



**Développement d'un modèle de prévision à court-terme de la
puissance produite à la centrale solaire photovoltaïque de
Nagréongo (30 MWc) dans le réseau national interconnecté du
Burkina Faso.**

MEMOIRE POUR L'OBTENTION DU DIPLOME D'INGENIEUR 2IE AVEC GRADE DE

**MASTER
SPECIALITE : GENIE ELECTRIQUE ET ENERGETIQUE**

Présenté et soutenu publiquement le 28/01/2025 par

Malika FOFANA (20190001)

Directeur de mémoire : Pr Y. Moussa SORO, Chef du département Génie Electrique, Energétique et Industriel (GEEI).

Encadrant 2iE : Dr ABOUBAKAR GOMNA, Enseignant-Chercheur.

Maître de stage : Ing. Yannick KABORE, Directeur des Opération (GreenYellow).

Structures d'accueil du stage : GreenYellow

Jury d'évaluation du stage :

Président : Dr. Kokouvi Edem N'TSOUKPOE

Membres et correcteurs : Dr. Hector TETE
Ing. Sédi AGBOKOU

Promotion [2024-2025]

REMERCIEMENTS

Je tiens à exprimer ma profonde gratitude envers l'Institut international d'Ingénierie de l'Eau et de l'Environnement pour la qualité de l'enseignement reçu et pour avoir contribué à l'aboutissement de ce travail de manière si précieuse.

Je remercie chaleureusement mon directeur de mémoire, le Professeur Moussa Y. SORO, pour sa guidance, ses conseils éclairés et son soutien constant tout au long de ce projet.

Un grand merci à mon encadrant interne, le Dr ABOUBAKAR GOMNA, pour ses conseils avisés et sa bienveillance. Ses remarques pertinentes ont grandement enrichi mes réflexions et m'ont permis de perfectionner mon travail. Sa disponibilité et son expertise ont été des atouts essentiels dans la réussite de ce travail.

Je tiens également à exprimer ma reconnaissance envers la structure GreenYellow, qui m'a offerte l'opportunité de réaliser ce stage et d'appliquer mes connaissances dans un contexte professionnel enrichissant.

Enfin, je remercie sincèrement mon maître de stage, M. Yannick KABORE, pour son accueil, son encadrement et son accompagnement tout au long de cette expérience. Sa disponibilité et ses conseils ont été d'une grande utilité pour mener à bien ce projet.

DEDICACES

A DIEU, le tout puissant, le tout miséricordieux,

Seigneur de l'univers, c'est à toi que revient toute la gloire, C'est par Ta grâce infinie que j'ai trouvé la paix dans les moments d'incertitude et la force dans les moments de faiblesse. Ce travail est avant tout une bénédiction de Ta part, un fruit de Ta guidance divine. Chaque réussite, chaque pas franchi, est une preuve de Ton amour et de Ta générosité. Je Te remercie pour chaque souffle, chaque épreuve et chaque opportunité que Tu m'as offerte. Que Ta lumière continue de m'éclairer, et que je puisse toujours marcher sur le chemin de la vérité, de la paix et de la bonté.

A mes parents Hadjirata et Zacharia FOFANA,

Vous avez été mon soutien inébranlable, mes premiers rêves. Votre amour et vos sacrifices ont été la force qui m'a permis d'aller de l'avant. Ce travail est le reflet de tout ce que vous m'avez donné durant ces cinq dernières années et je vous le dedie avec une gratitude infinie.

A mes frères et sœurs Abdallah, Djalila, Khaled et Khéïra,

Votre soutien et vos encouragements ont été ma lumière dans les moments de doute. Ce travail est tout aussi le vôtre que le mien, et je vous le dédie avec tout mon amour.

A mes amis,

Vous avez été mon réconfort et ma source d'énergie, je vous remercie pour votre présence constante.

CITATION

« Je crois au travail acharné et à la chance, et que le premier mène souvent
à la seconde. »

Thomas Jefferson

RESUME

Face aux contraintes énergétiques et environnementales comme le réchauffement climatique et l'épuisement des ressources fossiles, le solaire s'intègre de plus en plus dans les mix énergétiques des pays du monde entier. Au Burkina Faso, le développement rapide du solaire photovoltaïque, et son défi général d'intermittence, constituent de plus en plus une menace pour la stabilité du réseau. La prévision des centrales solaires devient dans ces conditions de plus en plus critique.

Ce mémoire s'inscrit dans le cadre du développement d'un modèle de prévision d'énergie à court terme d'une centrale solaire PV pour anticiper les éventuelles variations et optimiser la gestion de l'énergie. L'objectif principal était d'optimiser la gestion de la puissance injectée dans le réseau national. Initialement, des modèles classiques comme ARIMA et VECM ont été testés, mais leurs performances se sont avérées limitées en raison de la non-linéarité et de l'hétéroscédasticité des données.

Pour surmonter ces limites, nous avons adopté un modèle de régression quantile, robuste face aux valeurs extrêmes, un aspect crucial dans le contexte des énergies renouvelables. Ce modèle s'appuie sur plusieurs variables explicatives pour fournir des prévisions fiables, les performances ont été évaluées à l'aide de métriques de performances telles que l'erreur quadratique moyenne (MSE), l'erreur absolue moyenne (MAE) et l'erreur médiane absolue (MedAE).

En complément, un modèle basé sur les réseaux de neurones récurrents (LSTM) a été implémenté pour capturer des dépendances temporelles complexes. Les données, échantillonnées toutes les cinq (05) minutes et filtrées pour les heures diurnes (6h-18h), ont été organisées avec une fenêtre de trois (03) jours. Les hyperparamètres ont été optimisés via Keras Turner (une bibliothèque pour l'optimisation des performances du modèle) avec des techniques telles early stopping et dropout afin d'éviter le surapprentissage.

L'étude nous a permis d'obtenir des prévisions d'une précision de 96% lors de l'entraînement du modèle et 98% lors du test avec 3,10% d'erreur pendant la validation du modèle.

Enfin, une interface interactive a été développée afin de permettre la comparaison entre les prévisions et les valeurs réelles saisies. Cette interface offre une évaluation continue et permet d'intégrer de nouvelles données afin d'améliorer le modèle à long terme.

Mots Clés :

1 – Prévision de puissance

Développement d'un modèle de prévision à court terme de la puissance produite à la centrale solaire photovoltaïque de Nagréongo (30 MWc) dans le réseau national interconnecté du Burkina Faso.

2 – Centrale solaire photovoltaïque

3 – Modèle à court terme

4 – Réseau national interconnecté

5- Production énergétique solaire

ABSTRACT

In response to climate change and the depletion of fossil resources, solar energy is increasingly

v

integrated into national energy mixes. In Burkina Faso, its rapid expansion presents an intermittency challenge that threatens grid stability. To address this, a short-term forecasting model has been developed to anticipate fluctuations and optimize energy management.

We developed a short-term power production prediction model for the Ngréongo photovoltaic solar plant (30 MWc) in Burkina Faso, with the main objective of optimizing the management of power injected into national grid. Initially, classic models like ARIMA and VECM were tested, but their performance was limited due to non-linearity and heteroscedasticity of the data. To overcome these limitations, a quantile regression model was adopted, which is robust to extreme values a crucial aspect in the context of renewable energies. This model relies on several explanatory variables to provide reliable forecasts. Performance was evaluated using metrics such as Mean Squared Error (MSE), Mean Absolute Error (MAE) and Median Absolute Error (MedAE).

Additionally, a model based on Recurrent Neural Networks (LSTM) was implemented to capture complex temporal dependencies. The data, sampled every five (05) minutes and filtered for sunlight hours (6 AM to 6PM), was structured with a 3-day window. Hyperparameters were optimized using Keras Turner, with techniques such as Early stopping and dropout to prevent overfitting. Keras Tuner is a library that optimizes the search for the best hyperparameter combinations to enhance model performance. It allows exploring different values for parameters such as the number of neurons, learning rate, etc. The study resulted with an accuracy of 96% during training, 98% during testing and an error rate of approximately 3.10% during model validation.

Finally, an interactive interface was developed to enable comparisons between predictions and entered real values. This interface provides continuous evaluation and allow the integration of new data, to improve the model in long term.

Key words:

-
- 1 – Power forecasting**
 - 2 – Photovoltaic solar power plant**
 - 3 – Short-term model**
 - 4 – Interconnected power grid**
 - 5 – Solar energy production**

LISTE DES ABBREVIATIONS

2iE : Institut International d'Ingénierie de l'Eau et de l'Environnement

ACF : fonction d'autocorrélation [autocorrelation function].

ADF: test de Dickey-Fuller augmenté [augmented Dickey-Fuller].

AIC : critère d'information d'Akaike [Akaike information criterion].

ARIMA : modèle autoregressif intégrant la moyenne mobile [autoregressive integrated moving average].

BIC: critère d'information bayésien [bayesian information criterion].

KPSS: Kwiatkowski-Phillips-Schmidt-Shin

LTSM : mémoire à long et à court terme [long short-term memory].

MAE : erreur absolue moyenne [mean absolute error].

MedAE : erreur absolue médiane [median absolute error].

MSE : erreur quadratique moyenne [mean squared error].

PACF : fonction partielle d'autocorrélation [partial autocorrelation function].

RMSE : racine de l'erreur quadratique moyenne [root mean squared error].

SPES : Société de Production d'Energie Solaire

VAR : variance [vector autoregressive].

VAR : modèle autoregressif vectoriel [vector autoregressive model].

TABLE DE MATIERE

REMERCIEMENTS	i
DEDICACES	ii
RESUME	iv
LISTE DES ABBREVIATIONS.....	vii
<i>LISTE DE TABLEAUX</i>	4
<i>LISTE DE FIGURES</i>	5
<i>LISTE DES ANNEXES</i>	6
<i>Introduction</i>	7
<i>Contexte et justification</i>	7
<i>OBJECTIF DU TRAVAIL</i>	8
<i>ORGANISATION DU TRAVAIL</i>	8
<i>PRESENTATION DE LA STRUCTURE D'ACCUEIL</i>	8
<i>I. Types de prévision</i>	11
<i>II. Introduction au Machine Learning</i>	11
II-1 L'apprentissage supervisé (Supervised Learning)	12
II.2 L'apprentissage non supervisée (Unsupervised Learning)	12
II.3 L'APPRENTISSAGE PAR RENFORCEMENT (REINFORCEMENT LEARNING)	14
<i>III. APPROCHES DE PREDICTION</i>	14
III.1 L'APPROCHE STATISTIQUE	14
III.2 LA METHODE DE L'INTELLIGENCE ARTIFICIELLE.....	15
<i>IV. Logiciels utilisés en prévision solaire</i>	16
IV.1 LE LOGICIEL PYTHON.....	16
IV.2 LE LOGICIEL R	17
IV.3 LE CHOIX DU LOGICIEL A UTILISER.....	17
<i>V. METHODES UTILISES</i>	17

Développement d'un modèle de prévision à court terme de la puissance produite à la centrale solaire photovoltaïque de Nagréongo (30 MWc) dans le réseau national interconnecté du Burkina Faso.

VI. Généralités sur les séries temporelles 18

VI.1 QU'EST-CE QU'UNE SERIE TEMPORELLE ? 18

VI.2 ASPECTS FONDAMENTAUX D'UNE SERIE TEMPORELLE 18

PARTIE 2 : MODELISATION DE LA PRODUCTION DE LA CENTRALE SOLAIRE.. 19

I- Préparation des données 19

I.1 NETTOYAGE DES DONNEES 19

I.2 FILTRAGE DES DONNEES 20

I.3 STRUCTURATION DES VARIABLES 20

I.4 DIVISION DES DONNEES 22

II- Modélisation avec ARIMA..... 23

II.1 INTRODUCTION AU MODELE ARIMA..... 23

II.1.1 Concepts fondamentaux..... 23

II.1.2 Conditions d'application..... 23

II.2 IDENTIFICATION DU MODELE 24

II.2.1 Analyse des données 24

II.2.2 Test de stationnarité 27

II.2.3 ACF et PACF 29

II.2.4 Remarque 31

II.3 TRAITEMENT DE LA STATIONNARITE..... 32

II.4 ESTIMATION DU MODELE 33

II.3.1 Identification des paramètres (P, D, Q)..... 34

II.3.2 Critères d'information..... 34

II.3.2 Vérification du modèle 34

III- Modélisation avec la régression quantile 36

III.1 INTRODUCTION A LA REGRESSION QUANTILE 36

III.2 METHODOLOGIE 37

IV- Modélisation avec le modele long short-term memory (LSTM) 38

IV.1 INTRODUCTION AU MODELE LSTM..... 38

IV.2 METHODOLOGIE 38

IV.3 DEFINITION DU MODELE..... 39

IV.4 PREPARATION DES DONNEES POUR L'ENTRAINEMENT..... 40

IV.5 MODELE LSTM : PARAMETRES ET OPTIMISATION 40

PARTIE 3 : DEVELOPPEMENT DES MODELES DE PREDICTIONS DE LA PUISSANCE DELIVREE PAR LA CENTRALE 41

Développement d'un modèle de prévision à court terme de la puissance produite à la centrale solaire photovoltaïque de Nagréongo (30 MWc) dans le réseau national interconnecté du Burkina Faso.

I- Prévion avec ARIMA.....	41
II- Prévion avec Régression quantile	43
II.1 PREDICTION QUANTILE DE LA MEDIANE (QUANTILE 0,5)	43
II.2 PREVISION REGRESSION QUANTILE (QUANTILE 0,1)	44
II.3 PREVISION REGRESSION QUANTILE (QUANTILE 0,9).....	45
II.4 ANALYSE DES RESIDUS	46
II.5 METRIQUES DE PERFORMANCES	49
III- Prévion avec LSTM	51
III.1 ANALYSE DES PERFORMANCES.....	51
III.2 ANALYSE AVEC LES METRIQUES DE PERFORMANCES.....	53
PARTIE 4 : VALIDATION DU MODELE.....	56
I. Prévion sur de nouvelles données.....	56
II. Développement d'une interface de prévision	56
II.1 CONCEPTION DE L'INTERFACE	56
II.2 ETAPES DE VALIDATION DE L'INTERFACE	58
II.3 AMELIORATION CONTINUE	59
Conclusion partielle	60
CONCLUSION.....	61
RECOMMANDATIONS.....	63
BIBLIOGRAPHIE.....	64

LISTE DE TABLEAUX

Tableau I: Résultats du test de Dickey-Fuller	28
Tableau II : Résultats du test de KPSS.....	28
Tableau III : Test de KPSS après différenciation.....	33
Tableau IV : Test de ADF après différenciation.	33
Tableau V : Métriques de performances du modèle LSTM.....	54

LISTE DE FIGURES

Figure 1: Organigramme de SPES	10
Figure 2: Types de prévisions et leurs objectifs	11
Figure 3: Processus de regroupement (clustering) en apprentissage non supervisé	13
Figure 4: Différence entre Apprentissage supervisé et non supervisé	13
Figure 5: Etapes de la méthode statistique	14
Figure 6: Etapes de la méthode de l'intelligence artificielle.....	15
Figure 7: Différence entre la méthode statistique et l'intelligence artificielle.....	15
Figure 8 : Matrice de corrélation.....	22
Figure 9: Méthodologie de Box-Jenkins	24
Figure 10: Evolution de la Puissance en fonction du temps	25
Figure 11: Décomposition de la série.....	26
Figure 12: Graphes de l'ACF et du PACF de la série ARIMA	31
Figure 13: ACF et PACF après différenciation.....	33
Figure 14: Résultats de la vérification du modèle.....	35
Figure 15: Méthodologie du modèle LSTM	39
Figure 16: Prévision avec le modèle ARIMA.....	42
Figure 17: Prévision de la Puissance avec le quantile 0,5.....	44
Figure 18: Prévision de Puissance pour un quantile de 0,1.....	45
Figure 19: Prévision de la Puissance pour un quantile de 0,9.....	46
Figure 20: Résidus pour un quantile de 0,5.....	47
Figure 21: Résidus pour une prévision de quantile 0,1.	48
Figure 22: Résidus pour un quantile de 0,9.....	48
Figure 23: Ensemble d'entrainement modèle LSTM	52
Figure 24: Ensemble de test du modèle LSTM.....	52
Figure 25: Validation du modèle.....	58

LISTE DES ANNEXES

Annexe I : Synthèse bibliographique	I
Annexe II : Explication de la vérification de la stationnarité.....	VII
Annexe III : Explication sur la différenciation.....	VIII
Annexe IV : Formule calcul de l'AIC	IX
Annexe V : Formule calcul du BIC.....	X
Annexe VI: Exemple de validation du modèle sur des données nouveaux et comparaison avec les données réelles.....	XI

INTRODUCTION

CONTEXTE ET JUSTIFICATION

La transition énergétique mondiale, face aux défis du réchauffement climatique et l'épuisement des ressources fossiles entraîne une adoption fulgurante des énergies renouvelables notamment le solaire, dans les mix énergétiques nationaux. Au Burkina Faso, pays doté d'un fort potentiel solaire, cette dynamique se manifeste par une augmentation significative de la production photovoltaïque, portée par des initiatives publiques et privées. Selon le Plan Directeur National de Production-Transport-Distribution et de l'Electrification Rurale 2017-2025 [1] la demande en électricité, estimé à 281 MWc en 2016, atteindra 646 MWc d'ici 2025, justifiant la valorisation des ressources endogènes tel que le solaire. Bien que les données exactes pour 2025 ne soient pas encore disponibles, la tendance à la hausse de la demande énergétique, liée à l'urbanisation croissante et à l'industrialisation, confirme la nécessité de renforcer l'infrastructure énergétique. Dans ce contexte, l'exploitation des ressources endogènes, notamment l'énergie solaire, est primordiale pour répondre aux besoins énergétiques futurs et assurer la transition vers une énergie durable. Cette transition a conduit à une augmentation notable de la capacité solaire photovoltaïque nationale, qui est passée de 33 MWc produit par la centrale de Zagtoui à environ 150 MWC en 2023 grâce à l'ajout des centrales de Nagréongo et de Kodéni. Les prévisions montrent que cette tendance se produira avec l'entrée en service des nouvelles installations, comme Ouaga Nord-Ouest (42 MWc avec stockage) et d'autres projets en cours, ce qui pourrait atteindre 200 MWc. Cependant, la nature intermittente de la ressource solaire pose un défi important. Les variations soudaines de production dues aux conditions météorologiques, comme les couverts nuageux, peuvent perturber l'équilibre offre-demande et compromettre la stabilité du réseau. Dans ce contexte, la société GreenYellow, un acteur majeur de la transition énergétique et propriétaire de la centrale de Nagréongo, en collaboration avec le laboratoire LABEREE de 2iE, nous a proposé ce sujet afin de développer un système de prévision à court terme basé sur des données historiques. Ce système vise à anticiper les fluctuations de production avec une précision accrue et une marge d'erreur inférieure à 10 %.

Ce travail vise à développer un système de prévision. Il s'appuie sur l'analyse et le traitement des données existantes pour en extraire un jeu exploitable ainsi que sur l'évaluation de méthodes

Développement d'un modèle de prévision à court terme de la puissance produite à la centrale solaire photovoltaïque de Ngréongo (30 MWc) dans le réseau national interconnecté du Burkina Faso.

éprouvées pour la prévision à court terme. L'objectif final est d'identifier la solution la mieux adaptée aux spécificités techniques et météorologiques de la centrale de Ngréongo, contribuant ainsi à la planification efficace du système électrique national.

OBJECTIF DU TRAVAIL

L'objectif principal de cette étude est de concevoir un modèle de prévision à court terme de la puissance délivrée par la centrale solaire photovoltaïque de Ngréongo connectée au réseau burkinabé. Pour y parvenir, les objectifs suivants ont été définis :

- Modéliser la production de la centrale solaire.
- Développer un modèle de prévision de la puissance délivrée par la centrale solaire.
- Valider le modèle de prévision développé par la centrale.

ORGANISATION DU TRAVAIL

Notre étude vise à caractériser, prédire la production de la puissance de la centrale solaire photovoltaïque de la centrale en explorant différentes techniques de prévision.

Nous avons commencé par une revue bibliographique (voir [Annexe I](#)) des méthodes de prévision couramment utilisées, leurs avantages et limites, ainsi que les travaux antérieurs dans ce domaine. Cette analyse nous a permis d'identifier les approches adaptées à notre étude et de poser les bases méthodologiques. Des techniques variées comme les réseaux de neurones ou les modèles autorégressifs, ont été explorées et comparées à l'aide de test statistiques.

Nous avons ensuite décrit la méthodologie adoptée, pour développer le modèle de prévision en mettant en avant les approches expérimentées et leurs spécificités.

Enfin, nous avons présenté le modèle retenu, pour prédire efficacement la production solaire dans le contexte de la centrale solaire photovoltaïque de Ngréongo.

PRESENTATION DE LA STRUCTURE

D'ACCUEIL

GreenYellow, fondée en 2007 est une entreprise leader dans les solutions énergétiques durables, qui s'engage à relever les défis énergétiques mondiaux. Elle intervient dans la production

Développement d'un modèle de prévision à court terme de la puissance produite à la centrale solaire photovoltaïque de Nagréongo (30 MWc) dans le réseau national interconnecté du Burkina Faso.

d'énergie solaire photovoltaïque, l'efficacité énergétique, les services à l'énergie, et la mobilité énergétique.

Située à 37 km de Ouagadougou, la centrale solaire photovoltaïque de Nagréongo d'une puissance installée de 30 MWc alimente la SONABEL grâce à un contrat PPA (Power Purchase Agreement), participant ainsi à l'indépendance énergétique du Burkina Faso. Mise en service en 2022, elle couvre une superficie de 52 hectares et compte environ 67872 panneaux solaires. Grâce à son raccordement au réseau national via un poste de transformation de 33 kV, elle injecte annuellement près de 50 GWh d'électricité, contribuant ainsi à réduire la dépendance aux énergies fossiles et à renforcer la sécurité énergétique du pays. Son exploitation repose sur un suivi rigoureux des performances et une maintenance proactive afin d'optimiser sa production et sa durabilité.

Afin de gérer les actifs, les financements et limiter les risques des partenaires, une société de projet, SPES (Société de Production d'Energie Solaire) a été créée. SPES est détenue à 95% par GreenYellow et 5 % par African Energy Corporation (AEC).

Durant notre stage, nous avons intégré le service des opérations, sous la supervision directe du directeur des opérations. Ce service joue un rôle crucial dans l'exploitation et l'optimisation des performances de la centrale solaire. Il est chargé de piloter la production d'énergie, de suivre les indicateurs de performances (production réelles, pertes etc.) et de garantir le bon fonctionnement des équipements. De plus, ce service assure la coordination avec les sous-traitants techniques et veille à l'application de meilleures pratiques opérationnelles.

GreenYellow s'appuie également sur un responsable administratif et financier, un responsable RSE et des sous-traitants tels que African Energy Corporation (AEC) pour le pilotage de la centrale et GIGI PROTECT pour la maintenance et la réparation des équipements. Ces acteurs travaillent en étroite collaboration sous la supervision du Directeur des opérations, garantissant une gestion intégrée et efficace de la centrale.

