

THÈSE

en vue de l'obtention du : **DOCTORAT**

Centre de Recherche : Rabat Information Technology Center (RITC).
Structure de Recherche : Laboratoire de Recherche en Informatique et Télécommunications
Discipline : Sciences de l'ingénieur
Spécialité : Informatique et Télécommunications

Présentée et Soutenue le : 02/03/2024

par :

Mohammed SABRI

Modèles d'apprentissage profond pour la prédiction précise de la production d'énergie solaire photovoltaïque

Devant le JURY :

Mohamed EL HAZITI	PES	Ecole Supérieure de Technologie-Salé, Université Mohammed V	Président
Ahmed Drissi EL MALIANI	PH	Faculté des Sciences- Rabat, Université Mohammed V	Examinateur/Rapporteur
Mohamed EL AROUSSI	PH	Ecole Hassania Des Travaux Publics-Casablanca	Examinateur/Rapporteur
Khalid MINAOUI	PES	Faculté des Sciences-Rabat, Université Mohammed V	Examinateur/Rapporteur
Khadija RHOULAMI	PH	Faculté des Lettres et des Sciences Humaines Rabat, Université Mohammed V	Examinatrice
Mohammed EL HASSOUNI	PES	Faculté des Lettres et des Sciences Humaines Rabat, Université Mohammed V	Directeur de thèse

Année Universitaire : 2023 - 24

Dédicace

À mon directeur de thèse, Mohamed EL Hassouni pour son expertise, sa patience et sa guidance tout au long de cette aventure intellectuelle. Ses conseils éclairés ont façonné ma réflexion et ont enrichi ce travail de recherche.

À tous mes amis, famille et collègues qui ont partagé cette route avec moi, pour les échanges stimulants, les encouragements mutuels et les moments de détente qui ont allégé le fardeau des heures passées devant les écrans et les livres.

Enfin, à mes parents et à toutes les personnes qui croient en moi et en mon potentiel, cette thèse est dédiée à vous. Vos encouragements et votre soutien ont été les piliers sur lesquels je me suis appuyé pour atteindre ce but. Puissent les fruits de ce travail contribuer modestement à l'avancement de la connaissance dans notre domaine.

Remerciements

Cette thèse de doctorat a été menée pour l'obtention du grade de Docteur de l'Université Mohammed V - Rabat. Les travaux présentés dans ce mémoire ont été effectués au Laboratoire de Recherche en Informatique et Télécommunications (LRIT) de la Faculté des Sciences de Rabat (FSR) au Maroc. Je tiens à remercier tous les professeurs et membres du personnel pour leurs contributions remarquables à la création d'un cadre de recherche exceptionnel.

En premier lieu, je tiens à remercier mon directeur de thèse Professeur **Mohammed EL HAS-SOUNI**, Professeur d'Enseignement Supérieur à la faculté des lettres et sciences humaines de Rabat. Ses précieux conseils ont été déterminants pour la réalisation de cette thèse. En plus de ses compétences scientifiques indéniables, il se distingue surtout par sa générosité, ce qui rend les discussions, les échanges d'idées et les moments de détente en sa compagnie extrêmement plaisants. Je suis confiant que ces remarquables qualités continueront à exercer une influence durable sur moi.

Je tiens à remercier Professeur **Mohamed EL HAZITI**, Professeur d'Enseignement Supérieur à l'école supérieure de technologie de Salé, d'avoir accepté de présider le jury de ma thèse.

Je tiens aussi à remercier Professeur **Ahmed Drissi EL MALIANI**, Professeur Habilité à la faculté des sciences de Rabat, d'avoir accepté de juger la qualité de mon travail en tant que rapporteur.

Je voudrais également remercier Professeur **Mohamed EL AROUSSI**, Professeur Habilité à l'école hassania des travaux publics de Casablanca, d'avoir accepté de rapporter ce mémoire de thèse.

Je voudrais également remercier Professeur **Khalid MINAOUI**, Professeur d'Enseignement Supérieur à la faculté des Sciences de Rabat, qui m'a fait l'honneur de rapporter cette thèse.

Mes remerciements vont également à Professeur **Khadija RHOULAMI**, Professeur Habilité à la faculté des lettres et sciences humaines de Rabat, qui m'a fait l'honneur d'examiner cette thèse.

Table des matières

Dédicace	ii
Remerciements	iv
Résumé	xi
Abstract	xiii
Liste des tableaux	xv
Liste des figures	xviii
Liste d'Acronymes	xix
1 Introduction	1
1.1 Contexte générale	2
1.2 La problématique liée à l'utilisation à grande échelle de l'énergie solaire	4
1.3 Intérêt de la prévision photovoltaïque pour le réseau électrique	5
1.4 Objectifs et démarche de la thèse	6
1.5 Structure du manuscrit	7
2 Etat de l'art de la prévision de la production d'énergie solaire photovoltaïque	11
2.1 Introduction	12

TABLE DES MATIÈRES

2.2	Classification des prévisions de puissance photovoltaïque basée sur l'horizon de prévision	13
2.3	Classification des techniques de prévision de puissance photovoltaïque	14
2.3.1	Méthode physique	15
2.3.2	Méthode statistique	16
2.3.3	Méthode intelligences artificielle	16
2.4	Les réseaux de Neurones.	18
2.5	Différentes familles de réseaux de neurones	21
2.5.1	Réseaux de neurones convolutionnels	21
2.5.2	Résultats bibliographiques obtenus par application de réseaux de neurones convolutionnels pour la prévision dans le domaine PV.	22
2.5.3	Réseaux de neurones récurrents	24
2.5.4	Résultats bibliographiques obtenus par application de réseaux de neurones récurrents pour la prévision dans le domaine PV.	25
2.5.5	Le réseau récurrent à mémoire court-terme et long terme ou LSTM	27
2.5.6	Résultats bibliographiques obtenus par application de réseau récurrent à mémoire court-terme et long terme pour la prévision dans le domaine PV.	29
2.5.7	Les réseaux de neurones récurrents bidirectionnels à mémoire court terme et long-terme ou BLSTM	32
2.5.8	Résultats bibliographiques obtenus par application de réseaux de neurones récurrents bidirectionnels à mémoire court terme et long-terme pour la prévision dans le domaine PV.	33
2.5.9	Unités récurrentes avec portes ou GRU	34
2.5.10	Résultats bibliographiques obtenus par application des Unités récurrentes avec portes pour la prévision dans le domaine PV.	37
2.5.11	Autoencodeurs	38
2.5.12	Résultats bibliographiques obtenus par application de autoencodeur pour la prévision dans le domaine PV.	42

TABLE DES MATIÈRES

2.6	Métriques pour l'évaluation des modèles	43
2.7	conclusion	45
3	Exploration Complète du Site : Caractérisation, Acquisition, Sélection et Analyse des Données de Surveillance	47
3.1	Introduction	48
3.2	Analyse des performances pour optimiser les systèmes photovoltaïques	49
3.3	Production photovoltaïques	50
3.4	Caractérisation du site d'étude	54
3.5	Acquisition de données de surveillance	55
3.5.1	Imputation de l'irradiance en plan	58
3.5.2	Analyse des données de température	60
3.6	Sélection des données d'entrée	62
3.7	Présentation des données et analyse préliminaire	62
3.7.1	Collecte et préparation des données	62
3.7.2	Corrélation.	64
3.8	Prétraitement des données	69
3.8.1	Nettoyage de données.	69
3.8.2	Gestion des données manquantes.	69
3.8.3	Normalisation.	70
3.9	Conclusion	70
4	Modèles d'apprentissage profond pour une prédiction de la puissance Photovoltaïques	72
4.1	Introduction	73
4.2	Étude comparative des modèles LSTM et RNN pour la prévision de la puissance PV .	74
4.2.1	Collecte des données de la puissance PV	74
4.2.2	Paramètres et configurations des modèles	75

TABLE DES MATIÈRES

4.2.3	Comparaison des performances des modèles LSTM et RNN	75
4.2.4	Analyse des résultats et interprétation	78
4.3	Modèle hybride GRU-CNN pour la prédiction de la puissance PV	79
4.3.1	Description du modèle GRU-CNN proposé	79
4.3.2	Période des données de puissance PV	80
4.3.3	Configuration du modèle GRU-CNN	81
4.3.4	Présentation des résultats expérimentaux obtenus	82
4.3.5	Analyse des performances du modèle GRU-CNN	83
4.4	Prévision de la production d'énergie PV à l'aide d'un modèle hybride CNN-GRU . . .	83
4.4.1	Présentation du modèle hybride d'apprentissage en profondeur CNN-GRU . . .	84
4.4.2	Période de collecte de données sur la Puissance PV	85
4.4.3	Entraînement du CNN et du GRU sur les données collectées	85
4.4.4	Exposition des résultats expérimentaux obtenus	86
4.4.5	Évaluation des performances du modèle hybride	89
4.4.6	Comparaison avec d'autres approches de prévision de la production d'énergie PV	91
4.5	Conclusion	94
5	Amélioration de la précision de la prédiction de la puissance PV grâce à des modèles d'apprentissage profond avancés	96
5.1	Introduction	97
5.2	Prédiction de la génération d'énergie PV à l'aide d'un réseau double couche à mémoire à court terme récurrente bidirectionnelle-convolutionnelle	98
5.2.1	Description de l'architecture double-couche du réseau BLSTM-CNN	98
5.2.2	Description des données utilisées	99
5.2.3	Configuration des paramètres expérimentaux	100
5.2.4	Protocole d'évaluation de la performance	101

TABLE DES MATIÈRES

5.2.5	Analyse des résultats en fonction de différentes périodes	103
5.2.6	Comparaison des performances du modèle proposé avec d'autres approches . .	105
5.2.7	Discussion sur les avantages et les limites des modèles	108
5.3	Prévision de la production d'énergie PV à l'aide de réseaux autoencodeurs à mémoire à court terme	109
5.3.1	Rassemblement des données PV	110
5.3.2	Architecture et paramètres des modèles de prévision	110
5.3.3	Métriques d'évaluation de la performance	112
5.3.4	Comparaison des performances des modèles de prévision	116
5.3.5	Évaluation comparative de l'approche LSTM-AE par rapport à d'autres méthodes	117
5.4	Conclusion	119
6	Conclusions générales et perspectives	121
6.1	Conclusions générales	122
6.2	Perspectives	124
	Publications	125
	Bibliographie	126

Résumé

Cette thèse de doctorat se concentre sur la prédiction précise de la production d'énergie solaire photovoltaïque (PV) et son importance pour le fonctionnement sécurisé et efficace des systèmes d'énergie PV. Dans de nombreux pays développés, l'énergie solaire PV est considérée comme la source d'énergie renouvelable la plus rentable et représente une part importante de la production d'électricité. Cependant, la génération d'énergie PV présente une variabilité élevée et imprévisible en raison de facteurs météorologiques. Afin de réduire l'influence négative de l'utilisation de l'énergie solaire PV, la prédiction précise de la production d'énergie PV revêt une importance cruciale. Les variations élevées de la production d'énergie PV peuvent être causées par des éléments météorologiques tels que l'ensoleillement, la température et les nuages. En prédisant avec précision la production d'énergie PV, il est possible d'atténuer ces variations et d'assurer un fonctionnement sécurisé et efficace des systèmes d'énergie PV. Dans cette étude, les données de production d'énergie PV ont été collectées à Alice Springs, en Australie. Ces données ont été utilisées pour développer des modèles d'apprentissage profond afin de prédire la production d'énergie PV. Les modèles d'apprentissage profond sont des outils puissants pour capturer les relations complexes entre les variables météorologiques et la production d'énergie PV. Les résultats de cette recherche ont démontré que l'utilisation de modèles d'apprentissage profond permet une prédiction précise de la production d'énergie PV. Cette prédiction précise est essentielle pour le fonctionnement sécurisé et efficace des systèmes d'énergie PV, en permettant une meilleure planification et une gestion optimisée de l'énergie produite.

Mots-clés : prédiction, production d'énergie solaire photovoltaïque, efficacité, facteurs météorologiques, modèles d'apprentissage profond.

Abstract

This doctoral thesis focuses on the accurate prediction of photovoltaic solar power production and its importance for the secure and efficient operation of photovoltaic energy systems. In many developed countries, photovoltaic solar power is considered the most cost-effective renewable energy source and constitutes a significant portion of electricity production. However, the generation of photovoltaic power is characterized by high variability and unpredictability due to meteorological factors. To mitigate the negative impact of this variability, accurate prediction of photovoltaic power production is crucial. The fluctuations in photovoltaic power output can be attributed to meteorological elements such as sunlight, temperature, and clouds. By accurately predicting photovoltaic power production, it is possible to reduce these variations and ensure secure and efficient operation of photovoltaic energy systems. In this study, photovoltaic power data was collected in Alice Springs, Australia. Deep learning models were utilized to predict photovoltaic power production. Deep learning models are powerful tools for capturing the complex relationships between meteorological variables and photovoltaic power production. The findings of this research demonstrate that the use of deep learning models enables accurate prediction of photovoltaic power production. This precise prediction is essential for the secure and efficient operation of photovoltaic energy systems, enabling better planning and optimized management of generated energy.

Keywords : Prediction, photovoltaic solar energy production, efficiency, meteorological factors, deep learning models.

Liste des tableaux

3.1	Paramètre de surveillance mesuré divisé par paramètre météorologique et électrique [1].	56
3.2	Les informations spécifiques liées à ce système (1B) [2].	63
3.3	Coefficient de corrélation de Pearson entre la puissance PV et le prédicteur.	68
4.1	Spécification des paramètres pour l'entraînement.	75
4.2	Les résultats du modèle de prévision.	76
4.3	Les paramètres définis dans cette étude.	80
4.4	Erreur des différents modèles.	81
4.5	Les réglages des paramètres du modèle CNN-GRU dans cette étude.	87
4.6	Réglages des paramètres de la méthode de référence.	87
4.7	L'exactitude annuelle de prévision des modèles étudiés.	88
4.8	Valeurs statistiques des données expérimentales de puissance PV (KW).	90
4.9	L'exactitude de prévision saisonnière des modèles étudiés.	91
5.1	Réglages des paramètres de la méthode proposée.	101
5.2	Réglages des paramètres de la méthode proposée.	102
5.3	Valeurs statistiques des données expérimentales de puissance PV (KW).	104
5.4	Performance de prévision des différentes méthodes d'apprentissage profond pour chaque saison.	105
5.5	Descriptions de l'ensemble d'entraînement et de l'ensemble de test pour chaque période.	110

LISTE DES TABLEAUX

5.6	Réglages des paramètres de la méthode proposée dans cette étude.	111
5.7	Réglages des paramètres des modèles de référence simples dans cette étude.	111
5.8	Les résultats de prévision pour différentes périodes et différents modèles.	113
5.9	Pourcentage de promotion de la MAE, MSE et RMSE entre le modèle LSTM-AE et le modèle de réseau simple.. . . .	114
5.10	Le temps d'entraînement (en secondes) des différents modèles.	116
5.11	Réglages des paramètres des modèles hybrides de référence dans cette étude.	117
5.12	Performance de prévision des différentes méthodes d'apprentissage profond pour chaque saison.	118
5.13	Comparaison du modèle LSTM-AE proposé avec les modèles récentes.	119

Table des figures

1.1	Prévision de consommation mondiale d'énergie, en Mtep	2
2.1	Représentation d'un neurone artificiel.	18
2.2	Architecture de Réseaux de Neurones Convolutifs.	22
2.3	Processus de convolution et de mise en commun maximale.	23
2.4	Vue d'ensemble générale des réseaux neuronaux récurrents.	25
2.5	L'architecture des réseaux LSTM	27
2.6	Structure du réseau BLSTM.	33
2.7	La structure interne du GRU	36
2.8	Architecture du réseau autoencodeur.	39
2.9	Réseau LSTM-Autoencodeur	41
3.1	Représentation des différents rayonnements solaires terrestres reçus sur une installation donnée [3]	51
3.2	Carte de l'irradiation globale horizontale en Australie. Source : Solargis. Lecture : Répartition de la moyenne d'irradiation globale horizontale dans l'Australie (moyenne calculée sur la période 2007-2018)	52
3.3	L'inclinaison et l'orientation des panneaux d'un producteur PV (adapté de [4])	53
3.4	Visualisation de la carte du système d'alimentation DKASC.	63
3.5	(a) Diagramme de dispersion de la moyenne de phase de courant et de la puissance PV"; (b) Courbe de corrélation entre la moyenne de phase de courant et la puissance PV.	65

TABLE DES FIGURES

3.6	(a) Diagramme de dispersion du rayonnement global horizontal et de la puissance PV ; (b) Courbe de corrélation entre le rayonnement global horizontal et la puissance PV. . .	66
3.7	(a) Diagramme de dispersion du rayonnement horizontal diffus et de la puissance pv ; (b) Courbe de corrélation entre le rayonnement horizontal diffus et la puissance PV. . .	66
3.8	(a) Diagramme de dispersion de la température météorologique en degrés Celsius et de la puissance PV ; (b) Courbe de corrélation entre la température météorologique en degrés Celsius et la puissance PV.	67
3.9	(a) Diagramme de dispersion de l'humidité relative météorologique et de la puissance PV ; (b) Courbe de corrélation entre l'humidité relative météorologique et la puissance PV".	68
3.10	(a) Diagramme de dispersion de la direction du vent et de la puissance PV ; (b) Courbe de corrélation entre la direction du vent et la puissance PV.	69
4.1	Les résultats de prévision de deux modèles et la puissance PV réelle en hiver.	77
4.2	Les résultats de prévision de deux modèles et la puissance PV réelle au printemps. . .	77
4.3	Les résultats de prévision de deux modèles et la puissance PV réelle en été.	78
4.4	Les résultats de prévision de deux modèles et la puissance PV réelle en automne. . . .	78
4.5	Prévision de la puissance PV en utilisant toutes les méthodes dans la séquence d'entrée du test 3.	83
4.6	Comparaison des valeurs de MAE, MSE, RMSE et MBE sous forme de graphique en barres pour différents modèles.	88
4.7	Comparaison de la moyenne de R^2 entre la prévision et la mesure de la puissance PV en utilisant divers modèles.	89
4.8	Les résultats de prévision de la puissance PV en utilisant toutes les méthodes en automne.	91
4.9	Les résultats de prévision de la puissance PV en utilisant toutes les méthodes en hiver.	92
4.10	Les résultats de prévision de la puissance PV en utilisant toutes les méthodes au printemps.	92
4.11	Les résultats de prévision de la puissance PV en utilisant toutes les méthodes en été. .	93
4.12	Comparaison de modèle proposée avec d'autres études	94

TABLE DES FIGURES

5.1	Comparaison entre les prédictions du modèle et les valeurs réelles.	103
5.2	Erreur entre les valeurs prévues et les valeurs réelles.	103
5.3	Comparaison du modèle hybride proposé BLSTM-CNN avec des études précédentes.. .	108
5.4	Prévision de la puissance PV en utilisant toutes les méthodes pendant la période 1. . .	114
5.5	Prévision de la puissance PV en utilisant toutes les méthodes pendant la période 2. . .	115
5.6	Prévision de la puissance PV en utilisant toutes les méthodes pendant la période 3. . .	115
5.7	Prévision de la puissance PV en utilisant toutes les méthodes pendant la période 4. . .	116

Liste d'Acronymes

- AE** Autoencodeur
- AIE** Agence Internationale de L'énergie
- ALO** Algorithme d'Optimisation des Lions Fourmis
- ANFIS** Système d'Inférence Neuro-Flou Adaptatif
- ABPNN** Réseau de Neurones Artificiel à Rétropropagation
- ARMA** Modèles Autorégressifs et à Moyenne Mobile
- BPNN** Réseau Neuronal à propagation Arrière
- BLSTM** Réseau de Neurones à Mémoire à Long Terme Bidirectionnel
- CNN** Réseau Neuronal Convolutif
- DL** Apprentissage Profond
- DKASC** Centre de Connaissances du Désert en Australie Centrale
- ECMWF** Centre Européen pour les Prévisions Météorologiques à Moyen Terme
- EEMD** Décomposition en Modes Empiriques par Ensemble
- ENR** Energie Renouvelable
- ELM** Machine d'Apprentissage Extrême
- GA** Algorithme Génétique
- GNL** Gaz Naturel Liquéfié
- GRU** Unité Récurrente à Portes
- GHD** Rayonnement Horizontal Diffus
- GHI** Rayonnement Horizontal Global
- GCN** Réseau Convolutif sur Graphe
- HMM** Modèle de Markov Caché

LISTE D'ACRONYMES

NOAA Administration Nationale des Océans et de L'atmosphère
NASA Administration Nationale de L'espace et de L'aéronautique
IA Intelligence Artificielle
IMVO Optimisation Multi-verse
JMA Agence météorologique du Japon
Mtep Million de Tonnes équivalent Pétrole
MAPE Erreur Moyenne Absolue de Pourcentage
LSSVM Machine à Vecteurs de Support à Moindres Carrés
LSTM Réseau de Mémoire à Long Terme à Court Terme
MLP Réseau de Perception Multicouche
MSE Erreur quadratique moyenne
MAE Erreur Absolue Moyenne
MBE Erreur moyenne de biais
PIB Produit Intérieur Brut
PV photovoltaïque
RNN Réseau Neuronal Récurrent
SVM Machines à Vecteurs de Support
ReLU Unité Linéaire Rectifiée
RMSE Erreur quadratique moyenne
TCN convolution temporelle
WPD Décomposition en Paquets Ondelettes

Chapitre 1

Introduction

1.1 Contexte générale

Depuis les années 1970, la consommation mondiale d'énergie n'a cessé d'augmenter. Celle-ci a presque triplé en 50 ans pour atteindre environ 600 exajoules en 2019 [5]. La consommation mondiale d'énergie primaire est passée à 595,15 exajoules en 2021. Elle a augmenté de 5,8% par rapport à 2020 et se situe déjà à un niveau supérieur à la consommation d'avant la pandémie de Covid-19 [6], cette tendance à long terme devrait se poursuivre. Compte tenu des besoins énergétiques croissants des pays en développement et de la croissance démographique mondiale. Plus de 80% de la consommation mondiale d'énergie primaire provient désormais des énergies fossiles, dont 32% du pétrole. La prévision de consommation mondiale d'énergie est illustrée à la figure.1.1, ainsi que les différentes sources d'énergie primaire qui ont été utilisées pour répondre à la demande.

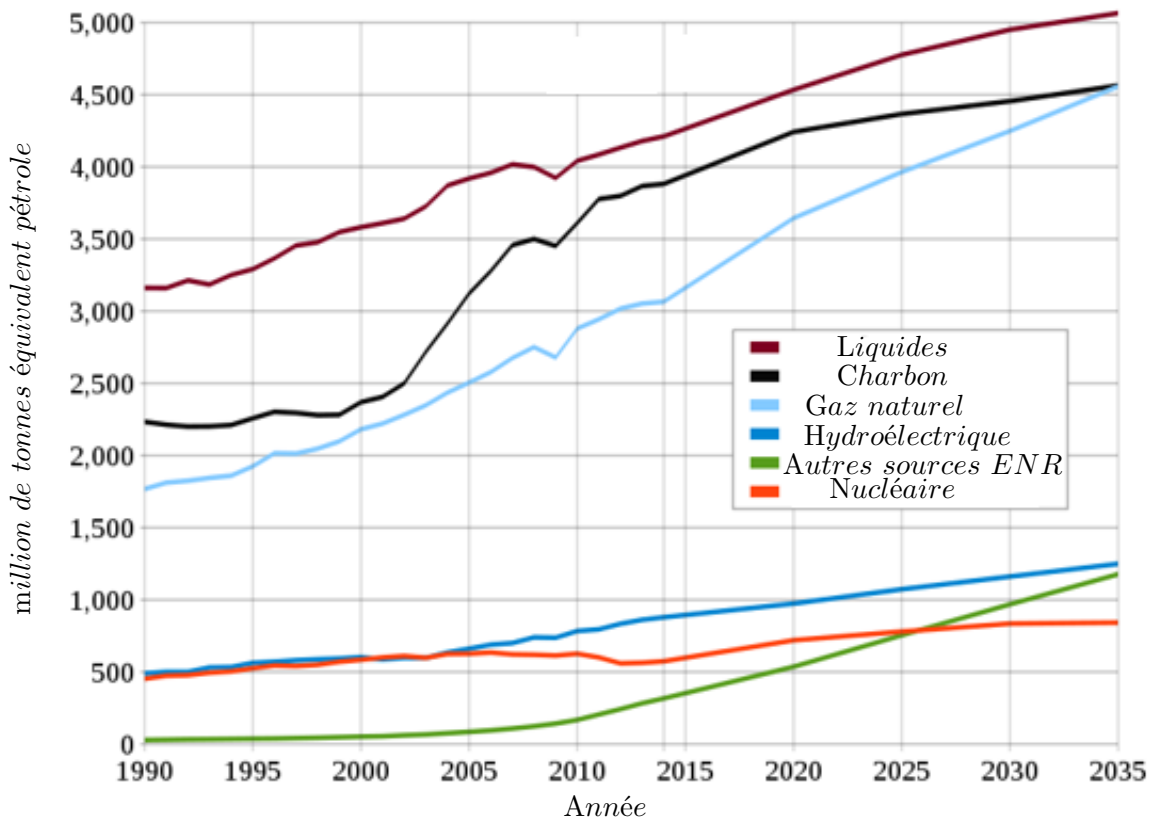


FIGURE 1.1 – Prévision de consommation mondiale d'énergie, en Mtep

Les chocs pétroliers de 1973 et 1979 ont montré à quel point la dépendance énergétique à l'échelle nationale est cruciale dans certains pays. Ces menaces de perturbation d'approvisionnement énergé-

1.1. CONTEXTE GÉNÉRALE

tique et l'impact de la volatilité des prix sur les balances des paiements ont incité les gouvernements à développer des politiques énergétiques et à proposer des stratégies à moyen et long terme pour le développement du secteur.

D'autre part, l'accélération du réchauffement climatique est devenue une préoccupation politique majeure au cours des dernières décennies. La Convention des Nations Unies sur les changements climatiques, qui devient le premier accord international traitant de ce sujet, est signée en 1992, clôturant le Sommet de la Terre à Rio de Janeiro. De nombreux autres sommets internationaux suivront, l'objectif principal étant de réduire les causes humaines du réchauffement climatique.

En décembre 2015, les représentants de 196 pays se sont réunis pour la 21^e Conférence des partis (COP) à Paris. L'objectif fondamental de l'Accord de Paris qui en résulte est de maintenir l'augmentation de la température mondiale nettement en dessous de 2°C d'ici l'an 2100 par rapport aux niveaux préindustriels. Les investissements dans la lutte contre le changement climatique, dont le coût moyen pourrait être limité à 1% du PIB mondial annuel, sont encouragés [7].

Il est nécessaire de modifier les modes de consommation et de se tourner vers l'utilisation d'énergies de substitution afin de réduire l'utilisation de ces énergies fossiles pour la production d'électricité. Parmi celles-ci, plusieurs ont un potentiel d'utilisation important mais ne sont plus utilisées qu'avec parcimonie. On peut citer l'énergie nucléaire ou les énergies renouvelables comme (géothermique, solaire, éolienne, hydraulique, etc.). L'utilisation massive de l'énergie nucléaire crée d'importants problèmes de gestion et de stockage des déchets. Être dépendant des approvisionnements en uranium, qui sont nécessaires pour utiliser l'énergie nucléaire, peut être problématique pour une région. C'est particulièrement sensible pour le secteur de l'énergie. De ce fait, l'Accord de Paris a été décliné en plusieurs objectifs, dont l'un est d'augmenter la part des énergies renouvelables dans la consommation finale d'énergie à 32% d'ici 2030 [8].

Alors que les pays de la région occidentaux recherchent des ressources pour remplacer le gaz russe, il est important de rappeler que l'Australie est le premier exportateur mondial de carbone, est également le premier exportateur mondial de gaz naturel liquéfié (GNL) [9]. De nombreuses critiques du rôle que joue le carbone dans le système énergétique australien se concentrent sur les importantes émissions de gaz à effet de serre produites lors de sa combustion. Malgré une baisse de 16% de la consommation nationale de carbone dans le secteur électrique entre 2016 et 2020, cette énergie représentait encore près de 54 % de la production électrique australienne en 2020 (contre 63% en 2015). L'énergie solaire

1.2. LA PROBLÉMATIQUE LIÉE À L'UTILISATION À GRANDE ÉCHELLE DE L'ÉNERGIE SOLAIRE

est désormais la première source renouvelable de production d'électricité du pays (près de 9% de la production électrique totale du pays en 2020) [9].

L'utilisation de l'énergie solaire et éolienne présente des défis, notamment ceux liés à la nature imprévisible des processus impliqués dans les conditions météorologiques. Le fait que ces sources d'énergie soient intermittentes pose un défi important pour la gestion du réseau électrique du pays, où le maintien d'un équilibre entre la consommation et la production d'électricité est essentiel. Cette intermittence pose également des problèmes avec la gestion indépendante des parcs éoliens ou solaires, qui peuvent ensuite être couplés à des installations de stockage ou de production supplémentaires pour fournir une puissance moins erratique et imprévisible. Dans cette thèse, l'accent est spécifiquement mis sur la production d'électricité résultant directement de la conversion de l'énergie solaire par effet PV.

1.2 La problématique liée à l'utilisation à grande échelle de l'énergie solaire

Le monde en est venu à croire que les sources d'énergie renouvelables étaient le seul substitut à l'énergie traditionnelle au cours des dernières années [10, 11, 12, 13]. Dans ce contexte, nombreux pays se concentrent sur le développement et l'utilisation de source d'énergie renouvelables pour remplacer partiellement les sources d'énergie traditionnelle pour une variété de raisons, notamment le fait qu'elles sont inépuisables, propres, et respectueuses de l'environnement [10, 14].

D'autre part, l'énergie solaire n'est généralement pas en phase avec ce dont les consommateurs ont besoin en termes d'énergie ; il peut être présent lorsque les besoins sont réduits ou au contraire absent lors des demandes d'énergie du réseau. Ainsi, on qualifie l'énergie solaire comme une énergie intermittente fatale, ce qui signifie que sa disponibilité est déterminée par la valeur actuelle de la ressource et ne peut être contrôlée. De plus, cette ressource présente une variabilité spatio-temporelle importante. Par exemple, des changements importants se produisent dans les réseaux électriques. Le recours croissant aux énergies renouvelables, notamment celles dont la production est sporadique et difficilement maîtrisable, suscite des inquiétudes quant à la sécurité de l'approvisionnement en électricité.

Le déploiement du PV à de nouveaux défis en raison des effets notables des systèmes PV distribués sur les opérations techniques et non techniques des réseaux électriques distants conventionnels. À mesure que les niveaux de pénétration du PV augmentent, des problèmes techniques importants sont

1.3. INTÉRÊT DE LA PRÉVISION PHOTOVOLTAÏQUE POUR LE RÉSEAU ÉLECTRIQUE

observés dans ces réseaux, notamment une diminution de la qualité de l'énergie, une complexité accrue du réseau dans l'exploitation et la gestion, etc. Les niveaux de pénétration du PV dans divers réseaux distants à travers le monde sont fortement influencés par le contrôle adopté par les services publics et leur stratégies opérationnelles [15].

L'électricité circule généralement des générateurs centralisés vers les sous-stations, puis vers les consommateurs, dans la plupart des systèmes de services publics d'électricité. Avec la production d'énergie solaire, l'énergie peut circuler dans les deux sens. Cependant, la majorité des systèmes de distribution électrique n'ont pas été créés pour prendre en charge un flux d'énergie dans les deux sens. Si la charge et la production PV ne sont pas correctement adaptées, même de petites quantités de PV peuvent avoir une influence sur les paramètres du système pour les circuits d'alimentation de distribution qui sont longs et desservent des zones rurales ou en développement [16]. La possibilité de dommages au réseau de distribution et d'effets sur les autres clients du service public desservis par le même circuit de distribution augmente lorsque la production PV dépasse la demande d'énergie locale [16]. Les lignes de transmission sont souvent nécessaires pour les projets ou les fermes PV à grande échelle, car la majorité d'entre elles sont situées loin des zones peuplées et doivent transporter l'électricité sur de grandes distances avant de pouvoir être utilisées. Cela nécessite des dépenses financières plus importantes dans la construction des lignes de transmission et conduit fréquemment à des "pertes de ligne" car une partie de l'énergie pendant le transport est convertie en chaleur et perdue.

1.3 Intérêt de la prévision photovoltaïque pour le réseau électrique

La quantité d'insolation, ou d'ensoleillement, à un endroit et à un moment donné détermine la quantité d'électricité produite par les systèmes PV. La surproduction et la sous-production peuvent entraîner une instabilité du réseau. La communauté scientifique a proposé différentes solutions pouvant se combiner par exemple utiliser de meilleurs outils de prévision pour permettre des prédictions plus précises du moment où la production solaire pourrait diminuer jusqu'à la capacité de pénétration minimale. L'installation d'énergie solaire sur une vaste zone géographique pour minimiser les effets de la fluctuation de la production provoquée par la couverture nuageuse locale. Periz et al. [17, 18] ont montré que l'utilisation d'une technique de prévision ferme mais à plus grande échelle était le moyen le moins coûteux de convertir le PV d'intermittent en effectivement dispatchable. Les besoins de stockage peuvent être considérablement réduits par la construction excessive d'installations PV et une réduction

proactive pour parvenir à une production d'énergie renouvelable ferme et économiquement viable. L'application du stockage seul sans sur construction et en déversant de manière proactive l'excès de PV serait d'un coût prohibitif. Les prévisions solaires font désormais partie intégrante des opérations des services publics, car la production d'énergie solaire - principalement PV- pénètre dans les réseaux électriques. Les modèles sous-jacents à ces prévisions s'améliorent [19, 20]. La raison sous-jacente des prévisions solaires est la nature intermittente de la ressource solaire non dispatchable. Anticiper avec précision la production solaire future peut minimiser les déséquilibres de charge, d'où la taille des marges de réserve et les pics de prix de l'électricité sur le marché au comptant. Les pénalités qui sont souvent imposées aux grands producteurs solaires visent à refléter le coût de ces déséquilibres de charge induits par l'énergie solaire. Les méthodes de prévision pour prévoir la production PV avec précision sont devenues un outil important pour résoudre la planification PV et problèmes de modélisation, qui peuvent atténuer les impacts négatifs sur l'ensemble du système électrique et améliorer la stabilité du système [21]. Avec le développement de compteurs d'électricité avancés dans le réseau électrique actuel, des données sources plus riches peuvent être utilisées pour construire des modèles de prévision plus sophistiqués afin d'obtenir des prévisions de puissance PV plus précises [22]. Le développement de méthodes capables de fournir une prévision précise de la puissance PV produite conduirait à une gestion plus efficace de l'énergie et à une fiabilité accrue du système lui-même.

1.4 Objectifs et démarche de la thèse

Cette thèse vise à développer des méthodes de prévision de la puissance PV en utilisant l'apprentissage profond, avec pour objectif d'améliorer la précision des prédictions et de faciliter l'intégration efficace de l'énergie solaire PV dans les systèmes d'alimentation électrique. Pour atteindre cet objectif, la démarche de cette thèse sera structurée en plusieurs étapes. Tout d'abord, une revue approfondie de la littérature sera réalisée pour examiner les travaux existants sur la prévision de la puissance PV en utilisant l'apprentissage profond. Cette étape permettra de comprendre les différentes architectures de réseaux de neurones profonds, les techniques d'entraînement et les approches de prétraitement des données qui ont été utilisées dans ce domaine. Cette étude utilise les données PV du Desert Knowledge Precinct (DKASC), un système PV situé à Alice Springs en Australie centrale, comme étude de cas. Des ensembles de données historiques de puissance PV seront collectés et préparés. Ces ensembles de données comprendront des mesures de puissance réelle ainsi que des données météorologiques asso-

ciées. Des étapes de prétraitement seront effectuées pour normaliser les données, traiter les valeurs manquantes et rendre les données prêtes à être utilisées dans les modèles d'apprentissage profond. La prochaine étape consistera à concevoir et construire un modèle de prédiction de la puissance PV en utilisant une architecture de réseau de neurones profonds adaptée. Différentes couches, fonctions d'activation et techniques seront expérimentées pour obtenir un modèle performant. Le modèle sera ensuite entraîné en utilisant les ensembles de données préparés, et les hyperparamètres seront ajustés pour améliorer les performances du modèle. Une fois le modèle entraîné, il sera validé et évalué en utilisant des ensembles de données de test indépendants. Les performances du modèle seront évaluées en utilisant des métriques telles que l'erreur absolue moyenne, l'erreur quadratique moyenne, la corrélation, etc. Les résultats obtenus seront comparés à ceux d'autres approches de prévision existantes afin d'évaluer l'efficacité de l'apprentissage profond dans la prédiction de la puissance PV. Une fois les résultats obtenus, une analyse approfondie sera effectuée pour interpréter les facteurs influençant la puissance PV. Les relations complexes entre les données météorologiques, les caractéristiques du système PV et la puissance générée seront identifiées et explorées.

1.5 Structure du manuscrit

Dans le chapitre 2, nous concentrerons sur l'état de l'art de la prévision de la production d'énergie solaire PV. nous nous commencerons sur la classification des prévisions de puissance PV en fonction de leur horizon temporel. L'objectif est de distinguer entre les prévisions à court terme, à moyen terme et à long terme. Enfin, nous explorerons les différentes techniques de prévision de puissance PV. Cela inclut les méthodes basées sur la compréhension des principes physiques, les approches statistiques utilisant des modèles probabilistes, ainsi que les méthodes d'intelligence artificielle, notamment les réseaux neuronaux et les algorithmes d'apprentissage profond. La comparaison approfondie de ces méthodes aidera à évaluer leur efficacité respective dans divers contextes et à choisir la meilleure méthode pour chaque situation. Ensuite, Nous examinerons les bases des réseaux de neurones, en comprenant leur structure et leur fonctionnement. Nous nous pencherons sur le concept fondamental d'un neurone. Nous nous intéresserons également aux fonctions d'activation usuelles, qui sont des outils essentiels pour introduire de la non-linéarité dans les réseaux de neurones, en comprenant leur rôle dans la modélisation des relations complexes entre les données. L'architecture d'un réseau de neurones sera un autre point clé de notre état de l'art. Nous explorerons la structure d'un réseau de neurones, composé

1.5. STRUCTURE DU MANUSCRIT

de couches d'entrée, de couches cachées et de couches de sortie, et nous comprendrons comment les informations se propagent à travers ces couches grâce à des connexions pondérées. En poursuivant notre exploration, nous découvrirons les différentes familles de réseaux de neurones, en mettant en évidence les réseaux de neurones convolutionnels, ainsi que les réseaux de neurones récurrents conçus pour traiter des données séquentielles et temporelles. Nous aborderons également les autoencodeurs, qui sont des réseaux de neurones utilisés pour la compression et la reconstruction de données. En particulier, nous nous intéresserons au LSTM-autoencodeur, qui intègre des mécanismes de mémoire à long terme pour capturer les dépendances temporelles dans les données. Dans la suite du chapitre nous présenterons les résultats qui ont été obtenue par l'utilisation de ces méthodes. Enfin, nous examinerons les métriques utilisées pour évaluer les performances des modèles de réseaux de neurones. Nous nous pencherons sur des mesures telles que l'erreur quadratique moyenne, l'erreur absolue moyenne et d'autres métriques qui nous permettent de quantifier l'adéquation entre les prédictions du modèle et les valeurs réelles. Ce chapitre de l'état de l'art sera donc essentiel pour jeter les bases théoriques nécessaires à notre recherche ultérieure sur les réseaux neuronaux et les algorithmes d'apprentissage profond.

Dans le chapitre 3, nous explorerons en détail les principaux axes de recherche liés aux caractérisation, acquisition, sélection des données photovoltaïques. Nous commencerons par une analyse des performances visant à optimiser les systèmes PV. L'objectif de cet axe est de plonger profondément dans les composants des systèmes PV afin de comprendre leur comportement et d'identifier les facteurs qui influent sur leur efficacité. L'axe de production PV se concentre sur l'étude des différentes sources de variabilité dans la production d'énergie solaire. Cela comprend l'impact des conditions météorologiques, de l'angle et de l'orientation des panneaux solaires, ainsi que d'autres variables environnementales. Ensuite, nous nous concentrerons sur le site d'étude et les données utilisées dans notre recherche. Une partie importante de notre travail consistera à caractériser le site d'étude. Cette caractérisation nous aidera à comprendre les conditions spécifiques dans lesquelles notre système PV fonctionne. Pour notre recherche, nous acquerrons des données de surveillance provenant des installations PV sur le site d'étude. Nous décrirons les différentes méthodes et technologies utilisées pour collecter ces données, telles que les capteurs, les enregistreurs de données et les systèmes de surveillance à distance. Une étape cruciale sera la sélection des données d'entrée appropriées pour notre modèle. Nous discuterons des différentes variables et paramètres à prendre en compte. Nous expliquerons les critères de sélection et les méthodes utilisées pour garantir la pertinence et la représentativité des

données. Enfin, nous présenterons les données collectées et effectuerons une analyse préliminaire pour comprendre les tendances, les variations et les relations entre les différentes variables. Nous aborderons également les étapes de prétraitement des données, telles que le nettoyage, la normalisation et la transformation, afin de les rendre adaptées à notre modèle de prédiction. Ce chapitre sur le site d'étude et les données utilisées sera donc essentiel pour établir les bases empiriques de notre recherche et assurer la qualité et la pertinence des informations sur lesquelles notre modèle sera construit.

Dans ce chapitre 4, nous nous pencherons sur l'utilisation des modèles d'apprentissage profond pour la prédiction de la puissance PV. Nous aborderons trois principaux axes de recherche : l'étude comparative des modèles LSTM et RNN pour la prévision de la puissance PV, le modèle hybride GRU-CNN pour la prédiction de la puissance PV, et la prévision de la production d'énergie PV à l'aide d'un modèle hybride CNN-GRU. Nous commencerons par une étude comparative approfondie des modèles LSTM et RNN pour la prévision de la puissance PV. Nous analyserons les différences entre ces deux modèles récurrents, en examinant leurs architectures, leurs capacités de mémoire à court terme et leur capacité à capturer les dépendances temporelles complexes. Ensuite, nous nous concentrerons sur le modèle hybride GRU-CNN pour la prédiction de la puissance PV. Nous expliquerons comment ce modèle combine les avantages des réseaux GRU et des CNN pour améliorer la précision des prévisions, et nous présenterons les résultats expérimentaux obtenus. Enfin, nous aborderons de la prévision de la production d'énergie PV en utilisant un modèle hybride CNN-GRU. Nous présenterons les détails de l'acquisition des mesures PV, la configuration du modèle et les résultats expérimentaux obtenus lors de son application. Nous concluons ce chapitre en analysant les performances des différents modèles avancés d'apprentissage profond utilisés. Nous évaluerons leurs avantages, leurs limites et leur capacité à améliorer la précision de la prédiction de la puissance PV. Ce chapitre sur les modèles avancés d'apprentissage profond pour la prédiction de la puissance PV sera donc essentiel pour évaluer et comparer différentes approches, et pour identifier les modèles les plus performants et prometteurs dans ce domaine.

Dans le chapitre 5, nous nous concentrerons sur l'amélioration de la précision de la prédiction de la puissance PV en utilisant des modèles d'apprentissage profond avancés. Nous présenterons deux approches principales pour atteindre cet objectif : la première approche est la prédiction de la génération d'énergie PV à l'aide d'un réseau double couche à mémoire à court terme récurrente bidirectionnelle-convolutionnelle et la deuxième approche est la prévision de la production d'énergie PV à l'aide de

réseaux autoencodeurs à mémoire à court terme. Nous décrivons la configuration du modèle, les paramètres expérimentaux utilisés et le protocole d'évaluation de la performance. Nous procéderons ensuite à une comparaison approfondie des performances du modèle proposé avec d'autres approches existantes. Nous analyserons les résultats obtenus en utilisant les métriques appropriées et évaluerons la précision, la stabilité et la capacité de généralisation de notre modèle. Nous discuterons des avantages et des limites des modèles avancés d'apprentissage profond utilisés, en mettant en évidence les aspects où ils surpassent les approches traditionnelles et les opportunités d'amélioration. Ce chapitre nous permettra donc de démontrer l'efficacité des modèles d'apprentissage profond avancés pour améliorer la précision de la prédiction de la puissance PV et d'établir les bases pour les conclusions finales de notre thèse.

Les conclusions de nos travaux, ainsi que certaines des perspectives qu'ils peuvent offrir, sont finalement présentées dans le chapitre 6.

Chapitre 2

Etat de l'art de la prévision de la production d'énergie solaire photovoltaïque

2.1 Introduction

Dans ce chapitre de l'état de l'art, nous nous aborderons sur la classification des prévisions de puissance PV, en particulier en fonction de l'horizon de prévision. Cette classification détaillée permettra de mieux comprendre les nuances temporelles des prévisions, allant des prévisions à court terme, qui sont cruciales pour la gestion en temps réel, aux prévisions à plus long terme, essentielles pour la planification stratégique. Notre deuxième axe s'attardera sur les techniques de prévision elles-mêmes. Nous explorerons les méthodes physiques, basées sur la compréhension des principes physiques sous-jacents, les méthodes statistiques, utilisant des modèles probabilistes, ainsi que les méthodes d'intelligence artificielle, telles que les réseaux neuronaux, pour prédire avec précision la puissance PV. Cette exploration comparative nous aidera à évaluer l'efficacité et la fiabilité de chaque méthode, et à identifier les contextes où elles sont les plus pertinentes. par la suite nous pencherons les réseaux de neurones, qui sont des modèles mathématiques largement utilisés en apprentissage automatique. Nous examinerons les différents aspects clés des réseaux de neurones, y compris la structure d'un neurone, les fonctions d'activation couramment utilisées, ainsi que l'architecture et les différentes familles de réseaux de neurones. Dans la première partie, nous présenterons le concept de neurone, qui est l'unité fondamentale d'un réseau de neurones. Nous examinerons comment un neurone reçoit des entrées pondérées, applique une fonction d'activation et génère une sortie. Ensuite, nous aborderons les fonctions d'activation usuelles, qui sont des fonctions mathématiques appliquées aux sorties des neurones pour introduire de la non-linéarité dans le modèle. Nous explorerons des fonctions d'activation populaires telles que la fonction sigmoïde, la fonction ReLU et la fonction tangente hyperbolique. Dans la section suivante, nous détaillerons l'architecture d'un réseau de neurones, en mettant l'accent sur les différentes couches qui composent un réseau typique. Nous expliquerons comment les neurones sont organisés en couches d'entrée, de sortie et de couches cachées, et comment les informations se propagent à travers le réseau via des connexions pondérées. Ensuite, nous explorerons les différentes familles de réseaux de neurones, en nous concentrant sur les réseaux de neurones convolutionnels et les réseaux de neurones récurrents. Les réseaux de neurones convolutionnels sont particulièrement adaptés à l'analyse d'images et de données spatiales, tandis que les réseaux de neurones récurrents sont conçus pour traiter des données séquentielles et temporelles. Nous aborderons également les autoencodeurs, qui sont des réseaux de neurones utilisés pour la compression et la reconstruction de données. Les autoencodeurs sont capables d'apprendre une représentation compressée d'un ensemble de données

2.2. CLASSIFICATION DES PRÉVISIONS DE PUISSANCE PHOTOVOLTAÏQUE BASÉE SUR L'HORIZON DE PRÉVISION

en utilisant un réseau de neurones avec une architecture spécifique. Enfin, nous discuterons des métriques couramment utilisées pour l'évaluation des modèles de réseaux de neurones. Nous examinerons des mesures telles que l'erreur quadratique moyenne, l'erreur absolue moyenne et d'autres métriques qui nous permettent de quantifier les performances d'un modèle par rapport aux valeurs réelles. Ce chapitre de l'état de l'art nous permettra de comprendre les principes fondamentaux des réseaux de neurones, leur architecture et leurs différentes applications.

