

RÉPUBLIQUE ALGÉRIENNE DÉMOCRATIQUE ET POPULAIRE  
MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE  
SCIENTIFIQUE  
UNIVERSITÉ ABDERRAHMANE MIRA DE BÉJAÏA



FACULTÉ DES SCIENCES EXACTES  
DÉPARTEMENT D'INFORMATIQUE

## MEMOIRE

EN VUE DE L'OBTENTION DU DIPLÔME DE  
MASTER

**Domaine :** MATHÉMATIQUES ET INFORMATIQUE      **Filière :** INFORMATIQUE

**Spécialité :** INTELLIGENCE ARTIFICIELLE

### Thème

APPRENTISSAGE MACHINE POUR  
LA PRÉDICTION D'ÉNERGIE  
SOLAIRE À BÉJAÏA

*Présenté par :*

M. MOHAND SAID CHEURFA      M. KHALED CHABANE

*Soutenu le 02/07/2023 devant le jury composé de :*

<i>Président</i>	M. Zoubeyr FARAH	M.C.A	U.A/MIRA DE BÉJAÏA
<i>Encadrante</i>	Mme Samiha AIT TALEB	M.C.B	ESTIN DE AMIZOUR
<i>Co-Encadrant</i>	M. Abderrezak SEBAA	M.C.A	ESTIN DE AMIZOUR
<i>Examineur</i>	M. Fatah BOUCHEBAH	M.C.B	U.A/MIRA DE BÉJAÏA

**Promotion 2022 – 2023**

# Remerciements

En tout premier lieu, nous remercions Dieu le tout-puissant pour nous avoir accordé la santé et la force nécessaires pour surmonter toutes les difficultés et mener à bien ce travail.

Nous souhaitons également adresser nos chaleureux remerciements à notre encadrante, **Madame AIT TALEB Samiha**, ainsi qu'à notre co-encadrant, **Monsieur SEBAA Abderrazak**. Leur encadrement, leurs orientations, leur soutien et leurs précieux conseils ont grandement contribué à la réalisation de cette recherche.

Que les membres de jury trouvent, ici, l'expression de mes sincères remerciements pour l'honneur qu'ils me font en prenant le temps de lire et d'évaluer ce travail.

Je souhaite aussi remercier l'équipe pédagogique et administrative de l'université pour leurs efforts dans le but de nous offrir une excellente formation.

Enfin, nous souhaitons exprimer notre reconnaissance à tous ceux qui, de près ou de loin, ont apporté leur soutien moral et leur encouragement tout au long de ce parcours.

# Résumé

Ce mémoire de recherche se concentre sur la prédiction de l'énergie solaire en utilisant diverses méthodes basé sur le machine learning. Il explore les travaux antérieurs et identifie les lacunes et les opportunités dans le domaine de la prédiction d'énergie solaire. La méthodologie détaille les techniques de collecte de données, les instruments de recherche et les procédures d'analyse, tout en abordant les considérations éthiques et les limitations de l'étude. Les résultats obtenus à partir des données de la NASA pour la ville de Bejaia, en Algérie, sont présentés en détail. Les différentes méthodes de prédiction d'énergie solaire sont décrites, ainsi que les outils et les données spécifiques utilisés. Les résultats de chaque méthode sont comparés en termes de précision et de performance. La discussion interprète les résultats à la lumière des objectifs de recherche et examine les implications théoriques et pratiques. Elle met en évidence les avantages et les limites de chaque méthode de prédiction et propose des recommandations pour de futures recherches.

---

**Mots clés :** Prédiction, Énergie solaire, Photovoltaïque, Apprentissage automatique, Méthodes statistiques.

---

# Abstract

This research paper focuses on the prediction of solar energy using various machine learning-based methods . It explores previous works and identifies gaps and opportunities in the field of solar energy prediction. The methodology details data collection techniques, research instruments, and analysis procedures, while addressing ethical considerations and study limitations. The results obtained from NASA data for the city of Bejaia, Algeria, are presented in detail. The different methods of solar energy prediction are described, along with the specific tools and data used. The results of each method are compared in terms of accuracy and performance. The discussion interprets the results in light of the research objectives and examines theoretical and practical implications. It highlights the advantages and limitations of each prediction method and provides recommendations for future research.

---

**Mots clés :** Prediction, Solar energy, Photovoltaics, Machine learning, Statistical methods.

---

# Table des matières

Remerciements	I
Résumé	II
Abstract	III
Introduction générale	1
<b>1 Généralités sur l'énergie solaire.</b>	<b>3</b>
1.1 Introduction . . . . .	3
1.2 Énergie Solaire . . . . .	3
1.2.1 L'énergie solaire photovoltaïque . . . . .	4
1.2.2 L'énergie solaire thermique . . . . .	4
1.2.3 L'énergie solaire thermodynamique . . . . .	4
1.3 Rayonnement solaire . . . . .	5
1.3.1 Composante déterministe et stochastique du rayonnement so- laire . . . . .	6
1.3.2 L'exploitation de rayonnement solaire . . . . .	7
1.3.2.1 La conversion photovoltaïque . . . . .	7
1.3.2.2 Les systèmes photovoltaïques (PV) . . . . .	8
1.3.2.3 Les capteurs solaires thermiques . . . . .	10
1.4 Les données . . . . .	10
1.5 L'analyse prédictive . . . . .	10
1.5.1 Définitions . . . . .	11
1.5.2 Les variables utilisés pour la prédiction . . . . .	11
1.5.2.1 Variables endogènes . . . . .	11
1.5.2.2 Les variables exogènes . . . . .	12
1.6 Conclusion . . . . .	13

<b>2</b>	<b>Prédiction d'énergie solaire</b>	<b>14</b>
2.1	Introduction . . . . .	14
2.2	Série temporelle . . . . .	15
2.2.1	Décomposition d'une série temporelle . . . . .	16
2.3	Les modèles et techniques de prédiction . . . . .	17
2.3.1	Modèle naïf . . . . .	19
2.3.1.1	Persistance . . . . .	19
2.3.1.2	La moyenne . . . . .	19
2.3.2	Modèle physique . . . . .	19
2.3.3	Techniques Statistiques . . . . .	20
2.3.3.1	Les chaînes de Markov . . . . .	20
2.3.3.2	Inférences Bayésiennes . . . . .	21
2.3.3.3	Les modèles autorégressifs . . . . .	21
2.3.3.4	Lissage exponentiel . . . . .	23
2.3.3.5	La régression . . . . .	23
2.3.3.6	Machines à vecteurs de support . . . . .	24
2.3.3.7	K- plus proches voisins (k-NN) . . . . .	24
2.3.3.8	Les arbres de décision . . . . .	25
2.3.3.9	Les forêts aléatoires . . . . .	25
2.3.3.10	Extra-Arbres . . . . .	26
2.3.3.11	Réseau de neurones artificiels . . . . .	26
2.3.4	Modèles hybrides . . . . .	41
2.4	Principaux facteurs influençant les prévisions d'énergie solaire . . . . .	42
2.4.1	L'horizon de prédiction . . . . .	42
2.4.1.1	Prédictions à court terme . . . . .	42
2.4.1.2	Prédiction à moyen terme . . . . .	43
2.4.1.3	Prédiction à long terme . . . . .	43
2.4.2	Les données d'entrées . . . . .	43
2.4.3	Les techniques de prévision . . . . .	45
2.5	Conclusion . . . . .	46
<b>3</b>	<b>Etat de l'art</b>	<b>47</b>
3.1	Introduction . . . . .	47
3.2	Travaux connexes . . . . .	47

3.3	Analyse et comparaison . . . . .	58
3.4	Conclusion . . . . .	59
<b>4</b>	<b>Approches proposée</b>	<b>61</b>
4.1	Introduction . . . . .	61
4.2	Environnement de développement . . . . .	61
4.2.1	Langage de programmation . . . . .	62
4.2.1.1	Python . . . . .	62
4.2.2	Bibliothèques de Python . . . . .	63
4.2.2.1	Numpy . . . . .	63
4.2.2.2	Pandas . . . . .	63
4.2.3	TensorFlow . . . . .	64
4.2.3.1	Matplotlib . . . . .	64
4.2.3.2	Sklearn . . . . .	65
4.2.3.3	Keras . . . . .	65
4.3	Méthodologie de travail . . . . .	66
4.3.1	Collecte des données . . . . .	66
4.3.2	Prétraitement des données . . . . .	68
4.3.2.1	Exploration des données et identification de relation	68
4.3.2.2	Normalisation . . . . .	72
4.3.3	Sélection de modèle . . . . .	73
4.3.4	Entraînement et test et validation des données . . . . .	74
4.3.4.1	La phase d'apprentissage . . . . .	74
4.3.4.2	La phase de test (Évaluation) . . . . .	75
4.4	Réalisation . . . . .	76
4.4.1	Le cas journalier . . . . .	76
4.4.1.1	Les méthodes de régression . . . . .	76
4.4.1.2	LSTM long short-term memory . . . . .	76
4.4.1.3	Stacking . . . . .	78
4.4.1.4	SARIMA . . . . .	80
4.4.2	Le cas horaire . . . . .	81
4.4.2.1	Les méthodes de régression . . . . .	81
4.4.2.2	LSTM long short-term memory . . . . .	81
4.5	Résultats et discussion . . . . .	82

4.6 Conclusion . . . . .	87
<b>Conclusion générale &amp; Perspectives</b>	<b>88</b>

# Table des figures

1.1	Interaction entre le rayonnement solaire et l'atmosphère terrestre [1].	6
1.2	Impact de la couverture nuageuse sur la série temporelle de rayonnement global [1] . . . . .	7
1.3	principe de l'effet photoélectrique [1]. . . . .	8
1.4	Principe du raccordement d'un module l'énergie PV au réseau électrique [1]. . . . .	9
2.1	Classification des techniques de prévision. . . . .	18
2.2	Représentation schématique d'un neurone biologique [2]. . . . .	29
2.3	Représentation mathématique d'un neurone artificiel [3]. . . . .	31
2.4	(a) Diagramme schématique d'une structure RNA composée d'une couche d'entrée, une couche cachée et une couche de sortie. (b) modèle mathématique du neurone artificiel [4]. . . . .	35
2.5	Exemple d'un perceptron à deux couches (quatre paramètres d'entrée et trois paramètres de sortie) [3]. . . . .	36
2.6	Réseau de neurones perceptrons multicouche [5]. . . . .	37
2.7	Réseau multicouche à connexions locales [3]. . . . .	38
2.8	L'architecture d'un réseau de neurones de type RBF [3]. . . . .	39
2.9	L'architecture d'un réseau de neurones à corrélation en cascade [3]. . . . .	40
2.10	L'architecture de Réseau neuronal récurrent(RNN) [6]. . . . .	41
2.11	Irradiation solaire et schéma de production d'énergie photovoltaïque pour un jour donné [7]. . . . .	44
2.12	Température atmosphérique et schéma de production d'énergie photovoltaïque pour un jour donné [7]. . . . .	44
2.13	Corrélation entre la température atmosphérique et la puissance photovoltaïque [7]. . . . .	45

2.14	Les courbes par unité de la puissance de sortie PV et de la vitesse du vent [5]. . . . .	45
3.1	Nombre d'articles publiés par an dans le domaine de la prévision des ressources solaires et de la production d'énergie solaire, retenus par SCOPUS [8]. . . . .	48
4.1	Processus de notre approche . . . . .	67
4.2	Moyenne mensuelle par mois de l'irradiation (wh/M2/H) mesurée a bejaia sur la période 2010-2021 . . . . .	69
4.3	Moyenne mensuelle par heure de l'irradiation (wh/M2/H) mesurée a bejaia sur la période 2010-2021 . . . . .	69
4.4	Moyenne mensuelle de l'irradiation (wh/M2/H) mesurée a bejaia sur la période 2010-2021 . . . . .	69
4.5	Moyenne mensuelle de temperature c mesurée a bejaia sur la période 2010-2021 par mois . . . . .	70
4.6	Moyenne mensuelle de temperature c mesurée a bejaia sur la période 2010-2021 . . . . .	70
4.7	Moyenne mensuelle par heure d'humidité mesurée a bejaia sur la période 2010-2021 . . . . .	70
4.8	Moyenne mensuelle par mois de vitesse de vent mesurée a bejaia sur la période 2010-2021 . . . . .	71
4.9	Matrice de corrélation . . . . .	72
4.10	Comparaison entre les modèles de régression cas journalier. . . . .	77
4.11	Les paramètres de modèle LSTMmonovaleur cas journalier. . . . .	77
4.12	Les paramètres de modèle LSTMmultivaleur cas journalier. . . . .	78
4.13	évolution du modèle LSTMmonovaleur cas journalier. . . . .	79
4.14	évolution du modèle LSTMmultivaleur cas journalier. . . . .	79
4.15	Stacking modèle [9]. . . . .	80
4.16	Meilleur SARIMA modèle. . . . .	81
4.17	Comparaison entre les modèles de régression cas horaire. . . . .	82
4.18	Les paramètres de modèle LSTMmonovaleur cas horaire . . . . .	83
4.19	Les paramètres de modèle LSTMmultivaleur cas horaire. . . . .	84
4.20	évolution du modèle LSTMmonovaleur cas horaire. . . . .	85
4.21	évolution du modèle LSTMmultivaleur cas horaire. . . . .	86

# Liste des tableaux

2.1	Fonctions de transfert $f(x)$ . . . . .	34
3.1	Résumé de quelques travaux récents sur la prévision d'énergie solaire. . . . .	59
4.1	évaluation de toutes les méthodologies de prédiction étudiées par jour j+1. . . . .	83
4.2	évaluation de toutes les méthodologies de prédiction étudiées par heure h+1. . . . .	84

# Liste des sigles et acronymes

<b>PV</b>	<i>PhotoVoltaïques</i>
<b>API</b>	<i>Application Programming Interface</i>
<b>MW</b>	<i>MégaWatt</i>
<b>UV</b>	<i>UltraViolet</i>
<b>V</b>	<i>Volt</i>
<b>Wc</b>	<i>Watt-crête</i>
<b>DC</b>	<i>Direct Current</i>
<b>AC</b>	<i>Alternate Current</i>
<b>IA</b>	<i>Intelligence Artificielle</i>
<b>ARIMA</b>	<i>Autoregressive Integrated Moving Average</i>
<b>VAR</b>	<i>Vector Autoregression</i>
<b>HMM</b>	<i>Hidden Markov Model</i>
<b>ST</b>	<i>Séries Temporelles</i>
<b>SARIMA</b>	<i>Seasonal Autoregressive Integrated Moving Average</i>
<b>SIM</b>	<i>Subscriber Identification Module</i>
<b>SVM</b>	<i>Support Vector Machine</i>
<b>ARMA</b>	<i>AutoRegressive Moving Average</i>
<b>AR</b>	<i>AutoRegressive</i>

<b>I</b>	<i>Integrated</i>
<b>MA</b>	<i>Moving Average</i>
<b>SVR</b>	<i>Support Vector Regression</i>
<b>SVM</b>	<i>Support Vector Machine</i>
<b>KNN</b>	<i>K-Nearest Neighbors</i>
<b>RNA</b>	<i>Réseau de Neurones Artificiels</i>
<b>ANN</b>	<i>Artificial Neural Network</i>
<b>XOR</b>	<i>Exclusive OR</i>
<b>PMC</b>	<i>Perceptron Multi-Couches</i>
<b>MLP</b>	<i>Multi-Layer Perceptron</i>
<b>MLPNN</b>	<i>Multi-Layer Perceptron Neural Network</i>
<b>RBF</b>	<i>Radial Basis Function</i>
<b>RNCC</b>	<i>Cascade Correlation Neural Network</i>
<b>RNN</b>	<i>Recurrent Neural Network</i>
<b>MNRE</b>	<i>Ministry of New and Renewable Energy</i>
<b>GHI</b>	<i>Global Horizontal Irradiance</i>
<b>DNI</b>	<i>Direct Normal Irradiance</i>
<b>DHI</b>	<i>Diffuse Horizontal Irradiance</i>
<b>BPT</b>	<i>Wet Bulb Temperature</i>
<b>DPT</b>	<i>Dew Point Temperature</i>
<b>PR</b>	<i>Pressure</i>
<b>WD</b>	<i>Wind Direction</i>
<b>WS</b>	<i>Wind Speed</i>

<b>HH</b>	<i>Hour</i>
<b>DD</b>	<i>Number of Days</i>
<b>MM</b>	<i>Number of Months</i>
<b>FFN</b>	<i>Feedforward Network</i>
<b>FFB</b>	<i>Feedback Forward Network</i>
<b>CFB</b>	<i>Cascade Forward Backpropagation Network</i>
<b>ELM</b>	<i>Elman Network</i>
<b>RB</b>	<i>Radial Basis Network</i>
<b>PNN</b>	<i>Probabilistic Neural Network</i>
<b>LSTM</b>	<i>Long Short-Term Memory</i>
<b>LR</b>	<i>Linear Regression</i>
<b>BPNN</b>	<i>Backpropagation Neural Network</i>
<b>WNN</b>	<i>Wireless Neural Network</i>
<b>NRMSE</b>	<i>Normalized Root Mean Squared Error</i>
<b>ETS</b>	<i>Exponential Smoothing State Space</i>
<b>CY</b>	<i>Castilla y León</i>
<b>UTC</b>	<i>Coordinated Universal Time</i>
<b>GSI</b>	<i>Global Sky Index</i>
<b>SR-FFNN</b>	<i>Stepwise Regression Feedforward Neural Network</i>
<b>MLR</b>	<i>Multiple Linear Regression</i>
<b>MBE</b>	<i>Mean Bias Error</i>
<b>FFNN</b>	<i>Feedforward Neural Network</i>
<b>GRNN</b>	<i>Generalized Regression Neural Network</i>

<b>DRNN</b>	<i>Deep Recurrent Neural Network</i>
<b>API</b>	<i>Application Programming Interface</i>
<b>FNN</b>	<i>Feedforward Neural Network</i>
<b>ARN</b>	<i>Artificial Recurrent Network</i>
<b>FA</b>	<i>Forêt Aléatoire</i>
<b>CNN</b>	<i>Convolutional Neural Network</i>
<b>RMSE</b>	<i>Root Mean Square Error</i>
<b>nRMSE</b>	<i>Normalized Root Mean Square Error</i>
<b>MAE</b>	<i>Mean Absolute Error</i>
<b>MSE</b>	<i>Mean Squared Error</i>
<b>MAPE</b>	<i>Mean Absolute Percentage Error</i>
<b>rMAPE</b>	<i>Relative Mean Absolute Percentage Error</i>
<b>rMSE</b>	<i>Relative Mean Squared Error</i>

# Introduction Générale

L'énergie solaire est une source d'énergie renouvelable qui suscite un intérêt croissant en tant que solution durable pour répondre aux besoins énergétiques mondiaux. Elle offre de nombreux avantages, tels que son abondance, sa disponibilité gratuite et son impact environnemental réduit. Afin d'exploiter pleinement le potentiel de l'énergie solaire, il est essentiel de comprendre ses caractéristiques, de prédire son rendement et d'améliorer son utilisation dans diverses applications.

La prédiction de la production d'énergie solaire revêt une importance capitale pour optimiser l'utilisation de cette source d'énergie et faciliter son intégration efficace dans le réseau électrique. Une prédiction précise de l'énergie solaire permet de planifier la production, de mieux gérer les ressources et d'optimiser les opérations des installations solaires.

Dans ce contexte, l'utilisation de techniques d'apprentissage machine et d'intelligence artificielle offre des opportunités prometteuses pour améliorer la prédiction de l'énergie solaire. Ces approches permettent d'exploiter les données historiques et en temps réel afin d'estimer de manière précise la production future d'énergie solaire.

Ce mémoire se concentre sur l'application de l'apprentissage machine pour la prédiction d'énergie solaire à Bejaia, en Algérie. Bejaia bénéficie d'un ensoleillement important, ce qui en fait une région propice à l'utilisation de l'énergie solaire. L'objectif principal de ce mémoire est de développer et d'évaluer des modèles prédictifs robustes pour estimer l'irradiation solaire à Bejaia, en utilisant des techniques d'apprentissage machine.

Le mémoire est structuré en quatre chapitres. Le premier chapitre constitue une introduction essentielle, présentant les bases théoriques et les enjeux de la prédiction de l'énergie solaire. Il met en évidence les différents types d'énergie solaire, le

rayonnement solaire et les données utilisées pour la prédiction, en insistant sur leur pertinence et leur fiabilité. De plus, il aborde l'analyse prédictive qui sera développée dans les chapitres suivants.

Le deuxième chapitre examine les avancées récentes dans le domaine de la prédiction d'énergie solaire. Il présente les modèles prédictifs basés sur l'apprentissage machine, en mettant l'accent sur les techniques telles que la régression linéaire, les réseaux de neurones et les modèles ARIMA. Il explore également les différents facteurs qui influencent la prédiction de l'énergie solaire.

Le troisième chapitre présente une revue des travaux récents réalisés dans le domaine de la prédiction d'énergie solaire, en mettant l'accent sur l'utilisation de l'intelligence artificielle et de l'apprentissage machine. Il met en évidence les avantages et les limitations des approches existantes et met en évidence la nécessité de développer de nouveaux modèles de prédiction et d'optimisation de l'énergie solaire.

Enfin, le quatrième chapitre se concentre sur la méthodologie utilisée pour prédire l'irradiation solaire à Bejaia. Il décrit en détail les différentes étapes suivies, de la collecte et de l'analyse des données à la sélection du modèle, ainsi que l'entraînement, le test et la validation des données. Les résultats obtenus pour différentes méthodes de prédiction sont également présentés et discutés.

# Chapitre 1

## Généralités sur l'énergie solaire.

### 1.1 Introduction

Le domaine de l'énergie solaire suscite un intérêt croissant dans le contexte actuel de la transition énergétique vers des sources d'énergie renouvelables et durables. Dans ce contexte, la prédiction de la production d'énergie solaire est un enjeu majeur pour optimiser l'utilisation de cette source d'énergie. C'est dans ce cadre que nous allons présenter les différents types d'énergie solaire ainsi que le rayonnement solaire, qui sont des éléments clés pour comprendre le fonctionnement des systèmes de production d'énergie solaire. nous allons également décrire les différents données qui vont être utilisées pour la prédiction de l'énergie solaire, en insistant sur leur pertinence et leur fiabilité. Enfin, nous allons aborder l'analyse prédictive qui sera développée dans les chapitres suivants.

Ce premier chapitre est donc une introduction essentielle qui permettra au lecteur de comprendre les bases théoriques et les enjeux de la prédiction de l'énergie solaire.

### 1.2 Énergie Solaire

Toute discussion sur l'énergie solaire devrait commencer par la source d'énergie elle-même. Le Soleil est avant tout l'élément essentiel rendant la vie sur Terre (chaleur, photosynthèse, et création des énergies fossiles).

L'énergie solaire est l'énergie renouvelable la plus importante dans l'échelle mondiale et la tendance actuelle.

L'énergie solaire est une source d'électricité renouvelable très prometteuse, mais

elle n'est pas encore pleinement exploitée. Récemment, une croissance rapide des systèmes solaires photovoltaïques (PV) installés à grande échelle dans les bâtiments résidentiels (sur les toits). Cela est dû à la baisse du coût des panneaux solaires photovoltaïques, à l'amélioration de la technologie et des performances et aux initiatives gouvernementales visant à encourager l'utilisation de ces systèmes [10, 11].

Il existe trois types d'énergie solaire :

### 1.2.1 L'énergie solaire photovoltaïque

L'un des moyens les plus courants d'utiliser l'énergie solaire consiste à utiliser des systèmes photovoltaïques qui génèrent de l'électricité directement à partir de la lumière du soleil.

Lorsque le rayonnement solaire entre en contact avec une face d'une cellule photovoltaïque, il crée une différence de tension entre les deux faces, provoquant un cycle d'électrons d'une face à l'autre, créant un courant électrique.

Ceci est réalisé en installant des panneaux solaires photovoltaïques, constitués de cellules au silicium qui convertissent la lumière et la chaleur du soleil en électricité. Ces panneaux solaires peuvent être installés sur des bâtiments et des maisons à l'échelle domestique, ou sur de plus grandes installations à plus grande échelle.

### 1.2.2 L'énergie solaire thermique

Les systèmes d'énergie solaire thermique collectent et concentrent la lumière du soleil pour générer la chaleur à haute température nécessaire à la production d'électricité. Tous les systèmes de production d'énergie solaire thermique sont équipés de capteurs solaires et se composent de deux éléments principaux : Des réflecteurs et Un système de stockage.

### 1.2.3 L'énergie solaire thermodynamique

Le chauffage par énergie solaire thermodynamique est une technologie qui repose sur un procédé simple et efficace pour chauffer l'eau domestique en utilisant l'énergie naturelle, propre et gratuite de l'environnement [12].

### 1.3 Rayonnement solaire

Le Soleil existe depuis plusieurs milliards d'années. Il est situé au centre de notre système solaire et est composé principalement d'hydrogène (75 %) et d'hélium (25 %). Les proportions varient au fur et à mesure que le Soleil convertit l'hydrogène en hélium par réaction de fusion nucléaire. L'énergie dégagée au sein du Soleil (386 milliards de milliards de MW) est produite par ces réactions de fusion nucléaire. Chaque seconde au cœur du Soleil, environ 700 millions de tonnes d'hydrogène sont converties en 695 millions de tonnes d'hélium et 5 millions de tonnes d'énergie sous forme de rayons gamma. Les conditions de température et de pression au centre du Soleil sont extrêmes : il y règne une température de 15 millions de °C et la pression est de 250 milliards d'atmosphère. Les gaz y sont comprimés à une densité équivalente à 150 fois celle de l'eau. Pendant qu'elle voyage vers la surface du Soleil, l'énergie est continuellement absorbée et réémise à des températures de plus en plus basses. De ce fait, lorsqu'elle en atteint la surface, elle est principalement constituée de lumière visible. Cette énergie émise est appelée rayonnement solaire [13, 14].

Les rayons solaires arrivent en permanence sur Terre et chauffent la surface de notre planète. Sans cet apport d'énergie, la Terre serait glaciale. Avant d'atteindre le sol, le rayonnement incident est fortement influencé par l'atmosphère terrestre (Figure 1.1), à travers des interactions de type absorption et réflexion [15].

L'énergie émise par le Soleil l'est d'abord sous la forme de rayonnements électromagnétiques dont l'ensemble forme le rayonnement solaire, qui constitue la seule source externe notable d'énergie pour l'atmosphère. Un rayonnement solaire se décompose en ondes radio et millimétriques, en émissions dans l'infrarouge, le visible et l'ultraviolet et, au-delà, en rayons X et gamma. Un corps froid comme la Terre émet principalement des ondes radio, millimétriques et infrarouges [16, 17].

Le rayonnement solaire joue un rôle crucial dans la vie sur terre, étant une source essentielle d'énergie solaire. Il permet la photosynthèse, oriente de nombreuses espèces et contribue à l'épuration chimique de l'eau en éliminant certaines molécules écotoxiques ou indésirables. De plus, les rayons UV solaires aident à éliminer les microbes présents dans les couches supérieures des eaux superficielles et océaniques. La chaleur apportée par le soleil a un effet synergique qui renforce l'effet [U+FB00] et

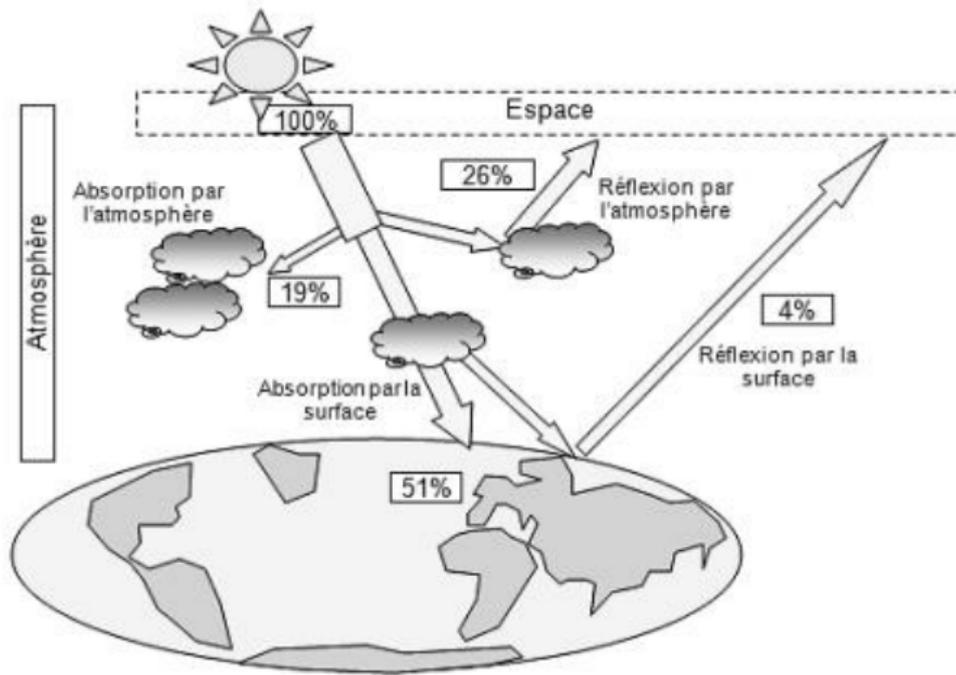


FIGURE 1.1 – Interaction entre le rayonnement solaire et l'atmosphère terrestre [1].

désinfectant des UV solaires [18].

### 1.3.1 Composante déterministe et stochastique du rayonnement solaire

La valeur de l'éclairement en un point du sol et à un instant donné dépend non seulement de la position du Soleil, mais aussi de l'état optique de l'atmosphère [19, 20]. Ce phénomène permet de scinder en deux parties distinctes le phénomène de rayonnement global au sol : une composante que l'on nommera « déterministe » et qui ne dépend que de la distance entre le point de mesure et le centre du Soleil, et une composante que l'on nommera « stochastique » qui dépend de l'état de la couverture nuageuse, du taux d'ozone, du taux d'humidité, etc. La première composante peut se déterminer aisément avec des règles de mécanique céleste et de physique énergétique. Concernant la seconde, elle est beaucoup plus difficile à appréhender du fait purement probabiliste des différents éléments qui la composent : occurrence de nuage, de concentration d'ozone, etc [21, 22, 23].

