جامعة سيدي محمد بن عبد الله بفاس
UNIVERSITÉ SIDI MOHAMED BEN ABDELLAH DE FES

كلية العلوم والتقنيات فاس
FACULTÉ DES SCIENCES ET TECHNIQUES DE FÈS

N° d'ordre : 63/2022

Centre des Etudes Doctorales : Sciences et Techniques et Sciences Médicales
Formation Doctorale : Sciences de l'ingénieur, sciences physiques, mathématiques et informatique.
Discipline : Physique
Spécialité* : Génie électrique/ Energies renouvelables
Laboratoire : Signaux, Systèmes et Composants (LSSC)

** voir la liste des spécialités qui correspond à la formation doctorale sur le lien : https://fst-usmba.ac.ma/formations-doctorales/*

Thèse pour l'obtention de DOCTORAT

Développement de nouveaux modèles à base des techniques de l'intelligence artificielle pour la prévision du rayonnement solaire

Préparée par : Mme El ALANI Omaima

Soutenue le 15/10/2022 devant le jury :

Nom et Prénom	Grade	Etablissement	Qualité
Farid ABDI	PES	Faculté des Sciences et Techniques de Fès	Président
Abdelmajid JAMIL	PES	Ecole Supérieure de Technologie de Fès	Rapporteur
Aumeur EL AMRANI	PES	Université Moulay Ismail de Meknès	Rapporteur
Ismail BERRADA	PH	Université Mohamed VI Polytechnique de BenGuérir	Rapporteur
Said NAJAH	PH	Faculté des Sciences et Techniques de Fès	Examinateur
Hicham GHENNIOUI	PH	Faculté des Sciences et Techniques de Fès	Directeur de thèse

Philippe BLANC	Professeur	Mines ParisTech, Sophia Antipolis	Invité

جامعة سيدي محمد بن عبد الله بفاس
UNIVERSITÉ SIDI MOHAMED BEN ABDELLAH DE FES
كلية العلوم والتقنيات فاس
FACULTÉ DES SCIENCES ET TECHNIQUES DE FÈS

Résumé de la thèse

L'énergie solaire est la plus importante source d'énergie sur Terre et elle est en passe de devenir la source d'énergie prédominante de l'avenir. La part de marché des technologies photovoltaïques (PV) augmente rapidement grâce à la réduction de leurs coûts. Cependant, la production PV dépend fortement du climat et elle est influencée par des facteurs météorologiques tels que le rayonnement solaire, la couverture nuageuse, la géographie et elle subit souvent à des fluctuations irrégulières. Les opérateurs de réseau électrique considèrent les variations soudaines de puissance comme une menace lors de l'intégration de l'énergie solaire produite dans le réseau électrique. Avec l'augmentation continue de taux de pénétration solaire, les méthodes de prévision du rayonnement solaire deviennent de plus en plus importantes. L'information sur le rayonnement solaire et la production PV pour divers horizons temporels est ainsi cruciale pour assurer la stabilité de réseau et permettre une gestion optimale de l'énergie produite. De nombreuses techniques de prévision de rayonnement solaire pour différents horizons apparaissent chaque année dans le but d'améliorer la précision des modèles de prévision.

Nous nous intéressons dans cette thèse à la prévision du rayonnement solaire global sur plan horizontal (GHI). Ceci, dans le but de gérer l'incertitude et les fluctuations de la production PV pour une intégration réussie de l'énergie produite dans le réseau électrique. L'objet de cette thèse est alors de développer des modèles de prévision du rayonnement solaire global comme étant une méthode indirecte pour la prévision du productible PV.

La revue de littérature nous a permis dans un premier temps de conclure que le court terme est l'horizon de prévision qui constitue une priorité pour beaucoup de chercheurs. La raison de considérer cet horizon est que la plupart de l'énergie est échangée sur les marchés à court terme, ce qui est susceptible d'intéresser un gestionnaire de réseau. Par la suite, cette étude bibliographique nous a facilité la classification des modèles de prévision en fonction de la résolution temporelle et spatiale. Nous avons choisi dans cette thèse de prévoir le GHI en utilisant les techniques de l'intelligence artificielle seulement ou bien sous forme hybride avec d'autres modèles. Pour la conception de notre modèle de prévision, nous avons analysé et pris en considération tous les aspects susceptibles d'impacter la précision, tels que : l'horizon de prévision, les métriques de performances, le prétraitement et le contrôle de qualité des données, l'optimisation de l'architecture, la classification météorologique et climatique, les conditions de ciel, la variabilité, et l'étape la plus importante qui consiste la sélection des variables d'entrées pertinentes. Plusieurs paramètres météorologiques et géographiques tels que l'angle solaire zénithal, l'emplacement géographique, la durée d'ensoleillement, le rayonnement au top de l'atmosphère, la température, l'humidité relative, la vitesse et la direction du vent, la pression, les précipitations, et les aérosols ont été analysés pour choisir la meilleure combinaison d'entrée permettant de fournir la meilleure précision possible. Pour le court terme, les modèles testés sont le modèle autorégressif SARIMA et les modèles de l'intelligence à savoir le MLP (perceptron à multicouche) et LSTM (Long Short-Term Memory). Une base de données expérimentale de rayonnement solaire et d'autres variables météorologiques collectée à partir de quatre stations météorologiques installées au sol avec

différentes conditions climatiques a été utilisée. Les résultats de validation montrent que les prédictions basées sur le MLP et LSTM sont plus précises que SARIMA et la persistance même dans des conditions météorologiques compliquées. La précision des modèles proposés varie d'un site à l'autre et dépend du type de jours. Pour les jours clairs avec une faible variabilité, la meilleure précision est obtenue avec le MLP avec un RMSE de 12.97 W/m^2, alors il est de 18.58 W/m^2 pour le LSTM, de 28.88 W/m^2 pour SARIMA et de 21.15 W/m^2 pour la persistance. Pour les jours nuageux à forte variabilité, les prévisions sont moins précises. En effet, les meilleures prévisions sont générées par le MLP avec un RMSE de 60.80 W/m^2, le RMSE de LSTM est plus grand avec une valeur de 103.75 W/m^2, alors que pour SARIMA et la persistance le RMSE est très élevé, il est de 302.36 W/m^2 et 353.16 W/m^2 respectivement.
Pour la prévision à très court terme, nous avons proposé un modèle hybride fondé sur le réseau de neurones convolutifs et de perceptron à multicouche (CNN-MLP) afin de pouvoir prévoir le GHI 15 minutes à l'avance. Le modèle se base sur des images provenant d'un imageur de ciel, des séries temporelles de GHI et des variables météorologiques. L'évaluation du modèle proposé par temps clair, mixte et couvert montre que le modèle proposé est plus performant que le modèle de persistance. Le coefficient de corrélation varie entre 0.99 et 0.94 pour le MLP-CNN et entre 0.98 et 0.79 pour la persistance. Les résultats montrent que le modèle proposé pourrait être un choix approprié pour la prévision à très court terme, même dans des conditions nuageuses. Les méthodes développées dans cette thèse peuvent être exploitées par les gestionnaires de centrales photovoltaïques et les opérateurs du marché de l'électricité pour la prise de décision. Ces méthodes permettent aussi l'optimisation de l'énergie photovoltaïque produite et facilitent ainsi l'intégration à grande échelle de l'énergie photovoltaïque dans le réseau électrique.

Mots clés : rayonnement solaire, prévision, contrôle de qualité, stations météorologiques, satellites, prévision numérique de temps, intelligence artificielle, réseau de neurones, photovoltaïque, variabilité, ciel clair.

Abstract

Solar energy is the most important source of energy on earth and is becoming the predominant energy source of the future. The market share of photovoltaic (PV) technologies is rapidly increasing due to their reduced costs. However, PV production is highly dependent on climate and influenced by meteorological factors such as solar radiation, cloud cover, and geography and is often subject to irregular fluctuations. Grid operators consider sudden power fluctuations a threat when integrating solar power generation into the electrical grid. With the continuous increase of solar penetration rates, solar radiation forecasting methods are becoming increasingly important. Information on solar radiation and PV production for various time horizons is crucial to ensure grid stability and to enable optimal management of the generated power. Several solar radiation forecasting techniques and models for different time horizons are emerging each year to improve the accuracy of forecasting models.

This thesis is articulated in the framework of global horizontal irradiance forecasting (GHI). Our thread stems from the need to forecast GHI in order to manage uncertainty and fluctuations in PV generation for successful integration of the generated energy into the power grid. The thesis admits the development of global solar radiation forecasting models as an indirect method for forecasting PV production that is a topic of primary interest today as solar energy begins to play a key role in the electricity markets. Energy markets operate over different time periods and therefore specific forecasts are needed for each time horizon. Throughout this thesis, we have focused largely on short-term forecasts and have also developed a model for the very short term. Initially, the state of the art of previous studies allowed us to conclude that the short term is forecasting horizon where most research has been done. The reason for this behavior is that most of the energy is traded in short-term markets, so we judged that this horizon is likely to be of interest to a system operator. Subsequently, it allowed us to classify forecasting models based on temporal and spatial resolution. Accurate forecasting is a challenge and disparate approaches involving physical, statistical, artificial intelligence, ensemble and hybrid models have been widely studied. In this thesis, we chose to predict the GHI using artificial intelligence techniques only or in hybrid form with other models. For the design of our forecasting model, we analyzed and considered all aspects that could impact the accuracy, such as the forecast horizon, performance metrics, data quality check and preprocessing, architecture optimization, weather and climate classification, sky conditions, variability, and the most important step which is the selection of relevant input variables. Several meteorological and geographical parameters such as solar zenith angle, geographical coordinates, sunshine duration, radiation at the top of the atmosphere, temperature, relative humidity, wind speed and direction, pressure, precipitation, and aerosols were analyzed to select the best combination of inputs for the best possible accuracy. For the short term, the models tested were the SARIMA autoregressive model, the artificial intelligence models, MLP (Multilayer Perceptron), and LSTM (Long Short-Term Memory). An experimental database of solar radiation and other meteorological variables collected from five ground-based weather stations with different climatic conditions were used. Validation results show that predictions based on MLP, and LSTM are more accurate than SARIMA and persist even under complicated weather conditions. The accuracy of the proposed models varies from site to site and depends on the type of days. For clear days with low variability, the best accuracy is obtained with MLP

with an RMSE of 12.97 W/m^2, while it is 18.58 W/m^2 for LSTM, 28.88 W/m^2 for SARIMA, and 21.15 W/m^2 for persistence. For cloudy days with high variability, the forecasts are less accurate. Indeed, the best forecasts are generated by the MLP with an RMSE of 60.80 W/m^2, the RMSE of LSTM is larger with a value of 103.75 W/m^2, while for SARIMA and persistence the RMSE is very high it is 302.36 W/m^2 and 353.16 W/m^2 respectively.

For very short-term forecasting, we proposed a hybrid convolution neural network-multilayer perceptron (CNN-MLP) model to forecast GHI 15 minutes ahead. The model is based on images from a hemispheric sky imager, GHI time series, and weather variables. Evaluation of the proposed model in clear, mixed, and overcast weather shows that the proposed model outperforms the persistence model. The correlation coefficient ranges from 0.99 to 0.94 for MLP-CNN and from 0.98 to 0.79 for persistence. The results show that the proposed model could be an appropriate choice for short-term forecasting even under cloudy conditions. The methods developed in this thesis can be exploited by photovoltaic power plant managers and electricity market operators for decision-making, optimization of photovoltaic energy produced, as well as to facilitate large-scale integration of photovoltaic energy into the power grid.

Keywords: solar radiation, forecasting, quality control, weather stations, satellite, numerical weather prediction, artificial intelligence, neural networks, photovoltaic, variability, clear sky.

Remerciements

Avant tout, je voudrais louer et remercier le Bon Dieu, tout-puissant, de m'avoir accordé d'innombrables bénédictions, connaissances et opportunités, de sorte que j'ai finalement pu accomplir cette thèse.

En dehors de mes efforts, le succès de cette thèse dépend largement des encouragements et des directives de nombreuses personnes. Je saisis cette occasion pour exprimer ma gratitude aux personnes qui m'ont encouragée et soutenue pour mener à bien cette thèse. Je ne sais pas par où commencer pour exprimer ma gratitude, mais il est certain que celle-ci va à toutes les personnes formidables, compétentes et aimables avec lesquelles j'ai eu l'honneur de travailler au cours de ces quatre années de thèse. Il serait impossible d'énumérer tous les noms, mais plusieurs personnes méritent ces remerciements sincères et spéciaux.

Je tiens à exprimer ma reconnaissance et mes remerciements à mon Directeur de thèse, **Professeur Hicham Ghennioui**, qui m'a toujours encouragé à faire cette thèse et m'a fait évoluer en tant que chercheur de la meilleure façon possible. Il m'a guidé pour explorer et clarifier ma direction de recherche au début de mon travail et m'a fourni des conseils constants avec une patience infinie tout au long de mon étude de doctorat. Je tiens aussi à saluer ses qualités humaines et sa disponibilité permanente. Je ne saurais trop le remercier pour son soutien et son aide considérable. Sans ses encouragements et ses conseils, cette thèse n'aurait pas vu le jour. Vous vous souciiez tellement de mon travail et étiez toujours si prompt à répondre à mes interrogations et mes demandes. Je suis très chanceuse d'avoir été l'une de vos doctorantes.

Je tiens à exprimer ma reconnaissance pour l'Institut de recherche sur l'énergie solaire et les énergies nouvelles (IRESEN) de m'avoir accueilli au sein de sa plateforme de recherche le Green Energy Park (GEP). J'exprime mes sincères remerciements à **M. Badr Ikken**, l'ex-directeur général de l'IRESEN, **M. Zakaria Naimi**, l'ex-directeur du GEP, pour leur accueil et bienveillance. Je suis redevable à **M. Abdellatif Ghennioui**, qui fut mon encadrant au sein de GEP et avec qui j'ai eu l'opportunité de travailler sur de nombreux projets. Je le remercie de m'avoir accordé sa confiance à chaque fois et su me mettre en avant sur les projets que nous conduisions. Grâce à lui, j'ai eu l'occasion de donner plusieurs cours de formations et d'encadrer plusieurs stagiaires venant de différentes écoles et universités. J'avais aussi l'occasion de travailler sur des projets avec l'**EDF**, **DLR**, **OCP**, **JESA** et autres entités. Qu'il soit aussi remercié pour sa gentillesse, sa disponibilité permanente et pour les nombreux encouragements qu'il m'a prodiguée.

Au cours de mes années de thèse, plusieurs collaborations ont eu lieu permettant de travailler avec des personnes d'autres instituts et universités, je tiens à remercier **Prof. Philippe Blanc** et **Dr. Yves-Marie Saint-Drenan** de Mines ParisTech de Sophia Antipolis pour leur aide et leur soutien technique et aussi pour leur accueil chaleureux au sein de leur département à Mines ParisTech. J'ai beaucoup appris avec eux sur le rayonnement solaire où nos discussions étaient un réel échange et ont permis de donner vie à de nombreuses idées. Je me sentais motivée et

encouragée chaque fois que je participe à leurs réunions. Je suis fière et reconnaissante d'avoir eu la possibilité de travailler avec eux et d'apprendre d'eux.

J’exprime ma profonde gratitude et mon immense respect à Professeur, pour l’honneur qu’il m’a fait en assurant la présidence de mon jury de thèse et le regard critique **Farid Abdi** de chercheur confirmé qu’il a porté sur l’ensemble de ce travail.

J’adresse mes plus vifs remerciements aux membres du jury qui m’ont fait l’honneur d’examiner ce travail. Les Professeurs **Abdelmajid Jamil**, **Ismail Berrada** et **Ameur El Amrani** ont accepté d’être les rapporteurs de ce travail de recherche. Je leur suis sincèrement reconnaissante de l’attention portée à mon travail, ainsi que des conseils prodigués. Je remercie également Professeur **Said Najah** pour avoir accepté d’être présent à ce jury pour examiner mes travaux de thèse et enrichir les discussions.

Je voudrais aussi remercier, ma famille pour son soutien sans faille et son amour sans réserve à chaque instant. Les mots ne sauraient exprimer à quel point je suis reconnaissante. J'apprécie tous les sacrifices qu'ils ont faits pour moi. Un merci spécial à mon mari pour m’encourager durant les moments où j’ai eu du mal avec mes travaux de recherche et j’ai perdu toute motivation durant les moments difficiles que les thésards connaissent épisodiquement, et notamment pour avoir eu la patience d’attendre que je dépose mon manuscrit sans aucune plainte ni regret. C'est votre amour tous et vos encouragements illimités qui m'ont aidé à surmonter toutes les difficultés.

Enfin, je tiens à remercier mes collègues de Green Energy Park et mes collègues du laboratoire Signaux, systèmes, et Composants de la Faculté des Sciences et Techniques de Fès pour leur compagnie, leurs encouragements et leur aide précieuse dans mes travaux de recherche. Sans leur soutien, je ne pourrais pas mener à bien cette thèse.

Il serait difficile de trouver les mots adéquats pour exprimer tout ce qu'elle doit aux gens. Beaucoup d'amour et merci à vous tous.

Dédicace

Louange à dieu le Tout-Puissant de nous avoir donné la santé et la volonté d'avoir entamé et de terminer cette thèse.

Je dédie cette thèse :

À ma chère mère **Saadia**. Source inépuisable de tendresse et de sacrifice, qui n'a jamais refusé mes demandes et qui ne ménageait pas ses efforts pour me rendre heureuse. Je ne pourrais, quoi que je dise et écrive, exprimer ma grande affection et ma profonde gratitude. J'espère ne jamais vous décevoir ni trahir votre confiance et vos sacrifices.

À mon cher papa **Ahmed**, tu été et tu seras ma précieuse offre du dieu, qui doit ma vie, ma réussite et tout mon respect. De tous les pères, tu es le meilleur. Dans cette thèse vous trouverez le fruit de tous vos efforts. Aucune dédicace ne saurait exprimer mes respects, ma reconnaissance et mon profond amour.

À mon adorable petite sœur **Ikram** et mon cher frère **Islam**, vous étiez à mes côtés pendant toutes les étapes de ce travail, je vous en suis très reconnaissante. Aucune formule dédicatoire ne saurait exprimer la densité des sentiments de fraternité et d'amour, que je ressens envers vous. Je n'oublierai jamais les merveilleux moments que nous avons passés ensemble.

À mon mari **Abdelkrim**, je remercie le Bon Dieu qui a croisé nos chemins. Dès notre rencontre, tu n'as jamais cessé de m'épauler durant mes périodes de stresse et d'inquiétude. Tu as partagé avec moi les meilleurs moments de ma vie, aux moments les plus pénibles de ma vie, tu étais toujours à mon chevet, je te remercie de ne m'avoir jamais déçu et de supporter notre situation de séparation pour que je puisse finir ma thèse.

À mon **futur enfant**, dans quelques semaines, tu seras parmi nous inchaallah , puisse dieu te protéger, te procurer santé et longue vie.

À ma chère grande mère **Khadija**, *tes prières et tes Bénédictions m'ont été d'un grand secours tout au long de ma vie. Puisse Dieu vous préserver et vous procurer santé et Bonheur.*

À ma belle-mère **Nezha**, *une belle-mère comme on ne peut trouver nulle part ailleurs, vous m'avez accueilli les bras ouverts qui Dieu vous apporte bonheur et santé.*

À mes **belles-sœurs**, *je vous remercie pour vos conseils et votre soutien. Je suis très chanceuse d'avoir une famille pareille dans laquelle je me sens toujours chez moi et que vous m'avez toujours considérée comme votre petite sœur.*

À tous mes **amis**, *Il me serait difficile de vous citer tous, je vous remercie pour votre soutien.*

Tables des matières

Liste de figures ..... **xiii**

Liste de tableaux ..... **xvii**

Abréviations/Acronymes ..... **xviii**

Symboles ..... **xx**

Introduction générale ..... **xxi**

Chapitre 1 -Mesure et modélisation du rayonnement solaire ***1***

Introduction **2**

1. Transition énergétique **2**

1.1. Contexte énergétique mondial 2

1.2. Contexte énergétique national 3

1.3. Potentiel solaire au Maroc 5

2. Rayonnement solaire : éléments de base et définitions **9**

2.1. Géométrie solaire 10

2.1.1. Coordonnées géographiques 10

2.1.2. Système de coordonnées horizontales 11

2.1.3. Système de coordonnées horaires 12

2.2. Rayonnement solaire 14

2.2.1. Constante solaire 14

2.2.2. Rayonnement extraterrestre 14

2.2.3. Rayonnement solaire à la surface terrestre 15

3. Mesure du rayonnement solaire **17**

3.1. Stations météorologiques au sol 17

3.1.1. Pyranomètres 18

3.1.2. Pyrhéliomètres 19

3.2. Modèles empiriques 20

3.3. Modèles physiques 20

3.4. Modèles satellitaires 21

3.5. Rayonnement par ciel clair 21

3.5.1. Indice de clarté K_t 22

3.5.2. Indice de ciel clair k_{cs} 23

4. Variabilité du rayonnement solaire **23**

Conclusion **25**

Chapitre 2 - État de l'art sur les méthodes de prévision du rayonnement solaire ***26***

Introduction **27**

1. Facteurs influençant la précision des prévisions **27**

1.1. Horizon de prévision 28

1.2. Métriques d'évaluation de la précision des prévisions 29

2. Techniques de prévision du rayonnement solaire 30
2.1. Prévision numérique du temps 31
2.2. Imagerie satellitaire 34
2.3. Imagerie du ciel 36
2.4. Méthodes de séries temporelles 38
2.4.1. Persistance 39
2.4.2. SARIMA 39
2.5. Techniques de l'intelligence artificielle 41
2.5.1. Réseaux de neurones artificiels 43
2.5.2. MLP 45
2.5.3. LSTM 47
2.5.4. CNN 50
2.6. Modèles hybrides 53
Conclusion 56
Chapitre 3 - Sites d'étude et contrôle de qualité des données de rayonnement solaire 59
Introduction 60
1. Mesures au sol et sites d'étude 60
1.1. Sites d'étude 61
1.2. Stations météorologiques au sol 61
2. Contrôle de qualité des données 64
2.1. État de l'art sur les techniques de contrôle de qualité 65
2.2. Procédure de contrôle de qualité 68
2.3. Tests à une composante 68
2.3.1. Inspection visuelle des séries temporelles 68
2.3.2. Vérification d'horodatage des mesures 69
2.3.3. Limites physiquement possibles (PPL) 70
2.3.4. Limites extrêmement rares (ERL) 71
2.4. Tests à deux composantes 72
2.4.1. Analyse de plausibilité BSRN à deux composantes 72
2.4.2. Tests des valeurs de K NREL-SERIS 74
2.5. Test à trois composantes 76
3. Validation des bases de données 78
3.1. Détection des instants de ciel clair 79
3.2. Résultats de validation 82
3.2.1. Helioclim3 82
3.2.2. McClear 85
4. Données de prévision de GFS 87
Conclusion 88
Chapitre 4 - Résultats de prévision du rayonnement solaire à court terme et à très court terme 90
Introduction 91

1. Préparation des données et climatologie des sites d'études 91
1.1. Climatologie des sites d'étude 92
1.1.1. Variabilité saisonnière 92
1.1.2. Indice de ciel clair journalier $\boldsymbol{kcs}$ 93
1.1.3. Variabilité intra-journalière horaire $\boldsymbol{\delta(kcs)}$ 94
1.2. Classification de l'irradiance solaire 94

2. Prévision à court terme 97
2.1. Prévision du rayonnement solaire avec les réseaux de neurones 97
2.1.1. Prétraitement des données 99
2.1.2. Sélection des entrées pertinentes 101
2.1.3. Le modèle MLP 103
2.1.4. Le modèle LSTM 108
2.2. Prévision avec le modèle statistique SARIMA 109
2.3. Prévision à très court terme 111
2.3.1. Site d'étude et prétraitement de données 111
2.3.2. Architecture du modèle CNN-MLP 113

3. Résultats de prévisions 115
3.1. Prévision du GHI 12 heures à l'avance 115
3.1.1. Analyse des performances prédictives des modèles de prévision en fonction du type de jour (clarté, et variabilité) 118
3.1.2. Analyse de la performance prédictive en fonction des saisons 123
3.1.3. Impact des aérosols et de la couverture nuageuse sur la précision des prévisions du GHI 129
3.2. Prévision à très court terme 134

Conclusion 139

Conclusion générale et perspectives 141
Publications dans des journaux indexés et chapitre de livre 147
Conférences 148

Références 150

Liste de figures

Figure 1-1. Répartition de la consommation totale d'énergie finale (TFEC) par source d'énergie en 2018 et 2050 [8]. 3
Figure 1-2. Émissions de CO2 par source d'énergie au Maroc du 1990 au 2018 [1]. 4
Figure 1-3. Le mix énergétique marocain 2013 – 2030. 4
Figure 1-4. La demande énergétique marocaine en Millions TEP 2008-2030 [3]. 6
Figure 1-5. La capacité totale installée et prévue pour 2018, 2021 et 2030. 6
Figure 1-6. Le potentiel solaire au Maroc pour le GHI (gauche), et le DNI (droite). 8
Figure 1-7. Mouvement orbital de la terre autour du soleil [20]. 10
Figure 1-8. Système de coordonnées géographiques..... 11
Figure 1-9. Systèmes des coordonnées horizontales. 12
Figure 1-10. Système de coordonnées horaires. 14
Figure 1-11. La variation du G_0 en fonction du mois de l'année. 15
Figure 1-12. Composantes du rayonnement solaire résultantes des interactions avec l'atmosphère. 16
Figure 1-13. Pyranomètres : thermopile Kipp & Zonen CMP21 (droite) installé à la station météo de Benguerir, et photodiode (gauche) installé à la station météo de Zagora..... 19
Figure 1-14. Réponse spectrale du pyranomètre photodiode de LI-COR, du pyranomètre thermopile CMP21 de Kipp & Zonen et du pyrhéliomètre CHP1 tracée à côté des spectres de GHI, DNI, DHI et GTI [35]..... 19
Figure 1-15. Pyrhéliomètre Kipp & Zonen CHP1 installé à la station météorologique de Benguerir. . 20
Figure 1-16. Le $GHIcls$ horaire pour le site de Benguerir, Maroc à titre d'exemple (32.12°,-7.94°) le 21 juin et le 21 décembre. 22
Figure 1-17. Les profils journaliers au pas de 2 minutes du GHI et de la production PV mesurés au GEP, Benguerir. 24
Figure 2-1. Les méthodes de prévision en fonction de la résolution spatiale et l'horizon de prévision.31
Figure 2-2. Résolution spatiale de Météosat. Source MINES ParisTech 2009. 35
Figure 2-3. Exemples de séries temporelles de GHI au pas journalier (gauche), et au pas horaire (droite). Station météorologique de GEP, Benguerir, Maroc. 38
Figure 2-4. Procédure générale du modèle SARIMA. 40
Figure 2-5. La différence entre la programmation classique et l'apprentissage automatique..... 42
Figure 2-6. (Gauche) schéma simplifié d'un neurone biologique, (droite) neurone artificiel avec des entrées $X1, X2, \dots, Xn$ pondérées par $w1, w2, \dots, wn$ et un biais b, une fonction d'entrée u, et une fonction d'activation f permettant de former la sortie y..... 44
Figure 2-7. Représentation schématique d'un MLP avec plusieurs couches cachées..... 45
Figure 2-8. La différence entre le réseau de neurones à action directe et le réseau de neurones récurrents. 47
Figure 2-9. La cellule mémoire du LSTM..... 49
Figure 2-10. L'architecture générale d'un réseau CNN. 52
Figure 3-1. Emplacements géographiques des sites d'étude. La taille des bulles correspond au GHI annuel estimé par satellite en kWh/m^2. 61
Figure 3-2. Les quatre sites d'étude et les capteurs pyranométriques associés. 63
Figure 3-3 Séries temporelles brutes du GHI, DHI, DNI, et des variables météorologiques collectés à la station météorologique de Benguerir à titre d'exemple pour la période allant de 01/01/2015 au 30/06 /2021 au pas horaire. 64

Figure 3-4. Séries temporelles des trois composantes du rayonnement solaire pour les sites de Benguerir et Erfoud. Le bleu représente les valeurs de 10 minutes, et le cyan correspond aux moyennes journalières. 69
Figure 3-5. Représentation bidimensionnelle des données au pas de 10 minutes de GHI, DNI et DHI à Benguerir et Erfoud. Les pointillées rouges correspondent au temps de lever et de coucher du soleil, l'axe des x correspond aux jours de l'année, et l'axe des y correspond aux heures de la journée..... 70
Figure 3-6. Représentations graphiques des tests PPL et ERL pour le GHI, le DHI, et le DNI au pas de 10 minutes pour Benguerir et Erfoud. 72
Figure 3-7. Représentation graphique du test BSRN à deux composants basés sur le ratio DHI/GHI pour les sites de Benguerir et Erfoud. 74
Figure 3-8. Nuage de points des valeurs de 10 minutes dans l'espace $Kt - Kn$, l'espace $Kt - Kd$ et les limites des tests pour deux sites. 76
Figure 3-9. Représentation graphique du test à trois composantes. Le taux du GHI mesuré et calculé en fonction de l'angle solaire zénithal (droit). Des nuages de points du GHI mesuré et calculé (gauche). Les pointillés rouges indiquent les limites du test. 77
Figure 3-10. Représentation bidimensionnelle des données marquées par le test BSRN à trois composantes pour la station de Benguerir. 78
Figure 3-11. Longueur de la ligne du GHI (W/m^2) par jour clair (à gauche) et par jour nuageux (à droite), ligne bleue : GHI, ligne cyan : GHI par ciel clair, ligne magenta : longueur de ligne du GHI mesuré, ligne verte : longueur de ligne du GHI par ciel clair. 81
Figure 3-12. Représentation visuelle des résultats de détection des instants de ciel clair pour les données d'une minute de Benguerir (haut) et Zagora (bas). Le bleu représente le GHI mesuré, les pointillés rouges indiquent les minutes identifiées comme claires. 82
Figure 3-13. Nuages de points des valeurs horaires du GHI mesuré au sol (axe horizontal) et estimé par HC3 (axe vertical) pour les quatre sites d'étude sous conditions de ciel normal. 83
Figure 3-14. Nuages de points des valeurs horaires du GHI mesuré au sol (axe horizontal) et estimé par HC3 (axe vertical) pour les quatre sites d'étude sous conditions de ciel clair. 84
Figure 3-15. Nuages de points des valeurs horaires du GHI mesuré au sol (axe horizontal) et estimé par McClear (axe vertical) pour les quatre sites d'étude sous conditions de ciel clair. 87
Figure 4-1. Variation de la moyenne mensuelle du GHI journalier (kWh/m^2), tout au long de l'année pour le site de Benguerir. 93
Figure 4-2. Classification du rayonnement selon la nébulosité et la variabilité dans l'espace $kcs - \delta(kcs)$ tout a long de la période d'étude pour les sites cibles. 96
Figure 4-3. Pourcentages des jours correspondants à chaque classe pour les quatre sites. 97
Figure 4-4. Organigramme de la démarche RNA utilisé pour la prévision du GHI. 99
Figure 4-5. Répartition des données pour l'apprentissage, la validation et le test des modèles de prévision. 100
Figure 4-6. Matrice de corrélation entre le GHI et les variables candidates pour Benguerir, Erfoud, Missour, et Zagora. 102
Figure 4-7. RMSE (W/m^2) obtenu sur la base de test en fonction de nombre de neurones dans la couche cachée pour le modèle M1 105
Figure 4-8. RMSE (W/m^2) obtenu sur la base de test en fonction de nombre de neurones dans la couche cachée pour le modèle M6 105
Figure 4-9. (a) architecture du réseau de neurones avec la fonction linéaire dans la couche de sortie, et (b) courbe de régression pour l'apprentissage, la validation, et le test pour le modèle M1. 107

Figure 4-10. (a) architecture du réseau de neurones avec la fonction sigmoïde dans la couche de sortie, et (b) courbe de régression pour l'apprentissage, la validation, et le test pour le modèle M1. 107
Figure 4-11. Structure générale du modèle M1 développé pour la prévision du GHI, basé sur des variables météorologiques et des calculs de la géométrie solaire. 107
Figure 4-12. Structure générale du modèle M6 développé pour la prévision du GHI, en se basant seulement sur des calculs de la géométrie solaire. 108
Figure 4-13. Architecture du réseau récurrent LSTM adopté pour la prévision du GHI. 109
Figure 4-14. La fonction d'autocorrélation ACF du GHI horaire pour les quatre sites d'étude..... 110
Figure 4-15. Vue d'ensemble de la méthode de prévision à court terme du GHI par le modèle hybride CNN-MLP. 113
Figure 4-16. Caméra hémisphérique utilisée pour l'acquisition des images et des exemples d'images collectées à différents moments de la journée et durant des jours différents. 113
Figure 4-17. Exemples des images enlevées de la base de données lors de l'étape de prétraitement.. 113
Figure 4-18. L'architecture originale de VGG-16. 114
Figure 4-19. L'architecture du VGG-16 modifiée, où les trois dernières couches ont été éliminées. . 114
Figure 4-20. Aperçu du modèle de prévision CNN-MLP. t représente l'instant d'observation, h représente l'horizon de prévision..... 115
Figure 4-21. Résultats de prévision du GHI horaire 12 heures à l'avance avec le modèle MLP M1 pour Benguerir, Erfoud, Missour, et Zagora sur la base de test..... 116
Figure 4-22. Résultats de prévision du GHI horaire 12 heures à l'avance avec le modèle MLP M6 pour Benguerir, Erfoud, Missour, et Zagora sur la base de test..... 117
Figure 4-23. Résultats de prévision du GHI horaire 12 heures à l'avance avec le modèle LSTM pour Benguerir, Erfoud, Missour, et Zagora sur la base de test..... 117
Figure 4-24. Profils journaliers du GHI mesuré et prédit et des métriques MBE et rRMSE obtenues par les différents modèles sous les quatre types de jour. 122
Figure 4-25. Skill Score des modèles de prévisions pour les quatre sites d'étude sous les différentes conditions météorologiques. 122
Figure 4-26. Distributions par diagramme en boîte à moustaches des rRMSE obtenus pour les différents modèles de prévision pour chaque type de jour sur l'ensemble du site d'étude. 123
Figure 4-27. Erreurs saisonnières des modèles de prévision du GHI horaire 12h à l'avance 125
Figure 4-28. Nuages de point du GHI saisonnier mesuré et prédit par le MLP M1 (bleu) et LSTM (orange) à l'horizon de 12h au pas horaires pour (a) Benguerir, (b) Erfoud, (c) Missour, et (d) Zagora. 128
Figure 4-29. Transport de poussières/aérosols du Sahara vers les îles Canaries (image de MODIS [298]). 130
Figure 4-30. Photos réelles prises au site de GEP (Benguerir) lors d'une tempête de poussière les 21 et 22 avril 2018..... 131
Figure 4-31. Moyennes saisonnières des AOD_{550} pour les sites d'études sur la période de test. 133
Figure 4-32. Histogramme de distribution de la couverture nuageuse pour les quatre saisons à (a) Benguerir, (b) Erfoud), (c) Missour, et (d) Zagora sur la période de test. 134
Figure 4-33. Exemples des images prises par l'imageur du ciel installé au GEP à différents instants (clair, mixte, nuageux) pour des journées choisies à partir de la base de données. 136
Figure 4-34. Exemple du profil journalier du GHI au pas d'une minute pour un jour mixte avec les images correspondantes aux quelques instants de la journée. 136

Figure 4-35. Profils journaliers du GHI prévu à l'horizon de 15 min par rapport au modèle de persistance, pour un jour clair, mixte, et nuageux..... 138
Figure 4-36. Résultats de prévision avec le modèle CNN-MLP en comparaison avec la persistance pour l'horizon de 15 min..... 138

Liste de tableaux

Tableau 1-1. Aperçu des projets solaires au Maroc..... 8
Tableau 2-1. Exemples d'applications des prévisions solaires pour chaque horizon de prévision..... 28
Tableau 2-2. Exemples des fonctions d'activation utilisées dans les réseaux de neurones artificiels. .. 44
Tableau 3-1. Caractéristiques des sites d'étude. 61
Tableau 3-2. Principales caractéristiques des capteurs des stations météorologiques..... 62
Tableau 3-3. Résumé des tests de contrôle des données appliqués aux mesures. 68
Tableau 3-4. Résultats de validation de GHI horaire estimé à partir de HC3 pour des conditions de ciel normal et ciel clair. 84
Tableau 3-5. Résultats de validation de GHI horaire estimé à partir de McClear pour des conditions de ciel clair. 87
Tableau 4-1. Répartition des données pour les modèles de prévision et le nombre des heures disponibles pour chaque ensemble. 100
Tableau 4-2. Les variables candidates en entrée du modèle de réseau de neurones et leurs corrélations avec le GHI..... 101
Tableau 4-3. Les modèles testés pour la prévision du GHI. 102
Tableau 4-4. Meilleures architectures pour chaque modèle de prévision. 103
Tableau 4-5. Paramètres adoptés pour le modèle LSTM. 109
Tableau 4-6. Valeurs d'AIC et de BIC pour les modèles SARIMA testés. 111
Tableau 4-7 performances des modèles de prévision. 126
Tableau 4-8. RMSE et CC saisonniers obtenue pour les quatre sites d'étude. 129
Tableau 4-9. Couverture nuageuse et AOD saisonniers pour les quatre d'étude..... 132
Tableau 4-10. Comparaison des performances statistiques entre le GHI mesuré et prévu par CNN-MLP et la persistance. 139

Abréviations/Acronymes

ACF	Fonction d'autocorrélation
AERONET	Aerosol Robotic NETwork
AIC	Critères d'information d'Akaike
AM	Masse d'air
AMEE	L'agence Marocaine Pour L'efficacité Energétique
ANFIS	Système à inférence neuro-floue adaptatif
ANN	Artificial Neural Network
AOD	Profondeur optique des aerosols
AR	Autorégressive
ARIMA	Moyenne mobile autorégressive intégrée
BIC	Critère d'information bayésien
BP	Pression Barométrique
Bsh	Climat semi-aride
BSRN	Baseline Surface Radiation Network
BWh	Climat chaud et désertique
BWK	Climat désertique et froid
CAMS	Copernicus Atmosphere Monitoring Service
CIBSE	Charted Institution of Building Services Engineers
CNN	Réseau de neurones convolutif
CPV	Concentrated Photovoltaics
CSP	Concentrated Solar Power (le solaire thermique à concentration)
DE	Durée d'Ensoleillement
DHI	Rayonnement diffuse sur plan horizontal
DL	Deep Leaning
DL	Durée d'ensoleillement
DNI	Rayonnement direct normal
ECMWF	European Centre for Medium-Range Weather Forecasts
ENR	Les Energies Renouvelables
ERL	Limites Extrêmement Rares
ESRA	European Solar Radiation Atlas
ETS	Lissage exponentiel
GEP	Green Energy Park
GES	Gaz A Effet De Serre
GES	Gaz à Effet de Serre
GFS	Global Forecast System
GHI	Rayonnement global sur plan horizontal
GSBP	Green Smart Building Park
GTI	Rayonnement global sur plan incliné
HC3	Helioclim3
HC3v5	Heclioclim3 version 5
HN	Hémisphère Nord
HS	Hémisphère Sud
IA	Intelligence Artificielle
IEA	International Energy Agency
IEA	International Energy Agency
IFS	Integrated Forecasting System
IRESEN	Institut de Recherche En Energie Solaire Et Energies Nouvelles
ISO	Organisation Mondiale de Normalisation
ISS	Irradiance Solaire à la Surface
K-NN	k-plus proches voisins
LM	Algorithme de Levenberg-Marquardt utilisé pour l'apprentissage du MLP

LSTM	Long Short Term Memory
MA	Moyenne mobile
MAE	Erreur moyenne absolue
MASEN	Moroccan Agency For Sustainable Energy
MBE	Erreur de biais moyen
MLP	Perceptron à multicouche
MLR	Régression linéaire multiple
MODIS	Moderate Resolution Imaging Spectroradiometer
MSG	Météosat Seconde Génération
NAM	North American Mesoscale
NCAR	National Center for Atmospheric Research
NDFD	National Digital Forecast Database
NOAA	National Oceanic and Atmospheric Administration
NOMADS	Operational Model Archive and Distribution System
NREL	Laboratoire national des énergies renouvelables
nRMSE	Normalized Root Mean Square Error
NWP	Numérical Weather Prédiction
OMM	Organisation Météorologique Mondiale
ONEE	Office National de L'electricite et de l'Eau Potable
PHANN	Physical Hybridized Artificial Neural Network
PNT	Prévision Numérique de
POPD	Daily Probability Of Persistence
PPL	Limites Physiquement Possibles
PV	Le Photovoltaique
QC	Contrôle de qualité
R	Rayonnement Réfléchi
RD&I	Recherche, Développement, et Innovation
ReLU	Unité linéaire rectifiée
RH	Humidité relative
RMSE	Erreur quadratique moyenne
RNA	Réseau de Neurones Artificiels
RNN	Réseau de neurones réccurents
RSI	Radiomètre rotatif a bande d'ombre
SARIMA	ARIMA saisonier
SIE	Société d'Investissement dans l'Energie
SS	Performance De Prévision (Skill Score)
STD	Déviation standard
SVM	Machine à vecteurs de support
SVR	Régression par vecteur de support
SZA	Solar Zénithal Angle
TL	Turbidité de Link
TSV	Temps Solaire Vrai
TU	Temps Universel
UTC	Temps universel coordonné
VGG	Visual Geometry Group
WD	Direction du vent
WNN	Réseaux d'ondelettes
WRF	Weather Research Forecasting Model
WS	Vitesse du vent
Et	Equation de Temps

Symboles

TOA	Rayonnement global horizontal au top de l'atmosphère	W/m^2, Wh/m^2
φ	Latitude	°
λ	Longitude	°
γ	Hauteur du soleil	°
Ψ	Azimut	°
θz	Angle solaire zénithal	°
δ	Déclinaison solaire	°
ω	Angle horaire	°
G_0	Rayonnement normal au top de l'atmosphère	W/m^2, Wh/m^2
I_0	Constante solaire	W/m^2
k_{cs}	Indice de ciel clair	-
k_t	Indice de clarté	-
K_n	Fraction normale	-
K_d	Fraction diffuse	-
f_t	Porte d'oubli	-
C_t	Etat de la cellule	-
i_t	Porte d'entrée	-
o_t	Porte de sortie	-
$\tilde{C}_t$	Vecteur candidat	-
R_c	Correction de la constante solaire	°
t_{lever}	Temps de lever de soleil	-
$t_{coucher}$	Temps du coucher de soleil	-
d	Nombre de jours depuis le 1er janvier.	-
GHI_{cls}	GHI par ciel clair	W/m^2
VI	Indice de variabilité	-
R^2	Coefficient de détermination	-
$\delta(k_{cs})$	Variabilité journalière	-
cov	Covariance	
σ	La fonction sigmoïde	
N_i	Nombre d'entrée du réseau LSTM	
N_h	Nombre de couches cachées du réseau LSTM	
CC	Coefficient de Corrélation	-
tanh	Tangente hyperbolique	

Introduction générale

L'énergie est essentielle au développement économique et social des pays et toute vie sur terre dépend de l'énergie. Aujourd'hui, le marché mondial de l'énergie est dominé par les énergies fossiles. Cependant, ces combustibles fossiles en plus qu'ils sont non renouvelables et finiront par s'épuiser, ils ont des effets négatifs sur l'environnement, ce qui a motivé le développement des énergies renouvelables. Pour des considérations de protection de l'environnement et de sécurité énergétique, la part des ressources renouvelables dans l'approvisionnement énergétique mondial progresse d'une manière fulgurante. En 2050, plus que 90 % des besoins totaux en électricité seraient couverts par les énergies renouvelables. D'ici 2050, la demande d'énergie pourrait doubler, voire tripler, en raison de la croissance de la population mondiale et de l'expansion économique des pays en développement [1]. Le Maroc est parmi les pays en développement, caractérisé par un mix énergétique dominé par les énergies fossiles. Vu que le changement climatique est une priorité pour le Maroc, le gouvernement s'est fixé un plan ambitieux avec un objectif d'atteindre 52 % de sa production d'électricité à partir des énergies renouvelables à l'horizon de 2030 [2].

Le Soleil est la principale source d'énergie sur Terre. La fraction de la puissance du rayonnement issu du Soleil interceptée par la terre à chaque instant est de l'ordre de $175x10^6$ GW [3]. L'énergie solaire, en plus qu'elle est le cœur battant du système solaire, elle est la principale source d'énergie de la terre qui alimente les processus physiques, biologiques, chimiques, climatiques, météorologiques, et photosynthétiques. Ces dernières années, la technologie photovoltaïque a connu un développement rapide et devrait être la technologie la plus déployée à l'avenir pour la production d'énergie solaire. Cependant, la production électrique PV est très variable dans l'espace et dans le temps et dépends fortement du rayonnement solaire et des conditions météorologiques telles que la température, les précipitations, et surtout la couverture nuageuse. Cette variabilité soulève des défis liés à la planification, la gestion des risques, et à l'intégration de l'électricité produite dans le réseau électrique pour les acteurs du secteur de l'énergie solaire [4]. Ces préoccupations émanent de la nature variable de la ressource solaire qui dépend de plusieurs variables de la géométrie solaire, climatiques, et atmosphériques. En plus, l'énergie solaire n'est disponible que pendant la journée, ce qui la rend une source d'énergie intermittente. Ce caractère variable et intermittent de l'énergie solaire fait obstacle à son intégration à grande échelle dans le réseau électrique, ce qui nuit à l'équilibre du réseau et augmente les coûts d'exploitation. Les solutions traditionnelles consistent à utiliser des batteries de stockage d'énergie permettant de gérer la variabilité de l'énergie solaire. Toutefois, cette solution est coûteuse. En revanche, la prévision de la production photovoltaïque à différents horizons temporels est un outil performant qui permet de minimiser les conséquences négatives dues à la variabilité de la ressource solaire et assurer une large pénétration de l'énergie produite dans le mix énergétique tout en maintenant une fiabilité élevée du réseau.

La prévision de la production PV pourra se faire d'une manière directe ou indirecte. Les méthodes indirectes prédisent en premier lieu le rayonnement solaire, puis convertissent cette prédiction en production d'énergie solaire en fonction des caractéristiques de la centrale PV [5]. Les modèles numériques de prévisions et l'imagerie satellitaire sont largement utilisés par cette méthode pour analyser les paramètres météorologiques afin de prédire le rayonnement solaire. La précision des méthodes indirectes dépend fortement de la qualité et la disponibilité des variables météorologiques utilisées qui ne sont pas disponibles d'une manière précise pour les mêmes emplacements des centrales PV. En revanche, les méthodes directes prédisent directement la production PV, sans avoir besoin de prédire tout d'abord l'irradiation solaire. Cette méthode utilise des données historiques de la production PV et des données météorologiques historiques supplémentaires pour améliorer la précision des prévisions. La limitation de cette méthode est qu'elle est applicable seulement pour les centrales où des données de productible PV indispensables pour la méthode sont disponibles. Cela favorise une application plus étendue des approches indirectes, par rapport aux approches directes. La prévision du rayonnement solaire pour assurer une large pénétration du PV reste une tâche très complexe. La précision des prévisions dépend fortement des variables météorologiques, des paramètres de la géométrie solaire, de l'emplacement, etc. Jusqu'à présent, il n'existe pas une technique standard permettant le choix ciblé de la méthode de prévision à utiliser. Le choix de la méthode de prévision dépend de l'horizon temporel, qui peut aller de quelques minutes à quelques jours et qui varie en fonction de l'application de la prévision, la durée et le type de données disponibles et plusieurs autres paramètres.

Cette thèse vise à développer et valider différentes méthodes de prévision du GHI à différents horizons temporels. Ces méthodes peuvent permettre aux opérateurs et gestionnaires de la centrale PV et du marché de l'électricité la prise des décisions. En particulier, nous nous concentrons sur la prédiction indirecte à travers la prévision du GHI, la prévision de la puissance PV ne sera pas entamée. Dans cette thèse, la tâche de prévision a été abordée en passant par plusieurs étapes dont le contrôle de qualité des données, la validation des bases de données satellitaires par des mesures aux sols à hautes précisions, la modélisation du rayonnement par ciel clair, l'étude de la climatologie par rapport à la couverture nuageuse, l'impact des différents paramètres de l'atmosphère claire sur le rayonnement, la conception et la modification des modèles de prévisions permettant d'améliorer la précision sont également étudiés. Tout au long de cette thèse, différentes sources de données ont été exploitées. Par exemple des données historiques de GHI, des variables météorologiques, des bases de données satellitaires, les prévisions des modèles numériques de prévisions, des modèles physiques de rayonnement par ciel clair, et des images hémisphériques. Les sources de données utilisées dans les différentes études de cas sont décrites dans les chapitres correspondants. Le GHI obtenu depuis différentes sources est la préoccupation majeure de cette étude.

Les aspects suivants sont mis en évidence dans cette thèse:

- **Collecte et contrôle de qualité des mesures au sol**: l'irradiance solaire est considérée comme l'une des variables climatiques les plus cruciales. Des données d'irradiation solaire

de haute qualité sont nécessaires pour assurer une meilleure précision de prédiction. Les mesures du rayonnement solaire directement depuis les stations au sol à long terme sont considérées comme la source la plus précise du rayonnement solaire. Cependant avec ces mesures au sol, il se peut y avoir toujours des incertitudes dues aux problèmes techniques des instruments, des erreurs opérationnelles, et des erreurs d'installation. Ces problèmes peuvent provoquer des erreurs systématiques et aléatoires dans les données. Différentes approches ont été proposées ces dernières années pour détecter efficacement les comportements anormaux, notamment lorsque deux ou trois composantes du rayonnement solaire sont mesurées. Si certaines approches proposent une représentation visuelle de la routine, d'autres approches ne produisent qu'une série temporelle de valeurs de drapeaux binaires et ne proposent aucune visualisation spécifique des données marquées par rapport aux données non marquées. Nos contributions à ce sens sont doubles. Premièrement, nous avons proposé une procédure de contrôle de qualité complète qui intègre les méthodes les plus pertinentes de l'état de l'art ainsi que la vérification issue des bonnes pratiques de l'ingénierie solaire (décalage temporel, détection des ombres, etc.). Deuxièmement, nous recommandons un support visuel à tous les contrôles proposés dans la littérature comme un moyen d'améliorer le taux de détection des données anormales ainsi qu'une première étape dans le diagnostic des causes d'erreur.

- **Validation des modèles satellitaires:** les mesures directes du rayonnement solaire disponibles à partir de radiomètres au sol ne sont généralement pas suffisantes. Lorsque les stations terrestres de rayonnement solaire sont clairsemées, il existe certaines méthodes pour obtenir des estimations du rayonnement solaire pour les endroits où il n'est pas directement mesuré. Les techniques satellitaires et les modèles numériques de prévision sont devenus la méthode la plus courante pour la prévision des ressources solaires à l'échelle régionale, continentale, ou mondiale. Avant d'utiliser ces données, il faut les valider par des données au sol. Dans ce sens, nous avons validé la base de données satellitaire HelioClim3 et le modèle de ciel clair McClear par rapport aux données mesurées disponibles.
- **Prévision du rayonnement solaire global**: l'objectif principal est de comparer différentes techniques de prévision du rayonnement solaire afin de développer un modèle capable de répondre aux différentes limitations des modèles existants. À cette fin, différentes techniques d'apprentissage automatique et de séries chronologiques ont été utilisées pour apprendre de manière dynamique la relation entre les différentes conditions météorologiques et le rayonnement solaire. Deux horizons de prévision ont été entamés : pour la prévision à court terme, trois techniques ont été utilisées, à savoir SARIMA (Moyenne mobile intégrée autorégressive saisonnière), MLP (perceptron à multicouche) et LSTM (Long Short-Term Memory). Des prévisions à partir d'un modèle numérique de prévision gratuitement accessibles ont été utilisées dans la phase de prédiction pour faciliter la reproductivité des résultats par les scientifiques sans avoir besoin à des données de hautes performances. Pour le très court terme, nous avons développé un modèle

hybride CNN-MLP qui se base sur le traitement des images d'une caméra hémisphérique afin d'extraire l'information sur la couverture nuageuse.

- **Analyse de l'impact des nuages et des aérosols sur le rayonnement solaire**: le potentiel solaire dépend de plusieurs paramètres atmosphériques. Les nuages sont considérés comme le facteur le plus important de réduction du rayonnement solaire. Cependant, en l'absence de nuages, les aérosols sont la principale source d'atténuation du rayonnement solaire. Dans un premier temps, nous avons quantifié la variabilité du rayonnement solaire global causée par les nuages, dans ce sens nous avons établi un schéma de classification des jours en fonction de la couverture nuageuse. Par la suite, nous avons examiné l'impact de l'AOD (Profondeur optique des aérosols) sur la prévision de GHI surtout dans les conditions de ciel clair. Pour faire, nous avons exploité les données AOD de MODIS Terra version 3.

Ce manuscrit est structuré en quatre chapitres :

- **Le chapitre 1** présente les différentes équations de la géométrie solaire décrivant le comportement du rayonnement solaire à un lieu donné, et la propagation du rayonnement solaire lors de sa traversée de l'atmosphère et les différentes composantes résultantes. Nous présentons également une description détaillée des stations météorologiques au sol pour la mesure des différentes composantes de rayonnement solaire ainsi que les modèles d'estimation du GHI. Par la suite, nous présentons une introduction du rayonnement par ciel clair vue sans importance pour la prévision du rayonnement solaire. Le chapitre finira par une discussion de la variabilité du rayonnement global causée par les nuages.
- **Le chapitre 2** présente l'état de l'art des travaux de recherche antérieurs sur la prévision du rayonnement solaire. Les sections de chapitre passent en revue les techniques de prévision de GHI basées sur les modèles numériques de prévision de temps, l'imagerie satellitaire, l'imagerie de ciel, les méthodes statistiques autorégressives et les méthodes d'apprentissages automatiques et finalement les modèles hybrides.
- **Le chapitre 3** présente les résultats des étapes qui précèdent la prévision. Dans une première partie, le chapitre présente un état de l'art sur les techniques de contrôle de qualité du rayonnement solaire et de la procédure complète que nous proposons pour le contrôle de qualité des composantes de rayonnement solaire mesurées au sol. Par la suite, les résultats de l'application des tests de contrôle de qualité pour les données de rayonnement solaire collectées sur les différentes stations aux sols disponibles sont présentés. Ces données seront utilisées pour la validation des bases de données HelcioClim3 et le modèle de ciel clair McClear pour des conditions tout ciel et des conditions de ciel clair après avoir appliqué des algorithmes de détection de ciel clair à partir des mesures de GHI.
- **Le chapitre 4** présente dans un premier temps une étude de la variabilité du rayonnement solaire par rapport à la couverture nuageuse permettant une classification du GHI par rapport à la variabilité et la clarté. Par la suite, les méthodologies utilisées pour la prévision à court terme et à très court terme sont discutées. Finalement, les modèles de

prévision sont évalués pour différentes saisons et diverses conditions météorologiques et leurs performances sont comparées aux modèles de références. Quatre sites avec des caractéristiques géographiques et climatiques différentes sont utilisés pour la validation des modèles de prévision.

- Enfin, la partie conclusions générale et perspectives présente un résumé et une critique des contributions globales de la thèse, en soulignant la pertinence pratique des méthodes présentées. Les perspectives de développement futur sont également présentées.

Chapitre 1

Mesure et modélisation du rayonnement solaire

Introduction **2**

1. Transition énergétique **2**

1.1. Contexte énergétique mondiale 2

1.2. Contexte énergétique national 3

1.3. Potentiel solaire au Maroc 5

2. Rayonnement solaire : éléments de base et définitions **9**

2.1. Géométrie solaire 10

2.1.1. Coordonnées géographiques 10

2.1.2. Système de coordonnées horizontales 11

Hauteur du soleil 11

Angle solaire zénithal 11

Azimut 12

Temps de lever et de coucher de soleil 12

2.1.3. Système de coordonnées horaires 12

Déclinaison (δ) 13

Angle horaire (ω) 13

2.2. Rayonnement solaire 14

2.2.1. Constance solaire 14

2.2.2. Rayonnement extraterrestre 14

2.2.3. Rayonnement solaire à la surface terrestre 15

3. Mesure du rayonnement solaire **17**

3.1. Stations météorologiques au sol 17

3.1.1. Pyranomètres 18

3.1.2. Pyrhéliomètres 19

3.2. Modèles empiriques 20

3.3. Modèles physiques 20

3.4. Modèles satellitaires 21

3.5. Rayonnement par ciel clair 21

3.5.1. Indice de clarté K_t 22

3.5.2. Indice de ciel clair k_{cs} 23

4. Variabilité du rayonnement solaire **23**

Conclusion **25**

Introduction

Aujourd'hui, l'énergie est l'un des principaux besoins de la société qui est au cours de tous les secteurs : l'industrie, le transport, le commerce, l'agriculture, la santé, le résidentiel, etc. L'énergie, l'environnement, et le développement économique d'un pays sont étroitement liés. Un approvisionnement énergétique fiable et suffisant est nécessaire pour répondre aux différents besoins énergétiques et augmenter la productivité. Cependant, les sources d'énergie fossile, telles que le pétrole, le gaz et le charbon, en plus qu'elles seront bientôt épuisées et non durables pour le développement futur, elles présentent de graves menaces pour l'environnement. Environ deux tiers des émissions mondiales de dioxyde de carbone proviennent de ces sources de combustibles, dont la part actuelle de la production d'énergie. Le défi de répondre au changement climatique est un problème environnemental mondial qui aura un impact sur tous les pays. Selon l'organisation météorologique mondiale (OMM), l'année 2019 a été marquée par une élévation de température extrême, un recul des glaciers et d'une élévation sans précédente du niveau des mers à cause des émissions de gaz effet de serre [6].

En effet, la croissance économique durable d'un pays n'est possible que par une utilisation bien planifiée et efficace des ressources naturelles disponibles localement, comme l'énergie solaire, l'énergie éolienne, l'énergie hydraulique et la biomasse. Cela améliore la qualité de vie de chacun sur la Terre. Pour des raisons de protection de l'environnement et de sécurité énergétique, la part des ressources renouvelables dans l'approvisionnement énergétique mondial augmente désormais à un rythme impressionnant. L'agence internationale de l'énergie (IEA) s'est fixé pour objectif d'atteindre une part de 90 % d'énergie provenant de sources renouvelables d'ici 2050 [1]. Le Maroc aussi a lancé un plan visant à augmenter la part des énergies renouvelables à 52% à l'horizon de 2030 dans le mix énergétique [7].

Dans la première partie de ce chapitre, nous résumons la stratégie énergétique mondiale/nationale et le potentiel solaire marocain. Par la suite, ce chapitre introduit les différentes notions sur le rayonnement solaire, en partant du rayonnement solaire au top de l'atmosphère jusqu'à son arrivée au sol terrestre, et les différentes composantes résultantes lors de sa traversée de l'atmosphère. Nous présentons ensuite les différents capteurs utilisés pour la mesure des différentes composantes du rayonnement solaire, ainsi que les différents modèles d'estimation du rayonnement au sol. Nous clôturons par une discussion sur la variabilité du rayonnement solaire et ses causes principales.

1. Transition énergétique

1.1. Contexte énergétique mondial

Selon l'IEA, la pandémie de Covid-19 a provoqué plus de perturbations dans le secteur de l'énergie que tout autre événement de l'histoire récente, laissant des impacts qui se feront sentir pendant des années. La demande mondiale d'énergie devrait augmenter de 4.6 % en 2021, ce qui devrait compenser la contraction de 4 % en 2020 et augmenter la demande de 0.5 % par

rapport 2019 [8]. La pandémie a durement affecté la demande de pétrole et de charbon, tandis que les énergies renouvelables ont augmenté et les émissions mondiales de CO2 ont diminué de 5.8 % en 2020, soit près de 2 Gt CO2 qui est la plus importante diminution jamais observée. En 2021, les émissions mondiales seraient inférieures d'environ 400 Mt de CO2, soit 1.2 %, au pic de 2019 [9].

En 2050, l'électricité serait le principal vecteur énergétique avec plus de 50 % de la consommation totale d'énergie finale, contre 21 % aujourd'hui. La Figure 1-1 représente la répartition de la consommation d'énergie en 2018 et 2050 [10].

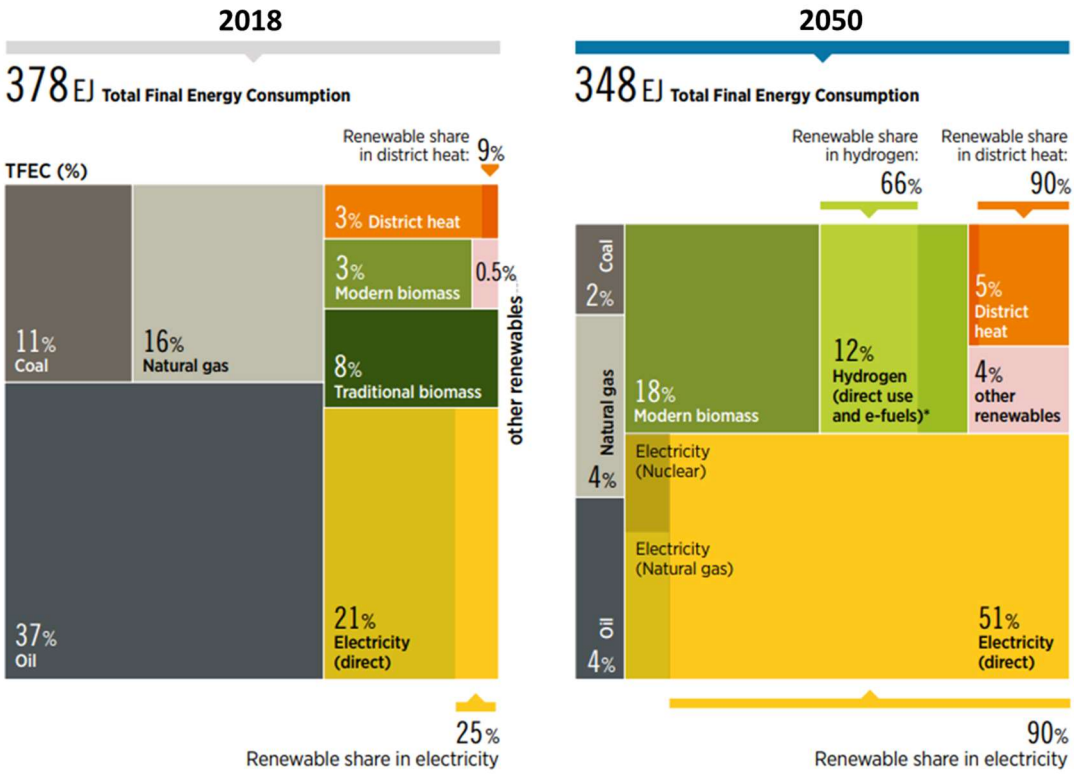


Figure 1-1. Répartition de la consommation totale d'énergie finale (TFEC) par source d'énergie en 2018 et 2050 [10].

Nous constatons qu'en 2050, plus que 90 % des besoins totaux en électricité seraient couverts par les énergies renouvelables par l'approvisionnement direct, l'électrification, l'efficacité énergétique et l'hydrogène qui atteindraient 66 % en 2050. Toutefois, les sources d'origines fossiles auront un rôle limé de 6 % par le gaz naturel et le reste par le nucléaire.

1.2. Contexte énergétique national

Le Maroc est le seul pays d'Afrique du Nord qui ne possède pas de réserves de pétrole et de gaz commercialement pertinentes. Par conséquent, le profil énergétique marocain a longtemps été dominé par l'importation de ses fournitures en ressources énergétiques. Le mix énergétique marocain est composé de plus que 90 % de combustible fossile de l'approvisionnement total en énergie primaire et de 80% de l'approvisionnement en électricité [11].

La consommation des énergies fossiles est la principale cause d'émissions des gaz à effet de serre (GES). Selon l'IEA, les émissions totales de CO2 provenant de l'énergie fossile au cours de 2006-2016 atteignaient 55.3 MTCO2, et de 181 % depuis 1990. En 2018, le pétrole était responsable de 63 % des émissions totale, le charbon de 34% et 3% du gaz naturel (Figure. 1-2).

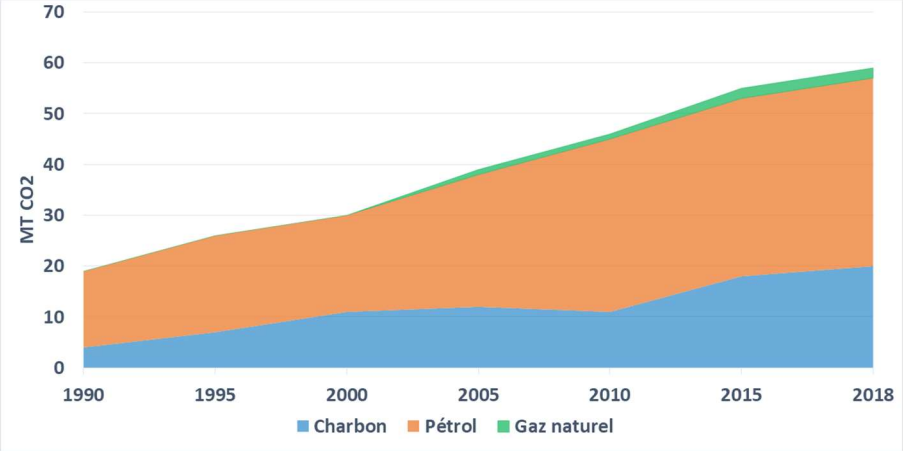


Figure 1-2. Émissions de CO2 par source d'énergie au Maroc du 1990 au 2018 [1].

Le changement climatique est une priorité pour le Maroc. A cet effet, le gouvernement marocain a été l'hôte de la COP7 en 2001, où l'accord de Marrakech a été conclu, et de la COP22 en 2016 à Marrakech, où l'accord de Paris a été ratifié. Dernièrement, le Maroc a agi comme un puissant défenseur de l'Afrique lors de la COP23 à Bonn et de la COP24 à Katowice. Le Maroc a lancé un plan visant à augmenter la part des énergies renouvelables dans le mix énergétique. Le pays vise à augmenter sa production électrique à partir du renouvelable de 52% à l'horizon de 2030 au lieu de 37 % en 2021 [12,13].

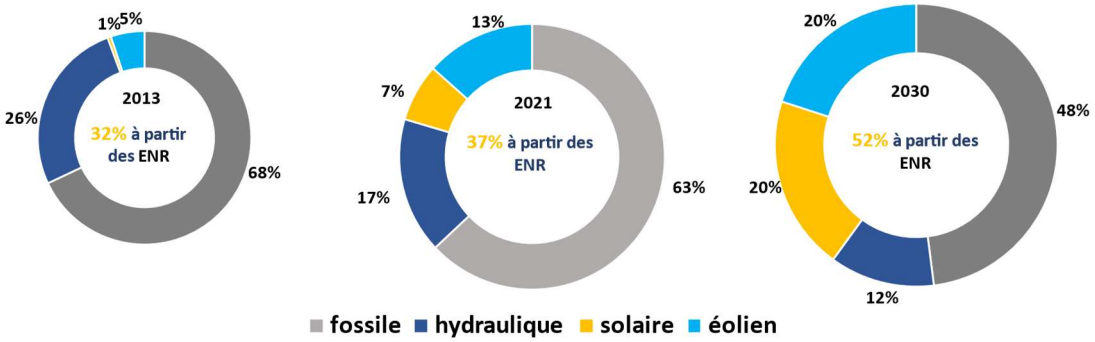


Figure 1-3. Le mix énergétique marocain 2013 – 2030.

En 2009, le Maroc a adopté une stratégie énergétique nationale comme feuille de route pour relever le défi de la sécurité énergétique et du développement durable pour une transition vers un système énergétique à faible émission de carbone. La stratégie énergétique a été axée sur cinq piliers [14] :

- Le développement d'un bouquet énergétique diversifié et optimisé.

- La mobilisation des ressources domestiques, y compris l'utilisation des énergies renouvelables (éolien/solaire).
- Faire de l'efficacité énergétique une priorité nationale.
- Le renforcement de la coopération régionale avec les marchés énergétiques européens et africains.
- L'intégration industrielle, qui favorise le développement des capacités industrielles locales à tous les niveaux de la chaîne de valeur des technologies vertes.

Depuis lors, plusieurs composantes de la stratégie énergétique ont été mises à jour :

- D'ici 2030, le Maroc s'est engagé à réduire les émissions de gaz à effet de serre à l'échelle de l'économie de 17 % par rapport au niveau de référence, et à les réduire de 25 % supplémentaires, sous réserve d'un soutien international.
- Atteindre 52 % de la capacité électrique installée à partir de sources renouvelables d'ici 2030 (Figure 1-3).
- La mise en œuvre d'un nouveau plan d'efficacité énergétique dans l'ensemble de l'économie, notamment dans l'industrie, les transports et les bâtiments. L'Agence Marocaine pour l'Efficacité Energétique (AMEE) a identifié un potentiel d'économies d'énergie de 25% d'ici 2030 [15].
- Des investissements importants dans le domaine de la recherche, du développement et de l'innovation (RD&I) en matière de technologies énergétiques.
- L'augmentation de la part du gaz naturel avec l'installation d'une capacité supplémentaire de 2400 MW de technologie à cycle combiné d'ici 2030.

1.3. Potentiel solaire au Maroc

La production d'électricité au Maroc a augmenté de 65% en une décennie en raison de la croissance démographique, du développement économique et des efforts du gouvernement pour assurer un accès global à l'électricité. La demande d'énergie primaire de l'électricité devrait doubler d'ici 2030. La Figure 1-4 montre l'évolution de la demande énergétique pour le Maroc [12]. L'augmentation et la diversification de la capacité électrique sont une priorité pour le Maroc, conformément à son objectif d'avoir 52% d'énergies renouvelables dans la capacité totale installée en 2030 pour un progrès vers une électricité propre et durable. Actuellement, le bouquet énergétique marocain est assez diversifié, avec une contribution remarquable des sources renouvelables. Comme montré par la Figure 1-5, en 2018 le Maroc a installé une capacité de 3700 MW (1170 MW de l'hydraulique, 1220 MW de l'éolien, et 710.8 MW du solaire), en 2021 la capacité installée a atteint 3950 MW (1770 MW de l'hydraulique, 1430 MW de l'éolien, et 750 MW de solaire), et cette puissance installée devrait atteindre 12 900 MW à l'horizon de 2030, répartie comme suit : 4800 MW de l'hydraulique, 5000 MW de l'éolien, et 3100 MW du solaire [16].

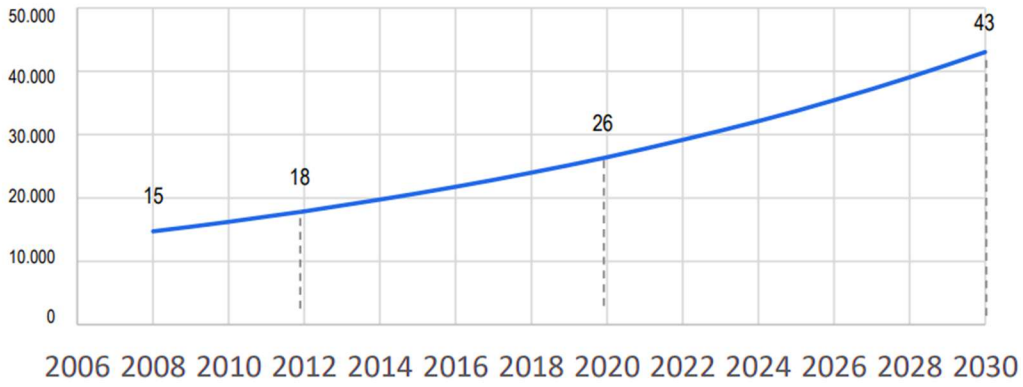


Figure 1-4. La demande énergétique marocaine en Millions TEP 2008-2030 [3].

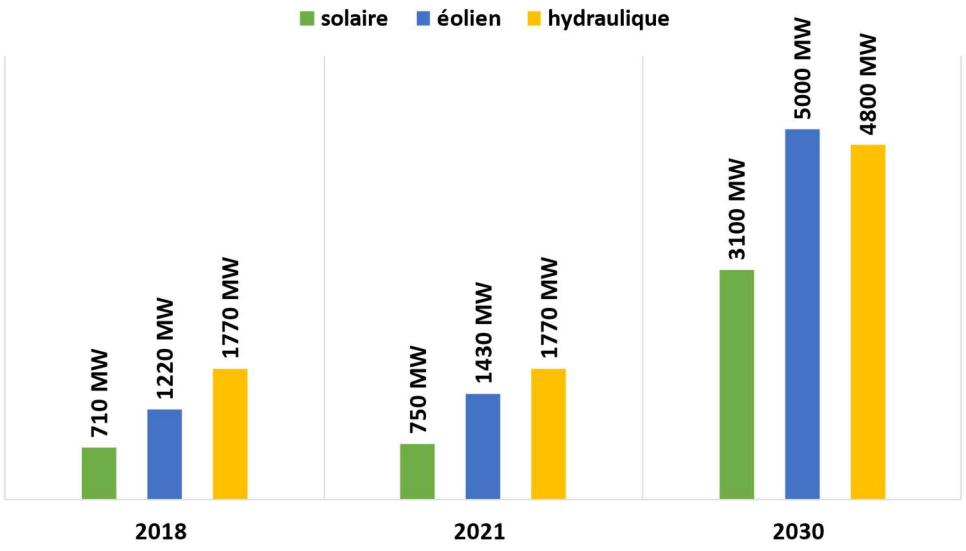


Figure 1-5. La capacité totale installée et prévue pour 2018, 2021 et 2030.

Le Royaume du Maroc est un pays situé au nord-ouest de l'Afrique géographiquement et climatiquement favorisé pour développer à grande échelle l'exploitation des énergies renouvelables. L'énergie solaire est considérée comme une alternative renouvelable prometteuse pour le Maroc. Le pays présent un fort potentiel d'ensoleillement avec une moyenne annuelle de rayonnement global d'environ 5.3 kWh/m^2/an et une durée d'insolation annuelle ente 2700 heures au Nord et arrivant jusqu'à 3500 heures au Sud [14]. Afin de pouvoir exploiter cette énergie, lors de la COP 21 le Maroc prévoyait l'augmentation de la capacité des énergies renouvelables à 52%, pour atteindre son objectif, le pays vise à ajouter une capacité des ENR de 10 GW entre 2018 et 2030, dont 3100 MW à partir du solaire. Le Tableau 1-1 énumère les principaux projets achevés, en cours ou prévus jusqu'à 2030 [7,12,17]. La Figure 1-6 représente le potentiel solaire au Maroc pour le GHI et le DNI, les cartes ont été générées sur la base des données satellitaires.

Dans le but de développer le fort potentiel national d'énergie solaire, L'Office National de l'Électricité et de l'Eau potable (ONEE) a lancé un programme permettant de déployer plusieurs

centrales solaires photovoltaïques, NOOR Tafilalet de 120 MW où la mise en service été prévue en 2021, NOOR Atlas pour le développement une centrale solaire photovoltaïque avec une capacité installée de 200 WM pour l'horizon 2024, et NOOR II d'une capacité de 400 & NOOR Boujdour II d'une puissance de 350 MW dont la mise en service est prévue en 2023.

