



Centre d'Etudes Doctorales : Sciences et Techniques de l'Ingénieur

N° d'ordre 98/2021

THESE DE DOCTORAT

Présentée par

Mme : Zineb BOUNOUA

Discipline : Génie électrique

Spécialité : Génie électrique

Sujet de la thèse :

**Estimation et Prédiction des irradiations solaires à l'aide des modèles
de Machine learning et de Deep learning**

Thèse présentée et soutenue le Samedi 18 Décembre 2021 à 10h devant le jury composé de :

Nom Prénom	Titre	Etablissement	
Ali AHAITOUF	PES	Faculté des Sciences et Techniques de Fès	Président
Abdelmajid JAMIL	PES	Ecole Supérieure de Technologie de Fès	Rapporteur
Rachid TADILI	PES	Faculté des Sciences de Rabat	Rapporteur
Sanaa HAYANI-MOUNIR	PH	Faculté Polydisciplinaire de Khouribga	Rapporteur
Fatima ERRAHIMI	PES	Faculté des Sciences et Techniques de Fès	Examineur
Youness IDRISSE KHAMLI	PH	Ecole Nationale des Sciences Appliquées de Fès	Examineur
Abdellah MECHAQRANE	PES	Faculté des Sciences et Techniques de Fès	Directeur de thèse

Laboratoire d'accueil : Systèmes intelligents, Géoressources, et Energies Renouvelables (SIGER).

Etablissement : Faculté des Sciences et Techniques de Fès (FSTF).



Centre d'Etudes Doctorales : Sciences et Techniques de l'Ingénieur

Titre de la thèse : Estimation et Prédiction des irradiances solaires à l'aide des modèles de Machine learning et de Deep learning.

Nom et prénom du candidat : Zineb BOUNOUA

Spécialité : Génie électrique

Résumé de la thèse

Une connaissance précise de l'énergie solaire incidente au sol en un site donné est un facteur majeur pour évaluer la rentabilité de tout projet d'installation solaire. En effet, la disponibilité des données fiables du rayonnement solaire permettrait de correctement dimensionner et de prédire la production des centrales solaires photovoltaïques ou thermodynamiques pour élaborer des stratégies d'optimisation et de contrôle pour une gestion optimale du mix énergétique. L'objectif principal de notre thèse consiste à développer des modèles efficaces pour l'estimation des irradiances solaires globales horizontales (GHI), diffuses horizontales (DHI) et normales directes (DNI). Aussi, il a été question d'élaborer des modèles d'apprentissage automatique (Machine Learning) et profond (Deep Learning) pour la prédiction à l'avance du GHI à différents horizons temporels (h+1 à h+6 et 10min+10 à 10min+60) susceptibles d'intéresser les gestionnaires des réseaux électriques.

En premier lieu, une procédure de contrôle de qualité des données a été présentée et appliquée aux mesures expérimentales provenant des six sites marocains considérés dans le présent travail (Fès, Erfoud, Missour, Oujda, Zagora, TanTan). Une estimation des irradiances solaires journalières GHI, DHI et DNI a été, ensuite, effectuée à l'aide de plusieurs modèles empiriques et d'apprentissage automatique (Machine Learning) n'utilisant, comme entrées, que les variables couramment mesurées dans les stations météorologiques. Parmi les modèles empiriques, il a été constaté que le modèle TG1 basé uniquement sur la connaissance du gradient de température journalier est le plus performant pour estimer le GHI. Pour le DHI et le DNI, les deux modèles basés sur les paramètres hybrides que nous avons proposés se sont avérés les plus appropriés. Aussi, il a été noté que les modèles d'apprentissage automatique sont plus performants que les modèles empiriques. En explorant le potentiel des trois méthodes d'ensemble (Boosting, Bagging, Random Forests (RF)) dans l'estimation du GHI, le modèle RF s'est avéré être le plus robuste pour toutes les stations. Une étude de catégorisation des modèles d'apprentissage automatique est ensuite effectuée, lorsque seulement des mesures météorologiques restreintes sont disponibles. Les résultats obtenus montrent que les modèles SVM, RF et PMC peuvent être utilisés pour estimer le GHI en disposant uniquement des valeurs extrêmes de la température ou de l'humidité relative de l'air et que le modèle RF est le plus efficace pour l'estimation du GHI, DHI, et DNI en utilisant comme variables d'entrées les paramètres hybrides.

Enfin, une prédiction à l'avance du GHI à différents horizons temporels, à très court, à court et à long termes (10min+10 à 10min+60, h+1 à h+6 et h+48), a été effectuée en utilisant les modèles d'apprentissage automatique (PMC, RF, NARX) et profond (LSTM). Les résultats montrent que le modèle LSTM est plus performant que les modèles PMC et RF pour les prédictions annuelles et saisonnières et pour les prévisions horaires et sub-horaires du GHI. En particulier, une grande amélioration en terme de précision a été observée lorsque l'horizon temporel de prévision augmente.

Mots clés : Irradiation solaire ; modèles empiriques ; Machine Learning ; Deep Learning.

“ Ce qui nous fait vivre, c'est l'hypothèse selon laquelle les problèmes, insurmontables de nuit, sont surmontables de jour ”

Thomas Bernhard (Perturbation (1967))

Remerciements

Ce travail n'est jamais une œuvre individuelle, il est le fruit de nombreuses expériences, rencontres, et collaborations.

Aussi, je saisis cette occasion pour exprimer mes sincères remerciements à toutes les personnes qui ont participé de près ou de loin à ce travail, qui ont croisé ma route, en m'accordant le temps, l'écoute, les conseils, l'expérience et parfois réconfort et m'auront permis à leur manière de progresser, avancer, mûrir, apprendre et à mieux me découvrir.

Tout d'abord, je tiens à exprimer mon immense respect et gratitude à Monsieur le Doyen de la Faculté des Sciences et Techniques de Fès, au cadre administratif, et tout le corps professoral du département Génie électrique.

Mes remerciements vont aussi, à Monsieur le Directeur du laboratoire Systèmes Intelligents, Géo-ressources, et Energies Renouvelables (SIGER), **Pr. Ali AHAITOUF**, pour sa disponibilité, son enthousiasme communicatif et sa bienveillance.

Mes sincères remerciements et les plus profonds à mon directeur de thèse, **Pr. Abdellah MECHAQRANE**, pour m'avoir accordé l'opportunité de découvrir le monde de la recherche en m'intégrant parmi son équipe de recherche et pour m'avoir fait confiance depuis le début. Je tiens à profiter de cette occasion pour lui exprimer ma profonde gratitude pour son encadrement, ses précieux conseils et également pour sa totale compréhension et efficacité tout au long de ce travail. Je le remercie aussi pour tout ce qu'il m'a appris, humainement et scientifiquement, et pour m'avoir assuré des conditions de travail entièrement favorables.

Je ne manquerais pas non plus de remercier infiniment les honorables membres du jury qui ont accepté d'évaluer cette thèse à sa juste valeur, et de me faire part de leurs remarques sûrement pertinentes qui contribueront, sans nul doute, au perfectionnement du présent travail.

Mes sincères remerciements à Monsieur **Ali AHAITOUF**, Professeur à la Faculté des Sciences et Techniques de Fès, qui m'a fait l'honneur de présider le Jury et de profiter encore de son ample expérience.

J'adresse également mes profonds remerciements aux rapporteurs de cette thèse, pour le grand intérêt qu'ils ont porté à mon travail et l'honneur qu'ils m'ont fait en acceptant de participer au Jury :

- Monsieur **Abdelmajid JAMIL**, Professeur à l'Ecole Supérieure de Technologie de Fès.
- Monsieur **Rachid TADILI**, Professeur à la Faculté des Sciences de Rabat.
- Madame **Sanaa HAYANI-MOUNIR**, Professeur à Faculté Polydisciplinaire de Khouribga.

Je souhaiterais aussi adresser ma gratitude aux examinateurs de cette thèse, qui ont accepté de juger ce travail. Je suis particulièrement très honorée par leur présence dans ce Jury :

- Madame **Fatima ERRAHIMI**, Professeur à la Faculté des Sciences et Techniques de Fès.
- Monsieur **Youness IDRISSE KHAMLICHI**, Professeur à l'Ecole Nationale des Sciences Appliquées de Fès.

Je tiens aussi à remercier chaleureusement tous les membres du laboratoire Systèmes Intelligents, Géo-ressources, et Energies Renouvelables (SIGER), enseignants chercheurs et doctorants. Il n'aurait pas été possible de mener à bien cette thèse dans les meilleures conditions sans leur sérieuse implication.

Je ne saurais terminer ce préliminaire, sans penser à l'appui et la sollicitude que j'ai toujours trouvé auprès de ma mère, mon père, mes deux sœurs, mon frère et tous mes ami(e)s et collègues.

Disponibilité des données expérimentales:

Les auteurs tiennent à exprimer leur gratitude particulière au Centre aérospatial allemand (DLR) pour son soutien scientifique en mettant à notre disposition les mesures des composants solaires et des paramètres météorologiques du réseau enerMENA utilisées durant cette thèse.

C'est à DIEU que revient le grand mérite.

Dédicaces

Louange à Dieu seul le tout Puissant, plein de Miséricorde.

Grâce à lui ce travail a pu être accompli.

Je dédie ce modeste travail :

A Mon très cher Père Zoubeir Bounoua
Mon Soutien moral, celui qui s'est toujours sacrifié pour me voir réussir.
Je tiens à honorer l'homme que tu es.

A Ma très chère Maman Asmae Alami
La lumière de mes jours, la source de mes efforts, la flamme de Mon cœur.

Ce modeste travail est le fruit de tous vos sacrifices que vous avez déployé pour mon éducation et ma formation. Aucune dédicace ne saurait exprimer l'amour l'estime et le respect que j'ai toujours eu pour vous. Que Dieu vous procure une longue vie pleine de Bonheur et de santé pour que je puisse vous rendre une partie de tout ce que je vous dois.

A Ma chère sœur Ghita
Mes remerciements ne pourront jamais égaler ton grand cœur qui m'a apporté du soutien au moment où j'avais besoin d'aide.

A Ma chère petite sœur Kenza et Mon cher petit frère El Mehdi
Mes sources de mon Bonheur, je vous souhaite une vie pleine de bonheur et de succès.
A tous les moments d'enfance passés avec vous. Puissent nos liens fraternels se consolider et se pérenniser encore plus.

A Mon cher Youssef Filali
Tu as toujours offert soutien et réconfort, j'exprime envers toi une profonde admiration, reconnaissance et attachement inconditionnels.

A toute ma famille (BOUNOUA, ALAMI)
Ceux qui m'ont toujours encouragé tout au long de cette thèse. Que Dieu le Tout Puissant vous garde et vous procure santé et bonheur.

A tous mes amis, collègues et membres du laboratoire SIGER
*Ceux qui m'ont toujours encouragé et soutenu avec amour et patience au cours de mon doctorat.
Ceux qui ont vécu avec moi tous les moments de joies, de tristesses, d'échecs, et de réussites.
Ceux auxquels je tiens à présenter tout mon respect et ma grande reconnaissance.*

Résumé et Mots clés

Une connaissance précise de l'énergie solaire incidente au sol en un site donné est un facteur majeur pour évaluer la rentabilité de tout projet d'installation solaire. En effet, la disponibilité des données fiables du rayonnement solaire permettrait de correctement dimensionner et de prédire la production des centrales solaires photovoltaïques ou thermodynamiques pour élaborer des stratégies d'optimisation et de contrôle pour une gestion optimale du mix énergétique. L'objectif principal de notre thèse consiste à développer des modèles efficaces pour l'estimation des irradiances solaires globales horizontales (GHI), diffuses horizontales (DHI) et normales directes (DNI). Aussi, il a été question d'élaborer des modèles d'apprentissage automatique (Machine Learning) et profond (Deep Learning) pour la prédiction à l'avance du GHI à différents horizons temporels ($h+1$ à $h+6$ et $10\text{min}+10$ à $10\text{min}+60$) susceptibles d'intéresser les gestionnaires des réseaux électriques.

En premier lieu, une procédure de contrôle de qualité des données a été présentée et appliquée aux mesures expérimentales provenant des six sites marocains considérés dans le présent travail (Fès, Erfoud, Missouri, Oujda, Zagora, TanTan). Une estimation des irradiances solaires journalières GHI, DHI et DNI a été, ensuite, effectuée à l'aide de plusieurs modèles empiriques et d'apprentissage automatique (Machine Learning) n'utilisant, comme entrées, que les variables couramment mesurées dans les stations météorologiques. Parmi les modèles empiriques, il a été constaté que le modèle TG1 basé uniquement sur la connaissance du gradient de température journalier est le plus performant pour estimer le GHI. Pour le DHI et le DNI, les deux modèles basés sur les paramètres hybrides que nous avons proposés se sont avérés les plus appropriés. Aussi, il a été noté que les modèles d'apprentissage automatique sont plus performants que les modèles empiriques. En explorant le potentiel des trois méthodes d'ensemble (Boosting, Bagging, Random Forests (RF)) dans l'estimation du GHI, le modèle RF s'est avéré être le plus robuste pour toutes les stations. Une étude de catégorisation des modèles d'apprentissage automatique est ensuite effectuée, lorsque seulement des mesures météorologiques restreintes sont disponibles. Les résultats obtenus montrent que les modèles SVM, RF et PMC peuvent être utilisés pour estimer le GHI en disposant uniquement des valeurs extrêmes de la température ou de l'humidité relative de l'air et que le modèle RF est le plus efficace pour l'estimation du GHI, DHI, et DNI en utilisant comme variables d'entrées les paramètres hybrides.

Enfin, une prédiction à l'avance du GHI à différents horizons temporels, à très court, à court et à long termes ($10\text{min}+10$ à $10\text{min}+60$, $h+1$ à $h+6$ et $h+48$), a été effectuée en utilisant les modèles d'apprentissage automatique (PMC, RF, NARX) et profond (LSTM). Les résultats montrent que le modèle LSTM est plus performant que les modèles PMC et RF pour les prédictions annuelles et saisonnières et pour les prévisions horaires et sub-horaires du GHI. En particulier, une grande amélioration en terme de précision a été observée lorsque l'horizon temporel de prévision augmente.

Mots clés : Irradiation solaire ; modèles empiriques ; Machine Learning ; Deep Learning.

Abstract & Keywords

An accurate knowledge of the incident solar energy on the ground at a given site is a major factor in evaluating the profitability of any solar installation project. Indeed, the availability of reliable solar radiation data would allow to properly size photovoltaic or thermodynamic solar power plants and to predict their production in order to develop optimization and control strategies for an optimal management of the energy mix. The main objective of our thesis is to develop efficient models for the estimation of global horizontal solar irradiation (GHI), diffuse horizontal solar irradiation (DHI) and direct normal solar irradiation (DNI). In addition, it was also a question of developing machine learning and deep learning models for the ahead prediction of GHI at different time horizons ($h + 1$ to $h + 6$ and $10\text{min} + 10$ to $10\text{min} + 60$), likely to be of interest to electric network managers.

Initially, a data quality control procedure is presented and applied to the experimental measurements of the six Moroccan sites considered in the present work (Fez, Erfoud, Missouri, Oujda, Zagora, TanTan). Estimations of the daily solar irradiations GHI, DHI and DNI were then performed using several empirical and machine learning models, employing only the commonly variables measured in meteorological stations. Among the empirical models, it has been that the TG1 model using the daily temperature gradient as inputs is the most recommended for estimating the GHI. In addition, the two proposed hybrid parameters-based models were the most appropriate for estimating DHI and DNI. Also, it has been noticed that machine learning models perform better than empirical models. By exploring the potential of the three ensemble methods (Boosting, Bagging, Random Forests (RF)) in estimating GHI, the results show that the RF model was found to be the most robust for all stations. A categorization study of the machine learning models, when only restricted meteorological measurements are available, is then performed. The obtained performances show that the SVM, RF, and MLP models can be used to estimate GHI when only extreme values of air temperature or relative humidity are available, and that the RF model is the most efficient for the estimation of GHI, DHI, and DNI using hybrid parameters as the input variables.

Finally, ahead predictions of GHI at different time horizons, very short, short, and long terms ($10\text{min}+10$ to $10\text{min}+60$, $h+1$ to $h+6$, and $h+48$) were performed using machine learning models (MLP, RF, and NARX) and the deep learning model (LSTM). The results show that the LSTM model performs better than the MLP and RF models for annual and seasonal predictions and for hourly and sub-hourly forecasts of the GHI. In particular, a great improvement in accuracy has been observed as the forecast time horizon increases.

Keywords: Solar irradiation; Empirical models; Machine Learning; Deep Learning.

Liste des publications

1. Bounoua, Z., Marzouq, M., & Mechaqrane, A. (2018, June). Assessment of a quality control procedure of hourly solar irradiations at Fez city, Morocco. In IOP Conference Series: Earth and Environmental Science (Vol. 161, No. 1, p. 012010). IOP Publishing (**IF :0.41**).
2. Bounoua, Z., & Mechaqrane, A. (2018, May). Statistical analysis of global horizontal solar irradiation GHI in Fez city, Morocco. In IOP Conference Series: Materials Science and Engineering (Vol. 353, No. 1, p. 012006). IOP Publishing (**IF : 0.51**).
3. Marzouq, M., Bounoua, Z., Mechaqrane, A., Fadili, H. E., Lakhliai, Z., & Zenkour, K. (2018, June). ANN-based modelling and prediction of daily global solar irradiation using commonly measured meteorological parameters. In IOP Conference Series: Earth and Environmental Science (Vol. 161, No. 1, p. 012017). IOP Publishing (**IF :0.41**).
4. Bounoua, Z., & Mechaqrane, A. (2018, December). Prediction of daily global horizontal solar irradiation using artificial neural networks and commonly measured meteorological parameters. In AIP Conference Proceedings (Vol. 2056, No. 1, p. 020024). AIP Publishing.
5. Marzouq, M., Bounoua, Z., El Fadili, H., Mechaqrane, A., Zenkour, K., & Lakhliai, Z. (2019). New daily global solar irradiation estimation model based on automatic selection of input parameters using evolutionary artificial neural networks. Journal of Cleaner Production, 209, 1105-1118 (**Q1 IF: 7.246**).
6. Bounoua, Zineb, and Abdellah Mechaqrane. "Prediction of Short-Term and Long-Term Hourly Global Horizontal Solar Irradiation Using Artificial Neural Networks Techniques in Fez City, Morocco." International Conference on Electronic Engineering and Renewable Energy. Springer, Singapore, 2020 (**IF: 0.32**).
7. Bounoua, Z., Chahidi, L. O., & Mechaqrane, A. (2021). Estimation of daily global solar radiation using empirical and machine-learning methods: a case study of five Moroccan locations. Sustainable Materials and Technologies, 28, e00261 (**Q1 IF:7.053**).
8. Bounoua, Z., & Mechaqrane, A. (2022). Estimation of Daily Direct Normal Solar Irradiation Using Machine-Learning Methods. In WITS 2020 (pp. 627-638). Springer, Singapore (**IF: 0.32**).
9. Bounoua, Zineb, and Abdellah Mechaqrane. "Categorized machine-learning models to estimate daily solar radiation in Zagora City, Southern Morocco" (**Under Review**).
10. Bounoua, Zineb, and Abdellah Mechaqrane. « Hourly and Sub-Hourly Ahead Global Horizontal Solar Irradiation Forecasting using a Long Short-Term Memory Neural Networks Deep Learning Approach: A Case Study" (**Under Review**).