2.2 Classification des prévisions de puissance photovoltaïque basée sur l'horizon de prévision

L'horizon de prévision fait référence à la période future pour laquelle les niveaux de puissance PV doivent être anticipés. L'efficacité et l'exactitude d'un modèle de prévision de la puissance PV dépendent de cette période de prévision. Lipperheide et al. [23] ont évalué la performance des prévisions de puissance PV sur diverses périodes de prévision, allant de 20 à 180 secondes, par exemple. L'erreur de prédiction (RMSE) du modèle envisagé varie de 3,2 à 15,5 % pour des prévisions s'étendant de 20 à 180 secondes. Dans le modèle conçu par Lonij et al. [24], les erreurs varient selon les horizons de prévision, de 15 minutes à 90 minutes. La précision de la prévision fluctue en fonction de l'horizon de prévision dans un même modèle, même avec des paramètres identiques. Ainsi, il est essentiel de prendre en considération la période temporelle de prévision lors de la conception du modèle de prédiction approprié. Aucun critère précis n'existe pour classer les modèles de prédiction en fonction de l'horizon de prévision. Toutefois, selon la plupart des études de chercheurs, les prévisions de production d'énergie PV peuvent être regroupées en trois catégories selon la période temporelle.

Prévision à court terme

La prévision de la production d'énergie PV pour des périodes d'une heure, de plusieurs heures, d'une journée ou jusqu'à sept jours est désignée sous le terme de prévision à court terme. Ce type de prévision assure la planification, l'organisation et la répartition de l'énergie électrique. Il est particulièrement utile pour élaborer des systèmes de gestion énergétique intégrant l'énergie PV et renforce la fiabilité des opérations du réseau électrique.

Prévision à moyen terme

Les prévisions de puissance PV à moyen terme s'étendent sur une période dépassant une semaine jusqu'à un mois. Ces prévisions facilitent la planification du réseau électrique et l'élaboration du calendrier de maintenance en anticipant la disponibilité de l'énergie électrique à venir.

Prévision à long terme

La prévision à long terme de la production d'énergie PV s'étend sur une période allant d'un mois à un an. Ce type de prévision joue un rôle essentiel dans la planification de la génération, de la transmission et de la distribution de l'électricité, ainsi que dans les processus d'enchères énergétiques et la garantie du bon fonctionnement du système. Cependant, certains chercheurs ont subdivisé l'horizon de prévision de la production d'énergie PV en quatre catégories [25]. La quatrième catégorie est désignée comme l' horizon de prévision très court terme. Les prévisions très court terme de la production d'énergie PV concernent les quelques secondes, une minute ou plusieurs minutes (<1 h) à venir. Ce type de prévision est réalisé dans le but d'atténuer les variations de puissance, d'assurer une distribution d'électricité en temps réel et d'optimiser la gestion des réserves.

2.3 Classification des techniques de prévision de puissance photovoltaïque

L'importance de l'énergie dans le processus d'industrialisation et de modernisation, ainsi que son rôle critique dans le progrès technologique et économique, ont été largement reconnus. Cependant, la crise énergétique mondiale s'est aggravée en raison de la croissance démographique massive, ce qui a entraîné une demande croissante en électricité ayant un impact négatif sur l'environnement [26]. En réponse à cette problématique, de nombreux gouvernements et régions ont pris des mesures législatives visant à encourager le développement des énergies renouvelables [27]. L'énergie solaire est devenue une solution cruciale pour résoudre les préoccupations environnementales et énergétiques en raison de ses propriétés propres et abondantes [28]. Cependant, l'intégration de la production d'énergie PV dans les réseaux énergétiques actuels est rendue extrêmement difficile en raison de la nature aléatoire et intermittente de la génération de puissance PV [29]. Cette caractéristique pose des défis majeurs en termes de stabilité et de fiabilité du réseau électrique. Dans ce contexte, la prédiction précise de la puissance PV est devenue un élément critique pour garantir la sécurité du réseau électrique et optimiser

2.3. CLASSIFICATION DES TECHNIQUES DE PRÉVISION DE PUISSANCE PHOTOVOLTAÏQUE

le stockage des sources d'énergie alternatives sur une durée raisonnable[20]. Des efforts de recherche importants ont été déployés pour développer des modèles et des techniques de prédiction avancés pour la puissance PV, en tenant compte de divers paramètres tels que la moyenne de phase de courant, le rayonnement horizontal global et le rayonnement horizontal diffus. L'objectif de cet état de l'art est de recenser et d'analyser les travaux de recherche pertinents qui se sont concentrés sur l'énergie, en mettant particulièrement l'accent sur l'énergie solaire PV. L'examen de la littérature scientifique permettra de comprendre les avancées réalisées dans le domaine de la prédiction de la puissance PV, ainsi que les défis et les limites actuelles de cette technologie. Différentes approches de prédiction ont été présentées au cours des dernières années pour prévoir la production d'énergie PV, et elles peuvent être regroupées en trois catégories : méthode physique [30], méthode statistique [31], méthode intelligences artificielle [32] . Ces trois types de techniques de prédiction de la puissance PV ont été développés en se basant sur différents principes de prévision.

2.3.1 Méthode physique

Les modèles physiques de prédiction de la puissance PV se basent sur des variables géologiques et des données météorologiques (telles que la pression atmosphérique, l'humidité, le rayonnement solaire, le volume des nuages, etc.) fournies par les stations météorologiques. À partir de ces données, un modèle physique est créé en utilisant les paramètres des panneaux solaires PV, ce qui permet de calculer directement la puissance générée par les panneaux solaires. Ce modèle physique nécessite moins de données historiques pour effectuer des prévisions, mais il est plus complexe à mettre en place en raison de la présence de plusieurs paramètres inconnus [33]. Lee et Lee [34] ont présenté un système d'apprentissage profond bayésien conscient de la confiance pour prédire les informations sur l'irradiance solaire. Alonso-Montesinos et al. [35] ont utilisé la technologie des caméras de ciel pour la prévision de l'irradiance solaire, en utilisant la technique de corrélation croisée maximale pour dériver les prévisions futures après avoir converti les niveaux d'image numérique en irradiances. Dans Geraldini et al. [36], une méthodologie avancée pour estimer le rayonnement solaire a été suggérée, incluant l'utilisation de capteurs supplémentaires qui ont considérablement amélioré la précision des prévisions. Les auteurs de Perez et al. [37] ont testé de nombreux modèles de prévision numérique du temps dans plusieurs endroits aux États-Unis, en Europe et au Canada. Les modèles physiques de prédiction de la puissance PV sont basés sur les principes scientifiques et les équations qui décrivent le comportement des cellules

2.3. CLASSIFICATION DES TECHNIQUES DE PRÉVISION DE PUISSANCE PHOTOVOLTAÏQUE

solaires et des installations PV. Ils prennent en compte les conditions météorologiques actuelles, telles que la pression atmosphérique, l'humidité, le rayonnement solaire et la couverture nuageuse, afin de calculer la puissance attendue des panneaux solaires. Ces modèles sont moins dépendants des données historiques, car ils se concentrent principalement sur les conditions météorologiques actuelles et les caractéristiques physiques des panneaux solaires. Cependant, leur mise en œuvre est plus complexe en raison de la nécessité de déterminer plusieurs paramètres inconnus, tels que les caractéristiques précises des panneaux solaires et les propriétés du site d'installation.

2.3.2 Méthode statistique

Les modèles de séries temporelles sont utilisés pour prédire la puissance PV en se basant sur des données statistiques. Dans ces modèles, l'impact des conditions météorologiques n'est pas pris en compte tout au long de la phase de prévision; seul le facteur temps est pris en considération. Une variété d'approches statistiques ont été utilisées pour prévoir la production d'énergie PV, telles que la méthode de la chaîne de Markov [38, 39], la théorie des gris [40] et les modèles autorégressifs et à moyenne mobile (ARMA) [41, 42]. Contrairement aux modèles physiques, les modèles statistiques ne nécessitent pas une compréhension préalable des aspects complexes de la conversion photoélectrique dans les systèmes PV. Ils reposent plutôt sur une analyse statistique des données pour identifier les tendances et les motifs temporels permettant de prédire la puissance PV. Cependant, ces modèles nécessitent une quantité importante de données pour les calculs et peuvent prendre plus de temps pour fournir des prévisions précises.

2.3.3 Méthode intelligences artificielle

Les méthodes d'apprentissage automatique basées sur l'intelligence artificielle exploitent la croissance rapide de la puissance de calcul en utilisant des algorithmes d'intelligence artificielle pour apprendre une relation de mappage entre les données d'entrée et de sortie, en mettant notamment l'accent sur les modèles de mappage non linéaire [43, 44]. Ces méthodes utilisent des algorithmes d'apprentissage automatique, tels que les réseaux neuronaux, les machines à vecteurs de support (SVM), les arbres de décision ou les algorithmes génétiques, pour analyser les données et découvrir des motifs et des relations complexes. Grâce à la puissance de calcul accrue et à l'efficacité des algorithmes d'apprentissage automatique, ces méthodes sont capables de modéliser et de prédire des relations non

2.3. CLASSIFICATION DES TECHNIQUES DE PRÉVISION DE PUISSANCE PHOTOVOLTAÏQUE

linéaires entre les variables d'entrée et de sortie. Les approches basées sur l'intelligence artificielle sont largement utilisées dans divers domaines, y compris la prévision. Dans l'étude de Mubiru [45], les auteurs ont proposé un modèle de réseau neuronal artificiel pour prédire l'irradiation solaire en se basant sur des variables physiques et environnementales. Pour la prévision de la puissance solaire pour le lendemain, Ghasvarian Jahromi et al. [46] ont proposé une nouvelle approche basée sur le modèle de Markov caché (HMM). Ikram et al. [47] ont proposé une nouvelle méthode LSSVM-IMVO basée sur la méthode d'optimisation multi-verse améliorée (IMVO) avec une machine à vecteurs de support à moindres carrés (LSSVM) pour prédire le rayonnement solaire de deux stations situées dans la région sud-est de la Chine. La méthode LSSVM-IMVO peut estimer avec succès le rayonnement solaire global en utilisant des données climatiques limitées (uniquement la température), comme le montrent les comparaisons avec des travaux antérieurs. Mellit et al. [48] ont développé plusieurs réseaux neuronaux à apprentissage profond (DLNN) pour la prévision en avance d'un pas et en avance de plusieurs pas de la génération de puissance PV, sur différents horizons temporels. Pour entraîner et évaluer les réseaux neuronaux, une base de données de la puissance PV générée par le microgrid installé à l'Université de Trieste (Italie) est utilisée. Les résultats ont démontré que les DLNN analysés offrent une très bonne précision, notamment pour un horizon temporel en avance d'un pas d'une minute. Zhou et al. [49] ont proposé un système de prédiction hybride révolutionnaire à variables multiples qui combine la décomposition du signal, des modèles d'apprentissage profond, des modèles d'intelligence artificielle et une technique d'optimisation pour l'intelligence de groupe. Cette méthode exploite pleinement les variables indépendantes, telles que la température des modules, pour améliorer la précision et l'efficacité de la prévision PV. Liu et al. [50] ont proposé un modèle de prédiction de la puissance PV basé sur le réseau neuronal à propagation arrière (BPNN) avec classification saisonnière de la météo. Huang et Kuo [51] ont proposé un puissant modèle de réseau neuronal profond pour prévoir la puissance de sortie du système PV. Un modèle ABPNN avec l'indice de qualité de l'air, le rayonnement solaire, l'humidité, la température, la vitesse du vent et l'indice de qualité de l'air en tant qu'attributs d'entrée a été utilisé pour prédire la puissance PV en cas de brouillard. Les résultats montrent que l'ajout de variables d'entrée supplémentaires, telles que l'indice d'aérosol et l'indice de qualité de l'air, améliore les performances par rapport au modèle NN proposé dans Khan et al.

Les méthodes hybrides combinent différentes approches de prédiction, telles que les méthodes physiques, statistiques et d'apprentissage automatique, afin de tirer parti des avantages de chacune.

Elles peuvent utiliser des techniques d'optimisation pour combiner les résultats de différentes méthodes ou utiliser des modèles en cascade pour améliorer la précision des prédictions.

2.4 Les réseaux de Neurones.

Le concept des réseaux de neurones est initialement basé sur le fonctionnement des neurones biologiques et vise à les modéliser. Un neurone artificiel est construit en s'inspirant des différentes composantes d'un neurone biologique : les dendrites (entrées du neurone), les synapses (points de connexion), l'axone (sortie) et le noyau (responsable de l'activation des sorties en fonction des entrées). Tout comme pour les neurones biologiques, les signaux d'entrée (dendrites), par exemple x_0 , sont pondérés par des poids (synapses), W_0 pour x_0 , ce qui permet de moduler leur influence. La combinaison de toutes les entrées pondérées est ensuite soumise à une étape de seuillage (noyau), réalisée à l'aide d'une fonction d'activation, afin de générer la sortie (axone). En résumé, chaque neurone réalise une opération de produit matriciel entre les poids et les entrées, et parfois applique une fonction non linéaire, appelée fonction d'activation, pour produire sa sortie. Il est possible d'ajouter un biais (b) après l'opération de produit matriciel. Cela permet de décaler artificiellement le seuil de la fonction d'activation. La figure.2.1 montre une représentation d'un neurone artificiel.

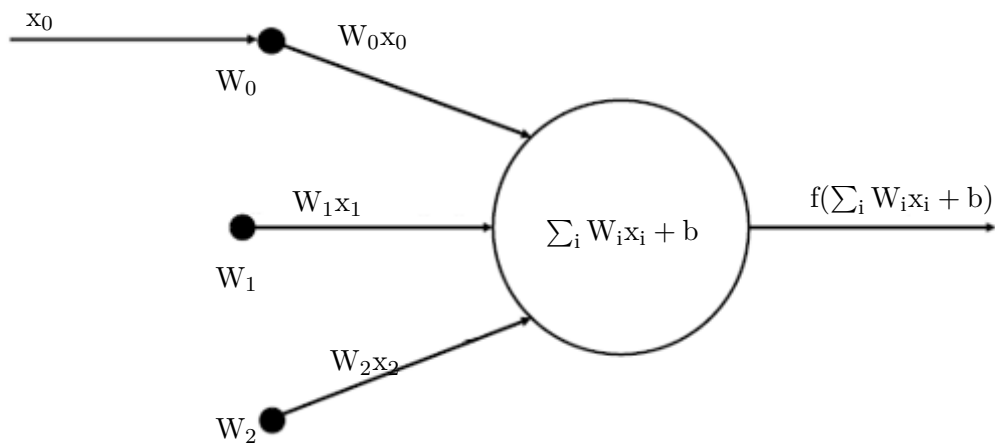


FIGURE 2.1 – Représentation d'un neurone artificiel.

Les fonctions d'activation usuelles

Les fonctions d'activation servent à établir un seuil qui, une fois franchi, provoque l'activation d'un neurone. Traditionnellement, la fonction sigmoïde a été privilégiée pour cela. Cependant, il semble que d'autres fonctions d'activation soient plus efficaces en fonction des tâches. Nous présentons ici une liste non exhaustive des fonctions d'activation les plus couramment utilisées.

Tanh La fonction tangente hyperbolique transforme n'importe quelle valeur réelle en une valeur comprise entre -1 et 1. Contrairement à la fonction sigmoïde, la tangente hyperbolique est centrée en 0. Il est intéressant de noter que la fonction tangente hyperbolique peut être exprimée en fonction de la fonction sigmoïde.

$$\tanh(x) = 2\sigma(2x) - 1 \quad (2.1)$$

Fonction sigmoïde La fonction sigmoïde, représentée mathématiquement par l'équation (2), réduit les valeurs réelles à une plage allant de 0 à 1. Ainsi, les valeurs fortement négatives sont associées à 0, tandis que les valeurs très grandes et positives sont ramenées à 1. Cependant, pour ces valeurs extrêmes, le gradient devient nul, ce qui signifie que le réseau ne peut pas apprendre à partir de ces données. En conséquence, la fonction sigmoïde est moins couramment utilisée de nos jours. Néanmoins, elle est encore utilisée dans certains cas, car la restriction des valeurs entre 0 et 1 permet d'établir un lien avec une mesure de probabilité.

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (2.2)$$

Relu La fonction d'activation Relu (Rectified Linear Unit) est définie par la formule $f(x) = \max(0, x)$. la fonction ReLU se résume à un simple seuil à zéro. Cependant, les unités ReLU peuvent être fragiles : le gradient peut mettre à jour les poids de manière à ce que le neurone ne soit plus activé par aucune donnée par la suite. En conséquence, le gradient pour ce neurone restera toujours nul à partir de ce moment.

Architecture d'un réseau de neurones

Les réseaux de neurones sont structurés en couches. La sortie d'un neurone peut devenir l'entrée d'un autre neurone, formant ainsi une hiérarchie de couches. Cependant, les cycles ne sont pas autorisés. Le type de couche le plus couramment utilisé est la couche dense, également appelée couche entièrement connectée, où chaque neurone de deux couches est associé les uns aux autres, mais aucune connexion n'est établie à l'intérieur de la même couche.

Par convention, la couche d'entrée n'est pas incluse lorsqu'on nomme un réseau. Par exemple, un réseau avec N couches aura $N-1$ couches dites "cachées" (hidden layer) et une couche de sortie (output layer), totalisant N couches. De plus, il y a une couche d'entrée (input layer). Ainsi, un réseau à une seule couche ne contient aucune couche cachée et relie directement l'entrée à la sortie. Un réseau de neurones est caractérisé par le nombre de ses couches, ainsi que par le nombre de paramètres qu'il contient, tels que les poids et les biais. Ces paramètres sont déterminés lors de la phase d'apprentissage. Il existe également d'autres paramètres, appelés hyperparamètres, qui doivent être définis avant l'apprentissage, tels que le nombre de couches ou le nombre de neurones par couche. Ces hyperparamètres sont des paramètres dont les valeurs doivent être fixées manuellement.

Régularisation

Il peut être bénéfique de prendre des mesures pour prévenir le surapprentissage. Voici quelques stratégies possibles qui peuvent être utilisées à cet effet.

Régularisation par la norme l_2 – Il s'agit probablement de la technique de régularisation la plus couramment utilisée. On ajoute un terme de pénalisation pour chaque poids à la loss function. En raison des interactions multiplicatives entre les poids, cela contraint le réseau à utiliser de manière équilibrée toutes les entrées plutôt que de se concentrer sur un petit nombre d'entre elles.

Régularisation par la norme l_1 – Contrairement à la régularisation l_2 , l'utilisation de la régularisation l_1 incite les neurones à se concentrer sur un petit nombre de leurs entrées.

Dropout – Le dropout est une technique récemment apparue, simple mais très efficace. Elle consiste à désactiver un certain nombre de sorties de neurones. Les neurones qui restent actifs sont déterminés par une probabilité p , qui est un hyperparamètre du réseau. Dans une couche dense, la plupart des paramètres sont connectés entre deux couches, ce qui peut entraîner une co-dépendance

entre les neurones. Cela peut conduire à un surapprentissage des données d'entraînement. Par conséquent, la régularisation basée sur le dropout permet au réseau d'apprendre des caractéristiques plus robustes. En effet, en désactivant aléatoirement des unités du réseau, il devient plus difficile pour les unités de compenser mutuellement leurs erreurs. Chaque unité du réseau contribue à l'erreur totale, et l'erreur d'une unité ne peut plus être compensée par l'erreur d'une autre unité. Cependant, cela augmente le nombre d'itérations d'apprentissage nécessaires pour atteindre la convergence.

2.5 Différentes familles de réseaux de neurones

2.5.1 Réseaux de neurones convolutionnels

Les réseaux de neurones convolutifs (CNN) sont un type spécifique d'architecture d'apprentissage profond qui s'est avéré efficace dans diverses applications [52, 53]. En général, les CNN sont considérés comme des extracteurs de caractéristiques hiérarchiques capables d'apprendre automatiquement des informations de haut niveau à partir de séquences originales. Un composant crucial des CNN est la couche de convolution. La technique de convolution est utilisée pour extraire des caractéristiques distinctes de l'entrée, et des couches supplémentaires peuvent extraire de manière récursive des caractéristiques complexes à partir des caractéristiques finales. Chaque couche de convolution est composée de nombreuses unités de convolution, chacune ayant son propre ensemble de paramètres qui est ajusté via une technique de rétropropagation. Dans la couche de convolution, le graphe de caractéristiques de la couche précédente réagit avec le noyau de convolution, ce qui donne le graphe de caractéristiques en sortie de la couche de convolution j . Chacun des graphes de caractéristiques de sortie j peut avoir une convolution avec plusieurs graphes de caractéristiques en entrée. La Figure.2.2 montre l'architecture des réseaux de neurones convolutifs. La couche de convolution est définie par les équations suivantes :

$$y_j^{(k)} = \sum_{i \in s_j} (p_i^{k-1} \otimes W_{ij}^{(k)}) + b_j^{(k)} \quad (2.3)$$

où $y_j^{(k)}$ est la sortie de la convolution, s_j est un ensemble de graphes d'entrée, $W_{ij}^{(k)}$ est le noyau de convolution et $b_j^{(k)}$ est le biais. p_j^k est le graphe de caractéristiques de la couche de convolution k , défini comme suit :

$$p_j^{(k)} = g(y_j^{(k)}) \quad (2.4)$$

La couche de pooling est utilisée principalement pour réduire le surajustement (overfitting) en comprimant la quantité de données et de paramètres. Entre les couches de convolution consécutives, on trouve différentes formes de processus de pooling, tels que le max-pooling et le average pooling. Le résultat final est calculé à l'aide d'une couche entièrement connectée (fully connected layer), qui fusionne toutes les caractéristiques locales en des caractéristiques globales. Le processus de pooling consiste à réduire la dimensionnalité spatiale des caractéristiques en prenant des valeurs maximales (max-pooling) ou moyennes (average pooling) sur des régions d'intérêt. Cela permet de réduire le nombre de paramètres et de rendre le modèle plus robuste aux variations mineures dans les caractéristiques. Entre les couches de convolution, le pooling est appliqué pour réduire la taille des caractéristiques et résumer les informations les plus importantes. Cela aide à réduire la quantité de données à traiter par les couches suivantes et à prévenir le surajustement en supprimant les détails redondants. Une fois que les caractéristiques ont été réduites par le pooling, elles sont ensuite passées à une couche entièrement connectée. Dans cette couche, les caractéristiques locales sont fusionnées en des caractéristiques globales qui capturent des informations plus abstraites et de plus haut niveau.

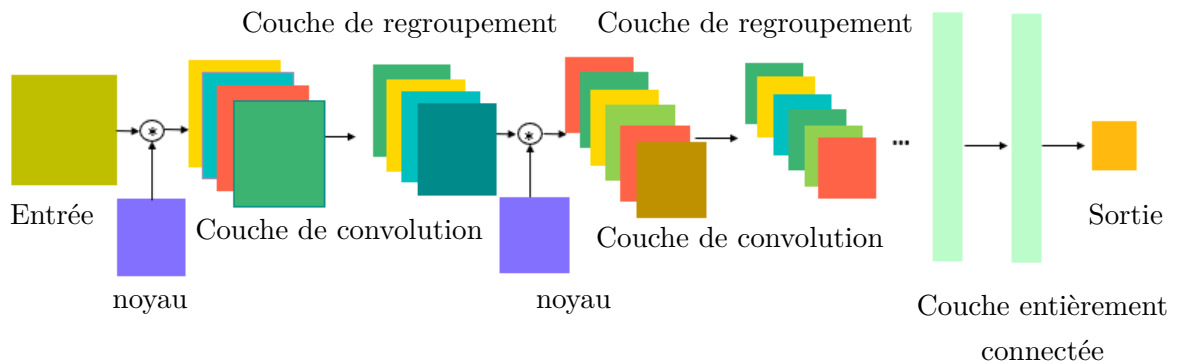


FIGURE 2.2 – Architecture de Réseaux de Neurones Convolutifs.

2.5.2 Résultats bibliographiques obtenus par application de réseaux de neurones convolutifs pour la prévision dans le domaine PV.

Korkmaz et al. [54] ont introduit un nouveau système de prévision de la production photovoltaïque qui utilise une structure de Réseau de Neurones Convolutifs profond et un algorithme de décompo-

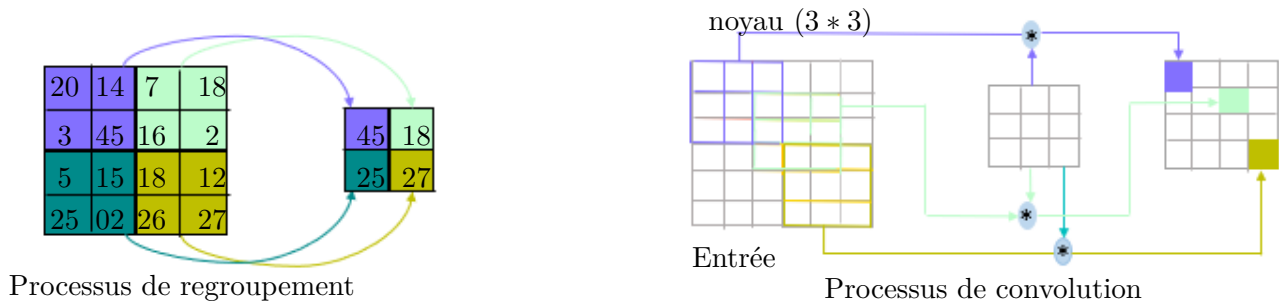


FIGURE 2.3 – Processus de convolution et de mise en commun maximale.

sition du signal d’entrée. En exploitant le modèle AlexNet basé sur l’apprentissage par transfert, le CNN extrait des caractéristiques profondes pour prédire la production d’énergie à court terme. Les expériences ont été menées sous différentes conditions météorologiques telles que partiellement nuageux, pluvieux, fortes pluies et ensoleillé pour démontrer l’efficacité de leur méthode. Les résultats obtenus ont été comparés à des algorithmes de régression de référence. L’analyse des résultats a montré que leur méthode avait le coefficient de corrélation le plus élevé et les erreurs les plus faibles en termes d’Erreur RMSE, d’Erreur MAE et de SMAPE pour tous les horizons temporels et conditions météorologiques. Pour les prévisions de 1 à 5 heures à l’avance, leur méthode a affiché des valeurs moyennes de R de 97,28 %, 95,77 %, 94,49%, 93,61 % et 92,62 % respectivement. Les valeurs moyennes de RMSE ont été observées à 4,90%, 6,30 %, 7,50 %, 8,00 % et 9,17 % pour les prévisions de 1 à 5 heures à l’avance.

Agga et al. [55] ont développé deux variantes étendues du réseau neuronal convolutif pour les prévisions à court terme d’une centrale photovoltaïque. Ces modèles sont utilisés pour anticiper la production d’énergie photovoltaïque un et deux jours à l’avance. Deux ensembles de données ont été utilisés : l’un univarié, ne contenant que les enregistrements de production d’énergie, et l’autre multivarié, incluant des variables météorologiques en plus des données de production d’énergie. La performance des modèles proposés a été évaluée au moyen d’une étude de cas utilisant des données réelles collectées à Rabat, au Maroc. Les résultats des trois mesures d’erreur, MAE, MAPE et RMSE, indiquent que les versions étendues du modèle CNN présentent une précision et une stabilité supérieures dans leurs prédictions.

Alam et al. [41] ont exploré différentes techniques de Réseaux Neuronaux Convolutifs telles que les CNN classiques, les CNN multi-têtes et les CNN-LSTM. Ils ont utilisé un algorithme de fenêtre glissante ainsi que d’autres méthodes d’extraction de caractéristiques et de prétraitement pour effectuer

des prédictions précises. Les paramètres météorologiques tels que l'irradiance solaire, la température de l'air, l'humidité, la direction du vent et la vitesse du vent ont été associés à la production des panneaux solaires. Par exemple, ils ont utilisé des données d'entrée sur une période de 5 ans pour prévoir la production d'un jour spécifique et d'une semaine. Les résultats ont été évalués en les comparant à des méthodes de prévision traditionnelles telles que l'Autoregressive Moving Average et la Régression Linéaire Multiple (MLR). L'efficacité des résultats a été évaluée à l'aide de mesures d'évaluation telles que le RMSE, le MAE et le MBE. Les techniques traditionnelles ainsi que les méthodes d'apprentissage automatique ont démontré leur efficacité dans la production de prévisions à court et à moyen terme.

Korkmaz [56] a proposé un nouveau modèle de réseau neuronal convolutif SolarNet pour la prévision à court terme de la production d'énergie photovoltaïque dans diverses conditions météorologiques et saisons. Ce modèle CNN proposé est structuré en parallèle avec des blocs de max-pooling et de average-pooling pour améliorer la performance de prévision. Les données d'entrée comprennent l'irradiation solaire historique mesurée, la température, l'humidité et les données de puissance active. Dans les expériences menées, le modèle proposé a atteint un coefficient de corrélation moyen de 0,9871, une erreur RMSE de 0,3090 et une erreur MAE de 0,1750 pour des conditions météorologiques différentes à 1 heure. Les résultats expérimentaux mettent en évidence la précision et la stabilité supérieures de la méthode de prévision profonde proposée pour la production d'énergie photovoltaïque à court terme, surpassant ainsi les autres méthodes d'apprentissage profond.

2.5.3 Réseaux de neurones récurrents

Le réseau de neurones récurrent (RNN) est un type de réseau neuronal utilisé pour prévoir des données séquentielles, dont la sortie dépend de l'entrée [57]. En conservant les informations des calculs précédents dans la mémoire interne, le RNN est capable de capturer la dynamique des données de séries chronologiques [58]. Le RNN a été utilisé dans des situations où les valeurs précédentes de la sortie ont un impact significatif sur l'avenir. En raison de sa capacité à analyser des données séquentielles de longueurs variables, il est principalement utilisé dans les applications de prévision. Le RNN fonctionne en tenant compte des entrées des neurones de la couche cachée, qui reçoivent les entrées des neurones de l'instant temporel précédent [59]. À cette fin, il utilise des cellules représentées par des portes pour contrôler la sortie, qui est générée en utilisant les observations de données passées. Le RNN excelle dans la compréhension des propriétés temporelles dynamiques qui émergent dans les séries chronologiques

[22]. La Figure 1 illustre un aperçu complet des réseaux de neurones récurrents.

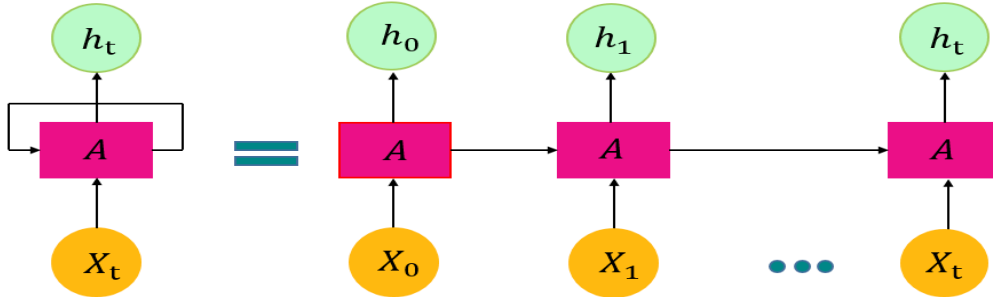


FIGURE 2.4 – Vue d'ensemble générale des réseaux neuronaux récurrents.

Pour une séquence d'entrée particulière x_t , le neurone caché h_t reçoit des informations en retour des autres neurones de l'instant temporel précédent, multipliées par W_{hh} , qui est le poids de l'état caché précédent h_{t-1} , et cela peut être déterminé séquentiellement selon l'équation 1. L'état de sortie y_t est calculé selon l'équation 2.

$$h_t = f(W_{hh}h_{t-1} + W_{xh}x_t + b_h) \quad (2.5)$$

$$y_t = g(W_{hy}h_t + b_y) \quad (2.6)$$

W_{xh} représente le poids de l'état d'entrée actuel, W_{hy} représente le poids à l'état de sortie, x_t correspond à l'entrée à l'instant de temps t , f et g représentent respectivement la fonction d'activation de la couche cachée et la fonction d'activation de la couche de sortie. Les réseaux de neurones récurrents (RNN) sont confrontés à un problème de disparition du gradient. Cela signifie que les poids ont des difficultés à se propager de manière efficace vers l'arrière lors de la rétropropagation du gradient.

2.5.4 Résultats bibliographiques obtenus par application de réseaux de neurones récurrents pour la prévision dans le domaine PV.

Park et al. [60] ont introduit un modèle basé sur un réseau de neurones récurrents pour prédire la production d'énergie PV. Dans leur étude, ils ont utilisé des données météorologiques et de production d'énergie PV préalablement collectées pour anticiper la production future d'énergie PV. Ils ont appliqué divers processus d'optimisation, tels que la normalisation des données, la classification des données d'apprentissage et le réglage des options de couche, afin de créer un modèle de prédiction. Ils ont créé

et comparé plusieurs configurations, notamment des modèles avec 500 neurones cachés et 1 à 3 couches cachées. Le taux d'apprentissage initial pour ces modèles à une seule couche et à plusieurs couches a été fixé à 0,05. Les résultats ont montré que le coefficient de variation de l'erreur quadratique moyenne normalisée ($Cv(RMSE)$), un indicateur statistique évaluant la précision globale des prédictions, était de 13,8% pour le modèle à une seule couche et de 13,2 % pour le modèle à mémoire à long terme avec plusieurs couches.

Liu et al. [61] ont présenté un modèle de Réseau de Neurones Récursifs Segmenté (SRNN) utilisé pour prédire la puissance photovoltaïque, assurant ainsi la capacité de calcul parallèle. Les résultats de leur étude démontrent que par rapport à d'autres modèles couramment utilisés, le SRNN peut accélérer considérablement l'entraînement des Réseaux de Deep Learning, avec une vitesse d'entraînement plus de 4 fois supérieure à celle des structures RNN classiques telles que LSTM et GRU pour la prédiction de la puissance photovoltaïque. De plus, la précision du modèle SRNN s'améliore de 0,1102 en termes d'erreur absolue moyenne, le plaçant nettement en avance sur les autres modèles. Cette architecture parallèle permet une mise à jour plus efficace des paramètres, conduisant ainsi à un résultat optimal dans la prédiction de la puissance photovoltaïque.

Park et al. [60] ont présenté un algorithme de prévision à court terme de la production photovoltaïque basé sur les réseaux de neurones récurrents profonds. Ils ont collecté des données sur l'énergie générée sur site ainsi que des informations météorologiques à l'aide de capteurs PIoT installés dans un système PV. Ils ont mené une analyse de corrélation entre les données de production d'énergie PV et les données météorologiques recueillies dans le système PV afin de sélectionner les paramètres météorologiques pertinents. Cette analyse de corrélation a confirmé que la production d'énergie solaire était influencée par l'irradiation solaire, la température ambiante et la température des modules. Les scores R^2 pour les prévisions à 1 heure et à 3 heures étaient respectivement de 0,963 et 0,927. Les résultats expérimentaux ont démontré que l'algorithme de prévision à court terme de la production photovoltaïque basé sur les RNN profonds qu'ils ont proposé offrait une précision de prédiction supérieure par rapport aux modèles ARIMA et SVR-RFN.

Wang et al. [29] ont présenté une nouvelle approche à court terme basée sur les réseaux de neurones récurrents pour prédire la puissance photovoltaïque. À la différence des méthodes de prévision classiques, ils ont alimenté le modèle RNN avec la puissance PV des jours adjacents, validant ainsi la similarité et la corrélation avec les données de puissance réelles en Flandre, en Belgique. Ce modèle de

prévision, doté d'une architecture récurrente et d'unités de mémoire, est capable d'apprendre efficacement les variations de puissance ainsi que les caractéristiques de la production PV inter-journalières et intra-journalières. Ils ont ensuite comparé cette méthode de manière exhaustive avec les méthodes LSTM, SVM, RBF, BPNN et Persistence. Certains résultats ont montré que l'erreur moyenne absolue en pourcentage, RMSE et MAE de la méthode proposée surpassent les références pour les prévisions à 15 minutes et 30 minutes.

2.5.5 Le réseau récurrent à mémoire court-terme et long terme ou LSTM

Le LSTM (Long Short-Term Memory) est une variante spéciale des réseaux de neurones récurrents. En raison du problème de disparition du gradient et du problème de dépendance à long terme, qui rendent le modèle inadapté à l'apprentissage des données séquentielles, ces problèmes peuvent être résolus en insérant une cellule de mémoire dans chaque neurone de la couche cachée et en utilisant des portes d'entrée, de sortie et d'oubli pour réguler l'état de la cellule de mémoire [62]. L'architecture du neurone LSTM est illustrée dans la figure.2.5.

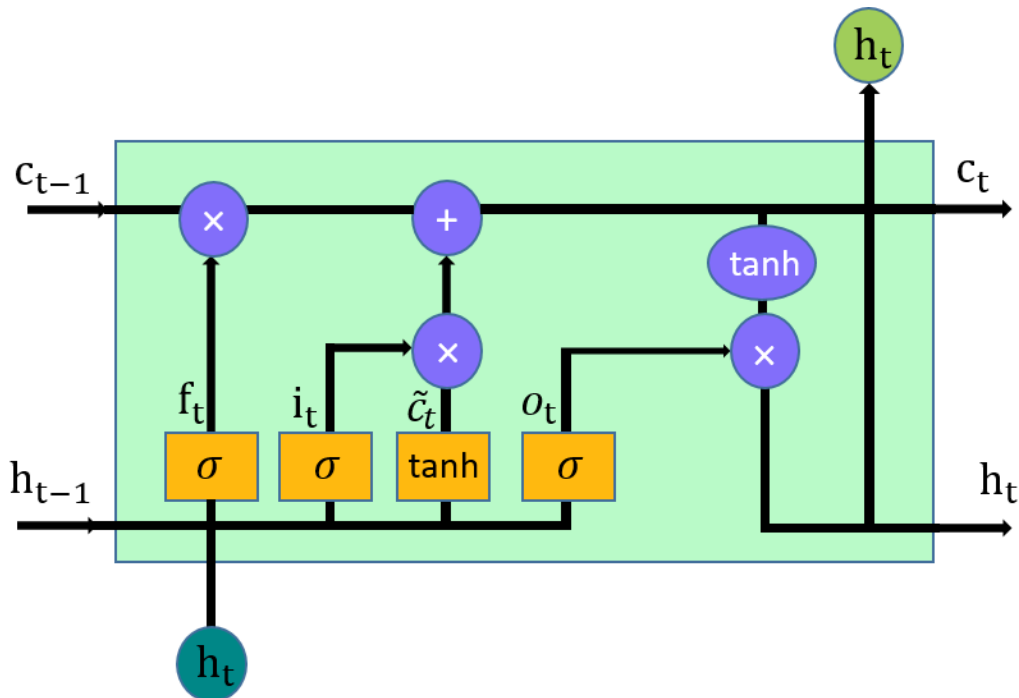


FIGURE 2.5 – L'architecture des réseaux LSTM

la cellule de mémoire et l'état caché peuvent mémoriser les informations historiques des données

séquentielles. Les informations contenues dans la cellule de mémoire sont contrôlées par trois unités de porte. Selon l'état caché précédent h_{t-1} et l'entrée actuelle x_t , la porte d'oubli (forget gate) supprime les informations présentes dans la cellule de mémoire. La porte d'oubli est calculée comme suit :

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (2.7)$$

En fonction de l'état caché précédent h_{t-1} à l'instant temporel précédent et de l'entrée actuelle x_t à l'instant temporel courant, la porte d'entrée (input gate) i_t ajoute des informations à la cellule de mémoire. Elle représente la porte d'entrée. C_t est l'état de la cellule et \tilde{C}_t est la valeur utilisée pour calculer l'état actuel de la cellule C_t . La porte d'entrée est responsable de la régulation des informations qui seront ajoutées à la cellule de mémoire en fonction de l'état caché précédent et de l'entrée actuelle. La valeur \tilde{C}_t est utilisée pour calculer la mise à jour de l'état de la cellule, qui sera ensuite utilisée dans les calculs ultérieurs.

Une fois que la porte d'oubli (forget gate) et la porte de sortie (output gate) ont été calculées, l'équation suivante est utilisée pour mettre à jour la cellule de mémoire :

$$C_t = f_t \cdot c_{t-1} + i_t \cdot \tilde{C}_t \quad (2.8)$$

Cette équation permet de mettre à jour la valeur de la cellule de mémoire en utilisant les informations de la porte d'oubli et de la porte de sortie précédemment calculées. Elle permet de contrôler le flux d'informations dans la cellule de mémoire en fonction des décisions prises par les portes d'oubli et de sortie. L'état caché actuel h_t est déterminé par la porte de sortie (output gate) en se basant sur l'état caché du moment précédent h_{t-1} , l'entrée du moment actuel x_t , et la cellule de mémoire mise à jour C_t .

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (2.9)$$

$$h_t = o_t \cdot \tanh(C_t) \quad (2.10)$$

La porte de sortie joue un rôle crucial dans le calcul de l'état caché actuel. Elle combine l'information de l'état caché précédent, de l'entrée actuelle et de la cellule de mémoire mise à jour pour

déterminer la contribution de chaque élément à la création de l'état caché actuel. Cette étape permet de réguler et de contrôler le flux d'informations pour générer l'état caché approprié en fonction des informations contextuelles précédentes et de l'entrée actuelle. W_f représente le poids de la porte d'oubli (forget gate), W_i et W_c représentent les poids de la porte d'entrée (input gate), et W_o représente le poids de la porte de sortie (output gate). De plus, b_f , b_i , b_c et b_o représentent les biais associés à chaque porte.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2.11)$$

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (2.12)$$

La fonction σ désigne la fonction sigmoïde, qui est utilisée pour transformer les valeurs en une plage entre 0 et 1. Ces poids et biais sont utilisés dans les équations de calcul des différentes portes du réseau LSTM pour ajuster l'influence de chaque entrée et guider le flux d'informations dans la cellule de mémoire et l'état caché.

2.5.6 Résultats bibliographiques obtenus par application de réseau récurrent à mémoire court-terme et long terme pour la prévision dans le domaine PV.

Gao et al. [63] ont développé un modèle efficace basé sur LSTM pour prédire la puissance produite dans des conditions météorologiques idéales. Pour obtenir des prévisions précises de l'énergie pour le lendemain, ils ont utilisé les données météorologiques moyennes quotidiennes prédites par les prévisions météorologiques numériques (NWP) comme entrées du modèle. Ils ont également pris en compte les facteurs saisonniers, concluant que les résultats de prévision sont les meilleurs au printemps, avec une valeur de RMSE de 4,62 %. En parallèle, d'autres algorithmes ont été utilisés pour la comparaison avec LSTM, obtenant des erreurs RMSE variant de 4,62 % à 17,2 %.

Konstantinou et al. [64] ont exploré l'utilisation d'un réseau de mémoire à long terme empilé (stacked LSTM), un composant clé des réseaux de neurones récurrents profonds, pour prédire la production d'énergie photovoltaïque 1,5 heures à l'avance. Ils ont utilisé les données historiques de production d'énergie photovoltaïque provenant d'une centrale à Nicosie, Chypre, comme données d'entrée pour leur modèle de prévision. Après avoir défini et entraîné le modèle, ils l'ont évalué qualitativement à

2.5. DIFFÉRENTES FAMILLES DE RÉSEAUX DE NEURONES

l'aide d'outils graphiques et quantitativement en calculant RMSE et en appliquant la méthode de validation croisée k-fold. Les résultats ont démontré que leur modèle peut effectuer des prédictions précises, avec un RMSE de 0,11368. Lors de l'application de la validation croisée k-fold, la moyenne des valeurs de RMSE obtenues s'est établie à 0,09394 avec un écart-type de 0,01616.

Hossain et al. [65] ont présenté un algorithme de prévision visant à prédire la production d'énergie photovoltaïque à l'aide d'un réseau de neurones à mémoire à long terme. Une prévision météorologique synthétique est générée pour l'emplacement spécifique de la centrale photovoltaïque en intégrant les connaissances statistiques des données historiques d'irradiance solaire avec le type de ciel disponible publiquement dans la ville hôte. À cette fin, un algorithme de K-moyennes est utilisé pour classifier les données historiques d'irradiance en groupes de types de ciel dynamiques qui varient d'heure en heure pendant la même saison. Les performances du modèle proposé sont évaluées pour différentes périodes de prévision intrajournalières pendant diverses saisons. Les résultats indiquent que l'utilisation de la prévision d'irradiance synthétique peut améliorer la précision jusqu'à 33 % par rapport à l'utilisation d'une prévision de type de ciel horaire, et jusqu'à 44,6 % par rapport à celle d'une prévision de type de ciel quotidien.

Gao et al. [66] ont examiné l'algorithme LSTM d'un point de vue théorique et l'ont appliqué à la prévision de séries temporelles de production d'énergie photovoltaïque. Sans altérer la couche d'entrée, l'optimiseur, la couche de sortie ou le taux d'apprentissage, ils ont étudié l'impact sur les performances de prévision en modifiant les paramètres et le nombre de couches cachées. De plus, ils ont décrit et comparé quatre modèles de prévision à court terme (LSTM, réseaux BP, réseaux WN et LSSVM) pour anticiper la production d'énergie une heure à l'avance, en utilisant les données météorologiques de l'heure suivante. Les valeurs RMSE des prévisions faites avec les modèles LSTM pour les quatre saisons étaient respectivement de 5,34 %, 9,57 %, 13,86% et 9,26 %. Tandis que les valeurs les plus basses obtenues avec les autres algorithmes étaient respectivement de 18,8 %, 13,03 %, 20,94 % et 17,74 %. Les résultats indiquent que la précision atteinte avec le modèle LSTM final dans les prévisions est suffisamment élevée pour permettre son utilisation dans le contrôle instantané des micro-réseaux.

Usharani [67] a présenté une fonction de perte améliorée dans le réseau neuronal LSTM pour prévoir la température de la surface de la mer dans différentes régions de l'Inde sur des horizons temporels quotidiens, hebdomadaires et mensuels. Le modèle LSTM avec la fonction de perte améliorée a atteint une précision de 98,7 %, surmontant les limitations des méthodes existantes et réduisant le temps de

traitement à moins de 0,35 seconde. Zhu et al. [49] ont proposé un modèle novateur utilisant le réseau LSTM (Long Short-Term Memory) optimisé par l'algorithme d'optimisation des lions fourmis (ALO) basé sur la décomposition en modes empiriques par ensemble (EEMD) et l'algorithme de regroupement K-means pour prédire la puissance PV. La précision MAPE (Mean Absolute Percentage Error) du modèle EEMD-Kmeans-LSTM est améliorée de 15,68 % par rapport au modèle EEMD-LSTM. La précision MAPE du modèle EEMD-Kmeans-ALO-LSTM est supérieure de 15,15 % à celle du modèle EEMD-Kmeans-LSTM. Limouni et al. [68] ont proposé une nouvelle approche pour prévoir la puissance PV en utilisant des modèles de réseau LSTM et de convolution temporelle (TCN). Les caractéristiques temporelles sont extraites des données d'entrée à l'aide du LSTM, puis les caractéristiques et les sorties sont connectées à l'aide du TCN. Dans différentes périodes de prévision de la puissance PV allant de 2 étapes à 7 étapes, le modèle LSTM-TCN a surpassé tous les modèles comparables en termes de performance. Dans Li et al. [69], les auteurs ont présenté un nouveau cadre d'apprentissage profond qui intègre la décomposition en paquets d'ondelettes (WPD) et les réseaux LSTM pour prévoir l'énergie PV une heure à l'avance en utilisant une fenêtre de 5 minutes. Le réseau LSTM est utile dans différentes applications. Pour prévoir la charge d'énergie agrégée et la génération d'énergie PV dans un microgrid communautaire, Wen et al.[70] ont développé un modèle de réseau neuronal récurrent profond (RNN) avec des unités LSTM, appelé DRNN-LSTM. Wang et al. [71] ont proposé un modèle hybride d'apprentissage profond combinant un réseau LSTM et un CNN pour la prévision de la puissance PV. Le modèle extrait d'abord les caractéristiques temporelles des données, puis les caractéristiques spatiales sont utilisées. Les résultats montrent que le modèle de prévision hybride offre de meilleurs résultats que la prédiction unique. Yildirim et al. [72] ont mis en œuvre quatre méthodes de séries temporelles, à savoir le réseau LSTM, le perceptron multicouche (MLP) et le système d'inférence neuro-flou adaptatif (ANFIS) avec partition en grille (GP) et c-means flou (FCM), pour prédire le rayonnement solaire à 1 heure à l'avance. Les résultats ont révélé que parmi les quatre modèles très précis, le modèle LSTM produit les meilleurs résultats pour la prévision du rayonnement solaire à 1 heure à l'avance. [73]. Dans Zhu et al. [74], un modèle amélioré de prévision basé sur le réseau neuronal BPNN (Back Propagation Neural Network) adaptatif a été développé pour prédire la puissance PV. Le modèle de prévision proposé a été testé sur deux ensembles de données du monde réel, et les résultats montrent que le modèle DRNN-LSTM surpasse un réseau de perception multicouche (MLP) et une machine à vecteurs de support (SVM). Alharkan et al. [75] ont développé un réseau

dual-stream CNN-LSTM basé sur l'apprentissage profond et un réseau de mécanisme d'auto-attention (DSCLANet). LSTM est utilisé pour extraire les caractéristiques temporelles et CNN est utilisé pour apprendre les motifs spatiaux. Les vecteurs de caractéristiques spatiales et temporelles de sortie sont ensuite combinés, et les meilleures caractéristiques sont choisies pour un traitement ultérieur à l'aide d'une méthode d'auto-attention.

