représentant la série chronologique, puis on trace une droite passant respectivement par les minima et par les maxima de chaque saison. Si ces

deux droites sont parallèles, nous sommes en présence d'un modèle additif. Dans le cas contraire, c'est un modèle multiplicatif.

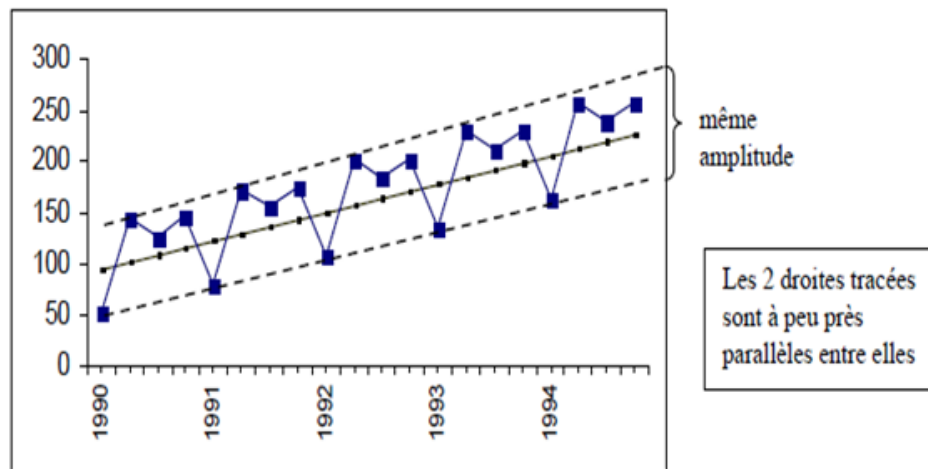


FIGURE 2.1 – Exemple de modèle additif

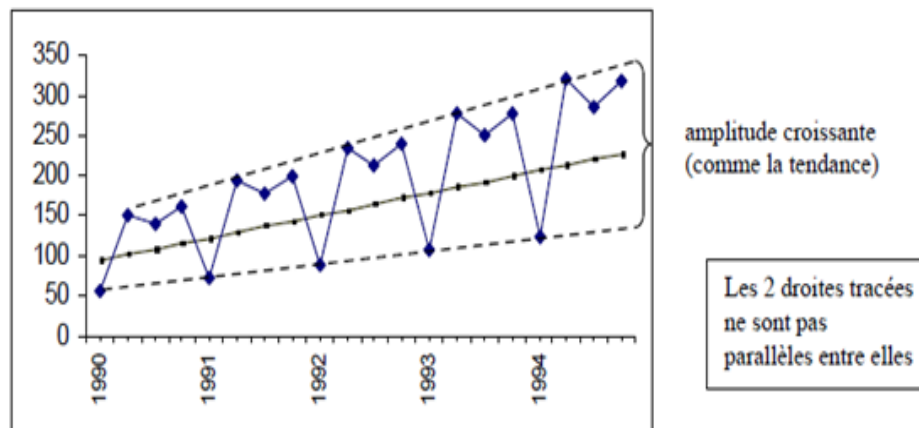


FIGURE 2.2 – Exemple de modèle multiplicatif

- **Méthode analytique (teste de Buys-Ballot) :**

On calcule, pour chacune des périodes, la moyennes \bar{Y}_i et l'écart-type σ_i , puis On estime par la méthode des moindres carrés ordinaires (MCO) les paramètres a_0 et a_1 de l'équation $\bar{Y}_i = a_1 \times Y_i + a_0 + \sigma_i$. dans le cas où le paramètre a_1 n'est pas significativement différent de zéro (teste de student) alors on accepte l'hypothèse d'un modèle additif dans le cas contraire nous retenons un modèle multiplicatif .

2.2 Analyse d'une série chronologique

Les modèles de prévision traditionnels sont basés sur l'extrapolation des tendances passées ie : utilisent le passé de la variable elle-même. Seul le passé de la variable est utilisé en vue de la prévoir sans apport d'information extérieure. La prévision classique est souvent faite sous l'hypothèse de stabilité du système en vue de dégager un scénario tendanciel.

2.2.1 Les opérateurs linéaires de retard, d'avance et de différenciation

- **L'opérateur Retard :**

On aura souvent à considérer une variable en fonction de son passé. Il est donc commode de définir un opérateur qui transforme une variable X_t sa valeur passée. C'est l'opérateur retard désigné par la lettre B et tel que :

$$B(X_t) = X_{t-1} \quad \text{et} \quad B^k(X_t) = X_{t-k}$$

- **L'opérateur d'avance :**

L'opérateur d'avance est l'inverse de l'opérateur de retard, noté F (Forward) il transforme une observation en sa valeur futur , il est défini par :

$$FX_t = X_{t+1}$$

- **L'opérateur de différence ordinaire :**

L'opérateur de différence ordinaire noté Δ associé au processus X_t ; $t \in T$ tel que :

$$\Delta X_t = X_t - X_{t-1} = X_t - (BX_t) = (1 - B)X_t \quad t \in T$$

Et par construction, nous obtiendrons l'opérateur de la d^{eme} différence noté d tel que :

$$\Delta^d X_t = (1 - B)^d X_t$$

- **L'opérateur de différence saisonnière Δ_s :**

L'opérateur s est défini par :

$$\Delta_s X_t = X_t - X_{t-s}$$

En d'autres termes :

$$\Delta_s = (1 - B_s)$$

2.2.2 Les effets des opérateurs

- L'opérateur Δ :

- permet d'éliminer la tendance de la série.
- peut être répété plusieurs fois si la tendance n'est pas linéaire.

- L'opérateur Δ_s :

- permet d'éliminer la saisonnalité de période s .
- On peut également l'appliquer plusieurs fois.

2.2.3 Processus aléatoires stationnaire :

- La stationnarité au sens strict (ou stationnarité forte) :

Définition :

Le processus Y_t est stationnaire au sens strict si pour tout (t_1, t_2, \dots, t_n) avec $t_i \in T_i = 1, 2, \dots, n$. Et si pour $h \in T$ avec $t_{i+h} \in T$, on a $(Y_{t_1}, \dots, Y_{t_n})$ a la même distribution de probabilité jointe que $(Y_{t_1+h}, \dots, Y_{t_n+h})$.

- stationnarité au second ordre (ou stationnarité faible) :

Définition :

Soit Y un processus aléatoire. On dit que Y est stationnaire à l'ordre 2 si la moyenne $m(t)$ et la covariance $\Gamma(s, t)$ sont invariantes par translation dans le temps, ie :

- si la moyenne est constante $E(Y_t) = m_t = m \forall t$
- si la covariance/corrélation dépend seulement de l'écart de temps $k = t - s$, c-à-d. il existe une fonction d'une variable $\gamma(k)$, paire, telle que :

$$Cov(Y_t, Y_s) = C(t, s) = \gamma(t - s) = \gamma(k), \forall k = -2, -1, 0, 1, 2, 3, ..$$

- Le processus Bruit Blanc (white noise) :

Parmi la classe des processus stationnaires, il existe des processus particuliers dits bruits blancs. Ils sont très souvent utilisés en analyse des processus temporels, car ils constituent en quelque sorte " les rubriques élémentaire " de l'ensemble des ces processus.

Définition : (bruit blanc)

Un processus bruit blanc $\epsilon_t; t \in Z$, est une suite de variables aléatoires indépendantes et identiquement distribués de moyenne nulle et de variance constante.

Définition : (bruit blanc gaussien)

Un bruit blanc gaussien s'agit d'un bruit blanc dont la distribution marginale suit une loi normale. $t \sim BBN(0, \delta^2)$.

Un bruit blanc gaussien est une série strictement stationnaire.

Remarque :

- 1 La plupart des processus ne sont pas stationnaires, mais on peut essayer quand-même de se ramener à ce cas par des transformations (logarithmes, Box-Cox, etc).
- 2 Pour un processus du second ordre, la stationnarité stricte implique la stationnarité au sens large (à l'ordre 2). La réciproque est fautive. Une suite Y de variable aléatoire indépendantes de même moyenne et même variance est toujours stationnaire à l'ordre 2; mais si les Y_n n'ont pas tous la même loi, Y n'est pas stationnaire au sens strict.
- 3 La stationnarité à l'ordre 2 est bien plus facile à étudier et vérifier que la stationnarité stricte. Son importance pratique tient surtout aux problèmes de prédiction ou de régression.

2.2.4 Processus aléatoires non stationnaire :

Les processus stochastiques non stationnaire sont caractérisés par des propriétés stochastiques qui évoluent en fonction du temps. Il existe deux sortes de non stationnarité : la non stationnarité de nature déterministe (TS) et la non stationnarité de nature stochastique (DS).

- **les processus TS (trend stationary) :**

Un processus TS s'écrit $Y_t = f_t + \epsilon_t$ où f_t est une fonction polynômiale du temps, linéaire ou non linéaire, et ϵ_t un processus stationnaire.

Le processus TS s'écrit comme :

$$Y_t = a_0 + a_1 t + \epsilon_t$$

Ce processus TS est non stationnaire car $E(Y_t)$ dépend du temps.

Connaissant \hat{a}_0 et \hat{a}_1 , le processus Y_t peut être stationnariser en retranchant de la valeur de Y_t , en t , la valeur estimée $\hat{a}_0 + \hat{a}_1 t$.

- **Les processus DS (difference stationary) :**

Le processus DS représente une non stationnarité de type aléatoire. le processus DS s'écrit comme un processus $AR(1)$:

$$Y_t = \rho Y_{t-1} + \beta + \epsilon_t \dots (*)$$

ϵ_t : est un processus stationnaire est β une constante réelle .

si : $\rho = 1$ est ϵ_t est un bruit blanc alors l'équation (*) s'écrit comme suit :

$$Y_t = Y_{t-1} + \beta + \epsilon_t$$

Le processus Y_t est un DS d'ordre " d " si le processus filtré, noté $(1 - B)^d Y_t$ est stationnaire où $(1 - B)^d Y_t = \beta + \epsilon_t$, et " d " l'ordre de filtre aux différences (de différence ou d'intégration) . On peut dire que le processus est intégré d'ordre noté (d).

À partir de la constante β , on distingue deux types de processus DS :

- pour $\beta = 0$: le processus DS est dit sans dérive, il porte aussi le nom de (marche au hasard ou marche aléatoire "Random walk model").

Une marche aléatoire est ainsi caractérisée par une présence d'une racine unitaire et par le fait que ϵ_t est un bruit blanc, dans ce cas la on parle de processus autorégressif d'ordre 1 avec $\phi = 1$, s'écrit :

$$Y_t = Y_{t-1} + \epsilon_t$$

Pour étudier les caractéristiques de ce modèle, écrivons-le sous sa forme moyenne mobile infinie :

$$Y_t = Y_{t-1} + \epsilon_t Y_{t-1} = Y_{t-2} + \epsilon_{t-1} \implies Y_t = Y_{t-2} + \epsilon_{t-1} + \epsilon_t$$

$$Y_{t-2} = Y_{t-3} + \epsilon_{t-2} \implies Y_t = Y_{t-3} + \epsilon_{t-2} + \epsilon_{t-1} + \epsilon_t.$$

...etc.

Si le premier terme de la série est Y_0 , le modèle s'écrit :

$$Y_t = Y_0 + \sum_{i=1}^t \epsilon_i$$

Les caractéristiques de ce processus sont :

- $E(Y_t) = Y_0$.
- $Var(Y_t) = t\sigma_\epsilon^2$.
- $cov(Y_t, Y_t) = \sigma_\epsilon^2 * \min(t, t)$ si $t \neq t$

Ce processus est non stationnaire en variance puisque elle dépend du temps. Cette non stationnarité est aléatoire. pour stationnariser la marche aléatoire il suffit de'appliquer le filtre aux différences premières :

$$Y_t = Y_{t-1} + \epsilon_t \implies (1 - B)^d Y_t + \epsilon_t$$

- pour $\beta \neq 0$ le processus est dit avec dérive, il s'agit d'un processus autoregressif d'ordre 1 avec constante s'écrit :

$$Y_t = Y_{t-1} + \beta + \epsilon_t.$$

Comme précédemment, on peut chercher sa forme équivalente développée :

$$Y_t = Y_{t-1} + \beta + \epsilon_t.$$

$$Y_{t-1} = Y_{t-2} + \beta + \epsilon_{t-1} \implies Y_t = Y_{t-2} + 2 * \beta + \epsilon_{t-1} + \epsilon_t.$$

$$Y_{t-2} = Y_{t-3} + \beta + \epsilon_{t-2} \implies Y_t = Y_{t-3} + 3 * \beta + \epsilon_{t-2} + \epsilon_{t-1}.$$

ect...

Si on suppose que la valeur d'origine est connue et déterministe on a alors :

$$Y_t = Y_0 + \beta t + \sum_{i=1}^t \epsilon_i$$

Les caractéristiques de ce processus sont :

- $E(Y_t) = Y_0 + \beta t$.
- $var(Y_t) = t\sigma_\epsilon^2$.
- $cov(Y_t, Y_t) = \sigma_\epsilon^2 * \min(t, t)$ si $t \neq t$

Ce processus est non stationnaire en moyenne et en variance.

La stationnarité de ce processus est réalisée en utilisant le filtre aux différences premières.

2.2.5 Modélisation des séries temporelle

a) Les processus linéaires stationnaire :

- **Les processus autorégressif $AR(p)$:**

Ces processus sont construits à partir de l'idée que l'observation au temps t s'explique linéairement par les observations précédentes.

Définition :

On appelle processus autorégressif d'ordre p , noté $AR(p)$, un processus stationnaire Y_t vérifiant une relation de type :

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \epsilon_t$$

- Les ϕ_i sont des réels positifs ou négatifs à déterminer.
- ϵ_t est un bruit blanc de variance σ^2 .

Le modèle autorégressif est stationnaire si et seulement si les racines Z de la fonction caractéristiques $\phi_B = 0$ sont en module supérieur à 1, c'est-à-dire : $\|Z\| > 1$. Où ϕ_B est le polynôme de retard de degré p donnée par :

$$\phi_B = 1 - \sum_{j=1}^p \phi_j B^j, \phi_j \in R \text{ et } \phi_p \neq 0$$

Un processus $(Y_t; t \in T)$ satisfaisant une représentation $AR(p)$ est toujours inversible.

- **Les processus moyenne mobile $MA(q)$:**

Ces processus sont construits à partir de l'idée que l'observation au temps t s'explique linéairement par les observations d'un bruit blanc.

Définition :

On appelle processus moyenne mobile d'ordre q , noté par $MA(q)$ un processus Y_t stationnaire vérifiant une relation du type :

$$Y_t = \epsilon_t - \sum_{j=1}^q \theta_j \epsilon_{t-j}$$

- les θ_j sont des réels positifs ou négatifs à déterminer.
- ϵ_t est un bruit blanc de variance σ^2

Remarque :

- Un processus $(Y_t; t \in T)$ satisfaisant une représentation $MA(q)$ est toujours stationnaire.
- Le processus $MA(q)$ est inversible si les racines de polynôme $(1 - \theta_1 Z - \theta_2 Z^2 - \dots - \theta_q Z^q)$ sont en module > 1 .

- **Les processus ARMA (p, q) :**

Ces processus constituent une extension naturelle des processus AR et MA .

Ce sont en effet des processus mixtes au sens où ils incorporent simultanément

des composantes *AR* et *MA* ce qui permet d'obtenir une description plus parcimonieuse des données.

Définition :

On appelle processus autorégressif moyenne mobile d'ordre (p, q) tout processus $(Y_t), t \in T$ stationnaire vérifiant une relation du type :

$$Y_t - \phi_1 Y_{t-1} - \phi_2 Y_{t-2} - \dots - \phi_p Y_{t-p} = \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \dots - \theta_q \epsilon_{t-q}$$

Cette équation peut aussi s'écrire :

$$\phi(B)Y_t = \theta(B)\epsilon_t$$

où :

$\phi(B)$: représente le polynôme autoregressif d'ordre p , définie par :

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)$$

$\theta(B)$: représente le polynôme moyenne mobile d'ordre q , définie par :

$$(1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q).$$

et ϵ_t est un Bruit Blanc .

Le processus ARMA (p, q) est stationnaire si toutes les racines de polynôme $\phi(B)$ sont à l'extérieur du cercle unité, et inversible si celles du polynôme $\theta(B)$ sont aussi à l'extérieur du cercle unité.