Liste des communications orales

1. Z. Bounoua et A. Mechaqrane « Quality control of hourly global and diffuse solar irradiation on horizontal surface at Fez city, Morocco » 2ème édition du Colloque Franco-Marocain sur les Energies Renouvelables et leur intégration aux réseaux de transport et de distribution (COFMER'2), 11-13 Avril 2017.
2. Zineb Bounoua, Manal Marzouq and Abdellah Mechaqrane « Assessment of a quality control procedure of hourly solar irradiations at Fez city, Morocco » The International Conference on Renewable Energies and Energy Efficiency (REEE'2017) FST Fès, 8 & 9 Novembre 2017.
3. Z. Bounoua et A. Mechaqrane « Statistical analysis of global horizontal solar irradiation GHI in Fez city, Morocco » Conférence Internationale des Bâtiments et des Villes Durables (BVD'17), UPF Fès, 5 & 6 Décembre 2017.
4. M. Marzouq, Z. Bounoua et A. Mechaqrane « Prediction of daily global solar irradiation using commonly measured meteorological parameters in the city of Fez » The International Conference on Renewable Energies and Energy Efficiency (REEE'2017) FST Fez, 8 & 9 Novembre 2017.
5. Zineb Bounoua and Abdellah Mechaqrane « Prediction of daily global horizontal solar irradiation using artificial neural networks and commonly measured meteorological parameters » The International Congress on solar energy research, technology and applications, Ouarzazate, 8-10 Mai 2018.
6. Zineb Bounoua and Abdellah Mechaqrane « Estimation of daily diffuse solar fraction using artificial neural networks and commonly measured meteorological parameters » 5ème Colloque International « Énergie & Ville du Futur » Fès, 25-26 Octobre 2018.
7. Zineb Bounoua and Abdellah Mechaqrane « Prediction of Short-Term and Long-Term Hourly Global Horizontal Solar Irradiation Using Artificial Neural Networks Techniques in Fez City, Morocco », 2nd International Conference on Electronic Engineering and Renewable Energy (ICEERE'2020), 13-15 April 2020, Saidia Morocco.
8. Zineb Bounoua and Abdellah Mechaqrane « Estimation of daily direct normal solar irradiation using Machine-learning methods », THE 6TH INTERNATIONAL CONFERENCE ON WIRELESS TECHNOLOGIES, EMBEDDED AND INTELLIGENT SYSTEMS (WITS-2020), OCTOBER 14-16, 2020, FEZ, MOROCCO.
9. Zineb Bounoua and Abdellah Mechaqrane « Analyse des variabilités temporelles du rayonnement solaire-Application à la gestion des centrales solaires », 5ème édition de la journée doctorale du laboratoire systèmes intelligents et énergies renouvelables (Décembre 2019), FEZ, MOROCCO.

Liste des abréviations

Bag	Bagging
Boost	Boosting
CIBMs	Modèles basés sur l'indice de clarté
DHI	Irradiance solaire diffuse horizontale (W/m^2)
DNI	Irradiance solaire directe normale (W/m^2)
EM	Modèles empiriques
GHI	Irradiance solaire globale horizontale (W/m^2)
HBMs	Modèles basés sur des paramètres hybrides
LSTM	Réseaux de neurones à longue mémoire à court terme
MAE	Erreur absolue moyenne
MAPE	Erreur moyenne absolue en pourcentage
MBE	Erreur de biais moyenne
NARX	Modèles Non linéaires AutoRégressifs avec des entrées eXogènes
nMAE	Erreur absolue moyenne normalisée
nMBE	Erreur de biais moyenne normalisée
nRMSE	Erreur quadratique moyenne normalisée
PAR	rayonnement photo-synthétiquement actif
PMC	Perceptron Multi-couches
QCP	Procédure de contrôle de qualité
R	Coefficient de corrélation
R²	Coefficient de détermination
REN21	Renewable Energy policy Network for the 21st Century
RF	Forêts aléatoires (Random Forest)
RHBMs	Modèles basés sur l'humidité relative
RNA	Réseaux de neurones artificiels
RMSE	Erreur quadratique moyenne
RSR	Radiomètre à bande d'ombre tournante
SG2	Solar Geometry 2
SPA	Solar Position Algorithm
SVM	Machine à vecteur de support (Support Vector Machine)
TBM	Modèles basés sur la température
TS	Capteurs thermiques

Liste des figures

Figure 1.1. Distribution spectrale du rayonnement solaire (Seidlitz et al., 2001).....	8
Figure 1.2. Coordonnées équatoriales.	10
Figure 1.3. Variation de la déclinaison solaire au cours de l'année.....	11
Figure 1.4. Coordonnées solaires horizontales.....	12
Figure 1.5. Spectre du rayonnement solaire (Dornelles et al., 2011).	17
Figure 1.6. Pyranomètres à thermopile.	21
Figure 1.7. Pyranomètres à photodiode.....	22
Figure 1.8. Un pyréliomètre (à gauche) et un héliographe (à droite).	22
Figure 1.9. Station météorologique placée sur le toit de la Faculté des Sciences et Techniques de Fès.....	24
Figure 1.10. Vue de l'ensemble des stations enerMENA installées au Maroc (encadré en vert), en Algérie, en Tunisie, en Égypte et en Jordanie.....	25
Figure 1.11. Station Thermique TS (gauche) et station RSR (droite) (Schüler et al., 2016).	26
Figure 1.12. Les variations horaires de GHI, DHI, mesurés à Fès (a), Erfoud (b), Missouri (c), Oujda (d), TanTan (e), et Zagora (f).	27
Figure 1.13. Sommes annuelles brutes des trois composantes solaires mesurées aux six sites marocains.	28
Figure 1.14. Variabilités horaires de la température ambiante mesurée aux six sites marocains.	29
Figure 1.15. Variabilités horaires de l'humidité relative de l'air mesurée aux six sites marocains.	29
Figure 1.16. Indice de clarté en fonction de l'altitude solaire pour les données solaires mesurées aux six stations étudiées.	35
Figure 1.17. Diagramme de dispersion de la fraction diffuse K_d en fonction de l'indice de clarté K_t après le premier test de contrôle de qualité.	36
Figure 1.18. Diagramme de dispersion de la fraction diffuse K_d en fonction de l'indice de clarté K_t après le deuxième test de contrôle de qualité (Les points en rouge sont les valeurs qui n'ont pas passé le test).	38
Figure 1.19. Diagramme de dispersion de la fraction diffuse K_d en fonction de l'indice de clarté K_t après le troisième test de contrôle de qualité (Les points en rouge sont les valeurs qui n'ont pas passé le test).	39
Figure 1.20. Les variations horaires de GHI, et PAR mesurés à Fès.	41
Figure 1.21. Régression linéaire entre GHI et PAR sans la période 3 Janvier – 23 Avril.	41
Figure 1.22. Corrélation entre les valeurs mesurées et calculées de GHI pour la période 3 Janvier – 23 Avril.	42
Figure 1.23. Variations horaires de GHI après correction des données manquantes au début de l'année 2013 à la ville de Fès.	43
Figure 2. 1. Graphique en Radar des valeurs de nRMSEs de chaque modèle empirique pour l'ensemble de données de validation.	63
Figure 2. 2. Nuages de points des valeurs quotidiennes mesurées et estimées de GHI en utilisant les modèles les plus performants pour les ensembles de données de validation.....	63

Figure 2. 3. Nuages de points des valeurs quotidiennes mesurées et estimées de GHI en utilisant les modèles les plus performants pour les ensembles de données de validation à la ville de Fès.	64
Figure 2. 4. Nuages de points des valeurs quotidiennes mesurées et estimées de GHI en utilisant le modèle proposé MH6 pour les ensembles de données de validation à la ville de Fès.	66
Figure 2. 5. Diagrammes de dispersion des valeurs quotidiennes mesurées et estimées de DHI et DNI en utilisant les modèles les plus performants dans chaque catégorie pour les ensembles de données de validation à Zagora.	68
Figure 3. 1. Représentation schématique d'un neurone biologique	85
Figure 3. 2. Analogie entre le neurone biologique et le neurone formel	86
Figure 3. 3. Architecture du PMC	87
Figure 3. 4. Architectures du réseau de neurones NARX	90
Figure 3. 5. Architecture des méthodes d'ensemble.....	92
Figure 3. 6. Diagramme du processus du Boosting.....	92
Figure 3. 7. Diagramme du processus du Bagging.....	94
Figure 3. 8. Diagramme du processus des forêts aléatoires (Random Forests).....	96
Figure 3. 9. Régression linéaire unidimensionnelle SVR (Awad and Khanna, 2015).	98
Figure 3. 10. Types de fonctions de perte : a) linéaire, b) quadratique, c) Huber (Awad and Khanna, 2015).	98
Figure 3. 11. Régressions obtenues par le meilleur modèle PMC.	103
Figure 3. 12. Régressions obtenues par le meilleur modèle NARX avec une récurrence d'un jour.	103
Figure 3. 13. Régressions obtenues par le meilleur modèle NARX avec une récurrence de deux jours.....	104
Figure 3. 14. Régressions obtenues par le meilleur modèle NARX avec une récurrence de trois jours.....	104
Figure 3. 15. Nuages de points des valeurs quotidiennes mesurées et estimées de GHI en utilisant les quatre modèles d'apprentissage automatique pour les ensembles de données de validation.....	111
Figure 3. 16. Humidité relative de l'air maximale journalière mesurée à Tan-Tan.....	113
Figure 3. 17. Graphique Radar des valeurs de nRMSE obtenues pour les estimations journalières de GHI pour les ensembles de données de validation et pour tous les modèles testés.	114
Figure 3. 18. Nuages de points des valeurs quotidiennes mesurées et estimées de GHI à la ville de Fès, en utilisant les quatre modèles d'apprentissage automatique pour les ensembles de données d'entraînement et de validation.	115
Figure 3. 19. Tracé des séries temporelles du GHI mesuré et estimé quotidiennement par le modèle SGT3, pour les ensembles d'entraînement et de validation.	118
Figure 3. 20. Séries temporelles des trois composantes solaires (GHI, DHI, et DNI) mesurées et estimées correspondant aux meilleurs modèles basés sur les paramètres hybrides (RGH4, RDH3 et RDNH1) pour les 120 jours de l'ensemble de validation.	121
Figure 3. 21. Séries temporelles de DHI et DNI mesurées et estimées correspondantes aux meilleurs modèles basés sur l'indice de clarté (RDC2 et PDNC1) pour les 120 jours de l'ensemble de validation.....	124
Figure 3. 22. La meilleure méthode d'apprentissage automatique par catégories d'entrées pour estimer le GHI, le DHI et le DNI quotidiens.....	125

Figure 3. 23. Nuages de points des valeurs quotidiennes mesurées et estimées de GHI (a), DHI (b) et DNI (c) correspondants au modèle le plus précis trouvé dans chaque catégorie d'entrées.	126
Figure 4.1. Structure d'une cellule à longue mémoire à court terme (Amey laddad, 2019).....	136
Figure 4. 2. La procédure des calculs de base effectués par la couche LSTM étape par étape (Jeff Hu, 2017).	137
Figure 4. 3. Régressions entre les valeurs mesurées et prédites par les modèles NARX (à gauche) et PMC (à droite) pour les ensembles de données de validation.	142
Figure 4. 4. Valeurs horaires prédites et mesurées de GHI pour une semaine choisie aléatoirement en utilisant les modèles NARX et PMC (MLP).	142
Figure 4. 5. Préviation horaire de l'irradiation globale horizontale pour deux jours à l'avance (h+48) par rapport aux données mesurées du 1 et 2 janvier 2015 au site de Fès.	143
Figure 4. 6. Régressions obtenues pour la ville de Zagora en utilisant le modèle NARX pour prédire l'irradiation globale horaire à l'horizon h+1 pour l'ensemble de données d'entraînement et de validation.	144
Figure 4. 7. Valeurs horaires prédites et mesurées de GHI pour une semaine choisie aléatoirement en utilisant le modèle NARX à Zagora (h+1).	144
Figure 4. 8. Prévvisions horaires de GHI pour deux jours à l'avance (h+48) par rapport aux données mesurées du 1 et 2 janvier 2015 pour chaque site enerMENA.	145
Figure 4. 9. Corrélations des valeurs passées des entrées endogènes avec les valeurs futures horaires (a) et sub-horaires (b) de GHI.	148
Figure 4. 10. Comparaison entre les performances annuelles des modèles de prédiction à l'avance du GHI en termes de nRMSE pour les différents horizons horaires.	151
Figure 4. 11. Nuages de points des valeurs mesurées et prédites de GHI pour h+1 (a), h+2 (b), h+3 (c), h+4 (d), h+5 (e) et h+6 (f) en utilisant les modèles LSTMs, pour l'ensemble de validation.	152
Figure 4. 12. Tracé des séries temporelles de 77 heures (7 jours) mesurées et prédites de GHI correspondantes aux modèles ANN, RF et LSTM pour l'ensemble de validation (h+6).	152
Figure 4. 13. Comparaison entre les performances saisonnières des modèles de prédiction à l'avance de GHI en termes du nRMSE pour les différents horizons horaires.	153
Figure 4. 14. Variabilités horaires mesurées et prédites (7 jours) de GHI correspondantes aux modèles PMC, RF et LSTM pour chaque saison et pour un horizon temporel allant jusqu'à 6h (h+6).	155
Figure 4. 15. Comparaison entre les performances annuelles des modèles de prédiction à l'avance du GHI en termes du nRMSE pour les différents horizons sub-horaires.	157
Figure 4. 16. Nuages de points des valeurs mesurées et prédites de GHI pour 10min+10 (a), 10min+20 (b), 10min+30 (c), 10min+40 (d), 10min+50 (e) et 10min+60 (f) en utilisant le modèle LSTM, pour l'ensemble de validation.	157
Figure 4. 17. Séries temporelles de 500 valeurs sub-horaires mesurées et prédites de GHI correspondantes aux modèles PMC, RF et LSTM pour l'ensemble de validation (10min+60).	158
Figure 4. 18. Comparaison entre les performances saisonnières des modèles de prédiction à l'avance de GHI en termes du nRMSE pour les différents horizons sub-horaires.....	160
Figure 4. 19. Séries temporelles de 500 valeurs sub-horaires mesurées et prédites de GHI correspondantes aux modèles PMC, RF et LSTM pour chaque saison et pour un horizon temporel 10min+60.	161

Liste des tableaux

Tableau 1.1. Variations horaires de l'azimut solaire au cours de la journée du 1 Mai 2014 à Fès.	16
Tableau 1.2. Variations horaires du zénith solaire au cours de la journée du 1 Mai 2014 à Fès.	16
Tableau 1.3. Variations horaires de la hauteur du soleil au cours de la journée du 1 Mai 2014 à Fès.....	17
Tableau 1.4. Les stations marocaines du réseau enerMENA.....	25
Tableau 1.5. Précision des instruments de mesure utilisés.	26
Tableau 1.6. Nombre des données solaires éliminées par la procédure de contrôle de qualité pour chaque site d'étude.....	40
Tableau 1.7. Performances entre les valeurs calculées et mesurées du GHI pour la période 3 Janvier – 23 Avril.....	43
Tableau 2. 1. Modèles empiriques testés pour estimer l'irradiation global horizontale dans les six villes marocaines considérées.....	52
Tableau 2. 2. Modèles empiriques testés pour l'estimation journalière de GHI.	55
Tableau 2. 3. Modèles empiriques testés pour l'estimation journalière de GHI pour la ville de Fès uniquement.	55
Tableau 2. 4. Modèles empiriques d'estimation des irradiations diffuses horizontales et directes normales journalières.	56
Tableau 2. 5. Coefficients de régression des 22 modèles empiriques appliqués aux stations de mesures du réseau enerMENA.....	57
Tableau 2. 6. Indicateurs statistiques de performances des 22 modèles empiriques pour l'estimation du GHI dans les six cas étudiés (Bounoua et al., 2021).....	60
Tableau 2. 7. Performances des quatre meilleurs modèles obtenus.....	62
Tableau 2. 8. Performances des quatre meilleurs modèles empiriques sur la ville de Fès.	62
Tableau 2. 9. Résultats des modèles empiriques du tableau 2 pour les 6 stations marocaines considérées.	64
Tableau 2. 10. Résultats des modèles empiriques appliqués à la ville de Fès uniquement. ...	66
Tableau 2. 11. Indicateurs de performances statistiques obtenus pour l'estimation des irradiations solaires diffuses journalières (DHI) à la ville de Zagora.	67
Tableau 2. 12. Indicateurs statistiques obtenus par les modèles empiriques d'estimation journalière des irradiations solaires directes d'un plan normal (DNI) à la ville de Zagora.....	68
Tableau 2. 13. Performances statistiques des modèles DH3 et DNH1 appliqués aux autres sites.	69
Tableau 2. 14. Indicateurs statistiques obtenus par les modèles CIBMs pour l'estimation des DHI et DNI aux autres sites.	69
Tableau 3. 1. Performances des modèles d'apprentissage automatique pour la prédiction de l'irradiation solaire globale journalière et horaire dans la littérature.....	82
Tableau 3. 2. Procédure générale du Boosting pour la régression (Gey and Poggi, 2006).....	93
Tableau 3. 3. Procédure générale du Bagging pour la régression (Agrégation de modèles, 2018).....	95

Tableau 3. 4. Procédure générale des forêts aléatoires pour la régression (Agrégation de modèles, 2018).	96
Tableau 3. 5. Performances du PMC pour l'estimation du GHI quotidienne à Fès (Bounoua and Mechaqrane, 2018).	100
Tableau 3. 6. Performances du modèle NARX pour l'estimation du GHI à différents horizons temporels (jour $j+1$, $j+2$ et $j+3$).	101
Tableau 3. 7. Meilleures architectures et leurs performances (Bounoua and Mechaqrane, 2018).	102
Tableau 3. 8. Performances statistiques des meilleures architectures des deux modèles PMC et NARX appliqués aux sites enerMENA, pour l'ensemble de validation.	105
Tableau 3. 9. Données météorologiques et astronomiques considérées (Bounoua et al., 2021).	109
Tableau 3. 10. Performances d'estimation du GHI quotidien des quatre méthodes d'apprentissage automatique pour tous les cas étudiés (ensembles de données d'entraînement et de validation) (Bounoua et al., 2021).	110
Tableau 3. 11. Performances des méthodes d'apprentissage automatique dans l'estimation du GHI quotidienne pour la ville de Fès (ensembles de données d'entraînement et de validation).	114
Tableau 3. 12. Détails des modèles d'apprentissage automatique développés.	116
Tableau 3. 13. Indicateurs statistiques des modèles TBMs pour estimer les valeurs journalières de GHI. Les lignes en surbrillance représentent les meilleures performances obtenues pour chaque méthode utilisée. En revanche, les chiffres en gras indiquent le meilleur modèle trouvé dans l'ensemble du groupe.	117
Tableau 3. 14. Indicateurs statistiques des modèles RHBMs pour estimer les valeurs journalières de GHI. Les lignes en surbrillance représentent les meilleures performances obtenues pour chaque méthode utilisée. En revanche, les chiffres en gras indiquent le meilleur modèle trouvé dans l'ensemble du groupe.	118
Tableau 3. 15. Indicateurs statistiques des modèles HBMs pour estimer les valeurs journalières de GHI. Les lignes en surbrillance représentent les meilleures performances obtenues pour chaque méthode utilisée. En revanche, les chiffres en gras indiquent le meilleur modèle trouvé dans l'ensemble du groupe.	119
Tableau 3. 16. Indicateurs statistiques des modèles HBMs pour estimer les valeurs journalières de DHI. Les lignes en surbrillance représentent les meilleures performances obtenues pour chaque méthode utilisée. En revanche, les chiffres en gras indiquent le meilleur modèle trouvé dans l'ensemble du groupe.	120
Tableau 3. 17. Indicateurs statistiques des modèles HBMs pour estimer les valeurs journalières de DNI. Les chiffres en gras indiquent le meilleur modèle trouvé dans l'ensemble du groupe.	120
Tableau 3. 18. Indicateurs statistiques des modèles CIBMs pour estimer les valeurs journalières de DHI. Les lignes en surbrillance représentent les meilleures performances obtenues pour chaque méthode utilisée. En revanche, les chiffres en gras indiquent le meilleur modèle trouvé dans l'ensemble du groupe.	123
Tableau 3. 19. Indicateurs statistiques des modèles CIBMs pour estimer les valeurs journalières de DNI. Les lignes en surbrillance représentent les meilleures performances obtenues pour chaque méthode utilisée. En revanche, les chiffres en gras indiquent le meilleur modèle trouvé dans l'ensemble du groupe.	123