2.5.7 Les réseaux de neurones récurrents bidirectionnels à mémoire court terme et long-terme ou BLSTM

BLSTM (Bidirectional LSTM) est un terme utilisé pour désigner un LSTM bidirectionnel couramment utilisé dans le traitement du langage naturel. En ce qui concerne la prévision des séries chronologiques, le BLSTM peut être plus performant que le LSTM. Le BLSTM est composé de deux LSTM fondamentaux [76] : un LSTM avant qui utilise les informations passées et un LSTM arrière qui utilise les informations futures, permettant ainsi d'utiliser les informations des instants temporels $t-1$ et $t+1$ à l'instant t . Généralement, le BLSTM est plus efficace que le LSTM et les RNN en général, car il peut utiliser à la fois les informations passées et futures. L'architecture du neurone BLSTM est illustrée dans la Fig.2.6. L'équation de calcul de y_t détermine la sortie du BLSTM à l'instant t . Cependant, l'équation précise de calcul de y_t n'est pas fournie dans le texte que vous avez mentionné. Veuillez fournir l'équation complète ou des détails supplémentaires pour une explication plus précise.

$$g_t = (W_y[h_t; h'_t] + b_y) \quad (2.13)$$

$$h_t = f(W_t[c_{t-1}; x] + b_t) \quad (2.14)$$

$$h'_t = f(W'_t[c'_{t-1}; x] + b'_t) \quad (2.15)$$

Dans cette équation, h'_t représente la sortie cachée de la cellule LSTM arrière (backward LSTM) et h_t représente la sortie cachée de la cellule LSTM avant (forward LSTM) à l'instant t . W'_t est la matrice de poids de la cellule LSTM arrière, W_t est la matrice de poids de la cellule LSTM avant à l'instant t . De plus, b'_t représente le vecteur de biais de la cellule LSTM arrière à l'instant t , et b_t représente le vecteur de biais de la cellule LSTM avant à l'instant t . En utilisant ces éléments, l'équation permet

de combiner les sorties des deux cellules LSTM (avant et arrière) pour obtenir la sortie finale y_t du modèle BLSTM à l'instant t . Cette approche bidirectionnelle permet au modèle de prendre en compte à la fois les informations passées et futures pour améliorer les performances de prédiction ou d'analyse des données.

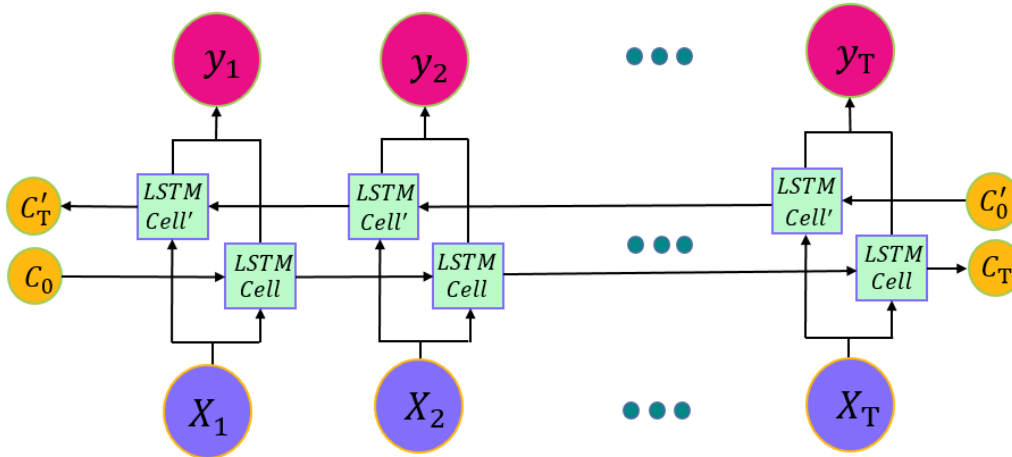


FIGURE 2.6 – Structure du réseau BLSTM.

2.5.8 Résultats bibliographiques obtenus par application de réseaux de neurones récurrents bidirectionnels à mémoire court terme et long-terme pour la prévision dans le domaine PV.

Alharbi et al. [77] ont mis au point un modèle à mémoire à court terme bidirectionnelle (BI-LSTM) qu'ils ont utilisé pour prédire les valeurs d'irradiance solaire des 169 prochaines heures. Ce modèle s'est basé sur des données horaires historiques provenant de la ville de Tabuk, recueillies sur la période allant du 01-01-1985 au 16-09-2020. Leurs résultats ont montré que le modèle BI-LSTM présentait des performances prometteuses en termes de classification et de considérations, avec une précision remarquable. De plus, le modèle a pu gérer efficacement la taille des données séquentielles sélectionnées. Les prévisions se sont avérées nettement plus précises et les performances du modèle BI-LSTM ont été considérablement améliorées lorsqu'ils ont intégré des données externes telles que la vitesse du vent et la température.

Zhen et al. [78] ont introduit un nouveau modèle de prédiction de la production photovoltaïque à ultra-court terme, basé sur le modèle bidirectionnel à mémoire à court terme amélioré par l'algorithme génétique (GA-BiLSTM), visant à améliorer les performances. Dans ce modèle, plusieurs séries de

production photovoltaïque sont innovamment utilisées comme entrées. Une étude de cas a été menée sur une station photovoltaïque réelle dans un micro-réseau. Une analyse de sensibilité des variables d'entrée a été effectuée et les performances du modèle GA-BiLSTM ont été comparées à celles d'autres modèles sur différents horizons temporels pour évaluer son efficacité. Les résultats mettent en évidence l'importance des séries de production des centrales photovoltaïques adjacentes, et le modèle proposé se démarque dans les prévisions à ultra-court terme, affichant des valeurs minimales de RMSE de 0.438, 0.806 et 1.118 respectivement pour les prédictions de sortie à 5 minutes, 15 minutes et 30 minutes à l'avance, sans recourir à des données météorologiques.

Xue et al. [79] ont développé une méthode de prévision de la production photovoltaïque à court terme en utilisant la technique MIC (coefficient d'information maximal) et le réseau neuronal à mémoire à court terme bidirectionnel. Ils ont d'abord analysé le degré de corrélation entre la production photovoltaïque prévue et les facteurs météorologiques à l'aide du coefficient d'information maximal (MIC) afin d'identifier les caractéristiques des données météorologiques en haute dimension dans le réseau électrique. Ensuite, ils ont sélectionné les jours historiques similaires en utilisant la somme pondérée du coefficient d'information maximal. Enfin, ils ont construit un réseau neuronal à mémoire à court terme bidirectionnel pour prédire la production photovoltaïque à court terme.

Peng et al. [80] ont proposé un modèle d'apprentissage profond basé sur le réseau neuronal à mémoire à court terme bidirectionnel, l'algorithme sinus-cosinus (SCA) et la décomposition empirique en modes empiriques complets avec bruit adaptatif (CEEMDAN) pour la prévision du rayonnement solaire. Dans un premier temps, le CEEMDAN est utilisé pour décomposer la série temporelle historique stochastique en certaines fonctions de mode intrinsèque périodique (IMF) et un résidu. Ensuite, les modèles antérieurs significatifs du rayonnement solaire des sous-modes décomposées sont identifiés à l'aide de deux techniques statistiques, à savoir la fonction d'autocorrélation (ACF) et la fonction d'autocorrélation partielle (PACF). Ensuite, toutes les sous-modes sont prédites à l'aide du modèle BiLSTM, et les paramètres du modèle BiLSTM sont optimisés grâce à l'algorithme SCA. Enfin, les sous-modes prédites sont agrégées pour générer le résultat final de la prévision.

2.5.9 Unités récurrentes avec portes ou GRU

Les couches GRU fonctionnent de la même manière que les LSTM, cependant, la cellule GRU utilise un seul état caché qui intègre les portes d'oubli et d'entrée en une seule porte de mise à jour.

Par conséquent, le GRU peut être considéré comme un mécanisme de régulation conceptuel avec un réseau plus efficace, moins de paramètres et un temps d'entraînement réduit par rapport au LSTM [81]. En d'autres termes, le GRU simplifie le processus en combinant certaines opérations des portes d'oubli et d'entrée en une seule porte de mise à jour. Cela réduit la complexité du modèle et le nombre de paramètres nécessaires, ce qui peut conduire à une formation plus rapide et plus efficace. la figure.2.7 montre la structure interne du GRU, qui comprend la porte de mise à jour unique et d'autres composants qui permettent au modèle de gérer les informations contextuelles et de prendre des décisions lors de la génération de séquences. Le module GRU reçoit deux entrées : l'état de la couche cachée h_{t-1} de l'instant précédent et l'entrée x_t de l'instant actuel. La structure du module GRU comprend deux fonctions de porte : la porte de mise à jour (update gate) z_t et la porte de réinitialisation (reset gate) r_t . Ainsi, le GRU est une cellule LSTM largement utilisée et simplifiée. L'état caché du module GRU est mis à jour en utilisant l'équation suivante (formule non visible) : Cette équation permet de calculer le nouvel état caché du GRU en utilisant l'état précédent h_{t-1} , l'entrée actuelle x_t , la porte de réinitialisation r_t et la porte de mise à jour z_t . Les portes de réinitialisation et de mise à jour contrôlent les informations à conserver ou à oublier lors de la mise à jour de l'état caché.

$$h_t = h_{t-1} * (1 - z_t) + z_t * \tilde{H}_t \quad (2.16)$$

Cette équation permet de calculer la valeur de la porte de mise à jour, qui détermine la quantité de l'unité GRU qui sera mise à jour lors de la propagation avant dans le réseau. La porte de mise à jour contrôle l'importance relative des informations précédentes et des nouvelles informations pour la mise à jour de l'état caché. Plus précisément, cette équation prend en compte l'état précédent de l'unité GRU h_{t-1} , l'entrée actuelle (x_t) ainsi que les poids et les biais appropriés pour calculer la valeur de la porte de mise à jour. Le résultat de cette équation est généralement une valeur entre 0 et 1, où une valeur proche de 0 indique que l'unité GRU est peu mise à jour et une valeur proche de 1 indique une mise à jour significative.

$$z_t = \sigma(W_z * [h_{t-1}, x_t]) \quad (2.17)$$

La porte de réinitialisation (reset gate) est calculée de la même manière que la porte de mise à

jour (update gate), en utilisant la formule suivante :

$$r_t = \sigma(W_r * [h_{t-1}, x_t]) \quad (2.18)$$

Cette équation permet de calculer la valeur de la porte de réinitialisation, qui détermine dans quelle mesure l'unité GRU réinitialise son état interne. Tout comme pour la porte de mise à jour, cette équation prend en compte l'état précédent de l'unité GRU (h_{t-1}), l'entrée actuelle (x_t) ainsi que les poids et les biais appropriés pour calculer la valeur de la porte de réinitialisation. Le résultat de cette équation est également généralement une valeur entre 0 et 1, où une valeur proche de 0 signifie une réinitialisation faible de l'unité GRU et une valeur proche de 1 indique une réinitialisation significative. La fonction tangente hyperbolique est appliquée à la porte de réinitialisation (reset gate), qui est définie par l'équation suivante, pour calculer le candidat pour l'état caché :

$$\tilde{H}_t = \tanh(W * [r_t * h_{t-1}, x_t]) \quad (2.19)$$

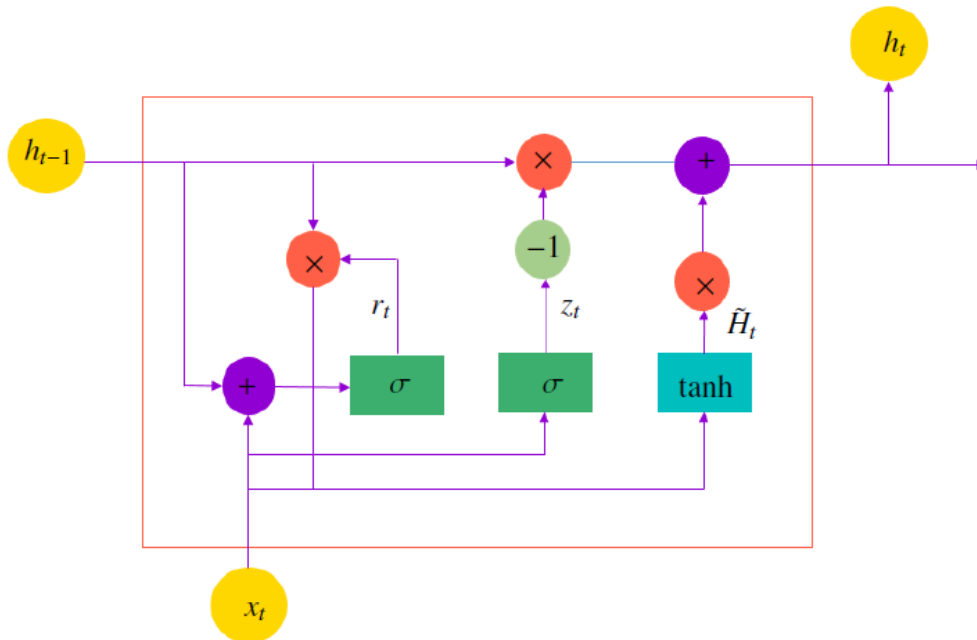


FIGURE 2.7 – La structure interne du GRU

Dans le contexte de la porte de réinitialisation (reset gate), l'activation ReLU est utilisée, représentée par la fonction σ , avec une plage de valeurs entre 0 et 1. Les poids W_r , W_z et W sont utilisés

pour régler respectivement la porte de réinitialisation, la porte de mise à jour et le coefficient de poids. Le symbole $*$ représente le produit scalaire entre vecteurs. En d'autres termes, les poids W_r , W_z et W sont des paramètres qui ajustent l'importance relative de la porte de réinitialisation, de la porte de mise à jour et des entrées dans le calcul de la porte de réinitialisation. Ces poids sont utilisés pour effectuer une combinaison linéaire des valeurs associées à la porte de réinitialisation, à la porte de mise à jour et aux entrées. La fonction d'activation ReLU est une fonction qui transforme les valeurs négatives en 0 et les valeurs positives restent inchangées. Cependant, dans le contexte de la porte de réinitialisation, cette fonction est modifiée pour que sa plage de valeurs soit entre 0 et 1, afin de représenter la probabilité de réinitialisation.

2.5.10 Résultats bibliographiques obtenus par application des Unités récurrentes avec portes pour la prévision dans le domaine PV.

Liao et al. [82] ont amélioré la précision des prévisions de production photovoltaïque à court terme en utilisant un réseau GRU. Ils ont d'abord employé le coefficient de Pearson pour identifier les principales caractéristiques influant sur la production photovoltaïque future, ce qui leur a permis d'analyser qualitativement la relation entre la production historique et celle à venir. Ensuite, les ensembles d'entraînement ont été regroupés en fonction des similarités de chaque caractéristique, avant d'appliquer l'entraînement du réseau GRU à chaque groupe. Les sorties de chaque réseau GRU ont été moyennées pour obtenir la production photovoltaïque du prochain moment.

Mellit et al. [48] ont exposé différentes architectures de réseaux de neurones profonds (DLNN) conçues pour prédire la puissance de sortie photovoltaïque sur divers horizons temporels (1 minute, 5 minutes, 30 minutes et 60 minutes), tant en prévision à un pas qu'en prévision à plusieurs pas en avant. Leur étude a mis en évidence qu'une architecture de DLNN simple, telle que LSTM ou GRU, peut offrir une grande précision ($r = 99\%$) pour la prévision à un pas en avant. Des résultats satisfaisants ont également été obtenus pour la prévision à plusieurs pas en avant ($r = 96,9\%$, par exemple, pour 8 pas en avant). Il est important de noter que des paramètres tels que la taille du lot, le nombre d'unités, le nombre de couches cachées, la taille des filtres, le taux d'abandon et la taille du noyau doivent être méticuleusement sélectionnés, car leur valeur varie d'un modèle à l'autre. Il a été confirmé qu'une vaste base de données est essentielle pour obtenir des résultats précis."

Kim et al. [83] Ont introduit deux modèles de prédiction de la production d'énergie PV en utilisant

les architectures LSTM et GRU, sans nécessiter d'informations météorologiques futures, contrairement aux méthodes précédentes. Dans leurs essais numériques, ils ont évalué les performances de ces modèles par rapport aux modèles conventionnels, y compris ceux basés sur les réseaux de neurones artificiels (ANN) et les réseaux de neurones profonds (DNN), en utilisant des ensembles de données du monde réel. Le modèle basé sur l'ANN montre des performances inférieures dans la plupart des configurations expérimentales. En revanche, leur modèles proposés affichent des performances très satisfaisantes. Notamment, le modèle basé sur GRU qu'ils ont présenté excelle dans la prédiction de la production PV, surtout dans les zones de pic, où les défis sont plus complexes.

Sodsong et al. [84] ont proposé un modèle de prévision de l'énergie solaire basé sur plusieurs réseaux à unités récurrentes à portes (GRU) afin d'améliorer significativement la précision des prédictions et de réduire le temps d'entraînement par rapport au réseau GRU classique. De plus, ils ont mis en œuvre d'autres algorithmes populaires d'apprentissage automatique pour la prédiction, tels que le réseau de neurones artificiels à propagation avant (ANN), la régression par vecteurs de support (SVR) et la méthode des k plus proches voisins (KNN), pour les comparer à leur modèle. Chacun de ces modèles a été évalué à l'aide de l'erreur quadratique moyenne normalisée (NRMSE). Les résultats ont montré que leur modèle proposé, ainsi que le GRU, l'ANN à propagation avant, le SVR et le KNN, ont affiché respectivement des taux de NRMSE de 9,64 %, 10,53 %, 11,62 %, 11,45 % et 11,89%."

2.5.11 Autoencodeurs

Les autoencodeurs [85] sont un type de réseau neuronal artificiel qui vise à produire une représentation complexe des connaissances ou une approximation de l'entrée donnée. Le réseau tente de produire une sortie \tilde{X} qui ressemble à l'entrée X et peut être divisée en deux parties ou phases [85, 86]. L'architecture du réseau d'autoencodeurs est illustrée dans la figure.2.8. Les autoencodeurs sont utilisés pour apprendre des caractéristiques latentes dans les données en comprimant et en décompressant l'information. Ils se composent généralement de deux parties : l'encodeur, qui transforme l'entrée en une représentation réduite, et le décodeur, qui tente de reconstruire l'entrée à partir de cette représentation réduite. L'objectif est de minimiser la différence entre l'entrée d'origine et la sortie reconstruite, ce qui permet au réseau d'apprendre à extraire les caractéristiques les plus importantes et de créer une représentation compacte de l'entrée. L'architecture spécifique de l'autoencodeur peut varier en fonction des besoins et des objectifs de l'application. Cependant, l'idée centrale est de créer un réseau

2.5. DIFFÉRENTES FAMILLES DE RÉSEAUX DE NEURONES

capable de représenter les données de manière compréhensible et utile, ce qui peut être utilisé pour des tâches telles que la compression de données, la réduction de dimensionnalité ou la génération de données similaires. L'encodeur (Encoder) a pour objectif de construire des caractéristiques génériques du vecteur d'entrée X . Sa fonction est définie dans l'équation. (2.20).

$$k = f(X) = \sigma(W_1 \cdot X + b_1) \quad (2.20)$$

Le décodeur (Decoder) a pour objectif de recréer l'entrée à partir de la représentation abstraite générée par la fonction de l'encodeur.

$$\tilde{X} = g(k) = \sigma(W_2 \cdot k + b_2) \quad (2.21)$$

W_1 représente le poids et b_1 représente le biais de l'encodeur, tandis que W_2 représente le poids et b_2 représente le biais du décodeur. Ces poids et biais sont des paramètres ajustables qui permettent au modèle d'apprendre et de représenter efficacement les données. Ce type de modèle, connu sous le nom d'autoencodeur empilé (stacked autoencoder), est utilisé pour améliorer les performances en modifiant les couches cachées de différentes manières. Lorsqu'un réseau neuronal est profond, c'est-à-dire qu'il comporte plusieurs couches cachées, un autoencodeur empilé est utilisé pour résoudre le problème de disparition du gradient en empilant des couches cachées.

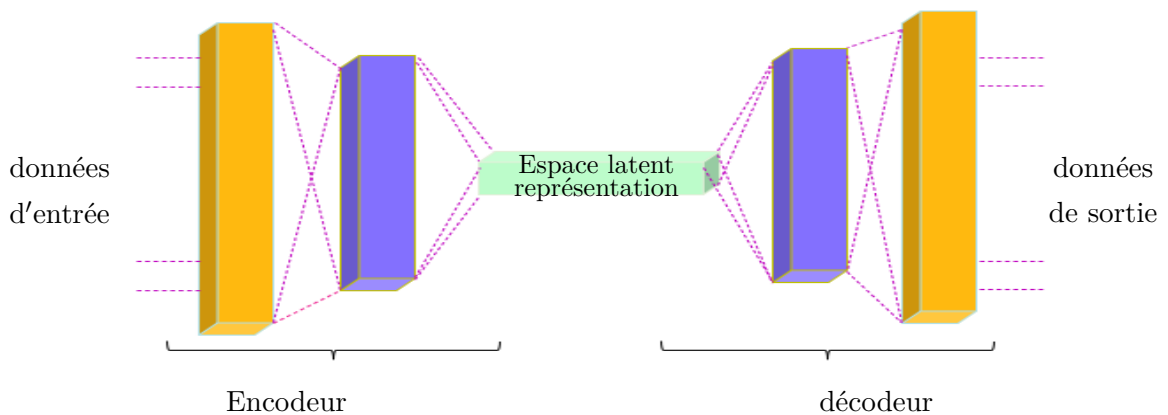


FIGURE 2.8 – Architecture du réseau autoencodeur.

LSTM- Autoencodeur

Le modèle LSTM-Autoencodeur [87] est implémenté pour les données séquentielles en utilisant deux réseaux LSTM : l'encodeur (encoder) et le décodeur (decoder). L'encodeur lit des séquences de longueur variable, tandis que le décodeur génère des séquences de longueur variable, à chaque instant t , à travers l'équation (2.22).

$$h_t, c_t = LSTM(x_t, h_{t-1}, c_{t-1}) \quad (2.22)$$

Ce modèle utilise l'architecture des réseaux LSTM pour capturer les dépendances à long terme dans les séquences de données. Les LSTM sont particulièrement efficaces pour modéliser et prédire des séquences complexes grâce à leur capacité à gérer les informations contextuelles et à mémoriser des informations importantes sur de longues périodes. L'encodeur prend une séquence en entrée et la transforme en une représentation latente de dimension réduite, capturant les caractéristiques les plus importantes de la séquence. Le décodeur prend ensuite cette représentation latente et tente de la reconstruire pour générer une séquence de sortie qui ressemble autant que possible à l'entrée d'origine. En utilisant cette architecture, le modèle LSTM-autoencodeur est capable de compresser et de décompresser des séquences de données, permettant ainsi de les représenter de manière plus compacte tout en conservant les informations essentielles. Cela peut être utile pour la réduction de dimensionnalité, la génération de séquences similaires ou la détection d'anomalies dans les données séquentielles. La figure.2.9 illustre le cadre général des autoencodeurs LSTM et montre comment l'encodeur et le décodeur interagissent pour traiter les données séquentielles. L'encodeur itère en entrant séquentiellement la séquence x_1, \dots, x_T de longueur T et met à jour le vecteur d'état de la mémoire cellulaire c_t et le vecteur d'état caché h_t à chaque étape. L'encodeur résume l'ensemble de la séquence d'entrée en utilisant ces mises à jour successives pour obtenir le vecteur d'état final de la mémoire cellulaire c_T et le vecteur d'état caché h_t après T étapes de temps. À chaque étape de temps t , l'encodeur traite l'entrée x_t et met à jour le vecteur d'état de la mémoire cellulaire c_t et le vecteur d'état caché h_t en fonction de l'entrée actuelle et des états précédents. Ces mises à jour permettent à l'encodeur de capturer les informations contextuelles et les dépendances temporelles dans la séquence d'entrée. À la fin des T étapes de temps, l'encodeur a parcouru toute la séquence d'entrée et a résumé les informations pertinentes dans le vecteur d'état final de la mémoire cellulaire

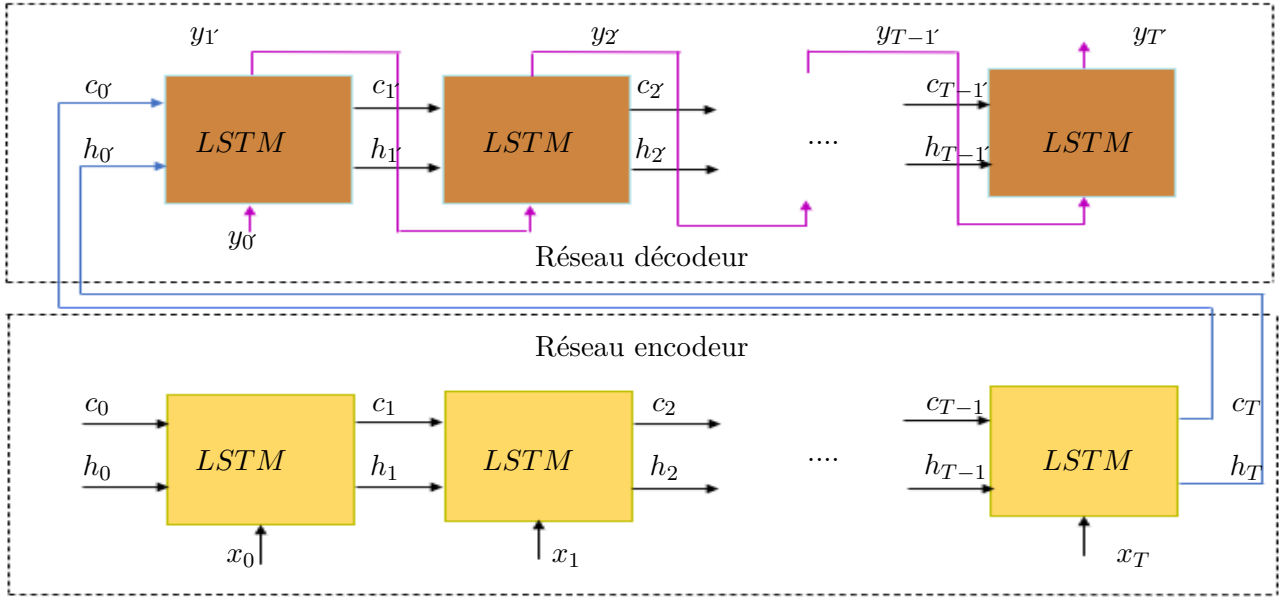


FIGURE 2.9 – Réseau LSTM-Autoencodeur .

c_T et le vecteur d'état caché h_t . Ces vecteurs représentent une représentation condensée de la séquence d'entrée, contenant les informations clés nécessaires pour la tâche ultérieure, telle que la génération de séquences, la classification ou la prédiction.

$$Y_t = W_l h_t, \text{ where } W_l \in R^{k \times 1}, Y_t = [y_1 \dots y_k]^T \quad (2.23)$$

$$y_t = \frac{e^{y_t}}{\sum_{j=1}^k e^{y_j}} \quad (2.24)$$

$$y_t = \max(y_t) \quad (2.25)$$

Pour la génération d'une séquence de longueur T , le décodeur utilise le vecteur d'état caché final de l'encodeur h_T et le vecteur d'état de mémoire cellulaire final c_T comme vecteurs initiaux ($h_0 = h_T$) et ($c_0 = c_T$). Le décodeur génère des séquences en utilisant d'abord une entrée factice y_0 et les vecteurs initiaux h_0 et c_0 pour produire h_1 et c_1 en utilisant l'équation (2.22). Ensuite, les équations (2.23), (2.24) and (2.25) sont utilisées pour calculer y_1 , qui représente l'endroit où une personne se rendra. Pour simplifier la représentation des équations (2.23), (2.24) and (2.25), nous utilisons la formule de l'équation (2.26). De même, la séquence de sortie y_1, \dots, y_T est créée de manière récursive en passant y_{t-1} , c_{t-1} et h_{t-1} aux équations (2.22) et Eq.(2.26). Il est important de noter que le décodeur LSTM génère y_t à partir de y_{t-1} . Si y_{t-1} est calculé incorrectement, l'estimation des valeurs suivantes peut

être affectée négativement.

$$y_t = f(y_t) \quad (2.26)$$

Les encodeurs-décodeurs LSTM sont conçus pour simuler la probabilité conditionnelle de la séquence de sortie étant donné la séquence d'entrée, c'est-à-dire $p(y_1, \dots, y_T | x_1, \dots, x_T)$. À travers l'état cellulaire LSTM c_T , l'encodeur offre un résumé de la séquence d'entrée x_1, \dots, x_T . Étant donné l'état cellulaire c_T de l'encodeur. La probabilité conditionnelle :

$$p(y_1, \dots, y_T | x_1, \dots, x_T) \approx \prod_{t=1}^T p(y_t | c_T, y_1, \dots, y_{t-1}) \quad (2.27)$$

Étant donné l'état cellulaire du décodeur c_{t-1} et le $(t - 1)$ -ième échantillon de la séquence de sortie y_{t-1} , le décodeur génère la distribution de probabilité $p(y_t | c_{t-1}, y_{t-1})$ étape par étape.

$$p(y_1, \dots, y_T | x_1, \dots, x_T) \approx \prod_{t=1}^T p(y_t | c_{t-1}, y_{t-1}) \quad (2.28)$$

la valeur réelle de l'échantillon de sortie précédent est inconnue du décodeur. En conséquence, le décodeur décide de y_t à chaque étape de décodage en se basant sur la distribution de probabilité $p(y_t | c_{t-1}, y_{t-1})$ obtenue à partir de la sortie du décodeur et utilise cette décision provisoire pour la mise à jour de l'état suivant du décodeur.

2.5.12 Résultats bibliographiques obtenus par application de autoencodeur pour la prévision dans le domaine PV.

Gensler et al. [88] ont utilisé différents algorithmes d'apprentissage en profondeur et de réseaux neuronaux artificiels, tels que les Deep Belief Networks, AutoEncoder et LSTM, pour introduire ces puissants algorithmes dans le domaine de la prévision de l'énergie renouvelable. Dans leurs expériences, ils ont combiné ces algorithmes pour démontrer leur puissance de prévision par rapport à un MLP standard et un modèle de prévision physique dans la prévision de la production d'énergie de 21 centrales solaires. Leurs résultats avec les algorithmes d'apprentissage en profondeur montrent des performances de prévision supérieures par rapport aux réseaux neuronaux artificiels ainsi qu'à d'autres modèles de référence tels que les modèles physiques.

Yank et al. [89] ont développé une approche basée sur un auto-encodeur à mémoire à court terme longue pour générer des scénarios représentatifs dans un système intégré de production d'énergie hydro-photovoltaïque. Cette méthode implique l'extraction des caractéristiques par l'encodeur LSTM, le regroupement des scénarios dans le domaine des caractéristiques en combinant la méthode des statistiques de l'écart et K-means++, ainsi que la reconstruction des scénarios représentatifs à l'aide du décodeur LSTM. Comparée aux méthodes traditionnelles de sélection et de génération de scénarios, cette approche permet de mieux capturer les motifs des données multivariées en séries temporelles dans les dimensions temporelles et spatiales. Une étude de cas menée dans le sud-ouest de la Chine démontre l'efficacité de cette méthode, qui surpasse les autres méthodes existantes en atteignant des indices SSE et DBI aussi bas que 0,89 et 0,12 respectivement. De plus, elle obtient les meilleurs scores SIL et CHI de 0,93 et 2,30 respectivement. L'étude de cas montre également que la configuration du modèle proposé fonctionne de manière plus stable pour la génération de scénarios.

Aksan et al. [90] ont relevé le défi de la prévision bidirectionnelle multistep du flux d'énergie en introduisant un modèle d'autoencodeur LSTM. Au cours de l'entraînement, le modèle proposé ainsi que les modèles de référence ont été développés en utilisant l'ajustement automatique des hyperparamètres pour affiner les modèles et optimiser leur performance. Ils ont exploité les données des 6 dernières heures précédant le moment actuel (24 intervalles de 15 minutes) pour prédire le flux d'énergie 1 heure à l'avance (4 intervalles de 15 minutes) à partir de l'instant présent. Lors de l'évaluation du modèle, le modèle proposé a affiché les scores de RMSE et de MAE les plus bas, avec des valeurs respectives de 32,243 MW et 24,154 MW.

2.6 Métriques pour l'évaluation des modèles

Dans le cadre de cette thèse, nous présenterons les métriques utilisées pour évaluer nos modèles.

Erreur absolue moyenne :

Elle peut être calculée en prenant la moyenne des erreurs absolues entre les valeurs réelles et prévues de la génération d'énergie solaire PV, comme indiqué ci-dessous :

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \tilde{y}_i| \quad (2.29)$$

Erreur quadratique moyenne :

2.6. MÉTRIQUES POUR L'ÉVALUATION DES MODÈLES

Elle peut être représentée par l'écart-type des valeurs réelles et prévues de la génération d'énergie solaire PV, comme indiqué ci-dessous :

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^n (y_i - \tilde{y}_i)^2} \quad (2.30)$$

Erreur moyenne de biais : Le MBE (Mean Bias Error) est une mesure de l'écart moyen entre les valeurs réelles et prévues de la génération d'énergie solaire PV, et il représente l'erreur systématique ou le biais. Les valeurs positives de MBE indiquent une surestimation du modèle, tandis que les valeurs négatives indiquent une sous-estimation du modèle. Le MBE est décrit comme suit :

$$MBE = \frac{1}{N} \sum_{i=1}^N (\tilde{y}_i - y_i) \quad (2.31)$$

Erreur quadratique moyenne : Elle peut être représentée par la différence quadratique moyenne entre les valeurs réelles et prévues de la génération d'énergie solaire PV, comme indiqué ci-dessous :

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \tilde{y}_i)^2 \quad (2.32)$$

Le coefficient de détermination peut être calculé mathématiquement en utilisant la formule suivante :

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \tilde{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y}_i)^2} \quad (2.33)$$

où N représente le nombre de prévisions, y_i est la puissance solaire PV réelle, \tilde{y}_i est la puissance solaire PV prévue, et \bar{y}_i est la moyenne de la puissance solaire PV réelle dans l'ensemble de test. Lorsque R^2 est proche de 1, le modèle de prédiction est plus précis. Plus les valeurs de MSE, MAE, MBE et RMSE sont petites, meilleures sont les performances de prédiction.

Ces indicateurs statistiques facilitent la comparaison objective des performances des différents modèles de prévision. Le critère du pourcentage d'erreur de promotion est utilisé pour comparer les performances des deux modèles de prédiction, qui sont décrits comme suit :

Le pourcentage d'amélioration du MAE (P_{MAE})

$$P_{MAE} = \frac{MAE_1 - MAE_2}{MAE_1} \times 100\% \quad (2.34)$$

Le pourcentage d'amélioration du MSE (P_{MSE})

$$P_{MSE} = \frac{MSE_1 - MSE_2}{MSE_1} \times 100\% \quad (2.35)$$

Le pourcentage d'amélioration du RMSE (P_{PRMSE})

$$P_{PRMSE} = \frac{RMSE_1 - RMSE_2}{RMSE_1} \times 100\% \quad (2.36)$$

MAE_1 , MAE_2 , MSE_1 , MSE_2 , $RMSE_1$, et $RMSE_2$ désigne la moyenne absolue erreur quadratique moyenne et erreur quadratique moyenne des modèles 1 et 2, respectivement.

2.7 conclusion

Dans ce chapitre de l'exploration approfondie des réseaux de neurones : typologies, familles et métriques d'évaluation des Modèles, nous avons exploré les principaux aspects des réseaux de neurones, un domaine clé de l'apprentissage automatique. Nous avons examiné les différentes composantes des réseaux de neurones, notamment les neurones, les fonctions d'activation, l'architecture et les différentes familles de réseaux de neurones. nous avons d'abord exploré la classification des prévisions de puissance PV en fonction de l'horizon de prévision a souligné l'importance de comprendre les différentes échelles de temps impliquées dans les prévisions. Cette compréhension fine permet d'ajuster les stratégies opérationnelles en conséquence, améliorant ainsi la réactivité des systèmes PV aux fluctuations de la demande et des conditions environnementales. par la suite, l'exploration des diverses techniques de prévision, allant des méthodes physiques aux méthodes statistiques et aux techniques d'intelligence artificielle, a offert une perspective complète sur les approches disponibles pour prédire la puissance PV. Chaque méthode a ses avantages et ses limites, et leur évaluation comparative fournit des indications cruciales pour choisir la méthode la plus appropriée en fonction du contexte et des objectifs spécifiques. Ensuite, nous avons étudié les neurones, les unités fondamentales des réseaux de neurones, qui reçoivent des entrées pondérées, appliquent des fonctions d'activation et génèrent des sorties. Ensuite, nous avons exploré les fonctions d'activation couramment utilisées, telles que la fonction sigmoïde, la fonction ReLU et la fonction tangente hyperbolique, qui introduisent de la non-linéarité dans les

2.7. CONCLUSION

modèles de réseaux de neurones. Nous avons ensuite abordé l'architecture des réseaux de neurones, en mettant en évidence les différentes couches et leur organisation. Nous avons également discuté des différentes familles de réseaux de neurones, en mettant l'accent sur les réseaux de neurones convolutionnels et les réseaux de neurones récurrents, qui sont adaptés à des types spécifiques de données et de problèmes. Nous avons également examiné les autoencodeurs, des réseaux de neurones utilisés pour la compression et la reconstruction de données, ainsi que le LSTM-autoencodeur, une variante qui intègre des mécanismes de mémoire à long terme pour capturer des dépendances temporelles. Dans la suite du chapitre nous avons présenté les résultats qui ont été obtenus par l'utilisation de ces méthodes. Enfin, nous avons exploré les métriques utilisées pour évaluer les performances des modèles de réseaux de neurones, notamment l'erreur quadratique moyenne, l'erreur absolue moyenne et d'autres métriques pertinentes. Dans l'ensemble, les réseaux de neurones offrent des capacités puissantes pour l'apprentissage automatique et la modélisation de données complexes. Ils permettent d'extraire des informations significatives à partir de grandes quantités de données et de résoudre des problèmes de prédiction et de classification. Les différentes familles de réseaux de neurones, tels que les CNN, RNN, LSTM, BLSTM et les GRU, ainsi que les techniques comme les autoencodeurs, offrent des outils spécifiques pour traiter différents types de données et résoudre différents types de problèmes. En conclusion, les réseaux de neurones continuent de jouer un rôle essentiel dans le domaine de l'apprentissage profond, et leur utilisation continue d'évoluer et de s'étendre. Ils offrent des perspectives prometteuses pour des applications dans divers domaines, de l'imagerie médicale à la traduction automatique, en passant par la reconnaissance de la parole et bien d'autres encore. La maîtrise des concepts et des techniques des réseaux de neurones est donc essentielle pour exploiter pleinement leur potentiel et repousser les limites de l'intelligence artificielle.

Chapitre 3

Exploration Complète du Site : Caractérisation, Acquisition, Sélection et Analyse des Données de Surveillance

3.1 Introduction

L'énergie solaire PV a connu une croissance spectaculaire ces dernières décennies en tant que source d'énergie renouvelable propre et durable. La conversion directe de la lumière solaire en électricité par le biais de panneaux PV offre une solution prometteuse pour répondre aux défis mondiaux de l'approvisionnement énergétique et de l'atténuation du changement climatique. Toutefois, pour tirer pleinement parti du potentiel des systèmes PV, il est essentiel de comprendre, d'analyser et d'optimiser leurs performances. Dans le cadre de cette thèse de doctorat, notre objectif est de développer des techniques de prévision de la production d'énergie PV en utilisant les réseaux de neurones, une approche basée sur l'intelligence artificielle qui a démontré son efficacité dans divers domaines de prédiction.

L'analyse des performances pour optimiser les systèmes PV est le premier axe, notre objectif est d'explorer en profondeur les performances des systèmes PV existants. Nous cherchons à comprendre les aspects techniques, structurels et environnementaux qui influent sur l'efficacité des panneaux solaires. En analysant ces performances de manière détaillée. Le deuxième axe de notre recherche se concentre sur la production effective d'énergie PV. Nous étudierons les facteurs qui influent sur la production, tels que les conditions météorologiques, l'inclinaison et l'orientation des panneaux solaires, ainsi que d'autres variables environnementales. En analysant ces paramètres en profondeur, nous visons à développer des modèles de prédiction plus précis, capables d'anticiper avec fiabilité la production d'énergie solaire dans diverses conditions. Avant d'approfondir dans la prévision des données, il est primordial de comprendre les caractéristiques du site d'étude. Le premier axe de notre recherche portera sur la caractérisation du site, qui englobe l'évaluation des conditions géographiques, météorologiques et environnementales spécifiques à l'installation PV. En analysant ces éléments, nous pourrions mieux appréhender les variations potentielles de la production d'énergie, fournissant ainsi des informations cruciales pour une prédiction précise et fiable. Pour mener à bien notre étude de prévision, il est indispensable de collecter des données de surveillance de haute qualité. Le deuxième axe de notre travail se focalisera sur l'acquisition de données de surveillance, où nous explorerons les différentes méthodes et technologies utilisées pour recueillir des données en temps réel auprès Par la suite, nous aborderons la sélection des données d'entrée, où nous discuterons des critères et des techniques utilisées pour choisir les variables pertinentes qui auront un impact significatif sur nos modèles de prévision de production d'énergie PV. Cette étape est cruciale pour éviter le surdimensionnement du système ou les pertes de

3.2. ANALYSE DES PERFORMANCES POUR OPTIMISER LES SYSTÈMES PHOTOVOLTAÏQUES

rendement. La présentation des données et l'analyse préliminaire constitueront une autre étape clé de notre investigation. Nous utiliserons des outils statistiques et des méthodes d'analyse de données pour explorer les tendances et les schémas de production, nous permettant ainsi de déceler d'éventuelles incohérences et d'identifier les principaux facteurs qui influencent les performances du système.

3.2 Analyse des performances pour optimiser les systèmes photovoltaïques

Lors de sa transition d'une source d'énergie de niche à une source d'énergie courante, l'énergie PV doit faire face à la concurrence économique et à la fiabilité des sources d'énergie établies et des techniques de fabrication existantes. La fiabilité opérationnelle du système englobe l'ensemble du cycle de vie des installations PV, allant de la production de matériaux de cellules solaires de haute qualité à la conception et à la construction des centrales solaires, puis jusqu'au démantèlement et au recyclage tant de la centrale que de ses composants individuels. En conséquence, la phase la plus longue du cycle de vie est de loin la période d'exploitation et de maintenance (E&M), qui garantit le fonctionnement quotidien optimal du système. Les activités d'E&M, ainsi que les études de performance PV qui y sont associées, reposent sur l'utilisation de mesures de données PV de haute qualité. La phase d'exploitation et de maintenance (E&M) se distingue ainsi comme la plus longue étape du cycle de vie, assurant de manière prédominante le fonctionnement optimal quotidien du système. Les activités d'E&M, conjointement avec les études de performance PV qui y sont liées, requièrent impérativement l'utilisation de mesures de données PV de qualité supérieure. Alors que cela devient de plus en plus fréquent pour les grandes centrales PV, la plupart des systèmes PV installés sont évalués en termes de performances à des intervalles de temps spécifiques, voire seulement lorsqu'un problème de production est détecté. Entre-temps, il est fréquent que les données soient enregistrées sans être soumises au processus de contrôle de qualité. L'étude intitulée "Analytical Monitoring of Grid-connected Photovoltaic Systems" [91], menée par la Tâche 13 du Programme de systèmes d'énergie solaire PV de l'AIE PVPS (Agence internationale de l'énergie - Programme de systèmes d'énergie solaire PV, offre une vue d'ensemble détaillée de l'analyse et de la visualisation des données de performances mesurées sur le terrain. Les chercheurs Van Sark et al. [92] récapitulent les éléments clés de la surveillance et de la caractérisation des systèmes PV. La société 3E propose, quant à elle, un service de vérification des données de capteurs avec des diagnostics avancés pour les capteurs de rayonnement solaire [93]. En conséquence, ils utilisent des données d'irradiance satellitaires pour effectuer des comparaisons. Des instituts de recherche renommés

se concentrent sur l'analyse statistique des données de surveillance et ont mis au point des algorithmes automatisés afin de repérer les données erronées et manquantes dans les ensembles de données PV. Des exemples notables incluent Pecos [94] des Laboratoires nationaux Sandia ou un rapport technique du NREL. Killinger et al. [95] ont également présenté un exemple dans lequel des modèles de ciel clair et de puissance PV sont utilisés pour détecter les données défectueuses de puissance de sortie. Une revue exhaustive des méthodes de surveillance des systèmes PV et de détection des défauts est proposée par Livera et al. [96], qui fournit des directives précises concernant les paramètres de mesure requis et leurs incertitudes maximales.

3.3 Production photovoltaïques

L'énergie PV est créée à partir du rayonnement solaire que la terre reçoit. Le rayonnement est une onde électromagnétique qui se propage par la surface du soleil à la suite de réactions de fusion de l'hydrogène à l'hélium dans le noyau. Le soleil émet environ 3.89 MJ d'énergie nucléaire par seconde, qui est rapidement convertie en énergie thermique [3]. Ce flux d'énergie est transporté à la surface et diffusé sous forme de rayonnement électromagnétique. La quantité d'énergie émise par le soleil est d'environ $64 \text{ MW}/m^2$. Seule une petite partie de cette énergie ($1.7 \text{ MW}/m^2$), après avoir parcouru plus de $150 \times 10^6 \text{ km}$, a atteint le sommet de l'atmosphère terrestre sans être absorbée par l'espace. La quantité reçue par la terre est connue sous le nom de constante solaire. Cependant, elle ne peut pas être pleinement utilisée, en raison de la réflexion de l'onde reçue et donc de l'énergie résultante qu'il transporte [3]. Le mot albédo fait référence à ce phénomène de réflexion. La part de rayonnement solaire dirigée vers l'atmosphère est désignée par le symbole albédo et est exprimée par un facteur entre 0 et 1. La nature de la surface réfléchissante affecte l'albédo. Par exemple, la neige ou les nuages ont un albédo de 0.40 à 0.90. tandis qu'une surface sombre a un albédo de 0.15, l'albédo moyen de la terre est d'environ 0.3 ce qui signifie que 30% de l'énergie solaire reçue par la terre sera réfléchi [97].

