



## Full Length Research Paper

# PREDICTION DE LA PRODUCTION D'UNE CENTRALE SOLAIRE EN UTILISANT L'INTELLIGENCE ARTIFICIELLE : CAS DE LA CENTRALE PHOTOVOLTAÏQUE DE ZAGTOULI AU BURKINA-FASO

Toussaint Tilado GUINGANE<sup>1,2\*</sup>, Mouhamadou Falilou NDIAYE<sup>3</sup>, Sosthène TASSEMBEDO<sup>1</sup>, Éric KORSAGA<sup>1</sup>, Dominique BONKOUNGOU<sup>1,2</sup>, Zacharie KOALAGA<sup>1</sup>

<sup>1</sup>Laboratoire des Matériaux et Environnement (LA.M.E.), Unité de Formation et de Recherche en Sciences Exactes et Appliquée (UFR/SEA), Université Pr KI-ZERBO, Ouagadougou, Burkina Faso

<sup>2</sup>Laboratoire de Sciences et Technologies (LaST), Unité de Formation et de Recherche en Sciences et Techniques (UFR/ST), Université Thomas SANKARA, Ouagadougou, Burkina Faso

<sup>3</sup>Centre International de Formation et de Recherche en Energie Solaire (C.I.F.R.E.S), ESP-UCAD: BP 5085, Dakar-Fann, Sénégal

Received April 2024 – Accepted June 2024



\*Corresponding author. [tilado88@yahoo.fr](mailto:tilado88@yahoo.fr)

Author(s) agree that this article remain permanently open access under the terms of the Creative Commons Attribution License 4.0 International License.

## Résumé :

Le Burkina Faso, comme de nombreux pays, s'engage résolument dans la transition vers des sources d'énergie renouvelable pour répondre à ses besoins croissants en électricité tout en réduisant son empreinte carbone. Au cœur de cette démarche, la centrale photovoltaïque de Zagtouli représente un jalon majeur dans la production d'énergie solaire dans la région. Cependant, la variabilité intrinsèque de l'énergie solaire pose des défis opérationnels, nécessitant des solutions innovantes pour maximiser son potentiel.

L'objectif de ce travail est de développer un système de prédiction de la production d'énergie solaire de la centrale de Zagtouli en utilisant l'Intelligence Artificielle (IA). Cette approche novatrice vise à améliorer la planification opérationnelle, à optimiser l'efficacité énergétique et à réduire les coûts associés à l'exploitation de la centrale. En intégrant des techniques avancées d'IA, nous cherchons à anticiper les variations de la production d'énergie solaire, offrant ainsi une meilleure gestion des ressources et une contribution accrue à la stabilité du réseau électrique.

**Mots clés :** énergie ; solaire, météorologiques, apprentissage, intelligences artificielles, prédiction.

## Cite this article:

Toussaint Tilado GUINGANE, Mouhamadou Falilou NDIAYE, Sosthène TASSEMBEDO, Éric KORSAGA, Dominique BONKOUNGOU, Zacharie KOALAGA (2024). Prédiction de la production d'une centrale solaire en utilisant l'intelligence artificielle : cas de la centrale photovoltaïque de zagtouli au BURKINA-FASO. Revue RAMReS – Sci. Appl. & de l'Ing., Vol. 6(1), pp. 37-42. ISSN 2630-1164.

## 1. Introduction

La prédiction de la production d'énergie solaire revêt une importance capitale pour garantir une exploitation optimale de la centrale. En anticipant les variations de production, nous visons à assurer une fourniture stable d'électricité, minimisant ainsi les interruptions potentielles et contribuant à la fiabilité du réseau.

L'Intelligence Artificielle offre des outils puissants pour analyser des données complexes et non linéaires, caractéristiques inhérentes à la production d'énergie

solaire. Son utilisation permet d'exploiter pleinement les informations recueillies par les capteurs de la centrale, ouvrant la voie à une prédiction plus précise et à une gestion proactive.

Achevée en 2017, la centrale solaire de Zagtouli [1] est constituée de 129600 modules (en silicium polycristallins) orientés vers le sud sous un angle de 15 degrés et s'étend sur une superficie de 60 ha.