Les perspectives pour le développement des énergies renouvelables sont prometteuses, et le pays représente un terrain fertile pour les investissements internationaux dans les énergies renouvelables. Pour promouvoir les ENR, le pays a mis en place plusieurs lois :

- La loi 13-09 qui concerne la production d'électricité d'origine renouvelable, elle permet : l'ouverture de la production à la concurrence, encourager les initiatives privées, l'accès au réseau électrique, l'exportation d'électricité verte et la construction d'une ligne directe pour l'exportation.
- La loi 57-09 votée en janvier 2010 qui vise la création de l'agence marocaine pour l'énergie durable (MASEN). Cette agence a pour mission l'étude et la conception de projets solaires. MASEN doit également se charger de la promotion, du placement, du financement et de la réalisation des projets solaires, contribuer au développement de l'expertise, de la recherche de l'industrie solaire.
- La loi n° 37-16 complétant et modifiant la loi 57-09, vise notamment à étendre les missions de MASEN à la construction de centrales électriques pour toutes les énergies renouvelables actuelles et les futurs projets, à l'exception des projets STEP, afin d'assurer la stabilité du réseau électrique et des projets développés dans le cadre de la loi 13-09.
- La loi 47-09 relative à l'efficacité énergétique a pour objet d'augmenter l'efficacité énergétique dans l'utilisation des sources d'énergie, d'éviter le gaspillage, d'atténuer le fardeau du coût de l'énergie sur l'économie nationale et de contribuer au développement durable.

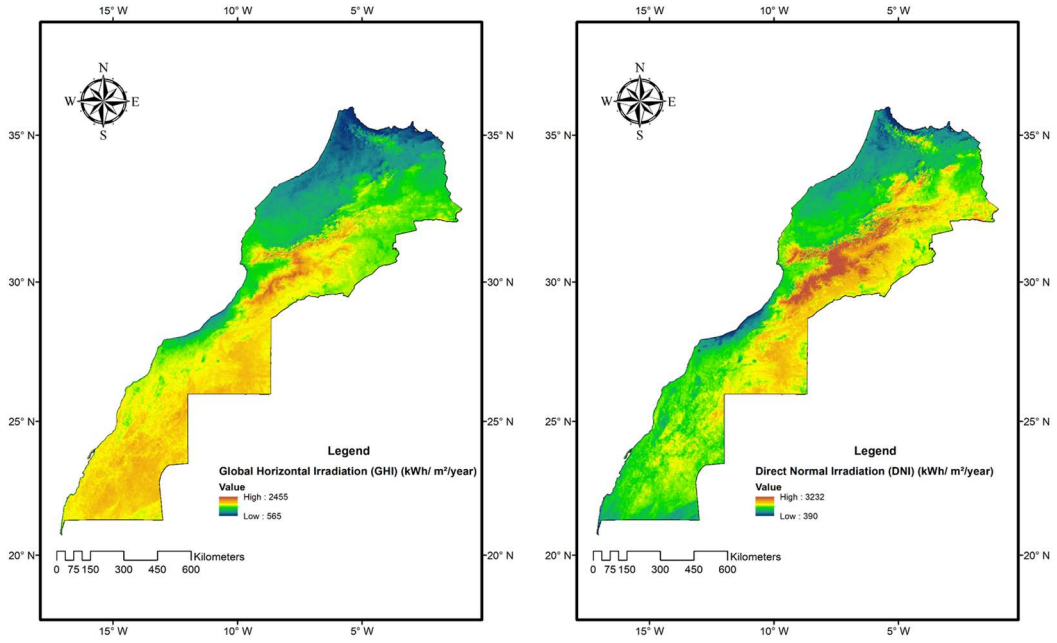


Figure 1-6. Le potentiel solaire au Maroc pour le GHI (gauche), et le DNI (droite)[1].

Tableau 1-1. Aperçu des projets solaires au Maroc.

	Site	Puissance en MW	Technologie
2018	NOOR Ouarzazate II	200 MW	CSP
	NOOR Ouarzazate III	150 MW	CSP
	NOOR Ouarzazate IV	72 MW	CSP
	NOOR Laayoune I	85 MW	PV
	NOOR Boujdour I	20 MW	PV
2020	NOOR Atlas	200 MW	PV
	NOOR Argana	200 MW	PV
	Nour Tafilalet-Erfoud	40 MW	PV
2021	NOOR Tafilalet	120 MW	PV
	Group OCP	18 MW	PV
2023	NOOR Boujdour II	350 MW	PV
	NOOR PV II	400 MW	PV
2024	NOOR Atlas	200MW	PV
	Maroc photovoltaïque	10 MW	PV

La politique du Maroc en matière de recherche, de développement et d'innovation (RD&I) dans le domaine de l'énergie vise à soutenir les priorités de sa politique énergétique nationale,

[1] Source:ARMINES /MINES ParisTech / TRANSVALOR / IRESEN

en stimulant l'utilisation de la technologie des énergies renouvelables et en aidant à développer une industrie manufacturière locale. Le Maroc dispose d'un paysage actif en matière de RD&I énergétique. Dans cette perspective, le gouvernement a créé une série d'agences pour les énergies renouvelables. Il s'agit notamment de : l'AMEE, MASEN, Société d'Investissement dans l'Energie (SIE), Institut de Recherche en Energie Solaire et Energies Nouvelles (IRESEN) [18,19].

- MASEN compte parmi ses missions principales la conduite d'une recherche préindustrielle, orientée vers le marché, sur toutes les sources d'ENR. Les initiatives de R&D à grande échelle de MASEN, orientées vers l'exploitation, ont abouti à des projets pilotes de production d'énergie renouvelable à l'échelle industrielle.
- IRESEN est l'une des institutions responsables de la mise en œuvre des projets de RD&I en soutien aux priorités du gouvernement. Il est chargé d'identifier les priorités et les projets de recherche qui sont alignés sur les objectifs de la politique énergétique nationale, tels que les énergies renouvelables et l'efficacité énergétique. L'IRESEN fonctionne à la fois comme une agence de financement qui organise des appels compétitifs pour des projets de R&D et comme un centre de recherche qui donne accès à des infrastructures de recherche et d'innovation, comme le Green Energy Park (GEP) et le Green Smart Buildings Park (GSBP) [20].

2. Rayonnement solaire : éléments de base et définitions

Le Soleil est la source absolue de toute l'énergie qui alimente le processus des cycles de vie de la planète et elle est la fournisseuse quasi unique d'énergie à la Terre. Le Soleil rayonne une quantité énorme d'énergie, appelée énergie solaire ou rayonnement solaire. Malgré la grande distance Terre-Soleil qui est d'environ 149.85 millions de km, les experts ont estimé qu'à tout moment, la Terre intercepte environ 180 10^6 GW de l'énergie solaire. En presque une demi-heure, la Terre reçoit une quantité d'énergie équivalente à la consommation annuelle mondiale en électricité [21]. Le rayonnement solaire est la principale source naturelle d'énergie sur Terre, seules des quantités mineures sont disponibles à partir d'autres sources, notamment l'énergie géothermique, l'énergie marémotrice et le rayonnement cosmique. En d'autres termes, le Soleil est responsable de toutes les sources d'énergie renouvelable sur Terre, qui répond aux besoins de l'être humain.

Le rayonnement solaire est une variable importante pour tout système de l'énergie solaire. L'information sur le rayonnement solaire est indispensable pour : l'étude de faisabilité et la sélection des sites favorables, l'optimisation de l'énergie produite à travers la prévision précise du rayonnement solaire, la minimisation de l'impact de variabilité de la puissance produite lors de son intégration dans le réseau électrique, l'estimation du coût financier de tout nouveau projet solaire.

2.1. Géométrie solaire

Le soleil est le plus grand membre du système solaire. Le soleil se situe à une distance moyenne de 14.9 10^7 km de la Terre, le Soleil a un diamètre de 13.9 10^5 km, ce qui est 110 fois supérieur à celui de la Terre qui est de 12.7 10^3 km. La terre fait un tour du Soleil tous les 365 jours. La trajectoire elliptique de la Terre nous rapproche du Soleil en janvier (périhélie) où la distance entre Terre-Soleil est de 14.7 10^7 km et nous éloigne en juillet (aphélie) où la distance entre les deux est de 15.2 10^7 Km. La Figure 1-7 représente l'orbite de la Terre en relation avec les saisons, causées par l'inclinaison moyenne de 23.5° de l'axe de rotation de la Terre par rapport au plan de l'orbite. La quantité et l'intensité du rayonnement solaire atteignant la surface de la Terre dépendent de la géométrie de la Terre par rapport au Soleil.

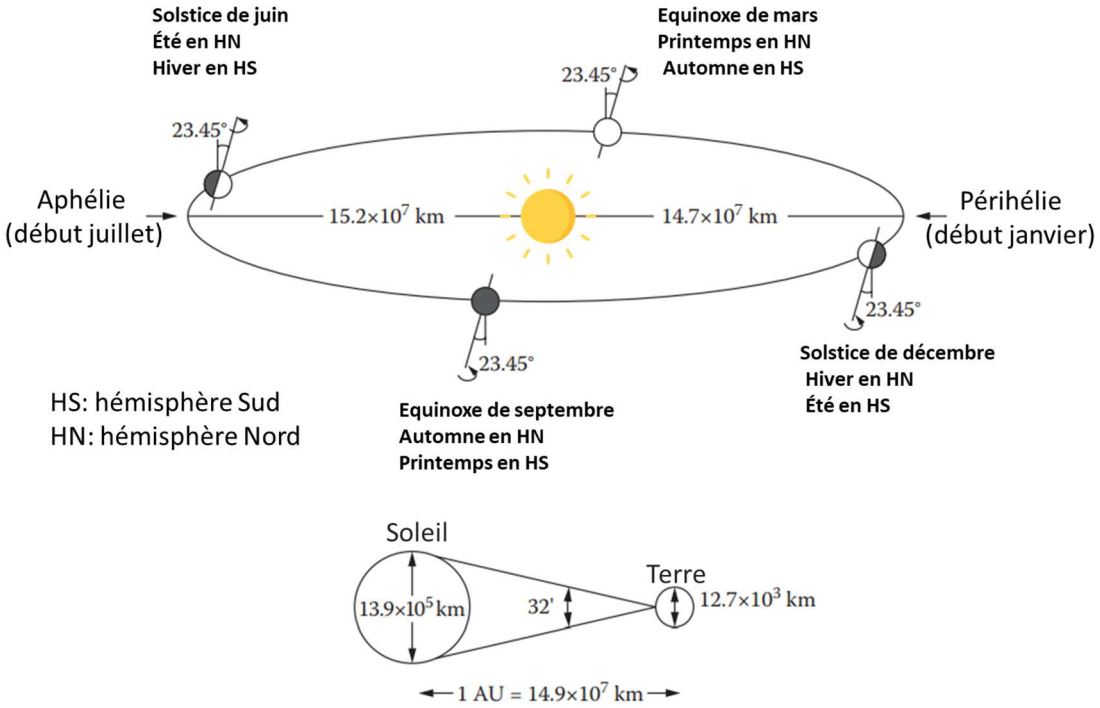


Figure 1-7. Mouvement orbital de la terre autour du soleil [22].

2.1.1. Coordonnées géographiques

Le système de coordonnées géographique permet de déterminer la localisation d'un endroit sur la Terre (Figure 1-8). Il est décrit par deux angles la latitude (φ) et la longitude (λ). La latitude correspond à l'angle d'élévation entre une ligne imaginaire allant du centre de la Terre à un point quelconque de la surface et sa projection sur le plan de l'équateur. Les valeurs de latitude sont comprises entre $90° < \varphi < -90°$. La latitude est positive pour l'hémisphère nord, négative pour l'hémisphère sud et nulle à l'équateur. La longitude d'un lieu est l'angle formé par le méridien du lieu avec le méridien de référence (Greenwich). Elle varie entre 0° et 180° vers l'est et entre 0° et -180° vers l'ouest.

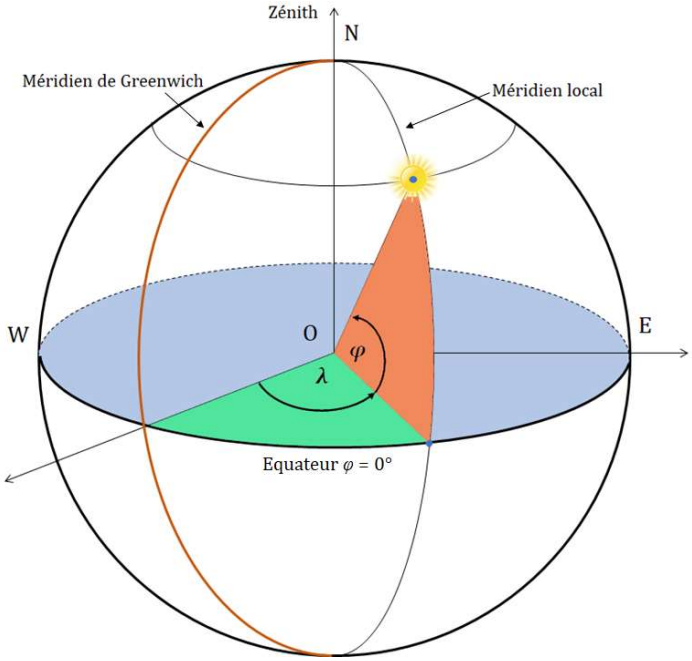


Figure 1-8. Système de coordonnées géographiques.

2.1.2. Système de coordonnées horizontales

La position du soleil par rapport à une surface horizontale sur la Terre est repérée par deux systèmes : le système de coordonnées horizontales et le système de coordonnées horaires.
Dans le système de coordonnées horizontales (Figure 1-9), la position de soleil est déterminée par deux grandeurs : la hauteur du soleil (γ) et l'azimut (Ψ). Les coordonnées horizontales dépendent du lieu de l'observation.

Hauteur du soleil

La hauteur de soleil (γ) est l'angle compris entre la direction apparente du soleil et la projection du faisceau de rayonnement sur le plan horizontal. Sa valeur varie entre -90° et 90° tout au long de la journée, elle dépend de la latitude et du jour de l'année.

Angle solaire zénithal

L'angle du Soleil par rapport à une ligne perpendiculaire à la surface de la Terre est appelé angle solaire zénithal (θ_z), de sorte que :

$$\gamma + \theta_z = 90° \tag{1.1}$$

L'angle solaire zénithal est défini par :

$$\cos(\theta_z) = \sin\gamma = \sin\varphi\sin\delta + \cos\delta\cos\varphi\cos\omega \tag{1.2}$$

δ représente la déclinaison solaire définie par l'équation (1.6)

Azimut

L'azimut (Ψ) est l'angle sur le plan horizontal entre la projection du rayonnement du faisceau et la ligne de direction nord-sud. Des valeurs positives de Ψ indiquent que le Soleil est à l'ouest du sud et des valeurs négatives indiquent que le Soleil est à l'est du sud. Sa valeur varie avec la latitude et la période de l'année.

L'azimut est calculé comme suit :

$$\psi = \cos^{-1}\left[\frac{\sin\delta\sin\varphi - \cos\delta\sin\varphi\cos\omega}{\cos\gamma}\right] \tag{1.3}$$

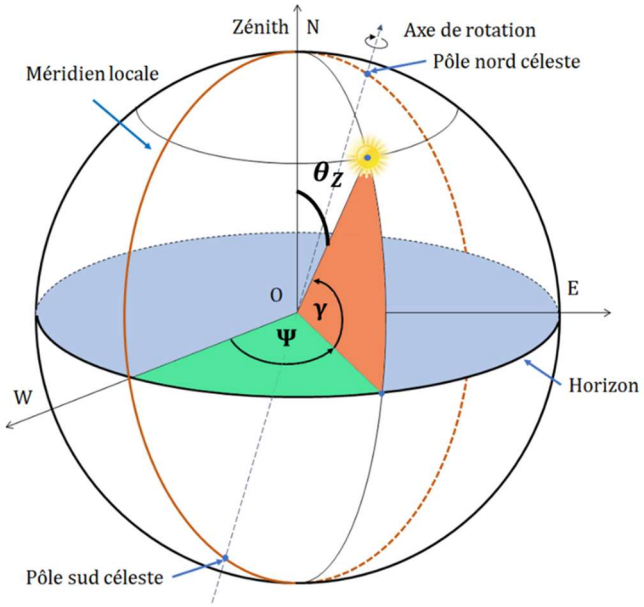


Figure 1-9. Systèmes des coordonnées horizontales.

Temps de lever et de coucher de soleil

Les temps du lever et du coucher du soleil correspondent à une hauteur solaire (γ) nulle, d'où :

- Temps du lever du soleil :

$$t_{lever} = 12 - \arccos(-tg(\delta)tg(\varphi))/15 \tag{1.4}$$

- Temps du coucher du soleil :

$$t_{coucher} = 12 + \arccos(-tg(\delta)tg(\varphi))/15 \tag{1.5}$$

2.1.3. Système de coordonnées horaires

Dans le système des coordonnées horaires (Figure 1-10), la position du soleil dépend de l'heure de l'observation et non pas à la position de l'observateur, le plan de référence pour ces

coordonnées est le plan équatorial. Dans ce système de coordonnées, la position du soleil est repérée par la déclinaison (δ), et l'angle horaire (ω).

Déclinaison (δ)

La déclinaison (δ) est définie par l'angle entre la ligne joignant les centres du Soleil et de la Terre, qui détermine également la direction des rayons directs provenant du Soleil et leur projection sur le plan équatorial. La déclinaison varie de -23.45° au solstice d'hiver à 23.45° au solstice d'été.

La déclinaison solaire est définie par l'équation :

$$\delta = 23.45\left[\sin\frac{2\pi(284+d)}{365}\right] \tag{1.6}$$

Angle horaire (ω)

L'angle horaire (ω) est la distance angulaire entre la position du Soleil à une heure donnée et sa position la plus élevée pour ce jour lorsqu'il traverse le méridien local au midi solaire. Il indique le temps écoulé depuis que le Soleil a été au méridien local. Comme la Terre tourne environ une fois toutes les 24 heures, l'angle horaire correspondant à 1 h est de 15°. L'angle horaire est nul à midi, négatif avant le passage du méridien et positif après le passage du méridien.

$$\omega = 15(TSV - 12) \tag{1.7}$$

TSV (Temps Solaire Vrai) se calcule par l'équation :

$$TSV = TU + \frac{\lambda}{15} + Et \tag{1.8}$$

TU corresponds au temps universel de Greenwich (GMT), et *Et* correspond à l'équation de temps définie comme suit :

$$Et = 9.87\sin(2B) - 7.53\cos(B) - 1.5\sin(B) \tag{1.9}$$

avec :

$$B = \frac{2\pi(d-81)}{365} \tag{1.10}$$

d est le nombre de jours depuis le début de l'année.

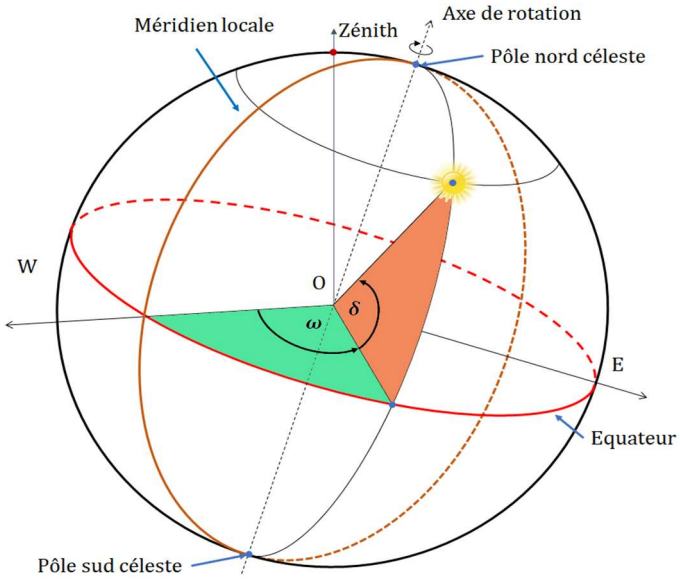


Figure 1-10. Système de coordonnées horaires.

2.2. Rayonnement solaire

2.2.1. Constante solaire

La quantité d'énergie solaire par unité de temps à la distance moyenne de la terre par rapport au soleil, reçue sur une unité de surface normale au soleil en dehors de l'atmosphère est appelée constante solaire ou bien la densité de flux énergétique au sommet de l'atmosphère I_0. La difficulté de mesure de cette quantité depuis la surface de la Terre en raison de l'effet de l'atmosphère a fait l'objet de plusieurs recherches scientifiques. La mise en orbite des satellites dotés d'outils de mesures modernes a permis de réaliser une mesure fiable de $I_0 = 1367 \text{W/m}^2$ avec une précision de 0.5% [23].

2.2.2. Rayonnement extraterrestre

Le rayonnement normal extraterrestre G_0 s'exprime en fonction de I_0 et de la correction de la constante solaire R_c par l'équation suivante :

$$G_0 = I_0 . R_c \tag{1.11}$$

où,

$$R_c = 1 + 0.033 \cos \frac{2\pi d}{365} \tag{1.12}$$

avec d est le numéro du jour de l'année à partir du 1er janvier.

Les variations du rayonnement normal extraterrestre G_0 sont dues à la variation de la distance Soleil-Terre au cours de l'année en raison de l'orbite elliptique du mouvement de la Terre autour du Soleil.
Pour une année normale, la variation G_0 à la surface de la Terre au début de chaque mois (le 1er janvier correspond à $d=1$) est illustrée par la Figure 1-11 en utilisant l'équation (1.11). Le niveau minimal du G_0 est de 1320 W/m² est atteint le jour de l'aphélie (5 juillet) et le niveau maximal est de 1415 W/m² est atteint le jour du périhélie (3 ou 4 janvier).

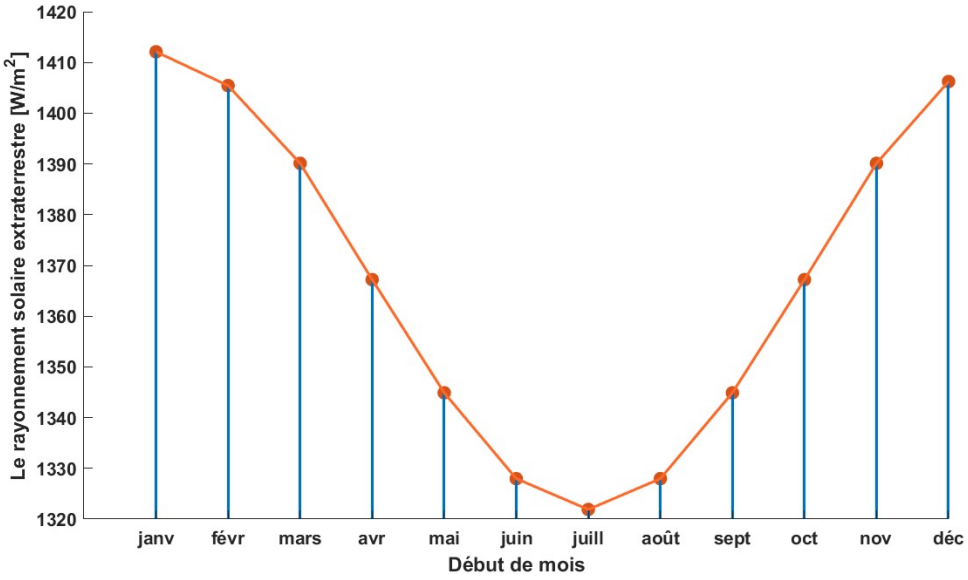


Figure 1-11. La variation du G_0 en fonction du mois de l'année.

Sur un plan horizontal, le rayonnement solaire au top de l'atmosphère TOA pour un site donné dépend du rayonnement normal extraterrestre G_0 et de la hauteur solaire (γ), et il est défini par l'équation suivante :

$$TOA = G_0.\sin(\gamma) \tag{1.13}$$

2.2.3. Rayonnement solaire à la surface terrestre

En dehors de l'atmosphère, le rayonnement solaire est pratiquement constant, sur Terre, il varie en fonction du jour de l'année, l'heure de la journée, la latitude et l'état de l'atmosphère. En traversant l'atmosphère, le rayonnement solaire interagit avec les différents constituants de l'atmosphère. Par conséquent, seule une partie du rayonnement extraterrestre atteint le sol, le reste est absorbé ou diffus vers l'espace par les composantes de l'atmosphère. Les rayonnements peuvent être transmis, absorbés, ou diffusés (Figure 1-12).

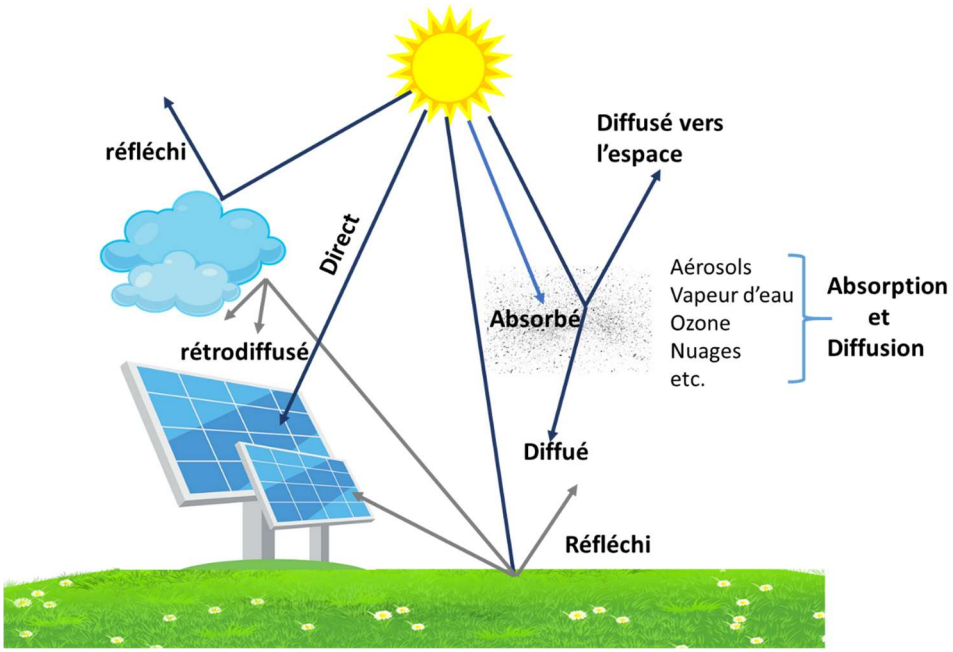


Figure 1-12. Composantes du rayonnement solaire résultantes des interactions avec l'atmosphère.

Les interactions du rayonnement solaire avec l'atmosphère terrestre donnent lieu à trois composantes fondamentales à large bande qui présentent un intérêt pour les technologies de conversion de l'énergie solaire :

- Rayonnement direct normal (DNI) : c'est le rayonnement solaire provenant directement du disque solaire, non absorbé et non diffusé. Sa valeur s'affaiblit avec la présence des nuages. Il est intéressant pour les technologies solaires à concentration (CSP) et le photovoltaïque à concentration (CPV).
- Rayonnement diffus sur plan horizontal (DHI): c'est le rayonnement résultant de la diffusion du rayonnement solaire par les différentes composantes de l'atmosphère (aérosols, vapeur d'eau, ozone, etc.). Il correspond au rayonnement diffusé à partir du disque solaire, hors le rayonnement direct. Plus le ciel est couvert plus sa valeur est grande.
- Rayonnement réfléchi R : il correspond au rayonnement réfléchi par l'environnement extérieur, en particulier le sol. Plus le sol est clair plus il est reflétant. Le coefficient de réflexion est appelé albédo.
- Rayonnement global sur plan horizontal (GHI) : il correspond au flux total du rayonnement incident sur une surface horizontale. C'est la somme du rayonnement direct et diffus, la composante réfléchie est nulle dans ce cas. Le rayonnement global est intéressant pour les technologies photovoltaïques. Le GHI est la quantité de rayonnement solaire à ondes courtes dans la gamme spectrale d'environ 280 à 4000 nm reçue par un champ de vision hémisphérique de 2π stéradians sur une surface quelconque.

- Rayonnement global sur plan incliné (GTI) : représente le rayonnement total reçu par une surface inclinée. C'est la combinaison du rayonnement direct, diffus et réfléchi par le sol.

Les composantes du rayonnement solaire sont liées entre eux par les équations suivantes :

$$GHI = DHI + DNI\cos(\theta_z) \tag{1.14}$$

3. Mesure du rayonnement solaire

La connaissance du rayonnement solaire disponible à un endroit donné est d'une grande importance pour plusieurs domaines scientifiques, tels que la biologie, l'agriculture, l'hydrologie, l'architecture, la climatologie, la météorologie et les énergies renouvelables. Les données de rayonnement solaire issues de mesures pyranométriques au sol sont particulièrement pertinentes pour le secteur de l'énergie solaire, où elles sont utilisées pour l'étude de faisabilité et de sélection de sites, l'étude de la variabilité de la ressource solaire [24], ou pour la validation et la calibration de modèles météorologiques satellitaires et numériques [25–27]. Pour ces applications, il est essentiel de s'assurer que les données sont de bonne qualité et satisfont à la fiabilité requise pour atteindre les objectifs scientifiques pour lesquels elles sont collectées. Les mesures directes du rayonnement solaire dans les stations météorologiques constituent la source la plus précise de données, à condition que l'équipement soit bien entretenu et régulièrement étalonné. Des estimations de rayonnement solaire peuvent aussi être obtenues à travers l'utilisation des méthodes d'interpolation spatiale à partir des stations voisines. Cependant, la densité de ces stations au sol reste insuffisante pour avoir une interpolation à cause des coûts élevés d'investissement et de maintenance [28–31]. Pour surmonter ce manque, la modélisation du rayonnement solaire et les observations par satellites se présentent comme des approches alternatives pour dériver les données du rayonnement solaire. Les modèles de rayonnement solaire peuvent être classifiés de différentes manières, et il n'existe pas une classification générale. Dans cette thèse, on va considérer les modèles empiriques, les modèles physiques et l'imagerie satellitaire.

3.1. Stations météorologiques au sol

Le champ de rayonnement à la surface de la Terre peut être divisé en deux zones spectrales à large bande distinctes : le rayonnement à ondes courtes ou bien rayonnement solaire avec des longueurs d'onde de 300 à 3000 nm, et le rayonnement à ondes longues nommé aussi rayonnement terrestre où les longueurs d'onde varient entre de 3000 nm et à 100 μm. La quantité de rayonnement atteignant la surface terrestre dépend d'interactions multiples du rayonnement terrestre avec les composantes de l'atmosphère telles que les nuages, les aérosols et les gaz atmosphériques.

Les pyranomètres et les pyrhéliomètres sont des capteurs utilisés pour mesurer le GHI, le DHI, le GTI et le DNI à travers la conversion de l'énergie électromagnétique reçue en énergie thermique par un matériau absorbant couvrant la surface d'une thermopile , ou à travers la

réponse d'un semi-conducteur aux photons arrivant à la surface du matériau, générant un courant ou une tension électrique.

Il existe deux types de capteurs : les pyranomètres à thermopiles et les pyranomètres à photodiode. Ils diffèrent les uns des autres par le type de capteur qu'ils utilisent pour déduire la quantité d'irradiation solaire reçue. La Figure 1-13 représente des images réelles des deux types de pyranomètres disponibles aux stations de Benguerir et Zagora. Les pyranomètres à thermopiles sont les meilleurs instruments, ils ont un revêtement noir qui permet une réponse spectrale large et uniforme à toutes les longueurs d'onde du rayonnement solaire d'environ 300-3000 nm (Figure 1-14). Les pyranomètres à thermopile mesurent indirectement le rayonnement en quantifiant la différence de température entre deux surfaces : l'une absorbante et l'autre non absorbante. Ces surfaces sont en contact avec deux métaux ayant des capacités thermiques différentes, appelés thermocouple. Lorsque la surface absorbante de la thermopile est chauffée par le rayonnement solaire incident, elle crée un flux thermique sur la jonction du thermocouple. Par conséquent, cette jonction produit une tension proportionnelle à la différence de température [3,32]. Les pyranomètres à photodiode mesurent indirectement l'irradiance en quantifiant le photocourant généré en réponse à un flux d'irradiance incident. Les pyranomètres à photodiode sont moins chers et ont un temps de réponse plus faible que les pyranomètres à thermopile. Le temps de réponse des pyranomètres à thermopile varie entre 1 et 240 s alors que le temps de réponse des pyranomètres à photodiode est d'environ 10 µs (Figure 1-14). Néanmoins, les pyranomètres à photodiode sont moins précis que la plupart des pyranomètres à thermopile. Une grande partie des imprécisions des pyranomètres à photodiode est liée à leurs réponses spectrales étroites [33].

3.1.1. Pyranomètres

Les pyranomètres mesurent le GHI, le DHI et le GTI. Les pyranomètres ont un champ de vision de 180◦. La mesure de DHI se fait en bloquant le rayonnement du faisceau à l'aide d'une bande ou d'une boule opaque placée au-dessus du pyranomètre installé sur suiveur de Soleil. Une autre méthode moins coûteuse pour la mesure de DHI consiste à fixer une bande circulaire opaque de manière à ombrager le pyranomètre tout au long de la journée d'une manière à empêcher le rayonnement direct. Cette deuxième méthode est loin d'être parfaite, car l'effet de blocage de la bande varie en fonction de la géométrie solaire, du temps et des conditions atmosphériques ce qui entraine des erreurs dans la mesure de DHI. Pour cette raison, le suivi de soleil est la méthode recommandée pour des études précises de la ressource solaire.



Figure 1-13. Pyranomètres : thermopile Kipp & Zonen CMP21 (droite) installé à la station météo de Benguerir, et photodiode (gauche) installé à la station météo de Zagora.

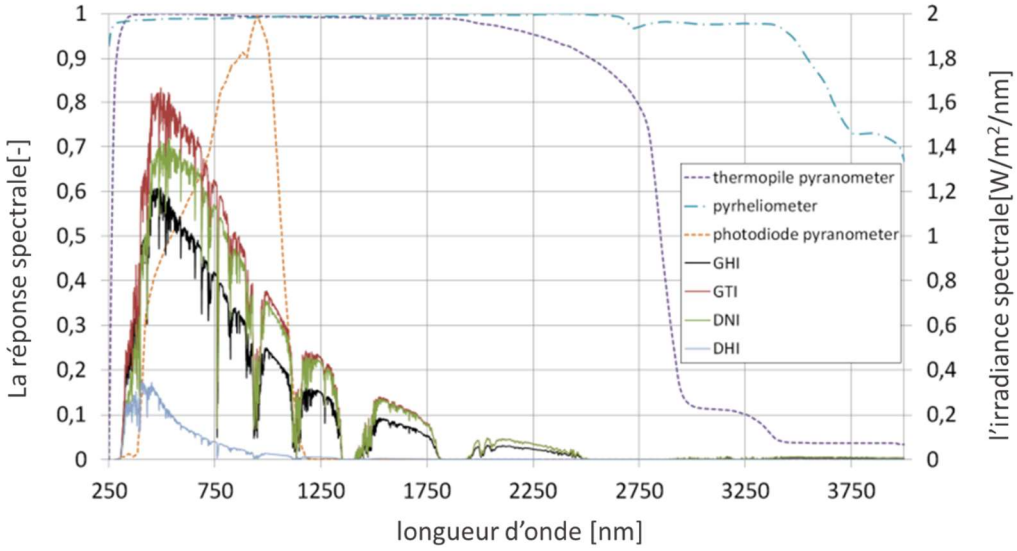


Figure 1-14. Réponse spectrale du pyranomètre photodiode de LI-COR, du pyranomètre thermopile CMP21 de Kipp & Zonen et du pyrhéliomètre CHP1 tracée à côté des spectres de GHI, DNI, DHI et GTI [35].

3.1.2. Pyrhéliomètres

Les pyrhéliomètres sont utilisés pour la mesure de DNI. Les pyranomètres utilisent des capteurs thermopiles, ils sont généralement montés sur des suiveurs solaires automatiques pour maintenir l'alignement de l'instrument avec le disque solaire et pour éclairer complètement le détecteur du lever au coucher du soleil. Les suiveurs solaires utilisent un logiciel pour calculer et suivre précisément la position du soleil. Les pyrhéliomètres se composent d'un élément de détection de rayonnement enfermé dans un boîtier (tube de collimation) qui a une petite ouverture à travers laquelle seuls les rayons solaires directs entrent. La Figure 1-15 représente une image réelle d'un pyrhéliomètre Kipp & Zonen CHP1installé sur un suiveur solaire à la station météorologique de GEP.



Figure 1-15. Pyrhéliomètre Kipp & Zonen CHP1 installé à la station météorologique de Benguerir.

Différents types de pyranomètres sont disponibles à partir de divers fabricants. L'organisation internationale de normalisation (ISO) et l'OMM ont établi des classifications d'instruments et des spécifications pour la mesure de l'irradiance solaire pour montrer la fiabilité des instruments selon leurs buts d'utilisation. Les instruments les plus utilisés sont ceux des sociétés Kipp & Zonen [34], Apogee [35], Campbell Scientific [36] et Vaisala [37].

3.2. Modèles empiriques

Les modèles empiriques de rayonnement solaire utilisent les relations entre la transmissivité atmosphérique et les variables météorologiques tels que la couverture nuageuse, la durée d'ensoleillement, la température, les précipitations, et l'humidité. La couverture nuageuse est la variable la plus importante, et elle est considérée le facteur principal atténuant le rayonnement solaire lors de sa traversée de l'atmosphère via le phénomène de réflexion et de rétrodiffusion. La précision des modèles empiriques dépend de la variable météorologique utilisée en entrée et ainsi que leurs coefficients dépendent de l'endroit où l'étude est faite, et ils ne peuvent être déterminés avec précision que pour les endroits où le rayonnement solaire est mesuré dans des stations météorologiques au sol [38–40].

3.3. Modèles physiques

Les modèles physiques se basent sur des algorithmes de transfert du rayonnement solaire à travers l'atmosphère sur la base de relations et de lois connues issues de la physique. Le rayonnement solaire transféré dans l'atmosphère est absorbé par les gaz et diffusé par les molécules, les aérosols et les particules des nuages. En raison de la complexité du transfert radiatif et de la grande variabilité temporelle et spatiale de la composition atmosphérique, les modèles basés sur la physique nécessitent de nombreux paramètres d'entrée afin de les régler localement. Les modèles basés sur la physique reposent généralement sur des estimations de la composition de l'atmosphère dérivée de modèles de prévision numérique du temps [41,42].

3.4. Modèles satellitaires

Les satellites géostationnaires sont la source la plus importante de données sur le rayonnement solaire dérivées des satellites. Parmi ces capteurs on trouve le radiomètre imageur SEVIRI embarqué sur le satellite météorologique géostationnaire Météosat Seconde Génération (MSG). Ces images sont caractérisées par une résolution spatiale élevée et comprise entre 3 et 5 km. En ce qui concerne la résolution temporelle, ces images sont disponibles toutes les 15 min.

Des observations fréquentes sont essentielles pour estimer la quantité de rayonnement solaire qui atteint la surface de la Terre. Ces images permettent d'estimer le rayonnement solaire de manière dense, régulière et persistante dans l'espace et le temps, surtout pour les endroits où il n'y a pas beaucoup de stations de mesure. De nos jours, plusieurs bases de données reposées sur les satellites géostationnaires sont disponibles, tels que EMPClimate [43], PVGIS [44], HelioClim3 [45], SolarGis [46], etc.

3.5. Rayonnement par ciel clair

Le rayonnement par ciel clair correspond à l'irradiation disponible au sol dans des conditions de ciel sans nuages. La connaissance du GHI par ciel clair atteignant le niveau du sol est un paramètre clé dans le domaine de la modélisation et l'évaluation de la ressource solaire et l'une des parties fondamentales des modèles de rayonnement solaire, après la modélisation des nuages. L'irradiation solaire terrestre par temps clair a été largement étudiée et dépend de l'angle d'élévation solaire, de l'altitude du site, de la concentration des aérosols, de la vapeur d'eau et de diverses autres conditions atmosphériques [47]. Comme l'irradiance est une mesure de la puissance de la lumière solaire (W/m^2), cette information peut être utilisée pour modéliser la production maximale générale d'un système solaire photovoltaïque pour un jour et un lieu donné. En effet, les différentes approches de modélisation du rayonnement solaire se basent sur le calcul des propriétés de nuages à partir des modèles satellitaires et de coupler ces informations avec un modèle de ciel clair. De même, les modèles de ciel clair sont utilisés pour des prévisions à court terme couplées à des prévisions de nuages à partir et d'imageurs de ciel. Les modèles de ciel clair sont nécessaires pour le calcul de l'indice de ciel clair qui est un paramètre important pour l'estimation du degré de nébulosité et la détection des instants de ciel clair à partir des mesures de rayonnement solaire brutes, pour l'étude de la variabilité du rayonnement solaire, ou pour le contrôle de qualité des mesures de rayonnement solaire . Les modèles de ciel clair peuvent être classés en deux grandes catégories, à savoir les approches empiriques [48–52] et les modèles physiques. Les modèles empiriques se basent sur les composantes atmosphériques qui décrivent les conditions atmosphériques locales pour fournir des expressions analytiques de l'éclairement par ciel clair pour un lieu donné. En revanche, les modèles physiques se basent sur les modèles de transfert radiatif qui nécessitent des paramètres d'entrée tels que la vapeur d'eau, l'ozone, l'eau précipitable et la profondeur optique des aérosols pour simuler l'atténuation de rayonnement à travers les différentes couches atmosphériques. Ils présentent l'avantage de la généralité et ils sont plus précis, mais leur coût de calcul est plus élevé [53]. Néanmoins, certains modèles, comme MAGIC [54], McClear [55],

ou SOLIS [56] utilisent une approche hybride qui combine les avantages des approches empiriques et physiques. Tous au long de cette thèse, pour le contrôle de qualité, la détection des instants de ciel clair et l'étude de la variabilité nous avons utilisé le modèle de ciel clair McClear après l'avoir validé par les différentes données au sol disponibles, les résultats de validation sont présentés dans le chapitre 3. La variation diurne du GHI par ciel clair (GHI_{cls}) dérivé de McClear est illustrée à la Figure 1-16, pour le site de GEP (Benguerir, Maroc) à titre d'exemple. La valeur du GHI_{cls} minimal est observée le 21 décembre (solstice d'hiver) et la plus élevée le 21 juin (solstice d'été).

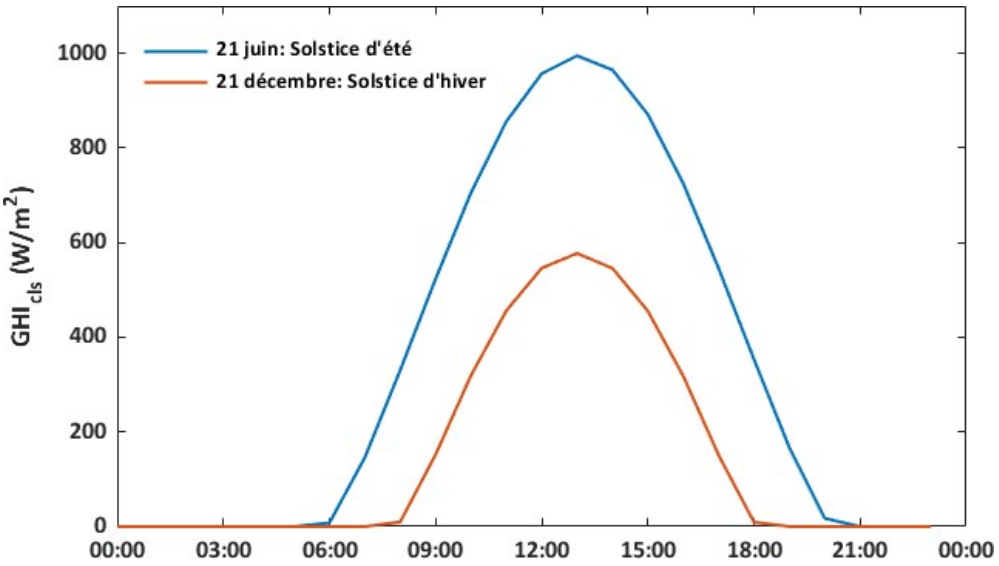


Figure 1-16. Le GHI_{cls} horaire pour le site de Benguerir, Maroc à titre d'exemple (32.12°,-7.94°) le 21 juin et le 21 décembre.

3.5.1. Indice de clarté K_t

L'indice de ciel clair k_{cs} et l'indice de clarté K_t sont largement utilisés pour décrire la transmission de l'atmosphère et pour classer les conditions météorologiques [57,58].
L'indice de clarté K_t est défini comme le rapport entre l'irradiance au niveau du sol et l'irradiance extraterrestre sur plan horizontal TOA et il ne prend en compte que les tendances causées par les effets géométriques sur la position solaire. Puisque l'extinction atmosphérique dépend de la longueur du trajet du rayonnement à travers l'atmosphère, elle est également régie par la géométrie solaire. Cet indice réduit fortement les variations saisonnières et journalières en considérant l'influence de l'angle zénithal, qui est modélisé par le rayonnement extraterrestre.

$$K_t = \frac{GHI}{TOA} \tag{1.15}$$

3.5.2. Indice de ciel clair k_{cs}

Pour tenir compte de l'influence des nuages sur l'irradiance solaire, l'indice de ciel clair k_{cs} est introduit. Le k_{cs} est défini comme étant le rapport entre le GHI au niveau du sol et le GHI_{cls}. L'indice de ciel clair permet d'éliminer les signaux diurnes et saisonniers d'un ensemble donné de données de rayonnement.

$$k_{cs} = \frac{GHI}{GHI_{cls}} \tag{1.16}$$

4. Variabilité du rayonnement solaire

L'énergie photovoltaïque est une source d'énergie intermittente, et sa production dépend fortement du GHI. La pénétration à grande échelle des installations PV dans le réseau est fortement limitée par la variabilité incontrôlable de l'irradiance solaire au niveau du sol, ce qui peut entraîner des problèmes de stabilité du réseau. La variabilité a deux causes, la première est causée par le mouvement de soleil et elle est prévisible vu qu'elle est le résultat de la géométrie solaire. En revanche, la deuxième est moins prévisible vu qu'elle est causée par le mouvement des nuages et des systèmes météorologiques. Un processus de classification de l'irradiation solaire en fonction des conditions nuageuses peut être utile pour différentes applications de l'énergie solaire, comme l'évaluation des modèles de prévision de l'irradiation solaire dans différentes conditions nuageuses, ce qui permet de réduire l'incertitude des prévisions de l'irradiation solaire et de permettre des prédictions précises, d'utiliser efficacement la production d'énergie fluctuante des systèmes PV, de comparer les caractéristiques de différents sites solaires et de déterminer la météorologie la plus dominante, et enfin de choisir la technologie solaire la plus appropriée pour un site donné en fonction des résultats de la classification. La Figure 1-17 représente un exemple de la variabilité causée par les nuages pour la production d'un système PV. La même figure montre la relation entre la puissance PV et le GHI pendent un jour nuageux (le 16/06/2020) et un jour clair (le 17/05/2020). Nous observons que la puissance PV et le GHI ont la même allure et chaque fluctuation de GHI provoque une fluctuation de la puissance PV. Ces fluctuations doivent être gérées pour assurer la fiabilité des opérations du réseau électrique.

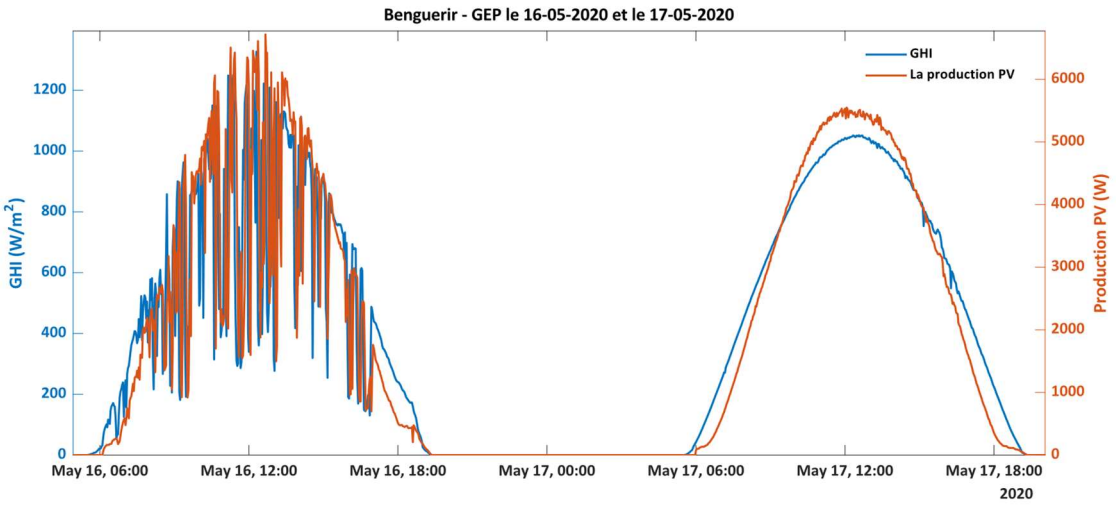


Figure 1-17. Les profils journaliers au pas de 2 minutes du GHI et de la production PV mesurés au GEP, Benguerir.

De nombreuses méthodes sont rapportées dans la littérature pour quantifier la variabilité de l'irradiance solaire, mais jusqu'à présent, il n'existe aucune méthode appropriée qui peut être employée à cette fin. Perez et al. [59] ont proposé un modèle empirique pour quantifier la variabilité de l'irradiation solaire pour les échelles de temps de 20 secondes, 1 minute, 5 minutes et 15 minutes, basé sur l'indice de ciel clair. Comme quantificateurs de la variabilité intrahoraire, l'écart-type de l'indice global d'irradiance du ciel clair et la variation moyenne de l'indice de ciel clair entre deux intervalles de temps consécutifs ont été utilisés. Stein et al. [60] ont proposé un nouvel indice de variabilité VI, défini comme le rapport de la somme des *n* différences des valeurs consécutives du GHI en W/m^2 et de la somme des n différences de l'indice de ciel clair pendant une période donnée. Kang et Tam [61] ont proposé une nouvelle méthode pour caractériser et classer les conditions de ciel quotidiennes en utilisant l'indice de ciel clair quotidien et une nouvelle métrique appelée probabilité de persistance quotidienne (POPD). Les auteurs ont constaté que la méthode proposée fournit des résultats intéressants et que les différentes classes obtenues sont significatives pour tous les sites d'étude. Lave et al. [62] ont introduit une nouvelle métrique appelée le score de variabilité de la distribution du taux de rampe pour quantifier la variabilité de l'irradiance solaire à petite échelle temporelle. Lauret et al. [63] ont analysé la variabilité intrajournalière de l'irradiance afin de caractériser un lieu donné en examinant la relation entre deux paramètres l'indice de ciel clair quotidien et la variabilité intrajournalière pour 20 sites présentant différents types de climat. Ils ont constaté que le modèle dépend de l'emplacement du site, et ont donc proposé deux modèles alternatifs pour les sites où la formation des nuages est influencée par l'orographie locale et les sites où la formation des nuages est uniquement liée à des évènements météorologiques. Dans leurs articles [64], Blaga et Paulescu ont présenté une caractérisation de la nature stochastique des séries d'irradiation solaire sous différents angles, en analysant six quantificateurs différents de la variabilité de l'irradiation solaire. Pour comparer ces différents quantificateurs, ils ont introduit une nouvelle procédure de classement multiparamétrique pour classer les jours selon le modèle de variabilité de l'irradiation solaire. Finalement, dans son article [65] Lohmann a

passé en revue les progrès récents dans la caractérisation de la variabilité de l'irradiance solaire sur de petites échelles spatiales et temporelles, selon l'auteur, bien qu'il existe de nombreux articles traitant de la question de la variabilité de l'irradiance solaire, il y a encore peu de données de mesure appropriée à haute résolution pour valider de manière robuste les modèles existants.

Conclusion

Dans ce chapitre, nous avons présenté les équations de la géométrie solaire qui seront utilisées pour le traitement de données et la modélisation du rayonnement solaire. Par la suite, nous avons présenté les instruments de mesure des différentes composantes de rayonnement solaire et nous avons également introduit d'autres sources d'estimation du GHI, telles que les modèles satellitaires et empiriques. Nous avons aussi discuté du rayonnement par ciel clair, qui est un élément important pour la prévision et l'estimation de la ressource solaire. Finalement, nous avons représenté la notion de variabilité du rayonnement solaire, la raison pour laquelle des prévisions précises sont nécessaires pour gérer l'énergie produite. Le chapitre suivant présente l'état de l'art sur les méthodes de prévision du rayonnement solaire et une classification de ces derniers selon la résolution spatiale et temporelle. Les métriques d'évaluation de la précision des prévisions et les modèles de références sont aussi abordés.

Chapitre 2

État de l'art sur les méthodes de prévision du rayonnement solaire

Introduction **27**

1. Facteurs influençant la précision des prévisions **27**

1.1. Horizon de prévision 28

1.2. Métriques d'évaluation de la précision des prévisions 29

2. Techniques de prévision du rayonnement solaire **30**

2.1. Prévision numérique du temps 31

2.2. Imagerie satellitaire 34

2.3. Imagerie du ciel 36

2.4. Méthodes de séries temporelles 38

2.4.1. Persistance 39

2.4.2. SARIMA 39

2.5. Techniques de l'intelligence artificielle 41

2.5.1. Réseaux de neurones artificiels 43

2.5.2. MLP 45

2.5.3. LSTM 47

2.5.4. CNN 50

2.6. Modèles hybrides 53

Conclusion **56**

Introduction

La prévision précise de la production PV à l'horizon souhaité est un élément essentiel pour l'industrie, l'électricité et l'énergie. Une prévision erronée de la production PV peut entraîner soit une surestimation de la charge, qui conduit à un excès d'approvisionnement et par conséquent à des coûts plus élevés. Elle peut également entraîner une sous-estimation de la charge, qui ne permet pas de rassembler suffisamment de provisions et donc des services supplémentaires plus coûteux. Ces raisons ont motivé de nombreuses recherches sur des modèles de prévision capables de réduire cette contrainte financière tout en augmentant la précision des prévisions, soit mêmes, d'un faible pourcentage. Des recherches considérables ont été menées pour le développement des méthodes de prévision de la ressource solaire et de la production PV au cours des dernières décennies et des milliers d'articles ont été publiés dans ce contexte.

Ce chapitre propose une revue de littérature sur les méthodes de prévision du GHI ou de la production PV. Nous enclenchons par une étude bibliographique de différentes méthodes qui ont été adoptées pour la prévision de la ressource solaire tout en soulignant leurs majeurs avantages et limitations. Par la suite, nous citons laconiquement les méthodes de prévision basée sur l'imagerie satellitaire, l'imagerie de ciel, et les méthodes numériques de prévision de temps. Une grande partie de cette étude bibliographique sera consacrée aux méthodes statistiques auxquelles nous nous intéressons dans cette thèse.

1. Facteurs influençant la précision des prévisions

La prévision de GHI est l'étape la plus essentielle de la majorité des systèmes de prévision de la puissance photovoltaïque. Les approches de prévision du GHI peuvent être classifiées en fonction de l'horizon de prévision ainsi selon les données d'entrée disponibles. Pour des horizons inférieurs à l'heure, les méthodes de l'imagerie du ciel au sol fournissant des informations sur le comportement temporel des nuages peuvent servir de base pour dériver des prévisions de GHI avec une résolution spatiale et temporelle précise. Les prévisions basées sur les images satellitaires sont performantes pour les horizons de 30 min à 6 h. Pour des horizons de prévision plus longs, des prévisions basées sur les modèles de prévision numérique de temps sont utilisées. Alors que les méthodes d'apprentissage automatique peuvent s'adapter à n'importe quel horizon de prévision. En plus de ces méthodes, il existe les méthodes hybrides qui combinent plusieurs techniques de prévision pour répondre aux limitations des modèles individuels, dérivant ainsi une prévision optimisée en fonction de l'horizon souhaité [66].

La précision des prévisions de GHI est principalement influencée par la variabilité des conditions météorologiques et climatologiques. La précision maximale réalisable est déterminée principalement par les facteurs suivants : les conditions climatiques et météorologiques locales, l'horizon de prévision et les métriques de performances utilisées. Pour pouvoir obtenir une meilleure précision, la prise en considération de ces différents facteurs est nécessaire.

Le type de climat et les conditions météorologiques sont parmi les facteurs qui agissent sur la précision des prévisions de GHI et de la production PV. Dans une étude conduite par Bayer et al. [67] en utilisant 26 stations de mesure au sol en Espagne pour un an, les auteurs ont témoigné que la précision des prévisions est fortement influencée par le climat régional. De même, Lorenz et al. [68] ont constaté que pour des stations d'Europe centrale, le rRMSE (l'erreur quadratique moyenne relative) variait entre 40 % et 60 %, contre des valeurs entre 20 % et 35 % pour des stations en Espagne. Dans une autre étude conduite en Suisse [69], les auteurs ont trouvé un RMSE plus faible dans les endroits où les conditions météorologiques étaient plus ensoleillées. Dans ce sens, plusieurs chercheurs ont déduit que la classification météorologique est considérée comme un outil efficace pour améliorer la précision de la prévision solaire [70–73].

1.1. Horizon de prévision

L'horizon de prévision est défini comme étant la durée entre le moment actuel et le moment de la prévision. Le choix de l'horizon dépend essentiellement de l'objectif de cette de prévision. Jusqu'à présent, aucune norme bien définie ne permet de classer les prévisions en fonction de l'horizon temporel. Certains chercheurs ont divisé les prévisions selon l'horizon temporel en trois catégories [74], certains d'entre eux l'ont divisée en quatre catégories [75], alors que d'autres chercheurs ont utilisé cinq catégories [76]. Dans cette thèse, nous considérons les quatre catégories en dessous vu qu'il n'existe pas encore de critères de classification universellement reconnus [76–78].

- Prévisions à très court terme (1 minute à plusieurs minutes à l'avance).
- Prévision à court terme (1 heure ou plusieurs heures à l'avance à 1 jour ou 1 semaine à l'avance).
- Prévisions à moyen terme (1 mois à 1 an).
- Prévisions à long terme (1 an à 10 ans).

Il a été déjà rapporté que l'horizon de prévision dépend de l'application de ces prévisions. Nous résumons alors au niveau du Tableau 2-1 les applications correspondantes à chaque horizon de prévision [76,79–82].

Tableau 2-1. Exemples d'applications des prévisions solaires pour chaque horizon de prévision.

Très court terme	Court terme	Moyen terme	Long terme
- Planification du réseau électrique et du réseau intelligent - Commercialisation et tarification de l'électricité - Surveillance de la répartition de	- Assurance de la qualité et la stabilité du réseau - Planification de l'exploitation des réseaux de distribution - Répartition économique de la	- Programmation de la maintenance des systèmes d'énergie solaire - Planification de la maintenance	- Production, transport et distribution d'électricité et rationnement de l'énergie solaire

l'électricité en temps réel - Contrôle du stockage PV	charge et exploitation du système électrique - Contrôle du système intégré de gestion		- Eudes des tendances saisonnières - Planification des investissements

1.2. Métriques d'évaluation de la précision des prévisions

Avec la diversité des méthodes de prévision, la comparaison des résultats et l'évaluation de l'efficacité relative de ces méthodes de prévision sont essentielles. Il n'y a pas une mesure de performance unique pour toutes les situations, le chercheur doit plutôt décider de la meilleure méthode selon la situation de prévision. Afin d'évaluer la précision des modèles de prévision développés, les indicateurs de performances les plus utilisés dans la littérature [66,81,83–85] ont été considérés dans ce travail.

- **Erreur de biais moyen (MBE),** caractérise la balance entre la sous-estimation et la surestimation de la prédiction, il est défini comme suit :

$$MBE = \frac{1}{N}\sum_{i=1}^{N}(y_t - \tilde{y}_t) \tag{2.1}$$

où $y_t$est le rayonnement solaire mesuré à l'instant *t*, $\tilde{y}_t$est la valeur du rayonnement solaire prédite à l'instant t, et N est le nombre de points de données.

- **L'erreur quadratique moyenne (RMSE)**, une mesure de la dispersion moyenne des erreurs, définie par l'équation suivante :

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y_t - \tilde{y}_t\right)^2} \tag{2.2}$$

- **Le RMSE relatif (rRMSE),** est également fréquemment utilisé pour afficher les erreurs. Le *rRMSE* est exprimé en pourcentage :

$$rRMSE = \frac{\sqrt{\frac{1}{N}\sum_{t=1}^{N}\left(y_t - \tilde{y}_t\right)^2}}{\frac{1}{N}\sum_{t=1}^{N} y_t} \times 100 \tag{2.3}$$

Toutes ces mesures peuvent être déclarées comme absolues, ou bien normalisées en les divisant par une valeur de référence pour faciliter les comparaisons.

- **Le coefficient de détermination R^2,** utilisé pour mesurer la capacité des valeurs prévisionnelles à prédire les tendances des valeurs mesurées. Il s'agit d'une comparaison

entre la variance des erreurs et la variance des données à modéliser comme le montre l'équation suivante :

$$R^2 = 1 - \frac{\mathrm{var}(y - \tilde{y})}{\mathrm{var}(y)} \tag{2.4}$$

- **Le coefficient de corrélation *CC*** dit aussi **le coefficient de Pearson**, permet de mesurer la correction entre les prévisions et les mesures (Eq. 2.5). Les meilleures prévisions sont celles qui possèdent un *CC* proche de 1.

$$CC = \frac{\mathrm{cov}(y, \tilde{y})^2}{\mathrm{var}(y)\mathrm{var}(\tilde{y})} \tag{2.5}$$

où *cov* et *var* représentent respectivement la covariance et la variance.

$$STD = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(Y_t - \overline{Y_t})^2} \tag{2.6}$$

où $\overline{Y_t} = \frac{1}{N}\sum_{t=1}^{N} Y_t$ et $Y_t = y_t - \tilde{y}_t$.

- **Le Skill Score (SS)** : les métriques précédentes peuvent aider à évaluer la précision des prévisions, toutefois, elles ne peuvent pas fournir une mesure de la variabilité des données de séries temporelles. Pour intégrer le sens de la variabilité dans les mesures précédentes, Marquez and Coimbra [86] ont proposé une métrique qui peut compléter les indicateurs statistiques fondamentaux, à savoir la compétence de prévision ou Skill Score (SS), définie comme le rapport entre l'incertitude de la prévision solaire et la variabilité solaire. Ce Skill Score a été adapté par les auteurs [86] afin d'évaluer les performances du modèle de prévision par rapport au modèle de persistance utilisé comme référence. Le SS peut s'écrire en fonction du *RMSE* de la manière suivante :

$$SS = 1 - \frac{RMSE}{RMSE_{pers}} \tag{2.7}$$

où le $RMSE_{pers}$ est le *RMSE* entre les mesures réelles et les prévisions du modèle de référence de persistance. Plus la valeur de *SS* est proche de 1, plus le modèle de prévision est performant.

2. Techniques de prévision du rayonnement solaire

Des prévisions solaires fiables sur plusieurs horizons temporels sont nécessaires pour que les opérateurs de systèmes indépendants ou les autorités équivalentes d'équilibrage du réseau soient en mesure d'intégrer avec succès des niveaux accrus de production d'énergie solaire tout en maintenant la fiabilité du réseau.

Le choix de la méthode de prévision solaire dépend fortement des échelles de temps concernées, qui peuvent varier entre des horizons à très court terme (quelques minutes à l'avance), de court terme (quelques heures ou jours à l'avance), ou de moyen terme (plusieurs jours à l'avance). L'ensemble de données disponibles et les méthodologies sont également liés aux compétences en matière de prévision et aux choix des méthodes de prévision.

Selon [87,88], les techniques de prévision peuvent être classées en cinq catégories lorsqu'on intègre aussi les stations de mesures aux sols : 1) la prévision à l'aide d'un réseau de capteurs pyranométriques, 2) l'imagerie de ciel, 3) l'imagerie satellitaire, 4) les modèles de prévision numérique de temps et 5) les modèles statistiques. La Figure 2-1 représente une classification des méthodes de prévision en fonction de la résolution spatiale et l'horizon de prévision.

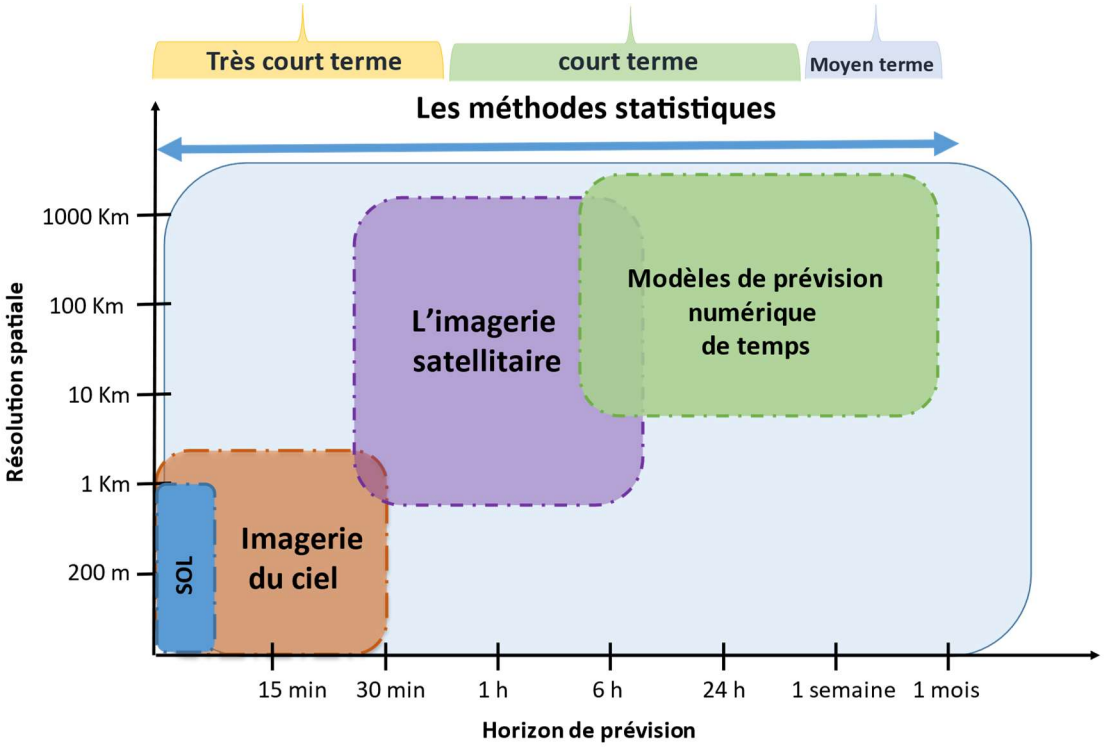


Figure 2-1. Les méthodes de prévision en fonction de la résolution spatiale et l'horizon de prévision.

2.1. Prévision numérique du temps

Les modèles de prévision numérique du temps (PNT) constituent un outil puissant pour la prévision du GHI à court et à moyen terme. Les PNT sont utilisés pour prévoir l'état de l'atmosphère pour plusieurs jours à venir. Les PNT sont généralement basés sur la physique de l'atmosphère et des équations décrivant l'état de l'atmosphère à partir d'un état initial observé. L'un des avantages majeurs des modèles numériques de prévision réside dans le fait qu'ils prévoient les variations transitoires des nuages, qui sont le principal modulateur de l'irradiation solaire au sol. Les PNT peuvent être classés en modèles globaux et en modèles à méso-échelle. Les variables de sortie sont fournies avec une résolution temporelle d'une heure pour les modèles régionaux et de trois à six heures pour les modèles globaux, alors que leur résolution spatiale se situe aujourd'hui entre 16 et 50 km. Pour les modèles à méso-échelle, les services

météorologiques exploitent généralement des modèles à méso-échelle avec une résolution spatiale comprise entre 5 et 20 km [89]. Parmi les modèles globaux, on trouve Global Forecast System (GFS) de National Oceanic and Atmospheric Administration (NOAA) des États-Unis et le Centre européen pour les prévisions météorologiques à moyen terme (ECMWF) [90]. Parmi les modèles méso-échelle, nous citons le MM5 [91] qui a été développé à l'Université d'État de Pennsylvanie, puis au National Center for Atmospheric Research (NCAR), et le Weather Research Forecast (WRF) qui est conçu pour être un modèle flexible pour la prévision du rayonnement solaire [92] .

Les modèles PNT sont généralement dédiés aux horizons de prévision au-delà de 6 heures. Ces méthodes peuvent fournir une bonne précision, mais dépendent principalement de la stabilité des conditions météorologiques. Pour les prévisions intra-horaires, seules quelques études ont été établies. Cependant, pour des prévisions au-delà d'une heure, de nombreux auteurs ont utilisé des prévisions à partir des PNT comme entrées de leurs modèles de prévision.

Heinemann et al. [93] ont montré que le modèle à méso-échelle MM5 peut prédire le GHI par ciel clair avec un MBE nul. Toutefois, le biais dépendait fortement des conditions nuageuses; par temps couvert, le MBE du modèle MM5 était de 129 %.

Perez et al. [94] ont évalué la précision du National Digital Forecast Database (NDFD) après l'application d'une fonction de correction locale. Les auteurs ont constaté que pour des horizons de prévision de 8 à 26 heures, le RMSE du GHI horaire est de 38 %.

Remund et al. [95] ont analysé les erreurs du ECMWF et GFS pour la prévision du GHI sur un seul site. Les auteurs ont obtenu un MBE de 19% pour les deux modèles et il est presque constant pour des horizons de prévision allant d'une heure jusqu'à trois jours à l'avance. De leur côté, Breitkreuz et al. [33] ont introduit les effets des aérosols sur le GHI dans le modèle ECMWF. Ils ont remarqué une réduction du rRMSE de 11.5% à 7.2% et du rMBE de 9.8 % à 5.1% pour la prévision du GHI horaire de 2 à 3 jours sous des conditions de ciel clair.

Lorenz et al. [97] ont exploité les prévisions d'ECMWF après avoir analysé différentes techniques d'interpolation spatiale et temporelle pour ramener les prévisions d'irradiance de 3 heures à 1 heure afin de prévoir la production PV pour un ensemble d'installations résidentielles. Les résultats moyennés jusqu'à 3 jours à l'avance et pour un site unique donnent un nRMSE de 13%. Ce dernier est réduit à 9% lorsque l'ensemble de 11 centrales est utilisé et encore inférieur à 5% pour le cas de 1400 centrales. Les auteurs ont conclu que l'augmentation de la précision des prévisions dépend fortement de la taille de la région étudiée. Dans une autre étude, Lorenz et al. [98] ont évalué plusieurs prévisions de GHI basées sur les PNT en Europe. Dans l'ensemble, les résultats obtenus indiquent des valeurs de rRMSE d'environ 40% pour l'Europe centrale et d'environ 30% pour l'Espagne.

Remund et al. [95] ont évalué différentes prévisions du GHI basées sur les NWP aux États-Unis. Ils ont obtenu des valeurs rRMSE allant de 20 à 40 % pour un horizon de prévision de 24 heures. Dans le même travail [95], les auteurs ont examiné les biais d'ECMWF et GFS par rapport à un site unique. Ils ont constaté que pour les prévisions du GHI pour le jour suivant, les modèles ont un MBE de 19 %. Les auteurs ont conclu que cette erreur moyenne est

approximativement constante pour les horizons de prévision intra-journaliers à savoir d'une heure jusqu'à 3 jours à l'avance.

Mathiesen et Kleissl [89] ont validé les prévisions de GHI à moyen terme de trois PNT, à savoir le modèle nord-américain (NAM), du GFS et du ECMWF pour les États-Unis continentaux en utilisant les données de mesure au sol. Pour les conditions de ciel clair, tous les modèles ont un MBE de moins de 50 W/m^2. Alors que, pour des conditions nuageuses, les biais dépassent 200 W/m^2. Par la suite, les auteurs ont appliqué une correction des statistiques de sorties de modèle via l'utilisation des MBE de chaque modèle en fonction de l'indice de ciel clair et l'angle solaire zénithal pour le développement de la fonction de correction. La correction a réduit le RMSE de 20 W/m^2, et 17W/m^2 pour le NAM et le ECMWF, respectivement. L'ECMWF fournissait la prévision la plus précise dans des conditions nuageuses, tandis que le GFS avait la meilleure prévision par ciel clair.

Dans une autre étude, Mathiesen et al. [99] ont déterminé un lien entre la précision des prévisions du modèle NAM et l'emplacement de la couverture nuageuse pour les données horaires en Californie. Les auteurs ont trouvé que les prévisions côtières estivales sont fortement biaisées avec un MBE >125 W/m^2. En revanche, pour les endroits internes où les conditions nuageuses sont moins probables, les prévisions NAM sont moins biaisées.

Cependant, selon [100], les modèles numériques sont limités par la non-linéarité des équations du domaine ainsi que par la résolution spatiale insuffisante de la grille d'intégration, de quelques km à 100 km, ce qui est trop large par rapport à la taille de la centrale PV. À cause de leurs résolutions grossières, les PNT ne permettent pas une cartographie précise des caractéristiques à petite échelle. Fanego et al. [92] ont comparé les prévisions intra-journalières de GHI horaire du modèle WRF à 3 km pour quatre sites de mesure au sol dans le sud de l'Espagne. Ils ont ontenu un MBE compris entre 49 W/m^2 et 64 W/m^2 et un RMSE entre 136 W/m^2 et 170 W/m^2. Les auteurs ont conclu que la sous-estimation de la couverture nuageuse démontre les limites des PNT actuelles pour la prévision de l'irradiance solaire.

Parmi les majeures limitations des PNT, on trouve que la précision des modèles numériques dépend de la disponibilité des enregistrements météorologiques. Ils ne sont performants que lorsqu'ils sont appliqués à des tâches de prévision à court terme. De plus, la mise en œuvre de ces modèles physiques est relativement difficile en général, car ils nécessitent un certain nombre de paramètres et des équipements coûteux qui ne sont pas toujours disponibles dans de nombreuses régions du monde. Une autre limitation réside dans leurs résolutions spatiales grossières, qui ne permettent pas d'évaluer la variabilité de la couverture nuageuse en fonction du temps afin de pouvoir prédire la position exacte de champs de nuages [100]. Bien que la résolution spatiale des PNT est améliorée ces dernières années, la gamme des résolutions se situe toujours entre 16 et 50 km selon les modèles, ce qui nuit à la précision des prévisions. Même si la résolution spatiale est plus fine, les intervalles de sortie temporels ne permettraient pas d'évaluer la variabilité de la couverture nuageuse en fonction du temps, élément important pour prévoir les taux de rampe et les plages de variabilité des centrales solaires [87].

2.2. Imagerie satellitaire

Un autre type de modèles météorologiques utilisés pour la prévision de l'énergie solaire est basé sur l'analyse des images capturées par les satellites géostationnaires. Ces satellites offrent grâce à leurs vastes résolutions spatiales et temporelles des informations continues sur la réflectance du sol et l'état de l'atmosphère. La télédétection par satellite offre un outil puissant pour estimer l'irradiance solaire à la surface (ISS) pour des horizons allant de 30 min à 6 h, car les satellites possèdent la capacité à saisir la distribution spatiale et l'évolution dynamique des nuages. En tant que tel, il est reconnu depuis longtemps que la télédétection par satellite est un outil puissant pour dériver l'ISS régionale et même globale que la modélisation numérique [101]. La base de cette méthode repose sur la détermination des structures nuageuses durant les étapes temporelles enregistrées précédemment. L'extrapolation de leur mouvement conduit à une prévision de la position des nuages et, par conséquent, de la situation locale du rayonnement. Les satellites géostationnaires MSG sont des satellites géostationnaires couvrant l'Europe, l'Afrique, l'océan Atlantique, et une partie de l'Asie et l'Amérique de sud. Ces satellites sont placés en orbite à 0° de longitude et de latitude, ils sont équipés de plusieurs capteurs de rayonnement couvrant des bandes spectrales spécifiques du rayonnement solaire et infrarouge [102].