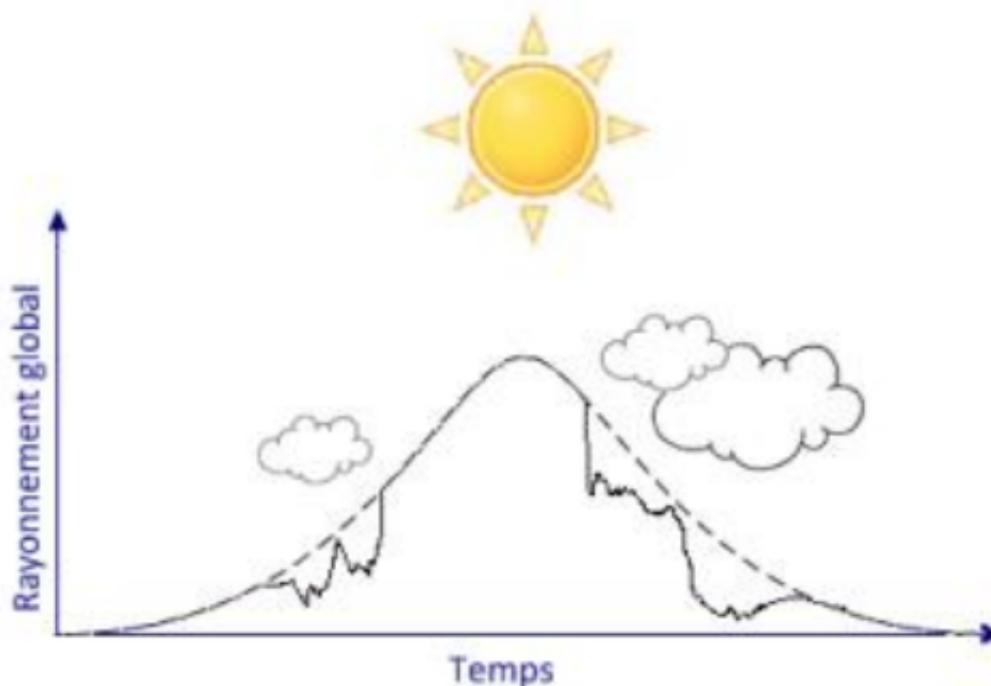


FIGURE 1.2 – Impact de la couverture nuageuse sur la série temporelle de rayonnement global [1]

### 1.3.2 L'exploitation de rayonnement solaire

Le rayonnement solaire peut être exploité à partir de panneaux et de réflecteurs :

#### 1.3.2.1 La conversion photovoltaïque

On peut exprimer la notion de conversion photovoltaïque comme étant la conversion de l'énergie lumineuse, sous forme de photons, en énergie électrique, par le biais d'un processus d'absorption de la lumière par la matière [24]. Lorsqu'un photon interagit avec la matière, il transfère une partie ou l'ensemble de son énergie lors de la collision, entraînant ainsi l'arrachement d'un électron de la matière (l'effet photoélectrique (Figure 1.3)). Grâce à la nature des cellules solaires on peut récupérer ces électrons, ce qui signifie la capacité de production des charges électriques [24].

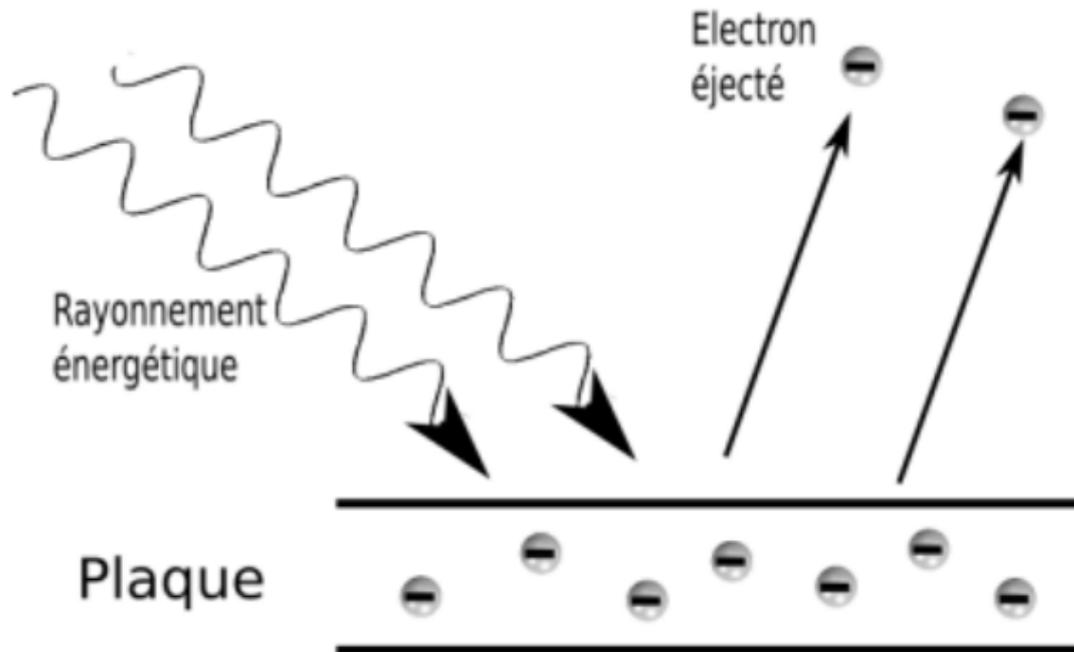


FIGURE 1.3 – principe de l'effet photoélectrique [1].

En général, lorsque toute l'énergie des photons ne peut pas être convertie en électricité, elle est absorbée par le matériau sous forme thermique. En conséquence, la température interne du matériau des capteurs photovoltaïques augmente proportionnellement à l'énergie solaire reçue. L'effet thermique est donc majoritaire sur la plupart des capteurs. Bien que l'effet électrique soit subordonné à l'effet thermique, la récupération totale ou partielle de l'énergie électrique reste l'objectif principal des capteurs photovoltaïques.

### 1.3.2.2 Les systèmes photovoltaïques (PV)

En utilisant des conditions d'ensoleillement standard ( $1000\text{W}/\text{m}^2$ ;  $25^\circ\text{C}$ ), une cellule en silicium de  $150\text{ cm}^2$  peut produire une puissance maximale d'environ  $2,3\text{ Wc}$  à une tension de  $0,5\text{ V}$ . En conséquence, une cellule photovoltaïque individuelle fournit une faible puissance électrique qui n'est pas suffisante pour répondre aux besoins de la plupart des applications domestiques ou industrielles [26]. De ce fait, les générateurs photovoltaïques sont fabriqués en combinant un grand nombre de cellules élémentaires en série et/ou en parallèle. En outre, les cellules photovoltaïques ne produisent que du courant continu en réponse au rayonnement solaire [25].

L'usage du courant continu varie selon l'installation, en fonction de son objectif. Il

existe principalement deux types d'utilisation, selon que l'installation photovoltaïque est connectée ou non à un réseau de distribution d'électricité.

Les installations non connectées peuvent utiliser directement l'électricité produite. Cela est courant pour les sites non reliés au réseau électrique, tels que les zones montagneuses, les îles, les voiliers, les satellites, etc. Ces sites peuvent également être équipés de batteries ou d'accumulateurs pour stocker l'électricité et l'utiliser pendant les périodes sans lumière, notamment la nuit.

Les installations photovoltaïques peuvent être connectées à un réseau de distribution, où elles produisent de l'électricité et la fournissent au réseau. Pour cela, ces installations convertissent le courant continu en courant alternatif qui convient aux caractéristiques du réseau (la fréquence par exemple). L'électricité est consommée instantanément par les consommateurs les plus proches sur le réseau, donc, il n'est pas nécessaire d'avoir des installations de stockage telles que des batteries. Ce raccordement nécessite l'utilisation d'un onduleur qui transforme le courant continu en courant alternatif identique à celui circulant sur le réseau. Bien qu'il existe des onduleurs de différentes technologies, leur principe et leur objectif sont les mêmes : convertir l'énergie du courant continu (DC) en courant alternatif (AC) [1].

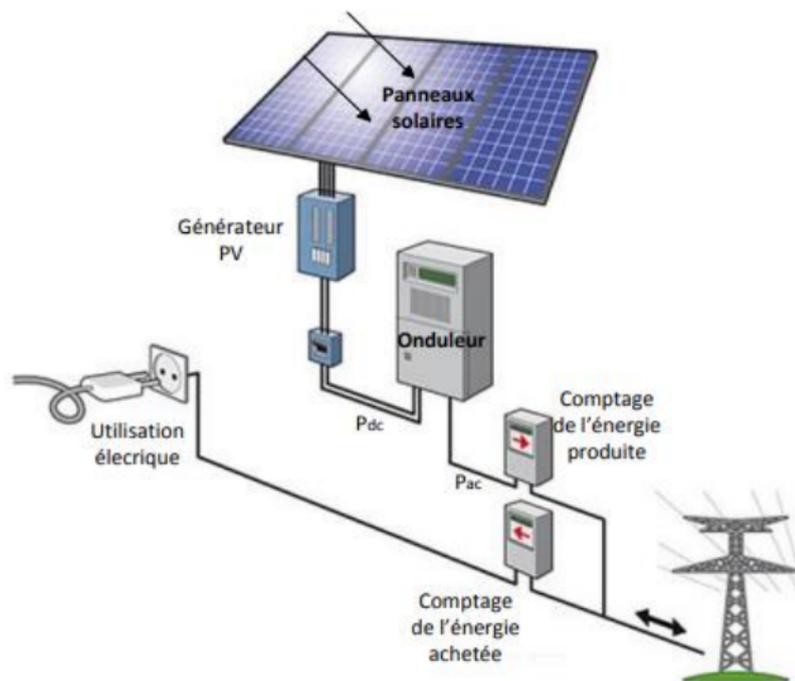


FIGURE 1.4 – Principe du raccordement d'un module l'énergie PV au réseau électrique [1].

### 1.3.2.3 Les capteurs solaires thermiques

Les capteurs solaires utilisent des miroirs ou des panneaux pour concentrer les rayons du soleil. Ces rayons chauffent un fluide qui crée de la vapeur pour actionner une turbine et produire de l'électricité [12].

## 1.4 Les données

IL existe différentes méthodes et approches sur l'utilisation des données pour l'extraction d'informations et la prédiction. Pour prendre des décisions basées sur des données on recense trois catégories d'analyse de données : descriptive, prédictive et prescriptive. Mais dans notre étude et recherche on concentre sur l'analyse prédictive qui est considérée comme un type d'exploration de données.

Les données jouent un rôle crucial dans la prédiction de l'énergie solaire. Les modèles de prédiction se basent sur des données météorologiques, des observations solaires, des images satellite, des données historiques et des modèles mathématiques. Plus les données sont précises et nombreuses, plus les prévisions sont fiables. Les techniques d'apprentissage automatique, telles que les réseaux de neurones, sont souvent utilisées pour analyser ces données et construire des modèles de prédiction.

Par la suite, nous allons voir qu'est-ce que l'analyse prédictive, aussi présenter les données en rapport direct ou indirect avec la prédiction. Ces données sont appelés les variables endogènes, et les variables exogènes [29 , 30].

## 1.5 L'analyse prédictive

L'analyse prédictive a fait l'objet d'une grande attention au cours des dernières années en raison des progrès enregistrés dans les technologies qui le supportent, en particulier dans les domaines du Big Data et du Machine Learning, et elle est de plus en plus utilisée par les entreprises et notamment par les départements marketing, les institutions financières et mêmes les organismes médicaux. En combinant une analyse de données historiques et nouvelles, elle aide les organisations à anticiper des tendances, à prévoir et évaluer des risques et ainsi à prendre des décisions optimales pour mener les bonnes actions auprès des bonnes personnes et au bon moment.

### 1.5.1 Définitions

L'analyse prédictive permet de prévoir des événements futurs à partir de données historiques. On utilise les données historiques ainsi que des techniques d'analyse, de statistiques et de Machine Learning pour créer un modèle mathématique permettant de capturer les tendances importantes. Ce modèle prédictif est ensuite utilisé sur les données actives pour prévoir ce qui va se produire, ou encore pour suggérer des mesures à prendre afin d'optimiser les résultats.

Le terme « analyse prédictive » décrit l'application d'une technique statistique ou de Machine Learning visant à établir une prévision quantitative du futur. Souvent, des techniques de machine Learning supervisé sont utilisées pour prévoir une valeur future.

Au départ, l'analyse prédictive est associée à un objectif commercial, qu'il s'agisse d'utiliser des données pour réduire le gaspillage, gagner du temps ou réduire les coûts. Le processus transforme des ensembles de données hétérogènes et souvent massifs en modèles capables de produire des résultats clairs et exploitables qui aident à atteindre cet objectif, qu'il s'agisse de réduire le gaspillage de matériau, de réduire l'inventaire en stock ou de fabriquer des produits conformes aux spécifications [26].

### 1.5.2 Les variables utilisés pour la prédiction

#### 1.5.2.1 Variables endogènes

Les variables dites endogènes sont directement liées à la prédiction désirée, par exemple dans la prédiction de l'énergie solaire, les variables endogènes peuvent inclure des mesures telles que :

- L'intensité de la lumière solaire : L'intensité de la lumière solaire est la quantité de lumière solaire qui atteint les panneaux solaires et qui est convertie en énergie électrique. C'est la variable la plus importante dans la prédiction de l'énergie solaire, car elle a un effet direct sur la production d'énergie solaire.

- Le rendement des panneaux solaires (ou la production d'énergie solaire cumulée sur une certaine période de temps) : Le rendement des panneaux solaires est la capacité des panneaux solaires à convertir la lumière solaire en énergie électrique. Un rendement plus élevé signifie que les panneaux solaires peuvent convertir plus d'énergie solaire en énergie électrique, ce qui peut augmenter la production d'énergie

solaire.

- La température des panneaux solaires : La température des panneaux solaires peut affecter la production d'énergie solaire, car une augmentation de la température peut entraîner une diminution de l'efficacité de conversion d'énergie des panneaux solaires.

- La surface des panneaux solaires : La surface des panneaux solaires est la quantité d'espace physique qui est utilisé pour capturer la lumière solaire. Plus la surface des panneaux solaires est grande, plus il est possible de capter de la lumière solaire, ce qui peut augmenter la production d'énergie solaire.

- L'angle d'incidence de la lumière solaire : L'angle d'incidence de la lumière solaire est l'angle entre la lumière solaire et les panneaux solaires. Lorsque la lumière solaire est perpendiculaire à la surface des panneaux solaires, elle est absorbée et convertie en électricité avec une efficacité maximale. Cependant, lorsque l'angle d'incidence est plus grand, une partie de la lumière est réfléchi, ce qui réduit l'efficacité des panneaux solaires. L'angle d'incidence dépend de la position du soleil dans le ciel, de l'heure de la journée, de la latitude et de l'orientation des panneaux solaires.

Ces variables sont influencées par d'autres variables à l'intérieur du modèle de prédiction de l'énergie solaire, telles que la position du soleil dans le ciel, la couverture nuageuse, l'heure de la journée et la saison, la topographie du site...etc [1].

### 1.5.2.2 Les variables exogènes

Les variables dites exogènes ne sont pas directement liées à la prédiction désirée. Elles sont les facteurs externes qui influencent la production d'énergie solaire, mais ne sont pas influencés par la production d'énergie solaire elle-même. Voici quelques exemples de variables exogènes dans le domaine de la prédiction de l'énergie solaire :

- La température ambiante : La température ambiante peut affecter la production d'énergie solaire, car une augmentation de la température peut entraîner une diminution de l'efficacité de conversion d'énergie des panneaux solaires.

- La couverture nuageuse : La couverture nuageuse peut réduire l'intensité de la lumière solaire qui atteint les panneaux solaires, ce qui peut réduire la production d'énergie solaire.

- La vitesse et la direction du vent : Bien que la vitesse et la direction du vent soient généralement considérées comme des variables exogènes, elles peuvent

également avoir un effet indirect sur la production d'énergie solaire, car elles peuvent affecter la température ambiante et donc le rendement des panneaux solaires.

- L'heure de la journée et la saison : La quantité de lumière solaire disponible varie en fonction de l'heure de la journée et de la saison, ce qui peut avoir un impact sur la production d'énergie solaire.

- Les caractéristiques géographiques : Les caractéristiques géographiques, telles que l'altitude, la latitude et l'ombrage, peuvent affecter la quantité de lumière solaire qui atteint les panneaux solaires, ce qui peut avoir un impact sur la production d'énergie solaire [1].

## 1.6 Conclusion

Dans ce chapitre, nous avons vu les différentes énergies en général, et en particulier l'énergie solaire. Nous avons examiné les différents types d'énergie solaire, notamment le photovoltaïque, le thermique et la thermodynamique, ainsi que la définition du rayonnement solaire et ses composants. Nous avons également exploré les méthodes permettant d'exploiter le rayonnement solaire pour produire de l'énergie utilisable. Ensuite, nous avons abordé l'analyse prédictive et les données, notamment les variables exogènes et endogènes, qui jouent un rôle important dans la prédiction de l'énergie solaire. Dans le chapitre suivant, nous approfondirons les séries temporelles, les différents travaux et méthodes sur la prédiction d'énergie solaire (rayonnement solaire), et nous effectuerons une comparaison entre ces méthodes.

# Chapitre 2

## Prédiction d'énergie solaire

### 2.1 Introduction

La prédiction de l'énergie solaire est un domaine de recherche en constante évolution au cours des dernières décennies, en particulier depuis l'avènement des technologies de l'information et des communications. Les premiers efforts pour prédire l'énergie solaire ont été basés sur des modèles simples qui ne prenaient en compte que des données météorologiques telles que la température, la pression atmosphérique et l'humidité relative. Cependant, ces modèles ont montré leurs limites, car ils ne prenaient pas en compte les variations saisonnières et diurnes de la luminosité solaire.

Au fil du temps, des modèles plus sophistiqués ont été développés, utilisant des techniques telles que l'apprentissage automatique et l'analyse statistique pour prédire l'énergie solaire. Ces modèles ont pris en compte des données plus complexes telles que la couverture nuageuse, la densité optique de l'aérosol, l'albédo du sol et l'inclinaison et l'orientation des panneaux solaires.

Avec l'avènement des réseaux de capteurs solaires, qui collectent des données en temps réel sur la production d'énergie solaire et les conditions météorologiques locales, les prévisions ont pu être affinées davantage. Les réseaux de capteurs solaires sont souvent intégrés dans des systèmes de prédiction de l'énergie solaire plus larges qui peuvent être utilisés pour planifier la production d'énergie et optimiser les opérations des centrales solaires.

Dans ce chapitre, nous allons parler sur les série temporelles ensuite, nous présenterons les différentes techniques de prédiction (classiques et modernes (techniques basées sur l'IA)) telles que la régression linéaire, les réseaux de neurones, les modèles ARIMA...etc, et les différents facteurs qui influencent la prédiction.

### 2.2 Série temporelle

Une série temporelle est définie comme étant une suite de mesures ou d'observations au cours du temps représentant un phénomène. Ces données sont souvent caractérisées par une évolution temporelle et sont utilisées dans de nombreux domaines, tels que les sciences économiques, la météorologie, la finance, la biologie, et bien sûr l'énergie solaire.

La prédiction des séries temporelles consiste à utiliser des données historiques pour prévoir les valeurs futures de la série. L'objectif est de créer un modèle mathématique qui puisse capturer les relations entre les observations antérieures et la valeur future de la série. Les techniques de prédiction de séries temporelles peuvent être utilisées pour prévoir des quantités telles que la production d'énergie solaire, les ventes d'une entreprise, les prix de marché, les conditions météorologiques futures, etc.

Les méthodes de prédiction de séries temporelles utilisent souvent des techniques de Machine Learning, telles que la régression linéaire, les réseaux de neurones, les modèles ARIMA (AutoRégressive Integrated Moving Average), les modèles à vecteur auto-régressif (VAR), les processus de Markov cachés (HMM) et d'autres algorithmes de prédiction basés sur ces modèles.

La prédiction des séries temporelles est une tâche complexe, car elle dépend de nombreux facteurs, tels que la qualité des données, la période de prédiction, le choix de l'algorithme de prédiction, etc. C'est pourquoi une analyse préalable des données et une évaluation régulière de la performance du modèle sont essentielles pour obtenir des prévisions précises et fiables [1, 27].

### 2.2.1 Décomposition d'une série temporelle

L'analyse des séries temporelles a souvent pour objectif de séparer la série en différentes composantes, afin de mieux comprendre son comportement et d'effectuer des prévisions plus précises. La séparation de la série est traditionnellement effectuée en trois composantes principales : la tendance, la composante saisonnière, la composante résiduelle [1, 27].

- **La tendance (le trend)  $T(t)$**  : Représente l'allure d'ensemble de la série temporelle sur une période relativement longue (quelques années, voire plusieurs décennies). Elle peut être croissante, décroissante ou constante. La tendance peut être influencée par des facteurs économiques ou environnementaux qui changent lentement avec le temps, comme la croissance démographique ou les évolutions technologiques [1, 27].
- **La composante saisonnière  $S(t)$**  : Est une composante cyclique relativement régulière de période intra-annuelle. Il existe différents types de saisonnalités pour le rayonnement global, on peut parler de périodicité intrinsèque « rigide », ou déterministe car elle est bien marquée et répétitive [1, 27].
- **La composante résiduelle (bruit ou résidu)  $R(t)$**  : Correspond à des fluctuations irrégulières, en général de faible intensité mais de nature aléatoire. C'est une composante qui existe par défaut, elle regroupe ce que les autres composantes n'ont pu intégrer [1, 27].

Afin d'améliorer la précision du modèle, il est important de choisir le schéma de décomposition qui convient le mieux à la nature du processus étudié. Ce schéma de décomposition consiste à combiner les trois composantes élémentaires d'une série temporelle (à savoir la tendance, la saisonnalité et la composante résiduelle) dans une structure qui reflète le mieux les propriétés inhérentes de la série temporelle [1, 27].

Il existe généralement trois grands schémas de décomposition qui peuvent être appliqués à la plupart des séries temporelles, chacun étant adapté à des types spécifiques de séries temporelles en fonction de leurs caractéristiques et de leur comportement :

- **Le schéma additive  $O(t)$**  : Cette méthode est utilisée lorsque la magnitude

de la saisonnalité est constante et ne varie pas en fonction de la tendance de la série. Dans ce cas, la série temporelle est décomposée en une somme de la tendance, de la saisonnalité et de la composante aléatoire [1, 27].

$$O(t) = T(t) + S(t) + R(t)$$

- **Le schéma multiplicative  $O(t)$**  : Nécessaire dès qu'il y a une interaction générale des trois composantes. Dans ce cas, la série temporelle est décomposée en une multiplication de la tendance, de la saisonnalité et de la composante résiduelle [1, 27].

$$O(t) = T(t) * S(t) * R(t)$$

### 2.3 Les modèles et techniques de prédiction

La prédiction des séries temporelles est un problème qui recouvre de nombreux domaines d'application. Les études et les travaux récents menées dans l'énergie, la finance et l'économétrie ont permis de dégager de nombreux modèles plus ou moins sophistiqués. Ces derniers ont été repris dans le cadre d'autres thématiques, dont la prédiction d'énergie solaire par modélisation des séries temporelles.

Il existe de nombreux modèles permettant de faire une prédiction de ST. Il est possible de les rassembler en quatre grands groupes [1]. :

- **Les modèles de type « naïf »** : Ces modèles sont simples et faciles à mettre en place. Ils sont souvent utilisés comme point de départ pour évaluer la pertinence des modèles plus complexes. On peut citer La persistance, la moyenne et Les k plus proches voisins..etc.
- **Les modèles à probabilités conditionnelles** : Ces modèles utilisent des concepts de probabilité conditionnelle pour prédire la valeur future de la série temporelle. Les modèles à probabilités conditionnelles incluent :  
Les chaînes de Markov, Les prédictions basées sur les inférences Bayésiennes.
- **Les modèles de référence** : Les modèles de référence sont souvent utilisés comme point de comparaison pour les nouveaux modèles de prédiction. Ils ont été largement étudiés et leur performance est bien documentée. Les modèles

de référence incluent : Les modèles ARIMA, SARIMA.

- **Les modèles de type connexionnistes** : Ces modèles sont basés sur les réseaux de neurones artificiels et sont très populaires en apprentissage automatique (réseau de neurone, apprentissage profonds...)

Il convient de noter que le nombre de modèles existants pour la prédiction de séries temporelles et d'énergie solaire est très important. Les méthodes de prédiction d'énergie solaire sont basées sur l'utilisation des données historiques de Rayonnement solaire et les variables météorologiques associées. Ces modèles sont les modèles naïfs, les modèles statistiques, modèles physiques et les modèles hybrides, comme indiqué à la figure 2.1 avec les sous-catégories et quelques techniques mais faut noter qu'il existe d'autres techniques mais on a essayé de mentionner les techniques les plus utilisées.

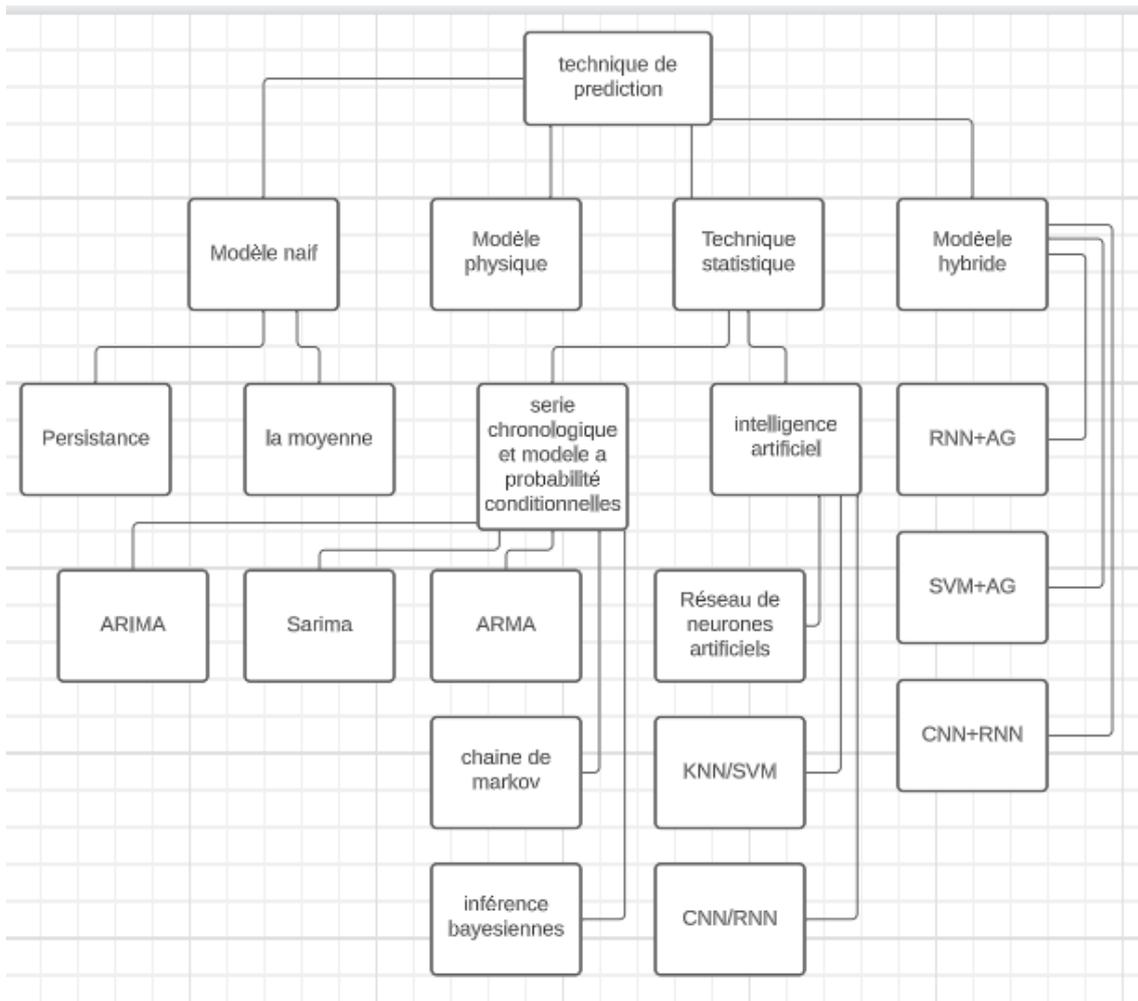


FIGURE 2.1 – Classification des techniques de prévision.

### 2.3.1 Modèle naïf

#### 2.3.1.1 Persistance

Le modèle de persistance est une méthode simple et naïve couramment utilisée pour les prévisions à court terme. Elle est basée sur l'hypothèse que les conditions climatiques resteront similaires d'un jour à l'autre et que la valeur prédite demeurera constante dans un futur proche. Cette technique peut être utile pour des prévisions à très court terme, car elle offre une précision raisonnable tout en ayant des coûts de calcul moindres et un délai plus court.

La persistance consiste à répéter la dernière mesure de l'instant  $t$  à l'instant  $t+h$  pour faire une prédiction de la valeur à l'instant  $t+h$  (où  $h$  est l'horizon de prédiction). Par exemple, pour prédire la valeur de la puissance PV à l'instant  $t+1$ , on utilise simplement la mesure de la puissance PV à l'instant  $t$ . Cependant, cette méthode peut ne pas être suffisamment précise pour des prévisions à plus long terme ou dans des conditions changeantes.

En outre, la persistance peut être le seul prédicteur utilisable dans certaines situations, car elle ne nécessite pas d'historique de la série temporelle, seule la valeur à l'instant  $t$  est nécessaire [28, 1].

#### 2.3.1.2 La moyenne

La moyenne est simple à mettre en œuvre mais nécessite un historique important pour être précise. Si l'historique de données est insuffisant, la moyenne risque de ne pas prendre en compte les fluctuations de mesure et la prédiction sera similaire à celle obtenue avec la méthode de persistance. Dans le cas d'une série temporelle périodique de  $T$  périodes avec un historique de  $N.T$  mesures (par exemple, des données de rayonnement solaire sur une année), la prédiction à l'instant  $t+h$  correspond à la somme des mesures précédentes pondérées par un facteur décroissant en fonction de l'écart entre la mesure et la dernière période [1].

### 2.3.2 Modèle physique

Le modèle physique est une méthode de modélisation de la production d'énergie solaire qui repose sur des équations mathématiques décrivant les lois physiques

régissant les phénomènes naturels impliqués dans la production d'énergie solaire.

Cette méthode prend en compte des paramètres tels que la position et l'orientation des panneaux solaires, l'angle d'incidence des rayons solaires, les propriétés optiques des surfaces et la transmission de l'air pour calculer les flux d'énergie solaire entrants sur les cellules photovoltaïques.