La centrale photovoltaïque est organisée en 32 sous champs. Chaque sous champ est raccordé à un

onduleur ayant les mêmes caractéristiques et de puissance 1,1 mégawatt chacun et regroupés dans un centre photovoltaïque intégré en abrégé CPI. Un CPI est un bâti électrique en acier divisé en trois compartiments.

Aujourd'hui, la centrale à une puissance effective de 33,7 MWC soit une production d'environ 55 GWh/an représentant 4% de la consommation annuelle burkinabè d'électricité. Elle est l'équivalent de la consommation de 660.000 personnes et son coût économique de production est estimé entre 30 et 40 FCFA/kWh. Un coût nettement inférieur au coût moyen de production de la SONABEL, qui s'établissait à 133CFA par kWh en 2016.

Inaugurée en 2017, la centrale a depuis joué un rôle central dans la fourniture d'électricité propre et renouvelable à la région. Cependant, son rendement pourrait être optimisé davantage par l'implémentation d'une solution de prédiction avancée.

Les conditions climatiques spécifiques de la région, les variations saisonnières et les fluctuations de la luminosité solaire posent des défis opérationnels significatifs. L'IA offre une réponse adaptative à ces défis, permettant une gestion plus fine de la production d'énergie solaire.

Le cœur de ce travail réside dans le développement d'un modèle de prédiction précis, exploitant les données en temps réel pour anticiper la production d'énergie solaire avec une haute précision.

En parallèle, nous cherchons à optimiser l'efficacité énergétique de la centrale en ajustant les opérations en fonction des prévisions, minimisant ainsi les pertes potentielles.

Cette initiative représente une avancée significative dans le domaine de l'énergie solaire au Burkina Faso, démontrant le potentiel de l'IA dans la gestion intelligente des installations photovoltaïques. Dans les sections suivantes, nous explorerons plus en détail les méthodes envisagées, l'état actuel des recherches dans le domaine, et les implications pratiques de notre approche.

## 2. Collecte de données

Les données disponibles sont sous le format Excel (.xlsx) et hébergés sur le cloud (google drive). Nous disposons de 128 sources de données (32 fichiers Excel par an sur 4 ans de 2018 à 2022).

Les 32 fichiers correspondent aux données relevées sur les 32 onduleurs.

La taille moyenne des fichiers est de 64,48 M.

Volume totale : 6 190,30 M soit autour de 6,19 G.

Plus particulièrement pour l'onduleur 8 on a une taille totale de 190 Mo. Toutes nos données sont dans des fichiers Excel et enregistré avec un pas de temps de 1 mn.

Le tableau 1 donne une désignation des features et des variables cible

**Tableau 1: Désignation des features et des variables**

	Variables	Description	Exemple
Features	Date	Comporte la date et l'heure d'enregistrement des données	01/01/2018 00:00:00.00
	Tempmodule	Température à la surface des modules et exprimée en degrés Celsius (°C) (Valeurs enregistrées par les capteurs)	18.825118
	angleinclinaison	Angle d'inclinaison des modules qui correspond en générale à la latitude du lieu (dans notre cas la latitude de notre site est autour de 12,35°) (Valeurs enregistrées par les capteurs)	14.193333
	Ensoleillement	Quantité et durée de la lumière solaire reçue sur nos modules exprimés en Watheure par mètre carre (Wh/m <sup>2</sup> ). (Valeurs enregistrées par les capteurs)	0.000000
	Direction_vent	Direction du vent (Valeurs enregistrées par les capteurs)	49.500126
	vitesse_vent	Vitesse du vent en mètre par seconde (m/s) (Valeurs enregistrées par les capteurs)	0.932827
	temp_amb	Température du milieu ambiant exprimée en degrés Celsius (°C) (Valeurs enregistrées par les capteurs)	21.424707
	Humidité	Teneur en vapeur d'eau exprimée en gramme par mètre carré (g/m <sup>2</sup> ) (Valeurs enregistrées par les capteurs)	19.80000
Variables cibles	Intensité	Intensité du courant exprimé en Ampère (A) (Valeurs enregistrées par les capteurs)	0.0
	tension	Tension aux bornes de chaque entrée d'onduleur exprimée en Volt(V)	32.5

Nous utiliserons les moyennes horaires des différentes variables de l'étude, cette approche offre plusieurs avantages :