Développement d'un modèle de prévision à court terme de la puissance produite à la centrale solaire photovoltaïque de Nagréongo (30 MWc) dans le réseau national interconnecté du Burkina Faso.

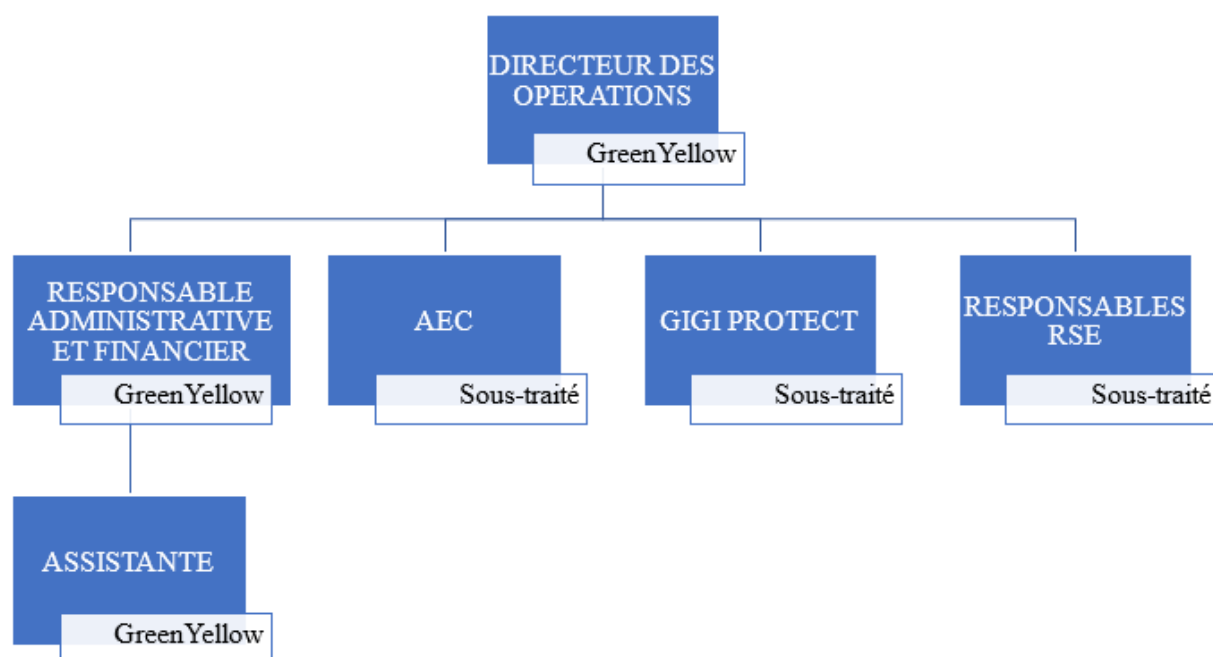


Figure 1: Organigramme de SPES

PARTIE 1 : GENERALITES SUR LA PREDICTION SOLAIRE

I. TYPES DE PREVISION

Les prévisions peuvent se faire en se basant sur l'horizon temporelle. En Machine Learning, il existe trois (03) type de prévision à savoir (voir Figure 2) :

- ❖ Court terme (quelques minutes à quelques heures)
- ❖ Moyen terme (quelques jours à quelques semaines)
- ❖ Long terme (quelques mois à quelques années)

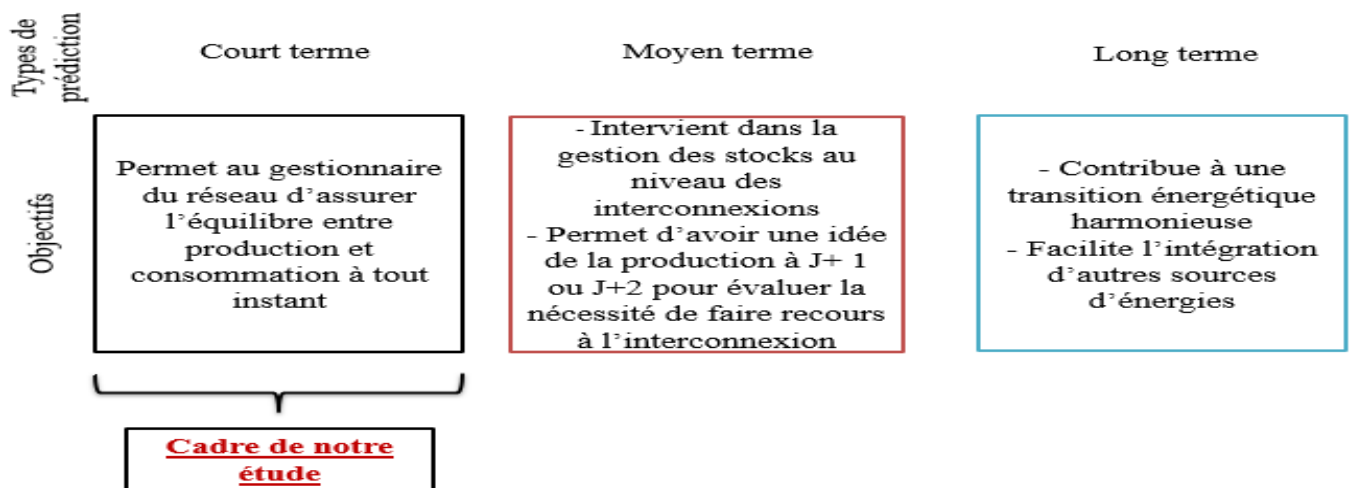


Figure 2: Types de prévisions et leurs objectifs

Pour ce qui est de notre étude, l'objectif étant de permettre au gestionnaire d'assurer l'équilibre entre la production et la distribution, nous opterons pour un modèle de prévisions à court-terme.

II. INTRODUCTION AU MACHINE LEARNING

Le Machine Learning consiste à donner la capacité à une machine d'apprendre sans la programmer de façon explicite. Plus précisément, on parle de Machine Learning lorsque la performance de la machine à exécuter une tâche s'améliore avec de nouvelles expériences.

En Machine Learning, il existe trois (03) méthodes à savoir :

- L'apprentissage supervisé (Supervised Learning)
- L'apprentissage non supervisé (Unsupervised Learning)
- L'apprentissage par renforcement (Reinforcement Learning)

II-1 L'APPRENTISSAGE SUPERVISÉ (SUPERVISED LEARNING)

L'apprentissage supervisé consiste à présenter à la machine une ou plusieurs variables et à lui demander de trouver l'association qui les relie. Tout d'abord, on met ces exemples dans une base de données (dataset). En apprentissage supervisé, la base de données contient deux types de variables, d'un côté nous avons la variable à prédire (Target variable) et de l'autre côté nous avons les facteurs (features) qui influencent la variable à prédire. Par convention, la Target variable est nommée m et les features n .

II.2 L'APPRENTISSAGE NON SUPERVISEE (UNSUPERVISED LEARNING)

En apprentissage non-supervisé, nous travaillons avec une base de données x sans valeurs associées y . L'objectif étant d'identifier des structures dans les données, cela implique quelques notions à savoir :

Clustering : regrouper les données en groupes en se basant sur leur similarité.

Détection d'anomalies : détecter les données qui ont une différence significative des autres.

La Figure 3 donne un aperçu du fonctionnement de l'apprentissage supervisé.

Développement d'un modèle de prévision à court terme de la puissance produite à la centrale solaire photovoltaïque de Ngréongo (30 MWc) dans le réseau national interconnecté du Burkina Faso.

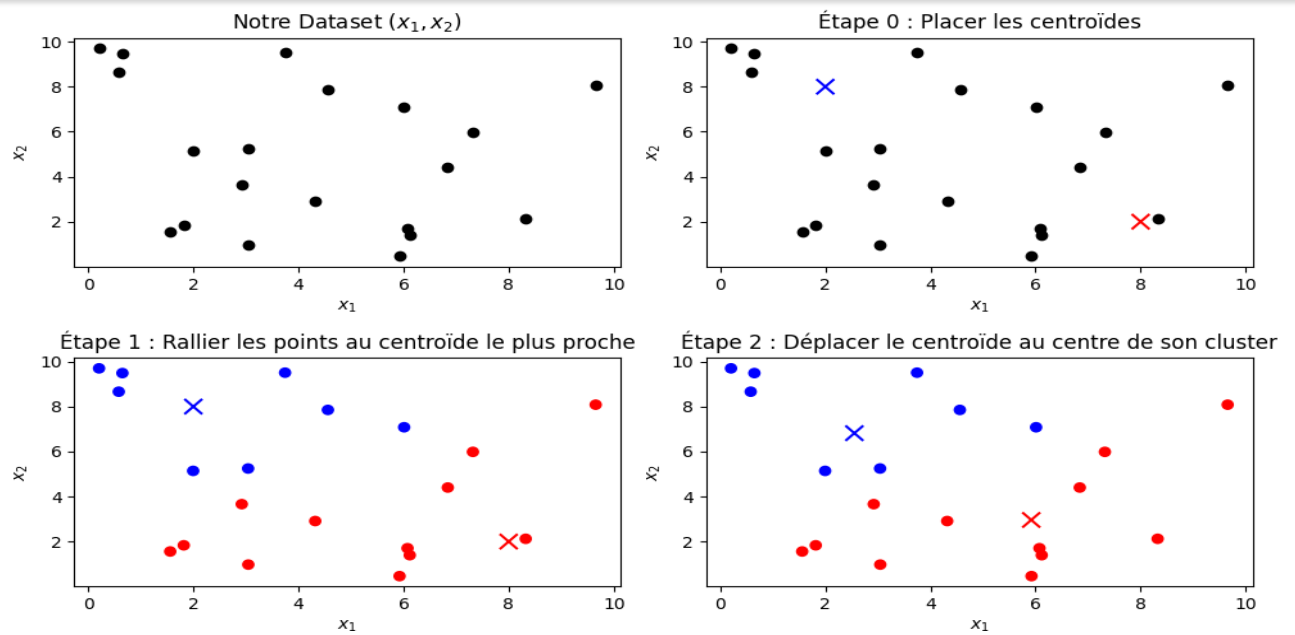


Figure 3: Processus de regroupement (clustering) en apprentissage non supervisé

En apprentissage supervisé, une base de données contient des valeurs d'entrée x associées à des cibles y (dataset étiqueté), permettant de modéliser une relation entre les deux. En revanche, en apprentissage non supervisé, il n'y a pas de cibles y , et l'objectif est d'explorer les structures des données sans cible prédéfinie. La Figure 4 illustre cette distinction :

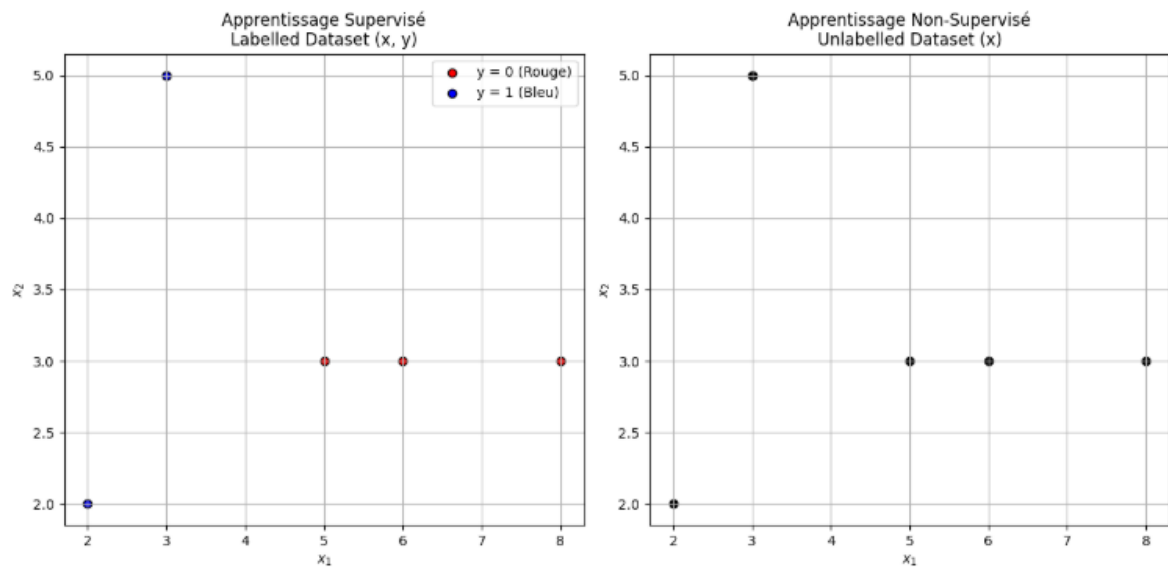


Figure 4: Différence entre Apprentissage supervisé et non supervisé

II.3 L'APPRENTISSAGE PAR RENFORCEMENT (REINFORCEMENT LEARNING)

L'apprentissage par renforcement est une branche du Machine Learning où un agent apprend à prendre des décisions en interagissant avec un environnement. Il reçoit des récompenses ou pénalités selon ses actions, l'objectif étant de maximiser les récompenses cumulées. Ce type d'apprentissage, basé sur l'optimisation d'une politique de décision est utilisé dans des domaines comme les jeux, nécessitant des prises de décisions en environnements dynamiques.

III. APPROCHES DE PREDICTION

Dans la littérature, il existe deux grandes approches de prévisions à savoir la méthode statistique et celle de l'intelligence artificiel.

III.1 L'APPROCHE STATISTIQUE

L'approche statistique repose sur l'analyse des données historiques pour établir des modèles mathématiques capables de prédire l'avenir. Elle s'appuie sur la relation entre variables et sur l'hypothèse que le passé peut prédire l'avenir, à condition que les données soient correctement analysées. La Figure 5 illustre les étapes essentielles de cette dans le cadre de notre étude approche.

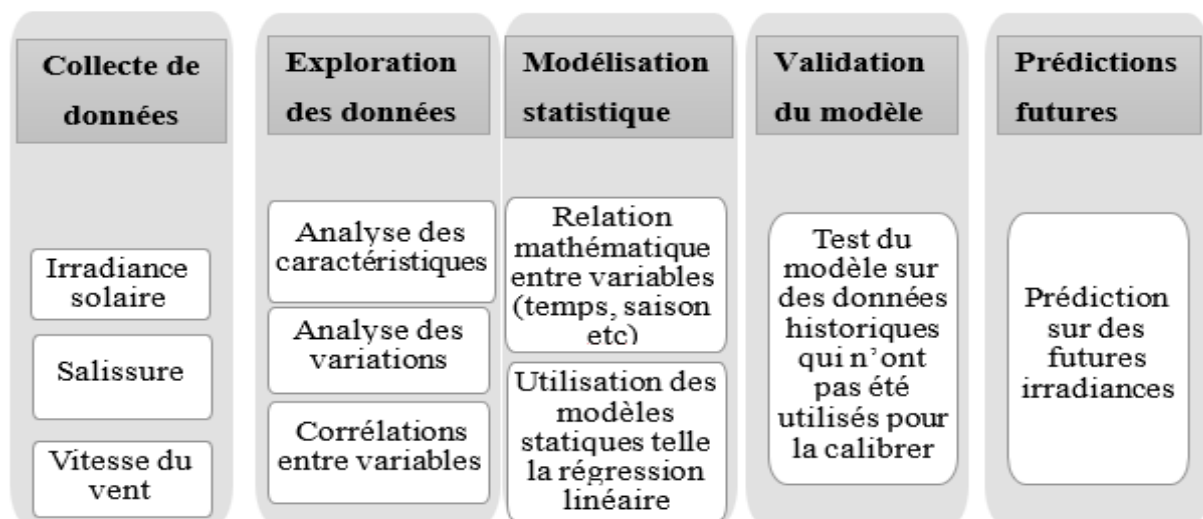


Figure 5: Etapes de la méthode statistique

III.2 LA METHODE DE L'INTELLIGENCE ARTIFICIELLE

Cette méthode utilise des techniques avancées pour modéliser les relations complexes entre les différentes données. A la différence de l'approche statistique, il n'est pas forcément nécessaire d'implémenter des relations établies pour chaque variable.

La Figure 6 montre les différentes étapes de la réalisation d'un modèle de prévision d'intelligence artificielle.

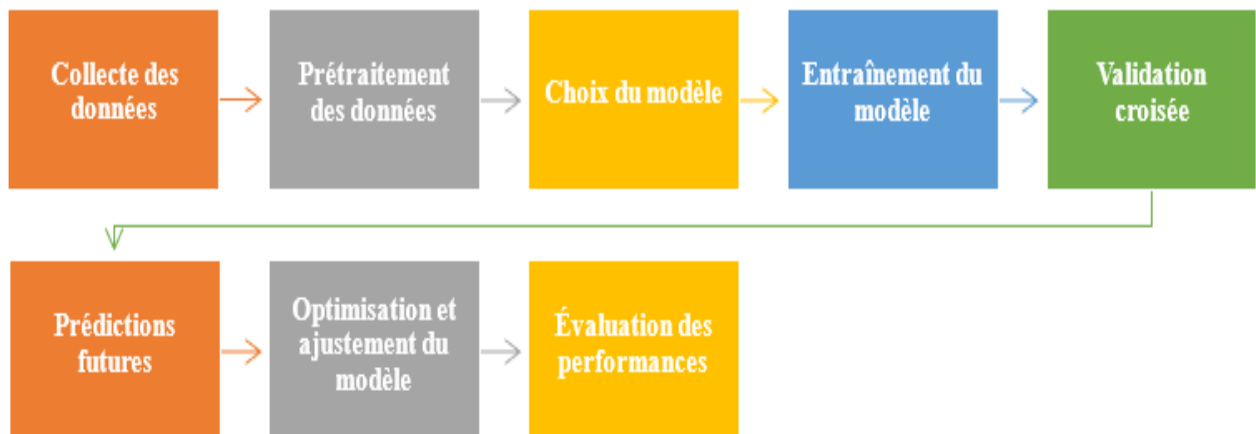


Figure 6: Etapes de la méthode de l'intelligence artificielle

La Figure 7 présente une synthèse de la différence entre la méthode statistique et la méthode de l'intelligence artificielle.

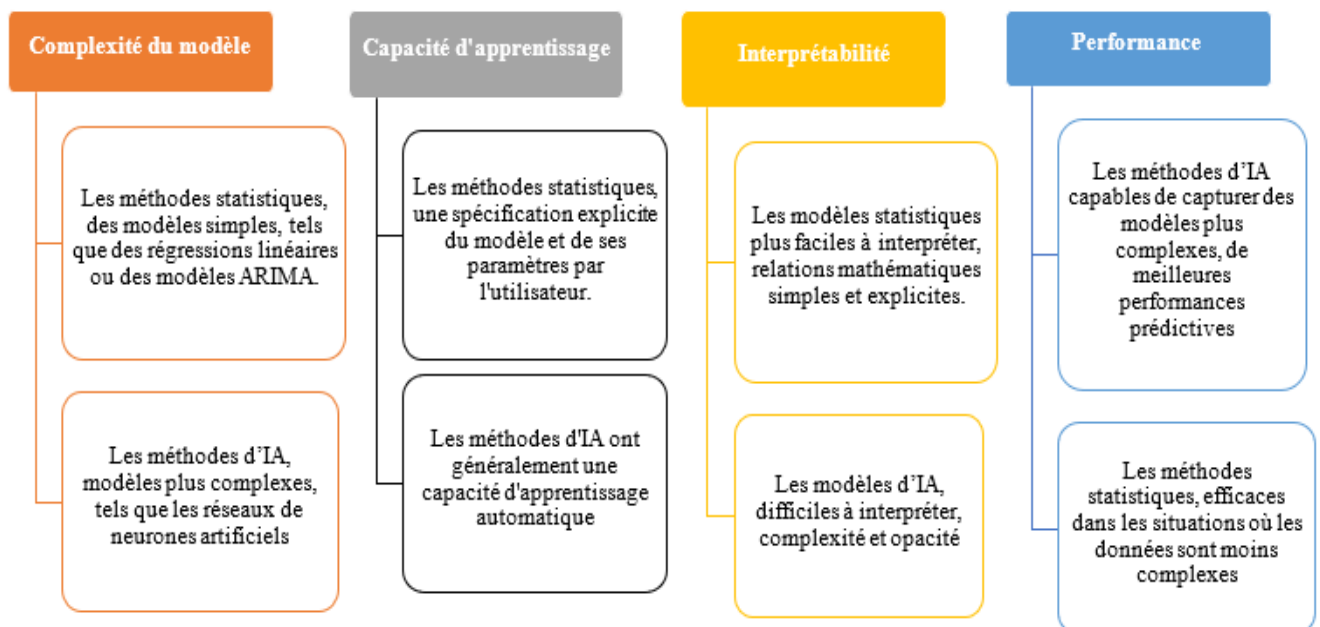


Figure 7: Différence entre la méthode statistique et l'intelligence artificielle

IV. LOGICIELS UTILISES EN PREVISION SOLAIRE

Lorsqu'il s'agit de la prévision solaire via le Machine Learning, les logiciels les plus utilisés sont Python et R. Bien qu'il soit généralement utilisé dans plusieurs approches de prévisions, leurs utilisation et fonctionnalités varient.

IV.1 LE LOGICIEL PYTHON

Python est un langage de programmation polyvalent et est largement utilisé dans plusieurs domaines y compris la prévision solaire. Python et Anaconda sont deux éléments liés dans l'écosystème de développement Python mais qui remplissent des rôles bien différents.

Anaconda est une distribution de Python qui lui fournit non seulement l'interpréteur Python de base, mais aussi de nombreux packages et outils utilisés.

Dans le langage Python, il existe quatre librairies essentielles à maîtriser pour faire du Machine Learning.



NumPy est la librairie qui permet de créer et manipuler des matrices simplement et avec efficacité. En Machine Learning, on insère le plus souvent notre Dataset dans des matrices. Il est important de le comprendre bien que les fonctions présentes dans **NumPy** font les calculs matriciels à notre place.



Matplotlib est la librairie qui permet de visualiser nos bases de données, nos fonctions, nos résultats sous forme de graphes, courbes et nuages de points.



Scikit learn est la librairie qui contient toutes les fonctions de l'état de l'art du Machine Learning. On y trouve les algorithmes les plus importants ainsi que diverses fonctions de pré-traitement.



Pandas est une excellente librairie pour importer vos tableaux Excel (et autres formats) dans Python dans le but de tirer des statistiques et de charger votre Dataset dans Sklearn.

IV.2 LE LOGICIEL R

R est un langage de programmation lui aussi présent dans le distribution Anaconda offrant des outils puissants pour l'analyse de données, avec des packages comme **solarEnergy** et **solarR** spécifiquement dédiés à la modélisation et à la prévision des systèmes solaires. La fonction **predict** de R permet de réaliser des prévisions tout en assurant précision et interprétabilité. Le package **solarEnergy** calcule la radiation solaire et prédit la production d'énergie, tandis que **solarR** simule la production d'énergie solaire en fonction des caractéristiques du site, des panneaux et des conditions météorologiques.