Les modèles physiques permettent ainsi de prédire avec précision la production d'énergie solaire à différents moments de la journée et dans des conditions météorologiques changeantes.

Cette technique est généralement plus fiable en prévision à long terme. Le modèle physique est une méthode de modélisation très précise pour la prédiction de la production d'énergie solaire. Cette méthode est particulièrement adaptée à la conception et à l'optimisation de systèmes solaires, mais peut être complexe et coûteuse en termes de ressources informatiques [29, 6].

### 2.3.3 Techniques Statistiques

Les techniques statistiques pour la prévision sont des méthodes utilisées pour prédire l'avenir en utilisant des données historiques. À travers plusieurs chercheurs, on peut diviser les techniques statistiques en deux groupes : techniques statistiques basées sur les séries chronologiques (modèles linéaires) et techniques statistiques basées sur l'intelligence artificielle [6, 7] (comportement non linéaire) comme l'apprentissage automatique et profond qui représentent une approche alternative face aux techniques conventionnelles classiques. Ils sont très efficaces pour traiter un ensemble de données incomplètes ou bruyantes, ainsi que résoudre des problèmes dont les dépendances existantes présentent des relations non linéaires entre les paramètres d'entrée et de sortie. Les méthodes les plus populaires et importantes sont décrites ci-dessous :

#### 2.3.3.1 Les chaînes de Markov

Les chaînes de Markov sont des modèles stochastiques qui permettent de modéliser des processus dynamiques discrets, tels que les séries chronologiques. Dans le contexte de la prédiction de l'énergie solaire, les chaînes de Markov peuvent être

utilisées pour modéliser la variation de l'irradiance solaire dans le temps.

L'idée est de considérer que les variations à un instant donné dépendent uniquement de l'état précédent et non pas de l'historique complet de la série chronologique. Cette propriété est appelée "absence de mémoire", elle est l'un des fondements théoriques des chaînes de Markov.

Cependant, il est important de noter que les chaînes de Markov sont des modèles simplistes qui ne prennent pas en compte les relations causales ou les tendances à long terme des données des séries chronologiques. Par conséquent, leur utilisation peut être limitée dans certains contextes de prédiction d'énergie solaire [30, 1].

### 2.3.3.2 Inférences Bayésiennes

L'inférence bayésienne pour la prédiction des séries chronologiques repose sur la notion de probabilité conditionnelle, qui est la probabilité qu'un événement se produise sachant que d'autres événements se sont déjà produits. Cette méthode suppose que les événements passés influencent les événements futurs, et permet de prendre en compte les informations historiques pour faire des prédictions probabilistes sur les valeurs futures de la série chronologique.

En d'autres termes, l'inférence bayésienne pour la prédiction des séries chronologiques consiste à utiliser les connaissances a priori sur les données historiques pour estimer la distribution de probabilité des valeurs futures de la série chronologique. Cela permet de prendre en compte l'incertitude associée à la prévision et de fournir des intervalles de confiance pour les prédictions.

Cette méthode est particulièrement utile pour les séries chronologiques où les valeurs observées sont influencées par des facteurs externes tels que la météo, les fluctuations économiques, etc. L'inférence bayésienne permet d'incorporer ces facteurs externes dans le modèle de prédiction [31, 1].

### 2.3.3.3 Les modèles autorégressifs

Ces techniques permettent d'estimer la relation entre une variable dépendante et des variables indépendantes. Nous trouvons ici Les méthodes ARMA (AutoRegressive Moving Average), ARIMA (AutoRegressive Integrated Moving Average) et

SARIMA (La moyenne mobile intégrée autorégressive saisonnière), elles sont des modèles de prévision de séries temporelles qui combinent les termes d'autorégression et de moyenne mobile pour capturer les relations de dépendance dans les données temporelles, avec la différence que les modèles ARIMA comprennent également une étape d'intégration pour stationnariser la série temporelle [7].

Le modèle ARMA est un modèle autorégressif stationnaire dans lequel les variables indépendantes suivent des tendances stochastiques et le terme d'erreur est stationnaire. En d'autres termes, le modèle ARMA intègre l'autocorrélation et le modèle de moyenne mobile dans sa régression. ARMA est divisé en deux parties :

- Autorégressif : La variable dépendante revient à elle-même dans un laps de tempst.
- Moyenne mobile : Les revers sont représentés par des processus aléatoires [32].

ARIMA (Moyenne Mobile Intégrée Autorégressive) est une méthode de prévision pour les données de séries chronologiques univariées, qui permet de modéliser à la fois les éléments autorégressifs et les moyennes mobiles. La partie "intégrée" de l'ARIMA fait référence à la différenciation, qui permet de prendre en compte les séries chronologiques avec une tendance en supprimant cette tendance dans les données.

Cependant, un inconvénient majeur d'ARIMA est qu'il ne prend pas en charge les données saisonnières. Les séries chronologiques avec des cycles répétitifs, tels que les données de production d'énergie solaire qui peuvent varier en fonction des saisons, nécessitent une méthode plus sophistiquée pour modéliser la saisonnalité. C'est là que l'extension de l'ARIMA appelée SARIMA (Moyenne Mobile Intégrée Autorégressive Saisonnière)

Le modèle SARIMA (Moyenne Mobile Intégrée Autorégressive Saisonnière) ou ARIMA saisonnier est une extension du modèle ARIMA qui est spécifiquement conçue pour traiter les séries chronologiques univariées avec une composante saisonnière. Il introduit trois nouveaux hyperparamètres qui permettent de spécifier l'autorégression (AR), la différenciation (I) et la moyenne mobile (MA) pour la

composante saisonnière de la série. De plus, un autre paramètre est ajouté pour indiquer la période de la saisonnalité. Cette extension du modèle ARIMA permet de mieux capturer les variations saisonnières régulières qui peuvent être présentes dans des séries chronologiques comme la production d'énergie solaire [33].

### 2.3.3.4 Lissage exponentiel

Brown a d'abord proposé la méthode de lissage exponentiel, communément appelée méthode de lissage exponentiel simple. Cette approche a été développée par Holt, qui l'a appelée la méthode Holt. De plus, Winter a modifié cette méthode et l'a nommée méthode Holt-Winter [7]. Il existe plusieurs variantes de cette méthode, notamment la méthode de lissage exponentiel simple, double et triple [34]. Dans la méthode de lissage exponentiel, un ensemble de pondérations des données historiques au lieu des pondérations égales est imposé aux données passées. Cependant, les poids des données passées diminuent de manière exponentielle des points de données les plus récents aux plus distants.

### 2.3.3.5 La régression

La régression est une méthode statistique couramment utilisée pour évaluer les relations entre des variables. Elle peut être utilisée pour détecter des tendances importantes dans de vastes ensembles de données et souvent utilisée pour déterminer l'influence de facteurs spécifiques, tels que le prix, sur les variations d'un actif.

La régression linéaire et la régression logistique sont deux types courants d'analyse de régression. Dans la régression linéaire, une variable indépendante est utilisée pour expliquer et/ou prédire le résultat de la variable de réponse, également appelée variable  $Y$ . La régression multiple, quant à elle, utilise deux variables indépendantes ou plus pour prédire ce résultat.

Dans la régression logistique, des variables inconnues d'une variable discrète sont prédites en fonction de la valeur connue d'autres variables. La variable de réponse est catégorique, ce qui signifie qu'elle ne peut prendre qu'un nombre limité de valeurs. Dans la régression logistique binaire, par exemple, la variable de réponse n'a que deux valeurs possibles, telles que 0 ou 1. Dans la régression logistique multiple, une variable de réponse peut comporter plusieurs niveaux, par exemple bas, moyen et

haut, ou 1, 2 et 3.

Cependant, il est important de noter que la régression est conçue pour les données continues censées suivre une distribution normale. Par conséquent, pour concevoir un modèle de prédiction basé sur la régression, un modèle mathématique et plusieurs variables explicatives sont nécessaires, ce qui peut être considéré comme une limitation de cette méthode [35].

### 2.3.3.6 Machines à vecteurs de support

Il s'agit d'une méthode de modélisation supervisée, développée pour être utilisée dans les problèmes de classification. Lorsqu'elles sont appliquées à des problèmes de régression, elles sont appelées machines de régression vectorielle (Support Vector Regression ou SVR). Elles se distinguent par leur forte capacité de généralisation et leur capacité à traiter des problèmes non linéaires. Elles fonctionnent comme une régression linéaire multiple utilisant des prédicteurs transformés tout en conservant une faible complexité et un bon ajustement des données. Trois paramètres principaux dominent les performances de la technique et doivent être ajustés : la précision, le paramètre de coût, qui concerne le compromis entre précision et complexité, et qui régule la fonction du noyau, utilisée pour transformer les prédicteurs en un espace de fonctions de plus grande dimension. Les SVM/SVR ont montré un grand potentiel dans plusieurs études [7].

### 2.3.3.7 K- plus proches voisins (k-NN)

La méthode des k plus proches voisins (kNN) est une technique d'apprentissage automatique supervisé et non paramétrique qui permet de classer des données. Elle est considérée comme l'une des méthodes les plus simples en apprentissage machine. L'algorithme de kNN fonctionne en comparant un point de données à des échantillons d'apprentissage dans un espace de fonctions et en déterminant les k échantillons les plus proches (en termes de distance euclidienne). Ensuite, les classes de ces k échantillons sont examinées et la classe majoritaire est attribuée au point de données [4].

### 2.3.3.8 Les arbres de décision

Les arbres de décision sont une des structures de données clés en apprentissage statistique. Leur fonctionnement repose sur des heuristiques intelligentes, qui en plus de satisfaire notre intuition, produisent des résultats remarquables dans la pratique, notamment lorsqu'ils sont utilisés en combinaison avec des techniques de forêts aléatoires. Leur caractéristique arborescente leur confère également une grande lisibilité pour les êtres humains, ce qui les distingue d'autres approches où les prédicteurs construits peuvent être considérés comme des boîtes noires hermétiques. Les arbres de décision offrent ainsi un équilibre unique entre performance prédictive et interprétabilité, permettant aux utilisateurs de comprendre le raisonnement sous-jacent des modèles et d'identifier les relations causales entre les variables. Cette transparence accrue facilite la prise de décisions éclairées et favorise la confiance dans les résultats obtenus [36].

### 2.3.3.9 Les forêts aléatoires

Les forêts aléatoires, introduites par L. Breiman au début des années 2000, se révèlent être des algorithmes extrêmement efficaces tant du point de vue computationnel que prédictif, même lorsqu'ils sont appliqués à de vastes ensembles de données. Les forêts aléatoires suivent le principe de la "diviser pour régner". La forêt est formée de plusieurs arbres, chacun étant construit à partir d'un sous-ensemble aléatoire du jeu de données. Ainsi, la prédiction de la forêt est obtenue en agrégeant simplement les prédictions de chaque arbre. Cette approche présente plusieurs avantages significatifs. Premièrement, elle réduit le risque de surapprentissage, car chaque arbre est construit sur une partie différente des données. Deuxièmement, elle permet une meilleure généralisation, car les arbres sont formés de manière indépendante et donc diversifiée. Enfin, elle offre une évaluation de l'importance des variables, permettant ainsi d'identifier les caractéristiques les plus influentes pour les prédictions. Dans l'ensemble, les forêts aléatoires sont devenues une méthode de choix pour de nombreux problèmes d'apprentissage automatique en raison de leur robustesse, de leur flexibilité et de leurs performances élevées [37].

### 2.3.3.10 Extra-Arbres

Extra-Arbres utilise un ensemble de décisions aléatoires (ou arbres supplémentaires) ajustés sur divers sous-échantillons de l'ensemble de données. Cette méthode vise à améliorer la précision prédictive tout en contrôlant le surajustement. Comparé à la méthode des forêts aléatoires, qui sélectionne la répartition optimale, Extra-Arbres effectue cette sélection de manière aléatoire. Cependant, une fois les points de partage choisis, les deux algorithmes examinent tous les sous-ensembles de fonctionnalités possibles pour choisir le meilleur. Par conséquent, Extra-Arbres introduit une randomisation supplémentaire tout en maintenant une optimisation. Son approche consiste à fortement randomiser le choix des attributs et des points de coupure lors de la division d'un nœud d'arbre. Dans certains cas extrêmes, des arbres totalement aléatoires sont construits, dont les structures ne dépendent pas des valeurs de sortie de l'échantillon d'apprentissage. Cette randomisation accrue peut conduire à une plus grande diversité d'arbres et potentiellement à une amélioration de la performance prédictive dans des situations où les données présentent des caractéristiques complexes ou des bruits élevés [38].

### 2.3.3.11 Réseau de neurones artificiels

Les réseaux de neurones artificiels ( RNA ) sont des ensembles organisés de neurones interconnectés qui sont utilisés dans le domaine de l'intelligence artificielle. Les RNA sont capables d'apprendre et de reconnaître des modèles dans les données, et ils ont été appliqués dans divers domaines, tels que la reconnaissance d'images et de la parole, le traitement du langage naturel et la modélisation prédictive [39, 40, 41].

Les principaux composants d'un ANN sont les couches d'entrée, caché et de sortie, les neurones et les connexions. La couche d'entrée recevant diverses informations d'entrée, la couche cachée analysant les informations d'entrée et la couche de sortie fournissant les résultats analysés. La connexion établit un lien entre les neurones dans différentes couches avec la mise à jour de son poids. La figure 3.3 montre l'architecture de base de l'ANN et un modèle schématique de la procédure de traitement dans une cellule neuronale, où la cellule a deux parties : la « fonction de combinaison » qui ajoute toutes les entrées et la « fonction d'activation ». Plusieurs

types de fonction d'activation sont utilisés, tels que la Fonction tout/rien (à seuil), Fonction linéaire, Fonction sigmoïde et Tangente hyperbolique..etc, servant à limiter l'amplitude du signal de sortie du neurone et reproduire l'effet de seuil observé sur les neurones biologiques.

### 1. Histoire

Peut-on reproduire artificiellement le comportement du cerveau humain ? Cette interrogation est à la base des réflexions ayant ensuite donné naissance aux réseaux de neurones artificiels. Les scientifiques espéraient ainsi recréer le fonctionnement cérébral, pour répliquer certaines capacités humaines au sein d'une machine.

En premier lieu, il a donc été question de comprendre le mécanisme opérant dans le cerveau humain. En 1881, l'anatomiste allemand Heinrich Wilhelm Waldeyer établit la théorie selon laquelle le système nerveux est constitué d'une multitude de cellules assurant le traitement des signaux. Il propose alors un nom pour ces éléments : « neurones ».

Des années plus tard, l'étude du système nerveux quitte le champ purement biologique. En effet les premières recherches dans le domaine des réseaux de neurones artificiels remontent à la fin du 19ème et au début du 20ème siècle. Ces travaux multidisciplinaires ont été menés par des scientifiques tels qu'Hermann von Helmholtz, Ernst Mach et Ivan Pavlov en physique, en psychologie et en neurophysiologie. À l'époque, il s'agissait principalement de théories générales sans modèle mathématique précis pour un neurone.

On considère que Mac Culloch et Pitts, ont donné naissance au premier modèle mathématique de neurone biologique, qu'ils ont appelé : "neurone formel" En 1948. Ils ont montré que ces réseaux étaient en principe capables de calculer n'importe quelle fonction arithmétique ou logique. Vers 1949, Donald Hebb a proposé une théorie fondamentale pour l'apprentissage dans ce domaine.

Les travaux des années 1940 ont ensuite inspiré Frank Rosenblatt, un psychologue et informaticien américain. La première application concrète des

réseaux de neurones artificiels est survenue en 1957 avec l'invention du réseau appelé "perceptron" par Frank Rosenblatt.

Malgré ces progrès significatifs, les recherches sur les réseaux de neurones artificiels ont connu un coup d'arrêt conséquent à partir de 1969. Cette année-là, Marvin Lee Minsky et Seymour Papert ont publié le livre "Perceptrons", dans lequel ils ont mis en évidence les limites théoriques des travaux de Frank Rosenblatt. Ils ont notamment montré l'impossibilité avérée de réaliser certaines fonctions logiques, comme le "OU exclusif" (XOR), à partir d'un seul neurone artificiel. L'ouvrage a eu un fort retentissement et les travaux sur les réseaux de neurones artificiels ont été petit à petit délaissés vers la fin des années 1970.

Heureusement, certains chercheurs ont continué à travailler en développant de nouvelles architectures et algorithmes plus puissants. Teuvo Kohonen et James Anderson ont créé indépendamment et simultanément de nouveaux réseaux qui peuvent servir de mémoires associatives.

Dans les années 1980, un obstacle majeur a été surmonté avec l'invention de l'algorithme de "rétropropagation des erreurs" et "perceptron multicouche". Ces inventions est la réponse aux critiques de Minsky et Papert formulées à la fin des années 1960. C'est ce nouveau développement, généralement attribué à David Rumelhart et James McClelland, mais aussi découvert plus ou moins en même temps par Paul Werbos et par Yann LeCun, qui a littéralement ressuscité le domaine des réseaux de neurones. Depuis ce temps les réseaux de neurones artificiels connaîtront un essor considérable et de nouvelles théories, de nouvelles structures et de nouveaux algorithmes. Grâce aux progrès cités précédemment, mais également au développement de l'informatique [42].

### 2. Neurone biologique

Le cerveau se compose d'environ 1012 neurones inter-connectés entre eux, avec 1000 à 10000 synapses par neurone. Les neurones ne sont pas tous identiques et ils n'ont pas le même comportement.

Un « neurone biologique » est une cellule faisant partie du système nerveux et

responsable de traiter l'information, d'analyser et de résoudre des problèmes, de se souvenir, de composer, de rêver et de ressentir dans tout le corps. Il existe des modèles mathématiques de neurones biologiques, connus sous le nom de modèles de neurones impulsionnels, qui décrivent les propriétés des neurones[2]. Du point de vue neurobiologique, un neurone se compose de quatre éléments clés : le noyau (soma), les dendrites, l'axone et les synapses (Figure 2.2)

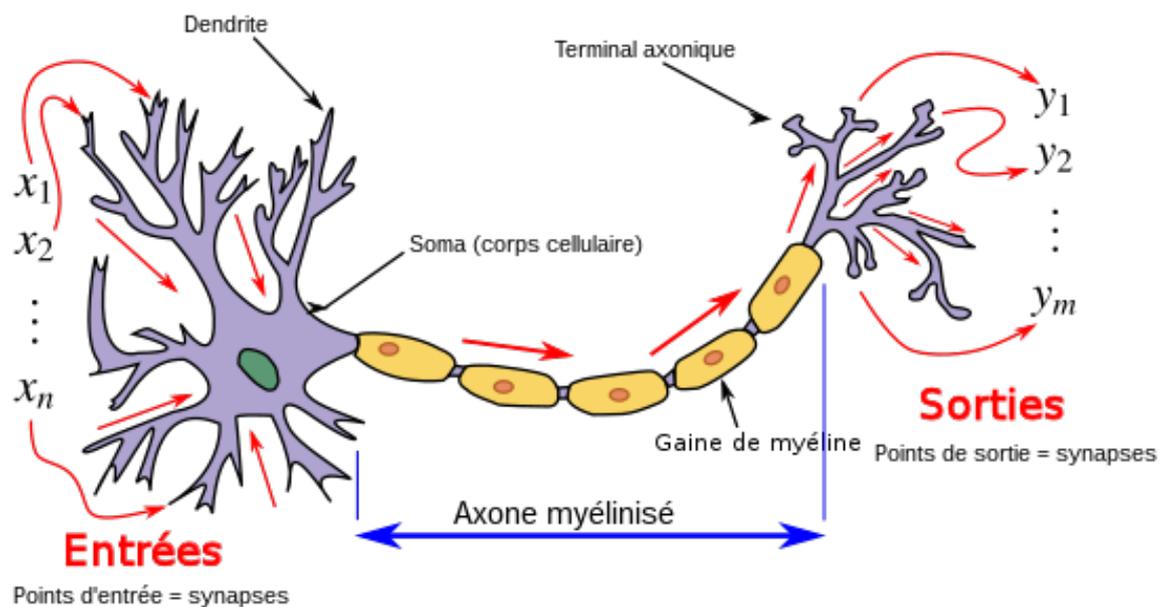


FIGURE 2.2 – Représentation schématique d'un neurone biologique [2].

- **Le corps cellulaire** : Il s'agit du "centre de commande" du neurone, qui reçoit les signaux en provenance des dendrites et effectue une sommation des influx nerveux par ses dendrites. Si la somme est supérieure à un seuil donné, le neurone répond par un flux nerveux ou potentiel d'action qui se propage le long de son axone. Si la somme est inférieure au seuil, le neurone reste inactif.
- **Les dendrites** : Ils sont des prolongements du corps cellulaire qui reçoivent les signaux en provenance d'autres cellules.
- **L'axone** : Il s'agit d'un prolongement du corps cellulaire qui diffuse le signal du neurone vers d'autres cellules.

- **Synapse** : Il s'agit d'une zone de communication entre les neurones, où l'axone d'un neurone peut se connecter aux dendrites d'un autre neurone. La synapse joue également un rôle important dans la modulation des signaux qui transitent le système nerveux.

### 3. Neurone artificiel (formel)

Pour mieux comprendre le fonctionnement des réseaux de neurones, il est utile d'expliquer le fonctionnement mathématique d'un neurone simple. Le neurone artificiel est un modèle de calcul conçu selon le fonctionnement d'un neurone biologique. Il peut être vu comme un opérateur qui reçoit des entrées provenant de l'environnement ou d'autres neurones, chaque entrée étant associée à un poids qui représente la force de la connexion. Chaque neurone a une sortie unique qui se connecte à plusieurs autres neurones situés en aval [3].

- **Un vecteur d'entrée** : Un neurone artificiel reçoit les informations provenant des entrées  $x_i (i = 1, 2, 3, \dots, n)$  Avec  $i$  : l'indice du signal entrant et  $n$  : le nombre de variables d'entrée.
- **Un biais d'entrée** : Le biais d'entrée  $b_i$  permet d'ajouter de la flexibilité au réseau.
- **Les poids synaptiques** : Facteur multiplicateur qui affecte l'influence de chaque entrée sur la sortie du neurone. Les valeurs des poids sont désignées par  $w_i$ , avec  $i$  l'indice du signal d'entrée ( $i=1, 2, 3, \dots, n$ ).
- **Le noyau** : Le noyau effectue la somme pondérée des entrées avec le biais et calcule la sortie du neurone  $y_i$  selon une fonction d'activation  $f$ . La sortie du neurone peut être modélisée par l'équation (2.1)

$$y_i = f(\text{net}) = f\left(b_i + \sum_{i=1}^n w_i x_i\right) \quad (2.1)$$

- **La sortie** : La sortie du neurone  $y_i$  représente la sortie du système qui peut

être distribuée vers d'autres neurones.

Dans la figure (2.3) nous avons présenté un schéma récapitulatif d'un neurone formel.

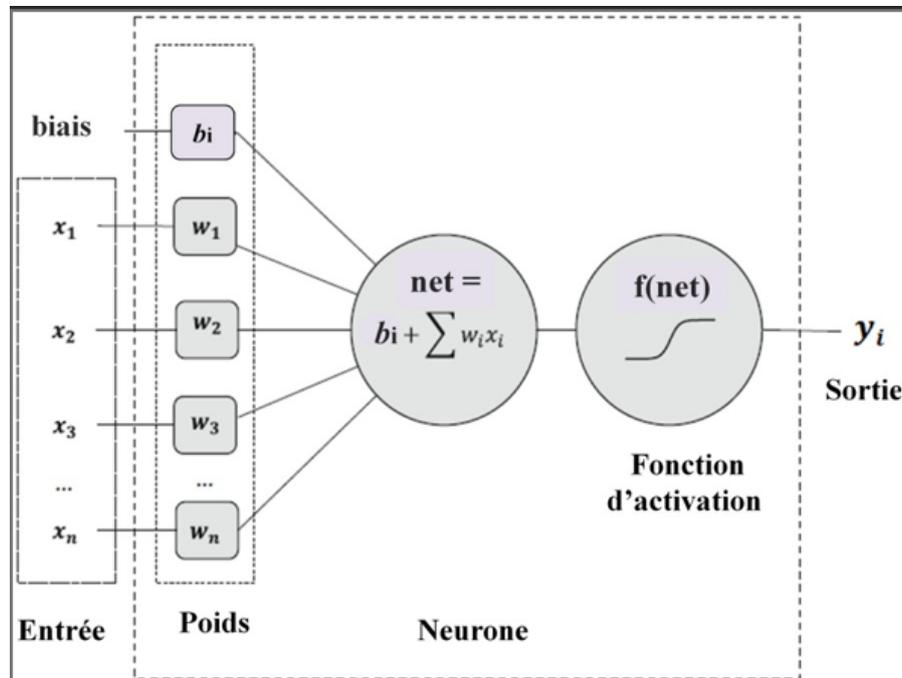


FIGURE 2.3 – Représentation mathématique d'un neurone artificiel [3].

4. **Fonction de transfert(activation)** Plusieurs types de fonction d'activation sont utilisés, servant à limiter l'amplitude du signal de sortie du neurone et reproduire l'effet de seuil observé sur les neurones biologiques. Le tableau (3.1) énumère ces différentes fonctions.

Cependant, certaines fonctions sont les plus utilisées, telles que : la fonction à seuil, la fonction linéaire, et la fonction sigmoïde.

- **La fonction à seuil** : Les versions originales utilisaient une fonction d'activation de type « seuil », c'est-à-dire une fonction qui prenait uniquement deux valeurs, 0 ou 1 Elle retourne 0 pour une entrée négative et 1 pour une entrée positive ou nulle (L'équation 3.2).

$$f(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases} \quad (2.2)$$

- **La fonction linéaire** : La fonction linéaire est très simple, elle affecte directement son entrée à sa sortie.(équation 3.3)

$$f(x) = x \quad (2.3)$$

- **Fonction sigmoïde** : Plus connue et plus populaire. Elle est souvent utilisée dans les réseaux de neurones. L'un de ses avantages est sa dérivabilité. Cette propriété est très importante, du fait qu'elle a permis de développer des algorithmes d'apprentissage à base du gradient pour les réseaux multicouches.

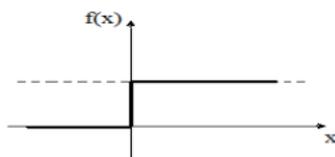
Deux formes de sigmoïdes sont utilisées :

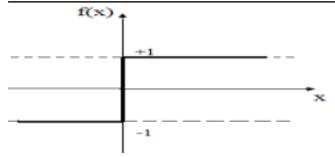
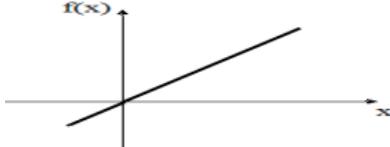
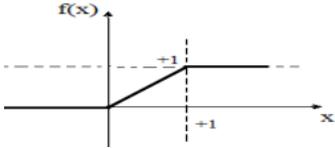
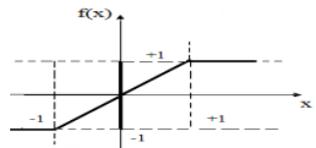
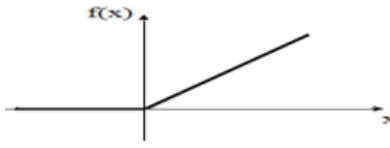
- Sigmoïde (logistique) : Ayant deux niveaux de saturation (0 et 1) est définie comme suit :

$$f(x) = \frac{1}{1 + e^{-x}} \quad (2.4)$$

- Tangente hyperbolique : Peut être exprimée par la tangente-hyperbolique suivante :

$$f(x) = \tanh(x) = \frac{1 - e^{-2x}}{1 + e^{-2x}} \quad (2.5)$$

Nom de la fonction	Relation entrée / sortie	Diagram
Seuil	$f(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases}$	

Seuil symétrique	$f(x) = \begin{cases} -1, & x < 0 \\ 1, & x \geq 0 \end{cases}$	
Linéaire	$f(x) = x$	
Linéaire saturée	$f(x) = \begin{cases} 0, & x < 0 \\ x, & 0 \leq x \leq 1 \\ 1, & x > 1 \end{cases}$	
Linéaire saturée Symétrique	$f(x) = \begin{cases} -1, & x < -1 \\ x, & -1 \leq x \leq 1 \\ 1, & x > 1 \end{cases}$	
Linéaire positive	$f(x) = \begin{cases} 0, & x < 0 \\ x, & 0 \leq x \end{cases}$	

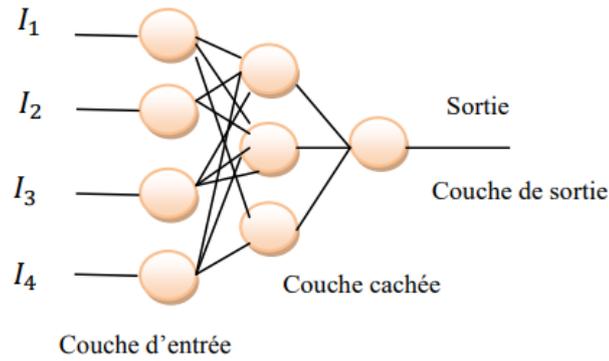
Sigmoïde	$f(x) = \frac{1}{1 + e^{-x}}$	
Tangente hyperbolique	$f(x) = \tanh(x) = \frac{1 - e^{-2x}}{1 + e^{-2x}}$	
compétitive	a = 1 si n maximum a = 0 autrement	pas de diagramme

TABLE 2.1 – Fonctions de transfert f(x).

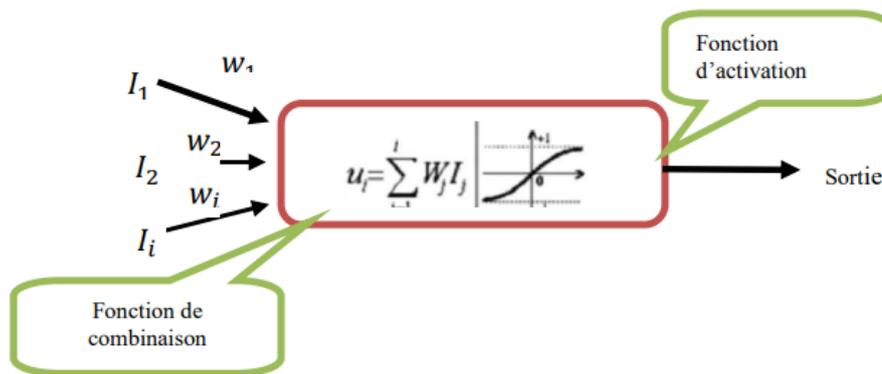
5. **Architecture des réseaux de neurones** : Les réseaux de neurones se déclinent en différentes architectures, chacune ayant ses propres avantages et applications spécifiques en fonction du problème à résoudre. L'architecture d'un réseau de neurones est définie par les connexions entre ses neurones, ce qui détermine sa topologie. En fonction de la topologie des connexions, les réseaux de neurones peuvent être classés en deux grandes catégories : les réseaux non bouclés et les réseaux bouclés.