- Réduction du bruit : Les données de capteurs peuvent être sujettes à des variations fréquentes ou à des fluctuations brusques. En calculant les moyennes horaires, on réduit le bruit et on obtient une vue plus lisse du comportement des variables, ce qui peut faciliter l'identification de tendances et de motifs.
- Visualisation simplifiée : Les données par minute peuvent souvent être difficiles à interpréter en raison de variations rapides. Les moyennes horaires peuvent simplifier la visualisation en réduisant le nombre de points de données, ce qui rend plus facile l'analyse et la comparaison.
- Corrélation entre variables : En calculant les moyennes horaires, on peut explorer la corrélation entre différentes variables sur une base horaire. Cela peut aider à comprendre comment ces variables interagissent au fil du temps.
- Facilitation de la modélisation : Lors de la création de modèles prédictifs ou de simulations, l'utilisation de moyennes horaires peut simplifier

- les entrées et améliorer la stabilité du modèle en réduisant l'impact des variations brusques.
- Comparaisons temporelles plus simples : Si vous avez des données sur une période prolongée, les moyennes horaires peuvent rendre plus facile la comparaison de différentes heures de la journée, jours de la semaine, ou saisons.
- Analyse de la performance de la centrale solaire : En moyennant les données sur des périodes spécifiques, comme les heures de pointe ou les heures creuses, on peut obtenir des informations précieuses sur la performance de la centrale solaire à différentes périodes.
- Réduction de la dimensionnalité : Dans le cas où vous avez plusieurs capteurs mesurant des variables similaires, les moyennes horaires peuvent être utilisées pour agréger les données et réduire la dimensionnalité de votre ensemble de données.
- Optimisation des opérations : Les moyennes horaires peuvent également être utiles dans l'optimisation des opérations, en particulier pour les systèmes d'énergie renouvelable où la production peut varier au fil du temps.

Après une analyse préliminaire nous avons constaté une incohérence qui s'expliquent principalement par des dysfonctionnements des capteurs. La typologie est donnée dans le tableau 2.

**Tableau 2: Typologie données incohérentes**

Type	Description
1	Les valeurs de toutes les variables sont égales à (0.0) simultanément Une ou plusieurs cellules contiguës affichent la valeur (0.0)
2	Les valeurs de toutes les variables sont manquantes simultanément Une ou plusieurs cellules contiguës ont des valeurs manquantes
3	La valeur de la tension et/ou de l'ensoleillement et/ou de l'intensité sont différentes de (0.0) entre 18 heures et 05 heures (Les valeurs de ces trois variables dépendent fortement des cycles de levée et de coucher du soleil)
4	La valeur de la tension et/ou de l'ensoleillement et/ou de l'intensité sont égales à (0.0) de 06 heures à 18 heures
5	Duplication de lignes d'information
6	Les variables ont des valeurs manquantes

Afin de nettoyer les données aberrantes identifiées dans notre jeu de données, nous avons mis en place un processus de nettoyage en quatre étapes, en développant des fonctions dédiées. Ce processus a été établi en fonction des différents types d'anomalies rencontrées lors de l'analyse préliminaire des données[2-5].

- ✓ La fonction `remplacer_doublon(df, p, q)`  
Cette fonction permet de remplacer les valeurs dupliquées dans un DataFrame par des valeurs NaN (valeurs manquantes).  
Les étapes sont :
  - Copier le DataFrame d'origine pour ne pas modifier l'original
  - Sélectionnez la plage de lignes entre p et q
  - Sélectionner toutes les colonnes sauf 'Date'
  - Remplacer les valeurs sélectionnées par NaN
  - Retourner le DataFrame modifié

Cette fonction permet de détecter et remplacer efficacement des plages de lignes en double (à l'exception de la colonne 'Date') dans un DataFrame, en conservant l'original initialement.

Les valeurs NaN permettent ensuite de repérer visuellement ou de traiter les données manquantes résultant du remplacement des doublons.

- ✓ La fonction `remplacer_par_nan(df)`

Cette fonction permet de remplacer efficacement toutes les valeurs nulles (0.0) dans un DataFrame par des valeurs NaN (valeurs manquantes).