Le rayonnement solaire incident relatif aux panneaux solaires est composé de trois types de rayonnement [3] voir figure.3.1 :

Le terme rayonnement diffus du ciel fait référence au rayonnement diffusé par la suspension dans l'air, les aérosols, etc. Le terme rayonnement diffus du sol fait référence au rayonnement solaire réfléchi par la surface terrestre (albédo). Le terme rayonnement direct fait référence au rayonnement non

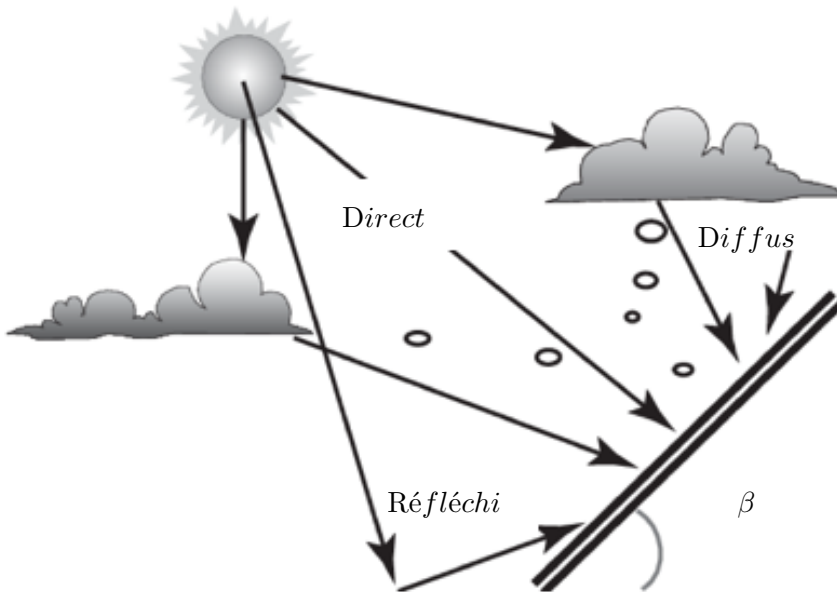


FIGURE 3.1 – Représentation des différents rayonnements solaires terrestres reçus sur une installation donnée [3]

réfléchissant et non diffusé qui frappe directement la surface des panneaux solaires en ligne droite.

Nous appelons le rayonnement global est la combinaison des trois types de rayonnement [3]. En Australie a un fort potentiel solaire, plus de la moitié du pays a plus de $2100 \text{ kWh}/\text{m}^2$, et une grande partie du nord-est a plus de $2300 \text{ kWh}/\text{m}^2$, un niveau que l'on ne trouve que dans les régions les plus ensoleillées du Sahara, de l'Arabie et du nord du Chili. La carte 3.2 illustre la répartition de cette irradiation sur tout le territoire de l'Australie. Pour l'évaluation des ressources solaires de la région, l'évaluation de la viabilité financière et l'analyse correcte des performances des centrales PV tout au long de leur durée de vie opérationnelle, la connaissance de la quantité et de la fiabilité de GHI est cruciale. Les données des satellites météorologiques géostationnaires (gérés par JMA) ainsi que les informations des modèles atmosphériques et météorologiques mondiaux sont utilisées pour calculer la quantité de rayonnement solaire dans la zone (exploités par l'ECMWF, la NOAA et la NASA).

Les cellules PV sont utilisées dans la conversion de cette énergie solaire en énergie électrique. Une cellule PV est un composant électrique constitué de matériaux semi-conducteurs qui absorbent une quantité importante de rayonnement solaire. Les cellules les plus courantes sont constituées de deux couches de silicium : une couche de type P dopée au bore qui a moins d'électrons que le silicium et une couche de type N dopée au phosphore qui a plus d'électrons que le silicium. Ainsi, lorsqu'un

3.3. PRODUCTION PHOTOVOLTAÏQUES

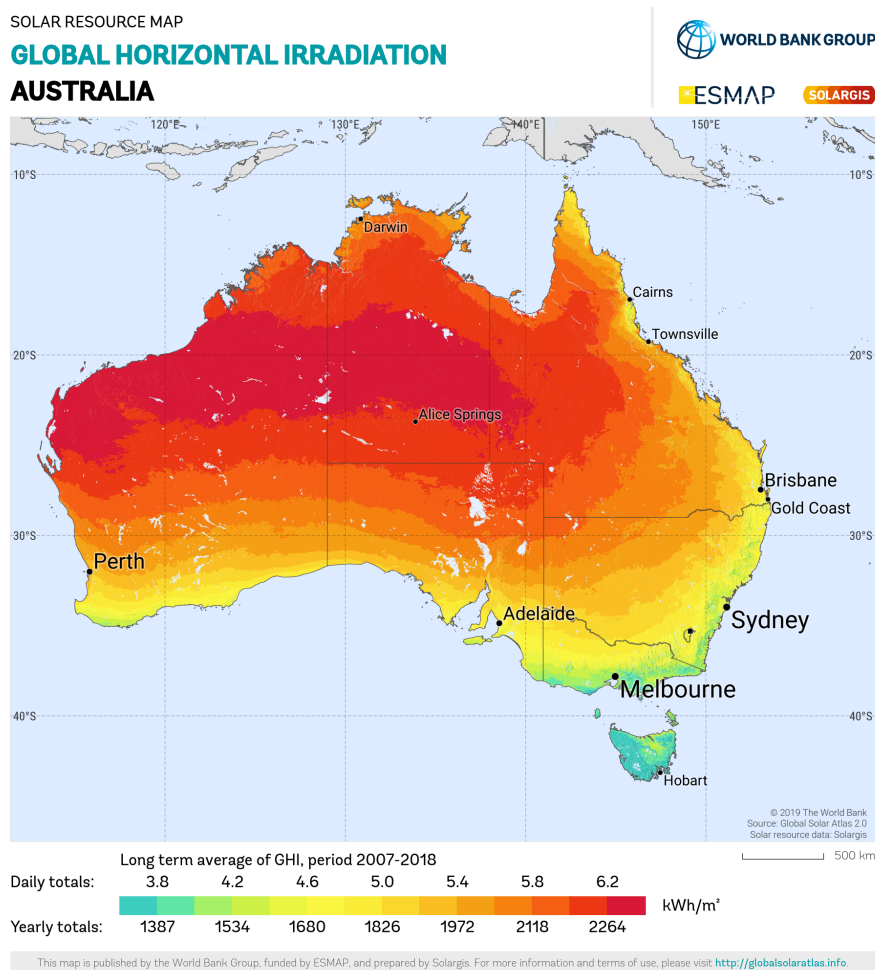


FIGURE 3.2 – Carte de l’irradiation globale horizontale en Australie. Source : Solargis. Lecture : Répartition de la moyenne d’irradiation globale horizontale dans l’Australie (moyenne calculée sur la période 2007-2018)

photon pénètre dans une cellule, son énergie est transférée aux électrons du silicium. Il y a une tension électrique créée par la libération de ces électrons, ce qui se traduit par une différence de potentiel entre les deux couches P et N. Ce concept est connu sous le nom d’effet PV. Le physicien français Antoine Becquerel l’a découvert en 1839. Un panneau solaire capable de générer de l’électricité à partir de l’énergie solaire tout en étant non polluant et dépourvu de composants mécaniques en assemblant plusieurs cellules [98].

De nombreux facteurs internes et externes affectent Le rendement d’un parc PV. Cette performance est en partie due à la composition et à la nature des panneaux PV. En fait, il existe une différence de

3.3. PRODUCTION PHOTOVOLTAÏQUES

performances selon les marques et les types de cellules solaires. De plus, l'orientation et l'inclinaison des panneaux affectent leurs performances. La majorité de la production en Australie est réalisée avec une direction sud et une inclinaison de 30 degrés pour tous les types de producteurs PV. Une production inclinée vers le nord ne fonctionne qu'à 30 % de sa capacité de production maximale. La figure.3.3 montre l'inclinaison et l'orientation des panneaux d'un producteur PV.

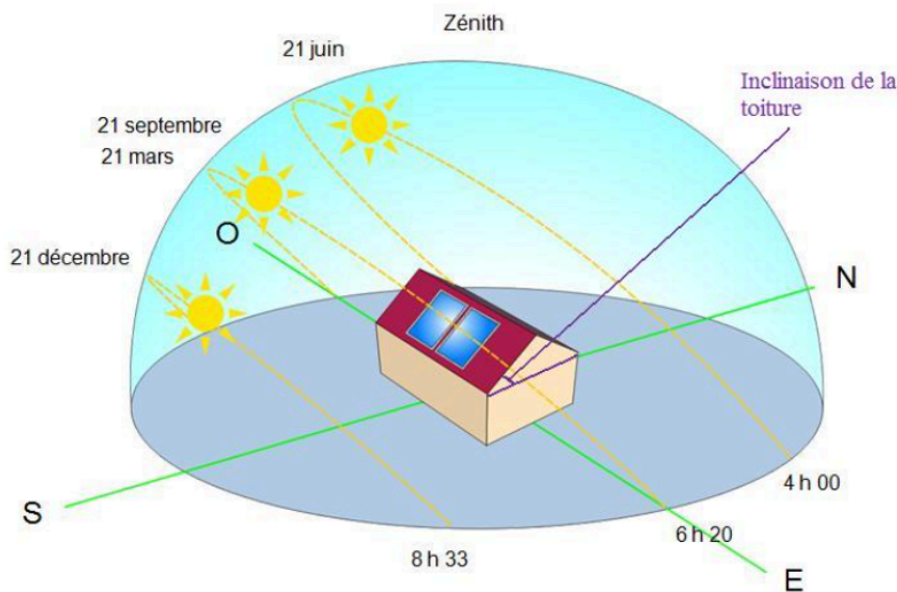


FIGURE 3.3 – L'inclinaison et l'orientation des panneaux d'un producteur PV (adapté de [4])

D'autre part, un certain nombre de facteurs externes résultant de phénomènes liés aux conditions météorologiques, tels que le mouvement des nuages et des aérosols, affecte la quantité d'énergie qu'un parc PV peut produire instantanément. De plus, la présence d'obstacles comme des bâtiments ou des arbres à côté d'un parc PV peut empêcher le rayonnement solaire direct ou diffus d'atteindre le panneau solaire. Le terme pour ce phénomène est l'ombrage. il est soit complètement, comme dans le cas d'un ensemble de panneaux solaires recouvert par une branche ou une particule tombée, soit partiellement, comme dans le cas d'un arbre [99]. Une Par conséquent, tous ces facteurs internes et externes se combinent pour former la courbe de production d'un producteur PV.

La puissance d'un panneau solaire est donnée par :

$$P_R = \eta SI[1 - 0.05(t_0 - 25)] \quad (3.1)$$

Où S est la surface du panneau en m^2 , t_0 représente la température extérieure en Celsius, I est l'irradiation solaire en kW/m^2 , η est le rendement de la cellule PV en pourcentage [100], La puissance de sortie maximale d'un panneau solaire est appelée sa puissance crête. Elle est calculée dans des conditions idéales de production lorsque la température t_0 est de 25 degrés Celsius et le rayonnement solaire I est égal à $1000 W/m^2$.

3.4 Caractérisation du site d'étude

En activité depuis 2008, le Desert Knowledge Australia Solar Centre (DKASC) est une démonstration concrète de technologies solaires regroupant différents types, marques, modèles et configurations. Le DKASC est situé dans le quartier Desert Knowledge à Alice Springs, une ville du Territoire du Nord bénéficiant de l'une des ressources solaires les plus élevées du pays, dans un environnement de désert aride. Cette initiative exceptionnelle a évolué depuis ses débuts en tant que projet financé par le Commonwealth pour devenir une entité commerciale autonome, accessible à un large éventail d'utilisateurs à travers le monde. En constante expansion, elle est considérée comme la plus grande installation de démonstration solaire multi-technologies de l'hémisphère sud, bénéficiant d'un emplacement unique dans un environnement naturel. Le DKASC offre une combinaison de :

- **Recherche** : Les données recueillies par les différents réseaux du DKASC sont exploitées par des chercheurs en Australie et dans le monde entier. Le téléchargement de ces données gratuites en libre accès, ils peuvent étudier diverses technologies solaires performantes dans un environnement réel. De plus, certains laboratoires de recherche privés disposent de leurs propres installations sur place.
- **Investissement** : Les ensembles de données fiables offrent une perspective à long terme et de qualité supérieure sur les performances des installations PV commerciales. Cela favorise le développement tant commercial que technique de l'industrie en Australie et à l'échelle internationale.
- **Développement** : Les fabricants de modules solaires manifestent un vif intérêt pour les données du DKASC, car celles-ci leur offrent la possibilité de tester, de comparer et d'améliorer leurs produits en vue de leur développement futur.
- **Éducation** : Le DKASC attire des visiteurs locaux, nationaux et internationaux qui apprécient de se promener dans ses installations, car il met en valeur le monde captivant de l'énergie

solaire.

- **Approvisionnement en énergie** : Le DKASC fournit de l'électricité aux utilisateurs du Desert Knowledge Precinct, leur offrant ainsi quotidiennement une source d'énergie propre. Cela permet aux organisations sur place de réduire leur dépendance au réseau électrique et de diminuer leur consommation d'électricité provenant de celui-ci.

Le DKASC utilise une matrice de silicium monocristallin Trina ainsi que des trackers à double axe à grande échelle. Ces trackers à deux axes ajustent la position des panneaux solaires tout au long de la journée afin de maximiser leur exposition au soleil. Pour optimiser l'utilisation de l'espace disponible sur les trackers, des panneaux monocristallins performants ont été utilisés. Les trackers assurent leur fonctionnement en surveillant en continu l'intensité lumineuse et en traitant ces données pour contrôler les supports motorisés des panneaux solaires. Au lever du soleil, la lumière détectée par les capteurs à l'arrière des unités de contrôle déclenche un mouvement de retour vers l'est. En cas de ciel totalement couvert, la surface des panneaux s'ajuste à l'horizontale. Les systèmes de contrôle et de mouvement nécessitent moins d'un watt de puissance continue en courant continu provenant de la sortie de la matrice. Le suivi devrait augmenter le rendement de cette installation de 35% [101].

3.5 Acquisition de données de surveillance

Les données de surveillance sont généralement collectées à partir de deux sources. D'un côté, on retrouve les paramètres électriques de la centrale PV, tels que la puissance au point de puissance maximale. De l'autre côté, on trouve les données climatiques provenant soit d'une station météo à proximité, soit de données satellites. La première partie de la norme CEI 61724 :2017 [1], intitulée "Performances des systèmes PV - Surveillance", offre des directives détaillées sur la manière adéquate de surveiller les installations PV. En conséquence, les systèmes de surveillance sont classés en trois catégories selon les variables mesurées et les types d'évaluations de performance réalisées. Le choix du système de surveillance dépend généralement de la taille du système PV. Le tableau.3.1 expose le paramètre de mesure requis pour chaque classe de surveillance, incluant les incertitudes déclarées conformément à la première partie de la norme CEI 61724 :2017. La classe A de surveillance correspond au niveau le plus élevé de surveillance des données, tandis que la classe C correspond au niveau le plus bas. L'irradiance dans le plan G_{POA} est mesurée dans le même plan que les modules PV. Pour les applications PV, l'irradiance peut être mesurée à l'aide de pyranomètres à thermopile ou

3.5. ACQUISITION DE DONNÉES DE SURVEILLANCE

de dispositifs de référence PV. Les procédures d'utilisation de ces appareils, les intervalles et les directives de calibration, ainsi que les éventuelles corrections de mesure sont spécifiés dans la norme CEI 60904 :2015 [102] et la norme CEI 61724 :2017 [1]. Si la réalisation de mesures sur site est difficile, il est possible d'acquérir des ensembles de données d'irradiance par le biais d'une modélisation du ciel clair ou de données dérivées de satellites. Généralement, ces mesures présentent des incertitudes plus élevées et peuvent sensiblement différer des mesures au sol. Un autre problème lié à l'utilisation des données satellitaires est la cohérence des données. Bien que la qualité des données satellitaires s'améliore constamment, les données collectées dans le passé peuvent avoir une précision différente de celle des données utilisées actuellement.

TABLE 3.1 – Paramètre de surveillance mesuré divisé par paramètre météorologique et électrique [1].

Paramètre	Classe A	Classe B	Classe C
Paramètre d'irradiance			
Irradiance en plan	*	*	*
Irradiance horizontale globale	*	*	
Irradiance normale directe			
Irradiance diffuse			
Incertitudes maximales	3%	8%	N'importe quel
Paramètre environnemental			
Température du module PV	*	*	
Température de l'air ambian	*	*	*
Wind speed	*	*	
Vitesse du vent	*		
Précipitations	*	*	
Humidité	*	*	
paramètre électrique			
Tension et courant du champ DC	*		
Puissance du champ DC	*		
Tension et courant du champ AC	*	*	
Puissance du champ AC	*	*	*
Énergie de sortie	*	*	*
Facteur de puissance de sortie	*	*	*
Incertitudes maximales	2%	3%	N'importe quel

Les données de température sont mesurées à l'aide de capteurs de température, généralement des capteurs à thermocouple ou à résistance tels qu'un Pt100 (Pt pour platine). Les capteurs de température de l'air ambiant doivent être bien ventilés et protégés contre le rayonnement solaire.

3.5. ACQUISITION DE DONNÉES DE SURVEILLANCE

Les capteurs de température des modules PV sont fixés à l'arrière du module. Il est très important de garantir une bonne adhérence entre le capteur et le module afin d'obtenir des mesures précises. Un adhésif thermiquement conducteur doit être utilisé de manière appropriée pour une utilisation prolongée en extérieur. Les incertitudes de mesure sont censées être inférieures à $2^{\circ}C$. Le nombre de capteurs par installation dépend de la taille de l'installation et est défini dans la première partie de la norme CEI 61724 :2017 [1].

Il ya différentes étapes nécessaires à la surveillance et à la préparation des données pour toute évaluation de performance. Cela comprend la collecte des données de surveillance, l'identification des éventuelles données corrompues, ainsi que l'application de paramètres de filtrage couramment utilisés. Par conséquent le filtrage des données est utilisé pour résoudre certains problèmes courants liés aux données.

La collecte de données et la vérification initiale de leur qualité constituent la première étape des études d'évaluation des performances. Des problèmes souvent négligés liés aux données comprennent la synchronisation et l'agrégation des données. Les données proviennent généralement de sources différentes. Une vérification minutieuse de la synchronisation des horodatages est essentielle pour garantir le calcul précis des indicateurs de performance clés (KPI). Il est également important de comprendre le type d'agrégation des données. En général, les données sont collectées à intervalles réguliers (par exemple, toutes les 1 minute, 15 minutes). Pour assurer la cohérence des données, il est crucial de savoir si les valeurs sont enregistrées sous forme de moyennes sur la période de collecte ou en tant que valeurs instantanées. L'analyse de la série chronologique des observations sur une journée permet de vérifier si la synchronisation et l'agrégation des données sont correctement réalisées.

Les procédures standardisées de contrôle de qualité comprennent l'élimination des mesures invalides et le traitement des données manquantes [1]. Par conséquent, il est recommandé d'identifier les données incorrectes et d'appliquer des seuils réalistes ainsi que des tests statistiques pour détecter les valeurs aberrantes. En cas de données manquantes, il est recommandé d'évaluer la possibilité de les compléter (imputation) de manière raisonnable, ainsi que le type d'approche à utiliser. Il existe plusieurs approches différentes pour l'imputation des données, telles que l'utilisation d'interpolations de différents types, de filtres de Kalman, de modèles autorégressifs ou de moyennes mobiles [103, 104]. Pour les données à haute résolution (telles que les mesures à la minute ou à l'heure) et les écarts de courte durée, l'interpolation est une approche raisonnable, cependant, pour les écarts plus importants

3.5. ACQUISITION DE DONNÉES DE SURVEILLANCE

ou les données à résolution plus basse (comme les données quotidiennes ou hebdomadaires), d'autres approches donnent de meilleurs résultats [103]. En fonction de la disponibilité d'autres paramètres mesurés (comme les mesures d'irradiance satellitaire ou les capteurs d'irradiance situés à différents endroits), il est également possible d'appliquer des modèles de régression multivariée ou univariée, ainsi que des techniques d'apprentissage automatique [104, 105]. Au sein de la communauté PV, il n'existe pas de consensus clair sur la manière de procéder au remplissage des données, et cela dépend souvent de la quantité de données à compléter et de l'ampleur des lacunes dans les données. Dans tous les cas où des données ont été imputées, il est recommandé de les étiqueter comme telles afin de les identifier clairement, et de documenter l'approche d'imputation utilisée.

Lors du processus de vérification de la qualité et de correction des données, il est nécessaire de considérer la température séparément des données d'éclairement et de puissance. Bien que l'utilisation de seuils et de tests statistiques pour détecter les valeurs aberrantes, qui seront abordés dans la section suivante, soit importante pour améliorer la qualité d'un ensemble de données, cela ne suffit pas toujours à détecter les mesures erronées. En examinant la série chronologique des données brutes, il est possible d'identifier des erreurs de mesure significatives. Ces erreurs peuvent être causées soit par des lectures incorrectes, soit par des problèmes lors de l'acquisition des mesures. Certains problèmes courants incluent l'occultation temporaire des capteurs d'irradiance par des objets ou le détachement des capteurs de température du module PV. Ces événements peuvent être facilement repérés lors de l'analyse des données disponibles. Si seules de petites portions des ensembles de données sont affectées, l'imputation des données peut être utilisée pour récupérer les valeurs erronées ou manquantes. En revanche, en cas de pannes de données plus longues, d'autres sources doivent être utilisées pour récupérer les données, comme l'utilisation de données satellitaires pour l'irradiance ou de modèles de température pour le module PV. Dans la suite, nous discuterons des données de température d'une part, et des données d'irradiance et de puissance PVs d'autre part, en donnant des exemples d'application de l'imputation de données. Ces exemples proviennent des problèmes rencontrés lors de l'analyse des données en vue de leur traitement ultérieur.

3.5.1 Imputation de l'irradiance en plan

Déterminer l'irradiance est un domaine extrêmement complexe. Idéalement, on mesure l'irradiance dans le plan (G_{POA} à l'aide d'un dispositif installé dans le même plan que le système PV étudié. Si

3.5. ACQUISITION DE DONNÉES DE SURVEILLANCE

aucun capteur d'irradiance n'est présent dans ce plan, il faut effectuer une transposition du G_{POA} à partir de l'irradiance horizontale globale GHI. Ainsi, les données d'irradiance peuvent être catégorisées en différentes classes de précision en fonction de la disponibilité des données [106]. Pour estimer le G_{POA} on utilise des méthodes de décomposition et de transposition à partir du GHI extrait, provenant d'une des sources suivantes : ensembles de données basés sur l'analyse ou ensembles de données modélisés pour les conditions de ciel clair.

1. **Faible précision** : Pour estimer le G_{POA} , on utilise des méthodes de décomposition et de transposition à partir du GHI extrait, provenant d'une des sources suivantes : ensembles de données basés sur l'analyse ou ensembles de données modélisés pour les conditions de ciel clair.
2. **Précision moyenne** : On effectue la mesure de l'éclairement horizontal GHI sur le site, puis on estime le G_{POA} en utilisant des méthodes de décomposition et de transposition.
3. **Haute précision** : le G_{POA} est mesuré directement sur site.

L'ordre de précision est déterminé en fonction de l'augmentation des incertitudes dans les ensembles de données. Bien que les valeurs d'irradiance mesurées puissent présenter des incertitudes aussi faibles que 2% [107], l'utilisation des approches de décomposition et de transposition, ainsi que l'estimation de GHI, introduit des incertitudes supplémentaires, certaines étant considérablement élevées. En règle générale, les mesures effectuées au sol sont toujours privilégiées en raison de leur plus grande précision, tant pour l'irradiance dans le plan que pour l'irradiance horizontale. Lorsqu'il est fait usage de stations météorologiques, la résolution spatiale peut ne pas être suffisamment élevée, ce qui entraîne des incertitudes importantes. Les ensembles de données provenant de satellites peuvent être inexacts en raison de leur faible résolution spatiale (et éventuellement temporelle) ainsi que du traitement des nuages, de la neige ou des aérosols. Il existe de nombreux modèles de ciel clair différents disponibles. Dans le cas le plus simple, ils se basent sur des calculs géométriques. Les modèles de ciel clair plus avancés prennent en compte divers paramètres atmosphériques mesurables tels que l'ozone, les aérosols et l'eau précipitable. Le problème réside dans le fait que ces données doivent être mesurées et fournies en tant qu'entrées du modèle. Afin d'obtenir des résultats pertinents, il est nécessaire de sélectionner la meilleure option parmi celles-ci pour chaque cas, et de l'évaluer attentivement. GHI représente la somme de l'éclairement diffus et direct, et sa définition est la suivante :

$$GHI[W/m^2] = DHI + DNI \cos(\theta_z) \quad (3.2)$$

DHI représente la composante diffuse de la luminosité horizontale, provenant de toutes les directions, tandis que DNI désigne la luminosité directe normale. θ_z est l'angle solaire par rapport au zénith.

G_{POA} est influencé par plusieurs facteurs tels que la position du soleil, l'albédo et l'ombrage, l'orientation du système, les différentes composantes d'irradiance. On peut le décrire comme la somme de la composante d'irradiance directe dans le plan $G_{b_{POA}}$ et des composantes d'irradiance diffuse dans le plan, comprenant une composante $G_{g_{POA}}$ réfléchié par le sol dans le plan et une composante $G_{d_{POA}}$ provenant de la diffusion dans le ciel dans le plan du réseau :

$$G_{POA}[W/m^2] = G_{b_{POA}} + G_{g_{POA}} + G_{d_{POA}} \quad (3.3)$$

En l'absence d'une irradiance mesurée dans le plan, les composantes individuelles sont calculées à partir du GHI, fourni par l'un des scénarios évoqués précédemment. La séparation des différentes composantes d'éclairement est nécessaire en raison de la complexité de la modélisation de la composante d'éclairement diffus. Une analyse approfondie des modèles peut être consultée dans la référence [108], tandis que les références [109, 110] comparent et évaluent différents modèles.

Si des capteurs G_{POA} sont disponibles sur les côtés, il est recommandé de les utiliser afin d'obtenir des données plus précises et exactes. En général, les lectures provenant de ces capteurs sont plus fiables. Pour garantir un fonctionnement optimal et fiable des capteurs, il est essentiel de les nettoyer et de les calibrer régulièrement, conformément à la première partie de la norme IEC 61724 :2017. Les problèmes potentiels tels que les erreurs de mesure, les problèmes d'alignement du capteur ou les dérives du capteur sont examinés plus en détail dans la section suivante.

Comme mentionné précédemment, les mesures incorrectes ou manquantes sur une courte période peuvent être remplacées par des techniques d'imputation de données. Cependant, si une mesure est défectueuse ou manquante pendant une longue période, l'utilisation d'approches simples telles que l'interpolation pour remplir ces valeurs n'est plus valide ou possible.

3.5.2 Analyse des données de température

La température du module T_{mod} dépend de plusieurs paramètres liés à l'irradiance solaire, tels que la température ambiante, la vitesse et la direction du vent, la configuration de montage, le comporte-

ment thermique et l'efficacité du module, ainsi que des paramètres au niveau du système, tels que les conditions d'encrassement ou d'ombrage. Généralement, la représentation graphique de la température ambiante et de la température du module au fil du temps permet d'estimer assez précisément la qualité des mesures. Si plusieurs capteurs de température de module sont disponibles, une comparaison entre eux est recommandée. Avec de tels graphiques, les valeurs aberrantes sont facilement détectables et peuvent être prises en compte. Si les relevés de température du module montrent des tendances inattendues, des modèles de température de module peuvent être appliqués et les valeurs mesurées peuvent être comparées aux valeurs modélisées. Si l'écart entre les deux est trop important, les valeurs mesurées devraient être remplacées par les valeurs modélisées.

Le modèle à choisir dépendra de la disponibilité d'autres données climatiques spécifiques à cet emplacement. Le modèle le plus simple à utiliser est l'équation de température de fonctionnement nominal de la cellule (modèle NOCT) [111].

$$T_{mod}[^{\circ}C] = T_{amb} + \frac{G_{POA}}{800W/m^2}(NOCT - 20^{\circ}C) \quad (3.4)$$

La température normale de fonctionnement de la cellule (NOCT) est déterminée pour un module orienté au sud avec un angle de 45 degrés, une irradiance incidente de $800 W/m^2$, une température ambiante de $20^{\circ}C$ et une vitesse du vent de $1 m/s$. G_{POA} représente l'irradiance mesurée dans le plan, tandis que T_{amb} désigne la température ambiante mesurée. La valeur de NOCT est généralement fournie dans la fiche technique spécifique du modèle. L'intégration de la vitesse du vent, si elle est disponible, améliore généralement la précision de l'estimation de la température du module. Un modèle bien performant est le modèle de température du module de Sandia (SMTM) [112].

La vitesse du vent mesurée est représentée par WS, tandis que les paramètres empiriques a et b dépendent de la configuration de montage, du matériau à l'arrière du module et du matériau de la cellule solaire. Des modèles plus avancés, qui intègrent des coefficients empiriques supplémentaires, la transmittance de la couverture du module et le coefficient d'absorption de la cellule, ont été développés par Skoplaki et al. [113] et Mattei et al. [114]. Le choix du modèle, et donc la précision de la température modélisée, dépendra toujours des paramètres d'entrée disponibles.

3.6 Sélection des données d'entrée

Il est crucial de minimiser autant que possible le nombre de variables d'entrée lors de la conception d'un modèle basé sur des mesures [115]. En effet, le nombre de paramètres dans le modèle est proportionnel au nombre de neurones cachés pour chaque entrée [116]. Par conséquent, il est essentiel que toutes les entrées contiennent les facteurs pertinents liés à la valeur à modéliser, tout en réduisant au minimum l'influence des perturbations ou des biais introduits par des paramètres superflus. Ces paramètres inutiles peuvent causer des variations non significatives dans les variables d'entrée, pouvant ainsi introduire des sources d'erreur dans le modèle. Avant de concevoir le modèle lui-même, il est donc nécessaire d'effectuer une sélection des entrées. Cette sélection comporte deux aspects distincts [117]. Cela implique de retirer les entrées qui ont une influence négligeable sur la sortie, et de diminuer la dimension de l'espace de représentation des variables du modèle. Afin de vérifier notre méthode, nous avons entrepris de prédire la production d'énergie PV. Notre étude s'est concentrée sur la région du Nord de l'Australie, en particulier sur le site d'Alice Springs, en utilisant des données directement ou indirectement liées à la puissance du PV. Notre objectif était de développer un outil de prédiction adaptable à la base de données. Les données que nous allons présenter peuvent être divisées en deux catégories : les variables endogènes et les variables exogènes. Les variables endogènes sont de même nature que la prédiction elle-même (telles que le rayonnement global ou la puissance PV), tandis que les variables exogènes sont d'une nature différente (comme la température, la pression, etc.).

3.7 Présentation des données et analyse préliminaire

3.7.1 Collecte et préparation des données

Cette étude utilise les données PVs de 1B Desert Knowledge Precinct à Central Australia (DKASC), le système PV d'Alice Springs, comme étude de cas. La configuration d'installation du système PV DKASC est présentée dans le tableau 3.2. Selon [101], les échantillons de l'étude de cas peuvent être facilement téléchargés dans l'espace de travail. Une série chronologique d'une résolution de 5 minutes a été capturée sur une période de Mai 2017 à Mai 2021. L'énergie PV est à son maximum pendant le midi, avec une quantité relativement faible au lever du soleil et en milieu d'après-midi, tendant à atteindre zéro au coucher du soleil. L'énergie PV entre des pas de temps consécutifs est influencée par divers facteurs et présente des variations significatives sans présenter une cohérence marquée.

3.7. PRÉSENTATION DES DONNÉES ET ANALYSE PRÉLIMINAIRE

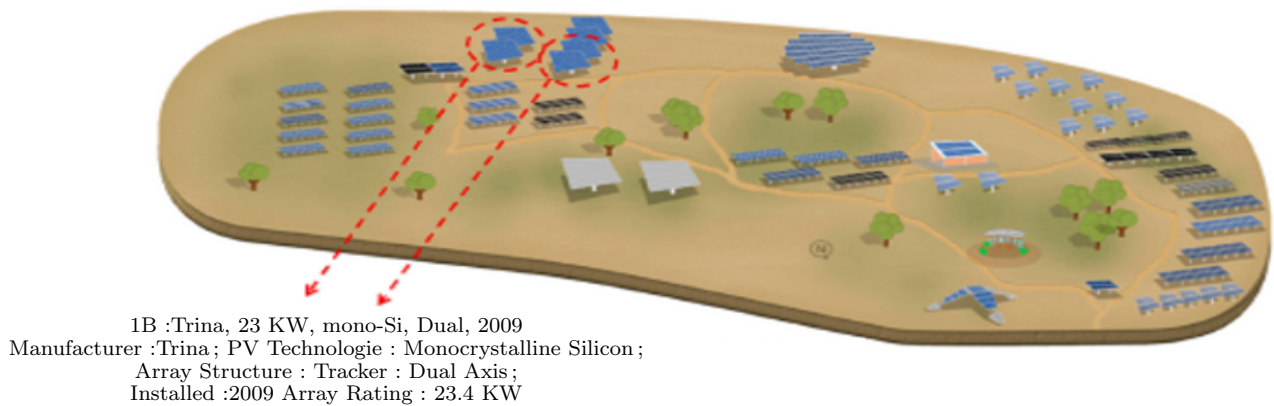


FIGURE 3.4 – Visualisation de la carte du système d’alimentation DKASC.

La figure.3.4 présente la carte des installations d’énergie solaire à Alice Springs, en Australie centrale, qui sont réparties sur 38 sites, dont un site (1) est subdivisé en groupes 1A et 1B. La série chronologique de l’énergie PV utilisée dans cette étude provient du secteur 1B. Des informations détaillées et les spécifications de tous les systèmes d’énergie solaire sont disponibles au public. Les données expérimentales concernent principalement la direction du vent (\hat{A}°); rayonnement horizontal diffus ($W/m^2/ \times sr$), rayonnement horizontal global ($W/m^2/ \times sr$), humidité relative du temps (%), température du temps Celsius ($^\circ C$), puissance active (kW), courant moyenne de phase (A), etc. Les données sont préparées avant d’être introduites dans le modèle en éliminant les échantillons anormaux et en remplaçant éléments manqués, tandis que les échantillons pour chaque attribut sont normalisés être mis à l’échelle dans la dimension égale.

TABLE 3.2 – Les informations spécifiques liées à ce système (1B) [2].

Spécifications du système	
Technologie PV	mono-Si
Zone de l’ensemble	4×38.37m ²
Taille/type de l’onduleur	4×6 kW, SMA SMC 6000A
Puissance nominale de l’ensemble	23.4 kW
Nombre de panneaux	4×30
Puissance nominale du panneau	23.4 kW
Installation terminée	jeu. 8 janv. 2009

3.7.2 Corrélation.

Afin de vérifier la corrélation entre la puissance de sortie PV et les facteurs météorologiques variables, cette étude a sélectionné de manière aléatoire un ensemble de données d'une journée à partir du site web du DKASC. Les facteurs météorologiques entrant dans le modèle de prévision de la puissance PV sont déterminés à l'aide de diagrammes de dispersion et de la valeur du coefficient de corrélation de Pearson (PCC). En statistiques, le PCC est une mesure de la corrélation linéaire entre les variables X et Y. Par conséquent, le PCC peut être défini mathématiquement comme suit :

$$R = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (3.5)$$

où $\bar{x} = \sum_{i=1}^n (x_i/N)$ et de manière analogue pour \bar{y} . Il est important de noter que le coefficient de corrélation de Pearson (PCC) est l'un des critères les plus souvent utilisés en pratique pour décrire la relation entre les variables.

La moyenne de phase de courant :

la figure.3.5 (a) présente un diagramme de dispersion qui présente la corrélation entre la moyenne de phase de courant et la puissance PV. La figure indique que la puissance PV est extrêmement fortement corrélée à la moyenne de phase de courant. En d'autres termes, il existe une relation très étroite entre ces deux variables. Cette forte corrélation est démontrée par le coefficient de corrélation (R) qui est égal à 0,9997. Un coefficient de corrélation proche de 1 indique une corrélation positive très forte. Par conséquent, la moyenne de phase de courant est un facteur d'entrée important dans la construction d'un modèle de prévision de la puissance PV. la figure.3.5 (b) présente la courbe de corrélation entre la moyenne de phase de courant et la puissance PV sur une période de 30 minutes, de 15h00 à 18h30. Elle montre comment la corrélation entre les deux variables évolue au fil du temps.

le rayonnement horizontal global :

la figure.3.6 (a) présente un diagramme de dispersion qui présente la corrélation entre le rayonnement horizontal global et la puissance PV. La figure indique que le rayonnement horizontal global est fortement corrélé à la puissance PV. Cela signifie qu'il existe une relation significative entre ces deux variables. La corrélation entre le rayonnement horizontal global et la puissance PV est démontrée par un coefficient de corrélation (R) élevé, qui est de 0,8396. Un coefficient de corrélation positif proche

3.7. PRÉSENTATION DES DONNÉES ET ANALYSE PRÉLIMINAIRE

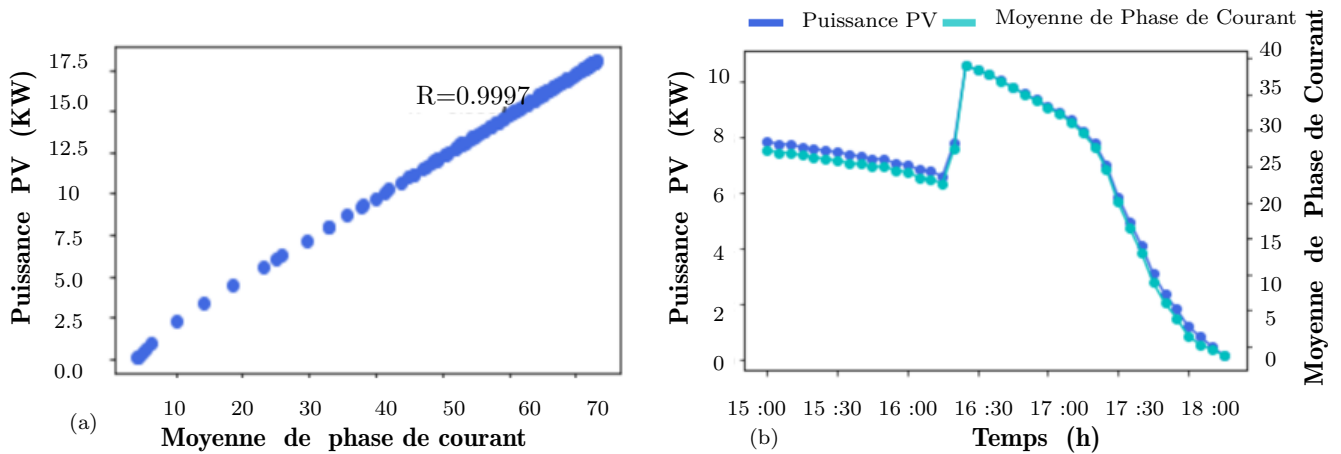


FIGURE 3.5 – (a) Diagramme de dispersion de la moyenne de phase de courant et de la puissance PV”; (b) Courbe de corrélation entre la moyenne de phase de courant et la puissance PV.

de 1 indique une corrélation forte. Par conséquent, le rayonnement horizontal global est un composant d’entrée clé dans la construction d’un modèle de prévision de la puissance PV. la figure.3.6 (b) présente la courbe de corrélation entre le rayonnement horizontal global et la puissance PV sur une période de 30 minutes, de 15h00 à 18h30. Elle montre comment la corrélation entre les deux variables évolue au fil du temps.

le rayonnement horizontal diffus

La figure.3.7 (a) présente un diagramme de dispersion qui présente la corrélation entre le rayonnement horizontal diffus et la puissance PV. La figure indique que la puissance PV est extrêmement fortement corrélée au rayonnement horizontal diffus. Cela signifie qu’il existe une relation très étroite entre ces deux variables. La corrélation entre le rayonnement horizontal diffus et la puissance PV est démontrée par un coefficient de corrélation (R) élevé, qui est de 0,8774. Un coefficient de corrélation positif proche de 1 indique une corrélation forte. Par conséquent, le rayonnement horizontal diffus est un composant d’entrée important dans la construction d’un modèle de prévision de la puissance PV. La figure.3.7 (b) présente la courbe de corrélation entre le rayonnement horizontal diffus et la puissance PV sur une période de 30 minutes, de 15h00 à 18h30.

la température météorologique

La figure.3.8 (a) présente un diagramme de dispersion qui présente la corrélation entre la température météorologique en degrés Celsius et la puissance PV. La figure indique que la puissance PV

3.7. PRÉSENTATION DES DONNÉES ET ANALYSE PRÉLIMINAIRE

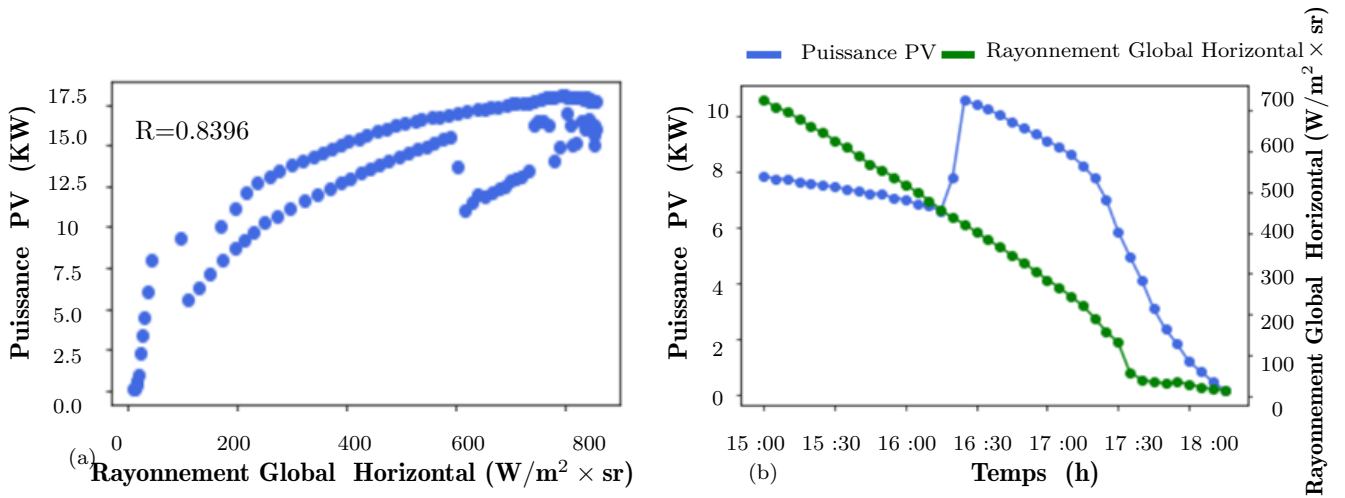


FIGURE 3.6 – (a) Diagramme de dispersion du rayonnement global horizontal et de la puissance PV ; (b) Courbe de corrélation entre le rayonnement global horizontal et la puissance PV.

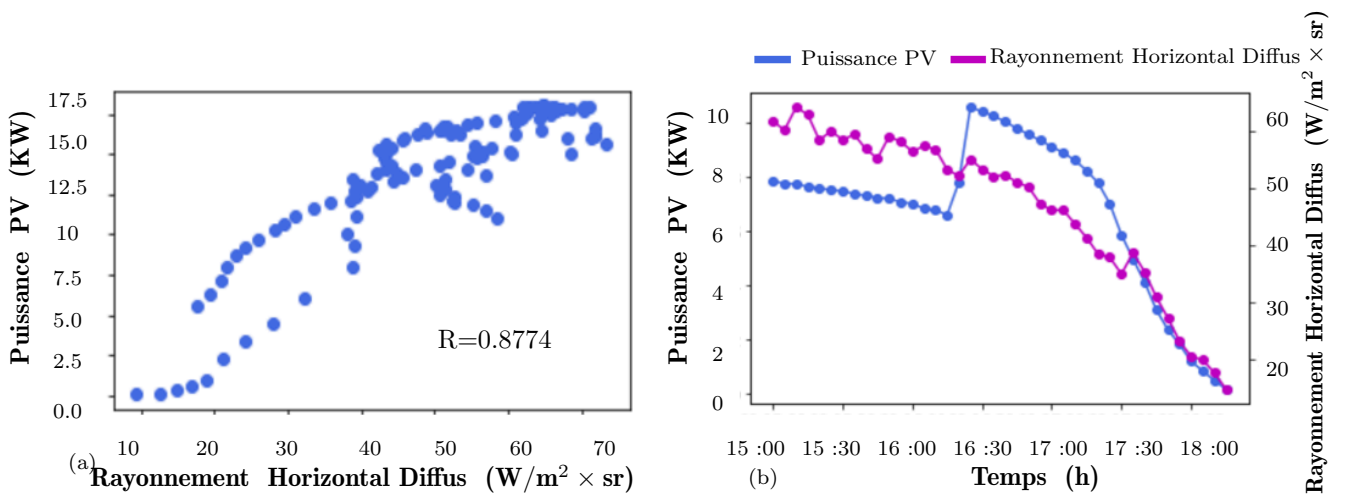


FIGURE 3.7 – (a) Diagramme de dispersion du rayonnement horizontal diffus et de la puissance pv ; (b) Courbe de corrélation entre le rayonnement horizontal diffus et la puissance PV.

est modérément corrélée à la température météorologique en degrés Celsius. Cela signifie qu'il existe une certaine relation entre ces deux variables, mais elle n'est pas aussi forte que dans les exemples précédents. La corrélation entre la température météorologique en degrés Celsius et la puissance PV est démontrée par un coefficient de corrélation (R) positif, qui est de 0,4177. Bien que ce coefficient soit plus faible que ceux mentionnés précédemment, il suggère tout de même une corrélation positive entre les deux variables. Par conséquent, la température météorologique en degrés Celsius est utilisée comme

3.7. PRÉSENTATION DES DONNÉES ET ANALYSE PRÉLIMINAIRE

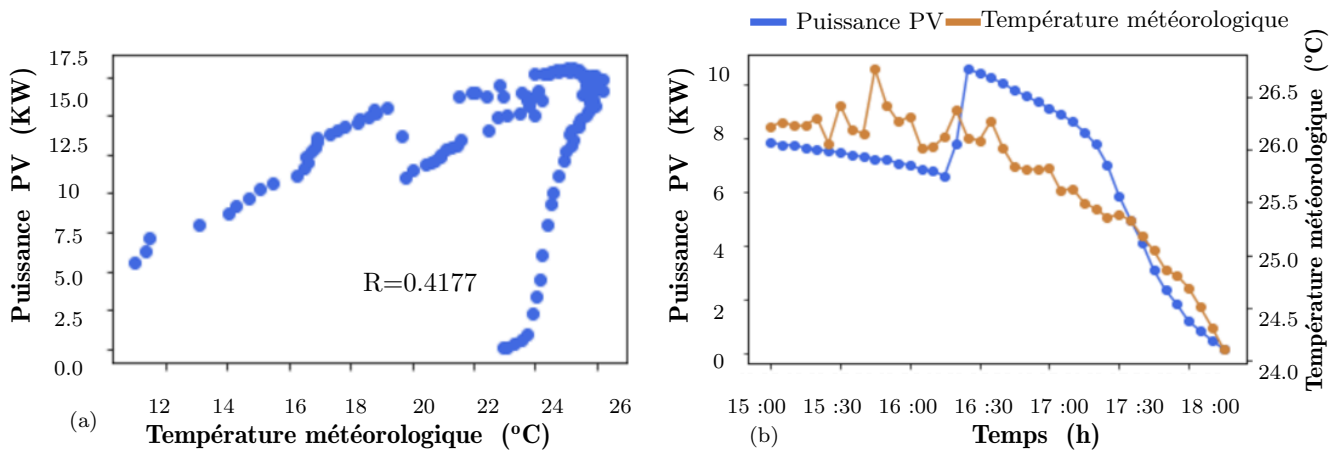


FIGURE 3.8 – (a) Diagramme de dispersion de la température météorologique en degrés Celsius et de la puissance PV ; (b) Courbe de corrélation entre la température météorologique en degrés Celsius et la puissance PV.

facteur d'entrée dans la construction d'un modèle de prévision de la puissance PV. La figure.3.8 (b) présente la courbe de corrélation entre la température météorologique en degrés Celsius et la puissance PV sur une période de 30 minutes, de 15h00 à 18h30.

l'humidité relative

La figure.3.9 (a) présente un diagramme de dispersion qui présente la corrélation entre l'humidité relative et la puissance PV. La figure indique que la puissance PV est faiblement corrélée à l'humidité relative. Cela signifie qu'il existe une certaine relation entre ces deux variables, mais elle est plutôt faible. La corrélation entre l'humidité relative et la puissance PV est démontrée par un coefficient de corrélation (R) négatif, qui est de -0,3735. Un coefficient de corrélation négatif indique une corrélation inverse entre les deux variables. Par conséquent, malgré une faible corrélation, l'humidité relative est utilisée comme facteur d'entrée dans la construction du modèle de prévision de la puissance PV. La figure.3.9 (b) présente la courbe de corrélation entre l'humidité relative et la puissance PV sur une période de 30 minutes, de 15h00 à 18h30 Le coefficient de corrélation de Pearson est une mesure statistique qui permet de quantifier la relation linéaire entre deux variables. Dans ce cas, les chercheurs ont calculé les coefficients de corrélation entre les concentrations (dont la nature n'est pas précisée dans le texte) et plusieurs variables mesurées.

la direction du vent

la figure.3.10 (a) montre un diagramme de dispersion de la direction du vent pour une journée

3.7. PRÉSENTATION DES DONNÉES ET ANALYSE PRÉLIMINAIRE

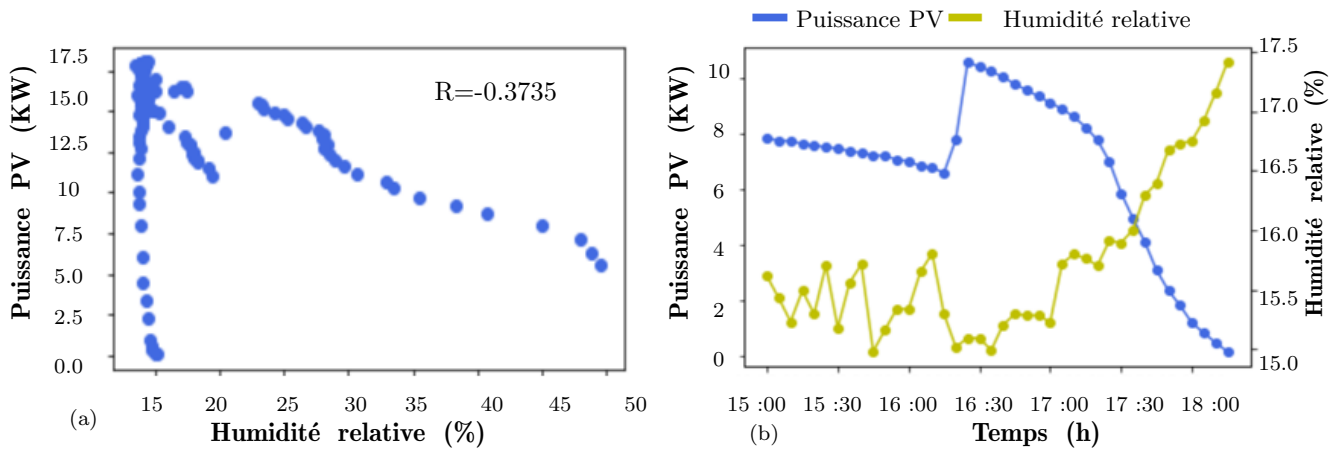


FIGURE 3.9 – (a) Diagramme de dispersion de l’humidité relative météorologique et de la puissance PV ; (b) Courbe de corrélation entre l’humidité relative météorologique et la puissance PV”.

particulière. La relation entre la courbe de direction du vent et la courbe de puissance PV est également non linéaire, ce qui montre que la puissance PV est faiblement corrélée avec la direction du vent. Par conséquent, elle est utilisée comme facteur d’entrée en raison de son coefficient de corrélation négatif $R = -0.0954$. La figure.3.10 (b) présente la courbe de corrélation entre la direction du vent et la puissance PV avec un pas de temps de 30 minutes de 15h00 à 18h30.