Tableau 3. 20. Indicateurs statistiques obtenus par les meilleurs modèles d'apprentissage automatique catégorisés pour l'estimation journalière de GHI aux autres sites d'étude. Les chiffres en gras indiquent la meilleure performance obtenue parmi ces modèles.....	127
Tableau 3. 21. Indicateurs statistiques obtenus par les meilleurs modèles d'apprentissage automatique catégorisés pour l'estimation journalière de DHI aux autres sites d'étude. Les chiffres en gras indiquent la meilleure performance obtenue parmi ces modèles.....	127
Tableau 3. 22. Indicateurs statistiques obtenus par les meilleurs modèles d'apprentissage automatique catégorisés pour l'estimation journalière de DNI aux autres sites d'étude. Les chiffres en gras indiquent la meilleure performance obtenue parmi ces modèles.....	128
Tableau 4. 1. Corrélation du GHI horaire avec les variables météorologiques.	139
Tableau 4. 2. Indicateurs statistiques obtenus par le modèle PMC en testant différentes combinaisons d'entrées pour l'ensemble de données de validation sur 5 exécutions.	140
Tableau 4. 3. Indicateurs statistiques obtenus par le modèle NARX en testant différentes combinaisons d'entrées pour l'ensemble de données de validation sur 5 exécutions.	141
Tableau 4. 4. Indicateurs statistiques obtenus en testant la meilleure combinaison d'entées et la meilleure architecture du modèle NARX (h+1) sur les sites enerMENA pour l'ensemble de données d'entraînement et de validation pour 5 exécutions.....	144
Tableau 4. 5. Optimisation des hyper-paramètres du modèle LSTM.	149
Tableau 4. 6. Performances annuelles des modèles pour la prédiction à l'avance des GHI horaires pour l'ensemble de validation. Les chiffres en gras représentent le meilleur prédicteur pour chaque horizon et chaque indicateur statistique.....	150
Tableau 4. 7. Performances saisonnières de la prédiction à l'avance des GHI pour l'ensemble de validation. Les chiffres en gras représentent le meilleur prédicteur pour chaque horizon et chaque indicateur statistique.	154
Tableau 4. 8. Performances annuelles pour la prédiction à l'avance des GHI sub-horaires pour l'ensemble de validation. Les chiffres en gras représentent le meilleur prédicteur pour chaque horizon et chaque indicateur statistique.	156
Tableau 4. 9. Performances saisonnières de la prédiction à l'avance des GHI sub-horaires pour l'ensemble de validation. Les chiffres en gras représentent le meilleur prédicteur pour chaque horizon et chaque indicateur statistique.	159
Tableau 4. 10. Performances annuelles pour la prédiction à l'avance du GHI à l'aide du modèle LSTM pour l'ensemble de validation. Les chiffres en gras représentent la meilleure précision obtenue pour chaque horizon.	163
Tableau 4. 11. Performances saisonnières de la prédiction à l'avance des GHI à l'aide du modèle LSTM pour l'ensemble de validation. Les chiffres en gras représentent la meilleure précision obtenue pour chaque horizon temporel et chaque saison.	163
Tableau 4. 12. Performances annuelles pour la prédiction à l'avance de GHI à l'aide du modèle LSTM pour l'ensemble de validation. Les chiffres en gras représentent la meilleure précision obtenue pour chaque horizon.	165
Tableau 4. 13. Performances saisonnières de la prédiction à l'avance des GHI à l'aide du modèle LSTM pour l'ensemble de validation. Les chiffres en gras représentent la meilleure précision obtenue pour chaque horizon temporel et chaque saison.	166

Sommaire

Introduction générale.....	1
Chapitre 1 : Rayonnement solaire : mesures expérimentales et contrôle de qualité	
1. Introduction	7
2. Rayonnement solaire.....	7
2.1. Constante solaire	8
2.2. Eclairement solaire extraterrestre	8
3. Coordonnées géographiques terrestres	9
3.1. Latitude ϕ	9
3.2. Longitude L	9
3.3. Altitude.....	9
4. Coordonnées solaires.....	10
4.1. Coordonnées équatoriales.....	10
4.1.1. Déclinaison solaire.....	10
4.1.2. Angle horaire	11
4.2. Coordonnées horizontales.....	12
4.2.1. Hauteur du soleil.....	13
4.2.2. Angle zénithal	13
4.2.3. Azimut du soleil	13
5. Temps du lever et du coucher-Durée d'ensoleillement maximale.....	13
6. Irradiations solaires extraterrestre.....	14
7. Algorithmes de la position solaire	15
7.1. Algorithme SPA (Solar Position Algorithm).....	15
7.2. Algorithme SG2 (Solar Geometry 2)	15
7.3. Algorithme Yallop	16
7.4. Comparaison entre les algorithmes de calcul de la position solaire.....	16
8. Rayonnement solaire au sol	17
8.1. Rayonnement solaire sur une surface horizontale.....	18
8.2. Rayonnement solaire sur une surface inclinée.....	18
9. Paramètres adimensionnels	19
9.1. Indice de clarté.....	19
9.2. Fraction diffuse.....	19

9.3.	Durée d'insolation	19
10.	Instruments de mesures	20
10.1.	Capteurs pyranométriques	20
10.1.1.	Pyranomètres à thermopile	20
10.1.2.	Pyranomètres à photodiode	21
10.2.	Pyrhéliomètres	22
10.3.	L'héliographe	22
11.	Sites d'étude et bases de données	23
12.	Contrôle de qualité et correction des données solaires	30
12.1.	Etude bibliographique sur les méthodes de contrôle de qualité	30
12.2.	Procédure et résultats du contrôle de qualité	33
12.2.1.	Premier test	34
12.2.2.	Deuxième test	34
12.2.3.	Troisième test	34
12.3.	Remplacement d'une longue série de données manquantes	40
13.	Conclusion	43

Chapitre 2 : Estimation des irradiations solaires journalières à l'aide des modèles empiriques

1.	Introduction	46
2.	Modèles empiriques d'estimation des irradiations solaires globales journalières sur plan horizontal	47
2.1.	Modèles empiriques basés sur la température ou l'humidité relative de l'air	48
2.2.	Modèles empiriques basés sur des paramètres hybrides	50
3.	Modèles empiriques d'estimation des irradiations solaires directes et diffuses journalières	55
4.	Résultats et discussions	57
4.1.	Performances des modèles empiriques d'estimation du GHI journalière	57
4.2.	Performances des modèles empiriques d'estimation journalière de DHI et DNI	67
5.	Conclusion	70

Chapitre 3 : Estimation des irradiations solaires journalières en utilisant les modèles d'apprentissage automatique

1.	Introduction	72
2.	Etat de l'art des modèles d'apprentissage automatique (Machine learning)	72
3.	Comparaison bibliographique des performances prédictives des modèles d'apprentissage automatique	78
4.	Réseaux de Neurones Artificiels (RNAs)	84
4.1.	Perceptron Multicouches (PMC)	86

4.2.	Modèles Non linéaires AutoRégressifs avec des entrées eXogènes (NARX)	88
5.	Méthodes d'ensemble	91
5.1.	Boosting	92
5.2.	Bagging	94
5.3.	Forêts aléatoires (Random Forests)	95
6.	Régression par machine à vecteur de support (SVM)	97
7.	Résultats et discussions	99
7.1.	Performances des réseaux de neurones artificiels pour l'estimation de GHI sur la base de données de la ville de Fès	99
7.1.1.	Performances du modèle PMC	100
7.1.2.	Performances du modèle NARX	101
7.1.3.	Optimisation du nombre de neurones cachés	102
7.2.	Performances des meilleures architectures obtenues des deux modèles PMC et NARX sur les bases de données enerMENA.	104
7.3.	Exploration du potentiel des méthodes d'ensemble dans l'estimation de GHI journalière.	106
7.3.1.	Procédure des modèles développés	106
7.3.2.	Performances obtenues par les méthodes d'ensemble et le modèle PMC pour l'estimation journalière de GHI	109
7.3.3.	Comparaison et analyse des performances	112
7.4.	Performances des méthodes d'ensemble pour l'estimation journalière de GHI à la ville de Fès	114
7.5.	Catégorisation des modèles d'apprentissage automatique pour l'estimation journalière des composantes solaires à la ville de Zagora	115
7.5.1.	Processus de catégorisation et évaluation des modèles	116
7.5.2.	Performances des modèles d'apprentissage automatique catégorisés	117
7.6.	Performances des meilleurs modèles catégorisés obtenus sur la base de données des autres villes.....	127
8.	Conclusion.....	128

Chapitre 4 : Prédiction à différents horizons temporels de l'irradiation solaire globale sur plan horizontal

1.	Introduction	132
2.	Modèles d'apprentissage profond (Deep Learning)	133
3.	Réseaux de neurones à longue mémoire à court terme (LSTM)	135
4.	Résultats et discussions	138
4.1.	Prédiction de l'irradiation solaire globale horizontale à court et à long termes à la ville de Fès	138

4.1.1.	Sélection de la meilleure combinaison de variables d'entrée.....	139
4.1.2.	Sélection du nombre de neurones dans la couche cachée	141
4.2.	Performances du modèle NARX sur les bases de données enerMENA.....	143
4.3.	Prédiction à l'avance des GHI horaires et sub-horaires à l'aide de la nouvelle approche d'apprentissage profond	146
4.3.1.	Sélection des données d'entrée	147
4.3.2.	Paramétrage des modèles et évaluation de la précision	148
4.3.3.	Prévisions horaires du GHI.....	150
4.3.4.	Prévisions sub-horaires du GHI.....	156
4.4.	Performances du modèle d'apprentissage profond LSTM sur les bases de données des autres villes.....	162
4.4.1.	Prédiction horaire du GHI	162
4.4.2.	Prédiction sub-horaire du GHI	164
5.	Conclusion.....	167
	Conclusion générale.....	168
	Références.....	172
	Annexe A : Algorithme SPA (Solar Position Algorithm)	183
	Annexe B: Algorithme SG2 (Solar Geometry 2)	186
	Annexe C: Algorithme Yallop.....	188
	Annexe D: Evaluation de la performance d'estimation et de prédiction.....	190

Introduction générale

Dans son dernier rapport, sur la situation mondiale des énergies renouvelables publié en 2021 (REN21, 2018), le réseau international REN21 (Renewable Energy Policy Network for the 21st Century) souligne que malgré les impacts de la pandémie de COVID-19, les énergies renouvelables (ERs) ont établi un record de 256 gigawatts (GW) de la nouvelle capacité électrique installée en 2020, la plus forte augmentation jamais enregistrée. Les ERs ont été, ainsi, la seule source de production d'électricité à enregistrer une augmentation nette de la capacité totale. En effet, et en raison des problèmes environnementaux, de l'épuisement accéléré des réserves de combustibles fossiles, de la forte hausse des prix de l'énergie, de l'augmentation soutenue de la demande en électricité due aux besoins de développement économique et à la croissance démographique, ..., de nombreux pays ont intensifié leurs efforts pour augmenter la part des ERs dans leur mix énergétique.

Dans sa politique énergétique, le Royaume du Maroc prévoit d'augmenter la part des ERs de 42% de capacité installée, fixée pour 2020, à 52% en 2030. Pour ce faire, il est prévu de développer une capacité supplémentaire de production d'électricité à partir de sources renouvelables d'environ 10.1 GW dont 4.56 GW de l'énergie solaire, 4.2 GW de l'énergie éolienne et 1.33 GW d'hydroélectricité (M. Amara, 2015). Il était également prévu d'installer 1.7 millions de m^2 de chauffe-eau solaires (1.2 GWth) en 2020 (REN21, 2018).

Grâce à sa position géographique aux latitudes moyennes (entre environ 20° N et 36° N), le Maroc bénéficie d'un potentiel considérable en énergie solaire (2600 kWh/m²/an) (A. Redouane et al, 2018). Sachant pertinemment que l'exploitation de ce potentiel lui permettrait de réduire fortement sa dépendance et sa facture énergétiques, le Maroc a mis en application un plan solaire ambitieux avec plusieurs centrales solaires NOOR (NOOR Ouarzazate (580 MW), NOOR Midelt (800 MW), NOOR Laâyoune I (85 MW), ...). L'objectif visé étant la production de 2 GW solaires à l'horizon 2020.

Néanmoins, l'intégration des centrales solaires dans le mix énergétique entraîne de sérieux problèmes dans la gestion du réseau électrique, en particulier pour maintenir son équilibre et éviter les blackouts. En effet, le caractère intermittent de l'énergie solaire reçue au sol entraîne des fluctuations conséquentes dans la production électrique des centrales solaires. Ceci pose de

sérieux problèmes car la stabilité du réseau est strictement liée à l'équilibre offre-demande d'électricité. Il est donc indispensable de disposer de données précises sur le rayonnement solaire. Ces données sont nécessaires pour la prévision du potentiel énergétique disponible et, donc, pour la production des centrales électriques.

En général, les principales composantes du rayonnement solaire au sol sont :

- Le rayonnement solaire global horizontal (GHI): quantité totale du rayonnement solaire reçue par une surface horizontale au sol.
- Le rayonnement solaire normal direct (DNI): quantité du rayonnement solaire provenant directement du soleil (sans subir de modifications lors de sa traversée de l'atmosphère terrestre) et reçue par une surface tenue perpendiculairement aux rayons du soleil.
- Le rayonnement solaire diffus horizontal (DHI): composante solaire horizontale formée par les rayons solaires diffusés par les molécules et les particules dans l'atmosphère et provenant de toute la voûte céleste.

La composante GHI est la somme de la composante DHI et de la projection horizontale du DNI.

La connaissance de ces composantes solaires est nécessaire pour dimensionner correctement les installations solaires. En fonction du système de conversion, il est nécessaire d'avoir une bonne connaissance de GHI pour les systèmes photovoltaïques et les chauffe-eau solaire et du DNI pour les systèmes solaires à concentration (CSP (centrales solaires thermodynamiques) et CPV (photovoltaïque à concentration)).

L'indisponibilité des mesures fiables de l'irradiation solaire et de ses composantes conduit à des erreurs dans la conception et le dimensionnement des systèmes solaires, ce qui entrainerait un impact négatif sur l'investissement. En revanche, il est difficile d'obtenir ces mesures dans la plupart des stations météorologiques en raison de plusieurs facteurs techniques et économiques. Par exemple au Maroc, il n'existe qu'un nombre limité de stations qui mesurent le rayonnement solaire à cause du prix élevé des appareils de mesures. En plus, même si ces mesures solaires existent, la plupart des bases de données contiennent des périodes manquantes ou des valeurs erronées dues aux pannes ou au mauvais suivi puisque la maintenance de toute station météorologique doit se faire d'une manière très minutieuse et soutenue dans le temps (nettoyage des appareils de mesure, corrections des dérèglements du système de poursuite du pyréliomètre, ...). Ce manque de données a incité plusieurs chercheurs à développer différents modèles et techniques pour prédire le rayonnement solaire pour un lieu donné à partir de variables communément mesurées dans les stations météorologiques (température et humidité

relative de l'air, vitesse et direction du vent, ...). Ces modèles sont depuis longtemps utilisés pour l'estimation du potentiel ou gisement solaire dans les sites où les mesures du rayonnement solaires ne sont pas disponibles. Une cartographie du rayonnement peut être ainsi élaborée pour une région ou un pays donnés pour aider les utilisateurs à dimensionner leurs installations solaires de production de l'électricité ou de pompage de l'eau en agriculture.

Comme nous l'avons mentionné précédemment, la gestion du réseau électrique est l'une des problématiques majeures de l'intégration de l'électricité solaire dans le mix énergétique d'un pays. En effet, pour maintenir la stabilité du réseau, il est nécessaire de connaître la consommation totale des charges ainsi que la production des sources d'énergie alimentant le réseau. Or, et à cause du caractère intermittent de l'énergie solaire, la production des centrales solaires est imprévisible. La solution à ce problème passe nécessairement par la recherche de modèles ou de techniques de prédiction à l'avance du rayonnement solaire à différents horizons temporels. En effet, pour un horizon de prédiction à long terme (d'un jour à une semaine), la ressource solaire disponible permettra d'anticiper la période pour laquelle la production d'énergie est faible afin de planifier l'entretien du système et gérer le stockage (Doucouré, 2015). En plus, ceci permettrait de négocier, à l'avance, l'achat ou la vente de l'électricité à des prix raisonnables en cas de sous ou de surproduction. Une prédiction à moyen terme (6 heures à un jour) de la ressource solaire représentera un outil primordial pour la gestion journalière du réseau, plus précisément, lorsque la demande d'énergie journalière est plus élevée (pics de demande journalière par exemple). La prédiction à court terme (30 minutes à 6 heures) de la ressource solaire permettra de prendre des décisions pour une gestion en temps-réel d'un réseau électrique. A la fin, une prédiction à très court terme (quelques secondes à 30 minutes) sert à anticiper les fluctuations de la ressource solaire pour une gestion du réseau électrique dans un temps très court (Doucouré, 2015).

Nos travaux dans le cadre de cette thèse portent sur deux champs de recherches principaux. Le premier s'intéresse à l'analyse des variabilités temporelles du rayonnement solaire au sol en estimant les différentes composantes (GHI, DHI, DNI) à l'échelle horaire et journalière. Dans notre étude, on s'est intéressé, en particulier, aux modèles mathématiques (empiriques et d'apprentissage automatique (Machine-learning)) basés uniquement sur les paramètres météorologiques couramment mesurés (température de l'air, humidité relative de l'air, vitesse du vent et précipitations pluviométriques). Le deuxième champ de recherche concerne les techniques de prédiction à l'avance de l'irradiation solaire globale horizontale (GHI) pour des horizons horaires et sub-horaires qui sont susceptibles d'intéresser un gestionnaire de réseau ou

un concepteur de réseaux électriques intelligents. Les horizons temporels considérés sont : $h+1$ à $h+6$ et $10\text{min}+10$ à $10\text{min}+60$. On se s'intéressera, en particulier, aux techniques d'apprentissage automatique (Machine-learning) et d'apprentissage profond (Deep learning), basées sur des variables endogènes (valeurs historiques de GHI). Tous les modèles ont été développés en se basant sur des données expérimentales mesurées au Maroc en six stations météorologiques de différents climats (Fès, Erfoud, Missour, Oujda, TanTan, Zagora).

Notre travail de recherche est structuré en quatre chapitres :

Le premier chapitre consiste en plusieurs étapes. Dans la première étape, il a été question de présenter les techniques de calcul du rayonnement solaire hors atmosphère à l'échelle horaire et journalier. Pour ce faire, il a été nécessaire de chercher un algorithme permettant de calculer avec précision les coordonnées solaires. Trois algorithmes (SPA, SG2 et Yallop) ont été développés et testés pour déterminer l'angle zénithal, l'azimut et la hauteur du soleil à chaque instant. Dans la troisième étape, nous avons procédé à la présentation des données expérimentales mesurées aux différents sites considérés en précisant les instruments et les périodes de mesures. A la fin, une procédure de contrôle de qualité a été appliquée aux différentes bases de données disponibles. Une procédure de remplacement des données manquantes de GHI, à partir des mesures du rayonnement photo-synthétiquement actif à la ville de Fès, sera aussi présentée.

Dans le second chapitre, nous présentons, en premier lieu, un état de l'art sur les différents modèles empiriques largement utilisés dans la modélisation des différentes composantes du rayonnement solaire incident au sol. A cet effet, nous nous sommes intéressés, en particulier, aux modèles basés sur des variables météorologiques couramment mesurées et des facteurs géographiques et/ou astronomiques connus ou facilement calculables. Nous présenterons, aussi, quelques modèles que nous avons, nous-mêmes, proposés et testés dans le présent travail. Les principaux résultats trouvés pour estimer les GHI, DHI et DNI journalières seront ensuite présentés et analysés.

Le troisième chapitre est consacré à la présentation et aux tests des différentes techniques d'apprentissage automatique (Machine-learning) utilisées durant notre travail de recherche. Nous commençons d'abord par la présentation d'un état de l'art pour évaluer l'efficacité de quelques modèles de la littérature. Ensuite, une description détaillée des modèles basés sur les Réseaux de Neurones Artificiels (RNAs), les Méthodes d'Ensemble (ME) et la Régression par Machine à Vecteur de Support (SVM) est effectuée. Les performances des modèles développés

pour l'estimation des GHI, DHI et DNI journalières en utilisant les variables couramment mesurées dans les stations météorologiques et quelques paramètres astronomiques facilement calculables sont, alors, présentées, analysées et discutées.

Dans le quatrième chapitre, nous présentons les résultats de la prédiction en avance de l'irradiation solaire globale horizontale (GHI) pour différents horizons temporels. On exposera d'abord une étude bibliographique sur les différentes techniques d'apprentissage profond (Deep learning) existantes pour prédire les irradiances solaires. La méthodologie de prédiction adoptée est, ensuite, décrite. Dans ce chapitre, deux études ont été effectuées. Une première étude pour prédire les GHI horaires à court ($h+1$) et à long termes ($h+48$) en appliquant deux types de réseaux de neurones artificiels (RNAs). Une deuxième étude permettant de prédire les GHI horaires et sub-horaires à l'aide du nouveau modèle d'apprentissage profond (LSTM), sera présentée et ses performances comparées avec celles de deux méthodes d'apprentissage automatique PMC (Perceptron Multi-Couches) et RF (Random Forests). Pour un horizon temporel allant de $h+1$ à $h+6$ et de $10\text{min}+10$ à $10\text{min}+60$, nous avons proposé deux scénarios de prédiction, un scénario des performances annuelles et un scénario des performances saisonnières.