Les étapes clés sont :

- Créer un masque de sélection des valeurs à remplacer
- Indexer le DataFrame avec ce masque
- Effectuer la substitution par `np.nan` sur les éléments sélectionnés
- Retourner le DataFrame modifié

Conformément à la spécificité de notre jeu de données, nous ne devons avoir la valeur 0.0 que pour les variables « intensite », « tension », « Ensoleillement » entre 18 heures et 06 heures, toute autre valeur 0.0 peut être considérée comme une valeur incohérente. Ainsi :

- Les valeurs NaN entre 18 heures et 06 heures pour les variables « intensite », « tension » et « Ensoleillement » seront remplacées par 0.0 par application de la fonction `traiter_int_tens_ens_tmod(df)`
- Les autres valeurs NaN seront imputées par la fonction `traiter_int_tens_ens_tmod(df)` ou la fonction `imputation_globale(df)`
- ✓ La fonction `traiter_int_tens_ens_tmod(df)`

Cette fonction permet de traiter et transformer les colonnes "intensite", "tension", "Ensoleillement" et "tempmodule" d'un DataFrame df contenant des données temporelles :

- Conversion de la colonne "Date" en date time
- Mise à 0 des valeurs la nuit (18h à 5h) pour "intensité", "tension" et "Ensoleillement"
- Séparation des données de jour (6h à 18h) dans `df_jour`
- Imputation des valeurs manquantes le jour pour "intensité", "tension" et "Ensoleillement" avec la méthode `SimpleFill`
- Remplacement des valeurs imputées dans le DataFrame d'origine `df`
- Imputation de la moyenne pour les valeurs manquantes de "tempmodule" sur les données de jour
- Retour du DataFrame `df` transformé et imputé.
- ✓ La fonction `imputation_globale(df)[6-7]`

Voici un résumé de ce que fait cette fonction :

- La fonction `imputation_globale` prend en entrée un DataFrame `df`
- Elle sélectionne uniquement les variables numériques dans `df_num`

- Elle utilise la méthode d'imputation KNN (K plus proches voisins) de la librairie fancyimpute pour créer un imputeur (imputer)
- Elle applique cet imputeur sur toutes les données numériques df\_num
- Elle convertit le résultat en DataFrame imputed\_df
- Elle réindexe les colonnes numériques imputées dans le DataFrame d'origine df
- Elle retourne le DataFrame df avec les valeurs manquantes imputées pour toutes les variables numériques.

- La méthode « head() » permet d'afficher les premières lignes d'un DataFrame.
- La méthode « info() » fournit un résumé concis du DataFrame, incluant le nombre d'entrées, le type de données de chaque colonne, le nombre de valeurs non nulles et l'utilisation de la mémoire.
- La méthode « describe() » est utilisée pour générer des statistiques descriptives résumant la tendance centrale, la dispersion et la forme de la distribution d'un ensemble de données, à l'exclusion des valeurs NaN. Par défaut, elle ne fonctionne que sur les colonnes numériques.

### 3. Évaluations du nettoyage par calcul de corrélation

Pour évaluer l'impact du nettoyage sur les relations entre les variables, nous avons calculé les matrices de corrélation avant et après nettoyage représenté respectivement sur la figure 1 et 2. La comparaison de ces matrices de corrélation nous permet de vérifier que le nettoyage n'a pas substantiellement modifié les coefficients de corrélation entre les différentes variables du jeu de données.