Le coefficient de corrélation de Pearson est une mesure statistique qui permet de quantifier la relation linéaire entre deux variables. Dans ce cas, les chercheurs ont calculé les coefficients de corrélation entre les concentrations (dont la nature n’est pas précisée dans le texte) et plusieurs variables mesurées. Le tableau 3.3 a révélé que les coefficients de corrélation les plus élevés étaient associés à trois variables spécifiques : la moyenne de phase de courant, le rayonnement horizontal global et le rayonnement horizontal diffus. Cela signifie que ces trois variables ont une influence significative sur les concentrations étudiées.

TABLE 3.3 – Coefficient de corrélation de Pearson entre la puissance PV et le prédicteur.

Prédicteur	Coefficient de corrélation
Moyenne de la phase de courant (A)	0.9997
Radiation horizontale globale ($W/m^2 \times sr$)	0.8396
Radiation horizontale diffuse ($W/m^2 \times sr$)	0.8774
Température météorologique en degrés Celsius ($^{\circ}C$)	0.4177
Humidité relative météorologique (%)	-0.3735
Direction du vent (\hat{A}°)	-0.0954

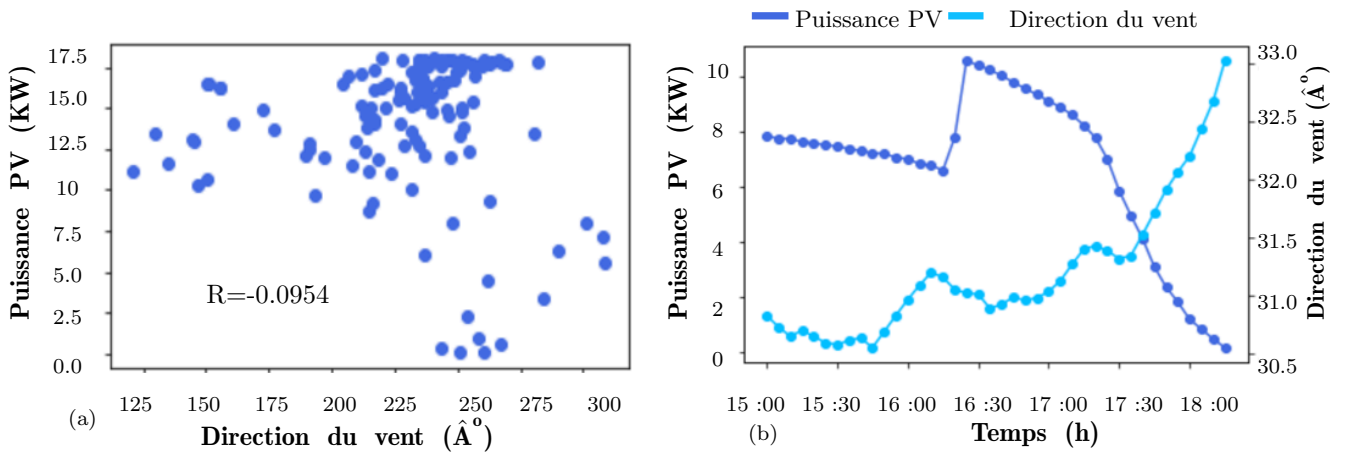


FIGURE 3.10 – (a) Diagramme de dispersion de la direction du vent et de la puissance PV ; (b) Courbe de corrélation entre la direction du vent et la puissance PV.

3.8 Prétraitement des données

Afin de pouvoir utiliser les données pour former des modèles basés sur l'apprentissage automatique, il est nécessaire de les soumettre à un processus de prétraitement. Ce prétraitement se compose de plusieurs étapes qui visent à préparer les données en vue de leur utilisation dans l'apprentissage des modèles prédictifs.

3.8.1 Nettoyage de données.

Le processus de nettoyage des données revêt une importance capitale lors de la création d'un modèle d'apprentissage, car les performances de ce dernier dépendent largement de la qualité des données en entrée. Il est crucial que les données d'entrée couvrent de manière adéquate et complète le phénomène étudié. Si les cas contenus dans les données d'apprentissage ne sont pas suffisamment représentatifs, le modèle sera tout simplement incapable de déterminer ou de prédire de manière fiable les variables de sortie.

3.8.2 Gestion des données manquantes.

Lors de cette étape de préparation des données, il est essentiel de veiller à ce qu'aucune donnée ne soit manquante dans l'ensemble d'apprentissage, car les données vides pourraient être interprétées comme une valeur distincte (0 ou null). Il convient également de noter que de nombreuses biblio-

3.9. CONCLUSION

thèques informatiques utilisées pour implémenter des modèles ne prennent pas en charge les valeurs manquantes, ce qui entraînerait une erreur si de telles valeurs étaient présentes.

Il existe deux approches pour gérer les valeurs manquantes :

- L'élimination : Inclut la suppression des observations avec un élément manquant.
- La complétion : Remplissez la variable manquante avec des données par copie directe ou par calcul. Dans notre cas, nous limitons dans la mesure du possible la copie de données provenant d'autres sources.

3.8.3 Normalisation.

La performance du modèle d'apprentissage profond est influencée par la qualité des données. Afin d'obtenir des prédictions précises, il est essentiel de procéder au nettoyage et à la préparation des données en supprimant les valeurs aberrantes et en comblant les données manquantes. Une fois les valeurs manquantes remplacées, il est nécessaire de les transformer pour les rendre utilisables. Ensuite, nous normalisons la puissance PV d'origine en utilisant la formule suivante :

$$x^{norm} = \frac{x - x^{min}}{x^{max} - x^{min}} \quad (3.6)$$

Lorsque x représente les données d'entrée d'origine x^{norm} représente les données normalisées, x^{max} et x^{min} correspondent respectivement aux valeurs maximale et minimale des données d'entrée.

3.9 Conclusion

L'objectif de cet état de l'art est de fournir une vue d'ensemble complète des différentes approches de prédiction de la puissance PV, en mettant en évidence les avancées réalisées, les défis rencontrés et les perspectives pour les travaux futurs. Cela contribuera à l'avancement des connaissances dans le domaine de la prédiction de la puissance PV et à l'identification des meilleures pratiques pour la mise en œuvre de ces techniques dans les systèmes énergétiques basés sur l'énergie solaire PV. Ce chapitre de thèse a été consacré à une exploration approfondie de divers aspects cruciaux dans le domaine des systèmes PV et de la prévision de la puissance solaire. À travers l'analyse des performances des systèmes PV, nous avons pu comprendre les paramètres clés qui influent sur leur efficacité et leur rendement. Cette connaissance approfondie offre des opportunités significatives d'optimisation, essentielles dans

3.9. CONCLUSION

le contexte de la transition vers des sources d'énergie renouvelables. Notre investigation approfondie sur la production PV a permis de mettre en lumière les nombreux facteurs environnementaux et météorologiques qui influent sur la quantité d'énergie générée.

Nous avons exploré plusieurs axes essentiels pour atteindre cet objectif, à savoir la caractérisation du site d'étude, l'acquisition de données de surveillance, la sélection des données d'entrée, la présentation des données et l'analyse préliminaire, ainsi que le prétraitement des données.

Tout au long de ce chapitre, nous avons mis en évidence l'importance primordiale de la caractérisation du site d'étude. En évaluant les caractéristiques géographiques, météorologiques et environnementales spécifiques à l'installation PV. Par ailleurs, la fiabilité des prévisions dépend fortement de la qualité des données de surveillance. Nous avons examiné différentes méthodes et technologies d'acquisition de données en temps réel pour garantir la collecte de données précises et fiables. La sélection des données d'entrée a été identifiée comme une étape critique pour la précision des modèles de prévision. En identifiant les variables pertinentes qui ont un impact significatif sur la production d'énergie PV. Dans la phase de présentation des données et d'analyse préliminaire, nous avons utilisé des outils statistiques et des méthodes d'analyse de données pour explorer les tendances et les schémas de production. Enfin, le prétraitement des données s'est avéré être une étape indispensable pour garantir la qualité et la fiabilité des données utilisées dans nos modèles de prévision. En éliminant les valeurs aberrantes, en comblant les lacunes et en corrigeant les erreurs potentielles, nous avons pu améliorer la robustesse et la précision de nos prédictions.

Dans la suite de cette thèse, nous approfondirons notre recherche en mettant en œuvre des modèles de réseaux de neurones avancés pour améliorer davantage les prévisions de l'énergie PV. En définitive, cette thèse contribue à l'avancement des connaissances dans le domaine de l'énergie solaire PV et de la prévision de la production d'énergie PV.

Chapitre 4

Modèles d'apprentissage profond pour une prédiction de la puissance Photovoltaïques

4.1 Introduction

L'intégration efficace des énergies renouvelables dans les réseaux électriques nécessite une prévision précise de la puissance PV. Cependant, les modèles de prévision traditionnels basés sur des méthodes statistiques et mathématiques présentent des limites face à la complexité et à la non-linéarité des données PVs. Ces dernières années, les avancées réalisées dans le domaine de l'apprentissage profond ont ouvert de nouvelles perspectives pour améliorer la précision des prévisions de la puissance PV. Dans ce chapitre, nous nous concentrons sur trois recherches qui proposent des approches novatrices basées sur l'apprentissage profond pour la prédiction de la puissance PV. La première recherche, intitulée "Étude comparative des modèles LSTM et RNN pour la prévision de la puissance PV", se concentre sur une étude comparative entre les modèles LSTM et RNN pour la prévision de la puissance PV. L'objectif est de déterminer le modèle le plus performant en termes de précision et d'efficacité de la prédiction. La deuxième recherche, intitulée "Une nouvelle approche d'apprentissage profond pour la prévision à court terme de la puissance PV basée sur le modèle GRU-CNN", propose une approche novatrice basée sur le modèle GRU-CNN pour la prévision à court terme de la puissance PV. Ce modèle combine les capacités de traitement séquentiel des GRU avec les capacités d'extraction de caractéristiques des CNN afin d'améliorer la précision des prévisions. La troisième recherche, intitulée "Modèles précis de prédiction de la puissance PV basés sur des réseaux neuronaux convolutifs profonds et des unités récurrentes avec portes" présente un modèle hybride combinant des réseaux de neurones convolutionnels profonds et des unités récurrentes à portes pour améliorer la précision des prévisions à court terme de la puissance PV. L'objectif de ce chapitre est de présenter en détail ces trois approches d'apprentissage profond pour la prédiction de la puissance PV, d'analyser leurs performances respectives et de mettre en évidence leurs avantages et limitations. En examinant ces travaux de recherche, nous visons à fournir un aperçu approfondi des progrès récents dans le domaine de la prédiction de la puissance PV grâce à l'utilisation de modèles d'apprentissage profond. Cette analyse comparative nous permettra d'identifier les modèles les plus performants et les meilleures pratiques pour améliorer la précision des prévisions, ouvrant ainsi la voie à des solutions plus fiables et efficaces pour l'intégration des énergies solaires dans les réseaux électriques.

4.2 Étude comparative des modèles LSTM et RNN pour la prévision de la puissance PV

La prévision précise de la puissance PV peut aider à atténuer les inconvénients de la génération d'énergie PV, ce qui est crucial pour la maintenance et la réparation des centrales électriques. Dans cette recherche, nous avons donc comparé les performances de deux modèles d'apprentissage profond, LSTM et RNN, pour déterminer lequel est le plus efficace dans la prévision de la puissance PV. Le choix du RNN s'explique par sa capacité à capturer la dynamique des données de séries temporelles en stockant les informations des calculs précédents dans sa mémoire interne. Quant à l'approche LSTM, elle est utilisée en raison de sa capacité à préserver et à entraîner les caractéristiques des données sur une période plus longue. LSTM et RNN sont tous deux des modèles d'apprentissage profond avantageux, chacun ayant ses propres caractéristiques et avantages. En comparant ces deux modèles, nous avons cherché à identifier lequel offre les meilleurs résultats en termes de précision et de fiabilité de la prévision de la puissance PV.

4.2.1 Collecte des données de la puissance PV

Dans cette étude nous avons cherché à comparer les performances des modèles LSTM et RNN dans la prédiction de données de puissance PV. Nous avons également examiné si les différents types de saisons et de conditions météorologiques avaient un impact sur l'exactitude des modèles entraînés à l'aide de techniques d'apprentissage profond. Ces questions de recherche visent à mieux comprendre quelle méthode de prédiction est la plus efficace et si des facteurs externes tels que les saisons et les conditions météorologiques influencent la précision des modèles. Dans le cadre de notre thèse, nous avons choisi d'étudier le système PV de 1B DKASC à Alice Springs comme cas spécifique. L'ensemble de données couvre une période allant du 1er mars 2020 au 28 février 2021 et a une résolution de 5 minutes, ce qui signifie que les mesures ont été prises toutes les 5 minutes pendant cette période. Les paramètres d'entrée comprennent des mesures telles que le rayonnement horizontal global, la température météorologique, le rayonnement horizontal diffus, la moyenne de la phase du courant, l'humidité relative et la direction du vent. Ces données sont essentielles pour notre recherche, car elles fournissent des informations détaillées sur les conditions météorologiques et les caractéristiques du système PV. Les données sont ensuite divisées en quatre situations en fonction des saisons : l'automne

4.2. ÉTUDE COMPARATIVE DES MODÈLES LSTM ET RNN POUR LA PRÉVISION DE LA PUISSANCE PV

(mars-mai), l'hiver (juin-août), le printemps (septembre-novembre) et l'été (décembre-février). Chaque saison contient trois mois de données, ce qui permet de prendre en compte les variations saisonnières. L'ensemble de données d'entraînement est composé des deux premiers mois de données, ce qui permet au modèle d'apprendre à partir de ces données pour faire des prédictions. L'ensemble de données de test est constitué des données restantes, qui servent à évaluer les performances du modèle en le testant sur des données qu'il n'a pas utilisées lors de l'apprentissage. L'objectif de notre étude est de prédire la puissance active du système PV (en kW) en fonction de ces paramètres.

4.2.2 Paramètres et configurations des modèles

Afin de comparer de manière cohérente les performances des deux modèles, LSTM et RNN, ils ont été configurés avec le même nombre de couches et d'entrées. Le premier modèle LSTM/RNN utilisé comportait une couche d'entrée, suivie d'une couche de dropout (pour la régularisation), puis d'une couche de sortie qui fournit une prévision unique. La couche cachée était initialement dotée de 100 unités de mémoire, et le modèle était compilé avec l'optimiseur ADAM. Le tableau 4.1 répertorie tous les paramètres qui ont été utilisés dans cette étude, ce qui permet de connaître les configurations spécifiques du modèle, tels que le nombre de couches, le nombre d'unités de mémoire, l'optimiseur utilisé, etc.

TABLE 4.1 – Spécification des paramètres pour l'entraînement.

Paramètres	Valeurs
Couche cachée	1 LSTM/RNN de 100 unités
Couche de désactivation (Dropout)	1 avec un taux de désactivation de 0.1
Couche de sortie	1
Nombre d'époques	100
Taille du lot (batch size)	80
Fonction de perte	MSE

4.2.3 Comparaison des performances des modèles LSTM et RNN

Dans chaque saison, les valeurs de MAE, MSE, MBE, RMSE et R^2 pour les modèles LSTM et RNN sont présentées dans le tableau ???. Les résultats montrent que les performances des modèles LSTM et RNN sont statistiquement différentes les unes des autres. Par exemple, pour le MAE, les valeurs de LSTM varient de 0.1039 à 0.2531, avec une moyenne de 0.1887, tandis que le MAE moyen

4.2. ÉTUDE COMPARATIVE DES MODÈLES LSTM ET RNN POUR LA PRÉVISION DE LA PUISSANCE PV

TABLE 4.2 – Les résultats du modèle de prévision.

Saison	Erreur	RNN	LSTM
Hiver	MAE	0.0914	0.1039
	MSE	0.0119	0.0266
	R ²	0.9986	0.9968
	RMSE	0.1093	0.1632
	MBE	0.0740	0.0930
Printemps	MAE	0.2410	0.2531
	MSE	0.0665	0.0703
	R ²	0.9946	0.9943
	RMSE	0.2580	0.2651
	MBE	-0.2408	-0.2530
Été	MAE	0.2536	0.1830
	MSE	0.0817	0.0461
	R ²	0.9931	0.9961
	RMSE	0.2860	0.2148
	MBE	-0.2469	-0.1651
Automne	MAE	0.2527	0.2149
	MSE	0.0808	0.0626
	R ²	0.9883	0.9909
	RMSE	0.2843	0.2503
	MBE	-0.2170	-0.1586
Moyenne	MAE	0.2096	0.1887
	MSE	0.0602	0.0514
	R ²	0.9936	0.9945
	RMSE	0.2344	0.2233
	MBE	-0.1576	-0.1207

de RNN est de 0.2096. De même, le MBE moyen du modèle LSTM est de -0.1207, meilleur que celui de RNN (-0.1576). Les indices RMSE obtenus à partir du modèle LSTM varient de 0.1632 à 0.2651, avec une moyenne de 0.2233, tandis que ceux du modèle RNN varient de 0.1093 à 0.2860, avec une moyenne de 0.2344. En ce qui concerne le MSE, les valeurs de LSTM varient de 0.0266 à 0.0703, avec une moyenne de 0.0514, tandis que RNN présente un minimum de 0.0119 et un maximum de 0.0817, avec une moyenne de 0.0602. Les résultats indiquent des différences statistiquement significatives entre les deux modèles, avec des performances légèrement supérieures pour le modèle LSTM dans la plupart des métriques analysées. Pour vérifier l'exactitude et la stabilité des modèles LSTM et RNN,

4.2. ÉTUDE COMPARATIVE DES MODÈLES LSTM ET RNN POUR LA PRÉVISION DE LA PUISSANCE PV

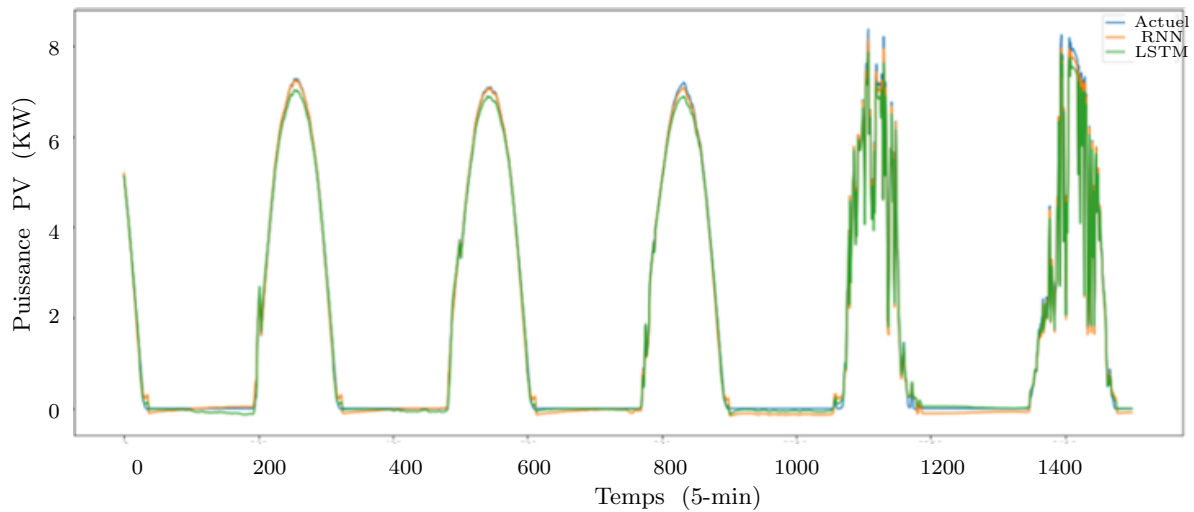


FIGURE 4.1 – Les résultats de prévision de deux modèles et la puissance PV réelle en hiver.

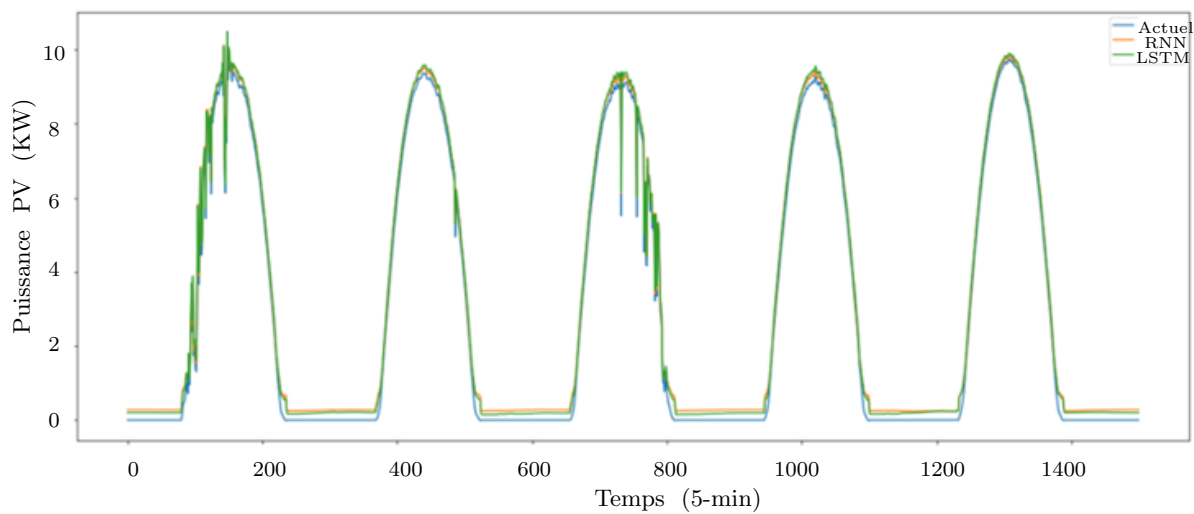


FIGURE 4.2 – Les résultats de prévision de deux modèles et la puissance PV réelle au printemps.

le coefficient de détermination R^2 est aussi utilisé pour mesurer quelle modèle indique de meilleures performances dans les quatre saisons. Les valeurs de R^2 obtenues avec le modèle LSTM sont de 0.9968 en hiver, 0.9943 au printemps, 0.9961 en été et 0.9909 en automne. On observe une amélioration allant de 0.9883 à 0.9986 par rapport aux valeurs typiques de R^2 obtenues avec les algorithmes RNN. En comparaison avec le modèle RNN, les résultats expérimentaux révèlent que l'approche de prédiction LSTM présente la meilleure précision de prédiction en raison de la structure unique de LSTM, qui dispose de fonctionnalités supplémentaires pour mémoriser la séquence des données. Cela indique que

4.2. ÉTUDE COMPARATIVE DES MODÈLES LSTM ET RNN POUR LA PRÉVISION DE LA PUISSANCE PV

le modèle LSTM offre une plus grande précision de prédiction dans différents types de conditions météorologiques et qu'il fonctionne mieux pour la prédiction de la puissance PV.

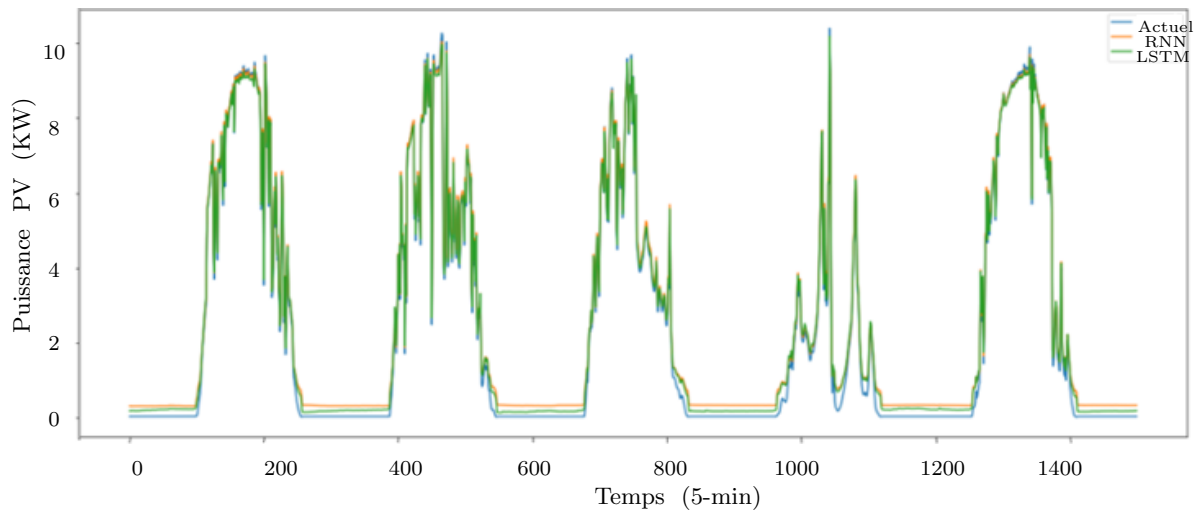


FIGURE 4.3 – Les résultats de prédiction de deux modèles et la puissance PV réelle en été.

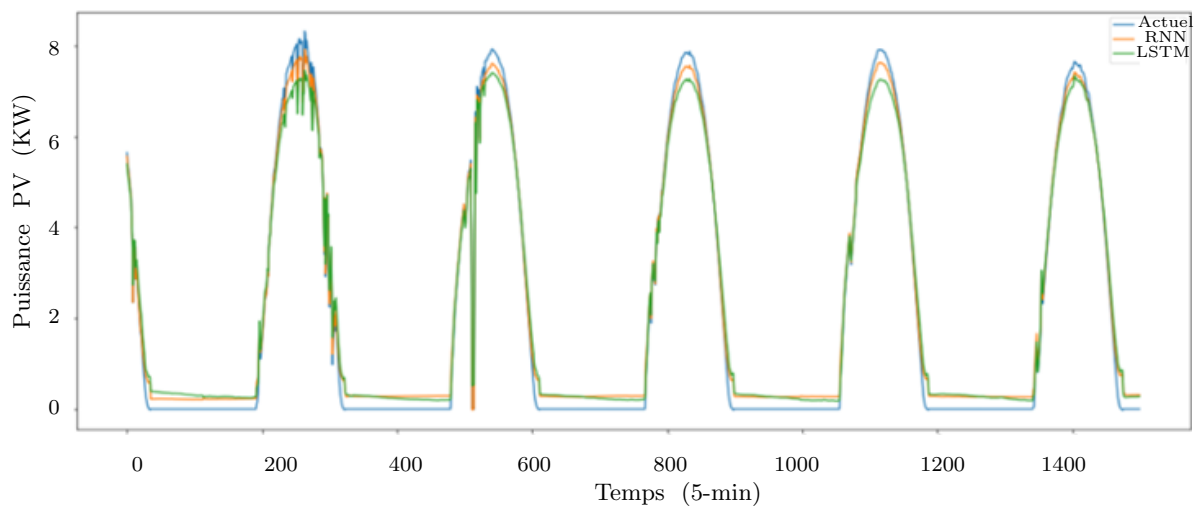


FIGURE 4.4 – Les résultats de prédiction de deux modèles et la puissance PV réelle en automne.

4.2.4 Analyse des résultats et interprétation

Pour une analyse plus détaillée, cinq jours provenant de différentes saisons ont été sélectionnés aléatoirement. Les résultats de prédiction de ces vingt jours, obtenus à l'aide des modèles LSTM et RNN, sont présentés dans les figures.4.1,-4.2, -4.3 et 4.4 . Il est possible de constater que les modèles

4.3. MODÈLE HYBRIDE GRU-CNN POUR LA PRÉDICTION DE LA PUISSANCE PV

ont une performance satisfaisante en termes de prévision. On observe que les résultats des modèles LSTM et RNN suivent la même tendance que les valeurs réelles, mais que le modèle LSTM se rapproche davantage des résultats réels de prévision. Cela suggère que le modèle LSTM offre une meilleure précision de prévision, indépendamment des saisons considérées. Les méthodes d'apprentissage profond, telles que LSTM et RNN, sont largement utilisées dans de nombreux domaines. Dans cet article, le modèle LSTM s'avère être plus précis dans ses prévisions par rapport au modèle RNN. Chaque LSTM est composé de cellules ou de modules système qui sont capables de capturer et de stocker des flux de données. Les cellules ressemblent à une ligne de transport, représentée par la ligne supérieure de chaque cellule, qui connecte un module au suivant et permet le transfert de données du passé vers le présent. Dans cette étude, l'ensemble de données est divisé en quatre cas en fonction des saisons afin de tester la précision et la stabilité des modèles LSTM et RNN. Les résultats démontrent que le modèle LSTM produit des erreurs moyennes absolues, des erreurs quadratiques moyennes, des erreurs absolues moyennes et des racines carrées des erreurs quadratiques moyennes inférieures à celles du modèle RNN. Cependant, les deux modèles semblent avoir produit des résultats solides et fiables.

4.3 Modèle hybride GRU-CNN pour la prédiction de la puissance PV

4.3.1 Description du modèle GRU-CNN proposé

Dans cette étude, un modèle hybride appelé GRU-CNN est présenté pour prédire la génération d'énergie PV. Ce modèle est composé d'une connexion en série de GRU et de CNN. La première moitié du modèle correspond au réseau GRU, qui est responsable de l'extraction des caractéristiques temporelles de la puissance PV. La deuxième moitié correspond au réseau CNN, qui est chargé d'extraire les caractéristiques spatiales. Le modèle proposé est hautement capable d'extraire des caractéristiques et des motifs complexes à partir des variables collectées pour la prédiction de la puissance PV. Le réseau GRU traite les caractéristiques temporelles dans les données historiques de puissance PV. L'idée principale du réseau GRU est d'utiliser des neurones spécifiques avec une structure de base pour stocker et transférer des informations sur une longue période afin de développer des souvenirs durables, réduire le taux de dégradation des informations et détecter les dépendances à long terme. Les informations temporelles résultantes collectées par le GRU à partir des séries temporelles de puissance PV sont ensuite transmises à la couche d'entrée du modèle CNN. La couche de convolution du CNN effectue

4.3. MODÈLE HYBRIDE GRU-CNN POUR LA PRÉDICTION DE LA PUISSANCE PV

une opération de convolution sur les données multivariées de séries temporelles afin d’extraire les caractéristiques spatiales à partir de différentes variables météorologiques. Un CNN contient souvent plusieurs niveaux de couches de convolution et de pooling, avec plusieurs exécutions de convolution à chaque niveau pour capturer des informations pertinentes. Dans cette procédure, le CNN attribue des poids aux paramètres météorologiques en fonction de leur impact sur la puissance PV. Enfin, une couche entièrement connectée (ou dense) est utilisée pour agréger les données et prévoir la puissance PV en utilisant les caractéristiques extraites.

4.3.2 Période des données de puissance PV

L’ensemble de données couvre la période de mai 2017 à mai 2021. Le jeu de données PVs contient un total de 414169 échantillons. Dans cette étude, l’ensemble d’entraînement est composé de 331329 échantillons prélevés dans ces bases de données. Cela signifie que ces échantillons seront utilisés pour entraîner le modèle et lui permettre d’apprendre les relations et les schémas dans les données. L’ensemble de test, quant à lui, comprend 82833 échantillons. Ces échantillons seront utilisés pour évaluer les performances du modèle une fois qu’il aura été entraîné.

TABLE 4.3 – Les paramètres définis dans cette étude.

Test	GRU	CNN
Test-1	unités = 60	filters = 50; taille du noyau = 2; pas (stride) = 2
Test-2	unités = 80	filters = 100; taille du noyau = 2; pas = 2
Test-3	unités = 80; Dropout = 0.2	filters = 100; taille du noyau = 2; pas = 2; Dropout = 0.2
Test-4	unités = 80,100	filters = 200; taille du noyau = 2; pas = 2
Test-5	unités = 100,150; Droupout=0.2	filters = 32,64; taille du noyau = 2; pas = 2; Dropout = 0.2
	LSTM	GRU-CNN
Test-1	unités = 60	units = 60; filters = 50; taille du noyau = 2; pas = 2
Test-2	unités = 80	units = 80; filters = 100; taille du noyau = 2; pas = 2
Test-3	unités = 80; Dropout = 0.2	units = 80; filters = 100; kernel-size = 2; pas = 2; Dropout = 0.2
Test-4	unités = 80,100	units = 80,100; filters = 200; taille du noyau = 2; pas = 2
Test-5	unités = 100,150; Droupout=0.2	units = 100,150; filters = 32,64; taille du noyau = 2; pas = 2; Dropout = 0.2

4.3.3 Configuration du modèle GRU-CNN

Dans cette simulation, les performances du modèle de prévision GRU-CNN sont évaluées en le comparant à trois modèles de prévision standard : le modèle GRU, le modèle CNN et le modèle LSTM. Les paramètres du réseau proposé dans cet article sont indiqués dans le tableau 4.3. Pour cette expérience, la taille du lot (batch size) est fixée à 80 et le nombre d'itérations dans chaque époque est de 100. L'objectif de cette évaluation est de déterminer si le modèle GRU-CNN présente des performances supérieures ou comparables à celles des autres modèles de prévision testés. Ces performances seront évaluées en utilisant des métriques appropriées telles que l'erreur de prévision, ou d'autres mesures de performance spécifiques au domaine.

TABLE 4.4 – Erreur des différents modèles..

Erreur	Test	GRU	LSTM	CNN	GRU-CNN
MAE	Test-1	0.1052	0.1033	0.1047	0.0878
	Test-2	0.0886	0.1196	0.1967	0.0852
	Test-3	0.1635	0.1050	0.1260	0.0709
	Test-4	0.0829	0.0958	0.1193	0.0755
	Test-5	0.1125	0.0959	0.1964	0.0872
	Test-Avg	0.1116	0.1039	0.1486	0.0813
MSE	Test-1	0.0366	0.0382	0.0487	0.0305
	Test-2	0.0226	0.0536	0.0763	0.0236
	Test-3	0.0492	0.0245	0.0459	0.0118
	Test-4	0.0220	0.0446	0.0493	0.0194
	Test-5	0.0291	0.0183	0.2735	0.0174
	Test-Avg	0.0319	0.0358	0.0987	0.0194
RMSE	Test-1	0.1914	0.1954	0.2207	0.1748
	Test-2	0.1505	0.2316	0.2763	0.1544
	Test-3	0.2219	0.1567	0.2143	0.1088
	Test-4	0.1484	0.2113	0.2222	0.1394
	Test-5	0.1708	0.1355	0.5229	0.1322
	Test-Avg	0.1766	0.1861	0.2912	0.1359
R^2	Test-1	0.9981	0.9980	0.9975	0.9984
	Test-2	0.9988	0.9973	0.9961	0.9988
	Test-3	0.9975	0.9987	0.9976	0.9994
	Test-4	0.9988	0.9977	0.9975	0.9990
	Test-5	0.9985	0.9990	0.9862	0.9991
	Test-Avg	0.9983	0.9981	0.9949	0.9989

4.3.4 Présentation des résultats expérimentaux obtenus

Comme le montre le tableau 4.4, le CNN présente le MAE, le MSE et le RMSE moyens les plus élevés parmi les quatre modèles. Les résultats du module GRU sont légèrement meilleurs que ceux du module LSTM, bien qu'ils soient tous les deux supérieurs à ceux du CNN. En détail, le CNN a obtenu des valeurs de 0.1486, 0.0987 et 0.2912 respectivement pour le MAE, le MSE et le RMSE moyens, tandis que le GRU a obtenu des valeurs de 0.1116, 0.0319 et 0.1766, et le LSTM a réduit son taux d'erreur à 0.1039, 0.0358 et 0.1861 pour le MAE, le MSE et le RMSE moyens. Le modèle suggéré GRU-CNN, qui combine les modèles GRU et CNN, produit les prévisions les plus précises. Les indices moyens des modèles GRU-CNN proposés sont les plus bas parmi tous les modèles : 0.0813, 0.0194 et 0.1359. Cependant, les résultats obtenus confirment la faisabilité du modèle proposé dans la prévision de la puissance du système PV. Cela indique que le modèle GRU-CNN présente les performances les plus prometteuses en termes MAE, MSE et RMSE par rapport aux autres modèles évalués. Il est capable de fournir des prévisions plus précises pour la génération d'énergie PV. Pour vérifier davantage la compétitivité du modèle proposé, le coefficient de détermination R^2 est utilisé pour mesurer les performances des quatre modèles. Il convient de noter que le modèle de prévision est plus efficace lorsque R^2 se rapproche de 1. Le tableau 4.4 présente les valeurs de R^2 pour les quatre modèles. La valeur R^2 moyenne du modèle GRU-CNN est de 0.9989, ce qui est plus élevé que le modèle GRU avec un R^2 de 0.9983, et meilleur que le modèle LSTM avec un R^2 de 0.9981. De même, il est meilleur que le modèle CNN avec un R^2 de 0.9983. Selon les indices moyens, comme le montrent les indices moyens, notre modèle suggéré GRU-CNN peut atteindre les meilleures performances en matière de prévision de la puissance PV. Cela indique que le modèle GRU-CNN présente un R^2 plus élevé, ce qui signifie qu'il est capable de capturer une plus grande proportion de la variance des données et de fournir des prévisions plus précises par rapport aux autres modèles évalués. Il démontre ainsi sa supériorité en termes de performance de prévision de la génération d'énergie PV. Comme on peut le constater à partir du tableau 4.4, parmi tous les modèles, la séquence d'entrée du test 3 sur l'ensemble de test montre le meilleur effet. Afin de mieux montrer les résultats de prévision du modèle proposé dans cet article, la figure 4.5 permet d'observer visuellement les performances de chaque modèle en comparant les prévisions générées par les modèles avec les valeurs réelles de la puissance PV. On peut constater que le modèle GRU-CNN se rapproche davantage des valeurs réelles, ce qui confirme son potentiel à fournir des prévisions précises.

4.4. PRÉVISION DE LA PRODUCTION D'ÉNERGIE PV À L'AIDE D'UN MODÈLE HYBRIDE CNN-GRU

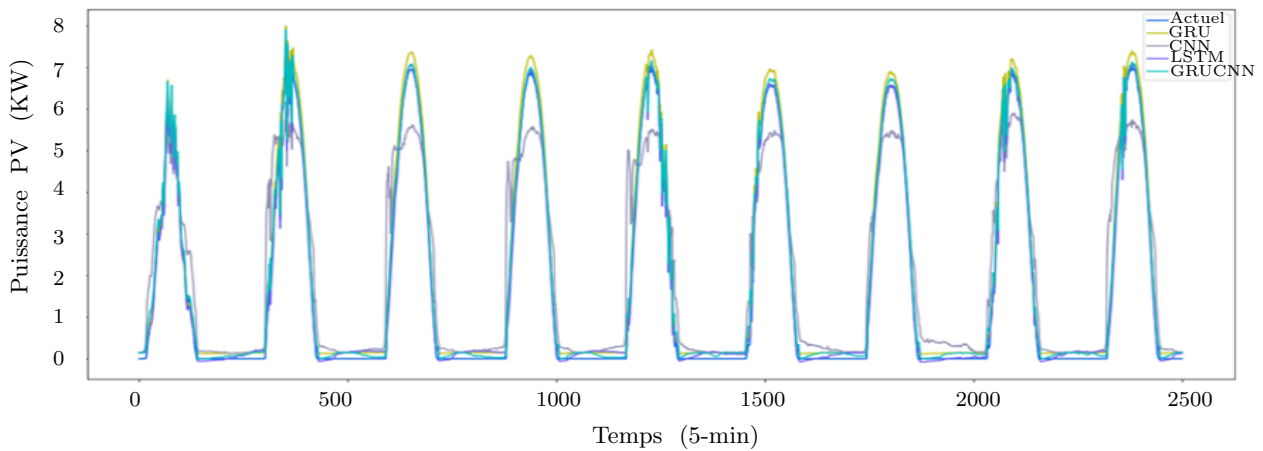


FIGURE 4.5 – Prédiction de la puissance PV en utilisant toutes les méthodes dans la séquence d'entrée du test 3.

4.3.5 Analyse des performances du modèle GRU-CNN

Dans cette étude, nous avons proposé un nouveau cadre de réseau neuronal profond hybride qui combine les modules GRU et CNN. Au cours des expériences, les résultats de GRU se sont avérés plus précis que ceux de LSTM et CNN, mais notre modèle GRU-CNN proposé présente le taux d'erreur le plus faible par rapport aux autres modèles de référence. De plus, l'ensemble de données est séparé en ensembles d'apprentissage et de test avec différentes séquences d'entrée afin de représenter de manière plus intuitive les performances du modèle proposé. Tous les tests montrent que le modèle proposé donne de bons résultats, ce qui démontre que la combinaison de GRU et le CNN peut améliorer la stabilité et l'efficacité des prévisions de puissance PV. Ce travail propose une nouvelle approche de recherche dans la prévision de séquences temporelles. Les recherches futures pourraient viser à améliorer la précision des prévisions à court terme de la puissance PV en intégrant de manière plus efficace les caractéristiques cachées de GRU et CNN.

4.4 Prédiction de la production d'énergie PV à l'aide d'un modèle hybride CNN-GRU

La prévision de la production d'énergie consiste à estimer avec précision la quantité d'énergie électrique qui sera générée par les panneaux solaires à un moment donné. Cette prévision joue un rôle essentiel dans l'optimisation de la fourniture d'énergie au réseau électrique en développant un modèle

4.4. PRÉVISION DE LA PRODUCTION D'ÉNERGIE PV À L'AIDE D'UN MODÈLE HYBRIDE CNN-GRU

hybride d'apprentissage en profondeur. Le modèle proposé, appelé réseau neuronal convolutionnel et unité récurrente à portes, est conçu pour capturer les caractéristiques complexes des données brutes et apprendre de nombreuses séquences parallèles de données d'entrée. Les couches de convolution sont utilisées pour extraire automatiquement les informations pertinentes des données brutes, tandis que les couches GRU permettent d'apprendre les dépendances temporelles des séquences de données.

4.4.1 Présentation du modèle hybride d'apprentissage en profondeur CNN-GRU

La méthode d'apprentissage profond hybride suggérée est une approche basée sur les données qui vise à capturer la forte non-linéarité présente dans les données de séries temporelles de puissance PV. Le module CNN est utilisé car il a la capacité d'apprendre automatiquement des caractéristiques complexes à partir des données brutes. Les couches de convolution du CNN permettent de détecter des motifs et des structures dans les données d'entrée, ce qui est particulièrement utile pour extraire des informations significatives des données PVs. D'autre part, les couches GRU sont capables d'apprendre directement à partir de plusieurs séquences parallèles de données d'entrée. Cela signifie qu'elles peuvent capturer les relations temporelles et les dépendances entre les différentes séquences de données, ce qui est essentiel pour prédire la production d'énergie PV. En combinant ces deux modules dans le réseau CNN-GRU, on bénéficie de l'apprentissage automatique des caractéristiques complexes par le CNN et de la capacité du GRU à apprendre les séquences parallèles de données d'entrée. Cela permet d'exploiter au mieux les informations présentes dans les données PVs et d'améliorer la précision des prédictions de production d'énergie solaire. La structure principale du module CNN contient deux couches de convolution. Les couches de convolution sont des convolutions unidimensionnelles. Chaque couche de convolution se compose d'une opération de convolution et de pooling. La fonction d'activation utilisée est RELU. La caractéristique cachée du module CNN est conçue pour fonctionner avec la caractéristique du module GRU dans le module de fusion des caractéristiques. Le module GRU, quant à lui, vise à capturer la dépendance à long terme, et la cellule de mémoire permet au module GRU de récupérer des informations pertinentes à partir des données passées sur une longue période, tandis que la porte d'oubli permet d'éliminer les informations inutiles. Enfin, les données sont agrégées et la puissance PV est prédite à l'aide d'une couche entièrement connectée.