De nombreuses études comparatives ont renforcé l'idée que la précision de l'ISS dérivée par les produits satellitaires est supérieure à celle des produits de réanalyse [103,104]. Les études menées dans les années 1990 ont révélé que les estimations de l'ISS peuvent être obtenues à 28 W/m^2 près par rapport au RMSE ou encore mieux à l'échelle mensuelle [105,106]. Des études récentes ont également démontré que l'incertitude de l'ISS dérivée du satellite pourrait être du même ordre que l'incertitude des observations au sol [107,108]. La majorité des divergences entre l'ISS dérivée par satellite et celle mesurée au sol sont imputables à des défauts d'instrumentation. Par conséquent, l'estimation à partir de satellites variés est prometteuse pour évaluer les conditions régionales de rayonnement de surface. En particulier, dans les régions aux environnements difficiles qui empêchent d'effectuer des mesures du rayonnement au sol. Les modèles de la récupération de l'ISS à partir d'images satellitaires peuvent être divisés en deux groupes. Le premier groupe correspond aux méthodes empiriques et semi-empiriques basées sur l'indice de nuage telle que la méthode Heliosat [109–111]. Le deuxième groupe correspond aux approches physiques basées sur la modélisation explicite du transfert radiatif [109,112].

Plusieurs services fournissent des données de GHI dérivées à partir des images satellitaires à l'échelle mondiale sont disponibles [113,114]. Les services HelioClim-3 (HC3) et Copernicus Atmosphere Monitoring Service (CAMS) sont les ensembles de données dérivées de satellites les plus largement utilisés et couvrent l'Europe, l'Afrique et l'Asie [115]. Cette base de données est celle que nous avons utilisée durant cette thèse. HC3 fournit des données de GHI même en intra-horaire en utilisant la méthode Heliosat-2 pour le traitement des images du satellite MSG pour estimer les valeurs de GHI. Les données de HC3 ont été validées dans plusieurs régions

climatiques avec des données au sol de haute qualité où de bonnes corrélations ont été obtenues [25,116–119].

Heliosat-2 est une méthode qui a été largement utilisée pour convertir les images satellitaires en GHI [120]. Heliosat-2 combine un modèle de ciel clair avec un « indice de nuage ». Ce dernier est quantifié par l'amplitude de changement de l'irradiance sur un pixel entre la réflectance qui devrait être observée sous un ciel sans nuages et celle actuellement observée. Météosat a une résolution spatiale qui varie de 3 km à plus de 12 km sur les bords de la planète avec une résolution temporelle de 15 min. Comme représenté sur la Figure 2-2, le Maroc possède une résolution spatiale de Météosat comprise entre 3 km et 4 km. Par conséquent, nous avons validé les données de GHI de la version 5 de HC3 (HC3v5) pour modéliser le GHI pour pouvoir les utilisées dans les sites où des mesures au sol ne sont pas disponibles.

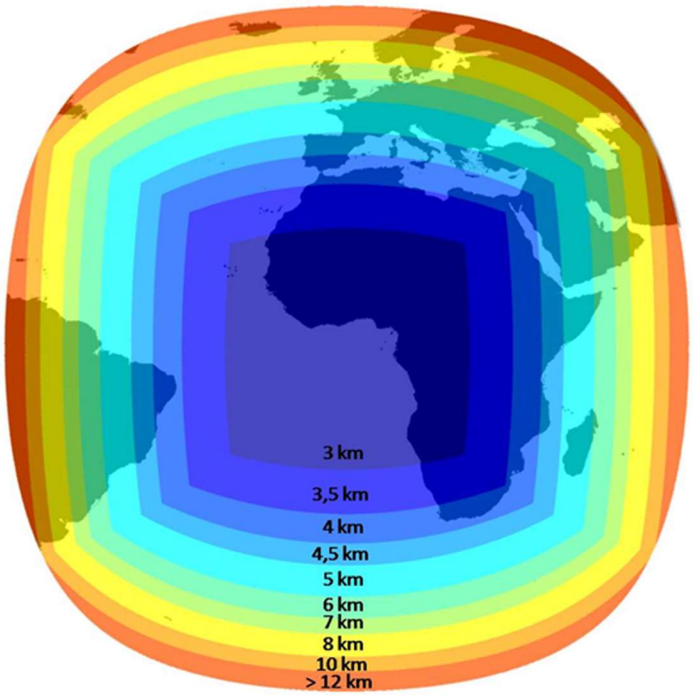


Figure 2-2. Résolution spatiale de Météosat. Source MINES ParisTech 2009.

L'utilisation directe de l'imagerie satellitaire est moins répandue pour les prévisions solaires en raison des résolutions temporelles et spatiales disponibles [101]. En revanche, les méthodes statistiques telles que les techniques de régression ou d'intelligence artificielle combinées avec des données satellitaires ont été largement utilisées pour prévoir le GHI à différents horizons temporels.

Şenkal et Kuleli [121] ont utilisé des données satellitaires et un modèle basé sur les réseaux de neurones artificiels (RNA) pour l'estimation du rayonnement solaire en Turquie. Les auteurs ont prouvé que l'utilisation de RNA et les données satellitaires sont des moyens peu coûteux et efficaces pour estimer le rayonnement solaire et construire une base de données solaires lorsque des données mesurées sur sol ne sont pas disponibles. Qin et al. [122] ont utilisé les RNA pour estimer le GHI en fonctions de plusieurs produits de télédétection à partir de deux bases de

données satellitaires, tels que la température, l'indice de végétation, et les précipitations. Les résultats de la validation indiquent que la méthode suggérée peut estimer le GHI mensuel à une résolution spatiale d'environ 5 km avec une grande précision.
Marquez et al. [123] ont développé une nouvelle méthode hybride qui combine des informations provenant des images satellitaires traitées avec des RNA pour prévoir le GHI à des horizons temporels de 30, 60, 90 et 120 minutes pour deux emplacements géographiques. La méthode proposée surpasse le modèle de persistance d'environ 5-19% pour les prévisions à un pas de temps et d'environ 10-25% lorsque les prévisions sont à plusieurs pas de temps.
Aguiar et al. [124] ont développé un modèle basé sur les RNA pour la prévision du GHI intrajournalier de 1 à 6 heures. Les auteurs ont utilisé des données au sol, des données de GHI du satellite à partir de Helioclim3, et les prévisions de GHI et de la couverture nuageuse fournies par l'ECMWF. Les auteurs ont montré que la combinaison des données de satellite et d'ECMWF avec les données au sol peut améliorer la performance de la prévision.

2.3. Imagerie du ciel

Les nuages sont les paramètres les plus importants qui influencent l'irradiation solaire. Les images satellitaires et la prévision numérique de temps sont utilisées pour caractériser les nuages et détecter leur variabilité. Cependant, ces deux techniques ne permettent pas de fournir des informations précises sur les fluctuations à haute fréquence de l'irradiation solaire [125]. Pour combler cette lacune, l'imagerie de ciel à partir du sol se présente comme une alternative fournissant des informations très détaillées sur l'étendue, la structure et le mouvement des nuages existants au moment où la prévision sera effectuée. Pour des horizons à très court terme, l'imagerie du ciel a montré son efficacité et peut apporter une contribution essentielle à la prévision solaire grâce à l'arrivée de caméras numériques de sécurité peu coûteuses et de capacités de traitement graphique plus robustes [126].

L'imageur de ciel est une caméra numérique installée au sol et orientée vers le ciel qui capture des images de l'ensemble du ciel à des intervalles de temps programmés. Les imageurs de ciel sont utilisés pour la détection des nuages, la mesure de la hauteur des nuages au-dessus du sol et la détermination de leurs mouvements [127]. Les imageurs du ciel permettent d'obtenir une vue hémisphérique du ciel.

Les prévisions basées sur l'imagerie de ciel se font en quatre étapes [78] :

- l'acquisition des images du ciel enregistrées par la caméra,
- le traitement des images pour la reconnaissance des nuages,
- l'estimation du vecteur de mouvements des nuages,
- et finalement l'utilisation des données des vecteurs de mouvement des nuages pour générer des prévisions de la couverture nuageuse et du rayonnement solaire.

Chow et al. [125] ont présenté une méthode basée sur un imageur de ciel pour la prévision immédiate des nuages et du GHI à l'horizon de cinq minutes et au-dessous du kilomètre à San Diego. Les auteurs ont utilisé des algorithmes de traitement d'image pour déterminer la

couverture nuageuse du ciel. Par la suite, à partir d'une carte de nuages bidimensionnelle générée par la transformation des coordonnées de la couverture du ciel, les ombres des nuages à la surface ont été estimées. Une validation limitée sur quatre jours partiellement nuageux a montré que les conditions nuageuses étaient correctement prévues dans 70 % des cas pour un réseau de six stations pyranométriques au sol réparti sur une zone de 2 km^2. Autrement, l'erreur de prévision des nuages augmentait avec l'horizon de prévision en raison de la grande variabilité de la couverture nuageuse sur le site côtier. Les auteurs ont conclu que l'utilisation de l'imagerie du ciel pour évaluer la ressource solaire pour les applications d'énergie solaire présente un fort potentiel pour augmenter la résolution spatiale et temporelle fournie par les satellites et les méthodes basées sur les PNT.
Urquhart et al. [128] ont utilisé des images du ciel dérivées de deux imageurs pour la prévision de la production PV d'une grande centrale solaire pour des horizons allant de 30 s à 15 minutes. Les résultats de validation ont montré que le MBE, le MAE (moyenne absolue de l'erreur) et le RMSE sont plus importants que la persistance. Les auteurs ont annoncé que même si les imageurs de ciel permettent de surveiller les conditions du ciel, ils présentaient des lacunes qui limitaient leur efficacité pour la prévision de l'énergie solaire.

Avec le développement des techniques de l'apprentissage profond pour le traitement des images, plusieurs auteurs ont opté à l'utilisation des CNN (Réseaux de neurones convolutifs) pour le traitement des images des caméras de ciel pour extraire les caractéristiques des images au lieu de se baser sur les vecteurs de mouvement des nuages.
De nombreux chercheurs ont essayé de corréler les images du ciel avec l'irradiance solaire en utilisant des techniques de l'apprentissage automatique au lieu du traitement classique des images. Fu et al. [129] ont proposé d'extraire les caractéristiques pertinentes des images à partir de l'apprentissage automatique au lieu du traitement des images pour extraire les champs de vecteurs de mouvement des nuages. Les auteurs ont utilisé les caractéristiques extraites pour l'apprentissage du modèle de prévision pour des horizons variant entre 5 et 15 minutes. Les résultats ont démontré l'intérêt de l'apprentissage pour la conversion d'information entre les images et le GHI. D'autres études sur l'utilisation des CNN pour le traitement des images du ciel sont énoncées dans la section 2.5.4.
Cependant, les approches actuelles basées sur les images de ciel ne tiennent pas compte du développement et de la dispersion des nuages ni des changements significatifs de leur géométrie. L'extrapolation des configurations nuageuses est également limitée à l'échelle spatiale définie par le champ de vision de l'imageur du ciel. Il est possible d'étendre l'échelle spatiale en utilisant plusieurs imageurs à différents endroits. Les couches de nuages multiples avec des vecteurs de mouvement caractéristiques différents peuvent également poser un problème puisque les nuages des niveaux supérieurs peuvent être partiellement obscurcis par les nuages des niveaux inférieurs. Le temps réel d'anticipation pour lequel cette méthode est efficace dépendra, entre autres, de la vitesse et de la hauteur des nuages. Pour les nuages bas et rapides, l'horizon de prévision ne peut être que de 3 minutes, tandis que pour les nuages élevés et lents, il peut arriver jusqu'à 30 minutes, mais en général, les horizons se situent entre 5 et 30 minutes. Même si la taille et la vitesse des nuages pouvaient être déterminées avec précision,

l'exactitude des prévisions dépend de la vitesse à laquelle le champ de nuages s'écarte de l'évolution définie par les vecteurs de mouvement des nuages (c'est-à-dire développement, dissipation, etc.).

2.4. Méthodes de séries temporelles

Les méthodes de séries temporelles sont des techniques statistiques basées sur l'analyse des séries chronologiques, qui permettent de prédire les valeurs futures du GHI en utilisant les valeurs passées de la variable ou la relation entre cette variable et d'autres paramètres météorologiques [100,130]. La modélisation et la prévision des séries temporelles constituent un domaine de recherche en pleine expansion qui a attiré l'attention des chercheurs dans de nombreux domaines différents : affaires, économie, ingénierie, et sciences physiques et environnementales [131]. L'objectif principal de la modélisation des séries chronologiques est de collecter et d'examiner avec précision les observations passées d'une série temporelle afin de développer un modèle approprié qui décrit la structure de cette série chronologique. Le modèle est ensuite utilisé pour générer des prévisions futures. Les méthodes de prédiction basées sur des séries chronologiques historiques peuvent inclure des méthodes statistiques telles que les modèles ARMA (autorégressif à moyenne mobile), ARIMA (autorégressif à moyenne mobile intégrée), SARIMA (autorégressif à moyenne mobile saisonnière), ETS (lissage exponentiel), et autres [132–135]. Dans le domaine de l'énergie, l'analyse des séries chronologiques pour la prédiction de l'irradiance solaire a été largement utilisée. Les séries temporelles sont définies comme une séquence d'observations mesurées dans le temps, avec des pas d'échantillonnage horaires, quotidiens, hebdomadaires ou mensuels. Un exemple d'une série temporelle de GHI mesuré au sol au pas horaire pour un an est représenté par la Figure 2-3.

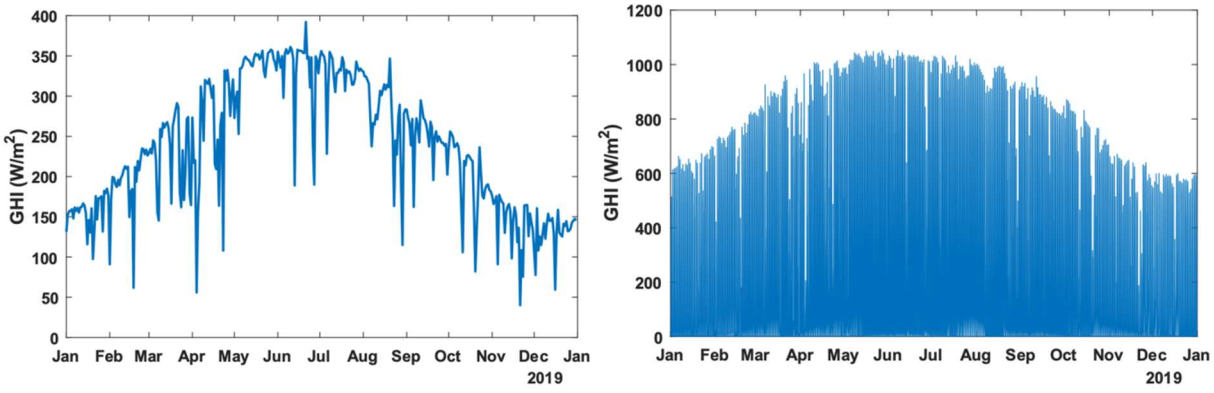


Figure 2-3. Exemples de séries temporelles de GHI au pas journalier (gauche), et au pas horaire (droite). Station météorologique de GEP, Benguerir, Maroc.

Le principe des méthodes basées sur les séries temporelles consiste à prédire l'irradiance ou la production PV solaire future à différentes échelles de temps sur la base des données observées dans le passé. Dans la partie suivante, nous passons en revue les modèles statistiques utilisés lors de développement de notre approche pour la prévision du GHI, à savoir les modèles SARIMA, le MLP (Perceptron multicouche), le LSTM (Long Short-Term Memory), et le CNN.

2.4.1. Persistance

Il est intéressant de vérifier la précision d'un modèle de prévision par rapport à un modèle de référence. Le modèle de persistance ou le prédicteur naïf est le modèle de référence largement utilisé dans la communauté de prévision solaire. Le modèle de persistance suppose que le GHI à l'instant $t+h$ est le même que le GHI à l'instant t. Par exemple, le GHI du lendemain sera le même que celui d'hier, et l'heure suivante sera similaire à l'heure actuelle. L'utilisation du modèle de persistance comme modèle de référence réside dans le fait qu'il est possible d'obtenir une bonne précision avec un modèle facile à mettre en œuvre, ne nécessitant pas un historique de données important et ayant un coût de calcul plus faible. Le modèle de persistance pour la prévision de GHI est représenté par l'équation [87]:

$$\widetilde{GHI}_{t+h} = GHI_t \tag{2.8}$$

où h est l'horizon de prévision.

2.4.2. SARIMA

L'un des modèles statistiques classiques de prévision est le modèle ARIMA qui est un modèle de prévision de séries chronologiques proposé au début des années 70 par Box et Jenkins [136]. Les modèles basés sur ARIMA ont excellé dans de nombreux domaines de prévision, tels que l'économie, l'ingénierie, le transport ou les études environnementales [131]. Le modèle SARIMA ou ARIMA saisonnier est un modèle de série chronologique autorégressive. Il s'agit d'une extension du modèle ARIMA qui combine un processus autorégressif AR (p) et une moyenne mobile MA (q). Le modèle AR exprime l'observation actuelle en fonction des comportements antérieurs d'un processus donné. Le modèle MA représente les données actuelles par rapport à la valeur antérieure des erreurs du processus sous la forme d'une régression linéaire.

Lors de la modélisation de séries chronologiques avec une composante saisonnière, le modèle est appelé SARIMA $(p,d,q)\ (P,D,Q)_s$, où (p,d,q) est la partie non saisonnière du modèle, (P,D,Q) est la partie saisonnière du modèle et s est la saisonnalité du modèle. Le modèle général de SARIMA $(p,d,q)\ (P,D,Q)_s$ peut être représenté par l'équation suivante :

$$\phi_p(B)\phi_p(B^s)(1-B)^d(1-B^s)Y_t = \theta_q(B)\Theta_q(B^s)e^t, \tag{2.9}$$

où Y_t est la valeur de la série chronologique observée à l'instant t, e^t est le résidu à l'instant t, B est l'opérateur de décalage $B^jY_t = Y_{t-j}$, s est la longueur de la période saisonnière, p,q sont l'ordre régulier des modèles AR et MA respectivement, P,Q sont les ordres saisonniers des modèles AR et MA, d est le nombre de différences régulières, D est le nombre de différences saisonnières, $\phi_p(B) = 1-\phi_1B-\phi_2B^2-\cdots-\phi_pB^p$ et $\theta_q(B) = 1-\theta_1B-\theta_2B^2-\cdots-\theta_qB^q$ sont les polynômes standards des modèles AR et MA respectivement.

$\Phi_p(B) = 1 - \Phi_1 B - \Phi_2 B^2 - \cdots - \Phi_p B^p$ et $\Theta_q(B) = 1 - \Theta_1 B - \Theta_2 B^2 - \cdots - \Theta_q B^q$ sont les polynômes saisonniers des modèles AR et MA, respectivement.

La procédure générale de l'utilisation de SARIMA pour la prévision des séries temporelles consiste à vérifier la stationnarité des données, à estimer les paramètres du modèle et les ajuster jusqu'à obtenir le modèle adéquat, et finalement à utiliser le modèle pour la prévision de GHI. La procédure générale de prévision avec SARIMA est illustrée par la Figure 2-4.

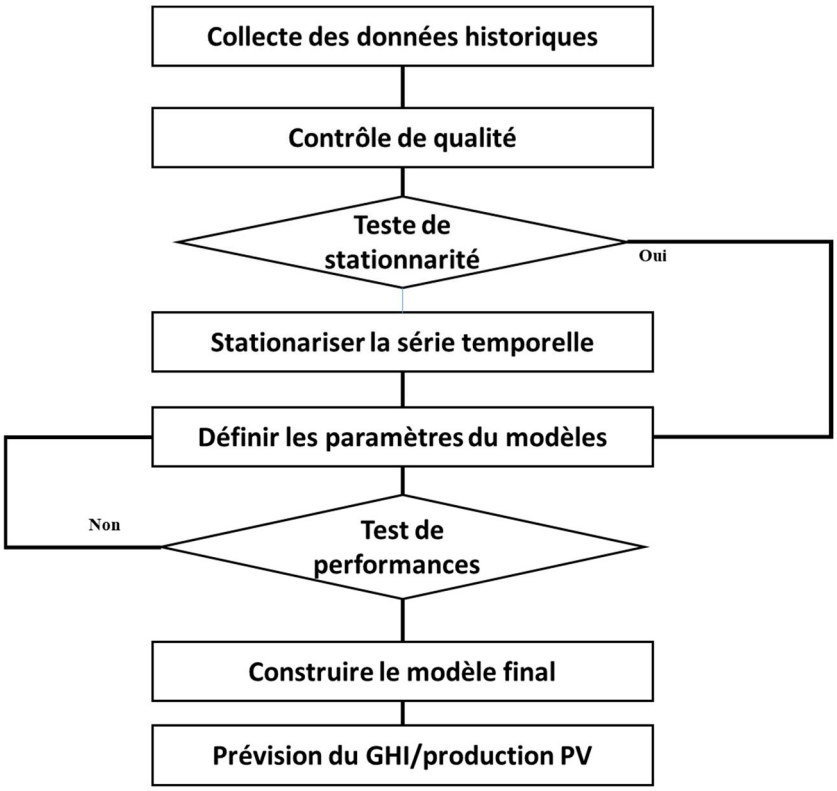


Figure 2-4. Procédure générale du modèle SARIMA.

Les modèles autorégressifs ont été la référence des méthodes de prédiction des séries temporelles pendant de nombreuses années [78]. Selon Box et al. [136], ces techniques nécessitent une certaine expertise, étant donné que ces modèles exigent la sélection correcte de l'ordre des entrées retardées afin d'identifier l'ordre adéquat du modèle. Une autre limitation réside dans le fait que dans de nombreux systèmes complexes, les propriétés de linéarité et même de stationnarité de la série chronologique analysée ne sont pas garanties. Pourtant, compte tenu de leur simplicité, les modèles autorégressifs ont été largement utilisés pour la mise en œuvre des systèmes de prévision de la ressource solaire.

Reikard [137] a comparé le modèle ARIMA et d'autres méthodes tels que les RNA, la régression, le modèle à fonction de transfert, les modèles à composantes non observées et le modèle hybride pour prédire l'irradiance solaire 24 heures à l'avance en utilisant le GHI de trois endroits aux États-Unis. Il affirma que pour les résolutions de 60, 30 et 15 minutes, le modèle ARIMA surpasse les autres méthodes. Pedro et Coimbra [138] ont évalué cinq techniques :

ARIMA, k-plus proches voisins (k-NN), RNA, et RNA optimisés par des algorithmes génétiques pour prédire la puissance d'une centrale photovoltaïque 1 heure et 2 heures à l'avance. De leur part, ils ont prouvé que les modèles RNA été plus performant que les modèles de persistance, ARIMA et k-NN.

Agoua et al. [139] ont construit une méthode statistique basée sur ARIMA pour prévoir la production électrique de quelques minutes jusqu'à 6 heures à l'avance. Arès la vérification de la stationnarité des données, les auteurs ont remarqué que le modèle proposé a obtenu une précision supérieure de 20 % comparé à un modèle de persistance, à la forêt aléatoire et au modèle autorégressif simple (AR).

Kardakos et al. [140] ont proposé une méthode basée sur SARIMA, implémentée avec les prévisions du GHI dérivées d'un modèle de prévision météorologique numérique de temps, et un modèle de RNA avec des entrées multiples, pour prédire la production PV horaire d'une centrale solaire pour l'horizon intrajournalier et journalier. Les résultats montrent que les valeurs moyennes de nRMSE sur une base annuelle étaient de 11.12 % pour SARIMA, 11.26% pour le modèle RNA, et 13.71 % pour le modèle de persistance.

Yang et al. [141] ont proposé des méthodes de prévision de GHI pour l'heure suivante reposées sur l'analyse des séries temporelles. Les modèles utilisent différents types d'entrées, à savoir le GHI, le DNI, le DHI et la couverture nuageuse. La première méthode utilise directement le GHI historique pour prévoir le GHI de l'heure suivante par une décomposition saisonnière additive suivie d'un modèle ARIMA. La deuxième méthode prévoit séparément le DHI et le DNI à l'aide d'une décomposition saisonnière additive suivie d'un modèle ARIMA, puis combine les deux prévisions pour prévoir le GHI à l'aide d'un modèle atmosphérique. La troisième méthode tient compte des effets de la couverture nuageuse où un modèle ARIMA est utilisé pour prévoir les passages des nuages. Les résultats ont montré que l'intégration de l'information sur la couverture nuageuse améliore la précision des prévisions, où un RMSE minimum de 29.73 W/m^2 a été atteint avec le troisième modèle.

2.5. Techniques de l'intelligence artificielle

L'intelligence artificielle (IA) n'est pas une science d'aujourd'hui, en effet depuis l'apparition des ordinateurs, les humains ont tenté la création d'une machine capable d'imiter le comportement humain. Cette quête de la création d'une machine capable d'égaler ou de dépasser l'intelligence humaine relève du domaine de l'IA. Les techniques de l'IA ont commencé à se déployer au début des années 1950. Herbert Simon a constaté que les ordinateurs pouvaient simuler les mécanismes de prise de décision et reproduire le processus de la pensée humaine lorsqu'il a découvert qu'une imprimante utilisait des symboles pour produire des images [142]. Les trois chercheurs, Herbert Simon, Allen Newell et Cliff Shaw ont rassemblé leur expertise pour développer un modèle mathématique et générer le code permettant de reproduire ce que faisait l'imprimante [143]. Pareil à la capacité humaine à apprendre de ses expériences passées, pour la machine, ces expériences passées sont les données. Alors l'apprentissage automatique découle de cette question : un ordinateur pourrait-il aller au-delà de ce qu'on l'ordonne de faire et apprendre par lui-même comment effectuer une tâche spécifiée ? Un ordinateur pourrait-il

nous surprendre ? Plutôt que de manuellement façonner des règles de traitement de données, un ordinateur pourrait-il automatiquement apprendre ces règles en observant les données ? Ces questions ont ouvert la porte à un nouveau paradigme de programmation. Dans la programmation classique, le paradigme de l'IA symbolique, les humains entrent des règles (un programme) et des données à traiter selon ces règles pour avoir des réponses. Avec l'apprentissage automatique, les humains introduisent des données ainsi que les réponses attendues des données, et soutirent les règles (Figure 2-5). Ces règles peuvent ensuite être appliquées à de nouvelles données pour produire des réponses originales. De ce fait, un système d'apprentissage automatique est entraîné plutôt que programmé explicitement.

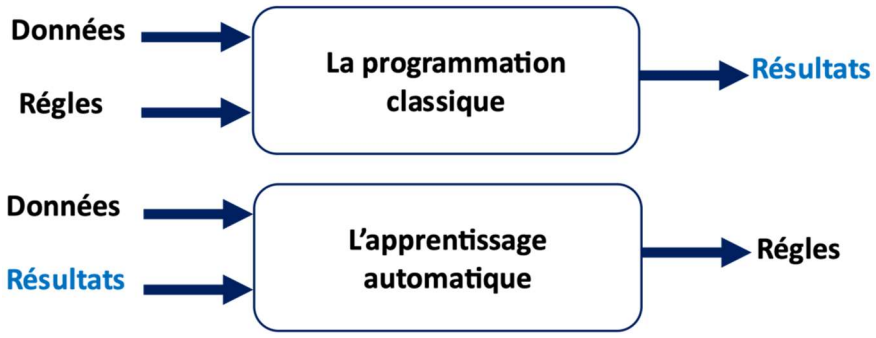


Figure 2-5. La différence entre la programmation classique et l'apprentissage automatique.

De nos jours, les techniques de l'IA ont révélé performantes dans plusieurs domaines et applications, tels que les sciences, la médecine, la comptabilité, la finance, l'économie, l'énergie et autres. L'IA a été utilisée dans plusieurs taches telles que le traitement du langage, la reconnaissance vocale et faciale, les prévisions boursières, la cybersécurité, les applications des Smartphones, les assistants numériques personnels, les infrastructures connectées, l'analyse des données, les voitures autonomes, la détection des maladies. L'IA a aussi été utilisée pour la lutte contre les pandémies, qu'il s'agit de diagnostiquer, de détecter les infections et prévoir leurs évolutions et d'accélérer d'autres aspects de la recherche médicale [143–145]. Les k-NN, SVM (Machine à vecteurs de support), SVR (Régression par vecteur de support), FL (logique floue), et les RNA sont des exemples des techniques de l'apprentissage automatique.

En plus des modèles autorégressifs, les techniques basées sur l'IA ont été largement utilisées dans le domaine de la prévision solaire. De nombreuses études concluent que les modèles d'apprentissage automatique sont meilleurs que les modèles statistiques pour prédire l'irradiation solaire. Certaines études ont favorisé l'application des techniques d'apprentissage automatique par rapport aux techniques régressives classiques. Mellit et al. [146] ont comparé un modèle d'apprentissage automatique à la régression linéaire multiple et aux méthodes analytiques pour la prévision de la production PV. Les auteurs ont montré que le modèle RNA ne requiert pas assez de variables ou de calculs complexes, contrairement aux modèles implicites. Darbellay et al. [147] ont favorisé les RNA par rapport aux modèles de prévision ARIMA standards pour la prévision de la charge électrique.

2.5.1. Réseaux de neurones artificiels

Il a été constaté que les méthodes classiques de la prévision des séries temporelles telles que les approches de Box-Jenkins sont insuffisantes pour représenter le comportement non linéaire entre les entrées et les sorties [148]. Cependant, des études ont montré le pouvoir des techniques de l'intelligence artificielle d'approximer avec pertinence n'importe quelle fonction continue. En particulier, les RNA ont reçu une attention particulière parmi les approches de l'IA de la part de la communauté scientifique après leurs succès dans la prévision financière [149]. Ils constituent un outil puissant pour l'estimation de fonctions, le regroupement et la classification de données, la détection d'anomalies, la reconnaissance vocale et la génération audio, l'analyse de séries temporelles, la traduction automatique, le traitement d'images et de nombreuses autres applications [144,145]. Les RNA regroupent plusieurs méthodes telles que les perceptrons multicouches (MLP), les réseaux de neurones convolutifs (CNN), les réseaux d'ondelettes (WNN), les réseaux de neurones récurrents (RNN), la quantification vectorielle d'apprentissage (LVQ), etc. Les RNA sont largement utilisés dans diverses applications en raison de leurs excellentes propriétés d'autoapprentissage, d'adaptabilité, de tolérance aux pannes, de non-linéarité et d'avancement dans la mise en correspondance des entrées et des sorties [150].

Les RNA sont un sous-domaine de l'apprentissage automatique dont le concept de fonctionnement est basé sur les neurones biologiques du cerveau humain. Dans le corps humain, les neurones sont des cellules responsables de la transmission de l'information à d'autres cellules nerveuses, musculaires et glandulaires. Les neurones reçoivent les signaux d'entrées à partir des dendrites qui peuvent également agir comme des sorties interconnectant les neurones, par la suite les neurones communiquent entre eux par des jonctions qui se trouvent à la fin des axones appelés synapses faisons l'effet de liaison et de pondération entre les neurones et permettent donc aux neurones de communiquer entre eux. Quant à l'axone est le fil conducteur qui conduit le signal de sortie du corps cellulaire vers d'autres neurones. L'axone a un potentiel électrique, s'il est excité au-delà d'un seuil, il transmet un signal électrique. Lorsque l'entrée électrique d'une synapse atteint un seuil, elle transmet le signal à la dendrite à laquelle elle est connectée [144]. Le neurone artificiel est un modèle mathématique développé pour modéliser le neurone biologique, qui consiste en des éléments de traitement appelés neurones et des connexions entre eux avec des coefficients appelés poids.

Le neurone artificiel a été développé dans le but de modéliser le neurone humain. Basé sur la représentation du neurone artificiel de la Figure 2-6 (droite). Les connexions d'entrées $(X_1, X_2, \dots, X_n)$ sont ajustés en fonction des poids associés $(w_1, w_2, \dots, w_n)$ et un biais ajouté (b), par la suite le signal d'entrée net agrégé du neurone est calculé à l'aide d'une fonction d'entrée qui est généralement la fonction somme $u = \sum_{i=1}^{n} X_i w_i + b$ où X et w sont les vecteurs d'entrée et de poids correspondants. Dès que la valeur d'entrée est calculée, une fonction d'activation f qui détermine le degré d'activation d'un neurone et transforme l'entrée du neurone en sa sortie qui représente l'entrée de la couche suivante est appliquée. La fonction d'activation fait référence à l'ensemble des fonctions de transfert utilisées pour atteindre la sortie souhaitée. Parmi les fonctions d'activation couramment utilisées, on trouve les fonctions d'activation

linéaire, la Heaviside (step), sigmoïde, tangente hyperbolique, unité linéaire rectifiée (ReLU), etc. Le Tableau 2-2 représente les fonctions d'activation largement utilisées. Par la suite, la valeur du signal de sortie émise par la sortie du neurone est calculée à l'aide d'une fonction de sortie $y = f(u)$.

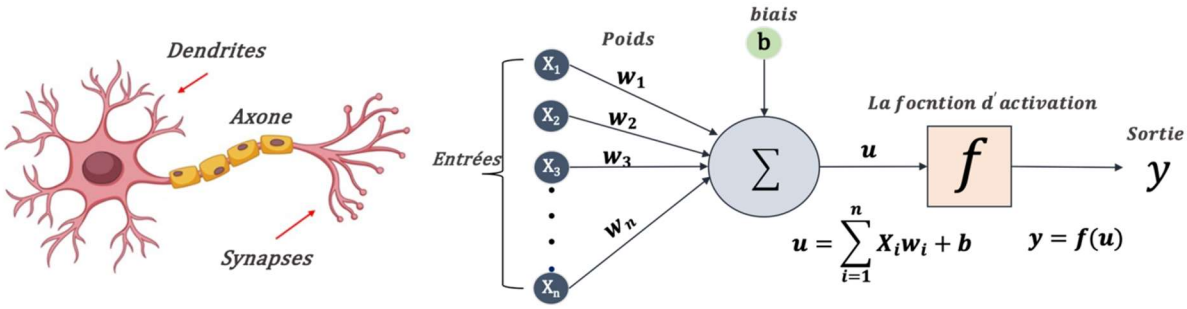
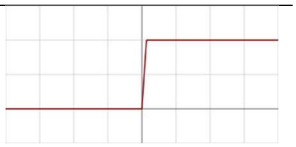
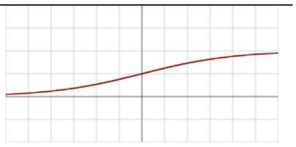
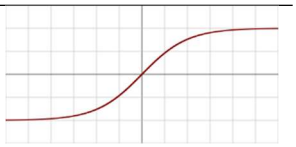
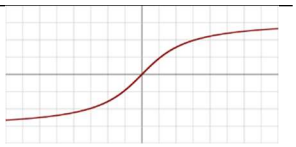
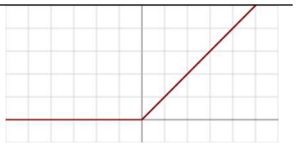


Figure 2-6. (Gauche) schéma simplifié d'un neurone biologique, (droite) neurone artificiel avec des entrées $X_1, X_2, \ldots, X_n$ pondérées par $w_1, w_2, \ldots, w_n$ et un biais b, une fonction d'entrée u, et une fonction d'activation f permettant de former la sortie y.

Tableau 2-2. Exemples des fonctions d'activation utilisées dans les réseaux de neurones artificiels.

Fonction	Équation	Graphe
Linéaire	$f(x) = x$	
Heaviside	$f(x) = \begin{cases} 0 & si \quad x < 0 \\ 1 & si \quad x \geq 0 \end{cases}$	
Sigmoïde	$f(x) = \frac{1}{1 + e^{-x}}$	
Tangente hyperbolique (tanh)	$f(x) = \frac{2}{1 + e^{-2x}} - 1$	
Arc tangente (arctan)	$f(x) = tan^{-1}(x)$	
Unité linéaire rectifiée (ReLU)	$f(x) = \begin{cases} 0 & si \quad x < 0 \\ x & si \quad x \geq 0 \end{cases}$	

2.5.2. MLP

Un certain nombre de problèmes complexes issus de différents domaines de recherche ne peuvent pas être résolus à l'aide d'un réseau de neurones à une seule couche à cause de la relation compliquée entre les différentes variables d'entrées et de sortie. Le MLP est un complément du réseau neuronal à action directe et est un estimateur qui peut être utilisé pour la classification et la régression. Le MLP est considéré parmi les réseaux perceptron les plus fréquemment utilisés pour les tâches d'approximation de fonction non linéaire [151]. Le MLP se compose de trois types de couches : une couche d'entrée, une ou plusieurs couches cachées, et la couche de sortie comme le montre la Figure 2-7. La couche d'entrée reçoit des valeurs de variables d'entrées, les couches cachées sont des couches intermédiaires qui effectuent tous les calculs et extraient les caractéristiques cachées dans les données. La couche de sortie prend les données des couches cachées précédentes et aboutit à une prédiction finale basée sur les apprentissages du modèle.

Pour obtenir des prédictions exactes, il est nécessaire d'entraîner le modèle pour apprendre les poids exacts. Un perceptron apprend comme suit : prends les données d'entrée insérées dans la couche d'entrée du perceptron, chaque variable d'entrée possède un poids qui est multiplié par chaque variable, puis la valeur pondérée est additionnée à l'aide d'une fonction de sommation simple. Ensuite, les résultats sont envoyés à la fonction d'activation, et le résultat final est la sortie du réseau. Les poids et les biais sont mis à jour par des passages itératifs en arrière à l'aide d'un algorithme de rétropropagation [152].

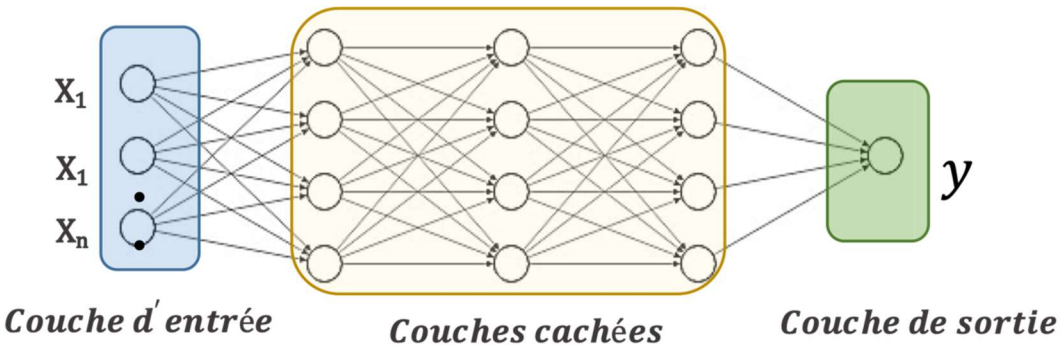


Figure 2-7. Représentation schématique d'un MLP avec plusieurs couches cachées.

Le MLP est le type de réseau de neurones le plus populaire pour la prévision solaire. Dans cette partie, nous passons en revue les études basées sur le MLP pour la prévision de GHI ou de la production PV les plus récentes.

Chow et al. [153] ont développé un modèle basé sur le MLP pour la prédiction en temps réel de la puissance photovoltaïque 10 à 20 minutes à l'avance. Les auteurs ont constaté que la performance des prévisions à court terme est comparable à celle des prédictions en temps réel lorsque des angles solaires sont utilisés pour les prévisions.
Renno et al. [154] ont développé un modèle MLP pour prédire le GHI en utilisant un ensemble de paramètres climatiques, astronomiques et radiométriques. Les meilleures configurations de

MLP ont été testées pour prédire le GHI. Les performances en termes de pourcentage d'erreur, RMSE et R^2 sont respectivement de 4.57%, 160.3 Wh/m^2 et 0.99.

El Alani et al. [155] ont développé un modèle MLP pour prédire le GHI un jour à l'avance en utilisant une base de données de mesures radiométriques collectées à partir d'une station météorologique au sol au Maroc. Plusieurs architectures ont été testées, et un modèle final avec la température, l'angle solaire, et le rayonnement au top de l'atmosphère a été expérimenté pour prédire le GHI pour des jours clairs et nuageux. Les MBE et les RMSE normalisés obtenus sont respectivement de -0.051%, 0.10% pour les jours clairs, et 0.14%, 0.39% pour les jours nuageux, respectivement.

Mellit et al. [156] ont proposé une architecture basée sur le MLP pour prédire le GHI 24 heures à l'avance pour une centrale en Italie. Les résultats de la validation ont démontré que les coefficients de corrélation variaient entre 98 % et 99% pour les jours ensoleillés et entre 94% et 96% pour les jours nuageux. Dans une autre étude, Mellit et al. [146] ont développé deux modèles basés sur le MLP pour prévoir la puissance des modules PV de 50 W_c pour l'horizon d'une heure à l'avance en Turquie. Le premier modèle prédisait la puissance PV pour les jours nuageux, pour ces types de jours, le RMSE est inférieur à 0.2% et le R^2 varie entre 93% et 97%. Le deuxième modèle est dédié à la prévision de la puissance PV pour les jours ensoleillés, pour ces jours, le RMSE est encore inférieur à 0.2% alors que le R^2 varie entre 96% et 97%.

Notton et al. [157] ont testé plusieurs architectures basées sur le MLP avec plusieurs combinaisons d'entrée pour prédire le GTI 10 minutes à l'avance en France. Les résultats de validation ont montré que pour la meilleure architecture le rRMSE été de 9%.

Izgi et al. [158] ont développé un MLP pour prévoir la puissance solaire d'un panneau solaire PV de 750 W pour des horizons entre 5 et 60 minutes à l'avance en Turquie. Les résultats de validation ont montré que la meilleure prévision de la puissance a été observée pour les horizons de prévision de 5 min et 35 min, respectivement, en avril. Cependant, les meilleures prévisions ont été étendues à 3 min et 40 min en août.

Kaushika et al. [159] ont appliqué un MLP pour prévoir le GHI, le DNI et le DHI une heure à l'avance en Inde. Les résultats ont montré que le modèle développé présente une excellente compatibilité avec des valeurs RMSE et MBE normales de 5.19% et -0.914%, respectivement, pour le GHI. Les auteurs ont reporté que le modèle se révèle utile pour prévoir les composantes du rayonnement solaire à un emplacement quelconque.

Gutierrez Corea, et al. [160] ont effectué plusieurs expériences avec différentes architectures et paramètres de MLP afin de déterminer lesquels généraient les meilleures prédictions de GHI pour la partie nord de la péninsule Ibérique. L'horizon de prévision était de 1 à 6 heures et les données d'une station météorologique voisine ont été utilisées comme entrées. Les résultats de l'étude ont permis de générer des modèles qui prédisent le GHI avec des taux d'erreur inférieurs à 20% en termes de nRMSE.

Voyant et al. [161] ont testé deux architectures de MLP pour prévoir le GHI 24 heures à l'avance en France. Les auteurs ont trouvé que le modèle de MLP utilisant des données d'entrée endogènes et exogènes est le plus pertinent. Les comparaisons avec ARMA et la persistance

ont montré une réduction du nRMSE de 1.3 % et 7.8% par rapport aux deux modèles, respectivement.

2.5.3. LSTM

Les réseaux de neurones récurrents (RNN) sont une classe de réseau de neurones artificiels caractérisé par une mémoire interne, ce qui le rend parfaitement adapté aux problèmes d'apprentissage impliquant des données séquentielles. Comme les réseaux de neurones à action directe, les réseaux de neurones récurrents utilisent des données d'apprentissage pour apprendre la corrélation entre l'entrée et la sortie. L'idée principale des RNN est d'utiliser non seulement les entrées, mais aussi les résultats précédents pour faire la prédiction actuelle. Ils se distinguent par leur "mémoire" qui prend les informations des entrées précédentes. Les réseaux de neurones à action directe ne contiennent aucune mémoire et ne considèrent que l'entrée actuelle, de plus l'information se disperse dans une seule direction ; de la couche d'entrée vers la couche de sortie à travers les couches cachées. Dans un RNN, les informations se propagent en boucle, la décision est prise en considérant l'entrée actuelle et également de ce qui a été appris des entrées précédentes (Figure 2-8). Par conséquent, la sortie du RNN à tout moment dépend de deux paramètres, à savoir le signal de rétroaction au pas de temps précédent et les signaux d'entrée.

Les auteurs ont affirmé que les RNN standards ne peuvent pas être utilisés pour résoudre des problèmes nécessitant un apprentissage de séquences profondes à long terme. Ceci est dû au problème de la disparition ou l'explosion du gradient qui décroit de façon exponentielle avec le temps. Même le cerveau humain est incapable de se souvenir de chaque évènement du passé [64]. Les RNN et leurs variantes ont été largement utilisés dans de nombreuses applications caractérisées par la dépendance temporelle des données. Parmi les applications notables des RNN, citons la traduction machine [103–105], la modélisation et la prédiction du langage [165–168], la reconnaissance vocale [169–171], résumé de texte et génération des rapports [172], la reconnaissance de l'écriture manuscrite [173,174], la reconnaissance et la caractérisation d'images[175,176], et pour l'analyse prédictive [177,178].

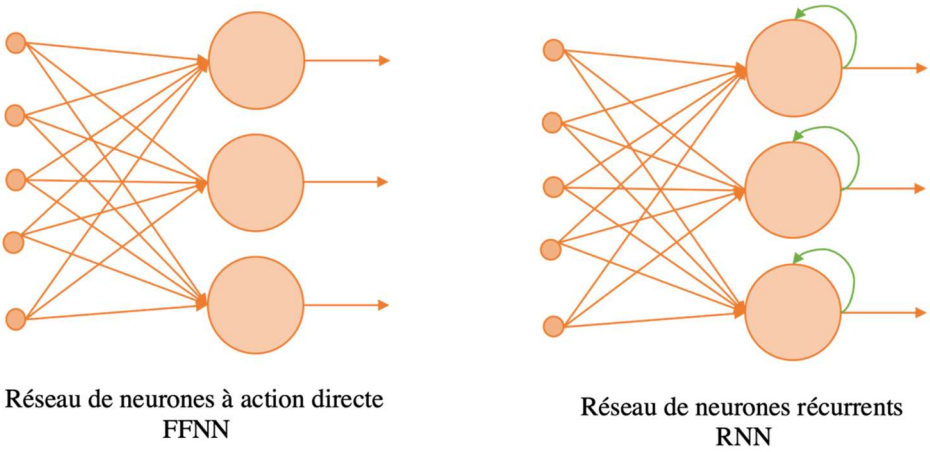


Figure 2-8. La différence entre le réseau de neurones à action directe et le réseau de neurones récurrents.

En 1997 Hochreiter & Schmidhuber ont introduit le LSTM pour pallier les inconvénients des RNN standards [179]. Le LSTM résout le problème du gradient en incorporant une série de cellules de mémoire y compris la porte d'entrée, la porte d'oubli, et la porte de sortie. Grâce à cette mémoire, le LSTM peut conserver des informations pour une longue période. Cela imite, une fois encore, la manière dont les humains se souviennent sélectivement des détails du passé. Le LSTM et ses adaptations ont montré leur potentiel pour résoudre une série de problèmes impliquant un apprentissage séquentiel. Similaire aux RNN standards les LSTM ont été employés dans de nombreuses applications d'apprentissage de séquences, notamment dans le domaine du traitement du langage naturel [182,184]. Cependant, LSTM n'a pas été utilisé pour la prédiction des séries temporelles autant que pour des problèmes avec des données séquentielles [185]. Pour la prévision du GHI, nous nous attendons à ce que le LSTM exploite la dépendance temporelle et spatiale des données, ce qui représente un atout majeur dans l'utilisation des informations contextuelles.

Les réseaux LSTM sont composés comme la plupart des architectures de réseaux neuronaux d'une couche d'entrée, d'une ou plusieurs couches cachées et d'une couche de sortie. La principale caractéristique des réseaux LSTM est contenue dans là où les couches cachées constituées de cellules mémoires. Chacune de ces cellules de mémoire possède trois portes qui maintiennent et ajustent l'état de la cellule (c_t): la porte d'oubli (f_t), la porte d'entrée (i_t) et la porte de sortie (o_t) [186]. La structure des cellules de mémoire est illustrée à la Figure 2-9. À chaque instant *t*, les trois portes sont présentées avec l'entrée x_t et la sortie h_{t-1} des cellules de mémoire à $t-1.i_t, f_t, o_t, c_t, et\ \tilde{c}_t$ sont la porte d'entrée, la porte d'oubli, la porte de sortie, l'état de la cellule et la valeur candidate à chaque instant *t*.

- La porte d'oubli f_t est un élément clé de l'architecture du LSTM qui définit les informations supprimées de l'état de la cellule. Elle fonctionne avec une fonction d'activation de type sigmoïde dure σ qui a une plage de 0 (tout oublier) à 1 (tout se rappeler) et elle est définie par l'équation (2.9).

$$f_t = \sigma(w_{f,x}x_t + w_{f,h}h_{t-1} + b_f) \tag{2.10}$$

- La porte d'entrée i_t permet d'appliquer dans un premier temps une fonction sigmoïde pour décider les valeurs à mettre à jour dans l'état de la cellule c_t, par la suite, elle utilise une fonction tangente hyperbolique pour créer le vecteur candidat $\tilde{c}_t$ contenant les nouvelles valeurs pour la cellule mémoire.

$$i_t = \sigma(w_{i,x}x_t + w_{i,h}h_{t-1} + b_i) \tag{2.11}$$

$$\tilde{c}_t = \tanh(w_{\tilde{c},x}x_t + w_{\tilde{c},h}h_{t-1} + b_{\tilde{c}}) \tag{2.12}$$

Les nouveaux états des cellules sont calculés sur la base des résultats des deux étapes précédentes, ◦ désignant le produit Hadamard (par éléments) :

$$c_t = f_t \circ c_{t-1} + i_t \circ \tilde{c}_t \tag{2.13}$$

- La porte de sortie o_t permet de produire une sortie h_t dépendant de l'état de la cellule c_t. Cette sortie sera filtrée par une couche sigmoïde σ qui sélectionne quelle information de l'état de la cellule sera utilisé en sortie.

$$o_t = \sigma(w_{o,x} x_t + w_{o,h} h_{t-1} + b_o) \tag{2.14}$$

$$h_t = o_t \circ \tanh(c_t) \tag{2.15}$$

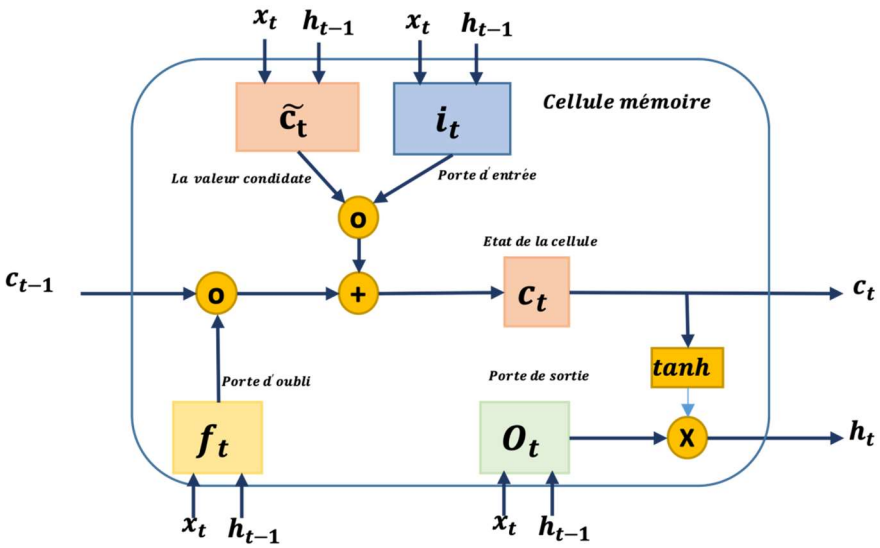


Figure 2-9. La cellule mémoire du LSTM.

La fonction d'activation sigmoïde σ et la fonction tangente hyperbolique $tanh$ sont définies comme suit :

$$\sigma(x) = \frac{1}{1+e^{-x}} \ , \tag{2.16}$$

$$\tanh(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \ , \tag{2.17}$$

avec:

- x_t est le vecteur d'entrée à l'instant t.
- $w_{f,x}, w_{f,h}, w_{\tilde{c},x}, w_{\tilde{c},h}, w_{i,x}, w_{i,h}, w_{o,x}, w_{o,h}$ sont les matrices de poids.
- $b_f, b_i, b_f, b_{\tilde{c}}$ et b_0 sont les vecteurs de biais.
- f_t, i_t, o_t sont la porte d'entrée, la porte d'oubli, la porte de sortie.
- $c_t, \tilde{c}_t$ sont des vecteurs pour les états des cellules et les valeurs candidates.
- h_t est le vecteur de sortie de la couche LSTM.

Lors du traitement d'une séquence d'entrée, ses caractéristiques sont présentées à chaque instant au réseau LSTM. L'entrée à chaque pas de temps *t* est traitée par le réseau, comme le montrent les équations ci-dessus. Une fois que le dernier élément de la séquence est traité, la sortie finale de la séquence entière est renvoyée. Les poids et les termes de biais sont ajustés pendant la phase d'apprentissage.

Ces dernières années, le LSTM a été appliqué pour la prédiction de l'irradiance solaire, et les résultats ont prouvé que les modèles basés sur le LSTM étaient aussi précis ainsi que les autres méthodes de prédiction.
Srivastava et Lessmann [187] ont comparé les prévisions de LSTM à d'autres méthodes de prévision pour prédire le GHI un jour à l'avance. Les auteurs ont utilisé des données de télédétection pour 16 endroits en Europe et 5 aux États-Unis. Les résultats de validation suggèrent que le modèle LSTM surpasse de nombreuses méthodes de 52.2 %.
Qing et Niu [188] ont utilisé des données de prévisions météorologiques pour prévoir le GHI horaire un jour à l'avance en se basant sur le réseau LSTM. Les auteurs ont démontré que dans le cas de l'utilisation de dix années de données historiques pour prédire une année de données de GHI, le RMSE de la prédiction à l'aide de l'algorithme LSTM proposé a diminué de 42.9 % par rapport au réseau neuronal à rétropropagation.
Brahma et Wadhvani [189] ont développé des modèles prédictifs basés sur des méthodologies d'apprentissage profond et des données provenant de plusieurs sites pour le GHI quotidien dans deux endroits en Inde. Les résultats indiquent que les modèles basés sur le LSTM peuvent être utilisés pour prédire les données d'irradiation quotidienne avec une bonne précision.
Récemment, Yu et al. [190] ont développé une approche basée sur le LSTM pour la prévision de GHI pour une heure et un jour à l'avance pour deux sites en Atlanta et Hawaii. Pour améliorer la précision des prévisions par temps nuageux, les auteurs ont introduit l'indice de clarté comme donnée d'entrée pour le modèle LSTM et pour la classification du type de jour. D'après les résultats des prévisions horaires, le R^2 du LSTM pour les jours nuageux et les jours mixtes dépasse 0.9, tandis que le R^2 du RNN simple n'est que de 0.70 et 0.79 pour des deux sites, respectivement. Pour les prévisions quotidiennes, le R^2 de tous les jours nuageux est d'environ 0.85. Les auteurs ont confirmé que le LSTM est très efficace et plus précis que les autres modèles RNN testés.
Huang et al. [191] ont testé divers modèles basés sur le LSTM pour la prévision du GHI une heure à l'avance. Les résultats expérimentaux ont montré que le modèle qui prend en compte les informations météorologiques de l'instant suivant présente les meilleures performances avec un RMSE de 62.16 W/m^2, et un SS de 0.44. Les auteurs ont reporté que l'algorithme basé sur le LSTM est plus précis de 19.19% que le réseau de neurones à rétropropagation en termes de RMSE.

2.5.4. CNN

Le CNN est un type particulier des réseaux de neurones traditionnels. Le CNN est une technologie qui associe les réseaux de neurones artificiels et les stratégies d'apprentissage profond les plus récentes. Dans le domaine de l'apprentissage profond, le réseau de neurones

convolutifs est au centre d'avancées surprenantes. Depuis sa première apparition au début des années 1980 [192], le CNN a démontré d'excellentes performances dans plusieurs tâches de reconnaissance d'images. Ces dernières années, avec le progrès de l'intelligence artificielle, le CNN a montré des performances prometteuses dans plusieurs tâches de vision par ordinateur et s'est imposé comme l'architecture de référence pour la reconnaissance d'images dans de nombreuses applications telles que les domaines de la médecine, de la santé, de l'automobile, de la sécurité et de l'énergie [193–196].

Les CNN sont spécialement conçus pour traiter les images présentées en entrée. Leur architecture est alors plus spécifique, les CNN standards sont composés de quatre couches, à savoir la couche de convolution, la couche non linéaire, la couche de mise en commun et les couches entièrement connectées. En général, dans les réseaux de neurones classiques, chaque neurone d'une couche est connecté à tous les neurones de la couche suivante, alors que chaque connexion est un paramètre du réseau. Cela peut donner lieu à un très grand nombre de paramètres. Au lieu d'utiliser des couches entièrement connectées, un CNN utilise une connectivité locale entre les neurones, c'est-à-dire qu'un neurone n'est connecté qu'aux neurones voisins de la couche suivante. Cela peut réduire considérablement le nombre total de paramètres dans le réseau [197].

La couche de convolution est l'élément clé des CNN. Son rôle est d'extraire des motifs trouvés dans les régions locales des images d'entrée. De manière générale, la couche de convolution reçoit en entrée plusieurs images, puis elle calcule la convolution de chacune d'entre elles avec chaque filtre. Ces filtres correspondent aux caractéristiques que l'on souhaite extraire de l'image. Contrairement aux RNA traditionnels, les caractéristiques ne sont pas prédéfinies, mais apprises par le réseau. Alors la force des CNN réside dans leurs capacités à déterminer les éléments discriminants d'une image. La couche de convolution génère à sa sortie une carte de caractéristique indiquant l'emplacement des caractéristiques de l'image. L'unité linéaire rectifiée généralement abrégée en ReLu vise à appliquer une fonction d'activation élémentaire telle que la sigmoïde à la sortie de l'activation produite par la couche précédente. La couche de mise en commun connue sous le nom de pooling est généralement placée entre deux couches de convolution. Les couches de mise en commun visent à réduire les dimensions des entrées tout en préservant autant d'informations possibles, et donc réduire le nombre de paramètres ainsi que la complexité de la procédure du modèle. La couche de pooling est capable d'introduire l'invariance spatiale dans le réseau, ce qui peut aider à améliorer la généralisation du modèle. La couche entièrement connectée n'est pas une caractéristique de CNN, mais elle constitue toujours la dernière couche d'un réseau de neurones. Les couches entièrement connectées sont généralement un type de réseau de neurones à action directe et suivent le principe de MLP. Ces couches reçoivent en entrée les sorties de la couche de convolution ou de mise en commun, qui est sous la forme d'un ensemble de métriques qui correspondent aux cartes de caractéristiques qui sont aplaties pour créer un vecteur, ensuite ce vecteur est introduit dans la couche pour générer la sortie finale du CNN comme le montre la Figure 2-10.

Pour la prévision solaire, les CNN ont démontré leurs supériorités pour le traitement des images satellitaires, les images de ciel, l'extraction des corrélations spatiales, et aussi pour extraire les caractéristiques de la couverture nuageuse [198].

Les images du ciel prises à proximité d'un système PV peuvent fournir des informations pertinentes pour la prévision de la production PV telles que la position du soleil, la transparence des nuages, le mouvement et la forme des nuages. Des informations pareilles sont difficiles à extraire avec des modèles explicites de traitement d'images. Le développement récent de l'apprentissage profond a permis aux CNN de montrer leur capacité d'apprendre la corrélation entre les images du ciel et la production PV.

Sun et al. [199] ont utilisé un réseau CNN pour prédire la production PV à partir des images du ciel en temps réel. Les auteurs ont prouvé que les images du ciel contiennent des informations abondantes sur le rayonnement solaire et que les CNN sont une structure appropriée pour interpréter les informations de ces images. En conséquence, leur modèle CNN atteint un RMSE relatif de 26.0 % à 30.2 % lorsqu'il est appliqué aux sorties de puissance de deux panneaux solaires. Zhang et al. [80] ont comparé trois modèles, MLP, CNN et LSTM, pour prévoir la production photovoltaïque à court terme à partir des images du ciel. L'évaluation des modèles de prévision a montré des améliorations de l'erreur par rapport à la persistance de 7% par rapport au le MLP, de 12% pour le CNN et de 21% pour le LSTM. Siddiqui et al. [200] ont appliqué des techniques d'apprentissage profond CNN-LSTM pour prédire le GHI d'une heure 1 à 4 à l'avance en traitant des vidéos du ciel. Les auteurs ont prouvé que l'approche d'apprentissage profond proposée réduit le pourcentage d'erreur à la fois pour la prévision immédiate et pour les 4 heures à venir par rapport à la prévision par satellite. Pour le même objectif, Paletta et Lasenby [201] ont utilisé des CNN pour prévoir le GHI à court terme à partir des images de caméra du ciel. L'étude a montré la performance des CNN avec un RMSE moyen de 40% pour un horizon de 10 minutes à l'avance. Al-lahham et al. [202] ont proposé une approche basée sur les CNN pour prévoir l'irradiation solaire à court terme à partir d'images du ciel. Les auteurs ont prouvé que leur approche permet d'obtenir des résultats compétitifs avec une complexité de calcul beaucoup plus faible pour les prévisions immédiates et les prévisions jusqu'à 4 heures à l'avance par rapport aux algorithmes traditionnels de traitement d'images.

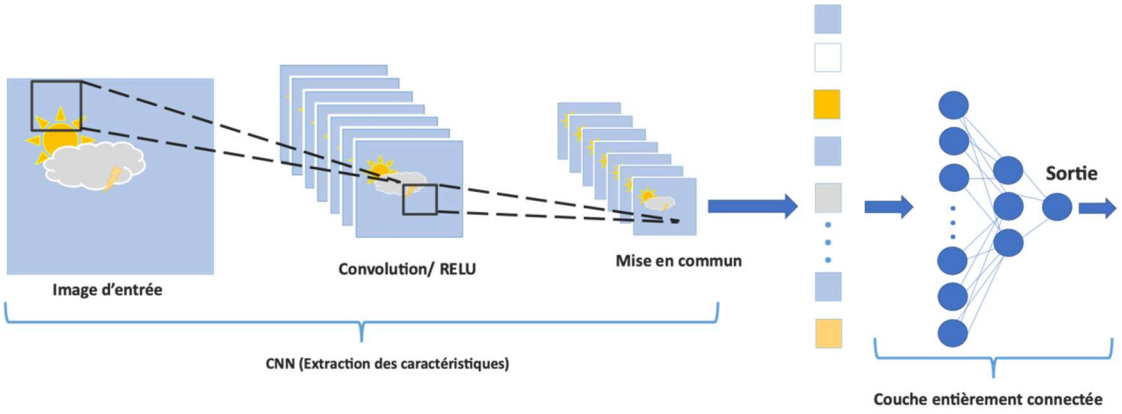


Figure 2-10. L'architecture générale d'un réseau CNN.

2.6. Modèles hybrides

Les modèles hybrides ou modèles d'ensembles sont le résultat de fusion de différentes méthodes de prévision. Les modèles hybrides ont été introduits pour pallier le manque d'utilisation d'un modèle individuel et explorer les possibilités de combinaison de divers algorithmes dans le but d'améliorer la précision des prévisions. Une autre motivation du système hybride est de réduire la complexité et le temps de calcul pour les applications de prévision en ligne. Les modèles hybrides peuvent combiner des modèles linéaires, des modèles non linéaires ou à la fois des modèles linéaires et non linéaires. Dans le domaine de la prévision solaire, de nombreuses études que nous allons citer dans la section suivante ont montré que les méthodes hybrides de prévision surpassent les prévisions individuelles [203].

Lu, et al. [204] ont évalué des techniques de l'apprentissage automatique combinées avec des PNT pour la prévision de la production PV pour les jours à venir. Leurs résultats ont montré que les performances du modèle hybride étaient supérieures aux modèles individuels avec une MAE inférieure à 30%.
Fernandez et al. [205] ont développé un système de prévision de la production horaire d'énergie électrique à court terme en combinant deux PNT à savoir le GFS et le MM5 et d'autres modèles tels qu'ARIMA, k-NN, et RNA. Les paramètres de sortie du premier modèle GFS sont définis comme entrées du second modèle MM5. Par la suite, les résultats sont injectés dans les autres méthodes de prévision. Les auteurs ont conclu que le modèle MLP-RNA s'est avéré le meilleur modèle de prévision parmi les autres. La valeur RMSE pour la puissance nominale de la centrale PV est de 11.79 % pour les échelles de temps de prévision de 16 et 39 heures.
Massidda et al. [206] ont prédit la production d'une centrale photovoltaïque un jour à l'avance en Allemagne, en combinant la régression multivariée et le modèle numérique de prévision GFS. Les auteurs ont prouvé d'après l'analyse de simulation, que les résultats été favorables, même si un nombre relativement faible d'échantillons et de caractéristiques a été pris en compte. Les erreurs de prévisions RMSE et MAE étaient respectivement de 177.8 kW et 125.9 kW.
Filipe et al. [207] ont développé un système hybride de prévision de l'énergie solaire en combinant une méthode statistique et un modèle numérique de prévision. Une comparaison du modèle hybride proposé avec deux modèles de références à savoir la persistance et le modèle diurne a montré une amélioration moyenne des prévisions de 57.4% et de 34.06% par rapport aux deux modèles de références, respectivement.
Zang et al. [208] ont appliqué un modèle hybride CNN-LSTM pour prévoir le GHI une heure à l'avance. Le modèle proposé applique d'abord un réseau CNN pour extraire les caractéristiques spatiales d'une matrice bidimensionnelle composée de paramètres météorologiques associés à un site cible et à ses sites voisins. Ensuite, un LSTM est appliqué pour extraire les caractéristiques temporelles des séries chronologiques historiques d'irradiation solaire associées au site cible. Finalement, les corrélations spatio-temporelles sont fusionnées pour prédire le GHI une heure à l'avance. Les résultats indiquent que le modèle proposé présente des avantages par rapport aux modèles de références et constitue une bonne alternative pour la prédiction du rayonnement solaire à court terme.

Bouzerdoum et al. [209] ont proposé un modèle hybride SARIMA-SVM pour prévoir la production PV à court terme de 5 minutes jusqu'à 1 heure à l'avance. Le modèle a été évalué à l'aide des données d'une installation PV de 20 kW et les performances ont été comparées aux prévisions générées par SARIMA et SVM indépendamment. Les résultats ont montré que le modèle hybride proposé était plus performant que les modèles simples avec un pourcentage d'erreur de 2.7%.

Yan et al. [210] ont développé un modèle hybride combinant le SVM, FL et un modèle de classification pour la prévision de la production d'une centrale PV de 5 kW un jour à l'avance. Le SVM a été utilisé pour l'apprentissage du modèle, et le modèle FL pour la sélection du meilleur modèle d'apprentissage généré par le SVM. Les résultats ont montré que le modèle proposé est plus performant que le modèle basé sur le RNA et le SVM individuellement.

Chu et al. [211] ont appliqué une méthode de reprévision basée sur les RNA pour optimiser trois modèles de prévision à savoir un modèle de suivi des nuages, un modèle ARIMA et un k-NN. La méthode est appliquée à des prévisions intra-horaire de la puissance d'une centrale photovoltaïque PV de 48 MWc. Les résultats ont montré que lorsqu'aucune optimisation n'était appliquée, la technique de suivi des nuages est la moins performante, avec des compétences de prévision de 71.3 % et 18.9 % pour 5 et 15 minutes. Cependant, lorsque l'optimisation a été effectuée, la méthode de suivi des nuages a surpassé la méthode ARIMA.

Wu et al. [212] ont utilisé un modèle hybride combinant les RNA, SVM, ARIMA et ANFIS (Système à inférence neuro-floue adaptatif) pour prévoir la production de trois centrales PV. Les résultats obtenus ont montré que le système hybride proposé était plus précis que les modèles individuels avec des nRMSE de 5.64 %, 3.43 % et 6.57 % pour les trois sites d'études, respectivement.

Cervone et al. [213] ont combiné un modèle basé sur les RNA et la méthode des analogues pour la prévision de la production PV de trois systèmes solaires 72 heures à l'avance. Les auteurs ont rapporté que le modèle hybride proposé est particulièrement adapté au calcul à grande échelle. Pour les trois systèmes étudiés, les valeurs de MAE sont de -1.85%, 0.38%, et -1.53%.

Dong et al. [214] ont combiné l'analyse des images satellitaires, le lissage exponentiel, et le MLP pour prévoir le GHI horaire. Les images satellitaires ont été analysées pour détecter le mouvement des nuages. Par la suite, à l'aide de l'indice de la couverture nuageuse dérivé à partir des images satellitaires, le modèle de lissage exponentiel a été mis en œuvre pour prévoir le prochain indice de couverture nuageuse. Finalement, le MLP a été appliqué pour dériver le GHI à partir de la prévision de l'indice de la couverture nuageuse. Comparée à d'autres modèles de séries chronologiques statistiques, la technique hybride s'avère plus performante que les modèles de prévision traditionnels.

Dolara et al. [215] ont proposé un système hybride, appelé Physical Hybridized Artificial Neural Network (PHANN), qui combine un modèle RNA avec un modèle physique du rayonnement par ciel clair pour générer des prévisions de puissance PV de 24 à 72 h à l'avance. Les résultats ont montré que le modèle hybride est capable de prévoir le profil quotidien de la puissance PV avec plus de précision que la méthode RNA avec un taux de diminution d'erreur qui atteint 5% pour les jours ensoleillés.

Wang et al. [216] ont proposé un nouveau modèle hybride déterministe basé sur la transformation en ondelettes (WT), le CNN et un modèle probabiliste combinant le modèle déterministe et la régression quantile (GR) pour la prévision de la puissance PV pour des horizons de 15 minutes à 2h à l'avance. L'évaluation des performances de modèle déterministe a montré que le modèle WT combiné au CNN présente une plus grande stabilité de la prévision. Alors que pour le modèle probabiliste, la précision s'est améliorée en moyenne de 49.10 % pour l'horizon de 15 minutes, de 54.08 % pour le 30 min, de 61.42 % pour 1 heure, et de 59.94 % pour 2 heures.

Gandelli et al. [217] ont prédit la puissance PV horaire d'une centrale solaire par la mise en œuvre d'un modèle hybride (PHANN) basé sur un modèle physique et un modèle RNA. Les résultats de validation ont indiqué que le modèle proposé est plus précis que les autres modèles RNA utilisés pour la comparaison. Le modèle hybride représente la précision la plus élevée pendant les jours ensoleillés avec une valeur normalisée MAE de 2.7 % alors que l'erreur augmente durant les jours nuageux.

Hussain et Al-Alili [218] ont proposé un modèle hybride combinant le filtre de Kalman et le MLP pour prédire le GHI un jour à l'avance. Le filtre de Kalman est utilisé pour ajuster les poids de MLP. Les résultats de validation ont montré la supériorité du modèle hybride sur les modèles de références à savoir le modèle non linéaire avec des entrées exogènes et la régression linéaire multiple (MLR) avec un RMSE de 269 Wh/m^2 et un R^2 de 95.97%.

Royer et al. [219] ont combiné un modèle basé sur le WT et un modèle de réseau de neurones artificiels pour prédire le GHI une heure à l'avance. Les auteurs ont souligné que le modèle hybride WT-RNA atteint de bonnes performances avec des RMSE variant entre 8.68 W/m^2 et 19.89 W/m^2 et un R^2 entre 99.65 % et 99.95% pour l'ensemble des régions étudiées.

Pour la prévision du GHI, une demi-heure à l'avance, Ghimire et al. [220] ont proposé un nouveau modèle hybride basé sur le CNN et le LSTM. Le CNN est utilisé pour extraire les caractéristiques des variables d'entrée, et le LSTM est utilisé pour générer les prévisions. Comparé aux méthodes individuelles, le modèle hybride présente les meilleures performances avec un rRMSE de 1.515 % et un pourcentage d'erreur de 4.67 %.

Dans une autre étude, Che et al.[221] ont combiné le modèle numérique de prévision WRF et le filtre de Kalman pour la prévision du GHI un jour à l'avance. Les résultats ont montré que la précision du modèle hybride surpasse celle du modèle WRF seul. En effet, l'amélioration des prévisions de l'approche combinée en termes de RMSE est de 42.2% et 26.8%, et de 95.7% et 90.9% en termes de MBE pour les deux sites d'étude, respectivement.

Voyant et al. [222] ont testé les modèles MLP, ARMA et la persistance afin de trouver la meilleure combinaison pour développer un modèle hybride dans le but de prévoir le GHI une heure à l'avance. Après plusieurs tests d'hybridation, les auteurs ont déduit que la combinaison MLP-ARMA-persistance génère les meilleures prévisions avec un nRMSE horaire de 36.59%.

Gairaa et al. [150] ont aussi proposé un autre modèle ARMA-MLP pour la prédiction du GHI pour l'horizon journalier. Le modèle ARMA est utilisé pour modéliser la composante linéaire du rayonnement solaire global et le modèle ANN pour la modélisation de la composante non

linéaire. Les résultats obtenus montrent que l'hybridation proposée surpasse le modèle autonome avec un nRMSE de 0.119% et 0.298% pour les deux sites d'études.
Shamshirband et al. [224] ont proposé une hybridation du modèle auto-régressif avec des entrées exogènes (ARX) avec un modèle RNA pour la prédiction journalière du GHI. Le modèle ARX-RNA surpasse les autres modèles de références avec un RMSE de 0.694 Mj/m^2 et un R^2 de 96.31%.
Dans le but de réduire les erreurs des autres modèles hybrides proposés dans la littérature pour la prévision du rayonnement solaire, Huang et al. [225] ont développé un modèle nommé WPD-CNN-LSTM-MLP pour la prévision de l'irradiance solaire 1 heure à l'avance. Le modèle hybride combine la décomposition en paquets d'ondelettes (WPD), le CNN, LSTM, et le MLP. Les auteurs ont souligné que le modèle développé surmonte les lacunes des modèles traditionnels et obtient des résultats de prévision plus précis. La comparaison de ce modèle avec plusieurs modèles de références individuels et hybrides révèle que le modèle WPD-CNN-LSTM-MLP proposé présente une meilleure précision de prévision de l'irradiance horaire avec un RMSE minimum de 32.1 W/m^2.

D'après l'analyse des résultats de divers modèles hybrides proposés dans la littérature, nous pourrions tirer conclusion que les modèles hybrides atteignent des performances élevées par rapport aux modèles individuels. D'autres combinaisons d'hybridation sont encore possibles pour améliorer la précision des prévisions du rayonnement solaire et enrichir les recherches dans cette thématique.

Conclusion

Dans ce chapitre, nous avons passé en revue les diverses méthodes de prévision du rayonnement solaire et de la production PV. Toutes les techniques que l'on vient de citer vont nous servir pour construire nos modèles de prévision. Les performances de diverses méthodes de prévision récentes ont été discutées et comparées à différents facteurs, tels que la précision, la fiabilité, et l'horizon.