IV.3 LE CHOIX DU LOGICIEL A UTILISER

Dans la suite de notre travail, nous avons utilisé le logiciel Python pour les raisons suivantes :

- Popularité du langage
- Langage généraliste
- Forte communauté

V. METHODES UTILISEES

En prévision, il est difficile de définir un modèle universel qui s'applique à tous les cas de figure. Il est donc important de composer avec tous les outils disponibles et finalement choisir ceux qui sont plus à même de répondre aux exigences. Pour cela, la bibliographie nous a été d'une importance cruciale (Voir [Annexe I](#)).

Ainsi, nous avons opté pour ceux qui nous paraissaient les mieux adaptés d'après la littérature à savoir les modèles statistiques à savoir **le modèle ARIMA et la régression linéaire** ainsi que **le modèle LSTM** qui est un modèle d'Intelligence Artificielles.

VI. GENERALITES SUR LES SERIES TEMPORELLES

VI.1 QU'EST-CE QU'UNE SERIE TEMPORELLE ?

En statistique une série temporelle est une collection de données qui a été observée à un intervalle de temps régulier. Par exemple, imaginons que l'on veuille mettre en place un modèle qui prédise les ventes dans une entreprise ces ventes seront observées par jour par mois par année à une fréquence donnée. Puisqu'il y a cette temporalité, pour faire la prévision des ventes des jours à venir il faut utiliser des propriétés des séries temporelles.

Afin de parvenir à faire la prévision d'une série temporelle il est important d'analyser certains aspects de la base de données.

Dans notre contexte, il est question de prédire la production de puissance de la centrale solaire photovoltaïque de Nagréongo à court terme.

VI.2 ASPECTS FONDAMENTAUX D'UNE SERIE TEMPORELLE

- **La tendance :** La tendance représente l'évolution globale d'une série temporelle à long terme. Il est important de savoir si la série est décroissante, croissante ou à tendance globale. Autrement dit il est essentiel de connaître la direction globale de notre série sur une longue période.
- **La saisonnalité :** La saisonnalité est le fait d'observer si une série se comporte de façon identique sur une période donnée. Elle permet aussi de savoir si sur une période spécifique l'on a une évolution particulière.
- **Les cycles :** Ils permettent de savoir si après une période donnée l'on observe la répétition d'un phénomène particulier.
- **La stationnarité :** La stationnarité fait référence au fait que dans le temps, les caractéristiques principales de la série ne changent (la moyenne, la variance et l'autocorrélation).

Développement d'un modèle de prévision à court terme de la puissance produite à la centrale solaire photovoltaïque de Nagréongo (30 MWc) dans le réseau national interconnecté du Burkina Faso.

PARTIE 2 : MODELISATION DE LA PRODUCTION DE LA CENTRALE SOLAIRE

La modélisation de la centrale solaire de Nagréongo repose sur une approche d'apprentissage supervisé, utilisant des données collectées directement sur site. Cette démarche vise à anticiper la production énergétique en fonction des paramètres climatiques et opérationnels, tout en adaptant les méthodologies à la nature des données et aux objectifs spécifiques.

Trois méthodes principales ont été utilisées : ARIMA, régression quantile et réseau de neurones LSTM. Chacune de ces approches a été choisie pour ses particularités en termes de modélisation temporelles, prévision et traitement dynamique non linéaire.

I- PREPARATION DES DONNEES

Avant de présenter chaque méthode, un travail essentiel a consisté à préparer les données pour garantir leur fiabilité et leur cohérence.

I.1 NETTOYAGE DES DONNEES

Pour faire de la prévision, le modèle nécessite une base de données sans valeurs manquantes d'où la nécessité d'un nettoyage de données. Le nettoyage des données consiste en la suppression des valeurs manquantes, incohérentes ou aberrantes (par exemple valeurs de puissance négatives). Pour y parvenir, deux méthodes sont généralement adoptées.

- L'imputation : consiste à remplacer les valeurs manquantes par des valeurs estimées afin d'éviter la perte d'informations.

Statistiques (moyenne, médiane) : remplace par une valeur centrale de la distribution, soit par la médiane ou la moyenne.

KNN (K-nearest neighbors) : remplace par les moyennes des valeurs des voisins similaires.

Interpolation : estime les valeurs manquantes en fonction des tendances des données.

L'avantage de l'imputation est qu'elle permet d'éviter les pertes de données, est facile à implémenter et maintient la continuité. Toutefois, cela entraîne une perte d'informations et la perte de la relation avec d'autres variables.

- Suppression des lignes : cela consiste à supprimer les lignes contenant les valeurs manquantes.

Développement d'un modèle de prévision à court terme de la puissance produite à la centrale solaire photovoltaïque de Ngréongo (30 MWc) dans le réseau national interconnecté du Burkina Faso.

L'avantage est que cela facilite l'implémentation et la préservation de la qualité des données mais aussi le nettoyage rapide. Toutefois, cela entraîne la réduction de la taille de l'échantillon, la perte d'informations et est inadapté pour les petites bases de données.

Toutefois, dans la littérature statistique, il généralement affirmé que, procéder à la suppression des lignes ***dans le cas où ceci ne dépasse pas 10% de la base de données n'altère pas la qualité des données.***

Pour ce qui est de notre étude, supprimer les lignes contenant des valeurs manquantes est l'option choisie.

Sur 131400 données, 130460 données restent après la suppression (environ 0,71% de données supprimés) soit moins de 10% de la base de données.

Nous allons donc opter pour la suppression des lignes contenant des valeurs manquantes.

I.2 FILTRAGE DES DONNEES

Dans le cadre de l'analyse de la base de données d'une centrale solaire, il est essentiel de se concentrer sur les heures où la production est effective, soit entre 6h et 18h (heures diurnes). Ce filtrage permet de garantir que l'analyse et la modélisation ne sont pas influencées par des données non pertinentes, comme celles enregistrées la nuit lorsque la centrale n'est pas active. En plus, en excluant les heures non productives, les modèles prédictifs se concentrent uniquement sur les données utiles.

I.3 STRUCTURATION DES VARIABLES

La base de données utilisée est une série temporelle représentant la puissance produite par la centrale solaire photovoltaïque de Ngréongo (30 MWc) à intervalle de cinq (05) minutes pour l'année 2023 (première année d'exploitation de la centrale). Elle contient :

La Date : représente l'année, le mois, le jour, l'heure et la minute de chaque observation.

La Puissance : la puissance active produite à intervalle de cinq (05) minutes.

Variable cible : puissance produite (en MWc).

Variables explicatives : jouent un rôle clé dans la prédiction de la puissance produite par la centrale solaire photovoltaïque de Ngréongo. Elles permettent de prendre en compte les différents facteurs qui influencent la production d'énergie et d'améliorer la précision du modèle.

Développement d'un modèle de prévision à court terme de la puissance produite à la centrale solaire photovoltaïque de Nagréongo (30 MWc) dans le réseau national interconnecté du Burkina Faso.

- **Irradiation (GHI, GTI) :** mesurée par les deux stations météo présente sur le site de la centrale, elle représente l'énergie solaire reçue par les panneaux. Une irradiation plus élevée entraîne une production d'électricité plus importante.
- **Température des modules :** captée par des capteurs sur les panneaux, elle influence leur rendement. Une température trop élevée réduit l'efficacité des cellules photovoltaïques.
- **Température ambiante :** relevée par les stations météo, elle affecte indirectement la température des modules et, donc, leur performance.
- **Vitesse du vent :** les stations météo enregistre la vitesse du vent, qui joue un rôle dans le refroidissement des panneaux et peut améliorer leur rendement en cas de forte chaleur.
- **Pourcentage de salissure :** mesuré par le **capteur DustLQ**, ce paramètre permet d'évaluer l'impact de la poussière sur la production. Le DustLQ compare la quantité de lumière reçue par un capteur propre et un capteur exposé, et estime ainsi la perte due à l'encrassement des panneaux.

Dans la figure 8, nous avons réalisé une matrice de corrélation afin d'analyser les relations entre les différentes variables et la puissance à prédire. Cet outil nous permet d'identifier les variables fortement corrélées ainsi que celles dont la corrélation avec la cible est plus faible.

Toutefois, nous choisissons de conserver l'ensemble des variables, y compris celles présentant une corrélation faible. En effet, la corrélation mesure uniquement des relations linéaires et ne capture pas les effets plus complexes que certaines variables peuvent avoir sur la prédiction. De plus, une variable qui semble peu liée à la cible lorsqu'elle est prise isolément peut devenir significative lorsqu'elle est combinée avec d'autres, permettant ainsi d'exploiter des interactions qui améliorent la précision du modèle. Contrairement à la corrélation, qui examine chaque variable indépendamment, notre modèle d'apprentissage automatique analyse l'ensemble des variables simultanément et ajuste leur importance en fonction du contexte global. Ainsi, même si certaines variables présentent une corrélation plus faible avec la cible, elles peuvent jouer un rôle essentiel dans la modélisation. Toutes ces raisons nous poussent à faire le choix de toutes les conserver afin de maximiser la performance et la robustesse de notre modèle.

Développement d'un modèle de prévision à court terme de la puissance produite à la centrale solaire photovoltaïque de Nagréongo (30 MWc) dans le réseau national interconnecté du Burkina Faso.

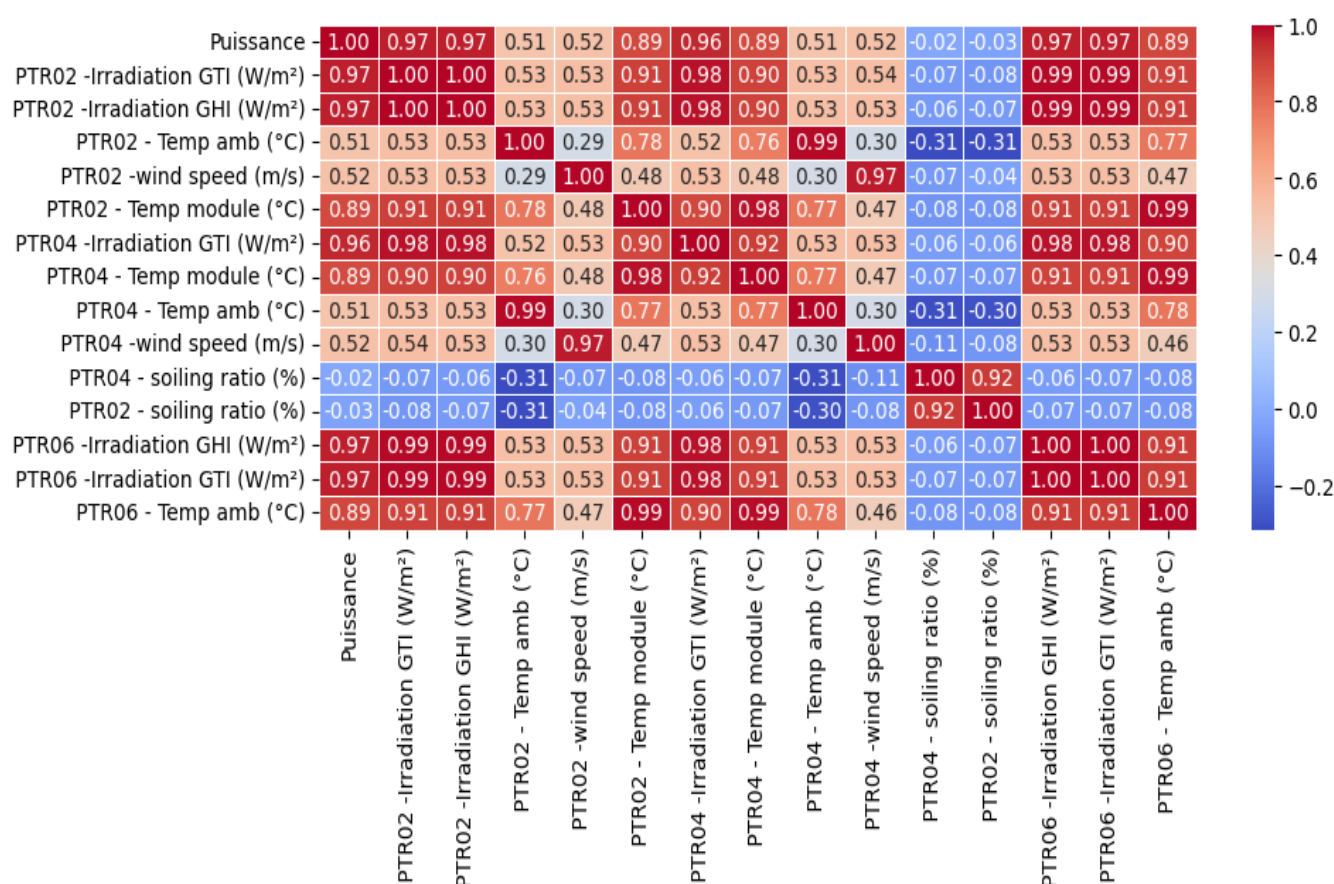


Figure 8 : Matrice de corrélation

I.4 DIVISION DES DONNEES

- Séparation simple

Elle consiste à diviser aléatoirement les données en ensemble de test et d'entraînement. Généralement, on opte pour 80% de données pour l'entraînement et 20% pour le test.

- Validation croisée

Elle est utilisée pour une évaluation plus robuste. Elle divise la base de données en plusieurs sous-ensembles (appelée folds) et utilise chaque sous-ensemble progressivement comme un ensemble de test tout en entraînant le modèle sur les autres sous-ensembles.

Pour ce qui est de notre étude, notre choix s'est porté sur la séparation simple pour les modèles ARIMA et régression quantile. Le LSTM quant à lui a été réalisé sur la base d'une validation croisée.

Développement d'un modèle de prévision à court terme de la puissance produite à la centrale solaire photovoltaïque de Nagréongo (30 MWc) dans le réseau national interconnecté du Burkina Faso.

II- MODELISATION AVEC ARIMA

Le modèle ARIMA se présente comme un outil assez performant pour la modélisation des séries temporelles. L'approche ARIMA a été utilisée pour capturer les relations dans les données de production.

II.1 INTRODUCTION AU MODELE ARIMA

II.1.1 Concepts fondamentaux

Le modèle ARIMA permet de faire à partir d'une seule variable quantitative des prévisions pertinentes sur des valeurs futures. Il existe trois (03) termes à connaître afin de mettre en place un modèle ARIMA.

La partie Autoregressive (p) : elle nous garantit la possibilité de prédire les valeurs futures sur la base de valeurs précédente.

La partie Integrated (d) : lorsque l'on souhaite prédire une variable, il faut s'assurer que la moyenne, la variance et la structure de la variable ne changent pas trop dans le temps ; on parle de **stationnarité**. Elle permet de trouver l'ordre de différenciation à mettre en place pour rendre la série stationnaire, c'est-à-dire moyenne, variance et structure constantes dans le temps.

La partie Moving Average (q) : il s'agit de faire la prévision sur les valeurs passées mais aussi d'utiliser les erreurs de prévision de notre modèle afin d'améliorer la performance dudit modèle.

II.1.2 Conditions d'application

Pour garantir des prévisions fiables, les modèles nécessitent des séries temporelles sans valeurs manquantes. De plus, la stationnarité est la condition clé pour plusieurs modèles comme ARIMA.

LA STATIONNARITE

Une série est dite stationnaire lorsque ses propriétés statistiques (moyenne, variance, autocorrélation) ne changent pas dans le temps. Autrement dit, la série est constante en moyenne, variance et structure dans le temps.

Elle est importante car elle facilite l'analyse et la modélisation de la série, améliore la précision des prévisions future et permet d'identifier les relations causales entre variables.

Vérification de la stationnarité

Analyse visuelle : examiner les graphiques pour détecter des tendances ou des variations de variance. L'analyse visuelle doit être confirmée par un test statistique afin d'éviter les erreurs.

Développement d'un modèle de prévision à court terme de la puissance produite à la centrale solaire photovoltaïque de Nagréongo (30 MWc) dans le réseau national interconnecté du Burkina Faso.

Tests statistiques : utiliser les test de Dickey-Fuller Augmenté (ADF) et Kwiatkowski-Phillips-Schmidt-Shin (KPSS) pour confirmer les observations visuelles.

➤ **Comprendre la méthodologie utilisée**

La méthodologie de Box-Jenkins est une approche couramment utilisée pour modéliser et prévoir les séries temporelles. Elle comprend les étapes suivantes :

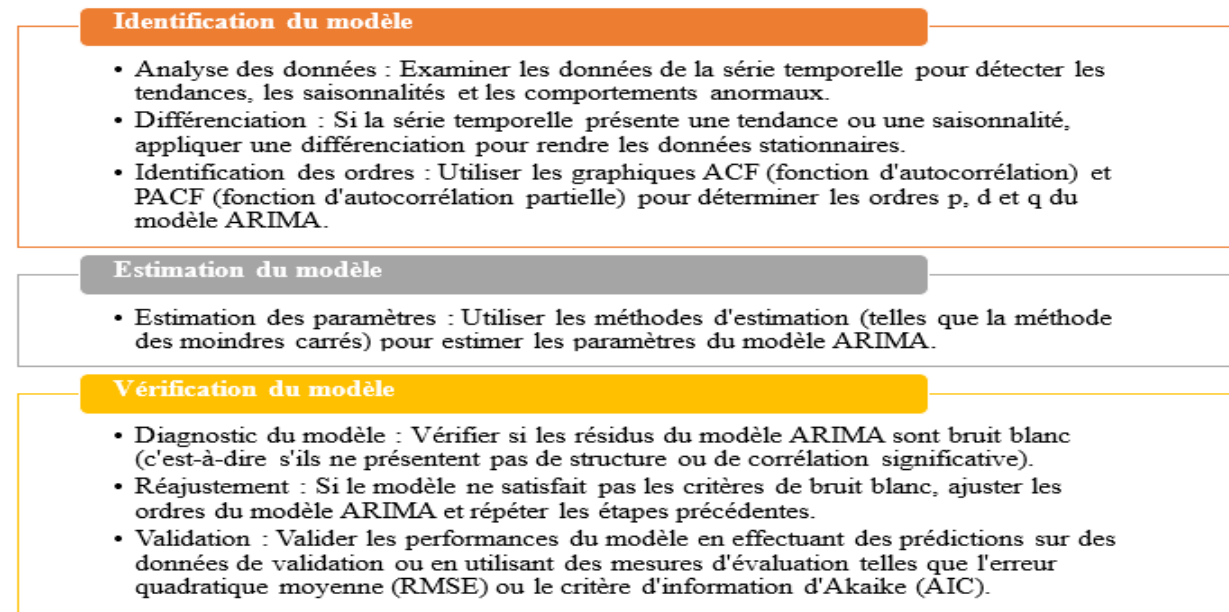


Figure 9: Méthodologie de Box-Jenkins

Source : Box, G. E. P., & Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control (Revised Edition)*. Holden-Day [2]

La méthodologie de Box-Jenkins est itérative, ce qui signifie que les étapes d'identification, d'estimation et de vérification peuvent être répétées plusieurs fois pour améliorer le modèle. L'objectif est de trouver le meilleur modèle ARIMA qui capture les motifs et les caractéristiques importantes de la série temporelle, et qui peut être utilisé pour effectuer des prévisions précises.

II.2 IDENTIFICATION DU MODELE

II.2.1 Analyse des données

L'analyse des données commence par une représentation graphique de la base de données. Elle permet d'identifier les tendances, saisonnalités et bien plus encore, la stationnarité.

Développement d'un modèle de prévision à court terme de la puissance produite à la centrale solaire photovoltaïque de Ngréongo (30 MWc) dans le réseau national interconnecté du Burkina Faso.

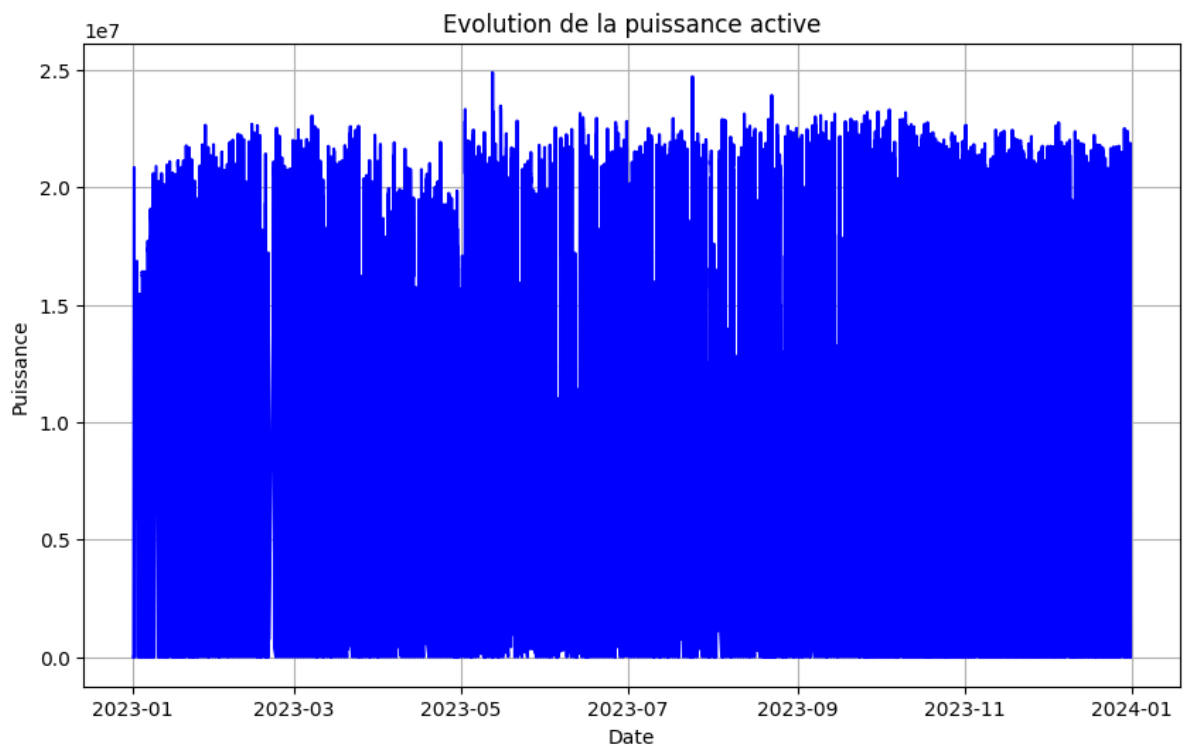


Figure 10: Evolution de la Puissance en fonction du temps

Dans cette série, on pourrait rechercher une tendance croissante ou décroissante globale, mais il semble que la série soit relativement stable. Il convient de faire la décomposition de la base de données pour des analyses plus poussées.

Développement d'un modèle de prévision à court terme de la puissance produite à la centrale solaire photovoltaïque de Ngréongo (30 MWc) dans le réseau national interconnecté du Burkina Faso.