4.4.2 Période de collecte de données sur la Puissance PV

le DKA Solar Center est un centre de mesure et de collecte de données qui fournit des informations historiques sur la production d'énergie solaire. Il offre un système de mesure standardisé et contrôlé pour évaluer l'efficacité des systèmes solaires existants et recueillir des données pertinentes pour le développement de grandes centrales solaires. Les données ont été collectées pendant une période d'un an, du 1er mars 2020 au 28 février 2021. La résolution des données est de 5 minutes, ce qui signifie que la production d'énergie solaire a été enregistrée toutes les 5 minutes. Cela permet d'obtenir des données détaillées sur la production d'énergie à intervalles de temps courts. Le DKASC utilise la technologie des panneaux solaires PV monocristallins. Les panneaux solaires monocristallins sont réputés pour leur haute efficacité de conversion de l'énergie solaire en électricité. Les panneaux solaires sont équipés d'un système de suivi à double axe. Ce système permet de modifier la direction des panneaux tout au long de la journée afin de maximiser leur exposition au soleil. Cela contribue à améliorer l'efficacité de la production d'énergie solaire. Avant d'être utilisées dans le modèle, les données sont soumises à un processus de prétraitement. Ce processus consiste à supprimer les valeurs aberrantes (valeurs anormales) et à remplacer les éléments manquants par des valeurs appropriées. Cela garantit que les données utilisées pour l'entraînement et l'évaluation du modèle sont de bonne qualité. Les données sont ensuite divisées en deux ensembles distincts : un ensemble d'entraînement et un ensemble de test. Dans ce cas, 80 % des données sont utilisées pour l'entraînement du modèle, tandis que 20 % sont réservées pour l'évaluation des prévisions générées par le modèle, ainsi que l'utilisation d'un modèle hybride CNN-GRU pour effectuer les prévisions.

Pour mener à bien les expériences, le langage de programmation Python version 3.7 a été utilisé. Les calculs ont été effectués sur un ordinateur personnel doté d'un système d'exploitation 64 bits. Les spécifications matérielles comprennent un processeur Intel (R) Core (TM) i7-4600 CPU@2.10GHZ 2.70GHZ et 8,00 Go de mémoire RAM. Ces spécifications matérielles permettent d'exécuter efficacement les algorithmes de prévision de la puissance PV et de traiter les données nécessaires à l'expérimentation.

4.4.3 Entraînement du CNN et du GRU sur les données collectées

L'architecture et les paramètres du modèle proposé, qui est une combinaison de couches de réseaux CNN et de réseaux GRU. Le modèle commence par deux couches CNN. La première couche a 128

4.4. PRÉVISION DE LA PRODUCTION D'ÉNERGIE PV À L'AIDE D'UN MODÈLE HYBRIDE CNN-GRU

filtres, tandis que la deuxième couche a 256 filtres. Chaque filtre a une taille spécifique. Les filtres sont utilisés pour extraire des caractéristiques pertinentes des données d'entrée. La taille du noyau (kernel size) est de 3, ce qui signifie que chaque filtre examine une fenêtre de 3 valeurs successives à la fois. Après l'application des filtres, une opération de pooling est effectuée avec une taille de 2, ce qui réduit la dimension des données. Après les couches CNN, deux couches GRU sont ajoutées et combinées avec la sortie du CNN. Chaque couche GRU contient un nombre spécifique de neurones, avec 32 neurones dans la première couche et 64 neurones dans la deuxième couche. Les couches GRU sont utilisées pour capturer les dépendances séquentielles des données et leur évolution dans le temps. La fonction d'activation utilisée dans ce modèle est la RELU. Cette fonction introduit une non-linéarité dans le modèle et aide à la représentation des caractéristiques complexes des données. Après les couches GRU, deux couches entièrement connectées reçoivent les sorties des couches GRU. Les couches entièrement connectées sont responsables de la prise de décision finale et de la génération des prédictions. Le MSE est utilisée comme fonction de perte dans cette étude. Le MSE mesure la différence entre les prédictions du modèle et les valeurs réelles, et elle est utilisée pour guider l'apprentissage du modèle en minimisant cette différence. Le nombre d'itérations dans chaque époque est fixé à 100, ce qui signifie que le modèle est entraîné sur 100 mini-batches de données à chaque époque. La taille du lot (batch size) est de 100, ce qui signifie que 100 échantillons de données sont utilisés à chaque itération pour mettre à jour les poids du modèle. L'optimisation Adam est choisie comme optimiseur dans cette étude. L'optimisation Adam est une méthode d'optimisation basée sur le gradient qui adapte les taux d'apprentissage pour chaque paramètre du modèle. Le tableau 4.5 présente les paramètres spécifiques du modèle CNN-GRU dans cette étude. La structure et les paramètres des modèles de référence sont répertoriés dans le Tableau 4.6.

4.4.4 Exposition des résultats expérimentaux obtenus

Les résultats des tests avec le CNN-GRU suggéré sur le ensemble de test de notre étude, comme on le voit dans le Tableau 4.7. Le RNN [118] présente la plus grande erreur de prévision (MSE : 0.3859; MAE : 0.4726; MBE : 0.4710; RMSE : 0.6212) et le plus faible R^2 de 0.9675. Cela peut être attribué à la portée limitée du contexte pouvant être récupéré avec les conceptions RNN classiques. La sensibilité du RNN à l'information provenant des phases précédentes dans les pas de temps diminue principalement pendant la phase d'entraînement. Même les informations d'entrée les plus éloignées

4.4. PRÉVISION DE LA PRODUCTION D'ÉNERGIE PV À L'AIDE D'UN MODÈLE HYBRIDE CNN-GRU

TABLE 4.5 – Les réglages des paramètres du modèle CNN-GRU dans cette étude.

Modèle Proposé		
Convolution 1	Filtres	128
	Taille du noyau	3
	Activation	relu
Max-pooling 1	Taille du pool	2
	Pas (stride)	2
	Convolution 2	Filtres
Max-pooling 2	Taille du noyau	3
	Activation	relu
	Taille du pool	2
GRU	Stride	2
	Unités	32
GRU	Unités	64
Entièrement connectés	Neurones	256
Entièrement connectés	Neurones	512

TABLE 4.6 – Réglages des paramètres de la méthode de référence.

Modèles	Optimisateur	Configuration
GRU	Adam	Unités GRU 32,64 ; Unités Dense 1 ;
RNN	Adam	Unités RNN 32,64 ; Unités Dense 1 ;
LSTM	Adam	Unités LSTM 32,64 ; Unités Dense 1 ;
CNN	Adam	Filtres=128,256 ; Taille du noyau=3;Taille du pool=2 ; Pas=2 ; Unités Dense 1 ;

peuvent être oubliées par le RNN. Le GRU affiche la plus grande erreur de prédiction (MAE : 0.2099 ; MBE : 0.2007 ; MSE : 0.1088 ; RMSE : 0.3299). Cela s'explique par la capacité du GRU [83] à capturer des dépendances à long terme dans les données en raison de ses unités de mémoire interne et de sa fonction de porte, malgré les défis liés à la disparition et à l'explosion du gradient lors de l'entraînement des réseaux neuronaux récurrents (RNN) classiques. Le CNN [48] présente également une erreur de prédiction élevée. L'incapacité des couches de convolution à conserver les représentations temporelles intrinsèques de la série temporelle PV historique s'explique par les caractéristiques des couches de convolution. Les couches de convolution sont davantage adaptées à la modélisation de représentations spatiales et peuvent donc avoir du mal à capturer les propriétés temporelles des données. Le LSTM présente une erreur de prévision meilleure que les autres modèles (MAE : 0.0953 ; MBE : 0.0266 ; MSE : 0.0174 ; RMSE : 0.1321 ; R^2 de 0,9985). Cela peut s'expliquer par la capacité intrinsèque du LSTM

4.4. PRÉVISION DE LA PRODUCTION D'ÉNERGIE PV À L'AIDE D'UN MODÈLE HYBRIDE CNN-GRU

TABLE 4.7 – L'exactitude annuelle de prévision des modèles étudiés.

Métriques	CNN	GRU	RNN	LSTM	CNN-GRU
MSE	0.0160	0.1088	0.3859	0.0174	0.0106
MAE	0.0905	0.2099	0.4726	0.0953	0.0686
MBE	0.0799	0.2007	0.4710	0.0266	0.0322
RMSE	0.1266	0.3299	0.6212	0.1321	0.1032
R^2	0.9986	0.9908	0.9675	0.9985	0.9991

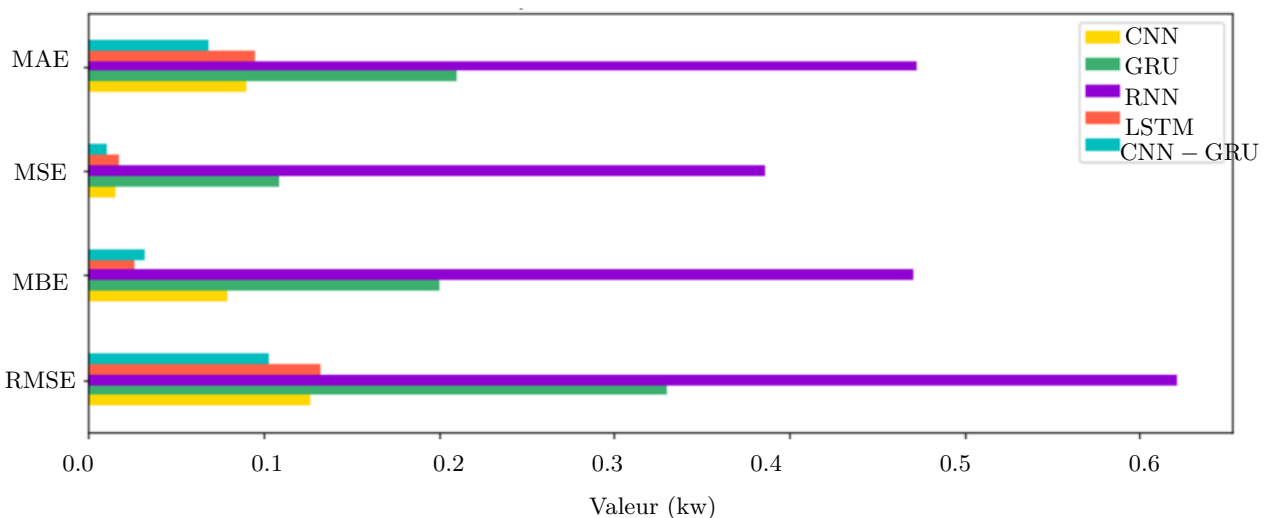


FIGURE 4.6 – Comparaison des valeurs de MAE, MSE, RMSE et MBE sous forme de graphique en barres pour différents modèles.

[119] à apprendre des dépendances à long terme grâce à ses unités de mémoire interne et à sa fonction de porte, ce qui permet de surmonter les problèmes de disparition et d'explosion du gradient dans l'entraînement des RNN classiques. Le modèle proposé CNN-GRU présente la plus faible erreur de prévision (MAE : 0.0686 ; MBE : 0.0322 ; MSE : 0.0106 ; RMSE : 0.1032) et le plus haut R^2 de 0.9991. Cela indique que parmi les cinq modèles, le modèle CNN-GRU développé offre la meilleure précision prédictive. Le modèle CNN-GRU est capable de collecter de manière fiable deux types d'informations, ce qui contribue à ses performances supérieures la figure.4.6 et figure.4.7 sont utilisées pour comparer les erreurs de prédiction du modèle proposé avec celles d'autres méthodes concurrentes. L'objectif de cette comparaison est d'obtenir une compréhension plus approfondie du comportement du modèle proposé.

4.4. PRÉVISION DE LA PRODUCTION D'ÉNERGIE PV À L'AIDE D'UN MODÈLE HYBRIDE CNN-GRU

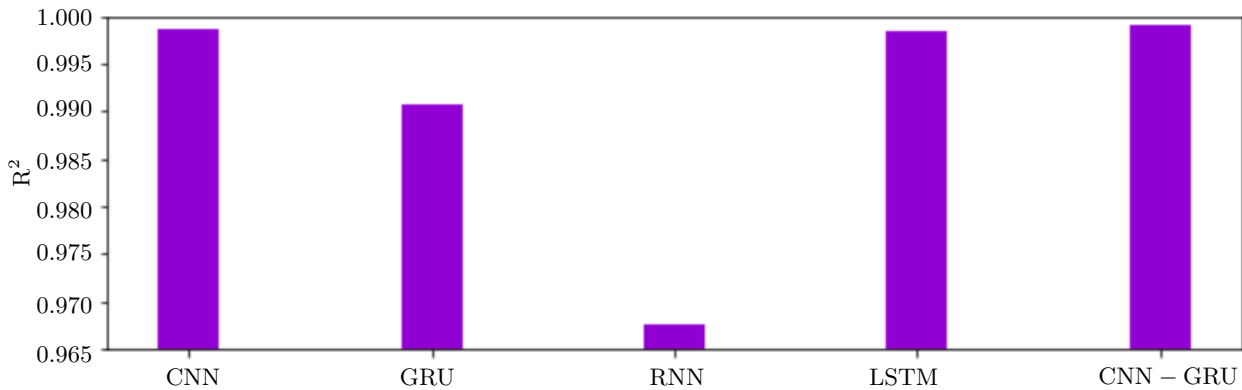


FIGURE 4.7 – Comparaison de la moyenne de R^2 entre la prévision et la mesure de la puissance PV en utilisant divers modèles.

4.4.5 Évaluation des performances du modèle hybride

Dans le cadre de cette recherche, cette étude vise à évaluer l'efficacité du modèle CNN-GRU proposé par rapport à d'autres techniques concurrentes, en tenant compte de l'impact significatif des variations saisonnières sur l'énergie PV. Pour ce faire, un ensemble de données original est utilisé, divisé en quatre sous-ensembles représentant les quatre saisons de l'Australie : l'automne (mars-mai), l'hiver (juin-août), le printemps (septembre-novembre) et l'été (décembre-février). L'idée sous-jacente est que chaque saison présente des caractéristiques distinctes en termes de production d'énergie solaire. Dans cette étude, les ensembles d'entraînement et de test diffèrent en taille pour chaque saison, compte tenu des spécificités saisonnières des données. La répartition des ensembles d'entraînement et de test est présentée dans le tableau 4.8. Pour évaluer les performances des différents modèles de prévision, les mesures d'erreur telles que MSE, MAE, MBE, RMSE et R^2 sont utilisées. Les résultats de chaque modèle sont rapportés dans le tableau 5.12 pour chaque saison. Ces analyses permettent d'évaluer la capacité du modèle CNN-GRU proposé à capturer les variations saisonnières de l'énergie PV par rapport aux autres techniques concurrentes. Les résultats obtenus contribuent à une meilleure compréhension de l'efficacité du modèle proposé et de sa pertinence dans le contexte de l'énergie solaire saisonnière en Australie. Les résultats des différents modèles (CNN, GRU, RNN, LSTM, CNN-GRU) pour les différentes saisons (automne, hiver, printemps, été) sont présentés dans le tableau 5.12. Les valeurs indiquent que la méthode proposée offre de bonnes performances de prévision et une robustesse supérieure aux modèles traditionnels dans les quatre saisons. Dans nos expériences, notre modèle proposé CNN-GRU obtient de meilleures performances de prédiction en termes de prévision à court

4.4. PRÉVISION DE LA PRODUCTION D'ÉNERGIE PV À L'AIDE D'UN MODÈLE HYBRIDE CNN-GRU

terme de l'énergie PV. Cela démontre que le modèle proposé a la capacité de prédire des données saisonnières/périodiques présentant des corrélations non linéaires complexes et des caractéristiques incohérentes au sein des données de puissance PV collectées. En termes de métriques de performance, le modèle proposé présente l'erreur minimale dans toutes les saisons, à l'exception du MSE et du RMSE en été. Le modèle LSTM affiche le deuxième taux d'erreur le plus faible. Il est important de mentionner que chaque indicateur a sa propre méthode unique d'évaluation de l'erreur, ce qui entraîne des différences de performance entre les modèles de prédiction.

TABLE 4.8 – Valeurs statistiques des données expérimentales de puissance PV (KW).

Saison	Échantillons	Nombre	Moyenne	Écart	Max
Automne	Échantillons d'entraînement	21196	2.32	3.08	10.92
	Échantillons de test	5300	1.63	2.44	7.91
	Tous les échantillons	26496	2.19	2.97	10.92
Hiver	Échantillons d'entraînement	21193	1.80	2.55	8.67
	Échantillons de test	5299	2.33	3.06	8.49
	Tous les échantillons	26492	1.91	2.67	8.67
Printemps	Échantillons d'entraînement	20964	2.70	3.44	11.04
	Échantillons de test	5241	2.82	3.50	10.42
	Tous les échantillons	26205	2.732	3.45	11.04
Été	Échantillons d'entraînement	17486	2.62	3.40	11.69
	Échantillons de test	4372	2.92	3.61	10.62
	Tous les échantillons	21858	2.68	3.44	11.69

La nuit, la puissance PV est approximativement nulle, comme le montrent les figures 4.8, 4.9, 4.10 et 4.11. La courbe de puissance prévue par le modèle CNN-GRU est cohérente avec la valeur observée. Le modèle CNN-GRU fournit une prédiction plus précise que les modèles autonomes, notamment la nuit et lorsque la puissance est à son maximum. De plus, le R^2 du modèle hybride proposé est d'environ 0.9975 en automne, 0.9998 en hiver, 0.9988 au printemps et 0.9988 en été, ce qui prouve que la tendance prévue par le modèle hybride proposé est significativement plus précise que les autres approches. Pour les différentes saisons de l'année, le modèle CNN-GRU affiche une meilleure précision prédictive. Ces résultats mettent en évidence l'efficacité du modèle hybride proposé, le CNN-GRU, dans la prédiction de la puissance PV, en particulier pendant la nuit et aux moments de puissance maximale. Les valeurs élevées du R^2 dans chaque saison attestent de l'exactitude de la prédiction du modèle proposé par rapport aux autres approches.

4.4. PRÉVISION DE LA PRODUCTION D'ÉNERGIE PV À L'AIDE D'UN MODÈLE HYBRIDE CNN-GRU

TABLE 4.9 – L'exactitude de prévision saisonnière des modèles étudiés.

Saison	Modèles	MSE	MAE	MBE	RMSE	R ²
Automne	CNN	0.0312	0.1393	0.0644	0.1768	0.9947
	GRU	0.0384	0.1586	0.1585	0.1961	0.9935
	RNN	0.0449	0.1564	0.0922	0.2120	0.9924
	LSTM	0.0185	0.1024	0.0989	0.1362	0.9975
	CNN-GRU	0.0149	0.0800	0.0573	0.1220	0.9975
Hiver	CNN	0.0028	0.0345	-0.0340	0.0532	0.9996
	GRU	0.0016	0.0249	-0.0219	0.0412	0.9998
	RNN	0.0015	0.0259	0.0165	0.0396	0.9998
	LSTM	0.0026	0.0335	-0.0329	0.0510	0.9997
	CNN-GRU	0.0010	0.0169	0.0026	0.0323	0.9998
Printemps	CNN	0.0209	0.0951	0.0940	0.1447	0.9982
	GRU	0.0312	0.1266	0.1250	0.1767	0.9974
	RNN	0.0256	0.1167	0.0938	0.1602	0.9979
	LSTM	0.0348	0.1287	0.1287	0.1867	0.9971
	CNN-GRU	0.0138	0.0692	0.0672	0.1178	0.9988
Été	CNN	0.0215	0.0964	0.0962	0.1467	0.9983
	GRU	0.0145	0.0822	0.0810	0.1206	0.9988
	RNN	0.0459	0.1736	0.0137	0.2142	0.9964
	LSTM	0.0166	0.0918	0.0917	0.1290	0.9987
	CNN-GRU	0.0153	0.0798	0.0798	0.1240	0.9988

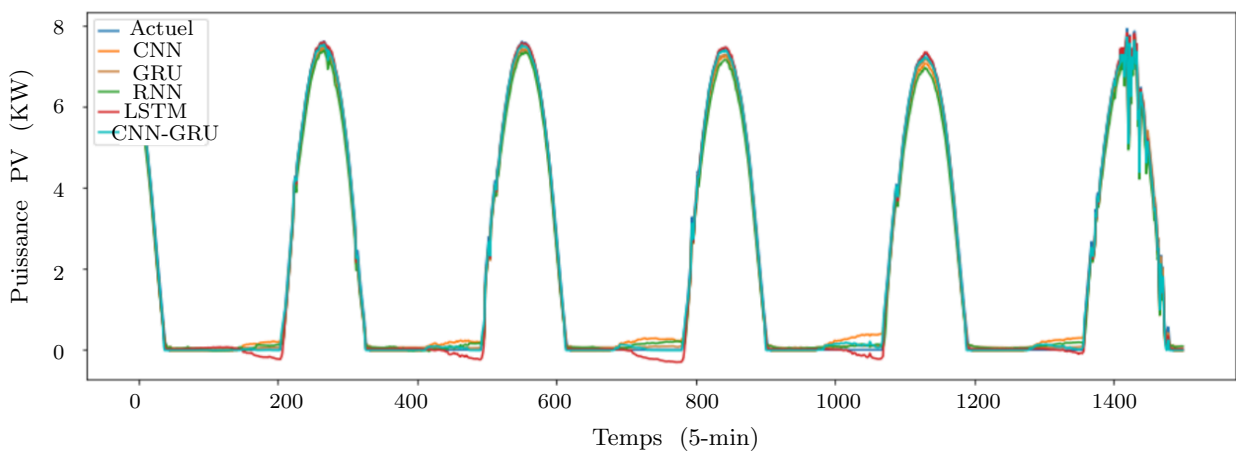


FIGURE 4.8 – Les résultats de prévision de la puissance PV en utilisant toutes les méthodes en automne.

4.4.6 Comparaison avec d'autres approches de prévision de la production d'énergie PV

Pour démontrer l'efficacité de la méthode proposée dans cette étude, des études récentes [71, 69, 120, 78] sur la prédiction de la puissance PV sont présentées afin de vérifier la précision prédictive du

4.4. PRÉVISION DE LA PRODUCTION D'ÉNERGIE PV À L'AIDE D'UN MODÈLE HYBRIDE CNN-GRU

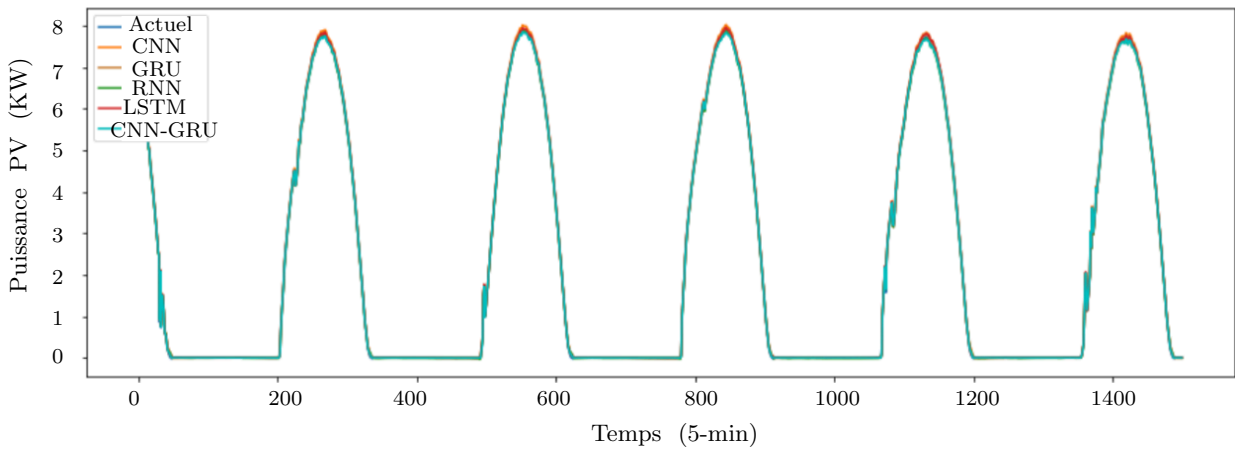


FIGURE 4.9 – Les résultats de prévision de la puissance PV en utilisant toutes les méthodes en hiver.

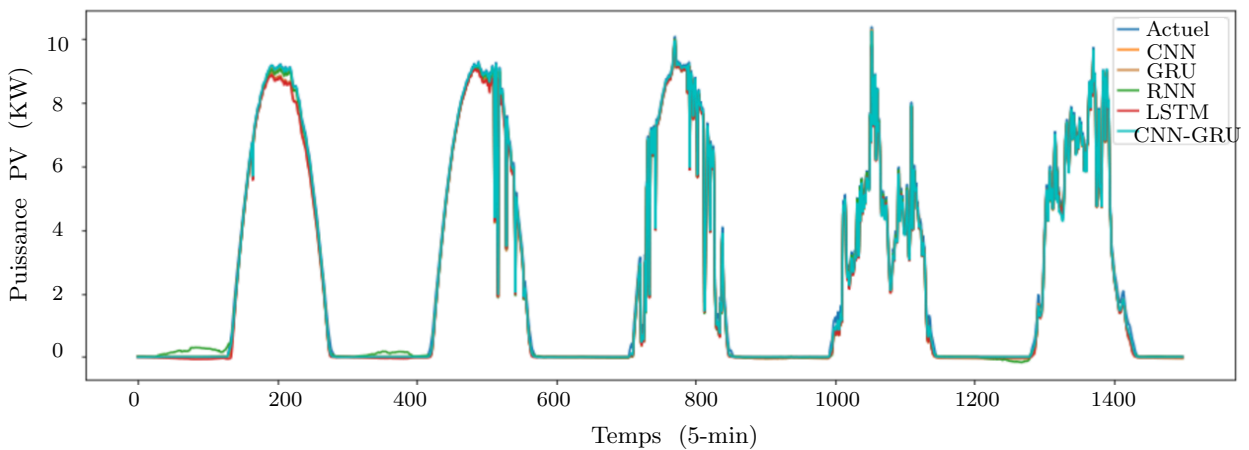


FIGURE 4.10 – Les résultats de prévision de la puissance PV en utilisant toutes les méthodes au printemps.

modèle proposé. Les résolutions des jeux de données dans les études de comparaison et notre étude sont toutes les deux de 5 minutes. La figure.4.12 illustre les résultats de comparaison avec d'autres études antérieures. Le modèle CNN-GRU présente une précision plus élevée que toutes les autres méthodes dans toutes les études. Ces résultats indiquent que le modèle CNN-GRU proposé dans cette étude offre une précision prédictive supérieure par rapport aux modèles utilisés dans les études antérieures. Cela souligne l'efficacité de la méthode proposée et renforce sa crédibilité en tant qu'outil de prédiction fiable pour la puissance PV. Ces résultats peuvent également contribuer à la validation et à l'acceptation de la méthode proposée par la communauté scientifique travaillant dans le domaine de la prédiction de la puissance PV. Il convient de noter que les études de comparaison utilisent des résolutions de

4.4. PRÉVISION DE LA PRODUCTION D'ÉNERGIE PV À L'AIDE D'UN MODÈLE HYBRIDE CNN-GRU

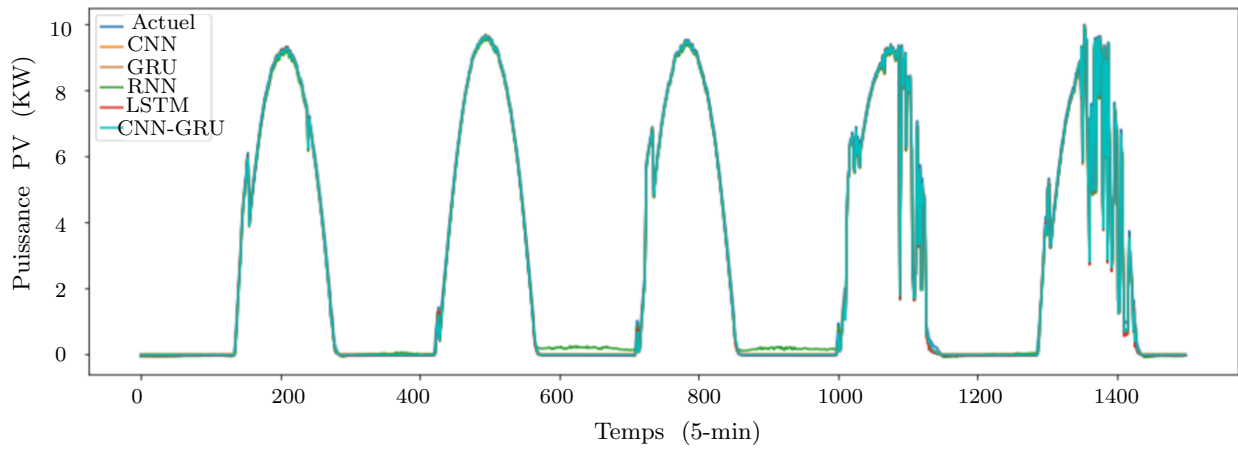


FIGURE 4.11 – Les résultats de prévision de la puissance PV en utilisant toutes les méthodes en été.

données similaires à notre étude, ce qui facilite la comparaison directe des performances des modèles. Cependant, il est important de prendre en compte les différences potentielles dans les caractéristiques des ensembles de données, les méthodologies et les paramètres utilisés dans chaque étude lors de l'interprétation et de la généralisation des résultats. Dans les études, notre modèle CNN-GRU suggéré surpasse toutes les autres méthodes. De plus, pour vérifier la robustesse du modèle proposé, l'ensemble de données est divisé en quatre parties : l'automne, l'hiver, le printemps et l'été. Dans certaines saisons, les modules CNN et GRU autonomes fonctionnent mieux, mais dans d'autres, ils ont de moins bonnes performances, tandis que le modèle suggéré performe bien dans toutes les saisons. La recherche montre que la combinaison des caractéristiques cachées des modèles CNN et LSTM dans la prévision de la production d'énergie PV est utile et peut augmenter la précision des prévisions. Ces résultats soulignent l'efficacité de la méthode proposée dans la prédiction de la production d'énergie PV à grande échelle. L'utilisation de l'approche hybride CNN-GRU permet de capturer les caractéristiques complexes et non linéaires des données, ce qui améliore la précision des prévisions. De plus, la division de l'ensemble de données en saisons permet de vérifier la robustesse du modèle dans des conditions variables. Ces résultats peuvent contribuer à la littérature existante en fournissant une méthode plus précise et fiable pour la prédiction de la production d'énergie PV, ce qui est essentiel pour la planification et l'optimisation des systèmes d'énergie renouvelable.

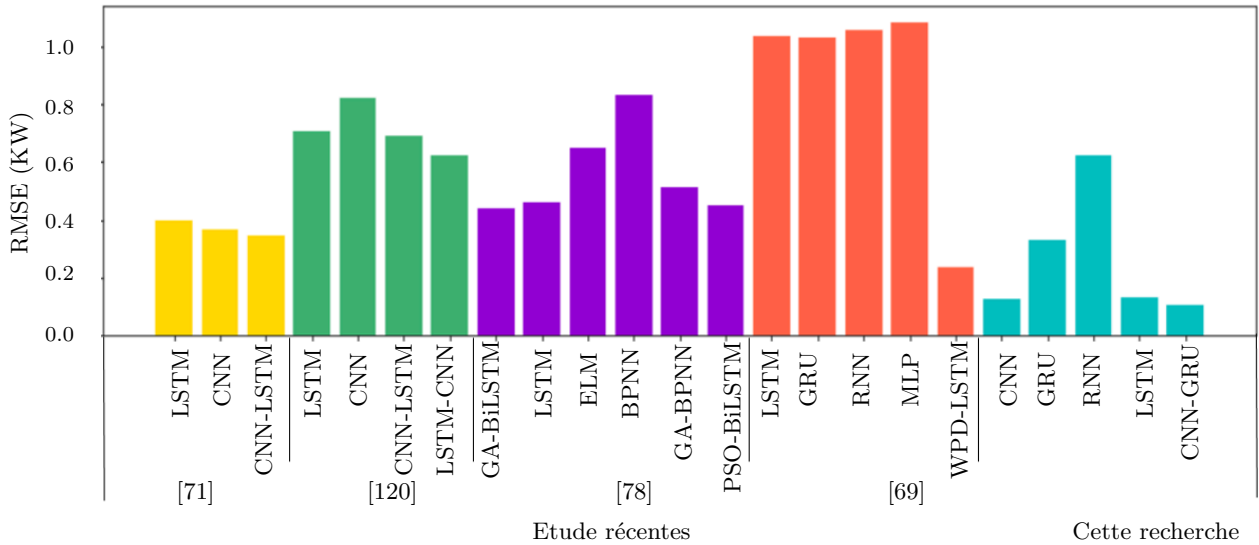


FIGURE 4.12 – Comparaison de modèle proposée avec d'autres études

4.5 Conclusion

L'objectif de cette recherche était de mettre en évidence l'importance croissante des approches d'apprentissage profond dans la prédiction précise de la puissance PV. La première recherche compare les modèles LSTM et RNN, et démontre que le modèle LSTM offre une meilleure précision de prédiction par rapport au modèle RNN. Cependant, des études futures devront inclure d'autres variables pour confirmer ces résultats. La deuxième recherche propose un nouveau modèle de prédiction basé sur GRU et CNN. Les résultats expérimentaux montrent que le modèle GRU-CNN offre les taux d'erreur les plus bas par rapport aux méthodes traditionnelles, grâce à l'utilisation de couches de convolution pour extraire les caractéristiques des données et de GRU pour stocker l'information en mémoire. Cette approche se révèle prometteuse pour prédire la puissance PV. La troisième recherche présente l'approche CNN-GRU pour la prédiction de la puissance PV. Les résultats de simulation montrent que le modèle proposé offre les plus petites erreurs de prédiction et les valeurs de R^2 les plus élevées parmi les méthodes de référence. Le modèle CNN-GRU se révèle capable de prédire la puissance PV dans différentes saisons avec une précision élevée. En conclusion, ces recherches démontrent que les approches d'apprentissage profond, telles que LSTM, GRU-CNN et CNN-GRU, offrent des résultats prometteurs pour la prédiction précise de la puissance PV. Cependant, des études futures devront approfondir ces approches et explorer l'intégration d'autres variables pour améliorer encore plus la

4.5. CONCLUSION

précision des prévisions à court terme et à moyen/long terme.

Chapitre 5

Amélioration de la précision de la prédiction de la puissance PV grâce à des modèles d'apprentissage profond avancés

5.1 Introduction

Ces dernières années, la prédiction précise de la production d'énergie PV est devenue de plus en plus importante pour garantir la sécurité, la stabilité et le bon fonctionnement des centrales solaires. La variabilité inhérente et l'intermittence de l'énergie solaire posent des défis significatifs pour obtenir des prévisions précises de la production d'énergie PV. Dans cette optique, ce chapitre présente deux études proposant des approches innovantes pour la prédiction de la production d'énergie PV en utilisant des techniques d'apprentissage profond. La première étude se concentre sur le développement d'un modèle hybride d'apprentissage profond combinant les architectures du réseau bidirectionnel à mémoire court-terme (BLSTM) et du réseau neuronal convolutif. L'objectif est de tirer parti des capacités complémentaires du BLSTM et du CNN pour améliorer la précision des prévisions de production d'énergie PV. Le modèle proposé est appliqué à un ensemble de données réel. En exploitant la puissance de l'apprentissage profond et en analysant les caractéristiques temporelles et spatiales des données, cette étude vise à améliorer la précision des prévisions de production d'énergie PV. La deuxième étude aborde l'imprévisibilité et l'imprécision de la production d'énergie PV, causées par les variations importantes des éléments météorologiques. Elle propose l'utilisation d'un modèle d'autoencodeur à mémoire court-terme (LSTM-AE) pour la prédiction de la production d'énergie PV. Ce modèle LSTM-AE est spécifiquement conçu pour la prédiction d'énergie PV, en exploitant les capacités de l'architecture LSTM pour extraire des caractéristiques temporelles significatives des données d'entrée. Les séquences encodées générées par l'encodeur LSTM sont ensuite décodées par le décodeur LSTM, permettant ainsi des prédictions précises de la production d'énergie PV. Les deux études reconnaissent l'importance cruciale d'une prédiction précise de la production d'énergie PV pour le bon fonctionnement et la sécurité des systèmes d'énergie solaire. En utilisant des techniques avancées d'apprentissage profond et en exploitant les caractéristiques uniques des données d'énergie PV, ces études visent à surmonter les défis liés à la variabilité et à l'intermittence de l'énergie solaire. Les modèles proposés présentent un grand potentiel pour améliorer la précision et la fiabilité des prévisions de production d'énergie PV, contribuant ainsi à l'avancement des systèmes d'énergie renouvelable. Dans les sections suivantes, nous examinerons en détail la méthodologie, la configuration expérimentale et les résultats de chaque étude, mettant en évidence leur contribution dans le domaine de la prédiction de la production d'énergie PV à l'aide de techniques d'apprentissage profond.

5.2 Prédiction de la génération d'énergie PV à l'aide d'un réseau double couche à mémoire à court terme récurrente bidirectionnelle-convolutionnelle

Dans le contexte des centrales solaires PVs, la précision de la prédiction de la puissance PV est d'une importance capitale pour assurer la sécurité et la stabilité de ces installations. Les principales contraintes qui affectent la précision des prévisions de puissance PV sont la variabilité et l'intermittence de l'énergie solaire. Par conséquent, cette étude propose un modèle hybride d'apprentissage profond pour la prédiction de la puissance PV. Ce modèle est développé avec succès en combinant le réseau de neurones récurrents bidirectionnels à mémoire à court terme et le réseau de neurones convolutifs, et il est appliqué à un ensemble de données réelles collectées sur le système PV DKASC. L'objectif de cette recherche est d'améliorer la précision des prévisions de puissance PV en utilisant un modèle hybride basé sur l'apprentissage profond. Le modèle hybride combine les avantages du BLSTM et du CNN pour capturer les caractéristiques temporelles et spatiales des données. Le BLSTM est utilisé pour modéliser les dépendances à long terme dans les séries temporelles de puissance PV, tandis que le CNN est utilisé pour extraire les caractéristiques spatiales des données. Cette combinaison permet de mieux tenir compte de la variabilité et de l'intermittence de l'énergie solaire, ce qui se traduit par une prédiction plus précise de la puissance PV.

5.2.1 Description de l'architecture double-couche du réseau BLSTM-CNN

Les modèles hybrides présentent en général de meilleures performances que les modèles simples. En prenant en compte les avantages du réseau BLSTM et du CNN, nous avons exploité leurs capacités complémentaires pour développer un nouveau modèle hybride permettant d'extraire efficacement les caractéristiques temporelles et spatiales et de prédire avec précision la génération d'énergie PV. Dans cet article, nous proposons une approche hybride appelée BLSTM-CNN, qui consiste en une connexion en série de BLSTM et de CNN. Cette approche se distingue par sa capacité à extraire des caractéristiques complexes et des motifs à partir des facteurs météorologiques utilisés pour la prédiction de la génération d'énergie PV. Les données historiques de génération d'énergie PV sont d'abord entrées dans le modèle BLSTM, qui est capable de traiter les séries temporelles et d'extraire les caractéristiques temporelles des données. Ces caractéristiques sont ensuite transmises à la couche d'entrée du modèle CNN pour extraire les caractéristiques spatiales des données. Un réseau CNN est généralement composé de plusieurs niveaux de couches de convolution et de pooling, effectuant de nombreuses

5.2. PRÉDICTION DE LA GÉNÉRATION D'ÉNERGIE PV À L'AIDE D'UN RÉSEAU DOUBLE COUCHE À MÉMOIRE À COURT TERME RÉCURRENTTE BIDIRECTIONNELLE-CONVOLUTIONNELLE

opérations de convolution à chaque niveau afin de capturer les informations importantes. Le CNN attribue des poids aux facteurs météorologiques en fonction de leur impact sur la génération d'énergie PV. Enfin, une couche entièrement connectée est utilisée pour regrouper les données et prédire la génération d'énergie PV en utilisant les caractéristiques extraites. De plus, une couche de dropout est introduite dans le modèle pour réduire le surapprentissage et améliorer sa généralisation.

5.2.2 Description des données utilisées

Dans le cadre de notre recherche, nous avons choisi d'étudier le système PV de 1B DKASC. Nous avons utilisé les données recueillies sur une période allant du 1er octobre 2020 au 27 janvier 2021, avec une résolution de 5 minutes. Les paramètres d'entrée comprennent des mesures telles que le rayonnement horizontal global, la température météorologique, le rayonnement horizontal diffus, la moyenne de la phase du courant, l'humidité relative et la direction du vent. Ces données sont essentielles pour notre recherche, car elles fournissent des informations détaillées sur les conditions météorologiques et les caractéristiques du système PV. L'objectif de notre étude est de prédire la puissance active du système PV (en kW) en fonction de ces paramètres. Pour améliorer l'efficacité et la précision des prévisions du modèle, il est nécessaire de prétraiter et filtrer les données avant de les utiliser. Le prétraitement des données comprend plusieurs étapes essentielles, telles que l'élimination des données aberrantes, la complétion des valeurs manquantes et la normalisation des données. Cette phase permet de garantir que les données sont propres et cohérentes avant d'être utilisées dans le modèle. Une fois les données prétraitées, elles sont séparées en deux ensembles distincts : 80 % des données sont réservées à l'entraînement du modèle et 20 % sont dédiées à son évaluation lors des tests. Cette division permet d'évaluer la performance du modèle sur des données qu'il n'a pas rencontrées pendant l'entraînement, ce qui fournit une estimation objective de ses capacités de prévision. Le modèle hybride BLSTM-CNN est composé de deux parties principales. La première partie utilise un outil de modélisation temporelle appelé BLSTM (Bidirectional Long Short-Term Memory) pour apprendre les dépendances à long terme dans les données après leur prétraitement. Le BLSTM est un type de réseau de neurones récurrents capable de prendre en compte les informations passées et futures pour effectuer des prédictions précises. La deuxième partie du modèle est un réseau de neurones convolutifs en une dimension (1D CNN) qui est utilisé pour extraire les caractéristiques spatiales des données. Cette architecture permet de détecter les schémas et les relations spatiales dans les données, ce qui

5.2. PRÉDICTION DE LA GÉNÉRATION D'ÉNERGIE PV À L'AIDE D'UN RÉSEAU DOUBLE COUCHE À MÉMOIRE À COURT TERME RÉCURRENTÉ BIDIRECTIONNELLE-CONVOLUTIONNELLE

peut être essentiel pour améliorer les prévisions. Dans le cadre de cette recherche, plusieurs modèles sont utilisés pour comparer les performances du modèle hybride BLSTM-CNN proposé. Les modèles individuels, tels que CNN, GRU, LSTM, RNN et BLSTM, sont entraînés et évalués pour prédire la puissance PV. De plus, deux modèles hybrides, LSTM-CNN et CNN-LSTM, sont également utilisés à des fins de comparaison. Ces modèles sont choisis en raison de leurs architectures différentes et de leurs capacités à capturer les caractéristiques des données temporelles. Les métriques utilisées pour évaluer l'efficacité et l'exactitude des prédictions des modèles sont le RMSE, le MSE, le MAE et le R^2 . Le RMSE mesure l'écart moyen entre les valeurs prédites et les valeurs réelles, le MSE quantifie la moyenne des erreurs au carré, le MAE mesure l'erreur moyenne absolue et le R^2 évalue la proportion de variance expliquée par le modèle. Ces métriques permettent d'évaluer la précision et la performance des différents modèles dans la prédiction de la puissance PV.

5.2.3 Configuration des paramètres expérimentaux

Cette recherche propose un modèle hybride BLSTM-CNN pour la prédiction de la puissance PV. Le modèle BLSTM est utilisé pour extraire les caractéristiques temporelles bidirectionnelles. Deux couches cachées ont été configurées en utilisant les données d'index filtrées dans le modèle BLSTM, avec des unités égales à 128 et 256 respectivement. Les caractéristiques temporelles ainsi obtenues sont ensuite envoyées à la couche d'entrée du modèle CNN, qui utilise des couches de convolution et des couches de pooling pour extraire les caractéristiques spatiales de l'ensemble de données. Dans le modèle CNN, deux couches de convolution et deux couches de pooling sont utilisées, avec respectivement 128 et 256 noyaux de convolution. Dans la couche de convolution, la taille du noyau est de 3*3. Une couche de dropout [121] est également incluse dans le modèle pour éviter les problèmes de surajustement lors de l'entraînement, ce qui pourrait réduire la précision des prédictions. La taille de lot du modèle proposé est de 500. Enfin, deux couches entièrement connectées avec 512 et 256 neurones respectivement sont utilisées pour produire les résultats de prévision de la génération de puissance PV. Les paramètres du modèle proposé dans cette étude sont résumés dans le tableau 5.1. Dans le cadre de cette recherche, un modèle hybride BLSTM-CNN a été développé pour la prédiction de la puissance PV. Le modèle combine les caractéristiques temporelles apprises par le modèle BLSTM avec les caractéristiques spatiales extraites par le modèle CNN. Des paramètres spécifiques ont été définis pour les différentes couches du modèle afin d'optimiser les performances de prédiction. L'inclusion

5.2. PRÉDICTION DE LA GÉNÉRATION D'ÉNERGIE PV À L'AIDE D'UN RÉSEAU DOUBLE COUCHE À MÉMOIRE À COURT TERME RÉCURRENTTE BIDIRECTIONNELLE-CONVOLUTIONNELLE

d'une couche de dropout vise à améliorer la généralisation du modèle en évitant le surajustement. Les détails des paramètres utilisés dans le modèle proposé sont présentés dans le tableau 5.1, fournissant ainsi une référence pour les expériences et les résultats obtenus. Cette approche hybride BLSTM-CNN pour la prédiction de la puissance PV constitue une contribution significative à la recherche dans ce domaine, et les résultats obtenus seront discutés et évalués dans le cadre de cette thèse.

TABLE 5.1 – Réglages des paramètres de la méthode proposée.

Type	Filtre	Taille du noyau	Pas
BLSTM (128)	-	-	-
BLSTM (256)	-	-	-
Convolution	128	3	1
Activation (Relu)	-	-	-
Max-Pooling	-	2	2
Convolution	256	3	1
Activation (Relu)	-	-	-
Max-Pooling	-	2	2
Dropout (0.2)	-	-	-
couche entièrement connectée (512)	-	-	-
couche entièrement connectée (256)	-	-	-

5.2.4 Protocole d'évaluation de la performance

Les résultats obtenus pour différents modèles d'apprentissage profond, tels que CNN, GRU, RNN, LSTM, BLSTM, CNN-LSTM et LSTM-CNN, ont été évalués en utilisant diverses mesures de performance, notamment le RMSE, le MAE, le MSE et le R^2 . Ces mesures permettent d'évaluer l'efficacité et l'exactitude des prédictions des modèles dans la prédiction de la puissance PV. Les résultats de prévision des huit modèles sont présentés dans cette tableau 5.2. Le modèle CNN a obtenu un RMSE de 0.1798, un MAE de 0.1325, un MSE de 0.0323 et un R^2 de 0.9975. Ces résultats indiquent une bonne performance du modèle CNN dans la prédiction de la puissance PV, avec une précision élevée et une faible erreur moyenne. Le modèle GRU a obtenu un RMSE de 0.2016, un MAE de 0.1752, un MSE de 0.1018 et un R^2 de 0.9968. Bien que légèrement moins performant que le modèle CNN, le modèle GRU présente toujours une précision satisfaisante et une erreur moyenne raisonnable. Le modèle RNN a obtenu un RMSE de 0.3191, un MAE de 0.2706, un MSE de 0.1018 et un R^2 de 0.9921. Les résultats indiquent que le modèle RNN présente une précision inférieure et une erreur moyenne plus élevée par rapport aux autres modèles évalués. Le modèle LSTM a obtenu un RMSE de 0.2812,

5.2. PRÉDICTION DE LA GÉNÉRATION D'ÉNERGIE PV À L'AIDE D'UN RÉSEAU DOUBLE COUCHE À MÉMOIRE À COURT TERME RÉCURRENTÉ BIDIRECTIONNELLE-CONVOLUTIONNELLE

un MAE de 0.2531, un MSE de 0,0791 et un R^2 de 0,9938. Ces résultats montrent que le modèle LSTM présente une performance raisonnable avec une précision acceptable et une erreur moyenne modérée. Le modèle BLSTM a obtenu un RMSE de 0.1609, un MAE de 0.1290, un MSE de 0.0259 et un R^2 de 0.9979. Ces résultats indiquent que le modèle BLSTM présente une précision élevée avec une erreur moyenne réduite par rapport aux autres modèles évalués. Le modèle CNN-LSTM a obtenu un RMSE de 0,1389, un MAE de 0.1064, un MSE de 0,0192 et un R^2 de 0.9985. Ces résultats témoignent d'une performance supérieure du modèle CNN-LSTM, avec une précision élevée et une erreur moyenne minimale. Le modèle LSTM-CNN a obtenu un RMSE de 0.1102, un MAE de 0.0794, un MSE de 0,0121 et un R^2 de 0.9990. Ces résultats indiquent une excellente performance du modèle LSTM-CNN, avec une précision très élevée et une erreur moyenne très faible. Enfin, le modèle BLSTM-CNN a obtenu les meilleurs résultats avec un RMSE de 0.0944, un MAE de 0.0531, un MSE de 0.0089 et un R^2 de 0.9993. Ces résultats confirment la supériorité du modèle hybride BLSTM-CNN.