Des prévisions précises du rayonnement solaire présentent un défi et des approches distinctes impliquant les modèles numériques de prévision du temps, l'imagerie satellitaire, l'imagerie du ciel, les méthodes de séries temporelles, des méthodes d'apprentissage automatiques, ou encore hybride ont été évaluées. Pour la conception efficace d'un modèle de prévision, les horizons de prévision, l'estimation des performances du modèle, la sélection des caractéristiques pertinentes, le contrôle de la qualité des données, la classification météorologique, le choix adéquat de l'architecture du réseau de neurones doivent être pris en compte. En outre, chaque groupe de ces méthodes présente des avantages et des limitations qui doivent être considérées lors de l'implémentation du modèle de prévision. Les modèles numériques de prévision sont pertinents pour prédire l'irradiance à des endroits où aucune donnée n'est disponible. En raison de leurs capacités de la modélisation et la prédiction à grande échelle de l'état de l'atmosphère, les PNT présentent une approche pertinente pour les prévisions à court et moyen terme avec des horizons temporels allant de 6 heures à 15 jours ou plus. Pour le court terme, l'imagerie

satellitaire constitue l'approche la plus adéquate pour des horizons allant de 30 min à 6 heures. Les images satellitaires se présentent comme une alternative aux coûteuses stations météorologiques au sol en fournissant directement des prévisions de l'irradiance solaire ou bien des prévisions de variables météorologiques telles que la température, la vitesse du vent et l'indice de couverture nuageuse qui seront utilisées en conjonction avec d'autres modèles pour la prévision de l'irradiance solaire. Cependant, l'utilisation des PNT et de l'imagerie satellitaire pour la prévision précise du rayonnement solaire est limitée par leurs résolutions grossières et leurs incapacités de détecter les nuages qui se disparaissent rapidement et qui sont considérés la cause principale derrière les fluctuations de la puissance produite. En plus, le succès des modèles de prévision basés sur ces deux modèles dépend de la précision de leurs prévisions météorologiques qui peuvent être imprécises pour tout emplacement de centrale solaire. Une autre limitation de ces méthodes réside dans la complexité des calculs qui nécessitent des connaissances solides dans le domaine de la métrologie et l'énergétique pour l'application et l'adaptation des modèles météorologiques pour chaque emplacement, ce qui est difficile à mettre en œuvre par des prévisionnistes sans une compréhension approfondie de la métrologie. Ce qui leur impose de supposer que ces modèles sont assez précis et utiliser leurs sorties pour faire des prévisions. Pour surmonter les lacunes des modèles numériques et satellitaires concernant la détection précise des nuages, l'imagerie du ciel a été largement utilisée. Cette imagerie de ciel permet de fournir des informations détaillées à haute fréquence sur la couverture nuageuse et d'autres conditions météorologiques locales qui sont utiles pour la prévision du GHI à très court terme de 3 à 30 minutes à l'avance. Toutefois, même avec les imageurs de ciel, les prévisions à très court terme ne sont pas toujours précises à cause de la dispersion rapide des nuages. Cependant, les prévisions à court terme de 6, 48, et 72 heures à l'avance sont les plus abordées dans la littérature. La raison de cet intérêt est que les prévisions à court terme permettent aux centrales photovoltaïques et aux gestionnaires d'énergies le temps de planification et de la prise de décisions soit pour la maintenance ou bien pour l'exploitation de l'énergie produite.

Pour les prévisions à court terme, les méthodes statistiques sont utilisées dans la plupart des cas. Pour des études qui présentent un pauvre potentiel au niveau de la diversité des variables météorologiques avec une archive raisonnable sur le GHI ou la production PV, les méthodes autorégressives telles que AR, MA, ARIMA, SARIMA sont amplement utilisés pour la prévision directe de la variable concernée. Ces méthodes ne nécessitent pas la disponibilité de données météorologiques, mais elles se basent principalement sur la corrélation entre les valeurs passées et futures de la variable étudiée. Cependant, ces modèles autorégressifs s'intéressent à l'analyse de la série temporelle de la variable cible GHI ou la puissance de sortie PV. Ils ne prennent pas en considération les changements des conditions météorologiques, ceci réduit leurs capacités à prédire avec précision le rayonnement solaire ou la production PV. Pour combler les lacunes des approches précédentes, les techniques d'apprentissage automatique ont été largement recommandées par les prévisionnistes soit sous forme individuelle ou en hybridation avec d'autres approches. Le succès incroyable des techniques de l'intelligence artificielle réside dans plusieurs avantages qu'offrent ces approches. Tout d'abord dans la

facilité de la modélisation qui ne nécessite pas de solides connaissances dans la météorologie, ce qui a permis la vaste implémentation de ces techniques dans le domaine de la prévision solaire. De plus, ces modèles sont flexibles par rapport à l'horizon de prévision. Elles sont capables de traiter les prévisions intra-horaires, horaires, journalières, et mensuelles, ce qui offre plus d'options aux gestionnaires des centrales photovoltaïques pour gérer l'énergie produite. En outre, ces techniques peuvent générer des prévisions sur la base seulement de l'historique de la variable pareille aux techniques autorégressives. Elles peuvent également intégrer facilement plus de caractéristiques en entrée, ce qui peut contribuer à l'amélioration des performances des modèles. Néanmoins, la précision des techniques d'intelligence artificielle dépend du degré de corrélation des caractéristiques d'entrée avec la variable cible, de l'architecture neurale, du prétraitement et la normalisation des données, et l'ajustement des hyperparamètres du modèle.
Des prévisions précises sont difficiles à obtenir et les chercheurs tentent toujours à les améliorer. Plusieurs auteurs ont mentionné que l'hybridation de diverses approches de prévision pourrait améliorer la précision des prévisions par rapport à l'utilisation des modèles individuels.

D'après cet état de l'art, nous pouvons constater que pour avoir des prévisions robustes qui couvrent plusieurs horizons temporels, il faut utiliser une variété de méthodes de prévision. Les résultats obtenus dans les diverses études de l'état de l'art sont pertinents, mais ils ne peuvent pas être généralisés et transposer sur un autre site vu qu'ils sont exclusifs aux conditions des expériences. Dans cette thèse, nous allons évaluer diverses approches de prévision dans le but de développer un modèle robuste pour la prévision du rayonnement solaire pour les horizons considérés, à savoir un 12 heures et 15 minutes à l'avance. Diverses techniques telles que SARIMA, MLP, LSTM combinées à un modèle numérique de prévision seront utilisées. L'imagerie du ciel sera aussi utilisée pour l'implémentation de l'information sur les nuages. Le chapitre 4 présentera en détail les méthodologies de prévisions que nous avons développées durant cette thèse. Alors que dans le prochain chapitre nous présentons les étapes qui précèdent le développement du modèle de prévision. Dans une première partie, nous présentons l'ensemble des données disponibles au sol. Par la suite, nous passons en revue les différentes techniques de contrôle de qualité et les résultats de leur implémentation pour les composantes solaires mesurées aux divers sites d'études. Ces données seront utilisées pour la validation des bases de données exploitées durant cette thèse, à savoir HeclioClim-3 et le modèle de ciel clair McClear et les résultats de validation sont représentés dans la dernière partie du chapitre.

Chapitre 3

Sites d'étude et contrôle de qualité des données de rayonnement solaire

Introduction **60**

1. Mesures au sol et sites d'étude **60**

1.1. Sites d'étude 61

1.2. Stations météorologiques au sol 61

2. Contrôle de qualité des données **64**

2.1. État de l'art sur les techniques de contrôle de qualité 65

2.2. Procédure de contrôle de qualité 68

2.3. Tests à une composante 68

2.3.1. Inspection visuelle des séries temporelles 68

2.3.2. Vérification d'horodatage des mesures 69

2.3.3. Limites physiquement possibles (PPL) 70

2.3.4. Limites extrêmement rares (ERL) 71

2.4. Tests à deux composantes 72

2.4.1. Analyse de plausibilité BSRN à deux composantes 72

2.4.2. Tests des valeurs de K NREL-SERIS 74

2.5. Test à trois composantes 76

3. Validation des bases de données **78**

3.1. Détection des instants de ciel clair 79

3.2. Résultats de validation 82

3.2.1. Helioclim3 82

3.2.2. McClear 85

4. Données de prévision de GFS **87**

Conclusion **88**

Introduction

Les instruments au sol constituent le moyen le plus précis pour évaluer le rayonnement solaire à un endroit donné. Cependant, ils sont peu nombreux en raison de leurs coûts élevés d'installation et de maintenance. Pour remédier au manque des enregistrements au sol, d'autres moyens ont été utilisés pour évaluer le rayonnement solaire en tout lieu et à tout moment. L'imagerie satellitaire se présente comme une alternative précise complémentaire aux capteurs radiométriques. Cependant, avant l'utilisation des bases de données satellitaires, l'analyse de leurs finesses à travers la validation des estimations par rapport aux mesures au sol est considérée comme cruciale. Plusieurs bases de données ont été construites pour fournir de l'information sur le rayonnement solaire à partir des satellites géostationnaires. Une d'entre elles est validée dans ce chapitre à savoir HelioClim-3 (version 5).

Au-delà de la modélisation du rayonnement pour les conditions de ciel normal, la modélisation de l'irradiance par ciel clair constitue une composante essentielle de l'évaluation de la ressource solaire. Elle est impliquée dans un grand nombre d'applications, y compris l'évaluation de l'irradiance sous conditions normales, basée sur des observations dérivées de satellites où leurs précisions dépendent d'une grande partie du modèle de ciel clair combiné à l'indice de nuages estimé par satellite. Dans ce contexte, le modèle de ciel clair McClear a aussi été validé pour les mêmes sites d'études sous des conditions de ciel clair. Bien que de nombreuses études de validation des modèles d'estimation du rayonnement solaire ont été menées dans la littérature, chacune d'entre elles souffre de certaines lacunes, notamment l'étendue géographique et la méthode de détection des instants de ciel clairs. L'objectif de la validation de ces bases de données est de tester leurs précisions spatiales pour la récupération des données de GHI pour divers types de climat au Maroc. Des données de références collectées à partir de quatre stations au sol au Maroc ont été utilisées pour la validation. Les erreurs des modèles sont discutées pour des conditions de ciel normal et de ciel clair. Dans ce même chapitre, les stations météorologiques au sol installées aux divers sites d'études, les données expérimentales collectées, et la procédure de contrôle de qualité appliquée aux trois composantes du rayonnement solaire (GHI, DHI, et DNI) seront discutées.

1. Mesures au sol et sites d'étude

Pour une évaluation complète des méthodes de prévision de GHI, une étude multi-sites est recommandée pour augmenter la précision des prévisions. Dans le but de développer un outil de prévision portable et adaptable à divers climats, des ensembles de données provenant de quatre sites présentant des conditions météorologiques différentes sont déployés pour valider les modèles de prévision développés. Les données que nous allons présenter dans ce chapitre peuvent se classer en trois catégories : des données mesurées au sol, des données estimées par des modèles physiques et satellitaires, et des prévisions métrologiques à partir des modèles numériques de prévision.

1.1. Sites d'étude

L'ensemble des données de GHI et des variables météorologiques sont collectées à partir de quatre stations météorologiques de haute précision installées à divers endroits au Maroc caractérisés par divers types de climat. À savoir, Benguerir, Erfoud, Missour et Zagora (Figure 3-1). Selon la classification climatique de Köppen Geiger [226], Erfoud et Zagora partagent un climat chaud et désertique (dans la classification de Köppen BWh), où l'irradiation solaire est très élevée, avec une durée moyenne d'insolation entre 3600 - 4000 h/an. Missour se caractérise par un climat désertique et froid BWK caractérisé par des étés chauds et secs avec une forte irradiation qui atteint 8.1 kWh en été et des hivers glacials et rigoureux. Alors que le site de Benguerir est caractérisé par un climat Bsh qui est semi-aride, sec avec une insolation de 3300 h/an.

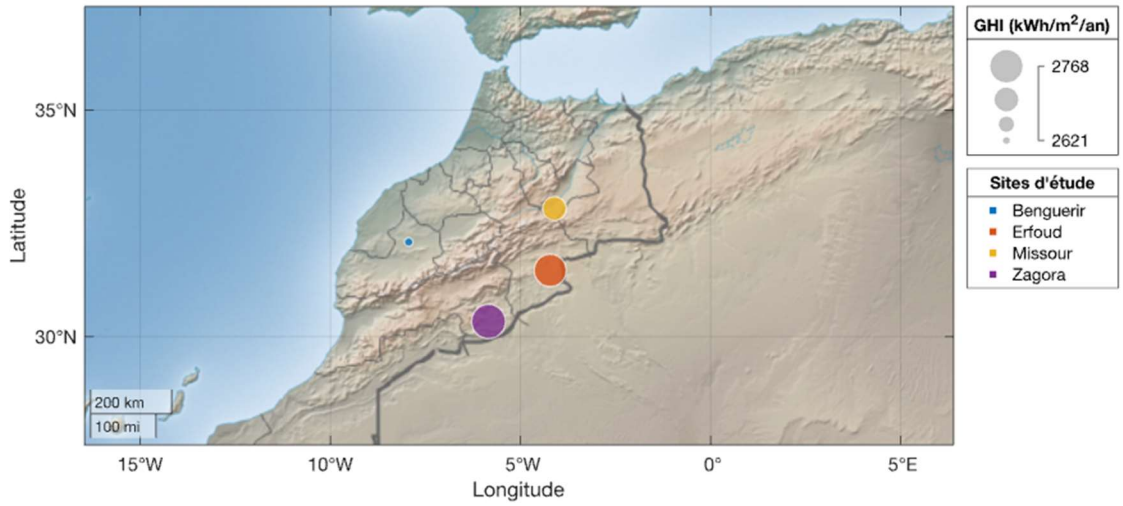


Figure 3-1. Emplacements géographiques des sites d'étude. La taille des bulles correspond au GHI annuel estimé par satellite en kWh/m^2.

Le Tableau 3-1 représente les principales caractéristiques des sites d'étude, leurs latitudes, longitudes, altitudes, le type de climat dominant, le GHI annuel (kWh/m^2) estimé à partir du satellite et la période des données disponibles.

Tableau 3-1. Caractéristiques des sites d'étude.

Station	Latitude (°)	Longitude (°)	Altitude (m)	Climat	Irradiance annuelle kWh/m²	Période disponible
Benguerir	32.12	-7.94	450	Bsh	2621	01/01/2015 – 30/06/2021
Erfoud	31.491	-4.218	859	BWh	2752	01/06/2013 – 30/06/2020
Missour	32.860	-4.107	1107	BWk	2688	01/06/2013 – 30/06/2020
Zagora	30.272	-5.852	783	BWh	2786	01/06/2013 – 30/06/2020

1.2. Stations météorologiques au sol

Les pyranomètres utilisés pour mesurer les composantes du rayonnement au sol (GHI, DHI et DNI) diffèrent selon les stations (thermopile ou photodiode). Les stations de Missour

et Benguerir sont de types thermopile partageant la configuration suivante : chaque station est équipée d'un pyrhéliomètre Kipp & Zonen CHP1 de première classe pour la mesure du DNI et de deux pyranomètres Kipp & Zonen CMP21 de seconde classe montés horizontalement pour la mesure du GHI et du DHI. Le pyrhéliomètre a un angle de vue à ouverture totale de 5° ± 0.2°, le domaine spectral s'étend de 200 à 4000 nm avec un temps de réponse de 1 s (95%) [227]. Les pyranomètres ont un angle de vue à ouverture totale de 180°, la gamme spectrale s'étend de 310 à 2800 nm avec un temps de réponse < 1 s (95%) [228]. Les termes "première classe" et ″deuxième classe" signifient que les appareils sont étalonnés conformément à la norme ISO 9060 [229], définie par l'OMM. Les capteurs sont montés sur un suiveur Kipp & Zonen à deux axes permettant le suivi de la trajectoire du soleil dans le ciel avec une précision supérieure à 0.1° [230]. Pour assurer un ombrage régulier du capteur DHI, une bille d'ombre est fixée sur le suiveur solaire.

Les sites d'Erfoud et de Zagora sont de types RSI équipés avec des pyranomètres à bande rotative de type photodiode. Chaque station est équipée d'un pyranomètre monté horizontalement et d'une bande d'ombre montée sur un moteur (Figure 3-2). Après avoir reçu un signal de démarrage, le moteur entraîne la bande d'ombre toutes les minutes en la faisant tourner une fois autour du capteur. Ainsi, quand le capteur est ombragé, il mesure le DHI. En outre, sans l'ombre, le GHI est mesuré. La précision est de ±4.1% et ±6.5% pour le GHI et le DHI, respectivement. Ensuite, le DNI est calculé à partir des mesures du GHI et du DHI avec une précision attendue de <± 4.1% [231].

En plus des mesures de l'irradiation solaire, les stations sont équipées des capteurs météorologiques mesurant les variables météorologiques nécessaires pour garantir une précision plus élevée des prévisions. Ces variables météorologiques sont la température (T, C°), l'humidité relative (RH, %), la pression barométrique (BP, hPa), la vitesse du vent (WS, m/s), la direction du vent (WD, deg °) et les précipitations (mm). Les caractéristiques des capteurs utilisés pour mesurer les variables météorologiques sont présentées dans le Tableau 3-2.

Tableau 3-2. Principales caractéristiques des capteurs des stations météorologiques.

Variables	Capteur	Plage de fonctionnement	Précision
Température	Campbell Scientific CS215	-40°C à +70°C	±0.3°C à 25°C ±0.4°C du +5°C à +40°C ±0.9°C du -40°C à +70°C
Humidité relative	Campbell Scientific CS215	0% à 100%	±2% du 10 à 90% ±4% du 0 à 100%
Vitesse du vent	Anemometer NRG #40C	1 à 96 m/s	< 0.1 m/s du 5 à 25 m/s
Direction du vent	Girouette NRG #200P	0° à 360°	< 1%
Précipitation	Young 52202/ 52203	Résolution de 0.1 mm par pointe	±2% jusqu'à 25 mm/h ±3% jusqu'à 50 mm/h

Pression barométrique	Campbell Scientific CS100	600 hPa à 1100 hPa	0.5 hPa à 20°C 1.0 hPa du 0°C à 40°C 1.5 hPa du -20°C à +50°C 2.0 hPa du -40°C à+60°C

La Figure 3-2 présente des images réelles des stations météorologiques installées aux sites d'étude. Benguerir et Missour sont des stations automatiques équipées par des capteurs installées sur des suiveurs de soleil, alors qu'Erfoud et Zagora sont des stations classiques équipées par des pyranomètres à bande rotative. L'ensemble de données est collecté au pas d'une minute et puis moyenné aux données sous horaires, horaires, ou mensuelles.



Figure 3-2. Les quatre sites d'étude et les capteurs pyranométriques associés.

La Figure 3-3 représente un exemple des valeurs de GHI et des variables météorologiques mesurées à la station météorologique de Benguerir et moyenné sur une heure.

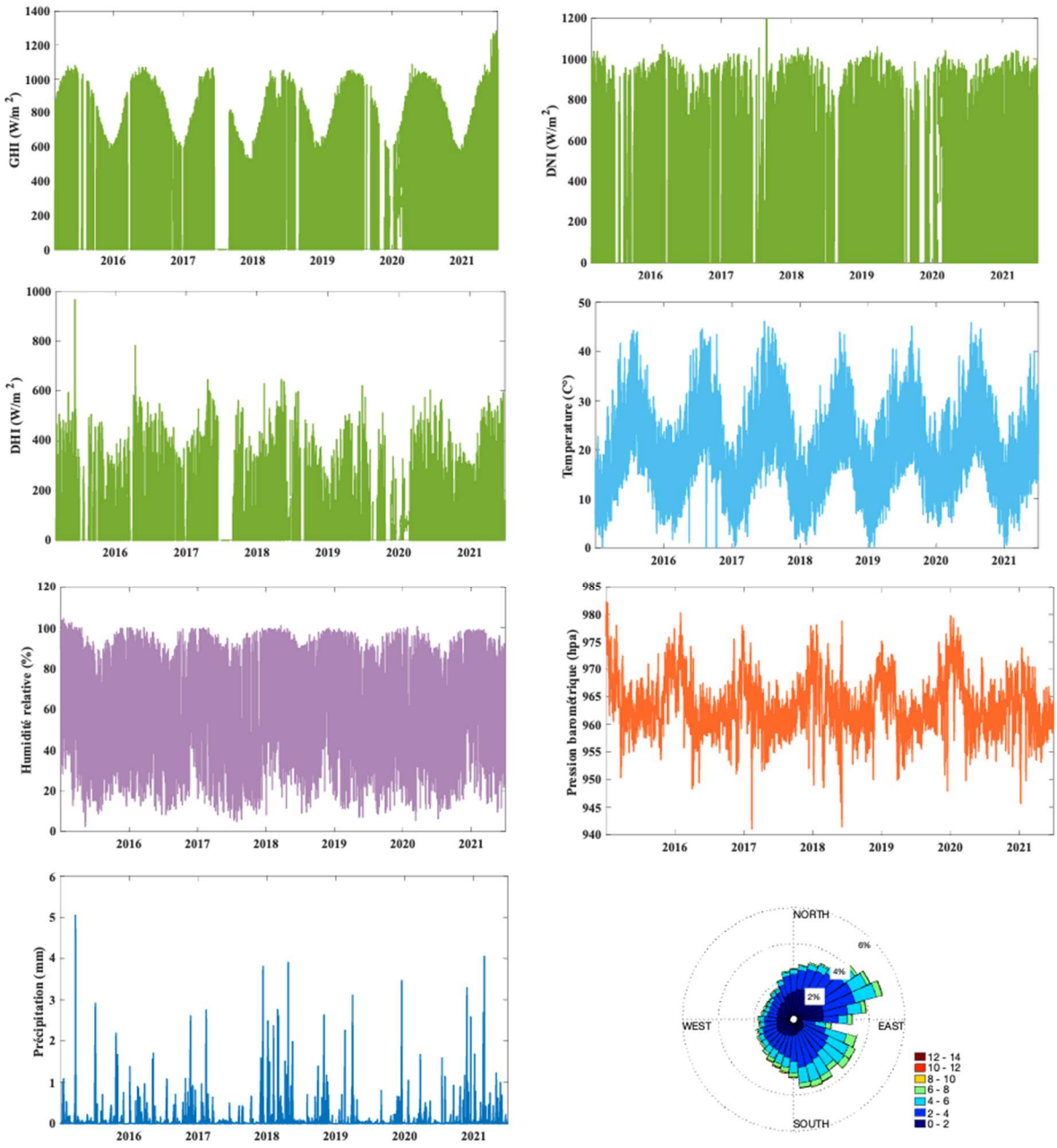


Figure 3-3 Séries temporelles brutes du GHI, DHI, DNI et des variables météorologiques collectés à la station météorologique de Benguerir à titre d'exemple pour la période allant de 01/01/2015 au 30/06 /2021 au pas horaire.

2. Contrôle de qualité des données

La mesure la plus précise du rayonnement solaire provient des stations météorologiques au sol. Les stations pyranométriques mesurent avec précision le rayonnement solaire réellement reçu à un emplacement géographique spécifique à un instant bien déterminé, avec plus ou moins une précision liée à l'incertitude du capteur. Avec toute mesure au sol, les données risquent toujours de contenir des incertitudes qui peuvent être systématiques ou générées par l'instrument lui-même. Selon [232,233], toute source d'erreurs ou de problèmes probables dans la mesure du rayonnement solaire peut être classifiée en trois types principaux :

- Erreurs et incertitudes liées à l'équipement (réponse en cosinus, réponse en azimut, réponse en température, sélectivité spectrale, et le désalignement de l'anneau d'ombrage).
- Problèmes et erreurs liés au fonctionnement (désalignement complet ou partiel de l'anneau d'ombrage, poussière, neige, rosée, gouttelettes d'eau, fientes d'oiseaux, etc.).
- Mise à niveau incorrecte des capteurs, ombrage causé par les structures des bâtiments, champs électriques à proximité des câbles, charge mécanique des câbles, orientation et/ou protection inadéquate des capteurs verticaux contre le rayonnement réfléchi par le sol, arrêt de la station, etc.

La qualité des données a constitué un thème de recherche important pour la majorité des organisations radiométriques [234–237]. Plusieurs études ont été menées et diverses méthodes de contrôle de la qualité des données ont été proposées. Certaines ont appliqué des tests sur le GHI uniquement, d'autres ont proposé des tests basés sur deux ou trois composantes de l'irradiance, tandis que d'autres études ont proposé des tests basés sur la représentation graphique du rayonnement selon l'indice de clarté. La section 2.1 présente un résumé d'un certain nombre de méthodes de contrôle de la qualité proposées dans la littérature.

2.1. État de l'art sur les techniques de contrôle de qualité

Le laboratoire national des énergies renouvelables (NREL) a développé un logiciel sous le nom de *"SERI QC"*. Ce logiciel permet l'évaluation de la qualité des données du rayonnement solaire en se basant sur les fractions directes, diffuses et l'indice de clarté [238].
Le Baseline Surface Radiation Network (BSRN) [3] a recommandé une procédure de contrôle de qualité basée sur plusieurs sous-tests : le test des limites physiquement possibles (PPL), le test des limites extrêmement rares (ERL) et le test de cohérence entre les composantes. La procédure a été proposée par Long et Dutton [239] et elle est devenue l'une des méthodologies les plus connues et utilisées dans la littérature [234,241–244].
D'autres tests ont été largement appliqués dans la littérature, Geiger et al. [245] ont développé leurs propres procédures pour le contrôle de qualité des données de GHI. Ils ont comparé les observations avec des prévisions faites sur la base de l'irradiation extraterrestre et d'une irradiation simulée dans des conditions de ciel clair. Cette procédure de contrôle de qualité fait partie d'un effort continu du groupe "Télédétection et Modélisation" de l'Ecole des Mines de Paris/Armines pour fournir des outils et des informations sur le rayonnement solaire à travers le web mondial. Ils ont créé un service en ligne pour effectuer le contrôle de qualité[2].

Sur la base des travaux effectués pour la production de l'outil ESRA (European Solar Radiation Atlas) [246] et du guide des données météorologiques et solaires de la CIBSE (Charted Institution of Building Services Engineers), Page [247] a proposé une procédure de contrôle qualité comprenant quatre niveaux de test basés sur le GHI et le DHI. Les tests vérifient que les valeurs de rayonnement se situent dans la plage des valeurs extrêmes par rapport au ciel

[2] http://www.helioclim.net/

clair et au ciel couvert pour le DHI, et par rapport au ciel clair pour le GHI. Le test prend également en considération la variabilité spatiale en se référant aux mesures des stations adjacentes.

Molineaux et Ineichen [248] ont développé un logiciel pour le contrôle de qualité des trois composantes du rayonnement solaire. La procédure consiste en trois tests : le premier test vérifie la cohérence entre les trois composantes de l'irradiance, le second compare les mesures avec des prévisions générées par un modèle de prévision, et le dernier test est basé sur la hauteur de la teneur en eau précipitable, et la température du point de rosée. Le programme vise à augmenter la précision des tests en fixant des seuils limites pour chaque test.

Muneer et Fairooz [249] ont proposé une procédure de contrôle de qualité basée sur les tests de la Commission Internationale de l'Éclairage (CIE) [250], et d'autres filtres. La procédure est composée de plusieurs examens : le premier répète le premier test proposé par la CIE qui vérifie les limites supérieures du GHI, DHI, et DNI. Le deuxième vérifie la cohérence entre les trois composantes et le rayonnement au top de l'atmosphère. Le troisième test est composé de deux parties ; la première partie vérifie que le DHI est conforme aux limites établies par l'enveloppe standard qui sont présentées dans l'espace ″indice de clarté - fraction diffuse", la deuxième partie est un contrôle de qualité supplémentaire de DHI qui vise à comparer la valeur de DHI dans deux conditions extrêmes, à savoir un ciel très clair et un ciel très nuageux.

Youness et al. [251] ont réalisé un système semi-automatique programmé sous le langage FORTRAN [252] pour le contrôle de la qualité des données de GHI. Les auteurs ont proposé un ensemble de métriques physiques et statistiques basées sur la création d'une enveloppe dans le domaine ″indice de clarté - fraction diffuse", la procédure utilise également les informations géographiques du site d'étude.

Journée et Bertrand [253] ont développé une procédure de contrôle qualité applicable aux données de GHI de 10 et 30 minutes. La procédure comprend plusieurs tests : les limites physiques, qui vérifient les limites supérieures du GHI par rapport au rayonnement au top de l'atmosphère et par rapport au ciel clair. Les tests de pas qui vérifient que la variation des mesures du rayonnement solaire entre deux horodatages successifs qui ne doivent pas dépasser certaines limites, et que les composantes solaires doivent varier de manière compatible les unes avec les autres. Le test de persistance pour évaluer la variabilité des mesures. Les tests d'enveloppe qui sont basés sur la représentation graphique à partir les fractions directe, diffuse et de l'indice de clarté. Le test de cohérence spatiale établi à partir de l'analyse des données en comparant les stations voisines entre elles. Finalement, le test de la durée d'ensoleillement.

En plus de la détection automatique des erreurs, Ineishen [254] a proposé la vérification de trois éléments. Tout d'abord, l'horodatage de la mesure qui permet de détecter le décalage horaire et de vérifier la symétrie des données par rapport au midi solaire en traçant les composantes globales et directes en fonction du sinus de l'élévation solaire par temps clair. L'horodatage est correct si la courbe du matin est située en dessous de celle de l'après-midi. Ensuite, il a évalué le coefficient d'étalonnage des capteurs utilisé pour convertir les données mesurées en valeurs physiques, en incluant l'indice de clarté avec un facteur de Turbidité de Linke (TL) fixée à 1.4. Enfin, le test de cohérence entre les composantes solaires est effectué en traçant les fractions diffuse et directe en fonction de l'indice de clarté ; et sur le même

graphique, les prédictions de ciel clair du modèle radiatif Solis [255] sont représentées. Toute différence significative entre les valeurs de ciel clair prédites et mesurées implique des incertitudes de calibration, un mauvais alignement du pyrhéliomètre, un encrassement, un ombrage, ou une classification erronée des conditions de ciel clair.
Norbert Geuder et al. [256] ont proposé une combinaison de filtrage automatique des données et de marquage manuel maintenu par un expert à une fréquence quotidienne. Le filtrage des données consiste à prouver que les données ne dépassent pas les limites inférieures et supérieures pour le GHI, le DHI, et le DNI. En outre, les tests BSRN sont appliqués à l'ensemble de données et un examen de la rosée sur les capteurs d'irradiance est également mis en œuvre. En plus des tests de qualité des données de rayonnement solaire, les auteurs recommandent la vérification de la température ambiante, le vent et les précipitations.
Récemment, Urraca et al. [257] ont proposé une procédure semi-automatique pour détecter les erreurs dans l'irradiation mesurée, basée sur l'analyse statistique des écarts quotidiens entre un ensemble d'observations au sol et les estimations d'irradiation par satellite. Cette méthode a été validée pour un ensemble de données de 313 stations au sol.

Après cette revue de littérature sur les procédures de contrôle de qualité des données de rayonnement solaire, nous pouvons affirmer que, dans la littérature et sur les plateformes web, il existe un large éventail de tests de qualité pour évaluer les séries temporelles d'irradiation solaire. Ces méthodes fournissent généralement un type de résultat booléen pour chaque mesure, signifiant soit la validation, soit le rejet de la valeur. Il s'agit d'un processus de suppression des données rapide. Cependant, elles ne fournissent aucune information sur la source du rejet et ne permettent pas de savoir comment remédier à la cause du problème potentiel.

Jusqu'à présent, une quantité importante de données a été générée et la communauté des données solaires est maintenant capable de manipuler de grandes séries temporelles grâce au développement récent des outils d'exploration de données. Notre contribution à cet égard réside dans le développement d'un support visuel des tests de contrôles de qualité les plus utilisés. Cet outil visuel permet un aperçu plus précis sur la qualité des mesures, et peut être envisagé comme un moyen permettant d'améliorer le taux de détection des données aberrantes, qui constitue la première étape du diagnostic permettant d'examiner les causes de l'erreur. L'objectif de ce contrôle de qualité est non seulement d'exclure les données jugées invraisemblables, mais aussi d'évaluer quantitativement et qualitativement l'incertitude des mesures en effectuant une série de tests et d'inspections visuelles des données. Ce travail est réalisé en parallèle d'une réflexion menée dans le cadre de la tâche 16 de l'IEA PVPS [258] et est donc novateur dans le domaine des mesures in situ pour l'énergie solaire. Tous les tests utilisés dans cette étude sont basés sur ceux proposés par BSRN et NREL [238,239] .

Dans la section suivante, nous présentons la procédure de contrôle de qualité appliquée aux trois composantes de l'irradiance solaire. Nous exploitons, à titre d'exemple, les données sur 10 minutes des mesures collectées sur deux sites d'étude équipés de types différents de capteurs pyranométriques à savoir Benguerir et Erfoud. La période des données utilisées est du

01/01/2015 au 30/06/2021 pour le site de Benguerir, et du 06/01/2013 au 30/06/2020 pour Erfoud.

2.2. Procédure de contrôle de qualité

Les différents tests de contrôle de qualité sont résumés dans le Tableau 3-3. Les lignes et les colonnes correspondent à la composante du rayonnement solaire impliquée dans le contrôle de qualité. Dans la diagonale (cellules bleues), les différents tests appliqués à une seule composante sont listés : l'inspection visuelle de la série temporelle, la vérification d'horodatage des mesures, et les tests ERL (Limites extrêmement rares) et PPL (Limites physiquement possibles) de la procédure de BSRN [239]. Ces différents tests sont décrits en détail dans la section 2.3. Dans la partie inférieure du tableau, les tests appliqués à deux composants sont résumés : les tests BSRN à deux composants et les tests des fractions du NREL. Ces tests sont décrits dans la section 2.4. Enfin, la partie supérieure droite du tableau (cellules orange) contient les tests impliquant toutes les composantes du rayonnement solaire. Cette dernière famille de tests est décrite dans la section 2.5.

Tableau 3-3. Résumé des tests de contrôle des données appliqués aux mesures.

<table>
<tr><th></th><th>GHI</th><th>DHI</th><th>DNI</th></tr>
<tr><td>GHI</td><td>Inspection visuelle de la série temporelle
Vérification d'horodatage
BSRN-ERL
BSRN-PPL</td><td colspan="2" rowspan="2">Test à 3 composantes.</td></tr>
<tr><td>DHI</td><td>BSRN- test à 2 composantes
Les tests de $K_t - K_d$</td><td>Inspection visuelle de la série temporelle
Vérification d'horodatage
BSRN-ERL
BSRN-PPL</td></tr>
<tr><td>DNI</td><td>Les tests de $K_t - K_n$</td><td></td><td>Inspection visuelle de la série temporelle
Vérification d'horodatage
BSRN-ERL
BSRN-PPL</td></tr>
</table>

Test à 1 composante

Test à 2 composantes

Test à 3 composantes

2.3. Tests à une composante

2.3.1. Inspection visuelle des séries temporelles

Avant de procéder à la routine de contrôle de qualité, une visualisation de la série temporelle est nécessaire. La présentation graphique des données ne permet pas de contrôler la qualité des données, mais plutôt d'examiner le type de données pour identifier leurs tendances et vérifier s'il y a des valeurs manquantes, des décalages temporels, etc. Les séries temporelles au pas de

10 min du GHI, DHI et DNI pour les sites de Benguerir et Erfoud sont représentées en fonction du temps à la Figure 3-4. En analysant ces graphiques, nous pouvons facilement remarquer l'existence de quelques données manquantes avec un pourcentage de 6.7% pour Erfoud et 17% pour Benguerir.

Ces valeurs manquantes sont conservées dans l'ensemble des données et aucune méthode de comblement des lacunes n'a été appliquée. Malgré le fait qu'un pourcentage substantiel de données soit manquant, il y a encore suffisamment de données pour obtenir une vue raisonnable de toutes les saisons. La majorité de ces données manquantes correspondent à une période continue allant de quelques jours à quelques mois. Les données manquantes sont généralement causées par les coupures de courant, des problèmes de communication qui durent trop longtemps de sorte que le stockage interne des données est dépassé, ou des problèmes de suivi de soleil où le suiveur s'est arrêté et a dû être redémarré et réparé.

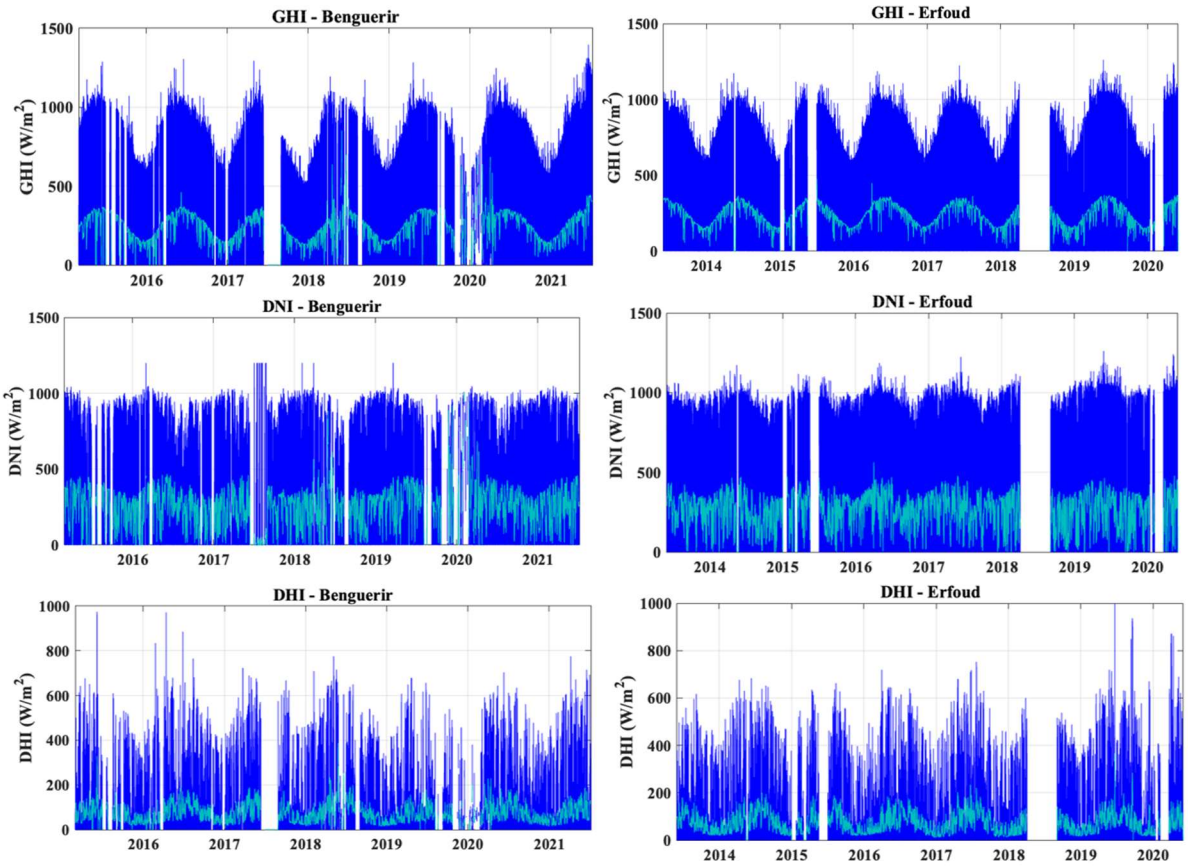


Figure 3-4. Séries temporelles des trois composantes du rayonnement solaire pour les sites de Benguerir et Erfoud. Le bleu représente les valeurs de 10 minutes, et le cyan correspond aux moyennes journalières.

2.3.2. Vérification d'horodatage des mesures

Afin d'approfondir l'analyse visuelle des données, nous représentons les mesures sous une forme bidimensionnelle, où l'axe des *x* représente les jours de l'année, l'axe des *y* représente les heures de la journée, les couleurs représentent la donnée considérée : le GHI, le DNI ou le DHI, et les bandes blanches indiquent les valeurs manquantes. Cette représentation permet de visualiser les décalages temporels, les valeurs manquantes, les valeurs anormales et dans le cas

échéant les problèmes d'ombrage. Les graphiques pour les stations considérées sont présentés à la Figure 3-5. Dans les différentes cellules de la Figure 3-5, les deux lignes pointillées rouges représentent le temps théorique de lever et de coucher du soleil. Ils ont été calculés à l'aide de l'équation du coucher du soleil en fonction de la latitude géographique des sites considérés, en utilisant l'algorithme SG2 [259]. Cette information permet de détecter les problèmes de références temporelles des données, ce qui n'est pas le cas pour le jeu de données considéré ici. D'après cette représentation, le rayonnement est nul lors du coucher de soleil puis augmente pour atteindre des valeurs maximales entre 10:00 et 15:00, par la suite il diminue jusqu'au s'annuler vers le coucher du soleil. D'une autre part, les valeurs de DNI sont supérieures au DHI dans la majorité de cas, ce qui revient à la dominance des jours clairs par rapport aux conditions nuageuses qui provoquent l'augmentation du rayonnement diffus. Néanmoins, il n'est pas possible de juger la plausibilité des données à ce stade en se basant uniquement sur la visualisation des séries temporelles. Une analyse de plausibilité des mesures sera effectuée ultérieurement.

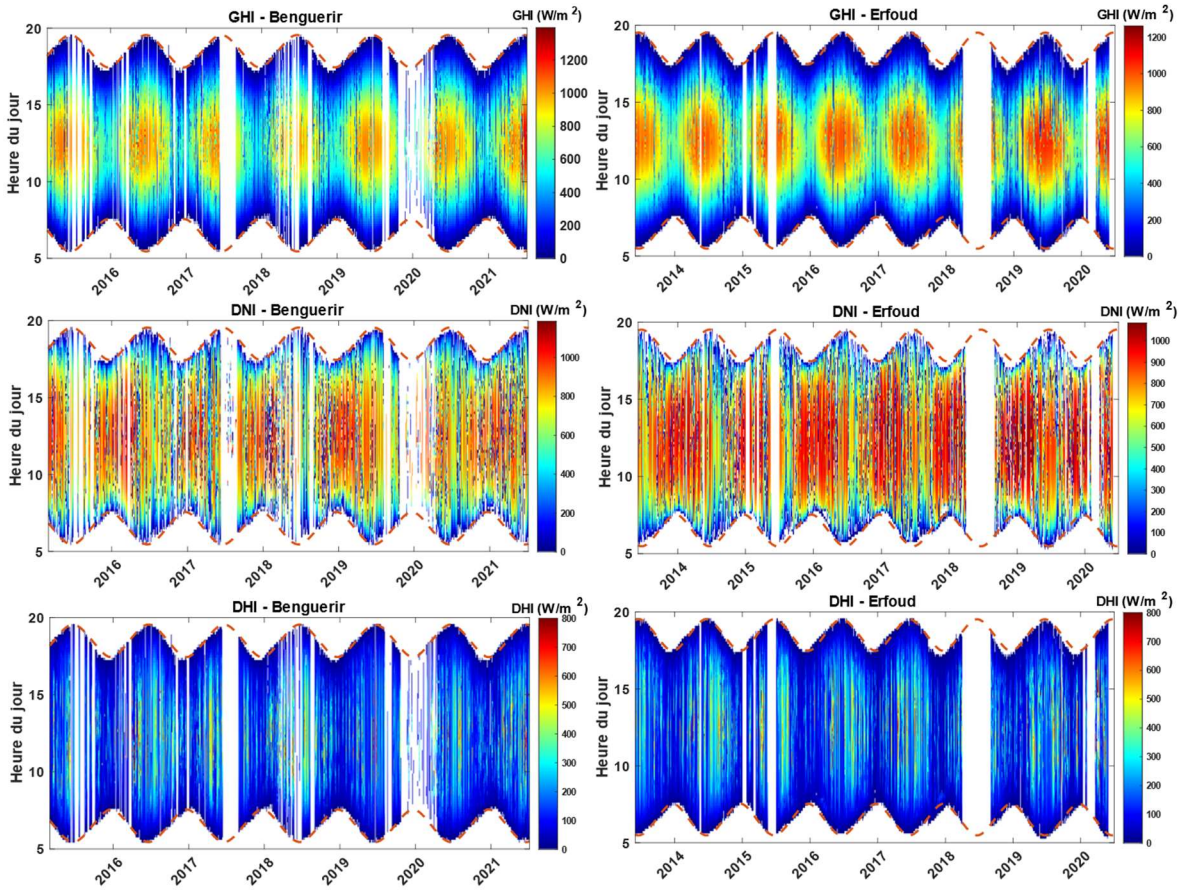


Figure 3-5. Représentation bidimensionnelle des données au pas de 10 minutes de GHI, DNI et DHI à Benguerir et Erfoud. Les pointillées rouges correspondent au temps de lever et de coucher du soleil, l'axe des x correspond aux jours de l'année, et l'axe des y correspond aux heures de la journée.

2.3.3. Limites physiquement possibles (PPL)

Le test PPL vérifie les limites maximales et minimales qui peuvent être physiquement atteintes par l'irradiance. Les limites supérieures dépendent de l'angle solaire zénithal (θ_z) et du

rayonnement normal au sommet de l'atmosphère (G_0). La valeur minimale de l'irradiance solaire devrait être de 0 W/m^2, mais en raison du refroidissement radiatif nocturne, la limite est fixée à - 4 W/m^2, le test s'applique indépendamment à chacun des trois composants comme suit :

$$-4 \le GHI \le 1.5\, G_0 \cos\left(\theta_z\right)^{1.2} + 100 \tag{3.1}$$

$$-4 \le DHI \le 0.95\, G_0 \cos\left(\theta_z\right)^{1.2} + 50 \tag{3.2}$$

$$-4 \le DNI \le G_0 \tag{3.3}$$

Toute mesure en dehors de ces intervalles de validité n'est pas plausible et doit être traitée en conséquence.

2.3.4. Limites extrêmement rares (ERL)

Les limites de test ERL sont encore plus strictes que celles du test PPL. L'ERL diffère du test PPL en ce que les mesures atteignent rarement ces limites, ou pour de courtes périodes allant de quelques secondes à une ou deux minutes, et aussi en ce que les mesures dépassant ces limites ne sont pas nécessairement incorrectes, mais leur plausibilité doit être vérifiée plus spécifiquement. Les limites de l'ERL sont définies comme suit :

$$-2 \le GHI \le 1.2\, G_0 cos\left(\theta_z\right)^{1.2} + 50 \tag{3.4}$$

$$-2 \le DHI \le 0.75\, G_0 cos\left(\theta_z\right)^{1.2} + 30 \tag{3.5}$$

$$-2 \le DNI \le 0.95\, G_0 cos\left(\theta_z\right)^{0.2} + 10 \tag{3.6}$$

Une représentation graphique de ces deux tests est possible en représentant les moyennes de 10 minutes de GHI, DHI ou DNI en fonction du rayonnement au top de l'atmosphère TOA. Cette représentation est illustrée à la Figure 3-6 où les tests PPL et ERL à une composante sont représentés par des lignes vertes et rouges respectivement. Les points orange correspondent aux valeurs dépassées par le test de PPL, alors que les points en violet indiquent les valeurs qui dépassent le test ERL.

Les différents graphes de la Figure 3-6 montrent qu'une très grande majorité des valeurs passent les tests. Cependant, on peut distinguer quelques groupes de données où la valeur des mesures est peu plausible. Pour le GHI et le DHI de Benguerir, nous pouvons observer un nombre limité de points qui dépassent les limites de PPL et ERL, alors que pour le DNI un faible nombre de points dépasse les limites de test ERL. Pour Erfoud, les données sont moins infectées, et seules quelques valeurs non remarquables dépassent les limites. En général, la plupart des données se situent en dessous du seuil des observations rares (vert), avec seulement un petit nombre de données au-dessus du seuil physiquement possible (rouge).

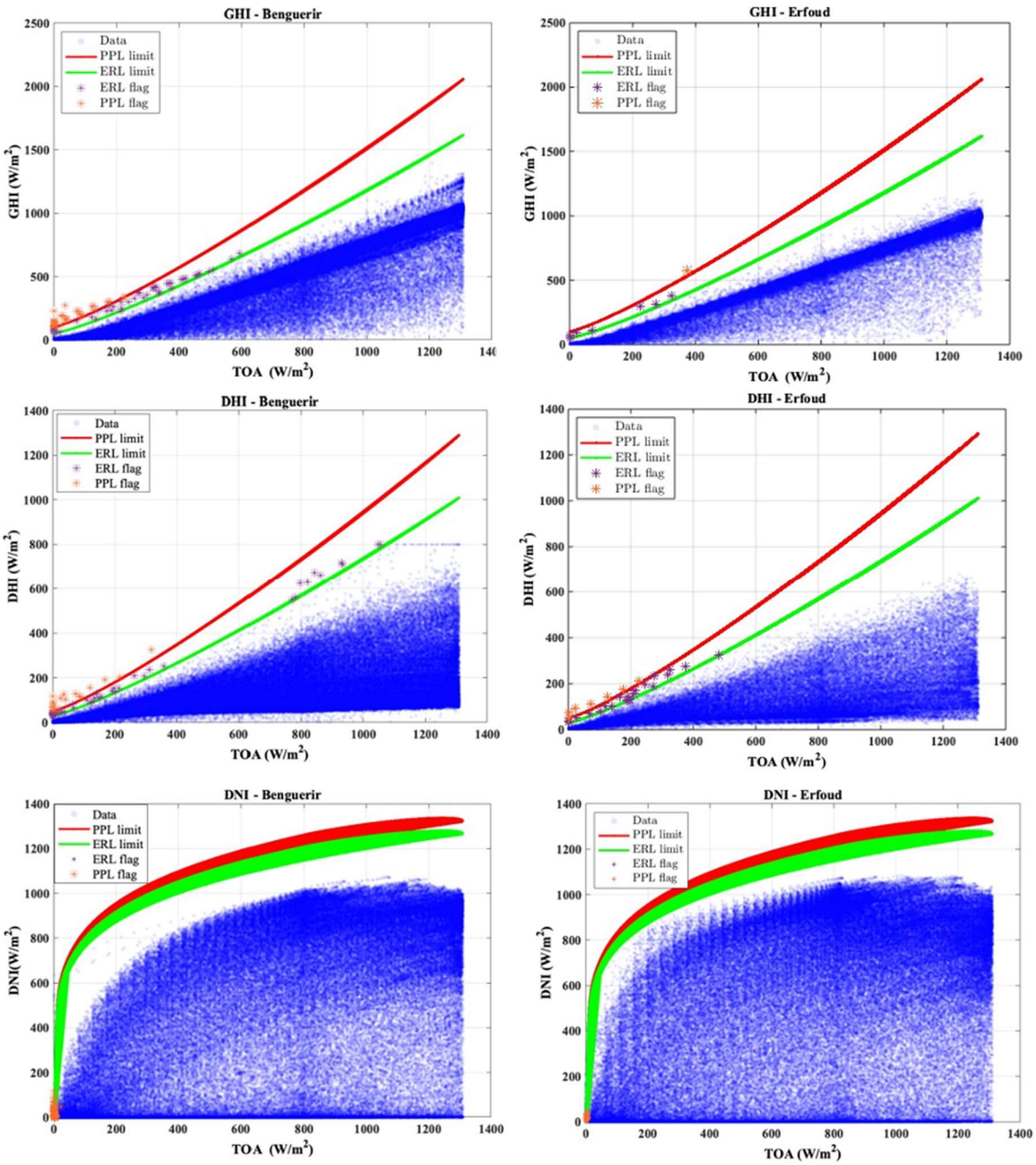


Figure 3-6. Représentations graphiques des tests PPL et ERL pour le GHI, le DHI et le DNI au pas de 10 minutes pour Benguerir et Erfoud.

2.4. Tests à deux composantes

2.4.1. Analyse de plausibilité BSRN à deux composantes

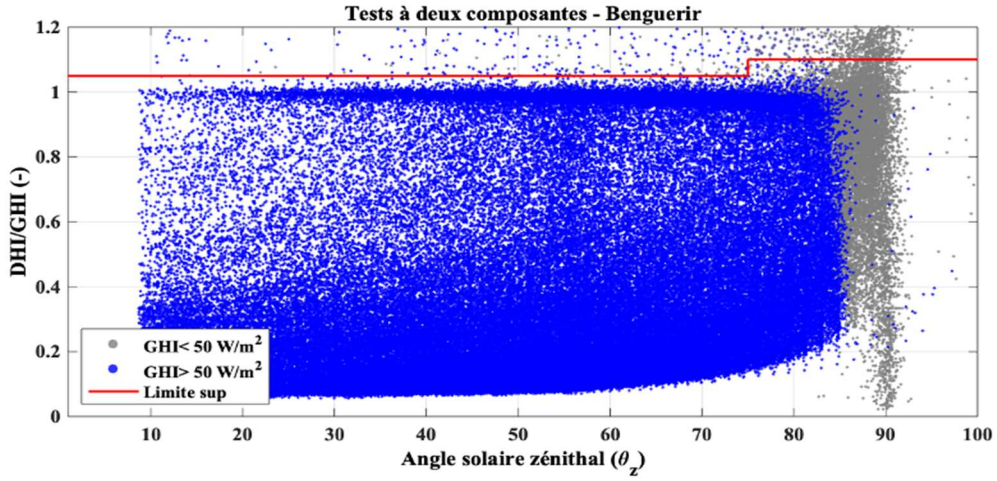
Un autre test recommandé par le BSRN est le test à deux composantes où la cohérence entre les composantes est testée. Lorsque le GHI et le DHI sont disponibles, la cohérence de la comparaison de ces deux mesures peut être vérifiée en appliquant pour le GHI > 50 W/m^2 les tests à deux composantes suivantes :

$$DHI \leq 1.01 GHI \text{ pour } \theta_z < 75° \quad (3.7)$$

$$DHI \leq 1.1 GHI \ pour \ 75° \leq \theta_z < 93° \quad (3.8)$$

Nous nous limiterons ici à une inspection visuelle des données correspondant à ce test à deux composantes. À cette fin, nous avons choisi de représenter le rapport DHI/GHI en fonction de l'angle solaire zénithal (θ_z). Sur les graphiques de la Figure 3-7, toutes les valeurs de GHI < 50 W/m^2 sont représentées par des points gris et les valeurs pour lesquelles GHI > 50 W/m^2 sont représentées par des points bleus. Les limites définies par les deux équations ci-dessus sont représentées par des lignes rouges.

Nous pouvons remarquer sur les graphes de la Figure 3-7 que pour le site d'Erfoud toutes les mesures remplissent les conditions du test à deux composants et aucun comportement suspect n'est observé. Alors que pour le site de Benguerir, nous observons quelques points qui dépassent les limites surtout pour la zone où le GHI > 50 W/m^2. En effet, pour la station de Benguerir, le GHI, le DHI et le DNI sont mesurés séparément avec deux pyranomètres et un pyrhéliomètre respectivement. Si le suiveur solaire ne suit pas bien le soleil, la boule d'ombre ne bloque pas le soleil pour le pyranomètre du DHI et donc ce pyranomètre mesure le GHI au lieu du DHI, ce qui peut entrainer des valeurs de DHI supérieur ou égale au GHI. En générale, les périodes de problèmes sur les mesures DHI sont très courtes et les différences entre les instruments de mesure ne sont observables qu'aux faibles angles d'élévation et restent marginales.



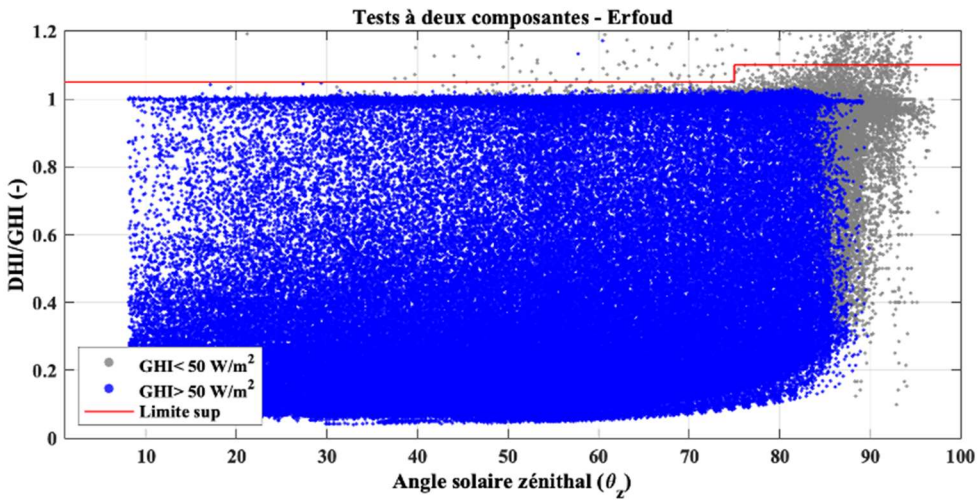


Figure 3-7. Représentation graphique du test BSRN à deux composants basés sur le ratio DHI/GHI pour les sites de Benguerir et Erfoud.

2.4.2. Tests des valeurs de K NREL-SERIS

Un autre ensemble de tests à deux composantes est souvent utilisé : le test des fractions ou bien "K NREL-SERIS" proposé par NREL [238].

Les limites supérieures exactes pour le GHI et le DHI sont difficiles à définir pour un site donné. En particulier dans des conditions instables ; avec la présence des nuages, des fumées ou des aérosols le rayonnement diffus peut dépasser les limites maximales entraînant ainsi une augmentation du GHI qui peut dans ces cas dépasser le rayonnement au top de l'atmosphère ou lorsque les nuages sont à une altitude qui permet au rayonnement direct d'augmenter pendant une brève période. À cette fin, une autre série de tests basés sur le calcul des fractions est souvent utilisée, le test comprend une vérification de la fraction diffuse K_d, la fraction directe K_n, et de l'indice de clarté K_t. Ces différents indices sont à l'origine du nom du test.

Les fractions sont calculées comme suit :

- La fraction directe :

$$K_n = DNI / G_0 \tag{3.9}$$

- La fraction diffuse :

$$K_d = GHI / DHI \tag{3.10}$$

- L'indice de clarté :

$$K_t = GHI / TOA \tag{3.11}$$

Il existe deux tests pour caractériser la procédure de NREL : un premier contrôle de plausibilité dans l'espace $K_t - K_n$ et un second dans l'espace $K_t - K_d$. Les tests effectués dans l'espace $K_t - K_n$ sont :

- $K_n > K_t$,
- $K_n < 0.8$,
- $K_t < 1$.

Pour la validation des mesures dans l'espace $K_t - K_d$, les tests suivants sont effectués :

- $K_d < 1.05$ *pour* $\theta_z < 75°$ et $K_d < 1.1$ *pour* $\theta_z > 75°$
- $K_d < 0.95$ *pour* $K_t > 0.6$

Pour visualiser ces tests, les fractions sont représentées dans les espaces $K_t - K_n$ et $K_t - K_d$ pour les deux sites Benguerir et Erfoud dans la Figure 3-8. Les limites des tests sont représentées par des pointillées rouges.

Pour le premier test $K_t - K_n$, la majorité des points passent le test et aucun comportement douteux n'est observé pour le site d'Erfoud. Alors que pour le site de Benguerir, nous observons quelques valeurs dépassant les limites et qui doivent être enlevées de l'ensemble.

Pour le test $K_t - K_d$, la majorité de mesures satisfont aux conditions de test sauf quelques points qui dépassent les limites surtout pour le site de Benguerir, ce qui est en conformité avec les observations faites précédemment au niveau du DHI. Autrement, pour le site d'Erfoud des valeurs de K_d égales à l'unité sont observées notamment pour les valeurs d'indice de clarté K_t supérieures à 0.6, ce qui n'a pas été détecté durant les tests précédents. Des valeurs de K_d proches de l'unité impliquent des problèmes de désalignement de la bande d'ombrage au niveau du pyranomètre et il mesure le GHI au lieu du DHI.

Nous pouvons également noter la différence des nuages de points pour $K_d \sim 1$ et $K_t < 0.4$ pour les deux stations. Ceci affirme que certaines mesures de DHI sont suspectes et doivent être enlevées de l'ensemble de données.

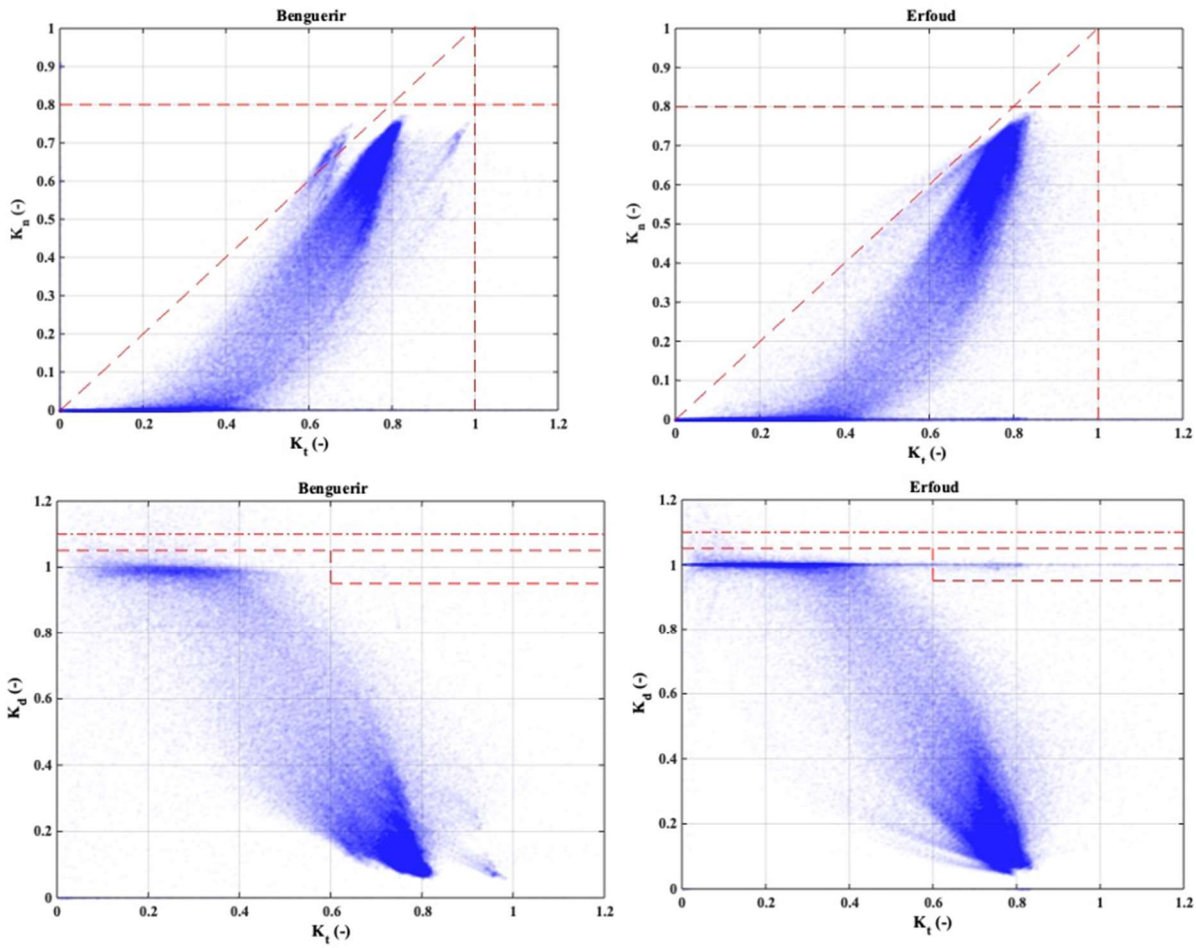


Figure 3-8. Nuage de points des valeurs de 10 minutes dans l'espace $K_t - K_n$, l'espace $K_t - K_d$ et les limites des tests pour les deux sites.

2.5. Test à trois composantes

Le test à trois composantes a pour but de comparer le GHI mesuré par le pyranomètre et calculé à partir du DHI et du DNI mesurés :

$$\left|\frac{GHI}{BHI + DHI} - 1\right| \leq 0.08, \; pour \; \theta_z \leq 75° \tag{3.12}$$

$$\left|\frac{GHI}{BHI + DHI} - 1\right| \leq 0.15, \; pour \; 75° < \theta_z \leq 93° \tag{3.13}$$

Ce test est également appliqué aux données de GHI > 50 W/m^2. En effet, le test en question est de toute évidence inutile pour les stations ne disposant pas de capteurs indépendants pour chacune des trois composantes. Nous allons appliquer le test pour la station de Benguerir.

Là encore, nous proposons une visualisation des données correspondant à ce test. Cela étant, il est intéressant de noter que le test à trois composantes consiste à comparer le GHI mesuré à l'aide du pyranomètre avec le GHI calculé à l'aide de l'équation (1.14) reliant le GHI, le DHI, le DNI, et le θ_z.

Idéalement, le rapport entre le GHI mesuré et estimé devrait être 1.0, mais en raison de l'imprécision des instruments, des valeurs un peu loin de l'unité sont souvent obtenues. Il est d'usage de considérer les trois mesures comme douteuses lorsque ce test n'est pas satisfait. Dans cette optique, ce test peut donc être assez vulnérable, car si une erreur de mesure se produit sur un instrument, elle peut considérablement falsifier les mesures des deux autres instruments.

L'évolution du rapport des valeurs de GHI en fonction de l'angle solaire zénithal est représentée dans la partie droite de la Figure 3-9, les points bleus représentent les valeurs de rapport correspondant à GHI > 50 W/m^2 et les linges pointillés rouges représentent les limites du test. Cette représentation montre que toutes les mesures sont cohérentes pour des angles zénithaux solaires inférieurs à 75°. Ce n'est pas le cas pour des angles plus élevés, où la dispersion des points est nettement plus importante. La dispersion entre le GHI mesuré et calculé est remarquable pour les angles zénithaux proches de 90°. La raison de cette dispersion peut être expliquée comme suit : idéalement, un pyranomètre a une réponse directionnelle qui est exactement la même que celle de la loi du cosinus. Cependant, dans un pyranomètre, la réponse directionnelle est influencée par le détecteur, la qualité, les dimensions et le processus de construction des dômes. Le CMP21 a une réponse directionnelle < 10 W/m2 jusqu'à 80° d'angle zénithal selon le manuel de Kipp & Zonen [260]. Alors que pour des conditions de très faible irradiance ($\theta_z > 80°$), les incertitudes relatives augmentent en raison de la sensibilité du capteur ou ce qui est connu sous le nom de la réponse en cosinus. Mais en général, ce type d'erreur n'a pas une influence remarquable sur les données puisque le rayonnement est plus important pour les faibles angles zénithaux. La Figure 3-10 présente les données marquées par le test à trois composantes en fonction des heures de la journée et des jours de l'année. Comme montré sur la Figure 3-10, la majorité des erreurs sont produites tôt le matin et vers le coucher du soleil, c'est-à-dire pour les valeurs élevées de l'angle solaire zénithal.

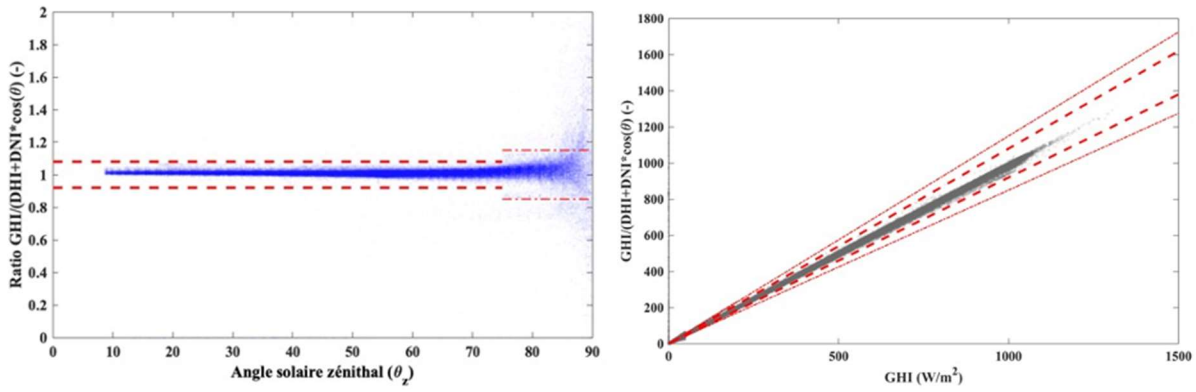


Figure 3-9. Représentation graphique du test à trois composantes. Le taux du GHI mesuré et calculé en fonction de l'angle solaire zénithal (droit). Des nuages de points du GHI mesuré et calculé (gauche). Les pointillés rouges indiquent les limites du test.

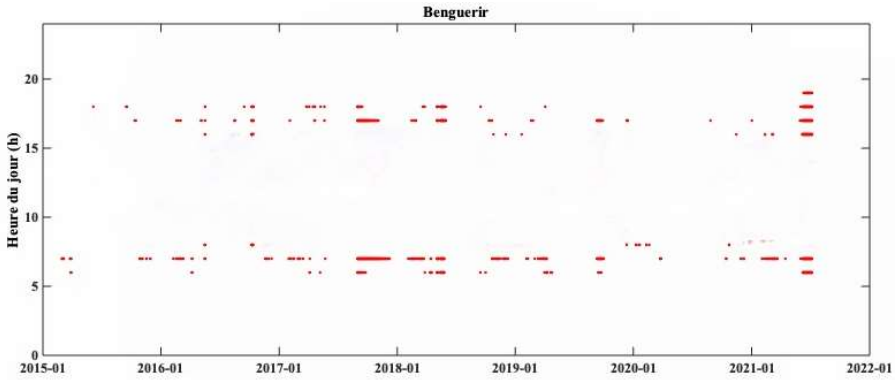


Figure 3-10. Représentation bidimensionnelle des données marquées par le test BSRN à trois composantes pour la station de Benguerir.

Après les divers tests de contrôle de qualité appliqués aux mesures, nous pouvons affirmer que la majorité des données passent les tests et leur qualité s'est maintenue tout au long de la période d'étude. Par ailleurs, pour la station d'Erfoud, un problème affectant les mesures de DHI a été détecté où les valeurs de DHI proches du GHI ont été observées pour une courte période. Les origines de ce problème peuvent être expliquées comme suit : si le GHI et le DHI sont identiques alors que la station n'est pas équipée par un pyrhéliomètre pour mesurer le DNI. Ceci dénote que la bande d'ombre n'est pas tournante et le pyranomètre ne mesure que le GHI. Un autre comportement a été observé : les points sont souvent dispersés pour les angles zénithaux supérieurs à 80°. Ce constat peut être expliqué par la réponse en cosinus des pyranomètres, qui peut introduire des erreurs au lever et au coucher du soleil. En général, l'impact des deux sources d'incertitude sur les mesures d'irradiance est très limité. Les périodes de problèmes sur les mesures DHI sont très courtes et les différences entre les instruments de mesure ne sont observables qu'aux faibles angles d'élévation et restent marginales. Les données marquées ont été éliminées de la base de données et aucune méthode de comblement des lacunes n'est considérée surtout pour des raisons d'utilisation de ces données pour la validation des modèles d'estimation du rayonnement solaire.

3. Validation des bases de données

La validation des modèles d'estimation du rayonnement solaire est considérée comme cruciale en raison de la couverture spatiale limitée des mesures au sol. Cette section présente les résultats de validation du GHI horaire estimé à partir du HC3 et du modèle de ciel clair McClear avec des mesures au sol de quatre stations au Maroc. Les performances de GHI estimé à partir de HC3 ont été évaluées pour les conditions tout ciel et de ciel clair. Pour la validation des modèles sous conditions de ciel clair, seuls les instants de ciel clair sont retenus. Ces instants de ciel clair ont été détectés à partir des mesures de GHI tout ciel en appliquant l'algorithme de détection des instants de ciel clair détaillé dans la section 3.1.

La période des données utilisées pour la validation est de 01/03/2013 au 30/06/2020 pour les sites d'Erfoud, Missour et Zagora est de 01/01/2015 au 30/06/2021 pour le site de Benguerir. Avant d'utiliser ces données, nous avons appliqué les divers tests de contrôle de qualité présenté dans la section 2 pour filtrer toute mesure suspecte. Le choix de HC3 et McClear comme base

de données complémentaires aux données mesurées au sol vient de leurs libres accessibilités. Contrairement à la plupart des modèles d'estimation de l'irradiance solaire, HC3 et McClear font partie d'un service en ligne à partir duquel l'utilisateur peut obtenir gratuitement des estimations d'irradiance de McClear pour une période illimitée et pour une période d'essai pour HC3. Les données sont disponibles via une interface conviviale sur le portal de Soda[3].

La validation est faite par rapport à la norme d'ISO [229] dans laquelle les différences sont obtenues en soustrayant les estimations des mesures effectuées au sol. Les ensembles de différences sont résumés par le MBE, le RMSE et le STD. Les pourcentages de ces quantités sont calculés par rapport à la moyenne du GHI sur un site donné. Les coefficients de corrélation *CC* sont également calculés. Les graphes de nuages de points entre les mesures et les estimations sont présentés.

3.1. Détection des instants de ciel clair

L'évaluation de la ressource solaire par des modèles de ciel clair est d'une grande importance dans le domaine de l'énergie solaire : vérification des performances des systèmes photovoltaïques dans des conditions stables, évaluation des effets des nuages, détermination des zones géographiques où l'irradiation est plus incertaine, et la prévision avec des caméras du ciel [47]. Toutefois, avant de pouvoir se servir de ces modèles, il faut les confronter à des mesures au sol de hautes performances. La difficulté de cette validation réside dans le fait que les observations sous conditions de ciel clair ne sont pas disponibles à partir des stations radiométriques, et l'irradiance par ciel clair n'est pas mesurée directement, mais plutôt extraite à partir des données du ciel entier. Du fait que lors de la validation des modèles de ciel clair, une présence accidentelle de nuages pendant une période supposée sans nuages pourrait avoir des conséquences négatives conduisant à des résultats erronés. Plusieurs études ont été menées dans ce cadre, mais jusqu'à présent la communauté scientifique n'a pas encore établi une méthode universelle adéquate pour la détection des instants de ciel clair.

Quelques auteurs ont utilisé uniquement le GHI mesuré et le GHI_{cls} estimé par des modèles de ciel clair pour identifier les instants de ciel clair [261,262], d'autres ont fait appel, à des calculs de géométrie solaire tels que l'angle zénithal [57], tandis que d'autres ont utilisé les composantes diffuse et/ou directe, l'angle zénithal et le rayonnement extraterrestre, et l'épaisseur optique des aérosols comme entrée.

Pour notre cas, pour objectif de valider HC3 et McClear sous les conditions de ciel clair, nous avons appliqué l'algorithme proposé par Reno et Hansen [263] pour identifier les instants de ciel clair à partir des mesures pyranométriques du GHI. Cet algorithme nécessite l'utilisation d'un modèle de ciel clair, et le modèle choisi est celui d'Ineichen [264]. Selon les auteurs, ce modèle de ciel clair s'est avéré généralement précis et assez facile à mettre en œuvre. La force de ce modèle est qu'en plus de l'utilisation du rayonnement normal au sommet de l'atmosphère

[3] https://www.soda-pro.com/

($G_0, W/m^2$), de l'angle solaire zénithal (θ_z, rad), l'élévation (h, m) de la masse d'air ($AM, sans\ unité$) comme entrée, il considère également le facteur de turbidité de Linke (TL), qui est un facteur représentant les effets de l'absorption et de la diffusion par les aérosols, et l'absorption par la vapeur d'eau sur le rayonnement solaire.