— **Le perceptron** : Le perceptron était la première signification complète de l'architecture adaptative, développé dans les années 1957 dans le cadre d'un effort du gouvernement américain pour développer un outil utile pour les photo-interprètes par Frank Rosenblatt. Il est limité à deux couches, une couche d'entrée et une couche de sortie, Les neurones de la couche d'entrée sont reliés au monde extérieur et reçoivent le vecteur d'entrée. Les données d'entrée sont transmises aux neurones de la couche cachée  $y_j$  qui calculent leur sortie selon la relation (3.6) [43, 44] :

$$y_j = f\left(b_j + \sum_{i=1}^n w_{ij}x_i\right) \tag{2.6}$$



(a)



(b)

FIGURE 2.4 – (a) Diagramme schématisé d'une structure RNA composée d'une couche d'entrée, une couche cachée et une couche de sortie. (b) modèle mathématique du neurone artificiel [4].

$i$  étant l'indice du neurone de la couche d'entrée,  $n$  : le nombre de variables d'entrée.  $w_{ij}$  : le poids reliant le  $i^{me}$  neurone d'entrée au  $j^{me}$  neurone de sortie.  $b_j$  est le biais associé au neurone de sortie  $j$ . La sortie du neurone  $j$  est calculée suivant la fonction de transfert 'f' à seuil [3].

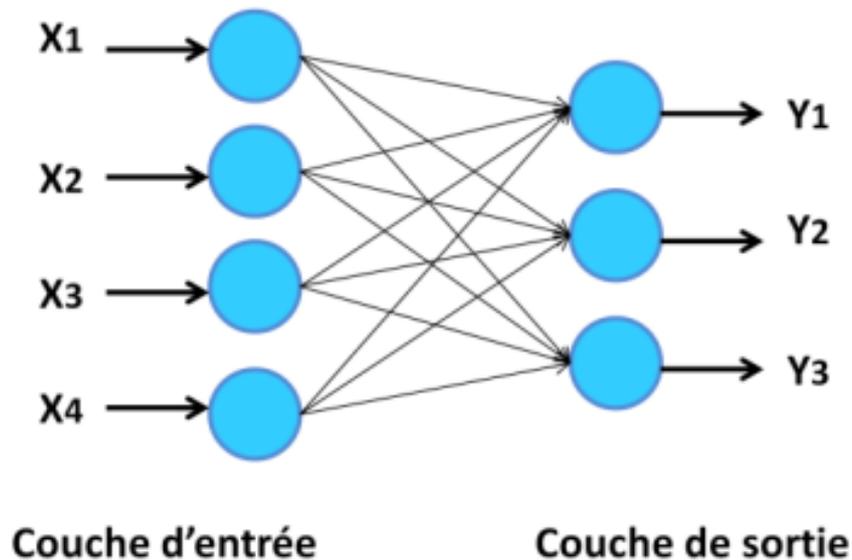


FIGURE 2.5 – Exemple d'un perceptron à deux couches (quatre paramètres d'entrée et trois paramètres de sortie) [3].

- **L'Adaline** : L'architecture ADALINE (Adaptive Linear Element) a été inventée par Bernard Widrow en 1959. Dans cette architecture, les neurones sont connectés de la même manière que les perceptrons, mais avec une différence clé : les ADALINE utilisent une fonction de transfert à seuil bipolaire. Cela signifie que la sortie de l'ADALINE peut prendre les valeurs  $\pm 1$ , plutôt que seulement 0 ou 1 comme dans les perceptrons traditionnels. Cette fonction de transfert à seuil bipolaire permet une plus grande flexibilité dans la représentation des données et facilite l'apprentissage des poids des connexions du réseau [3].
- **Le perceptron multicouche (PMC)** : Il s'agit d'une technique d'approche élémentaire et efficace basée sur des réseaux de neurones artificiels (ANN) pour la conception et la prédiction. Ce modèle est si puissant qu'il est utilisé dans l'approximation universelle, la modélisation non linéaire et les problèmes complexes qui ne peuvent pas être résolus par un réseau de neurones à une seule couche ordinaire. En général, le MLP est constitué d'au moins trois couches de nœuds activés de manière non linéaire. Les nœuds dans chaque couche sont connectés aux nœuds de la couche suivante par des poids. Grâce à un apprentissage adéquat, ce mo-

dèle peut corrélérer la relation d'entrée et de sortie. La corrélation entre le nombre de nœuds et la couche cachée est essentielle pour le MLP. Certaines recherches ont montré qu'une seule couche cachée est suffisante pour concevoir une fonction non linéaire compliquée si le nombre de nœuds est suffisant. Cependant, augmenter le nombre de nœuds peut entraîner des difficultés de sur-apprentissage et d'entraînement. La figure 3.5 montre la structure simplifiée d'un MLPNN avec huit entrées utilisé pour la prédiction de l'énergie solaire [6].

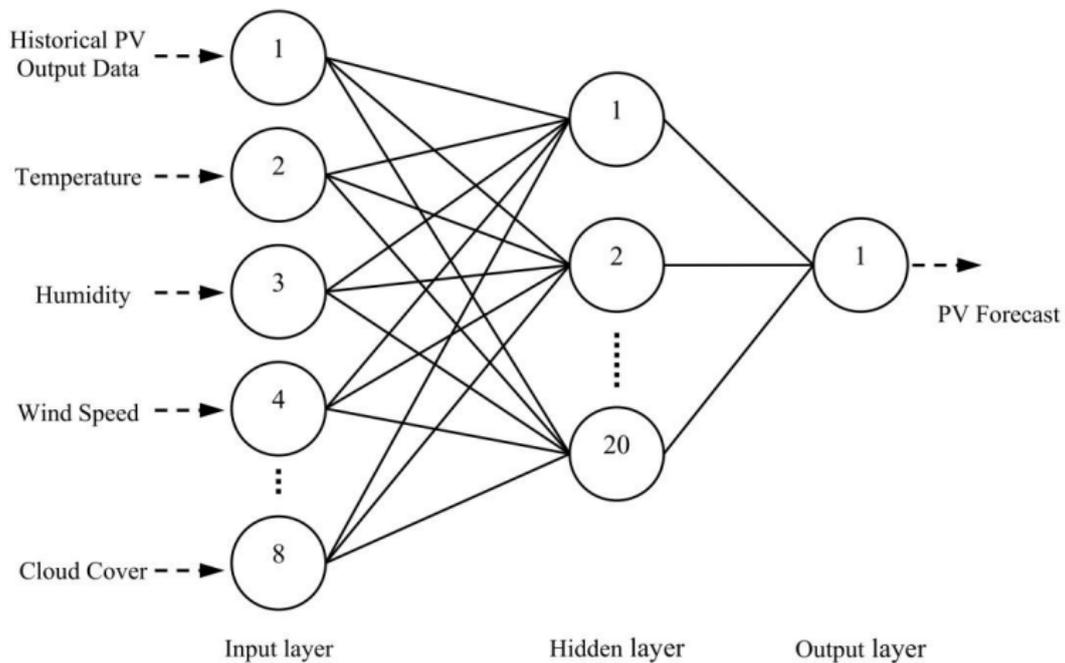


FIGURE 2.6 – Réseau de neurones perceptrons multicouche [5].

- **Multicouche à connexions locales** : Dans ce type de réseau de neurones, appelé réseau de neurones multicouche avec connectivité partielle, tous les neurones d'une couche amont ne sont pas connectés à tous les neurones de la couche aval (figure 3.6). Contrairement au réseau de neurones multicouche classique où chaque neurone d'une couche est connecté à tous les neurones de la couche suivante, la connectivité partielle réduit le nombre de connexions synaptiques dans le réseau. Cela permet de réduire la complexité du réseau et d'optimiser les ressources, tout en maintenant la capacité de représenter et d'apprendre des modèles complexes [4].
- **Réseau à fonction radial (RBF)** : Le réseau RBF (Radial Basis Func-

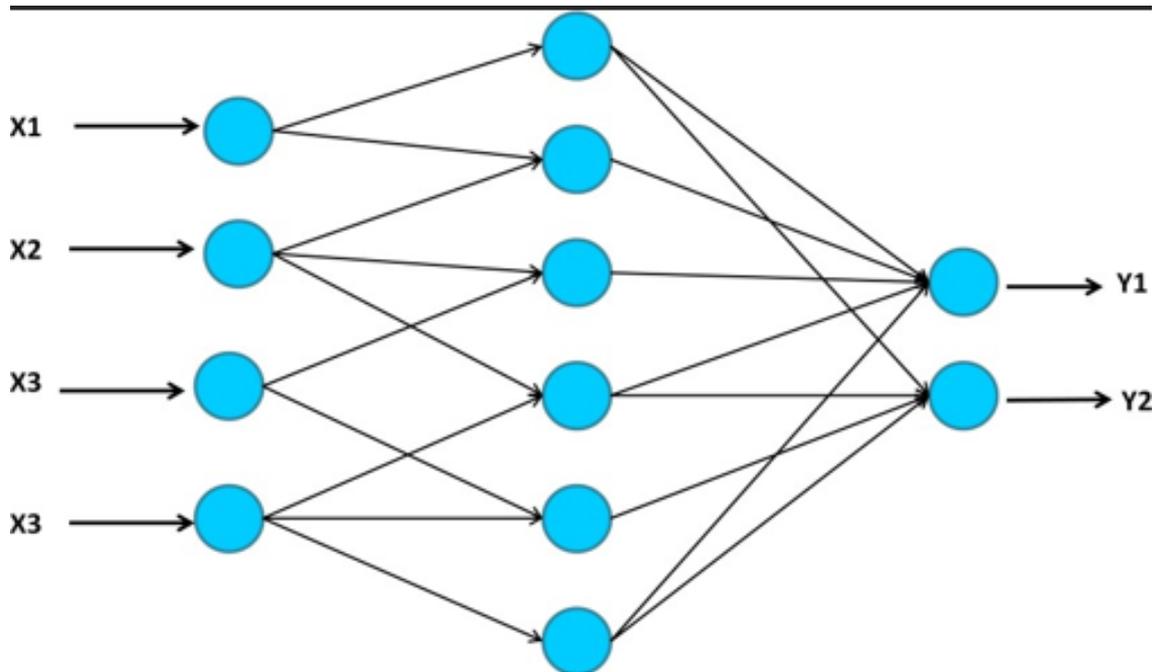


FIGURE 2.7 – Réseau multicouche à connexions locales [3].

tion), a été introduit par Powell et Broomhead en 1988. Il est composé de trois couches (Figure 3.7) : une couche d'entrée, une couche cachée contenant des neurones RBF généralement basés sur des fonctions gaussiennes, et une couche de sortie dont les neurones sont généralement activés par une fonction linéaire [3].

Le réseau RBF est utilisé pour résoudre des problèmes similaires à ceux des Perceptrons Multi-Couches (PMC). Pour un réseau RBF avec  $n$  entrées et  $m$  neurones cachés, la sortie du réseau  $y_j$  est exprimée par la relation (3.7) :

$$y_j = b_j + \sum_{i=1}^m w_{ij} \phi_i \quad (2.7)$$

$$\phi_i = \exp \frac{(x - c_i)^2}{2\sigma_i^2} \quad (2.8)$$

Avec :

$\phi_i$  : la sortie du  $i$  ème neurone de la couche cachée.

$w_{ij}$  : le poids reliant le  $i$  ème neurone de la couche cachée au  $j$  ème neurone de sortie.

$b_j$  : le biais associé au  $j$  ème de la sortie.

$c_i$  : le centre du  $i$  ème neurone de la couche cachée.  $c_i = (w_{i1}, w_{i2}, \dots, w_{in})$ .

$\sigma_i^2$  : la largeur gaussienne du  $i$  ème neurone de la couche cachée.

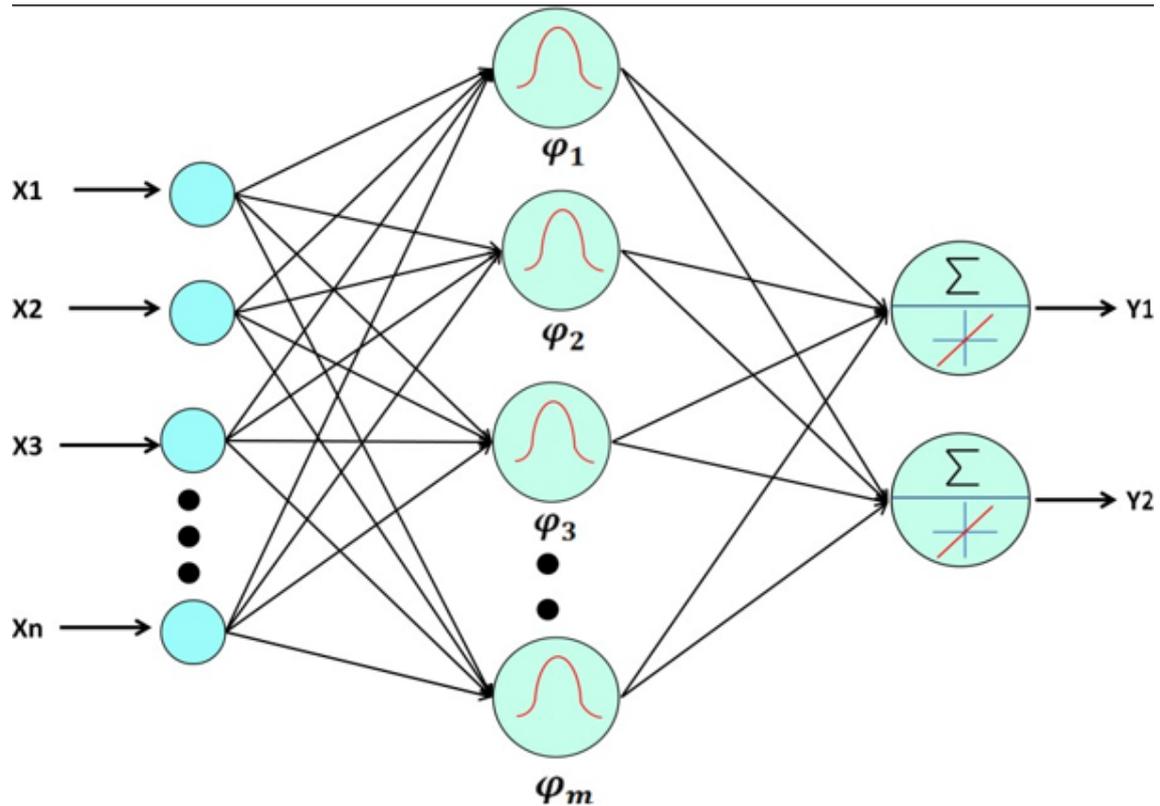


FIGURE 2.8 – L'architecture d'un réseau de neurones de type RBF [3].

- **Réseau de neurones à corrélation en cascade (RNCC)** : Les réseaux neuronaux à corrélation en cascade (Cascade-Forward Backpropagation Network) sont composés de trois couches : une couche d'entrée, une couche cachée et une couche de sortie. Dans ce type de réseau, les variables d'entrée sont reliées aux neurones de la couche cachée et de sortie par des connexions directes (Figure 3.8). En d'autres termes, un neurone de la couche cachée  $y_j$  effectue la somme pondérée des signaux provenant de tous les neurones d'entrée et les neurones cachés [3]. La sortie du réseau  $y_j$  est exprimée par la relation (3.9) :

$$y_j = b_j + \sum_{k=1}^n w_{kj}x_k + \sum_{i=1}^m \phi_i w_{ij} \quad (2.9)$$

Avec :

$x_k$  : le  $k$  éme neurone d'entrée.

$w_{kj}$  : le poids reliant le  $k$  éme neurone d'entrée au  $j$  éme neurone de la couche de sortie.

$\phi_i$  : la sortie du  $i$  ème neurone de la couche cachée.

$w_{ij}$  : le poids reliant le  $i$  ème neurone de la couche cachée au  $j$  ème neurone de sortie.

$b_j$  : le biais associé au  $j$  ème neurone de la sortie.

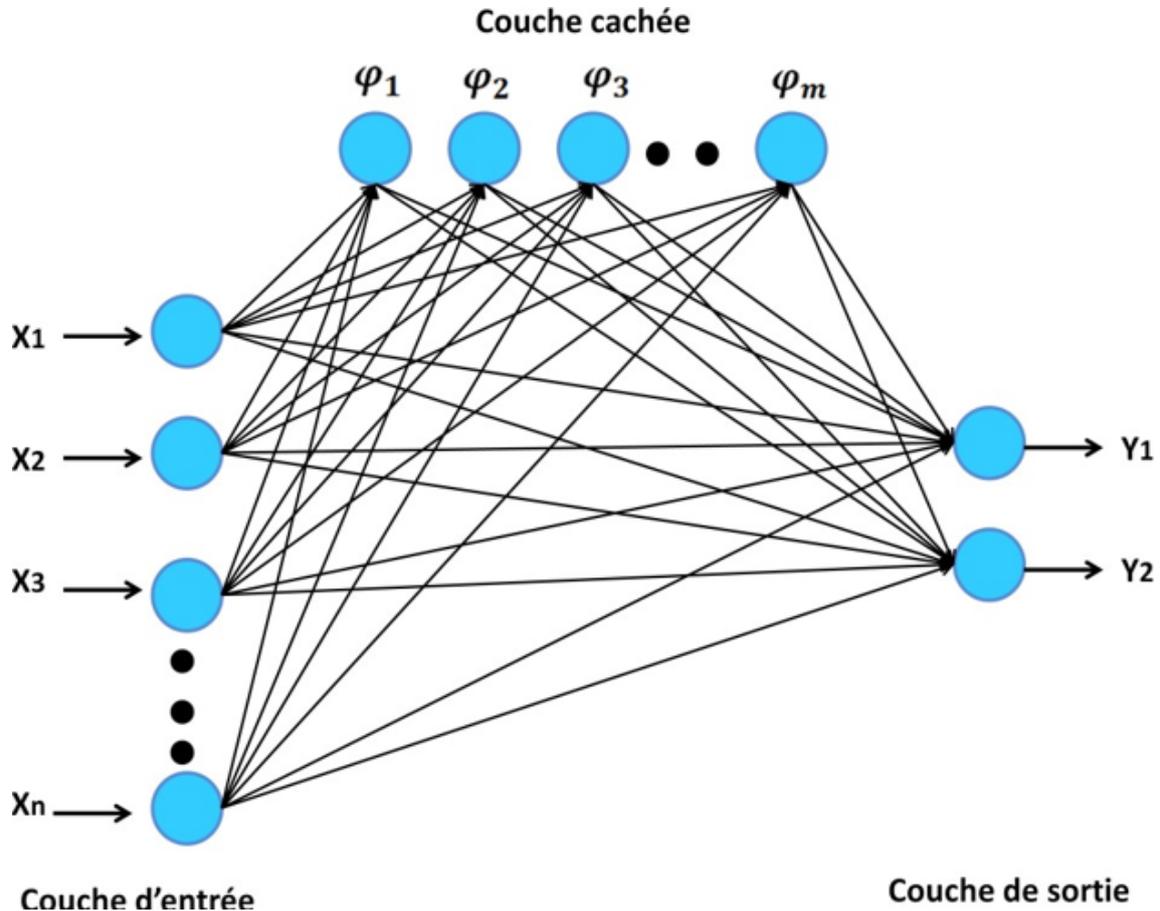


FIGURE 2.9 – L'architecture d'un réseau de neurones à corrélation en cascade [3].

- **Réseau de neurones bouclés (récurrents) :** Le réseau neuronal récurrent (RNN) est un type de réseau neuronal qui se distingue des réseaux neuronaux traditionnels par le fait que la sortie de l'étape précédente est utilisée comme entrée à l'étape courante (figure 3.9). Contrairement aux réseaux neuronaux classiques où toutes les entrées et sorties sont indépendantes les unes des autres, RNN est conçu pour prendre en compte les dépendances entre les éléments d'une séquence. Par exemple, pour prédire le mot suivant dans une phrase, il est nécessaire de considérer les mots précédents. RNN a été développé pour résoudre ce problème en utilisant une couche cachée qui permet de stocker des informations sur la séquence. La caractéristique clé de RNN est son état caché, qui conserve en mémoire

des informations importantes sur la séquence.[45]

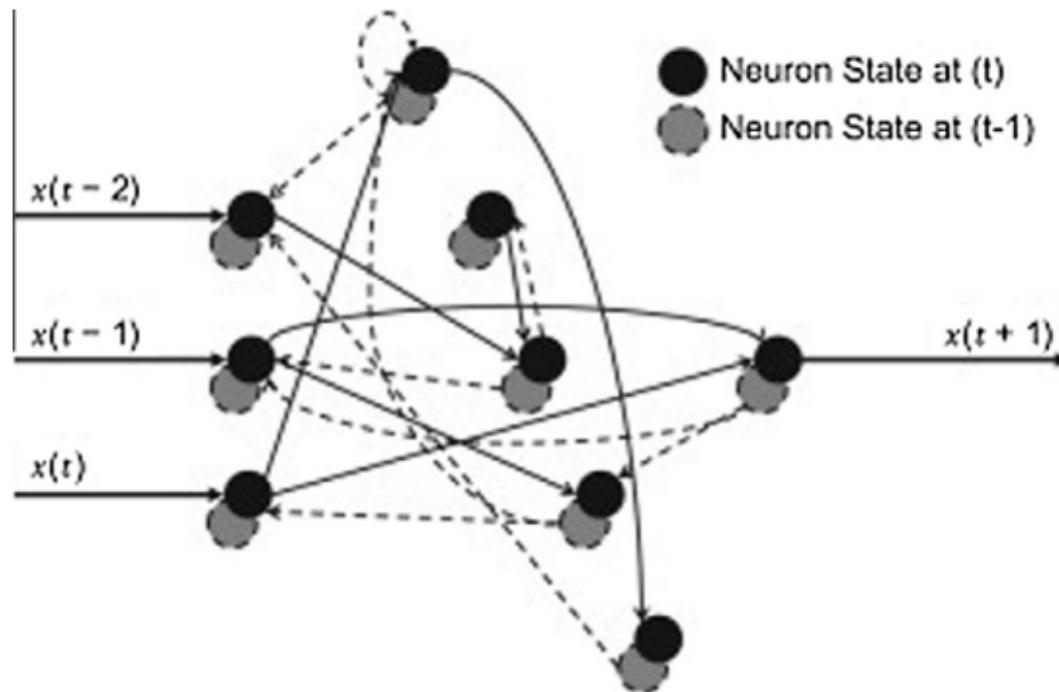


FIGURE 2.10 – L'architecture de Réseau neuronal récurrent(RNN) [6].

### 2.3.4 Modèles hybrides

Il est important de noter que la prédiction de la production d'énergie photovoltaïque peut être particulièrement difficile en raison de l'influence de facteurs tels que la météo, la saisonnalité et les caractéristiques propres à chaque site. Ainsi, l'utilisation d'un modèle unique peut ne pas être suffisante pour obtenir des résultats précis dans différents cas. C'est pourquoi les modèles hybrides, qui combinent deux ou plusieurs techniques de prédiction, ont été largement étudiés et utilisés pour améliorer les prévisions de production d'énergie photovoltaïque. Les modèles hybrides peuvent combiner des techniques de réseaux de neurones artificiels, des modèles de régression, des arbres de décision, des algorithmes génétiques et bien d'autres méthodes. En général, les modèles hybrides ont montré de meilleures performances pour la prédiction de la production d'énergie photovoltaïque que les modèles autonomes [7].

Plusieurs modèles de prédiction du rayonnement solaire existent dans la littérature et d'après notre bibliographie, la précision de prévision dépend fortement des différents facteurs : la nature du modèle utilisé, l'horizon temporel, le climat local,

ainsi que d'autres caractéristiques relatives aux données.

## 2.4 Principaux facteurs influençant les prévisions d'énergie solaire

Il existe différentes facteurs qui influencent la précision du prédiction d'énergie solaire.

### 2.4.1 L'horizon de prédiction

La précision d'un modèle de prédiction dépend de l'horizon de prédiction. Lippeheide et al [46] et d'autre chercheurs, Lonij et al [47] ont analysé les performances de l'énergie photovoltaïque sur différents horizons de prédiction et ils ont conçu des modèle de prédiction d'énergie solaire où les erreurs changent avec les horizons de prédiction. La précision d'une prédiction peut varier en fonction de l'horizon de prédiction, même si le modèle et les paramètres restent les mêmes. Il est donc important de prendre en compte l'horizon de prédiction lors de la conception d'un modèle de prédiction approprié. Bien qu'il n'existe pas de critères précis pour classier les modèles de prédiction en fonction de l'horizon de prédiction, la plupart des rapports de recherche suggèrent que la prédiction de l'énergie solaire peut être divisée en trois catégories selon l'horizon temporel (prédiction à long terme, à moyen terme et à court terme) mais Selon certains chercheurs, l'horizon de prédiction de l'énergie solaire peut être divisé en quatre catégories. La quatrième catégorie est appelée "horizon de prédiction à très court terme", qui prend en compte les prévisions pour quelques secondes seulement [6, 7].

#### 2.4.1.1 Prédiction à court terme

La prédiction à court terme est populaire sur le marché de l'électricité y compris la répartition économique de la charge et l'exploitation du système électrique. Elle est également utile dans le contrôle des systèmes de gestion des énergies renouvelables.

L'horizon de prédiction à court terme peut varier de 30 à 360 minutes, mais pour certains, il peut inclure une à plusieurs heures, un jour ou jusqu'à sept jours [6].

### 2.4.1.2 Prédiction à moyen terme

En prévoyant la disponibilité de l'énergie électrique à l'avenir, cette forme de prédiction contribue à la planification du système d'alimentation et à l'établissement du calendrier de maintenance.

Les prévisions à moyen terme ont une durée de 6 à 24 heures, selon certains rapports. Cependant, d'autres considèrent que cette catégorie peut inclure des prévisions pour un jour, une semaine ou même un mois à venir [6].

### 2.4.1.3 Prédiction à long terme

Les prévisions à long terme prédisent des scénarios pour plus de 24 heures à l'avance, selon certains rapports. Cependant, d'autres ont inclus des périodes allant d'un mois à un an dans la catégorie des prévisions à long terme. Cet horizon de prévision est adapté à la production, au transport, à la distribution et au rationnement à long terme de l'énergie solaire, ainsi qu'à la prise en compte des tendances saisonnières. Cependant, la précision de ces modèles est limitée car les fluctuations météorologiques qui s'étendent sur un ou quelques jours ne peuvent pas être prédites avec des horizons aussi longs [6, 7].

## 2.4.2 Les données d'entrées

La variation des paramètres météorologiques a un impact différent sur la production d'électricité photovoltaïque selon la situation géographique et les conditions météorologiques. La corrélation entre les entrées météorologiques telles que l'irradiation solaire, la température atmosphérique, la vitesse et la direction du vent, ainsi que l'humidité et la production d'énergie photovoltaïque peut être positive ou négative et varie selon les endroits. Pour un modèle de prévision précis, il est important d'étudier la corrélation entre les différentes entrées météorologiques et la production d'énergie photovoltaïque, et d'utiliser les variables d'entrée fortement corrélées. Les données vectorielles d'entrée faiblement corrélées doivent être refusées [5, 7].

La figure 2.3 présente le schéma de l'irradiation solaire et de la production d'énergie photovoltaïque au cours d'une journée normale. La production d'énergie photovoltaïque est fortement corrélée avec l'irradiation solaire, mais cette corrélation diminue lors de conditions météorologiques anormales.

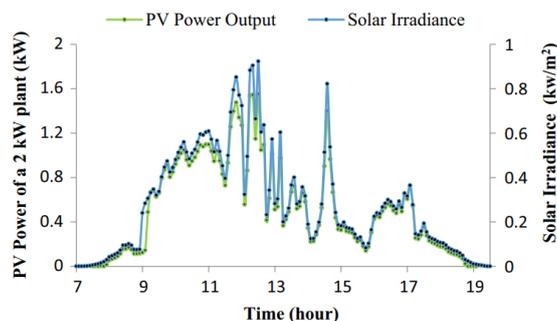


FIGURE 2.11 – Irradiation solaire et schéma de production d'énergie photovoltaïque pour un jour donné [7].

La figure 2.4 montre une forte corrélation positive entre Température atmosphérique et la puissance PV, ce qui en fait un vecteur d'entrée important pour le développement d'un modèle approprié de prévision de la puissance PV.

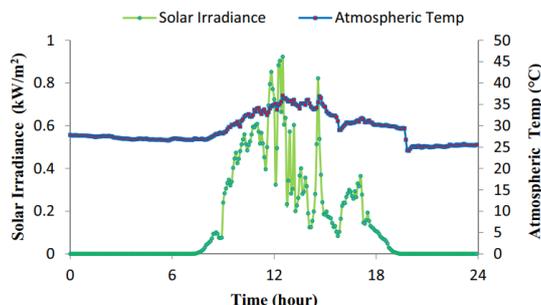


FIGURE 2.12 – Température atmosphérique et schéma de production d'énergie photovoltaïque pour un jour donné [7].

La figure 2.4 montre que la température atmosphérique a peu d'impact sur la production d'énergie photovoltaïque pendant la période de lumière du jour, mais n'a aucun impact pendant l'absence de lumière du jour. La figure 2.5 indique que la corrélation entre la température atmosphérique et la production d'énergie photovoltaïque n'est pas forte, mais reste un vecteur d'entrée potentiel pour l'établissement d'un modèle de prévision de la production d'énergie photovoltaïque.