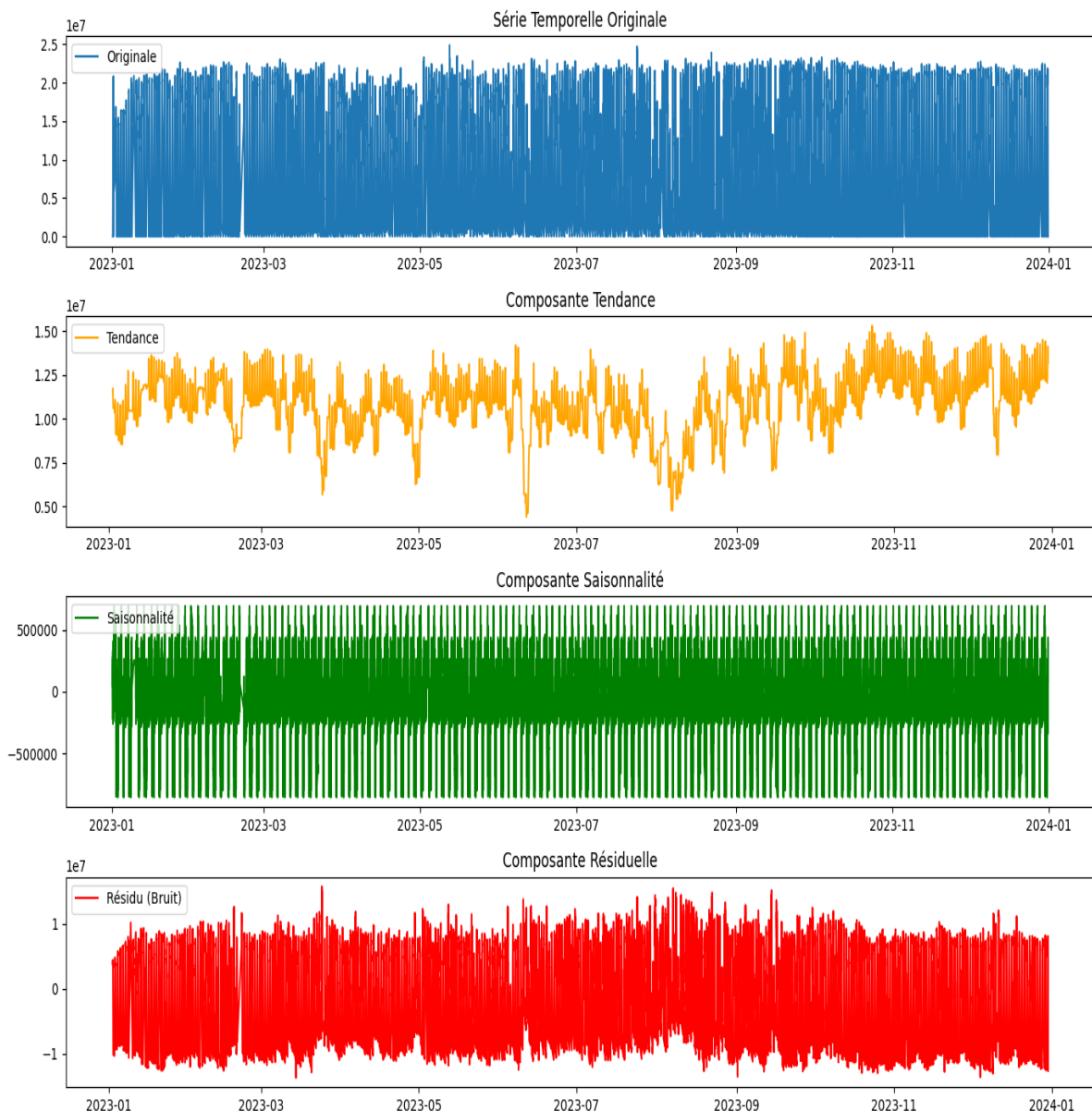


Figure 11: Décomposition de la série

Tendance : on note une variation non linéaire, avec une légère baisse en début d'année suivie d'une remontée progressive. La tendance présente une certaine cyclicité sur plusieurs mois.

Saisonnalité : la composante saisonnière montre des fluctuations périodiques. On observe une répétition régulière des motifs, ce qui indique un phénomène récurrent (probablement journalier ou hebdomadaire). L'amplitude des cycles semble relativement stable au fil du temps.

Résidus : ce graphique représente les résidus (bruit aléatoire) après avoir retiré la tendance et la saisonnalité. On observe une forte dispersion, ce qui montre la présence d'une variabilité

Développement d'un modèle de prévision à court terme de la puissance produite à la centrale solaire photovoltaïque de Nagréongo (30 MWc) dans le réseau national interconnecté du Burkina Faso.

importante.

II.2.2 Test de stationnarité

Le modèle ARIMA suppose qu'une série doit être stationnaire afin de pouvoir réaliser des prévisions exactes. Les tests couramment utilisés pour vérifier la stationnarité sont au nombre deux : le test de Dickey-Fuller (ADF) et le test de Kwiatkowski-Phillips-Schmidt-Shin (KPSS). Une série doit répondre aux spécificités des deux tests pour être stationnaire.

Test de Dickey-Fuller (ADF)

Il s'agit d'un test statistique utilisé pour vérifier l'existence d'une racine unitaire dans une série temporelle.

Hypothèse nulle : la série est non stationnaire.

Pour effectuer le test, il faut comparer la statistique de test Dickey-Fuller avec les valeurs critique tabulées pour différents niveaux de signification (généralement 5% pour des conventions statistiques appliqués dans plusieurs domaines).

Le test de Dickey-Fuller repose sur l'estimation d'une équation de la forme suivante pour une série temporelle y_t :

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} - 1 + \delta^2 \Delta y_{t-2} + \dots + \delta p \Delta y_{t-p} - p + \varepsilon_t \quad (1)$$

Avec :

Δy_t est la différence première de la série temporelle ($y_t - y_{t-1}$)

y_t représente la variable à prédire

α est une constante et représente la valeur moyenne de la première différence lorsque toutes les autres variables sont nulles.

β_t est un terme de tendance temporelle

γ est le coefficient de régression associé à la valeur retardée de la série temporelle y_{t-1} . Il est utilisé pour tester la présence d'une racine unitaire. Si $\gamma=0$, cela suggère que la série possède une racine unitaire et n'est donc pas stationnaire.

δ_i est le coefficient associé aux termes de différence retardée Δy_{t-1} . Ces termes représentent les valeurs différenciées retardée de la série temporelle.

ε_t est le terme d'erreur et représente la différence entre les valeurs observées et prédites.

Le test de ADF évalue l'hypothèse nulle H_0 selon laquelle la série temporelle y_t possède une racine unitaire ($\gamma=0$) non stationnaire. L'autre alternative H_1 est que la série est stationnaire et n'est donc pas stationnaire.

Si la statistique du test est inférieure à la valeur critique (p-value) et plus la valeur de test

Développement d'un modèle de prévision à court terme de la puissance produite à la centrale solaire photovoltaïque de Nagréongo (30 MWc) dans le réseau national interconnecté du Burkina Faso.

est petite, on rejette l'hypothèse nulle et on conclut que la série est stationnaire.

Tableau I: Résultats du test de Dickey-Fuller

Métrique	Valeur
Valeur de test	-39.69177927902888
P-valeur	0.0
Conclusion	La série est stationnaire

La statistique du test étant de -39,69 et la p-valeur de 0,0, cela indique que la série est stationnaire.

Le test de KPSS

Il s'agit d'un test pour évaluer la stationnarité d'une série temporelle. Contrairement à celui de Dickey-Fuller,

Hypothèse nulle : la série est stationnaire.

Le test repose sur une statistique qui suit la forme suivante :

$$KPSS = \frac{1}{T} \sum_{t=1}^n (\varepsilon_t^2 - \frac{1}{T} \sum_{t=1}^n \varepsilon_t^2) \quad (2)$$

T est le nombre d'observations dans la série temporelle

ε_t^2 sont les résidus de la régression de la série temporelle sur une constante (ou une tendance, selon le type de test KPSS)

Le terme $\frac{1}{T} \sum_{t=1}^n \varepsilon_t^2$ est la moyenne des carrées des résidus.

Tableau II : Résultats du test de KPSS

Métrique	Valeur
Valeur de test	0.517671179654083
P-valeur	0.03768667124908039
Nombre de retard	131
Conclusion	La série est non stationnaire

Si la statistique du test est inférieur à la valeur critique (p-value), on rejette l'hypothèse

Développement d'un modèle de prévision à court terme de la puissance produite à la centrale solaire photovoltaïque de Nagréongo (30 MWc) dans le réseau national interconnecté du Burkina Faso.

nulle et on conclut que la série est non stationnaire.

La p-valeur de 0,038, on conclut que la série n'est pas stationnaire.

Remarque :

Le test de KPSS est souvent utilisé conjointement à celui de l'ADF pour obtenir une évaluation plus complète de la stationnarité.

Le test de ADF stipule que la série est stationnaire et celui de KPSS le démentant, cela signifie qu'il y a des éléments comme une tendance ou une variance non constante que le test ADF ne détecte pas bien. Dans ce cas, l'on conclut que **la série n'est pas stationnaire** car KPSS est plus sensible aux changements de variance ou de tendances (*"Time Series Analysis: Forecasting and Control"* par George E.P. Box et Gwilym M. Jenkins) [2].

II.2.3 ACF et PACF

L'ACF (Autocorrelation Function) et la PACF (Partial Autocorrelation Function) sont deux outils essentiels en analyse de séries temporelles. Elles permettent de comprendre les dépendances temporelles dans une série de données.

➤ **ACF (Autocorrelation Function)**

L'ACF mesure la corrélation entre une séquence et elle-même à différentes périodes. Autrement dit, elle permet d'évaluer la similitude entre les observations en fonction du décalage de temps entre elles.

Pour un décalage 'k', l'ACF mesure la corrélation entre la série temporelle et elle-même décalée de 'k' périodes.

Formule

Pour une série X_t , l'ACF au décalage 'k' est donnée par la formule

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)} \quad (3)$$

Avec :

$\gamma(k)$ la fonction d'autocovariance au décalage 'k'

$\gamma(0)$ l'autocovariance au décalage '0'

$$\gamma(k) = \frac{1}{N} \sum_{t=1}^{N-k} (X_t - \bar{X})(X_{t+k} - \bar{X}) \quad (4)$$

Avec :

$\bar{X}$ la moyenne

N le nombre total d'observations

Développement d'un modèle de prévision à court terme de la puissance produite à la centrale solaire photovoltaïque de Nagréongo (30 MWc) dans le réseau national interconnecté du Burkina Faso.

Si $\rho(k)$ supérieur à 0, il existe une corrélation positive entre les observations

Si $\rho(k)$ inférieur ou égale à 0, la corrélation est négative

L'ACF est utilisée pour calculer l'ordre q du modèle. Généralement, les coefficients de l'ACF au-delà de l'ordre q tendent à être nul, aidant ainsi à identifier l'ordre q du modèle.

➤ **PACF (Partial Autocorrelation Function)**

La PACF mesure la corrélation entre une observation et une autre à un certain décalage h , tout en ajustant ou en éliminant les effets des autres observations intermédiaires.

La PACF est une corrélation qui exclut l'effet des termes intermédiaires. C'est-à-dire qu'elle est une mesure de la corrélation entre des observations à un certain intervalle, en tenant compte des valeurs à des intervalles plus courts.

Soit h , le décalage de la série.

Pour $h=1$, la PACF est égale à 1 car il s'agit de la corrélation de la série avec elle-même au même point dans le temps.

Lorsque h est supérieur à 1, la PACF $\alpha(h)$ est représenté par le coefficient ϕ_{hh} . Ce coefficient est le dernier composant de ϕ_h qui est calculé en utilisant la matrice d'autovariance Γ_h et le vecteur d'auto-variance jusqu'au lag h γ_h .

Formules

$$\Phi_h = \Gamma_h^{-1} \cdot \Gamma_h \quad (5)$$

Avec :

Γ_h une matrice symétrique

γ_h une vecteur.

$$\gamma(h) = \text{Cov}(X_t, X_{t-h})$$

$$\Gamma_h^{-1} = \frac{1}{\det(\Gamma_h)} \text{Adj}(\Gamma_h) \quad (6)$$

La PACF est utilisée pour calculer l'ordre p du modèle. Généralement, les coefficients de la PACF au-delà de l'ordre p tendent à être nul, aidant ainsi à identifier l'ordre p du modèle.

Commentaire

En sommes, les graphiques ACF et PACF sont couramment utilisés pour aider à choisir les paramètres d'un modèle ARIMA en analyse de séries temporelles. Le graphique ACF peut être utilisé pour identifier le terme MA (moyenne mobile p), tandis que le graphique PACF peut

Développement d'un modèle de prévision à court terme de la puissance produite à la centrale solaire photovoltaïque de Nagréongo (30 MWc) dans le réseau national interconnecté du Burkina Faso.

aider à identifier le terme AR (autorégressif q).

Toutefois, l'ACF permet aussi d'étudier la stationnarité de façon visuelle. En effet, lorsque le graphe de l'ACF présente une décroissance lente et régulière, la série n'est pas stationnaire. Il y a stationnarité lorsqu'il y a une variation brusque de l'ACF.

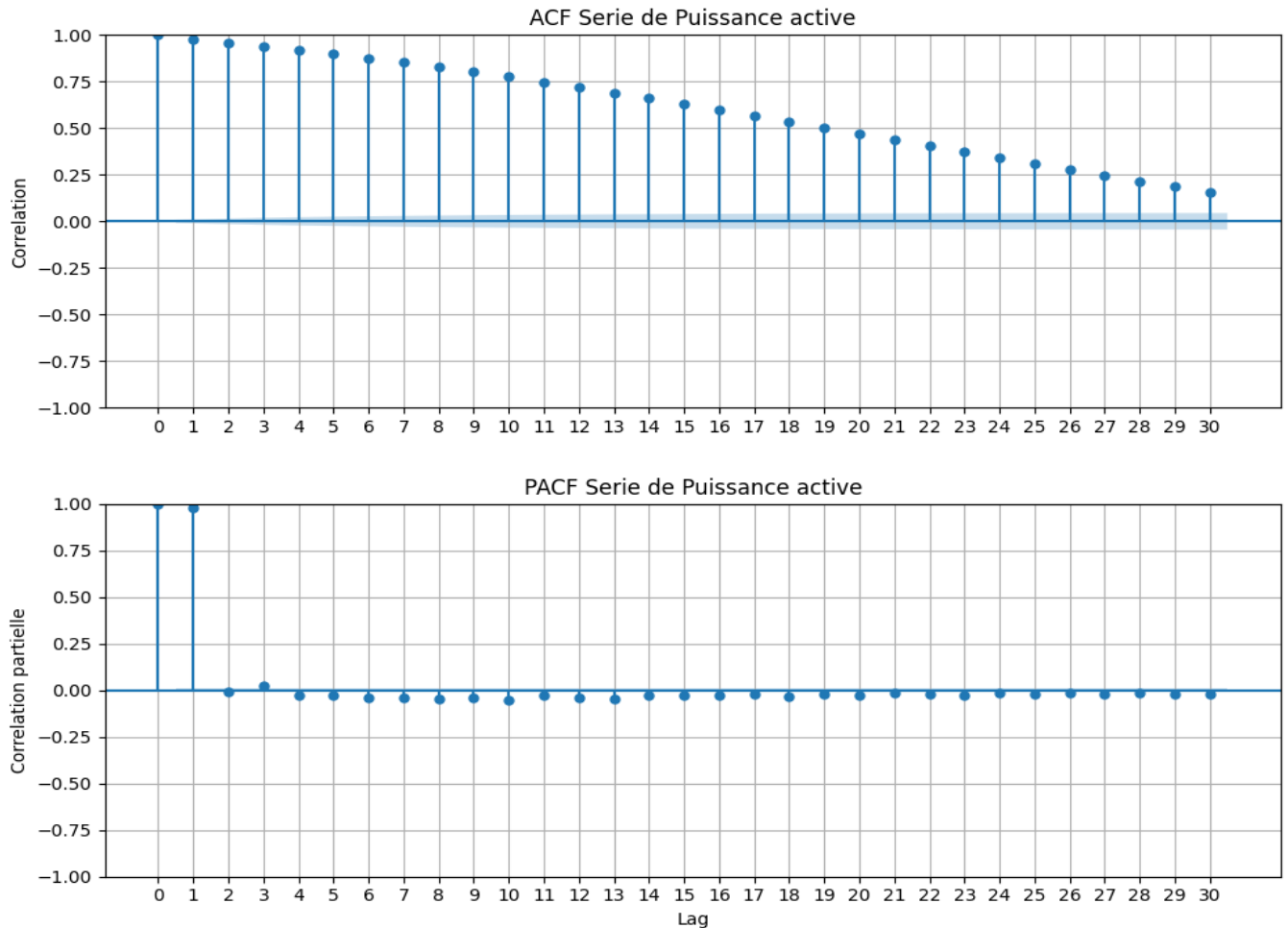


Figure 12: Graphes de l'ACF et du PACF de la série ARIMA

II.2.4 Remarque

Le test de Dickey-Fuller (ADF) indique que la série est stationnaire, tandis que le test de KPSS réfute cette stationnarité. Lorsqu'il y a conflit entre les deux tests, la série est traitée comme non stationnaire et doit subir une transformation. L'ACF présente une décroissance lente et régulière, ce qui, selon les études de Box et Jenkins (1976) dans leur ouvrage intitulé *Time Series Analysis : Forecasting and Control* [2], George Box et Gwilym Jenkins expliquent que les séries non-stationnaires présentent souvent une autocorrélation qui décroît lentement contrairement aux séries stationnaires où l'ACF tend à se stabiliser ou montre une décroissance rapide et cela, après quelques décalages. **On conclut alors que la série n'est pas stationnaire.**

Développement d'un modèle de prévision à court terme de la puissance produite à la centrale solaire photovoltaïque de Nagréongo (30 MWc) dans le réseau national interconnecté du Burkina Faso.

II.3 TRAITEMENT DE LA STATIONNARITE

Afin de résoudre le problème de stationnarité, trois méthodes sont principalement utilisées.

Différenciation : consiste à soustraire chaque valeur de la série par la précédente ($Y_t - Y_{t-1}$). Cela aide à supprimer les tendances et à stabiliser les données.

Découpage de la tendance : consiste à identifier et enlever la tendance générale (croissance ou baisse) en ajustant un modèle mathématique. Cela permet d'obtenir des données plus stables.

Transformation logarithmique : consiste à appliquer un logarithme ($\log Y_t$) sur les valeurs pour réduire les grandes variations et rendre la série plus régulière. Cette méthode est utile lorsque les écarts entre les valeurs sont trop importants.

Pour la suite de notre travail, nous avons opté pour la différenciation, particulièrement efficace pour éliminer les tendances et rendre les séries stationnaires.

Types de différenciation

- ❖ Différenciation simple : utilisée pour éliminer les tendances linéaires.
- ❖ Différenciation d'ordre supérieur : appliquée lorsque la première ne suffit pas, permet de traiter des tendances plus complexes.
- ❖ Différenciation saisonnière : permet de supprimer les composantes saisonnières d'une série se répétant à intervalles réguliers.
- ❖ Différenciation fractionnaire : une approche qui permet de stabiliser les séries sans les rendre stables. Elle est utilisée pour des séries complexes où une simple différenciation ne rend pas la série stationnaire et une seconde la rend trop instable.

Remarque

Pour ce qui est de notre étude, nous avons opté pour une différenciation simple qui a donné des résultats concluants. Cette méthode étant la plus simple de différenciation a suffi pour rendre la série stationnaire.

Cela ayant donné des résultats concluant au niveau du test de stationnarité, fort est de constater que les résultats des tests de ADF, KPSS et de l'ACF ramène à une seule conclusion : **La série différenciée est stationnaire.**

Développement d'un modèle de prévision à court terme de la puissance produite à la centrale solaire photovoltaïque de Nagréongo (30 MWc) dans le réseau national interconnecté du Burkina Faso.

Tableau III : Test de KPSS après différenciation.

KPSS Statistic	0.001288
p-value	0.100000
Conclusion	La série est stationnaire (hypothèse nulle non rejetée).

Le test de ADF après différenciation donne les résultats suivants :

Tableau IV : Test de ADF après différenciation.

ADF Statistic	-41.332082
p-value	0.000000
Conclusion	La série est stationnaire (hypothèse nulle rejetée).

Aussi, le constat est fait au niveau de la courbe de l'ACF qui évolue jusqu'à présenter une diminution brusque au niveau de la corrélation à un certain niveau de décalage qui est une caractéristique des séries stationnaire (d'après les travaux de Box et Jenkins) [2].

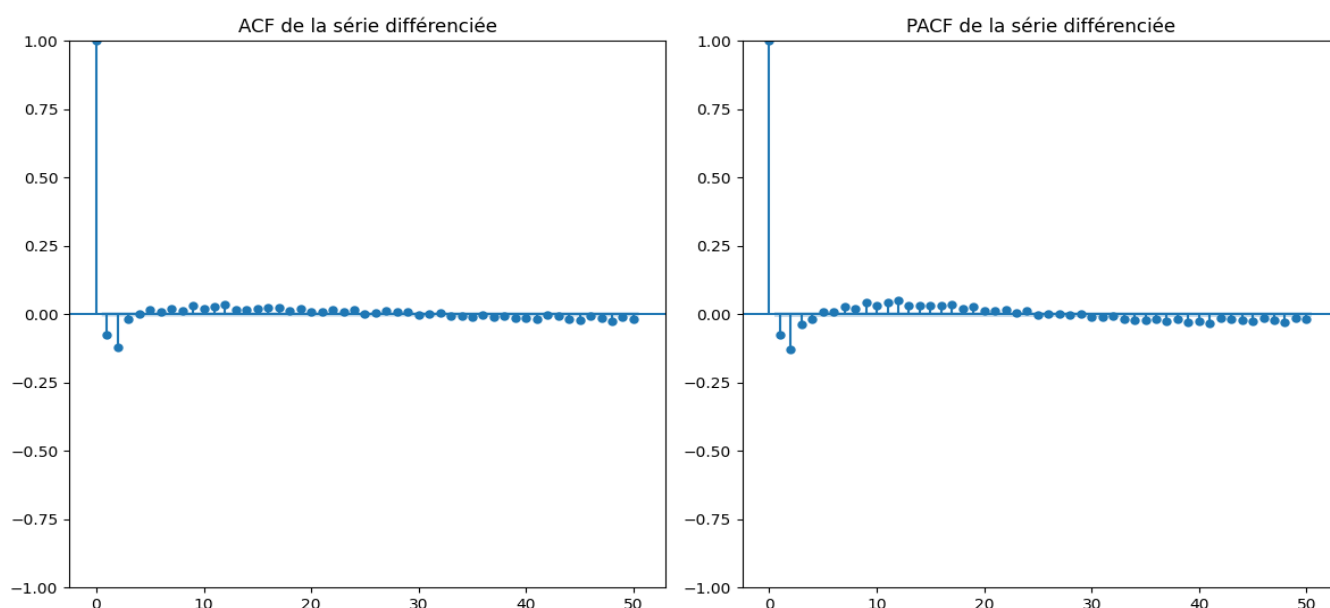


Figure 13: ACF et PACF après différenciation

II.4 ESTIMATION DU MODELE

L'estimation des paramètres d'un modèle ARIMA implique plusieurs étapes. Pour ce qui est du modèle ARIMA, il passe non seulement par l'identification des paramètres p , d et q du modèle mais aussi l'estimation de ces paramètres.