Le modèle de ciel clair proposé par Ineichen [17] est défini par l'équation suivante :

$$GHI_{cls} = c_{g1} \times G_0 \times \cos(\theta_z) \times \exp(-c_{g2} \times AM \times (f_{h1} + f_{h2}(TL-1))) \times \exp(0.01 \times AM^{1.8}) \qquad (3.14)$$

avec :

$f_{h1} = \exp(-h/8000)$,

$f_{h2} = \exp(-h/1250)$,

$c_{g1} = 5.09 \times 10 - 5 \times h + 0.868$,

$c_{g2} = 3.920 \times 10 - 5 + 0.0387$.

La masse d'air est calculée par l'équation de Kasten and Young [265] définie comme suit :

$$AM = \frac{1}{\cos(\theta_z) + 0.50572(96.07995 - \cos(\theta_z))^{-1.6364}} \qquad (3.15)$$

Pour identifier les instants de ciel clair, l'algorithme se base sur cinq statistiques calculées à partir du GHI mesuré et le GHI_{cls} estimé par le modèle de ciel clair est celui d'Ineichen:

- **La valeur moyenne du GHI sur toute la période** : vérifie la valeur moyenne du GHI, qui devrait être faible dans les conditions de ciel nuageux que dans les conditions de ciel clair
- **La valeur maximale du GHI** : il est possible que les nuages provoquent un GHI plus élevé que le GHI_{cls} par ciel clair lors de certains phénomènes pouvant conduire à une moyenne de GHI proche du GHI_{cls} par ciel clair. Le calcul de la valeur maximale du GHI est le moyen le plus simple d'identifier de tels évènements.
- **L'indice de variabilité** : il est utilisé pour définir la variabilité du rayonnement solaire pour une fenêtre de temps bien précise. L'indice de variabilité VI est défini comme le rapport entre la longueur de la séquence de la série temporelle de GHI en W/m^2 et la longueur de la séquence de la série temporelle du GHI_{cls} par ciel clair sur une période donnée. L'indice de variabilité est défini par :

$$AM = \frac{1}{\cos(\theta_z) + 0.50572(96.07995 - \cos(\theta_z))^{-1.6364}} \qquad (3.16)$$

avec N est la longueur de la série temporelle, et Δt est le pas de temps entre deux valeurs consécutives de GHI.

Un exemple de la longueur de la ligne calculée pour un jour clair et autre nuageux pour les données d'irradiance d'une minute est présenté à la Figure 3-11. Nous pouvons constater que toute variabilité mesurée du GHI conduit à une augmentation de la longueur de ligne du GHI mesuré par rapport à la longueur de la ligne du GHI_{cls} estimé par le modèle de ciel clair.

- **L'écart type pour quantifier la variation du rayonnement solaire**.
- **La différence maximale entre les variations des séries temporelles GHI et GHI_{cls}**: pendant une période de ciel clair, la variation entre les mesures successives doit être similaire à la variation du rayonnement attendue par le modèle de ciel clair.

Nous avons appliqué l'algorithme de Reno et Hansen pour détecter les instants de ciel clair à partir de l'ensemble des données disponibles pour les quatre sites d'études au pas d'une minute. Nous présentons dans la Figure 3-12 les résultats obtenus pour les sites de Benguerir et Zagora à titre d'exemple. Nous pouvons observer que pour les jours complètement clairs où le soleil n'est pas masqué par les nuages, nous obtenons une courbe lisse et tous les instants clairs sont bien détectés par l'algorithme. Pour les jours fortement couverts, il peut y avoir de brèves périodes pendant lesquelles le soleil n'est pas obscurci par les nuages et où nous pouvons observer la présence de quelques instants clairs. Et vice versa, pour les jours clairs où le passage de certains nuages entraîne l'apparition de brèves périodes nuageuses. Comme le montrent les graphes de la Figure 3-12, l'algorithme utilisé réussit à détecter les rares moments de ciel clair qui surviennent lors des journées nuageuses.

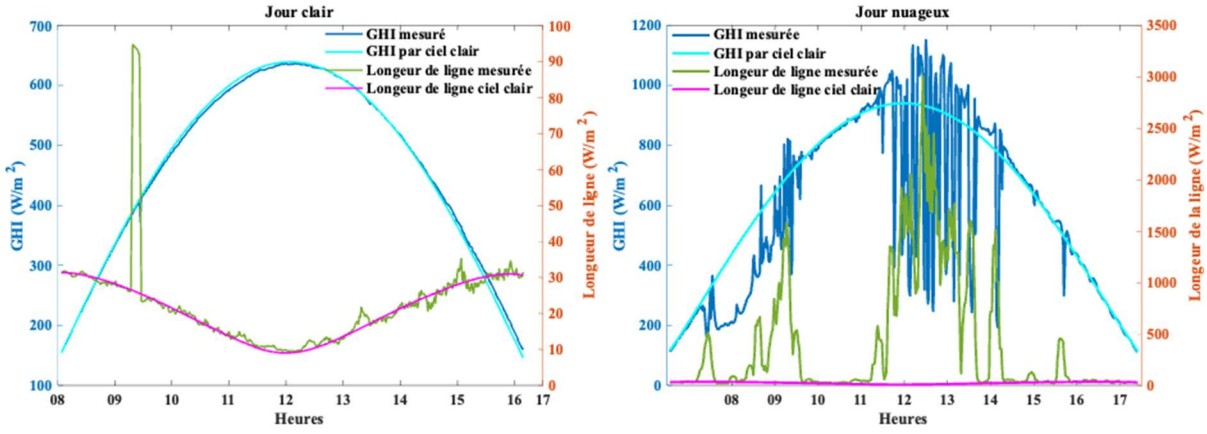


Figure 3-11. Longueur de la ligne du GHI (W/m^2) par jour clair (à gauche) et par jour nuageux (à droite), ligne bleue : GHI, ligne cyan : GHI par ciel clair, ligne magenta : longueur de ligne du GHI mesuré, ligne verte : longueur de ligne du GHI par ciel clair.

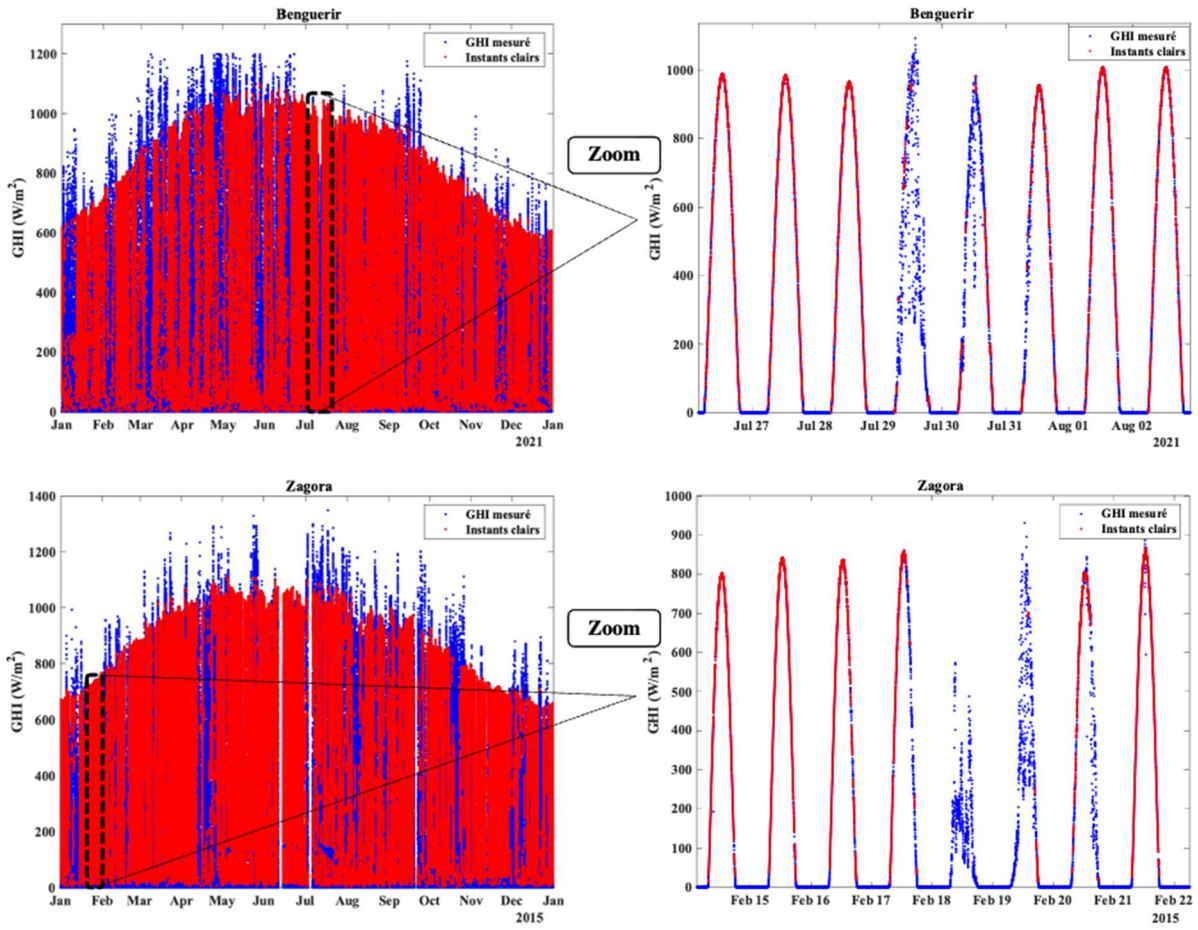


Figure 3-12. Représentation visuelle des résultats de détection des instants de ciel clair pour les données d'une minute de Benguerir (haut) et Zagora (bas). Le bleu représente le GHI mesuré, les pointillés rouges indiquent les minutes identifiées comme claires.

3.2. Résultats de validation

3.2.1. Helioclim3

Comme déjà introduit dans le chapitre 2, HC3 est une base de données fournissant des enregistrements de l'irradiation solaire couvrant l'Europe, l'Afrique, l'océan Atlantique et le Moyen-Orient [266] pour différents pas de temps allant de 1 min à 1 mois depuis 2004. Ces données sont obtenues par la combinaison d'un modèle de ciel clair et d'un indice de nuage obtenue par le traitement des images du satellite Meteosat. Deux versions de Helioclim3 sont disponibles à savoir HC3v4 et HC3v5. La version HC3v4 utilise le modèle ESRA [12] comme modèle de ciel clair avec une climatologie mensuelle de la turbidité [269]. Tandis que HC3v5 [270,271] est une version proposée pour améliorer HC3v4 en utilisant McClear comme modèle de ciel clair qui prend en considération les informations sur le contenu de l'atmosphère claire (les aérosols, la vapeur d'eau et l'ozone). Pour nos applications, nous avons utilisé HC3v5 que nous allons nommer HC3 dans ce qui suit.

La Figure 3-13 et la Figure 3-14 représentent les nuages de points du GHI au sol (axe horizontal) et des estimations de HC3 (axe vertical) pour les quatre sites d'étude pour les

conditions de ciel normal et de ciel clair, respectivement. Le MBE, le RMSE, le STD et les coefficients de corrélation sont indiqués dans le coin supérieur gauche. Les valeurs relatives en % sont obtenues par la division des métriques par la moyenne des valeurs mesurées. Le Tableau 3-4 représente les résultats de validation du GHI horaire estimé par HC3 pour les quatre sites d'étude sous les conditions de ciel normal et de ciel clair.

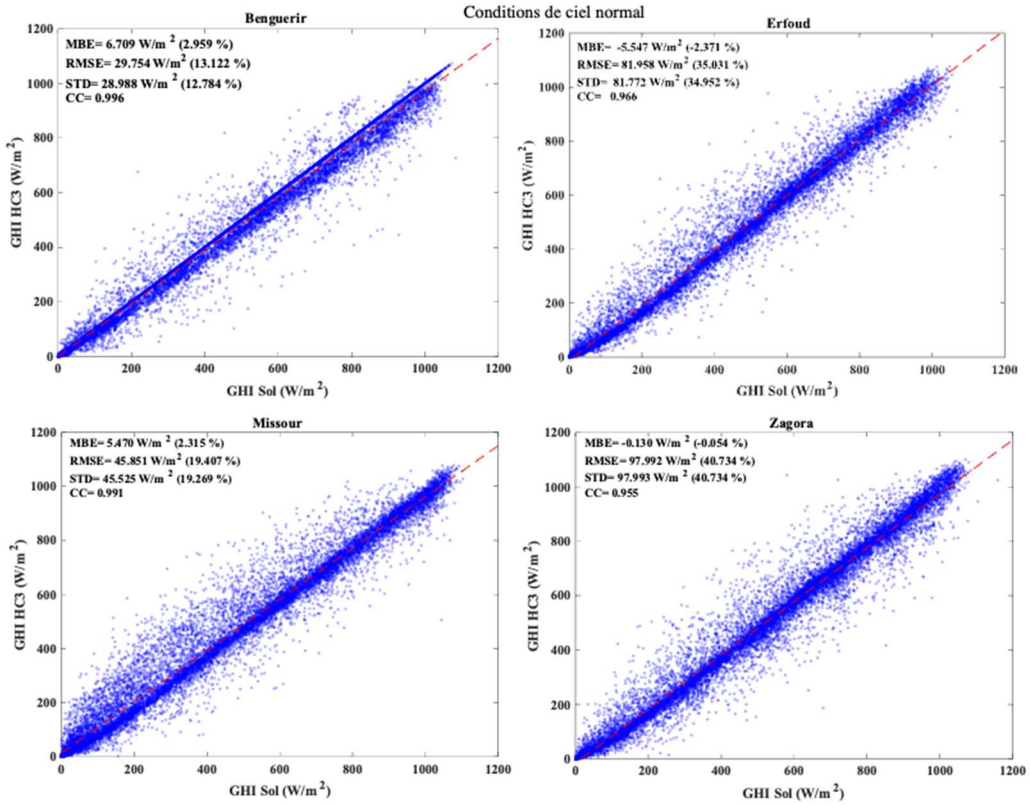


Figure 3-13. Nuages de points des valeurs horaires du GHI mesuré au sol (axe horizontal) et estimé par HC3 (axe vertical) pour les quatre sites d'étude sous conditions de ciel normal.

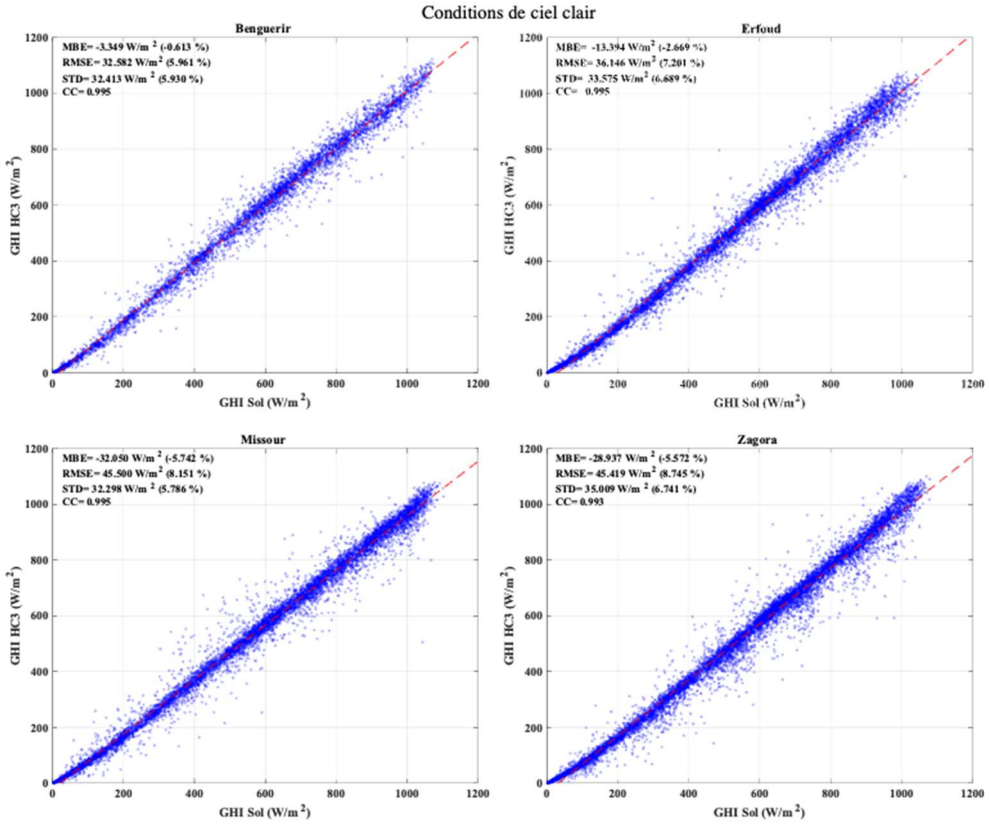


Figure 3-14. Nuages de points des valeurs horaires du GHI mesuré au sol (axe horizontal) et estimé par HC3 (axe vertical) pour les quatre sites d'étude sous conditions de ciel clair.

Tableau 3-4. Résultats de validation de GHI horaire estimé à partir de HC3 pour des conditions de ciel normal et ciel clair.

			Benguerir	**Erfoud**	**Missour**	**Zagora**
Ciel normal (clair+nuageux)	MBE	W/m2 (%)	6.71 2 (2.96)	-5.54 (-2.37)	5.47 (2.31)	-0.31 (-0.05)
	RMSE	W/m2 (%)	29.74 (13.12)	81.96 (35.03)	45.85 (19.41)	97. 99 (40.73)
	STD	W/m2 (%)	28.99 (12.78)	81.77 (34.95)	45.52 (19.27)	97.99 (40.73)
	CC	[-]	0.99	0.96	0.99	0.95
Ciel clair	MBE	W/m2 (%)	-3.35 (-0.61)	-13.39 (-2.66)	-32.05 (-5.74)	-28.93 (-5.57)
	RMSE	W/m2 (%)	32.58 (5.96)	36.14 (7.20)	45.50 (8.15)	45.41 (8.74)
	STD	W/m2 (%)	32.41 (5.93)	33.57 (6.68)	32.29 (5.78)	35.00 (6 .71)
	CC	[-]	0.99	0.99	0.99	0.99

Pour le ciel normal, nous observons que, pour l'ensemble des sites, HC3 montrent des résultats satisfaisants. La plus faible valeur de MBE est observée pour les sites de Zagora avec une sous-estimation de -0.31 W/m^2. Une sous-estimation de -5.54 W/m^2 a été aussi observée sur le site d'Erfoud. Alors que pour Benguerir et Missour, nous observons des surestimations de 6.71 W/m^2 et 5.47 W/m^2, respectivement. Contrairement au MBE, la plus grande valeur de RMSE est obtenue pour le site de Zagora avec une valeur de 97.88 W/m^2. Concernant les autres sites, la plus faible valeur de RMSE a été enregistrée sur le site de Benguerir avec une valeur de 29.74 W/m^2 et a augmenté jusqu'à la valeur de 81.96 W/m^2 pour Erfoud, alors que sa valeur est de 45.85 W/m^2 pour Missour. Des valeurs importantes de coefficient de corrélation variant entre 0.95 et 0.99 ont été obtenues pour la plupart sites. Les importants coefficients de corrélation sont dus au fait que les modèles satellitaires estiment avec une grande précision l'angle solaire zénithal pendant la journée. Ceci entraîne de fortes corrélations entre les mesures au sol et les estimations dérivées du satellite, ce qui dissimule les lacunes potentielles des modèles.

Pour les conditions de ciel clair, aucune amélioration remarquable n'est observée sauf pour le site de Benguerir où le MBE était de 6.71 W/m^2 alors qu'il a diminué à -3.35 W/m^2 sous conditions de ciel clair. Cependant, pour les autres sites, les valeurs de MBE ont augmenté pour atteindre -13.39 W/m^2 contre -5.54 W/m^2 pour Erfoud, -32.05 W/m^2 au lieu de 5.47 W/m^2 pour Missour, et -28.93 W/m^2 contre -0.31 W/m^2 pour Zagora. Sous ciel clair, les valeurs de RMSE sont diminuées pour les sites d'Erfoud, Missour et Zagora avec des valeurs de 36.14 W/m^2, 45.40 W/m^2, et 45.41 W/m^2 contre des valeurs de 81.96 W/m^2, 45.85 W/m^2, et 77.99 W/m^2 pour les trois sites, respectivement. Alors que pour le site de Benguerir, le RMSE a augmenté de 29.74 W/m^2 à 32.58 W/m^2. Les coefficients de corrélation sont de 0.99 pour l'ensemble des sites.

Les résultats de validation varient d'un site à l'autre. Cependant, HC3 a montré d'excellentes performances mêmes sous conditions compliquées où ces performances sous ciel normal surpassent celle obtenue sous conditions de ciel clair. Cela est dû aux conditions climatiques du Maroc qui est caractérisé par la dominance des instants de ciel clair. Pour l'ensemble des sites, les valeurs de MBE sont inférieures à 3 % pour les conditions de ciel normal et inférieures à 6% pour le ciel clair. Les valeurs de RMSE varient entre 30% et 40% pour le ciel normal et entre 5% et 9% pour le ciel clair. Alors que les coefficients de corrélation sont de l'ordre de 0.99 pour la majorité des sites.

3.2.2. McClear

La modélisation de l'irradiance du ciel clair représente un aspect fondamental pour l'évaluation de la ressource solaire en tout lieu, et intervient dans un grand nombre d'applications. Le contrôle de qualité des données mesurées, le choix des sites approprié pour l'adoption des centrales solaires dans des conditions stables, la prévision de la production PV des centrales pour des conditions de fonctionnement stables, la prévision à très court terme par des caméras hémisphériques sont parmi plusieurs applications des modèles de ciel clair dans le domaine de l'énergie solaire. De ce fait, ces modèles de rayonnement par ciel clair doivent être évalués

régulièrement par la comparaison avec des observations au sol de haute qualité afin de garantir leur qualité et de définir le domaine de validité de ces simplifications. Tout au long de cette thèse, nous avons utilisé le modèle McClear pour la modélisation du rayonnement par ciel clair.

McClear est un modèle physique d'estimation de l'irradiance solaire de surface dans des conditions de ciel clair [272]. Le modèle est basé sur le modèle de transfert radiatif libRadtran [273] qui exploite les propriétés optiques de l'atmosphère claire (y compris les profondeurs optiques partielles et totales des aérosols -AOD- à différentes longueurs d'onde, la teneur totale en vapeur d'eau et en ozone) fournies par CAMS [274]. McClear admis comme entrées l'angle solaire zénithal, l'albédo du sol, l'élévation du sol, l'altitude au-dessus du sol, les profondeurs optiques partielles des aérosols à 550 nm pour différentes espèces d'aérosols (matière organique, sel marin, poussière, carbone noir, etc.), le contenu total de la colonne d'ozone et de vapeur d'eau, et les profils atmosphériques. Pareille à HC3, ce modèle fournit des séries temporelles à n'importe quel endroit depuis 2004, avec un retard de 2 jours, dont la résolution temporelle peut aller jusqu'à 1 min (interpolation).

Les résultats de validation de McClear pour les instants de ciel clair sont représentés dans le Tableau 3-5 et sous forme de nuages de points à Figure 3-15 pour les quatre sites d'étude.

Comme le montre la Figure 3-15, les points sont majoritairement alignés avec une dispersion très limitée. Pour l'ensemble des sites, le MBE varie entre -0.15% et 3.22%, sa plus faible valeur est obtenue pour le site de Missour avec une valeur de -0.84 W/m^2 alors qu'une valeur de 16.14 W/m^2 a été obtenue pour Erfoud. Pour les RMSE des valeurs de l'ordre de 6% ont été obtenues, un RMSE de 31.84 W/m^2 a été obtenu pour Zagora puis il a augmenté à 35.10 W/m^2 pour Erfoud. Pour Benguerir et Missour, les RMSE sont de 33.82 W/m^2 et 32.75 W/m^2, respectivement. Les coefficients de corrélation sont de l'ordre de 0.99 pour les quatre sites, ce qui signifie que les changements temporels de GHI sont bien reproduits par McClear.

Nous pouvons remarquer que McClear ne fournit pas de meilleures performances par rapport à HC3 pour les conditions de ciel clair et leurs performances sont plus au moins similaires surtout par rapport aux RMSE et aux coefficients de corrélation. Cela pourrait être expliqué par le fait que HC3 repose sur McClear combiné avec l'indice des nuages pour l'estimation de GHI. Lorsqu'un algorithme de détection de ciel clair bien précis est utilisé, HC3 et McClear peuvent fournir les mêmes performances.

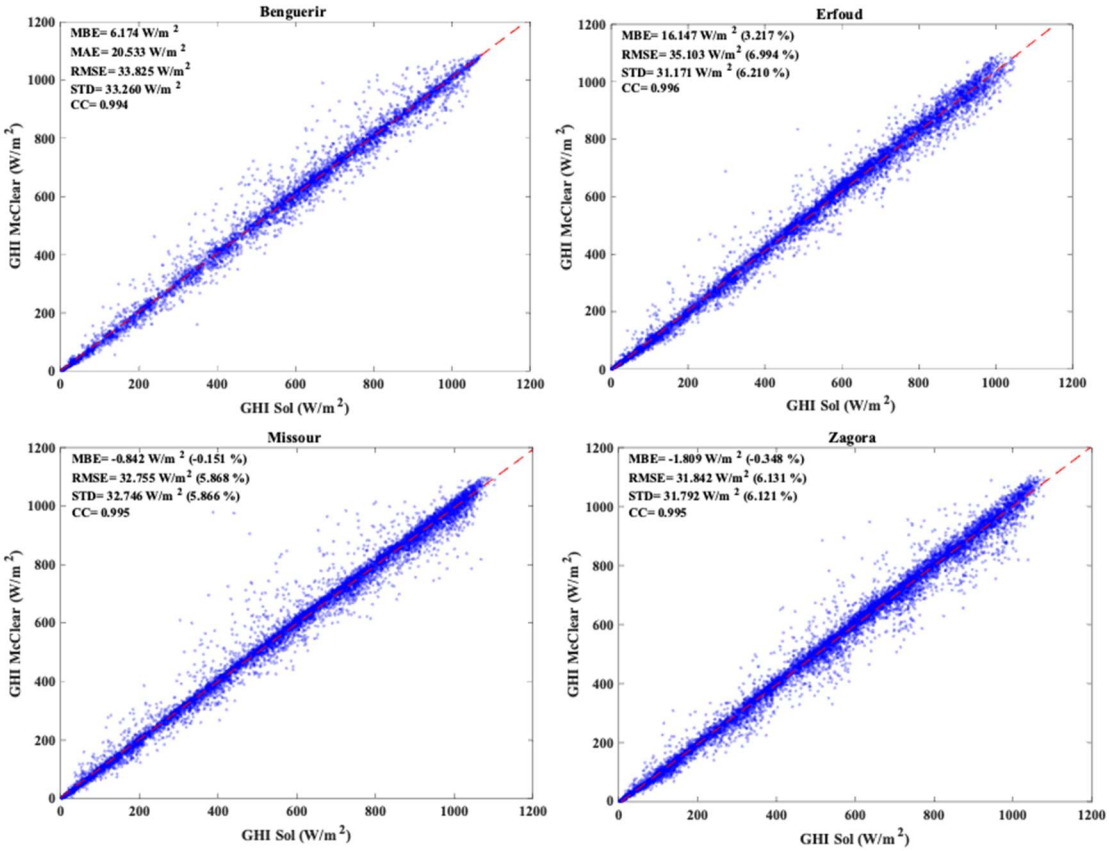


Figure 3-15. Nuages de points des valeurs horaires du GHI mesuré au sol (axe horizontal) et estimé par McClear (axe vertical) pour les quatre sites d'étude sous conditions de ciel clair.

Tableau 3-5. Résultats de validation de GHI horaire estimé à partir de McClear pour des conditions de ciel clair.

			Benguerir	**Erfoud**	**Missour**	**Zagora**
Ciel clair	MBE	W/m2 (%)	6.17 (1.13)	16.14 (3.22)	-0.84 (-0.15)	-1.81 (-0.35)
	RMSE	W/m2 (%)	33.82 (6.18)	35.10 (6.99)	32.75 (5.86)	31.84 (6.13)
	STD	W/m2 (%)	32.26 (6.08)	31.71 (6.21)	32.74 (5.86)	31.79 (6.12)
	CC	[-]	0.99	0.99	0.99	0.99

4. Données de prévision de GFS

Au-delà des variables météorologiques collectées à partir des stations installées au sol, nous avons également fait appel aux prévisions météorologiques pour les besoins de prévision. Nous avons choisi le GFS qui est un modèle de numérique de prévision de temps opéré par NCEP (National Centers for Environmental Prediction). Le GFS est l'un des modèles de prévision météorologique mondiaux les plus connus, en raison de son accès gratuit et de sa couverture mondiale [275]. Les prévisions GFS couvrent des dizaines de variables atmosphériques dans une grille avec une résolution horizontale 0.5ºx0.5º. GFS est exécuté quatre fois par jour à 00,

06, 12 et 18 UTC (temps universel coordonné) et les prévisions sont disponibles jusqu'à 2 semaines à l'avance.

Les prévisions historiques sont disponibles gratuitement depuis janvier 2015 [276]. Les ensembles de données d'archives GFS peuvent être récupérés sur le site web du NOAA - Operational Model Archive and Distribution System (NOMADS)[4] et prétraités à l'aide de rNOMADS qui est un package R open source (Bowman et Lees, 2015). Dans notre cas, Mines ParisTech, nous a permis un accès automatique via le logiciel Matlab aux données GFS prétraitées avec une résolution de 0.25° et avec des pas temporels entre 1 minute et 1 mois jusqu'à 15 jours à l'avance. Les variables requises sont la température, l'humidité relative, la pression barométrique, la vitesse et la direction du vent.

Conclusion

Ce chapitre clôture l'ensemble des connaissances et des outils à acquérir au préalable pour développer des modèles de prévision du rayonnement solaire.

Dans la première partie de ce chapitre, nous avons présenté les quatre sites d'études et l'ensemble des données qui nous seront intéressantes pour le développement du modèle de prévision. Les mesures au sol de l'irradiance sont précieuses pour de nombreux domaines de recherche des applications énergétiques. Cependant, ces stations météorologiques au sol peuvent présenter des données de rayonnement discutables. Dans la deuxième partie de ce chapitre, nous avons présenté une routine complète impliquant plusieurs procédures standards pour le contrôle de qualité des données de rayonnement solaire. La procédure s'est basée sur l'inspection visuelle des données pour l'identification des valeurs manquantes, des problèmes de décalages temporels, etc., des tests des enveloppes pour de chaque composante, et les tests de cohérence entre les différentes composantes du rayonnement. Le processus de qualité a été mis en œuvre aux mesures au pas de 10 minutes du rayonnement solaire collectées à partir de deux stations pyranométriques (Benguerir et Erfoud).

En général, à l'exception de quelques cas la qualité des données brutes des deux stations est jugée excellente. Les sources d'erreurs dans les données peuvent être dues à des problèmes de la réponse en cosinus du capteur, au désalignement des disques d'ombrage des capteurs diffus, à la présence de gouttelettes d'eau ou de fientes d'oiseaux sur les dômes des capteurs ou bien aux problèmes au niveau du capteur lui-même. Pour les valeurs manquantes enregistrées pour l'ensemble des stations, elles sont principalement liées à des coupures de courant, des problèmes de capteurs, ou à la déconnexion de la station météorologique pour des raisons de calibration des capteurs. Ces tests de contrôle de qualité ont été aussi appliqués aux données de Missour et Zagora pour avoir une base de données complète et sans mesures aberrantes. Les ensembles de données examinées ont été utilisés pour la validation du GHI estimé à partir du satellite.

[4] http://nomads.ncep.noaa.gov/

Dans la troisième partie du chapitre, nous avons validé le GHI horaire de deux bases de données à savoir HC3 et le modèle de ciel clair McClear pour les quatre sites d'études pour des conditions de ciel normal et de ciel clair. D'une façon générale, les résultats des comparaisons entre les estimations de HC3, McClear et les mesures de GHI au sol pour les quatre stations montrent qu'une grande corrélation est atteinte. Nous pouvons en conclure d'une part que la qualité des estimations répond en grande partie au critère de la norme OMM, dans laquelle 95 % des erreurs ne doivent pas dépasser 20 W/m^2 [33]. D'autre part, les bases de données ne présentent pas de large changement géographique où des corrélations élevées ont été obtenues pour l'ensemble des sites. La chose qui montre la capacité de HC3 et McClear à capturer la variabilité temporelle et spatiale du champ d'irradiance. Au final, nous pouvons assumer que les deux bases de données sont des sources fiables des données du rayonnement solaire pour le Maroc et peuvent être utilisées comme alternatives aux mesures pyranométriques.

Enfin, nous pourrions conclure qu'il est difficile de définir un modèle universel qui s'appliquerait à tous les cas de figures (comme souvent en prédiction). Nous avons essayé de manipuler tous les outils et les données disponibles pour finalement développer un modèle de prévision précis et portable. Dans le chapitre prochain, nous présentons les diverses expérimentations effectuées durant cette thèse pour la prévision du GHI.

Chapitre 4

Résultats de prévision du rayonnement solaire à court terme et à très court terme

Introduction **91**

1. Préparation des données et climatologie des sites d'études **91**

1.1. Climatologie des sites d'étude 92

1.1.1. Variabilité saisonnière 92

1.1.2. Indice de ciel clair journalier $\boldsymbol{kcs}$ 93

1.1.3. Variabilité intra-journalière horaire $\boldsymbol{\delta(kcs)}$ 94

1.2. Classification de l'irradiance solaire 94

2. Prévision à court terme **97**

2.1. Prévision du rayonnement solaire avec les réseaux de neurones 97

2.1.1. Prétraitement des données 99

2.1.2. Sélection des entrées pertinentes 101

2.1.3. Le modèle MLP 103

2.1.4. Le modèle LSTM 108

2.2. Prévision avec le modèle statistique SARIMA 109

2.3. Prévision à très court terme 111

2.3.1. Site d'étude et prétraitement de données 111

2.3.2. Architecture du modèle CNN-MLP 113

3. Résultats de prévisions **115**

3.1. Prévision du GHI 12 heures à l'avance 115

3.1.1. Analyse des performances prédictives des modèles de prévision en fonction du type de jour (clarté, et variabilité) 118

3.1.2. Analyse de la performance prédictive en fonction des saisons 123

3.1.3. Impact des aérosols et de la couverture nuageuse sur la précision des prévisions du GHI 129

3.2. Prévision à très court terme 134

Conclusion **139**

Introduction

La prévision de l'irradiance solaire représente un élément clé pour renforcer l'intégration future de l'énergie solaire dans les structures d'approvisionnement en énergie existantes et réduire la dépendance aux combustibles fossiles. Cependant, son caractère fluctuant a une influence significative sur l'énergie produite et une utilisation efficace de cette énergie doit tenir en compte ce comportement stochastique avec des stratégies d'exploitation anticipatives. Les méthodes de prévision présentées dans le chapitre 3 nous ont permis de choisir les techniques les plus recommandées et d'entreprendre plusieurs manipulations pour prévoir le GHI horaire à court et à très court terme.

Ce chapitre est le cœur de notre travail puisque nous allons y présenter les résultats des modèles développés au cours de cette thèse pour la prévision du GHI à court et à très court terme. La première partie de ce chapitre présente une caractérisation des sites d'étude par rapport à la variabilité et à la couverture nuageuse pour savoir la climatologie dominante à chaque site. La deuxième partie présente les résultats expérimentaux de chaque modèle de prévision. Nous présenterons dans un premier temps l'ensemble des étapes de construction des modèles développés. Ces étapes comportent le prétraitement des données, la sélection des caractéristiques d'entrée pertinentes, le choix de l'architecture du réseau et l'optimisation de la structure neuronale. Les performances prédictives des modèles proposés ont été évaluées sur une année entière, sous diverses saisons et conditions de ciel. Quatre ensembles de données collectées à partir de quatre stations météorologiques installées au sol au Maroc sont exploités pour l'évaluation des prévisions. En parallèle, les performances des modèles de prévision ont été comparées aux modèles de référence en se basant sur les métriques d'évaluation les plus connues.

Deux études de cas vont être successivement mises en œuvre :

- La prévision du GHI horaire pour le court terme (une journée à l'avance), en se basant sur des techniques de l'intelligence artificielle (MLP et LSTM) et en exploitant les données météorologiques collectées au sol et des calculs de la géométrie solaire pour construire les modèles. Les performances prédictives des modèles développés seront comparées au modèle statistique SARIMA et au modèle de la persistance.
- La prévision du GHI à très court terme (15 minutes à l'avance), dans ce cas, nous avons développé un modèle hybride CNN-MLP. Ce modèle exploite en plus des paramètres météorologiques les images d'une caméra hémisphérique installée au sol pour extraire les informations sur la couverture nuageuse. Les performances prédictives du modèle sont aussi comparées au modèle référence de la persistance.

1. Préparation des données et climatologie des sites d'études

L'évaluation de la précision des modèles de prévision est réalisée sur la base des mesures effectuées dans les quatre stations présentées dans le chapitre 3. Les données ont une résolution temporelle d'une heure pour l'horizon du court terme et d'une minute pour le très court terme.

1.1. Climatologie des sites d'étude

Étant donné que la production des systèmes photovoltaïques dépend essentiellement des conditions météorologiques, la caractérisation de la climatologie dominante dans un site donné est une étape fondamentale. La variabilité est principalement causée par deux facteurs : la position du soleil et la couverture nuageuse. Le premier facteur est déterministe et peut être modélisé par des équations physiques [277]. En revanche, le second représente les variations stochastiques qui sont imprévisibles et ont un impact important sur l'incertitude des prévisions. La couverture nuageuse représente la partie la plus critique à considérer lors de la prévision de l'irradiance solaire [278]. Cependant, la quantification de la variabilité de l'irradiance solaire due aux nuages constitue un élément important pour une intégration réussite de l'énergie produite dans le réseau électrique.

Un schéma de classification du GHI par rapport aux conditions nuageuses peut être utile pour différentes applications de l'énergie solaire. À savoir,

- l'évaluation des modèles de prévision de l'irradiation solaire sous différentes conditions nuageuses, ce qui permet de réduire l'incertitude des prévisions et assurer des prédictions précises. Ceci permet une utilisation efficace de la production d'énergie fluctuante des systèmes PV,
- la comparaison des caractéristiques de différents sites solaires,
- la détermination de la météorologie dominante dans un site donné,
- et enfin la sélection de la technologie solaire la plus appropriée pour un site en fonction des résultats de la classification.

Dans cette section, nous allons présenter une caractérisation des sites d'étude par rapport à la variabilité intra-journalière du GHI causée par les nuages afin de déterminer la climatologie prédominante dans chaque site. L'étude de la climatologie des sites par rapport à la variabilité due aux nuages nous permettra de tester les modèles de prévisions sous différentes situations climatiques et pouvoir améliorer les performances des prévisions. En particulier, sous les conditions de ciel nuageux et variable, pour finalement choisir le modèle le plus approprié pour chaque condition de ciel.

Nous avons effectué une étude de la climatologie par rapport aux conditions nuageuses et au degré de stabilité de la nébulosité tout au long de la journée que nous allons nommer variabilité. La classification de l'irradiance solaire est faite sur la base de l'analyse des critères suivants : (1) l'indice de ciel clair journalier k_{cs} utilisé pour définir les conditions météorologiques d'un jour donné et (2) la variabilité intra-journalière horaire $\delta(k_{cs})$ définie par l'écart type des variations de l'indice de ciel clair.

1.1.1. Variabilité saisonnière

La variabilité saisonnière du rayonnement solaire est causée par les nuages, ainsi que par le cycle solaire journalier et saisonnier. Le cycle solaire journalier est responsable de la forme parabolique de l'irradiation solaire : l'irradiation augmente après le lever du soleil jusqu'à midi

et diminue jusqu'au coucher du soleil. Et le cycle solaire saisonnier est responsable du changement des valeurs de rayonnement d'une saison à l'autre. La variabilité saisonnière est bien observée dans la variation de la moyenne mensuelle de l'irradiance solaire (Figure 4-1).

Étant donné que nous sommes soucieux de quantifier la variabilité causée par les nuages, l'indice de ciel clair k_{cs} est largement utilisé dans la littérature pour séparer les effets des nuages de la variabilité saisonnière de l'irradiation solaire. Ainsi, pour éliminer le cycle solaire saisonnier et quotidien, nous considérerons la variabilité de l'indice de ciel clair $\delta(k_{cs})$, qui représente uniquement la variabilité causée par les nuages.

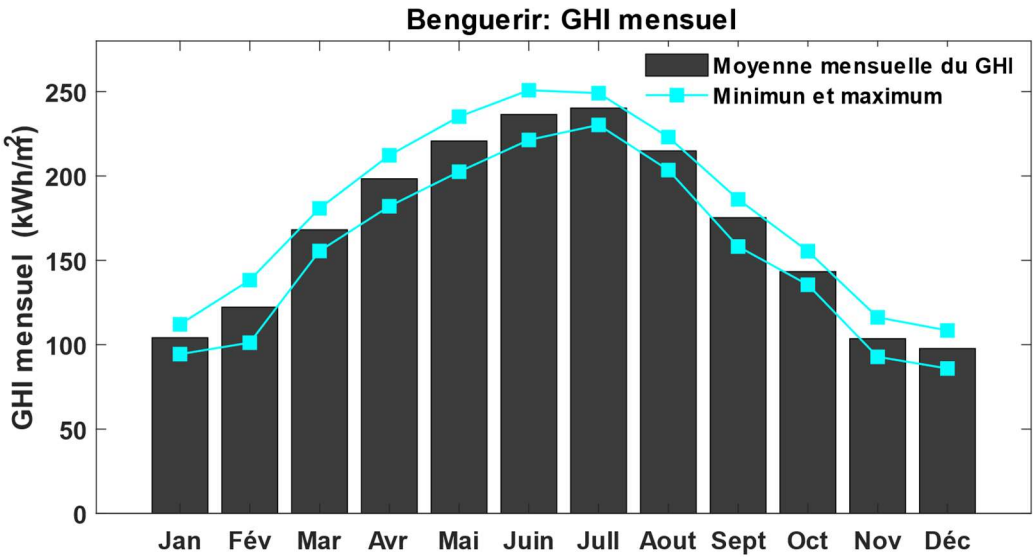


Figure 4-1. Variation de la moyenne mensuelle du GHI journalier (kWh/m^2), tout au long de l'année pour le site de Benguerir.

1.1.2. Indice de ciel clair journalier k_{cs}

Comme déjà mentionné, pour quantifier la variabilité causée par les nuages, il est nécessaire de séparer l'effet des nuages de la variabilité déterministe saisonnière et diurne de la série temporelle du GHI. Pour faire, nous calculons l'indice de ciel clair k_{cs} définie par le rapport entre l'irradiation solaire GHI et l'irradiation modélisée de ciel clair GHI_{cls}. Pour le rayonnement par ciel clair GHI_{cls}, nous avons utilisé celui estimé par le modèle de ciel clair McClear que nous avons déjà analysé ces performances dans le chapitre 3. L'indice de ciel clair journalier est défini par l'équation suivante :

$$k_{cs} = \frac{\sum_{n=1}^{N} GHI(n)}{\sum_{n=1}^{N} GHI_{cls}(n)} \tag{4.1}$$

Avec N est le nombre des heures éclairées de la journée.

1.1.3. Variabilité intra-journalière horaire $\delta(k_{cs})$

Différentes méthodes existantes permettent de quantifier la variabilité du rayonnement solaire par rapport aux nuages. Dans cette étude, nous avons utilisé la méthode proposée dans [279] qui définit l'indice de variabilité $\delta(k_{cs})$ comme l'écart type de k_{cs} dans un intervalle de temps Δt,

$$\delta(k_{cs}) = \sqrt{\frac{1}{N}\sum_{n=1}^{N}\left(\Delta k_{cs}\right)}, \tag{4.2}$$

avec

$$\Delta k_{cs} = k_{cs}(n+1) - k_{cs}(n)$$

Pour calculer la variabilité intra-journalière horaire, nous avons appliqué l'équation aux valeurs horaires du k_{cs} sur une journée.

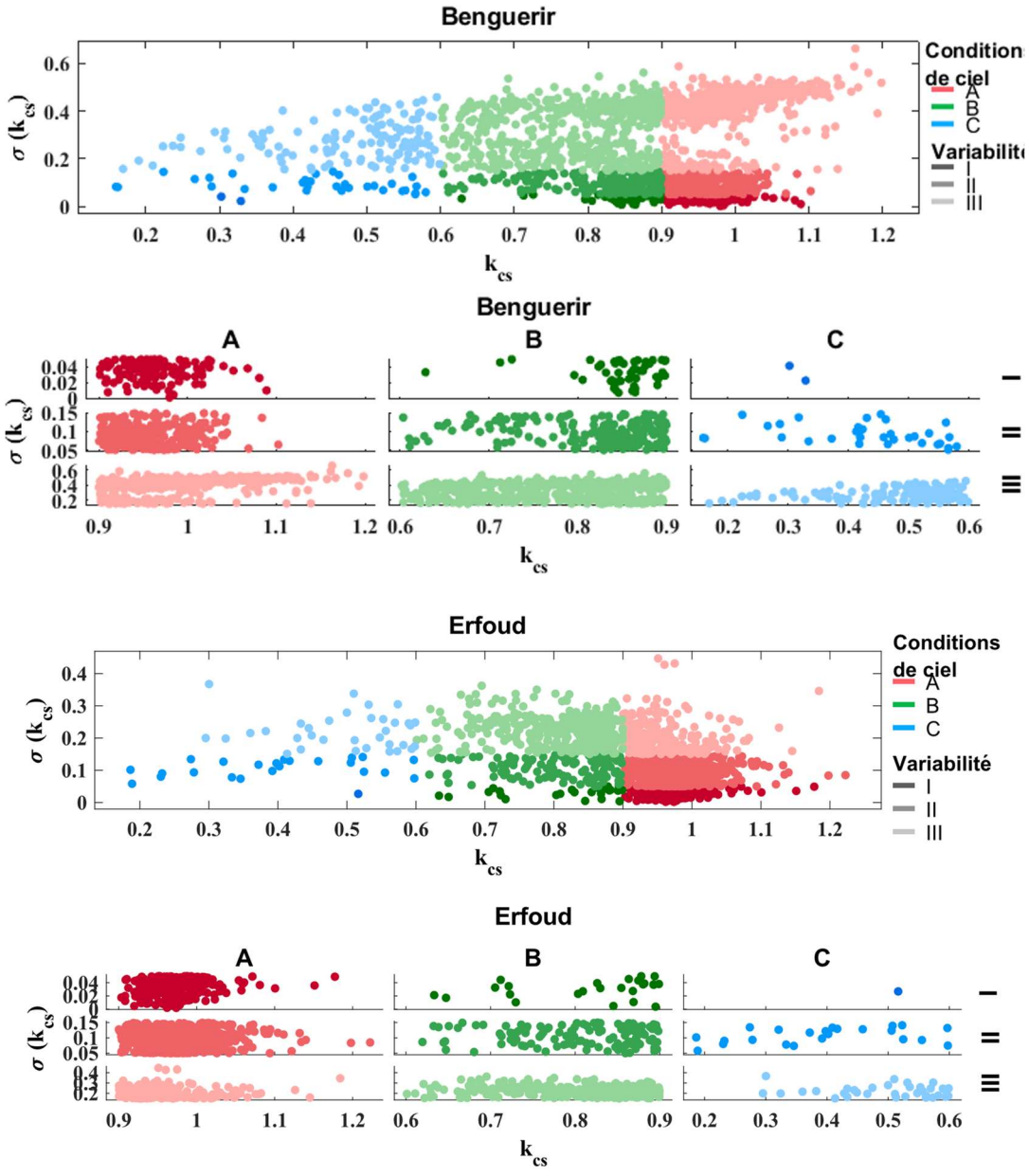
1.2. Classification de l'irradiance solaire

En se basant sur les deux critères k_{cs} et $\delta(k_{cs})$ nous obtenons un plan de classification du GHI délimitant un espace bidimensionnel présentant la série temporelle de l'indice de ciel clair journalier k_{cs} en fonction de la variabilité intra-journalière $\delta(k_{cs})$. Les classes obtenues sont également des indicateurs de l'état des nuages.

L'espace de classification est dévissé en neuf zones où A, B et C représentent le degré de nébulosité d'une journée, la zone A étant la plus claire et la zone C la plus nuageuse. D'autre part, les zones I, II et III reflètent la stabilité de la nébulosité ou la variabilité tout au long de la journée. Par conséquent, les jours dans la zone CIII représentent des jours nuageux avec une forte variabilité et elles sont les plus difficiles à prévoir. Cependant, les jours dans la zone AI sont plus faciles à prévoir comme ils représentent des jours clairs avec une faible variabilité.

Pour la classification des jours, nous nous sommes servis de l'ensemble des données disponibles pour les quatre sites d'études. La Figure 4-2 représente les résultats de classification pour les sites cibles. L'axe des *x* représente la moyenne journalière de l'indice de ciel clair k_{cs} qui présente les trois classes correspondantes au degré de nébulosité : ciel clair (A), ciel mixte (B) et ciel nuageux (C). L'axe des *y* représente les $\delta(k_{cs})$ désignant le degré de la variabilité dû aux nuages : faible variabilité (I), variabilité moyenne (II) et forte variabilité (III). Les nuances de gris indiquent le degré de la variabilité journalière, le rose, le vert et le bleu correspondent aux valeurs de k_{cs} pour des conditions respectivement claires, mixtes et nuageuses. La Figure 4-3 illustre le pourcentage des jours correspondant à chaque classe pour les quatre sites. Les résultats de classification soulignent la dominance des jours clairs et mixtes, alors que les conditions fortement nuageuses se produisent à moins de 1.3 % de temps. Les conditions de ciel clair avec une faible variabilité ne présentent que 2.5% de jours, alors que les conditions de ciel clair avec une variabilité moyenne (AII) se produisent à environ 32% du temps. Cette situation est tout à fait normale, en effet il est très rare d'avoir un ciel clair sans

aucun passage nuageux même pour une courte durée. Les conditions nuageuses pour les classes AII et AIII sont généralement causées un passage ou une dispersion des nuages, ce qui provoque une instabilité de la nébulosité d'où le passage d'un ciel non variable à un ciel variable. Les conditions de ciel mixte avec une faible variabilité (BI) sont relativement faibles et ne représentent que 1% des jours. Alors que les conditions de ciel mixte avec une variabilité moyenne (BIII) et le ciel mixte avec une forte variabilité (BIII) présentent environ 23% et 10%, respectivement. Les conditions de ciel nuageux présentent le plus faible pourcentage, avec des valeurs de 0.1% pour le CI, 1.3% pour CII et 2.7% pour CIII. Une telle classification nous sera utile pour évaluer la précision des modèles de prévision sous diverses conditions météorologiques et notamment pour des conditions de ciel les plus dominantes.



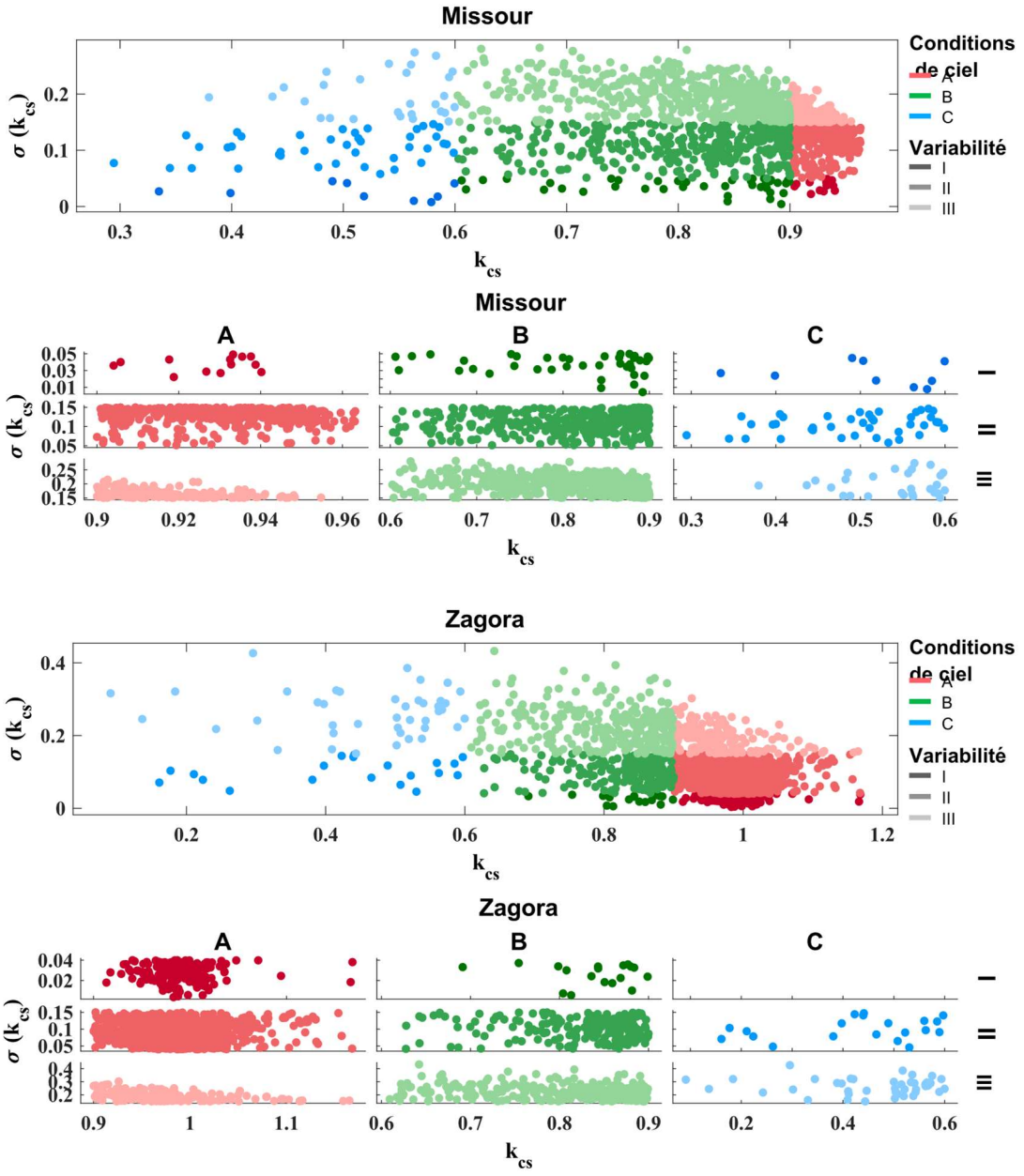


Figure 4-2. Classification du rayonnement selon la nébulosité et la variabilité dans l'espace $k_{cs} - \delta(k_{cs})$ tout a long de la période d'étude pour les sites cibles.

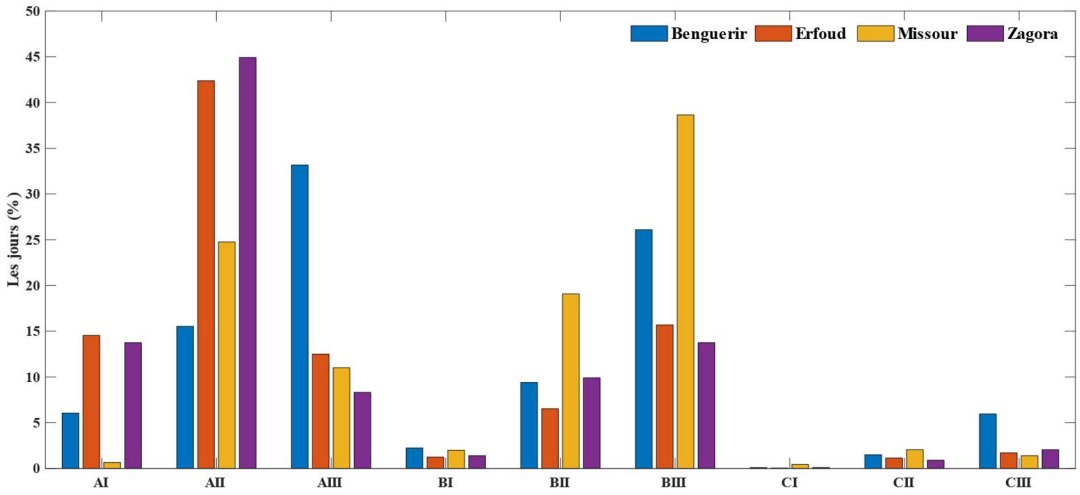


Figure 4-3. Pourcentages des jours correspondants à chaque classe pour les quatre sites.

2. Prévision à court terme

Comme l'a montré l'étude bibliographie, le court terme est l'horizon de prévision le plus important pour la gestion de l'énergie produite. Dans ce contexte, nous nous sommes intéressés à la prévision du GHI à court terme à savoir une journée à l'avance (h+12 sans valeurs nocturnes), pour cela, deux modèles basés sur le MLP et autre un modèle basé sur LSTM ont été développés. Ces trois modèles sont comparés au modèle statistique SARIMA et au modèle de persistance afin de montrer le rôle complémentaire des RNA aux approches statistiques classiques dans l'estimation des relations non linéaires entre les caractéristiques d'entrée et la variable cible. Les étapes de développement de chaque modèle de prévision pour le court terme sont présentées dans la section suivante.

2.1. Prévision du rayonnement solaire avec les réseaux de neurones

Du fait de leurs succès dans le domaine de la prévision solaire et comme l'a montré l'étude bibliographique présentée dans le chapitre 3, les techniques de l'intelligence artificielle présentent un fort intérêt pour améliorer les prévisions de la ressource solaire. Nous avons testé deux techniques de l'IA à savoir le MLP et LSTM pour la prévision du GHI horaire une journée à l'avance.

Dans cette partie, nous présentons les étapes derrière le développement de chaque modèle de prévision. Dans un premier temps, nous décrivons la procédure adoptée pour la sélection des caractéristiques pertinentes pour l'entrée des modèles à partir d'un ensemble de données disponibles. Pour la suite, nous choisissons l'architecture optimale des modèles, c'est-à-dire le choix des meilleures combinaisons des variables d'entrée, les fonctions de transfert, ainsi que le nombre optimal de neurones dans la couche cachée. Pour évaluer la précision des modèles RNA, nous avons comparé leurs prévisions avec le modèle statistique SARIMA et le modèle de référence représenté par le modèle de persistance.

Pour la mise en œuvre des modèles de prévision développés, nous avons utilisé Matlab [280] qui est un logiciel de calcul numérique. Matlab possède une boîte à outils appelée 'Deep Learning Toolbox™' permettant la mise en œuvre, la conception, la formation, l'apprentissage, la visualisation, et la simulation des RNA. La conception d'un modèle basé sur les RNA passe par les étapes suivantes :

1. Le choix du type de réseau de neurones.
2. Le choix des entrées pertinentes.
3. Le prétraitement des entrées.
4. Le choix de nombres de couches cachées et le nombre de neurones pour chaque couche.
5. La définition des fonctions de transfert pour l'entrée et la sortie.
6. Le choix de l'algorithme d'apprentissage.
7. Le choix de la méthode d'ajustement des poids et des biais.
8. Initialisation des poids et du biais.
9. Le choix de la méthode d'arrêt de l'apprentissage pour éviter le surapprentissage (arrêt prématuré, seuils d'arrêt, etc.).
10. La définition de la fonction coût utilisée durant l'apprentissage.
11. Définition de l'indicateur statistique du test.
12. La répartition des données en données d'apprentissage, de validation, et de test.

La Figure 4-4 schématise la démarche retenue pour la prévision du GHI avec le MLP, LSTM, et le CNN.

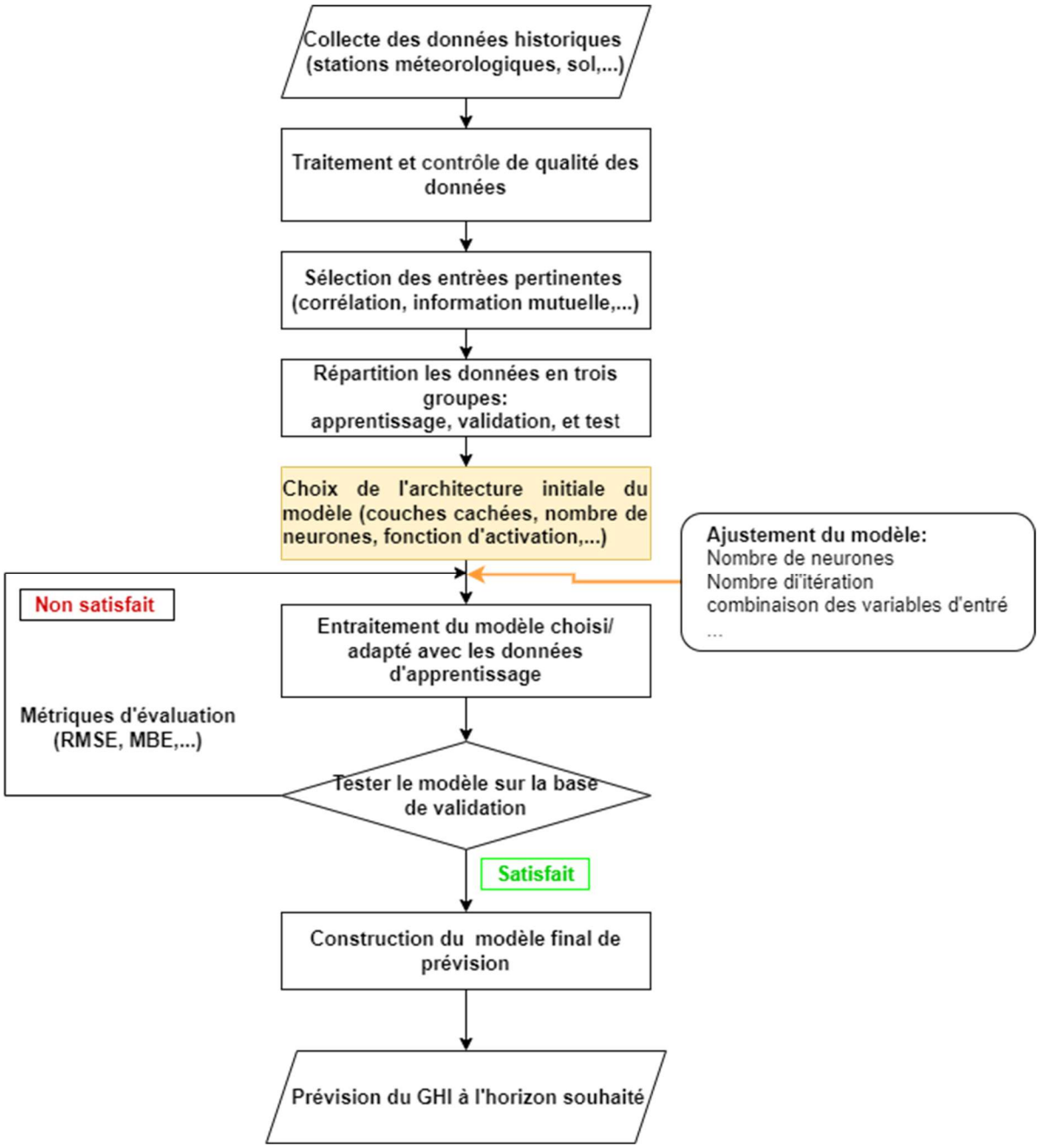


Figure 4-4. Organigramme de la démarche RNA utilisé pour la prévision du GHI.

2.1.1. Prétraitement des données

Le succès d'un algorithme de RNA dépend de la qualité des données d'entrée. Selon [281], pour obtenir des prévisions précises du GHI, les données doivent couvrir au moins une année et de préférence deux années consécutives ou plus. Toutes les données doivent être de haute qualité, avec des intervalles minimums de données manquantes et des considérations de contrôle de qualité. Les données utilisées dans cette thèse ont été prétraitées comme suit :

- Contrôle de la qualité : les données collectées à partir des stations météorologiques peuvent contenir des valeurs aberrantes même si la station est de hautes performances.

Nous avons appliqué à l'ensemble des données les algorithmes de contrôle de qualité détaillés dans le chapitre 2, où toutes les valeurs anormales ou aberrantes ont été exclues.

- Valeurs nocturnes: en raison de problèmes causés par une réponse inadéquate des pyranomètres [282], les heures nocturnes ont été supprimées. Nous avons appliqué un filtre aux données pour ne garder que les observations entre 07:00 et 19:00.
- La stationnarité : la série temporelle du GHI est caractérisée par certaines propriétés telles que la saisonnalité et la périodicité. La méthode SARIMA utilisée dans notre étude exige la stationnarité de la série temporelle. La procédure utilisée pour la stationnarisation de la série temporelle du GHI est détaillée dans la section 2.2.
- La répartition des données : la répartition des données est une phase importante de la préparation des données. Les données sont généralement réparties en trois groupes : les données d'apprentissage qui servent pour paramétrer les modèles surtout ceux basés sur les RNA. Les données de validation sont utilisées pour éviter le surapprentissage du modèle de prévision qui arrête l'apprentissage lorsque l'erreur de validation commence à augmenter, c'est ce qu'on appelle le « early stopping ». Et finalement, les données de test qui servent pour évaluer les performances et la fiabilité du modèle une fois paramétré. Ces données de tests n'ont pas été utilisées lors de la conception du modèle. Dans cette étude, nous avons utilisé 80% des données réparties entre l'apprentissage et la validation et 10% pour le test. La Figure 4-5 et le Tableau 4-1 représentent la répartition des données disponibles selon les bases d'apprentissage, de validation, de test, et les heures disponibles pour chaque ensemble.

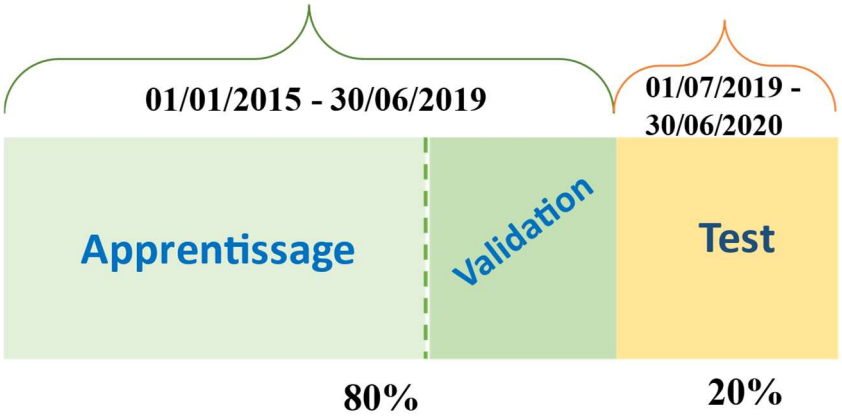


Figure 4-5. Répartition des données pour l'apprentissage, la validation et le test des modèles de prévision.

Tableau 4-1. Répartition des données pour les modèles de prévision et le nombre des heures disponibles pour chaque ensemble.

	Période	Nombres des heures
Données d'apprentissage	01/01/2015 – 31/12/2019	17424
Données de validation	01/01/2019 – 30/06/2019	2127
Données de test	01/07/2019 - 30/06/2020	4380

2.1.2. Sélection des entrées pertinentes

Dans le chapitre 3, nous avons présenté l'ensemble des données disponibles et qui sont supposées déterminantes pour la prévision du GHI. À partir de cet ensemble de données qui combine des paramètres météorologiques et des calculs de la géométrie solaire, il est nécessaire de limiter le nombre de caractéristiques à utiliser en entrées des modèles de prévision pour ne pas alourdir le calcul et considérer des paramètres qui n'ont aucun impact sur le phénomène étudié. Pour la sélection des caractéristiques pertinentes, nous nous sommes basés sur la corrélation de Pearson entre la variable cible GHI et les diverses variables candidates. Le coefficient de corrélation de Pearson (CC) est défini par le rapport entre la covariance de deux variables et le produit de leurs écarts types, il est représenté par l'équation (2.5).

Le CC est compris entre +1 et -1. Une valeur de +1 signifie une corrélation linéaire positive complète entre X et Y. Une valeur de -1 implique que X a une corrélation linéaire complètement négative avec Y. Enfin, une valeur de 0 indique qu'il n'y a pas une corrélation entre X et Y.

La Figure 4-6 et le Tableau 4-2 représentent les résultats de corrélation entre les variables météorologiques, les variables de la géométrie solaire et le GHI pour les quatre sites d'études. Les résultats de corrélations sont proches d'un site à l'autre, pour cette raison les variables sélectionnées pour un site seront utilisées pour l'ensemble de sites. Comme la montre figure, le rayonnement par ciel clair, le rayonnement au top de l'atmosphère, et l'angle solaire zénithal sont largement corrélés avec le GHI avec des coefficients de corrélation supérieurs à 0.94 pour le TOA et le GHI_{cls} , et de l'ordre de 0.85 pour le SZA (θ_z). La température est modérément corrélée avec le GHI avec une corrélation de l'ordre de 0.5. Les autres variables sont faiblement corrélées avec le GHI.

Pour notre modèle de prévision, la sélection des caractéristiques n'est pas faite sur la base de la corrélation uniquement, mais au moyen des tets de plusieurs modèles afin de trouver la combinaison permettant des prévisions précises. Cela s'explique par le fait qu'une caractéristique peut ne pas être pertinente en soi, mais peut apporter une amélioration significative des performances lorsqu'elle est associée à d'autres caractéristiques [188].

Nous disposons en général de 10 variables candidates, ce qui implique 1023 possibilités de combinaison de ces variables en entrée. Du fait que le test du modèle avec toutes ces combinaisons est très difficile, nous avons considéré des combinaisons que nous avons jugé être les plus intéressantes. Le Tableau 4-3 présente les modèles testés.

Tableau 4-2. Les variables candidates en entrée du modèle de réseau de neurones et leurs corrélations avec le GHI.

Type	Variables	Abréviation	CC			
			Benguerir	Erfoud	Missour	Zagora
Météorologique	Température	T	0.55	0.51	0.52	0.55
	Humidité relative	RH	-0.59	-0.51	-0.49	-0.62
	Vitesse du vent	WS	0.17	0.16	0 .32	0.34
	Direction du vent	WD	0.11	0.06	-0.024	0.021

	Précipitations	Preci	-0.076	-0.04	-0.031	-0.05
	Pression barométrique	BP	-0.11	-0.06	-0.071	-0.05
Modèle de ciel clair	Rayonnement par ciel clair	GHI_{cls}	0.95	0.99	0.99	0.98
Géométrie solaire	Rayonnement au top de l'atmosphère	TOA	0.94	0.98	0.98	0.97
	L'angle solaire zénithal	$SZA\ (\theta_z)$	-0.85	-0.87	-0.87	-0.84
	Durée d'insolation	DL	0.001	-0.003	-0.00	-0.002

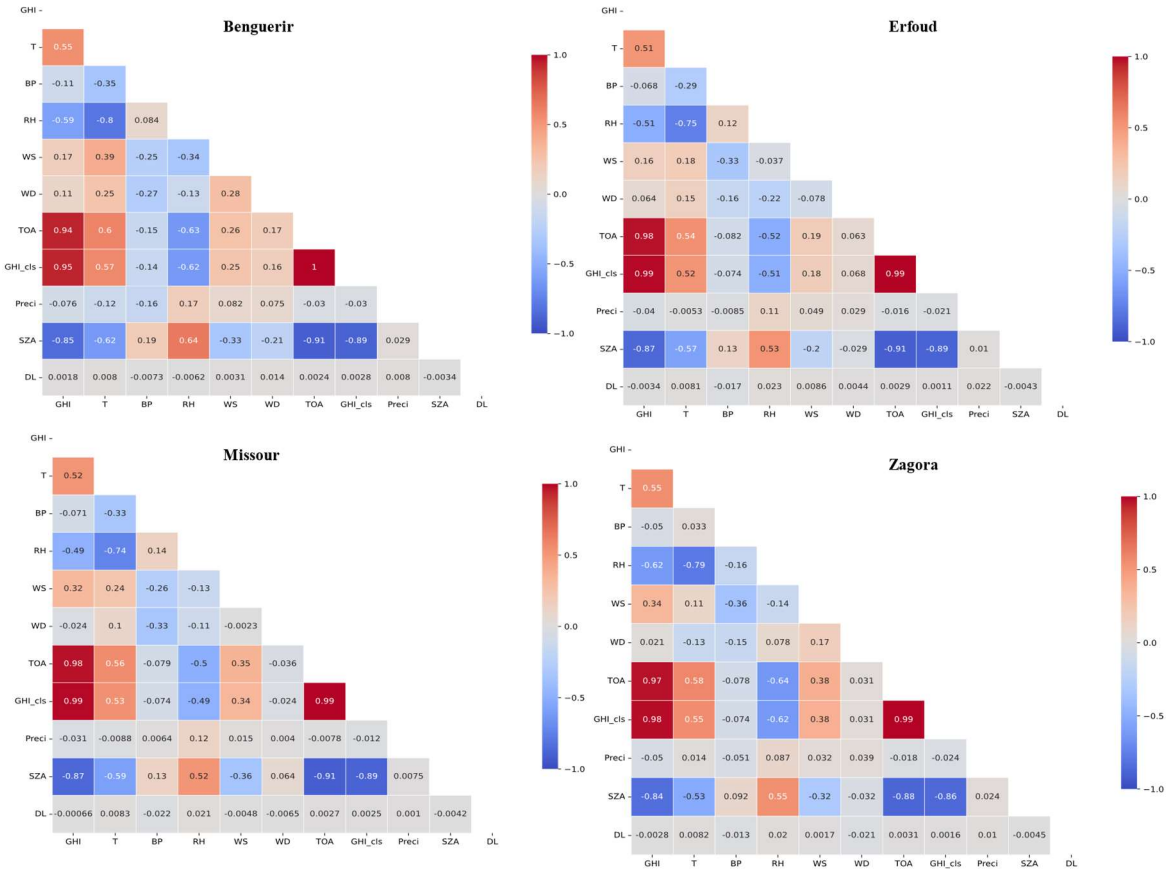


Figure 4-6. Matrice de corrélation entre le GHI et les variables candidates pour Benguerir, Erfoud, Missour, et Zagora.

Tableau 4-3. Les modèles testés pour la prévision du GHI.

Modèle	Caractéristiques d'entrée									
	T	**RH**	**BP**	**WS**	**WD**	**TOA**	**GH_{cls}**	**Preci**	**θ_z**	**DL**
M1	*	*	*	*	*	*			*	*
M2	*					*	*		*	
M3	*					*			*	*
M4	*	*				*			*	*
M5	*	*	*	*	*			*		

M6							*			*	*
M7		*		*	*	*				*	
M8	*	*	*	*	*	*	*	*	*	*	
M9	*	*							*	*	

Le prétraitement et la sélection des entrées pertinentes sont des étapes communes entre les trois modèles de l'intelligence artificielle utilisés, à savoir MLP, LSTM et CNN. La différence entre ces modèles réside généralement dans l'architecture du réseau neuronal. Pour le LSTM et la partie MLP du modèle hybride MLP-CNN, les mêmes entrées sélectionnées pour le meilleur modèle MLP seront utilisées, alors que pour la partie CNN des informations extraites à partir du traitement des images de la caméra hémisphérique seront concaténées avec les entrées du MLP.