TABLE 5.2 – Réglages des paramètres de la méthode proposée..

Modèle	RMSE	MAE	MSE	R^2
CNN [120]	0.1798	0.1325	0.0323	0.9975
GRU [122]	0.2016	0.1752	0.0406	0.9968
RNN [83]	0.3191	0.2706	0.1018	0.9921
LSTM [63]	0.2812	0.2531	0.0791	0.9938
BLSTM [80]	0.1609	0.1290	0.0259	0.9979
CNN-LSTM [123]	0.1389	0.1064	0.0192	0.9985
LSTM-CNN [71]	0.1102	0.0794	0.0121	0.9990
BLSTM-CNN	0.0944	0.0531	0.0089	0.9993

Les données de génération de puissance PV pour une journée en 2021 ont été choisies aléatoirement comme données de validation pour tester l'efficacité du modèle BLSTM-CNN. La figure.5.1 montre une comparaison entre les valeurs prévues et les valeurs réelles. Pour toutes les heures dans la plage considérée, les courbes de prévision montrent une forte cohérence avec les données réelles. Comme le montre la figure.5.2, l'erreur entre les valeurs prévues et les valeurs réelles est illustrée par la courbe en forme de rose.

Ces résultats visuels permettent d'évaluer la performance du modèle BLSTM-CNN dans la prédiction de la puissance PV. Les courbes de prévision montrent une adéquation étroite avec les données réelles, indiquant une précision élevée du modèle. De plus, la courbe en forme de rose illustre l'erreur

5.2. PRÉDICTION DE LA GÉNÉRATION D'ÉNERGIE PV À L'AIDE D'UN RÉSEAU DOUBLE COUCHE À MÉMOIRE À COURT TERME RÉCURRENTTE BIDIRECTIONNELLE-CONVOLUTIONNELLE

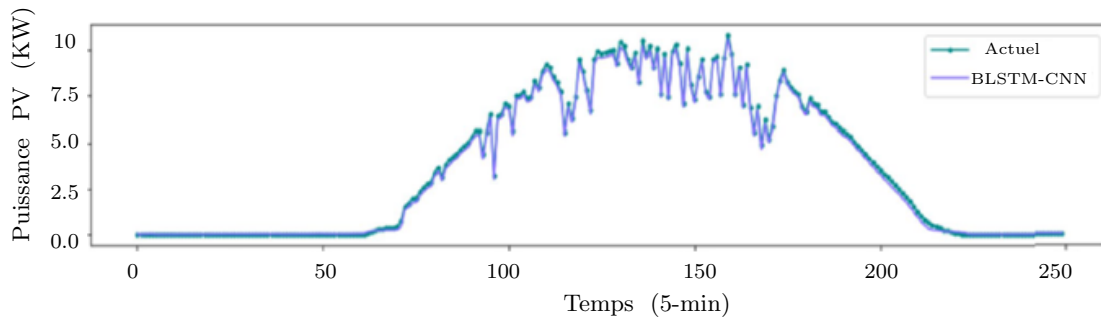


FIGURE 5.1 – Comparaison entre les prédictions du modèle et les valeurs réelles.

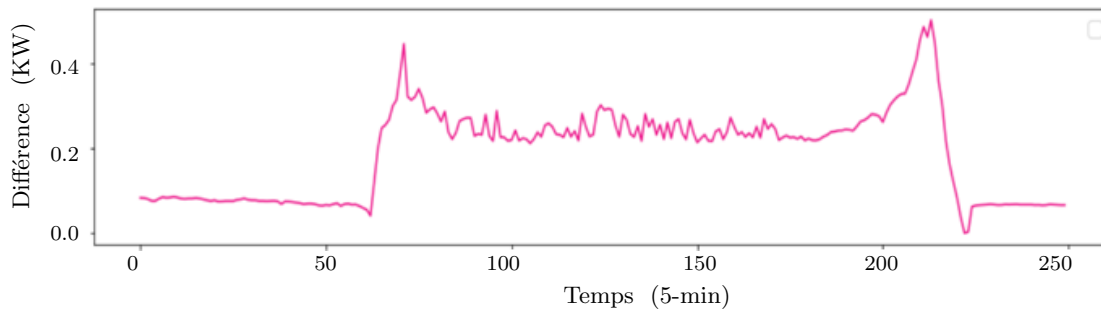


FIGURE 5.2 – Erreur entre les valeurs prévues et les valeurs réelles.

entre les valeurs prévues et les valeurs réelles, offrant ainsi une mesure visuelle de la précision du modèle. Ces résultats et visualisations confirment l'efficacité du modèle BLSTM-CNN dans la prédiction de la puissance PV. Ces informations peuvent être utilisées pour évaluer et valider la performance du modèle dans des scénarios réels, ce qui est essentiel pour la fiabilité et la pertinence des prévisions de la puissance PV.

5.2.5 Analyse des résultats en fonction de différentes périodes

les performances et la stabilité du modèle de prévision proposé sont testées sur une période de quatre mois afin de garantir la fiabilité et l'efficacité de la prévision BLSTM-CNN. Les données collectées sont séparées en quatre mois : octobre, novembre, décembre et janvier. Les données de chaque mois sont divisées en deux parties : 80% pour l'entraînement et 20% pour les tests. Le tableau 5.3 présente la répartition des ensembles d'entraînement et de test sur les quatre mois. Cette répartition permet de s'assurer que le modèle est entraîné sur une période suffisamment longue et variée pour capturer les différentes tendances et variations de la génération de puissance PV. Les ensembles d'entraînement

5.2. PRÉDICTION DE LA GÉNÉRATION D'ÉNERGIE PV À L'AIDE D'UN RÉSEAU DOUBLE COUCHE À MÉMOIRE À COURT TERME RÉCURRENTÉ BIDIRECTIONNELLE-CONVOLUTIONNELLE

et de test sont soigneusement sélectionnés pour garantir une représentativité adéquate des données et une évaluation précise des performances du modèle. Cette approche de partitionnement des données en mois et en ensembles d'entraînement/test permet d'évaluer la capacité du modèle à généraliser et à fournir des prévisions précises sur des périodes distinctes. Elle permet également d'identifier d'éventuelles variations de performances du modèle au fil du temps. Le tableau 5.4 présente les résultats de différents modèles, tels que CNN, GRU, RNN, LSTM, BLSTM, CNN-LSTM, LSTM-CNN et BLSTM-CNN, pour la prévision de la puissance PV sur les mois d'octobre, novembre, décembre et janvier. Les modèles hybrides surpassent les modèles simples en termes de précision de prédiction. De plus, l'effet de prédiction du modèle hybride BLSTM-CNN est meilleur que celui des deux autres modèles hybrides (LSTM-CNN et CNN-LSTM). Les résultats indiquent que le modèle BLSTM-CNN surpasse les autres modèles en termes de précision de prédiction sur les quatre mois. Bien que différents modèles puissent être utilisés pour prévoir la génération de puissance PV, aucun modèle unique ne se distingue toujours des autres de manière cohérente. Ces résultats confirment la capacité du modèle proposé à extraire des caractéristiques temporelles et spatiales, lui permettant ainsi de créer une relation complexe entre les données d'entrée et la génération de puissance PV cible. Les modèles hybrides, en combinant les avantages des différents modèles, parviennent à améliorer la précision de la prédiction. Dans le cas particulier du modèle BLSTM-CNN, il se distingue des autres modèles hybrides en termes de précision de prédiction.

TABLE 5.3 – Valeurs statistiques des données expérimentales de puissance PV (KW).

Mois	Échantillons	Nombre	Moy	Écart	Min	Max
Octobre	Tous les échantillons	8928	2.7547	3.5559	-0.0215	10.5396
	Échantillons d'entraînement	7142	2.6843	3.5070	-0.0203	10.5396
	Échantillons de test	1786	3.0359	3.7330	-0.0215	10.0366
Novembre	Tous les échantillons	8639	2.8974	3.5440	-0.0199	11.0413
	Échantillons d'entraînement	6911	2.8447	3.5168	-0.0199	11.0413
	Échantillons de test	1728	3.1084	3.6440	-0.0184	10.4283
Décembre	Tous les échantillons	8908	2.3949	3.2452	-0.0188	11.6942
	Échantillons d'entraînement	7126	2.2086	3.0685	-0.0187	10.8382
	Échantillons de test	1782	3.1400	3.7824	-0.0188	11.6942
Janvier	Tous les échantillons	7658	2.9545	3.6139	-0.0222	11.4737
	Échantillons d'entraînement	6126	3.0698	3.6956	-0.0222	11.4737
	Échantillons de test	1532	2.4936	3.2268	-0.0158	10.5551

5.2. PRÉDICTION DE LA GÉNÉRATION D'ÉNERGIE PV À L'AIDE D'UN RÉSEAU DOUBLE COUCHE À MÉMOIRE À COURT TERME RÉCURRENTÉ BIDIRECTIONNELLE-CONVOLUTIONNELLE

TABLE 5.4 – Performance de prévision des différentes méthodes d'apprentissage profond pour chaque saison.

Ensemble de données	Modèle	RMSE	MAE	MSE	R ²
Octobre	CNN	0.2139	0.1514	0.0457	0.9967
	GRU	0.1205	0.0803	0.0145	0.9989
	RNN	0.1928	0.1335	0.0371	0.9973
	LSTM	0.1517	0.1188	0.0230	0.9983
	BLSTM	0.1240	0.1085	0.0153	0.9988
	CNN-LSTM	0.1083	0.0843	0.0117	0.9991
	LSTM-CNN	0.0795	0.0516	0.0063	0.9995
	BLSTM-LSTM	0.0471	0.0350	0.0022	0.9998
Novembre	CNN	0.3121	0.2626	0.0975	0.9926
	GRU	0.4720	0.3851	0.2228	0.9832
	RNN	0.5490	0.4462	0.3014	0.9772
	LSTM	0.3941	0.3425	0.1553	0.9882
	BLSTM	0.3443	0.3265	0.1185	0.9910
	CNN-LSTM	0.1986	0.1844	0.0394	0.9970
	LSTM-CNN	0.1524	0.1265	0.0232	0.9982
	BLSTM-CNN	0.1333	0.1114	0.0177	0.9986
Décembre	CNN	0.1430	0.1268	0.0225	0.9984
	GRU	0.2656	0.2069	0.0705	0.9950
	RNN	0.1696	0.1220	0.0287	0.9979
	LSTM	0.3569	0.2846	0.1273	0.9910
	BLSTM	0.1896	0.1411	0.0359	0.9974
	CNN-LSTM	0.1300	0.1051	0.0169	0.9988
	LSTM-CNN	0.1158	0.0797	0.0134	0.9990
	BLSTM-CNN	0.0728	0.0591	0.0053	0.9996
Janvier	CNN	0.1153	0.0909	0.0133	0.9987
	GRU	0.4503	0.3246	0.2027	0.9805
	RNN	0.4104	0.3155	0.1684	0.9838
	LSTM	0.1512	0.1164	0.0228	0.9978
	BLSTM	0.1541	0.1042	0.0237	0.9977
	CNN-LSTM	0.1020	0.0664	0.0104	0.9989
	LSTM-CNN	0.0914	0.0642	0.0083	0.9991
	BLSTM-CNN	0.0939	0.0608	0.0088	0.9991

5.2.6 Comparaison des performances du modèle proposé avec d'autres approches

Les données historiques utilisées dans cette étude proviennent du DKASC. Différentes méthodologies ont été appliquées et examinées sur des centrales solaires du DKASC dans des études précédentes. Les résultats de comparaison sont présentés dans la figure.5.3. Le DKASC, en tant que source de

5.2. PRÉDICTION DE LA GÉNÉRATION D'ÉNERGIE PV À L'AIDE D'UN RÉSEAU DOUBLE COUCHE À MÉMOIRE À COURT TERME RÉCURRENTTE BIDIRECTIONNELLE-CONVOLUTIONNELLE

données, a été largement utilisé dans la recherche en raison de sa fiabilité et de sa disponibilité. Les méthodologies précédemment utilisées ont permis de comparer les performances des différents modèles de prévision de la génération d'énergie solaire. Cette comparaison vise à évaluer l'efficacité des modèles proposés dans cette étude en les confrontant aux méthodologies de référence précédemment utilisées. La figure 5.3 offre une visualisation claire des résultats de cette comparaison. Elle permet de comparer les performances des différentes méthodologies en termes de précision de la prédiction de la génération d'énergie solaire. Ces résultats sont essentiels pour évaluer l'efficacité du modèle proposé par rapport aux approches antérieures et pour démontrer les améliorations apportées par la nouvelle méthodologie.

L'étude menée par Chen et al. [124] met en avant l'utilisation d'un modèle RCC-LSTM pour prédire la puissance PV. Leur approche repose sur la méthode RCC, qui permet de regrouper les périodes de temps similaires, et sur l'utilisation de LSTM pour extraire les caractéristiques des données de puissance PV. Les résultats obtenus montrent une valeur MAE moyenne de 0.587, ce qui indique une certaine précision dans les prédictions réalisées par ce modèle. Les données utilisées dans cette étude ont été collectées sur une période de deux ans à Yulara, Alice Springs, et ont été enregistrées avec une résolution de 5 minutes. Cela permet d'avoir une granularité temporelle fine pour la prédiction de la puissance PV. Les résultats obtenus fournissent des informations sur l'efficacité et la performance du modèle RCC-LSTM dans ce contexte spécifique. Zhou et al. [69] ont proposé un modèle hybride d'apprentissage profond (WPD-LSTM) pour la prévision à court terme de la puissance PV. Les données ont été collectées à partir du DKASC, à Alice Springs, du 1er juin 2014 au 12 juin 2016. La valeur MAE moyenne obtenue est de 0.2357. Zhou et al. [125] ont proposé un modèle hybride (SDA-GA-ELM) basé sur l'extrême learning machine (ELM), l'algorithme génétique (GA) et l'analyse de jours similaires personnalisée (SDA) pour prévoir la génération horaire de puissance PV. L'ensemble de données a été collecté du 14 janvier 2017 au 15 octobre 2018, avec une résolution d'une heure à partir du DKASC. La valeur MAE moyenne obtenue est de 0,2367. Dans ces deux études menées par Zhou et al., différentes approches de modélisation ont été utilisées pour la prévision de la puissance PV. Dans le modèle WPD-LSTM, une combinaison d'apprentissage profond hybride a été utilisée, tandis que dans le modèle SDA-GA-ELM, une combinaison d'ELM, d'algorithme génétique et d'analyse de jours similaires personnalisée a été appliquée. Les données utilisées dans ces études ont été collectées sur des périodes de temps spécifiques à partir du DKASC, à Alice Springs, et ont été enregistrées avec différentes résolutions (5 minutes dans la première étude et 1 heure dans la

5.2. PRÉDICTION DE LA GÉNÉRATION D'ÉNERGIE PV À L'AIDE D'UN RÉSEAU DOUBLE COUCHE À MÉMOIRE À COURT TERME RÉCURRENTTE BIDIRECTIONNELLE-CONVOLUTIONNELLE

deuxième étude). Les résultats obtenus, mesurés par la valeur MAE moyenne, indiquent une précision raisonnable dans les prévisions réalisées par les modèles proposés. Les travaux de Zhou et al. Apportent donc une contribution significative à la recherche sur la prévision de la puissance PV en proposant des approches hybrides basées sur l'apprentissage profond et des techniques telles que l'ELM et l'algorithme génétique. Ces résultats peuvent être utiles aux chercheurs et praticiens du domaine de l'énergie solaire pour améliorer la précision des prévisions de la puissance PV à court terme. Wang et al. [71] ont proposé un modèle hybride (LSTM-CNN) pour la prévision de la puissance PV. Les données ont été collectées à partir du DKASC, à Alice Springs, sur une période de six mois avec une résolution de l'ensemble de données historiques de 5 minutes. La valeur MAE moyenne obtenue est de 0.2210. Zhen et al. [78] ont proposé un modèle hybride (GA-BLSTM) pour la prédiction ultra-court terme de la puissance PV. Les données ont été collectées à partir de 8 centrales PV sur une période s'étalant de 2017 à 2019, avec une résolution de l'ensemble de données historiques de 5 minutes. La valeur MAE moyenne obtenue est de 0.242. Abdel-Basset et al. [126] ont proposé une nouvelle architecture d'apprentissage profond appelée PV-Net, permettant l'extraction efficace de caractéristiques positionnelles et temporelles dans les séquences de puissance PV. Les portes du GRU sont modifiées en utilisant des couches convolutives (nommées Conv-GRU) pour la prévision à court terme de la production d'énergie PV. Les données ont été collectées à partir du DKASC, à Alice Springs, sur une période de cinq ans (2015-2019) avec une résolution de l'ensemble de données historiques de 5 minutes. La valeur MAE moyenne obtenue est de 0.398. Ces trois études menées respectivement par Wang et al., Zhen et al., et Abdel-Basset et al. proposent des modèles hybrides pour la prévision de la puissance PV, en combinant différentes architectures d'apprentissage profond et techniques spécifiques. Les données utilisées dans ces études ont été collectées à partir de diverses sources, notamment le DKASC à Alice Springs, et enregistrées avec une résolution de 5 minutes. Les valeurs MAE moyennes obtenues indiquent une précision raisonnable dans les prévisions réalisées par les modèles proposés. Ces travaux de recherche contribuent à l'amélioration des méthodes de prévision de la puissance PV en explorant des approches hybrides qui combinent les avantages de différentes architectures d'apprentissage profond et techniques spécifiques. Les résultats obtenus peuvent être utiles pour les chercheurs et les praticiens du domaine de l'énergie solaire afin d'améliorer la précision des prévisions de la puissance PV à court terme. Lorsque les résultats des recherches mentionnées ci-dessus sont comparés, le modèle BLSTM-CNN suggéré présente la valeur MAE minimale. Il est évident que le modèle proposé surpasse les recherches

5.2. PRÉDICTION DE LA GÉNÉRATION D'ÉNERGIE PV À L'AIDE D'UN RÉSEAU DOUBLE COUCHE À MÉMOIRE À COURT TERME RÉCURRENTÉ BIDIRECTIONNELLE-CONVOLUTIONNELLE

antérieures et offre de meilleures performances de prédiction de la génération PV.

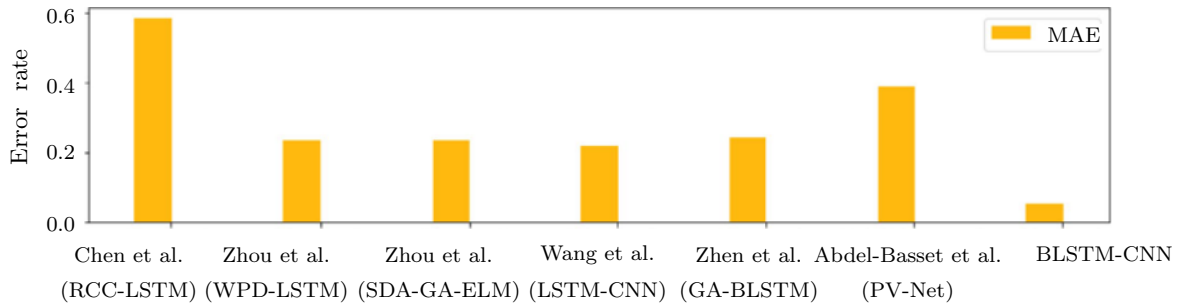


FIGURE 5.3 – Comparaison du modèle hybride proposé BLSTM-CNN avec des études précédentes..

5.2.7 Discussion sur les avantages et les limites des modèles

La prévision précise de la production d'énergie PV joue un rôle crucial dans la maintenance, le contrôle, la gestion et l'exploitation des systèmes de génération d'énergie solaire. Dans cette recherche, nous avons proposé un nouveau modèle hybride de prévision de la production d'énergie PV basé sur un algorithme d'apprentissage profond, le BLSTM-CNN, afin d'augmenter la précision et la fiabilité de la prévision de la production d'énergie solaire. Plus spécifiquement, le modèle BLSTM extrait automatiquement les caractéristiques de corrélation temporelle bidirectionnelle des données PVs, tandis que le modèle CNN extrait les caractéristiques de corrélation spatiale des données PVs pour produire les résultats finaux de prévision de la production d'énergie solaire. Pour valider les performances de prévision du modèle proposé, nous avons utilisé quatre mesures d'évaluation et cinq modèles d'apprentissage profond individuels (CNN, GRU, RNN, LSTM, BLSTM), ainsi que deux modèles hybrides (LSTM-CNN, CNN-LSTM), dans le cadre d'une étude expérimentale. Le modèle BLSTM-CNN, proposé de manière novatrice, a obtenu les meilleurs résultats avec une valeur R^2 de 0.9993, une valeur RMSE de 0.0944, une valeur MAE de 0.0531 et une valeur MSE de 0.0089. En termes de précision de prévision, les résultats indiquent que le modèle proposé surpasse les autres modèles classiques traditionnels. Dans les études futures, nous envisageons de combiner le modèle hybride avec des modèles d'apprentissage profond plus sophistiqués afin d'extraire séparément les caractéristiques temporelles et spatiales, ce qui permettra d'obtenir des résultats de prévision de la production d'énergie solaire plus précis. De plus, le modèle proposé peut également être amélioré et utilisé dans d'autres domaines, tels que la prévision de la vitesse du vent et la prévision de la charge résidentielle. En conclusion, notre

5.3. PRÉVISION DE LA PRODUCTION D'ÉNERGIE PV À L'AIDE DE RÉSEAUX AUTOENCODEURS À MÉMOIRE À COURT TERME

recherche a démontré l'importance de la prévision précise de la production d'énergie PV et a proposé un modèle hybride novateur, le BLSTM-CNN, qui améliore significativement l'exactitude et la fiabilité des prévisions. Les résultats obtenus montrent que ce modèle surpasse les modèles classiques existants et ouvre la voie à de nouvelles opportunités d'application dans d'autres domaines. En utilisant des approches d'apprentissage profond plus avancées, il est possible d'améliorer davantage les performances de prévision et d'obtenir des résultats plus précis pour la prévision de la production d'énergie solaire. Cependant, le modèle suggéré présente quelques défauts qui doivent être approfondis. Par exemple, dans cette recherche, la structure et les hyperparamètres d'entraînement du modèle ont été déterminés par expérimentation, ce qui est chronophage. Par conséquent, des approches d'estimation automatisée des paramètres, telles que des algorithmes d'optimisation heuristique, seront utilisées dans nos futures recherches afin de sélectionner et d'améliorer de manière plus efficace les paramètres du réseau neuronal. Cette observation souligne l'importance de continuer à améliorer et à optimiser le modèle proposé. Les chercheurs reconnaissent la nécessité de recherches futures visant à automatiser le processus de sélection des paramètres du modèle, afin de réduire le temps et les efforts nécessaires pour obtenir des performances optimales. Ces efforts permettront de renforcer la fiabilité et la précision des prédictions de génération PV, ce qui est essentiel pour soutenir le déploiement efficace des systèmes d'énergie solaire.

5.3 Prévision de la production d'énergie PV à l'aide de réseaux autoencod- deurs à mémoire à court terme

Dans de nombreux pays développés, l'énergie solaire PV occupe une position prépondérante dans la production d'électricité en tant que source d'énergie renouvelable rentable. Cependant, la génération d'énergie PV est caractérisée par une grande variabilité, souvent influencée par des facteurs météorologiques, ce qui rend sa prévision imprévisible et peu précise. Afin de réduire l'influence négative de l'utilisation de l'énergie PV, une prédiction précise de la puissance PV revêt une importance cruciale pour assurer le fonctionnement sécurisé et efficace du système d'alimentation PV. Dans cette recherche, nous proposons un autoencodeur à mémoire à court et long terme pour la prévision de la puissance PV. Tout d'abord, pour générer des séquences encodées, le LSTM-encodeur extrait les caractéristiques des données d'entrée. Ensuite, le LSTM-décodeur décode les séquences encodées pour les faire avancer vers la dernière couche dense afin de prédire la puissance PV. Cette approche de l'autoencodeur LSTM

5.3. PRÉVISION DE LA PRODUCTION D'ÉNERGIE PV À L'AIDE DE RÉSEAUX AUTOENCODEURS À MÉMOIRE À COURT TERME

offre une méthode prometteuse pour la prédiction de la puissance PV en exploitant les caractéristiques clés extraites des données d'entrée. En utilisant cette méthode, il est possible d'améliorer la précision des prévisions de la puissance PV, ce qui contribue à une meilleure gestion et utilisation de l'énergie PV dans les systèmes d'alimentation.

5.3.1 Rassemblement des données PV

Dans le cadre de cette recherche, la base de données couvre la période de décembre 2017 à mars 2021, avec une résolution de 5 minutes. Cette base de données est divisé en quatre périodes, à savoir la période 1, la période 2, la période 3 et la période 4. Pour chaque période, la base de données est séparé en deux sous-ensembles : l'ensemble d'entraînement et l'ensemble de test. L'ensemble d'entraînement et l'ensemble de test sont utilisés respectivement pour entraîner les modèles de prédiction et évaluer leurs performances en matière de prévisions, la base de données de chaque période a été divisé en deux ensembles, où 80% des données ont été utilisées pour l'entraînement et 20% pour les tests. Les différentes divisions des données sont présentées dans le tableau 5.5.

TABLE 5.5 – Descriptions de l'ensemble d'entraînement et de l'ensemble de test pour chaque période.

Période	Ensemble d'entraînement	Ensemble de test
Période 1	(2017/12/01–2018/04/26)	(2018/04/27–2018/06/01)
Période 2	(2018/06/01–2019/05/02)	(2019/05/03–2019/07/01)
Période 3	(2019/07/01–2020/06/09)	(2020/06/10–2020/09/01)
Période 4	(2020/09/01–2021/01/13)	(2021/01/14–2021/03/01)

5.3.2 Architecture et paramètres des modèles de prévision

Cette section se concentre sur la discussion des résultats expérimentaux visant à comparer les performances de plusieurs modèles, à savoir BPNN [127], GRU [83], RNN [29], LSTM [119], ainsi que le modèle LSTM-AE proposé. Dans notre approche, l'encodeur LSTM est configuré avec 64 et 32 unités de mémoire respectivement, permettant ainsi d'encoder les séquences d'entrée. Pour décoder les cartes de caractéristiques obtenues à partir des couches de l'encodeur, nous utilisons une couche "repeat vector". Le rôle principal de cette couche est de répéter le vecteur de sortie final de la couche d'encodage, de manière à fournir une entrée constante pour chaque étape de décodage. Les séquences ainsi encodées sont ensuite introduites dans un autre LSTM, qui les décode. Chaque couche LSTM

5.3. PRÉVISION DE LA PRODUCTION D'ÉNERGIE PV À L'AIDE DE RÉSEAUX AUTOENCODEURS À MÉMOIRE À COURT TERME

est composée de 32 et 64 unités de mémoire respectivement. Enfin, une couche dense est utilisée pour générer les prédictions de sortie pour la séquence d'entrée. L'architecture comprend donc 2 couches LSTM pour l'encodeur, 1 couche "repeat vector", 2 couches LSTM pour le décodeur, et enfin 1 couche entièrement connectée, soit un total de 6 couches. Les paramètres du modèle LSTM-AE proposé sont présentés dans le tableau 5.6, tandis que les structures et les paramètres des modèles de référence sont répertoriés dans le tableau 5.7. Les hyperparamètres ont été ajustés manuellement par essais et erreurs, en sélectionnant les meilleurs résultats en fonction de leur score de validation. Pour réduire l'erreur entre les vecteurs d'entrée et de sortie, nous avons utilisé la fonction d'activation ReLU. Pour cette expérience, la taille du lot est de 300, le nombre d'itérations dans chaque époque est de 100, le MSE est utilisée pour optimiser les paramètres du réseau, et le modèle a été compilé avec l'optimiseur ADAM.

TABLE 5.6 – Réglages des paramètres de la méthode proposée dans cette étude.

	Description des paramètres	Valeurs
LSTM Encodeur	Le nombre de couches denses de la LSTM	64
	La fonction d'activation de la LSTM	Relu
	Le nombre de couches denses de la LSTM	32
	La fonction d'activation de la LSTM	Relu
	Longueur du vecteur ReaptVector	6
LSTM Decodeur	Le nombre de couches denses de la LSTM	32
	La fonction d'activation de la LSTM	Relu
	The number of the Dense of LSTM	64
	La fonction d'activation de la LSTM	Relu
Dense	Le nombre de couches cachées	1
Other	Taille du lot	300
	Époques	100
	Optimiseur	Adam
	Fonction de perte	MSE

TABLE 5.7 – Réglages des paramètres des modèles de référence simples dans cette étude.

Modèles	Configuration
GRU	Unités GRU 64,32,32,64 ; Unités Dense 1 ; Activation 'relu'
RNN	Unités RNN 64,32,32,64 ; Unités Dense 1 ; Activation 'relu'
LSTM	Unités LSTM 64,32,32,64 ; Unités Dense 1 ; Activation 'relu'
BPNN	Unités Dense 64,32,32,64,1 ; Activation 'relu'

5.3.3 Métriques d'évaluation de la performance

Le tableau 5.8 présente la comparaison en mettant en évidence les erreurs de prédiction obtenues par le modèle LSTM-AE par rapport aux autres méthodes. Les résultats obtenus ont permis de vérifier que le modèle LSTM-AE proposé permet de réaliser des prévisions plus précises. Les indices moyens utilisés confirment que le modèle LSTM-AE proposé présente de bonnes performances dans le cadre de la prédiction de la puissance PV. En d'autres termes, les évaluations et les mesures réalisées indiquent que le modèle LSTM-AE est capable de fournir des prévisions de puissance PV de qualité, avec une précision satisfaisante. Cela renforce la validité et la pertinence du modèle LSTM-AE dans le contexte de la prédiction de la puissance PV. Le modèle proposé présente une plus grande précision et une meilleure stabilité dans ses prévisions, notamment lorsqu'il est comparé à quatre autres méthodes de prédiction. Cette constatation est étayée par les résultats obtenus pour différentes périodes d'évaluation. Le tableau 5.8 présente une comparaison précise entre le modèle de prédiction LSTM-AE et les quatre autres modèles de prédiction, en ce qui concerne le R^2 . La valeur moyenne de R^2 du modèle proposé atteint jusqu'à 0.993, ce qui est le maximum parmi tous les modèles. Cela indique que le modèle LSTM-AE présente une meilleure précision et une plus grande stabilité dans ses prévisions pour différentes périodes. En utilisant le R^2 comme mesure, on peut conclure que le modèle LSTM-AE offre une excellente adéquation aux données et une capacité de prédiction supérieure par rapport aux autres modèles évalués.

Les pourcentages d'amélioration de l'erreur absolue moyenne, de l'erreur quadratique moyenne et de l'erreur quadratique moyenne racine pour les modèles BPNN, GRU, RNN et LSTM-AE, présente dans le tableau 5.9. Par exemple, dans la période 1, les valeurs de P_{MAE} , P_{MSE} et P_{RMSE} de BPNN par rapport au modèle LSTM-AE sont respectivement de 63,73%, 69,78% et 44,85%. Pour GRU par rapport au modèle LSTM-AE, les valeurs sont de 46,74%, 61,43% et 37,80%. Pour RNN par rapport au modèle LSTM-AE, les valeurs sont de 50%, 48,40% et 27,91%. Enfin, pour LSTM par rapport au modèle LSTM-AE, les valeurs sont de 34,30%, 39,24% et 21,90%. Lorsque le modèle LSTM est comparé au modèle LSTM-AE, il est observé que la structure de l'autoencodeur a amélioré la précision de prévision du modèle LSTM. Ces pourcentages mettent en évidence l'amélioration significative de la précision des prévisions obtenue grâce à l'utilisation du modèle LSTM-AE par rapport aux autres modèles évalués.

5.3. PRÉVISION DE LA PRODUCTION D'ÉNERGIE PV À L'AIDE DE RÉSEAUX AUTOENCODEURS À MÉMOIRE À COURT TERME

TABLE 5.8 – Les résultats de prévision pour différentes périodes et différents modèles.

Période	Erreur	LSTM-AE	GRU	RNN	LSTM	BPNN
Période 1	MAE	0.0540	0.1014	0.1080	0.0822	0.1489
	MSE	0.0113	0.0293	0.0219	0.0186	0.0374
	RMSE	0.1066	0.1714	0.1480	0.1365	0.1933
	MBE	-0.0010	0.0924	0.0089	0.0376	-0.0936
	R ²	0.9995	0.9992	0.9992	0.9993	0.9986
Période 2	MAE	0.0522	0.1337	0.2196	0.0842	0.2475
	MSE	0.0074	0.0290	0.0574	0.0170	0.0889
	RMSE	0.0863	0.1704	0.2397	0.1305	0.2982
	MBE	-0.0440	-0.1337	-0.0442	-0.0837	-0.2280
	R ²	0.9990	0.9961	0.9924	0.9977	0.9882
Période 3	MAE	0.0311	0.0508	0.0926	0.0421	0.3461
	MSE	0.0055	0.0060	0.0157	0.0056	0.1002
	RMSE	0.0739	0.0780	0.1256	0.0747	0.3338
	MBE	-0.0166	0.0145	-0.0160	-0.0182	-0.1536
	R ²	0.9992	0.9991	0.9978	0.9992	0.9722
Période 4	MAE	0.0186	0.0482	0.0216	0.0303	0.0223
	MSE	0.0014	0.0053	0.0015	0.0024	0.0016
	RMSE	0.0381	0.0731	0.0394	0.0498	0.0400
	MBE	-0.0078	-0.0380	-0.0060	-0.0182	-0.0135
	R ²	0.9998	0.9995	0.9998	0.9998	0.9998
Moyenne	MAE	0.0389	0.0835	0.1104	0.0597	0.1912
	MSE	0.0064	0.0174	0.0241	0.0108	0.0820
	RMSE	0.0762	0.1232	0.1381	0.0978	0.2163
	MBE	-0.0197	-0.0162	-0.0143	-0.0232	-0.1228
	R ²	0.9993	0.9984	0.9973	0.9990	0.9897

Les figures.5.4-5.7 présentent la comparaison des résultats de prévision de la puissance PV sur une période de cinq jours avec les valeurs réelles de la puissance PV pour le modèle proposé. Les figures mettent en évidence que la puissance PV est proche de zéro pendant la nuit. Les prévisions des cinq modèles montrent une précision satisfaisante, cependant la courbe de puissance prédite par le modèle LSTM-AE se rapproche davantage de la valeur réelle. Les résultats de prévision du modèle proposé surpassent ceux des modèles BPNN, GRU, RNN et LSTM. La courbe représentant les valeurs réelles et la courbe du modèle LSTM-AE sont les plus similaires et présentent une meilleure performance de prédiction, notamment pendant la nuit et aux moments de puissance maximale. Ces observations confirment que le modèle LSTM-AE offre des prévisions plus précises et plus proches de la réalité par

5.3. PRÉVISION DE LA PRODUCTION D'ÉNERGIE PV À L'AIDE DE RÉSEAUX AUTOENCODEURS À MÉMOIRE À COURT TERME

TABLE 5.9 – Pourcentage de promotion de la MAE, MSE et RMSE entre le modèle LSTM-AE et le modèle de réseau simple..

Période	Indices	BPNN vs LSTM-AE	GRU vs LSTM-AE	RNN vs LSTM-AE	LSTM vs LSTM-AE
Période 1	P_{MAE}	63.73	46.74	50.00	34.30
	P_{MSE}	69.78	61.43	48.40	39.24
	P_{RMSE}	44.85	37.80	27.91	21.90
Période 2	P_{MAE}	87.43	60.95	76.22	38.00
	P_{MSE}	94.51	74.48	87.10	56.44
	P_{RMSE}	71.05	49.35	63.99	33.86
Période 3	P_{MAE}	91.01	38.77	66.41	26.12
	P_{MSE}	94.51	8.33	64.96	1.78
	P_{RMSE}	88.58	5.25	41.16	1.07
Période 4	P_{MAE}	16.52	61.41	13.88	38.61
	P_{MSE}	12.50	73.58	6.66	41.66
	P_{RMSE}	4.75	47.87	3.29	29.69

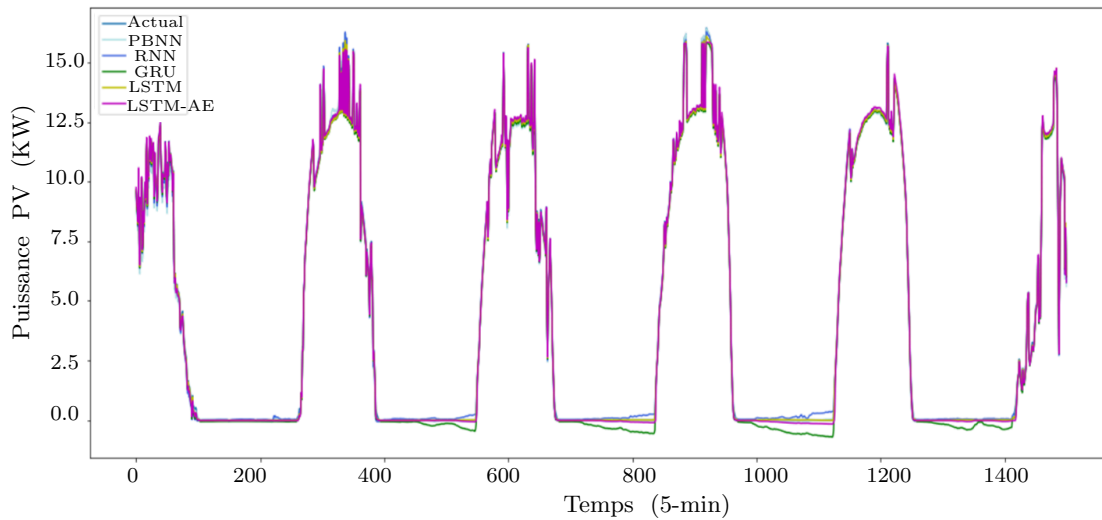


FIGURE 5.4 – Prédiction de la puissance PV en utilisant toutes les méthodes pendant la période 1.

rapport aux autres modèles évalués, en particulier dans des situations de faible puissance et de pointe. Le tableau 5.10 présente les durées d'entraînement des différents modèles pendant différentes périodes. Le modèle BPNN affiche le temps d'entraînement moyen le plus court, à savoir 33.72 secondes. Les modèles GRU, RNN et LSTM ont respectivement des temps d'entraînement moyens de 743.9 secondes,

5.3. PRÉVISION DE LA PRODUCTION D'ÉNERGIE PV À L'AIDE DE RÉSEAUX AUTOENCODEURS À MÉMOIRE À COURT TERME

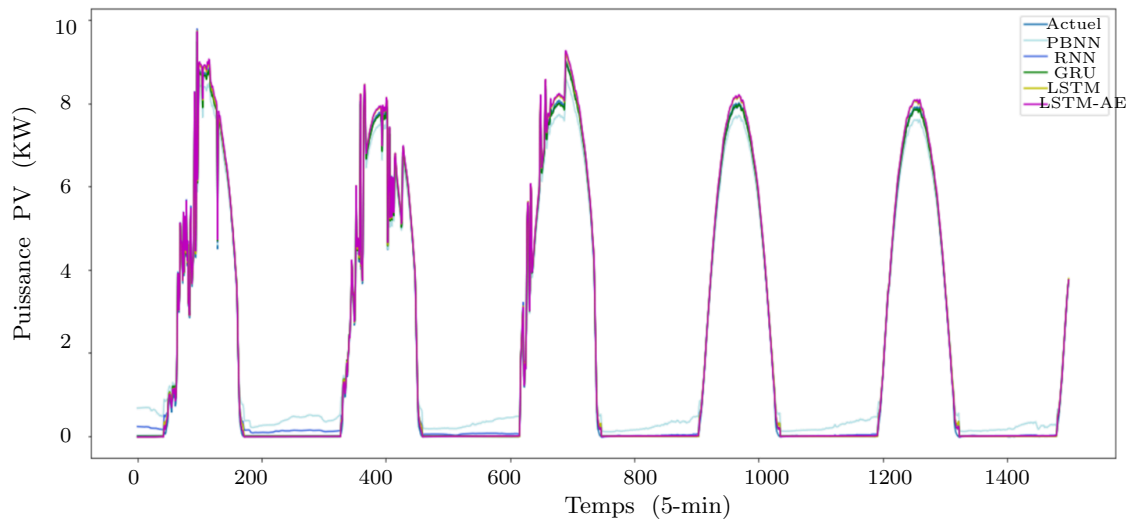


FIGURE 5.5 – Prédiction de la puissance PV en utilisant toutes les méthodes pendant la période 2.

235.63 secondes et 913.22 secondes. Le modèle LSTM-AE présente un temps d'entraînement moyen de 789.77 secondes. Cette différence s'explique principalement par le fait que le modèle LSTM est capable de modéliser les dépendances temporelles dans une séquence, et l'autoencodeur est efficace pour extraire des caractéristiques, ce qui nécessite un temps d'entraînement plus long. Le tableau.5.10 illustre les temps d'entraînement des différents modèles.

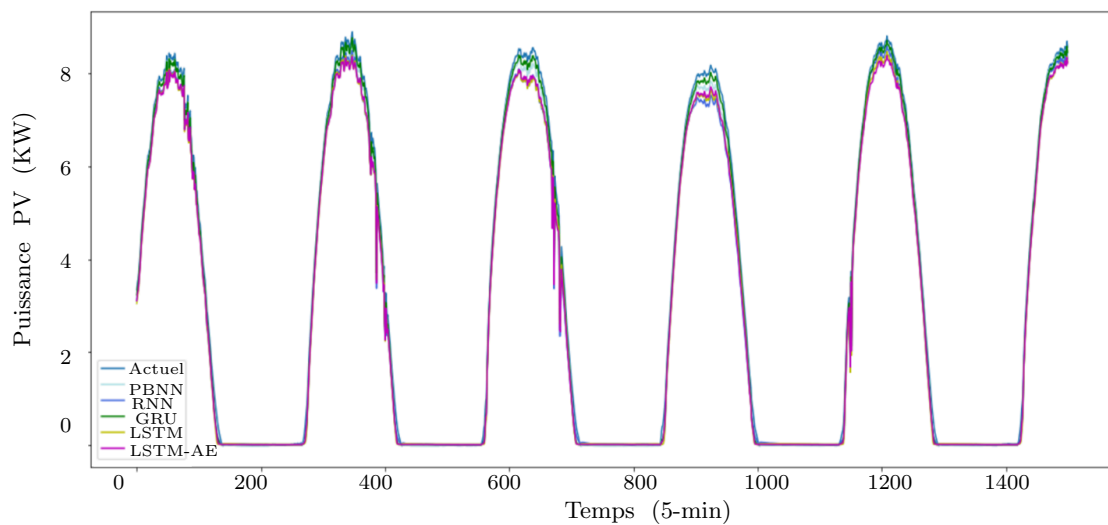


FIGURE 5.6 – Prédiction de la puissance PV en utilisant toutes les méthodes pendant la période 3.

5.3. PRÉVISION DE LA PRODUCTION D'ÉNERGIE PV À L'AIDE DE RÉSEAUX AUTOENCODEURS À MÉMOIRE À COURT TERME

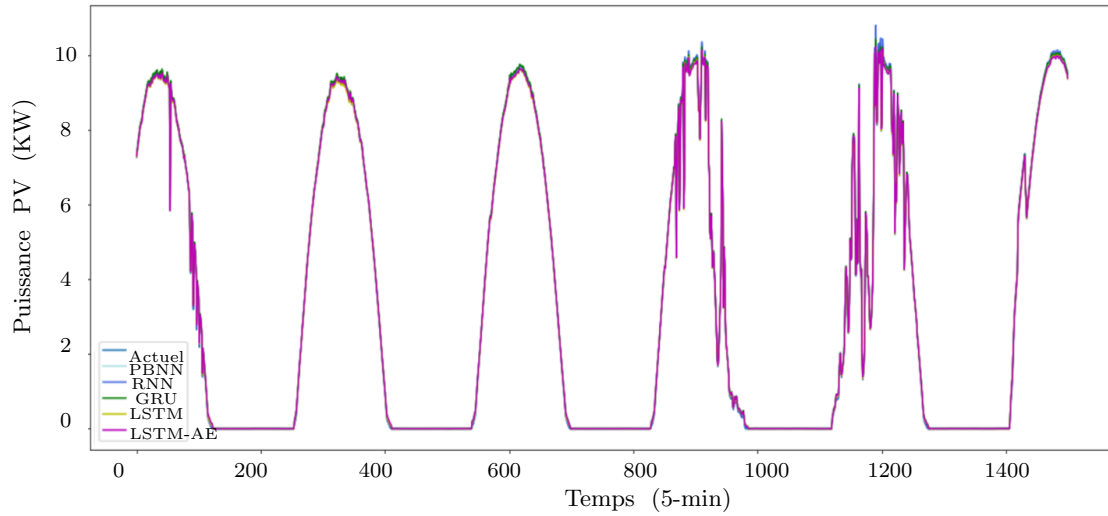


FIGURE 5.7 – Prédiction de la puissance PV en utilisant toutes les méthodes pendant la période 4.

TABLE 5.10 – Le temps d'entraînement (en secondes) des différents modèles.

Période	LSTM-AE	GRU	RNN	LSTM	BPNN
Période 1	519.18	526.30	174.55	505.75	24.21
Période 2	816.00	886.68	266.79	819.04	29.58
Période 3	1360.59	1112.44	353.15	1914.78	60.98
Période 4	463.32	450.18	148.04	413.34	20.14
Average	789.77	743.9	235.63	913.22	33.72

5.3.4 Comparaison des performances des modèles de prédiction

L'objectif de cette étude est d'évaluer les performances du modèle LSTM-AE proposé en le comparant à d'autres techniques concurrentes telles que BPNN, GRU, RNN et LSTM, ainsi qu'à deux modèles hybrides de réseaux neuronaux à apprentissage profond (GRU-RNN [128] and LSTM-RNN [129]). L'accent est mis sur la validation complète des performances du modèle LSTM-AE en utilisant des données de test saisonnières, car il est bien reconnu que les variations saisonnières ont un impact significatif sur la production d'énergie PV. Les modèles sont évalués à l'aide d'un jeu de données ayant une résolution de 5 minutes et couvrant la période allant du 28 février 2020 au 28 février 2021. Le jeu de données est divisé en quatre sous-groupes représentant les saisons en Australie : le printemps (septembre-novembre), l'été (décembre-février), l'automne (mars-mai) et l'hiver (juin-août). Chaque saison est composée de trois mois de données, les deux premiers mois étant utilisés comme ensemble

5.3. PRÉVISION DE LA PRODUCTION D'ÉNERGIE PV À L'AIDE DE RÉSEAUX AUTOENCODEURS À MÉMOIRE À COURT TERME

d'entraînement et les mois restants comme ensemble de test.

TABLE 5.11 – Réglages des paramètres des modèles hybrides de référence dans cette étude.

Modèles	Configuration
GRU-RNN	Unités GRU 64,32 ; Unités RNN 32,64 ; Unités Dense 1 ; Activation 'relu'
LSTM-RNN	Unités LSTM 64,32 ; Unités RNN 32,64 ; Unités Dense 1 ; Activation 'relu'

Les résultats détaillés de la comparaison entre les différents modèles pour chaque saison sont présentés dans le tableau 5.12. Les paramètres des modèles hybrides de référence sont indiqués dans le tableau 5.11. Les résultats présentés dans le tableau 5.12 soulignent la supériorité du modèle proposé par rapport aux autres modèles. En effet, le modèle LSTM-AE affiche les plus faibles erreurs selon les quatre indicateurs et le coefficient de détermination R^2 le plus élevé. Ces résultats démontrent la capacité du modèle proposé à modéliser de manière précise les variations saisonnières et les relations non linéaires complexes présentes dans les données de puissance PV collectées.