• **Les processus SARMA (ARMA saisonniers) :**

Les processus SARMA ou ARMA saisonniers représentent les mêmes caractéristiques qu'un modèle ARMA (p, q) mais il contient l'effet saisonnier, on les note $ARMA_{s,s}(p, q)$ et il s'écrit :

$$(1 - \phi_{1s} \beta^s - \dots - \phi_{ps} \beta^{ps}) Y_t = (1 - \theta_{1s} \beta^s - \dots - \theta_{qs} \beta^{qs}) \epsilon_t \iff \Phi_s(B^s) Y^t = \Theta_s(B^s) \epsilon_t$$

avec :

$\Theta_s(B^s)$ représente le polynôme moyen mobile saisonnier.

$\Phi_s(B^s)$ représente le polynôme autorégressif saisonnier.

s et s : représente respectivement la période de la saisonnalité.

b) **Les processus linéaire non stationnaire :**

• **Les processus ARIMA (p,d,q) et sa version saisonnière SARIMA :**

Les processus ARIMA (AutoRegressive Integrated Moving Average) sont la généralisation des modèles ARMA pour des processus non stationnaires, admettant une tendance.

Un modèle ARIMA est d'ordre (p, d, q) , p correspondant au nombre de termes autorégressifs, d au nombre de différenciations, et q au nombre de termes de moyenne mobile, il vérifie la relation suivante :

$$\Phi(B)(1 - B)^d Y_t = \Theta(B)\epsilon(t)$$

ou : $\Phi(B) = 1 - \Phi_1(B) - \Phi_2(B)^2 - \dots - \Phi_p(B)^p$.
 et : $\Theta(B) = 1 - \Theta_1(B) - \Theta_2(B)^2 - \dots - \Theta_q(B)^q$.

ϵ_t : est un bruit blanc et d : est l'ordre d'intégration.

les processus SARIMA est la version saisonnière des processus ARIMA .

2.2.6 Les modèles ARMAX

Les modèles ARMAX (Auto Regressive Moving Average with eXogeneous inputs) sont des modèles de régression linéaire avec erreur ARMA, noté aussi *REGARMA*.

La forme du modèle ARMAX est :

$$y_t = \beta_0 + \beta_1 x_{1,t} + \dots + \beta_k x_{k,t} + \mu_t, t = 1, 2, \dots, T$$

$$u_t = \frac{1 + \Theta_1 B + \dots + \Theta_q B^q}{1 - \phi_1 B - \dots - \phi_p B^p}$$

2.2.7 Étude de la stationnarité d'une serie temporelle :

Il existe deux méthodes pour étudier la stationnarité des série temporelle :

Méthode graphique :

La représentation graphique permet de détecter la présence d'une saisonnalité, tendance ou d'une rupture (modification), ainsi l'observation du correlogramme d'après l'étude de la fonction d'autocorrélation.

- dans le correlogramme de la fonction d'autocorrélation partiel (PACF) des pics important apparaissent aus retard S, S_2, \dots , ce qui fait signe de la présence d'une saisonnalité de période S .
- La manière dans la quelle la fonction d'autocorrélation décroît montre la présence ou non d'une tendance.

La représentation graphique de la série peut (dans certains cas) nous indiquer que la série n'est pas stationnaire, mais elle ne nous permet pas de discriminer entre les deux types de non stationnarité (la présence de non stationnarité de type stochastique peut produire une série à l'allure croissante, de la même manière que la non stationnarité de type déterministe).

Méthode analytique :• **Analyse de tendance :**

Les tests de stationnarité permettent de vérifier si une série est stationnaire ou non. Il y a deux types de test différents : les tests de stationnarité, comme le test Kwiatkowski–Phillips–Schmidt–Shin (KPSS), pour lesquels l’hypothèse nulle H_0 est que la série est stationnaire, et les tests de racine unitaire comme le test de Dickey–Fuller, le test augmenté de Dickey–Fuller (ADF), 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.

La perception des processus non stationnaire TS et DS n’est pas une tâche facile, c’est pour cette raison qu’on a recours aux tests de racine unitaire. Ces tests, et comme l’indique leur nom, portent sur l’existence ou non d’une racine unitaire de la fonction caractéristique du modèle de série chronologique candidat à décrire le processus. Si une telle racine existe, alors la transformation adéquate pour la Stationnarisation de ce processus est bien la différence ordinaire, et dans ce cas on dit qu’on est face à un processus non stationnaire de type DS . Le principe de ces tests consiste à tester l’hypothèse nulle de racine unitaire (le processus est non stationnaire DS) contre l’hypothèse alternative d’absence de racine unitaire.

Tests de racine unitaire :

Il existe plusieurs tests de racine unitaire, parmi lesquels nous citons les tests de Dickey–Fuller.

Tests de Dickey–Fuller simple :

Fuller (1976) et Dickey (1976) sont les premiers à fournir un ensemble d’outils statistiques formels pour détecter la présence d’une racine unitaire dans un processus purement autorégressif du premier ordre. Cette procédure de test, maintenant bien connue, est fondée sur l’estimation par les *MCO*, sous l’hypothèse alternative, de trois modèles autorégressifs du premier ordre dont les erreurs sont identiquement et indépendamment distribuées : le modèle sans constante, le modèle avec constante et le modèle avec constante et tendance.

L’hypothèse nulle du test est la présence de racine unitaire, soit la non stationnarité de type stochastique. Dickey et Fuller considèrent trois modèles de base pour la série $Y_t; t = 1, \dots, T$. Les trois modèles de base sont :

Modèle [1] : modèle sans constante ni tendance déterministe :

$$(1 - \Phi_1 B)Y_t = \epsilon_t \implies Y_t = \Phi_1 Y_{t-1} + \epsilon_t; \Phi \in R$$

Modèle [2] : modèle avec constante sans tendance déterministe :

$$(1 - \Phi_1 B)(Y_t - \mu) = \epsilon_t \implies Y_t = \Phi_1 Y_{t-1} + c + \epsilon_t$$

tel que : $c = (1 - \Phi_1)\mu$

modèle [3] : modèle avec constante et tendance déterministe :

$$(1 - \Phi_1 B)(Y_t - \alpha - \beta t) = \epsilon_t \implies Y_t = \Phi_1 Y_{t-1} + c + \beta t + \epsilon_t$$

tel que : $c = \alpha(1 - \Phi_1) + \beta\Phi_1$
 $b = (1 - \Phi_1)\beta$

Tests de Dickey-Fuller Augmenté :

Dans les modèles précédents utilisés pour les tests de Dickey-Fuller simple, le processus ϵ_t est par hypothèse un bruit blanc. Or les tests de Dickey-Fuller augmentés prennent en compte l'hypothèse d'existence d'autocorrélation des erreurs (ce test est utilisé pour blanchir les résidus). Les différents modèles retenus pour construire les statistiques de tests sont les suivants :

modèle [4] : $\Delta Y_t = \rho Y_{t-1} + \sum_{i=1}^p (\lambda_i \Delta Y_{t-1}) + \epsilon_t$

modèle [5] : $\Delta Y_t = \rho Y_{t-1} + c + \sum_{i=1}^p (\lambda_i \Delta Y_{t-1}) + \epsilon_t$

modèle [6] : $\Delta Y_t = \rho Y_{t-1} + c + b t + \sum_{i=1}^p (\lambda_i \Delta Y_{t-1}) + \epsilon_t$

où : $\rho = (1 - \Phi_1)$; $c = (1 - \Phi_1)\mu$; $b = (1 - \Phi_1)\beta$.

- **Analyse de la saisonnalité :**

Analyse de la variance et teste de Fisher :

Cette méthode est basée sur le test de Fisher, on a recours à ce test pour détecter l'existence d'une éventuelle saisonnalité et/ou une tendance.

conclusion

L'analyse des séries temporelles est un objet fondamental de la statistique, qui permet de connaître son concepts en définissant toute les caractéristiques des processus pour pouvoir prévoir la production future de l'électricité.

3

Méthodes de prévisions

Introduction

Nous examinons dans ce chapitre la notion de prévision, la nécessité de prévoir la production des énergies renouvelables ainsi les différentes approches de prévisions pour ce type d'énergie.

Ce chapitre traite aussi les notions essentielles de la régression et discute en détail la méthode des moindres carrés et traite ces propriétés en tant que méthode de prévision ainsi que le modèle *ARMAX*.

3.1 Quelques questions sur une prévision ?

Prévoir, c'est observer un ensemble de données qui permet d'envisager une situation future et d'entreprendre des actions pour y parer concrètement autrement dit c'est porter un jugement sur les événements ou évolutions possibles à venir en utilisant comme outils le passé et le présent.

Il en résulte que les prévisions sont toujours entachées d'erreur et qu'il est possible d'en établir plusieurs pour un même événement à venir (qui constitue l'objet de la prévision). Des outils d'analyse doivent donc être développés afin de comparer et de hiérarchiser les prévisions pour discerner ce qui fait qu'on puisse, ou non, en qualifier certaines de "bonnes".

3.2 Ce qui caractérise un problème de prévision

La première approche de la prévision consiste à en mesurer les spécificités. Pour envisager une typologie des problèmes de prévision en termes d'horizon, de type de produit

ou de secteur, ou en termes de but opérationnel, il est utile de dégager quelques traits qui différencient fondamentalement :

- le secteur d'activité.
- l'utilisation opérationnelle.
- la (ou les) fonction(s) utilisatrice(s) de la prévision.
- l'horizon.

L'approche est très dépendante du secteur d'activité : on ne prévoit pas des livraisons de ciment par les mêmes méthodes que des ventes de savons. Les causalités économiques sous-jacentes sont différentes suivant que le secteur est plus ou moins en amont dans le circuit industriel donc plus ou moins proche de la demande finale, suivant que le produit est stockable ou non, qu'il donne lieu à un marché de renouvellement (télévision) ou non (acier), que le produit est standardisé ou non.

3.3 Les méthodes de prévisions

- **Méthodes extrapolatives**

Ces méthodes utilisent le passé de la variable elle-même.

Seul le passé de la variable est utilisé en vue de la prévoir sans apport d'information extérieure, nous citons à titre d'exemple : lissage par les moyennes mobiles, modélisation *ARMA*.

- **Méthodes explicatives**

Celles ci utilisent les valeurs passées et présentes d'une ou de plusieurs variables pour prévoir y . L'ensemble d'information utilisé comporte des facteurs extérieurs qui peuvent influencer le futur de y en plus du passé de la variable y elle-même.

Parmi ces méthodes on peut citer : la régression linéaire, modèle *ARMAX*.

3.4 La prévision à court terme

Fondamentalement l' horizon d'une prévision à court terme dépend du contexte étudié :

- En gestion, marketing, le court terme est de l'ordre de quelques mois ;
- En météo, le court terme est de l'ordre d'une ou deux journée(s).

Avec un horizon à court terme : on suppose que le phénomène dépend de ses valeurs passées et présent.

3.5 Critères de validité de la méthode de prévision

La prévision des valeurs futures d'une variable y peut se faire en utilisant différentes méthodes. Les prévisions obtenues peuvent être comparées et appréciées selon plusieurs critères parmi lesquels :

1. **L'erreur moyenne** (mean error) : $\bar{e} = \frac{1}{n} \sum e_t$
2. **Le carre moyen des erreurs** (mean square error) : $MSE = \frac{1}{n} \sum e_t^2$
3. **L'erreur quadratique moyenne** (Le root mean square error) : $RMSE = \sqrt{MSE}$

e_t étant l'erreur de prévision pour un instant t (la différence entre la valeur observée et la valeur prévue par une méthode quelconque).

La meilleure méthode est celle qui fournit les valeurs les plus faibles pour ces critères.

3.6 Prévision des énergies renouvelables

Les méthodes de prévision de production d'énergie photovoltaïque et éolienne peuvent être classifiées en quatre approches, à savoir, approche statistique, approche d'intelligence artificielle (*AI*), approche physique et approche hybride.

L'approche statistique : Est fondée sur des modèles statistiques de séries temporelles comme les modèles ARMA, mais ces derniers ne peuvent pas prendre en compte l'information de climatique (les données météorologiques à savoir l'irradiation pour l'énergie photovoltaïque et la vitesse de vent pour l'énergie éolienne). Intuitivement, ces informations sont nécessaires pour améliorer les prévisions. Pour cela, un autre modèle est proposé qui est le modèle ARMAX, ce modèle considère les données météorologiques comme variables exogènes.

L'approches (*AI*) : Utilise des techniques avancées d'intelligence artificielle, tels que les réseaux de neurones artificiels (*ANN*) pour la construction de l'énergie prévisionnistes, qui peuvent également être classés dans la catégorie de l'approche statistique.

L'approche physique : utilise des modèles basés sur les images satellite.

les approches hybrides : sont une combinaison des trois méthodes ci-dessus.

3.7 Régression linéaire par la méthode des moindres carrés

Un modèle de régression linéaire et un modèle de régression d'une variable expliquée sur une ou plusieurs variables explicatives dans la quel on fait l'hypothèse que la fonction qui relie les variables explicatives aux variables expliqués est linéaire.

3.7.1 Principe de la régression linéaire

La régression linéaire décompose une variable aléatoire y en son espérance (ou moyenne), exprimée linéairement en fonction d'autres variables non aléatoires : $x_1, x_2, x_3, \dots, x_k$, plus une erreur aléatoire. K est donc le nombre de variables explicatives (ou co-variables), constante comprise. Par exemple, si $K = 3$ nous avons le modèle linéaire suivant :

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + u_t \quad t = 1, 2, \dots, T \quad (4.1)$$

Les erreurs u_t doivent vérifier un certain nombre de présupposés :

- **P1** être d'espérance nulle, $E(u_t) = 0$;
- **P2** avoir la même variance pour tout t , $var(u_t) = \sigma^2$;
- **P3** être non corrélées entre elles, $corr(u_t, u_s) = 0$, $t \neq s$;
- **P4** être normalement distribuées.

Sous ces hypothèses, l'espérance de y_t est $\beta_1 + \beta_2 * x_{2t} + \beta_3 * x_{3t}$. Nous pouvons aussi écrire matriciellement le modèle et les présupposés :

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ y_T \end{pmatrix} \quad X = \begin{bmatrix} 1 & x_{21} & x_{31} \\ 1 & x_{22} & x_{32} \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 1 & x_{2T} & x_{3T} \end{bmatrix} \quad U = \begin{bmatrix} u_1 \\ u_2 \\ \cdot \\ \cdot \\ u_T \end{bmatrix} \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \cdot \\ \cdot \\ \beta_T \end{bmatrix}$$

Avec ces notations (4.1) s'écrit :

$$Y_{T*1} = X_{K*T} \beta_{T*1} + U_{T*1} \quad (4.2)$$

Et les présupposés s'expriment :

$$\mathbf{P1} : \quad E(Y) = X\beta,$$

$$\mathbf{P2+P3+P4} : \quad U \sim N(0, \sigma_u^2 I_T)$$

I_t est la matrice identité d'ordre T . On peut encore les formuler :

$$Y \sim N(X\beta, \sigma_u^2 I_T) \quad (4.3)$$

La méthode de MCO (moindre carrés ordinaires) estime β par la valeur qui minimise

$$(Y - X * \beta)'(Y - X * \beta)$$

Où X' est la transposée de X .

Le minimum est atteint pour :

$$\hat{\beta} = (X'X)^{-1}X'Y$$

Si les présupposés $P1 + P2 + P3$ sont vérifiés, cet estimateur est sans biais et de variance minimum (théorème de Gauss-Markov). Il est normalement distribué de matrice de covariance $(\sigma_u^2(X'X)^{-1})$:

$$\hat{\beta} \sim N(\beta, \sigma_u^2(X'X)^{-1}) \quad (4.4)$$

Les vecteurs des valeurs ajustées et des résidus son respectivement définis par :

$$\hat{Y} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \cdot \\ \hat{y}_T \end{bmatrix} = X\hat{\beta} \qquad \hat{U} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \cdot \\ \hat{y}_T \end{bmatrix} \qquad (4.5)$$

La variance de l'erreur est estimée sans biais par :

$$s_u^2 = \frac{1}{T-K} \sum_t (y_t - \hat{y}_t)^2 = \frac{1}{T-K} \sum_t \hat{u}_t^2 \qquad (4.6)$$

On estime la matrice des covariances des paramètres en remplaçant dans $(\sigma_u^2(X'X)^{-1})$ la variance inconnue (σ_u^2) par son estimateur s_u^2 . On note $s_u^2(\hat{\beta}_i)$, l'estimation de la variance de $\hat{\beta}_i$ ainsi obtenue, c'est l'élément (i, i) de la matrice $s_u^2(X'X)^{-1}$.

Rappelons que le coefficient $\hat{\beta}_i$ est la quantité dont augmente y quand la variable x_i augmente de 1, toutes choses égales par ailleurs ; dans certains milieux scientifiques, on appelle d'ailleurs $\hat{\beta}_i$, coefficient de régression partielle de x_i .