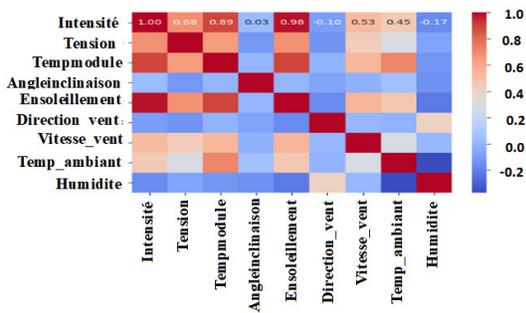


Figure 1: Corrélations avant le nettoyage

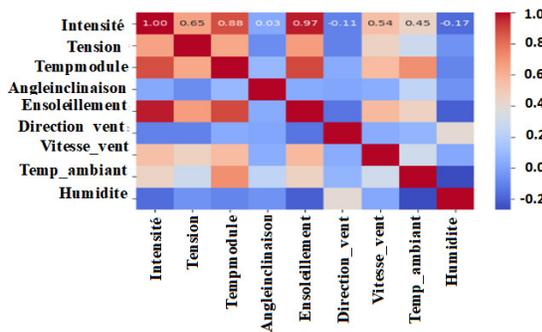


Figure 2: Corrélations après le nettoyage

Au terme du processus de nettoyage détaillé précédemment, nous disposons du jeu de données final nommé *dataset* qui sera utilisé pour l'entraînement du modèle. Nous avons utilisé les méthodes « head() », « info() », « describe() » pour présenter les caractéristiques.

### 4. Modelisation

Dans le cadre de ce travail, nous cherchons à prédire la puissance de production de la centrale PV en fonction de variables météorologiques [8-12]. Il s'agit donc d'un problème de régression. Nous allons tester et comparer différents algorithmes de régression comme la régression linéaire, les forêts aléatoires, le Réseau de Neurones Récurrent (RNN) et les SVM afin de déterminer le modèle optimal pour nos données.

Le dataset nettoyé comporte 28470 lignes. Pour l'entraînement de nos modèles nous utiliserons 80 % de notre dataset et le reste sera réservé pour le test [12], [13-17].

### 5. Résultats

Le tableau 3 donne une comparaison des différents modèles.

Tableau 3 : Tableau comparatif des méthodes

Méthodes	R <sup>2</sup>	MSE
<b>Régression Linéaire avec Utilisation de Corrélation :</b>	0.83	12243683482.4900
<b>Random Forest :</b>	0.94	0.0027
<b>K Plus Proches Voisins (KNN) :</b>	0.92	5423005520.2700
<b>Réseau de Neurones Récurrent (RNN) :</b>	0.89	7892349291.5200

#### ✚ Régression Linéaire :

- Performances moyennes avec un R2 de 0.83, mais le MSE est élevé, indiquant des erreurs significatives.

- ✚ **Random Forest :**
  - Performances élevées avec un faible MSE (0.0027) et un R2 élevé (0.9474). Bonne capacité à capturer des relations complexes.
- ✚ **KNN :**
  - Précision élevée (R2 : 0.9254) mais MSE élevé, indiquant une certaine sensibilité aux valeurs aberrantes ou bruit.
- ✚ **RNN :**
  - MSE élevé et R2 relativement bas par rapport aux autres méthodes, peut nécessiter un ajustement ou une optimisation des paramètres.

De la comparaison de toutes ces méthodes Random Forest a montré des performances globales solides dans ce cas. La figure 3 donne une représentation des données réelles et prédites. Cette figure vient encore confirmer de la robustesse du modèle.

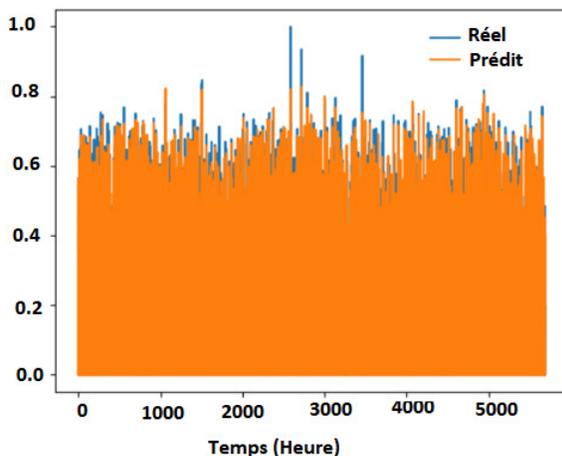


Figure 3 : Données Réelles et prédites avec Random forest

## 6. Conclusion

Dans le cadre de ce travail, notre objectif principal était de développer un système de prédiction de la production d'énergie solaire pour la centrale de Zagtoui en utilisant des techniques d'Intelligence Artificielle (IA). Notre démarche a commencé par un nettoyage approfondi du dataset, visant à préparer les données pour l'entraînement des modèles.