Développement d'un modèle de prévision à court terme de la puissance produite à la centrale solaire photovoltaïque de Nagréongo (30 MWc) dans le réseau national interconnecté du Burkina Faso.

II.3.1 Identification des paramètres (P, D, Q)

Les paramètres du modèle sont indiqués à l'aide de :

ACF pour le paramètre q (autocorrélation).

PACF pour le paramètre p (autocorrélations partielles).

Différenciation pour le paramètre d.

Afin de simplifier cette étape, la fonction `auto_arima` de la bibliothèque `pmdarima` peut être utilisée pour tester automatiquement différentes combinaisons en se basant sur les critères d'information. Cette méthode nous a donné des valeurs ARIMA (3,0,3)

II.3.2 Critères d'information

AIC (Akaike Information Criterion) : développé par Hirotugu Akaike en 1973 dans son article intitulé '*Information Theory and an Extension of the Maximum Likelihood Principle*' [3], il mesure l'équilibre entre qualité d'ajustement et complexité du modèle.

Un AIC faible indique un meilleur modèle.

Bayesian Information Criterion (BIC) : développé par Gideon Schwarz en 1978, il est similaire à l'AIC mais avec une pénalisation davantage les modèles complexes. Le BIC favorise des modèles plus simples, surtout pour les grands échantillons.

Sur la base du BIC et de l'AIC, les paramètres choisis sont : ARIMA (3,0,3).

II.3.2 Vérification du modèle

Une fois le modèle ajusté, il est important de le vérifier. La vérification permet d'observer la différence entre les valeurs observées et celles prédites. Lorsque le modèle est bien ajusté, les résidus se comportent comme des '**bruits blancs**' (normalité, absence d'autocorrélation, moyenne de zéro). La qualité du modèle est vérifiée à l'aide des tests suivants :

Test de L-Jung Box : vérifie l'absence d'autocorrélation dans les résidus.

Test de Jarque-Bera (JB) : évalue la normalité des résidus.

Test d'hétéroscédasticité : vérifie si les erreurs de prévision ont une non variance constante (erreur de prévision non constante).

Les différents test ont données les résultats suivants :

Développement d'un modèle de prévision à court terme de la puissance produite à la centrale solaire photovoltaïque de Nagréongo (30 MWc) dans le réseau national interconnecté du Burkina Faso.

SARIMAX Results						
=====						
Dep. Variable:	Puissance_diff1	No. Observations:	42883			
Model:	ARIMA(3, 0, 3)	Log Likelihood	-675864.759			
Date:	Mon, 02 Sep 2024	AIC	1351745.518			
Time:	15:25:24	BIC	1351814.848			
Sample:	0	HQIC	1351767.391			
	- 42883					
Covariance Type:	opg					
=====						
	coef	std err	z	P> z	[0.025	0.975]

const	432.6120	1.07e+04	0.040	0.968	-2.05e+04	2.14e+04
ar.L1	1.0086	0.025	40.352	0.000	0.960	1.058
ar.L2	0.3064	0.043	7.105	0.000	0.222	0.391
ar.L3	-0.3563	0.021	-17.239	0.000	-0.397	-0.316
ma.L1	-1.1228	0.025	-45.660	0.000	-1.171	-1.075
ma.L2	-0.3227	0.045	-7.171	0.000	-0.411	-0.234
ma.L3	0.4945	0.023	21.974	0.000	0.450	0.539
sigma2	2.899e+12	0.005	5.93e+14	0.000	2.9e+12	2.9e+12
=====						
Ljung-Box (L1) (Q):	0.35	Jarque-Bera (JB):	862686.89			
Prob(Q):	0.55	Prob(JB):	0.00			
Heteroskedasticity (H):	1.92	Skew:	-0.94			
Prob(H) (two-sided):	0.00	Kurtosis:	24.89			
=====						
Warnings:						
[1] Covariance matrix calculated using the outer product of gradients (complex-step).						
[2] Covariance matrix is singular or near-singular, with condition number 3.82e+28. Standard errors may be unstable.						

Figure 14: Résultats de la vérification du modèle

- Test de Ljung-Box (Q)

Q=0,35 et Prob(Q)= 0,55

Une p valeur Q supérieur à 0,05 signifie que nous n'avons pas assez de preuves pour rejeter l'hypothèse nulle d'absence d'autocorrélation. Cela suggère que les résidus ne présentent pas de corrélation significative et que le modèle capte bien la dynamique des données.

- Test de Jarque Bera

Le test de **Jarque-Bera (JB)** repose principalement sur la **p-valeur** et les statistiques de **skewness** (asymétrie) et **kurtosis** (aplatissement).

Statistique JB = 862686.89 (valeur très élevée, ce qui indique une forte déviation par rapport à la normalité).

p-valeur = 0.00 Comme $p < 0,05$, on rejette l'hypothèse de normalité : les résidus ne sont pas normalement distribués.

Skew=-0.94 (une asymétrie négative) et **Kurtosis=24.89** (valeur très élevée indiquant des

Développement d'un modèle de prévision à court terme de la puissance produite à la centrale solaire photovoltaïque de Nagréongo (30 MWc) dans le réseau national interconnecté du Burkina Faso.

queues épaisses, donc la présence d'extrêmes).

Test d'hétéroscédasticité :

$H > 1 \rightarrow$ Hétéroscédasticité (variance variable)

Si $p < 0.05$ cela implique une présence d'hétéroscédasticité

$H = 1.92 \rightarrow$ Supérieur à 1 $\rightarrow$ hétéroscédasticité probable.

$p = 0.00 (< 0.05) \rightarrow$ On rejette l'hypothèse d'homoscédasticité $\rightarrow$ hétéroscédasticité confirmée.

Malgré l'application de plusieurs transformations, notamment la transformation logarithmique, Box-Cox, la différenciation, ainsi que l'ajustement des paramètres du modèle ARIMA et l'intégration de composantes saisonnières, les problèmes de non-normalité des résidus, d'hétéroscédasticité et d'autocorrélation persistent, indiquant que le modèle ne parvient pas à capturer adéquatement la dynamique des données.

Nous avons donc décidé d'effectuer quand bien même la prévision du modèle.

III- MODELISATION AVEC LA REGRESSION

QUANTILE

La Régression Quantile est une méthode statistique qui permet de modéliser les relations entre une ou plusieurs variables explicatives. Contrairement aux modèles de régression simples qui se concentrent uniquement sur la prévision de la moyenne de la variable dépendante (à prédire), la régression quantile s'intéresse à l'ensemble de la distribution.

III.1 INTRODUCTION A LA REGRESSION QUANTILE

Conditions d'application

Contrairement aux autres modèles statistiques tel ARIMA et VECM, la régression quantile est un modèle assez flexible dans laquelle les hypothèses applicables aux autres modèles (normalité, homoscédasticité, etc.) ne sont pas respectés. Toutefois, il nécessite une grande base de données pour obtenir des estimations fiables à différents quantiles.

Concepts fondamentaux

La régression quantile est une méthode assez flexible qui s'applique dans bien de situations et cela, même lorsque les hypothèses de la régression classique ne sont pas appliquées.

Toutefois, il existe quelques considérations à prendre en compte pour une bonne utilisation de

Développement d'un modèle de prévision à court terme de la puissance produite à la centrale solaire photovoltaïque de Nagréongo (30 MWc) dans le réseau national interconnecté du Burkina Faso.

cette méthode notamment la notion de Quantile.

Définition :

Un Quantile est une valeur qui divise un ensemble de valeur en intervalles de probabilités égales. Il permet la répartition des données selon leurs positions dans une distribution.

Exemples de quantiles :

- Quartile : divise les données en quatre intervalles de probabilités égales
- Déciles : divise les données en dix intervalles de probabilités égales
- Percentiles : divise les données en cent intervalles de probabilités égales

Importance :

Ils permettent d'analyser facilement la répartition des données. En effet, cela permet d'analyser les données et voir si elles sont concentrées autour d'une certaine valeur ou dispersée.

Les quantiles donnent une vue plus complète de la relation entre les variables à prédire et cela, en modélisant la distribution à plusieurs niveaux, pas uniquement autour de la moyenne.

III.2 METHODOLOGIE

La régression quantile est différente de la méthodologie de Box-Jenkins car elle n'utilise pas cette méthode en tant que tel. Elle consiste dans cette situation en plusieurs étapes.

Définition du modèle de régression quantile : l'ajustement du modèle avec les quantiles les mieux adaptés

Prévision : utiliser le modèle ajusté afin de faire la prévision des ensembles d'entraînement et de test.

Calcul des résidus : évaluer la différence entre les valeurs observées et prédites.

Calcul des métriques d'observations : évaluer la qualité du modèle en utilisant les métriques tel MSE (Mean Squared Error), MAE (Mean Absolute Error), MedAE (Median Absolute Error).

Pour ce qui est du prétraitement des données, nous en avons déjà parler ci-dessus (Voir [Préparation des données](#)) est le même pour la plupart des modèles de séries temporelles.

La modélisation étant achevée pour la régression quantile, le modèle est fin prêt pour être utilisé pour des prévisions.

Développement d'un modèle de prévision à court terme de la puissance produite à la centrale solaire photovoltaïque de Nagréongo (30 MWc) dans le réseau national interconnecté du Burkina Faso.

IV- MODÉLISATION AVEC LE MODELE LONG SHORT-TERM MEMORY (LSTM)

Le modèle LSTM (Long Short-Term Memory) est un type de réseau de neurones conçu pour traiter des données séquentielles, comme les séries temporelles. Ils sont particulièrement efficaces pour capturer les dépendances temporelles, ce qui les rend adaptés aux séries temporelles, objet de notre étude.

IV.1 INTRODUCTION AU MODELE LSTM

Conditions d'application

Le modèle LSTM est adapté aux séries chronologiques complexes, notamment celles avec des dépendances à long terme. Les LSTM surpassent les Réseaux de Neurones Récurrents (RNN- Recurrent neural Network) classiques en résolvant le problème de *vanishing gradient*¹, ce qui leur permet de mieux capturer les relations à long terme.

Concepts fondamentaux

Les Réseaux de Neurones Récurrents (RNN) traitent les données séquentielles mais sont limités pour capturer des relations à long terme. Les LSTM, une extension des RNN, utilisent des portes (input, forget, output) pour réguler le flux d'informations et mémoriser ou oublier des données en fonction de leur pertinence. Cette architecture est idéale pour modéliser les interactions complexes entre des variables comme le GHI, la température et la vitesse du vent, en lien avec la production énergétique.

IV.2 METHODOLOGIE

Pour appliquer le modèle LSTM à la prévision énergétique, les étapes suivantes ont été suivies :

¹ Vanishing gradient : Problème dans les réseaux de neurones où les ajustements de paramètres deviennent trop faibles, empêchant les couches profondes d'apprendre efficacement.

Développement d'un modèle de prévision à court terme de la puissance produite à la centrale solaire photovoltaïque de Nagréongo (30 MWc) dans le réseau national interconnecté du Burkina Faso.

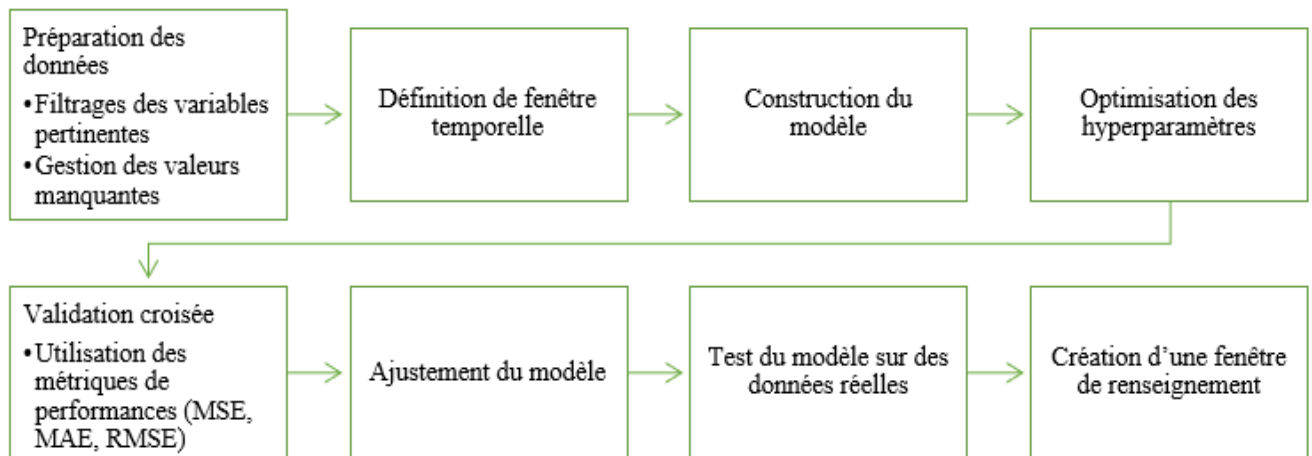


Figure 15: Méthodologie du modèle LSTM

Ces étapes permettront de construire un modèle LSTM pour prédire la puissance générée par la centrale solaire.

IV.3 DEFINITION DU MODELE

Architecture du modèle

Le modèle LSTM (Long Short-Term Memory) est conçu pour traiter des dépendances temporelles longues, en évitant les problèmes de vanishing gradient rencontrés par les RNN classiques. Son architecture comprend :

Entrée : données temporelles séquentielles.

Couches cachées : cellules traitant les informations à chaque étape.

Porte d'oubli : décide quelles informations doivent être oubliées.

Porte d'entrée : détermine quelles informations seront ajoutées à la mémoire.

Porte de sortie : décide quelles informations sont pertinentes pour la sortie.

Sortie : information transmise à la couche suivante ou à l'étape suivante.

Justification du choix de LSTM

LSTM est préféré pour les séries temporelles complexes car elle a la capacité de capturer les dépendances temporelles longues. Contrairement aux RNN classiques sujets au vanishing gradient, LSTM utilise ses fenêtres pour ajuster l'information. Bien que les Gated Recurrent Units (GRU²) soit plus rapides et légers, ils sont moins efficaces pour les relations temporelles longues, faisant de LSTM un choix robuste pour notre travail.

² GRU : Les GRU (Gated Recurrent Units) sont une version simplifiée des RNN qui utilisent deux portes pour gérer l'information, réduisant ainsi les paramètres et rendant l'entraînement plus rapide.

Développement d'un modèle de prévision à court terme de la puissance produite à la centrale solaire photovoltaïque de Nagréongo (30 MWc) dans le réseau national interconnecté du Burkina Faso.

IV.4 PREPARATION DES DONNEES POUR L'ENTRAINEMENT

Traitement de données

Pour ce qui est du prétraitement des données, nous en avons déjà parlé ci-dessus (voir [Préparation des données](#)) est le même pour la plupart des modèles de séries temporelles.

Sélection d'une fenêtre temporelle

Le choix d'une fenêtre de temps de 3 jours (432 points) permet de capturer les dépendances temporelles sur une période significative, tout en conservant une granularité³ suffisante pour prédire la production de puissance avec précision. Cette fenêtre permet au modèle LSTM d'analyser un historique représentatif tout en minimisant le bruit⁴.

Division des données

Les données ont été divisées en 80% pour l'entraînement et 20% pour le test, afin de garantir une bonne formation et évaluation du modèle. La validation croisée⁵ a été utilisée pour éviter les biais⁶ et tester plusieurs sous-ensembles, ce qui est essentiel dans les contextes énergétiques où les tendances saisonnières et irrégularités peuvent influencer les données.

IV.5 MODELE LSTM : PARAMETRES ET OPTIMISATION

Les hyperparamètres ont été choisis pour optimiser précision et régularisation. Le modèle est entraîné sur 14 **epochs** (qui constitue le nombre de fois que le modèle sera analysé), suffisant pour une bonne convergence sans surajustement. Un dropout de 20% aide à la généralisation. Le dropout est une technique de régularisation utilisée avec LSTM qui consiste à désactiver à chaque itérations un certain pourcentage de neurones. L'**early stopping** arrête l'entraînement si les performances ne s'améliorent plus. L'optimisation a été réalisée avec **Keras Tuner**, qui a testé différentes combinaisons de paramètres, comme la taille des lots et le taux d'apprentissage, afin de maximiser la précision sur les données de validation. Après cela, le modèle a été entraîné avec les meilleurs paramètres obtenus.

³ Granularité : finesse des détails

⁴ Bruit : variations aléatoires qui n'apportent pas de valeur ou de pertinence pour le modèle

⁵ Validation croisée : méthode pour évaluer un modèle en l'entraînant plusieurs fois sur des sous-ensembles différents des données pour tester sa performance de manière plus fiable.

⁶ Biais : lorsque le modèle commet des erreurs en raison de données mal équilibrées ou d'un entraînement insuffisant.

Développement d'un modèle de prévision à court terme de la puissance produite à la centrale solaire photovoltaïque de Nagréongo (30 MWc) dans le réseau national interconnecté du Burkina Faso.

PARTIE 3 : DEVELOPPEMENT DES MODELES

DE PREDICTIONS DE LA PUISSANCE

DELIVREE PAR LA CENTRALE

Après avoir établi les bases théoriques et méthodologiques, nous abordons désormais le développement des modèles de prévision de la puissance délivrée par la centrale solaire photovoltaïque de Nagréongo. Cette étape vise à appliquer les techniques de modélisation identifiées pour analyser les données disponibles et fournir des prévisions précises de la production énergétique.

I- PRÉVISION AVEC ARIMA

La prévision est un processus qui consiste dans le cas d'un apprentissage supervisé à utiliser des valeurs passées d'une base de données afin de prédire les valeurs futures de la même variable. Pour ce qui est de notre situation, nous avons la variable puissance à prédire dans le temps : on parle de série temporelle.

La transformation de Box-Cox a été appliquée à la puissance pour stabiliser la variance et rendre les données plus adaptées à la modélisation ARIMA. Ainsi, l'axe des ordonnées affiche la puissance transformée, et une conversion inverse serait nécessaire pour retrouver l'unité originale (MWc).

Développement d'un modèle de prévision à court terme de la puissance produite à la centrale solaire photovoltaïque de Nagréongo (30 MWc) dans le réseau national interconnecté du Burkina Faso.

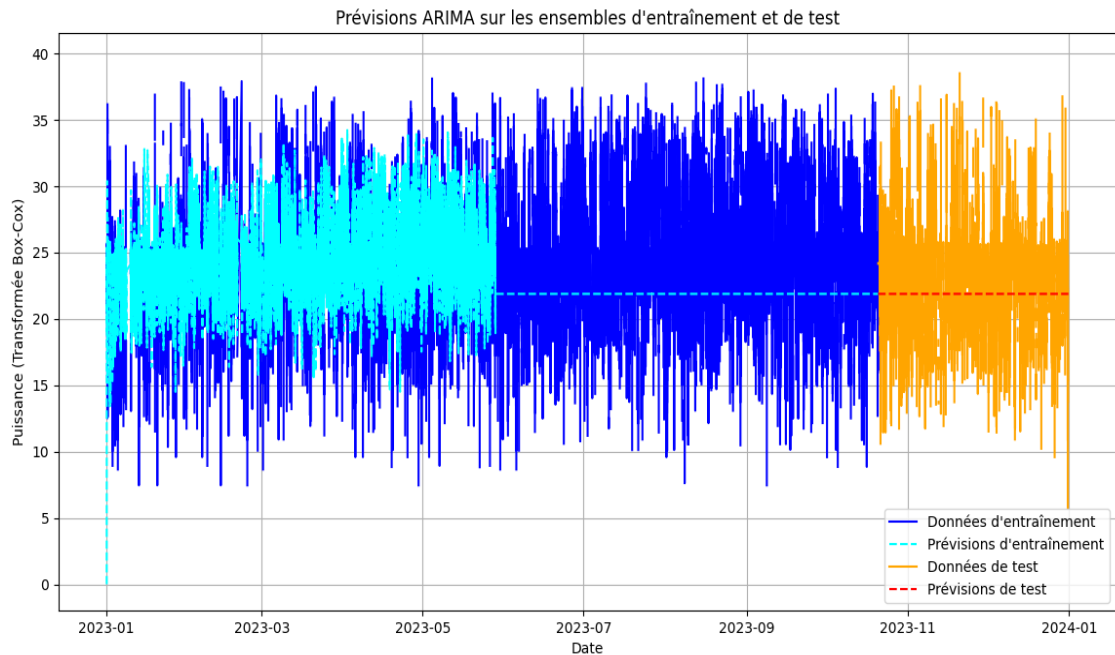


Figure 16: Prévision avec le modèle ARIMA

Remarque :

Comme nous pouvons le constater, le modèle produit des résultats insatisfaisants. En Machine Learning, lorsque les modèles ne sont pas des '**Bruits Blancs**', il devient difficile pour les méthodes statistiques traditionnelles de faire la prévision illustrant le principe de '**Garbage in, Garbage out**'. Cela implique que si les données ne sont pas bien prétraitées et modélisées, les prévisions seront de mauvaise qualité.

Il est également crucial de débiter par des modèles simples et interprétables avant d'explorer des techniques avancées, selon le principe de progression graduelle attribué à Robert Park.

Conclusion partielle

Dans un premier temps, nous avons développé le modèle ARIMA afin de capturer l'évolution de notre série. Cependant, ARIMA s'est révélé inadapté aux caractéristiques de la série, notamment en raison de sa non-normalité et son hétéroscédasticité. Ces deux éléments ont compromis la capacité du modèle à fournir des prévisions fiables.

À la suite de cet échec, un modèle VAR (Vector Autoregressive Model) a été utilisé. Il s'agit tout comme ARIMA, d'une approche statistique qui à la différence de ARIMA lui analyse les relations entre plusieurs variables, contrairement aux modèles univariés comme ARIMA. Toutefois, ce modèle a rencontré tout comme ARIMA, les mêmes limites. Ainsi, nous avons

Développement d'un modèle de prévision à court terme de la puissance produite à la centrale solaire photovoltaïque de Nagréongo (30 MWc) dans le réseau national interconnecté du Burkina Faso.

décidé de ne pas détailler davantage cette approche dans ce document.

Face à ces obstacles, nous avons jugé utile de nous orienter vers un modèle plus adapté, capable de contourner les hypothèses strictes de normalité et d'homoscédasticité. ; s'est alors imposé comme la méthode la mieux adaptée la Régression Quantile. Le développement de ce modèle est détaillé dans les sections suivantes de notre étude.