On trouvera ce vocabulaire dans les séries temporelles quand on étudiera la régression d'une série sur son passé. Une fois effectuée une régression, il faut s'assurer que les présupposer sur les erreurs sont vérifiés, on effectue cette vérification sur les résidus. Le diagramme de dispersion des résidus ou des résidus normalisés contre les valeurs ajustées permet de voir si la distribution de ces résidus est bien indépendante de la valeur ajustée. Le **QQ-plot** de normalité des résidus permet de vérifier l'hypothèse de normalité.

Le logiciel *R* offre de nombreux diagnostics, en particulier dans **lm()**, fonction de base de la régression linéaire.

Dans le cas que nous traiterons, **P4**, la normalité, sera systématiquement examinée, **P3**, la non-corrélation des erreurs, sera souvent rejetée. Il faudra donc comprendre le mécanisme de cette corrélation pour en tenir compte. Une fois la corrélation modélisée, la méthode des moindres carrés généralisés (*MCG*) permet de l'intégrer dans l'estimation de la régression. La connaissance du mécanisme de la corrélation des erreurs au cours du temps permet quant à elle de mieux prédire la série.

Moindres carrés généralisés

Dans la régression MCO d'une série temporelle telle que (4.1), on constate souvent que

les résidus \hat{u}_t sont autocorrélés (si l'on calcule le coefficient de corrélation entre la série des résidus et la série des résidus retardés d'un instant, on obtient une valeur élevée). C'est signe que les u_t le sont eux-mêmes, donc **P3** ne tient pas. Dans d'autres cas, le chronogramme de la série y_t montre que les y_t ont une variance non constante (cas d'hétéroscédasticité), donc **P2** ne tient pas. Si les erreurs sont conjointement normalement distribuées $N(0, \Omega)$, le modèle devient en notation matricielle :

$$Y \sim N(X\beta, \Omega) \qquad (4.7)$$

Où Ω est la matrice des covariances de l'erreur.
La méthode des *MCG* fournit l'estimation de β :

$$\tilde{\beta} = (X' \Omega^{-1} X)^{-1} x' \Omega^{-1} Y \quad \tilde{\beta} \sim N(\beta, X' \Omega^{-1} X)^{-1} \quad (4.7)$$

On ne connaît généralement pas la matrice Ω , mais l'étude des résidus de l'ajustement *MCO* permet d'en découvrir la structure.

Dans ce travail, les résidus de la méthode des *MCO* montrent une dynamique dont on s'inspire pour définir la matrice Ω .

Même si les présupposés sont vérifiés ; il n'est pas sûr qu'une régression linéaire sur un certain ensemble de données soit pertinente, significative. On dispose heureusement d'outils pour apprécier l'intérêt d'une régression linéaire.

3.7.2 Significativité de la régression

Le coefficient de détermination, R^2 , noté dans les sorties, **Multiple R-squared**, est défini de façon unique quand il y a une constante dans la régression par :

$$R^2 = \frac{\sum_{i=1}^T (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^T (y_i - \bar{y})^2} = 1 - \frac{\sum_{i=1}^T \hat{u}_i^2}{\sum_{i=1}^T (y_i - \bar{y})^2} \quad (4.9)$$

C'est le rapport de la variabilité expliquée par la régression : $\sum_{i=1}^T (\hat{y}_i - \bar{y})^2$ sur la variabilité totale des $y_i = \sum_{i=1}^T (y_i - \bar{y})^2$. Le R^2 augmente avec le nombre de variables explicatives. Ce n'est donc pas un indicateur pertinent de significativité de la régression. Il n'est intéressant que pour comparer des régressions portant sur un même nombre d'observations et de variables.

Le coefficient de détermination ajusté, **Adjusted R-squared**, est :

$$R_{adj}^2 = 1 - (1 - R^2) \frac{T-1}{T-k}$$

A T donné, c'est une fonction décroissante de K , le nombre de variables explicatives, constante comprise. L'effet de l'augmentation du nombre de variables sur le R^2 est pénalisant pour la valeur de ce coefficient.

On peut calculer un R^2 pour tout ajustement d'un modèle linéaire de l'espérance, qu'il soit obtenu par *MCO* ou par autre méthode. Les deux expressions de R^2 dans (4.9) ne coïncident généralement pas en dehors de la méthode des *MCO* avec constante dans la régression on estime l'espérance de y par *MCG* ; la variabilité expliquée par la régression peut se calculer par :

$$\sum_{i=1}^T (\tilde{y}_i - \bar{y}_i)^2$$

où \tilde{y}_i désigne l'estimation de ces valeurs ajustées.

Significativité globale .

La régression est dite significative, si au moins un des coefficients des variables autres que la constante est non nul. L'hypothèse nulle est : « la régression n'est pas significative », c'est-à-dire :

$H_0 : \beta_2 = \dots = \beta_K = 0$ contre $H_1 : \text{au moins un de ces coefficients est non nul.}$

Statistique de test .

H_0 exprime une contrainte sur les paramètres. Notons SCR_{libre} et $SCR_{contrainte}$ les sommes des carrés des résidus dans l'estimation libre et dans l'estimation contrainte, c'est-à-dire sous H_0 , on a $SCR_{libre} = \sum_t (y_t - \bar{y})^2$ et $SCR_{contrainte} = \sum_t (y_t - \hat{y}_t)^2$.

La statistique pour tester H_0 contre H_1 est

$$F = \frac{SCR_{contrainte} - SCR_{libre} / (K-1)}{SCR_{libre} / (T-K)}$$

Sous H_0 , F suit une loi de Fisher à $(K-1, T-K)$ ddl : $F \sim F(K-1, T-K)$

On rejette l'hypothèse nulle pour de grandes valeurs de la statistique de test.

Significativité d'un coefficient de régression.

On veut tester qu'un coefficient β_i vaut une certaine valeur de β_{i0} :

$H_0 : \beta_i = \beta_{i0}$ contre $H_1^A : \beta_i \neq \beta_{i0}$ ou contre $H_{1B} : \beta_i < \beta_{i0}$ ou contre $H_1^C : \beta_i > \beta_{i0}$

Statistique de test.

Considérons l'expression de $\hat{\beta}$, (4.4). Un sous vecteur d'un vecteur gaussien est lui-même gaussien, donc $H_0, \hat{\beta}_i \sim N(\beta_{i0}, \sigma_0^2)$ où σ_0^2 est l'élément (i, i) de $(\sigma_u^2 (X'X)^{-1})$.

Sous les quatre présumées et si H_0 est vérifiée, $T_0 = (\hat{\beta}_i - \beta_{i0}) / s(\hat{\beta}_i) \sim T(T_K)$, loi de Student à $T-K$ ddl.

Si $T-K > 10$, la loi de T_0 est très proche d'une loi $N(0, 1)$ et nous nous contenterons

de cette approximation par loi normale.

Sous H_0

$$T_0 = (\hat{\beta}_i - \beta_{i0}) / s(\hat{\beta}_i) \sim AN(0, 1), \quad (4.11)$$

Si l'on choisit $\beta_{i0} = 0$ et que H_0 est rejetée, on dit que β_i est significatif. En toute rigueur, T_0 suit sous H_0 une loi de student à $T-k$ ddl. La statistique $\hat{\beta}_i / s(\hat{\beta}_i)$ est appelée *t-statistique*. Des statistiques semblables sont utilisées dans les modèles de séries temporelles, ou les estimateurs suivent seulement des lois approximativement normales. Nous traiterons la prédiction à la faveur de l'exemple de la section : l'étude d'une production électrique.

3.7.3 Comparaison de modèles et critères d'information

En régression linéaire comme en modélisation de séries temporelles, on est amené à choisir entre différents modèles, emboîtés ou non. Il est classique d'utiliser pour cela un critère d'information.

Le principe de ces critères est le suivant. Les modèles sont estimés par maximum de vraisemblance, mais la comparaison des valeurs des vraisemblances n'est pertinente que si les modèles ont de même nombre de paramètres. Par des arguments théoriques qui peuvent être très divers, les critères d'information fournissent une fonction de l'opposé de la log-vraisemblance augmenté (pénalisé) d'une fonction croissante du nombre de paramètres contenus dans le modèle. Cette dernière peut éventuellement croître avec le nombre d'observations, étant donné plusieurs modèles et leurs estimations, et un critère étant choisi, on retient le modèle pour lequel le critère est minimum. Nous donnons quelques détails sur deux de ces critères.

AIC (Critère d'information d'akaike). L'AIC (akaike's information criterion) est :

$$AIC(k) = -2 \ln(L) + 2k$$

Où L est la fonction de vraisemblance évaluée en les estimations par maximum de vraisemblance (MV) des paramètres, et k est le nombre de paramètres estimés.

(on rencontre quelquefois une autre expression de l'AIC : $(-2 \ln(L) + 2k)/T$ ou T est le nombre d'observations.) dans le cas de l'ajustement par maximum de vraisemblance de (4.3), l'AIC prend la forme :

$$AIC(k) = T \ln(\hat{\sigma}^2) + 2k;$$

Où σ^2 est l'estimation MV de σ_v^2 .

SBC (Critère bayésien de Schwartez). Le SBC (Schwartz' Bayesian Criterion) ou BIC (Bayésien Information Criterion) est :

$$SBC(k) = -2 \ln(L) + 2k \ln(T)$$

Si les erreurs sont normalement distribuées, il prend la forme

$$SBC(k) = T \ln(\hat{\sigma}^2) + k \ln(T)$$

On voit que ces critères (AIC et SBC) sont formés de deux termes :

- le terme $-2 \ln(L)$ qui est d'autant plus faible que l'ajustement par maximum de vraisemblance est bon ;
- le terme $2k$ ou $2k \ln(T)$ qui pénalise cette faible valeur en l'augmentant par une fonction croissante du nombre de paramètres estimés.

3.7.4 Intervalle de confiance (IC)

Un intervalle de confiance est un intervalle dont les bornes sont aléatoires et qui contient une constante avec une probabilité qu'on se fixe. Par exemple si β_i est estimé par $\hat{\beta}_i \sim N(\beta_i, \sigma_i^2)$, un IC à $100(1 - \alpha)\%$ pour β_i est :

$$\hat{\beta}_i \sim q(1 - \alpha/2)\sigma_i$$

où $(1 - \alpha/2)$ désigne le quantile d'ordre $1 - \alpha/2$ d'une v.a. $N(0, 1)$. si σ_i est inconnu, on peut approcher cet IC par $\beta_i \pm q(1 - \alpha/2)s_i$ ou s_i est l'estimation de σ_i .

Si maintenant on s'intéresse à une combinaison linéaire des paramètres $c'\beta$, avec β estimé par $\hat{\beta} \sim N(\beta, \sum_{\beta})$, un IC à $100(1 - \alpha)\%$ pour $c'\beta$:

$$c'\hat{\beta} \pm q(1 - \alpha/2)c'\sum_{\beta c} \text{ approché par } c'\hat{\beta} \pm q(1 - \alpha/2)c'\sum_{\beta c} \quad (4.12)$$

3.7.5 Prédiction

Etant donné par X_F matrice $m * (K + 1)$ de valeurs des explications, pour m observations vérifiant (4.3) et indépendant des premières observations, les y associés vérifiant :

$$Y_F = X_F\beta + U_F \quad U_F \sim N(0, \sigma_u^2 I_m)$$

Le terme « prédiction » recouvre deux situations.

1. Estimation ponctuelle ou estimation par intervalle des composants de $E(Y_F)$.

le terme $E(Y_F) = X_F\beta$ est un vecteur de combinaison linéaire des paramètres.

On estime sans biais ce vecteur certain par

$$E(\hat{Y}_F) = X_F\hat{\beta}$$

De matrice des covariances : $\sigma_u^2 X_F(X'X)^{-1}X_F'$ Sous l'hypothèse de normalité, on peut calculer des IC pour chacune des composantes de $E(Y_F)$. Ainsi, nous prenons x_f une ligne de X_F et y_f la valeur correspondante de y : l'IC à $(1 - \alpha)\%$ pour $E(y_f)$ est

$$x_f' \pm q(1 - \alpha/2)\sigma_u \sqrt{x_f(X'X)^{-1}x_f'} \quad (4.15)$$

Une représentation simultanée de ces IC est appelée bande de confiance.

2. Prédiction ponctuelle ou prédiction par intervalle des composantes de Y_F .

On prédit le vecteur aléatoire $Y_F = X_F\beta + U_F$ par $E(\hat{Y}_F) = x_f\hat{\beta}$ c'est-à-dire par la quantité qui a servi à estimer son espérance (ou moyenne.) Ce prédicteur est sans

biais : $E(E(\hat{Y}_F) - Y_F) = 0$.
L'erreur de prédiction est

$$X_F \hat{\beta} - Y_F = X_F(\hat{\beta} - \beta) - U_F;$$

Sa matrice de covariance est $\sigma_u^2(X_F(X'X)^{-1}X'_F + I_m)$, elle est évidemment plus \ll grande \gg que la matrice des covariances de l'estimateur de la moyenne, en ce sens que $(X_F(X'X)^{-1}X'_F + I_m) - ((X_F(X'X)^{-1}X'_F) = I_m$ est défini positive.

On peut calculer des intervalles de prédiction (*IP*) pour les composantes de Y_F .

La représentation simultanée de ces (*IP*) est appelée bande de prédiction.

L'(*IP*) à $(1 - \alpha)\%$ pour y_f est

$$x'_f \pm q(1 - \alpha/2)\sigma_u \sqrt{1 + x'_f(X'X)^{-1}x'_f} \quad (4.16)$$

3.8 Régression non linéaire

La régression non linéaire a pour but d'ajuster un modèle non linéaire pour un ensemble de valeurs afin de déterminer la courbe qui se rapproche le plus de celle des données de Y en fonction de x .

Le modèle de régression linéaire s'écrit :

$$Y_i = f(X_i, \Theta) + \epsilon_i$$

- La loi de probabilité sur ϵ_i est une loi normale, centrée réduite et de variance σ_i^2 finie;
- Les ϵ_i sont indépendants entre eux;
- Y_i représente l'observation i de la variable dépendante;
- Θ représente un vecteur à p composantes de paramètres généralement inconnus;
- La fonction f est la fonction de régression, la plupart du temps non linéaire. Elle dépend d'une variable réelle x et de paramètres Θ .

Dans les cas où le modèle mathématique est autre non linéaire; il faudra linéariser les données et par le fait même le modèle mathématique. Les principaux modèles que nous pourrions linéariser sont :

- le modèle logarithmique : $Y = a_0 + a_1 \ln X$
- le modèle exponentiel : $Y = a_0 * a_1^X$
- le modèle puissance : $Y = a_1 * X^{a_1}$

3.9 Prévision par le modèle *ARMAX*

Les modèles *ARMAX* sont des modèles de Box-Jenkins pour les cas multivariés. Comme les régressions multiples, ils permettent l'établissement de prévisions d'une variable dépendante à partir de variables indépendantes.

La forme du modèle ARMAX est :

$$y_t = \beta_0 + \beta_1 x_{1,t} + \dots + \beta_k x_{k,t} + \mu_t, \quad t = 1, 2, \dots, T$$

$$u_t = \frac{1 + \Theta_1 B + \dots + \Theta_p B^p}{1 - \phi_1 B - \dots - \phi_p B^p} z_t$$

Où μ_t est l'erreur de modèle de régression, z_t est un bruit blanc de variance σ_z^2 .

Le modèle ARMAX est un modèle linéaire avec erreur corrélées, son estimation est un problème de moindres carrés généralisés.

L'estimation d'un modèle ARMAX consiste à identifier le modèle de l'erreur μ_t , elle se déroule en deux étapes.

3.9.1 Identification des résidus par la méthode des moindres carrés ordinaires MCO

Les résidus sont déterminés à partir d'une régression linéaire par la MCO de y_t sur les variables explicatives, puis un modèle ARMA doit être identifié pour ces résidus.

Modélisation des résidus par un modèle ARIMA

Le choix du modèle résulte d'une procédure itérative qui comporte trois étapes : identification, estimation et validation du modèle.

a) Identification du modèle

Cette phase est très difficile, car le choix des paramètres p , d et q se fait de manière a priori, où éventuellement plusieurs modèles seront examinés.

la fonction d'autocorrélation partielle (PAC) nous donne une idée sur l'ordre du modèle autorégressif, et celle des autocorrélations simple (AC) nous donne une idée sur l'ordre du modèle moyenne mobile.

Le paramètre d représente le nombre de différenciations pour stationnariser la série.

b) Estimation des paramètres :

Une fois l'étape de l'identification terminée, il faut estimer les paramètres qui sont les coefficients des polynômes AR et MA ainsi que les polynômes saisonniers SAR et SMA .