La figure 2.6 décrit comment la puissance de sortie des cellules photovoltaïques varie en fonction de la vitesse du vent, qui est mesurée par le système de gestion des données solaires. La vitesse du vent est importante car elle aide à dissiper la chaleur produite par les cellules photovoltaïques, ce qui réduit leur température. Lorsque la température de la cellule est plus basse, la puissance de sortie des cellules

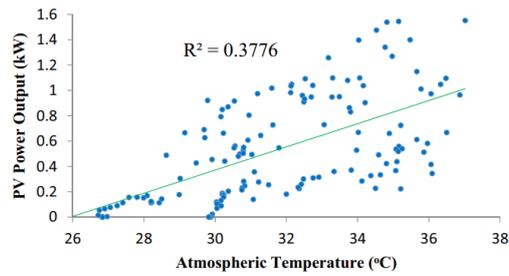


FIGURE 2.13 – Corrélation entre la température atmosphérique et la puissance photovoltaïque [7].

photovoltaïques sera réduite. La figure 8 montre que le schéma de la puissance de sortie PV ne suit pas exactement le schéma de la vitesse du vent. Pendant la journée, la puissance de sortie PV est généralement plus élevée que la vitesse du vent. Cependant, lorsque la vitesse du vent augmente, la puissance de sortie PV par unité diminue. Il n'y a pas de schéma clair de la puissance de sortie PV par rapport à la vitesse du vent observé sur toute l'année, ce qui montre une corrélation relativement faible entre la puissance de sortie PV et la vitesse du vent par rapport à la température.

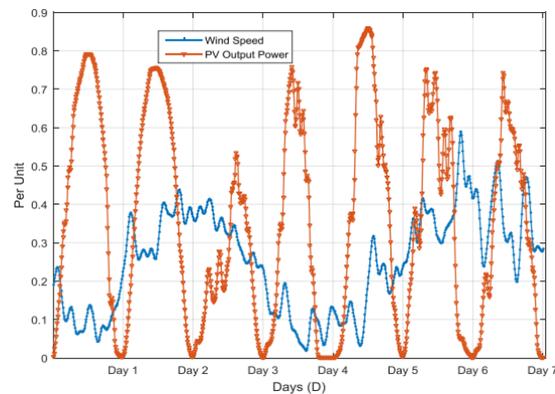


FIGURE 2.14 – Les courbes par unité de la puissance de sortie PV et de la vitesse du vent [5].

### 2.4.3 Les techniques de prévision

Plusieurs approches de modélisation : physique, statistique, intelligence artificielle (y compris l'apprentissage automatique et l'apprentissage profond), ensemble et modèles de prédiction hybrides ont été utilisées pour les prévisions d'énergie solaire (ils sont détaillé dans ce chapitre). Il est évident que la précision des prévisions d'énergie solaire varie en fonction du modèle de modélisation utilisé.

### 2.5 Conclusion

Ce chapitre a jeté les bases nécessaires pour comprendre et aborder la prédiction des séries temporelles, en fournissant des définitions clés et en explorant diverses techniques et approches utilisées dans le domaine de prediction. Comme mentionné dans ce chapitre, il existe un large éventail de méthodes et de types de méthodes disponibles pour la prédiction. Chaque méthode a ses propres caractéristiques, avantages et limitations, et peut être choisie en fonction des spécificités des données et des objectifs, puis nous nous sommes intéressés aux facteurs principaux qui influence la prédiction et plus particulièrement la prédiction d'énergie solaire.

Le chapitre suivant sera consacré aux travaux récents existants dans la littérature concernant l'optimisation et la prédiction d'énergie solaire. Ce chapitre permettra de se familiariser avec les avancées les plus récentes et les techniques émergentes utilisées dans ce domaine.

# Chapitre 3

## Etat de l'art

### 3.1 Introduction

L'utilisation de l'énergie solaire est devenue une solution incontournable pour répondre aux défis énergétiques mondiaux et réduire notre dépendance aux combustibles fossiles. Cependant, la production d'énergie solaire est fortement influencée par les conditions météorologiques et la disponibilité du rayonnement solaire. La prédiction précise de l'énergie solaire est donc essentielle pour optimiser son intégration dans le réseau électrique et assurer une gestion efficace des ressources.

Dans ce contexte, les avancées récentes dans le domaine de l'intelligence artificielle et plus précisément l'apprentissage machine offrent des opportunités prometteuses pour améliorer la prédiction de l'énergie solaire. En utilisant des modèles prédictifs basés sur des algorithmes d'apprentissage automatique, il est possible d'exploiter les données historiques et en temps réel pour estimer de manière précise la production future d'énergie solaire, cela encourage les chercheurs à encore développer des modèles de prédiction et d'optimisation d'énergie solaire.

Dans ce chapitre, nous présenterons un examen des derniers travaux réalisés sur la prédiction d'énergie solaire.

### 3.2 Travaux connexes

La prédiction d'énergie solaire est un sujet très actuel qui a suscité un nombre croissant d'articles, comme en témoigne la figure 3.1

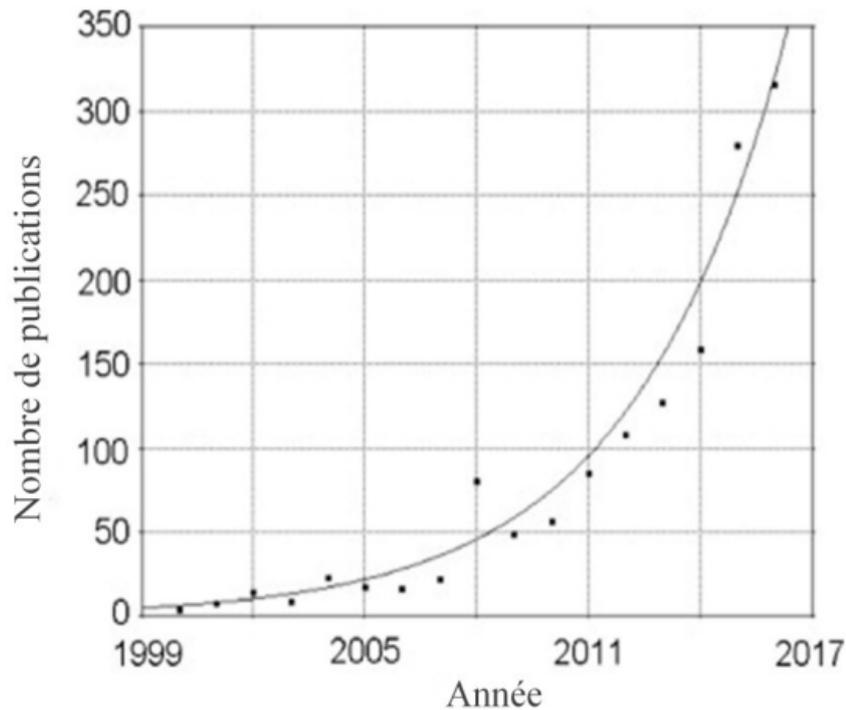


FIGURE 3.1 – Nombre d’articles publiés par an dans le domaine de la prévision des ressources solaires et de la production d’énergie solaire, retenus par SCOPUS [8].

Il existe plusieurs recherches qui se sont intéressées à la prédiction des séries temporelles de l’énergie solaire en utilisant des différentes méthodes. La littérature a recensé plusieurs approches pour aborder ce sujet.

**Kashyap et al.** [48] dans leur approche ont utilisé les données fournies par le satellite "Sunny" du ministère indien des énergies renouvelables (MNRE) pour collecter les données. Les données horaires ont été recueillies pendant 5000 heures en 2008 à partir des coordonnées 31.70691N, 76.93171E en Inde. Les paramètres utilisés comprennent l’irradiance horizontale globale (GHI), le rayonnement normal direct (DNI), l’irradiance horizontale diffuse (DHI), la température du point bulbe (BPT), la température du point de rosée (DPT), la pression (PR), la direction du vent (WD), la vitesse du vent (WS), l’heure (HH), le nombre de jours (DD) et le nombre de mois (MM).

Dans cette approche les auteurs proposent l’utilisation de huit modèles de réseaux neuronaux artificiels (ANN) différents, notamment le réseau d’anticipation (FFN), le réseau de rétropropagation à rétroaction directe (FFB), le réseau de rétropropagation en cascade (CFB), le réseau de rétropropagation d’Elman (ELM), le

réseau à base radiale (RB), le réseau de neurones probabiliste (PNN) et le réseau personnalisé (C-1 et C-2).

Les performances de chaque réseau neuronal ont été évaluées individuellement en utilisant les valeurs horaires de rayonnement solaire. L'erreur quadratique moyenne des valeurs de rayonnement solaire (RMSE) a été utilisée comme mesure pour comparer les performances des huit modèles de réseaux neuronaux. L'erreur RMSE est couramment utilisée pour évaluer la précision des prédictions par rapport aux valeurs réelles.

**Qing and Niu** [49], leur données sur l'irradiance solaire provenant d'une centrale solaire située sur l'île de Santiago au Cap-Vert. L'ensemble de données couvre une période de 30 mois, divisée en deux périodes distinctes : de mars 2011 à août 2012, puis de janvier 2013 à décembre 2013. Le Cap-Vert, situé à l'extrême sud-ouest de la région paléarctique occidentale, connaît deux saisons principales : une période venteuse d'octobre à mi-juillet, et une saison des pluies en août et septembre. Les températures au Cap-Vert varient de 21°C à 29°C, ce qui les rend relativement douces par rapport au continent africain. Avec environ 350 jours d'ensoleillement, le Cap-Vert offre un potentiel solaire important.

La collecte des données d'irradiance s'est concentrée sur les heures comprises entre 8h00 et 18h00, qui correspondent à la période d'ensoleillement et au temps solaire réel. Toutefois, en raison de valeurs manquantes causées par un dysfonctionnement du système de contrôle et d'acquisition des données (SCADA) de la centrale solaire, l'ensemble de données ne couvre que les données d'irradiance horaire sur une période de 875 jours. Les données météorologiques telles que la température, le point de rosée, l'humidité, la visibilité, la vitesse du vent et les résumés météorologiques descriptifs ont été obtenues auprès d'un service météorologique.

Dans cet article, une méthode basée sur l'utilisation d'un réseau neuronal récurrent LSTM (Long Short-Term Memory) est employée pour prédire l'énergie solaire. Parallèlement, l'algorithme de régression linéaire (LR) est également utilisé pour minimiser les différences au carré entre les valeurs réelles d'irradiance solaire horaire et les valeurs prédites par une approximation linéaire en utilisant neuf variables caractéristiques extraites du jeu de données d'entraînement. L'évaluation de la précision des prédictions se fait ensuite en utilisant un jeu de données de test et le modèle de régression linéaire appris. L'avantage de cet algorithme réside dans sa capacité à

fonctionner sans nécessiter de réglage de paramètres spécifiques.

En outre, l'algorithme de réseau neuronal à rétropropagation (BPNN), disponible dans MATLAB, est également exploité dans l'étude. Ce réseau comprend une couche d'entrée, une couche cachée et une couche de sortie. Après une série d'expérimentations pour déterminer le nombre optimal de neurones dans la couche cachée, celui-ci est fixé à 50. Les vecteurs caractéristiques horaires du jeu de données d'entraînement sont utilisés comme entrée pour la couche d'entrée, tandis que la couche de sortie fournit les valeurs prédites d'irradiance solaire horaire. La fonction de transfert "transig" est appliquée aux trois couches du réseau, et l'algorithme d'entraînement "traingd" (descente de gradient) est sélectionné. Le nombre maximum d'époques est fixé à 2500.

De plus, l'algorithme LSTM est mis en œuvre à l'aide de la bibliothèque de deep learning Keras. La couche d'entrée du LSTM est composée de neuf caractéristiques et onze pas de temps (timesteps). Le nombre de neurones cachés est fixé à 30, et la couche de sortie, dotée d'une fonction d'activation linéaire, contient un unique neurone. Le nombre maximum d'époques est fixé à 50.

Étant donné que l'échelle de l'ensemble de données était petite, l'ensemble de données de validation n'a pas été défini pour les algorithmes ANN et LSTM. Le RMSE (racine Mean Square Error) a été sélectionné comme mesure pour la comparaison des performances.

Les résultats montrent que l'algorithme d'apprentissage LSTM proposé est 18,34 l'algorithme d'apprentissage BPNN en termes de RMSE.

La méthode de collecte des données utilisée par **Sharma et al.** [50] repose sur l'utilisation de six ensembles de données de test répartis sur toute l'année. Les données horaires et quinze minutes de Singapour (1.3000 N, 103.8000 E) sur une période de douze mois en 2014 ont été utilisées. Les données sont collectées chaque heure à partir de 25 capteurs solaires répartis dans tout Singapour, puis agrégées pour calculer l'irradiance solaire totale disponible pour chaque heure. Les capteurs en silicium sont calibrés par l'Institut Fraunhofer pour les systèmes d'énergie solaire afin d'atteindre une incertitude de 5%. Les prévisions sont réalisées à partir des données observées après l'entraînement, en utilisant deux mois de données pour l'entraînement du WNN (Wireless Neural Network) et le mois suivant de données non vues pour tester les capacités de prédiction du modèle de prévision. Dans cette

étude, le jeu de données utilisé est de deux mois pour l'entraînement du WNN (le réseau de neurones utilisé dans cet article). Les sept derniers jours sont ensuite utilisés comme ensemble de validation.

Deux mesureurs de performances sont utilisés : erreur de biais moyenne (MBE) et erreur quadratique moyenne normalisée (NRMSE) pour évaluer et comparer les performances de WNN.

Pour évaluer l'efficacité de la technique proposée, le modèle développé est comparé à un modèle de persistance, ETS, ARIMA et des modèles basés sur MLP standard. Le modèle de persistance est un modèle de référence standard où la valeur prévue est donnée par la valeur actuelle de la série temporelle.

L'approche de prévision implique des prévisions d'un pas en avant, c'est-à-dire la prévision des 15 prochaines minutes ou de la prochaine heure en utilisant le réseau entraîné. La prévision est effectuée de manière itérative pour l'ensemble du mois de test, en calculant les erreurs cumulatives. Le modèle WNN fournit de véritables prévisions hors échantillon, en utilisant uniquement des données antérieures. Le nombre d'entrées requis pour une prévision optimale varie selon les saisons. Les données historiques des heures précédentes sont utilisées pour prédire la valeur d'irradiance pour l'heure suivante.

Les performances du modèle WNN sont testées pour différentes semaines représentant différentes saisons, avec des dynamiques et une couverture nuageuse variables. Les prévisions sont généralement plus précises les jours sans nuages par rapport aux jours nuageux.

À partir des résultats de cet article, on peut en déduire qualitativement que, en présence d'une variabilité à court terme relativement faible, WNN offre de meilleures performances que les trois autres méthodes.

**Gutierrez-Corea et al.** [51] leur étude a été réalisée dans la région de Castille et León (CYL), qui couvre une superficie de 94 226 km<sup>2</sup> et se situe dans la partie nord du plateau de la péninsule ibérique. Cette région présente une altitude moyenne de 800 m et est assez plate, à l'exception de ses limites géographiques. Les données utilisées ont été fournies par les stations météorologiques automatiques officielles du gouvernement régional de Castille et León (AWS CYL) via l'Institut de technologie agricole de Castille et León (ITACYL, 2014). L'AWS-CYL est composé de 50 stations météorologiques homogènes, avec une densité moyenne de 11

884 km<sup>2</sup> par station et une distance moyenne de 25,94 km entre les stations voisines (Gutiérrez-Corea et al., 2014). L'AWS-CYL propose librement des observations de ses stations météorologiques sur Internet avec une résolution temporelle de valeurs moyennes sur 30 minutes. Chacune des stations météorologiques de l'AWS-CYL dispose d'une référence horaire correspondant à l'heure locale moyenne. Afin de faire coïncider toutes les observations à la même heure, deux processus ont été effectués. Premièrement, toutes les observations ont été converties en heure UTC en décalant l'heure d'origine de quelques minutes. Deuxièmement, un processus d'interpolation temporelle linéaire a été effectué pour convertir les valeurs en multiples de 15 minutes (0, 15, 30 et 45 minutes). Les données expérimentales utilisées sont des observations effectuées par l'AWS-CYL au cours de l'année 2011. L'apprentissage a été réalisé uniquement pour les valeurs diurnes, en tenant compte du fait que l'ANN apprend facilement que lorsque le soleil est en dessous de l'horizon, l'indice de ciel clair (GSI) est égal à zéro. Les valeurs nocturnes ont également été utilisées, mais uniquement en tant qu'entrées lors de la prédiction des premières heures de la journée.

Dans cette étude, une architecture de réseau de neurones artificiels MLP (Perceptron Multi-Couches) a été utilisée pour expérimenter la prédiction à court terme de l'indice de ciel clair (GSI). Le MLP est l'architecture de réseau de neurones la plus populaire utilisée pour résoudre des problèmes scientifiques et est donc largement utilisée dans les études liées à la prédiction du GSI.

Les auteurs ont étudiés 6 ANN, ils sont utilisés pour les prédictions dans des délais de 15 minutes. Les ANN (1- 3) le font en incluant les stations voisines, tandis que les ANN (4-6) font la prédiction en se basant uniquement sur les observations précédentes de la station cible elle-même. 15 ANN sont utilisés pour prédire le GSI pour une période d'une heure. Dans ce tableau, les premiers ANN (7-14) utilisent les données des stations voisines, tandis que les autres (15-21) utilisent uniquement les observations précédentes de la station cible elle-même. D'autres ANN sont utilisés aussi pour prédire les périodes plus d'une heure. Les données d'entraînement et de test ont été réparties respectivement en 70 % et 30 %. Pour les mesures de performances RMSE est utilisé.

Les ANN qui prédisent le GSI sur des périodes de 1 à 3 heures et intègrent les données des stations voisines dans leurs entrées affichent des taux d'erreur plus faibles, contrairement aux ANN qui utilisent uniquement les données de la station cible de prédiction elle-même. Pour les périodes de 4 à 6 heures, les résultats sont

inversés, le deuxième type d'ANN présentant des taux d'erreur inférieurs par rapport aux valeurs obtenues par le premier type d'ANN. Cela a permis de trouver une relation entre les tranches horaires (1-3 heures) et la distance de référence (55 km), dans laquelle les prédictions sont meilleures lors de l'utilisation des stations voisines. Cela signifie que l'effet des conditions météorologiques enregistrées par les stations voisines situées à la distance de référence n'est valable que pour les tranches horaires de 1 à 3 heures. Cela nous pousse à croire qu'il pourrait être possible d'améliorer les prédictions pour les tranches horaires de 4 à 6 heures en utilisant des ANN avec des stations voisines situées au-delà de 55 km.

Cinq emplacements ont été sélectionnés par **pedro-2015** [52] pour étudier l'effet des conditions météorologiques locales sur les performances de prévision. Merced et Davis, situés dans la vallée centrale de Californie, ont une faible variabilité solaire en raison de leurs étés secs et sans nuages. San Diego et Bellingham, proches de l'océan Pacifique, ont une irradiance plus variable en raison de phénomènes météorologiques tels que la couche marine. Ewa Beach à Hawaï présente une irradiance élevée en raison de formations météorologiques locales. Ces emplacements représentent une large gamme de microclimats et de topologies du sol.

Après avoir conclu l'analyse des données ci-dessus, nous avons exploré des techniques pour améliorer la prévision de l'irradiance solaire globale (GHI) pour des horizons de prévision courts, allant de 15 à 120 minutes. Les données des cinq emplacements ont été divisées en trois sous-ensembles distincts : l'ensemble d'entraînement, de validation et de test. L'ensemble d'entraînement a été utilisé pour créer une base de données de caractéristiques pour les méthodes kNN et ANN, l'ensemble de validation a été utilisé pour déterminer l'ensemble optimal de paramètres libres pour les différents modèles de prévision, et l'ensemble de test a été utilisé comme jeu de données indépendant pour évaluer la précision des prévisions. Afin de garantir que les trois ensembles de données incluent des données provenant de périodes d'irradiance similaires, nous avons sélectionné la première semaine de chaque mois comme données de test, la deuxième semaine comme données de validation et le reste comme données d'entraînement (ou historiques), ce qui a abouti à une répartition approximative de 25 %, 25 % et 50 % des données.

Les prédictions pour les différents emplacements et horizons de prévision (15, 30, 45, 60, 90, 120 minutes) ont été obtenues à l'aide de quatre modèles : la persistance, la

persistance optimisée, le kNN optimisé et l'ANN optimisée. Ces modèles n'utilisent aucune variable exogène.

Différentes mesures d'erreur ont été utilisées pour évaluer la performance des différents modèles : l'erreur quadratique moyenne (RMSE), l'erreur absolue moyenne (MAE), l'erreur de biais moyen (MBE) et la compétence de prévision, qui mesure l'amélioration en pourcentage du RMSE.

En termes de MAE, les modèles optimisés kNN et ANN présentent les meilleures performances, à l'exception des prévisions de 15, 45, 60 et 90 minutes pour San Diego, où la MAE obtenue par la persistance optimisée est légèrement plus petite. En ce qui concerne le RMSE, la variable pour laquelle les modèles ont été optimisés, les modèles kNN et ANN montrent les valeurs les plus petites, indépendamment de l'emplacement et de l'horizon de prévision. Par conséquent, ils obtiennent toujours la meilleure compétence de prévision.

En comparant uniquement les résultats des modèles kNN et ANN, on observe que les performances de ces modèles sont comparables. En général, il semble y avoir une tendance suggérant que le modèle ANN donne de meilleurs résultats que le modèle kNN pour les horizons de prévision plus courts ( $\leq 60$  min) et moins bons pour les horizons plus longs.

**Ramsamo et al.** [53] ont étudiés un système photovoltaïque qui constitué de seize panneaux solaires Suntech STP250S d'une puissance de 250 W chacun, ainsi que d'un onduleur Power One PV13.6 OUTD. Il est installé sur le toit d'un bâtiment à Newquay, en Cornouailles, au Royaume-Uni. Les prévisions photovoltaïques sont basées sur des données météorologiques et la production d'énergie du système. Un modèle de prédiction a été développé en utilisant plusieurs variables météorologiques, telles que la température de l'air, la pression atmosphérique, l'humidité, les précipitations, l'irradiance solaire, la direction et la vitesse du vent. Les performances du modèle ont été évaluées en utilisant un ensemble de données historiques, divisé en deux parties : une pour l'entraînement et une pour l'évaluation. Les résultats ont montré que le modèle de prédiction basé sur l'intégration de ces variables météorologiques améliore les performances de prévision par rapport au modèle de persistance.

Dans cette étude trois types de modèles sont utilisés, à savoir Modèle de régression linéaire multiple (MLR), Modèles de réseaux de neurones artificiels et Modèles

Hybrides.

Les modèles de prévision étudiés affichent des performances satisfaisantes, avec des indicateurs tels que le RMSE, le MBE, le MAE et le coefficient de corrélation  $R$  se situant dans des plages acceptables.

Les modèles basés sur les réseaux de neurones artificiels (ANN) surpassent ceux basés sur les moindres carrés ordinaires (MLR), en prenant en compte à la fois les relations linéaires et non linéaires entre les facteurs météorologiques et la production d'énergie photovoltaïque. Parmi les modèles ANN, le FFNN se distingue en termes de performances, surpassant le GRNN tant pour les modèles à un seul stade que pour les modèles hybrides. Les modèles hybrides présentent des résultats comparables à ceux des modèles à un seul stade, et le SR-FFNN est identifié comme le meilleur modèle en termes de dispersion moyenne et de biais global des erreurs de prédiction. Ces résultats sont confirmés par la comparaison des valeurs prédites de la production mensuelle d'énergie avec les mesures réelles, où les modèles hybrides montrent une meilleure capacité de prévision à long terme que les modèles à un seul stade. En conclusion, le modèle SR-FFNN est identifié comme le meilleur modèle de prédiction, et les modèles hybrides surpassent les modèles à un seul stade en termes de précision de prévision.

**Zhang et al.** [54] ont utilisé Les données de mai 2015 jusqu'à la fin de février 2016 comme ensemble de données d'entraînement et de validation, et les données de mars 2016 comme ensemble de données de test. Après plusieurs expériences, nous avons sélectionné 17 caractéristiques météorologiques parmi les 42 disponibles comme ensemble final de caractéristiques utilisées dans cet article. Par conséquent, un exemple de données est un tenseur 4D de taille  $6 \times 17 \times 258 \times 172$ , représentant une heure de 17 caractéristiques météorologiques pour une grille de  $258 \times 172$ . La répartition des données d'entraînement est 19.935, de validation est 2.215 et de test est 2700.

Les modèles de référence utilisés dans cette étude comprennent le modèle persistant, le modèle SVR et le modèle SVR basé sur la similarité. Dans cette étude, trois schémas d'apprentissage profond ont été proposés pour analyser l'évolution temporelle des données météorologiques et leur impact sur la production d'énergie solaire. Différents modèles de connectivité temporelle ont été examinés, notamment le modèle d'une seule heure, le modèle journalier et le modèle hybride. Le modèle hybride

combine des réseaux de convolution et des réseaux neuronaux récurrents (LSTM) pour prédire la production d'énergie solaire à intervalles horaires. De plus, des caractéristiques supplémentaires ont été ajoutées, démontrant leur efficacité dans un modèle basé sur SVR.

Les résultats montrent que les modèles d'apprentissage profond surpassent les modèles de référence SVR conventionnels. L'ajout de caractéristiques de similarité s'est avéré efficace pour les modèles SVR, mais moins évident pour les modèles d'apprentissage profond. Les modèles d'une seule heure basés sur AlexNet ont obtenu les meilleurs résultats, avec un taux d'erreur sans précédent de 11,8%. L'extension de la dimension temporelle à un modèle journalier n'a pas amélioré les performances, indiquant que l'expansion d'une heure est suffisante pour cette tâche. Les tentatives d'utiliser des structures plus profondes, comme ResNet, n'ont pas apporté d'améliorations significatives. Il est également souligné que l'apprentissage profond permet d'obtenir de bien meilleurs résultats sans nécessiter de connaissances spécifiques du domaine, contrairement aux approches basées sur des modèles physiques sophistiqués.

**Alzahrani, et al.** [55] présentent des méthodes basées sur les réseaux neuronaux profonds pour prédire l'irradiance solaire à court terme. Les données utilisées proviennent d'une ferme solaire canadienne. Les étapes de prétraitement comprennent le nettoyage des données en supprimant les valeurs aberrantes et en normalisant les données. L'algorithme de réseau neuronal récurrent profond (DRNN) est utilisé, avec des unités à mémoire à court et long terme (LSTM), deux couches cachées et 35 neurones cachés. Le modèle est implémenté en utilisant l'API Keras et MATLAB. Un autre modèle de référence basé sur un réseau neuronal à propagation arrière (FNN) est également mis en place. Les données sont divisées en ensembles d'entraînement, de test et de validation, avec des proportions de 70%, 15% et 15% respectivement. Après l'entraînement, les données sont renormalisées à l'aide de la fonction `mapminmax`.

L'évaluation des performances se fait en utilisant plusieurs mesures telles que l'erreur principale de biais (MBE), l'erreur quadratique moyenne (MSE) et l'erreur quadratique moyenne racine (RMSE). Les résultats montrent que le modèle DRNN offre les meilleures performances, avec des valeurs de RMSE les plus basses. Une comparaison avec d'autres méthodes telles que le FNN et la machine à vecteurs de

support (SVM) est effectuée, et le modèle DRNN se révèle supérieur aux autres.

En conclusion, l'article démontre que Le modèle de prédiction utilisant un réseau neuronal récurrent profond avec des couches LSTM a obtenu de bons résultats. L'erreur quadratique moyenne (RMSE) moyenne pour l'entraînement est de 0.0513 et de 0.068 pour les tests. Pour l'entraînement, l'RMSE moyen est de 0.0746, et de 0.086 pour les tests. Ces résultats sont prometteurs, car l'RMSE est le plus bas. Comparé à d'autres méthodes telles qu'un réseau neuronal à propagation avant (FNN) et une machine à vecteurs de support (SVM), le modèle LSTM de deep learning offre de meilleures performances avec un RMSE de 0.086.

**Benali et al.** [56] ont mené une étude comparative pour prédire l'irradiance horizontale globale et ses deux composantes, directe et diffuse, pour des horizons temporels allant d'une heure à six heures à l'avance sur le site d'Odeillo en France. Un modèle de persistance intelligent, un modèle ARN et un modèle FA ont été utilisés comme prédicteurs. L'étude a montré que le modèle FA surpassait le modèle ARN à mesure que l'horizon de prédiction augmentait. Les résultats ont également révélé que la précision des prévisions au printemps et en automne est inférieure à celle en hiver et en été.

**Dong et al.** [57] présente une méthode pour prédire l'irradiance solaire à court terme en utilisant un modèle de lissage exponentiel d'espace d'état. Les résultats montrent que cette méthode est précise et pourrait aider à optimiser la production d'énergie solaire en prévoyant les conditions météorologiques à court terme.

**Dazhi Yang et al.** [58] ont utilisé la méthode de lissage exponentiel pour prédire l'irradiance horizontale globale en utilisant des décompositions. Cette méthode s'est avérée précise pour prédire l'irradiance future et pourrait être utilisée pour optimiser la production d'énergie solaire.

**Gairaa et al.** [59] ont proposé une approche qui combine les modèles ARN et ARMA pour prédire le rayonnement global quotidien dans deux zones climatiques différentes en Algérie (Alger et Ghardaïa). Le modèle combiné a montré une amélioration de l'erreur absolue moyenne par rapport aux modèles ARMA et ARN séparément, d'environ 18,1 % et 2,7% pour le site d'Alger et d'environ 27,26% et 1,39% pour le site de Ghardaïa.

**Wu et al.** [60] ont développé une approche qui combine l'algorithme génétique avec l'algorithme de regroupement k-means pour trouver le segment optimal de la série temporelle du rayonnement solaire à Singapour. La séquence segmentée optimale est ensuite utilisée comme entrée pour un modèle RNA récurrent, afin de générer des prédictions de 5 minutes d'irradiation globale horizontale. Le modèle développé a présenté les meilleures performances par rapport aux modèles ARMA et au modèle RNA récurrent.

**Marzouq et al.** [61] ont mis au point un modèle d'apprentissage évolutif de réseau neuronal (RNA) pour prédire l'irradiation solaire horaire jusqu'à six heures à l'avance en utilisant une base de données de 28 villes appartenant à différentes zones climatiques au Maroc. L'approche évolutionnaire optimise l'architecture et les paramètres du RNA représentés par un chromosome développé par un ensemble d'opérateurs d'algorithme génétique, incluant la sélection, le croisement et la mutation. Trois scénarios ont été étudiés, les deux premiers utilisant les mêmes sites dans l'ensemble de données pour l'apprentissage et le test, tandis que le troisième prédit l'irradiation solaire pour des sites différents de ceux de l'ensemble de données d'apprentissage. Les résultats ont montré que le modèle développé surpasse les modèles de persistance, AR et FA. L'analyse saisonnière du modèle a montré son efficacité pour prédire l'irradiation solaire dans différentes conditions météorologiques, avec de meilleurs résultats obtenus pendant la saison estivale.