II- PRÉVISION AVEC RÉGRESSION QUANTILE

La régression quantile modélise les relations entre les variables en se basant sur différents quantiles d'une distribution, capturant ainsi les relations à divers points. Le quantile 0,5 correspond à la médiane, 0,9 et 0,1 respectivement aux points les plus élevés et les plus bas. Dans notre étude, nous avons testé plusieurs quantiles pour identifier celui qui convient le mieux à nos données.

Cette méthode minimise la fonction de perte pour chaque quantile τ en cherchant à déterminer les coefficients β qui l'optimisent.

$$L_{\tau} = (y, y') = \sum_{i=1}^n \rho_{\tau}(y_i - y_i') \quad (7)$$

Avec :

y_i la valeur observée

y_i' la valeur la valeur prédite

$\rho_{\tau}(u)$ la fonction de perte

$\rho_{\tau}(u) = u(\tau - 1)$ ($u < 0$)

II.1 PREDICTION QUANTILE DE LA MEDIANE (QUANTILE 0,5)

Le quantile 0,5 offre une estimation centrale, minimisant l'impact des valeurs extrêmes pour une représentation équilibrée. Nous avons divisé les données qui nous ont été fournies (données de production de la centrale de Nagréongo durant l'année 2023 avec un pas de temps de cinq minutes) en 80% pour l'entraînement et 20% pour les tests. Les résultats de la prévision de la Puissance en Watt sont présentés dans la Figure 17.

Développement d'un modèle de prévision à court terme de la puissance produite à la centrale solaire photovoltaïque de Nagréongo (30 MWc) dans le réseau national interconnecté du Burkina Faso.

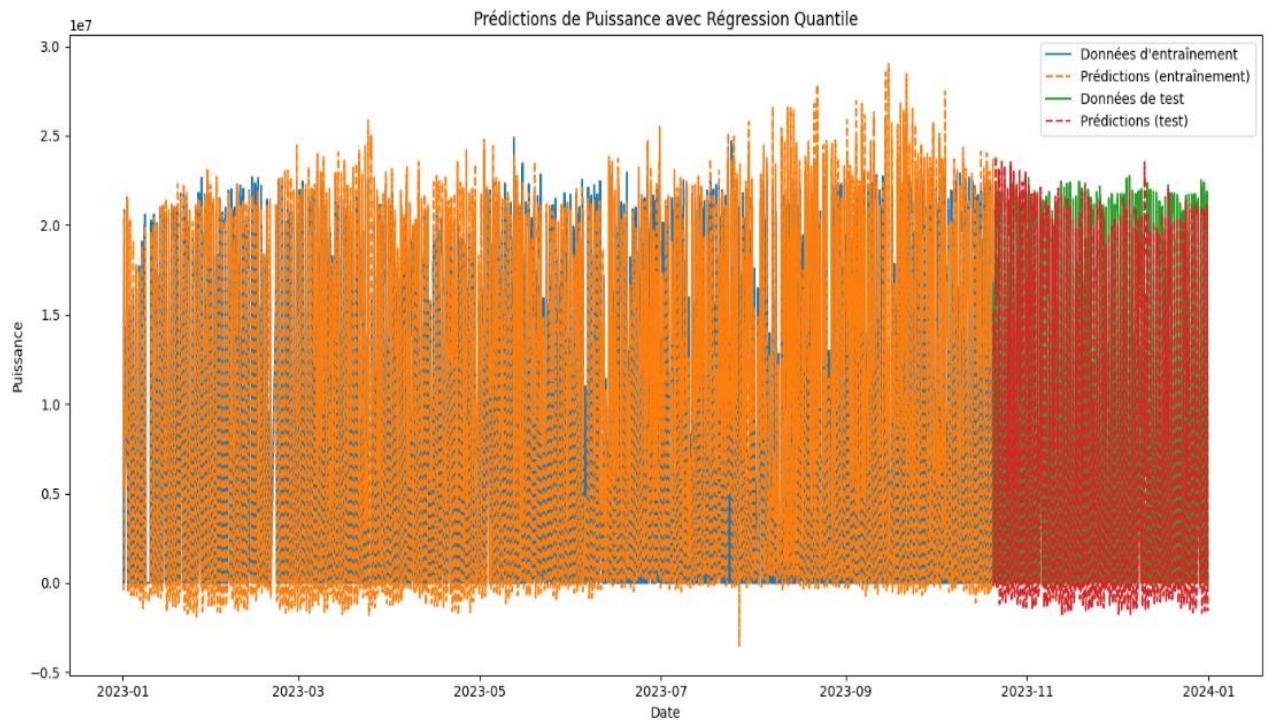


Figure 17: Prédiction de la Puissance avec le quantile 0,5

Les données de prévisions et d'entraînement étant assez proches indiquent une bonne correspondance. Les données de test aussi semblent avoir une assez bonne correspondance avec la prévision bien qu'il y ait quelques écarts plus importants que ceux de l'entraînement. **Une analyse de résidus permettra de comprendre les résultats et manquements observés.**

II.2 PREVISION REGRESSION QUANTILE (QUANTILE 0,1)

Le modèle vise à prédire le 10e percentile, soit une valeur inférieure à 90% des observations. Ce quantile est utile pour analyser les valeurs minimales, souvent utilisé pour modéliser les pires scénarios ou les valeurs faibles. Les résultats de la prévision sont les suivants:

Développement d'un modèle de prévision à court terme de la puissance produite à la centrale solaire photovoltaïque de Nagréongo (30 MWc) dans le réseau national interconnecté du Burkina Faso.

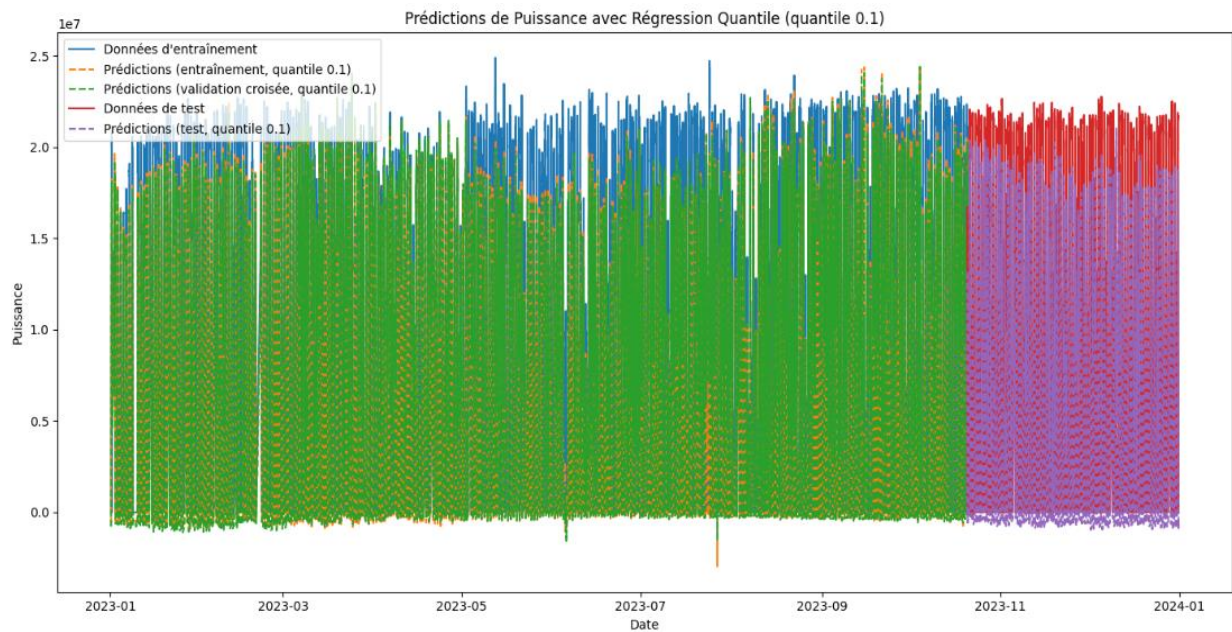


Figure 18: Prédiction de Puissance pour un quantile de 0,1

Les prévisions sont généralement orientées vers des valeurs plus faibles que les valeurs réelles. Même si la prévision a la même tendance que les données initiales, elle n'est pas assez précise. **Une analyse de résidus permettra de comprendre les résultats et manquements observés.**

II.3 PREVISION REGRESSION QUANTILE (QUANTILE 0,9)

Pour ce quantile, la régression s'articule autour de la partie supérieure de la base de données. Le modèle est conçu pour prédire les valeurs élevées, les 10% des valeurs les plus élevées de la distribution.

La prévision a donné les résultats suivants :

Développement d'un modèle de prévision à court terme de la puissance produite à la centrale solaire photovoltaïque de Nagréongo (30 MWc) dans le réseau national interconnecté du Burkina Faso.

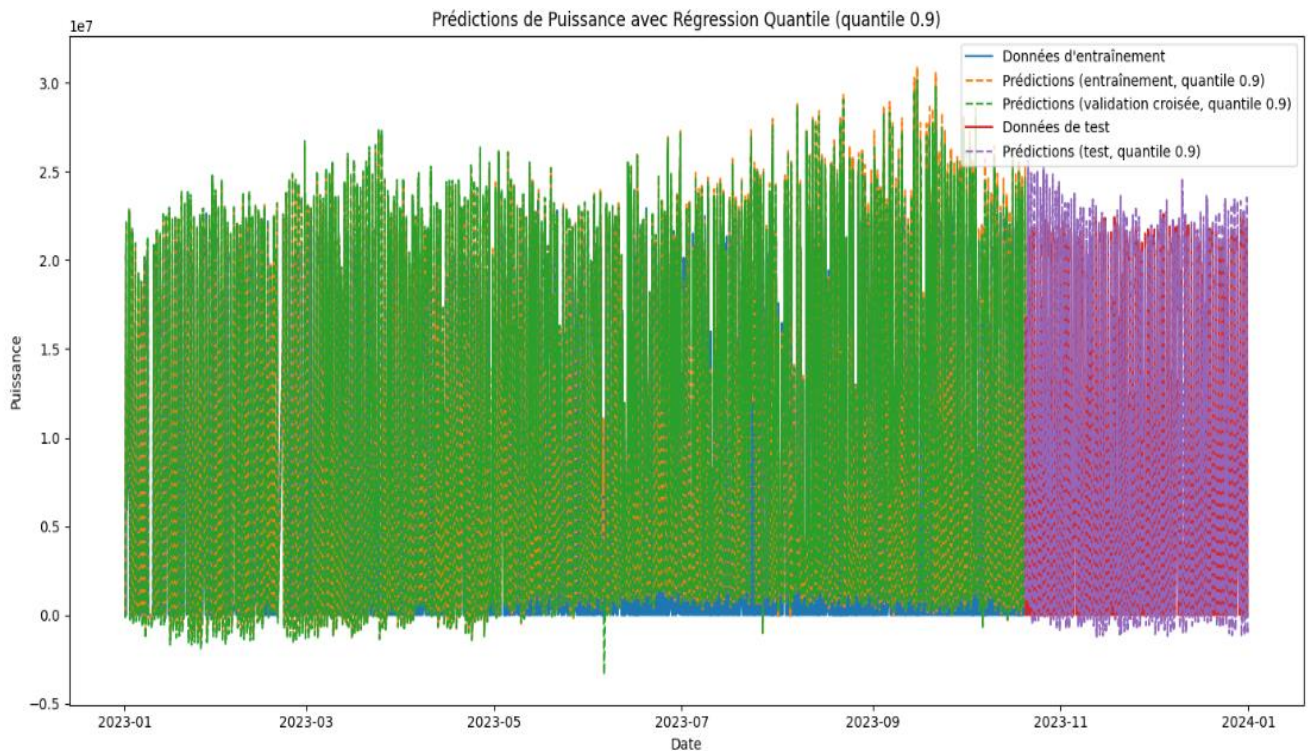


Figure 19: Prédiction de la Puissance pour un quantile de 0,9

Les prévisions pour le quantile 0,9 suivent globalement la tendance des données d'entraînement, s'alignant sur les valeurs élevées. Cependant, le modèle surestime parfois les valeurs, surtout pendant la phase de test. Cette difficulté à généraliser pour les valeurs extrêmes peut expliquer ces erreurs.

Une analyse des résidus sera nécessaire pour mieux comprendre les échecs de prévision.

II.4 ANALYSE DES RESIDUS

L'analyse des résidus dans une régression quantile aide à évaluer les prévisions pour différents quantiles. Les résidus représentent la différence entre les valeurs réelles et celles prédites et permet de voir comment le modèle gère les différents niveaux de la distribution (médian, bas, haut). Pour chaque quantile, les résidus nous ont permis de savoir si le modèle surestime ou sous-estime les données.

Résidus quantile de médiane (quantile 0,5)

Développement d'un modèle de prévision à court terme de la puissance produite à la centrale solaire photovoltaïque de Nagréongo (30 MWc) dans le réseau national interconnecté du Burkina Faso.

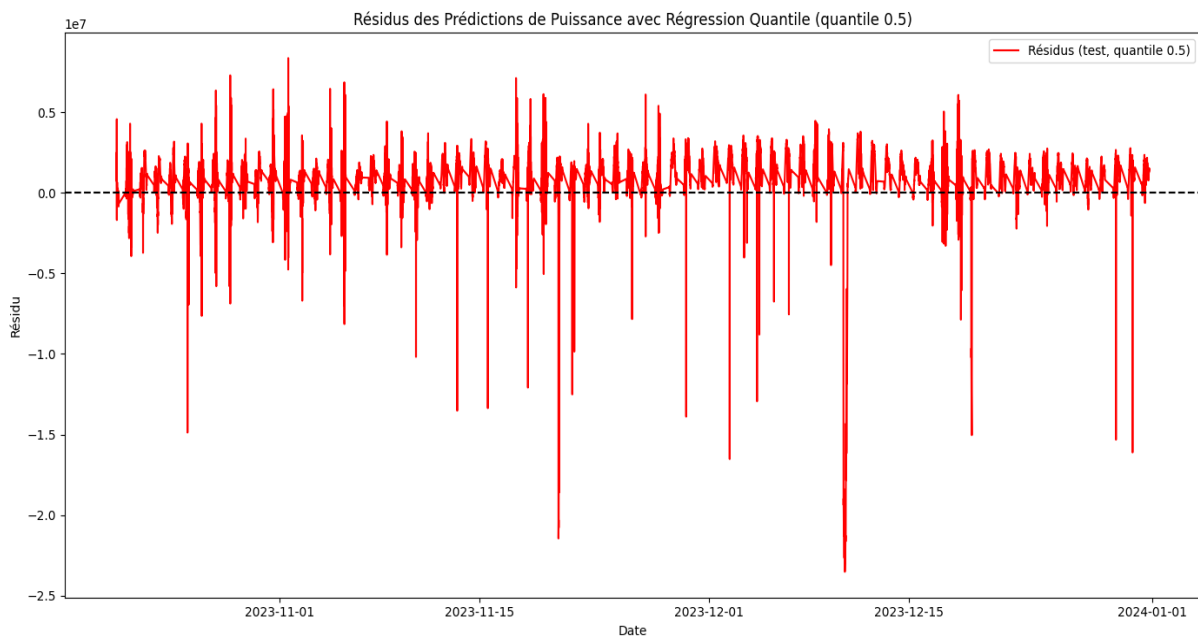


Figure 20: Résidus pour un quantile de 0,5

En statistique, il est expliqué que les résidus montrant un biais symétrique autour de zéro, sont ceux qui attendu pour une régression quantile de la médiane. Dans notre cas, des pics et des chutes sont observés, probablement dus à la non-normalité et l'hétéroscédasticité du modèle. Globalement, le graphique suit bien les prévisions médianes, mais peut rencontrer des difficultés avec les variations importantes.

Résidus Régression Quantile 0,1

Développement d'un modèle de prévision à court terme de la puissance produite à la centrale solaire photovoltaïque de Ngréongo (30 MWc) dans le réseau national interconnecté du Burkina Faso.

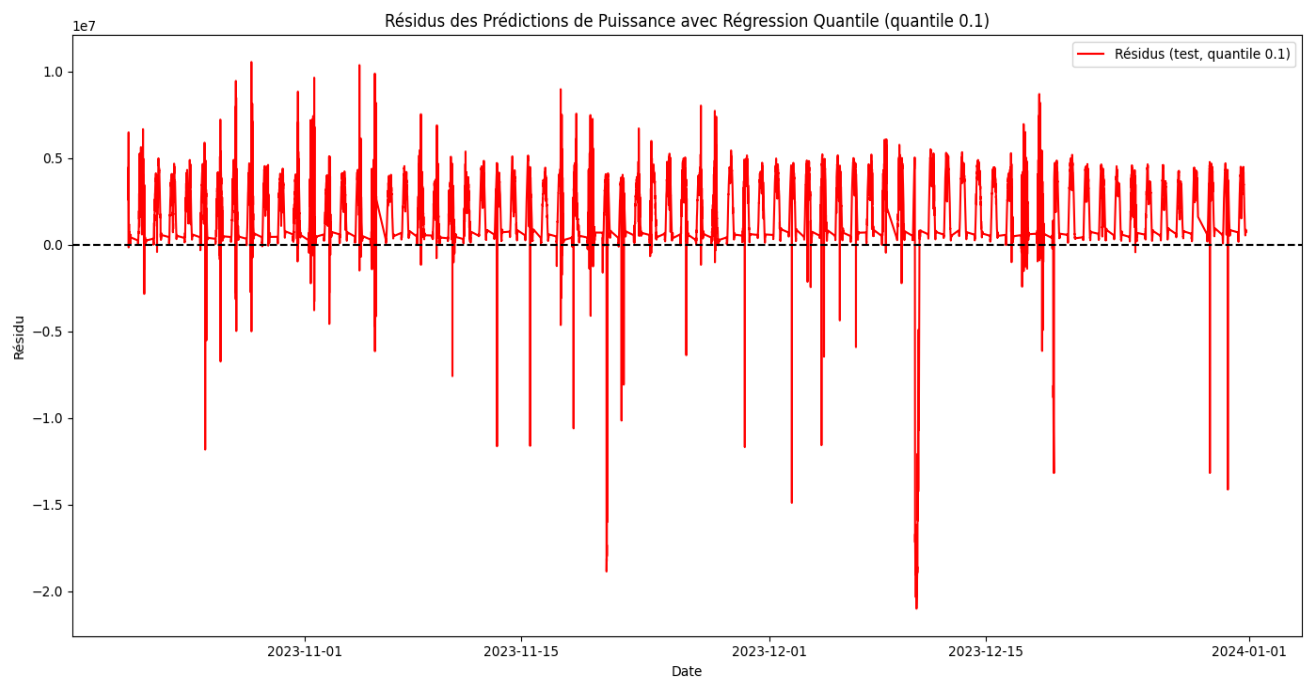


Figure 21: Résidus pour une prévision de quantile 0,1.

Pour ce qui est du quantile 0,1, les valeurs prédites sont largement inférieures à celles du modèle. Cela montre que le modèle sous-estime la puissance pour sur la majorité des données mais la surestime pour les valeurs situées à l'extrémité de la distribution.

Résidus quantile 0,9

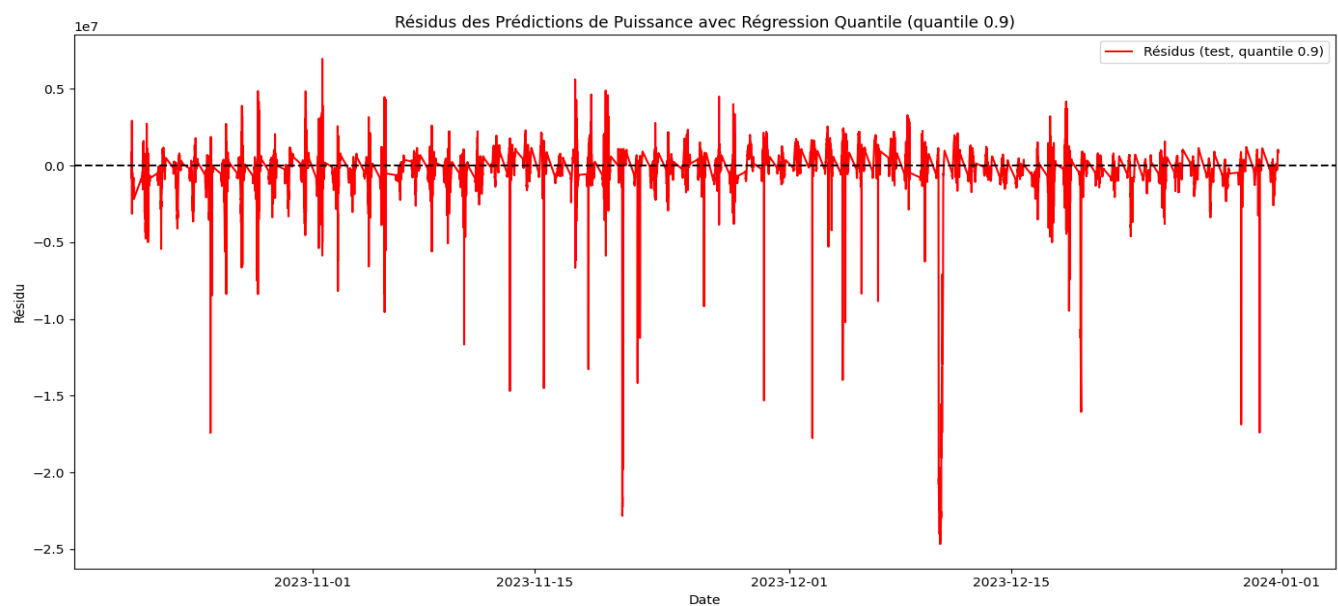


Figure 22: Résidus pour un quantile de 0,9

Pour le quantile 0,9, les valeurs observées sont généralement négatives, indiquant que le modèle

Développement d'un modèle de prévision à court terme de la puissance produite à la centrale solaire photovoltaïque de Nagréongo (30 MWc) dans le réseau national interconnecté du Burkina Faso.

surestime souvent les valeurs de la puissance. Quelques résidus positifs montrent que le modèle sous-estime rarement la partie élevée de la distribution. Pour choisir le quantile le mieux adapté, il est nécessaire d'évaluer leurs performances avec des métriques.

II.5 METRIQUES DE PERFORMANCES

Les résultats de nos différentes Régressions Quantiles fournissent un aperçu des performances des modèles de prévision de Puissance. Toutefois, il convient de prime abord, d'expliquer le rôle des différentes métriques utilisées.