La méthode d'estimation la plus utilisée est celle des moindres carrés ou bien la méthode du maximum de vraisemblance. Cette dernière a donné beaucoup de satisfaction aussi bien pour la précision des résultats obtenus que pour la stabilité et la rapidité des calculs.

Quand l'étape d'estimation est achevée, l'étape suivante va nous permettre de valider le(s) modèle(s) estimé(s).

c) validation de modèle :

Il en existe de très nombreux test permettant de valider le modèle retenu, nous citons :

• **Test de Student des paramètres :**

Afin que le modèle soit valide, il faut que tous les coefficients soient significativement différent de zéro (leurs probabilités soient inférieures à 0,05. Pour vérifier cela, on applique le test de Student.

Si un coefficient n'est pas significativement différent de zéro, on envisage une nouvelle spécification du modèle en supprimant l'ordre *AR* ou *MA* qui n'est pas valide ou bien on calcule la statistique de student :

$$t_c = \frac{|\hat{\phi}_p|}{\sqrt{\text{var}(\hat{\phi}_p)}}$$

avec $\hat{\phi}_p$ estimé au seuil $\alpha = 5\%$.

si $t_c > 1,96$ (la valeur tabulée de student) ,le paramètre est valide.

• **Test de Ljung-Box sur les résidus :**

Ljung et Box ont proposé une modification qui améliore l'approximation Q . cette statistique est donnée par la formule :

$$Q_K^K = n(n+2) \sum_{k=1}^K \left(\frac{\varphi_k^2}{n-k} \right) < \chi_{0,05}^2(k)$$

– Si $Q_k < \chi_{1-\alpha}^2(k-p-q)$ alors la série comporte comme un bruit blanc.

– Si $Q_k > \chi_{1-\alpha}^2(k-p-q)$ alors la série ne comporte pas comme un bruit blanc.

3.9.2 Estimation du modèle ARMAX

Cette étape nécessite l'identification de la variable expliquée et les variables explicatives. Une fois l'étape de l'identification est terminée, il faut estimer un modèle de régression de la variable expliquée sur les variables explicatives en spécifiant que l'erreur suit un modèle ARMA .

3.9.3 Validation du modèle ARMAX

La validation d'un modèle ARMAX est pareille à celle d'un modèle ARMA , elle concerne les tests sur les paramètres et les tests sur les résidus.

3.9.4 Prévion du modèle ARMAX

Une fois terminer les étapes précédents ,on arrive à l'étape de prévion.

conclusion

Ce chapitre nous a donné une introduction aux méthodes de prévisions appliquer dans le prochaine chapitre, mettant l'accent sur des concepts fondamentaux,

4

Application

Introduction

Nous allons présenter dans ce chapitre une étude prévisionnelle des centrales éoliennes et photovoltaïques .

Ressources :

Avant de commencer l'application, on a besoin d'installer le package *caschrono*, par la suite d'autres packages seront appelés comme : car, Chrono, des, forecast... de logiciel R.

4.1 Etude prévisionnelle de la centrale éolienne de kaberten

Identification :

En étude la dépendance de la production d'électricité de type éolienne sur la variable décrit la vitesse de vent. La série VENT représente l'évolution de la vitesse de vent sur une période allant du 08 avril 2016 jusqu'au 14 avril 2016 avec un pas de temps de 15 minutes. Donc, la série sous étude comportera 672 observations.

La série *EOL* représente l'évolution de la production électrique éolienne sur la même période (du 8 avril 2016 jusqu'à 14 avril 2016) avec un pas de temps de 15 minutes.

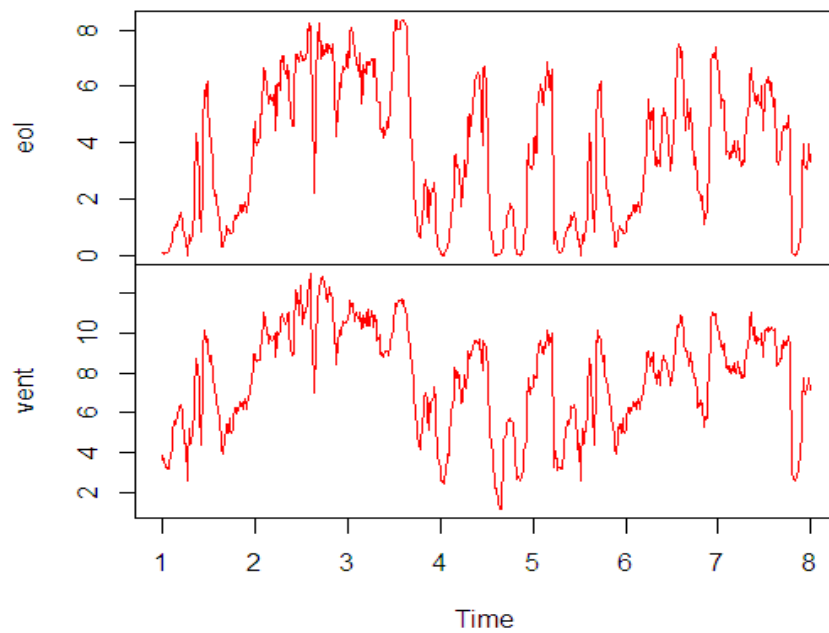


FIGURE 4.1 – Chronogramme de la production d'électricité et la vitesse de vent

Le graphique montre que les deux séries ont presque la même allure, ce qui montre l'influence de la vitesse de vent sur la production d'électricité.

4.1.1 Régression linéaire par la méthode des moindres carrés

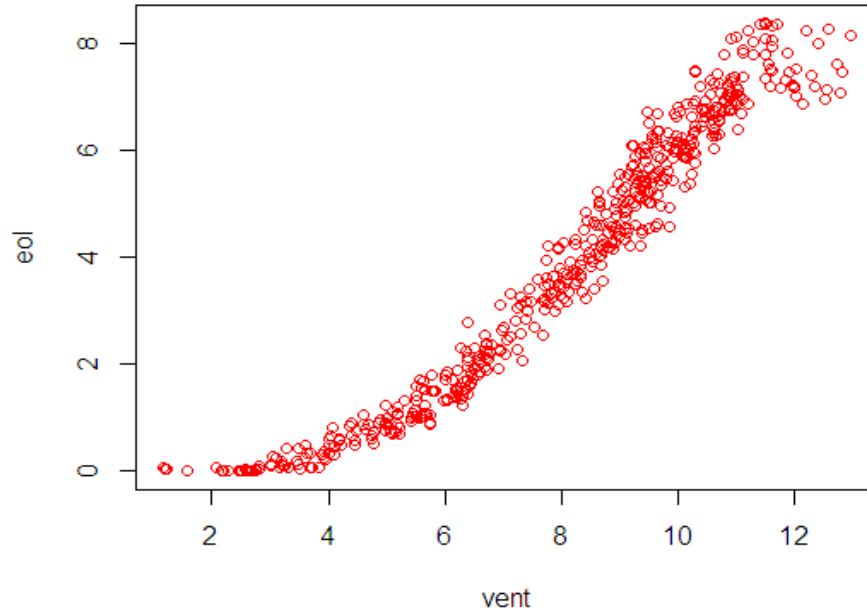


FIGURE 4.2 – Nuage de points de la production en fonction de vitesse de vent

Le nuage de point nous indique qu'il y a une forte corrélation entre la production d'électricité et la vitesse de vent.

D'après le graphe de nuage de points de la production en fonction de vitesse de vent, on voit une courbe qui croît lentement ce qui nous a poussé à proposer un modèle puissance de la forme :

$$eol_t = a \times (vent_t)^b + \mu_t$$

Pour but de simplification, on passe à la linéarisation du modèle.

a)) **Linéarisation du modèle :**

$$eol_t = a \times (vent_t)^b + \mu_t \implies \log(eol_t) = \log(a) + b \times \log(vent_t) + \mu_t$$

le modèle sera :

$$leol_t = \alpha + b \times lven_t + \mu_t$$

tel que : $leol = \log(eol)$; $lvent = \log(vent)$ et $\alpha = \log(a)$.

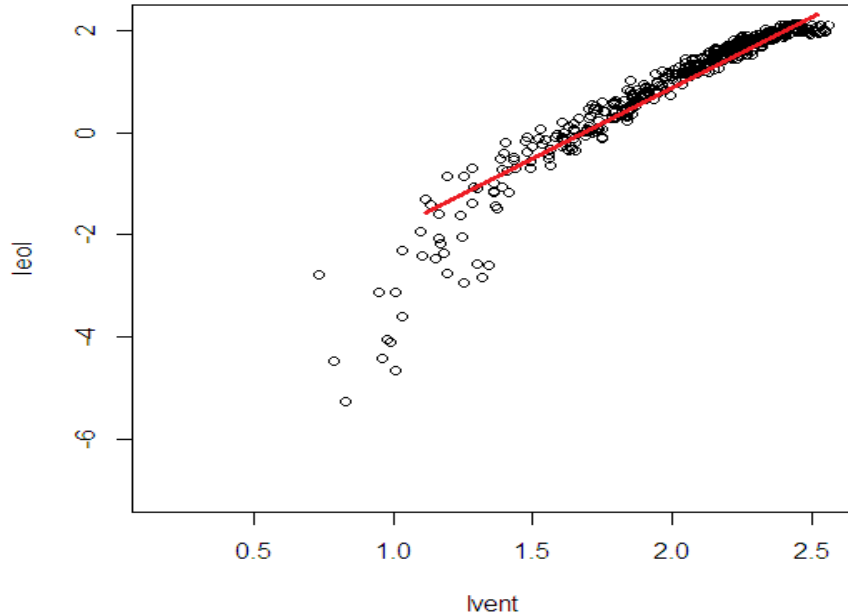


FIGURE 4.3 – le nuage de points des deux séries après la linéarisation

La figure présente le nuage de points des deux séries après la linéarisation, on voit clairement que la plupart des points sont proches d'une droite ce qui confirme la validation de la transformation logarithmique qu'on a effectué sur le modèle. Toutes les étapes qui suit seront effectués sur le modèle linéaire obtenu.

b) Estimation des paramètres du modèle :

L'estimation des paramètres du modèle est basée sur la méthode des moindres carrés ordinaires (*MCO*), elle est obtenue par la fonction $lm()$ fournit par le logiciel *R*.

```
> mod=lm(leol~lvent)
> mod

Call:
lm(formula = leol ~ lvent)

Coefficients:
(Intercept)          lvent
   -6.014           3.432
```

FIGURE 4.4 – Résultats de l'estimation

Suivant les résultats de l'estimation, le modèle de régression s'écrit comme suit :

$$leol_t = -6.014 + 3.432 \times lvent_t$$

c) **Test des résidus :**

Avant toute analyse des estimations, il faut tester les résidus $\hat{\mu}_t$ pour voir si les présupposés de la régression sont vérifiés, pour cela, trois graphiques sont à examiner :

- Graphiques des résidus contre les valeurs ajustées ou contre le temps. Ils peuvent montrer des erreurs dans la spécification de la moyenne.
- lag plot des résidus, ils peuvent indiquer une éventuelle autocorrélation des erreurs μ_t .
- *QQ* plot de normalité.

• **Graphiques des résidus :**

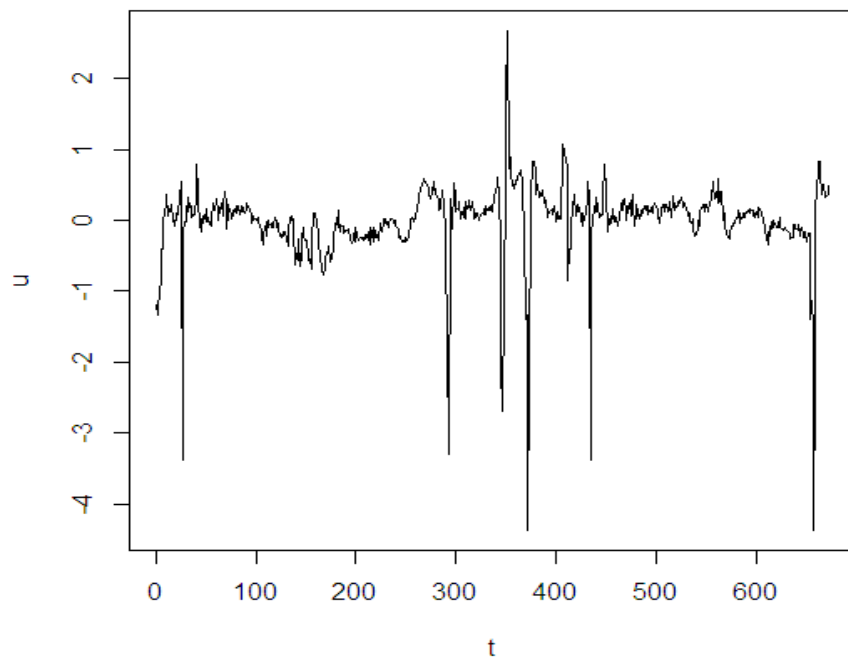


FIGURE 4.5 – graphique des résidus

Calculons la moyenne des résidus

```
> moy=mean(u)
> moy
[1] -9.470635e-19
```

FIGURE 4.6 – Résultat de calcul de moyenne des résidus

A partir de la représentation graphique des résidus et le résultat fourni sous R , on en déduit que les résidus sont centrés.

- **lag plot des résidus :**

La commande suivante fournir le graphe des résidus.

```
lag.plot(rev(residuals(mod)),12,layout=c(4,3),diag.col="red",cex.main=.7)
```

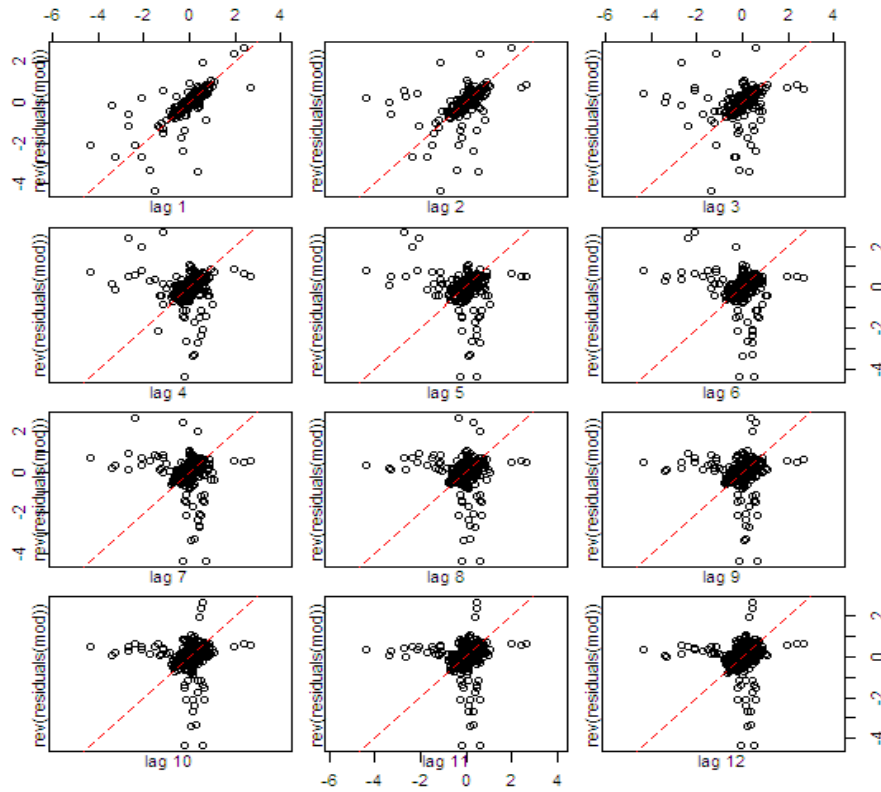


FIGURE 4.7 – lag plot des résidus

On observe sur ce *lag plot* une forte corrélation aux différents retards de 1 jusqu'à 12, donc la présupposé $P3$ (la non corrélation des résidus) ne tient pas mais nous poursuivons la démarche de régression comme si tous les présupposés étaient vérifiés.