2.1.3. Le modèle MLP

Le choix de l'architecture du modèle neuronal est une étape importante pour atteindre la précision souhaitée. Dans notre cas, pour la couche d'entrée, le nombre de variables d'entrée a été choisi tout d'abord en se basant sur les résultats de corrélation. Par la suite, plusieurs combinaisons en entrée sont testées à partir des variables sélectionnées par le test de corrélation. Pour la couche de sortie, le nombre de neurones est imposé par le nombre de variables à prédire. Dans notre cas, la couche de sortie contient un seul neurone vu qu'on souhaite prédire une seule variable qui est le GHI. Comme montré dans la littérature [156,283,284], un réseau de neurones multicouche avec une seule couche cachée, une fonction d'activation de type sigmoïde en entrée, et de type linéaire dans la couche de sortie est capable d'identifier n'importe quelle fonction non linéaire et ses dérivées à condition que le nombre de neurones dans la couche cachée soit suffisant. Alors, la structure du réseau neuronal que nous avons utilisé contient une seule couche cachée. Pour le choix de nombre de neurones dans cette couche cachée, nous avons construit plusieurs modèles avec diverses architectures et le meilleur modèle a été retenu comme modèle final de prévision.

Le Tableau 4-4 représente la meilleure architecture pour chacun des modèles candidats (Tableau 4-3) et l'erreur associée à chaque modèle. Nous avons commencé avec une architecture simple avec 3 neurones en entrées et nous avons incrémenté ce nombre jusqu'à obtenir une erreur minimale entre les observations et les prévisions sur la base de test. Nous avons testé au minimum 10 architectures pour chaque modèle, mais seulement les meilleures architectures trouvées sont présentées.

Tableau 4-4. Meilleures architectures pour chaque modèle de prévision.

Modèles	**Architecture [entrées, neurones dans la couche cachée, sortie]**	CC	**rRMSE (%)**
M1	[8, 10, 1]	0.99	3.70
M2	[4, 4, 1]	0.95	12.04
M3	[4, 5, 1]	0.88	32.16

M4	[5, 12, 1]	0.92	23.97
M5	[6, 8, 1]	0.84	35.78
M6	[3, 4, 1]	0.88	8.78
M7	[5, 6, 1]	0.85	18.59
M8	[10, 14, 1]	0.82	13.21
M9	[4,7,1]	0.92	20.12

D'après plusieurs tests et selon les résultats présentés dans le Tableau 4-4, nous pouvons constater que les modèles qui se base seulement sur les paramètres météorologiques sont moins performants par rapport à ceux qui introduisent le rayonnement au top de l'atmosphère et l'angle zénithal solaire, chose logique vu que l'angle solaire est relatif à la position du soleil dans le ciel, ainsi que le rayonnement au top de l'atmosphère représente la quantité maximale de rayonnement solaire incident. La température ambiante présente aussi une caractéristique importante et améliore la précision des prévisions, cela est dû à la relation linéaire entre cette dernière et le GHI. L'humidité relative a aussi montré sa capacité dans l'amélioration des prévisions surtout lorsqu'elle est associée avec les autres paramètres météorologiques. En effet, l'humidité est largement impactée par la température, et elle est inversement corrélée avec la température et le GHI. Lorsque l'humidité augmente, la température et le GHI diminuent et vice versa. Par exemple, pour le M4, nous avons utilisé la température et l'humidité relative en plus d'autres variables en entrée, alors que pour le modèle M3, nous avons gardé les mêmes caractéristiques et enlevé l'humidité relative. Nous pouvons le remarquer que la précision diminue de 0.92 pour M4 et de 0.8 pour M3. D'après les résultats présentés dans le Tableau 4-4, le modèle M6 est celui qui génère les meilleures performances. Nous pouvons aussi remarquer que l'utilisation de plusieurs variables en entrées n'améliore pas nécessairement la précision des prévisions comme le cas du modèle M8. En revanche, des modèles avec peu de caractéristiques en entrée fournissent également de meilleures performances tels que les modèles M2 et M4 surtout lorsque le rayonnement au top de l'atmosphère et l'angle solaire zénithale sont inclus. D'après les résultats de test de divers modèles, nous pouvons remarquer que le modèle M1 est le plus performant en termes de variables d'entrée, alors ce modèle sera considéré pour la prévision du GHI. En plus du modèle M1, nous allons aussi évaluer le modèle M6 qui ne nécessite aucune variable météorologique en entrée et qui peut être facilement reproduit par autres utilisateurs même s'ils n'ont pas accès aux données météorologiques.

Après le choix des entrées pertinentes, il est nécessaire de choisir l'architecture optimale pour chaque modèle de prévision. En considérant les modèles M1 et M6, la procédure utilisée pour déterminer le nombre de neurones optimal dans la couche cachée est expliquée ci-dessous.

Pour chaque modèle, nous avons fait varier le nombre de neurones dans la couche cachée et calculé le RMSE pour chaque test. La Figure 4-7 et la Figure 4-8 illustrent la variation du RMSE en fonction du nombre de neurones dans la couche cachée sur la base des données de test pour les modèles M1 et M6, respectivement.

Pour le modèle M1, nous constatons que la valeur de RMSE diminue jusqu'à 8 neurones puis sa valeur reste presque constante et qu'une erreur minimale de 18.58 W/m^2 est obtenue pour un nombre de neurones égal à 10. La même allure se reproduit pour le modèle M6, l'erreur diminue jusqu'à 4 neurones où une erreur minimale de 45.07 W/m^2 est obtenue. Ce sont donc ces structures qui seront retenues pour la prévision du GHI. La même procédure a été utilisée pour déterminer la meilleure architecture pour les modèles cités dans le Tableau 4-4.

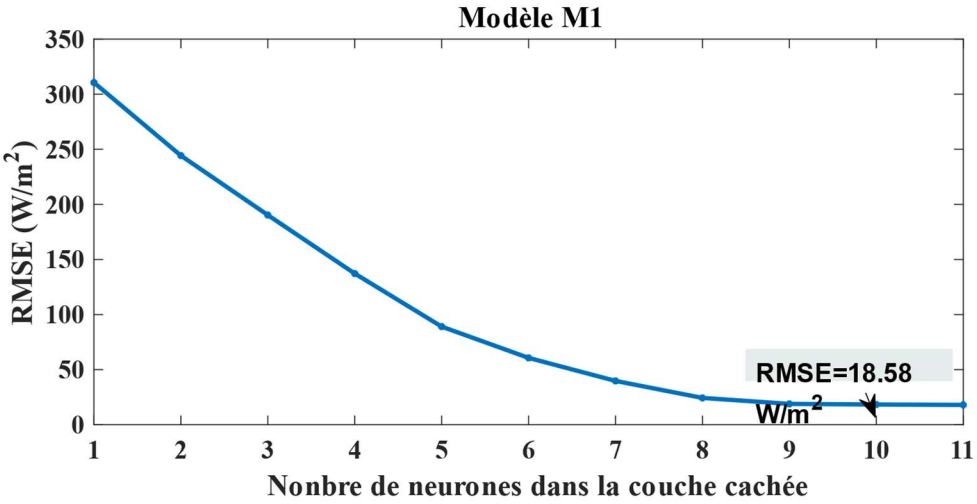


Figure 4-7. RMSE (W/m^2) obtenu sur la base de test en fonction de nombre de neurones dans la couche cachée pour le modèle M1.

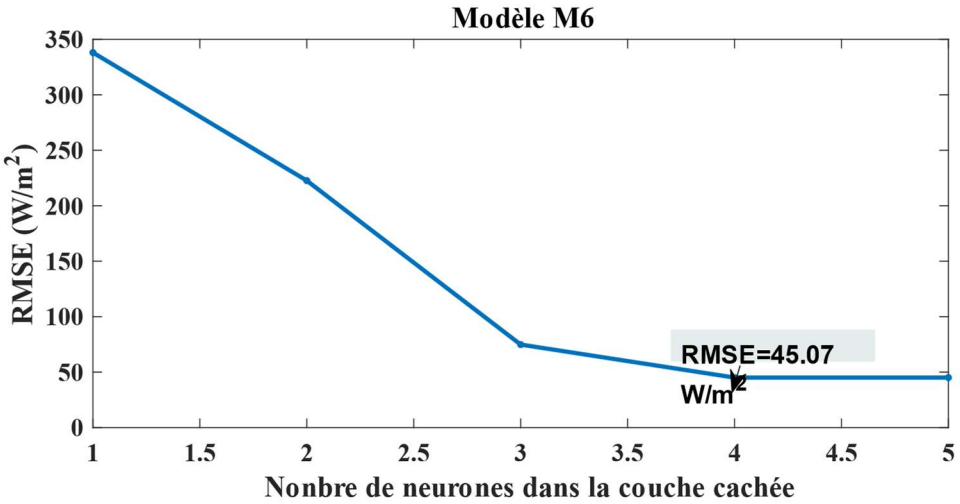


Figure 4-8. RMSE (W/m^2) obtenu sur la base de test en fonction de nombre de neurones dans la couche cachée pour le modèle M6.

L'objectif de la phase d'apprentissage d'un réseau de neurones est de trouver une fonction qui permet d'approximer la fonction de régression entre les entrées et la sortie. Lors de l'utilisation des RNA, on ne connait que les valeurs observées avec un bruit ajouté, mais on ne connait pas la fonction de régression. Pour trouver une approximation, il faut définir une fonction coût qui mesure l'écart entre la sortie prédite et désirée. Cette fonction coût est une fonction scalaire composée qui dépend du vecteur de paramètres du modèle qui est constitué par les poids du

modèle. Plus que la fonction coût est petite plus que le modèle est précis. Pour cette raison, les algorithmes d'apprentissage cherchent à minimiser la fonction coût à l'aide des algorithmes d'optimisation.

Concernant l'algorithme d'apprentissage, nous avons utilisé l'algorithme de Levenberg Marquardt (LM) [285,286]. Cette méthode présente des avantages par rapport à d'autres algorithmes, puisqu'elle peut être utilisée sans avoir besoin à choisir le pas et le nombre d'itérations. Les poids du modèle sont initialisés aléatoirement. Pour la fonction de transfert, il est souvent reporté dans la littérature qu'une fonction de transfert de type sigmoïde pour la couche cachée et de type linéaire pour la couche de sortie permet d'obtenir de meilleurs résultats par rapport aux autres fonctions de transfert. Pour nos modèles, nous avons testé diverses fonctions de transfert, où les architectures comparées étaient les mêmes sauf pour les fonctions de transfert. Après les simulations, nous avons trouvé que la fonction sigmoïde pour la couche cachée, et aussi la fonction sigmoïde pour la couche de sortie au lieu de la fonction linéaire présente les meilleurs résultats.

Les résultats de comparaison des mêmes architectures avec une différence dans la fonction de transfert dans la couche de sortie sont illustrés par la Figure 4-10 et la Figure 4-9. Comme nous pouvons le constater, lors de l'utilisation d'une fonction de transfert de type linéaire dans la couche de sortie, nous avons obtenu des nuages de points largement dispersés et mal corrélés surtout pour les faibles valeurs du GHI avec une corrélation de 0.95. Cependant, l'utilisation de la fonction de transfert sigmoïde dans la couche de sortie au lieu de la fonction linéaire a amélioré notablement les résultats, les points sont bien corrélés avec un coefficient de corrélation de 0.99.

Pour éviter le phénomène de surapprentissage, nous avons utilisé l'arrêt prématuré qui permet d'arrêter l'apprentissage lorsqu'une erreur minimale de validation est atteinte et quand l'erreur d'apprentissage devient inférieure à celle de test.

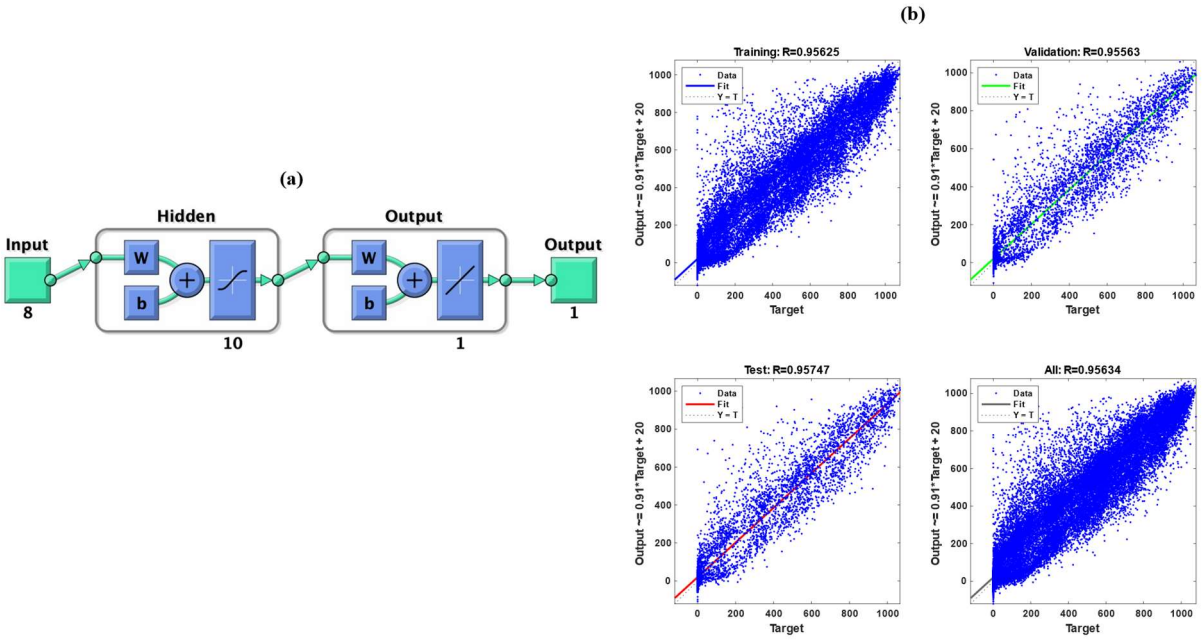


Figure 4-9. (a) architecture du réseau de neurones avec la fonction linéaire dans la couche de sortie, et (b) courbe de régression pour l'apprentissage, la validation, et le test pour le modèle M1.

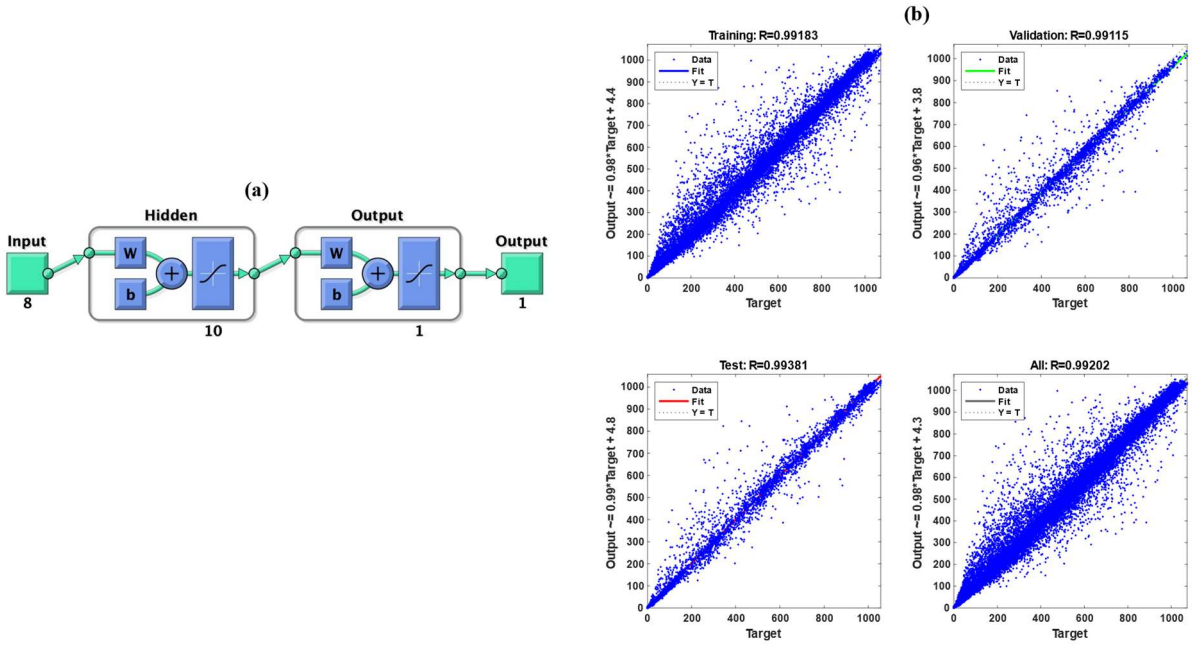


Figure 4-10. (a) architecture du réseau de neurones avec la fonction sigmoïde dans la couche de sortie, et (b) courbe de régression pour l'apprentissage, la validation, et le test pour le modèle M1.

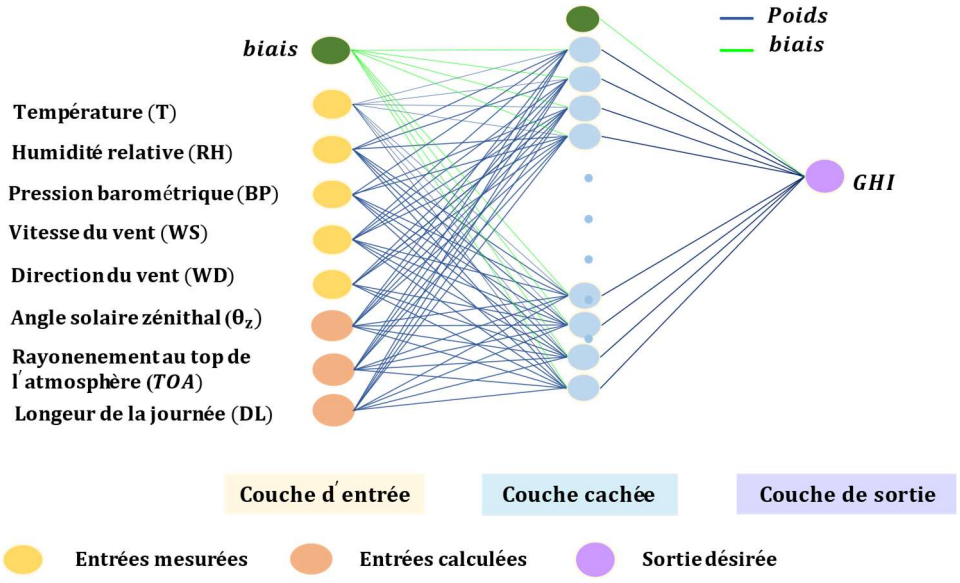


Figure 4-11. Structure générale du modèle M1 développé pour la prévision du GHI, basé sur des variables météorologiques et des calculs de la géométrie solaire.

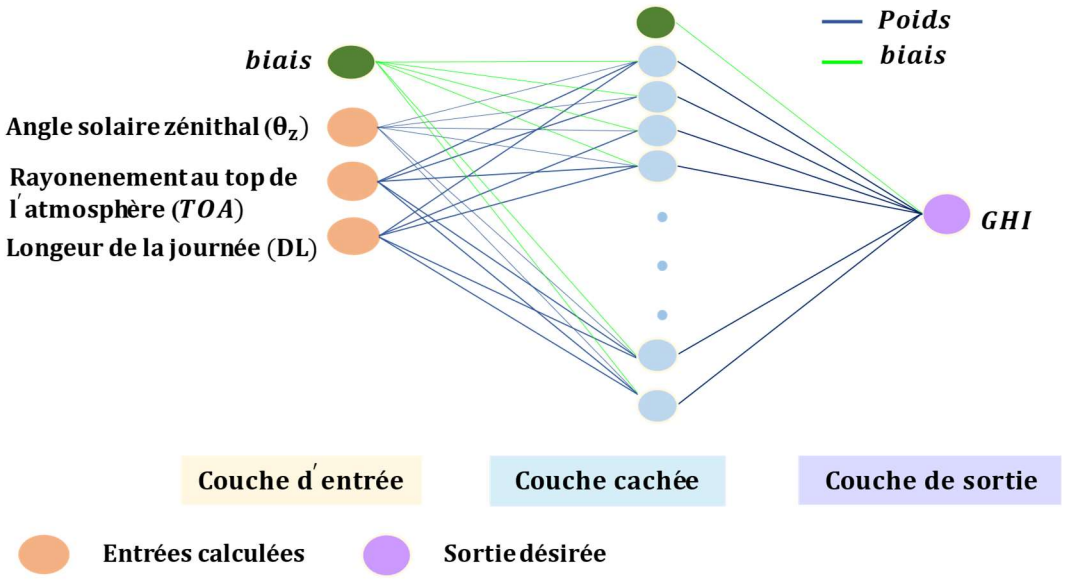


Figure 4-12. Structure générale du modèle M6 développé pour la prévision du GHI, en se basant seulement sur des calculs de la géométrie solaire.

2.1.4. Le modèle LSTM

Le modèle LSTM adopté pour cette étude se compose d'une couche d'entrée, de couches LSTM cachées, d'une couche entièrement connectée et d'une couche de sortie. La couche d'entrée alimente le réseau en données de séries temporelles. Les couches cachées apprennent les dépendances à long terme entre les étapes de données de séries temporelles. Toutefois, les prédictions sont transmises par la couche de sortie. Nous avons utilisé au niveau du modèle LSTM les mêmes variables d'entrée choisies pour le MLP M1 (Figure 4-11). En ce qui concerne l'algorithme d'optimisation, nous avons utilisé l'algorithme Adam, qui est l'algorithme le plus populaire pour la prévision avec les réseaux de neurones récurrents [287].

Pour démarrer le processus d'apprentissage, les biais et les poids des connexions du réseau neuronal doivent être initialisés. Pour assurer la stabilité et la rapidité de l'apprentissage du réseau, les poids d'entrée sont initialisés avec la méthode Glorot [288] qui permet d'initialiser les poids à partir d'une distribution uniforme avec une moyenne égale à zéro et une variance de $\frac{2}{N_i+4N_h}$, avec N_i et N_h étant respectivement le nombre d'entrées et le nombre de couches cachées. Les poids récurrents avec la méthode orthogonale [289] permet d'initialiser les poids récurrents à partir d'une matrice orthogonale issue d'une distribution normale unitaire. Le biais de la porte d'oubli avec 1 et les autres biais avec des zéros. Le Tableau 4-5 résume les paramètres du réseau récurrent LSTM adopté pour la prévision du GHI. Nous résumons l'architecture générale du modèle LSTM au niveau de La Figure 4-13.

Tableau 4-5. Paramètres adoptés pour le modèle LSTM.

Paramètre	Valeur
Algorithme d'optimisation	Adam
Itérations	200
Taux d'apprentissage	0.0003
Type de couches	Régression
Unités cachées	250
Entrées	8
Sortie	1

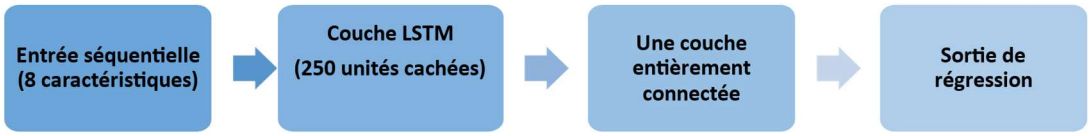


Figure 4-13. Architecture du réseau récurrent LSTM adopté pour la prévision du GHI.

2.2. Prévision avec le modèle statistique SARIMA

Les modèles autorégressifs ARIMA sont l'un des modèles classiques les plus populaires pour la prévision des séries temporelles. Comme détaillé dans le chapitre 3, ARIMA est basé sur trois composantes linéaires paramétriques : AR, qui est le terme autorégressif, I est le terme d'intégration et MA est le terme de moyenne mobile. Une fois les paramètres du modèle sont identifiés, il peut prédire la valeur future de la variable cible à partir des valeurs passées de la série temporelle avec une erreur aléatoire. Pour les séries temporelles avec une composante saisonnière, le modèle est appelé $SARIMA(p,d,q)(P,D,Q)_s$, où (p,d,q) présente la partie non saisonnière du modèle (P,D,Q) est la partie saisonnière du modèle, et s présente la saisonnalité du modèle.

La stationnarité des séries temporelles est essentielle dans l'analyse et la prévision des séries temporelles avec les méthodes statistiques classiques telle que SARIMA. Avant de procéder à la stationnarisation de la série temporelle, il est intéressant de déterminer sa saisonnalité. Pour déterminer la saisonnalité, nous nous sommes basés sur l'analyse de la fonction d'autocorrélation ACF. Une variable saisonnière aurait un pic ACF à un certain délai. Les données sont ensuite différenciées par le décalage S qui représente la valeur de la période saisonnière jusqu'à ce qu'aucune saisonnalité ne soit détectée.

Les résultats de l'ACF de la série temporelle du GHI horaire pour chaque station sont illustrés à la Figure 4-14. Pendant la période d'observation, la série temporelle conserve une saisonnalité quotidienne. Nous remarquons la présence des pics toutes les 12, 24, 38 ... alors nous pourrions conclure que la saisonnalité S est de l'ordre 12.

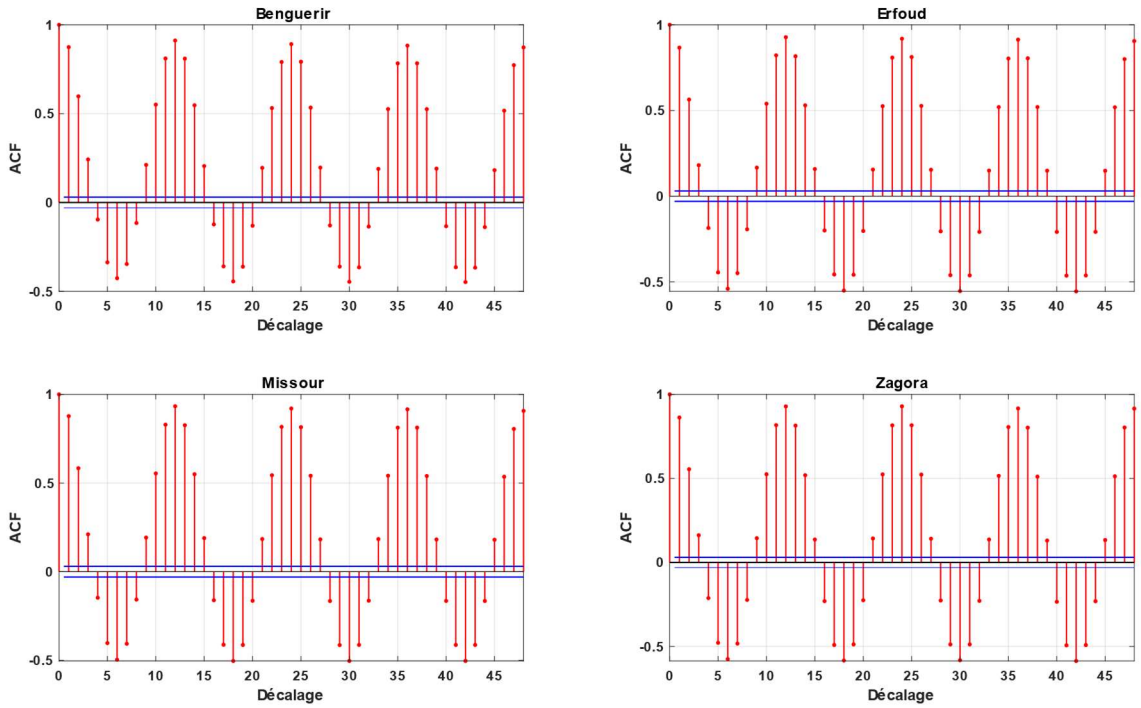


Figure 4-14. La fonction d'autocorrélation ACF du GHI horaire pour les quatre sites d'étude.

Après avoir choisi l'ordre de saisonnalité approprié, il convient maintenant de vérifier la stationnarité de la série temporelle. Une série temporelle est dite stationnaire si son espérance et sa variance ne dépendent pas du temps. Pour vérifier la stationnarité de notre série temporelle, nous avons utilisé le test de racine unitaire de Dickey-Fuller pour lequel la présence d'une racine unitaire implique que la série temporelle est non stationnaire [290]. Dans le cas de la présence d'une racine unitaire, nous appliquons une différenciation à la série temporelle du GHI avec un décalage de 1, et nous vérifions à chaque fois l'existence de la racine unitaire et répétons la boucle jusqu'à ce qu'elle devienne stationnaire. Le nombre de différences saisonnières détermine la valeur de D, tandis que le nombre de différences pour obtenir la stationnarité détermine la valeur de d. Dans notre cas, nous avons appliqué une différenciation d'ordre 1 avec une saisonnalité $S = 12$ pour éliminer la saisonnalité, et une différenciation d'ordre 0 pour avoir une série temporelle stationnaire, alors $d = 0$ et $D = 1$.

Après la stationnarisation et la différenciation de la série temporelle, il faut déterminer les poids optimaux du modèle SARIMA. Pour choisir l'ordre adéquat du modèle SARIMA, nous nous sommes basés sur le critère d'information d'Akaike (AIC) et le critère d'information bayésien (BIC). L'AIC et le BIC appliquent une fonction de vraisemblance pour estimer les paramètres optimaux du modèle. Le modèle avec les valeurs minimales d'AIC et du BIC est le modèle le mieux ajusté [136].

Pour estimer les poids des modèles, nous avons utilisé l'environnement Matlab, qui utilise la méthode du maximum de vraisemblance basée sur l'hypothèse du résidu gaussien pour estimer ces paramètres [291]. Cette technique permet de trouver les paramètres qui maximisent la probabilité d'obtenir les données que nous avons observées.

L'AIC et le BIC sont définis par l'équation suivante :

$$AIC = \ln(\hat{\sigma}^2) + r\frac{2}{N} \tag{4.3}$$

$$BIC = \ln(\hat{\sigma}^2) + r\frac{\ln(N)}{N} \tag{4.4}$$

où $\hat{\sigma}^2$ représente le maximum de vraisemblance pour le modèle, $r = p + q + 1$ est le nombre de paramètres libres dans le modèle incluant un terme constant, et N est le nombre des observations de la série temporelle.

En considérant différentes valeurs de p et q, le Tableau 4-6 représente les valeurs d'AIC et de BIC obtenues pour les modèles SARIMA testés pour la prévision à court terme du GHI pour les quatre sites d'étude. Avec une différence saisonnière fixée à $D = 1$, une saisonnalité $S = 12$ et une différence de stationnarité $d = 0$. Les meilleurs AIC et BIC pour chaque station sont présentés en gras.

Comme nous pouvons le voir, les valeurs les plus faibles de l'AIC et du BIC sont obtenues pour le modèle (1, 0, 2) $(1, 1, 1)_{12}$ pour la station de Benguerir (2, 0, 3) $(1, 1, 1)_{12}$ pour Missour. Les stations d'Erfoud et de Zagora partagent le même modèle (3, 0, 3) $(1, 1, 1)_{12}$. Ces modèles seront évalués pour la prévision du GHI horaire un jour à l'avance.

Tableau 4-6. Valeurs d'AIC et de BIC pour les modèles SARIMA testés.

	Benguerir		**Erfoud**		**Missour**		**Zagora**	
Modèles	Critère d'information							
SARIMA	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
$(1,0,1)(1,1,1)_{12}$	68854	68887	68600	68633	63232	63265	56817	56848
$(1,0,2)(1,1,1)_{12}$	**68527**	**68588**	68584	68624	63021	63061	56566	56604
$(2,0,3)(1,1,1)_{12}$	69192	69239	68560	68607	**62861**	**62915**	56299	56350
$(3,0,3)(1,1,1)_{12}$	68787	68828	**68255**	**68315**	62913	62973	**56105**	**56161**
$(4,0,3)(1,1,1)_{12}$	68731	68785	68278	68327	63038	63048	56714	56764

2.3. Prévision à très court terme

2.3.1. Site d'étude et prétraitement de données

Pour des horizons de prévision à très court terme, l'imagerie hémisphérique a montré son efficacité et peut apporter une contribution essentielle à la prévision de l'irradiation solaire [126], grâce à l'arrivée de caméras de sécurité numériques peu coûteuses et de capacités de traitement graphique plus robustes.

Pour la prévision à très court terme (15 minutes à l'avance), nous avons exploité les images d'une caméra hémisphérique installée au sol au Green Energy Park (GEP, Benguerir). Pour faire, nous avons développé un modèle hybride CNN-MLP. Le CNN adopté dans cette étude possède la capacité d'extraire les caractéristiques intrinsèques contenues dans les images. Par

la suite, les caractéristiques extraites des images sont utilisées à côté d'autres variables météorologiques comme entrée du modèle MLP pour prédire le GHI 15 minutes à l'avance. La Figure 4-15 représente une vue d'ensemble de la méthode considérée. Le modèle de prévision est conçu en quatre étapes décrites comme suit :

- **Étape 1** : consiste l'acquisition des images du ciel et des paramètres météorologiques supplémentaires. Les images utilisées dans cette étude ont été acquises par une caméra hémisphérique de type MOBOTIX Q25 disposant d'un champ de vision de 182°. Elle occupe une zone circulaire d'un diamètre de 1616 pixels, et chaque image est capturée toutes les minutes avec une taille de 2048x1536 pixels. Les paramètres météorologiques sont collectés à partir de la station météorologique installée au sol au site de Benguerir, GEP, les caractéristiques des capteurs de la station sont décrites en détail dans les chapitres 1 et 3. La Figure 4-16 représente la caméra utilisée pour l'acquisition des données et des exemples des images acquises.
- **Étape 2** : cette étape consiste le prétraitement des images acquises. Même si la caméra est souvent nettoyée, il faut s'assurer que les images sont de meilleure qualité. Les différents algorithmes de prétraitement nécessaires sont déjà intégrés dans le dispositif d'acquisition d'images. De notre part, dans cette étape, nous éliminons les images nocturnes et nous inspectons visuellement la qualité des images. Parfois, nous remarquons l'existence des images qui doivent être éliminées de la base de données, telles que les images floues, et les images avec des objets indésirables comme les insectes, les fientes d'oiseaux, etc. La Figure 4-17 montre des exemples d'images qui doivent être exclues lors de l'étape de prétraitement. Sur l'image (a), nous pouvons voir que le dôme de la caméra est encrassé, ce qui pourra influencer le processus de l'extraction des caractéristiques de l'image où l'encrassement pourra être considéré comme des nuages par les CNN. Sur les images (b) et (c), nous observons la présence des oiseaux et des fientes d'oiseaux qui peuvent aussi influencer négativement le modèle de prévision.
- **Étape 3** : cette étape a pour objectif d'extraire les informations contenues dans les images et par la suite les utiliser à côté des variables météorologiques pour prévoir le GHI à l'aide du réseau MLP. Les caractéristiques extraites par le CNN sont : la couverture nuageuse (mouvement, couleur, position, etc.), la position du soleil, type de ciel et le temps de la journée. Les séries temporelles du GHI, la température, l'humidité relative, la vitesse et la direction du vent sont ensuite concaténées avec la matrice aplatie résultant de la sortie du réseau CNN. Après cela, l'ensemble des données est transmis au MLP pour générer les prévisions du GHI à l'horizon de 15 minutes.
- **Étape 4** : dans cette étape, les modèles de prévision sont développés dans l'étape 3 et les prévisions sont générées. Les résultats de la prévision sont ainsi comparés avec des observations au sol du GHI pour des jours clairs, mixtes et nuageux.

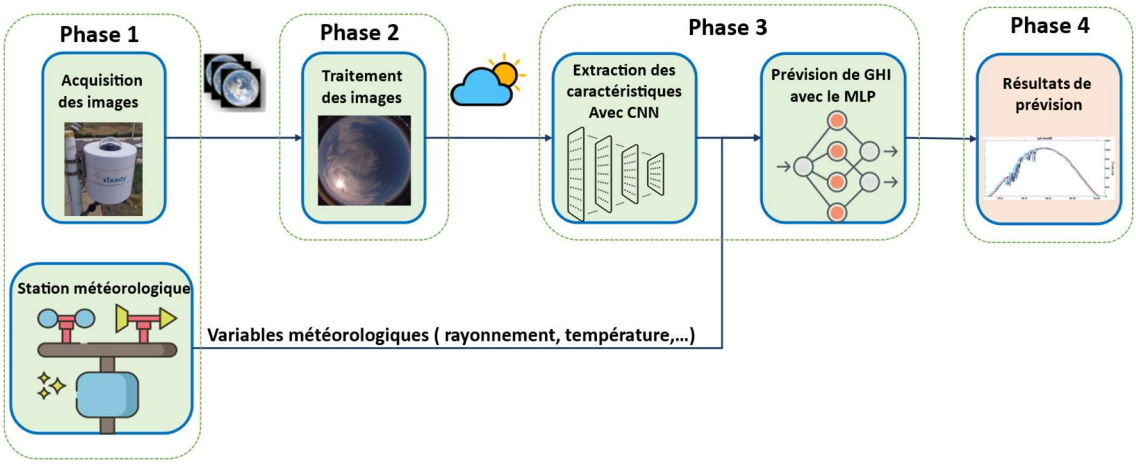


Figure 4-15. Vue d'ensemble de la méthode de prévision à court terme du GHI par le modèle hybride CNN-MLP.



La caméra hémisphérique

Exemple des images acquises par la caméra

Figure 4-16. Caméra hémisphérique utilisée pour l'acquisition des images et des exemples d'images collectées à différents moments de la journée et durant des jours différents.

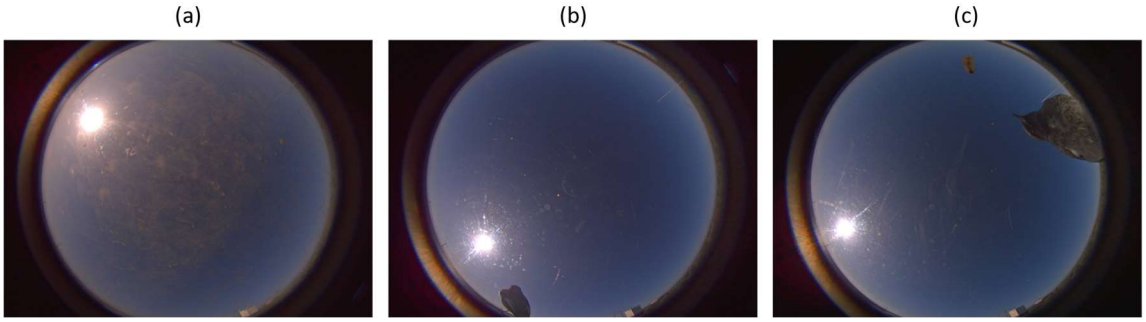


Figure 4-17. Exemples des images enlevées de la base de données lors de l'étape de prétraitement.

2.3.2. Architecture du modèle CNN-MLP

De nombreuses architectures du réseau CNN ont été développées pour la reconnaissance et la classification des images, telles que AlexNet, LeNet-5, VGG-16, Inception, ResNet, etc. [292]. La plupart des architectures CNN utilisent le même concept général consistant à appliquer des opérations de convolution successives à l'image d'entrée pour réduire les dimensions spatiales tout en augmentant le nombre de cartes de caractéristiques extraites.

Dans cette étude, nous avons utilisé l'architecture CNN VGG-16 [293]. Le choix de cette architecture revient à ses performances éprouvées dans plusieurs applications industrielles et scientifiques complexes [294,295]. Le VGG-16 a atteint une précision de 92.7 % sur ImageNet, qui est une base de données contenant plus de 14 millions d'images. Cette base de données a joué un rôle déterminant dans l'avancement de la vision par ordinateur et de la recherche sur l'apprentissage profond [296]. La Figure 4-18 représente l'architecture originale du VGG-16.

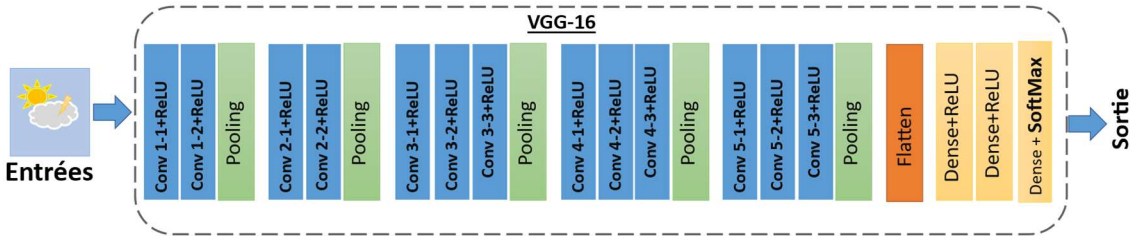


Figure 4-18. L'architecture originale de VGG-16.

L'architecture originale du VGG-16 a été développée pour des tâches de classification. Puisque notre objectif est de mettre en œuvre cette architecture pour la régression, nous avons modifié l'architecture originale pour l'adapter à notre application. Pour se faire, nous avons éliminé les trois dernières couches du réseau (Figure 4-19), notamment la fonction d'activation softmax [297] utilisée pour des raisons de classification et conservé la couche aplatie (Flatten) pour obtenir un vecteur qui peut être inséré dans le réseau MLP. Lors du développement de ce modèle, de nombreuses expériences ont été réalisées pour choisir la meilleure architecture. Nous pouvons finalement conclure que le VGG-16 offre un compromis équitable entre les performances et le coût de calcul pour les besoins en mémoire d'apprentissage et le temps d'exécution. Une normalisation par lot a été appliquée après chaque bloc de convolution dans le but d'accélérer l'apprentissage du réseau. Pour le réseau MLP, nous avons utilisé les mêmes variables météorologiques sélectionnées pour le MLP pour la prévision à court terme, les calculs de la géométrie solaire telle la position du soleil sont supposées extraites par le réseau CNN. Pour le choix de la meilleure architecture du réseau MLP, plusieurs tests ont été conduits et la même procédure détaillée dans la section 2.1.3 est réutilisée. Finalement, nous avons trouvé qu'un réseau MLP avec les sorties de CNN, les cinq variables météorologiques en entrée, une couche cachée avec 10 neurones avec une fonction d'activation de type sigmoïde et une sortie avec la fonction d'activation sigmoïde génère les meilleures performances.

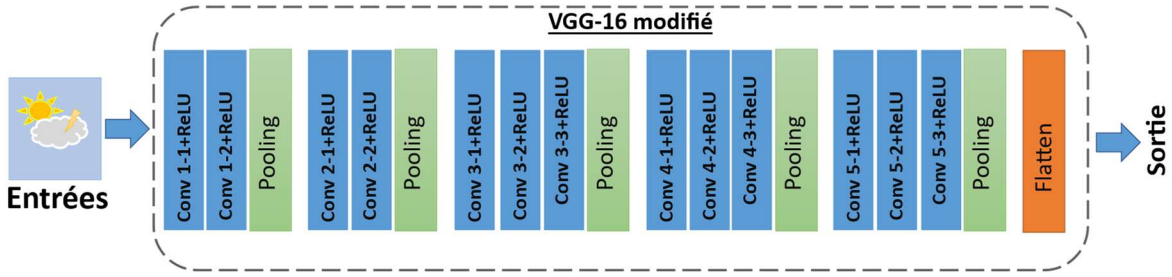


Figure 4-19. L'architecture du VGG-16 modifiée, où les trois dernières couches ont été éliminées.

La taille de l'image d'entrée est (128*128pxl*3 canaux), et la taille de la matrice de sortie du CNN est (4 * 4 * 512), ce qui génère après la couche d'aplatissement un vecteur de 8192 éléments. Ce vecteur est ensuite concaténé avec les séries temporelles des variables météorologiques (T, GHI, RH, WS et WD), d'où un vecteur de 8197 éléments est injecté à l'entrée du réseau MLP. La vue d'ensemble de l'architecture CNN-MLP est illustrée par Figure 4-20. Contrairement aux autres modèles, le modèle CNN-MLP a été développé sous Python.

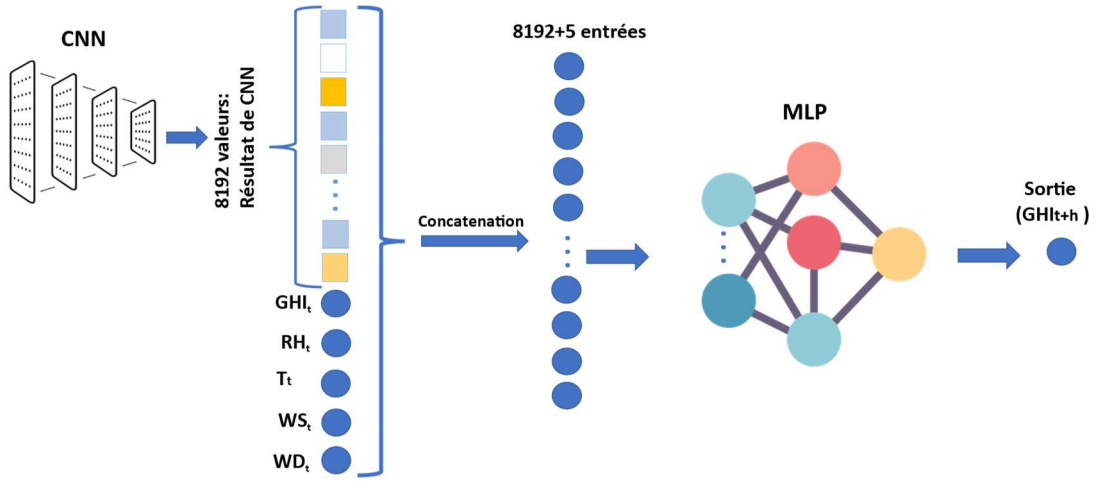


Figure 4-20. Aperçu du modèle de prévision CNN-MLP. t représente l'instant d'observation, h représente l'horizon de prévision.

3. Résultats de prévisions

3.1. Prévision du GHI 12 heures à l'avance

Après l'identification et le choix des architectures des modèles de prévision, nous présenterons dans cette partie les résultats des simulations pour la prévision du GHI horaire une journée à l'avance. Les trois modèles : SARIMA, MLP et LSTM sont évalués pour prévoir le GHI et leurs précisions ont été comparées au modèle de référence (la persistance). Nous présentons dans un premier temps les résultats de prévision du GHI horaire 12 heures à l'avance pour les modèles MLP M1, MLP M6 et LSTM sur la base de test du 01/07/2019 au 30/06/2020 (soit 4380 heures avec les valeurs nocturnes exclues) pour les quatre sites d'études et pour les quatre types de jours. Pour les modèles de persistance et SARIMA, les résultats seront représentés seulement pour les profils journaliers. En effet, les modèles de persistance et SARIMA adoptés consistent à prédire les futures valeurs de GHI en se basant seulement sur l'historique des données. Pour le modèle SARIMA, les données d'apprentissage ont été utilisées pour le choix des paramètres du modèle, cependant, le jeu de test n'est pas assez signifiant comme pour les modèles basés sur les RNA.

La Figure 4-21, la Figure 4-22 et la Figure 4-23 représentent des courbes de nuages de points des prévisions sur la base des données de test pour MLP M1, MLP M6 et LSTM, respectivement. Nous pouvons remarquer que les coefficients de corrélation sont souvent élevés pour les trois modèles (>0.95), ce qui montre que les prédictions faites par les modèles sont exactes. Cependant, les coefficients de corrélation seuls ne peuvent pas être utilisés pour

juger la qualité des modèles. En termes de rRMSE, les résultats révèlent que les modèles MLP M1 et LSTM présentent les meilleures performances. Pour le MLP M6, les prévisions sont moins bonnes et les incertitudes obtenues sont souvent faibles. Le rRMSE du MLP M1 varie entre 10% et 17% pour l'ensemble des sites. Pour le LSTM les valeurs de rRMSE sont plus grandes par rapport au MLP M1 avec des valeurs qui varient entre 11% et 17%. Cependant, ces erreurs sont encore plus importantes pour le modèle MLP M6 avec des valeurs de rRMSE atteignant 26% (102 W/m^2).

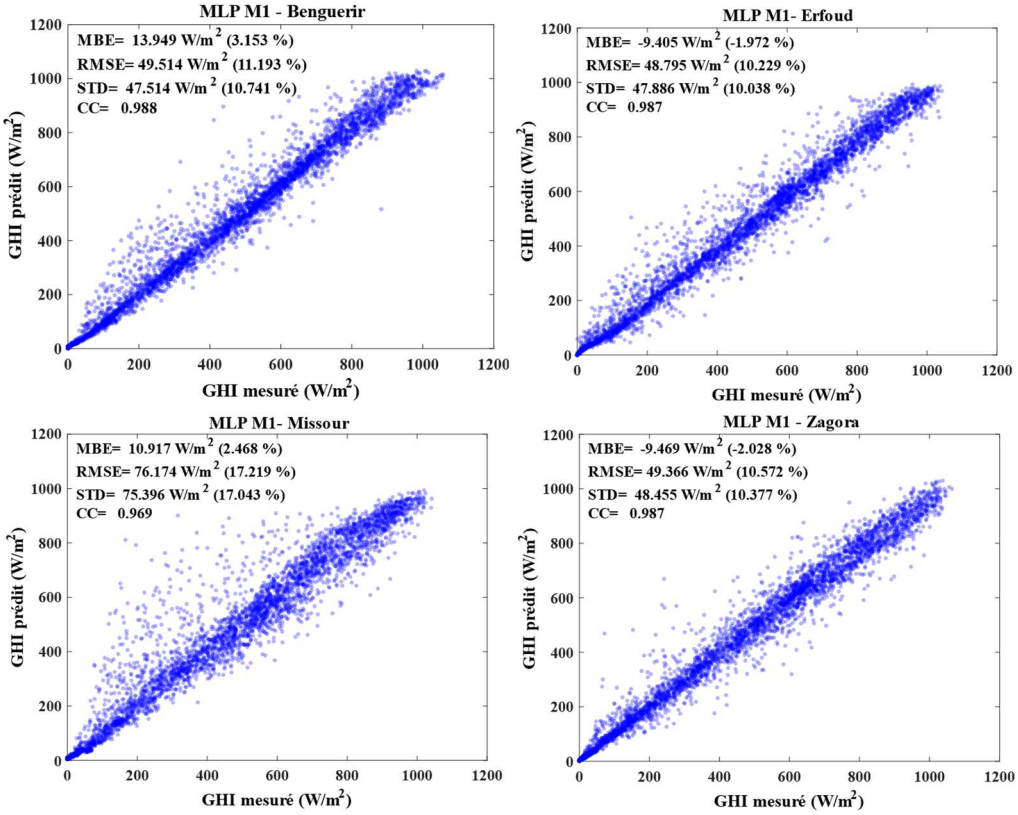


Figure 4-21. Résultats de prévision du GHI horaire 12 heures à l'avance avec le modèle MLP M1 pour Benguerir, Erfoud, Missour, et Zagora sur la base de test.

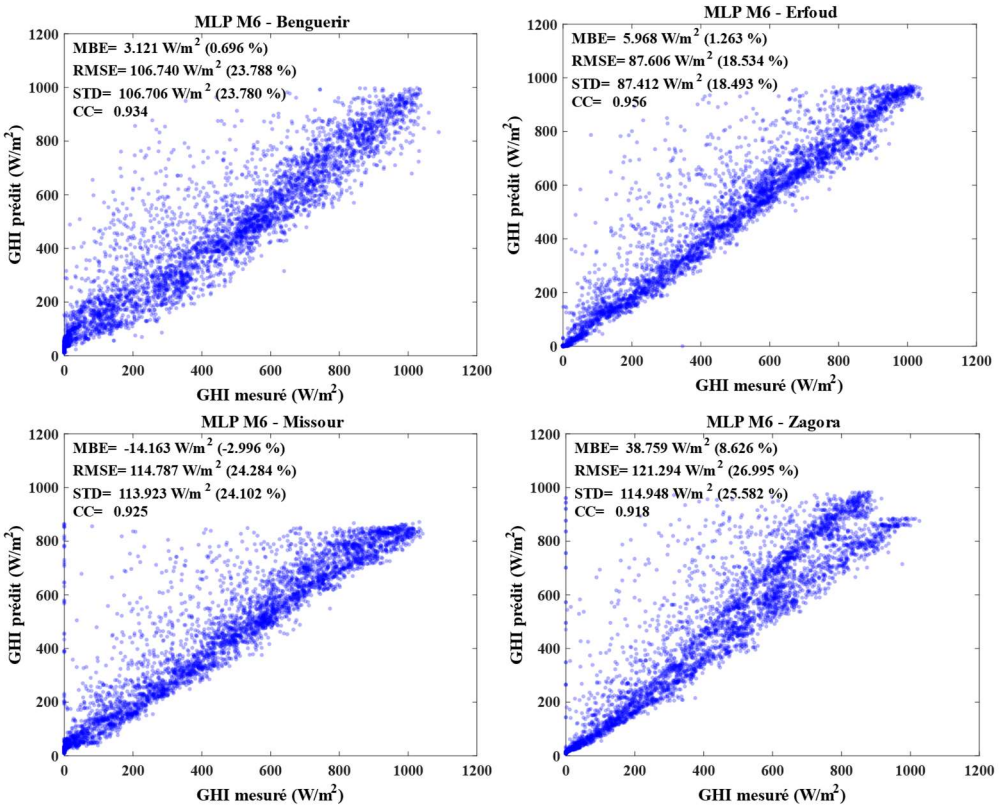


Figure 4-22. Résultats de prévision du GHI horaire 12 heures à l'avance avec le modèle MLP M6 pour Benguerir, Erfoud, Missour, et Zagora sur la base de test.

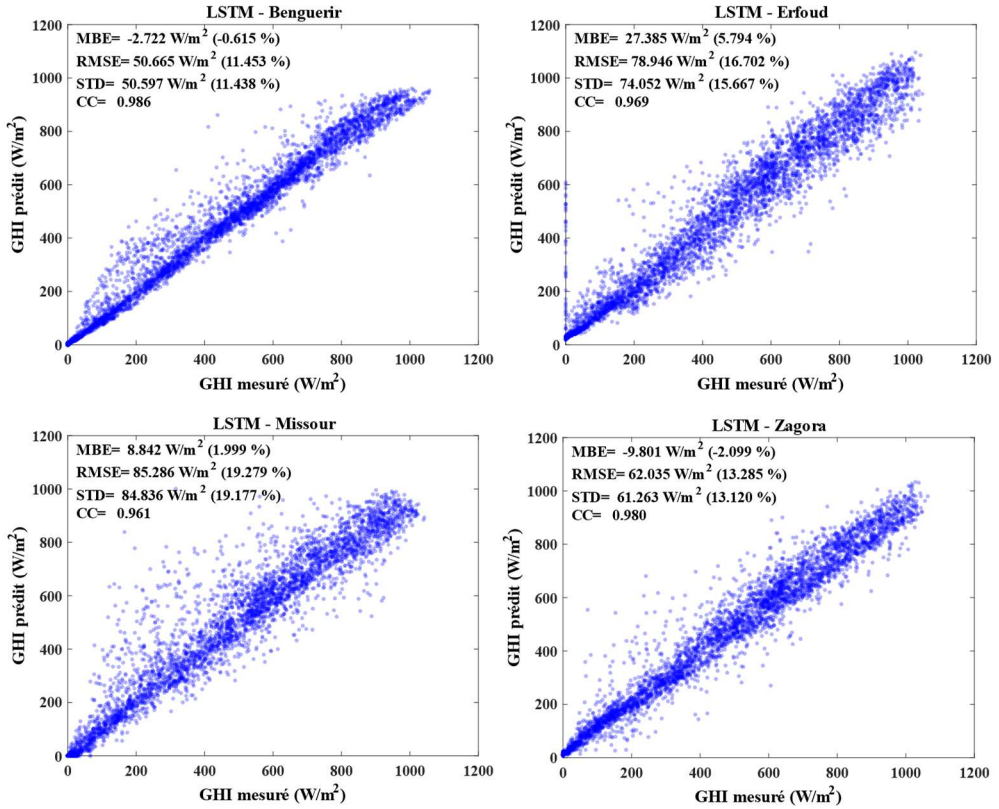


Figure 4-23. Résultats de prévision du GHI horaire 12 heures à l'avance avec le modèle LSTM pour Benguerir, Erfoud, Missour, et Zagora sur la base de test.

3.1.1. Analyse des performances prédictives des modèles de prévision en fonction du type de jour (clarté, et variabilité)

Afin de mettre en évidence l'influence de la variabilité du rayonnement solaire sur la précision des prévisions, nous avons analysé les performances prédictives des modèles pour différentes classes de jours. Les profils journaliers du GHI prédit une journée à l'avance par rapport aux valeurs mesurées pour Benguerir (a), Erfoud (b), Missour (c) et Zagora (d) sont représentés sur la Figure 4-24. Les diagrammes à barres de MBE et de rRMSE sont également représentés sur la Figure 4-24. Les prédictions sont comprises entre 07:00 et 19 :00. Les résultats de prévision sont présentés pour quatre types de jours: AI (clair, faible variabilité), AII (clair, variabilité moyenne), BII (mixte, variabilité moyenne) et CII (couvert, forte variabilité). Le Tableau 4-7 résume les indicateurs de performances obtenus par les divers modèles de prévision du GHI horaire à l'horizon de 12 heures pour les quatre sites d'étude.

Les graphiques et les indicateurs statistiques montrent que le modèle de persistance génère de mauvaises prévisions pour l'ensemble des sites et pour tous les types de jours. Cela pourrait s'expliquer par le fait que le GHI du jour précédent est négligemment corrélé avec celui du jour suivant. Le modèle SARIMA a également révélé de faibles compétences pour la prévision du GHI, étant donné qu'il ne se base que sur les données historiques du GHI pour prédire le jour suivant, il ne peut donc pas être considéré comme un prédicteur puissant. Pour les autres modèles, la précision varie d'un site à l'autre et en fonction du type du jour. Sur les graphes à barre de MBE, nous pouvons observer des alternances des valeurs positives et négatives du MBE journalier, ceci implique une surestimation ou une sous-estimation du GHI par le modèle de prévision. Pour l'ensemble des modèles, le MBE est très souvent négatif pour les jours clairs (AI, et AII), ce qui montre que le modèle sous-estime le GHI pour ce type de jour, alors que pour les jours mixtes et nuageux le GHI est souvent surestimé avec des MBE positifs. Les modèles qui génèrent les grandes valeurs de MBE sont dans la plupart des cas la persistance et SARIMA. Cela s'explique par le fait que ces modèles ne considèrent pas les variations du GHI et de la couverture nuageuse, ce qui conduit à des surestimations conséquentes du GHI arrivant jusqu'à 353.16 W/m^2 pour la persistance et à 388.53 W/m^2 pour SARIMA. Les jours qui présentent de faibles valeurs d'erreur sont le plus souvent les jours clairs AI et AII.

Les jours de la classe AI représentent les plus faibles erreurs de prévision. Ces jours sont caractérisés par un fort rayonnement avec des passages nuageux très rares. Pour ce type de jours, le modèle de persistance est le moins performant avec des rRMSE entre 18.59% et 28.62%. Alors que les meilleurs indicateurs statistiques sont obtenus avec le MLP M1 et LSTM avec des rRMSE entre 2.32% et 7.08 % pour le MLP M1 et entre 2.99% et 9.99% pour LSTM sur l'ensemble des sites. Le modèle SARIMA a également montré une performance acceptable en dépassant le MLP M6 pour tous les sites pour ce type de jours.

La classe AII représente des jours avec un rayonnement solaire important avec des passages nuageux épisodiques brutaux au cours de la journée, ce qui conduit parfois à augmenter les erreurs de prévisions, comme nous pouvons le voir sur les courbes, où un passage nuageux survient vers 13:00 à Erfoud et Zagora et vers 14:00 à Missour. Pour ce type de jours, les

modèles MLP M1, MLP M6 et LSTM ont démontré la même excellente performance prédictive même pour les jours clairs et non variables. Pour l'ensemble des sites, le rRMSE du modèle LSTM varie entre 9.56% et 15.62% et le MBE entre -12.75 W/m^2 et -19.95 W/m^2. Pour MLP M1, le rRMSE varie entre 10.77% et 17.15% et le MBE entre -10.97 W/m^2 et 50.52 W/m^2. Le rRMSE du MLP M6 varie entre12.96% et 20.89% et le MBE entre -28.07 W/m^2 et 60.60 W/m^2.

La classe BII est représentative des jours avec faible un rayonnement solaire et des fluctuations remarquables causées par plusieurs passages nuageux tout au long de la journée. Pour cette classe, nous pouvons remarquer que les erreurs de prévisions sont augmentées par rapport aux autres classes (AI, et AII). Les valeurs d'erreur les plus importantes ont été obtenues pour les modèles de persistance et SARIMA. Pour la persistance, le MBE atteint 238.91 W/m^2 sur le site de Benguerir, et au même site, le modèle SARIMA atteint sa plus grande valeur de MBE qui est de 203.62 W/m^2. Pour ce type de jour, le modèle LSTM génère la meilleure performance de prédiction dans la plupart des cas. Les valeurs de rRMSE sont de 41.12%, 32.51%, 44.43% et 14.43% pour Benguerir, Erfoud, Missour et Zagora, respectivement. Les prévisions de MLP M1 sont assez proches de celles générées par LSTM avec un rRMSE minimal de 20.28% et maximal de 74.62%.

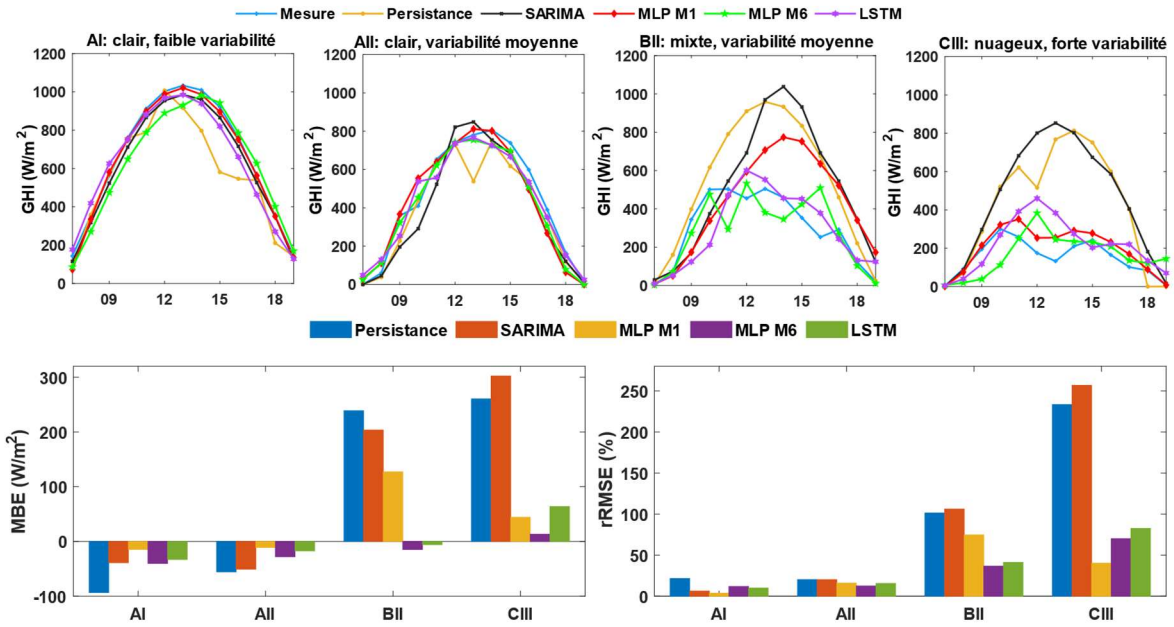
Les jours de la classe CIII sont caractérisés avec des fluctuations importantes du rayonnement solaire dû à un nombre considérable de passages nuageux qui peuvent être de petites masses avec une vitesse de passage rapide. L'erreur de prédiction pour cette classe est encore plus importante que pour les autres types de jours. Sur les graphiques correspondants à cette classe, nous pouvons observer que les modèles n'arrivent pas à suivre l'allure du GHI mesuré et capturer sa variabilité temporelle. Encore une fois, les meilleures prédictions sont obtenues par le MLP M1 et LSTM, alors que les erreurs de persistance et SARIMA sont énormes pour la plupart des sites. Pour les quatre sites, les meilleures prédictions par rapport au RMSE sont générées par le MLP M1 avec un RMSE de 60.80 W/m^2, 141.17 W/m^2, 76.05 W/m^2 et 82.57 W/m^2 pour Benguerir, Erfoud, Missour et Zagora, respectivement. Le LSTM a montré également ses performances de prédiction pour ces jours nuageux et variables. Les valeurs d'erreur les plus faibles de LSTM ont été obtenues pour le site de Zagora avec un MBE de 36.76 W/m^2 et un RMSE de 112.07 W/m^2. L'erreur du LSTM est plus importante pour la station d'Erfoud, avec un MBE de -98.30 W/m^2 et un RMSE de 147.24 W/m^2.

Les graphiques à barres du MBE et du rRMSE révèlent que le rRMSE et le MBE de tous les modèles augmentent en passant des jours de la classe AI vers les jours de la classe CIII. Pour les jours AI, tous les modèles ont montré de bonnes performances, sauf le modèle de persistance, en raison de la corrélation négligeable entre le jour actuel et le jour précédent. Pour ce type de jour, les trois modèles: MLP M1, MLP M6 et LSTM, génèrent de bonnes prédictions. Lorsque les conditions météorologiques sont plus compliquées, par temps nuageux, les résultats montrent que le MLP M1 fournit les meilleures performances prédictives pour presque tous les sites étudiés. La plus petite valeur MBE du MLP M1 a été obtenue pour Missour avec

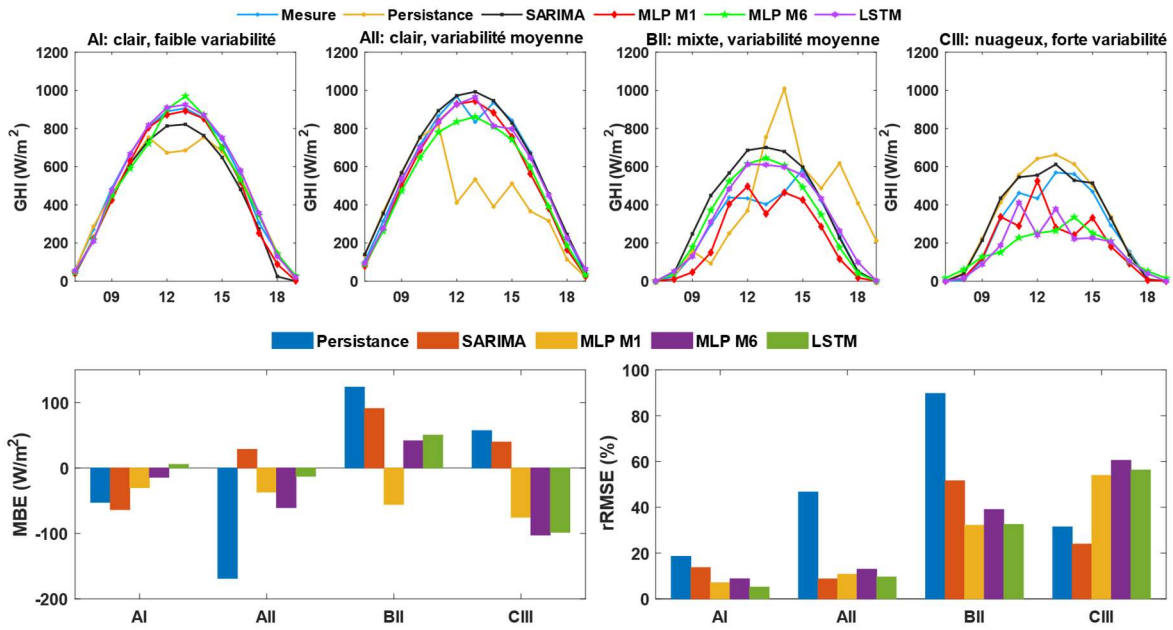
un MBE de 34.31 W/m^2 et un rRMSE de 46.24%. Le LSTM génère également de bonnes prédictions, et un MBE de minimal de 36.76 W/m^2 qui a été enregistré à Zagora.

En plus du MBE et du RMSE, nous avons utilisé le Skill Score permettant une évaluation des modèles de prévision par rapport à un modèle de référence qui est la persistance dans notre cas. La Figure 4-25 illustre les graphes à barres du Skill Score en fonction du type de jour. Nous pouvons observer que pour la majorité des cas, les modèles MLP M1 et LSTM présentent les meilleures performances, avec des SS arrivant jusqu'à 88.27% pour MLP M1 et à 89.57% pour LSTM. Les faibles valeurs de SS sont obtenues par SARIMA et MLP M6.

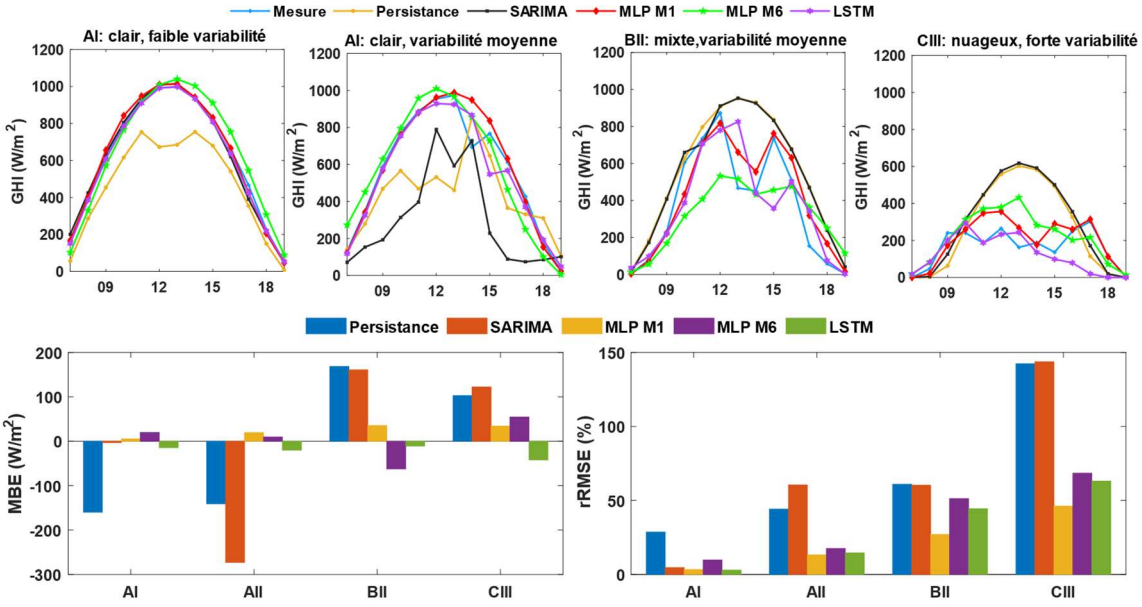
Pour les jours AI, MLP M1 et LSTM ont les meilleures performances, où le SS atteint 88.27% pour MLP M1 et 89.57% pour LSTM. Pour les jours AII, MLP M1 et LSTM sont également les plus efficaces, avec un SS de 79.54% pour LSTM et de 76.94% pour MLP M1, sauf pour Benguerir, où le MLP M6 est le plus performant avec un SS de 38.22%. Toutefois, pour les jours BII, pour Benguerir et Zagora, le MLP M6 est celui qui a les valeurs les plus élevées du SS avec des pourcentages de 63.76% pour Benguerir et de 71.82% pour Zagora. En revanche, pour Erfoud et Missour, c'est le MLP M1qui est le plus efficace, avec des SS de 66.92% et 55.72%, pour les deux sites, respectivement. Tandis que pour les jours de la classe CIII, le MLP M1 et LSTM sont ceux qui présentent de meilleurs pourcentages de SS, avec respectivement, 82.79%, 69.96% pour Benguerir, de -71.59% et -78.96% pour Missour, de 76.53% et 55.70% pour Erfoud et de 69.50% et 58.60% pour le site Zagora pour les deux modèles, respectivement.



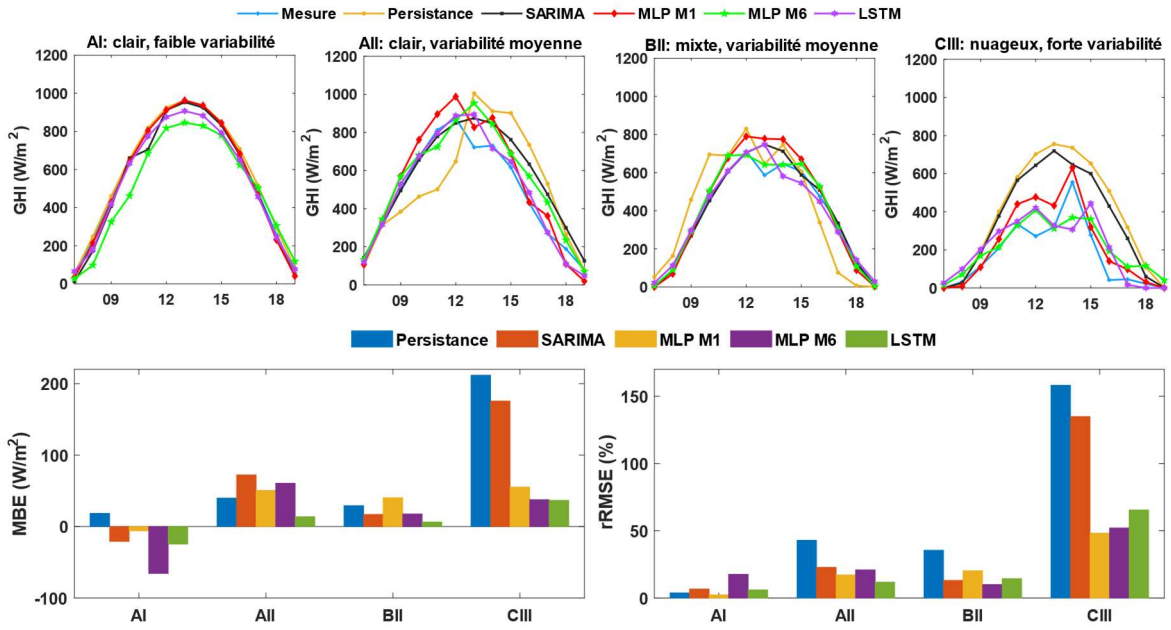
(a) Benguerir



(b) Erfoud



(c) Missour



(d) Zagora

Figure 4-24. Profils journaliers du GHI mesuré et prédit et des métriques MBE et rRMSE obtenues par les différents modèles sous les quatre types de jour.

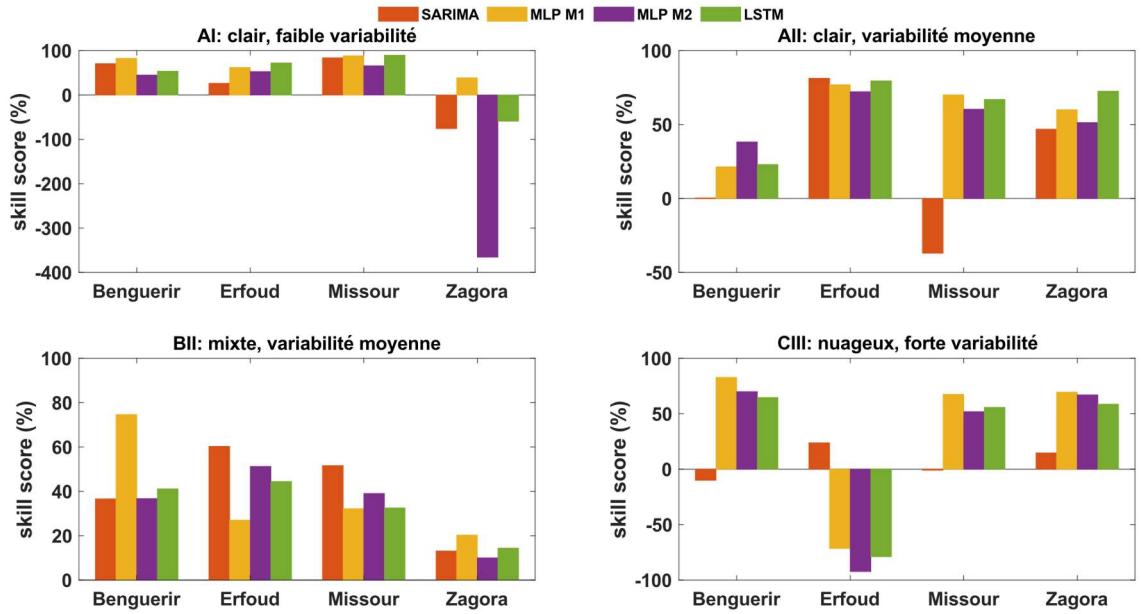


Figure 4-25. Skill Score des modèles de prévisions pour les quatre sites d'étude sous les différentes conditions météorologiques.