### 3.3 Analyse et comparaison

Dans le tableau ci-dessous nous effectuerons une étude comparative des approches proposées ci-dessus selon les 3 facteurs suivants :

- **Méthode** : désigne la méthode utilisée par les auteurs dans leur études.
- **Horizon temporel** : L'horizon utilisé pour la prédiction.
- **Métrique d'évaluation** : L'indicateur statistique utilisés dans l'étude pour évaluer la qualité et l'efficacité de modèles.

Approche	Méthode	Horizon temporel	Métrique d'évaluation
Huang et al. (2010)	RNA et statistique model	Horaire	RMSE :10%/ et 13 %
Voyant et al. (2013)	RNA	Quotidien	nRMSE :23.4%
Mazorra-Aguiar et al. (2015)	RNA	Horaire	nRMSE : 15.30%
Pedro and Coimbra (2015)	RNA et KNN	15 min jusqu'à 2 h	rMSE<15%
Alzahrani, et al. (2017)	RNN	Horaire	RMSE 8.6%
Qing and Niu (2018)	LSTM	Horaire	RMSE : 18.34%
Fouilloy et al. (2018)	ARIMA	Horaire	nRMSE :18.35%
Fouilloy et al. (2018)	ARMA	Horaire	nRMSE : 17.71%
Fouilloy et al. (2018)	AR	Horaire	nRMSE : 18.72%
Zhang, et al. (2018)	CNN	Horaire	rMAPE 11.8%
Zhang, et al. (2018)	LSTM	Horaire par minute	RMSE< 21%
Takilate et al. (2019)	SVM	Quotidien	nRMSE :20.42%

TABLE 3.1 – Résumé de quelques travaux récents sur la prévision d'énergie solaire.

### 3.4 Conclusion

En conclusion, ce chapitre a examiné différents travaux et méthodes de prédiction de l'énergie solaire en se basant sur des données provenant de différentes régions géographiques. Les résultats ont montré que l'utilisation de techniques basées sur des algorithmes d'apprentissage automatique tels que les réseaux de neurones artificiels (ANN) et les réseaux neuronaux récurrents LSTM (Long Short-Term Memory) permet d'obtenir des prévisions plus précises de l'énergie solaire. Ces méthodes offrent des prévisions plus précises, ce qui peut être utile pour optimiser l'utilisation de l'énergie solaire dans différentes applications.

Cependant, il convient de noter que les performances de prédiction peuvent varier en fonction de la topologie et les paramètres d'un modèle et des caractéristiques spécifiques de chaque région géographique qui consiste un défi pour un problème

spécifique. Des recherches supplémentaires sont nécessaires pour affiner les modèles et les techniques de prédiction afin d'améliorer encore davantage la précision des prévisions d'irradiance solaire.

Le chapitre suivant sera consacré à l'élaboration des modèles de prédiction d'irradiation solaire à Béjaia ainsi les différents étape applique durant notre travaille.

# Chapitre 4

## Approches proposées

### 4.1 Introduction

Ce chapitre se concentre sur la méthodologie utilisée pour prédire l'irradiation solaire à la ville de Bejaia en Algérie en utilisant des différentes méthodes sur deux horizons H+1 et J+1. Il met en évidence les différentes étapes suivies, y compris la collecte et l'analyse des données, la corrélation entre les variables, la normalisation des données, la sélection du modèle, ainsi que l'entraînement, le test et la validation des données. Les résultats obtenus pour différentes méthodes de prédiction sont également présentés et discutés.

### 4.2 Environnement de développement

- **Anaconda** Anaconda a été fondée en 2012 par Peter Wang et Travis Oliphant dans le but de rendre Python et R appliqués accessibles à l'analyse de données en entreprise, qui connaissait une transformation rapide en raison des tendances technologiques émergentes. De plus, la communauté open-source ne disposait pas d'une entité capable de l'organiser et de la rassembler pour maximiser son impact. Depuis lors, l'écosystème Python s'est considérablement développé, Python étant aujourd'hui le langage de programmation le plus populaire. Parallèlement à cette expansion, Anaconda a apporté une valeur ajoutée aux étudiants qui apprennent Python et la science des données, aux praticiens individuels, aux petites équipes et aux entreprises de grande envergure. leur objectif est de répondre aux besoins de chaque utilisateur

à chaque étape de son parcours en science des données. Anaconda compte désormais plus de 300 employés à temps plein répartis aux États-Unis, au Canada, en Allemagne, au Royaume-Uni, en Australie [62].

- **Jupyter notebook** Jupyter notebook est un environnement de développement interactif basé sur le web pour les notebooks, le code et les données. Son interface flexible permet aux utilisateurs de configurer et d'organiser des flux de travail dans les domaines de la science des données, du calcul scientifique, du journalisme computationnel et de l'apprentissage automatique. Sa conception modulaire permet d'ajouter des extensions pour étendre et enrichir ses fonctionnalités. Le Jupyter Notebook est l'application web d'origine pour la création et le partage de documents computationnels. Il offre une expérience simple, rationalisée et centrée sur le document [63].

### 4.2.1 Langage de programmation

#### 4.2.1.1 Python

Python est un langage de programmation interprété, orienté objet et de haut niveau, réputé pour sa puissance et sa rapidité. Il est largement utilisé par les informaticiens et les développeurs dans divers domaines tels que la gestion de l'infrastructure, l'analyse de données, le développement de logiciels et l'intelligence artificielle. La popularité de Python provient de sa capacité à permettre aux développeurs de se concentrer sur ce qu'ils font plutôt que sur la manière de le faire. Il les libère des contraintes des langages plus anciens et favorise le développement rapide de code.

L'attrait de Python réside également dans sa facilité d'accès, ce qui en fait un choix adapté aux débutants prêts à investir du temps dans son apprentissage.

Python est couramment utilisé pour diverses tâches, telles que la programmation d'applications, le développement de services web, science de données, la génération de code et la métaprogrammation. Son écosystème étendu, notamment l'index des packages Python (PyPI), offre une vaste collection de modules et de bibliothèques tiers qui contribuent à sa polyvalence et ouvrent des possibilités infinies.

Python est développé sous une licence open source gérée par la Python Software Foundation, ce qui permet son utilisation et sa distribution gratuites, même à des fins commerciales. La communauté Python organise des conférences, des rencontres

et des initiatives de codage collaboratif, favorisant un environnement dynamique et solidaire pour le développement et l'avancement des technologies open source liées au langage de programmation Python [64, 65].

### 4.2.2 Bibliothèques de Python

#### 4.2.2.1 Numpy

NumPy est le package fondamental pour le calcul scientifique en Python. C'est une bibliothèque Python qui fournit un objet de tableau multidimensionnel, divers objets dérivés (tels que les tableaux masqués et les matrices) ainsi qu'une collection de fonctions pour des opérations rapides sur les tableaux. Ces opérations incluent des calculs mathématiques, des opérations logiques, des manipulations de forme, des tris, des sélections, des opérations d'entrée/sortie, des transformations de Fourier discrètes, de l'algèbre linéaire de base, des opérations statistiques de base, des simulations aléatoires, et bien plus encore. NumPy est essentiel pour les scientifiques et les programmeurs Python, car il offre une base solide pour les calculs numériques et la manipulation de données [66].

#### 4.2.2.2 Pandas

Pandas est une bibliothèque Python open source largement utilisée dans les domaines de la science des données, de l'analyse de données et de l'apprentissage automatique. Elle repose sur la bibliothèque Numpy, qui offre une prise en charge des tableaux multidimensionnels. En tant que l'une des bibliothèques les plus populaires pour la manipulation de données, Pandas s'intègre parfaitement à de nombreux autres modules de science des données de l'écosystème Python. On la retrouve généralement dans toutes les distributions Python, qu'elles soient fournies avec le système d'exploitation ou proposées par des fournisseurs commerciaux tels qu'ActiveState's ActivePython.

Avec Pandas, vous pouvez accomplir de nombreuses tâches fastidieuses et répétitives liées à la manipulation des données, telles que le nettoyage des données, le remplissage des données, la normalisation des données, les fusions et les jointures, la visualisation des données, l'analyse statistique, l'inspection des données, le chargement et l'enregistrement des données, et bien d'autres encore.

En fait, grâce à Pandas, vous pouvez effectuer toutes les opérations qui font de

cette bibliothèque l'outil privilégié des data scientists du monde entier pour l'analyse et la manipulation de données [67].

### 4.2.3 TensorFlow

TensorFlow est une bibliothèque logicielle qui permet aux ingénieurs logiciels et aux data scientists de construire, entraîner et travailler avec des modèles d'apprentissage profond pour effectuer des prédictions à partir de données.

TensorFlow est utilisé pour simplifier le développement et l'entraînement de modèles d'apprentissage automatique. Il est principalement utilisé pour la classification, la perception, la compréhension, la découverte, la prédiction et la création, avec des applications concrètes dans le monde réel.

TensorFlow a été créé par Google Brain en 2011 pour offrir aux data scientists et aux ingénieurs un logiciel puissant permettant de simplifier les tâches ardues liées à l'entraînement des modèles d'apprentissage automatique, telles que le développement et l'exécution d'applications d'analyse avancée. La bibliothèque utilise des graphes de flux de données pour construire des modèles et créer des réseaux neuronaux multicouches à grande échelle. Les capacités de classification, de perception, de compréhension, de découverte, de prédiction et de création de TensorFlow facilitent l'utilisation de technologies concrètes telles que la reconnaissance d'images et de la voix, ainsi que les algorithmes de séries temporelles [68].

#### 4.2.3.1 Matplotlib

Matplotlib est une bibliothèque Python open source qui permet la création de graphiques et de diagrammes de haute qualité. Elle a été initialement développée en 2002 par le neurobiologiste John Hunter dans le but de visualiser les signaux électriques du cerveau des patients épileptiques. Son objectif était de reproduire les fonctionnalités de création graphique de MATLAB en utilisant Python.

Depuis le décès de John Hunter en 2012, Matplotlib a connu de nombreuses améliorations grâce à la contribution de nombreux membres de la communauté open source. Cette bibliothèque est largement utilisée pour générer des graphiques et des diagrammes de grande qualité et constitue une alternative open source à MATLAB. Avec Matplotlib, il est possible de créer différents types de graphiques, tels que des tracés, des histogrammes, des diagrammes à barres, et bien d'autres, en utilisant

seulement quelques lignes de code. Elle est particulièrement utile pour les personnes travaillant avec Python ou NumPy.

Matplotlib est utilisée dans divers environnements, tels que les serveurs d'application web, les shells et les scripts Python. Les développeurs peuvent également intégrer des graphiques dans des applications dotées d'interfaces graphiques en utilisant les API de Matplotlib. Les principaux concepts de Matplotlib reposent sur la notion de "figure", qui représente une illustration complète, et sur les "axes", qui correspondent à chaque tracé au sein de cette figure. La création d'un graphique, appelée "plotting", nécessite l'utilisation de données sous forme de paires clé/valeur pour les axes X et Y. Des fonctions telles que "scatter", "bar" et "pie" sont utilisées pour créer différents types de schémas.

Matplotlib permet de réaliser des graphiques de base, tels que des diagrammes à barres ou des histogrammes, ainsi que des figures plus complexes en trois dimensions. Elle offre un outil puissant pour la visualisation et l'analyse des données, et permet de créer des représentations visuelles détaillées [69].

### 4.2.3.2 Sklearn

Scikit-Learn, également connu sous le nom de sklearn, est une bibliothèque de machine learning gratuite pour Python. Elle offre un large éventail d'algorithmes pour la classification, la régression, le regroupement et la réduction de la dimensionnalité, et prend en charge à la fois l'apprentissage supervisé et non supervisé. Scikit-Learn est construit en utilisant des bibliothèques populaires telles que NumPy et SciPy, et il est compatible avec d'autres bibliothèques couramment utilisées comme Pandas et Seaborn. Cette bibliothèque offre aux développeurs Python un ensemble d'outils puissants et flexibles pour l'exploration et l'analyse des données, ainsi que pour la création de modèles de machine learning [70].

### 4.2.3.3 Keras

Keras est une bibliothèque Python minimaliste pour le deep learning qui peut s'exécuter sur Theano ou TensorFlow. Elle a été développée dans le but de faciliter et d'accélérer la mise en œuvre de modèles de deep learning, que ce soit pour la recherche ou le développement. Keras fonctionne avec Python 2.7 ou 3.5 et peut s'exécuter de manière transparente sur des GPU et des CPU en utilisant les fra-

metworks sous-jacents. Elle offre une interface conviviale pour créer et entraîner des réseaux de neurones profonds, permettant aux utilisateurs de se concentrer sur leur tâche sans se soucier des détails techniques complexes [71].

### 4.3 Méthodologie de travail

Notre travaille est résumé dans la figure 4.1.

#### 4.3.1 Collecte des données

L'objectif de cette étape consiste à collecter une quantité adéquate de données afin de constituer une base de données représentative. Cette base de données sera utilisée pour l'apprentissage et le test de différentes méthodes basées sur l'apprentissage machine.

Les paramètres météorologiques de Béjaïa (Latitude 36.7557, Longitude 5.0844), y compris la température, l'humidité relative, la pression, la vitesse du vent et l'irradiation solaire, ont été utilisés dans notre étude. Ces données ont été obtenues à partir de la base de données de la NASA [72], qui fournit des données météorologiques mensuelles sur une période de 10 ans, avec une résolution horaire et journalier, allant du 1er janvier 2010 au 1er janvier 2021.

Il y a exactement 87696 collectes de données soit une fréquence de collecte à chaque une heure et 3654 chaque jour Voici les caractéristiques contenues par notre base :

- **Date** : ça représente la date dans laquelle on a effectué notre collecte, elle est sous la forme mm/jj/aaaa.
- **hour** : ça représente l'heure de la date dans laquelle on a effectué notre collecte.
- **ghirradiation** : c'est l'irradiation solaire, elle est mesurée en Wh/m<sup>2</sup>/hr.
- **vitessevent** : c'est la vitesse du vent, mesurée metre par heure (m/s).
- **temperature** : elle est mesurée en (C).

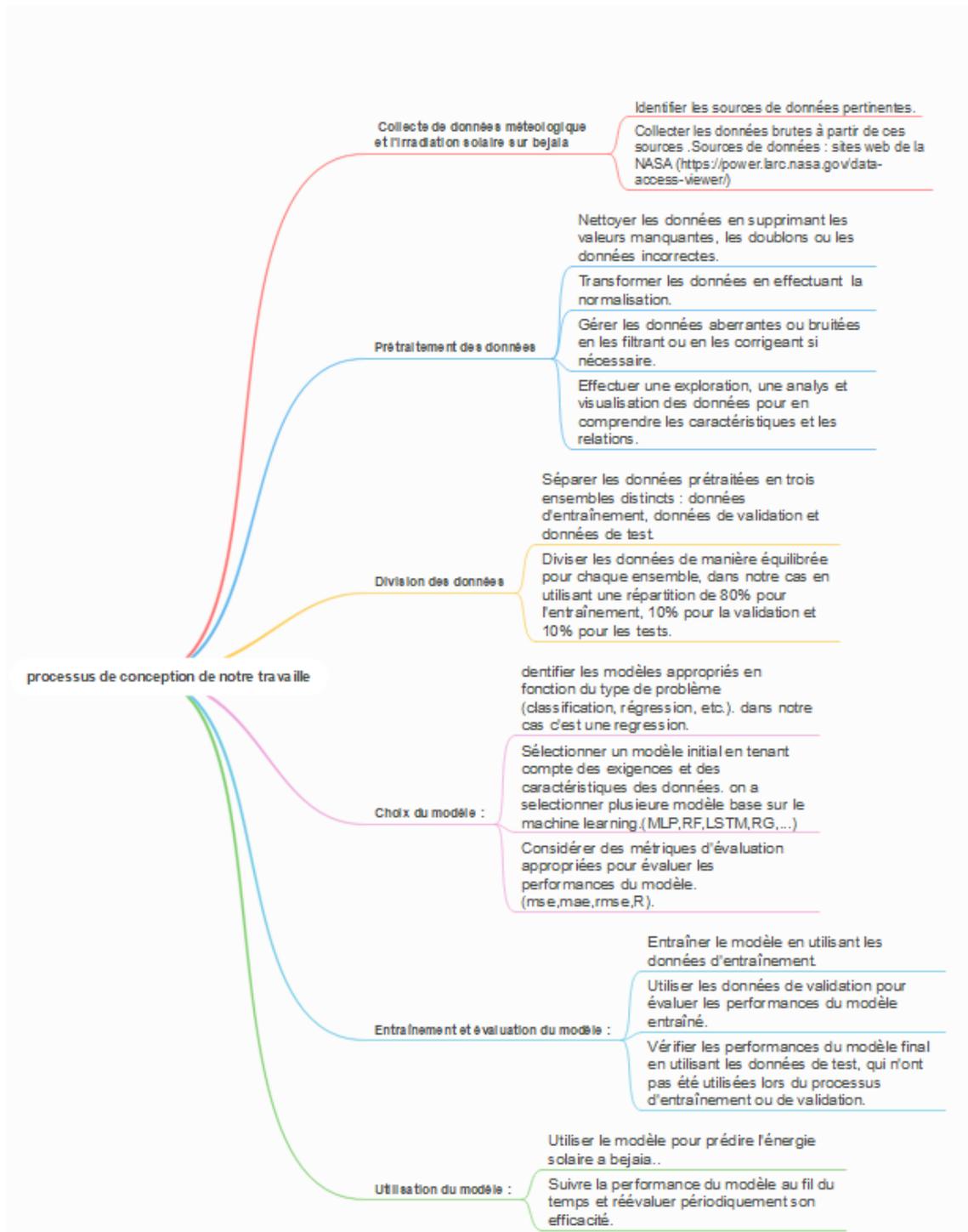


FIGURE 4.1 – Processus de notre approche

- **humidité** : c'est la présence d'eau ou de vapeur d'eau dans l'air ou dans une substance, elle est mesurée en pourcentage (%).
- **ps** : c'est la pression qu'exerce le mélange gazeux constituant l'atmosphère considérée (sur Terre : de l'air) sur une surface quelconque en contact avec elle, elle est mesurée en kPa.

### 4.3.2 Prétraitement des données

#### 4.3.2.1 Exploration des données et identification de relation

Dans cet étape nous utilisons la visualisation des données et la corrélation. Il convient de noter que la visualisation et la corrélation ne sont pas des étapes strictes du prétraitement des données, mais plutôt des techniques et des outils utilisés pendant le processus de prétraitement pour mieux comprendre les données. Elles aident à identifier les problèmes de qualité des données, les tendances, les schémas et les relations.

##### 1. Évolutions et visualisation des données

###### — L'irradiation solaire

La figure (4.2) présente les variations de l'irradiation solaire à Bejaïa au cours de l'année. Les mois de décembre et janvier montrent une moyenne mensuelle d'environ  $100 \text{ Wh/m}^2/\text{h}$ , tandis que les mois de juillet et août affichent une moyenne d'environ  $320 \text{ Wh/m}^2/\text{h}$ . La figure 4.3 illustre la variation de l'irradiation solaire tout au long de la journée, commençant à environ  $30 \text{ Wh/m}^2/\text{h}$  pendant 5 heures, puis atteignant un pic d'environ  $615 \text{ Wh/m}^2/\text{h}$  vers 11/12 heures du matin, avant de diminuer progressivement. Ainsi, l'énergie solaire est disponible toute l'année à Bejaïa, mais les quantités dépendent de la saison (figure 4.4), du mois et du cycle jour/nuit. Sur une période de 10 ans (2010-2020), la moyenne annuelle du rayonnement global est de  $197.39 \text{ Wh/m}^2/\text{h}$ .

— **La température** Entre 2010 et 2021, on observe que les mois de juillet et août présentent les moyennes mensuelles de température les plus élevées, tandis que les mois de décembre et janvier ont les moyennes les plus basses (figure 4.5). La moyenne annuelle de la température sur cette période de 10 ans est de  $19.35^\circ\text{C}$  (figure 4.6).

— **L'humidité relative** En raison de sa proximité avec la Méditerranée, la ville de Bejaïa est caractérisée par une humidité élevée. La moyenne annuelle de l'humidité sur une période de 10 ans est de  $72.96\%$ . Une ob-

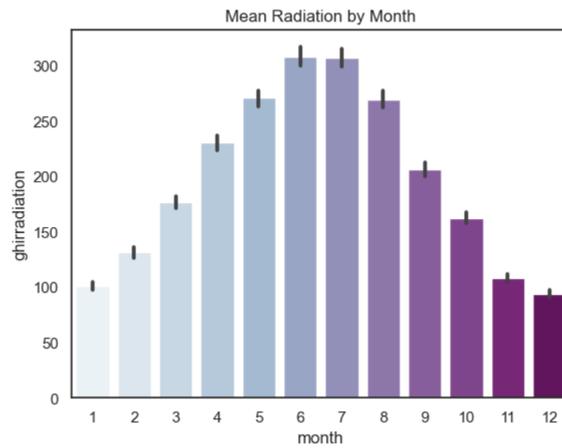


FIGURE 4.2 – Moyenne mensuelle par mois de l'irradiation (wh/M2/H) mesurée a bejaia sur la période 2010-2021

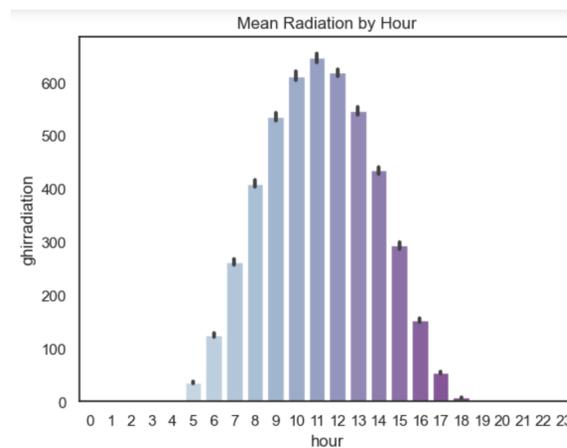


FIGURE 4.3 – Moyenne mensuelle par heure de l'irradiation (wh/M2/H) mesurée a bejaia sur la période 2010-2021

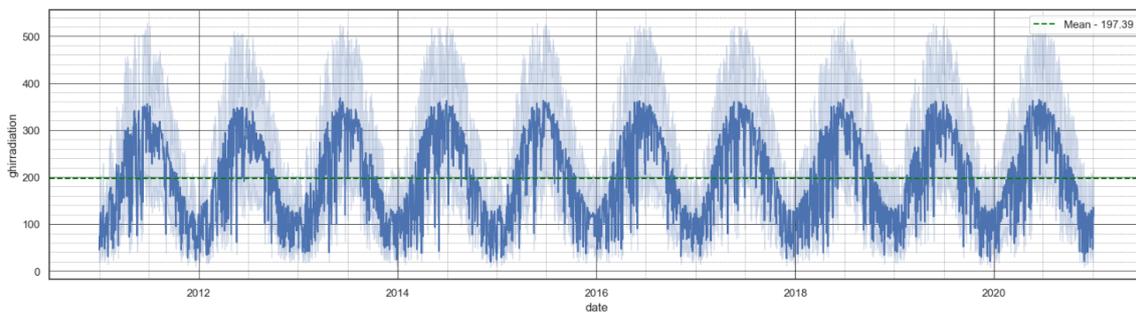


FIGURE 4.4 – Moyenne mensuelle de l'irradiation (wh/M2/H) mesurée a bejaia sur la période 2010-2021

servation intéressante est que l'évolution de l'humidité est inverse à celle de la température et de l'irradiation solaire. Cela signifie que lorsque la température est élevée, l'humidité a tendance à être plus basse et vice versa. Cette tendance est particulièrement visible dans les variations horaires, comme le montre la Figure 4.7.

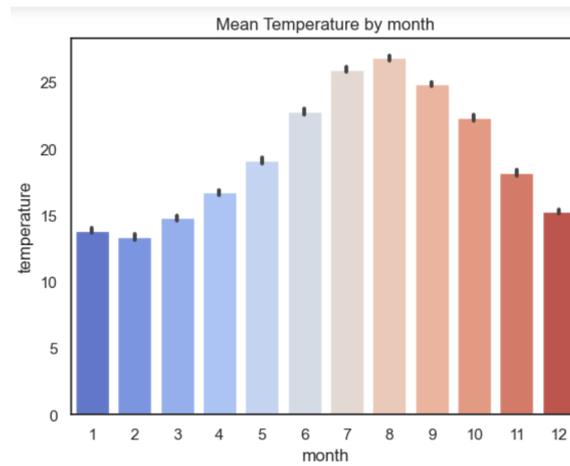


FIGURE 4.5 – Moyenne mensuelle de temperature c mesurée a bejaia sur la période 2010-2021 par mois

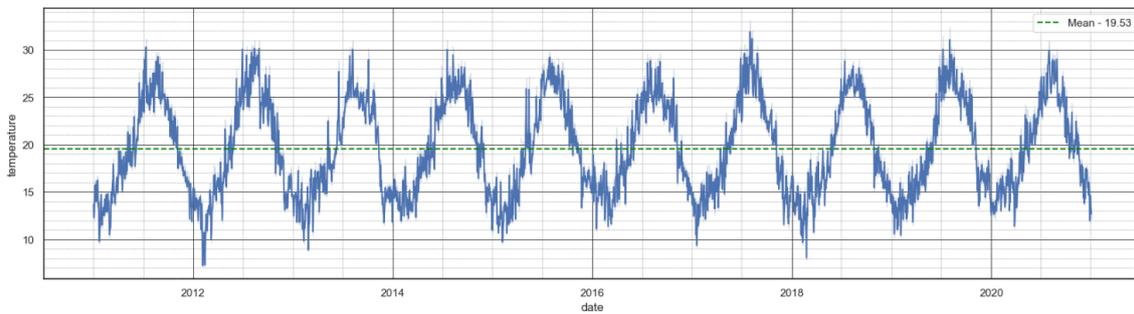


FIGURE 4.6 – Moyenne mensuelle de temperature c mesurée a bejaia sur la période 2010-2021

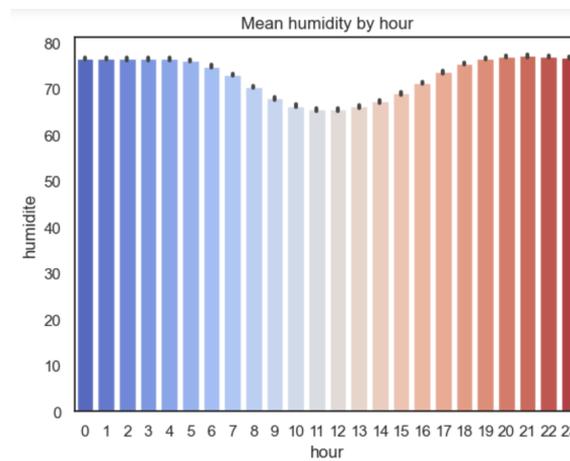


FIGURE 4.7 – Moyenne mensuelle par heure d’humidité mesurée a bejaia sur la période 2010-2021

— **La pression** Sur une période de 10 ans, la moyenne de la vitesse du vent à Bejaia est de 4.53 m/s. La valeur maximale de la vitesse du vent est observée en janvier et février, pendant la saison hivernale, tandis que la valeur minimale est enregistrée en juin, pendant la saison estivale. Cela suggère une augmentation de la vitesse du vent pendant les mois d’hiver

et une diminution pendant les mois d'été à Bejaia.(figure 4.8)

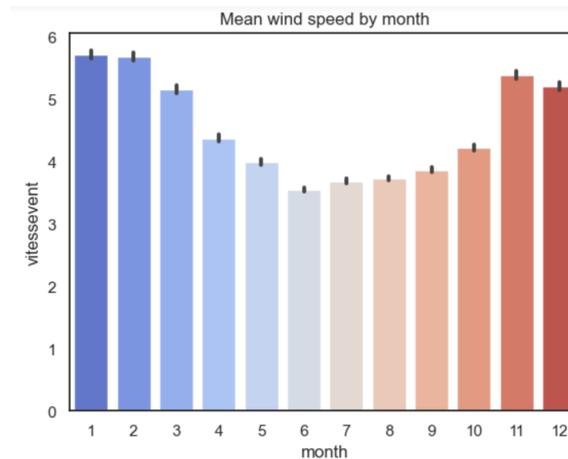


FIGURE 4.8 – Moyenne mensuelle par mois de vitesse de vent mesurée a bejaia sur la période 2010-2021

### 2. Corrélation

Un bon ensemble de données est caractérisé par des variables qui sont fortement corrélées à la classe cible et fortement non corrélées entre elles. Dans notre cas, en analysant visuellement les données, nous avons observé des relations entre les variables de température, d'humidité, de vitesse du vent, pression et d'irradiation solaire.

Cela signifie qu'il existe une corrélation ou une relation entre ces variables. Par exemple, il peut y avoir une corrélation négative entre la température et l'humidité, ce qui signifie que lorsque la température est élevée, l'humidité a tendance à être plus basse. De même, il peut y avoir une corrélation positive entre la température et l'irradiation solaire, où une augmentation de la température est associée à une augmentation de l'irradiation solaire.

Cependant, il est important de noter que la corrélation entre les variables peut varier en fonction des spécificités du jeu de données. Pour une analyse plus précise, il peut être utile d'utiliser des mesures de corrélation statistique, telles que le coefficient de corrélation de Pearson, pour quantifier les relations entre les variables.

La Figure 4.9 met en évidence que la température et l'humidité sont des caractéristiques importantes pour prédire l'irradiation solaire.

Cela indique que la température et l'humidité jouent un rôle significatif dans

la prédiction de l'irradiation solaire et qu'ils peuvent être utilisés comme des indicateurs importants pour estimer l'énergie solaire disponible.



FIGURE 4.9 – Matrice de corrélation

#### 4.3.2.2 Normalisation

La normalisation des données est une étape de prétraitement qui vise à réduire et simplifier la complexité des modèles. Elle consiste à ajuster les valeurs des données de sorte qu'elles se situent dans une plage de 0 à 1. La normalisation permet d'éviter les problèmes liés à l'échelle des variables, facilitant ainsi la convergence des algorithmes d'apprentissage automatique. En ramenant les valeurs des données dans une plage commune, la normalisation facilite également la comparaison et l'interprétation des différentes caractéristiques. les paramètres d'entrée ( $X_i$ ) et de sortie ( $Y_i$ ) sont normalisés selon les relations suivantes :

$$X_n = X_{min} + \frac{x_i - x_{min}}{x_{max} - x_{min}}(X_{max} - X_{min}) \quad (4.1)$$

$$Y_n = Y_{min} + \frac{y_i - y_{min}}{y_{max} - y_{min}}(Y_{max} - Y_{min}) \quad (4.2)$$

où  $x_{min} \leq x_i \leq x_{max}$  et  $y_{min} \leq y_i \leq y_{max}$  et  $X_n$  et  $Y_n$  sont les paramètres d'entrée et de sortie normalisés respectivement.