- **Test de normalité des résidus :**

La commande suivante fournir le graphe des résidus.

```
plot(mod, which = 2, col = "red")
```

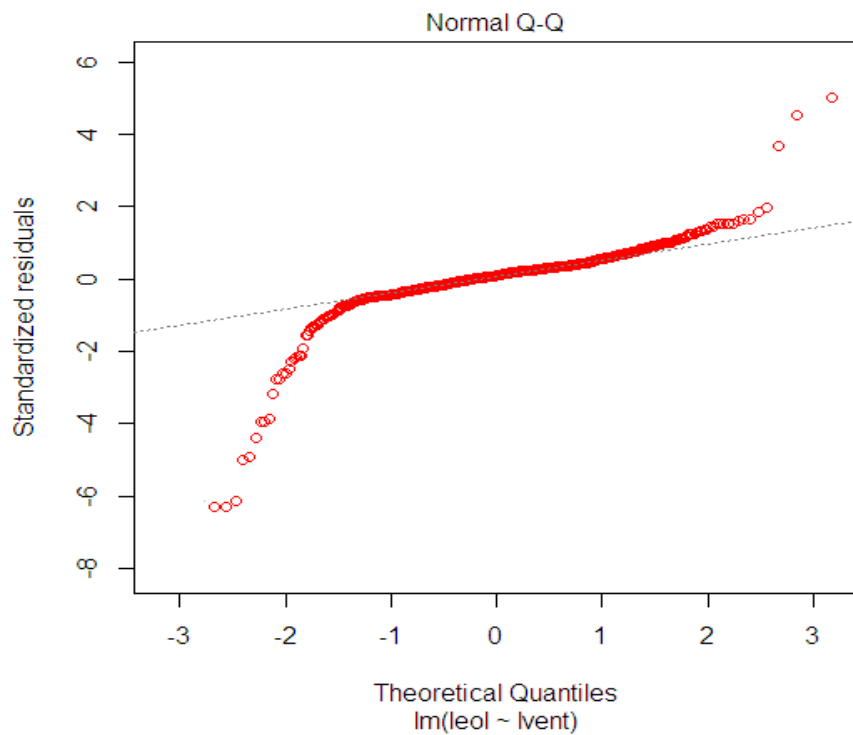


FIGURE 4.8 – Q-Q plot de normalité

La représentation Q-Q plot nous indique qu'effectivement les résidus sont distribués normalement.

d))Significativité du modèle :

Le modèle s'écrit come suit :

$$leol_t = \alpha + b \times lven_t + \mu_t$$

```

> summary(mod)

Call:
lm(formula = leol ~ lvent)

Residuals:
    Min       1Q   Median       3Q      Max
-4.3661 -0.1263  0.0523  0.2014  2.6670

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -6.01436    0.10316  -58.30  <2e-16 ***
lvent        3.43184    0.05113   67.12  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5416 on 670 degrees of freedom
Multiple R-squared:  0.8705,    Adjusted R-squared:  0.8703
F-statistic: 4505 on 1 and 670 DF,  p-value: < 2.2e-16

```

FIGURE 4.9 – L'estimation du modèle

- **Significativité des paramètres :**

1. **Test de la constante :**

$$H_0 : \alpha = 0 \text{ vs } H_1 : \alpha \neq 0$$

$t - stat = -58.30$, $|t_{stat}| > 1.96$ de plus la $p - value = 2 \times 10^{-16} < 0.05$

Donc , H_0 est rejeté $\Rightarrow \alpha$ est significativement différent de zéros au seuil 5%

2. **Test de significativité du modèle :**

$$H_0 : b = 0 \text{ vs } H_1 : b \neq 0$$

$t - stat = 67.12$; $|t_{stat}| > 1.96$ de plus la $p - value = 2 \times 10^{-16} < 0.05$

on rejete $H_0 \Rightarrow b$ est significativement différent de zéros au seuil 5%

De plus le coefficient de détermination $R^2 = 0,87$ donc le modèle est significative, on peut l'écrire comme suit :

$$leolt = -6.01436 + 3.43184 \times lventt \quad (5.1)$$

On peut passer maintenant à l'étape de prévision.

d) Prévisions :

Une fois que le modèle est validé, nous pouvons faire des prévisions des valeurs futures pour l'horizon h .

A cet effet nous n'avons qu'emporté les valeurs de la variable explicative dans L'équation(5.1) Nous allons faire des prévisions pour une journée :le 15 avril 2016 de (00 : 00 : 00) à (23 : 45 : 00) .Le tableau suivant représente les valeurs réels et prédites du la production électrique par la MCO ainsi que les écarts correspondants.

| Quard'heure | valeurs réelles | valeurs prédites | Ecart |
|-------------|-----------------|------------------|--------------|
| 00 :00 :00 | 1,85 | 2.040194811 | -0.190194811 |
| 00 :15 :00 | 1,6786926 | 1.763410474 | -0.084717874 |
| 00 :30 :00 | 1,32764 | 1.426244294 | -0.098600294 |
| 00 :45 :00 | 0,7231688 | 0.833904309 | -0.110735509 |
| 01 :00 :00 | 1,0758745 | 1.380329102 | -0.304454602 |
| 01 :15 :00 | 1,0904008 | 1.288876660 | -0.198475860 |
| 01 :30 :00 | 0,8556713 | 0.962006843 | -0.106335543 |
| 01 :45 :00 | 0,6699372 | 0.657781986 | 0.012155214 |
| 02 :00 :00 | 0,4807481 | 0.465250840 | 0.015497260 |
| 02 :15 :00 | 0,3877253 | 0.415822161 | -0.028096861 |
| 02 :30 :00 | 0,2063047 | 0.249720508 | -0.043415808 |
| 02 :45 :00 | 0,156582 | 0.225388938 | -0.068806938 |
| 03 :00 :00 | 0,1495947 | 0.228571994 | -0.078977294 |
| 03 :15 :00 | 0,2943525 | 0.312323196 | -0.017970696 |
| 03 :30 :00 | 0,3277686 | 0.374726977 | -0.046958377 |
| 03 :45 :00 | 0,2506848 | 0.257211642 | -0.006526842 |
| 04 :00 :00 | 0,2506422 | 0.277149342 | -0.026507142 |
| 04 :15 :00 | 0,3405026 | 0.363975721 | -0.023473121 |
| 04 :30 :00 | 0,1860299 | 0.230669555 | -0.044639655 |
| 04 :45 :00 | 0,2420614 | 0.268920862 | -0.026859462 |
| 05 :00 :00 | 0,2931593 | 0.317412719 | -0.024253419 |
| 05 :15 :00 | 0,35357 | 0.382710711 | -0.029140711 |
| 05 :30 :00 | 0,3159274 | 0.328442937 | -0.012515537 |
| 05 :45 :00 | 0,2872253 | 0.310760899 | -0.023535599 |
| 06 :00 :00 | 0,3871253 | 0.367856960 | 0.019268340 |
| 06 :15 :00 | 0,5262387 | 0.416877452 | 0.109361248 |
| 06 :30 :00 | 0,4605273 | 0.199464340 | 0.261062960 |
| 06 :45 :00 | 0,8229634 | 0.491355366 | 0.331608034 |
| 07 :00 :00 | 1,5960379 | 1.830100098 | -0.234062198 |
| 07 :15 :00 | 2,3833645 | 2.549932360 | -0.166567860 |
| 07 :30 :00 | 4,3253708 | 5.237611883 | -0.912241083 |
| 07 :45 :00 | 4,3096115 | 5.071392618 | -0.761781118 |

| | | | |
|------------|-----------|-------------|--------------|
| 08 :00 :00 | 4,6177547 | 5.220895483 | -0.603140783 |
| 08 :15 :00 | 3,3673593 | 3.869958330 | -0.502599030 |
| 08 :30 :00 | 3,2436581 | 3.167984462 | 0.075673638 |
| 08 :45 :00 | 3,8661659 | 4.626100409 | -0.759934509 |
| 09 :00 :00 | 3,6532074 | 4.353651332 | -0.700443932 |
| 09 :15 :00 | 3,0916695 | 3.459068040 | -0.367398540 |
| 09 :30 :00 | 3,449974 | 4.055101585 | -0.605127585 |
| 09 :45 :00 | 2,826232 | 3.374650247 | -0.548418247 |
| 10 :00 :00 | 2,3653039 | 2.765503513 | -0.400199613 |
| 10 :15 :00 | 1,7461173 | 1.909168493 | -0.163051193 |
| 10 :30 :00 | 1,3394805 | 1.821531263 | -0.482050763 |
| 10 :45 :00 | 0,937874 | 1.168649687 | -0.230775687 |
| 11 :00 :00 | 0,7627247 | 0.800185263 | -0.037460563 |
| 11 :15 :00 | 0,5426954 | 0.472696368 | 0.069999032 |
| 11 :30 :00 | 0,4557708 | 0.495213909 | -0.039443109 |
| 11 :45 :00 | 0,299442 | 0.374934768 | -0.075492768 |
| 12 :00 :00 | 0,1792787 | 0.243695533 | -0.064416833 |
| 12 :15 :00 | 0,1890125 | 0.258737881 | -0.069725381 |
| 12 :30 :00 | 0,0989773 | 0.177590841 | -0.078613541 |
| 12 :45 :00 | 0,0591794 | 0.197440375 | -0.138260975 |
| 13 :00 :00 | 0,164498 | 0.170235729 | -0.005737729 |
| 13 :15 :00 | 0,113326 | 0.221132617 | -0.107806617 |
| 13 :30 :00 | 0,1091347 | 0.181822242 | -0.072687542 |
| 13 :45 :00 | 0,0324493 | 0.081006185 | -0.048556885 |
| 14 :00 :00 | 0,0097647 | 0.174808982 | -0.165044282 |
| 14 :15 :00 | 0,0527433 | 0.096081354 | -0.043338054 |
| 14 :30 :00 | 0 | 0.024873197 | -0.024873197 |
| 14 :45 :00 | 0 | 0.008817720 | -0.008817720 |
| 15 :00 :00 | 0 | 0.019815624 | -0.019815624 |
| 15 :15 :00 | 0 | 0.021832208 | -0.021832208 |
| 15 :30 :00 | 0 | 0.023125278 | -0.023125278 |
| 15 :45 :00 | 0 | 0.020869806 | -0.020869806 |
| 16 :00 :00 | 0 | 0.049272339 | -0.049272339 |
| 16 :15 :00 | 0 | 0.090173423 | -0.090173423 |
| 16 :30 :00 | 0,002498 | 0.085440288 | -0.082942288 |
| 16 :45 :00 | 0 | 0.069542560 | -0.069542560 |
| 17 :00 :00 | 0 | 0.055964051 | -0.055964051 |
| 17 :15 :00 | 0 | 0.033420459 | -0.033420459 |
| 17 :30 :00 | 0 | 0.014442387 | -0.014442387 |
| 17 :45 :00 | 0 | 0.014187015 | -0.014187015 |

| | | | |
|------------|--------|-------------|--------------|
| 18 :00 :00 | 0.0001 | 0.028680268 | -0.028680268 |
| 18 :15 :00 | 0 | 0.005846391 | -0.005846391 |
| 18 :30 :00 | 0 | 0.001596601 | -0.001596601 |
| 18 :45 :00 | 0 | 0.002421716 | -0.002421716 |
| 19 :00 :00 | 0 | 0.004043903 | -0.004043903 |
| 19 :15 :00 | 0 | 0.062155097 | -0.062155097 |
| 19 :30 :00 | 0 | 0.053287773 | -0.053287773 |
| 19 :45 :00 | 0 | 0.027237005 | -0.027237005 |
| 20 :00 :00 | 0 | 0.007718512 | -0.007718512 |
| 20 :15 :00 | 0 | 0.001307204 | -0.001307204 |
| 20 :30 :00 | 0 | 0.001737260 | -0.001737260 |
| 20 :45 :00 | 0 | 0.002238505 | -0.002238505 |
| 21 :00 :00 | 0 | 0.001185970 | -0.001185970 |
| 21 :15 :00 | 0 | 0.001452573 | -0.001452573 |
| 21 :30 :00 | 0 | 0.003901886 | -0.003901886 |
| 21 :45 :00 | 0 | 0.008104119 | -0.008104119 |
| 22 :00 :00 | 0 | 0.012424293 | -0.012424293 |
| 22 :15 :00 | 0 | 0.035198061 | -0.035198061 |
| 22 :30 :00 | 0 | 0.035756497 | -0.035756497 |
| 22 :45 :00 | 0 | 0.018470655 | -0.018470655 |
| 23 :00 :00 | 0 | 0.013860123 | -0.013860123 |
| 23 :15 :00 | 0 | 0.018400421 | -0.018400421 |
| 23 :30 :00 | 0 | 0.018942555 | -0.018942555 |
| 23 :45 :00 | 0 | 0.023427688 | -0.023427688 |

4.1.2 Prévision par le modèle ARMAX

On va modéliser l'influence de la série $vent(t)$ sur la série $eol(t)$ par un modèle *ARMAX*.

Comme on a dit au chapitre précédent, l'estimation d'un modèle *ARMAX* se fait en deux étapes, la première consiste à faire une régression *MCO* de la série $eolt$ sur la variable explicative $vent_t$ et la deuxième se réalise en identifiant le modèle des résidus.

a) Identification de la série des résidus par la méthode des moindres carrés ordinaires MCO :

Le modèle que nous avons retenu pour expliquer la production d'électricité par la variable vitesse de vent est :

$$leol_t = -6.01436 + 3.43184 \times lvent_t + \mu_t$$

> $mod = lm(leol \sim lvent)$

> $u = ts(residuals(mod), start = 1, frequency = 96)$

les résidus $\hat{\mu}_t$ montraient une autocorrélation significative, nous allons identifier leur modèle, mais tout d'abord on doit les récupérer et les transformer en série temporelle pour ne pas avoir à la préciser au cours de l'emploi de modèle *ARIMA*

et d'autres fonctions.

b) modélisation des résidus par un modèle ARIMA :

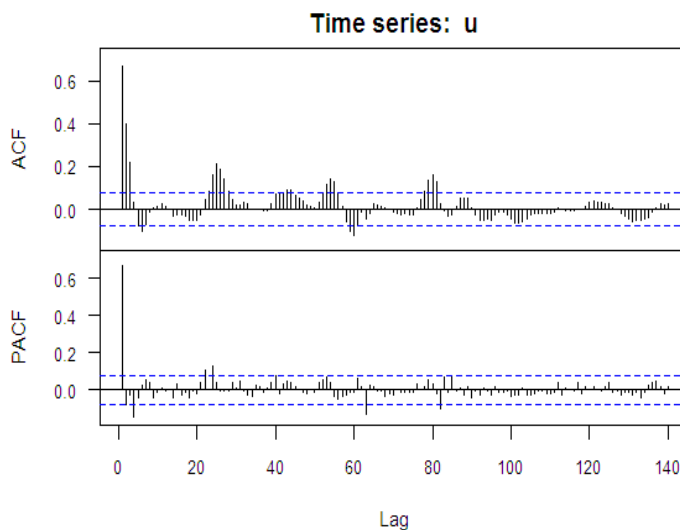


FIGURE 4.10 – Fonction d'autocorrélation de la série des résidus

L'analyse de graphes de la fonction d'autocorrélation (ACF) et celui de la fonction d'autocorrélation partielle ($PACF$) nous permet de déterminer les ordres ($p; d; q$) du modèle.

Utilisons la procédure d'identification automatique $auto.arima()$ de package *forecast* pour estimer le modèle le plus adéquat pour la série des résidus.

```
> model=auto.arima(u,max.p=4,max.q=4,max.P=1,approximation=FALSE)
> model
Series: u
ARIMA(2,0,1) with zero mean

Coefficients:
      ar1      ar2      ma1
      1.3918  -0.5325  -0.6719
s.e.  0.1312   0.0842   0.1385

sigma^2 estimated as 0.1599: log likelihood=-336.4
AIC=680.81  AICc=680.87  BIC=698.85
```

FIGURE 4.11 – Résultat d'estimation du modèle

```
> t_stat(model)
              ar1      ar2      ma1
t.stat 10.61086 -6.326452 -4.851181
p.val   0.000000 0.000000 0.000001
```

le modèle $ARIMA(2, 0, 1)$ est sélectionné pour la série des résidus μ_t .

- **Test de student des paramètres**

| Paramètre | p-value |
|-----------|----------|
| ar(1) | 0 |
| ar(2) | 0 |
| ma(1) | 0.000001 |

Pour tous les paramètres la p -value est inférieure à 0.05, donc ces paramètres sont significative au seuil 5%.

- **Test de Ljung-Box sur les résidus :**

Le test de Ljung-Box est bâti sur le fait que la fonction d'autocorrélation des erreurs ne doit pas être significativement différente de 0.

```
> llag=seq(4,140,4)
> t(Box.test.2(residuals(model),llag,type="Ljung-Box",decim=2,fitdf=3))
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13] [,14]
Retard  4.00  8.00 12.00 16.00 20.00 24.00 28.00 32.00 36.00 40.00 44.00 48.00 52.00 56.00
p-value 0.15 0.09 0.27 0.41 0.63 0.32 0.08 0.15 0.24 0.24 0.29 0.43 0.56 0.45
      [,15] [,16] [,17] [,18] [,19] [,20] [,21] [,22] [,23] [,24] [,25] [,26] [,27]
Retard 60.00 64.00 68.00 72.00 76.00 80.00 84.0 88.00 92.00 96.00 100.00 104.00 108.00
p-value 0.35 0.13 0.17 0.27 0.36 0.25 0.1 0.12 0.13 0.16 0.24 0.29 0.38
      [,28] [,29] [,30] [,31] [,32] [,33] [,34] [,35]
Retard 112.00 116.00 120.00 124.00 128.00 132.00 136.00 140.00
p-value 0.45 0.55 0.64 0.73 0.78 0.84 0.89 0.91
```

FIGURE 4.12 – Résultats du Test de Ljung Box

$\forall p = 1 \dots 140$, la p -value pour chaque retard est supérieure à 0.05, donc la série des résidus est un bruit blanc.