Enfin, une conclusion générale est proposée pour récapituler les principaux résultats trouvés et dégager une synthèse sur les différents modèles de prédiction utilisés. Suite aux différentes études effectuées durant notre thèse, nous présenterons, à la fin, les perspectives envisageables pour la poursuite des présents travaux de recherche.

Chapitre 1

Rayonnement solaire : mesures expérimentales
et contrôle de qualité

1. Introduction

Le rayonnement solaire reçu au sol dépend de nombreux paramètres tels que, la position géographique du site, les conditions météorologiques (turbidité atmosphérique, nuages, ...), les grandeurs astronomiques (saison), l'inclinaison et l'orientation de la surface, l'heure de la journée et la hauteur du soleil.

L'énergie solaire disponible au sol est un facteur très important dans de nombreux domaines : prévision de la production des systèmes de conversion solaires photovoltaïques ou thermiques, prévisions météorologiques et climatiques, efficacité énergétique dans les bâtiments, agriculture et environnement, Une connaissance précise du rayonnement solaire incident au sol est nécessaire pour concevoir et réaliser des études technico-économiques afin d'évaluer la rentabilité de tout projet d'installation solaire. Aussi, un modèle correct de prévision du rayonnement solaire permettrait d'élaborer des stratégies d'optimisation et de contrôle des centrales solaires et la gestion optimale du mix énergétique.

Ce chapitre est consacré à la description des notions de base sur le rayonnement solaire : ses caractéristiques, ses composantes et les différents paramètres géographiques et astronomiques qui peuvent influencer sur l'énergie solaire incidente au sol. A la fin, on présentera les sites d'étude et les instruments de mesures des données solaires et météorologiques utilisés ainsi que les techniques de contrôle de qualité de ces mesures.

2. Rayonnement solaire

Le soleil émet une énergie sous forme d'ondes électromagnétiques qui se propagent à la vitesse de la lumière. La distribution spectrale du rayonnement solaire contient une partie visible, située au domaine de longueurs d'onde λ comprises entre 400 nm et 770 nm (~ 48%), une partie dans le domaine ultraviolet ($\lambda < 400$ nm) (~ 8%) et une partie infrarouge ($\lambda > 770$ nm) (~ 42%). La distribution spectrale du rayonnement solaire extraterrestre ressemble approximativement à la distribution d'un corps noir à une température égale à 5780 K (Figure 1.1). L'énergie solaire reçue par la surface de la terre varie en fonction de plusieurs facteurs au cours de la journée et dépend des saisons, de la situation géographique, du climat, de l'angle d'incidence et de l'orientation des surfaces.

2.1. Constante solaire

La constante solaire est définie comme étant l'éclairement solaire reçue, hors atmosphère, par une surface de 1 m^2 perpendiculaire au rayonnement solaire et placée à la distance moyenne Terre-Soleil. D'après les résultats de mesures faites entre 1969 et 1980 (Bessemoulin and Oliiviéri, 2000), la valeur de la constante solaire a été estimée à 1367 W/m^2 par Claus Fröhlich et Christoph Wehrli (Fröhlich, 1977) au Centre radiométrique mondial de Davos (Suisse) et adoptée par la Commission des instruments et des méthodes d'observation (CIMO) de l'Organisation météorologique mondiale (OMM) en octobre 1981 (Bessemoulin and Oliiviéri, 2000). La constante solaire peut être mesurée à partir des satellites situés dans la partie supérieure de l'atmosphère. Sa valeur varie légèrement, au cours de l'année, en fonction de la position de la Terre par rapport au Soleil (Trouillet and Vallée, 2008).

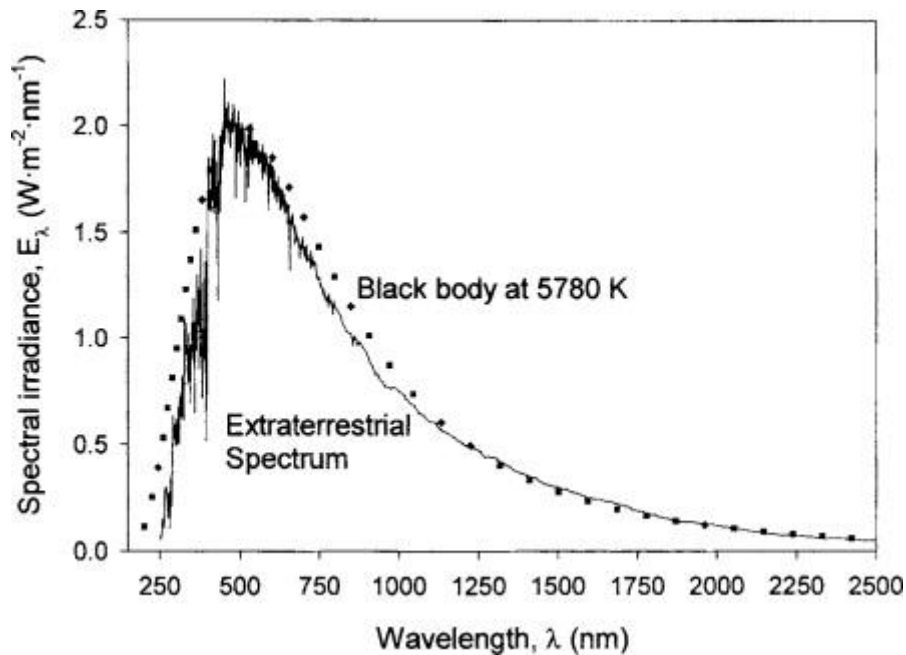


Figure 1.1. Distribution spectrale du rayonnement solaire (Seidlitz et al., 2001).

2.2. Eclairement solaire extraterrestre

L'éclairement solaire extraterrestre [W/m^2] est la puissance du rayonnement solaire reçu à la limite supérieure de l'atmosphère terrestre par unité de surface. En absence de l'atmosphère terrestre, l'éclairement d'une surface normale aux rayons solaires est donné par la relation suivante :

$$I_{0n} = I_0 \left[1 + 0.033 \cos \left(\frac{360 J}{365} \right) \right] \quad (1-1)$$

où I_0 est la constante solaire égale à 1367 W/m^2 et J est le numéro du jour de l'année compté à partir du 1^{er} Janvier.

Pour une surface d'inclinaison quelconque, l'éclairement solaire reçu dépend de l'angle d'incidence θ des rayons solaires :

$$I_{0i} = I_{0n} \cos(\theta) \quad (1-2)$$

I_{0i} peut être calculée connaissant la position géographique du lieu considéré, le jour de l'année et l'heure de la journée. En effet, l'angle d'incidence des rayons solaires sur une surface d'inclinaison quelconque en un lieu donné de la surface de la terre peut être déterminée, avec précision, grâce à quelques relations astronomiques (Voir section 4.2.2).

3. Coordonnées géographiques terrestres

La position géographique d'un lieu sur terre est déterminée par sa latitude, sa longitude et son altitude.

3.1. Latitude ϕ

La latitude ϕ est une coordonnée géographique qui représente l'angle d'inclinaison d'un point sur la surface de la terre par rapport au plan de l'équateur (Figure 1.2). C'est une mesure angulaire qui varie entre -90° et 90° . Il est compté positivement vers le pôle Nord et négativement vers le pôle Sud.

3.2. Longitude L

En géométrie solaire, la longitude L d'un lieu à la surface de la terre représente l'angle que fait le méridien passant par le lieu considéré avec le méridien de Greenwich qui est le méridien de référence de longitude 0° (Figure 1.2). La longitude varie entre -180° et 180° . Il est compté positivement vers l'est et négativement vers l'ouest.

3.3. Altitude

L'altitude est l'élévation verticale d'un lieu par rapport au niveau moyen de la mer, mesurée en mètre (m). Plus l'altitude d'un lieu est importante plus la pression atmosphérique diminue.

4. Coordonnées solaires

Pour un observateur en un lieu quelconque à la surface de la terre, la position du soleil dans le ciel peut être déterminée à l'aide de deux systèmes de coordonnées : coordonnées équatoriales ou coordonnées horizontales.

4.1. Coordonnées équatoriales

Dans ce système de coordonnées, la position du soleil est déterminée en utilisant deux coordonnées angulaires : la déclinaison solaire δ et l'angle horaire ω (Figure 1.2).

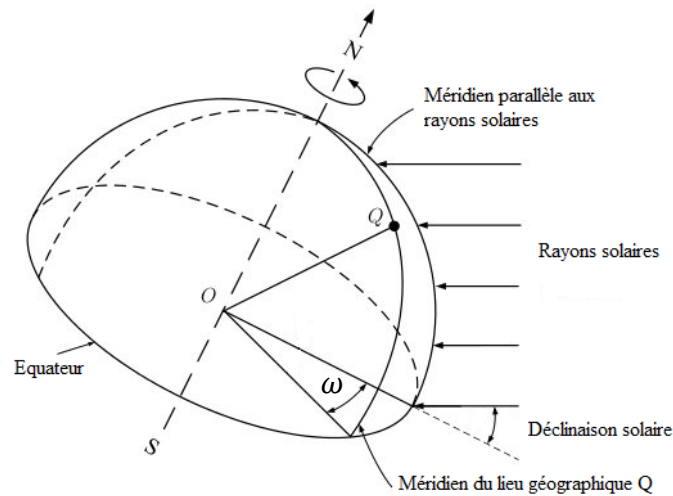


Figure 1.2. Coordonnées équatoriales.

4.1.1. Déclinaison solaire

La déclinaison solaire δ représente l'angle d'inclinaison des rayons solaires par rapport au plan de l'équateur terrestre. Au cours de l'année, cet angle varie entre $-23,45^\circ$ (solstice d'hiver) et $23,45^\circ$ (solstice d'été) et s'annule aux équinoxes (Figure 1.3). Pour un site donné, la hauteur du soleil varie en fonction de la variation de la déclinaison solaire. Par exemple en été, à l'hémisphère nord, le soleil est plus haut sur l'horizon par rapport à l'hiver pour une même heure de la journée, ce qui entraîne une longue durée de la journée et, donc, une longue durée d'insolation et une quantité d'irradiation journalière plus importante.

Plusieurs formules mathématiques ont été développées pour exprimer approximativement la déclinaison solaire δ . L'équation adoptée pour nos calculs est celle décrite ci-dessous :

$$\delta = 23.45 \sin\left(\frac{360}{365} (J + 284)\right) \quad (1-3)$$

où J est le numéro du jour de l'année compté à partir du premier janvier.

Dans la relation ci-dessus, la déclinaison solaire est considérée comme constante sur une journée.

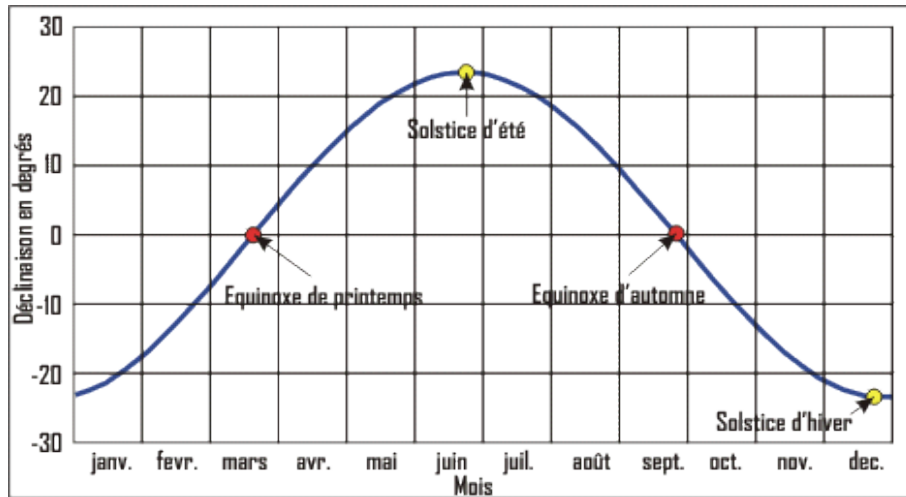


Figure 1.3. Variation de la déclinaison solaire au cours de l'année.

4.1.2. Angle horaire

L'angle horaire ω désigne l'angle entre le méridien du lieu géographique considéré et le méridien de la position actuelle du soleil. ω (en degrés) est exprimé en fonction du temps solaire vrai TSV, à l'aide de la relation suivante :

$$\omega = 15 (\text{TSV} - 12) \quad (1-4)$$

On peut noter que, dans sa trajectoire apparente autour de la terre, le soleil parcourt 15° pendant une heure. L'angle horaire ω est négatif le matin lorsque le soleil est vers l'Est et positif l'après-midi lorsque le soleil est vers l'Ouest. Il vaut 0° à midi TSV lorsque le méridien du soleil coïncide avec le méridien du lieu considéré. Pour cette heure, le soleil atteint son point le plus haut de sa trajectoire journalière dans le ciel.

Le temps solaire vrai (TSV), en heure, peut être calculé à l'aide de la relation (Kalogirou, 2013):

$$\text{TSV} = \text{TL} + \frac{E_t}{60} \pm \frac{(L_{st} - L)}{15} - \Delta h \quad (1-5)$$

où TL est le temps légal, indiqué par les montres, L est la longitude du lieu considéré, L_{st} est le méridien standard du fuseau horaire local. Les longitudes sont exprimées en degrés ouest, c'est-à-dire entre 0° et 360° . L_{st} est donné par la relation (Zobaa and Bansal, 2011):

$$LSM = 15^\circ \times int\left(\frac{L}{15^\circ}\right) \quad (1-6)$$

La fonction int représente la partie entière.

Le signe algébrique \pm dans l'éq. (1-5) doit être pris comme positif pour les L situées à l'Est de L_{st} et vice versa.

Δh est l'heure d'été. Pour le Maroc, $\Delta h = 1$ sur toute l'année.

L'équation du temps (E_t) est exprimée par la formule suivante :

$$E_t = -9.87 \sin(2N) + 7.53 \cos(N) + 1.5 \sin(N) \quad (1-7)$$

Où N est défini en fonction du numéro du jour de l'année J :

$$N = \frac{2\pi}{365} (J - 81) \quad (1-8)$$

4.2. Coordonnées horizontales

Ces coordonnées ont comme plans de référence, le plan horizontal du lieu d'observation. La position du soleil dans le ciel est définie par deux coordonnées qui sont la hauteur h , l'angle zénithal du soleil O_z et l'azimut solaire A_z (Figure 1.4).

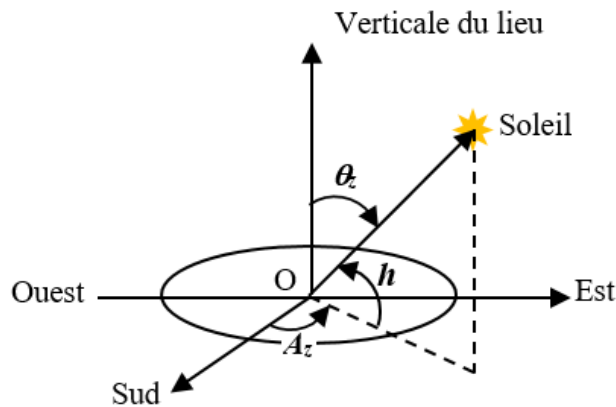


Figure 1.4. Coordonnées solaires horizontales.

4.2.1. Hauteur du soleil

La hauteur h du soleil, appelé aussi élévation ou altitude solaire, est définie par l'angle que forme le vecteur observateur local-soleil avec le plan horizontal du lieu considéré. Cette hauteur varie de 0° à 90° (de l'horizon au zénith pour l'hémisphère Nord) et de 0° à -90° (de l'horizon au nadir pour l'hémisphère Sud). Pour un jour et un site géographique donnés, la hauteur solaire est à sa valeur maximale à midi TSV et nulle au lever et au coucher du soleil.

La hauteur solaire varie au cours de la journée en fonction de l'angle horaire ω , de la déclinaison solaire δ et de la latitude ϕ du lieu géographique. Elle est donnée par la relation suivante :

$$\sin(h) = \sin(\delta)\sin(\phi) + \cos(\delta)\cos(\phi)\cos(\omega) \quad (1-9)$$

4.2.2. Angle zénithal

L'angle zénithal θ_z est l'angle que fait la verticale en un point donné avec la direction du Soleil. C'est l'angle d'incidence des rayons solaires sur une surface horizontale. Sa valeur varie entre 0 et 90° . Un angle de 0° indique que le Soleil est au zénith, alors qu'un angle de 90° indique que le Soleil est à l'horizon. θ_z est le complément à 90° de la hauteur solaire. Il est donné par la relation :

$$\cos(\theta_z) = \sin(h) \quad (1-10)$$

4.2.3. Azimut du soleil

L'azimut solaire A_z est l'angle compris entre la projection de la direction du soleil sur le plan horizontal et la direction sud du lieu géographique considéré. A_z est compté négativement vers l'Est et positivement vers l'Ouest. Sa valeur varie entre -180 et 180° . L'azimut solaire A_z peut être calculé à l'aide de la relation suivante :

$$\sin(A_z) = \frac{\cos(\delta)\sin(\omega)}{\cos(h)} \quad (1-11)$$

5. Temps du lever et du coucher-Durée d'ensoleillement maximale

Le temps du lever ($t_{SV\text{lever}}$) et du coucher ($t_{SV\text{coucher}}$) du soleil correspondent à une hauteur solaire (h) nulle. Ils peuvent être obtenus à partir de l'équation (1-9) en prenant $h = 0$:

$$t_{SVlever} = 12 - \arccos(-tg(\delta)tg(\phi))/15 \quad (1-12)$$

$$t_{SVcoucher} = 12 + \arccos(-tg(\delta)tg(\phi))/15 \quad (1-13)$$

La durée d'ensoleillement théorique (durée pendant laquelle le soleil est visible dans le ciel) est donnée par la relation :

$$S_0 = t_{SVcoucher} - t_{SVlever} = 2 \arccos(-\tan \phi \tan \delta) / 15 \quad (1-14)$$

6. Irradiations solaires extraterrestre

L'éclairement solaire reçu par une surface horizontale hors atmosphère peut être calculé à partir de la relation :

$$I_{0h}[W/m^2] = I_{0n} \sin(h) = I_0[\sin(\phi)\sin(\delta) + \cos(\phi)\cos(\delta)\cos(\omega)] \quad (1-15)$$

Entre deux instants donnés d'angles horaires ω_1 et ω_2 ($\omega_2 > \omega_1$), l'irradiation solaire [Wh/m^2] d'une surface horizontale hors atmosphère peut être obtenue en intégrant l'équation ci-dessus entre ω_1 et ω_2 :

$$I_{0ext}[Wh/m^2] = \frac{12}{\pi} I_o \left(1 + 0.033 \cos\left(\frac{360}{365} J\right) \right) \left[\cos(\phi)\cos(\delta)(\sin(\omega_2) - \sin(\omega_1)) + \frac{\pi(\omega_2 - \omega_1)}{180} \sin(\phi)\sin(\delta) \right] \quad (1-16)$$

En particulier, les irradiations horaire I_{0h} et journalière H_0 d'une surface horizontale hors atmosphère sont données par les relations suivantes :

$$I_{0h}[Wh/m^2] = \frac{12}{\pi} I_o \left(1 + 0.033 \cos\left(\frac{360}{365} J\right) \right) \left[\cos(\phi)\cos(\delta)(\sin(\omega_2) - \sin(\omega_1)) + \frac{\pi}{12} \sin(\phi)\sin(\delta) \right] \quad (1-17)$$

$$H_0[Wh/m^2] = \frac{24}{\pi} I_o \left(1 + 0.033 \cos\left(\frac{360}{365} J\right) \right) \left[\cos(\phi)\cos(\delta)\sin(\omega_s) + \frac{\pi\omega_s}{180} \sin(\phi)\sin(\delta) \right] \quad (1-18)$$

où I_0 est la constante solaire et ω_s est l'angle horaire du coucher du soleil (en degrés) pour le jour J considéré.

L'irradiations horaire I_{0h} peut aussi être calculée à l'aide de la relation (Iqbal, 2012) :

$$I_{0h}[Wh/m^2] = I_o \left(1 + 0.033 \cos\left(\frac{360}{365} J\right) \right) \left[\cos(\phi)\cos(\delta)\cos(\omega_i) + \sin(\phi)\sin(\delta) \right] \quad (1-19)$$

où ω_i est l'angle horaire évalué au milieu de ω_1 et ω_2 ($\omega_i = \omega_1 + \frac{15}{2}$, en degrés (le soleil parcourt 15° par heure dans son mouvement apparent autour de la terre)).