Importance des métriques utilisées

MSE (Mean Squared Error) : mesure l'erreur quadratique moyenne. Elle prend en compte une grande pénalisation pour les grandes erreurs puisque élevée au carré. Il s'agit de la différence entre la prévision et les données réelles. Plus le MSE est petit, plus le modèle est précis.

$$\text{Formule : } \text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - y_i')^2 \quad (8)$$

Avec :

y_i : valeurs réelles de la distribution

y_i' : valeurs prédites par le modèle

n : nombres d'observations

Le calcul de la métrique de performances MSE nous a donné les résultats suivants :

Quantile 0,1 : 10.393.832.158.671,64

Quantile 0,5 : 5.562.870.449.671,63

Quantile 0,9 : 5.081.421.515.376,48

Le MSE a une valeur plus élevée pour le quantile 0,1. Les quantiles 0,5 et 0,9 ont des valeurs plus faibles indiquant ainsi une meilleure performance pour ces derniers.

MAE (Mean Absolute Error) : contrairement au MSE, le MAE calcule la moyenne des erreurs absolues et est moins sensibles aux extrêmes. Il donne une idée de l'erreur moyenne de prévision. Une faible valeur du MAE est un bon signe de l'ajustement du modèle.

$$\text{Formule : } \text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - y_i'| \quad (9)$$

NB : il s'agit des mêmes paramètres que MSE.

Le calcul de la métrique de performances MAE nous a donné les résultats suivants :

- Quantile 0,1 : 2.491.177,23

Développement d'un modèle de prévision à court terme de la puissance produite à la centrale solaire photovoltaïque de Nagréongo (30 MWc) dans le réseau national interconnecté du Burkina Faso.

- Quantile 0,5 : 1.384.151,49
- Quantile 0,9 : 986.948,96

L'erreur moyenne par point de chaque quantile est :

- Quantile 0,1 : 41,52
- Quantile 0,5 : 23,07
- Quantile 0,9 : 16,45

Le MAE a une valeur plus élevée pour le quantile 0,1. Les quantiles 0,5 et 0,9 ont des valeurs plus faibles indiquant ainsi une meilleure performance pour ces derniers.

MedAE : elle calcule la médiane des valeurs absolues des résidus, ce qui permet la performance centrale du modèle sans qu'elle en soit influencée par les valeurs aberrantes. Comme pour les autres métriques, une valeur faible est préférable.

Formule : $\text{MedAE} = \text{médiane } |y_i - y_i'|$

Le calcul de la métrique de performances MedAE nous a donné les résultats suivants :

Quantile 0,1 : 2.511.023,75

Quantile 0,5 : 1.090.465,48

Quantile 0,9 : 611.334,09

Le MAE a une valeur plus élevée pour le quantile 0,1. Les quantiles 0,5 et 0,9 ont des valeurs plus faibles indiquant ainsi une meilleure performance pour ces derniers.

NB : il s'agit des mêmes paramètres que les métriques ci-dessus.

R² : Le **R carré** (ou coefficient de détermination) est une mesure qui permet de savoir à quel point un modèle est capable d'expliquer la variabilité des données. Il donne une idée de la "qualité" de la prévision du modèle par rapport aux données réelles.

$$\text{Formule : } R^2 = 1 - \frac{\sum_{i=1}^n (y_i - y_i')^2}{\sum_{i=1}^n (y_i - y_m)^2} \quad (10)$$

Avec :

y_i : valeurs réelles de la distribution

y_i' : valeurs prédites par le modèle

n : nombres d'observations

y_m : Moyenne des valeurs réelles

Le calcul de la métrique de performances R² nous a donné les résultats suivants :

Quantile 0,1 : 0.8449

Quantile 0,5 : 0.9170

Développement d'un modèle de prévision à court terme de la puissance produite à la centrale solaire photovoltaïque de Nagréongo (30 MWc) dans le réseau national interconnecté du Burkina Faso.

Quantile 0,9 : 0.9242

Cette métrique représente le pourcentage de précision ou la qualité de l'ajustement du modèle.

R² et la précision du modèle : Le R² mesure la proportion de la variabilité totale de la variable dépendante (la variable que nous prédisons) qui est expliquée par le modèle. Il indique donc combien de pourcentage de la variabilité des données est "capturé" par le modèle. Puisque nous avons **R² = 0,9242**, cela signifie que **92,42%** de la variabilité des données est expliquée par le modèle, et les **7,58%** restants ne sont pas expliqués par le modèle. Les **7,58% restants** représentent les erreurs (ou les résidus) qui ne peuvent pas être expliquées par le modèle. Autrement dit, ces erreurs correspondent à la variance qui reste dans les données, mais qui n'a pas été capturée par les prévisions du modèle.

Conclusion partielle :

En conclusion, le quantile 0,9 est celui qui capture le mieux les paramètres du modèle. Nous l'utiliserons pour les tests avec les données réelles non utilisées dans la modélisation. Après avoir exploré ARIMA, VAR et la régression quantile, nous allons maintenant explorer le LSTM, un modèle d'apprentissage profond (Deep Learning) adaptés aux séries temporelles, connus pour leur capacité à capturer des relations complexes et non linéaires. L'objectif est de comparer leurs performances et de déterminer le modèle le mieux adapté à nos données.

III- PRÉVISION AVEC LSTM

III.1 ANALYSE DES PERFORMANCES

L'analyse des résultats montre que le modèle LSTM prédit efficacement la production énergétique. Les figures ci-dessous comparent les valeurs prédites et réelles sur l'ensemble de test, offrant une visualisation des écarts et des tendances capturées par le modèle.

Développement d'un modèle de prévision à court terme de la puissance produite à la centrale solaire photovoltaïque de Nagréongo (30 MWc) dans le réseau national interconnecté du Burkina Faso.

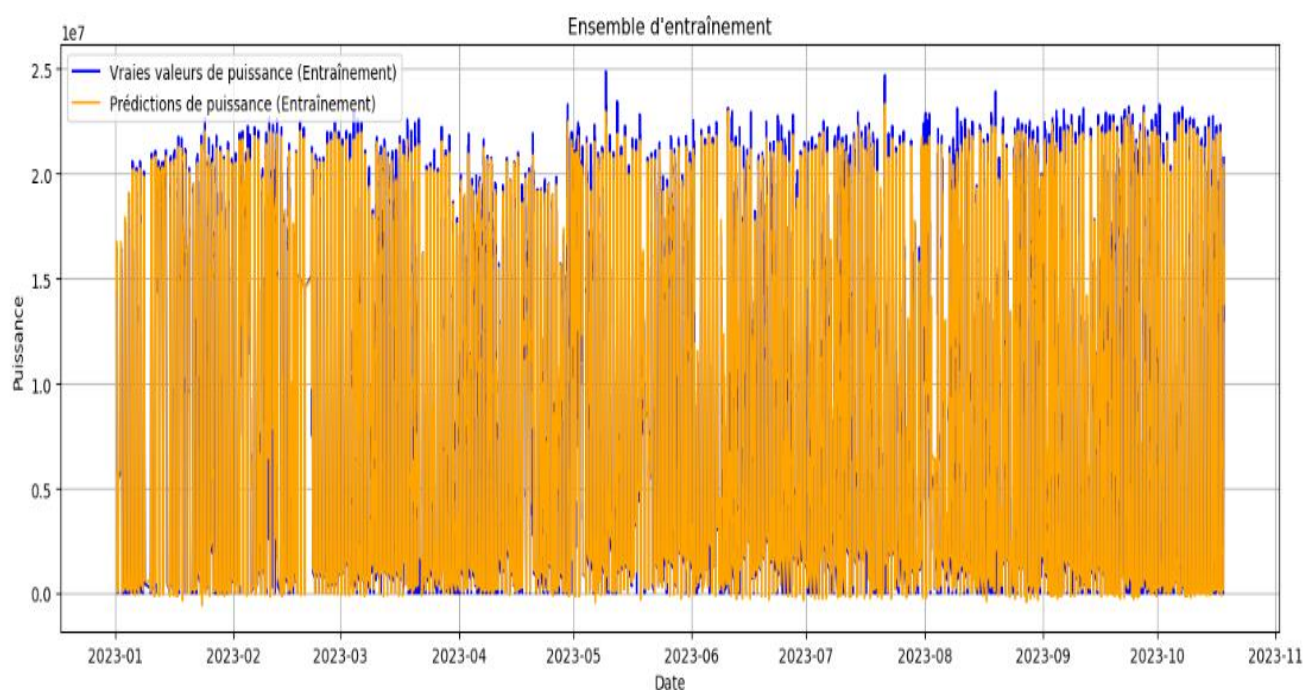


Figure 23: Ensemble d'entraînement modèle LSTM

La Figure 23 montre les prévisions et les valeurs réelles sur l'ensemble d'entraînement. Les prévisions (ligne orange) collent de près aux valeurs réelles (ligne bleue), ce qui indique une très bonne adaptation du modèle sur ces données.

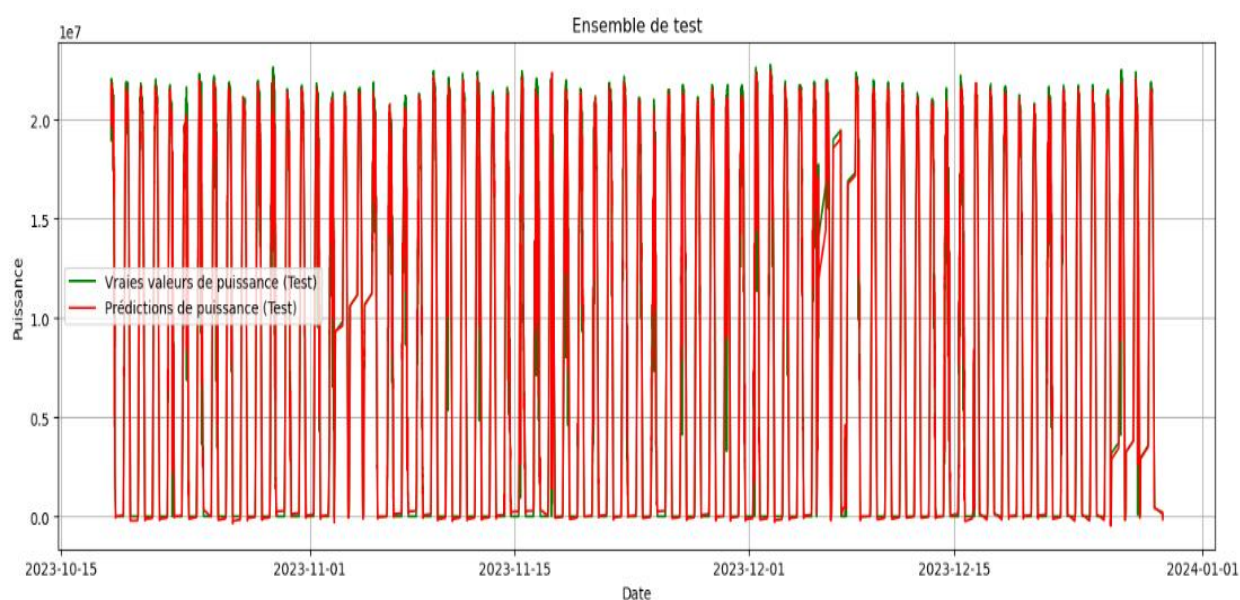


Figure 24: Ensemble de test du modèle LSTM

La Figure 24 présente les prévisions du modèle LSTM comparées aux valeurs réelles de puissance sur l'ensemble de test. Les prévisions (ligne rouge) suivent globalement les tendances

Développement d'un modèle de prévision à court terme de la puissance produite à la centrale solaire photovoltaïque de Ngréongo (30 MWc) dans le réseau national interconnecté du Burkina Faso.

des valeurs réelles (ligne verte).

Ces visualisations confirment que le modèle est performant dans la reproduction des tendances générales, mais elles mettent également en évidence des marges d'amélioration, surtout sur l'ensemble de test, pour mieux capturer les variations locales.

III.2 ANALYSE AVEC LES METRIQUES DE PERFORMANCES

Les performances du modèle LSTM ont été évaluées à l'aide des métriques de référence (MSE, MAE et R^2) sur les ensembles d'entraînement et de test. Ces résultats permettent de quantifier la précision et la robustesse du modèle dans la prévision de la puissance.

Les métriques utilisées pour évaluer les performances du modèle LSTM sont le MAE et le R^2 .

MAE (Mean Absolute Error) : Il mesure la moyenne des erreurs absolues, offrant une interprétation claire de l'erreur moyenne. Une valeur faible indique des prévisions proches des valeurs réelles.

Développement d'un modèle de prévision à court terme de la puissance produite à la centrale solaire photovoltaïque de Nagréongo (30 MWc) dans le réseau national interconnecté du Burkina Faso.

R² (Coefficient de détermination) : Il mesure la proportion de la variance expliquée par le modèle. Une valeur proche de 1 indique un bon ajustement, tandis qu'une valeur proche de 0 montre un faible pouvoir explicatif.

Tableau V : Métriques de performances du modèle LSTM.

➤ **Interprétation :**

Métrique	Rôle	Ensemble d'entraînement	Ensemble de test
MAE	Mesure la moyenne des valeurs absolues des écarts.	692 188,09	555 087,37
R²	Indique la proportion de variance expliquée par le modèle.	0,96	0.98

La MAE (Erreur Absolue Moyenne) est une métrique qui permet de mesurer l'écart moyen entre les valeurs prédites et réelles, sans tenir compte de la direction de l'erreur. Elle est calculée de la manière suivante :

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_{réel,i} - y_{prédit,i}| \quad (11)$$

Total de points = 105 120

Ensemble d'entraînement : 80% de 105 120 = 84 096 points

Ensemble de test : 20% de 105 120 = 21 024 points

MAE de l'entraînement : 692 188,09 W

MAE du test : 555 087,37 W

$$MAE \text{ Ensemble d'entraînement} = \frac{708349,66873}{84096} \approx 8,42 \text{ W}$$

$$MAE \text{ Ensemble de test} = \frac{493385,247171}{21024} \approx 23,46 \text{ W}$$

Calcul global de la MAE :

$$MAE \text{ totale} = (8,42 \times 0,8) + (23,4 \times 0,2) = 6,58 + 5,28 = 11,45 \text{ W}$$

Conclusion partielle:

La **MAE globale** (en tenant compte de la proportion des points d'entraînement et de test) est de **11,46 W**.

Développement d'un modèle de prévision à court terme de la puissance produite à la centrale solaire photovoltaïque de Ngréongo (30 MWc) dans le réseau national interconnecté du Burkina Faso.

Un R^2 plus élevé sur le test indique que le modèle capture bien les grandes tendances des données, même si des erreurs plus importantes se produisent à certains moments.

Le **MAE plus élevé** sur le test peut simplement refléter des erreurs locales plus importantes à certains moments, mais cela n'empêche pas le modèle de suivre globalement les tendances.

C'est donc un phénomène normal qui peut se produire, particulièrement lorsque le modèle capte bien les grandes tendances mais a des erreurs sur des moments précis ou des événements rares.

A la suite d'une analyse comparative des performances des différents modèle développé, **notre choix s'est finalement porté sur le modèle LSTM** en raison de sa capacité à capturer la complexité et la non-linéarité des données, offrant ainsi de meilleures performances prédictives.

Développement d'un modèle de prévision à court terme de la puissance produite à la centrale solaire photovoltaïque de Ngréongo (30 MWc) dans le réseau national interconnecté du Burkina Faso.

PARTIE 4 : VALIDATION DU MODELE

La validation du modèle de prévision s'articule autour de quelques étapes, intégrant la mise en place d'une interface. Elle inclut non seulement une évaluation basée sur des données testées, mais aussi une évaluation dynamique à travers une interface interactive permettant de comparer les prévisions aux valeurs réelles. Cette partie décrit les différentes étapes de validation et de mise en œuvre de l'interface, ainsi que l'importance d'une amélioration continue du modèle pour garantir sa précision et sa pertinence au fil du temps.

I. PREVISION SUR DE NOUVELLES DONNEES

Une étape clé dans la validation d'un modèle de prévision est de tester ses performances sur des données qui n'ont pas été utilisées lors de l'entraînement ou de la validation initiale. Ainsi, des prévisions ont été réalisées sur des **données nouvelles** couvrant la période de début 2024, afin de vérifier la capacité du modèle à **généraliser** ses résultats à de nouvelles situations.

Données nouvelles : Ces données, non incluses dans l'ensemble d'entraînement ou de test, permettent de simuler des conditions réelles de prévision. Elles permettent de tester la réactivité et la robustesse du modèle face à des conditions imprévues.

Comparaison des résultats : Les résultats obtenus ont été comparés aux **valeurs réelles** mesurées pour la période 2024. Ce processus de **validation externe** permet de s'assurer que le modèle ne s'est pas "surajusté" (overfitting) aux données d'entraînement, mais qu'il peut prédire avec précision des données non vues précédemment.

Les détails de cette validation sont disponibles en Annexe (Voir [Annexe XII](#)).

II. DEVELOPPEMENT D'UNE INTERFACE DE PREVISION

II.1 CONCEPTION DE L'INTERFACE

Pour faciliter la validation continue et la mise en production du modèle, une **interface interactive** a été développée. Cette interface permet aux utilisateurs de comparer **en temps réel** les prévisions du modèle LSTM avec les **valeurs réelles** qu'ils saisissent manuellement.

L'interface utilisateur a été pensée pour être à la fois **intuitive** et **dynamique**, avec des fonctionnalités clés permettant une interaction simple et une validation rapide des prévisions.

Les éléments suivants ont été intégrés dans l'interface :

Zone de saisie de la date : Cette zone permet à l'utilisateur d'entrer une ou plusieurs **dates**

Développement d'un modèle de prévision à court terme de la puissance produite à la centrale solaire photovoltaïque de Nagréongo (30 MWc) dans le réseau national interconnecté du Burkina Faso.

spécifiques (au format Année-Mois-Jour-Heure-Minute). Ces dates peuvent correspondre à des moments précis pour lesquels l'utilisateur souhaite obtenir des prévisions de production de puissance. Ce mécanisme de saisie permet d'effectuer des prévisions sur n'importe quelle période de production.

Bouton "Prédire" : Ce bouton, une fois cliqué, lance le modèle LSTM préalablement entraîné pour effectuer la **prévision** de la production de puissance pour les dates fournies. Cela permet de simuler des prévisions en temps réel, sans avoir à relancer des calculs complexes à chaque fois.

Zone d'entrée des valeurs réelles : Une fois la prévision générée, l'utilisateur peut entrer les **valeurs réelles** mesurées pour la même période (si ces données sont disponibles). Cela permet une comparaison directe entre les valeurs prédites et réelles, essentielle pour évaluer la précision du modèle dans un environnement opérationnel.

Zone d'affichage des résultats : Après la prévision, cette section affiche les résultats sous forme de texte et de graphiques. Les valeurs **prédites** et **réelles** sont affichées côte à côte, et un **écart** entre les deux valeurs est calculé et présenté. Cela permet à l'utilisateur de voir rapidement l'adéquation entre la prévision et la réalité, et de mieux comprendre la performance du modèle.

Visualisation graphique : Un graphique interactif est affiché pour donner une vue d'ensemble de la **comparaison entre les prévisions et les valeurs réelles**. Cela permet aux utilisateurs d'avoir une vue visuelle claire de la précision du modèle sur des périodes spécifiques, facilitant ainsi l'analyse des résultats.

La figure 25 est une capture d'écran de l'interface développée.

Développement d'un modèle de prévision à court terme de la puissance produite à la centrale solaire photovoltaïque de Nagréongo (30 MWc) dans le réseau national interconnecté du Burkina Faso.

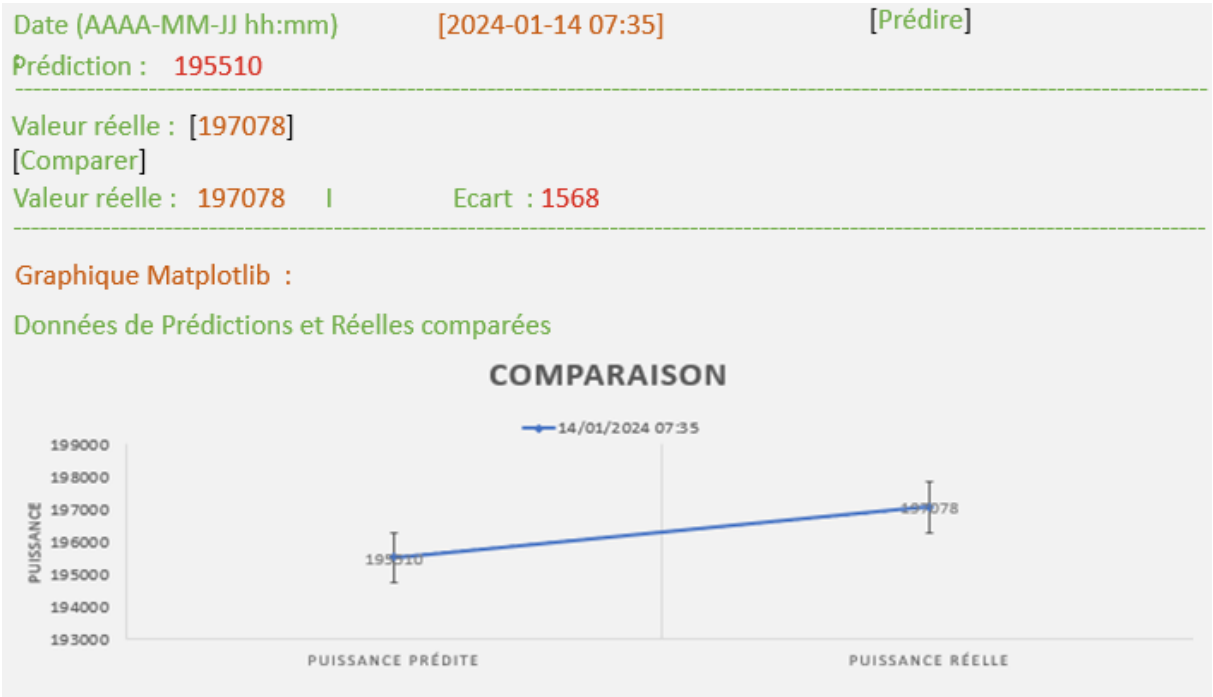


Figure 25: Validation du modèle

II.2 ETAPES DE VALIDATION DE L’INTERFACE

La validation de l’interface implique plusieurs étapes clés :

Saisie de la date : L’utilisateur entre la date souhaitée pour laquelle une prévision est à réaliser. Le format précis de la date (Année-Mois-Jour-Heure-Minute) est important pour garantir la précision de la prévision sur des périodes spécifiques.

Génération des prévisions : Le modèle LSTM, déjà chargé dans l’interface, génère la **prévision** pour la date saisie. Le modèle prend en compte les tendances historiques et les conditions réelles de la période choisie pour fournir une estimation de la production d’énergie.

Entrée des valeurs réelles : Après l’affichage de la prévision, l’utilisateur peut saisir les **valeurs réelles** observées pendant la même période. Si ces données réelles sont disponibles, elles sont introduites dans l’interface pour être comparées aux valeurs prédites.