Nous pouvons passer à l'estimation du modèle ARMAX.

c) **Estimation du modèle ARMAX :**

Afin d'estimer le modèle *ARMAX* on doit définir la variable à expliquer (la série *leol*) et la variable explicative (la série $lvent_t$) pour cela en utilise la commande suivante :

```
> modelx = Arima(leol, order = c(2, 0, 1), xreg = lvent)
> modelx
```

```
Series: leol
ARIMA(2,0,1) with non-zero mean

Coefficients:
          ar1      ar2      ma1  intercept      V1
      1.3819 -0.5256 -0.6614   -6.0597  3.4546
s.e.  0.1420  0.0919  0.1499    0.1851  0.0918

sigma^2 estimated as 0.1604: log likelihood=-336.37
AIC=684.75  AICc=684.87  BIC=711.81
```

FIGURE 4.13 – Résultat d'estimation

• **Test de student des paramètres :**

| Paramètre | p-value |
|-----------|---------|
| ar1 | 0.00000 |
| ar2 | 0.00000 |
| ma1 | 0.00001 |
| intercept | 0.00000 |
| v1 | 0.00000 |

On voit que tous les paramètres du modèle sont significativement différent de zéros au seuile 5%(la p-value pour chaque paramètre est inferieure a 0.05).

• **Test de Ljung-Box sur les résidus :**

```

> t(Box.test.2(residuals(modelx),seq(1,140,4),type="Ljung-Box",decim=2,fitdf=5))

      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12]
Retard 1.00 5.00 9.00 13.0 17.00 21.00 25.00 29.00 33.00 37.00 41.00 45.0
p-value 0.77 0.09 0.12 0.3 0.46 0.54 0.07 0.09 0.14 0.25 0.24 0.3
      [,13] [,14] [,15] [,16] [,17] [,18] [,19] [,20] [,21] [,22] [,23] [,24]
Retard 49.00 53.00 57.00 61.00 65.0 69.00 73.00 77.00 81.00 85.00 89.00 93.00
p-value 0.42 0.53 0.44 0.34 0.1 0.17 0.27 0.35 0.11 0.09 0.09 0.12
      [,25] [,26] [,27] [,28] [,29] [,30] [,31] [,32] [,33] [,34]
Retard 97.00 101.00 105.00 109.00 113.00 117.00 121.00 125.00 129.00 133.00
p-value 0.16 0.22 0.28 0.37 0.44 0.55 0.63 0.72 0.78 0.83
      [,35]
Retard 137.00
p-value 0.88

```

FIGURE 4.14 –

Le modèle obtenu finalement est :

$$leol_t = -6.0597 + 3.4546 \times lven_t + \mu_t, t = 1 \dots 672$$

avec
$$\mu_t = \frac{1+0.6614}{(1-1.3819B-0.5266B^2)} z_t.$$

d) Prévision par le modèle ARMAX :

La prévision est faite par la commande suivante :

```
> prev = forecaste(modelx, level = c(95, 90), fan = FALSE, xreg = log(new))
```

Nous allons faire des prévisions pour une journée :le 15 avril 2016 de (00 : 00 : 00) à (23 : 45 : 00) Le tableau suivant représente les valeurs réels et prédites de la production électrique par le modèle ARMAX ainsi que les écarts correspondants.

| Quard'heure | Valeurs réelles | Valeurs prédites | Ecart |
|-------------|-----------------|------------------|--------------|
| 00 :00 :00 | 1,85 | 2,724804902 | -0,874804902 |
| 00 :15 :00 | 1,6786926 | 2,04484837 | -0,36615577 |
| 00 :30 :00 | 1,327644 | 1,501294551 | -0,173650551 |
| 00 :45 :00 | 0,7231688 | 0,825346492 | -0,102177692 |
| 01 :00 :00 | 1,0758745 | 1,330042571 | -0,254168071 |
| 01 :15 :00 | 1,0904008 | 1,227683647 | -0,137282847 |
| 01 :30 :00 | 0,8556713 | 0,915042616 | -0,059371316 |
| 01 :45 :00 | 0,6699372 | 0,628199373 | 0,041737827 |
| 02 :00 :00 | 0,4807481 | 0,44721672 | 0,03353138 |
| 02 :15 :00 | 0,3877253 | 0,402886534 | -0,015161234 |
| 02 :30 :00 | 0,2063047 | 0,24292797 | -0,03662327 |
| 02 :45 :00 | 0,156582 | 0,220348999 | -0,063766999 |
| 03 :00 :00 | 0,1495947 | 0,224357331 | -0,074762631 |
| 03 :15 :00 | 0,2943525 | 0,307945063 | -0,013592563 |
| 03 :30 :00 | 0,3277686 | 0,370405073 | -0,042636473 |
| 03 :45 :00 | 0,2506848 | 0,253754176 | -0,003069376 |
| 04 :00 :00 | 0,2506422 | 0,273572314 | -0,022930114 |
| 04 :15 :00 | 0,3405026 | 0,359845851 | -0,019343251 |
| 04 :30 :00 | 0,1860299 | 0,227297131 | -0,041267231 |
| 04 :45 :00 | 0,2420614 | 0,265168761 | -0,023107361 |
| 05 :00 :00 | 0,2931593 | 0,31323464 | -0,02007534 |
| 05 :15 :00 | 0,35357 | 0,378052403 | -0,024482403 |
| 05 :30 :00 | 0,3159274 | 0,324067607 | -0,008140207 |
| 05 :45 :00 | 0,2872253 | 0,306478345 | -0,019253045 |
| 06 :00 :00 | 0,3871253 | 0,363169079 | 0,023956221 |
| 06 :15 :00 | 0,5262387 | 0,41189284 | 0,11434586 |
| 06 :30 :00 | 0,4605273 | 0,196126108 | 0,264401192 |
| 06 :45 :00 | 0,8229634 | 0,486005533 | 0,336957867 |
| 07 :00 :00 | 1,5960379 | 1,825902837 | -0,229864937 |
| 07 :15 :00 | 2,3833645 | 2,549666815 | -0,166302315 |
| 07 :30 :00 | 4,3253708 | 5,261952997 | -0,936582197 |
| 07 :45 :00 | 4,3096115 | 5,096180818 | -0,786569318 |
| 08 :00 :00 | 4,6177547 | 5,245129858 | -0,627375158 |
| 08 :15 :00 | 3,3673593 | 3,880296856 | -0,512937556 |
| 08 :30 :00 | 3,2436581 | 3,172280913 | 0,071377187 |
| 08 :45 :00 | 3,8661659 | 4,644039642 | -0,777873742 |
| 09 :00 :00 | 3,6532074 | 4,368678497 | -0,715471097 |
| 09 :15 :00 | 3,0916695 | 3,465764974 | -0,374095474 |
| 09 :30 :00 | 3,449974 | 4,067197023 | -0,617223023 |
| 09 :45 :00 | 2,826232 | 3,380631918 | -0,554399918 |
| 10 :00 :00 | 2,3653039 | 2,766782935 | -0,401479035 |
| 10 :15 :00 | 1,7461173 | 1,905406211 | -0,159288911 |
| 10 :30 :00 | 1,3394805 | 1,817379994 | -0,477899494 |
| 10 :45 :00 | 0,937874 | 1,162590831 | -0,224716831 |

| | | | |
|------------|-----------|-------------|--------------|
| 11 :00 :00 | 0,7627247 | 0,794058018 | -0,031333318 |
| 11 :15 :00 | 0,5426954 | 0,467457193 | 0,075238207 |
| 11 :30 :00 | 0,4557708 | 0,489874918 | -0,034104118 |
| 11 :45 :00 | 0,299442 | 0,370215089 | -0,070773089 |
| 12 :00 :00 | 0,1792787 | 0,239947689 | -0,060668989 |
| 12 :15 :00 | 0,1890125 | 0,254858982 | -0,065846482 |
| 12 :30 :00 | 0,0989773 | 0,174496455 | -0,075519155 |
| 12 :45 :00 | 0,0591794 | 0,194135236 | -0,134955836 |
| 13 :00 :00 | 0,164498 | 0,167223023 | -0,002725023 |
| 13 :15 :00 | 0,113326 | 0,217592838 | -0,104266838 |
| 13 :30 :00 | 0,1091347 | 0,178681778 | -0,069547078 |
| 13 :45 :00 | 0,0324493 | 0,079185256 | -0,046735956 |
| 14 :00 :00 | 0,0097647 | 0,171745258 | -0,161980558 |
| 14 :15 :00 | 0,0527433 | 0,094026935 | -0,041283635 |
| 14 :30 :00 | 0 | 0,024126174 | -0,024126174 |
| 14 :45 :00 | 0 | 0,008494815 | -0,008494815 |
| 15 :00 :00 | 0 | 0,01919181 | -0,01919181 |
| 15 :15 :00 | 0 | 0,021158379 | -0,021158379 |
| 15 :30 :00 | 0 | 0,022420014 | -0,022420014 |
| 15 :45 :00 | 0 | 0,02021969 | -0,02021969 |
| 16 :00 :00 | 0 | 0,048007671 | -0,048007671 |
| 16 :15 :00 | 0 | 0,083548935 | -0,083548935 |
| 16 :30 :00 | 0,002498 | 0,083548935 | -0,081050935 |
| 16 :45 :00 | 0 | 0,0679112 | -0,0679112 |
| 17 :00 :00 | 0 | 0,054573273 | -0,054573273 |
| 17 :15 :00 | 0 | 0,032479709 | -0,032479709 |
| 17 :30 :00 | 0 | 0,013958687 | -0,013958687 |
| 17 :45 :00 | 0 | 0,01371026 | -0,01371026 |
| 18 :00 :00 | 0 | 0,02784495 | -0,02784495 |
| 18 :15 :00 | 0 | 0,005617109 | -0,005617109 |
| 18 :30 :00 | 0 | 0,00152096 | -0,00152096 |
| 18 :45 :00 | 0 | 0,002313307 | -0,002313307 |
| 19 :00 :00 | 0 | 0,003875912 | -0,003875912 |
| 19 :15 :00 | 0 | 0,060652263 | -0,060652263 |
| 19 :30 :00 | 0 | 0,052060108 | -0,052060108 |
| 19 :45 :00 | 0 | 0,695372523 | -0,695372523 |
| 20 :00 :00 | 0 | 0,007429358 | -0,007429358 |
| 20 :15 :00 | 0 | 0,001243638 | -0,001243638 |
| 20 :30 :00 | 0 | 0,001655873 | -0,001655873 |
| 20 :45 :00 | 0 | 0,002137193 | -0,002137193 |

| | | | |
|------------|---|-------------|--------------|
| 21 :00 :00 | 0 | 0,001127577 | -0,001127577 |
| 21 :15 :00 | 0 | 0,001382896 | -0,001382896 |
| 21 :30 :00 | 0 | 0,003738916 | -0,003738916 |
| 21 :45 :00 | 0 | 0,010221677 | -0,010221677 |
| 22 :00 :00 | 0 | 0,011996313 | -0,011996313 |
| 22 :15 :00 | 0 | 0,034218924 | -0,034218924 |
| 22 :30 :00 | 0 | 0,034765421 | -0,034765421 |
| 22 :45 :00 | 0 | 0,017880922 | -0,017880922 |
| 23 :00 :00 | 0 | 0,013392302 | -0,013392302 |
| 23 :15 :00 | 0 | 0,017812484 | -0,017812484 |
| 23 :30 :00 | 0 | 0,018340148 | -0,018340148 |
| 23 :45 :00 | 0 | 0,022715141 | -0,022715141 |

4.1.3 Comparaison et interprétation des résultats

Après avoir appliqué la régression linéaire par la méthode des moindres carrés ordinaires et modèle *ARMAX* sur la production d'électricité éolienne, les résultats obtenus sont jugés acceptables .

La figure suivante présente les variations des valeurs des prédictions. elle sont comparés avec les valeurs réels.

On observant la figure on vois que la régression par *MCO* fournit des valeurs prochent a celles obtenus par le *ARMAX* mais elles sont plus proche de la réalité.

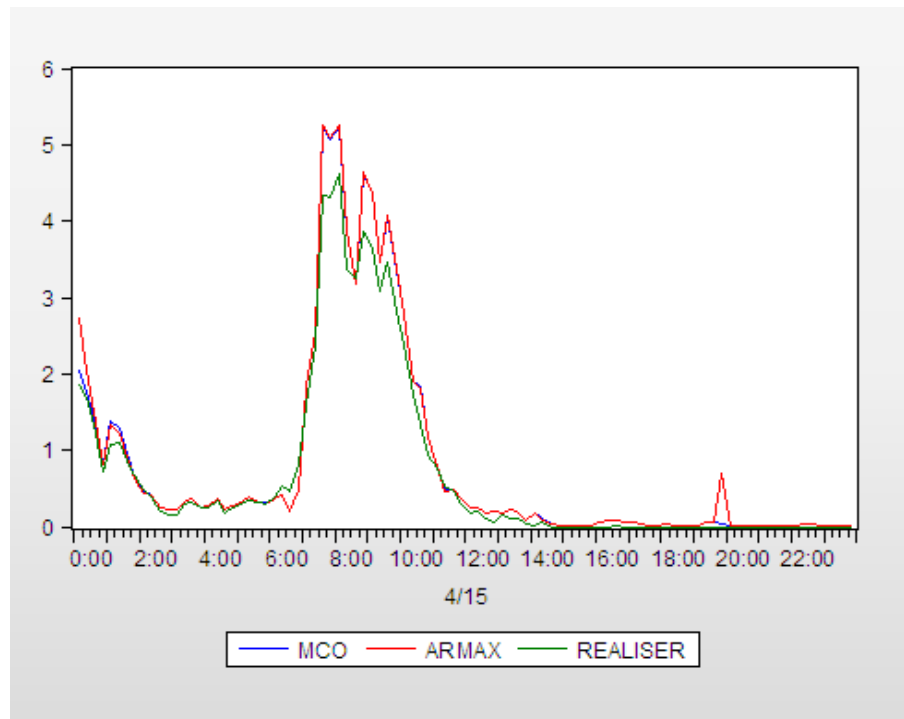


FIGURE 4.15 – Comparaison entre les valeurs réel et les valeurs prédites de la production électrique pour la journée 15 avril 2016

4.2 Etude prévisionnelle de la centrale photovoltaïque de Adrar

Identification

En étude la dépendance de la production d'électricité de type photovoltaïque sur la variable décrit l'irradiation solaire. La série IRA représente l'évolution d'irradiation sur une période allant du 08 avril 2016 jusqu'au 14 avril 2016 avec un pas de temps de 15 minutes. Donc, la série sous étude comportera 315 observation. la série PV représente l'évolution de la production électrique photovoltaïque sur la même période (du 8 avril 2016 jusqu'à 14 avril 2016) avec un pas de temps de 15 minutes.