7. Algorithmes de la position solaire

Un calcul précis du rayonnement solaire hors atmosphère nécessite une connaissance précise de la position solaire et donc de l'angle zénithal et de l'azimut du soleil. Plusieurs algorithmes permettant d'effectuer ces calculs existent dans la littérature, mais avec un degré d'incertitude et une validité temporelle différents (certains algorithmes ne sont valables que pour un certain nombre d'années bien défini). A titre d'exemple, l'algorithme de Michalsky (Michalsky, 1988) est limité à une période allant de 1950 à 2050 et donne la position solaire avec une incertitude supérieure à ± 0.01 . L'algorithme proposé par Blanco-Muriel et al. (Blanco-Muriel et al., 2001) présente, quant à lui, une validité temporelle limitée à la période entre 1999 et 2015.

Dans cette thèse, trois algorithmes ont été testés et comparés pour calculer la position du soleil à chaque instant : l'algorithme SPA (Solar Position Algorithm), l'algorithme SG2 (Solar Geometry 2) et l'algorithme Yallop.

7.1. Algorithme SPA (Solar Position Algorithm)

L'algorithme SPA (Solar Position Algorithm) permet de calculer les zénith et azimut solaires avec des incertitudes de l'ordre de $\pm 0,0003^\circ$ pour la période allant de l'année -2000 jusqu'à 6000. La procédure mathématique de cet algorithme est définie par des équations astronomiques qui sont basées sur les variations séculaires des orbites planétaires (VSOP), développés par Bretagnon en 1982 (Bretagnon, 1982), puis modifié en 1987 par Bretagnon et Francou (Bretagnon and Francou, 1988). L'algorithme SPA a été développé par Reda and Andreas (Reda and Andreas, 2004) en modifiant le "The Astronomical Algorithms", basés sur le VSOP87, et développés par Meeus (Meeus, 1998), pour les adapter à la géométrie solaire. Le tableau en Annexe A représente la procédure de calcul de l'algorithme SPA.

7.2. Algorithme SG2 (Solar Geometry 2)

SG2 (Solar Geometry 2) est un algorithme proposé par Blanc et Wald (Blanc and Wald, 2012) pour déterminer la position du soleil dans le ciel avec une bonne précision sur la période 1980-2030. Le SG2 donne une précision inférieure par rapport à l'algorithme SPA, mais son avantage est qu'il est plus rapide en exécution. En effet, le SG2 est obtenu en appliquant plusieurs approximations aux équations originales de l'algorithme SPA pour diminuer le nombre d'opérations. Toutefois, la précision obtenue dans les calculs des azimut et élévation solaires est de l'ordre de $\pm 0.0015^\circ$, ce qui est inférieur à la précision $\pm 0.01^\circ$ requise pour l'étalonnage des pyranomètres et les suiveurs solaires ((Reda and Andreas, 2004), (Stafford et al., 2009)).

Le tableau en Annexe B montre les formules utilisées la procédure de calcul de l'algorithme SG2.

7.3. Algorithme Yallop

L'algorithme de Yallop (1992) est connu par sa haute précision dans le calcul de l'équation du temps et de la déclinaison, l'azimut, le zénith et l'élévation solaires. Cet algorithme est valable pour une période qui s'étend de 1980 à 2050. La précision de cet algorithme est telle que l'erreur absolue maximale de déclinaison est de $\pm 0.005^\circ$ dans le calcul de la déclinaison solaire et $\pm 0.008^\circ$ dans le calcul de l'angle horaire de Greenwich (Toledo et al., 2020). Le tableau en Annexe C résume la procédure de calcul de l'algorithme Yallop.

7.4. Comparaison entre les algorithmes de calcul de la position solaire

Pour comparer les résultats des trois algorithmes de la position solaire, nous avons tracé pendant une journée la variation de l'azimut solaire, l'angle zénithal, et la hauteur solaire calculés respectivement par l'algorithme Yallop, SPA, et SG2. D'après les tableaux 1.1, 1.2, et 1.3, On constate que les évolutions de l'azimut solaire, l'angle zénithal, et la hauteur solaire sont les mêmes pour les trois algorithmes pendant la journée du 01/05/2014 à la ville de Fès. De plus, le tableau 1.3 montre que la hauteur du soleil arrive à sa valeur maximale à midi solaire et atteint sa valeur minimale au temps du lever et du coucher du soleil.

Tableau 1.1. Variations horaires de l'azimut solaire au cours de la journée du 1 Mai 2014 à Fès.

Heure	5	6	7	8	9	10	11
Yallop	66.6498	75.1423	83.0885	91.243	100.553	112.9502	132.6489
SPA	66.6592	75.1383	83.0897	91.2327	100.5632	112.9467	132.6573
SG2	66.6593	75.1384	83.0897	91.2327	100.5632	112.9467	132.6574
Heure	12	13	14	15	16	17	18
Yallop	167.4576	209.8478	237.3005	253.0642	263.816	272.4761	280.4567
SPA	167.4596	209.8561	237.2915	253.0654	263.8105	272.4815	280.4561
SG2	167.4601	209.8568	237.2921	253.0658	263.8108	272.4818	280.4563

Tableau 1.2. Variations horaires du zénith solaire au cours de la journée du 1 Mai 2014 à Fès.

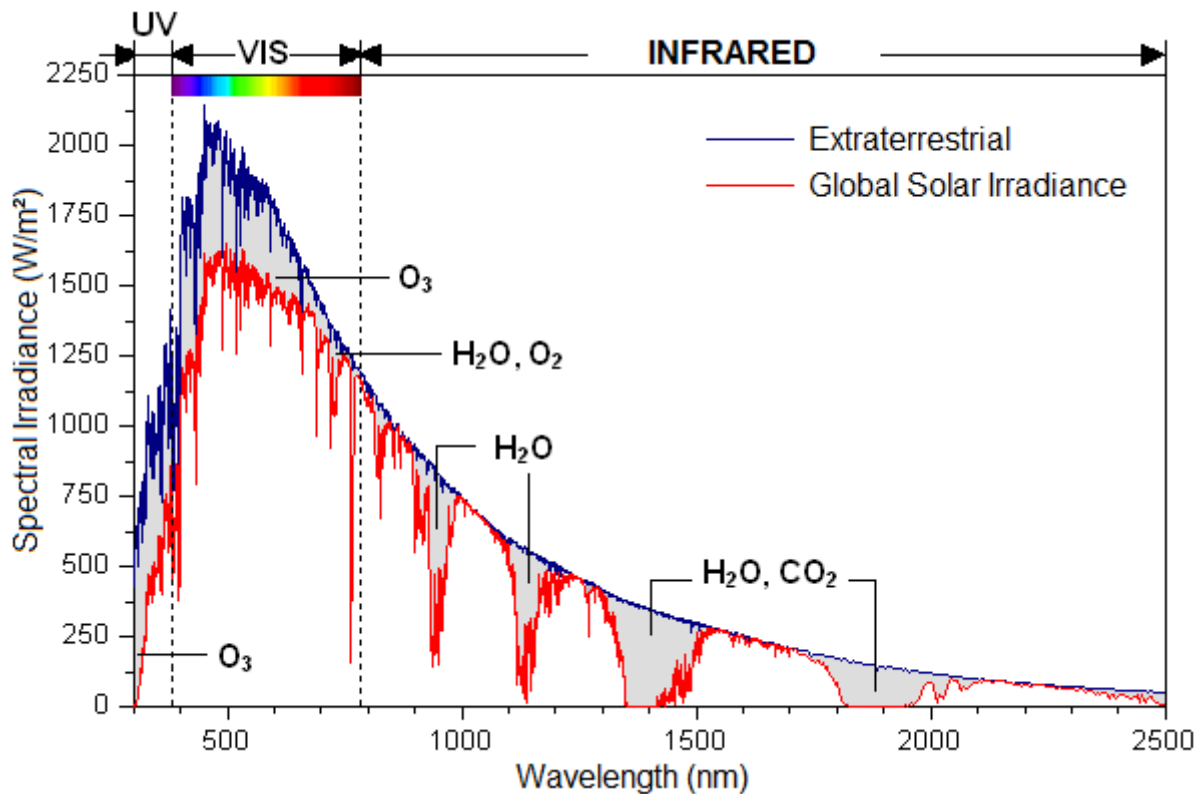
Heure	5	6	7	8	9	10	11
Yallop	96.8336	85.0798	72.8691	60.4521	48.0848	36.1685	25.6527
SPA	96.8322	85.0784	72.8677	60.4505	48.0829	36.1663	25.6501
SG2	96.8321	85.0783	72.8675	60.4503	48.0827	36.1661	25.6499
Heure	12	13	14	15	16	17	18
Yallop	19.2519	21.2003	29.8901	41.1902	53.3610	65.7736	78.1143
SPA	19.2498	21.2001	29.8916	41.1926	53.3639	65.7768	78.1175
SG2	19.2496	21.2000	29.8916	41.1926	53.3640	65.7769	78.1176

Tableau 1.3. Variations horaires de la hauteur du soleil au cours de la journée du 1 Mai 2014 à Fès.

Heure	5	6	7	8	9	10	11
Yallop	0	4.9202	17.1309	29.5479	41.9152	53.8315	64.3473
SPA	0	4.9216	17.1323	29.5495	41.9171	53.8337	64.3499
SG2	0	4.9217	17.1325	29.5497	41.9173	53.8339	64.3501
Heure	12	13	14	15	16	17	18
Yallop	70.7481	68.7997	60.1099	48.8098	36.6390	24.2264	11.8857
SPA	70.7502	68.7999	60.1084	48.8074	36.6361	24.2232	11.8825
SG2	70.7504	68.8000	60.1084	48.8074	36.6360	24.2231	11.8824

8. Rayonnement solaire au sol

Avant d'arriver au sol, le rayonnement solaire subit des interactions avec les molécules gazeuses et les particules présentes dans l'atmosphère. Ces interactions provoquent des atténuations du rayonnement solaire par trois phénomènes principaux : La réflexion, l'absorption et la diffusion. La quantité de rayonnement solaire réfléchi vers l'espace dépend des conditions atmosphériques. Elle est plus forte en présence des nuages. Le rayonnement solaire peut aussi être absorbé principalement par les molécules gazeuses, en raison de la présence de l'ozone (absorption des rayons ultraviolets nocifs), de la vapeur d'eau et du dioxyde de carbone (Figure 1.5).



Lorsque le rayonnement solaire n'est ni réfléchi, ni absorbé, il peut être soit transmis au sol (rayonnement solaire direct) soit diffusé par les molécules, les poussières et les aérosols en suspension dans l'atmosphère. Ce phénomène de diffusion dépend de plusieurs facteurs : la longueur d'onde du rayonnement, la densité et la taille des particules et des molécules atmosphériques et l'épaisseur de la couche atmosphérique traversée.

En général, le rayonnement solaire reçu par la surface de la terre varie en fonction de la position du soleil et des conditions météorologiques.

8.1. Rayonnement solaire sur une surface horizontale

Le rayonnement solaire reçu par une surface horizontale au sol est la somme de deux composantes :

- **Le rayonnement solaire direct** : L'éclairement le plus puissant provenant directement du soleil lorsque le ciel est clair sans subir de diffusion par l'atmosphère. Il dépend de la hauteur du soleil et de l'angle d'inclinaison et d'orientation de la surface considérée. Le rayonnement solaire direct est la seule composante utile pour les systèmes solaires à concentration (seul le rayonnement solaire direct peut être concentré par les systèmes actuels).
- **Le rayonnement solaire diffus** : il provient des multiples interactions entre une partie du rayonnement solaire et les constituants de l'atmosphère. Il représente la lumière du jour qui nous permet de voir même quand le temps est couvert.

Pour une surface horizontale au sol, la somme des composantes solaires directe et diffuse constitue ce que l'on appelle le rayonnement solaire global. Cette composante peut s'écrire sous la forme :

$$I_h = I_d + I_{bh} = I_d + I_{bn} \sin(h) \quad (1-20)$$

I_d et I_{bh} représentent respectivement les composantes solaires diffuse et directe horizontales. I_{bn} est la composante solaire directe reçue par unité de surface normale aux rayons et h est la hauteur du soleil.

8.2. Rayonnement solaire sur une surface inclinée

Le rayonnement solaire global reçu par une surface inclinée peut être estimé à l'aide de la relation suivante (Liu and Jordan, 1961):

$$I_{\beta} = R_b I_{bn} + R_d I_d + \rho R_r I_h \quad (1-21)$$

où R_b , R_d et R_r sont les facteurs de conversions pour les composantes directe, diffuse et réfléchiée, respectivement. Dans les conditions isotropes, ces coefficients sont donnés par les relations (Liu and Jordan, 1961):

$$R_b = \frac{I_{b\beta}}{I_{bh}} = \frac{\cos(\theta)}{\cos(\theta_z)} \quad (1-22)$$

$$R_d = \frac{1 + \cos(\beta)}{2} \quad (1-23)$$

$$R_r = \frac{1 - \cos(\beta)}{2} \quad (1-24)$$

où $I_{b\beta}$ est la composante du rayonnement solaire direct sur une surface inclinée d'un angle β par rapport à l'horizontale, θ est l'angle d'incidence des rayons solaires sur la surface inclinée, θ_z est l'angle zénithal ou l'angle d'incidence des rayons solaires sur une surface horizontale et ρ est le coefficient de réflexion ou albédo de la surface avoisinant notre surface inclinée.

9. Paramètres adimensionnels

9.1. Indice de clarté

L'indice de clarté K_t est défini comme étant le rapport entre le rayonnement solaire global au sol G et le rayonnement solaire extraterrestre H_0 , reçus par un plan horizontal :

$$K_t = \frac{G}{H_0} \quad (1-25)$$

9.2. Fraction diffuse

Le rapport entre le rayonnement solaire diffus D et le rayonnement solaire global G reçus par un plan horizontal est appelé fraction diffuse K_d qui est définie par la relation suivante :

$$K_d = \frac{D}{G} \quad (1-26)$$

9.3. Durée d'insolation

Pour une position géographique donnée, la durée d'insolation ou d'ensoleillement S est définie comme étant la durée pendant laquelle le soleil est visible dans le ciel et le rayonnement solaire

direct sur plan normal est supérieur à 120 W/m^2 selon l'OMM (Organisation Météorologique Mondiale).

En absence de l'atmosphère, la durée d'insolation théorique S_0 est définie comme étant la durée du jour et est obtenue à partir de l'équation suivante :

$$S_0 = \frac{2}{15} \cos^{-1}(-\tan(\phi) \tan(\delta)) \quad (1-27)$$

où ϕ et δ représentent respectivement, la latitude d'un lieu et la déclinaison solaire.

La fraction d'ensoleillement (SF) est définie comme le rapport entre les durées d'ensoleillement réelles (S) et théoriques (S_0) :

$$SF = \frac{S}{S_0} \quad (1-28)$$

10. Instruments de mesures

Le dimensionnement et la simulation de tout système solaire nécessitent la disponibilité de mesures fiables du rayonnement solaire à différentes résolutions temporelles. Les instruments de mesures les plus utilisés dans le domaine du rayonnement solaire sont : les pyranomètres pour mesurer le rayonnement global, les pyrhéliomètres pour mesurer le rayonnement direct et les héliographes pour déterminer la durée d'ensoleillement.

10.1. Capteurs pyranométriques

10.1.1. Pyranomètres à thermopile

Un pyranomètre à thermopile est un capteur conçu pour mesurer une large bande du flux de rayonnement solaire incident sur une surface plane sous un angle solide de 180 sr (stéradian). Ce pyranomètre se compose essentiellement d'un capteur thermopile qui possède une réponse spectrale constante pour le spectre solaire reçu au sol. La première génération des pyranomètres à thermopile avait la jonction chaude du capteur divisée en deux secteurs identiques (noirs et blancs). La mesure différentielle entre la température des secteurs noirs et blancs fournit la valeur de l'irradiation solaire globale. Pour toutes les technologies à thermopiles, l'irradiation est proportionnelle à la différence entre la température du secteur exposé au soleil et la température du secteur ombré ("Thermopile Pyranometer calibration by ITC India Pvt Ltd," 2019). La figure 1.6 montre quelques types de pyranomètres à thermopiles existants au marché.

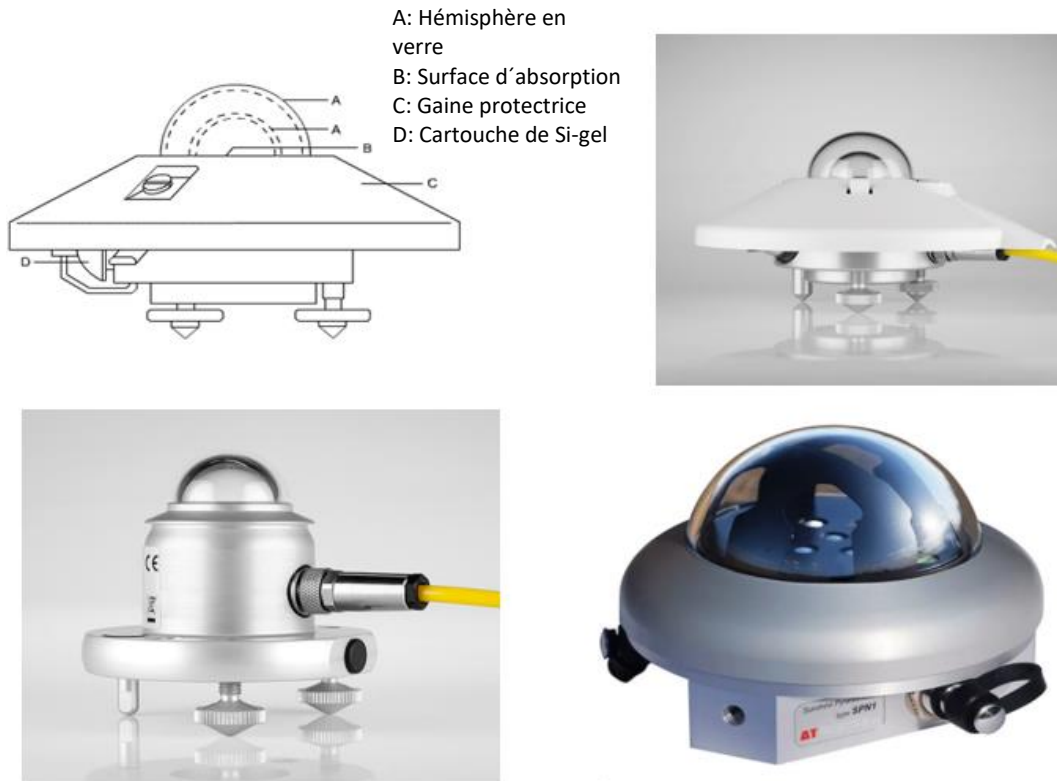


Figure 1.6. Pyranomètres à thermopile.

10.1.2. Pyranomètres à photodiode

Un pyranomètre à photodiode est un capteur qui permet de mesurer le rayonnement solaire global (400nm-1100nm) en utilisant une photodiode au silicium. Un pyranomètre à photodiode en silicium SP Lite 2 possède un diffuseur conique qui fournit une excellente réponse directionnelle, qui permet à la pluie de s'écouler pour un entretien et un nettoyage facile. Pour une journée claire et dégagée, le rendement de SP Lite 2 est comparé favorablement à un pyranomètre à thermopile ISO9060 seconde classe, même si sa gamme spectrale est limitée par le détecteur à photodiode (Brochure of the SP Lite2 Silicon Pyranometer - Kipp & Zonen, 2014). Le principe de fonctionnement de la Si-photodiode est le même qu'un semi-conducteur pour transformer l'onde solaire en signal électrique. La photodiode génère un signal de sortie proportionnel à la quantité de la radiation solaire reçue. Les types des pyranomètres à photodiode existants au marché sont présentés à la figure 1.7.