Pour résumer, nous représentons sur la Figure 4-26 sous forme de boîte à moustache la variation du rRMSE pour chaque modèle de prévision en fonction du type de jours pour l'ensemble des sites. Comme le montre la Figure 4-26, les valeurs d'erreur les plus faibles sont obtenues pour les jours de type AI et l'erreur augmente d'une manière remarquable pour les jours CIII. Comme pour les jours AI, tous les modèles sont performants, avec la supériorité du

modèle MLP M1 avec un rRMSE entre 2.83% et 5.39%. Toutefois, pour les jours AII, les modèles MLP M1, MLP M6 et LSTM sont plus performants avec l'avantage du modèle LSTM avec un rRMSE entre 10.67% et 15.09%. Tandis que pour BII, la meilleure précision est obtenue avec LSTM avec un rRMSE entre 23.47% et 42.77%. Enfin, pour les jours CIII, l'erreur de prévision est plus importante, et les meilleurs résultats sont obtenus par le MLP M1 avec un rRMSE variant entre 43.21% et 51.1 %.

Selon les résultats obtenus, nous pouvons conclure que les performances des modèles décroissent quand les fluctuations du rayonnement solaire causées par les nuages augmentent. Cela peut être expliqué par le fait que les modèles n'ont pas la capacité à bien capturer les variations rapides et brutales du GHI même s'ils ont montré de bonnes capacités prédictives pour les autres jours clairs. Néanmoins, les divers indicateurs statistiques montrent que les modèles MLP M1 et LSTM ont les meilleures capacités prédictives pour l'ensemble des classes. Alors que la persistance, SARIMA et le MLP M6 ne peuvent pas être considérés comme prédicteurs performants surtout sous les conditions de ciel complexes.

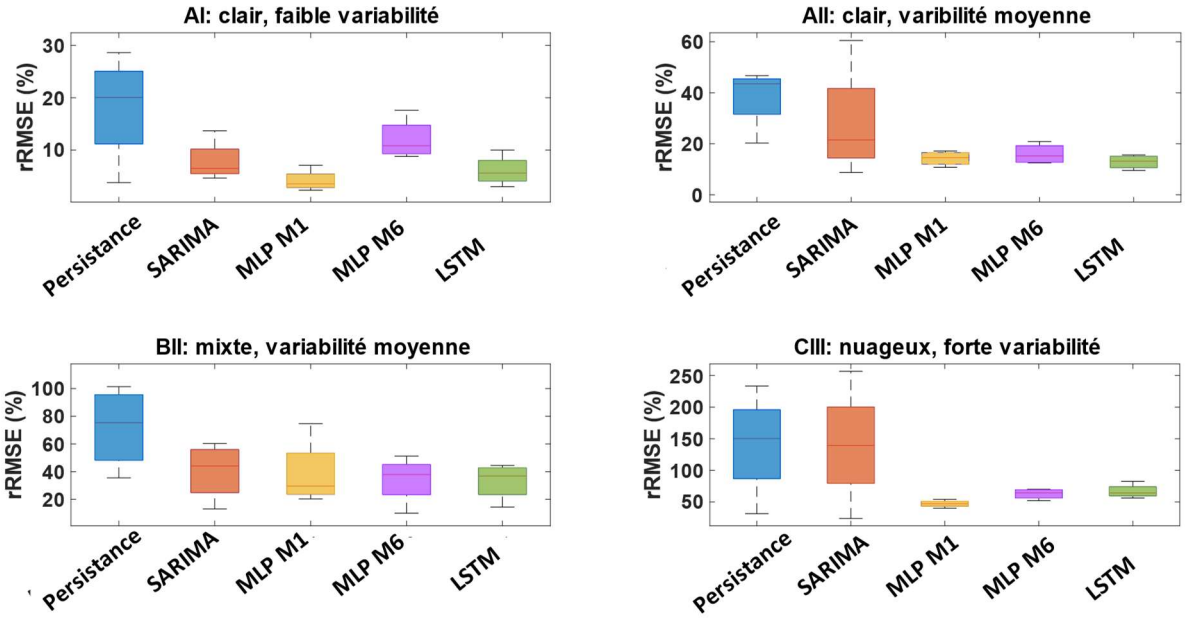


Figure 4-26. Distributions par diagramme en boîte à moustaches des rRMSE obtenus pour les différents modèles de prévision pour chaque type de jour sur l'ensemble du site d'étude.

3.1.2. Analyse de la performance prédictive en fonction des saisons

Le potentiel solaire varie tout au long de l'année, alors il est aussi intéressant de vérifier l'évolution des erreurs de prévision pour chaque saison. Pour cela, nous avons analysé les performances prédictives des modèles de prévision développés pour chaque saison. La Figure 4-27 représente la répartition saisonnière des erreurs pour le MLP M1, MLP M6, LSTM et la persistance. Le Tableau 4-8 résume les indicateurs statistiques saisonniers obtenus pour chaque site. Nous pouvons constater que les meilleures prévisions sont cette fois-ci aussi obtenues par le MLP M1 et LSTM pour les quatre sites et pour toutes les saisons. Les performances de ces deux modèles pour chaque saison sont représentées sous forme de nuages de point sur la base

des données de test par la Figure 4-28. Les indicateurs de performances sont également représentés sur la Figure 4-2 en bleu pour le MLP M1 et en orange pour LSTM. D'après l'analyse des résultats représentés dans la Figure 4-27, la Figure 4-28 et le Tableau 4-8, nous pouvons constater que la persistance et le MLP M6 ne peuvent pas être considérés comme prédicteurs performants et ils ne seront pas considérés dans l'analyse des performances saisonnières. Contrairement à ce que nous prévoyons et à ce qui a été obtenu dans plusieurs études précédentes, où la performance des modèles de prévision augmente en été et en printemps par rapport à l'hiver et l'automne. Dans notre cas, nous pouvons observer que l'automne et l'hiver présentent des résultats très satisfaisants, alors que le printemps et l'été semblent des saisons difficiles à appréhender et présentent une large gamme d'erreurs. Pour le MLP M1, nous pouvons remarquer que le RMSE a augmenté de 34.87 W/m^2 en hiver et de 43.41 W/m^2 durant l'automne à 51.78 W/m^2 pour l'été et à 65.83 W/m^2 pour le printemps pour le site de Benguerir. Pour le LSTM, en hiver le RMSE est de 41.89 W/m^2 alors que sa valeur a augmenté jusqu'à 72.64 W/m^2 durant l'été pour le même site de Benguerir. D'après les barres de la Figure 4-26, nous pouvons conclure que les quatre sites présentent les mêmes tendances saisonnières où l'erreur augmente de l'automne vers l'été. Cette tendance est un peu imprévue vu que dans la majorité des études la saison hivernale présente la saison la plus difficile à prévoir à cause de la présence d'une forte couverture nuageuse et des conditions météorologiques instables. En revanche, les meilleures prévisions sont obtenues durant la saison estivale qui est connue par l'absence des nuages et la stabilité des conditions météorologiques. Ce qui est conforme avec ce que nous avons obtenu lors de l'évaluation des prévisions en fonction de type de jours, où les meilleures performances prédictives ont été obtenues pour les jours clairs sans nuages et la précision démunie avec la présence de la couverture nuageuse. Pour interpréter les résultats saisonniers obtenus et savoir la raison derrière les mauvaises performances durant la saison estivale, nous avons approfondi l'analyse de la climatologie des sites par rapport à la couverture nuageuse et les aérosols, l'analyse conduite est détaillée dans la section suivante.

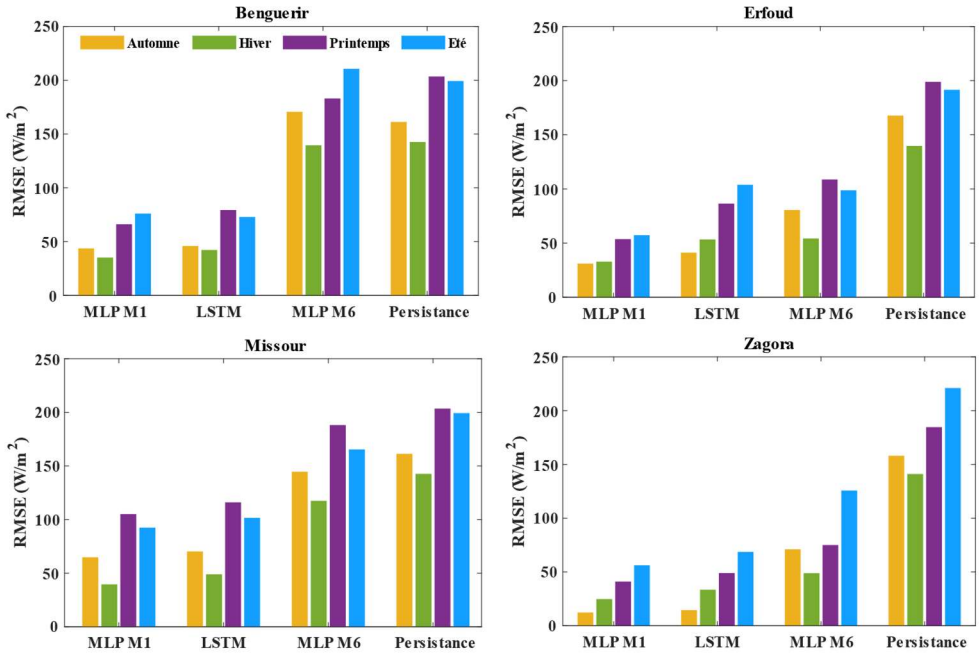
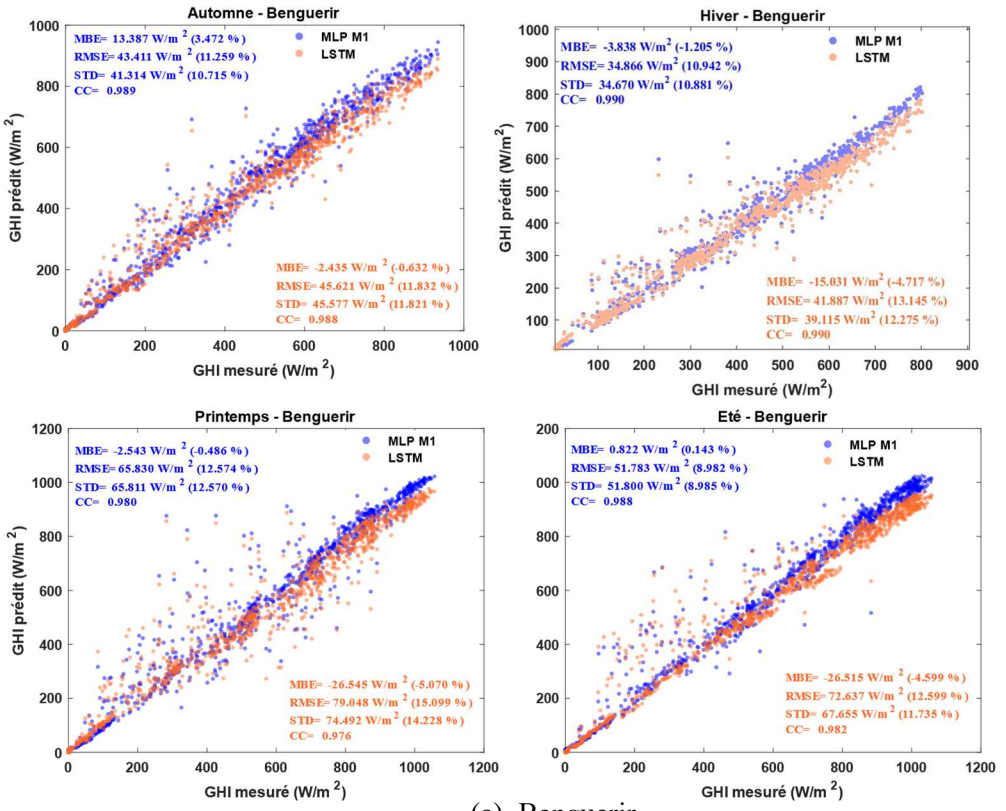


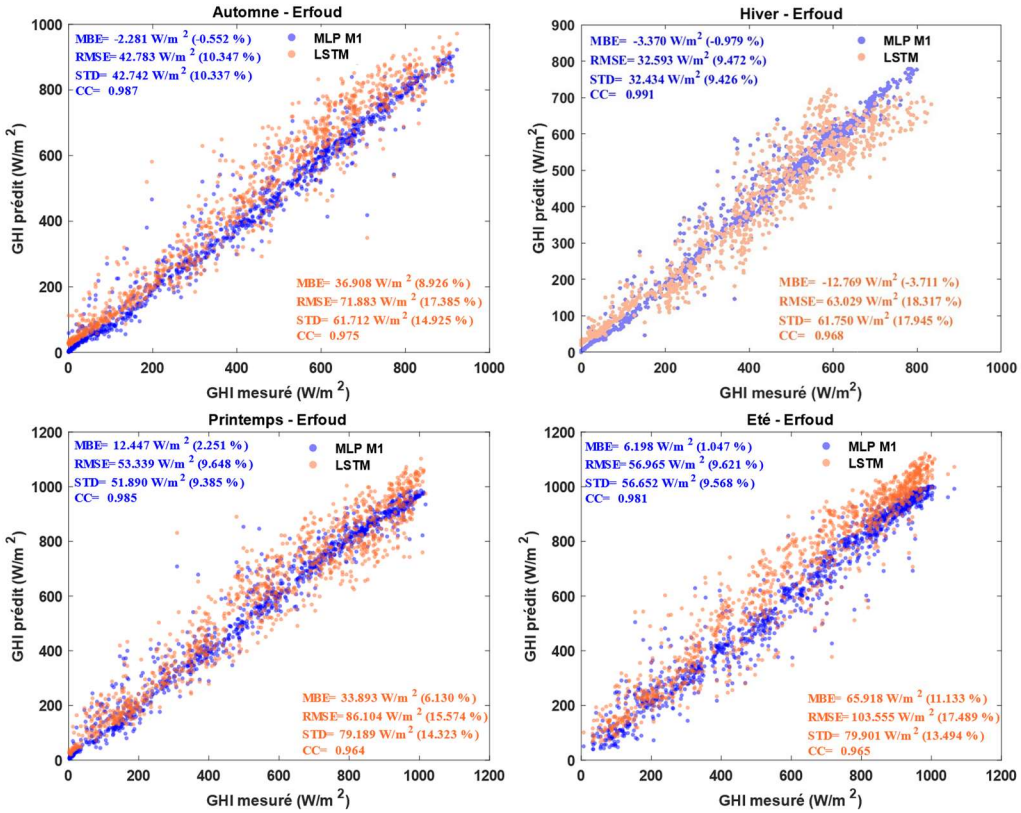
Figure 4-27. Erreurs saisonnières des modèles de prévision du GHI horaire 12h à l'avance

Tableau 4-7 performances des modèles de prévision.

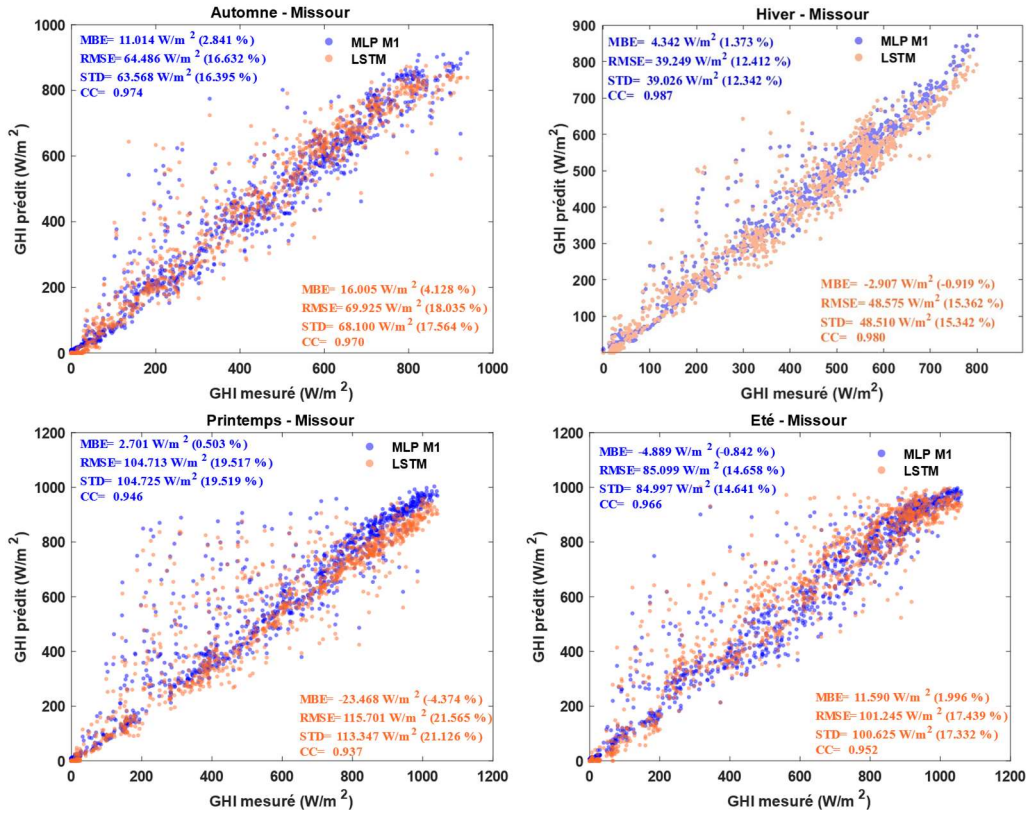
Type de jour		Benguerir				Erfoud				Missour				Zagora			
		MBE (W/m^2)	RMSE (W/m^2)	rRMSE (%)	SS (%)	MBE (W/m^2)	RMSE (W/m^2)	rRMSE (%)	SS (%)	MBE (W/m^2)	RMSE (W/m^2)	rRMSE (%)	SS (%)	MBE (W/m^2)	RMSE (W/m^2)	rRMSE (%)	SS (%)
AI		Moyenne journalière (W/m^2) : 665.27				Moyenne journalière (W/m^2) : 514.99				Moyenne journalière (W/m^2) : 622.30				Moyenne journalière (W/m^2) : 559.56			
	Persistance	-93.44	140.96	21.51		-52.46	95.73	18.59		-159.77	178.09	28.62		18.47	21.15	3.78	
	SARIMA	-38.84	41.45	6.32	70.60	-63.67	70.45	13.68	26.41	-2.92	28.88	4.64	83.78	-20.76	37.19	6.65	-75.84
	MLP M1	-14.56	24.27	3.70	82.78	-29.99	36.47	7.08	61.91	5.16	20.89	3.36	88.27	-6.10	12.97	2.32	38.66
	MLP M6	-40.14	77.62	11.84	44.94	-14.15	45.20	8.78	52.79	19.85	60.91	9.79	65.80	-65.63	98.47	17.60	-365.59
	LSTM	-32.95	65.47	9.99	53.56	5.62	26.63	5.17	72.18	-14.32	18.58	2.99	89.57	-24.51	33.65	6.01	-59.13
AII		Moyenne journalière (W/m^2) : 439.51				Moyenne journalière (W/m^2) : 576.72				Moyenne journalière (W/m^2) : 566.30				Moyenne journalière (W/m^2) : 486.24			
	Persistance	-55.49	89.14	20.28		-168.73	269.31	46.70		-140.89	250.10	44.16		39.75	208.75	42.93	
	SARIMA	-50.69	88.81	20.21	0.37	28.47	50.32	8.72	81.32	-272.86	342.64	60.50	-37.00	72.24	110.90	22.81	46.87
	MLP M1	-10.97	70.04	15.94	21.43	-36.80	62.10	10.77	76.94	19.44	74.81	13.21	70.09	50.52	83.40	17.15	60.05
	MLP M6	-28.07	55.07	12.53	38.22	-60.67	74.73	12.96	72.25	9.57	99.14	17.51	60.36	60.60	101.59	20.89	51.33
	LSTM	-17.36	68.65	15.62	22.99	-12.75	55.11	9.56	79.54	-19.95	82.45	14.56	67.03	13.69	57.27	11.78	72.57
BII		Moyenne journalière (W/m2) : 298.73				Moyenne journalière (W/m^2) : 269.02				Moyenne journalière (W/m^2) : 377.93				Moyenne journalière (W/m^2) : 378.51			
	Persistance	238.91	302.85	101.38		123.69	261.52	89.73		168.38	230.39	60.96		29.30	134.39	35.51	
	SARIMA	203.62	317.31	106.22	-4.77	90.96	138.78	51.59	46.93	160.87	227.82	60.28	1.11	17.03	49.57	13.10	63.11
	MLP M1	127.20	222.92	74.62	26.39	-55.79	86.52	32.16	66.92	35.29	102.01	26.99	55.72	40.33	76.78	20.28	42.87
	MLP M6	-14.81	109.74	36.74	63.76	41.75	105.04	39.05	59.83	-62.18	193.63	51.23	15.95	17.68	37.87	10.01	71.82
	LSTM	-5.86	122.82	41.12	59.44	50.33	87.46	32.51	66.56	-10.99	167.93	44.43	27.11	6.35	54.60	14.43	59.37
CIII		Moyenne journalière (W/m2) : 151.31				Moyenne journalière (W/m^2) : 261.63				Moyenne journalière (W/m^2) : 164.46				Moyenne journalière (W/m^2) : 171.17			
	Persistance	260.45	353.16	233.40		57.26	82.27	31.45		102.85	234.19	142.40		211.80	270.70	158.14	
	SARIMA	302.36	388.53	256.77	-10.01	39.84	62.70	23.97	23.79	122.18	236.40	143.74	-0.94	175.68	231.02	134.96	14.66
	MLP M1	44.30	60.80	40.18	82.79	-75.40	141.17	53.96	-71.59	34.31	76.05	46.24	67.53	55.28	82.57	48.24	69.50
	MLP M6	13.21	106.08	70.11	69.96	-102.49	158.25	60.49	-92.34	54.67	112.56	68.44	51.94	37.67	89.17	52.09	67.06
	LSTM	63.74	124.82	82.49	64.66	-98.30	147.24	56.28	-78.96	-41.89	103.75	63.08	55.70	36.76	112.07	65.47	58.60



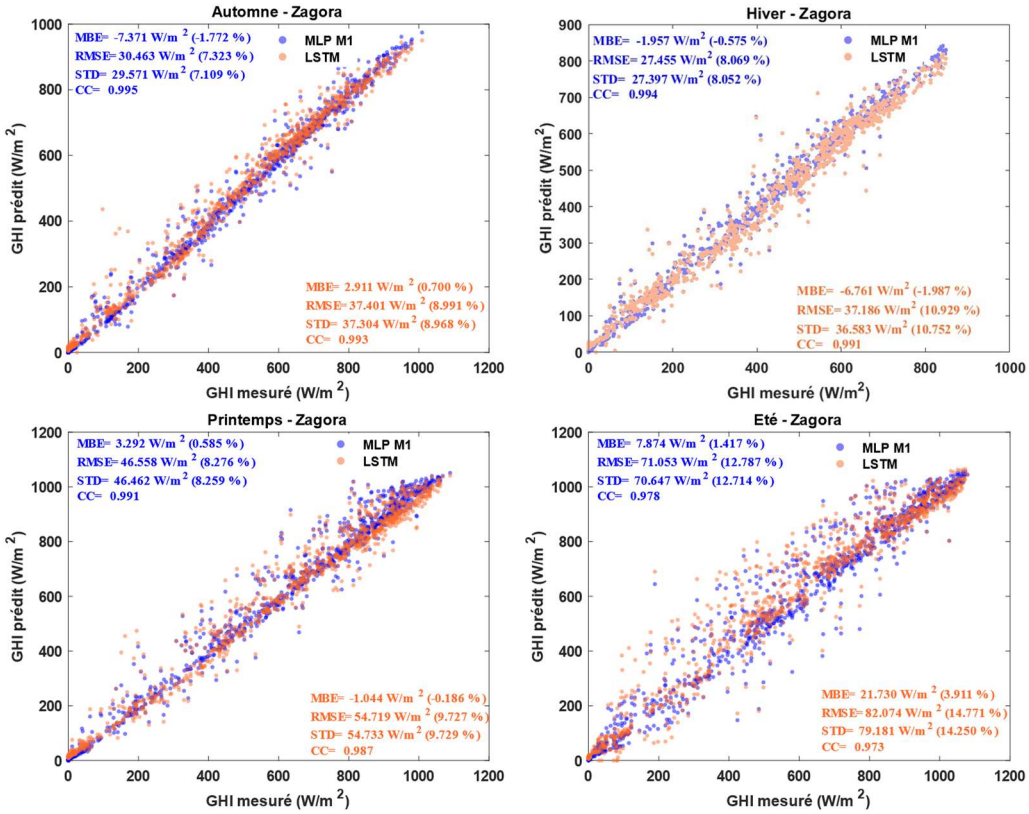
(a) Benguerir



(b) Erfoud



(c) Missour



(d) Zagora

Figure 4-28. Nuages de point du GHI saisonnier mesuré et prédit par le MLP M1 (bleu) et LSTM (orange) à l'horizon de 12h au pas horaires pour (a) Benguerir, (b) Erfoud, (c) Missour, et (d) Zagora.

Tableau 4-8. RMSE et CC saisonniers obtenue pour les quatre sites d'étude.

Modèle	Benguerir		Erfoud		Missour		Zagora	
	RMSE (W/m^2)	CC -	RMSE (W/m^2)	CC -	RMSE (W/m^2)	CC -	RMSE (W/m^2)	CC -
Automne								
MLP M1	43.41	0.98	42.78	0.99	64.49	0.97	43.01	0.99
LSTM	45.62	0.98	71.88	0.98	69.92	0.97	50.15	0.98
MLP M6	170.35	0.79	80.23	0.95	144.23	0.86	70.74	0.97
Persistance	160.95	0.83	167.45	0.81	165.24	0.82	157.69	0.84
Hiver								
MLP M1	34.87	0.99	32.59	0.99	39.25	0.99	24.47	0.99
LSTM	41.89	0.99	63.03	0.97	48.58	0.98	33.12	0.99
MLP M6	139.23	0.83	53.92	0.97	117.15	0.88	48.53	0.98
Persistance	142.30	0.83	139.39	0.83	130.91	0.85	140.77	0.85
Printemps								
MLP M1	65.83	0.98	53.34	0.98	104.71	0.95	40.72	0.99
LSTM	79.05	0.97	86.10	0.96	187.76	0.82	48.63	0.99
MLP M6	182.73	0.83	108.38	0.93	115.70	0.94	74.73	0.97
Persistance	203.17	0.80	198.51	0.78	210.21	0.78	184.37	0.83
Eté								
MLP M1	51.78	0.98	56.96	0.98	85.10	0.96	55.89	0.98
LSTM	72.64	0.98	103.56	0.96	210.21	0.78	68.30	0.98
MLP M6	210.32	0.78	98.41	0.94	165.06	0.87	125.38	0.92
Persistance	198.96	0.81	191.23	0.79	190.46	0.82	220.70	0.76

3.1.3. Impact des aérosols et de la couverture nuageuse sur la précision des prévisions du GHI

Les prévisions solaires dépendent de plusieurs paramètres météorologiques, climatiques, et atmosphériques. En l'absence des nuages, les aérosols sont la principale source d'atténuation du rayonnement solaire, tandis que sous les conditions de ciel clair, le rayonnement solaire est fortement impacté par les aérosols [298].

Les aérosols sont un mélange de minuscules particules solides et liquides en suspension dans l'atmosphère. Les aérosols comprennent la poussière minérale, la poussière transportée par le vent, les cendres volcaniques, la brume, le brouillard, la fumée des incendies, le sel marin et la pollution particulaire due aux activités naturelles et anthropiques [299]. Cette définition est suffisamment large pour englober des particules de nature, de taille, et de formes très variées. Étant donné la grande diversité d'aérosols atmosphériques, il est nécessaire de les classer pour pouvoir décrire plus précisément le phénomène souhaité. Dans notre cas, nous utiliserons la profondeur optique des aérosols (AOD) à 550 nm qui est un paramètre clé qui reflète les caractéristiques optiques des aérosols; elle représente le degré auquel les aérosols inhibent la transmission de la lumière par des phénomènes de diffusion et d'absorption.

Les informations fiables sur les aérosols sont limitées en raison de leur grande variabilité spatiale et temporelle. Les données sur l'AOD peuvent être mesurées au niveau du sol à l'aide des photomètres solaires tels que pour le réseau Aerosol Robotic NETwork (AERONET). Bien

qu'AERONET fournit des informations de haute précision sur l'AOD par rapport aux données satellitaires, sa couverture spatiale est encore très limitée. Cependant, dans le monde entier, de nombreuses études ont été consacrées à l'évaluation de produits satellitaires tels que le Moderate Resolution Imaging Spectroradiometer (MODIS) par rapport aux mesures au sol d'AERONET, afin de les utiliser comme alternatives aux mesures au sol.

Dans notre cas, pour évaluer l'impact des AOD sur la précision des prévisions, nous utiliserons les données dérivées de la base de données satellitaire MODIS. Avant d'exploiter les données de MODIS, cette base de données a été validée par des données mesurées au sol à partir du réseau AERONET pour six sites au Maroc (Ouarzazate, Oujda, Oukaimden, Ras El Ain, Saada et Dakhla).

Variations saisonnières des aérosols

Compte tenu de sa situation géographique, le Maroc possède un potentiel solaire important et est également caractérisé par la dominance des conditions de ciel clair (Figure 4-3). En parallèle, le pays est très sensible aux aérosols élevés et souffre d'épisodes de poussière du désert saharien et de tempêtes de sable [300]. Comme nous pouvons l'observer sur la Figure 4-29 où un grand panache de poussière du désert africain soufflant vers l'Ouest sur l'océan Atlantique de la côte marocaine, et décrivant un arc vers le Nord-Est jusqu'aux îles Canaries dans l'Atlantique oriental [301]. La Figure 4-30 illustre des images réelles prises au site GEP pendant une tempête de poussière.

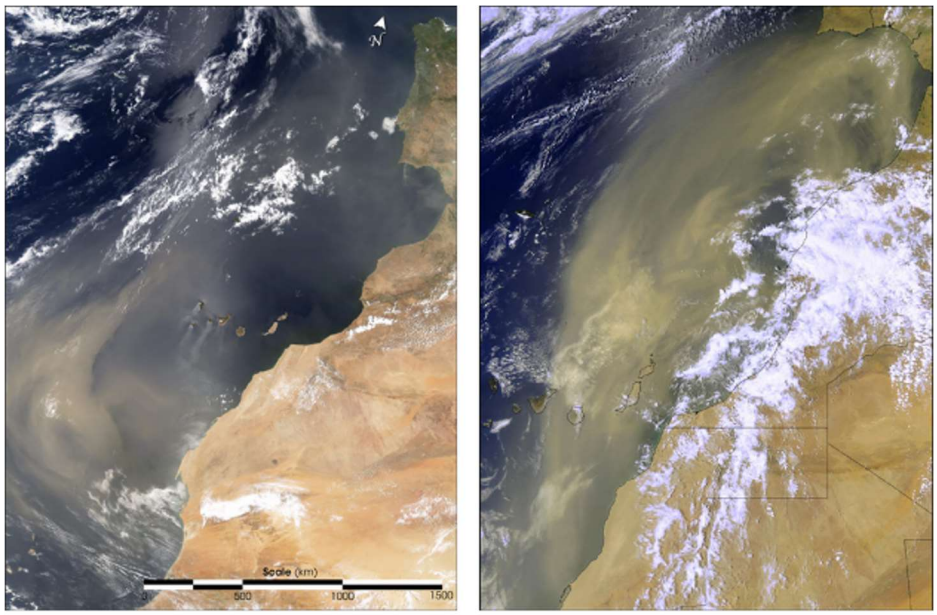


Figure 4-29. Transport de poussières/aérosols du Sahara vers les îles Canaries (image de MODIS [301]).



Figure 4-30. Photos réelles prises au site de GEP (Benguerir) lors d'une tempête de poussière les 21 et 22 avril 2018.

Pour approfondir l'analyse sur l'incertitude des prévisions saisonnières, nous présenterons les variations saisonnières des aérosols ainsi que de la couverture nuageuse pour les sites d'étude pour la même période du test des prévisions. Les données de la couverture nuageuse ont été aussi obtenues à partir de MODIS, les données de la couverture nuageuse n'ont été pas validé vue le manque des mesures au sol de cette dernière. La Figure 4-31 représente les variations mensuelles de l'AOD à 550 nm pour les quatre sites du 01/01/2017 au 31/12/2020, collectées à partir de MODIS. La Figure 4-32 représente des histogrammes de la couverture nuageuse pour chaque saison. Les jours sont classifiés clairs lorsque la couverture nuageuse est entre 0 et 0.5, alors que les jours avec une couverture nuageuse supérieure à 0.5 sont considérés comme étant des jours nuageux. Le Tableau 4-9 représente les pourcentages des jours clairs et nuageux pour chaque site en fonction des saisons ainsi que les valeurs saisonnières de l'AOD. D'après la Figure 4-32 et le Tableau 4-2, nous pouvons remarquer que les sites sont caractérisés par la dominance des jours clairs même en hiver avec des pourcentages variant entre 67% 86% pour l'ensemble des sites. La saison d'automne est aussi caractérisée par un pourcentage élevé des jours clairs variant entre 70% et 94% alors que les jours nuageux présentent seulement des pourcentages entre 6% et 30%. Les plus grands pourcentages des jours clairs sont obtenus pour la saison d'été arrivant jusqu'à 87%. D'autre part, si nous observons les moyennes saisonnières de l'AOD, nous pouvons donc remarquer que l'AOD est faible durant l'automne et l'hiver et sa valeur augmente d'une manière spectaculaire durant le printemps et l'été. À partir de l'analyse de ces deux paramètres, nous pourrions interpréter les performances saisonnières des prévisions; où les meilleures performances prédictives ont été obtenues durant l'hiver et

l'automne et commencent à diminuer vers l'été. Lorsque nous considérons le site de Benguerir à titre d'exemple, en hiver le site est caractérisé par 67 % des jours clairs et 33 % de jours nuageux, alors que pour l'été nous avons 87% de jours clairs et seulement 13 % de jours nuageux, alors que l'erreur de prévision pour le MLP M1 était de 34.87 W/m^2 pour l'hiver et a augmenté à 51.75 W/m^2 en été, et jusqu'à 65.83W/m^2 au printemps même si les jours clairs présentent 75% pour cette saison. Normalement, nous devrions trouver le contraire, mais les résultats obtenus sont tout à fait logiques ; si nous observons les valeurs de l'AOD, nous pouvons remarquer que l'AOD a augmenté de 0.10 en hiver à 0.21 en été sur le même site de Benguerir. En effet, en absence de nuages, les aérosols jouent un rôle important dans l'atténuation et la variation du rayonnement solaire et peuvent parfois avoir un impact similaire à celui des nuages dans les conditions de ciel clair, ce qui explique les bonnes performances prédictives obtenues durant l'hiver qui est caractérisé par un grand pourcentage de jour clair et de faibles valeurs d'AOD, alors que pour l'été même si nous avons un grand pourcentage de jours clairs, mais aussi de grandes valeurs d'AOD. Cette constatation est plus remarquable sur le site de Zagora. Pour l'hiver, nous avons obtenu un RMSE de 24.47 W/m^2 avec 94% de jours clairs et un AOD de 0.19. Cependant, durant l'été l'AOD a pu atteindre 0.37 et le RMSE a aussi augmenté à 55.89 même si les conditions nuageuses ne présentaient que 12%. Nous pourrions conclure que malgré le fort potentiel solaire dans les divers sites d'étude et l'absence des nuages, la présence de particules d'aérosol affecte considérablement le rayonnement solaire que ce soit mesuré ou prédit.

Tableau 4-9. Couverture nuageuse et AOD saisonniers pour les quatre d'étude

	Automne			**Hiver**			**Printemps**			**Été**		
	Clair	Non clair	AOD	Clair	Non Clair	AOD	Clair	Non clair	AOD	Clair	Non clair	AOD
	(%)	(%)	(-)	(%)	(%)	(-)	(%)	(%)	(-)	(%)	(%)	(-)
Benguerir	70	30	0.17	67	33	0.10	75	25	0.19	87	13	0.21
Erfoud	75	25	0.19	85	15	0.09	75	25	0.26	87	13	0.34
Missour	76	24	0.18	79	21	0.10	66	34	0.24	84	16	0.25
Zagora	94	6	0.19	86	14	0.09	82	18	0.22	88	12	0.37

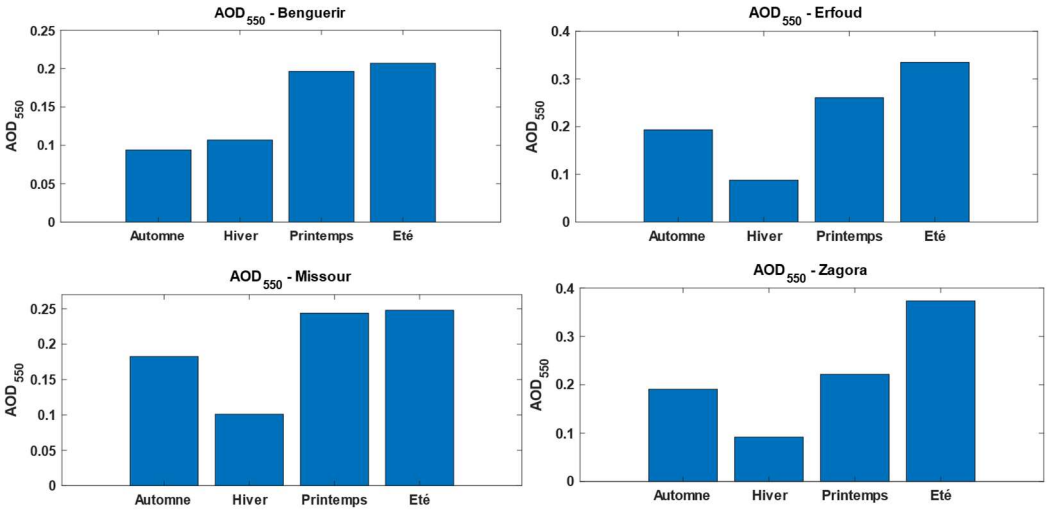
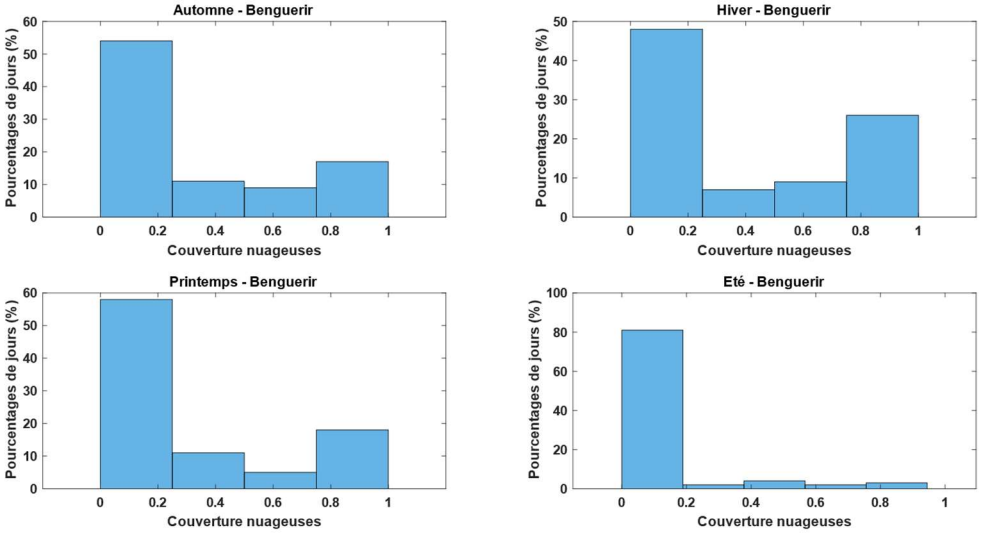
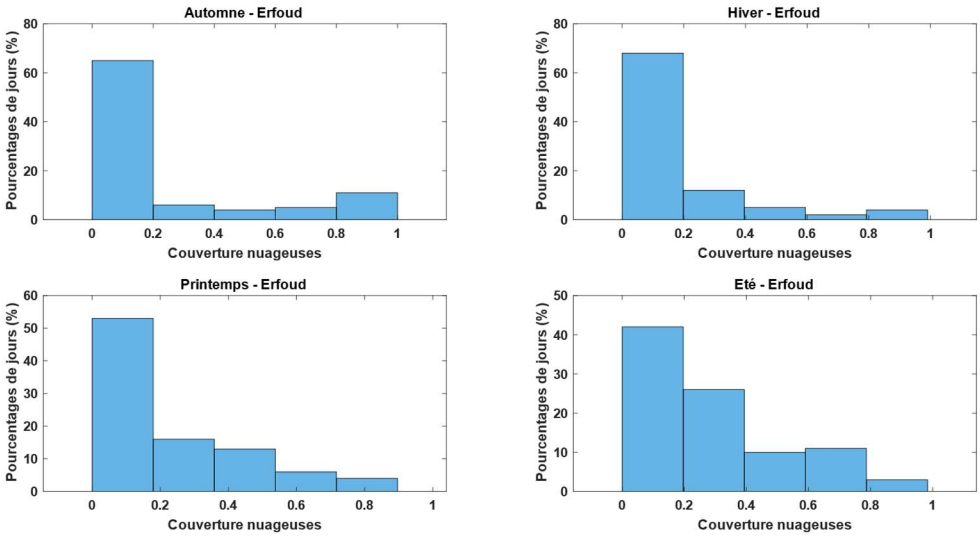


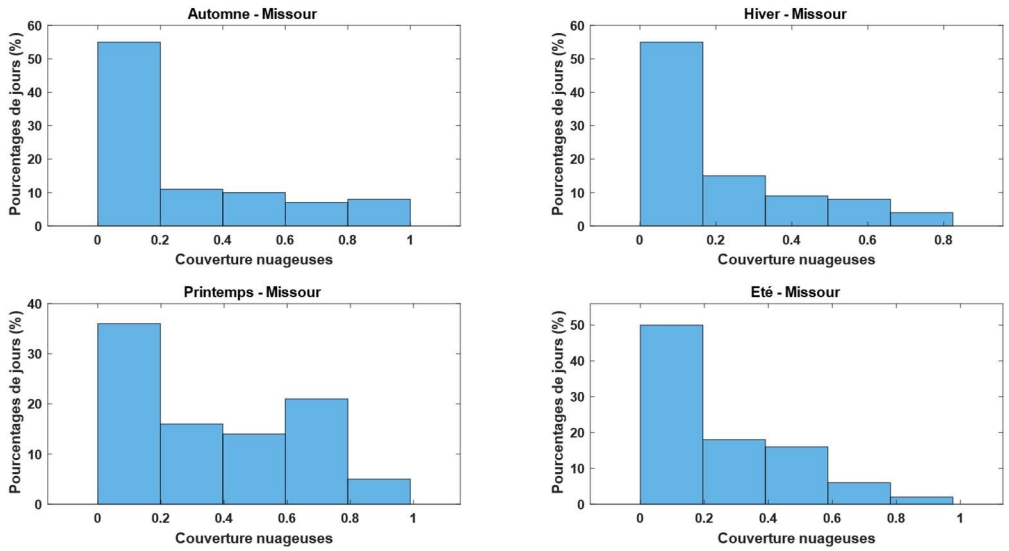
Figure 4-31. Moyennes saisonnières des AOD_{550} pour les sites d'études sur la période de test.



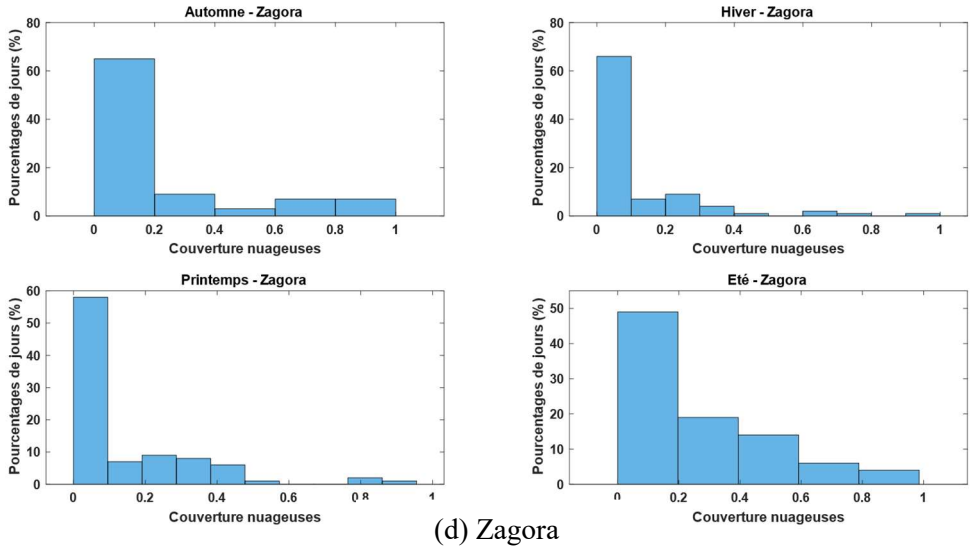
(a) Benguerir



(b) Erfoud



(c) Missour



(d) Zagora

Figure 4-32. Histogramme de distribution de la couverture nuageuse pour les quatre saisons à (a) Benguerir, (b) Erfoud), (c) Missour, et (d) Zagora sur la période de test.

3.2. Prévision à très court terme

Bien que la prévision à très court terme du GHI n'été pas prioritaire pour nous, mais la possibilité de prévoir le rayonnement solaire quelques minutes à l'avance présente un enjeu pour la gestion de la variabilité de l'énergie intermittente. En effet, pour le très court terme, les nuages ont un impact significatif sur la baisse et les fluctuations de l'énergie produite. Des prévisions à quelques minutes à l'avance sont nécessaires pour atténuer l'effet de la baisse de production dû aux nuages. L'horizon de ces prévisions est caractérisé par sa haute fréquence d'échantillonnage des mesures permettant ainsi de capturer la dynamique des nuages. En effet, à 15 minutes le ciel a une grande possibilité de conserver la même couverture nuageuse, ce qui

nous permettra de capturer la variation des nuages vu que les données utilisées pour la prévision ont une résolution d'une minute. Comme détaillé dans la section 2.3, la méthodologie utilisée pour la prévision à cet horizon diffère de ce que nous avons appliqué pour les prévisions à court terme. Avec le développement récent des techniques de la photogrammétrie, les caméras hémisphériques installées au sol pour capturer le mouvement des nuages sont maintenant largement utilisées. Les images capturées sont ensuite exploitées grâce à des outils de traitement d'image pour extraire les caractéristiques contenues afin de prévoir le GHI à très court terme.

Dans cette partie, nous présentons les résultats du modèle CNN-MLP discuté dans la section 2.3 pour prévoir le GHI 15 minutes à l'avance en utilisant les images prises par la caméra hémisphérique ainsi que les paramètres météorologiques. Les données relatives aux enregistrements des images sont du 01/01/2020 au 21/12/2020 avec un pas d'échantillonnage d'une minute. Les séries temporelles du GHI et d'autres variables météorologiques utilisées en entrée ont aussi un pas d'une minute (Figure 4-20).

Les performances du modèle CNN-MLP pour la prévision à très court terme sont présentées pour trois types de jours : clair, partiellement nuageux, et nuageux. La Figure 4-33 représente des images prises par l'imageur installé au GEP et correspondant aux divers instants : clair, partiellement nuageux et nuageux . La Figure 4-34 présente un exemple du profil journalier du GHI à 1 minute et les images correspondantes à quelques instants de la journée. Nous observons sur la Figure 4-34 que la première tranche de jour est claire avec une courbe de rayonnement lisse et parallèlement sur les images, nous observons l'absence des nuages, alors qu'à partir de 11:00 le rayonnement commence à fluctuer où nous observons la présence de grandes masses nuageuses jusqu'à la fin de la journée. Ceci justifie que l'information sur la couverture nuageuse extraite à partir des images présente une caractéristique importante pour l'amélioration de la précision des prévisions. Ces instants de ciel clair peuvent être détectés à partir de la série temporelle du GHI à l'aide de la procédure présentée dans la section 3.1 du chapitre 3, mais malheureusement la procédure n'est pas facilement applicable en mode prévision. Dans notre cas, pour l'identification des nuages, nous nous sommes basés sur les techniques d'apprentissage automatique en exploitant les images des caméras et les données météorologiques sans passer par la procédure de détection des instants de ciel clair.

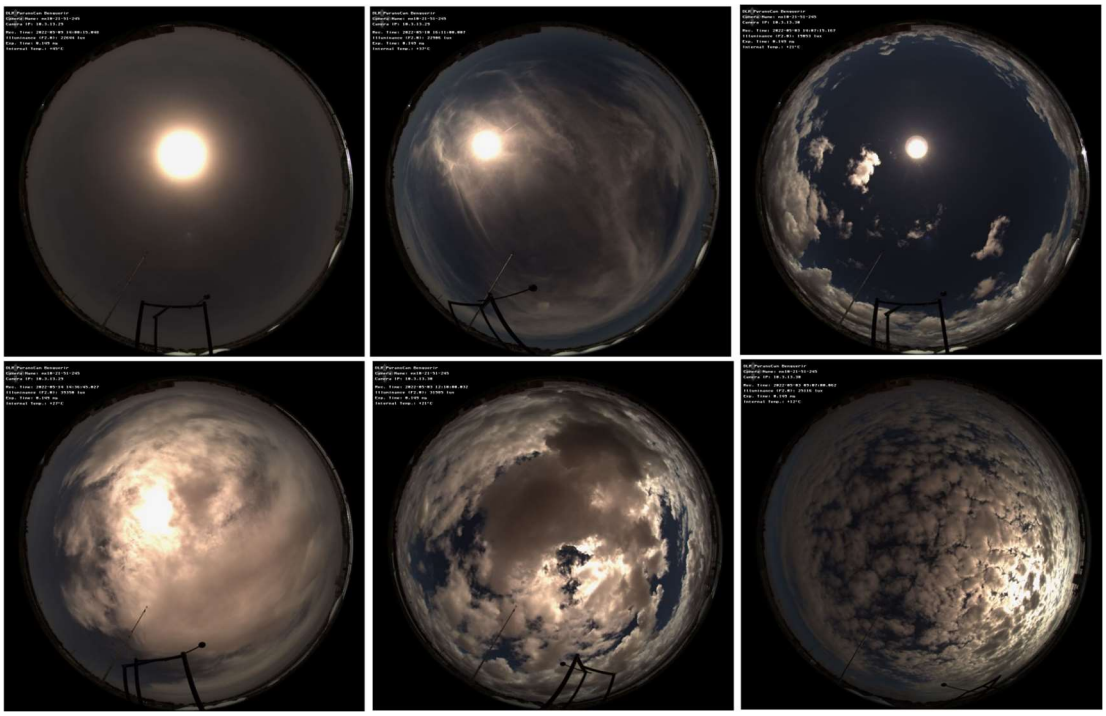


Figure 4-33. Exemples des images prises par l'imageur du ciel installé au GEP à différents instants (clair, mixte, nuageux) pour des journées choisies à partir de la base de données.

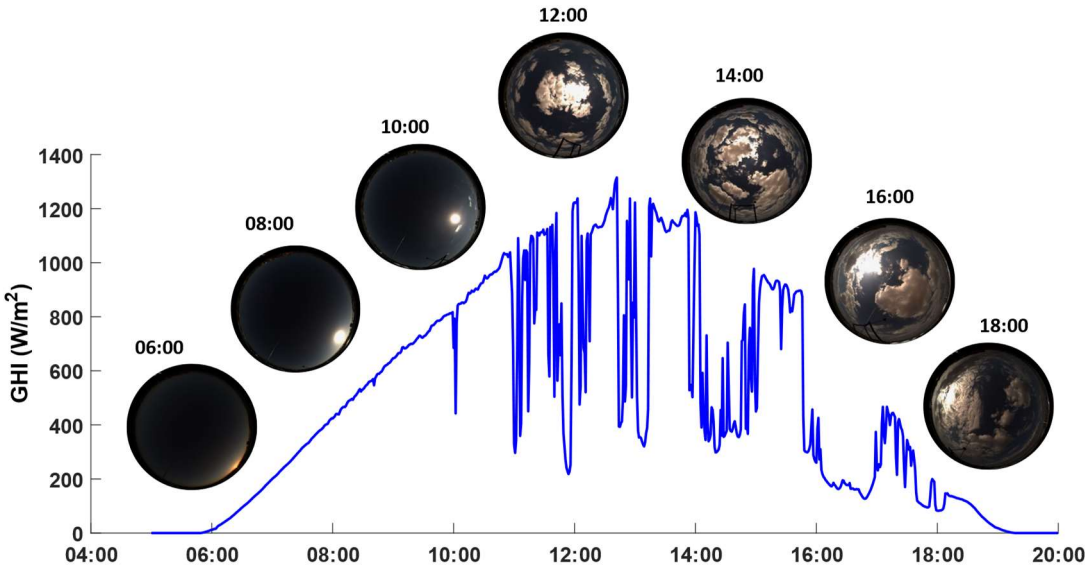


Figure 4-34. Exemple du profil journalier du GHI au pas d'une minute pour un jour mixte avec les images correspondantes aux quelques instants de la journée.

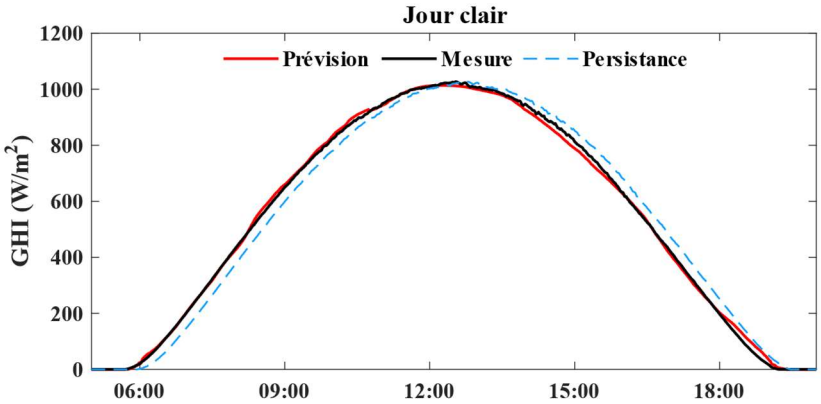
La Figure 4-35 illustre la comparaison du profil journalier du GHI prédit par le modèle CNN-MLP, et par la persistance qui est le modèle de référence utilisé, par rapport aux mesures expérimentales pour les trois types de jours. Le Tableau 4-10 résume les MBE, les RMSE, les rRMSE, et les coefficients de corrélation obtenus par le modèle CNN-MLP et la persistance. La Figure 4-36 représente une illustration graphique de ces indicateurs de performance. D'après les performances obtenues, nous pouvons remarquer que les erreurs de prévision à très court terme sont significatives par rapport à la prévision à court terme, ce qui prouve la

complexité de la prévision à très terme. D'autre part, le modèle de persistance présente les valeurs d'erreur les plus élevées surtout pour les jours mixtes et nuageux, alors le modèle de persistance ne peut pas être considéré comme un prédicteur performant même pour la prévision à très court terme.

En revanche, les résultats de comparaison indiquent que le modèle développé présente des résultats satisfaisants. Pour les jours clairs, le modèle sous-estime légèrement le GHI avec un MBE de -0.87 W/m^2, alors qu'une surestimation de 3.85 W/m^2 est obtenue avec la persistance. A son rôle, le RMSE augmente d'une valeur de 13.05 W/m^2 avec le CNN-MLP à 45.76 W/m^2 avec la persistance. Pour les jours mixtes, le modèle CNN-MLP a également démontré sa capacité prédictive, avec un MBE de -1.66 W/m^2, et un RMSE de 41.7W/m^2 alors que pour la persistance, le MBE est de 9.93 W/m^2 et le RMSE a augmenté jusqu'à 88.10 W/m^2.

Pour les jours nuageux, les erreurs sont plus significatives. Le CNN-MLP sous-estime le GHI avec un MBE de -7.66 W/m^2, un RMSE de 49.16 W/m^2, et un CC de 0.94 qui restent des valeurs acceptables vu la complexité des conditions météorologiques. Concernant la persistance, les erreurs de prévision sont plus significatives, avec un RMSE de 114.19 W/m^2 et un CC de 0.70, la faible valeur de MBE obtenues malgré les fortes fluctuations peut être expliquée par le fait que pour ce type de jours il y avait de grands nuages avec une vitesse de dispersion très lente ce qui a permis le ciel de conserver sa couverture nuageuse tout au long de l'horizon de prévision d'où résulte la faible valeur de MBE.

En conclusion, les résultats ont montré que le CNN-MLP développé est très approprié pour prédire les jours ensoleillés et mixtes avec des CC de 0.99 et 0.98, respectivement. Alors qu'il fournit également des résultats acceptables pour les jours nuageux (CC=0.94) par rapport à la persistance (CC=0.70). D'après les variations des profils journaliers, nous pouvons observer que le modèle tente à capturer les variations à très court terme, cela revient au fait que le modèle prend en considération en plus des variables météorologiques, l'information sur la couverture nuageuse (mouvement, évolution, type, etc.) qui est la raison fondamentale derrière les fluctuations du rayonnement solaire à très court terme.



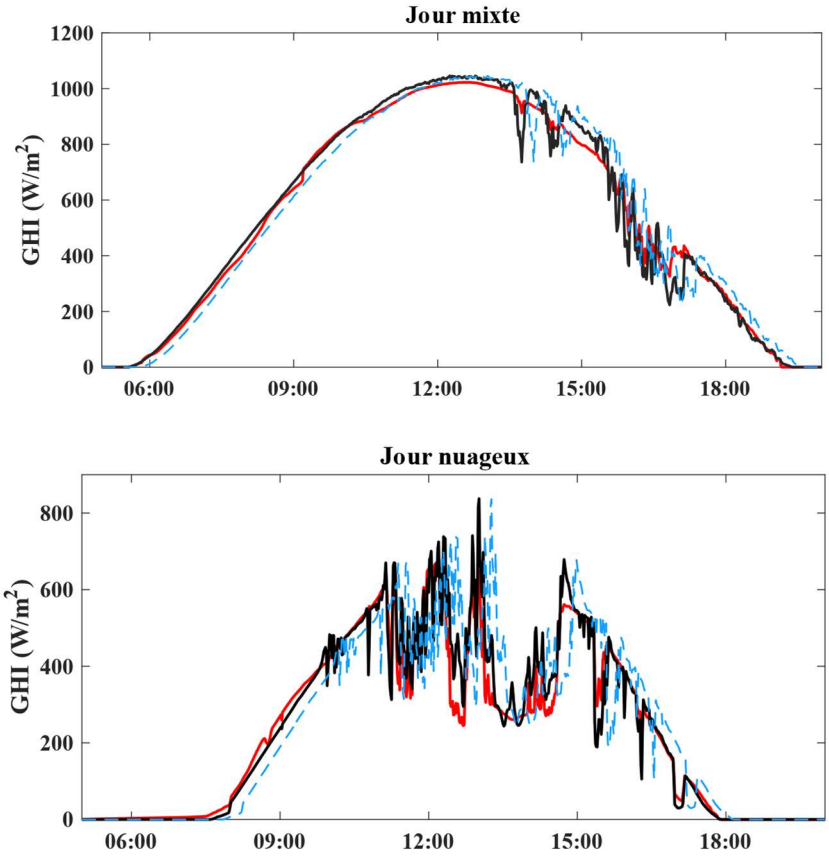


Figure 4-35. Profils journaliers du GHI prévu à l'horizon de 15 min par rapport au modèle de persistance, pour un jour clair, mixte et nuageux.

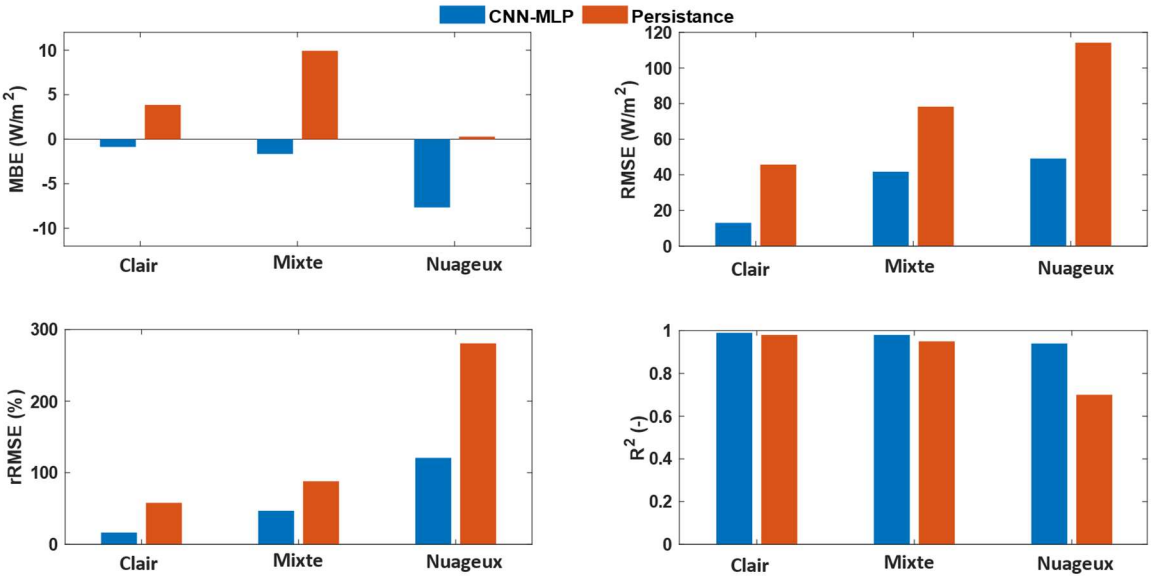


Figure 4-36. Résultats de prévision avec le modèle CNN-MLP en comparaison avec la persistance pour l'horizon de 15 min.

Tableau 4-10. Comparaison des performances statistiques entre le GHI mesuré et prévu par CNN-MLP et la persistance.

	CNN-MLP				Persistance			
	MBE W/m^2	RMSE W/m^2	rRMSE %	CC -	MBE W/m^2	RMSE W/m^2	rRMSE %	CC -
Jour clair	-0.87	13.05	16.54	0.99	3.85	45.76	57.99	0.98
Jour mixte	-1.66	41.74	46.98	0.98	9.93	78.27	88.10	0.95
Jour nuageux	-7.66	49.16	120.83	0.94	0.29	114.19	280.66	0.70

L'étude bibliographique a montré l'importance de l'utilisation des images au sol pour la prévision à très court terme afin de soutenir la répartition et la gestion optimales des systèmes électriques avec une forte pénétration de la production d'énergie photovoltaïque. Dans ce contexte, nous avons proposé un modèle hybride CNN-MLP. Le CNN est utilisé pour extraire les caractéristiques des nuages qui sont considérés la principale cause des fluctuations du GHI et le MLP pour extraire les caractéristiques contenues dans les diverses variables météorologiques considérées. Les performances du modèle proposé sont rigoureusement évaluées en analysant les résultats de prévision pour différentes conditions de ciel. Les résultats de prévision indiquent que le modèle proposé fournit généralement de bonnes performances prédictives par rapport au modèle de référence (la persistance) avec un coefficient de corrélation de 0.99 pour les jours clairs et de 0.94 pour des conditions de ciel plus complexe.

Conclusion

Ce chapitre décrit les modèles de prévision développés au cours de cette thèse pour la prévision du GHI à court terme et à très court terme. Lors de développement des modèles, nous avons considéré l'horizon h+12 pour le court terme et 15 min pour le très court terme. Pour le court terme, nous avons évalué deux modèles basés sur le MLP et un modèle basé sur LSTM. Après plusieurs tests et simulations, la meilleure architecture a été choisie pour chaque modèle. Le MLP M1 et LSTM utilisent plusieurs variables météorologiques ainsi que des calculs de la géométrie solaire (température, humidité relative, vitesse et direction du vent, angle solaire zénithal, rayonnement au top de l'atmosphère et la durée d'ensoleillement), alors que le MLP M6 ne considère que des calculs de la géométrie solaire (angle solaire zénithal, rayonnement au top de l'atmosphère et la durée d'ensoleillement). Les performances prédictives des modèles proposés ont été évaluées sur une année entière, sous diverses saisons et condition de ciel. Quatre ensembles de données collectés à partir de quatre emplacements au Maroc ayant des climats différents ont été impliqués pour l'évaluation des modèles proposés. En plus, les performances des modèles de prévision ont été comparées aux performances obtenues avec les modèles SARIMA et la persistance sur la base de plusieurs métriques d'évaluation. Les résultats d'évaluation des prévisions par rapport aux types du jour ont montré que la précision des prévisions dépend largement au type du jour. Les plus faibles valeurs d'erreurs ont été obtenues pour les jours clairs avec une faible variabilité puis l'erreur augmente pour les jours nuageux et variables. Le MLP M1 et LSTM bénéficient des meilleures performances

prédictives durant les jours clairs et mixtes avec une variabilité moyenne, alors qu'une légère détérioration des performances est observée dans les conditions nuageuses avec une forte variabilité. Pour les jours clairs, SARIMA et le MLP M6 présentent de bonnes performances vu leurs simplicités, mais ils ne peuvent pas être utilisés dans des conditions de ciel plus complexe. D'après divers scénarios d'évaluation des modèles de prévision, les modèles MLP M1 et LSTM s'avèrent être de bons choix pour la prévision à court terme même sous des conditions de ciel nuageux et partiellement nuageux.

Par la suite, les résultats de la prédiction saisonnière démontrent que la distribution saisonnière des erreurs de prédiction change en fonction des saisons. À l'issue de divers tests, un élément peut nous surprendre : selon l'emplacement considéré, les résultats de prévision peuvent être totalement différents et les modèles de prévision ne peuvent pas être facilement portables et généralisables. Dans la majorité des études menées dans le cadre de la prévision du GHI, les performances prédictives étaient bonnes en été et diminuent durant la saison hivernale à cause de la forte couverture nuageuse. Dans notre cas, nous avons obtenu l'inverse où les performances prédictives hivernales étaient meilleures que celles de la saison estivale, ce qui nous a menés à conduire une analyse approfondie pour pouvoir interpréter les résultats obtenus. Nous avons pu constater que pour le Maroc les nuages ne présentent qu'un faible pourcentage même en hiver et en automne, en revanche le pays est fortement impacté par la poussière. En absence des nuages, les aérosols sont le paramètre le plus important qui affecte le rayonnement solaire. Après la comparaison des résultats nous avons pu conclure que durant l'été nous avons un grand pourcentage de jours clairs, mais un fort taux d'aérosols, alors que durant l'hiver les sites sont caractérisés par un pourcentage remarquable des jours clairs et un faible taux d'aérosols ce qui justifie les résultats obtenus. Les résultats des prévisions saisonniers démontrent aussi que les modèles MLP M1 et LSTM proposés fournissent des résultats de prédiction fortement améliorés par rapport aux résultats des autres modèles considérés pendant les saisons où les erreurs de prédiction sont relativement importantes. Concernant la prévision à très court terme (15 min), une technique tout à fait différente a été utilisée en exploitant les images d'un imageur de ciel installé au sol, permettant de capturer le mouvement des nuages dans le ciel. De ce fait, nous avons développé un modèle hybride CNN-MLP permettant tout d'abord à travers le modèle CNN d'extraire les informations contenues dans les images surtout par rapport à la couverture nuageuse ; la première cause derrière la variabilité du GHI, pour par la suite les utiliser avec d'autres variables météorologiques et prévoir le GHI à travers le MLP à l'horizon souhaité. Les résultats obtenus montrent l'intérêt de l'utilisation des images du ciel et de l'information sur la couverture nuageuse pour la prévision à très court terme.

Conclusion générale et perspectives

Avec l'augmentation de la part de la production renouvelable dans le mix énergétique, un défi central dans l'évolution de ces ressources renouvelables consiste à élaborer des prévisions efficaces et la mise à disposition des résultats des prévisions. Sans prévisions précises, le taux de pénétration de l'énergie solaire dans le réseau électrique ne permettra pas de répondre aux attentes engendrées par le changement climatique. Par conséquent, les démarches présentées dans ce travail de recherche sont liées à cette problématique cruciale qui consiste la conception des méthodes de prévision du rayonnement solaire pour le court et le très court terme en mettant l'accent sur les conditions météorologiques complexes fréquemment rencontrées dans la pratique. Ces travaux de recherche ont été une occasion de contribuer à l'amélioration des prévisions solaires pour une meilleure gestion de l'énergie produite, étant donné que le pays suscite un intérêt exceptionnel à l'énergie solaire.

Dans cette thèse, nous nous sommes intéressés à l'application de méthodes d'apprentissage automatique pour la prévision du GHI pour le court (15 minutes) et le très court terme (12 heures). Tout au long de cette thèse, des données collectées à partir de quatre stations au sol au Maroc, des prévisions météorologiques d'un modèle numérique de prévision, des calculs de la géométrie solaire, des données satellitaires sur la couverture nuageuse et les aérosols et des images de ciel, ont été exploitées pour le développement et l'analyse des modèles de prévision.

Une connaissance spécifique du sujet a été indispensable pour pouvoir développer de différents prédicteurs du GHI. Dans un premier temps et avant t'entamer le problème de la prévision, une analyse du positionnement énergétique du Maroc en ce qui concerne les énergies renouvelables a été faite. Les résultats de cette analyse montrent que le Maroc dispose d'un fort potentiel énergétique renouvelable, notamment le solaire permettant ainsi un terrain fertile pour l'exploitation de cette énergie pour la production de l'énergie nationale. Nous avons aussi pu introduire les paramètres importants régissant le rayonnement solaire qui nous sauront utiles pour le développement des modèles de prévision. Par la suite, les techniques et les capteurs utilisés pour la modélisation et la mesure du rayonnement solaire ont été décrits. En plus, nous avons pu identifier l'ensemble des notions pertinentes permettant d'appréhender le rayonnement solaire.

La prévision du rayonnement solaire et de la production PV est un domaine de R&D mature, avec également une évolution énorme de l'état de l'art en matière de prévision des énergies renouvelables. Le choix d'une méthode de prévision est une tâche délicate, il a fallu dégager des connaissances précises sur les diverses méthodes de prévision existantes. Nous avons présenté un aperçu de l'état de l'art sur les méthodes de prévision du GHI et de la production PV. Cet état de l'art nous a permis de classer les méthodes de prévision en fonction de l'horizon temporel et de la résolution spatiale. Diverses méthodes ont été employées pour générer des prévisions solaires, allant des modèles statistiques et d'intelligence artificielle exploitant des données météorologiques et photovoltaïques mesurées au sol aux observations des nuages par satellite et par imagerie du ciel, en passant par les modèles de prévision numérique du temps

et finalement des modèles hybrides combinant deux ou plusieurs approches. L'état de l'art a montré que les PNT sont un bon choix pour la prévision du GHI à long terme pour des horizons temporels allant de quelques heures à quelques semaines dans des endroits où aucune donnée n'est disponible. L'imagerie satellitaire se présente aussi comme une alternative aux réseaux pyranométriques coûteux en permettant la prévision de GHI pour des horizons allant de 3 heures à heures dans des environnements où des mesures au sol ne sont pas disponibles. Cependant, les PNT et l'imagerie satellitaire présentent des limitations à cause de leurs résolutions grossières et leurs incapacités de capturer la variabilité temporelle du rayonnement causée par les nuages. Pour des horizons de prévision à très court terme, l'imagerie de ciel semble très prometteuse. Les imageurs de ciel permettent d'acquérir l'information sur la couverture nuageuse locale, ce qui permet la modélisation des fluctuations à haute fréquence du rayonnement solaire pour des horizons allant de 3 à 30 minutes. Néanmoins, même avec les imageurs de ciel, les prévisions à très court terme ne sont pas toujours précises en raison de la dispersion rapide des nuages. Pour notre cas d'application, nous nous sommes intéressés plus à la prévision à court terme, pour cet horizon, l'état de l'art a montré que les méthodes statistiques sont les plus déployées. Pour des emplacements pauvres en données avec un historique limité, les méthodes stochastiques autorégressives telles qu'ARIMA et SARIMA sont largement utilisées. Ces méthodes se basent généralement sur la nature corrélée des observations futures avec les valeurs passées de la variable en question. La précision de ces méthodes reste limitée vu qu'elles ne considèrent pas les changements des conditions météorologiques ce qui augmente leurs marges d'erreur. Récemment, avec le succès incroyable des techniques de l'intelligence artificielle dans le domaine des prévisions solaires, ces méthodes constituent une approche alternative à la modélisation physique des données de GHI et permettent de combler les lacunes des méthodes précédentes. Le succès spectaculaire de ces techniques réside dans leurs capacités à appréhender avec une grande précision la relation entre la variable cible et les divers prédicteurs considérés. De plus, les RNA ne nécessitent pas de connaissances solides dans le domaine de la métrologie ou bien les paramètres internes du système ceci a permet une vaste implémentation de ces techniques pour plusieurs problèmes non linéaires, stochastiques, et multivariés. Contrairement aux autres techniques de prévision, les RNA sont flexibles par rapport à l'horizon de prévision et ne sont pas limités dans le temps et elles ont pu modéliser avec succès le GHI pour des horizons allant de quelques minutes à une année. L'état de l'art a aussi montré que l'hybridation de plusieurs approches de prévision pourrait améliorer la précision des prévisions en tirant profit des avantages de chaque technique. Après cette synthèse bibliographique sur les méthodes de prévision, nous avons pu orienter nos travaux vers les techniques de l'IA. Plus précisément, nous avons considéré la tâche d'utiliser le MLP et LSTM pour la prévision du GHI horaire à court terme une journée à l'avance, et un modèle hybride CNN-MLP pour la prévision à très court terme à l'horizon de 15 minutes.