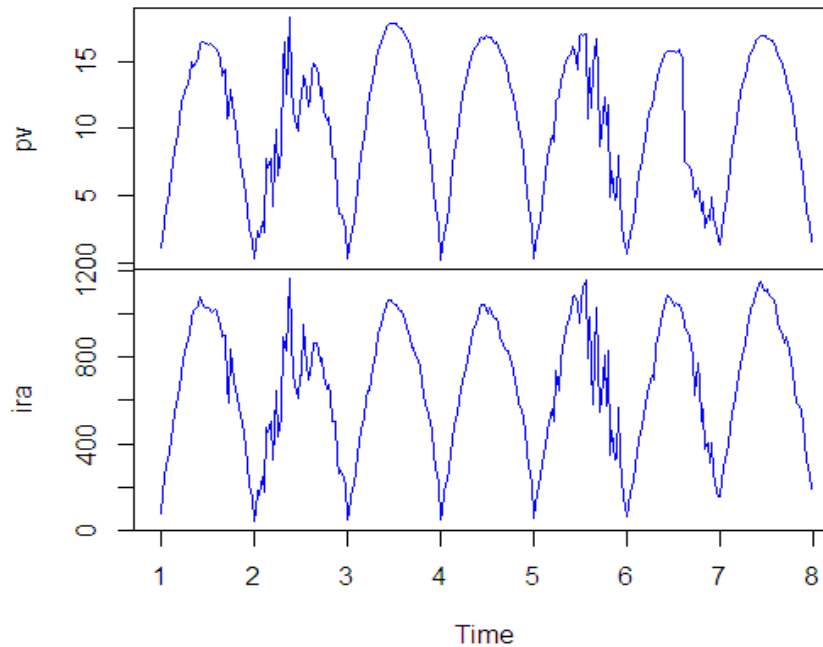


FIGURE 4.16 – Chronogramme de la production électrique et l'irradiation solaire

Le graphique montre que les deux séries ont presque la même allure, se qui montre l'influence de l'irradiation sur la production d'électricité

4.2.1 Régression linéaire par la méthode des moindres carrés

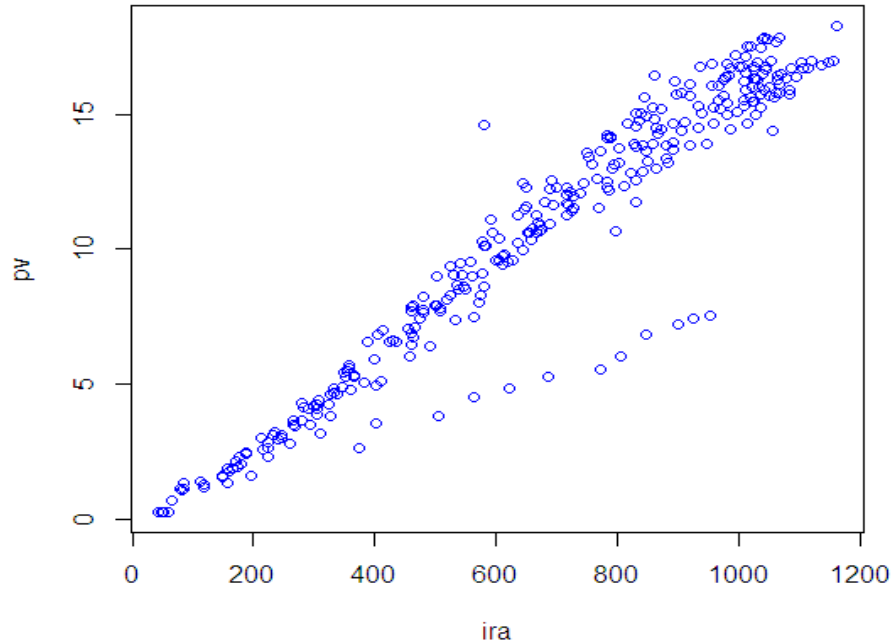


FIGURE 4.17 – Nuage de points de la production en fonction de l’irradiation solaire

D’après le graphe de nuage de point, on constate une forte corrélation entre la production d’électricité et l’irradiation solaire. Cette corrélation peut être modéliser par un modèle linéaire de la forme suivante :

$$pv_t = a + b(ira_t) + \mu_t$$

b) Estimation des paramètres du modèle :

L’estimation des paramètres du modèle est basée sur la méthode des moindres carrés ordinaires (*MCO*), elle est obtenue par la fonction *lm()* fournit par le logiciel *R*.

```

> modpv=lm(pv~ira)
> modpv

Call:
lm(formula = pv ~ ira)

Coefficients:
(Intercept)          ira
   -0.56595         0.01642

```

FIGURE 4.18 – Résultats de l’estimation

Suivant les résultats de l’estimation, le modèle de régression s’écrit comme suit :

$$PV_t = -0.56595 + 0.01642 \times IRA_t$$

c) Test des résidus :

• Graphiques des résidus :

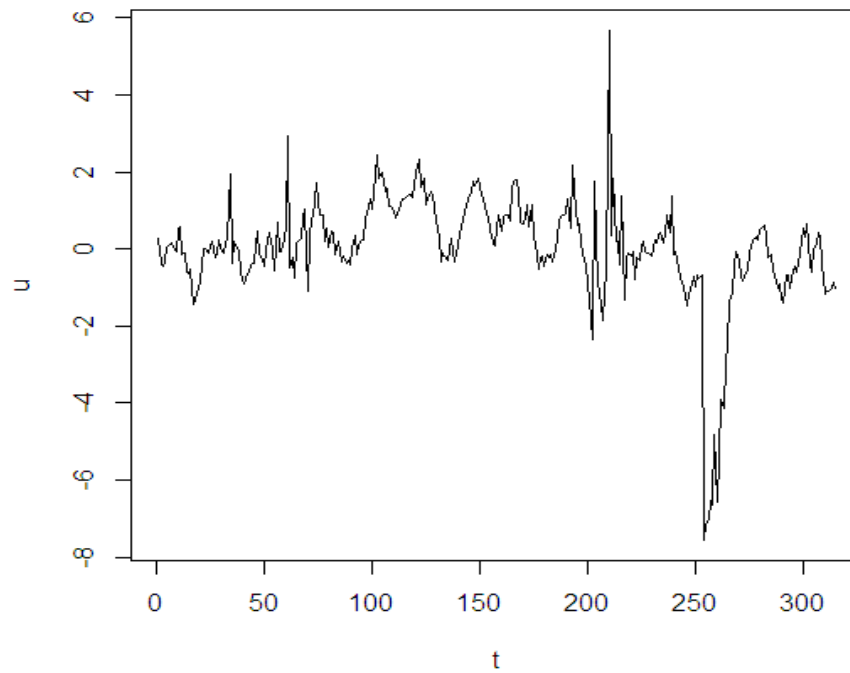


FIGURE 4.19 – graphique des résidus

Calculons la moyenne des résidus

```
> moyen=mean(u)
> moyen
[1] 1.183846e-17
```

FIGURE 4.20 – Résultat de calcul de moyenne des résidus

A partir de la représentation graphique des résidus et le résultat fournit sous R , on en déduit que les résidus sont centrés.

- lag plot des résidus :

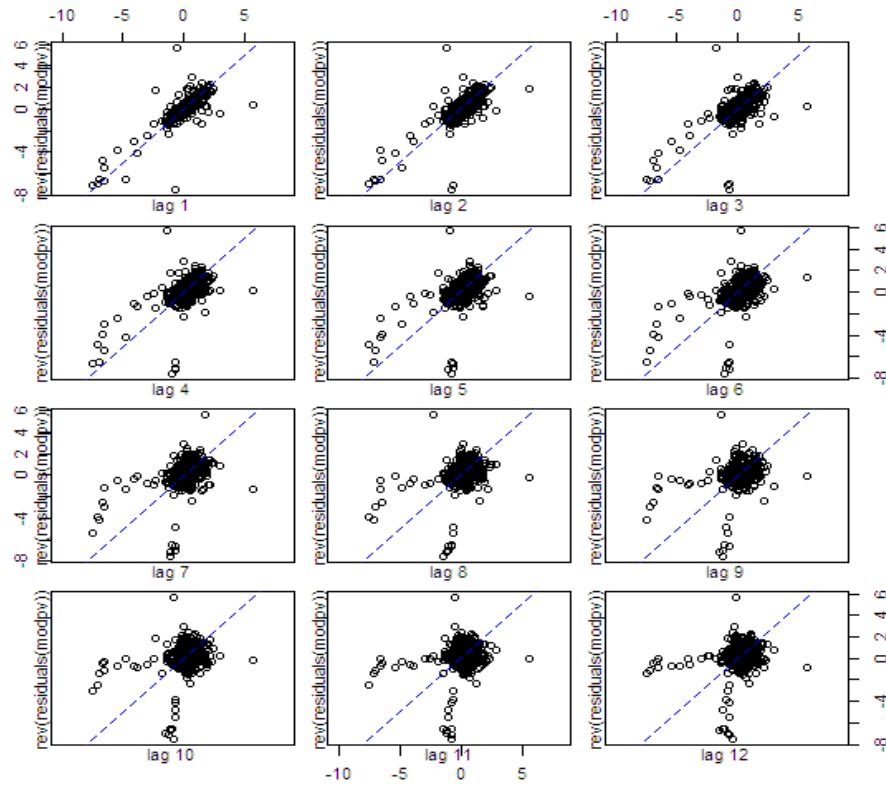


FIGURE 4.21 – lag plot des résidus

les résidus montrent une forte corrélation aux différents retards de 1 jusqu'à 12, donc la présupposé $P3$ (la non corrélation des résidus) ne tient pas mais nous poursuivons la démarche de régression comme si tous les présupposés étaient vérifiés.

- **Test de normalité des résidus :**

La commande suivante fournir le graphe des résidus.

```
plot(modpv, which = 2, col = "red")
```

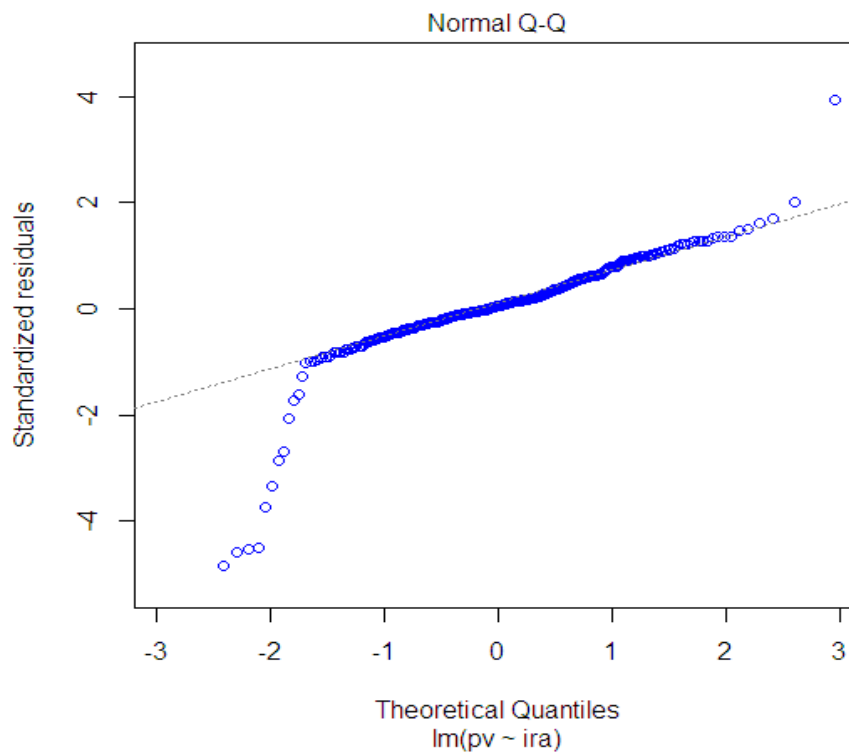


FIGURE 4.22 – Q-Q plot de normalité

La représentation Q-Q plot nous indique qu'effectivement les résidus sont distribués normalement.

d))Significativité du modèle :

Le modèle s'écrit comme suit :

$$pv_t = a + b \times ira_t$$

```

> summary(modpv)

Call:
lm(formula = pv ~ ira)

Residuals:
    Min       1Q   Median       3Q      Max
-7.5450 -0.4509  0.0913  0.7549  5.6663

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.5659462  0.1999889   -2.83  0.00496 **
ira          0.0164172  0.0002709   60.61 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.447 on 313 degrees of freedom
Multiple R-squared:  0.9215,    Adjusted R-squared:  0.9212
F-statistic: 3674 on 1 and 313 DF,  p-value: < 2.2e-16

```

FIGURE 4.23 – L'estimation du modèle

- **Significativité des paramètres :**

1. **Test de la constante :**

$$H_0 : a = 0 \text{ vs } H_1 : a \neq 0$$

$t - stat = -2.83$, $|t - stat| > 1.96$ de plus la $p - value = 0.00496 < 0.05$.
donc H_0 est rejeté $\implies a$ est significativement différent de zero au seuil 5%.

2. **Test de significativité des modèle :**

$$H_0 : b = 0 \text{ vs } H_1 : b \neq 0$$

$t - stat = 60.61$, $|t - stat| > 1.96$ de plus la $p - value = 2 \times 10^{-16} < 0.05$.
donc H_0 est rejeté $\implies b$ est significativement différent de zero au seuil 5%.

De plus le coefficient de détermination $R^2 = 0,92$ donc le modèle est significative, on peut l'écrire comme suit :

$$pv_t = -0.56594 + 0.01641 \times ira_t$$

on peut passé maintenant à l'étape de prévision.

e) **Prévision :**

Nous allons faire des prévision pour une journée : le 15 avril 2016. le tableau suivant représente les valeurs réel et prédit de la production électrique photovoltaïque par ma méthode des moindres carrés ordinaires, ainsi que les écarts correspondants.

| quard'heure | valeur réel | valeur prédit | écart | | | | |
|-------------|-------------|---------------|----------|------------|------|----------|----------|
| 07 :30 :00 | 1,3 | 1,88063 | -0,58063 | | | | |
| 07 :45 :00 | 2,3 | 3,09571 | -0,79571 | | | | |
| 08 :00 :00 | 3,5 | 3,91671 | -0,41671 | | | | |
| 08 :15 :00 | 5,1 | 5,88711 | -0,78711 | | | | |
| 08 :30 :00 | 6,3 | 6,59317 | -0,29317 | | | | |
| 08 :45 :00 | 3,7 | 3,17781 | 0,52219 | | | | |
| 09 :00 :00 | 9 | 7,90677 | 1,09323 | | | | |
| 09 :15 :00 | 9,7 | 9,76223 | -0,06223 | | | | |
| 09 :30 :00 | 10,7 | 10,66533 | 0,03467 | | | | |
| 09 :45 :00 | 11,5 | 11,58485 | -0,08485 | 15 :30 :00 | 12,4 | 12,34017 | 0,05983 |
| 10 :00 :00 | 12 | 12,47153 | -0,47153 | 15 :45 :00 | 13,3 | 14,06427 | -0,76427 |
| 10 :15 :00 | 13,3 | 12,84919 | 0,45081 | 16 :00 :00 | 13,3 | 13,11191 | 0,18809 |
| 10 :30 :00 | 13,8 | 13,93291 | -0,13291 | 16 :15 :00 | 11,6 | 11,79831 | -0,19831 |
| 10 :45 :00 | 14,3 | 14,58971 | -0,28971 | 16 :30 :00 | 10,6 | 10,46829 | 0,13171 |
| 11 :00 :00 | 14,8 | 15,55849 | -0,75849 | 16 :45 :00 | 9,9 | 9,72939 | 0,17061 |
| 11 :15 :00 | 15,1 | 16,19887 | -1,09887 | 17 :00 :00 | 8,7 | 9,31889 | -0,61889 |
| 11 :30 :00 | 15,4 | 17,08555 | -1,68555 | 17 :15 :00 | 6,9 | 7,52911 | -0,62911 |
| 11 :45 :00 | 15,7 | 17,54531 | -1,84531 | 17 :30 :00 | 6,2 | 7,54553 | -1,34553 |
| 12 :00 :00 | 16,1 | 18,00507 | -1,90507 | 17 :45 :00 | 4,5 | 5,04969 | -0,54969 |
| 12 :15 :00 | 16,3 | 18,33347 | -2,03347 | 18 :00 :00 | 3,5 | 4,42573 | -0,92573 |
| 12 :30 :00 | 16,6 | 18,61261 | -2,01261 | 18 :15 :00 | 2,2 | 2,40607 | -0,20607 |
| 12 :45 :00 | 17,8 | 20,30387 | -2,50387 | 18 :30 :00 | 1,6 | 2,07767 | -0,47767 |
| 13 :00 :00 | 14,7 | 17,41395 | -2,71395 | | | | |
| 13 :15 :00 | 11 | 13,11191 | -2,11191 | | | | |
| 13 :30 :00 | 11 | 11,76547 | -0,76547 | | | | |
| 13 :45 :00 | 10 | 8,43221 | 1,56779 | | | | |
| 14 :00 :00 | 16,4 | 17,16765 | -0,76765 | | | | |
| 14 :15 :00 | 15,1 | 16,62579 | -1,52579 | | | | |
| 14 :30 :00 | 15,8 | 16,10035 | -0,30035 | | | | |
| 14 :45 :00 | 15,7 | 15,88689 | -0,18689 | | | | |
| 15 :00 :00 | 13,7 | 14,03143 | -0,33143 | | | | |
| 15 :15 :00 | 14,1 | 13,81797 | 0,28203 | | | | |

4.2.2 Prévision par le modèle ARMAX

On va modéliser l'influence de la série $IRA(t)$ sur la série $PV(t)$ par un modèle *ARMAX*.

a) **Identification de la série des résidus par la méthode des moindres carrés ordinaires MCO :**

Le modèle que nous avons retenu pour expliquer la production d'électricité par la variable irradiation est :

$$PV_t = -0.56595 + 0.01642 \times IRA_t + \mu_t$$

> `modpv = lm(PV ~ IRA)`

> `u = ts(residuals(modpv), start = 1, frequency = 45)`

les résidus $\hat{\mu}_t$ montraient une autocorrélation significative, nous allons identifier leur modèle, mais tout d'abord on doit les récupérer et les transformer en série temporelle

b) **modélisation des résidus par un modèle ARIMA :**

Utilisons la procédure d'identification automatique `auto.arima()` de package `forecast` pour estimer le modèle le plus adéquat pour la série des résidus.