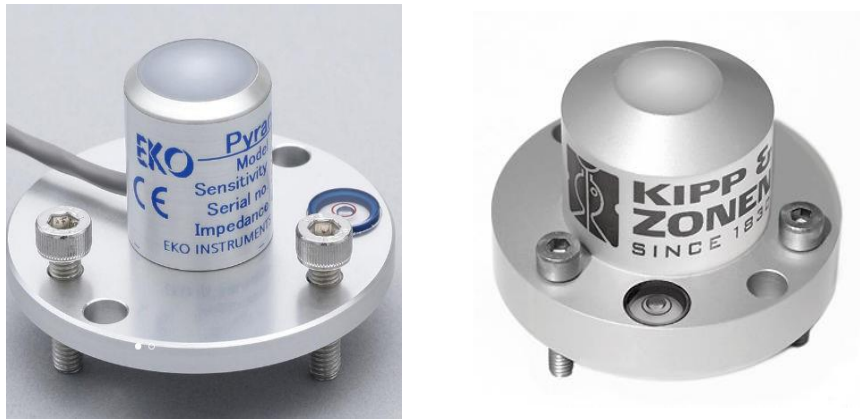


Figure 1.7. Pyranomètres à photodiode.

10.2. Pyrhéliomètres

Le pyrhéliomètre est destiné à la mesure de l'éclairement solaire direct. Il est monté sur un système de poursuite solaire pointant, à chaque instant de la journée, vers le soleil (Figure 1.8 à gauche). L'élément sensible du pyrhéliomètre est généralement constitué par une thermopile noire très absorbante. La lumière arrivant du soleil pénètre l'instrument par une fenêtre et ensuite dirigée sur la thermopile qui convertit la chaleur en un signal électrique qui est enregistré et converti en une mesure équivalente exprimée en W/m^2 .

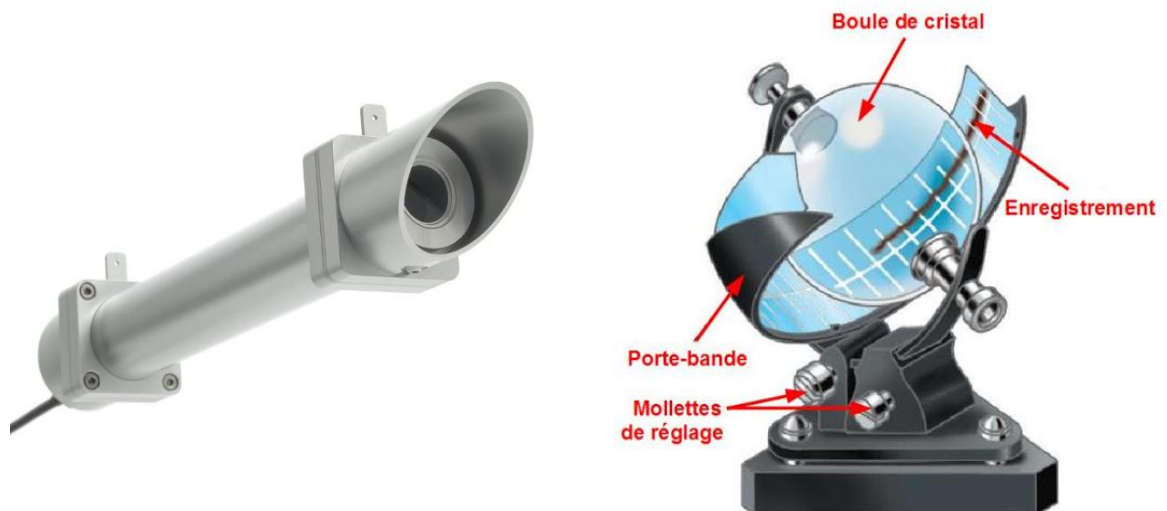


Figure 1.8. Un pyrhéliomètre (à gauche) et un héliographe (à droite).

10.3. L'héliographe

L'héliographe est un ancien appareil qui permet de mesurer la durée d'insolation pendant une journée. Il se compose d'une sphère de cristal qui joue le rôle d'une loupe pour focaliser le

rayonnement solaire (Figure 1.8 à droite). A l'opposé du soleil, une bande graduée en papier (héliogramme) traitée contre l'inflammation et l'humidité est placée à la distance focale de la sphère. Cette bande se consume et sa brûlure progresse avec le déplacement apparent du soleil pendant toute la journée (tant que le soleil brille) ce qui permet de mesurer la durée quotidienne d'insolation (Principe de la mesure du rayonnement solaire incident, 2021).

11. Sites d'étude et bases de données

Des mesures fiables du rayonnement solaire ne sont pas facilement disponibles dans la plupart des endroits du monde, en particulier dans les pays en voies de développement, très probablement en raison des coûts élevés des installations et des difficultés d'entretien des instruments de mesure. Toutefois, et par rapport aux variabilités spatiales du rayonnement solaire, le réseau des stations de mesures au sol reste très peu dense même dans les pays développés. Le recours à des modèles de simulation, des techniques d'interpolation ou à des données satellitaires a toujours été une solution pour l'estimation des gisements des sites isolés qui s'appêtent bien à l'installation des centrales solaires.

Dans le présent travail, nous avons utilisé plusieurs bases de données enregistrées dans des sites météorologiques marocains de différents climats pour estimer le rayonnement solaire pour chaque région et ainsi construire un outil d'estimation adaptable à différentes localités. Nous avons ainsi utilisé les mesures solaires et météorologiques de notre laboratoire SIGER, et des mesures solaires et météorologiques effectuées dans le cadre du projet EnerMENA. Chaque base de données contient deux types de variables : des variables endogènes qui sont de mêmes natures que l'estimation ou la prédiction (composantes du rayonnement solaire) et des variables exogènes de natures différentes tels que, la température et l'humidité relative de l'air, la vitesse du vent,

Durant cette thèse, le premier site étudié a été celui de la ville de Fès, située dans la plaine du Saïs (entre le Rif au nord et le Moyen Atlas au sud). Cette ville se caractérise par un climat continental avec des variations extrêmes entre les différentes saisons : hivers froids et pluvieux, étés secs et chauds, printemps et automnes doux.

Les mesures des deux composantes solaires (GHI et DHI) et des données météorologiques (température, humidité relative, ...) ont été collectées par une station météorologique (Figure 1.9) placée sur le toit de la Faculté des Sciences et Techniques (Latitude : 33°59'58" N, Longitude : 4°59'22" W et Altitude : 450 m). Un pyranomètres Kip & Zonen (modèle CM11)

a été utilisé pour mesurer l'irradiation solaire horaire globale sur plan horizontal (GHI); un pyranomètre identique, équipé d'un anneau d'ombrage, mesure, quant à lui, l'irradiation solaire horaire diffuse (DHI). La station contient également un thermo-hygromètre (HMP45C) de marque Campbell Scientific pour mesurer la température et l'humidité relative de l'air, un anémomètre (Young 05103) pour mesurer la vitesse et la direction du vent, ainsi qu'un pluviomètre pour mesurer les précipitations. De plus, cette station contient des capteurs radiométriques pour la mesure des composantes spectrales du rayonnement solaire, en particulier l'ultraviolet (UV) et le rayonnement photo-synthétiquement actif (PAR). La période de mesure s'étend du 01/01/2009 00:00:00 GMT au 11/06/2015 à 12:00:00 GMT. Après cette dernière date, la chaîne d'acquisition est tombée en panne.

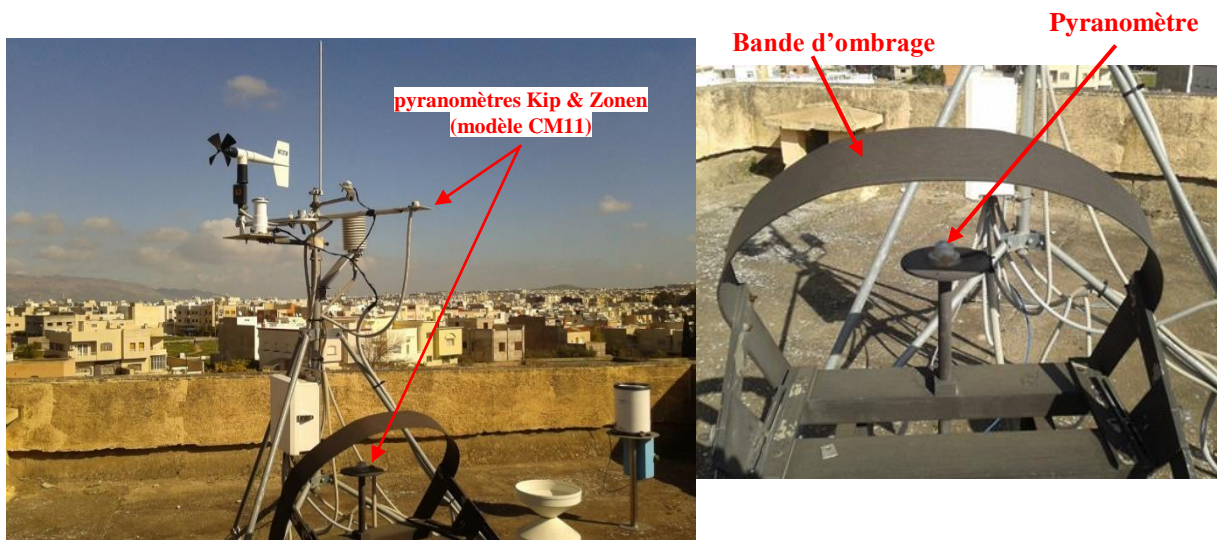


Figure 1.9. Station météorologique placée sur le toit de la Faculté des Sciences et Techniques de Fès.

En plus des données mesurées par notre station, nous avons réussi à obtenir d'autres bases de données enregistrées dans d'autres sites météorologiques marocains de différents climats. Ces bases de données nous ont été fournies par le Centre Aérospatial Allemand (DRL: Deutsches Zentrum für Luft- und Raumfahrt). La figure 1.10 montre les emplacements géographiques des dix stations météorologiques installées dans la région MENA (Moyen-Orient et Afrique du Nord) dans le cadre du projet enerMENA et parmi lesquelles les cinq villes marocaines étudiées. Chaque base de données contient les mesures solaires et météorologiques quotidiennes, horaires et sub-horaires mesurées dans cinq sites marocains : Oujda, Missour, Erfoud, Zagora et Tan-Tan. Le tableau 1.4 indique les coordonnées géographiques, les composantes solaires mesurées, les périodes de mesure et certaines caractéristiques des instruments utilisés pour chaque lieu. Les stations d'Oujda, Missour et Tan-Tan sont équipées de "Capteurs Thermiques" (TS), à savoir un pyrhéliomètre CHP1 pour les mesures du DNI, un

pyranomètre CMP21 pour les mesures du GHI et un pyranomètre CMP21 avec une boule d'ombrage pour les mesures du DHI. Ces capteurs sont montés sur un suiveur solaire SOLYS 2 entièrement automatique (Figure 1.11-Gauche)(Schüler et al., 2016). Les stations d'Erfoud et Zagora sont équipées d'un "radiomètre à bande d'ombre tournante" (RSR) (Figure 1.11-Droite). Cette technique offre une option moins coûteuse pour mesurer simultanément les composantes solaires GHI et DHI. En connaissant ces deux composantes et l'angle d'incidence des rayons solaires, on peut calculer l'éclairement direct d'une surface horizontale.

Toutes les stations sont également équipées de capteurs de température, d'humidité relative et de pression, ainsi que d'un anémomètre pour mesurer la vitesse et la direction du vent à 10 m de hauteur. Les précisions des instruments de mesure du réseau enerMENA et de la station de Fès sont présentées dans le tableau 1.5.



Figure 1.10. Vue de l'ensemble des stations enerMENA installées au Maroc (encadré en vert), en Algérie, en Tunisie, en Égypte et en Jordanie.

Tableau 1.4. Les stations marocaines du réseau enerMENA.

Site	Latitude [°N]	Longitude [°W]	Altitude [m]	Période des mesures	Type de station	Composantes solaires mesurées
OUJDA	34.650	1.900	617 m	18/08/2011 - 30/09/2015	TS	GHI DHI DNI
MISSOUR	32.860	4.107	1107 m	01/01/2014 - 30/09/2015	TS	GHI DHI DNI
ERFOUD	31.491	4.218	859 m	02/06/2013 - 30/09/2015	RSR	GHI DHI
ZAGORA	30.272	5.852	783 m	03/06/2013 - 30/09/2015	RSR	GHI DHI
TAN-TAN	28.498	11.322	75 m	01/06/2013 - 30/09/2015	TS	GHI DHI DNI



Figure 1.11. Station Thermique TS (gauche) et station RSR (droite) (Schüler et al., 2016).

Tableau 1.5. Précision des instruments de mesure utilisés.

Variables mesurées	Station TS	Station RSR	Station Fès
GHI (Irradiation globale horizontale en W/m ²)	CMP21 Précision instantanée : $\pm 2\%$	RSP-4G Précision instantanée : $\pm 4.7\%$	CM11 Précision instantanée : $\pm 5\%$
DNI (Irradiation normale directe en W/m ²)	CHP1 Précision instantanée : $\pm 1\%$	-	-
DHI (Irradiation diffuse horizontale en W/m ²)	CMP21 + boule d'ombrage Précision instantanée : $\pm 2\%$	RSP-4G Précision instantanée : $\pm 6.5\%$	CM11+ bande d'ombrage Précision instantanée : $\pm 5\%$
T_{amb} (Température ambiante en °C)	CS215 Précision : $\pm 0.3^\circ\text{C}$ en 25°C $\pm 0.4^\circ\text{C}$, plus de $+5^\circ\text{C}$ à $+40^\circ\text{C}$ $\pm 0.9^\circ\text{C}$, plus de -40°C à $+70^\circ\text{C}$	CS215 Précision : $\pm 0.3^\circ\text{C}$ en 25°C $\pm 0.4^\circ\text{C}$, plus de $+5^\circ\text{C}$ à $+40^\circ\text{C}$ $\pm 0.9^\circ\text{C}$, plus de -40°C à $+70^\circ\text{C}$	Précision : $\pm 0.3^\circ\text{C}$ pour $T > 20^\circ\text{C}$
RH (Humidité relative en %)	CS215 Précision : $\pm 2\%$ plus de 10 à 90% $\pm 4\%$ plus de 0 à 100%	CS215 Précision : $\pm 2\%$ plus de 10 à 90% $\pm 4\%$ plus de 0 à 100%	Précision : $\pm 3\%$
WS (Vitesse du vent en m/s)	NRG 40C Anémomètre Précision : < 0.1 m/s plus de 5 à 25 m/s	NRG 40C Anémomètre Précision : < 0.1 m/s (entre 5 à 25 m/s)	Anémomètre Précision : $\pm 1\%$
BP (Pression en hPa)	CS100 Précision : 0.5 hPa en 20°C 1.0 hPa plus de 0°C à 40°C 1.5 hPa plus de -20°C à $+50^\circ\text{C}$ 2.0 hPa plus de -40°C à $+60^\circ\text{C}$	CS100 Précision : 0.5 hPa en 20°C 1.0 hPa plus de 0°C à 40°C 1.5 hPa plus de -20°C à $+50^\circ\text{C}$ 2.0 hPa plus de -40°C à $+60^\circ\text{C}$	-

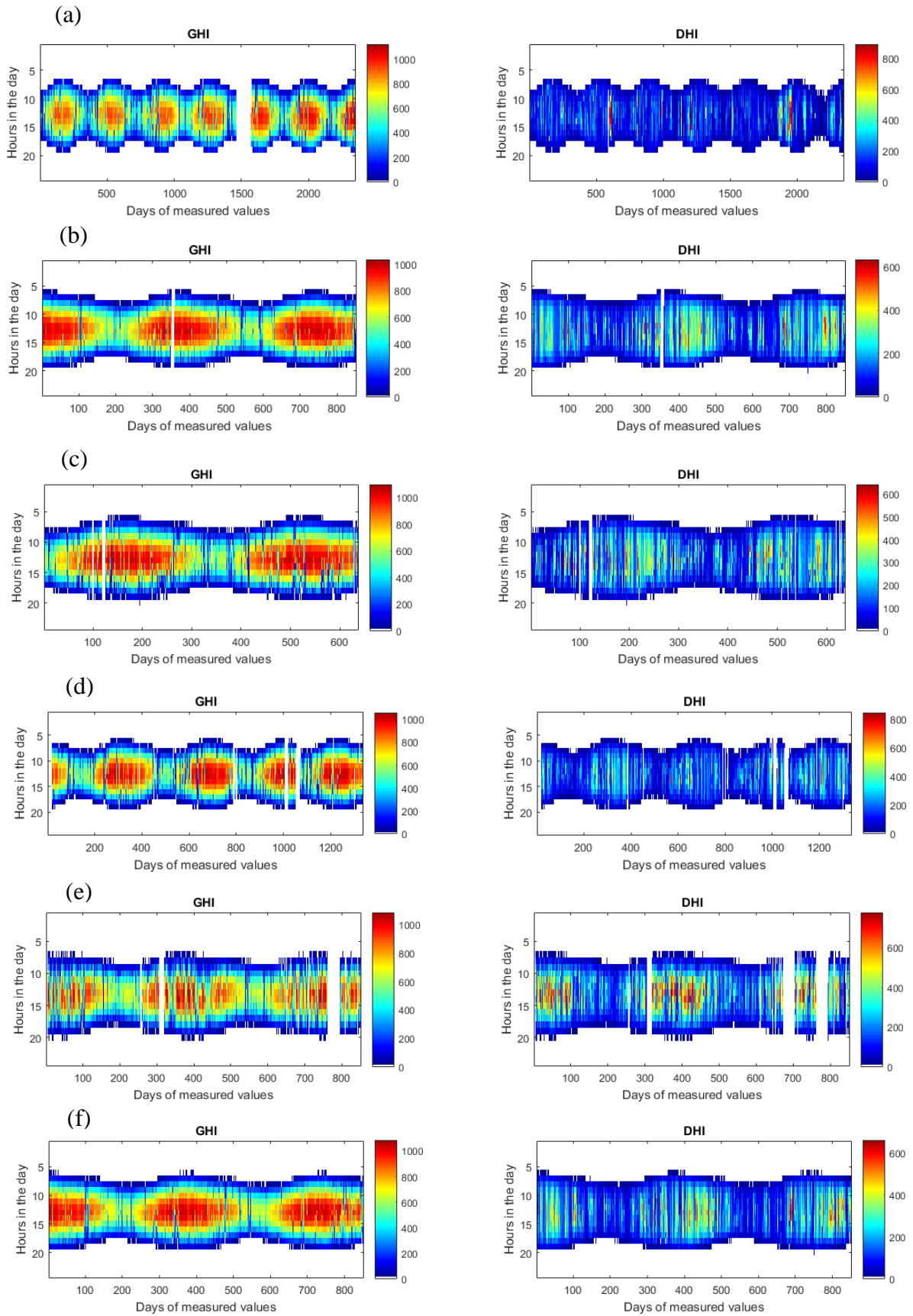


Figure 1.12. Les variations horaires de GHI, DHI, mesurés à Fès (a), Erfoud (b), Missouri (c), Oujda (d), TanTan (e), et Zagora (f).

Sur la figure 1.12., nous avons représenté les variations horaires brutes des composantes solaires GHI et DHI pour la période considérée pour chaque site d'étude. Sur cette figure, les valeurs significatives (en couleur rouge) de DHI ont été causées par des désalignements, de temps en temps, de la bande ou de la boule d'ombrage solaire. Ceci montre qu'une procédure de contrôle de qualité des mesures est nécessaire pour éliminer les mesures erronées.

Pour les mesures de la ville de Fès (Figure 1.12. (a)), on remarque qu'il existe une longue période de données manquantes (trou de données) dans les mesures de GHI au début de l'année 2013 (3 Janvier – 23 Avril 2013). Ceci a été dû à un problème de connexion du pyranomètre avec la chaîne d'acquisition de données. En fait, il s'agit d'une période où le fil de la masse du pyranomètre était cassé. Aussi, des valeurs manquantes ont été observées dans les autres villes et, en particulier, à TanTan (Figure 1.12. (e)), où on remarque l'absence d'une grande partie de mesures de GHI et DHI en 2015.

La figure 1.13 montre les sommes annuelles brutes des trois composantes solaires (GHI, DHI, et DNI) mesurées aux six sites marocains pour chaque année. En plus, les variabilités horaires de la température, et l'humidité relative mesurées à chaque site pour chaque période considérée, sont présentées simultanément dans les figures 1.14 et 1.15. Selon la figure 1.15, nous avons remarqué que la ville de TanTan se caractérise par des variations rapides et élevées de l'humidité relative (la plupart des valeurs dépassent les 80%) par rapport aux autres villes pendant toute la période de mesure.