À la fin du processus d'apprentissage, les valeurs de sortie calculées par le modèle peuvent être dénormalisées selon la relation suivante :

$$y_i^d = \frac{Y_i - Y_{min}}{Y_{max} - Y_{min}}(y_{max} - y_{min}) + y_{min} \quad (4.3)$$

Il est également important de noter qu'il existe d'autres méthodes de normalisation disponibles dans sklearn une bibliothèque de Python, telles que StandardScaler, MinMaxScaler et bien d'autres, qui peuvent être utilisées en fonction des besoins spécifiques du modèle et des données.

Dans notre cas nous avons utilisé la méthodes de normalisation StandarScaler.

### 4.3.3 Sélection de modèle

Lorsqu'il s'agit de prédire l'énergie solaire, la sélection du modèle approprié pour effectuer les prédictions est une étape cruciale. Cela nécessite d'explorer différents algorithmes d'apprentissage automatique (Machine Learning) ou statistique afin de trouver le modèle le plus adapté au problème spécifique.

Dans cette étude, plusieurs modèles ont été considérés pour prédire l'énergie solaire pour différents horizon horaire et journalier. Parmi ceux-ci, nous avons utilisé des modèles base sur le « machine learning » les classes de régresseurs correspondantes à partir de la bibliothèque scikit-learn tels que l'Extra Trees, la Régression linéaire (Linear Regression), l'Arbre de décision (Decision Tree), la Forêt aléatoire (Random Forest), le Gradient Boosting, et les K-plus proches voisins (KNeighbors) et ne méthode d'ensemble le Stacking qui permet de combiner les prédictions de plusieurs modèles pour améliorer les performances de prédiction. De plus, nous avons également utilisé d'autres modèles base sur le « deep learning » avec de reseaux de neurone tels que RNN plus précisément les cellules LSTM qu'on a utilise avec deux méthodes une avec entrées mono-valeur(irradiation solaire seulement) et avec multi-valeur(données métrologique et irradiation solaire) et nous avons utilisé aussi une méthodes statistique qui est SARIMA . Ces modèles comportent différents hyperparamètres qu'on a sélectionner pour optimiser la performance de chaque modèle.

### 4.3.4 Entraînement et test et validation des données

La division de l'ensemble de données en parties d'entraînement, de test et de validation est une étape essentielle dans le processus de création de modèles d'apprentissage automatique. Cette division permet d'évaluer les performances du modèle sur des données indépendantes et d'estimer sa capacité à généraliser les prédictions sur de nouvelles données.

**Partie d'entraînement** : La partie d'entraînement est utilisée pour ajuster les paramètres du modèle en exposant le modèle à des exemples d'entrée avec leurs sorties correspondantes. Le modèle apprend à partir de ces exemples et essaie de minimiser l'erreur entre les sorties prédites et les sorties réelles. Cette partie représente la majorité des données. Dans notre travail nous avons utilisé 80% des données pour l'ensemble d'entraînement.

**Partie de test** : La partie de test est utilisée pour évaluer les performances du modèle sur des données qu'il n'a pas encore vues. Elle permet de mesurer à quel point le modèle est capable de généraliser et de faire des prédictions précises sur de nouvelles données. Pour cette partie dans notre travail nous avons utilisé 10% du dataset.

**Partie de validation** : La partie de validation est utilisée pour ajuster les hyperparamètres du modèle et effectuer des ajustements supplémentaires pour améliorer ses performances. Nous avons utilisé 10% des données aussi pour cette partie.

Une fois que les données ont été divisées, il devient essentiel de procéder à l'entraînement et le test du modèle, ce qui implique deux étapes distinctes : l'apprentissage et l'évaluation en vue de définir un modèle.

#### 4.3.4.1 La phase d'apprentissage

Les propriétés désirables pour un réseau de neurones, la plus fondamentale est sûrement la capacité d'apprendre de son environnement, d'améliorer sa performance à travers un processus d'apprentissage.

L'apprentissage est un processus dynamique et itératif permettant de modifier les paramètres d'un modèle ML en réaction avec les stimuli qu'il reçoit de son environnement.

#### 4.3.4.2 La phase de test (Évaluation)

Une fois qu'un modèle a été développé, il est essentiel d'évaluer ses performances et sa capacité de généralisation. Cette étape finale consiste à estimer la qualité du modèle en lui présentant des exemples qui ne font pas partie de l'ensemble d'apprentissage. Si les performances du réseau ne sont pas satisfaisantes, des ajustements doivent être effectués soit en modifiant les hyperparamètres, soit en révisant la base d'apprentissage [3].

L'évaluation du réseau repose sur des indicateurs statistiques spécifiques qui permettent de mesurer sa performance. Ces indicateurs servent à évaluer la précision, la robustesse et la capacité de généralisation du modèle. Dans les titres suivant, nous présenterons en détail les indicateurs statistiques utilisés dans cette étude pour évaluer et comparer les performances de chaque modèles.

1. La racine carrée de l'erreur quadratique moyenne (RMSE) et sa valeur normalisée (nRMSE) :

$$RMSE = \sqrt{\sum_{i=1}^p (y_i^d - y_i)^2 / p}, \quad nRMSE = RMSE / \bar{y} \quad (4.4)$$

la précision du modèle est considérée excellente si  $nRMSE < 10\%$ ; bonne si  $10\% < nRMSE < 20\%$ ; raisonnable si  $20\% < nRMSE < 30\%$ ; faible si  $nRMSE > 30\%$  [3]. L'erreur du RMSE ont été recommandées récemment par l'agence internationale de l'énergie (IEA) pour déclarer la précision des modèles

2. L'erreur absolue moyenne (MAE) et sa valeur normalisée (nMAE) :

$$MAE = \sum_{i=1}^p |y_i^d - y_i| / p \quad nMAE = MAE / \bar{y} \quad (4.5)$$

3. Le coefficient de détermination  $R^2$  :

$$R^2 = 1 - \frac{\sum_{i=1}^p (y_i^d - y_i)^2}{\sum_{i=1}^p (y_i^d - \bar{y}_i)^2} \quad (4.6)$$

Le modèle est considéré efficace si la valeur de  $R^2$  est proches de 1. Avec

$$\bar{y} = \sum_{i=1}^p (y_i/p).$$

### 4.4 Réalisation

Dans cette étude, nous avons utilisé différentes méthodes de prédiction, à savoir le Decision tree, Random Forest, Extra trees, Gradient Boosting, KNN, LSTM-monovalues, LSTMmultivalues, SARIMA et une méthode d'ensemble (Stacking qui permet de combiner les prédictions de plusieurs modèles pour améliorer les performances de prédiction), pour prédire l'irradiation solaire à la ville de Bejaia-Algérie. Deux ensembles de données distincts ont été considérés, l'un contenant des données par jour et l'autre contenant des données par heure. Les prédictions ont été réalisées pour deux horizons temporels différents : une heure après (H+1) et un jour après (J+1).

#### 4.4.1 Le cas journalier

Pour le cas du jeu de données horaires, nous avons développé différentes méthodes d'apprentissage machine, statistique et d'ensemble : méthodes de régression, LSTM, Stacking.

Pour ce cas nous allons présenter les modèles de prédiction utilisés et une analyse comparative de leur performances.

##### 4.4.1.1 Les méthodes de régression

Les méthodes de régression ont utilisées dans cette étude comprennent la régression linéaire, les arbres de décision, les forêts aléatoires, les Extra Trees, le Gradient Boosting et les K plus proches voisins. Chaque méthode offre une approche unique pour prédire l'énergie solaire en fonction des caractéristiques du système. Ces méthodes ont été évaluées à l'aide de la validation croisée, en utilisant des métriques telles que le coefficient de détermination, MAE et RMSE (Figure 4.10).

##### 4.4.1.2 LSTM long short-term memory

Est un modèle d'apprentissage en profond séquentiel. le modèle LSTM (Long Short-Term Memory) composée de plusieurs couches LSTM et Droupout. Il prend

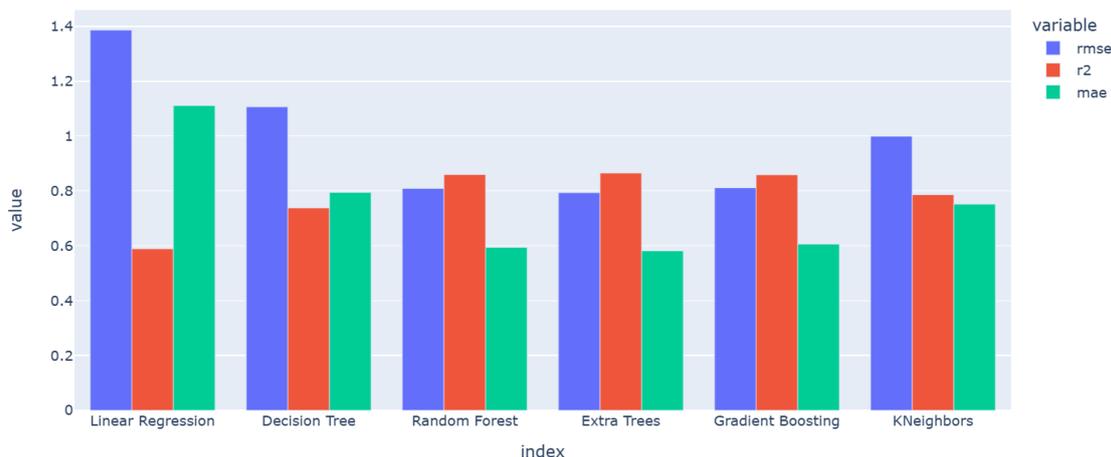


FIGURE 4.10 – Comparaison entre les modèles de régression cas journalier.

en entrée des séquences de longueur 7 avec une dimension 1 pour mono valeur et 4 pour le deuxième modèle multi valeur. Les premières couches LSTM traitent les séquences d’entrée, puis les couches Dropout sont utilisées pour la régularisation du modèle. La dernière couche LSTM renvoie une sortie unidimensionnelle, qui est ensuite alimentée dans des couches Dense. Le modèle a un total de 168,001 et 16,961 paramètres entraînables est utilisé pour effectuer des prédictions sur des données séquentielles, en particulier des séries temporelles (figure 4.11 et 4.12)

Layer (type)	Output Shape	Param #
lstm_47 (LSTM)	(None, 64)	16896
dense_86 (Dense)	(None, 1)	65

=====  
 Total params: 16,961  
 Trainable params: 16,961  
 Non-trainable params: 0  
 =====

FIGURE 4.11 – Les paramètres de modèle LSTMmonovaleur cas journalier.

- **LSTMmonovaleur** :Le modèle est entraîné pendant 32 époques avec l’algorithme d’optimisation NAdam. Nous avons remarqué que la courbe de perte d’entraînement (train loss) est décroissante, ce diminue de 0.35 vers 0.017 et la même chose Mise en œuvre et évaluation pour la courbe de perte de validation, qui diminuait de 0.20 à 0.018 comme montre la figure 4.13.

Layer (type)	Output Shape	Param #
input_2 (InputLayer)	[(None, 24, 4)]	0
lstm_3 (LSTM)	(None, 24, 64)	17664
dropout_23 (Dropout)	(None, 24, 64)	0
lstm_4 (LSTM)	(None, 24, 128)	98816
dropout_24 (Dropout)	(None, 24, 128)	0
lstm_5 (LSTM)	(None, 64)	49408
dropout_25 (Dropout)	(None, 64)	0
dense_27 (Dense)	(None, 32)	2080
dense_28 (Dense)	(None, 1)	33

=====  
Total params: 168,001  
Trainable params: 168,001  
Non-trainable params: 0

FIGURE 4.12 – Les paramètres de modèle LSTMmultivaleur cas journalier.

- **LSTMmonovaleur** :Le modèle est entraîné pendant 32 époques avec l’algorithme d’optimisation NAdam. Nous avons remarqué que la courbe de perte d’entraînement (train loss) est décroissante, ce di-minuit de 0.35 vers 0.017 et la même chose Mise en œuvre et évaluation pour la courbe de perte de validation, qui diminuait de 0.20 à 0.018 comme montre la figure 4.14

#### 4.4.1.3 Stacking

Dans le modèle de stacking nous avons créé une liste estimators comprenant différents modèles tels que MLP, DecisionTreeRegressor, GradientBoostingRegressor, RandomForestRegressor et XGBoostRegressor et un modèle final de GradientBoostingRegressor est également spécifié. Le modèle est ensuite ajusté aux données d’entraînement. Ce modèle utilise les prédiction issues des différents modèles (estimators) sont utilisées en entrée d’un modèle final (méta-learner), qui va alors combiner les prévisions afin d’obtenir la meilleure prédiction finale prévue (4.15).

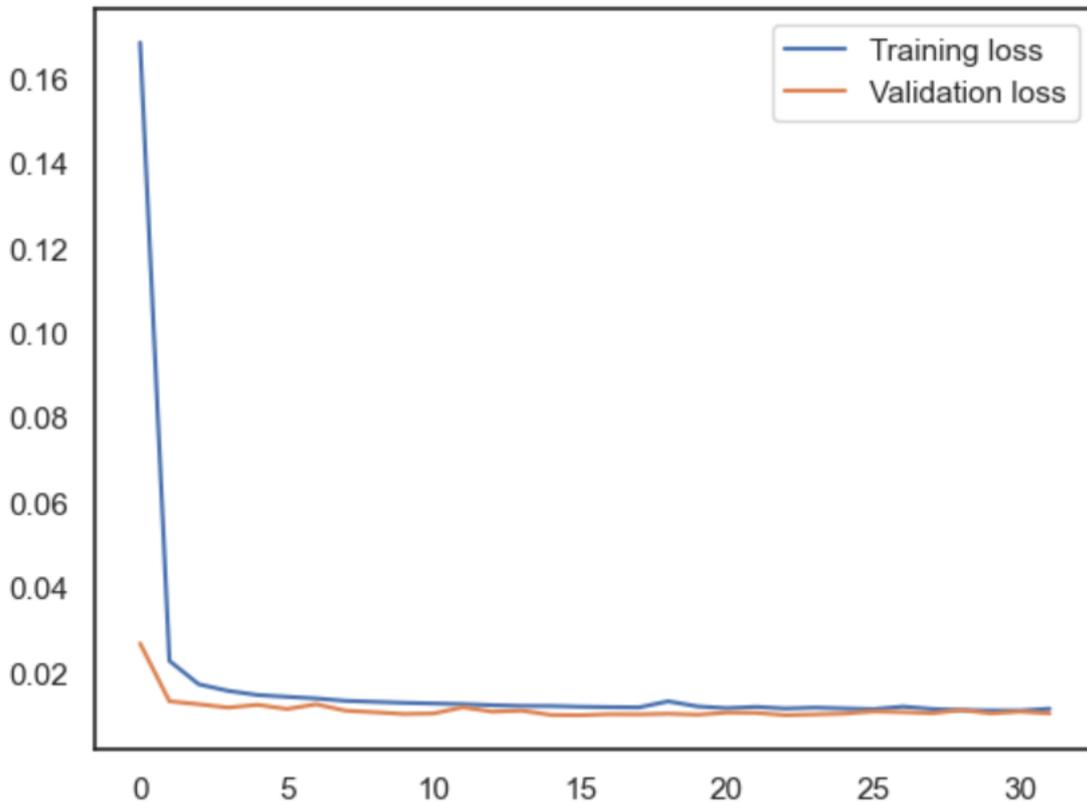


FIGURE 4.13 – évolution du modèle LSTMmonovaleur cas journalier.

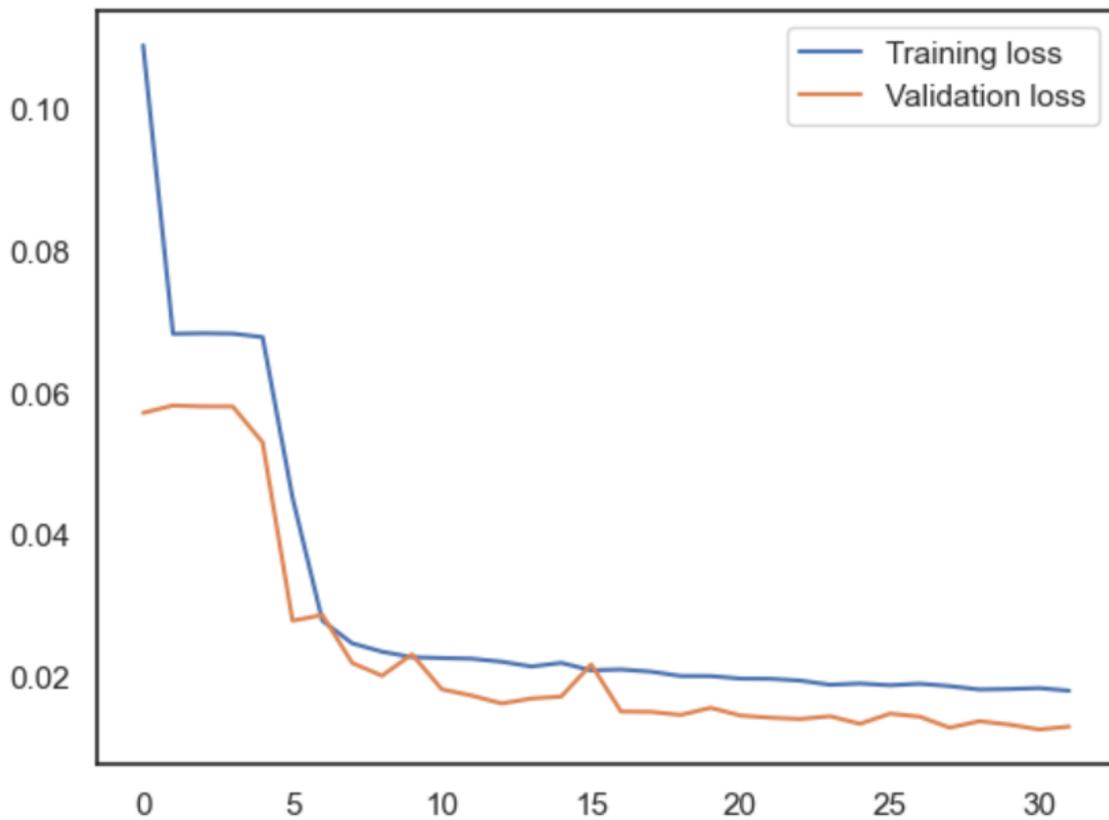


FIGURE 4.14 – évolution du modèle LSTMmultivaleur cas journalier.

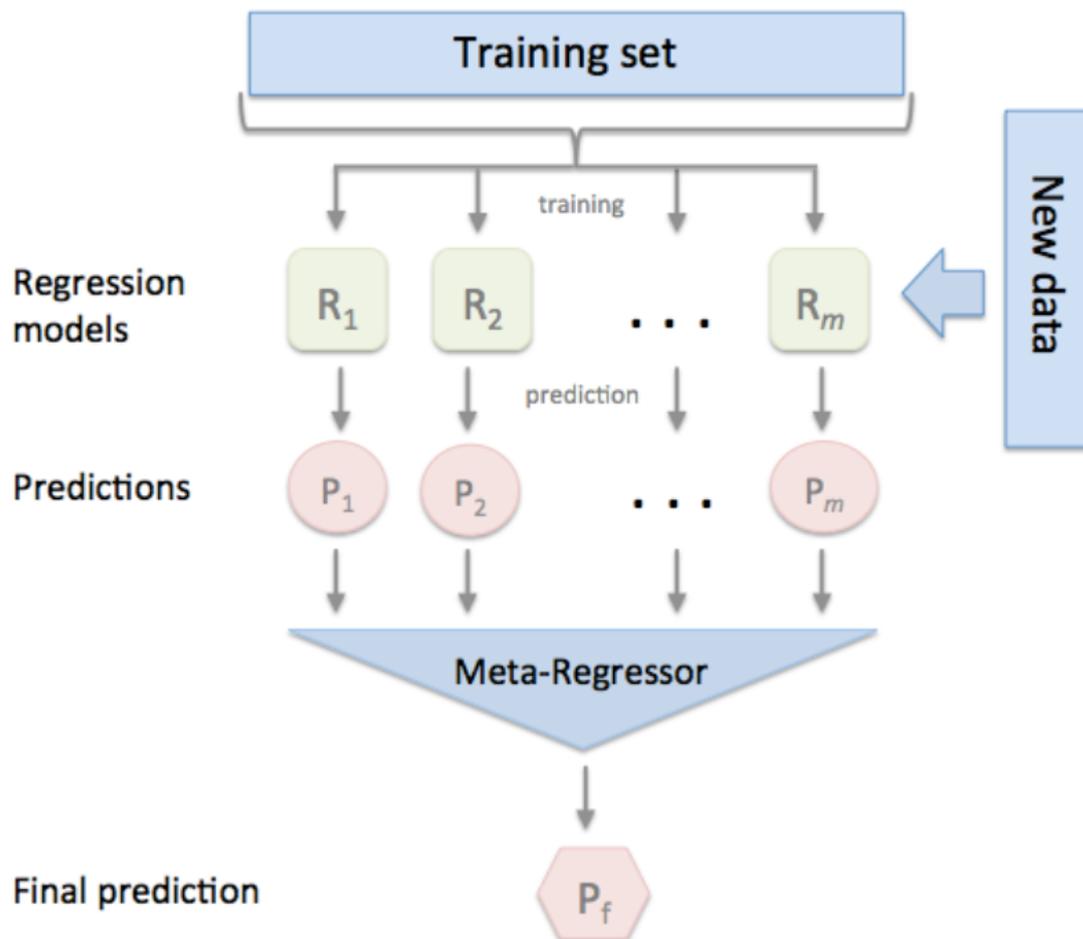


FIGURE 4.15 – Stacking modèle [9].

#### 4.4.1.4 SARIMA

Pour SARIMA nous avons effectué une recherche pas à pas pour sélectionner le meilleur modèle (stepwise search) en minimisant le critère d'information d'Akaike (AIC) (Figure 4.16). Le modèle choisi comme le meilleur est un SARIMA(3,0,2). Cela signifie qu'il utilise 3 termes autoregressifs (AR) non saisonniers, pas de différenciation saisonnière, et 2 termes de moyenne mobile (MA) saisonniers. De plus, une période saisonnière de 12 a été choisie, ce qui implique que le modèle prend en compte des motifs saisonniers qui se répètent tous les 12 pas de temps. Ces choix ont été faits pour mieux ajuster le modèle aux données et capturer les variations saisonnières.

```

Performing stepwise search to minimize aic
ARIMA(2,0,2)(0,0,0)[0] intercept : AIC=11016.885, Time=2.34 sec
ARIMA(0,0,0)(0,0,0)[0] intercept : AIC=16065.452, Time=0.04 sec
ARIMA(1,0,0)(0,0,0)[0] intercept : AIC=11727.420, Time=0.22 sec
ARIMA(0,0,1)(0,0,0)[0] intercept : AIC=13839.574, Time=0.32 sec
ARIMA(0,0,0)(0,0,0)[0] : AIC=22440.327, Time=0.02 sec
ARIMA(1,0,2)(0,0,0)[0] intercept : AIC=11033.088, Time=1.12 sec
ARIMA(2,0,1)(0,0,0)[0] intercept : AIC=11016.766, Time=1.52 sec
ARIMA(1,0,1)(0,0,0)[0] intercept : AIC=11241.523, Time=0.77 sec
ARIMA(2,0,0)(0,0,0)[0] intercept : AIC=11507.327, Time=0.42 sec
ARIMA(3,0,1)(0,0,0)[0] intercept : AIC=11016.874, Time=2.31 sec
ARIMA(3,0,0)(0,0,0)[0] intercept : AIC=11346.047, Time=0.53 sec
ARIMA(3,0,2)(0,0,0)[0] intercept : AIC=11015.112, Time=1.54 sec
ARIMA(4,0,2)(0,0,0)[0] intercept : AIC=11016.996, Time=1.90 sec
ARIMA(3,0,3)(0,0,0)[0] intercept : AIC=11017.071, Time=3.77 sec
ARIMA(2,0,3)(0,0,0)[0] intercept : AIC=11036.229, Time=3.37 sec
ARIMA(4,0,1)(0,0,0)[0] intercept : AIC=11018.874, Time=2.80 sec
ARIMA(4,0,3)(0,0,0)[0] intercept : AIC=11020.510, Time=4.36 sec
ARIMA(3,0,2)(0,0,0)[0] : AIC=11023.383, Time=1.62 sec

Best model: ARIMA(3,0,2)(0,0,0)[0] intercept
Total fit time: 28.999 seconds

```

FIGURE 4.16 – Meilleur SARIMA modèle.

## 4.4.2 Le cas horaire

Différents modèles ont été utilisés pour la prédiction d'énergie solaire à Bejaia pour l'horizon horaire : méthodes de régression et LSTM.

### 4.4.2.1 Les méthodes de régression

Les méthodes de régression utilisées dans cette étude comprennent la régression linéaire, les arbres de décision, les forêts aléatoires, les Extra Trees, le Gradient Boosting et les K plus proches voisins. Chaque méthode offre une approche unique pour prédire l'énergie solaire en fonction des caractéristiques du système.

Ces méthodes ont été évaluées à l'aide de la validation croisée, en utilisant des métriques telles que le coefficient de détermination, MAE et RMSE (Figure 4.17).

### 4.4.2.2 LSTM long short-term memory

Est un modèle d'apprentissage en profond séquentiel. Le modèle LSTM (Long Short-Term Memory) composée de plusieurs couches LSTM et Dropout. Il prend en entrée des séquences de longueur 24 avec une dimension 1 pour une seule valeur et



FIGURE 4.17 – Comparaison entre les modèles de régression cas horaire.

4 pour le deuxième modèle multi valeur. Les premières couches LSTM traitent les séquences d’entrée, puis les couches Dropout sont utilisées pour la régularisation du modèle. La dernière couche LSTM renvoie une sortie unidimensionnelle, qui est ensuite alimentée dans des couches Dense. Le modèle a un total de 85 057 paramètres entraînaables et est utilisé pour effectuer des prédictions sur des données séquentielles, en particulier des séries temporelles (figure 4.18 et 4.19).

- **LSTMmono Valeur** : Le modèle est entraîné pendant 32 époques avec l’algorithme d’optimisation NAdam. Nous avons remarqué que la courbe de perte d’entraînement (train loss) est décroissante, ce qui diminue de 0.25 vers 0.009 et la même chose Mise en œuvre et évaluation pour la courbe de perte de validation, qui diminuait de 0.11 à 0.008 comme montre la figure 4.20.
- **LSTMmulti Valeur** : Le modèle est entraîné pendant 32 époques avec l’algorithme d’optimisation NAdam. Nous avons remarqué que la courbe de perte d’entraînement (train loss) est décroissante, ce qui diminue de 0.16 vers 0.019 et la même chose Mise en œuvre et évaluation pour la courbe de perte de validation, qui diminuait de 0.03 à 0.0017 comme montre la figure 4.21.

## 4.5 Résultats et discussion

Dans les deux cas d’horizon, nous avons également évalué les performances des différentes méthodes de prédiction à l’aide de RMSE, MAE et R2. Les résultats obtenus sont les suivants (Table 4.2 et Table 4.1)

En analysant ces résultats, nous pouvons tirer les observations suivantes :

Layer (type)	Output Shape	Param #
input_1 (InputLayer)	[(None, 24, 1)]	0
lstm (LSTM)	(None, 24, 64)	16896
dropout (Dropout)	(None, 24, 64)	0
lstm_1 (LSTM)	(None, 24, 64)	33024
dropout_1 (Dropout)	(None, 24, 64)	0
lstm_2 (LSTM)	(None, 64)	33024
dropout_2 (Dropout)	(None, 64)	0
dense (Dense)	(None, 32)	2080
dense_1 (Dense)	(None, 1)	33

=====  
 Total params: 85,057  
 Trainable params: 85,057  
 Non-trainable params: 0

FIGURE 4.18 – Les paramètres de modèle LSTMmonovaleur cas horaire

Méthode	RMSE	R2	MAE
Decision tree	1.08	75%	0.78
Random Forest	0.79	86%	0.58
Extra trees	0.78	86%	0.57
Gradient Boosting	0.80	86%	0.59
KNN	0.99	79%	0.75
LSTMmonovalues	1.06	67%	0.84
LSTMmultivalues	1.06	75%	0.84
Sarima	1.03	77%	0.79
<b>Méthode d'ensemble stacking</b>	<b>0.76</b>	<b>88%</b>	<b>0.56</b>

TABLE 4.1 – évaluation de toutes les méthodologies de prédiction étudiées par jour j+1.

Layer (type)	Output Shape	Param #
input_2 (InputLayer)	[(None, 24, 4)]	0
lstm_3 (LSTM)	(None, 24, 64)	17664
dropout_23 (Dropout)	(None, 24, 64)	0
lstm_4 (LSTM)	(None, 24, 128)	98816
dropout_24 (Dropout)	(None, 24, 128)	0
lstm_5 (LSTM)	(None, 64)	49408
dropout_25 (Dropout)	(None, 64)	0
dense_27 (Dense)	(None, 32)	2080
dense_28 (Dense)	(None, 1)	33

=====  
 Total params: 168,001  
 Trainable params: 168,001  
 Non-trainable params: 0

FIGURE 4.19 – Les paramètres de modèle LSTMmultivaleur cas horaire.

Méthode	RMSE	R2	MAE
Decision tree	79.88	91.62%	36.95
Random Forest	56.84	95.75%	27.75
Extra trees	56.88	95.75%	27.85
Gradient Boosting	73.60	92.89%	46.97
KNN	66.25	94.24%	39.43
LSTMmonovalues	26.95	99%	13.02
<b>LSTMmultivalues</b>	<b>25.91</b>	<b>99.1%</b>	<b>11.88</b>

TABLE 4.2 – évaluation de toutes les méthodologies de prédiction étudiées par heure h+1.

- Les méthodes basées sur les arbres de décision (Decision tree, Random Forest, Extra trees) ont obtenu des performances comparables pour les deux Horizons. Elles ont produit des résultats précis, en particulier dans le cas horaire.