Calcul des écarts : L’interface calcule l’écart entre la **prévision** et la **valeur réelle**. Cette différence est ensuite affichée dans un tableau récapitulatif ou sous forme graphique, permettant

Développement d'un modèle de prévision à court terme de la puissance produite à la centrale solaire photovoltaïque de Nagréongo (30 MWc) dans le réseau national interconnecté du Burkina Faso.

à l'utilisateur de vérifier immédiatement la précision de la prévision du modèle.

Affichage des résultats : Le calcul des écarts entre les valeurs réelles et les valeurs prédites est présenté sous forme visuelle (graphiques) et textuelle, ce qui permet de faire une analyse immédiate de la performance du modèle sur un échantillon donné de données.

II.3 AMELIORATION CONTINUE

Une fois l'interface mise en place et validée, il est essentiel de mettre en place un processus d'**amélioration continue** pour garantir que le modèle reste performant au fil du temps.

Intégration de nouvelles données : Pour maintenir la précision des prévisions, des données réelles issues de la production d'énergie sont continuellement intégrées dans le modèle. Cette mise à jour régulière permet de recalibrer le modèle afin qu'il puisse s'adapter à de nouveaux schémas ou à des changements dans les conditions de production (par exemple, des modifications dans l'ensoleillement ou l'efficacité des panneaux solaires).

Réajustement des hyperparamètres : En fonction des nouvelles données collectées, il est possible d'ajuster les **hyperparamètres** du modèle LSTM. Par exemple, si de nouvelles tendances sont observées dans les données (comme des variations saisonnières), l'ajustement des paramètres pourrait améliorer la capacité du modèle à prédire avec plus de précision.

Adaptation aux évolutions saisonnières : Le modèle peut être réévalué à la lumière des évolutions saisonnières (par exemple, l'ensoleillement changeant entre les saisons sèches et pluvieuses). L'ajustement des données et des hyperparamètres en fonction de ces changements permet d'optimiser le modèle pour chaque période de l'année.

Développement d'un modèle de prévision à court terme de la puissance produite à la centrale solaire photovoltaïque de Ngréongo (30 MWc) dans le réseau national interconnecté du Burkina Faso.

CONCLUSION PARTIELLE

Ce travail a abouti au développement d'un modèle LSTM performant pour la prévision de la production énergétique solaire. En exploitant les caractéristiques des séries temporelles, le modèle a montré sa capacité à fournir des prévisions précises, surpassant des approches classiques comme ARIMA et la régression quantile.

Grâce à des outils comme Keras Tuner et des techniques de régularisation, le modèle a été optimisé pour éviter le surapprentissage et garantir sa robustesse. L'ajout d'une interface interactive facilite son utilisation et permet de comparer les prévisions avec des valeurs réelles, renforçant son applicabilité.

Des pistes d'amélioration, notamment l'intégration de nouvelles données ou variables explicatives, offriront des opportunités pour affiner encore les performances et répondre aux besoins futurs en gestion énergétique.

Développement d'un modèle de prévision à court terme de la puissance produite à la centrale solaire photovoltaïque de Ngréongo (30 MWc) dans le réseau national interconnecté du Burkina Faso.

CONCLUSION

La présente étude avait pour but de prédire la production de la puissance de la centrale solaire photovoltaïque de Ngréongo, en explorant trois méthodologies complémentaires : le modèle ARIMA, la régression quantile et les réseaux de neurones LSTM. Chacune des approches ayant apporté un aperçu distinct sur la dynamique des données, offrant ainsi des perspectives variées de l'optimisation des prévisions.

ARIMA, conçu à la base pour des séries temporelles stationnaires, a montré ses limites face à la complexité non linéaire et à la variabilité hétéroscédastique des données de production de la centrale. Bien qu'il soit performant dans la capture de tendances et de comportements cycliques simples, il peine à s'adapter aux fluctuations plus complexes inhérentes aux séries temporelles photovoltaïques, réduisant ainsi son efficacité pour les prévisions à long terme.

La régression quantile, de son côté, a permis de mieux appréhender l'incertitude associée aux prévisions, en fournissant des intervalles de prévision robustes. Toutefois, sa capacité à modéliser les dépendances temporelles entre les observations successives reste limitée, ce qui compromet sa performance pour des prévisions précises sur des horizons longs et complexes.

En revanche, les réseaux neuronaux LSTM ont démontré une remarquable supériorité, grâce à leur aptitude à saisir les relations non linéaires et les dépendances temporelles présentes dans les séries de données. L'utilisation de mécanismes avancés tels que le dropout pour la régularisation et l'early stopping pour prévenir le surapprentissage a permis d'optimiser la performance du modèle, tout en garantissant une capacité de généralisation exceptionnelle. Lors de l'entraînement, le modèle LSTM a atteint une performance de **96%** de précision, tandis que sur l'ensemble de test, cette précision a atteint **98%**, ce qui atteste de sa robustesse et de sa capacité à généraliser efficacement sur de nouvelles données.

Ces résultats ont été validés non seulement sur les ensembles d'entraînement et de test, mais aussi sur des données extérieures, confirmant ainsi la robustesse du modèle face à de nouvelles informations. Une autre avancée notable a été le développement d'une interface interactive permettant aux utilisateurs de comparer les prévisions du modèle avec les valeurs réelles, saisies

Développement d'un modèle de prévision à court terme de la puissance produite à la centrale solaire photovoltaïque de Nagréongo (30 MWc) dans le réseau national interconnecté du Burkina Faso.

manuellement. Cette fonctionnalité simplifie l'évaluation en temps réel et rend l'utilisation du modèle plus accessible et intuitive.

En conclusion, les résultats obtenus montrent que, des méthodes choisies, les réseaux neuronaux LSTM constituent des trois, la méthode la plus appropriée pour la prévision de la production de puissance de la centrale solaire photovoltaïque de Nagréongo.

Développement d'un modèle de prévision à court terme de la puissance produite à la centrale solaire photovoltaïque de Nagréongo (30 MWc) dans le réseau national interconnecté du Burkina Faso.

RECOMMANDATIONS

Pour améliorer davantage les performances du modèle et répondre aux besoins évolutifs, plusieurs pistes sont recommandées :

- 1- **Intégration de nouvelles variables explicatives** : Ajouter des données telles que l'humidité, la pression atmosphérique ou la qualité de l'air pour enrichir les entrées du modèle.
- 2- **Collecte et utilisation des données pour les jours et semaines à venir** : Ces données permettront de mieux capturer les tendances récentes et d'adapter le modèle à des évolutions potentielles des conditions de production.
- 3- **Expérimentation avec des architectures avancées** : Explorer des modèles hybrides combinant LSTM avec d'autres approches comme les réseaux de neurones convolutionnels (CNN) pour améliorer la capture des patterns complexes.
- 4- **Optimisation continue** : Affiner régulièrement les hyperparamètres à mesure que de nouvelles données deviennent disponibles et que les besoins évoluent.
- 5- **Validation croisée plus étendue** : Tester le modèle sur des ensembles de données variés pour s'assurer de sa robustesse dans différents contextes.

Ces recommandations visent à renforcer la pertinence, la précision et la fiabilité des prévisions, soutenant ainsi une gestion optimale de la production d'énergie solaire.

Développement d'un modèle de prévision à court terme de la puissance produite à la centrale solaire photovoltaïque de Nagréongo (30 MWc) dans le réseau national interconnecté du Burkina Faso.

BIBLIOGRAPHIE

- [1] Ministère de l'énergie, « PLAN DIRECTEUR NATIONAL PRODUCTION-TRANSPORT-DISTRIBUTION ET D'ELECTRIFICATION RURALE 2017 – 2025 ». 2017.
<https://faolex.fao.org/docs/pdf/bkf219602.pdf>
- [2] Box, G. E. P., & Jenkins, G. M. «Time Series Analysis: Forecasting and Control (Revised Edition) ». San Francisco: Holden-Day. ISBN: 978-0816211043
- [3] Akaike, H. (1998). «Information Theory and an Extension of the Maximum Likelihood Principle». In E. Parzen, K. Tanabe, & G. Kitagawa (Eds.), *Selected Papers of Hirotugu Akaike* (pp. 199–213). New York, NY: Springer. doi: 10.1007/978-1-4612-1694-0_15
- [4] C. Voyant, « Prévision de séries temporelles de rayonnement solaire global et de production d'énergie photovoltaïque à partir de réseaux de neurones artificiels », vol. 1, p. 258, oct. 2011.
<https://theses.hal.science/tel-00635298v1/document>
- [5] A. Fouilloy, « Comparaison de méthodes d'apprentissage automatique de prévision de la ressource solaire pour une application à une gestion optimisée des réseaux intelligents ». <https://theses.hal.science/tel-03479238v1/document>
- [6] C.-J. Huang et P.-H. Kuo, « Multiple-Input Deep Convolutional Neural Network Model for Short-Term Photovoltaic Power Forecasting », *IEEE Access*, vol. 7, p. 74822-74834, 2019, doi: 10.1109/ACCESS.2019.2921238.
<https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8732333>
- [7] J. Li et Q. Liu, « Short-term Photovoltaic Power Forecasting Using SOM-based Regional Modelling Methods », *Chin. J. Electr. Eng.*, vol. 9, n° 1, p. 158-176, mars 2023, doi: 10.23919/CJEE.2023.000004.
<https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10093784>

Développement d'un modèle de prévision à court terme de la puissance produite à la centrale solaire photovoltaïque de Nagréongo (30 MWc) dans le réseau national interconnecté du Burkina Faso.

[8] M. Alaraj, A. Kumar, I. Alsaidan, M. Rizwan, et M. Jamil, « Energy Production Forecasting From Solar Photovoltaic Plants Based on Meteorological Parameters for Qassim Region, Saudi Arabia », *IEEE Access*, vol. 9, p. 83241-83251, 2021, doi: 10.1109/ACCESS.2021.3087345.

https://www.researchgate.net/publication/352243603_Energy_Production_Forecasting_From_Solar_Photovoltaic_Plants_Based_on_Meteorological_Parameters_for_Qassim_Region_Saudi_Arabia#:~:text=Here%2C%20in%20this%20paper%2C%20the%20ensemble%20trees%20approach-based,Qassim%2C%20Saudi%20Arabia%20is%20used%20in%20this%20research.

[9] Bastien Dussap, Antoine Faul, « Projet : Prévision solaire ». 2020.

https://www.imo.universite-paris-saclay.fr/~yannig.goude/Materials/ProjetMLF/Rapports2021/Solar_Projet_ML-3.pdf

[10] Brockwell, P. J., & Davis, R. A. (1991). *Time Series: Theory and Methods* (2nd ed.). Springer.

[11] Guillaume Saint-Cirgue, « Apprendre le Machine Learning en une semaine. ». Eyrolles, 2020.

[12] J. Heaton, « Ian Goodfellow, Yoshua Bengio, and Aaron Courville: Deep learning: The MIT Press, 2016, 800 pp, ISBN: 0262035618 », *Genet. Program. Evolvable Mach.*, vol. 19, n° 1-2, p. 305-307, juin 2018, doi: 10.1007/s10710-017-9314-z.

<https://link.springer.com/article/10.1007/s10710-017-9314-z>

ANNEXES

Annexe I : Synthèse bibliographique

Auteur	Thème	Contexte	Objectifs	Dans quelle mesure s'apparente-il au thème de l'étude	Méthodologie	Résultats	Degré de précision (sur 5 étoiles)
Cyril Voyant [4]	Prévision de séries temporelles de rayonnement solaire global et de production d'énergie photovoltaïque à partir de réseaux de neurones artificiels	La Corse faisant partie des petits réseaux non-interconnectés, son approvisionnement énergétique est très particulier du fait qu'elle soit une île. Une solution souvent utilisée pour pallier cet isolement est les	- Caractérisation de la ressource solaire - Quantifier l'énergie disponible. - Fournir pour chaque horizon de prévision considéré une méthode avec	L'auteur fait des prévisions dans différents horizons temporels et fourni pour chaque horizon un modèle de prévision adapté. Il s'apparente à notre étude puisqu'il s'agit	Pour l'horizon j+1, réseaux de neurones. Pour l'horizon h+1, une méthodologie hybride ARIMA permet d'obtenir des résultats très satisfaisants. Pour le cas h+24, les réseaux de	Pour j+1, la méthode ARIMA et RNA donne de bons résultats. Toutefois, la stationnarisation apporte plus de précision au modèle RNA. Pour h+1, la combinaison de RNA et ARIMA	****

		énergies renouvelables. Ainsi, afin de pallier aux intermittences, il convient de quantifier l'énergie disponible et permettre une gestion optimale.	un degré de précision élevé	de la prévision dans les deux cas et nous pouvons grâce à son étude connaître le modèle le mieux adapté à notre horizon temporel très court terme.	neurones à sorties multiples donnent de très bons résultats. Concernant l'horizon m+5, les conclusions sont moins catégoriques.	diminue l'erreur de 3,5% Pour m+5, l'utilisation de K-mean permet de diminuer l'erreur de 0,7%	
Alexis Fouilloy [5]	Comparaison de méthodes d'apprentissage automatique de prévision de la ressource solaire pour une application à une gestion optimisée des réseaux	Les enjeux relatifs à la production énergétique future, notamment en termes d'utilisation de ressources locales et « propres », conduisent les producteurs d'électricité à se tourner de plus en	Connaître à l'avance l'énergie disponible et de permettre une gestion optimale du couplage entre systèmes de production conventionnels	L'auteur teste des méthodes de prévision à court terme dans le but de choisir celui avec le moins d'erreur Nrsme et de ce fait adapté aux conditions météorologiques	-Traitement des données météorologiques avec la méthode de ciel clair -Répartition de série temporelle en deux groupes (80% pour l'apprentissage et 20% pour le	Pour des pas de temps de 1 à 10 min, la régression linéaire donnait de meilleurs valeurs (10% nRSME) Pour des pas de 15min à 2h, le RNA reste la	***

	intelligents	plus vers les sources renouvelables d'énergie et particulièrement les sources intermittentes	et intermittents pour le projet TILOS qui consiste en l'installation d'une centrale hybride solaire, éolienne et stockage par batteries sur une petite île de l'archipel du Dodécanèse	de l'île de Dodécanèse.	test) -Prévision avec RNA, SVM et régression linéaire	meilleure méthode Pour 2 à 6h, le modèle PMC est le meilleur. Ainsi, le PMC a été choisi pour la prévision	
GHIOU JYE HUANG and PINGHUAN KUO [6]	Multiple-Input Deep Convolutional Neural Network Model for Short-Term Photovoltaic Power	Du fait de l'intermittence des énergies renouvelables, il devient nécessaire de pouvoir estimer leur production afin de les inclure dans	Utiliser le Deep Learning afin de fournir un modèle de prévision avec une marge d'erreur faible.	Il teste une variante de machine Learning et nous permet ainsi d'avoir une idée du comportement de	La méthodologie est basée sur un modèle de Deep Learning en utilisant Pvnnet.	La méthodologie donne une grande incertitude et de nombreuses irrégularités	*

	Forecasting	le mix énergétique.		la machine dans certaines situations et ainsi bien choisir la méthodologie à adopter pour notre travail.			
Jun Li and Qibo Liu [7]	Short-term Power Forecasting Using SOM-based Regional Modelling Methods	Du fait de l'intermittence des énergies renouvelables, il devient nécessaire de pouvoir estimer leur production afin de les inclure dans le mix énergétique.	Tester une méthode hybride de machine Learning appelé KELM afin d'avoir une marge d'erreur très faible pour une prévision à court terme.	Il teste une variante de machine Learning et nous permet ainsi d'avoir une idée du comportement de la machine pour ladite variante.	Utilisation du clustering of the self organizing map (SOM) pour le traitement des données. -Utilisation du c-mean pour le traitement des données -Utilisation du KELM pour établir la prévision	La comparaison avec les autres méthodes donne une plus grande précision pour celle-ci. En plus, pour différente saisons, la marge d'erreurs varie faiblement.	****

Muhannad Alaraj, Ibrahim Alsaidan, Astitva Kumar, Mohammad Rizwan, Majid Jamil [8]	Energy Production Forecasting From Solar Photovoltaic Plants Based on Meteorological Parameters for Qassim Region, Saudi Arabia	Avec la COVID 19, l'exploitation des énergies fossiles a diminué drastiquement. Toutefois, la demande d'électricité ayant augmentée, il est devenu primordiale d'intégrer une grande part des énergies renouvelables dans le mix énergétique de l'Arabie Saudia.	Trouver un modèle de prévision adapté à la région de Qassim et avec une marge d'erreur faible. Cette étude pourra servir à adapter le modèle à l'échelle national.		-Prévision sur la base du support vector machine (SVM) -Prévision sur la base de la régression linéaire. -Utilisation de l'erreur quadratique afin de déterminer la méthode la plus adéquate		***
Bastien Dussap et Antoine Faul [9]	Projet de Machine Learning and Forecasting	Dans le cadre du projet de machine Learning pour la prévision, il a été	Prévoir la prévision à court terme d'une centrale	Il utilise plusieurs méthodes de prévisions afin	-Prévision sur la base de la régression linéaire	Individuellement, Il revient que le meilleur modèle pour la prévision	****

		question d'étudier la radiation solaire reçue sur un point donné	solaire photovoltaïque sur l'île de Hawaï	de trouver la mieux adapté pour la prévision à court terme pour un emplacement donné	-Prévision sur la base du modèle ARIMA	dans cet article est le modèle ARIMA	
--	--	---	--	--	--	--	--

Annexe II : Explication de la vérification de la stationnarité

Examen visuel

Tracez la série temporelle et évaluer si elle présente des variations de moyenne de variance ou de structure au fil du temps (on observe au niveau de la courbe une décroissance lente et régulière)

Recherchez des tendances des cycles ou des motifs non aléatoires

Test statistique

Utiliser des tests statistiques pour évaluer la présence de stationnarité

Test de Dickey-Fuller augmenté (ADF) est l'un des tests couramment utilisés

L'hypothèse nulle du test est que la série temporelle est non stationnaire

Si la valeur p du test est inférieure à un seuil prédéfini nous rejetons l'hypothèse nulle en concluant que la série est stationnaire

Annexe III : Explication sur la différenciation

La différenciation simple est un concept fondamental en analyse des séries temporelles, surtout pour rendre les séries stationnaires. Elle consiste à soustraire la valeur précédente d'une observation pour obtenir la différence entre elles.

Principe de la différenciation simple :

Pour une série temporelle Y_t , la différenciation à l'instant Y_t' à l'instant t est donnée par :

$$Y_t' = Y_t - Y_{t-1} \quad (12)$$

Développement d'un modèle de prévision à court terme de la puissance produite à la centrale solaire photovoltaïque de Nagréongo (30 MWc) dans le réseau national interconnecté du Burkina Faso.

Annexe IV : Formule calcul de l'AIC

$$AIC + 2k - 2\ln(L) \quad (13)$$

Avec :

K le nombre de paramètres estimés du modèle

L la valeur maximale de la fonction vraisemblance du modèle.

Annexe V : Formule calcul du BIC

$$\text{BIC} = k \ln(n) - 2 \ln(L_{\max}) \quad (14)$$

Avec :

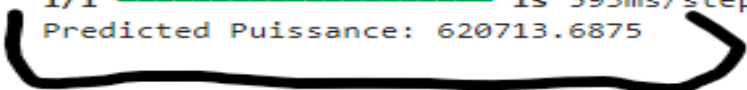
k le nombre de paramètres estimés dans le modèle

n : le nombre total d'observations dans les données

$\ln(L_{\max})$: log-vraisemblance maximisée pour le modèle.

Annexe VI: Exemple de validation du modèle sur des données nouveaux et comparaison avec les données réelles

```
1295/1295 ————— 17s 8ms/step - loss: 0.0173
Epoch 2/14
1295/1295 ————— 11s 8ms/step - loss: 0.0052
Epoch 3/14
1295/1295 ————— 11s 9ms/step - loss: 0.0046
Epoch 4/14
1295/1295 ————— 10s 8ms/step - loss: 0.0050
Epoch 5/14
1295/1295 ————— 12s 9ms/step - loss: 0.0046
Epoch 6/14
1295/1295 ————— 14s 11ms/step - loss: 0.0045
Epoch 7/14
1295/1295 ————— 14s 11ms/step - loss: 0.0044
Epoch 8/14
1295/1295 ————— 13s 10ms/step - loss: 0.0045
Epoch 9/14
1295/1295 ————— 12s 9ms/step - loss: 0.0044
Epoch 10/14
1295/1295 ————— 12s 9ms/step - loss: 0.0043
Epoch 11/14
1295/1295 ————— 12s 9ms/step - loss: 0.0045
Epoch 12/14
1295/1295 ————— 12s 9ms/step - loss: 0.0045
Epoch 13/14
1295/1295 ————— 11s 9ms/step - loss: 0.0047
Epoch 14/14
1295/1295 ————— 12s 9ms/step - loss: 0.0045
1/1 ————— 1s 593ms/step
Predicted Puissance: 620713.6875
```



Développement d'un modèle de prévision à court terme de la puissance produite à la centrale solaire photovoltaïque de Nagréongo (30 MWc) dans le réseau national interconnecté du Burkina Faso.

1/1/24 6:30	14,7795	15,2144	15,2681	18,1	18,1	0,7305	0,8949	1,2908	1,82	2,1	0
1/1/24 6:35	14,8487	15,2721	15,2482	18,08	18,1	2,2804	2,4871	2,8281	1,9	2,18	0
1/1/24 6:40	14,8357	15,2614	15,0954	18	18,06	4,5675	4,8656	5,1844	1,9	2,2	0
1/1/24 6:45	14,7495	15,2216	15,0415	17,9	17,86	10,7945	11,1052	12,3159	1,8	2,16	0
1/1/24 6:50	14,7332	15,1893	15,0634	17,78	17,72	21,525	21,5644	23,4946	1,78	2,06	68000
1/1/24 6:55	14,7566	15,3117	15,1645	17,66	17,56	31,3056	32,4237	35,5319	1,6	1,9	392667
1/1/24 7:00	14,9101	15,5135	15,3792	17,56	17,5	44,2812	45,2323	51,2726	1,44	1,64	602000
1/1/24 7:05	15,2795	15,8544	15,7341	17,58	17,5	58,8425	59,9662	69,1436	1,36	1,42	899667
1/1/24 7:10	15,8013	16,3764	16,1843	17,76	17,58	75,4823	76,5396	89,6786	1,42	1,44	1236670
1/1/24 7:15	16,4999	17,154	16,7706	17,98	17,8	92,1339	94,9943	110,4348	1,5	1,52	1621670
1/1/24 7:20	17,2506	17,9649	17,4384	18,26	18,06	109,1442	111,4214	129,8002	1,42	1,4	2070330
1/1/24 7:25	18,0733	18,9166	18,1162	18,6	18,42	127,3566	130,8902	151,6388	1,3	1,22	2506330
1/1/24 7:30	19,0583	19,951	18,9755	19	18,86	147,7334	150,5386	173,6856	1,44	1,28	3002000
1/1/24 7:35	20,1044	21,0511	19,7887	19,42	19,28	164,2386	169,2998	192,8958	1,6	1,6	3496000