5.3.5 Évaluation comparative de l'approche LSTM-AE par rapport à d'autres méthodes

Dans cette section, une comparaison a été effectuée entre le modèle LSTM-AE proposé et les modèles de pointe existants pour la prédiction de la génération d'énergie solaire PV. Plusieurs chercheurs ont examiné les ensembles de données DKASC et évalué les techniques de modélisation prédictive avec différentes résolutions de données et horizons de prévision dans la littérature. Les résultats de cette comparaison sont présentés dans le tableau 5.13. Lorsque les résultats des études précédentes sont examinés, on constate que le modèle LSTM-AE proposé affiche les valeurs les plus faibles de MAE et de RMSE. Il est clairement supérieur aux études précédentes et offre de meilleures performances de prédiction de la génération d'énergie solaire PV (tableau 5.13). Les techniques d'apprentissage profond telles que GRU, RNN et LSTM se sont révélées très précieuses pour les data scientists qui traitent de grands volumes de données. Dans cette étude, les modèles LSTM et GRU surpassent les modèles RNN et BPNN en termes de précision. Le modèle LSTM, en particulier, utilise trois portes la porte d'entrée, la porte de sortie et la porte d'oubli pour conserver uniquement les informations pertinentes et significatives dans son état de mémoire interne. La porte d'oubli joue un rôle crucial en éliminant les informations moins pertinentes, ce qui améliore l'efficacité du réseau. Les expériences menées dans cette étude démontrent également l'utilité d'autres méthodes de prévision pour la prédiction de la

5.3. PRÉVISION DE LA PRODUCTION D'ÉNERGIE PV À L'AIDE DE RÉSEAUX AUTOENCODEURS À MÉMOIRE À COURT TERME

TABLE 5.12 – Performance de prévision des différentes méthodes d'apprentissage profond pour chaque saison.

Saison	Modèles	MAE	MSE	R ²	RMSE	MBE
Automne	BPNN	0.1638	0.0622	0.9895	0.2494	0.1303
	GRU	0.1210	0.0254	0.9957	0.1594	0.1163
	RNN	0.1459	0.0426	0.9928	0.2064	0.0547
	LSTM	0.0991	0.0182	0.9969	0.1351	0.0685
	GRU-RNN	0.0687	0.0100	0.9983	0.1003	0.0132
	LSTM-RNN	0.0545	0.0079	0.9986	0.0891	0.0510
	LSTM-AE	0.0495	0.0070	0.9988	0.0840	0.0366
Hiver	BPNN	0.0370	0.0025	0.9997	0.0501	-0.0356
	GRU	0.0315	0.0024	0.9997	0.0495	-0.0301
	RNN	0.0162	0.0012	0.9998	0.0353	-0.0080
	LSTM	0.0265	0.0016	0.9998	0.0403	0.0173
	GRU-RNN	0.0214	0.0012	0.9998	0.0355	-0.0166
	LSTM-RNN	0.0192	0.0015	0.9998	0.0393	0.0038
	LSTM-AE	0.0167	0.0008	0.9998	0.0283	0.0117
Printemps	BPNN	0.1421	0.0297	0.9975	0.1724	0.1418
	GRU	0.1156	0.0207	0.9983	0.1441	0.1024
	RNN	0.1206	0.0306	0.9975	0.1750	0.1184
	LSTM	0.1119	0.0267	0.9978	0.1634	0.0832
	GRU-RNN	0.1046	0.0214	0.9982	0.1465	0.0916
	LSTM-RNN	0.1032	0.0238	0.9980	0.1543	0.0995
	LSTM-AE	0.0838	0.0164	0.9986	0.1283	0.0831
Été	BPNN	0.1049	0.0208	0.9984	0.1442	0.0764
	GRU	0.0754	0.0097	0.9992	0.0989	0.0031
	RNN	0.0923	0.0195	0.9985	0.1399	0.0916
	LSTM	0.0731	0.0160	0.9987	0.1264	0.0687
	GRU-RNN	0.0709	0.0119	0.9990	0.1092	0.0327
	LSTM-RNN	0.0690	0.0116	0.9991	0.1078	0.0059
	LSTM-AE	0.0616	0.0111	0.9991	0.1053	0.0298
Moyenne	BPNN	0.1119	0.0288	0.9962	0.1540	0.0782
	GRU	0.0858	0.0140	0.9982	0.1129	0.0479
	RNN	0.0937	0.0226	0.9971	0.1391	0.0641
	LSTM	0.0776	0.0140	0.9983	0.1163	0.0615
	GRU-RNN	0.0664	0.0156	0.9988	0.0978	0.0302
	LSTM-RNN	0.0614	0.0112	0.9988	0.0976	0.0400
	LSTM-AE	0.0529	0.0088	0.9991	0.0864	0.0403

puissance PV. Cependant, afin d'améliorer la précision de la prédiction de la génération d'énergie PV, une approche LSTM-AE est proposée. Comparé aux méthodes existantes décrites dans la littérature,

le modèle LSTM-AE présente une précision de prévision supérieure.

TABLE 5.13 – Comparaison du modèle LSTM-AE proposé avec les modèles récentes.

Étude	Méthode	RMSE	MAE
Korkmaz [56]	SolarNet	0.309	0.175
Zhou et al. [125]	SDA-GA-ELM	-	0.236
Zhou et al. [69]	WPD-LSTM	-	0.242
Wang et al. [120]	CNN-LSTM	0.343	0.126
Cheng et al. [130]	GCN	0.336	0.117
Proposé	LSTM-AE	0.086	0.052

5.4 Conclusion

Les deux études présentées dans ce chapitre ont proposé des approches novatrices pour la prédiction de la production d'énergie PV en utilisant des techniques d'apprentissage profond. Elles ont toutes deux démontré l'importance de prévisions précises de la production d'énergie PV pour la gestion des systèmes de génération d'énergie solaire. La première étude a introduit un modèle hybride de prévision de la génération d'énergie PV basé sur l'algorithme d'apprentissage profond BLSTM-CNN. Ce modèle combine les capacités du réseau bidirectionnel à mémoire court-terme et du réseau neuronal convolutif pour améliorer la précision et la fiabilité des prévisions de production d'énergie PV. En exploitant les caractéristiques temporelles et spatiales des données, le modèle propose une approche complète pour la prédiction de la production d'énergie PV. Les résultats des expériences ont confirmé les performances supérieures du modèle proposé par rapport à d'autres modèles d'apprentissage profond. La deuxième étude a présenté un modèle de prévision de la génération d'énergie PV basé sur l'autoencodeur à mémoire court-terme (LSTM-AE). Ce modèle utilise l'architecture LSTM pour extraire les caractéristiques temporelles significatives des données d'entrée, ce qui permet des prévisions précises de la production d'énergie PV. Les expériences menées sur des données historiques de production d'énergie PV ont démontré l'efficacité du modèle LSTM-AE par rapport à d'autres modèles de référence. Dans l'ensemble, les deux études ont souligné l'importance de l'utilisation de techniques d'apprentissage profond pour améliorer la précision des prévisions de production d'énergie PV. Les modèles proposés ont montré des performances supérieures par rapport aux modèles simples et ont exploité les caractéristiques uniques des données d'énergie PV pour obtenir des prévisions plus précises. Ces résultats sont encourageants pour le développement de systèmes d'énergie solaire plus ef-

5.4. CONCLUSION

ficaces et fiables. Cependant, il convient de noter que les deux études ont leurs limites et peuvent être améliorées. Des recherches supplémentaires pourraient se concentrer sur l'exploration de différentes architectures de modèles, l'optimisation des paramètres du modèle et l'utilisation de données supplémentaires pour améliorer encore la précision des prévisions de production d'énergie PV. En conclusion, les deux études ont apporté des contributions significatives à la prédiction de la production d'énergie PV en utilisant des approches d'apprentissage profond. Ces travaux ouvrent la voie à de nouvelles possibilités pour l'amélioration de la gestion des systèmes d'énergie solaire et pour l'exploitation plus efficace des ressources d'énergie renouvelable.

Chapitre 6

Conclusions générales et perspectives

6.1 Conclusions générales

L'intégration de l'énergie PV offre une solution clé pour une énergie propre. Cependant, la part croissante de l'énergie PV dans les systèmes électriques, en raison de la nature aléatoire et intermittente des ressources solaires, pose des difficultés aux opérateurs de système pour la gestion des centrales PV. Afin de réduire l'impact négatif de l'utilisation de l'énergie PV, il est essentiel de prédire avec précision la production d'énergie PV. La prédiction précise de l'énergie PV revêt une grande importance pour atténuer l'incertitude et la variabilité de la production d'énergie renouvelable. Elle permet aux opérateurs de système de planifier et d'optimiser efficacement l'utilisation de l'énergie PV, améliorant ainsi la stabilité et la fiabilité des systèmes électriques. En disposant de prévisions fiables, les opérateurs peuvent prendre des décisions éclairées concernant la gestion de l'énergie PV, favorisant une meilleure intégration des énergies renouvelables dans le réseau existant. Le développement de modèles et de techniques de prédiction avancés joue un rôle crucial dans l'obtention d'une prédiction précise de l'énergie PV. Ces modèles utilisent des données historiques, des informations météorologiques et d'autres facteurs pertinents pour prévoir la production des centrales PV. En utilisant des algorithmes sophistiqués tels que l'apprentissage automatique et l'intelligence artificielle, ces modèles peuvent s'adapter et apprendre à partir des données passées afin d'améliorer progressivement la précision des prédictions. Dans le chapitre 2, l'étude approfondie se focalise sur les réseaux de neurones, un domaine essentiel de l'apprentissage automatique. Il explore divers aspects des réseaux de neurones, notamment les neurones, les fonctions d'activation, l'architecture et les différentes catégories de réseaux. L'architecture des réseaux de neurones est ensuite abordée, en mettant en évidence les différentes couches et leur organisation. Les différentes catégories de réseaux de neurones, tels que les réseaux de neurones convolutionnels et les réseaux de neurones récurrents, adaptés à des types spécifiques de données et de problèmes, sont également discutées. Les autoencodeurs, qui sont des réseaux de neurones utilisés pour la compression et la reconstruction de données, sont explorés, tout comme le LSTM-autoencodeur, une variante qui intègre des mécanismes de mémoire à long terme pour capturer des dépendances temporelles. Le chapitre explore également les métriques utilisées pour évaluer les performances des modèles de réseaux de neurones, telles que l'erreur quadratique moyenne, l'erreur absolue moyenne et d'autres métriques pertinentes. Ces métriques permettent de quantifier la précision et l'efficacité des réseaux de neurones. le chapitre 3 a exploré les principaux aspects liés à la prévision de l'énergie PV en utilisant les réseaux de neurones. Nous avons d'abord souligné l'import-

tance de l'analyse des performances pour optimiser les systèmes PVs, en identifiant les facteurs clés qui influent sur leur efficacité énergétique. Ensuite, nous avons mis en évidence l'intérêt crucial de la prévision des données de production d'énergie PV, permettant une gestion proactive des fluctuations de production. La caractérisation du site d'étude a été soulignée comme une étape essentielle pour comprendre les caractéristiques géographiques et environnementales spécifiques. Nous avons également abordé l'acquisition de données de surveillance, la sélection des données d'entrée, la présentation des données et l'analyse préliminaire, ainsi que le prétraitement des données, qui sont tous des éléments clés dans le processus de prévision. Ces axes de recherche contribuent à améliorer l'efficacité énergétique des systèmes PVs et à faciliter l'intégration de l'énergie solaire dans le réseau électrique. Dans le chapitre 4, l'objectif est de mettre en évidence l'importance croissante des approches d'apprentissage profond pour prédire avec précision la puissance PV. Trois recherches distinctes ont été menées pour évaluer différentes approches et modèles. La première recherche compare les modèles LSTM et RNN et démontre que le modèle LSTM offre une meilleure précision de prédiction que le modèle RNN. Cependant, des études futures devront confirmer ces résultats en incluant d'autres variables dans l'analyse. La deuxième recherche présente un nouveau modèle de prédiction à court terme basé sur GRU et CNN. Les résultats expérimentaux révèlent que le modèle GRU-CNN offre les taux d'erreur les plus bas par rapport aux méthodes traditionnelles. Cette approche utilise des couches de convolution pour extraire les caractéristiques des données et GRU pour stocker l'information en mémoire, ce qui conduit à des performances prometteuses pour la prédiction de la puissance PV. La troisième recherche présente l'approche CNN-GRU pour la prédiction de la puissance PV. Les résultats de simulation démontrent que ce modèle génère les plus petites erreurs de prédiction et les valeurs de R^2 les plus élevées parmi les méthodes de référence. Le modèle CNN-GRU se révèle hautement précis pour prédire la puissance PV dans différentes saisons. Dans le chapitre 5, Deux études présentées proposent des approches innovantes pour prédire la production d'énergie PV en utilisant des techniques d'apprentissage profond. Elles mettent toutes deux en évidence l'importance de prévisions précises de la production d'énergie PV pour la maintenance, le contrôle et la gestion des systèmes de génération d'énergie solaire. La première étude introduit un modèle hybride de prédiction de la génération d'énergie PV basé sur l'algorithme d'apprentissage profond BLSTM-CNN. Ce modèle combine les capacités du réseau bidirectionnel à mémoire court-terme et du réseau neuronal convolutif pour améliorer la précision et la fiabilité des prévisions de production d'énergie PV. En utilisant les caractéristiques temporelles et

spatiales des données, ce modèle propose une approche complète pour prédire la production d'énergie PV. Les résultats expérimentaux confirment les performances supérieures du modèle proposé par rapport à d'autres modèles d'apprentissage profond. La deuxième étude présente un modèle de prédiction de la génération d'énergie PV basé sur l'autoencodeur à mémoire court-terme. Ce modèle utilise l'architecture LSTM pour extraire les caractéristiques temporelles significatives des données d'entrée, permettant ainsi des prévisions précises de la production d'énergie PV. Les expériences menées sur des données historiques de production d'énergie PV ont démontré l'efficacité du modèle LSTM-AE par rapport à d'autres modèles de référence. Dans l'ensemble, ces deux études soulignent l'importance de l'utilisation des techniques d'apprentissage profond pour améliorer la précision des prévisions de production d'énergie PV. Les modèles proposés ont montré des performances supérieures par rapport aux modèles simples et ont exploité les caractéristiques uniques des données d'énergie PV pour obtenir des prévisions plus précises. Ces résultats sont encourageants pour le développement de systèmes d'énergie solaire plus efficaces et fiables. Cependant, il est important de noter que ces deux études présentent certaines limites et peuvent être améliorées. Des recherches supplémentaires pourraient se concentrer sur l'exploration de différentes architectures de modèles, l'optimisation des paramètres du modèle et l'utilisation de données supplémentaires pour améliorer encore la précision des prévisions de production d'énergie PV.

La prédiction précise de l'énergie PV profite non seulement aux opérateurs de système, mais contribue également à la durabilité globale des systèmes énergétiques. Elle facilite l'intégration efficace des sources d'énergie renouvelable, réduit la dépendance à l'égard de la production d'énergie conventionnelle et contribue à une transition vers un mix énergétique plus vert et plus durable. La prédiction précise de l'énergie PV est un élément essentiel pour surmonter les défis liés à l'intégration de l'énergie PV. En utilisant des modèles et des techniques de prédiction avancés, nous pouvons optimiser l'utilisation de l'énergie PV, améliorer la stabilité des systèmes électriques et progresser vers un avenir énergétique propre et durable.

6.2 Perspectives

Cependant, les modèles proposés présentent quelques lacunes qui doivent être examinées plus en détail. Par exemple, dans cette recherche, la structure et les hyperparamètres d'entraînement du modèle ont été déterminés par expérimentation, ce qui est chronophage. Par conséquent, des approches

d'estimation automatique des paramètres, telles que des algorithmes d'optimisation heuristique, seront utilisées dans nos futures recherches pour sélectionner et améliorer de manière plus efficace les paramètres du réseau neuronal. Les perspectives de recherche futures incluront donc :

- Utilisation d'algorithmes d'optimisation heuristique : Les approches d'estimation automatique des paramètres, comme les algorithmes génétiques ou les algorithmes d'optimisation par essaim de particules, seront explorées pour déterminer de manière plus efficace la structure et les hyperparamètres du modèle de réseau neuronal. Cela permettra de réduire le temps nécessaire à l'expérimentation manuelle et d'obtenir des résultats plus précis.
- Exploration de nouvelles architectures de réseau neuronal : En plus de l'optimisation des hyperparamètres, il sera intéressant d'explorer de nouvelles architectures de réseau neuronal pour améliorer les performances de prédiction de l'énergie PV.
- Intégration de sources de données supplémentaires : Actuellement, les modèles de prédiction de l'énergie PV sont principalement basés sur des données météorologiques et historiques de production d'énergie. Dans nos futures recherches, il serait intéressant d'explorer l'intégration de nouvelles sources de données, telles que les données de capteurs IoT ou les données d'occupation du bâtiment, afin d'améliorer la précision des prévisions et d'adapter les modèles aux spécificités du site d'étude.
- Validation et évaluation des modèles : Pour assurer la fiabilité des modèles de prédiction de l'énergie PV, il sera important de les valider et de les évaluer rigoureusement. Cela pourrait inclure des comparaisons avec d'autres méthodes de prédiction existantes, ainsi que des études de sensibilité pour évaluer l'impact des différentes variables d'entrée sur les performances du modèle.

En résumé, les perspectives de recherche future dans le domaine de la prédiction de l'énergie PV avec des réseaux de neurones incluront l'utilisation d'algorithmes d'optimisation heuristique, l'exploration de nouvelles architectures de réseau neuronal, l'intégration de sources de données supplémentaires, ainsi que la validation et l'évaluation approfondies des modèles. Ces efforts contribueront à améliorer la précision et l'efficacité des prévisions de l'énergie PV, facilitant ainsi une meilleure intégration de l'énergie solaire dans les systèmes électriques.

Publications

Publication en revue internationale

- **Sabri, M.**, El Hassouni, M. Accurate photovoltaic power prediction models based on deep convolutional neural networks and gated recurrent units, *Energy Sources, Part A : Recovery, Utilization, and Environmental Effects*, 44 :3, 6303-6320 (2022).
- **Sabri, M.**, El Hassouni, M. Predicting photovoltaic power generation using double-layer bidirectional long short-term memory-convolutional network. *Int J Energy Environ Eng* 14, 497–510 (2023).
- **Sabri, M.**, El Hassouni, M. Photovoltaic power forecasting with a long short-term memory autoencoder networks. *Soft Comput* 27, 10533–10553 (2023).

Publication en conférence internationale

- **Sabri, M.**, El Hassouni, M. A Comparative Study of LSTM and RNN for Photovoltaic Power Forecasting. In : Saidi, R., El Bhiri, B., Maleh, Y., Mosallam, A., Essaaidi, M. (eds) *Advanced Technologies for Humanity. ICATH 2021. Lecture Notes on Data Engineering and Communications Technologies*, vol 110. Springer, Cham (2022).
- **Sabri, M.**, El Hassouni, M., A Novel Deep Learning Approach for Short Term Photovoltaic Power Forecasting Based on GRU-CNN Model *E3S Web Conf.*, 336 (2022) 00064.

Bibliographie

- [1] Commission électrotechnique internationale, “Systèmes de conversion photovoltaïque de l’énergie solaire,” https://webstore.iec.ch/p-preview/info_iec61724%7Bed1.0%7Db.pdf, 2023.
- [2] DKASC Alice Springs, “1B : Trina,” <http://dkasolarcentre.com.au/locations/alice-springs?source=1B>, 2021.
- [3] M. K. Islam *et al.*, “Analysis of maximum possible utilization of solar radiation on a solar photovoltaic cell with a proposed model,” *International journal of modeling and optimization*, vol. 1, n^o. 1, p. 66, 2011.
- [4] G. Prinsloo et R. Dobson, “Solar tracking,” *Stellenbosch : SolarBoo7s. ISBN 978Y0Y620Y61576Y1*, p. 1–542, 2015.
- [5] F. Birol *et al.*, “Key world energy statistics 2017,” *International Energy Agency (IEA)*, vol. 3, 2017.
- [6] BP, “Bp statistical review of world energy 2022,” 2022.
- [7] C. Cany, “Interactions entre énergie nucléaire et énergies renouvelables variables dans la transition énergétique en france : adaptations du parc électrique vers plus de flexibilité,” Thèse de doctorat, Université Paris Saclay (COmUE), 2017.
- [8] E. Commission, “Europe leads the global clean energy transition : commission welcomes ambitious agreement on further renewable energy development in the eu,” *Accessed on February*, vol. 8, p. 2019, 2018.
- [9] E. I. Administration, *Country Analysis Executive Summary : Australia*, 2022.
- [10] T. M. C. Le, “Couplage onduleurs photovoltaïques et réseau, aspects contrôle/commande et rejet de perturbations,” Thèse de doctorat, Université de Grenoble, 2012.
- [11] O. Gergaud, “Modélisation énergétique et optimisation économique d’un système de production éolien et photovoltaïque couplé au réseau et associé à un accumulateur,” Thèse de doctorat,

- École normale supérieure de Cachan-ENS Cachan, 2002.
- [12] B. Boukezata, “Étude et commande d’une chaîne de conversion d’énergie d’un système solaire photovoltaïque,” Thèse de doctorat, Université Ferhat Abbas, 2017.
- [13] H. G. Arantzamendi, “Étude de structures d’intégration des systèmes de génération décentralisée : application aux microréseaux,” Thèse de doctorat, Institut National Polytechnique de Grenoble-INPG, 2006.
- [14] T. L. Le, “Analyses dynamiques du réseau de distribution en présence des productions décentralisées,” Thèse de doctorat, Grenoble INPG, 2008.
- [15] T. Jamal *et al.*, “Technical challenges of pv deployment into remote australian electricity networks : A review,” *Renewable and Sustainable Energy Reviews*, vol. 77, p. 1309–1325, 2017.
- [16] M. Coddington, M. Miller et J. Katz, “Grid-integrated distributed solar : Addressing challenges for operations and planning, greening the grid,” National Renewable Energy Lab.(NREL), Golden, CO (United States), Rapport technique, 2016.
- [17] M. Perez *et al.*, “Overbuilding & curtailment : The cost-effective enablers of firm pv generation,” *Solar Energy*, vol. 180, p. 412–422, 2019.
- [18] —, “Achieving 100% renewables : supply-shaping through curtailment,” *PVTECH Power*, vol. 19, p. 56–61, 2019.
- [19] D. Yang *et al.*, “History and trends in solar irradiance and pv power forecasting : A preliminary assessment and review using text mining,” *Solar Energy*, vol. 168, p. 60–101, 2018.
- [20] R. Blaga *et al.*, “A current perspective on the accuracy of incoming solar energy forecasting,” *Progress in energy and combustion science*, vol. 70, p. 119–144, 2019.
- [21] J. Antonanzas *et al.*, “Review of photovoltaic power forecasting,” *Solar energy*, vol. 136, p. 78–111, 2016.
- [22] H. Wang *et al.*, “A review of deep learning for renewable energy forecasting,” *Energy Conversion and Management*, vol. 198, p. 111799, 2019.
- [23] M. Lipperheide, J. Bosch et J. Kleissl, “Embedded nowcasting method using cloud speed persistence for a photovoltaic power plant,” *Solar Energy*, vol. 112, p. 232–238, 2015.
- [24] V. P. Lonij *et al.*, “Intra-hour forecasts of solar power production using measurements from a network of irradiance sensors,” *Solar energy*, vol. 97, p. 58–66, 2013.

- [25] N. Amral, C. Ozveren et D. King, "Short term load forecasting using multiple linear regression," dans *2007 42nd International universities power engineering conference*. IEEE, 2007, p. 1192–1198.
- [26] Y. Wang, "The analysis of the impacts of energy consumption on environment and public health in china," *Energy*, vol. 35, n^o. 11, p. 4473–4479, 2010.
- [27] J. Liu, "China's renewable energy law and policy : a critical review," *Renewable and Sustainable Energy Reviews*, vol. 99, p. 212–219, 2019.
- [28] Y. Shi *et al.*, "Expected output calculation based on inverse distance weighting and its application in anomaly detection of distributed photovoltaic power stations," *Journal of Cleaner Production*, vol. 253, p. 119965, 2020.
- [29] G. Li *et al.*, "Recurrent neural networks based photovoltaic power forecasting approach," *Energies*, vol. 12, n^o. 13, p. 2538, 2019.
- [30] E. Ogliari *et al.*, "Physical and hybrid methods comparison for the day ahead pv output power forecast," *Renewable energy*, vol. 113, p. 11–21, 2017.
- [31] M. G. De Giorgi, P. M. Congedo et M. Malvoni, "Photovoltaic power forecasting using statistical methods : impact of weather data," *IET Science, Measurement & Technology*, vol. 8, n^o. 3, p. 90–97, 2014.
- [32] C. N. Obiora *et al.*, "Forecasting hourly solar radiation using artificial intelligence techniques," *IEEE Canadian Journal of Electrical and Computer Engineering*, vol. 44, n^o. 4, p. 497–508, 2021.
- [33] A. Dolara, S. Leva et G. Manzolini, "Comparison of different physical models for pv power output prediction," *Solar energy*, vol. 119, p. 83–99, 2015.
- [34] H. Lee et B.-T. Lee, "Confidence-aware deep learning forecasting system for daily solar irradiance," *IET Renewable Power Generation*, vol. 13, n^o. 10, p. 1681–1689, 2019.
- [35] J. Alonso-Montesinos, F. Batlles et C. Portillo, "Solar irradiance forecasting at one-minute intervals for different sky conditions using sky camera images," *Energy Conversion and Management*, vol. 105, p. 1166–1177, 2015.
- [36] E. Galdi, F. Romano et E. Ricciardelli, "An advanced model for the estimation of the surface solar irradiance under all atmospheric conditions using msg/seviri data," *IEEE transactions on geoscience and remote sensing*, vol. 50, n^o. 8, p. 2934–2953, 2012.

- [37] R. Perez *et al.*, “Comparison of numerical weather prediction solar irradiance forecasts in the us, canada and europe,” *Solar Energy*, vol. 94, p. 305–326, 2013.
- [38] S. Miao *et al.*, “Markov chain model for solar farm generation and its application to generation performance evaluation,” *Journal of Cleaner Production*, vol. 186, p. 905–917, 2018.
- [39] M. J. Sanjari et H. Gooi, “Probabilistic forecast of pv power generation based on higher order markov chain,” *IEEE Transactions on Power Systems*, vol. 32, n^o. 4, p. 2942–2952, 2016.
- [40] F. Yamada *et al.*, “Prediction of next day solar power generation by gray theory and neural networks,” *TRANSACTIONS-INSTITUTE OF ELECTRICAL ENGINEERS OF JAPAN B*, vol. 134, n^o. 6, p. 494–500, 2014.
- [41] A. M. Alam *et al.*, “Solar pv power forecasting using traditional methods and machine learning techniques,” dans *2021 IEEE Kansas power and energy conference (KPEC)*. IEEE, 2021, p. 1–5.
- [42] B. Singh et D. Pozo, “A guide to solar power forecasting using arma models,” dans *2019 IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe)*. IEEE, 2019, p. 1–4.
- [43] H.-z. Wang *et al.*, “Deep learning based ensemble approach for probabilistic wind power forecasting,” *Applied energy*, vol. 188, p. 56–70, 2017.
- [44] R. De Leone, M. Pietrini et A. Giovannelli, “Photovoltaic energy production forecast using support vector regression,” *Neural Computing and Applications*, vol. 26, n^o. 8, p. 1955–1962, 2015.
- [45] J. Mubiru, “Predicting total solar irradiation values using artificial neural networks,” *Renewable Energy*, vol. 33, n^o. 10, p. 2329–2332, 2008.
- [46] K. Ghasvarian Jahromi, D. Gharavian et H. Mahdiani, “A novel method for day-ahead solar power prediction based on hidden markov model and cosine similarity,” *Soft computing*, vol. 24, p. 4991–5004, 2020.
- [47] R. M. A. Ikram *et al.*, “Application of improved version of multi verse optimizer algorithm for modeling solar radiation,” *Energy Reports*, vol. 8, p. 12 063–12 080, 2022.
- [48] A. Mellit, A. M. Pavan et V. Lughi, “Deep learning neural networks for short-term photovoltaic power forecasting,” *Renewable Energy*, vol. 172, p. 276–288, 2021.
- [49] Y. Zhou *et al.*, “Short-term photovoltaic power forecasting based on signal decomposition and machine learning optimization,” *Energy Conversion and Management*, vol. 267, p. 115944, 2022.

- [50] J. Liu *et al.*, “An improved photovoltaic power forecasting model with the assistance of aerosol index data,” *IEEE Transactions on Sustainable Energy*, vol. 6, n^o. 2, p. 434–442, 2015.
- [51] C.-J. Huang et P.-H. Kuo, “Multiple-input deep convolutional neural network model for short-term photovoltaic power forecasting,” *IEEE access*, vol. 7, p. 74 822–74 834, 2019.
- [52] I. Abouelaziz *et al.*, “Convolutional neural network for blind mesh visual quality assessment using 3d visual saliency,” dans *2018 25th IEEE International Conference on Image Processing (ICIP)*. IEEE, 2018, p. 3533–3537.
- [53] M. Sabri et M. El Hassouni, “A novel deep learning approach for short term photovoltaic power forecasting based on gru-cnn model,” dans *E3S Web of Conferences*, vol. 336. EDP Sciences, 2022, p. 00064.
- [54] D. Korkmaz, H. Acikgoz et C. Yildiz, “A novel short-term photovoltaic power forecasting approach based on deep convolutional neural network,” *International Journal of Green Energy*, vol. 18, n^o. 5, p. 525–539, 2021.
- [55] A. Agga *et al.*, “Convolutional neural network (cnn) extended architectures for photovoltaic power production forecasting,” dans *2021 9th International Conference on Smart Grid and Clean Energy Technologies (ICSGCE)*. IEEE, 2021, p. 104–108.
- [56] D. Korkmaz, “Solarnet : A hybrid reliable model based on convolutional neural network and variational mode decomposition for hourly photovoltaic power forecasting,” *Applied Energy*, vol. 300, p. 117410, 2021.
- [57] A. Alzahrani *et al.*, “Solar irradiance forecasting using deep neural networks,” *Procedia Computer Science*, vol. 114, p. 304–313, 2017.
- [58] ———, “Solar irradiance forecasting using deep recurrent neural networks,” dans *2017 IEEE 6th international conference on renewable energy research and applications (ICRERA)*. IEEE, 2017, p. 988–994.
- [59] Y. LeCun, Y. Bengio et G. Hinton, “Deep learning,” *nature*, vol. 521, n^o. 7553, p. 436–444, 2015.
- [60] M. K. Park *et al.*, “Predictive model for pv power generation using rnn (lstm),” *Journal of Mechanical Science and Technology*, vol. 35, p. 795–803, 2021.
- [61] S. Liu *et al.*, “Application of srnn-gru in photovoltaic power forecasting,” dans *E3S Web of Conferences*, vol. 256. EDP Sciences, 2021, p. 02001.

- [62] Y. Bengio, N. Boulanger-Lewandowski et R. Pascanu, “Advances in optimizing recurrent networks,” dans *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2013, p. 8624–8628.
- [63] M. Gao *et al.*, “Day-ahead power forecasting in a large-scale photovoltaic plant based on weather classification using lstm,” *Energy*, vol. 187, p. 115838, 2019.
- [64] M. Konstantinou, S. Peratikou et A. G. Charalambides, “Solar photovoltaic forecasting of power output using lstm networks,” *Atmosphere*, vol. 12, n^o. 1, p. 124, 2021.
- [65] M. S. Hossain et H. Mahmood, “Short-term photovoltaic power forecasting using an lstm neural network and synthetic weather forecast,” *Ieee Access*, vol. 8, p. 172 524–172 533, 2020.
- [66] M. Gao *et al.*, “Short-term forecasting of power production in a large-scale photovoltaic plant based on lstm,” *applied sciences*, vol. 9, n^o. 15, p. 3192, 2019.
- [67] B. Usharani, “If-lstm : Enhanced loss function in lstm to predict the sea surface temperature,” *Soft Computing*, p. 1–13, 2022.
- [68] T. Limouni *et al.*, “Accurate one step and multistep forecasting of very short-term pv power using lstm-tcn model,” *Renewable Energy*, vol. 205, p. 1010–1024, 2023.
- [69] P. Li *et al.*, “A hybrid deep learning model for short-term pv power forecasting,” *Applied Energy*, vol. 259, p. 114216, 2020.
- [70] L. Wen *et al.*, “Optimal load dispatch of community microgrid with deep learning based solar power and load forecasting,” *Energy*, vol. 171, p. 1053–1065, 2019.
- [71] K. Wang, X. Qi et H. Liu, “Photovoltaic power forecasting based lstm-convolutional network,” *Energy*, vol. 189, p. 116225, 2019.
- [72] A. Yildirim, M. Bilgili et A. Ozbek, “One-hour-ahead solar radiation forecasting by mlp, lstm, and anfis approaches,” *Meteorology and Atmospheric Physics*, vol. 135, n^o. 1, p. 10, 2023.
- [73] I. Khan *et al.*, “Hybrid power forecasting model for photovoltaic plants based on neural network with air quality index,” *International Journal of Photoenergy*, vol. 2017, 2017.
- [74] H. Zhu *et al.*, “An improved forecasting method for photovoltaic power based on adaptive bp neural network with a scrolling time window,” *Energies*, vol. 10, n^o. 10, p. 1542, 2017.
- [75] H. Alharkan, S. Habib et M. Islam, “Solar power prediction using dual stream cnn-lstm architecture,” *Sensors*, vol. 23, n^o. 2, p. 945, 2023.

- [76] Y.-L. He *et al.*, “Novel double-layer bidirectional lstm network with improved attention mechanism for predicting energy consumption,” *ISA transactions*, vol. 127, p. 350–360, 2022.
- [77] F. R. Alharbi et D. Csala, “Short-term solar irradiance forecasting model based on bidirectional long short-term memory deep learning,” dans *2021 International Conference on Electrical, Communication, and Computer Engineering (ICECCE)*. IEEE, 2021, p. 1–6.
- [78] H. Zhen *et al.*, “Photovoltaic power forecasting based on ga improved bi-lstm in microgrid without meteorological information,” *Energy*, vol. 231, p. 120908, 2021.
- [79] Y. Xue *et al.*, “Short-term photovoltaic power prediction of bi-lstm based on mic,” dans *2021 6th International Conference on Power and Renewable Energy (ICPRE)*. IEEE, 2021, p. 1107–1111.
- [80] T. Peng *et al.*, “An integrated framework of bi-directional long-short term memory (bilstm) based on sine cosine algorithm for hourly solar radiation forecasting,” *Energy*, vol. 221, p. 119887, 2021.
- [81] J. Chung *et al.*, “Empirical evaluation of gated recurrent neural networks on sequence modeling,” *arXiv preprint arXiv :1412.3555*, 2014.
- [82] Y. Wang, W. Liao et Y. Chang, “Gated recurrent unit network-based short-term photovoltaic forecasting,” *Energies*, vol. 11, n^o. 8, p. 2163, 2018.
- [83] D. Lee et K. Kim, “Pv power prediction in a peak zone using recurrent neural networks in the absence of future meteorological information,” *Renewable Energy*, vol. 173, p. 1098–1110, 2021.
- [84] N. Sodsong, K. M. Yu et W. Ouyang, “Short-term solar pv forecasting using gated recurrent unit with a cascade model,” dans *2019 International Conference on Artificial Intelligence in Information and Communication (ICAIIIC)*. IEEE, 2019, p. 292–297.
- [85] I. Goodfellow, Y. Bengio et A. Courville, *Deep learning*. MIT press, 2016.
- [86] J. Bedi et D. Toshniwal, “Energy load time-series forecast using decomposition and autoencoder integrated memory network,” *Applied Soft Computing*, vol. 93, p. 106390, 2020.
- [87] S. H. Park *et al.*, “Sequence-to-sequence prediction of vehicle trajectory via lstm encoder-decoder architecture,” dans *2018 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2018, p. 1672–1678.
- [88] A. Gensler *et al.*, “Deep learning for solar power forecasting—an approach using autoencoder and lstm neural networks,” dans *2016 IEEE international conference on systems, man, and cybernetics (SMC)*. IEEE, 2016, p. 002 858–002 865.

- [89] J. Yang *et al.*, “Lstm auto-encoder based representative scenario generation method for hybrid hydro-pv power system,” *IET Generation, Transmission & Distribution*, vol. 14, n^o. 24, p. 5935–5943, 2020.
- [90] F. Aksan *et al.*, “Multistep forecasting of power flow based on lstm autoencoder : A study case in regional grid cluster proposal,” *Energies*, vol. 16, n^o. 13, p. 5014, 2023.
- [91] A. Woyte *et al.*, “Analytical monitoring of grid-connected photovoltaic systems,” *IEA-PVPS T13-03*, vol. 2014, 2014.
- [92] S. Lindig *et al.*, “Outdoor pv system monitoring—input data quality, data imputation and filtering approaches,” *Energies*, vol. 13, n^o. 19, p. 5099, 2020.
- [93] W. van Sark *et al.*, “Pv system monitoring and characterization,” *Photovoltaic Solar Energy : From Fundamentals to Applications*, p. 553–563, 2016.
- [94] K. A. Klise et J. S. Stein, “Automated performance monitoring for pv systems using pecos,” dans *2016 IEEE 43rd Photovoltaic Specialists Conference (PVSC)*. IEEE, 2016, p. 3431–3435.
- [95] S. Killinger, N. Engerer et B. Müller, “Qcpv : A quality control algorithm for distributed photovoltaic array power output,” *Solar Energy*, vol. 143, p. 120–131, 2017.
- [96] A. Livera *et al.*, “Recent advances in failure diagnosis techniques based on performance data analysis for grid-connected photovoltaic systems,” *Renewable energy*, vol. 133, p. 126–143, 2019.
- [97] M. Dumont, “Détermination de l’albédo des surfaces enneigées par télédétection : application à la reconstruction du bilan de masse du glacier de saint sorlin,” Thèse de doctorat, Université Paris-Est, 2010.
- [98] A. ., “L’électricité solaire,” <https://www.ademe.fr/sites/default/files/assets/documents/guide-pratique-electricite-solaire.pdf>, 2021, [Online ; accessed 02-novembre-2022].
- [99] E. . Lexique, “Ombrages,” <https://www.edfenr.com/lexique/ombrages/>, 2021, [Online ; accessed 02-novembre-2022].
- [100] X. Wang *et al.*, “Experimental comparison of representation methods and distance measures for time series data,” *Data Mining and Knowledge Discovery*, vol. 26, n^o. 2, p. 275–309, 2013.
- [101] DKASC Alice Springs, “1B : Trina,” <https://dkasolarcentre.com.au/source/alice-springs/dka-m9-a-c-phases>, 2023.
- [102] Commission électrotechnique internationale, “Requirements for photovoltaic reference devices,” <https://webstore.iec.ch/publication/22266>, 2023.

- [103] H. Demirhan et Z. Renwick, “Missing value imputation for short to mid-term horizontal solar irradiance data,” *Applied Energy*, vol. 225, p. 998–1012, 2018.
- [104] C. C. Turrado *et al.*, “Missing data imputation of solar radiation data under different atmospheric conditions,” *Sensors*, vol. 14, n° 11, p. 20 382–20 399, 2014.
- [105] G. M. Yagli *et al.*, “Can we justify producing univariate machine-learning forecasts with satellite-derived solar irradiance?” *Applied Energy*, vol. 259, p. 114122, 2020.
- [106] J. Ascencio-Vásquez *et al.*, “Advanced pv performance modelling based on different levels of irradiance data accuracy,” *Energies*, vol. 13, n° 9, p. 2166, 2020.
- [107] F. Mariottini *et al.*, “Assessment and improvement of thermoelectric pyranometer measurements,” 2019.
- [108] M. Sengupta *et al.*, “Best practices handbook for the collection and use of solar resource data for solar energy applications,” National Renewable Energy Lab.(NREL), Golden, CO (United States), Rapport technique, 2021.
- [109] A. Gracia Amillo et T. Huld, “Performance comparison of different models for the estimation of global irradiance on inclined surfaces,” *Validation of the model implemented in PVGIS ; JRC Technical Report ; Publications Office of the European Union : Brussels, EUR*, vol. 26075, 2013.
- [110] C. Toledo *et al.*, “Evaluation of solar radiation transposition models for passive energy management and building integrated photovoltaics,” *Energies*, vol. 13, n° 3, p. 702, 2020.
- [111] R. Ross Jr, “Flat-plate photovoltaic array design optimization,” dans *14th Photovoltaic Specialists Conference*, 1980, p. 1126–1132.
- [112] J. A. Kratochvil, W. E. Boyson et D. L. King, “Photovoltaic array performance model.” Sandia National Laboratories (SNL), Albuquerque, NM, and Livermore, CA ..., Rapport technique, 2004.
- [113] E. Skoplaki, A. Boudouvis et J. Palyvos, “A simple correlation for the operating temperature of photovoltaic modules of arbitrary mounting,” *Solar energy materials and solar cells*, vol. 92, n° 11, p. 1393–1402, 2008.
- [114] M. Mattei *et al.*, “Calculation of the polycrystalline pv module temperature using a simple method of energy balance,” *Renewable energy*, vol. 31, n° 4, p. 553–567, 2006.
- [115] Y. Zhao *et al.*, “Artificial intelligence-based fault detection and diagnosis methods for building energy systems : Advantages, challenges and the future,” *Renewable and Sustainable Energy*

- Reviews*, vol. 109, p. 85–101, 2019.
- [116] S. Simsek, S. Uslu et H. Simsek, “Proportional impact prediction model of animal waste fat-derived biodiesel by ann and rsm technique for diesel engine,” *Energy*, vol. 239, p. 122389, 2022.
- [117] C. Jiang *et al.*, “An interpretable framework of data-driven turbulence modeling using deep neural networks,” *Physics of Fluids*, vol. 33, n^o. 5, p. 055133, 2021.
- [118] V. Gundu et S. P. Simon, “Short term solar power and temperature forecast using recurrent neural networks,” *Neural Processing Letters*, vol. 53, n^o. 6, p. 4407–4418, 2021.
- [119] H. Sharadga, S. Hajimirza et R. S. Balog, “Time series forecasting of solar power generation for large-scale photovoltaic plants,” *Renewable Energy*, vol. 150, p. 797–807, 2020.
- [120] K. Wang, X. Qi et H. Liu, “A comparison of day-ahead photovoltaic power forecasting models based on deep learning neural network,” *Applied Energy*, vol. 251, p. 113315, 2019.
- [121] N. Srivastava *et al.*, “Dropout : a simple way to prevent neural networks from overfitting,” *The journal of machine learning research*, vol. 15, n^o. 1, p. 1929–1958, 2014.
- [122] J. Wojtkiewicz *et al.*, “Hour-ahead solar irradiance forecasting using multivariate gated recurrent units,” *Energies*, vol. 12, n^o. 21, p. 4055, 2019.
- [123] M. Tovar, M. Robles et F. Rashid, “Pv power prediction, using cnn-lstm hybrid neural network model. case of study : Temixco-morelos, méxico,” *Energies*, vol. 13, n^o. 24, p. 6512, 2020.
- [124] B. Chen *et al.*, “Very-short-term power prediction for pv power plants using a simple and effective rcc-lstm model based on short term multivariate historical datasets,” *Electronics*, vol. 9, n^o. 2, p. 289, 2020.
- [125] Y. Zhou *et al.*, “Prediction of photovoltaic power output based on similar day analysis, genetic algorithm and extreme learning machine,” *Energy*, vol. 204, p. 117894, 2020.
- [126] M. Abdel-Basset *et al.*, “Pv-net : An innovative deep learning approach for efficient forecasting of short-term photovoltaic energy production,” *Journal of Cleaner Production*, vol. 303, p. 127037, 2021.
- [127] D. Niu *et al.*, “Short-term photovoltaic power generation forecasting based on random forest feature selection and ceemd : A case study,” *Applied Soft Computing*, vol. 93, p. 106389, 2020.
- [128] M. Massaoudi *et al.*, “Performance evaluation of deep recurrent neural networks architectures : Application to pv power forecasting,” dans *2019 2nd International Conference on Smart Grid and Renewable Energy (SGRE)*. IEEE, 2019, p. 1–6.

BIBLIOGRAPHIE

- [129] M. Abdel-Nasser et K. Mahmoud, “Accurate photovoltaic power forecasting models using deep lstm-rnn,” *Neural Computing and Applications*, vol. 31, n^o. 7, p. 2727–2740, 2019.
- [130] L. Cheng *et al.*, “Multi-meteorological-factor-based graph modeling for photovoltaic power forecasting,” *IEEE Transactions on Sustainable Energy*, vol. 12, n^o. 3, p. 1593–1603, 2021.

Résumé

Cette thèse de doctorat se concentre sur la prédiction précise de la production d'énergie solaire photovoltaïque (PV) et son importance pour le fonctionnement sécurisé et efficace des systèmes d'énergie PV. Dans de nombreux pays développés, l'énergie solaire PV est considérée comme la source d'énergie renouvelable la plus rentable et représente une part importante de la production d'électricité. Cependant, la génération d'énergie PV présente une variabilité élevée et imprévisible en raison de facteurs météorologiques. Afin de réduire l'influence négative de l'utilisation de l'énergie solaire PV, la prédiction précise de la production d'énergie PV revêt une importance cruciale. Les variations élevées de la production d'énergie PV peuvent être causées par des éléments météorologiques tels que l'ensoleillement, la température et les nuages. En prédisant avec précision la production d'énergie PV, il est possible d'atténuer ces variations et d'assurer un fonctionnement sécurisé et efficace des systèmes d'énergie PV. Dans cette étude, les données de production d'énergie PV ont été collectées à Alice Springs, en Australie. Ces données ont été utilisées pour développer des modèles d'apprentissage profond afin de prédire la production d'énergie PV. Les modèles d'apprentissage profond sont des outils puissants pour capturer les relations complexes entre les variables météorologiques et la production d'énergie PV. Les résultats de cette recherche ont démontré que l'utilisation de modèles d'apprentissage profond permet une prédiction précise de la production d'énergie PV. Cette prédiction précise est essentielle pour le fonctionnement sécurisé et efficace des systèmes d'énergie PV, en permettant une meilleure planification et une gestion optimisée de l'énergie produite

Mots-clefs (5) : prédiction, production d'énergie solaire photovoltaïque, efficacité, facteurs météorologiques, modèles d'apprentissage profond.

Abstract

This doctoral thesis focuses on the accurate prediction of photovoltaic solar power production and its importance for the secure and efficient operation of photovoltaic energy systems. In many developed countries, photovoltaic solar power is considered the most cost-effective renewable energy source and constitutes a significant portion of electricity production. However, the generation of photovoltaic power is characterized by high variability and unpredictability due to meteorological factors. To mitigate the negative impact of this variability, accurate prediction of photovoltaic power production is crucial. The fluctuations in photovoltaic power output can be attributed to meteorological elements such as sunlight, temperature, and clouds. By accurately predicting photovoltaic power production, it is possible to reduce these variations and ensure secure and efficient operation of photovoltaic energy systems. In this study, photovoltaic power data was collected in Alice Springs, Australia. Deep learning models were utilized to predict photovoltaic power production. Deep learning models are powerful tools for capturing the complex relationships between meteorological variables and photovoltaic power production. The findings of this research demonstrate that the use of deep learning models enables accurate prediction of photovoltaic power production. This precise prediction is essential for the secure and efficient operation of photovoltaic energy systems, enabling better planning and optimized management of generated energy.

Keywords (5) : Prediction, photovoltaic solar energy production, efficiency, meteorological factors, deep learning models