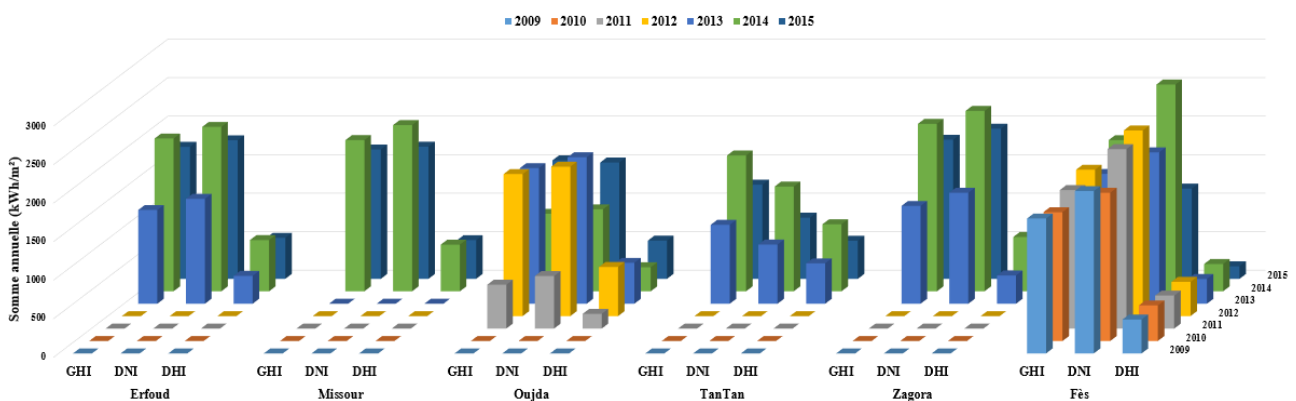


Figure 1.13. Sommes annuelles brutes des trois composantes solaires mesurées aux six sites marocains.

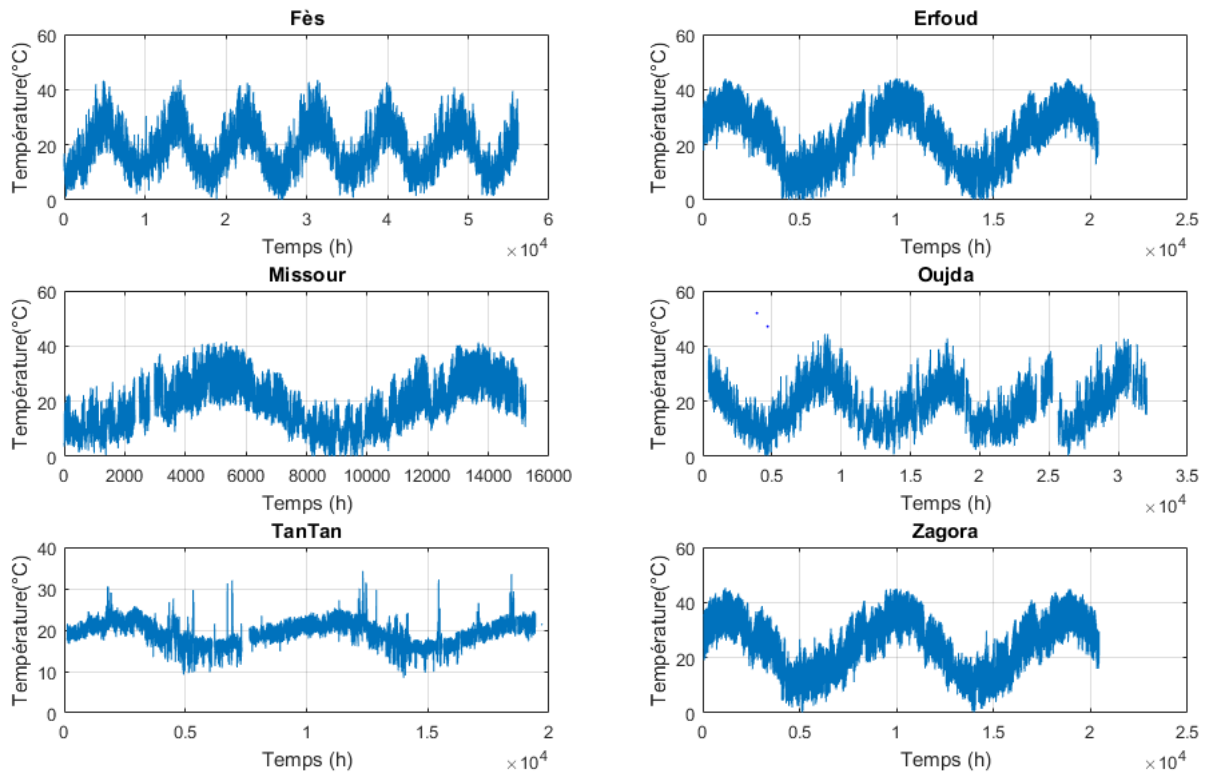


Figure 1.14. Variabilités horaires de la température ambiante mesurée aux six sites marocains.

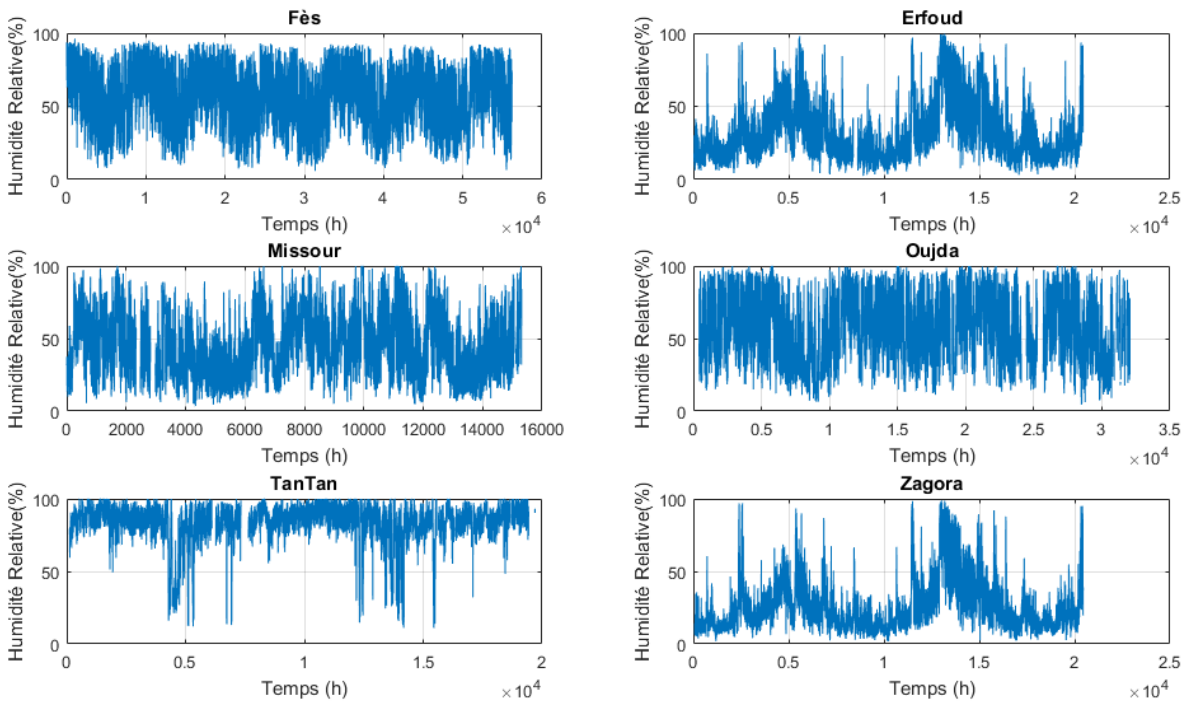


Figure 1.15. Variabilités horaires de l'humidité relative de l'air mesurée aux six sites marocains.

12. Contrôle de qualité et correction des données solaires

La mesure au sol des composantes du rayonnement solaire est une tâche difficile qui nécessite une grande maintenance, en raison des erreurs et des incertitudes contenues dans le processus (Calça et al., 2019). Les causes les plus courantes de ces erreurs sont liées, premièrement, aux équipements et à leurs sensibilités respectives (effet cosinus, réponse en azimut, réponse en température, sélectivité spectrale, non-linéarité, stabilité de l'instrument); deuxièmement, aux problèmes d'exploitation connexes (désalignement complet ou partiel de l'anneau ou boule d'ombrage; poussière, neige, rosée, gouttelettes d'eau, fientes d'oiseaux, charge mécanique sur les câbles, coupures d'électricité, arrêt de la station, etc) ((Younes et al., 2005), (Muneer et al., 2007)). Pour ces raisons, la validation et le contrôle de la qualité des données du rayonnement solaire sont aussi importants que les techniques de mesure elles-mêmes (Hay, 1993). En effet, ces étapes de vérification sont essentielles pour garantir l'exactitude et l'utilité des données.

Le contrôle de la qualité des données solaires a préoccupé plusieurs chercheurs qui ont proposé différentes procédures, afin d'évaluer et d'éliminer les valeurs erronées dans les mesures des irradiances solaires. Dans cette section, une étude bibliographique sur les diverses techniques de contrôle de qualité proposées par de nombreux chercheurs est présentée. Ensuite, la procédure de contrôle de qualité suivie durant notre travail est décrite étape par étape, en l'appliquant aux irradiances horizontales globales et diffuses horaires de nos bases de données. Une solution de remplacement des données manquantes des mesures de GHI au début de l'année 2013 à la ville de Fès a, aussi, été proposée.

12.1. Etude bibliographique sur les méthodes de contrôle de qualité

La précision et l'utilité des données solaires ne peuvent être assurées qu'après avoir passé les étapes de contrôle de qualité. Hay (Hay, 1993) a identifié cinq catégories de procédures pour contrôler la qualité des données du rayonnement solaire. La première étape consiste en des contrôles utilisant des systèmes de mesure indépendants. Les autres étapes consistent en des contrôles de cohérence utilisant les irradiances de surface horizontale observées, les irradiances de surface horizontale calculées, les irradiances de surface inclinée mesurées et les irradiances de surface inclinée calculées.

Selon Geiger et al.(Geiger et al., 2002), le processus de contrôle de la qualité des données solaires est une tâche longue et fastidieuse. Pour cette raison, ils ont décidé de rendre cette procédure accessible à tout le monde, en utilisant un service web. Elle est appliquée aux données quotidiennes d'irradiation globale et basée sur les directives de l'Organisation Météorologique

Mondiale (WMO (World Meteorological Organization) 1981) et met en œuvre certaines des procédures utilisées dans la construction de l'Atlas européen du rayonnement solaire (ESRA 2000) (Scharmer et al., 2000).

Younes et al. (Younes et al., 2005) ont proposé une procédure semi-automatique basée sur quatre tests. Dans le premier test, les données représentant une altitude solaire inférieure à 7° ont été éliminées. Le deuxième test est un test logique appliqué à l'indice de clarté K_t et à la fraction diffuse K_d , qui doivent être positifs et compris entre 0 et 1. Le troisième test est une comparaison des irradiances globale et diffuse avec leurs limites supérieures et inférieures correspondant au modèle de Page (Page et al., 2001). Le dernier test est statistique et repose sur une procédure d'écart-type dans un diagramme de dispersion de K_t - K_d . Cette approche a été appliquée dans onze endroits de l'hémisphère nord couvrant deux continents. De plus, elle a été testée et validée par plusieurs chercheurs ((Ihya et al., 2015), (Ruiz-Arias et al., 2010)). Cette méthode est considérée comme générale et peut être efficace pour tout ensemble de données terrestres. En 2007, Muneer et al. (Muneer et al., 2007) ont appliqué une procédure de contrôle de qualité sur les données solaires mesurées à Bahreïn en suivant quatre tests. Le premier test consiste à imposer des limites physiques pour les irradiances globales et diffuses. Le deuxième test est divisé en deux tests de cohérence entre les différentes composantes du rayonnement. Le troisième test est basé sur l'indice de clarté (K_t) et la fraction diffuse (K_d) et une vérification supplémentaire de l'irradiation diffuse en comparant sa valeur aux deux conditions extrêmes lorsque le ciel est clair et couvert. Le dernier test représente les calculs du facteur de turbidité pour la série temporelle des données et ses limites en fonction du type de ciel.

Long et Shi (Long and Shi, 2008) ont développé un algorithme automatisé pour tester les mesures quotidiennes du rayonnement solaire reçu par une surface, en utilisant les données du programme ARM (Atmospheric Radiation Measurement) du département américain de l'énergie. Cette procédure est basée sur les limites physiquement possibles déterminées par le réseau de référence du rayonnement de surface du WMO (World Meteorological Organization Baseline Surface Radiation Network) (Ohmura et al., 1998), et utilisait des limites configurables qui dépendent de l'analyse climatologique des données mesurées du site. Un an plus tard, Moradi (Moradi, 2009) a développé une méthode automatique pour contrôler la qualité des données du rayonnement solaire global quotidien en utilisant la période d'ensoleillement. Les données utilisées représentent trois stations avec différents types de climat exploitées par l'Organisation météorologique iranienne (IRIMO). Cette approche est décrite en trois tests : le premier test consiste à comparer l'irradiation globale quotidienne avec le

rayonnement extraterrestre quotidien reçu par une surface horizontale. Le deuxième test comprend la comparaison entre les mesures de l'irradiation globale quotidienne et leurs valeurs correspondantes aux conditions du ciel clair. Alors que, le dernier test se base sur la relation entre l'irradiation globale quotidienne et la durée d'ensoleillement en heures.

Le projet ENDORSE (Espinari et al., 2011) a proposé une procédure de contrôle de qualité (QCP) basée sur des observations extrêmes et rares, prenant en compte les trois composantes de l'irradiation solaire: l'irradiation horizontale globale (GHI), l'irradiation horizontale diffuse (DHI) et l'irradiation normale directe (DNI) pour des périodes mensuelles, quotidiennes, horaires et sub-horaires. Ce QCP n'a été appliqué que pour les données d'altitude solaire supérieure à 7° afin d'éviter les erreurs intrinsèques en cosinus des pyranomètres. Pour surmonter le problème des valeurs erronées dans une base de données solaires, les chercheurs de l'Institut Royal Météorologique de Belgique (IRMB) (Journée and Bertrand, 2011) ont développé des procédures et un logiciel pour vérifier la qualité des données solaires mesurées. Ces procédures peuvent être appliquées aux données mesurées toutes les dix ou trente minutes en suivant les tests suivants : Tests de seuil physique, tests d'étape, tests de persistance, tests d'enveloppe de qualité, tests de cohérence spatiale et tests d'ensoleillement.

La procédure de contrôle de qualité adoptée par Mirás-Avalos et al. (Mirás-Avalos et al., 2012) est identique à celle présentée par Younes et al. (Younes et al., 2005) et comprend quatre étapes, la différence principale étant que les tests sont utilisés spatialement et pas seulement temporellement. Dans leur étude, Laitia et al. (Laitia et al., 2014) ont appliqué la procédure de contrôle de qualité automatique (QC) proposée par Journée et Bertrand (Journée and Bertrand, 2011), aux données d'irradiation collectées dans les stations radiométriques de la région de Trentino (Alpes italiennes) pour une période de mesure définie entre 1987 et 2012. Cette procédure inclut les tests de seuil physique, les tests d'étape, les tests de persistance et les tests de cohérence spatiale.

Moreno-Tejera et al. (Moreno-Tejera et al., 2015) ont appliqué une procédure de contrôle de qualité proposée par Long et Dutton (Long and Dutton, 2010) appliquée à quatre bases de données enregistrées dans différents endroits en Espagne. Cette procédure est décrite en trois tests pour vérifier si les données respectent les limites définies. Le premier test consiste à imposer des limites physiques de base non strictes aux différentes composantes du rayonnement solaire, le deuxième test a des limites plus restrictives que celles du premier test. Dans le troisième test, le rayonnement direct mesuré est comparé à sa valeur calculée à partir des valeurs mesurées du rayonnement global et diffus et en fonction de la valeur de l'angle zénithal.

Pashiardis et Kalogirou (Pashiardis and Kalogirou, 2016) ont utilisé une procédure pour tester la qualité des mesures du rayonnement solaire effectuées dans deux stations à Chypre pendant deux ans. Cette procédure est basée sur des limites physiquement possibles pour toutes les composantes du rayonnement et une comparaison entre le rayonnement global et la somme des rayonnements direct et diffus. Ce processus de contrôle de la qualité est valable pour chaque irradiation mesurée toutes les dix minutes, toutes les heures et tous les jours.

Pour assurer la fiabilité des bases de données solaires horaires et sub-horaires collectées au Brésil, Lemos et al. (Lemos et al., 2017) ont utilisé le modèle BRL de Ridley et al. (Ridley et al., 2010). Ce modèle est ajusté aux caractéristiques spécifiques du climat brésilien en utilisant les mesures des irradiances globale, diffuse et directe maintenues par l'INPE dans le cadre du projet SONDA dans neuf stations. Cette procédure consiste à appliquer un ensemble de filtres à l'altitude solaire, au facteur de persistance, aux indices de clarté et de diffusion, l'ensemble des données adapté aux contrôles de plausibilité, de cohérence et de défauts d'équipement. Les variations extrêmes des valeurs mesurées consécutives de l'irradiation globale sont éliminées et les valeurs quotidiennes de l'irradiation globale considérées pour estimer la fraction diffuse ne doivent pas être inférieures à 3% de l'irradiation extraterrestre quotidienne, puis les valeurs de l'irradiation globale mesurées sont comparées à leurs valeurs dans les conditions du ciel clair. Ensuite, cette procédure est finalisée par une comparaison entre l'irradiation diffuse et les limites inférieures de Rayleigh.

Récemment, Perez-Astudillo et al. (Perez-Astudillo et al., 2018) ont proposé un ensemble de tests exhaustifs basés sur le BSRN (Baseline Surface Radiation Network) et d'autres vérifications adaptées qui permettent une identification automatisée en temps réel des valeurs aberrantes dans les mesures d'irradiation globale, diffuse et directe. Les résultats de ces tests proposés ont montré une amélioration de la détection des problèmes non identifiables par les tests de BSRN (Ohmura et al., 1998).

12.2. Procédure et résultats du contrôle de qualité

Comme rapporté ci-dessus, plusieurs tests ont été développés pour garantir la fiabilité des mesures de toutes les composantes solaires. Dans le présent travail, la procédure de contrôle de qualité utilisée pour éliminer les éventuelles données erronées est basée sur trois principaux tests proposés par (Younes et al., 2005) et (Muneer et al., 2007). Cette procédure est appliquée aux mesures horaires des irradiances solaires globales et diffuses. Nous n'avons conservé que les mesures qui ont passé avec succès ce processus de contrôle de qualité.

12.2.1. Premier test

Les mesures solaires effectuées à de faibles altitudes solaires sont considérées comme douteuses en raison de la réponse en cosinus du pyranomètre. Il s'agit de la réponse du capteur à l'angle auquel le rayonnement solaire touche la zone de détection. Cet angle est généralement important lors du lever et du coucher du soleil, ce qui introduit une erreur de cosinus plus élevée. Pour cette raison, certains chercheurs ((Younes et al., 2005), (Muneer, 2007), (Ruiz-Arias et al., 2010)) ont recommandé d'éliminer toutes les données solaires correspondantes à une altitude solaire inférieure à 7° . De plus, en traçant l'indice de clarté K_t en fonction de l'altitude solaire pour nos sites d'étude (Figure 1.16), on a constaté que ce choix d'une limite de 7° est très plausible (Bounoua et al., 2018). Selon cette illustration, on voit clairement l'effet cosinus sur les réponses des pyranomètres ($K_t > 1.5$ pour des altitudes solaires inférieures à 7°). La figure 1.17 montre le diagramme de dispersion de la fraction diffuse K_d en fonction de l'indice de clarté K_t après application du premier test. On voit clairement la nécessité d'application d'autres tests de contrôle de qualité.