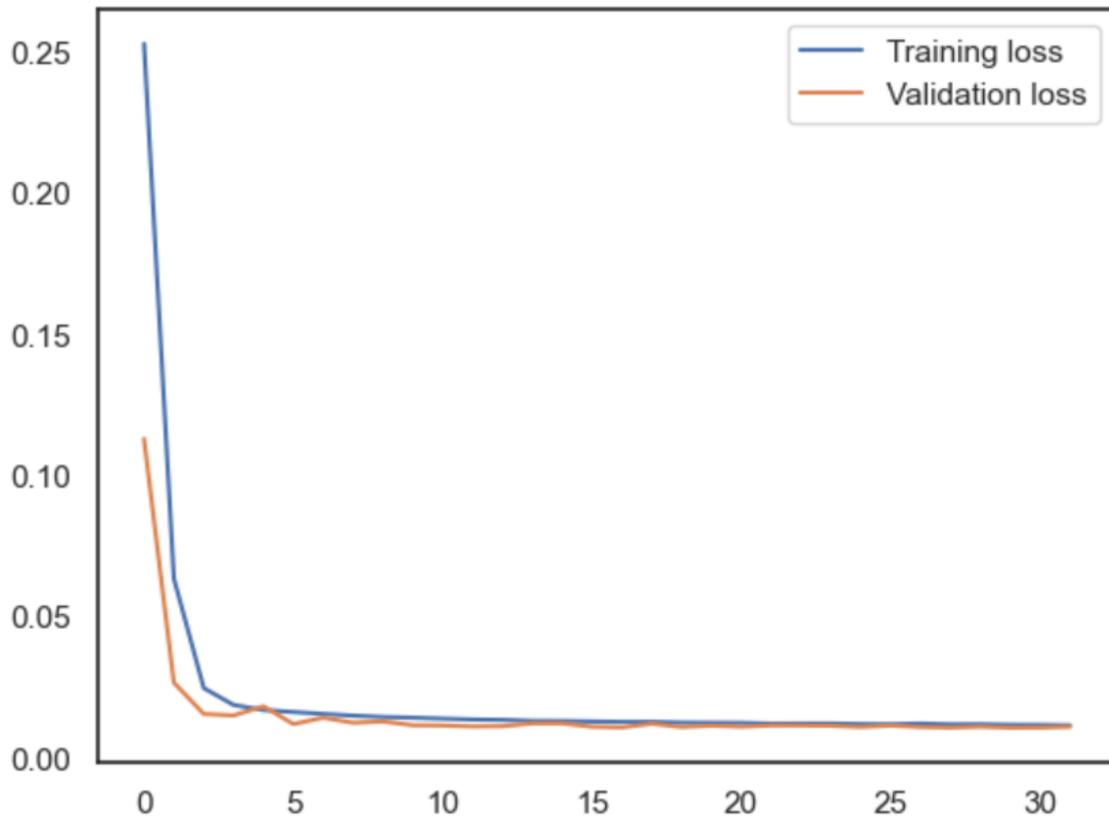


FIGURE 4.20 – évolution du modèle LSTMmonovaleur cas horaire.

- Le Gradient Boosting s’est également révélé être une méthode efficace avec des performances solides pour les deux cas de jeu de données.

- Les méthodes basées sur les réseaux de neurones récurrents, LSTMmonovalues et LSTMmultivalues, ont démontré une capacité prometteuse à capturer les motifs temporels des données, en particulier dans le cas horaire, où elles ont surpassé les autres méthodes en termes de précision.

- Les méthodes de séries temporelles, comme SARIMA, ont donné des performances variables, ce qui suggère que leur efficacité peut être influencée par la nature spécifique des données et les tendances saisonnières.

Selon ces résultats, le meilleur modèle dans le cas horaire serait LSTMmultivalues, qui présente le RMSE le plus bas, le R-square le plus élevé et le MAE le plus faible. Il est capable de capturer efficacement les motifs temporels des données et fournit des prédictions très précises.

Dans le cas journalier, Random Forest et Extra trees se démarquent avec des

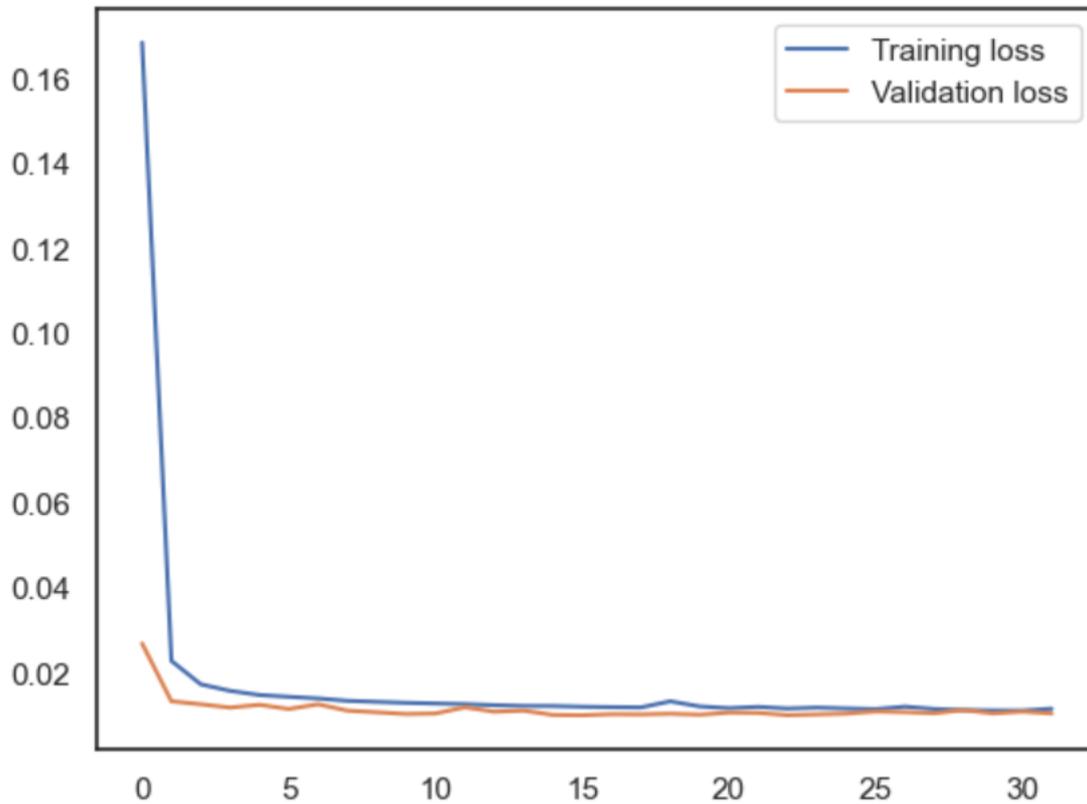


FIGURE 4.21 – évolution du modèle LSTMmultivaleur cas horaire.

valeurs de RMSE, R-square et MAE très proches et meilleures que celles des autres modèles. Ces deux modèles sont robustes et offrent une bonne précision pour la prédiction de l'irradiation solaire sur une base quotidienne. Cependant, nous avons également utilisé une méthode d'ensemble appelée stacking, qui combine les prédictions des modèles Random Forest et Extra Trees et d'autres modèles (MLPRegressor, DecisionTreeRegressor...etc). Cette approche nous a permis d'obtenir les meilleurs résultats en termes de performance. Le stacking a donné un RMSE de 0.76, un R-square de 87.6 et un MAE de 0.56, ce qui est encore meilleur que les résultats obtenus avec les modèles individuels. Cela confirme l'efficacité du stacking en tant qu'approche prometteuse pour améliorer les performances de prédiction en combinant plusieurs modèles.

### 4.6 Conclusion

Dans cette étude, nous avons exploré plusieurs méthodes de prédiction pour estimer l'irradiation solaire à Bejaia, en Algérie. Nous avons utilisé deux ensembles de données distincts, l'un avec des données horaires et l'autre avec des données journalières, et avons évalué les performances des différents modèles pour prédire l'irradiation solaire à des horizons temporels différents.

Dans le cas des données horaires, nous avons constaté que les méthodes basées sur les arbres de décision, telles que Decision tree, Random Forest et Extra trees, ont produit des résultats précis. Le Gradient Boosting s'est également révélé être une méthode efficace. Cependant, les méthodes basées sur les réseaux de neurones récurrents, telles que LSTMmonovalues et LSTMmultivalues, ont surpassé les autres méthodes en termes de précision, en capturant efficacement les motifs temporels des données.

Pour les données journalières, les méthodes basées sur les arbres de décision, en particulier Random Forest et Extra trees, ont obtenu de bons résultats avec des performances comparables. Cependant, en utilisant une méthode d'ensemble appelée stacking, qui combine les prédictions de plusieurs modèles, nous avons pu améliorer les performances de prédiction, obtenant ainsi les meilleurs résultats en termes de précision.

Cependant, il est important de noter que le choix de la méthode de prédiction dépendra de la nature spécifique des données et des objectifs de l'étude. Une analyse approfondie des caractéristiques des données et des objectifs de l'étude permettra de choisir la méthode de prédiction la plus adaptée pour estimer l'irradiation solaire à Bejaia, en vue d'une utilisation optimale des ressources solaires dans cette région.

# Conclusion générale & Perspectives

En conclusion, ce mémoire a examiné en détail les différents aspects de l'énergie solaire, en mettant particulièrement l'accent sur la prédiction de l'irradiation solaire. Le premier chapitre a permis d'explorer les différentes formes d'énergie, en se concentrant sur l'énergie solaire et ses multiples applications. Le deuxième chapitre a posé les bases nécessaires à la compréhension et à l'approche de la prédiction des séries temporelles, en présentant diverses techniques et approches utilisées dans ce domaine. Le troisième chapitre a ensuite examiné des travaux et des méthodes spécifiques de prédiction pour l'énergie solaire, mettant en évidence l'efficacité des techniques d'apprentissage automatique pour obtenir des prévisions plus précises. Enfin, le quatrième chapitre a appliqué ces méthodes de prédiction à des ensembles de données spécifiques pour prédire l'irradiation solaire à Bejaia, en Algérie, évaluant ainsi les performances de différentes approches.

Les résultats de cette étude ont clairement démontré que les méthodes basées sur les réseaux de neurones récurrents, tels que LSTM, surpassaient les autres méthodes en termes de précision de prédiction. Toutefois, les méthodes basées sur les arbres de décision, notamment Random Forest et Extra Trees, ont également produit des résultats prometteurs. De plus, l'utilisation d'une méthode d'ensemble, comme le stacking, a permis d'améliorer les performances de prédiction.

Il est important de souligner que le choix de la méthode de prédiction dépendra des caractéristiques spécifiques des données et des objectifs de l'étude. Une analyse approfondie de ces éléments permettra de sélectionner la méthode de prédiction la plus adaptée pour estimer l'irradiation solaire et optimiser ainsi l'utilisation des ressources solaires dans une région donnée. Cette recherche offre des perspectives significatives pour l'intégration efficace de l'énergie solaire dans divers domaines d'application.

En somme, ce mémoire apporte une contribution importante à la compréhension de l'énergie solaire et de la prédiction de l'irradiation solaire. Les résultats obtenus soulignent l'efficacité des méthodes d'apprentissage automatique dans la prédiction précise de l'énergie solaire.

Les résultats de cette étude ouvrent la voie à plusieurs perspectives prometteuses pour la prédiction de l'irradiation solaire et l'utilisation efficace de l'énergie solaire. Voici quelques perspectives à considérer :

- Amélioration des modèles existants : Bien que les méthodes basées sur les réseaux de neurones récurrents aient montré de bonnes performances, il est possible d'explorer et d'améliorer davantage ces modèles. Des recherches supplémentaires pourraient se concentrer sur l'optimisation des architectures de réseaux de neurones, la sélection des hyperparamètres et l'utilisation de techniques d'apprentissage en ligne pour des mises à jour en temps réel des prédictions.
- Intégration de données supplémentaires : Pour améliorer la précision des prévisions, il peut être intéressant d'intégrer des données supplémentaires, telles que des données météorologiques, des données de capteurs solaires et des données géographiques. L'utilisation de ces données complémentaires peut permettre de capturer des relations plus complexes et d'améliorer la capacité des modèles à prédire l'irradiation solaire.
- Expérimentation de nouvelles techniques : Outre les modèles traditionnels d'apprentissage automatique, il serait intéressant d'explorer de nouvelles techniques émergentes telles que les réseaux de neurones convolutifs, les modèles de séquence à séquence (Seq2Seq), les modèles hybride et les modèles de transfert d'apprentissage. Ces approches pourraient apporter de nouvelles perspectives et des améliorations significatives en termes de précision de prédiction.
- Validation sur différents sites : Bien que cette étude se soit concentrée sur la prédiction de l'irradiation solaire à Bejaia, en Algérie, il serait bénéfique de valider les modèles et les méthodes sur d'autres sites géographiques. Cela permettrait de mieux comprendre la généralisation des modèles et leur applicabilité dans différentes régions du monde.
- Applications pratiques : Les résultats de cette étude peuvent être utilisés pour

soutenir la planification énergétique, l'optimisation des systèmes solaires et la gestion de la demande d'électricité. Des efforts supplémentaires pourraient être consacrés à la mise en œuvre pratique de ces modèles de prédiction dans des contextes réels, en collaborant avec des acteurs de l'industrie solaire et des organismes de régulation.

# Bibliographie

- [1] Cyril Voyant. *Prédiction de séries temporelles de rayonnement solaire global et de production d'énergie photovoltaïque à partir de réseaux de neurones artificiels*. Theses, Université Pascal Paoli, November 2011.
- [2] Contributeurs aux projets Wikimedia. Modèles du neurone biologique. [https://fr.wikipedia.org/wiki/Modèles\\_duneuronebiologique](https://fr.wikipedia.org/wiki/Modèles_duneuronebiologique), 5 2023.
- [3] AMIRI BILLEL. *Estimation Temporelle du Rayonnement Solaire au Nord et au Sahara Algérien à Partir de Données de Mesures Sélectionnées. Application de la Méthode RNA*. Theses, Université des Sciences et de la Technologie HOUARI BOUMEDIENE , March 2021.
- [4] Nahed ZEMOURI. *UTILISATION DES METHODES D'APPRENTISSAGE MACHINE POUR LA PREDICTION DES SIGNAUX ENERGETIQUES*. Theses, Université Mohamed Boudiaf - M'sila, November 2020.
- [5] Shafaqat Ali, Nadarajah Mithulananthan, and Chandima Ekanayake. On recent advances in PV output power forecast. *Solar Energy*, 136 :125–144, 10 2016.
- [6] Razin Ahmed, Victor Sreeram, Yateendra Mishra, and M. D. Arif. A review and evaluation of the state-of-the-art in PV solar power forecasting : Techniques and optimization. *Renewable Sustainable Energy Reviews*, 124 :109792, 5 2020.
- [7] Utpal Das, Kok Soon Tey, Mehdi Seyedmahmoudian, Saad Mekhilef, Moh. Idris, Willem Van Deventer, Ben Horan, and Alex Stojcevski. Forecasting of photovoltaic power generation and model optimization : A review. *Renewable Sustainable Energy Reviews*, 81 :912–928, 1 2018.
- [8] Yong Shuai, Andreea Sabadus, Nicoleta Stefu, Ciprian Dughir, Marius Paulescu, and Viorel Badescu. A current perspective on the accuracy of incoming solar energy forecasting. *Progress in Energy and Combustion Science*, 70 :119–144, 1 2019.

- [9] Sebastian Raschka. StackingCVRegressor : stacking with cross-validation for regression - mlxtend. [https://rasbt.github.io/mlxtend/user\\_guide/regressor/StackingCVRegressor/](https://rasbt.github.io/mlxtend/user_guide/regressor/StackingCVRegressor/).
- [10] José F Torres, Alicia Troncoso, Irena Koprinska, Zheng Wang, and Francisco Martínez-Álvarez. Big data solar power forecasting based on deep learning and multiple data sources. *Expert Systems*, 36(4) :e12394, 2019.
- [11] N L Panwar, S C Kaushik, and Surendra Kothari. Role of renewable energy sources in environmental protection : A review. *Renewable and sustainable energy reviews*, 15(3) :1513–1524, 2011.
- [12] Energie solaire : fonctionnement, types, avantages, chiffres en France. <https://climate.selectra.com/fr/energie-verte/solaire>, 4 2023.
- [13] RJ Aguiar, Manuel Collares-Pereira, and JP Conde. Simple procedure for generating sequences of daily radiation values using a library of markov transition matrices. *Solar Energy*, 40(3) :269–279, 1988.
- [14] Viorel Badescu. *Modeling Solar Radiation at the Earth's Surface : Recent Advances*. Springer Nature, 11 2014.
- [15] Koray Ulgen and Arif Hepbasli. Diffuse solar radiation estimation models for Turkey's big cities. *Energy Conversion and Management*, 50(1) :149–156, 1 2009.
- [16] Le climat de la Terre - Soleil et effet de serre - Le rayonnement solaire - CNRS sagascience. [https://www.cnrs.fr/cw/dossiers/dosclim/contenu/alternative/alter\\_etape1\\_1.html](https://www.cnrs.fr/cw/dossiers/dosclim/contenu/alternative/alter_etape1_1.html).
- [17] Pierre Bessemoulin and Jean Oliviéri. Le rayonnement solaire et sa composante ultraviolette. *La météorologie*, 2000(31) :42–59, 2000.
- [18] Contributeurs aux projets Wikimedia. Rayonnement solaire. [https://fr.wikipedia.org/wiki/Rayonnement\\_solaire](https://fr.wikipedia.org/wiki/Rayonnement_solaire), 4 2023.
- [19] Shafiqur Rehman. Empirical model development and comparison with existing correlations. *Applied Energy*, 9 1999.
- [20] Francisco Javier Batlles, Monica Rubio, Juan A. Tovar, F.J. Olmo, and Lucas Alados-Arboledas. Empirical modeling of hourly direct irradiance by means of hourly global irradiance. *Energy*, 25(7) :675–688, 7 2000.

- [21] Christoph Marty and Rolf Philipona. The clear-sky index to separate clear-sky from cloudy-sky situations in climate research. *Geophysical Research Letters*, 27(17) :2649–2652, 9 2000.
- [22] Adel Mellit, Soteris A Kalogirou, S Shaari, Hassen Salhi, and A Hadj Arab. Methodology for predicting sequences of mean monthly clearness index and daily solar radiation data in remote areas : Application for sizing a stand-alone pv system. *Renewable Energy*, 33(7) :1570–1590, 2008.
- [23] Armel Oumbe. Exploitation des nouvelles capacités d’observation de la Terre pour évaluer le rayonnement solaire incident au sol. 11 2009.
- [24] Effet photoélectrique. <https://www.studysmarter.fr/resumes/physique-chimie/physique/effet-photoelectrique/>.
- [25] Lucien Wald. *Solar radiation energy (fundamentals)*. Le Centre pour la Communication Scientifique Directe, 1 2007.
- [26] Talend. Analyse prédictive : définition, enjeux, exemples. <https://www.talend.com/fr/resources/analyse-predictive/>.
- [27] La série chronologique statistique : cours avec exemples. <https://xn--apprendreconomie-jqb.com/serie-chronologique/>, 5 2022.
- [28] Shreya Dutta, Yan-Ling Li, Aditya Venkataraman, Luís Costa, Tianxiang Jiang, Robert Plana, Philippe Tordjman, Fook Hoong Choo, Chek Fok Foo, and Hans Björn Püttgen. Load and Renewable Energy Forecasting for a Micro-grid using Persistence Technique. *Energy Procedia*, 143 :617–622, 12 2017.
- [29] Ma Lei, Luan Shi-Yan, Jiang Chuanwen, Liu Hongling, and Zhang Yan. A review on the forecasting of wind speed and generated power. *Renewable Sustainable Energy Reviews*, 13(4) :915–920, 5 2009.
- [30] Benoit Cayla. Les chaînes de Markov. <https://datacorner.fr/markov/>, 11 2020.
- [31] Les techniques algorithmiques de l’IA | L’inférence bayésienne. <https://www.cyberjustice.ca/2020/12/16/les-techniques-algorithmiques-de-lia-linference-bayesienne/>, 12 2020.
- [32] ARMA Model - Qu’est-ce que c’est, définition et concept - 2021 - Économie-Wiki.com. = <https://economy-pedia.com/11031156-arma-model>.

- [33] Jason Brownlee. A Gentle Introduction to SARIMA for Time Series Forecasting in Python. <https://machinelearningmastery.com/sarima-for-time-series-forecasting-in-python/>, 8 2019.
- [34] Simplilearn. An Introduction to Exponential Smoothing for Time Series Forecasting in Python. <https://www.simplilearn.com/exponential-smoothing-for-time-series-forecasting-in-python->, 2 2023.
- [35] Analyse prédictive : présentation et atouts. [https://www.sas.com/fr\\_fr/insights/analytics/predictive-analytics.html](https://www.sas.com/fr_fr/insights/analytics/predictive-analytics.html).
- [36] Stéphane Caron. Une introduction aux arbres de décision. *Stéphane Caron*, 31, 2011.
- [37] Erwan Scornet. *Learning with random forests*. Theses, Université Pierre et Marie Curie - Paris VI, November 2015.
- [38] scikit-learn - sklearn.tree.ExtraTreeRegressor Un régresseur d'arbre extrêmement aléatoire. [https://runebook.dev/fr/docs/scikit\\_learn/modules/generated/sklearn.tree.extratreeRegressor](https://runebook.dev/fr/docs/scikit_learn/modules/generated/sklearn.tree.extratreeRegressor).
- [39] Basel Solaiman and R. Lepage. Les réseaux de neurones artificiels et leurs applications en imagerie et en vision par ordinateur. [https://www.researchgate.net/publication/326657820\\_Les\\_reseaux\\_de\\_neurones\\_artificiels\\_et\\_leurs\\_applications\\_en\\_imagerie\\_et\\_en\\_vision\\_par\\_ordinateur](https://www.researchgate.net/publication/326657820_Les_reseaux_de_neurones_artificiels_et_leurs_applications_en_imagerie_et_en_vision_par_ordinateur), 7 2018.
- [40] Réseau de neurones artificiels : quelles sont leurs capacités? <https://www.ionos.fr/digitalguide/web-marketing/search-engine-marketing/quest-ce-quun-reseau-neuronal-artificiel/>, 10 2020.
- [41] Réseau de neurones artificiels (artificial neural network) | CNIL. <https://www.cnil.fr/fr/definition/reseau-de-neurones-artificiels-artificial-neural-network>.
- [42] Bastien Botella. Introduction aux réseaux de neurones – 1/3 : Histoire. <https://blog.clevy.io/fr/introduction-aux-reseaux-de-neurones-1-3-histoire>, 9 2022.
- [43] Wikipedia contributors. Perceptron. <https://en.wikipedia.org/wiki/Perceptron>, 5 2023.

- [44] Sagar Sharma. What the Hell is Perceptron? - Towards Data Science. <https://towardsdatascience.com/what-the-hell-is-perceptron-626217814f53>, 10 2019.
- [45] GeeksforGeeks. Introduction to Recurrent Neural Network. <https://www.geeksforgeeks.org/introduction-to-recurrent-neural-network/>, 5 2023.
- [46] Moritz Lipperheide, Jaime Bosch, and Jan Kleissl. Embedded nowcasting method using cloud speed persistence for a photovoltaic power plant. *Solar Energy*, 112 :232–238, 2 2015.
- [47] Vincent P. A. Lonij, Adria E. Brooks, Alexander D. Cronin, Michael Leuthold, and Kevin M. Koch. Intra-hour forecasts of solar power production using measurements from a network of irradiance sensors. *Solar Energy*, 97 :58–66, 11 2013.
- [48] Yashwant Kashyap, Ankit Bansal, and Anil Kumar Sao. Solar radiation forecasting with multiple parameters neural networks. *Renewable Sustainable Energy Reviews*, 49 :825–835, 9 2015.
- [49] Xiangyun Qing and Yugang Niu. Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM. *Energy*, 148 :461–468, 4 2018.
- [50] Vishal Sharma, John Boland, Wilfred M. Walsh, and Thomas Reindl. Short term solar irradiance forecasting using a mixed wavelet neural network. *Renewable Energy*, 90 :481–492, 5 2016.
- [51] Federico-Vladimir Gutierrez-Corea, Miguel-Ángel Manso-Callejo, María Pilar Moreno-Regidor, and Maria Teresa Manrique-Sancho. Forecasting short-term solar irradiance based on artificial neural networks and data from neighboring meteorological stations. *Solar Energy*, 134 :119–131, 9 2016.
- [52] Hugo T.C. Pedro and Carlos E. A. Coimbra. Short-term irradiance forecastability for various solar micro-climates. *Solar Energy*, 122 :587–602, 12 2015.
- [53] Pamela Ramsami and Vishwamitra Oree. A hybrid method for forecasting the energy output of photovoltaic systems. *Energy Conversion and Management*, 95 :406–413, 2015.
- [54] Rui Zhang, Minwei Feng, Wei Zhang, Siyuan Lu, and Fei Wang. Forecast of Solar Energy Production - A Deep Learning Approach. *2018 IEEE International Conference on Big Knowledge (ICBK)*, 11 2018.

- [55] Ahmad Alzahrani, Pourya Shamsi, Cihan H. Dagli, and Mehdi Ferdowsi. Solar Irradiance Forecasting Using Deep Neural Networks. *Procedia Computer Science*, 114 :304–313, 11 2017.
- [56] L. Benali, Gilles Notton, Alexis Fouilloy, Yong Shuai, and Rabah Dizene. Solar radiation forecasting using artificial neural network and random forest methods : Application to normal beam, horizontal diffuse and global components. *Renewable Energy*, 132 :871–884, 3 2019.
- [57] Zibo Dong, John Boland, Thomas Reindl, and Wilfred M. Walsh. Short-term solar irradiance forecasting using exponential smoothing state space model. *Energy*, 55 :1104–1113, 6 2013.
- [58] John Boland, Vishal Sharma, Zhen Ye, Li Lim, Lu Zhao, and Aloysius W. Aryaputera. Forecasting of global horizontal irradiance by exponential smoothing, using decompositions. *Energy*, 81 :111–119, 3 2015.
- [59] Kacem Gairaa, Abdallah Khellaf, Youcef Messlem, and Farouk Chellali. Estimation of the daily global solar radiation based on Box–Jenkins and ANN models : A combined approach. *Renewable Sustainable Energy Reviews*, 57 :238–249, 5 2016.
- [60] Ji Wu, Chee Seng Chan, Yu Zhang, Bin Xiong, and Qinghai Zhang. Prediction of solar radiation with genetic approach combing multi-model framework. *Renewable Energy*, 66 :132–139, 6 2014.
- [61] Manal Marzouq, Hakim El Fadili, Khalid Zenkouar, Zakia Lakhliai, and Mohammed Amouzg. Short term solar irradiance forecasting via a novel evolutionary multi-model framework and performance assessment for sites with no solar irradiance data. *Renewable Energy*, 157 :214–231, 9 2020.
- [62] Anaconda. Anaconda | The World’s Most Popular Data Science Platform. <https://www.anaconda.com/>, 6 2023.
- [63] Project Jupyter. <https://jupyter.org/>.
- [64] Python ; : deacute ;fnition et utilisation de ce langage informatique. <https://www.journaldunet.fr/web-tech/dictionnaire-du-webmastering/1445304-python-definition-et-utilisation-de-ce-langage-informatique/>, 3 2020.
- [65] Welcome to Python.org. <https://www.python.org/>, 6 2023.

- [66] What is NumPy? — NumPy v1.24 Manual. <https://numpy.org/doc/stable/user/whatisnumpy.html>.
- [67] ActiveState. What Is Pandas in Python? Everything You Need to Know - ActiveState. <https://www.activestate.com/resources/quick-reads/what-is-pandas-in-python-everything-you-need-to-know/>, 8 2022.
- [68] Anthony Corbo. What Is TensorFlow? <https://builtin.com/data-science/tensorflow>, 12 2022.
- [69] Matplotlib : Tout savoir sur la bibliothèque Python de Dataviz. <https://datascientest.com/matplotlib-tout-savoir>, 2023.
- [70] Nik. Introduction to Scikit-Learn (sklearn) in Python. <https://datagy.io/python-scikit-learn-introduction/>, 4 2023.
- [71] Jason Brownlee. Introduction to Python Deep Learning with Keras. <https://machinelearningmastery.com/introduction-python-deep-learning-library-keras/>, 9 2019.
- [72] NASA POWER | Data Access Viewer. <https://power.larc.nasa.gov/data-access-viewer/>.
- [73] Yong Shuai, Christophe Paoli, Marc Muselli, and Marie Laure Nivet. Multi-horizon solar radiation forecasting for Mediterranean locations using time series models. *Renewable Sustainable Energy Reviews*, 28 :44–52, 12 2013.
- [74] Abdelatif Takilalte and Samia Harrouni. Daily Direct Normal Irradiance Forecasting by Support Vector Regression Case Study : in Ghardaia-Algeria. *2019 International Conference on Advanced Electrical Engineering (ICAEE)*, 11 2019.
- [75] Alexis Fouilloy, Yong Shuai, Gilles Notton, Fabrice Motte, Christophe Paoli, Marie Laure Nivet, Emmanuel Guillot, and Jean-Laurent Duchaud. Solar irradiation prediction with machine learning : Forecasting models selection method depending on weather variability. *Energy*, 165 :620–629, 12 2018.
- [76] Ahmed F. Mashaly, Abdulrahman Ali Alazba, Alhussein M. Al-Awaadh, and Mohamed A. Mattar. Predictive model for assessing and optimizing solar still performance using artificial neural network under hyper arid environment. *Solar Energy*, 118 :41–58, 8 2015.

- [77] Yuehui Huang, Jing Lu, Chun Liu, Xiaoyan Xu, Weisheng Wang, and Xiao-Xin Zhou. Comparative study of power forecasting methods for PV stations. *IEEE International Conference on Power System Technology*, 12 2010.
- [78] Jin-Song Zhang, Rodrigo Verschae, Shohei Nobuhara, and Jean-François Lalonde. Deep photovoltaic nowcasting. *Solar Energy*, 176 :267–276, 12 2018.
- [79] Bastien L. Reinforcement Learning : qu'est-ce que l'apprentissage par renforcement? <https://www.lebigdata.fr/reinforcement-learning-definition>, 6 2021.
- [80] DataCamp Team. Gradient Descent Tutorial. <https://www.datacamp.com/tutorial/tutorial-gradient-descent>, 3 2022.
- [81] Guillaume Saint-Cirgue. Descente de Gradient 8211 ; Gradient Descent. <https://machinelearningia.com/descente-de-gradient/>, 2 2020.