> `modelpv = auto.arima(u, max.p = 5, max.q = 4, max.P = 1, approximation = FALSE)`

le modèle $ARIMA(2, 1, 1)$ est sélectionné pour la série des résidus μ_t . on passe aux tests de validation pour valider le modèle.

• **Test de student des paramètres**

utilisons la commande :

> `t - stat(modelpv)`

on aura les résultats :

| Paramètre | p-value |
|-----------|----------|
| ar(1) | 0.000000 |
| ar(2) | 0.000001 |
| ma(1) | 0.000000 |

Pour tous les paramètres la $p - value$ est inférieure à 0.05, donc ces paramètres sont significative au seuil 5%.

• **Test de Ljung-Box sur les résidus :**

```

> t(Box.test.2(residuals(modelpv),llag,type="Ljung-Box",decim=2,fitdf=3))
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12]
Retard  4.0 8.00 12.00 16.00 20.00 24.00 28.00 32.00 36.00 40.00 44.00 48.00
p-value 0.9 0.11 0.17 0.27 0.49 0.63 0.72 0.84 0.91 0.96 0.53 0.67
      [,13] [,14] [,15] [,16] [,17] [,18] [,19] [,20] [,21] [,22] [,23] [,24]
Retard 52.00 56.00 60.00 64.0 68.00 72.00 76.00 80.00 84.00 88.00 92.00 96.00
p-value 0.32 0.38 0.49 0.6 0.71 0.81 0.88 0.91 0.95 0.97 0.98 0.99
      [,25] [,26] [,27] [,28] [,29] [,30] [,31] [,32] [,33] [,34] [,35]
Retard 100.00 104 108 112 116 120 124 128 132 136 140
p-value 0.99 1 1 1 1 1 1 1 1 1 1

```

FIGURE 4.24 – Résultats du Test de Ljung Box

$\forall p = 1 \dots 140$, la p -value pour chaque retard est supérieure à 0.05, donc la série des résidus est un bruit blanc.

Nous pouvons passer à l'estimation du modèle ARMAX.

c) Estimation du modèle ARMAX :

Afin d'estimer le modèle *ARMAX*, on doit définir la variable à expliquer (la série PV) et la variable explicative (la série IRA).

pour cela en utilise la commande suivante :

```

> modxpv = Arima(PV, order = c(1, 1, 1), xreg = IRA)

```

```

> modxpv=Arima(pv,order=c(1,1,1),xreg=ira)
> modxpv
Series: pv
ARIMA(1,1,1)

Coefficients:
      ar1      ma1      V1
    -0.6283  0.3912  0.0138
s.e.   0.1381  0.1656  0.0005

sigma^2 estimated as 0.6242:  log likelihood=-370.09
AIC=748.18  AICc=748.31  BIC=763.17

```

FIGURE 4.25 – Résultat d'estimation

- **Test de student des paramètres :**

avec la commande :

```
> t - stat(modxpv)
```

on aura les résultats suivants :

| Paramètre | p-value |
|-----------|----------|
| ar(1) | 0.000005 |
| ma(1) | 0.018184 |
| IRA | 0.000000 |

On voit que tous les paramètres du modèle sont significativement différent de zéros au seuile 5%(la p-value pour chaque paramètre est inférieure a 0.05).

- **Test de Ljung-Box sur les résidus :**

```

> t(Box.test.2(residuals(modxpv),llag,type="Ljung-Box",decim=2,fitdf=3))
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12]
Retard  4.0 8.00 12.00 16.00 20.00 24.00 28.00 32.00 36.00 40.00 44.00 48.00
p-value 0.9 0.19 0.26 0.36 0.53 0.68 0.78 0.88 0.87 0.92 0.36 0.45
      [,13] [,14] [,15] [,16] [,17] [,18] [,19] [,20] [,21] [,22] [,23] [,24]
Retard 52.00 56.00 60.00 64.00 68.00 72.00 76.00 80.00 84.00 88.00 92.00 96.00
p-value 0.17 0.24 0.32 0.37 0.49 0.61 0.71 0.78 0.85 0.89 0.89 0.93
      [,25] [,26] [,27] [,28] [,29] [,30] [,31] [,32] [,33] [,34] [,35]
Retard 100.00 104.00 108.00 112.00 116.00 120 124 128 132 136 140
p-value 0.94 0.97 0.98 0.99 0.99 1 1 1 1 1 1

```

FIGURE 4.26 –

D'après les résultats des tests, le modèle est validé.
Le modèle obtenu finalement est :

$$PV_t = -0.56595 + 0.01642 \times IRA_t + \mu_t, \quad t = 1 \dots 315$$

avec $\mu_t = \frac{1-0.3912}{1+0.6289B} z_t$.

d) Prévision par le modèle ARMAX :

La prévision est faite par la commande suivante :

```
> prevpv = forecaste(modxpv, level = c(95, 90), fan = FALSE, xreg = nwira)
```

Nous allons faire des prévisions pour une journée : le 15 avril 2016 de (07 : 30 : 00) à (18 : 30 : 00) Le tableau suivant représente les valeurs réels et prédites de la production électrique par le modèle ARMAX ainsi que les écarts correspondants.

| Quard'heure | Valeurs réelles | Valeurs prédites | Ecart |
|-------------|-----------------|------------------|----------|
| 07 :30 :00 | 1,3 | 1,038534 | 0,261466 |
| 07 :45 :00 | 2,3 | 2,012001 | 0,287999 |
| 08 :00 :00 | 3,5 | 2,728445 | 0,771555 |
| 08 :15 :00 | 5,1 | 4,362266 | 0,737734 |
| 08 :30 :00 | 6,3 | 4,965249 | 1,334751 |
| 08 :45 :00 | 3,7 | 2,09548 | 1,60452 |
| 09 :00 :00 | 9 | 6,063702 | 2,936298 |
| 09 :15 :00 | 9,7 | 7,61618 | 2,08382 |
| 09 :30 :00 | 10,7 | 8,374898 | 2,325102 |
| 09 :45 :00 | 11,5 | 9,144547 | 2,355453 |
| 10 :00 :00 | 12 | 9,888449 | 2,111551 |
| 10 :15 :00 | 13,3 | 10,204572 | 3,095428 |
| 10 :30 :00 | 13,8 | 11,113218 | 2,686782 |
| 10 :45 :00 | 14,3 | 11,663577 | 2,636423 |
| 11 :00 :00 | 14,8 | 12,475716 | 2,324284 |
| 11 :15 :00 | 15,1 | 13,012416 | 2,087584 |
| 11 :30 :00 | 15,4 | 13,755673 | 1,644327 |
| 11 :45 :00 | 15,7 | 14,141018 | 1,558982 |
| 12 :00 :00 | 16,1 | 14,526405 | 1,573595 |
| 12 :15 :00 | 16,3 | 14,80166 | 1,49834 |
| 12 :30 :00 | 16,6 | 15,035641 | 1,564359 |
| 12 :45 :00 | 17,8 | 16,453252 | 1,346748 |
| 13 :00 :00 | 14,7 | 14,030923 | 0,669077 |
| 13 :15 :00 | 11 | 10,42495 | 0,57505 |
| 13 :30 :00 | 11 | 9,296365 | 1,703635 |
| 13 :45 :00 | 10 | 6,502425 | 3,497575 |
| 14 :00 :00 | 16,4 | 13,824473 | 2,575527 |
| 14 :15 :00 | 15,1 | 13,370286 | 1,729714 |
| 14 :30 :00 | 15,8 | 12,929862 | 2,870138 |
| 14 :45 :00 | 15,7 | 12,75094 | 2,94906 |
| 15 :00 :00 | 13,7 | 11,195693 | 2,504307 |
| 15 :15 :00 | 14,1 | 11,016771 | 3,083229 |
| 15 :30 :00 | 12,4 | 9,778078 | 2,621922 |
| 15 :45 :00 | 13,3 | 11,223219 | 2,076781 |

| | | | |
|------------|------|-----------|----------|
| 16 :00 :00 | 13,3 | 10,424951 | 2,875049 |
| 16 :15 :00 | 11,6 | 9,323891 | 2,276109 |
| 16 :30 :00 | 10,6 | 8,209068 | 2,390932 |
| 16 :45 :00 | 9,9 | 7,589722 | 2,310278 |
| 17 :00 :00 | 8,7 | 7,245641 | 1,454359 |
| 17 :15 :00 | 6,9 | 5,745447 | 1,154553 |
| 17 :30 :00 | 6,2 | 5,75921 | 0,44079 |
| 17 :45 :00 | 4,5 | 3,667196 | 0,832804 |
| 18 :00 :00 | 3,5 | 3,144193 | 0,355807 |
| 18 :15 :00 | 2,2 | 1,451314 | 0,748686 |
| 18 :30 :00 | 1,6 | 1,176049 | 0,423951 |

4.2.3 Comparaison et interprétation des résultats

Après avoir appliqué la régression linéaire par la méthode des moindres carrés ordinaires et le modèle *ARMAX* sur la production d'électricité photovoltaïque, les résultats obtenus sont jugés acceptables .

La figure suivante présente les variations des valeurs des prédictions. elle sont comparés avec les valeurs réels.

On observant la figure, on voit que les valeurs obtenus par la régression par *MCO* sont plus proches de la réalité que celles obtenus par le modèle *ARMAX* .

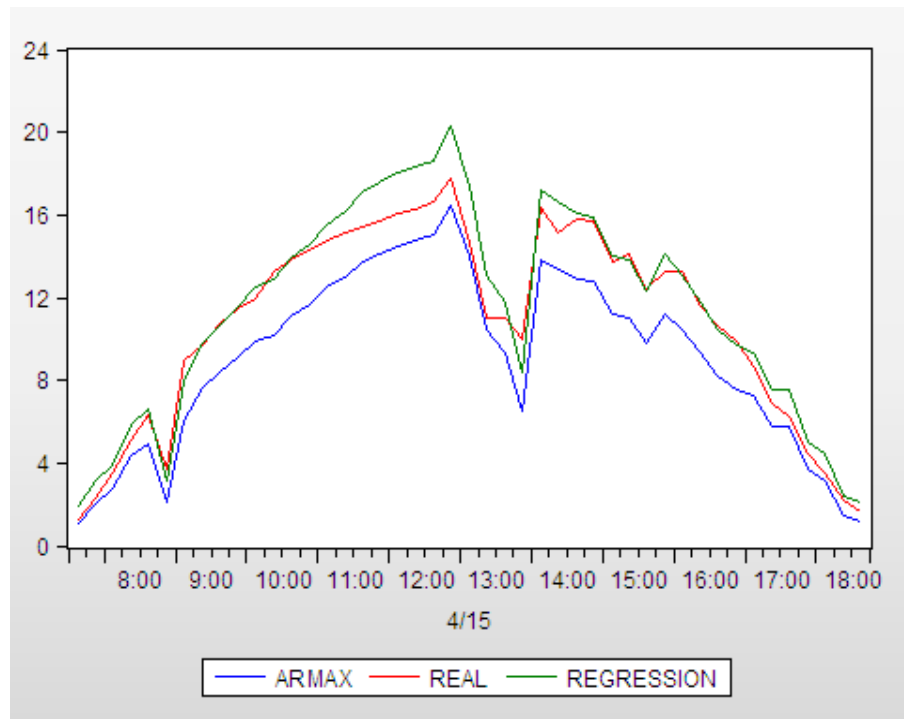


FIGURE 4.27 – Comparaison entre les valeurs réels et les valeurs prédites de la production électrique pour la journée 15 avril 2016

Conclusion

Dans ce chapitre, deux méthodes de prévision sont appliqués sur la production électrique, la régression par la méthode des moindres carrés ordinaires et la méthodes *ARMAX*.

Ces méthodes présentent des résultats acceptables car les écarts entre les prévisions et les réalisations sont faibles.

Conclusion générale

La transition énergétique est devenue incontournable pour répondre aux besoins croissantes en énergie, atténuer les risques liés à la volatilité des prix du pétrole et la tendance à la baisse des réserves des énergies fossiles qui affectent les marchés. Les énergies renouvelables contribuent à répondre aux besoins énergétique actuels pour soutenir une croissance durable.

Au cours de ce travail, nous avons traité le problème de prédiction de la production de ces énergies renouvelables par des séries temporelles.

Dans cette étude, nous avons utilisé les données journalières par quart d'heure de la production d'électricité de type ENR's et de variables climatiques (vitesse de vent du centrale éolienne de Kaberten et irradiance (nébulosité) du pôle d'Adrar) pour tester des modèles de prévision de cette production à court terme.

Nous avons commencé par introduire les séries temporelles avec leurs caractéristiques. Dans un second plan, nous avons présenté les différents processus de modélisation par la suite décrire les méthodes de résolution pour avoir une bonne prédiction.

Nous avons effectué des prévisions des échantillons avec deux types de méthodes : la méthode des moindres carrés et un ARMAX avec variables exogènes climatiques.

Nous avons proposé des prédictions sur horizons temporels de 24h (1 jour). La méthode des moindres carrés se montrent généralement plus performants que Les modèles ARMAX. La qualité des prévisions est tout de même globalement satisfaisante, avec l'erreur absolue moyenne proche de zero dans les deux modèles.

La prévision de la production d'électricité de type ENR's dépend des changements climatiques pour cette raison, les chercheurs n'ont pas pu trouver un modèle général pour la prédire d'où chaque série dépend de l'étude faite sur ses valeurs, pour tirer les caractéristiques qui aident à mieux la modéliser.

On a essayé d'appliquer des méthodes de prédiction pour la production électrique éolienne et photovoltaïque. Il reste encore du travail dans ce contexte afin d'améliorer

les résultats obtenus. Cette amélioration peut être obtenue en développant les aspects suivants :

- Pendant une année le climat change de l'été vers l'hiver et vice-versa. Une classification des données peut se faire pour que la modélisation se réalise par la suite selon les saisons.
- élargir l'historique pour avoir un résultat plus proche de la réalité

Nous constatons qu'il n'est plus difficile de parler d'un après-pétrole qui semble lointain dans les dernières années.

La maîtrise de la technologie de ENR's est la seule garantie pour mettre en confiance l'utilisateur potentiel, encore il faut que la ressource humaine spécialisée soit formée.

Bibliographie

- [1] Can Wan , Juan Zhao , Youghua Song , Zhao Xu , Jin Lin , Zechun Hu, Photovoltaic and solar power forecasting for smart grid energy management, CSEE journal of power and energie systemes Vol 1 , (2015),
- [2] Julien Najac, La prévision de production éolienne et photovoltaïque à EDF, Siminaire In'Tech,LNRIA,Grnoble, (27/09/2012),
- [3] Said Bentourba , Aek Slimani , Med Sghir Boucherit , Messaoud Hammouda, L'énergie renouvelable en Algérie et l'impacte sur l'environnement, , ,
- [4] Emmanuel Neau, Photovoltaïques : des prévisions météo aux prévisions de production, , (15/11/2015),
- [5] Mehdi Merad Boudia , Injection d'électricité produit par les énergies renouvelables dans le reseau électrique, Mémoir majistter université Abou Beker Belkaid, Tlemcen, ,
- [6] J.J Daudin , C. DUBY , S. Robin , P. Trécourt, Analyse des séries chronologiques, , mai (1996),
- [7] Agnès Langroux, Séries chronologiques , Université de Toulouse le mirail, (2011-2012),
- [8] D. Eilstein , H. Prouvost , P. Saviuc , C. Declercq , A. Lefranc , S. Cassadou , A. Zeghnoun , A. Tertre , P. Quénel , M. D'Helf , C. Nunes , P. Fabre , L. Filleul , L. Pascal , L. Pascal , J.F. Jusot , B Chardon , S. Médina , Séries temporelles et modèles de Regression Application à l'analyse des associations à court terme entre la pollution atmosphérique et la santé, Département santé-environnement, Institut de veille sanitaire, France (2003),
- [9] R. Bourbounnais , M. Tirraza, Analyse des séries temporelles en économie, Edition Dunod, Paris (2014),
- [10] Yves Aragon , Séries temporelles avec R, méthodes et cas, Edition Springer, (2011),