Nous avons exploré différentes méthodes de régression pour modéliser la relation entre les variables indépendantes, telles que la température, l'inclinaison du module solaire, l'ensoleillement, et la variable dépendante, la puissance solaire. La Régression Linéaire, avec l'utilisation de la corrélation, a donné un coefficient de détermination (R2) de 0.83 et une Erreur Quadratique Moyenne (MSE) de 12243683482.4900. Ensuite, nous avons appliqué la méthode Random Forest, qui a émergé comme le modèle le plus

performant avec un MSE remarquablement bas de 0.0027 et un R2 de 0.94. Cette méthode s'est avérée particulièrement efficace pour prédire la production d'énergie solaire.

Le modèle des K Plus Proches Voisins (KNN) a également été testé, montrant une précision appréciable avec un MSE de 5423005520.2700 et un R2 de 0.92. Cependant, des ajustements potentiels peuvent être nécessaires pour optimiser davantage les performances. Le Recurrent Neural Network (RNN) a été utilisé pour modéliser des dépendances temporelles dans les données, obtenant un MSE de 7892349291.5200 et un R2 de 0.8925. Bien que performant, des explorations supplémentaires de l'architecture ou des paramètres du modèle pourraient être bénéfiques.

## REFERENCES

- [1] Centrale solaire de Zagtoui : <https://www.alliance-sahel.org/actualites/centrale-solaire-de-zagtoui/>
- [2] Python : <https://www.python.org/>,
- [3] Pandas : <https://pandas.pydata.org/>,
- [4] Numpy : <https://numpy.org/>,
- [5] Matplotlib : <https://matplotlib.org/>,
- [6] Fancyimpute : <https://pypi.org/project/fancyimpute/>
- [7] SKLearn : <https://scikit-learn.org/stable/>
- [8] A.Jäger-Waldau and H.Ossenbrink "Progress of electricity from biomass, wind and photovoltaics in the European Union" Renewable and Sustainable Energy Reviews, Elsevier, vol. 8(2), pages 237-182, April 2004. [2]
- [9] S.Masri and D.Eltigani "Challenges of integrating renewable energy sources to smart grids: A review", Renewable and Sustainable Energy Reviews, Elsevier, Volume 52, Pages 770-780, December 2015.
- [10] H.Zhu, W.Lian, L.Lu, S.Dai and Y.Hu "An Improved Forecasting Method for Photovoltaic Power Based on Adaptive BP Neural Network with a Scrolling Time Window", Energies, p.10-2342, 2017.
- [11] E. ZAZOUA, Modèles De Prédiction À Apprentissage Automatique Pour Les Réseaux De Capteurs Sans Fil À Énergie Renouvelable, Maîtrise en génie électrique, Université du QUÉBEC à MONTRÉAL, Avril 2022
- [12] Glorot Xavier., Apprentissage des réseaux de neurones profonds et applications en traitement automatique de la langue naturelle, Thèse présentée à la Faculté des arts et des sciences de l'Université de Montréal en vue de l'obtention du grade de Philosophiæ Doctor en informatique ; Février 2014.
- [13] K. Almi'ani, A. Viglas and L. Libman, . Energy-efficient data gathering with tour length-constrained mobile elements in wireless sensor networks, Proc. - Conf. Local Comput. Networks LCN, pp. 582-589, 2010.
- [14] M. Al-Omary, K.Hassini, A. Fakhfakh and O. Kanoun . Prediction of Energy in Solar Powered Wireless Sensors Using Artificial Neural Network. 16th International Multi-Conference on Systems, Signals & Devices (SSD), Istanbul, Turkey, S,2019.
- [15] Aram, L. Mesin and E. Pasero. Improving lifetime in wireless sensor networks using neural data prediction. World Symposium on Computer Applications & Research (WSCAR), Sousse, Tunisia, 2014.
- [16] A. Cammarano, C. Petrioli, and D. Spenza. (2012, octobre). Pro-Energy: A novel energy prediction model for solar and wind energy harvesting wireless sensor networks. in Proceedings of the 9th IEEE International Conference on Mobile Ad-Hoc and Sensor Systems (MASS '12), pp. 75-83, USA, Octobre 2012.

[17] A. Cammarano, C. Petrioli, and D. Spenza. . Poster abstract: Improving energy predictions in EH-WSNs with pro-energy-VLT. in Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems (SenSys '13), Novembre 2013.