Ensuite, nous avons présenté dans un premier temps l'ensemble des données disponibles pour la prévision du GHI. Ces données peuvent être d'une qualité discutable, c'est pour cela le contrôle de qualité des données est une étape importante lors de l'utilisation de ces données

pour des raisons de prévision. Dans un second temps, nous avons présenté d'abord l'état de l'art sur les procédures de contrôle de qualité des données de rayonnement solaire les plus connues, et par la suite une nouvelle procédure a été présentée. La principale contribution sur ce sujet réside dans la procédure de contrôle de qualité (QC) appliquée aux données pour détecter les erreurs systématiques et les données douteuses. Dans cette contribution, nous avons proposé une routine complète impliquant les procédures les plus connues sous forme de standard visuel pour le contrôle de qualité des composantes de rayonnement solaire. Le processus de contrôle de qualité a été appliqué aux trois composantes du rayonnement solaire (GHI, DHI et DNI) collectées aux sols au pas de 10 minutes. Les résultats de QC ont révélé que les données collectées sont d'une excellente qualité à l'exception de quelques cas où les données ont montré un comportement suspect. Malgré leurs qualités élevées, la couverture spatiale de ces données mesurées aux sols est toujours limitée à cause des couts élevés de l'investissement et la maintenance des stations métrologiques. Pour surmonter ce manque, l'estimation du rayonnement par les modèles satellitaires et les modèles numériques de prévision se présentent comme une solution alternative aux données mesurées. Par conséquent, nous avons contribué aux études précédentes en validant deux modèles d'estimation du GHI pour de nouveaux emplacements (quatre sites au Maroc), à savoir la base de données satellitaire HC3 et le modèle de ciel clair McClear. L'étude a évalué la capacité de chaque modèle à reproduire les données horaires du GHI dans des conditions de ciel clair et de ciel nuageux. Les résultats de validation ont montré que HC3 et McClear peuvent être des bonnes alternatives aux mesures pyrométriques dans la zone d'étude pour les conditions claires ainsi que pour les conditions nuageuses. Nous avons également étudié une méthode de détection des instants de ciel clair à partir des mesures de GHI tout ciel. Cette identification des instants de ciel clairs nous a servis pour valider le modèle de ciel clair McClear et la HC3 sous conditions de ciel clair.

Nous avons ensuite établi des approches de prévision du GHI pour le court et le très court terme. Nous avons axé notre travail sur trois thèmes : la sélection des entrées pertinente pour les modèles de prévision, l'optimisation de l'architecture neuronale et l'évaluation des prévisions sous divers climat et pour différentes conditions météorologiques en se basant sur plusieurs métriques de performance. De plus, il nous a paru primordial de se concentrer sur plusieurs horizons de prévision en appliquant divers modèles au lieu de se limiter à une seule méthodologie. La méthodologie développée nous a permis de prévoir le GHI pour le court terme et le très court terme. Avant toute chose, une analyse de la variabilité du GHI causé par les nuages a été effectuée sur chaque site pour l'identification des classes de journée. Cette classification a été basée sur la clarté du jour et le degré de stabilité de la nébulosité où neuf classes ont été définies. Cette classification a permis de montrer que les quatre sites partagent les mêmes caractéristiques avec la dominance des jours clairs et mixtes avec une variabilité modérée pour les quatre sites et que les conditions de ciel nuageux avec une forte variabilité se produisent très rarement. Les résultats de classification nous ont été utiles pour évaluer les modèles de prévision sous diverses conditions de ciel où quatre classes de jours ont été testées

: AI (jour clair avec une faible variabilité), AII (jour clair avec une variabilité moyenne), BII (jour mixte avec une variabilité moyenne) et CIII (jours nuageux avec une forte variabilité).

Nous avons par la suite présenté les méthodologies de la prévision. Ces méthodologies développées nous ont permis de prévoir le GHI pour le court terme et le très court terme. La première est relative à l'horizon de h+12 hormis les valeurs de nuit. Comme prédicteurs basés sur les RNA, nous avons considéré le MLP et LSTM. Une étude comparative avec SARIMA et la persistance a été menée pour pouvoir illustrer l'efficacité des approches proposées.

Plusieurs manipulations ont été effectuées pour mettre en place des modèles de prévisions précis. Nous avons tout d'abord commencé par le prétraitement des données pour éliminer toutes les données aberrantes. Par la suite, nous avons sélectionné les variables d'entrée les plus pertinentes qui impactent le comportement réseau de neurones et la prévision du GHI. Puis, plusieurs tests ont été effectués pour le choix de la meilleure architecture du réseau neuronal, en termes de combinaison d'entrée, le nombre de neurones dans la couche cachée et les fonctions de transfert. Après plusieurs tests, trois modèles ont été retenus pour la prévision à court terme : le MLP M1 et LSTM ; tous les deux utilisent les mêmes entrées (température, humidité relative, vitesse et direction du vent, angle solaire, etc.), et le MLP M6 qui utilise seulement des calculs de la géométrie solaire (TOA, SZA et DL). Nous avons pu montrer qu'un modèle de prévision avec plusieurs paramètres en entrée n'augmente absolument pas les performances du modèle de prévision, et que le rayonnement au top de l'atmosphère et l'angle solaire zénithal améliorent les résultats des modèles d'une manière significative. Les modèles de prévisions ont été évalués sur une année, sous diverses saisons et condition de ciel. Les performances prédictives des modèles développés ont été comparées à SARIMA et au modèle de la persistance. Les résultats d'évaluation des modèles de prévision ont montré que suivant le type de jour, les résultats peuvent totalement diverger. Les meilleures prévisions ont été obtenues sous des conditions de ciel clair et l'erreur augmente pour les jours nuageux avec une forte variabilité. Le MLP M1 et LTSM ont prouvé leurs performances prédictives surtout pour les jours clairs et mixtes. Pour les jours clairs et pour l'ensemble des sites, le MLP M1 présente un rRMSE 2.83% et 5.39%, et LSTM un rRMSE entre et 4.07 % et 8%. Pour ces types de jours, même le modèle SARIMA a montré de bonnes compétences prédictives avec un rRMSE variant entre 5.48% et 10.16%, alors que les mauvaises prévisions sont obtenues avec la persistance avec un rRMSE entre 11.18% et 25.06%. Pour les jours nuageux, les erreurs de prévision sont plus importantes, où le rRMSE cette fois-ci varie entre 43.21% et 51.09% pour le MLP M1 et entre 59.68% et 73.97% pour le LSTM. Pour les jours nuageux, les erreurs de SARIMA et les persistances sont médiocres avec rRMSE variant entre 86.92% et 195.76% pour SARIMA et entre 79.46% et 200.25% pour la persistance. Le MLP M6 même s'il se base seulement sur des calculs de la géométrie solaire pour prévoir les GHI, il présente des résultats meilleurs que SARIMA et la persistance surtout sous les conditions de ciel clair. D'après les résultats obtenus, nous pouvons affirmer que les performances des modèles décroissent avec l'augmentation des fluctuations du GHI causées par les nuages. Cela est expliqué par le fait que les modèles n'arrivent pas à bien capturer les variations rapides du GHI même s'ils ont montré de bonnes capacités prédictives pour les autres jours clairs. L'évaluation saisonnière

des modèles de prévision a aussi prouvé la pertinence des modèles MLP M1 et LSTM. Les meilleures prévisions ont été obtenues durant la saison hivernale et l'erreur a augmenté d'une manière spectaculaire durant la saison estivale. Ce résultat nous a été surprenant surtout que dans la majorité des études reportées dans l'état de l'art ont obtenu le contraire. Cette intention nous a poussés à approfondir notre analyse pour pouvoir interpréter les résultats. Nous avons pu constater après une évaluation des sites d'études au regard de la couverture nuageuse et le taux des aérosols qu'en été, les conditions nuageuses sont rares au Maroc, alors que les aérosols présentent des pourcentages élevés par rapport à l'hiver et ces aérosols jouent un rôle crucial dans l'atténuation du rayonnement solaire en absence des nuages. En revanche, la période d'hiver dans le site étudié est caractérisée par une faible occurrence des nuages et un faible taux d'aérosols. Ce qui explique les résultats obtenus.

La dernière contribution de cette thèse concerne la prévision à très court terme où nous avons considéré l'horizon de 15 minutes. Le fait d'évaluer cet horizon de prévision modifie un peu la démarche suivie. En effet, cet horizon de prévision est difficile à appréhender et l'amélioration de la précision des prévisions sur 15 minutes doit envisager plusieurs paramètres pour espérer d'obtenir des prévisions précises. Dans cette optique, nous avons mis en œuvre un modèle hybride CNN-MLP pour développer le modèle de prévision. Ce dernier utilise dans un premier temps le modèle CNN pour extraire l'information sur la couverture nuageuse et les conditions météorologiques locales à partir des images d'une caméra hémisphérique installée au sol. Par la suite, les caractéristiques extraites sont concaténées avec les observations mesurées au niveau du site de GEP à une résolution d'une minute de la température, l'humidité relative, la vitesse et la direction du vent. Le GHI est injecté à un réseau MLP pour prévoir le GHI à l'horizon souhaité. Les résultats de validation des prévisions pour trois types de jours (clairs, mixtes et nuageux) ont montré que ce modèle améliore d'une manière remarquable la qualité des prévisions par rapport au modèle de référence à savoir la persistance.

À l'issue de tous les tests menés, il ressort qu'il est très difficile de développer un modèle de prévisions généralisable et portable d'un site à l'autre vu que les précisions de prévision sont largement dépendantes aux conditions météorologiques locales. Cependant, selon le type de jours et les conditions atmosphériques, un basculement d'une situation à l'autre pourrait pénaliser largement la précision du modèle. De ce fait, il est nécessaire d'être attentif aux méthodes appliquées et aux résultats attendus. Pour conclure, nous pourrions affirmer que les méthodes proposées dans cette thèse peuvent être utilisées par les ingénieurs des centrales photovoltaïques à des fins pratiques pour piloter des unités de stockage d'énergie, de gérer efficacement la consommation d'électricité à partir de panneaux solaires hors réseau, d'optimiser la production d'énergie solaire et d'effectuer un suivi intelligent et d'aider à la prise de décision pour la gestion de l'énergie produite ou adoptées par les chercheurs dans de futures études liées à l'établissement de prédictions et de prévisions précises du rayonnement global ou de la production photovoltaïque.

Perspectives

Les résultats de cette thèse ouvrent plusieurs pistes à envisager dans les futurs travaux :

- Appliquer et étendre les méthodes que nous avons proposées pour prévoir la production PV. Cela permettra d'évaluer la robustesse de nos méthodes pour différents types de technologies PV.
- Tests des modèles de prévision entrainés pour la prévision du GHI dans d'autres emplacements au Monde, cela permettra d'améliorer la précision des modèles et augmenter leurs portabilités surtout dans des endroits qui manquent des données de haute qualité nécessaires pour le développement des modèles de prévision.
- L'implémentation de l'information sur la couverture nuageuse et le masque solaire dans les modèles de prévision en faisant appel aux outils de Système d'Information Géographique (SIG).
- Présentation des méthodes de correction de biais pour les modèles d'estimation du rayonnement solaire pour pouvoir les utiliser comme des alternatives aux mesures au sol.
- Implémentation des modèles de prévision dans une architecture de Digital Twin pour optimiser le flux d'énergie des centrales solaires.
- Pour le climat marocain, l'intégration de l'information sur l'AOD dans les modèles de prévisions pourrait améliorer les performances prédictives d'une manière considérable comme le cas pour la couverture nuageuse pour d'autres endroits au monde.
- Les aérosols sont également responsables de la dégradation de la performance des panneaux PV et de la diminution de la réflectance des miroirs CSP qui affectent directement la production d'énergie. Il est intéressant de pouvoir modéliser l'impact des aérosols sur la prédiction de la production solaire.

Publications dans des journaux indexés et chapitre de livre

El Alani, O, Hajjaj, C, Ghennioui, H, Ghennioui, A, Philippe, B, Yves-Marie, & Monady, M (2022).
Performance assessment of SARIMA, MLP and LSTM models for short-term solar irradiance prediction under different climates in Morocco. *International Journal of Ambient Energy, Taylor & Francis.* doi.org/10.1080/01430750.2022.2127889.

Abraim, M., Salihi, M., **El Alani, O**., Hanrieder, N., Ghennioui, H., Ghennioui, A., Ghennioui, H., El Ydrissi, M., Ghennioui, H., Azouzoute, A. (2022).
Techno-economic assessment of soiling losses in CSP and PV solar power plants: A case study for the semi-arid climate of Morocco, *Energy Conversion and Management*. Elsevier. doi.org/10.1016/j.enconman.2022.116285

El Alani, O., Abraim, M., Ghennioui, H., Ghennioui, A., & A, Azouzoute. (2022).
Evaluation of Aerosols Impact on global and direct irradiance attenuation under clear sky condition: A case study in Morocco (Accepted and under production). *Journal of Sustainable Development of Energy, Water and Environment System*s. doi.org/10.13044/j.sdewes.d10.0433

Abraim, M., El Gallassi, H., **El Alani, O.**, Ghennioui, H., Ghennioui, A., Hanrieder, N., & Wilbert. S (2022).
Comparative Study of Soiling Effect on CSP and PV Technologies under Semi-arid Climate in Morocco, *Solar Energy Advances.* Elsevier. doi.org/10.1016/j.seja.2022.100021

Abdelmoula, I, Elhamaoui, S, El Alani, O, Ghennioui, A, & El Arouddi, M. (2022).
A Photovoltaic Power Prediction Approach Enhanced by Feature Engineering and Stacked Machine Learning Model. *Energy Reports*. Elsevier. doi.org/10.1016/j.egyr.2022.07.082

El Alani, O., Ghennioui, H., Abraim, M., Ghennioui, A., Blanc, P., Saint-Drenan, Y. M., & Naimi, Z. (2021).
Solar Energy Resource Assessment Using GHI and DNI Satellite Data for Moroccan Climate. *Lecture Notes on Data Engineering and Communications Technologies, Springer*. doi.org/10.1007/978-3-030-94188-8_26

El Alani, O., Abraim, M., Ghennioui, H., Ghennioui, A., Ikenbi, I., & Dahr, F. E. (2021).
Short term solar irradiance forecasting using sky images based on a hybrid CNN–MLP model. *Energy Reports*. Elsevier. doi.org/10.1016/j.egyr.2021.07.053

El Alani, O., Ghennioui, H., Ghennioui, A., Saint-Drenan, Y. M., Blanc, P., Hanrieder, N., & Dahr, F. E. (2021).
A visual support of standard procedures for solar radiation quality control. *International Journal of Renewable Energy Development*. doi.org/10.14710/ijred.2021.34806

El Alani, O., Ghennioui, H., & Ghennioui, A. (2020).
Intra-day variability quantification from ground-based measurements of global solar irradiance. *International Journal of Renewable Energy Research*. doi.org/10.20508/ijrer.v10i4.11308.g8042

El Alani, O., Ghennioui, H., Ghennioui, A., & Dahr, F. E. (2020).
Detection of clear sky instants from high frequencies pyranometric measurements of global horizontal irradiance. *E3S Web of Conferences.* doi.org/10.1051/e3sconf/202122901008

Dahr, F. E., **El Alani, O**., Bah, A., & Ghennioui, A. (2020).
Simple method for the estimation of solar spectral radiation. *E3S Web of Conferences.* doi.org/10.1051/e3sconf/202122901007

El Alani, O. E., Ghennioui, A., Ghennioui, H., Saint-Drenan, Y. M., & Blanc, P. (2020).
Evaluation of 24-Hours forecasts of global solar irradiation from IFS, GFS and McClear models. *AIP Conference Preceeding.* doi.org/10.1063/5.0032744

El Alani, O., Ghennioui, H., & Ghennioui, A. (2019).
Short term solar irradiance forecasting using artificial neural network for a semi-arid climate in Morocco. *IEE Explore.* doi.org/10.1109/WINCOM47513.2019.8942412

El Alani, O. E., Ghennioui, A., Merrouni, A. A., Ghennioui, H., Saint-Drenan, Y. M., & Blanc, P. (2018).
Validation of surface solar irradiances estimates and forecast under clear-sky conditions from the CAMS McClear model in Benguerir, Morocco. *AIP Conference Proceedings*. doi.org/10.1063/1.5117702

Conférences

El Alani, O. Abrain, M, Ghennioui, & H., Ghennioui. (2022).

Comparison of two models for PV power output prediction using ground measurements. *Rencontre Nationale des Jeunes Chercheurs en Sciences de l'Ingénieur (RNJC-SI'22)*, *18 -19 Mai 2022, Fès, Maroc*.

Alani, O. E., Ghennioui, H., Abraim, M., Ghennioui, A., Blanc, P., Saint-Drenan, Y. M., & Naimi, Z. (2021).

Solar Energy Resource Assessment Using GHI and DNI Satellite Data for Moroccan Climate. *The 3th International Conference on Advanced Technologies for Humanity* (*ICATH'21), 26 - 27 November 2021, Rabat, Maroc*. (*Driss Aboutajdine Best Paper award*).

El Alani, O., Abraim, M., Ghennioui, H., Ghennioui, A., Ikenbi, I., & Dahr, F. E. (2021).
Short term solar irradiance forecasting using sky images based on a hybrid CNN–MLP model. *TMREES Conference Series: Technologies and Materials for Renewable Energy, Environment and Sustainability* (*TMREES 20), 28 -30 Mai 2021, Athens, Greece. (Best paper).*

El Alani, O., Ghennioui, H., Ghennioui, A., & Dahr, F. E. (2020).
Detection of clear sky instants from high frequencies pyranometric measurements of global horizontal irradiance. *International Conference of Computer Science and Renewable Energies (ICCSRE'20), 22 – 24 December 2020, Agadir, Maroc.*

Dahr, F. E., **El Alani, O**., Bah, A., & Ghennioui, A. (2020).
Simple method for the estimation of solar spectral radiation. *International Conference of Computer Science and Renewable Energies (ICCSRE'20), 22 – 24 Décembre 2020, Agadir, Maroc.*

El Alani, O., Ghennioui, A., Ghennioui, H., Saint-Drenan, Y. M., & Blanc, P. (2020).
Evaluation of 24-Hours forecasts of global solar irradiation from IFS, GFS and McClear models. *TMREES Conference Series: Technologies and Materials for Renewable Energy, Environment and Sustainability* (*TMREES 20), 25-27 Juin 2020, Athens, Greece (Best paper).*

El Alani, O., Ghennioui, H., & Ghennioui, A. (2019).
Short term solar irradiance forecasting using artificial neural network for a semi-arid climate in Morocco. *The 7th International Conference on Wireless Networks and Mobile Communications* (*WINCOM'19), 29 Octobre -1 Novembre 2019, Fès, Maroc*.

Alani, O. E., Ghennioui, A., Merrouni, A. A., Ghennioui, H., Saint-Drenan, Y. M., & Blanc, P. (2018).
Validation of surface solar irradiances estimates and forecast under clear-sky conditions from the CAMS McClear model in Benguerir, Morocco. The international conference on Concentrating Solar Power and Chemical Energy Systems (*SolarPACES'18), 2-5 Octobre 2018, Casablanca, Maroc*.

El Alani. O, Ghennioui, H, Ghennioui, A. (2018)
Forecasting of Global Horizontal Irradiance one Day-ahead under Clear-Sky conditions. *Rencontre Nationale des Jeunes Chercheurs en Sciences de l'Ingénieur (RNJC-SI'18)*, 04-05 Décembre 2018, *Fès, Maroc*.
.

Références

[1] Transformation GE. A Roadmap to 2050. International Renewable Energy Agency, Abu Dhabi 2018.

[2] Choukri K, Naddami A, Hayani S. Renewable energy in emergent countries: lessons from energy transition in Morocco. Energy, Sustainability and Society 2017;7:1–11.

[3] Badescu V. Modeling Solar Radiation at the Earth's Surface. vol. 53. 2008. https://doi.org/10.1007/978-3-540-77455-6.

[4] Wan C, Zhao J, Song Y, Xu Z, Lin J, Hu Z. Photovoltaic and solar power forecasting for smart grid energy management. CSEE Journal of Power and Energy Systems 2015;1:38–46. https://doi.org/10.17775/CSEEJPES.2015.00046.

[5] Das UK, Tey KS, Seyedmahmoudian M, Mekhilef S, Idris MYI, Van Deventer W, et al. Forecasting of photovoltaic power generation and model optimization: A review. Renewable and Sustainable Energy Reviews 2018;81:912–28.

[6] World Energy Outlook 2019 – Analysis. IEA n.d. https://www.iea.org/reports/world-energy-outlook-2019 (accessed June 1, 2021).

[7] Morocco Renewable Energy Target 2030 – Policies. IEA n.d. https://www.iea.org/policies/6557-morocco-renewable-energy-target-2030 (accessed May 27, 2021).

[8] Global Energy Review 2021 – Analysis. IEA n.d. https://www.iea.org/reports/global-energy-review-2021 (accessed June 3, 2021).

[9] Global Energy Review: CO2 Emissions in 2020 – Analysis. IEA n.d. https://www.iea.org/articles/global-energy-review-co2-emissions-in-2020 (accessed October 13, 2021).

[10] Global energy transformation: A roadmap to 2050 (2019 edition). /Publications/2019/Apr/Global-Energy-Transformation-A-Roadmap-to-2050-2019Edition n.d. /publications/2019/Apr/Global-energy-transformation-A-roadmap-to-2050-2019Edition (accessed November 19, 2020).

[11] Kousksou T, Allouhi A, Belattar M, Jamil A, El Rhafiki T, Zeraouli Y. Morocco's strategy for energy security and low-carbon growth. Energy 2015;84:98–105.

[12] Moroccan Energy Outlook: Achievements and Opportunities. Africa Energy Forum 2019. https://www.africa-energy-forum.com/article/moroccan-energy-outlook-achievements-and-opportunities (accessed June 1, 2021).

[13] ÉNERGIES RENOUVELABLES n.d. https://www.mem.gov.ma/Pages/secteur.aspx?e=2 (accessed July 4, 2022).

[14] Kousksou T, Allouhi A, Belattar M, Jamil A, El Rhafiki T, Arid A, et al. Renewable energy potential and national policy directions for sustainable development in Morocco. Renewable and Sustainable Energy Reviews 2015;47:46–57.

[15] Plaquette institution AMEE-Fr.pdf n.d.

[16] Choukri K, Naddami A, Hayani S. Renewable energy in emergent countries: lessons from energy transition in Morocco. Energy, Sustainability and Society 2017;7:1–11.

[17] El Iysaouy L, El Idrissi N, Tvaronavičienė M, Lahbabi M, Oumnad A. Towards energy efficiency: Case of Morocco. Insights into Regional Development 2019;1:259–71.

[18] Stratégue Nationale de l'Efficacité énergétique à l'horizon 2030.pdf n.d.

[19] Parema M. Les énergies Renouvelables et L'efficacité énergétique au Maroc. Int Zusammenarbeit Gmbh En 2016.

[20] IRESEN – Institut de Recherche en Energie Solaire et Energies Nouvelles n.d. https://www.iresen.org/?lang=en (accessed June 7, 2021).

[21] Tiwari G, Tiwari A. Handbook of solar energy. vol. 498. Springer; 2016.

[22] Wald L. Fundamentals of Solar Radiation. CRC Press; 2021.

[23] Sen Z. Solar energy fundamentals and modeling techniques: atmosphere, environment, climate change and renewable energy. Springer Science & Business Media; 2008.

[24] Reviewed P, Merced ME. University of California Solar Irradiance Forecasting at Multiple Time Horizons and Novel Methods To Evaluate Uncertainty A dissertation submitted in partial satisfaction 2012;6.

[25] Eissa Y, Korany M, Aoun Y, Boraiy M, Abdel Wahab MM, Alfaro SC, et al. Validation of the Surface Downwelling Solar Irradiance Estimates of the HelioClim-3 Database in Egypt. Remote Sensing 2015;7:9269–91. https://doi.org/10.3390/rs70709269.

[26] Polo J, Fernández-Peruchena C, Salamalikis V, Mazorra-Aguiar L, Turpin M, Martín-Pomares L, et al. Benchmarking on improvement and site-adaptation techniques for modeled solar radiation datasets. Solar Energy 2020;201:469–79.

[27] Polo J, Wilbert S, Ruiz-Arias JA, Meyer R, Gueymard C, Suri M, et al. Preliminary survey on site-adaptation techniques for satellite-derived and reanalysis solar radiation datasets. Solar Energy 2016;132:25–37.

[28] Bechini L, Ducco G, Donatelli M, Stein A. Modelling, interpolation and stochastic simulation in space and time of global solar radiation. Agriculture, Ecosystems & Environment 2000;81:29–42.

[29] Ertekin C, Evrendilek F. Spatio-temporal modeling of global solar radiation dynamics as a function of sunshine duration for Turkey. Agricultural and Forest Meteorology 2007;145:36–47.

[30] Hay JE. An assessment of the mesoscale variability of solar radiation at the earth's surface. Solar Energy 1984;32:425–34.

[31] Perez R, Seals R, Zelenka A. Comparing satellite remote sensing and ground network measurements for the production of site/time specific irradiance data. Solar Energy 1997;60:89–96.

[32] Myers DR. Solar radiation: practical modeling for renewable energy applications. CRC press; 2017.

[33] Guide to Instruments and Methods of Observation (WMO-No. 8) | World Meteorological Organization n.d. https://community.wmo.int/activity-areas/imop/wmo-no_8 (accessed March 18, 2022).

[34] Solar Radiation Measurement - Kipp & Zonen n.d. https://www.kippzonen.com/ (accessed October 12, 2021).

[35] Apogee Instruments | Official Site n.d. https://www.apogeeinstruments.com/ (accessed October 12, 2021).

[36] Automated Weather Stations - Research-grade stations for reliable monitoring n.d. http://www.campbellsci.com/automated-weather-stations (accessed October 12, 2021).

[37] Weather Stations and Sensors. Vaisala n.d. https://www.vaisala.com/en/products/instruments-sensors-and-other-measurement-devices/weather-stations-and-sensors (accessed October 12, 2021).

[38] Foukal P v, Lean J. An empirical model of total solar irradiance variation between 1874 and 1988. Science 1990;247:556–8.

[39] Tobiska WK, Woods T, Eparvier F, Viereck R, Floyd L, Bouwer D, et al. The SOLAR2000 empirical solar irradiance model and forecast tool. Journal of Atmospheric and Solar-Terrestrial Physics 2000;62:1233–50.

[40] Krivova NA, Solanki SK. Models of solar irradiance variations: Current status. Journal of Astrophysics and Astronomy 2008;29:151–8.

[41] Kasten F, Czeplak G. Solar and terrestrial radiation dependent on the amount and type of cloud. Solar Energy 1980;24:177–89. https://doi.org/10.1016/0038-092X(80)90391-6.

[42] Qingyuan Z, Huang J, Siwei L. Development of typical year weather data for Chinese locations. ASHRAE Transactions 2002;108:1063–75.

[43] MESoR - Management and Exploitation of Solar Resource Knowledge n.d. http://www.mesor.org/ (accessed October 12, 2021).

[44] JRC Photovoltaic Geographical Information System (PVGIS) - European Commission n.d. https://re.jrc.ec.europa.eu/pvg_tools/fr/#MR (accessed October 12, 2021).

[45] HOME - www.soda-pro.com n.d. http://www.soda-pro.com/ (accessed August 30, 2021).

[46] s.r.o © Solargis. Solargis :: iMaps n.d. https://solargis.info/imaps/ (accessed August 5, 2021).

[47] Antonanzas-Torres F, Urraca R, Polo J, Perpiñán-Lamigueiro O, Escobar R. Clear sky solar irradiance models: A review of seventy models. Renewable and Sustainable Energy Reviews 2019;107:374–87.

[48] Threlkeld J, Jordan R. Direct solar radiation available on clear days. Heat, Piping Air Cond 1957;29.

[49] Biga AJ, Rosa R. Contribution to the study of the solar radiation climate of Lisbon. Solar Energy 1979;23:61–7. https://doi.org/10.1016/0038-092X(79)90044-6.

[50] Daneshyar M. Solar radiation statistics for Iran. Sol Energy;(United States) 1978;21.

[51] Schulze RE. A physically based method of estimating solar radiation from suncards. Agricultural Meteorology 1976;16:85–101. https://doi.org/10.1016/0002-1571(76)90070-4.

[52] Sharma MR, Pal RS. Interrelationships between total, direct, and diffuse solar radiation in the tropics. Solar Energy 1965;9:183–92. https://doi.org/10.1016/0038-092X(65)90045-9.

[53] Yang D. Choice of clear-sky model in solar forecasting. Journal of Renewable and Sustainable Energy 2020;12:026101.

[54] Demain C, Journée M, Bertrand C. Sensitivity of the RMI's MAGIC/Heliosat-2 method to relevant input data. Advances in Science and Research 2013;10:7–13.

[55] Lefèvre M, Oumbe A, Blanc P, Espinar B, Gschwind B, Qu Z, et al. McClear: a new model estimating downwelling solar radiation at ground level in clear-sky conditions 2013. https://doi.org/10.5194/amt-6-2403-2013.

[56] Ineichen P. A broadband simplified version of the Solis clear sky model. Solar Energy 2008;82:758–62.

[57] Alia-Martinez M, Antonanzas J, Urraca R, Martinez-De-Pison FJ, Antonanzas-Torres F. Benchmark of algorithms for solar clear-sky detection. Journal of Renewable and Sustainable Energy 2016;8. https://doi.org/10.1063/1.4950948.

[58] Antonanzas-Torres F, Urraca R, Polo J, Perpiñán-Lamigueiro O, Escobar R. Clear sky solar irradiance models: A review of seventy models. Renewable and Sustainable Energy Reviews 2019;107:374–87. https://doi.org/10.1016/j.rser.2019.02.032.

[59] Perez R, Kivalov S, Schlemmer J, Hemker Jr K, Hoff TE. Short-term irradiance variability: Preliminary estimation of station pair correlation as a function of distance. Solar Energy 2012;86:2170–6.

[60] Stein J, Hansen C, Reno MJ. The variability index: A new and novel metric for quantifying irradiance and PV output variability. Sandia National Laboratories; 2012.

[61] Kang BO, Tam K-S. A new characterization and classification method for daily sky conditions based on ground-based solar irradiance measurement data. Solar Energy 2013;94:102–18.

[62] Lave M, Reno MJ, Broderick RJ. Characterizing local high-frequency solar variability and its impact to distribution studies. Solar Energy 2015;118:327–37.

[63] Lauret P, Perez R, Aguiar LM, Tapachès E, Diagne HM, David M. Characterization of the intraday variability regime of solar irradiation of climatically distinct locations. Solar Energy 2016;125:99–110.

[64] Blaga R, Paulescu M. Quantifiers for the solar irradiance variability: A new perspective. Solar Energy 2018;174:606–16.

[65] Lohmann GM. Irradiance variability quantification and small-scale averaging in space and time: A short review. Atmosphere 2018;9:264.

[66] Pelland S, Remund J, Kleissl J, Oozeki T, De Brabandere K. Photovoltaic and solar forecasting: state of the art. IEA PVPS, Task 2013;14:1–36.

[67] Beyer HG, Polo Martinez J, Suri M, Torres JL, Lorenz E, Müller SC, et al. D 1.1. 3 "Report on Benchmarking of Radiation Products." Report under Contract 2009.

[68] Lorenz E, Hurka J, Heinemann D, Beyer HG. Irradiance forecasting for the power prediction of grid-connected photovoltaic systems. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 2009;2:2–10.

[69] Müller SC, Remund J. Advances in radiation forecast based on regional weather models MM5 and WRF, 2010, p. 6–10.

[70] Shi J, Lee W-J, Liu Y, Yang Y, Wang P. Forecasting power output of photovoltaic systems based on weather classification and support vector machines. IEEE Transactions on Industry Applications 2012;48:1064–9.

[71] Yang H-T, Huang C-M, Huang Y-C, Pai Y-S. A weather-based hybrid method for 1-day ahead hourly forecasting of PV power output. IEEE Transactions on Sustainable Energy 2014;5:917–26.

[72] Pani AK, Nayak N. Forecasting solar irradiance with weather classification and chaotic gravitational search algorithm based wavelet kernel extreme learning machine. International Journal of Renewable Energy Research (IJRER) 2019;9:1650–9.

[73] Zafar R, Vu BH, Husein M, Chung I-Y. Day-Ahead Solar Irradiance Forecasting Using Hybrid Recurrent Neural Network with Weather Classification for Power System Scheduling. Applied Sciences 2021;11:6738.

[74] Raza MQ, Khosravi A. A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings. Renewable and Sustainable Energy Reviews 2015;50:1352–72.

[75] Rahman S, Hazim O. A generalized knowledge-based short-term load-forecasting technique. IEEE Transactions on Power Systems 1993;8:508–14.

[76] Raza MQ, Nadarajah M, Ekanayake C. On recent advances in PV output power forecast. Solar Energy 2016;136:125–44.

[77] Nespoli A, Ogliari E, Leva S, Massi Pavan A, Mellit A, Lughi V, et al. Day-ahead photovoltaic forecasting: A comparison of the most effective techniques. Energies 2019;12:1621.

[78] Sobri S, Koohi-Kamali S, Rahim NAbd. Solar photovoltaic generation forecasting methods: A review. Energy Conversion and Management 2018;156:459–97. https://doi.org/10.1016/j.enconman.2017.11.019.

[79] Das UK, Tey KS, Seyedmahmoudian M, Mekhilef S, Idris MYI, Van Deventer W, et al. Forecasting of photovoltaic power generation and model optimization: A review. Renewable and Sustainable Energy Reviews 2018;81:912–28. https://doi.org/10.1016/j.rser.2017.08.017.

[80] Zhang J, Verschae R, Nobuhara S, Lalonde J-F. Deep photovoltaic nowcasting. Solar Energy 2018;176:267–76.

[81] Antonanzas J, Osorio N, Escobar R, Urraca R, Martinez-de-Pison FJ, Antonanzas-Torres F. Review of photovoltaic power forecasting. Solar Energy 2016;136:78–111. https://doi.org/10.1016/j.solener.2016.06.069.

[82] Pedro HT, Coimbra CF. Assessment of forecasting techniques for solar power production with no exogenous inputs. Solar Energy 2012;86:2017–28.

[83] Zhang J, Hodge B-M, Florita A, Lu S, Hamann HF, Banunarayanan V. Metrics for evaluating the accuracy of solar power forecasting. National Renewable Energy Lab.(NREL), Golden, CO (United States); 2013.

[84] Marquez R, Coimbra CFM. Proposed Metric for Evaluation of Solar Forecasting Models. J Sol Energy Eng 2012;135:011016-011016–9. https://doi.org/10.1115/1.4007496.

[85] Vallance L, Charbonnier B, Paul N, Dubost S, Blanc P. Towards a standardized procedure to assess solar forecast accuracy: A new ramp and time alignment metric. Solar Energy 2017;150:408–22. https://doi.org/10.1016/j.solener.2017.04.064.

[86] Marquez R, Coimbra CF. Proposed metric for evaluation of solar forecasting models. Journal of Solar Energy Engineering 2013;135.

[87] Diagne M, David M, Lauret P, Boland J, Schmutz N. Review of solar irradiance forecasting methods and a proposition for small-scale insular grids. Renewable and Sustainable Energy Reviews 2013;27:65–76. https://doi.org/10.1016/j.rser.2013.06.042.

[88] Inman RH, Pedro HTC, Coimbra CFM. Solar forecasting methods for renewable energy integration. Progress in Energy and Combustion Science 2013;39:535–76. https://doi.org/10.1016/j.pecs.2013.06.002.

[89] Mathiesen P, Kleissl J. Evaluation of numerical weather prediction for intra-day solar forecasting in the continental United States. Solar Energy 2011;85:967–77.

[90] Setchell H. ECMWF. ECMWF 2017. https://www.ecmwf.int/en/forecasts (accessed June 24, 2020).

[91] Zamora RJ, Dutton EG, Trainer M, McKeen SA, Wilczak JM, Hou Y-T. The accuracy of solar irradiance calculations used in mesoscale numerical weather prediction. Monthly Weather Review 2005;133:783–92.

[92] Lara-Fanego V, Ruiz-Arias J, Pozo-Vázquez D, Santos-Alamillos F, Tovar-Pescador J. Evaluation of the WRF model solar irradiance forecasts in Andalusia (southern Spain). Solar Energy 2012;86:2200–17.

[93] Heinemann D, Lorenz E, Girodo M. Forecasting of solar radiation. Solar Energy Resource Management for Electricity Generation from Local Level to Global Scale Nova Science Publishers, New York 2006:83–94.

[94] Perez R, Moore K, Wilcox S, Renné D, Zelenka A. Forecasting solar radiation–Preliminary evaluation of an approach based upon the national forecast database. Solar Energy 2007;81:809–12.

[95] Remund J, Perez R, Lorenz E. Comparison of solar radiation forecasts for the USA. vol. 14, 2008.

[96] Breitkreuz H, Schroedter-Homscheidt M, Holzer-Popp T, Dech S. Short-range direct and diffuse irradiance forecasts for solar energy applications based on aerosol chemical transport and numerical weather modeling. Journal of Applied Meteorology and Climatology 2009;48:1766–79.

[97] Lorenz E, Hurka J, Karampela G, Heinemann D, Beyer HG, Schneider M. Qualified forecast of ensemble power production by spatially dispersed grid-connected PV systems. Measurement 2007:1–7.

[98] Lorenz E, Remund J, Müller SC, Traunmüller W, Steinmaurer G, Pozo D, et al. Benchmarking of different approaches to forecast solar irradiance, Hamburg Germany; 2009, p. 21–5.

[99] Mathiesen P, Brown JM, Kleissl J. Geostrophic wind dependent probabilistic irradiance forecasts for coastal California. IEEE Transactions on Sustainable Energy 2012;4:510–8.

[100] Cornaro C, Pierro M, Bucci F. Master optimization process based on neural networks ensemble for 24-h solar irradiance forecast. Solar Energy 2015;111:297–312.

[101] Huang G, Li Z, Li X, Liang S, Yang K, Wang D, et al. Estimating surface solar irradiance from satellites: Past, present, and future perspectives. Remote Sensing of Environment 2019;233:111371.

[102] Kleissl J. Solar energy forecasting and resource assessment. Academic Press; 2013.

[103] Zhang X, Liang S, Wang G, Yao Y, Jiang B, Cheng J. Evaluation of the reanalysis surface incident shortwave radiation products from NCEP, ECMWF, GSFC, and JMA using satellite and surface observations. Remote Sensing 2016;8:225.

[104] Vindel JM, Navarro AA, Valenzuela RX, Ramírez L. Temporal scaling analysis of irradiance estimated from daily satellite data and numerical modelling. Atmospheric Research 2016;181:154–62.

[105] Pinker R, Frouin R, Li Z. A review of satellite methods to derive surface shortwave irradiance. Remote Sensing of Environment 1995;51:108–24.

[106] Li Z, Whitlock CH, Charlock TP. Assessment of the global monthly mean surface insolation estimated from satellite measurements using global energy balance archive data. Journal of Climate 1995;8:315–28.

[107] Müller R, Pfeifroth U, Träger-Chatterjee C, Trentmann J, Cremer R. Digging the METEOSAT treasure—3 decades of solar surface radiation. Remote Sensing 2015;7:8067–101.

[108] Qin J, Chen Z, Yang K, Liang S, Tang W. Estimation of monthly-mean daily global solar radiation based on MODIS and TRMM products. Applied Energy 2011;88:2480–9.

[109] Qu Z, Oumbe A, Blanc P, Espinar B, Gesell G, GSCHwIND Beno, et al. Fast radiative transfer parameterisation for assessing the surface solar irradiance: The Heliosat-4 method. Meteorologische Zeitschrift 2017;26:33–57.

[110] Moradi I, Mueller R, Alijani B, Kamali GA. Evaluation of the Heliosat-II method using daily irradiation data for four stations in Iran. Solar Energy 2009;83:150–6.

[111] Dürr B, Zelenka A. Deriving surface global irradiance over the Alpine region from METEOSAT Second Generation data by supplementing the HELIOSAT method. International Journal of Remote Sensing 2009;30:5821–41.

[112] Liu Q, van Delst P, Chen Y, Groff D, Han Y, Collard A, et al. Community radiative transfer model for radiance assimilation and applications, IEEE; 2012, p. 3700–3.

[113] Polo J, Wilbert S, Ruiz-Arias JA, Meyer R, Gueymard C, Suri M, et al. Preliminary survey on site-adaptation techniques for satellite-derived and reanalysis solar radiation datasets. Solar Energy 2016;132:25–37.

[114] Yang D, Kleissl J, Gueymard CA, Pedro HTC, Coimbra CFM. History and trends in solar irradiance and PV power forecasting: A preliminary assessment and review using text mining. Solar Energy 2018;168:60–101. https://doi.org/10.1016/j.solener.2017.11.023.

[115] Espinar B, Blanc P, Wald L, Gschwind B, Ménard L, Wey E, et al. HelioClim-3: a near-real time and long-term surface solar irradiance database, 2012.

[116] Marchand M, Ghennioui A, Wey E, Wald L. Comparison of several satellite-derived databases of surface solar radiation against ground measurement in Morocco. Advances in Science and Research, vol. 15, Copernicus GmbH; 2018, p. 21–9. https://doi.org/10.5194/asr-15-21-2018.

[117] Thomas C, Wey E, Blanc P, Wald L. Validation of three satellite-derived databases of surface solar radiation using measurements performed at 42 stations in Brazil. Advances in Science and Research 2016;13:81–6.

[118] Marchand M, Saint-Drenan Y-M, Saboret L, Wey E, Wald L. Performance of CAMS Radiation Service and HelioClim-3 databases of solar radiation at surface: evaluating the spatial variation in Germany. Advances in Science and Research 2020;17:143–52.
[119] Marchand M, Lefèvre M, Saboret L, Wey E, Wald L. Verifying the spatial consistency of the CAMS Radiation Service and HelioClim-3 satellite-derived databases of solar radiation using a dense network of measuring stations: the case of The Netherlands. Advances in Science and Research 2019;16:103–11.
[120] Rigollier C, Lefèvre M, Wald L. The method Heliosat-2 for deriving shortwave solar radiation from satellite images. Solar Energy 2004;77:159–69.
[121] Şenkal O, Kuleli T. Estimation of solar radiation over Turkey using artificial neural network and satellite data. Applied Energy 2009;86:1222–8.
[122] Qin J, Chen Z, Yang K, Liang S, Tang W. Estimation of monthly-mean daily global solar radiation based on MODIS and TRMM products. Applied Energy 2011;88:2480–9.
[123] Marquez R, Pedro HT, Coimbra CF. Hybrid solar forecasting method uses satellite imaging and ground telemetry as inputs to ANNs. Solar Energy 2013;92:176–88.
[124] Aguiar LM, Pereira B, Lauret P, Díaz F, David M. Combining solar irradiance measurements, satellite-derived data and a numerical weather prediction model to improve intra-day solar forecasting. Renewable Energy 2016;97:599–610. https://doi.org/10.1016/j.renene.2016.06.018.
[125] Chow CW, Urquhart B, Lave M, Dominguez A, Kleissl J, Shields J, et al. Intra-hour forecasting with a total sky imager at the UC San Diego solar energy testbed. Solar Energy 2011;85:2881–93.
[126] Deo RC, Şahin M. Forecasting long-term global solar radiation with an ANN algorithm coupled with satellite-derived (MODIS) land surface temperature (LST) for regional locations in Queensland. Renewable and Sustainable Energy Reviews 2017;72:828–48.
[127] Tuohy A, Zack J, Haupt SE, Sharp J, Ahlstrom M, Dise S, et al. Solar forecasting: methods, challenges, and performance. IEEE Power and Energy Magazine 2015;13:50–9.
[128] Urquhart B, Ghonima M, Nguyen D, Kurtz B, Chow C. Sky imaging systems for short-term solar forecasting, chapter 9 in: Solar Energy Forecasting and Resource Assessment, edited by: Kleissl, J 2013.
[129] Fu C-L, Cheng H-Y. Predicting solar irradiance with all-sky image features via regression. Solar Energy 2013;97:537–50.
[130] Zhu H, Li X, Sun Q, Nie L, Yao J, Zhao G. A power prediction method for photovoltaic power plant based on wavelet decomposition and artificial neural networks. Energies 2016;9:11.
[131] Aladag CH. Advances in time series forecasting. vol. 2. Bentham Science Publishers; 2017.
[132] Yang D, Sharma V, Ye Z, Lim LI, Zhao L, Aryaputera AW. Forecasting of global horizontal irradiance by exponential smoothing, using decompositions. Energy 2015;81:111–9.
[133] Hyndman R, Koehler AB, Ord JK, Snyder RD. Forecasting with exponential smoothing: the state space approach. Springer Science & Business Media; 2008.
[134] Sbrana G, Silvestrini A. Random switching exponential smoothing and inventory forecasting. International Journal of Production Economics 2014;156:283–94.
[135] Dong Z, Yang D, Reindl T, Walsh WM. Short-term solar irradiance forecasting using exponential smoothing state space model. Energy 2013;55:1104–13.
[136] Box GE, Jenkins GM, Reinsel GC. Time series analysis: forecasting and control. vol. 734. John Wiley & Sons; 2011.
[137] Reikard G. Predicting solar radiation at high resolutions: A comparison of time series forecasts. Solar Energy 2009;83:342–9.
[138] Pedro HT, Coimbra CF. Assessment of forecasting techniques for solar power production with no exogenous inputs. Solar Energy 2012;86:2017–28.
[139] Agoua XG, Girard R, Kariniotakis G. Short-term spatio-temporal forecasting of photovoltaic power production. IEEE Transactions on Sustainable Energy 2017;9:538–46.
[140] Kardakos EG, Alexiadis MC, Vagropoulos SI, Simoglou CK, Biskas PN, Bakirtzis AG. Application of time series and artificial neural network models in short-term forecasting of PV power generation, IEEE; 2013, p. 1–6.

[141] Yang D, Jirutitijaroen P, Walsh WM. Hourly solar irradiance time series forecasting using cloud cover index. Solar Energy 2012;86:3531–43.
[142] Simon HA. Artificial intelligence: an empirical science. Artificial Intelligence 1995;77:95–127.
[143] Krishnamoorthy C, Rajeev S. Artificial intelligence and expert systems for engineers. vol. 11. CRC press; 1996.
[144] Abiodun OI, Jantan A, Omolara AE, Dada KV, Mohamed NA, Arshad H. State-of-the-art in artificial neural network applications: A survey. Heliyon 2018;4:e00938.
[145] Schmidhuber J. Deep learning in neural networks: An overview. Neural Networks 2015;61:85–117.
[146] Mellit A, Saglam S, Kalogirou S. Artificial neural network-based model for estimating the produced power of a photovoltaic module. Renewable Energy 2013;60:71–8. https://doi.org/10.1016/j.renene.2013.04.011.
[147] Darbellay GA, Slama M. Forecasting the short-term demand for electricity: Do neural networks stand a better chance? International Journal of Forecasting 2000;16:71–83.
[148] Tealab A. Time series forecasting using artificial neural networks methodologies: A systematic review. Future Computing and Informatics Journal 2018;3:334–40.
[149] Ashour MAH, Jamal A, Helmi RAA. Effectiveness of artificial neural networks in solving financial time series. International Journal of Engineering & Technology 2018;7:99–105.
[150] Wang D, He H, Liu D. Intelligent optimal control with critic learning for a nonlinear overhead crane system. IEEE Transactions on Industrial Informatics 2017;14:2932–40.
[151] Cabaneros SM, Calautit JK, Hughes BR. A review of artificial neural network models for ambient air pollution prediction. Environmental Modelling & Software 2019;119:285–304.
[152] Wythoff BJ. Backpropagation neural networks: a tutorial. Chemometrics and Intelligent Laboratory Systems 1993;18:115–55.
[153] Chow SK, Lee EW, Li DH. Short-term prediction of photovoltaic energy generation by intelligent approach. Energy and Buildings 2012;55:660–7.
[154] Renno C, Petito F, Gatto A. ANN model for predicting the direct normal irradiance and the global radiation for a solar application to a residential building. Journal of Cleaner Production 2016;135:1298–316.
[155] El Alani O, Ghennioui H, Ghennioui A. Short term solar irradiance forecasting using artificial neural network for a semi-arid climate in Morocco, IEEE; 2019, p. 1–7.
[156] Mellit A, Pavan AM. A 24-h forecast of solar irradiance using artificial neural network: Application for performance prediction of a grid-connected PV plant at Trieste, Italy. Solar Energy 2010;84:807–21.
[157] Notton G, Paoli C, Ivanova L, Vasileva S, Nivet ML. Neural network approach to estimate 10-min solar global irradiation values on tilted planes. Renewable Energy 2013;50:576–84. https://doi.org/10.1016/j.renene.2012.07.035.
[158] Izgi E, Öztopal A, Yerli B, Kaymak MK, Şahin AD. Short–mid-term solar power prediction by using artificial neural networks. Solar Energy 2012;86:725–33.
[159] Kaushika ND, Tomar RK, Kaushik SC. Artificial neural network model based on interrelationship of direct, diffuse and global solar radiations. Solar Energy 2014;103:327–42. https://doi.org/10.1016/j.solener.2014.02.015.
[160] Gutierrez-Corea F-V, Manso-Callejo M-A, Moreno-Regidor M-P, Manrique-Sancho M-T. Forecasting short-term solar irradiance based on artificial neural networks and data from neighboring meteorological stations. Solar Energy 2016;134:119–31.
[161] Voyant C, Randimbivololona P, Nivet ML, Paoli C, Muselli M. Twenty four hours ahead global irradiation forecasting using multi-layer perceptron. Meteorological Applications 2014;21:644–55.
[162] Vathsala M, Holi G. RNN based machine translation and transliteration for Twitter data. International Journal of Speech Technology 2020;23:499–504.
[163] Singh SP, Kumar A, Darbari H, Singh L, Rastogi A, Jain S. Machine translation using deep learning: An overview, IEEE; 2017, p. 162–7.
[164] Liu S, Yang N, Li M, Zhou M. A recursive recurrent neural network for statistical machine translation 2014.
[165] Williams W, Prasad N, Mrva D, Ash T, Robinson T. Scaling recurrent neural network language models, IEEE; 2015, p. 5391–5.

[166] Wang T, Cho K. Larger-context language modelling. ArXiv Preprint ArXiv:151103729 2015.
[167] Melis G, Dyer C, Blunsom P. On the state of the art of evaluation in neural language models. ArXiv Preprint ArXiv:170705589 2017.
[168] Kłosowski P. Deep learning for natural language processing and language modelling, IEEE; 2018, p. 223–8.
[169] Graves A, Mohamed A, Hinton G. Speech recognition with deep recurrent neural networks, Ieee; 2013, p. 6645–9.
[170] Graves A, Jaitly N. Towards end-to-end speech recognition with recurrent neural networks, PMLR; 2014, p. 1764–72.
[171] Islam J, Mubassira M, Islam MR, Das AK. A speech recognition system for Bengali language using recurrent neural network, IEEE; 2019, p. 73–6.
[172] Nallapati R, Zhai F, Zhou B. Summarunner: A recurrent neural network based sequence model for extractive summarization of documents, 2017.
[173] Ghosh R, Vamshi C, Kumar P. RNN based online handwritten word recognition in Devanagari and Bengali scripts using horizontal zoning. Pattern Recognition 2019;92:203–18.
[174] Dutta K, Krishnan P, Mathew M, Jawahar C. Improving cnn-rnn hybrid networks for handwriting recognition, IEEE; 2018, p. 80–5.
[175] Wang M, Zhang YD, Cui G. Human motion recognition exploiting radar with stacked recurrent neural network. Digital Signal Processing 2019;87:125–31.
[176] Dong Y. An application of deep neural networks to the in-flight parameter identification for detection and characterization of aircraft icing. Aerospace Science and Technology 2018;77:34–49.
[177] Qin K, Li C, Pavlu V, Aslam JA. Adapting RNN sequence prediction model to multi-label set prediction. ArXiv Preprint ArXiv:190405829 2019.
[178] Selvin S, Vinayakumar R, Gopalakrishnan E, Menon VK, Soman K. Stock price prediction using LSTM, RNN and CNN-sliding window model, IEEE; 2017, p. 1643–7.
[179] Hochreiter S, Schmidhuber J. Long short-term memory. Neural Computation 1997;9:1735–80.
[180] Soutner D, Müller L. Application of LSTM neural networks in language modelling, Springer; 2013, p. 105–12.
[181] Yao L, Guan Y. An improved LSTM structure for natural language processing, IEEE; 2018, p. 565–9.
[182] Wang S, Jiang J. Learning natural language inference with LSTM. ArXiv Preprint ArXiv:151208849 2015.
[183] Chen Q, Zhu X, Ling Z, Wei S, Jiang H, Inkpen D. Enhanced lstm for natural language inference. ArXiv Preprint ArXiv:160906038 2016.
[184] Yao L, Guan Y. An improved LSTM structure for natural language processing, IEEE; 2018, p. 565–9.
[185] Bianchi FM, Maiorino E, Kampffmeyer MC, Rizzi A, Jenssen R. Recurrent neural networks for short-term load forecasting: an overview and comparative analysis 2017.
[186] Fischer T, Krauss C. Deep learning with long short-term memory networks for financial market predictions. European Journal of Operational Research 2018;270:654–69.
[187] Srivastava S, Lessmann S. A comparative study of LSTM neural networks in forecasting day-ahead global horizontal irradiance with satellite data. Solar Energy 2018;162:232–47.
[188] Qing X, Niu Y. Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM. Energy 2018;148:461–8.
[189] Brahma B, Wadhvani R. Solar Irradiance Forecasting Based on Deep Learning Methodologies and Multi-Site Data. Symmetry 2020;12:1830.
[190] Yu Y, Cao J, Zhu J. An LSTM short-term solar irradiance forecasting under complicated weather conditions. IEEE Access 2019;7:145651–66.
[191] Huang X, Zhang C, Li Q, Tai Y, Gao B, Shi J. A comparison of hour-ahead solar irradiance forecasting models based on LSTM network. Mathematical Problems in Engineering 2020;2020.
[192] Fukushima K, Miyake S. Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition. Competition and cooperation in neural nets, Springer; 1982, p. 267–85.

[193] Khan S, Rahmani H, Shah SAA, Bennamoun M. A guide to convolutional neural networks for computer vision. Synthesis Lectures on Computer Vision 2018;8:1–207.
[194] Bhatt D, Patel C, Talsania H, Patel J, Vaghela R, Pandya S, et al. CNN Variants for Computer Vision: History, Architecture, Application, Challenges and Future Scope. Electronics 2021;10:2470.
[195] Kebria PM, Khosravi A, Salaken SM, Nahavandi S. Deep imitation learning for autonomous vehicles based on convolutional neural networks. IEEE/CAA Journal of Automatica Sinica 2019;7:82–95.
[196] Koprinska I, Wu D, Wang Z. Convolutional neural networks for energy time series forecasting, IEEE; 2018, p. 1–8.
[197] Sakib S, Ahmed N, Kabir AJ, Ahmed H. An overview of convolutional neural network: its architecture and applications 2019.
[198] Ye L, Cao Z, Xiao Y. DeepCloud: Ground-based cloud image categorization using deep convolutional features. IEEE Transactions on Geoscience and Remote Sensing 2017;55:5729–40.
[199] Sun Y, Szűcs G, Brandt AR. Solar PV output prediction from video streams using convolutional neural networks. Energy & Environmental Science 2018;11:1811–8.
[200] Siddiqui TA, Bharadwaj S, Kalyanaraman S. A deep learning approach to solar-irradiance forecasting in sky-videos, IEEE; 2019, p. 2166–74.
[201] Paletta Q, Lasenby J. Convolutional Neural Networks applied to sky images for short-term solar irradiance forecasting. ArXiv Preprint ArXiv:200511246 2020.
[202] Al-lahham A, Theeb O, Elalem K, A Alshawi T, A Alshebeili S. Sky Imager-Based Forecast of Solar Irradiance Using Machine Learning. Electronics 2020;9:1700.
[203] Diagne M, David M, Lauret P, Boland J, Schmutz N. Review of solar irradiance forecasting methods and a proposition for small-scale insular grids. Renewable and Sustainable Energy Reviews 2013;27:65–76.
[204] Lu S, Hwang Y, Khabibrakhmanov I, Marianno FJ, Shao X, Zhang J, et al. Machine learning based multi-physical-model blending for enhancing renewable energy forecast-improvement via situation dependent error correction, IEEE; 2015, p. 283–90.
[205] Fernandez-Jimenez LA, Muñoz-Jimenez A, Falces A, Mendoza-Villena M, Garcia-Garrido E, Lara-Santillan PM, et al. Short-term power forecasting system for photovoltaic plants. Renewable Energy 2012;44:311–7.
[206] Massidda L, Marrocu M. Use of Multilinear Adaptive Regression Splines and numerical weather prediction to forecast the power output of a PV plant in Borkum, Germany. Solar Energy 2017;146:141–9.
[207] Filipe JM, Bessa RJ, Sumaili J, Tomé R, Sousa JN. A hybrid short-term solar power forecasting tool, IEEE; 2015, p. 1–6.
[208] Zang H, Liu L, Sun L, Cheng L, Wei Z, Sun G. Short-term global horizontal irradiance forecasting based on a hybrid CNN-LSTM model with spatiotemporal correlations. Renewable Energy 2020;160:26–41.
[209] Bouzerdoum M, Mellit A, Pavan AM. A hybrid model (SARIMA–SVM) for short-term power forecasting of a small-scale grid-connected photovoltaic plant. Solar Energy 2013;98:226–35. https://doi.org/10.1016/j.solener.2013.10.002.
[210] Yona A, Senjyu T, Funabashi T, Kim C-H. Determination method of insolation prediction with fuzzy and applying neural network for long-term ahead PV power output correction. IEEE Transactions on Sustainable Energy 2013;4:527–33.
[211] Chu Y, Urquhart B, Gohari SMI, Pedro HTC, Kleissl J, Coimbra CFM. Short-term reforecasting of power output from a 48 MWe solar PV plant. Solar Energy 2015;112:68–77. https://doi.org/10.1016/j.solener.2014.11.017.
[212] Wu Y-K, Chen C-R, Abdul Rahman H. A novel hybrid model for short-term forecasting in PV power generation. International Journal of Photoenergy 2014;2014.
[213] Cervone G, Clemente-Harding L, Alessandrini S, Delle Monache L. Short-term photovoltaic power forecasting using Artificial Neural Networks and an Analog Ensemble. Renewable Energy 2017;108:274–86.
[214] Dong Z, Yang D, Reindl T, Walsh WM. Satellite image analysis and a hybrid ESSS/ANN model to forecast solar irradiance in the tropics. Energy Conversion and Management 2014;79:66–73.

[215] Dolara A, Grimaccia F, Leva S, Mussetta M, Ogliari E. A Physical Hybrid Artificial Neural Network for Short Term Forecasting of PV Plant Power Output. Energies 2015;8:1138–53. https://doi.org/10.3390/en8021138.

[216] Wang H, Yi H, Peng J, Wang G, Liu Y, Jiang H, et al. Deterministic and probabilistic forecasting of photovoltaic power based on deep convolutional neural network. Energy Conversion and Management 2017;153:409–22.

[217] Gandelli A, Grimaccia F, Leva S, Mussetta M, Ogliari E. Hybrid model analysis and validation for PV energy production forecasting, IEEE; 2014, p. 1957–62.

[218] Hussain S, AlAlili A. Online Sequential Learning of Neural Networks in Solar Radiation Modeling Using Hybrid Bayesian Hierarchical Approach. Journal of Solar Energy Engineering 2016;138.

[219] Royer JC, Wilhelm VE, Junior LAT, Franco EMC. Short-term solar radiation forecasting by using an iterative combination of wavelet artificial neural networks. Independent Journal of Management & Production 2016;7:271–88.

[220] Ghimire S, Deo RC, Raj N, Mi J. Deep solar radiation forecasting with convolutional neural network and long short-term memory network algorithms. Applied Energy 2019;253:113541.

[221] Che Y, Chen L, Zheng J, Yuan L, Xiao F. A novel hybrid model of WRF and clearness index-based Kalman filter for day-ahead solar radiation forecasting. Applied Sciences 2019;9:3967.

[222] Voyant C, Darras C, Muselli M, Paoli C, Nivet M-L, Poggi P. Bayesian rules and stochastic models for high accuracy prediction of solar radiation. Applied Energy 2014;114:218–26.

[223] Gairaa K, Khellaf A, Messlem Y, Chellali F. Estimation of the daily global solar radiation based on Box–Jenkins and ANN models: A combined approach. Renewable and Sustainable Energy Reviews 2016;57:238–49.

[224] Shamshirband S, Mohammadi K, Piri J, Petković D, Karim A. Hybrid auto-regressive neural network model for estimating global solar radiation in Bandar Abbas, Iran. Environmental Earth Sciences 2016;75:1–12.

[225] Huang X, Li Q, Tai Y, Chen Z, Zhang J, Shi J, et al. Hybrid deep neural model for hourly solar irradiance forecasting. Renewable Energy 2021;171:1041–60.

[226] Methodology of Köppen-Geiger-Photovoltaic climate classification and implications to worldwide mapping of PV system performance. Solar Energy 2019;191:672–85. https://doi.org/10.1016/j.solener.2019.08.072.

[227] CMP21. CMP21 spectrally flat Class A pyranometer - Kipp & Zonen n.d. https://www.kippzonen.com/Product/14/CMP21-Pyranometer#.X2m3Y2hKiM8 (accessed September 22, 2020).

[228] Improved accuracy and reliability with CHP 1 Pyrheliometer - Kipp & Zonen n.d. https://www.kippzonen.com/Product/18/CHP1-Pyrheliometer#.YN7Ul-hKiM8 (accessed July 2, 2021).

[229] ISO 9060. ISO 9060:2018(en), Solar energy — Specification and classification of instruments for measuring hemispherical solar and direct solar radiation 2018. https://www.iso.org/obp/ui/#iso:std:iso:9060:ed-2:v1:en (accessed September 22, 2020).

[230] SOLYS2 sun tracker - Kipp & Zonen n.d. https://www.kippzonen.com/Product/20/SOLYS2-Sun-Tracker#.YkV_IG7P1hE (accessed March 31, 2022).

[231] Schüler D, Wilbert S, Geuder N, Affolter R, Wolfertstetter F, Prahl C, et al. The enerMENA meteorological network – Solar radiation measurements in the MENA region. AIP Conference Proceedings 2016;1734:150008. https://doi.org/10.1063/1.4949240.

[232] Kopp G, Lean J. A New, Lower Value of Total Solar Irradiance: Evidence and Climate Significance. Geophysical Research Letters - GEOPHYS RES LETT 2011;38. https://doi.org/10.1029/2010GL045777.

[233] Solar and Infrared Radiation Measurements | Taylor & Francis Group n.d. https://www.taylorfrancis.com/books/9781315217376 (accessed December 2, 2019).

[234] Ameen B, Balzter H, Jarvis C. Quality Control of Global Horizontal Irradiance Estimates through BSRN, TOACs and Air Temperature/Sunshine Duration Test Procedures. Climate 2018;6:69. https://doi.org/10.3390/cli6030069.

[235] Sengupta M, Kurtz S, Dobos A, Wilbert S, Lorenz E, Renné D, et al. Best Practices Handbook for the Collection and Use of Solar Resource Data for Solar Energy Applications. IEA Solar Heating and Cooling Programme; 2015. https://doi.org/10.18777/ieashc-task46-2015-0001.

[236] Shi G-Y, Hayasaka T, Ohmura A, Chen Z-H, Wang B, Zhao J-Q, et al. Data Quality Assessment and the Long-Term Trend of Ground Solar Radiation in China. J Appl Meteor Climatol 2008;47:1006–16. https://doi.org/10.1175/2007JAMC1493.1.

[237] Tang W, Yang K, He J, Qin J. Quality control and estimation of global solar radiation in China. Solar Energy 2010;84:466–75. https://doi.org/10.1016/j.solener.2010.01.006.

[238] Maxwell E, Wilcox S, Rymes M. Users manual for seri qc software, assessing the quality of solar radiation data. Solar Energy Research Institute, Golden, CO 1993.

[239] Long CN, Dutton EG. BSRN Global Network recommended QC tests, V2.x. EPIC3Bremerhaven, PANGAEA 2010. https://epic.awi.de/id/eprint/30083/ (accessed November 21, 2019).

[240] García RD, Cuevas E, Ramos R, Cachorro VE, Redondas A, Moreno-Ruiz JA. Description of the Baseline Surface Radiation Network (BSRN) station at the Izaña Observatory (2009–2017): measurements and quality control/assurance procedures. Geosci Instrum Method Data Syst 2019;8:77–96. https://doi.org/10.5194/gi-8-77-2019.

[241] Moreno S, Silva M, Santigosa L. A proposed methodology for quick assessment of timestamp and quality control results of solar radiation data. Renewable Energy 2015;Volume 78,:Pages 531-537. https://doi.org/10.1016/j.renene.2015.01.031.

[242] Ntsangwane L, Sivakumar V, Mabasa B, Zwane N, Ncongwane K, Botai J. Quality Control of Solar Radiation Data within the South African Weather Service Solar Radiometric Network 2018. https://doi.org/10.20944/preprints201808.0363.v1.

[243] Roesch A, Wild M, Ohmura A, Dutton EG, Long CN, Zhang T. Assessment of BSRN radiation records for the computation of monthly means. Atmospheric Measurement Techniques 2011;4:339–54. https://doi.org/10.5194/amt-4-339-2011.

[244] Zo I-S, Jee J-B, Kim B-Y, Lee K-T. Baseline Surface Radiation Network (BSRN) quality control of solar radiation data on the Gangneung-Wonju National University radiation station. Asia-Pacific J Atmos Sci 2017;53:11–9. https://doi.org/10.1007/s13143-016-0029-5.

[245] Geiger M, Diabaté L, Ménard L, Wald L. A Web service for controlling the quality of measurements of global solar irradiation. Solar Energy 2002;73:475–80.

[246] Scharmer K, Page J, Wald L, Albuisson M, Czeplak G, Bourges B, et al. The European Solar Radiation Atlas Vol.1: Fundamentals and maps, 2000.

[247] Tham YW. SOLAR RADIATION RELATED CLIMATE CHANGE PROJECTIONS FOR THE UK n.d.:263.

[248] Molineaux B, Ineichen P. AQC- B. Molineaux- P. Ineichen p. 1 Automatic Quality Control of Daylight Measurements: Software for IDMP Stations. n.d.

[249] Muneer T, Fairooz F. Quality control of solar radiation and sunshine measurements – lessons learnt from processing worldwide databases. Building Services Engineering Research and Technology 2002;23:151–66. https://doi.org/10.1191/0143624402bt038oa.

[250] Guide to recommended practice of daylight measurement + Disk D006 | CIE n.d. http://www.cie.co.at/publications/guide-recommended-practice-daylight-measurement-disk-d006 (accessed December 9, 2019).

[251] Younes S, Claywell R, Muneer T. Quality control of solar radiation data: Present status and proposed new approaches. Energy 2005;30:1533–49. https://doi.org/10.1016/j.energy.2004.04.031.

[252] Fox G, Hiranandani S, Kennedy K, Koelbel C, Kremer U, Tseng C-W, et al. Fortran D language specification 1990.

[253] Journée M, Bertrand C. Quality control of solar radiation data within the RMIB solar measurements network. Solar Energy 2011;85:72–86. https://doi.org/10.1016/j.solener.2010.10.021.

[254] Ineichen P. Solar radiation resource in Geneva: measurements, modeling, data quality control, format and accessibility 2013.

[255] A broadband simplified version of the Solis clear sky model - ScienceDirect n.d. https://www.sciencedirect.com/science/article/abs/pii/S0038092X08000406 (accessed January 15, 2020).
[256] Geuder N, Wolfertstetter F, Wilbert S, Schüler D, Affolter R, Kraas B, et al. Screening and flagging of solar irradiation and ancillary meteorological data. Energy Procedia 2015;69:1989–98.
[257] Urraca R, Gracia-Amillo AM, Huld T, Martinez-de-Pison FJ, Trentmann J, Lindfors AV, et al. Quality control of global solar radiation data with satellite-based products. Solar Energy 2017;158:49–62.
[258] iea-pvps.org - Task 16: Solar Ressource n.d. http://www.iea-pvps.org/index.php?id=389 (accessed January 15, 2020).
[259] Blanc Ph, Wald L. The SG2 algorithm for a fast and accurate computation of the position of the Sun for multi-decadal time period. Solar Energy 2012;86:3072–83. https://doi.org/10.1016/j.solener.2012.07.018.
[260] CMP21- Kipp & Zonen. CMP21 spectrally flat Class A pyranometer - Kipp & Zonen n.d. https://www.kippzonen.com/Product/14/CMP21-Pyranometer#.X1YSF3lKiM8 (accessed September 7, 2020).
[261] Ellis BH, Deceglie M, Jain A. Automatic Detection of Clear-sky Periods Using Ground and Satellite Based Solar Resource Data. 2018 IEEE 7th World Conference on Photovoltaic Energy Conversion, WCPEC 2018 - A Joint Conference of 45th IEEE PVSC, 28th PVSEC and 34th EU PVSEC 2018:2293–8. https://doi.org/10.1109/PVSC.2018.8547877.
[262] Zhang W, Kleiber W, Florita AR, Hodge B-M, Mather B. A stochastic downscaling approach for generating high-frequency solar irradiance scenarios. Solar Energy 2018;176:370–9.
[263] Reno MJ, Hansen CW. Identification of periods of clear sky irradiance in time series of GHI measurements. Renewable Energy 2016;90:520–31. https://doi.org/10.1016/j.renene.2015.12.031.
[264] Ineichen P, Perez R. A new airmass independent formulation for the Linke turbidity coefficient. Solar Energy 2002;73:151–7.
[265] Kasten F, Young AT. Revised optical air mass tables and approximation formula. Applied Optics 1989;28:4735–8.
[266] Schmetz J, Pili P, Tjemkes S, Just D, Kerkmann J, Rota S, et al. An introduction to Meteosat second generation (MSG). Bulletin of the American Meteorological Society 2002;83:977–92.
[267] Diabaté L, Blanc Ph, Wald L. Solar radiation climate in Africa. Solar Energy 2004;76:733–44. https://doi.org/10.1016/j.solener.2004.01.002.
[268] Rigollier C, Bauer O, Wald L. On the clear sky model of the ESRA—European Solar Radiation Atlas—with respect to the Heliosat method. Solar Energy 2000;68:33–48.
[269] Diabaté L, Remund J, Wald L. Linke turbidity factors for several sites in Africa. Solar Energy 2003;75:111–9. https://doi.org/10.1016/j.solener.2003.07.002.
[270] Qu Z, Gschwind B, Lefèvre M, Wald L. Improving HelioClim-3 estimates of surface solar irradiance using the McClear clear-sky model and recent advances in atmosphere composition 2014.
[271] Thomas C, Wey E, Blanc P, Wald L, Lefèvre M. Validation of HelioClim-3 Version 4, HelioClim-3 Version 5 and MACC-RAD Using 14 BSRN Stations. Energy Procedia 2016;91:1059–69. https://doi.org/10.1016/j.egypro.2016.06.275.
[272] Lefèvre M, Oumbe A, Blanc P, Espinar B, Gschwind B, Qu Z, et al. McClear: a new model estimating downwelling solar radiation at ground level in clear-sky conditions. Atmospheric Measurement Techniques 2013;6:2403–18. https://doi.org/10.5194/amt-6-2403-2013.
[273] Emde C, Buras-Schnell R, Kylling A, Mayer B, Gasteiger J, Hamann U, et al. The libRadtran software package for radiative transfer calculations (version 2.0. 1). Geoscientific Model Development 2016;9:1647–72.
[274] Schroedter-Homscheidt M. The Copernicus Atmosphere Monitoring Service (CAMS) Radiation Service in a nutshell n.d.:6.
[275] Global Forecast System (GFS) n.d. https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/global-forcast-system-gfs (accessed February 10, 2020).
[276] Verbois H, Huva R, Rusydi A, Walsh W. Solar irradiance forecasting in the tropics using numerical weather prediction and statistical learning. Solar Energy 2018;162:265–77.

[277] Peerlings E. Cloud gazing and catching the sun's rays 2019.
[278] Nevins MG, Apell JN. Emerging investigator series: quantifying the impact of cloud cover on solar irradiance and environmental photodegradation. Environmental Science: Processes & Impacts 2021;23:1884–92.
[279] Coimbra CF, Kleissl J, Marquez R. Overview of solar-forecasting methods and a metric for accuracy evaluation. Solar Energy Forecasting and Resource Assessment 2013:171–94.
[280] MATLAB n.d. https://matlab.mathworks.com/ (accessed June 23, 2022).
[281] Pedro HT, Larson DP, Coimbra CF. A comprehensive dataset for the accelerated development and benchmarking of solar forecasting methods. Journal of Renewable and Sustainable Energy 2019;11:036102.
[282] Badescu V. Modeling solar radiation at the earth's surface. vol. 1. Springer; 2014.
[283] Notton G, Paoli C, Vasileva S, Nivet ML, Canaletti J-L, Cristofari C. Estimation of hourly global solar irradiation on tilted planes from horizontal one using artificial neural networks. Energy 2012;39:166–79.
[284] Gazela M, Tambouratzis T. Estimation of hourly average solar radiation on tilted surface via ANNs. International Journal of Neural Systems 2002;12:1–13.
[285] Bilski J, Kowalczyk B, Marchlewska A, Zurada JM. Local Levenberg-Marquardt algorithm for learning feedforwad neural networks. Journal of Artificial Intelligence and Soft Computing Research 2020;10.
[286] Moré JJ. The Levenberg-Marquardt algorithm: implementation and theory. Numerical analysis, Springer; 1978, p. 105–16.
[287] Sakinah N, Tahir M, Badriyah T, Syarif I. LSTM with adam optimization-powered high accuracy preeclampsia classification, IEEE; 2019, p. 314–9.
[288] Sainath TN, Vinyals O, Senior A, Sak H. Convolutional, long short-term memory, fully connected deep neural networks, IEEE; 2015, p. 4580–4.
[289] Vorontsov E, Trabelsi C, Kadoury S, Pal C. On orthogonality and learning recurrent networks with long term dependencies, PMLR; 2017, p. 3570–8.
[290] Dickey DA, Bell WR, Miller RB. Unit roots in time series models: Tests and implications. The American Statistician 1986;40:12–26.
[291] Melard G. Algorithm AS 197: A fast algorithm for the exact likelihood of autoregressive-moving average models. Journal of the Royal Statistical Society Series C (Applied Statistics) 1984;33:104–14.
[292] Khan A, Sohail A, Zahoora U, Qureshi AS. A survey of the recent architectures of deep convolutional neural networks. Artificial Intelligence Review 2020;53:5455–516.
[293] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. ArXiv Preprint ArXiv:14091556 2014.
[294] Ke H, Chen D, Li X, Tang Y, Shah T, Ranjan R. Towards brain big data classification: Epileptic EEG identification with a lightweight VGGNet on global MIC. IEEE Access 2018;6:14722–33.
[295] Chen X, Guhl J. Industrial robot control with object recognition based on deep learning. Procedia CIRP 2018;76:149–54.
[296] Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L. Imagenet: A large-scale hierarchical image database, Ieee; 2009, p. 248–55.
[297] Dunne RA, Campbell NA. On the pairing of the softmax activation and cross-entropy penalty functions and the derivation of the softmax activation function. vol. 181, Citeseer; 1997, p. 185.
[298] Gueymard CA, Habte A, Sengupta M. Reducing uncertainties in large-scale solar resource data: the impact of aerosols. IEEE Journal of Photovoltaics 2018;8:1732–7. https://doi.org/10.1109/jphotov.2018.2869554.
[299] Boucher O. Atmospheric aerosols. Atmospheric Aerosols, Springer; 2015, p. 9–24.
[300] Conceição R, Merrouni AA, Lopes D, Alae A, Silva HG, Bennouna EG, et al. A comparative study of soiling on solar mirrors in Portugal and Morocco: preliminary results for the dry season. vol. 2126, AIP Publishing LLC; 2019, p. 220001. https://doi.org/10.1063/1.5117760.
[301] MODIS Web n.d. https://modis.gsfc.nasa.gov/gallery/ (accessed April 12, 2022).