12.2.2. Deuxième test

Le deuxième test adopté, dans notre procédure de contrôle de qualité, consiste en l'application des limites physiques de l'indice de clarté K_t et de la fraction diffuse K_d . Ces deux paramètres doivent avoir des valeurs comprises entre 0 et 1. Le résultat de ce test est illustré dans la figure 18. Il convient de préciser que toutes les valeurs de K_t et K_d (points rouges) dépassant 1 ont été éliminées par ce test.

$$0 < K_t < 1 \quad (1-29)$$

$$0 < K_d < 1 \quad (1-30)$$

12.2.3. Troisième test

Ce troisième test est basé sur une analyse statistique des données solaires et il est décrit comme suit :

- On divise l'intervalle [0-1] de l'indice de clarté K_t en vingt intervalles équidistants (au lieu de 10 intervalles proposés par (Younes et al., 2005), pour augmenter la précision de détection des points aberrants.
- Pour chaque intervalle, on calcule la moyenne de la fraction diffuse K_d (K_{dm}) et son écart-type (σ_{dm}).
- On ne retient que les valeurs qui respectent l'intervalle :

$$(K_{dm} - 2\sigma_{dm}) \leq K_d \leq (K_{dm} + 2\sigma_{dm}) \quad (1-31)$$

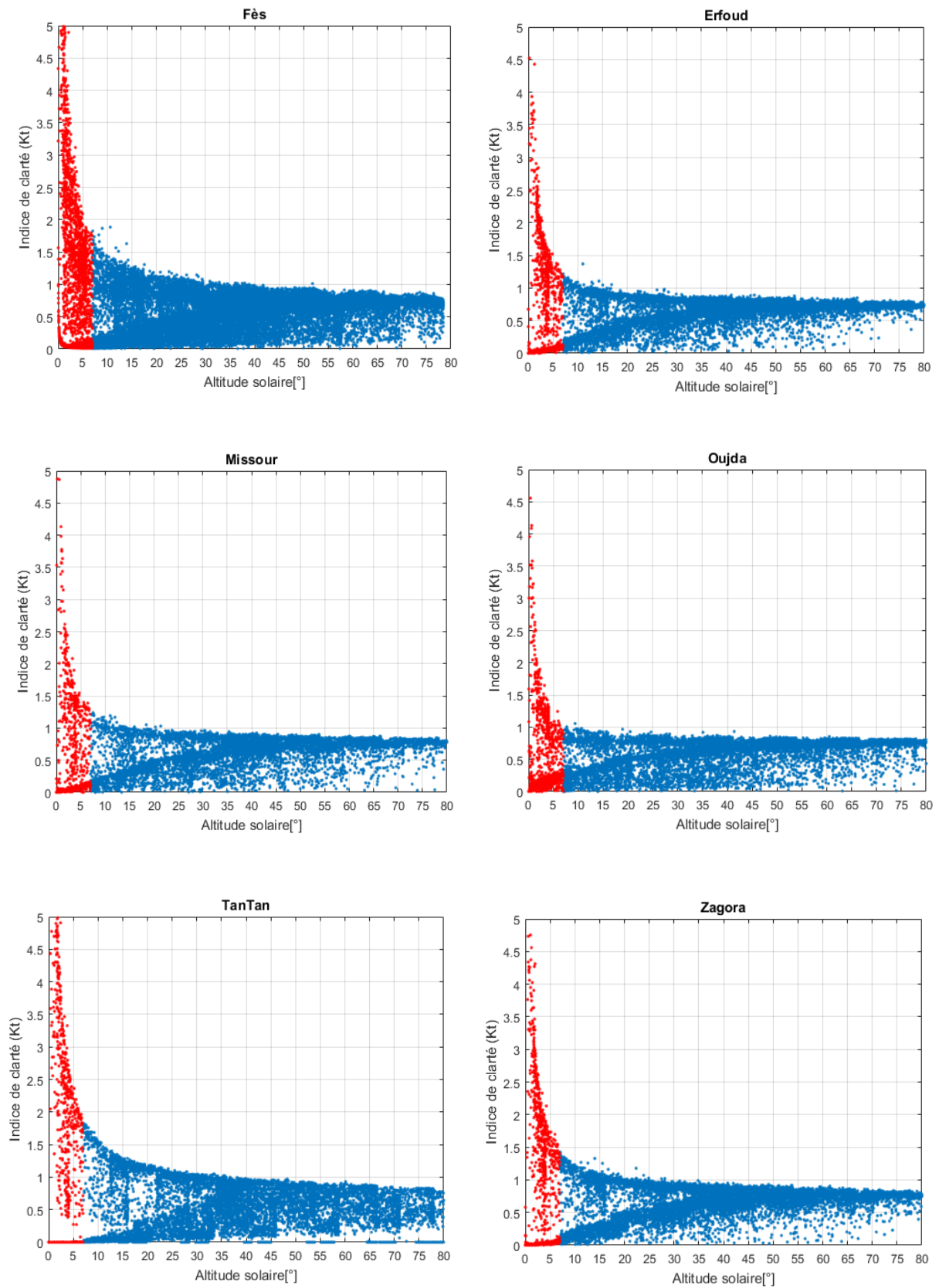


Figure 1.16. Indice de clarté en fonction de l'altitude solaire pour les données solaires mesurées aux six stations étudiées.

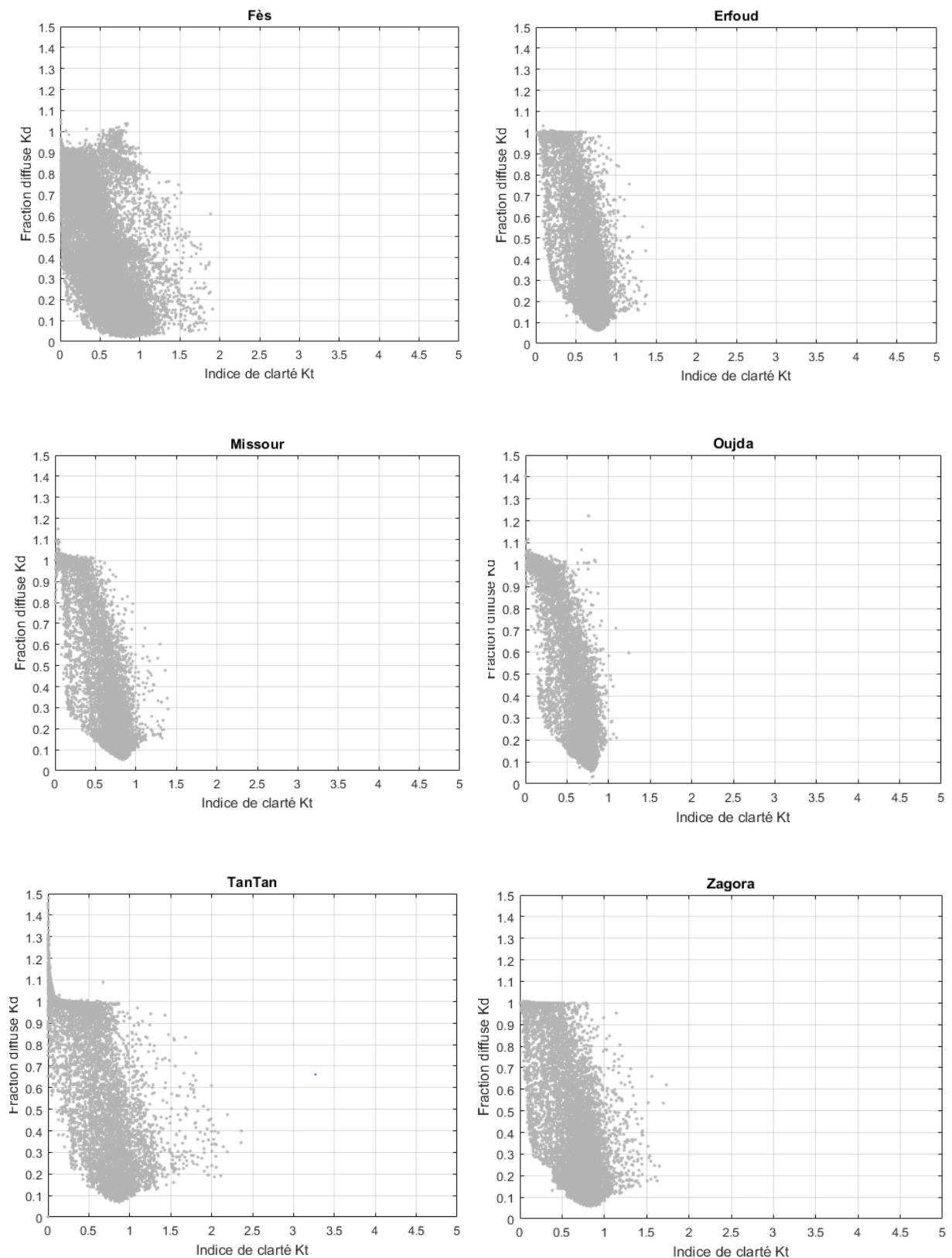


Figure 1.17. Diagramme de dispersion de la fraction diffuse K_d en fonction de l'indice de clarté K_t après le premier test de contrôle de qualité.

Une description mathématique a été faite pour représenter la courbe inférieure $F_{low}(K_t)$ et la courbe supérieure $F_{upp}(K_t)$ qui expriment respectivement la marge inférieure ($K_{dm} - 2\sigma_{dm}$) et la marge supérieure ($K_{dm} + 2\sigma_{dm}$) que les valeurs des données solaires ne doivent pas dépasser. Ces deux courbes sont exprimées par une équation polynomiale du second degré et extraites d'une interpolation des vingt valeurs de $(m_{K_t interval}, K_{dm} - 2\sigma_{dm})$ et $(m_{K_t interval}, K_{dm} + 2\sigma_{dm})$ pour chaque intervalle m donné de K_t .

Les équations supérieures et inférieures du second degré obtenues pour chaque base de données utilisées sont :

$$\text{Fès :} \quad F_{upp}(K_t) = -0.132*K_t^2 + 0.710*K_t + 1.086 \quad F_{low}(K_t) = 0.635*K_t^2 - 1.145*K_t + 0.454 \quad (1-32)$$

$$\text{Erfoud :} \quad F_{upp}(K_t) = -1.742*K_t^2 + 0.632*K_t + 1.017 \quad F_{low}(K_t) = 1.205*K_t^2 - 2.184*K_t + 0.882 \quad (1-33)$$

$$\text{Missour :} \quad F_{upp}(K_t) = -1.515*K_t^2 + 0.280*K_t + 1.156 \quad F_{low}(K_t) = 1.420*K_t^2 - 2.277*K_t + 0.854 \quad (1-34)$$

$$\text{Oujda :} \quad F_{upp}(K_t) = -1.413*K_t^2 + 0.171*K_t + 1.152 \quad F_{low}(K_t) = 1.599*K_t^2 - 2.688*K_t + 1.022 \quad (1-35)$$

$$\text{TanTan :} \quad F_{upp}(K_t) = -1.524*K_t^2 + 0.582*K_t + 1.042 \quad F_{low}(K_t) = 0.467*K_t^2 - 1.560*K_t + 0.896 \quad (1-36)$$

$$\text{Zagora :} \quad F_{upp}(K_t) = -1.579*K_t^2 + 0.536*K_t + 1.060 \quad F_{low}(K_t) = 1.083*K_t^2 - 1.963*K_t + 0.860 \quad (1-37)$$

Pour respecter les limites physiques de K_t et K_d , nous avons considéré que :

$$F_{low}(K_t) = \max(0, F_{low}(K_t)) \quad (1-38)$$

$$F_{upp}(K_t) = \min(1, F_{up}(K_t)) \quad (1-39)$$

La figure 1.19 représente le diagramme de dispersion de la fraction diffuse K_d en fonction de l'indice de clarté K_t après application de ce dernier test de contrôle de qualité pour chaque base de données utilisées. Les points en gris sont les valeurs qui ont passé ce test, alors que les points en rouge sont les valeurs qui n'ont pas passé ce test. La courbe supérieure $F_{upp}(K_t)$ et inférieure $F_{low}(K_t)$ sont représentées respectivement par une marge bleue et verte.

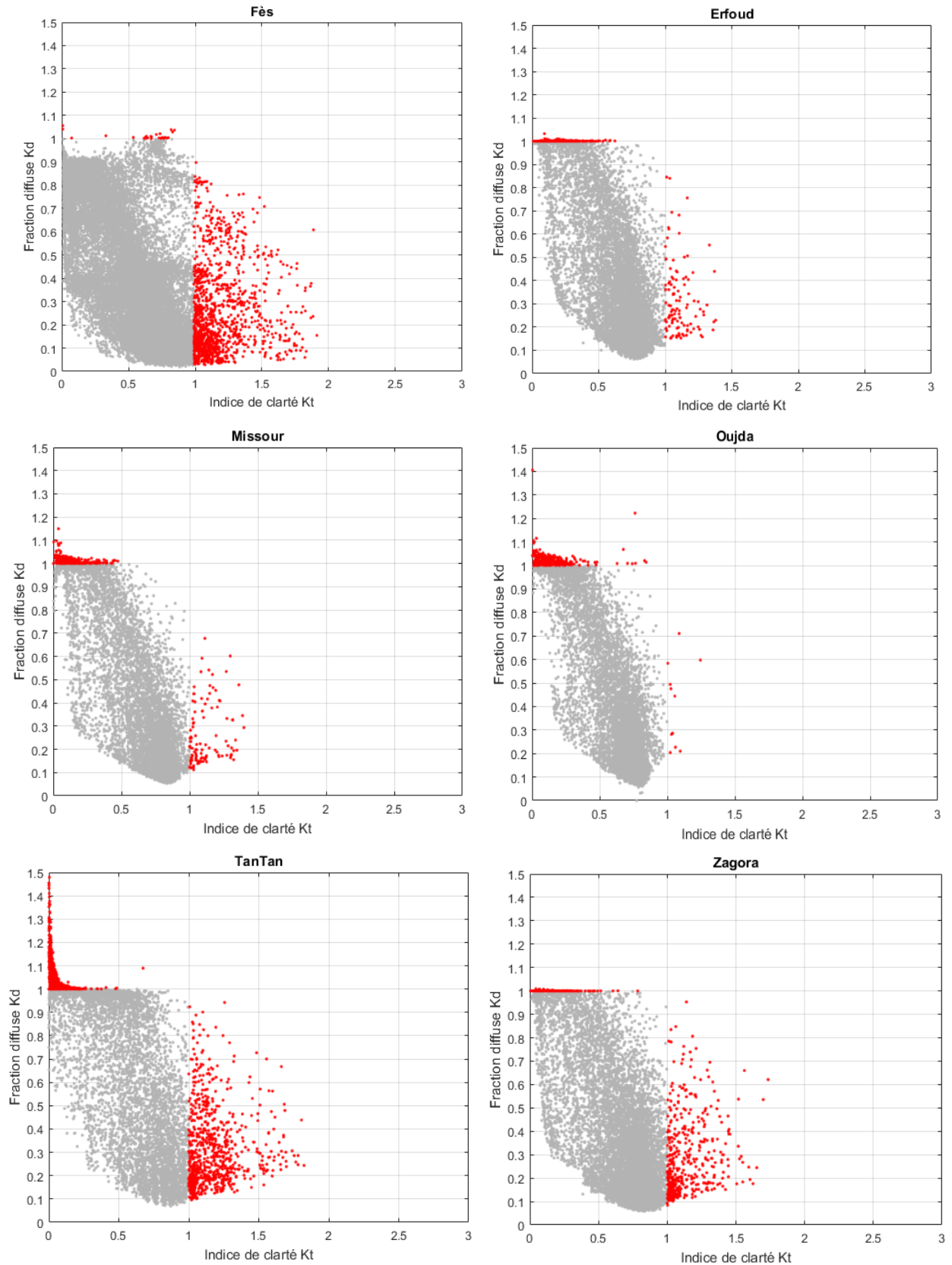


Figure 1.18. Diagramme de dispersion de la fraction diffuse K_d en fonction de l'indice de clarté K_t après le deuxième test de contrôle de qualité (Les points en rouge sont les valeurs qui n'ont pas passé le test).

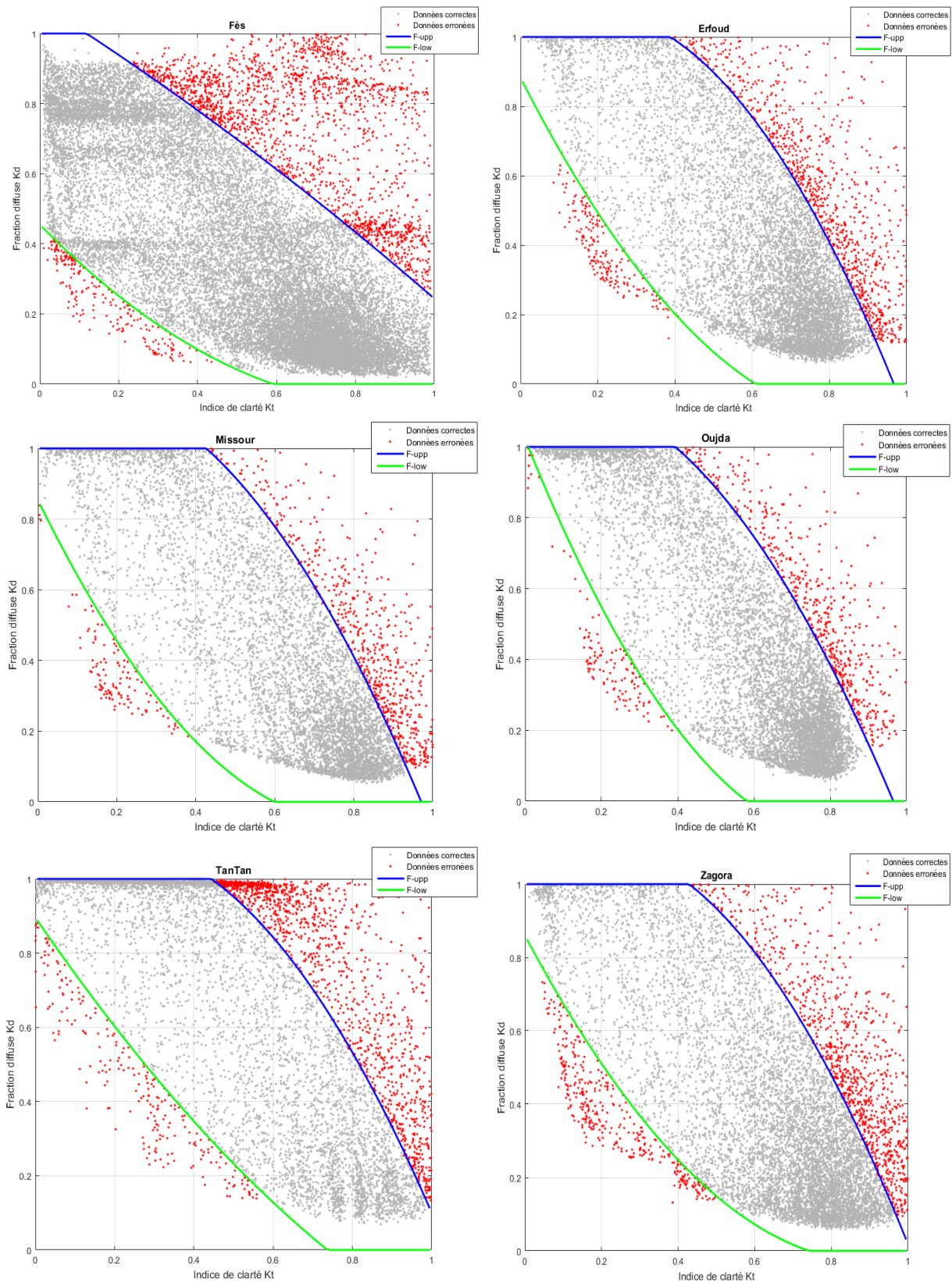


Figure 1.19. Diagramme de dispersion de la fraction diffuse K_d en fonction de l'indice de clarté K_t après le troisième test de contrôle de qualité (Les points en rouge sont les valeurs qui n'ont pas passé le test).

Le tableau 1.6 montre la totalité des données solaires éliminées par la procédure de contrôle de qualité suivie durant notre travail, en indiquant le nombre des données éliminées par chaque test pour chaque station. D'après ce tableau, on note que le plus faible pourcentage d'élimination des données est obtenu à Erfoud (31.50%), suivi par Fès (31.84%), Zagora (34.19%), Oujda (37.55%), et Missour (39.59%). Par contre, plus de 50% des données ont été éliminées à TanTan (54.48%). Cela prouve que les mesures à TanTan sont moins précises que celle des autres villes.

Tableau 1.6. Nombre des données solaires éliminées par la procédure de contrôle de qualité pour chaque site d'étude.

	Fès	Erfoud	Missour	Oujda	TanTan	Zagora
Nombre de données brutes sans compter les heures de nuit	31868	11498	8585	11563	11288	11435
Données éliminées par le premier test	5446	1837	1316	2186	2160	1774
Données éliminées par le deuxième test	2511	684	639	1150	2475	774
Données éliminées par le troisième test	2190	1101	1444	1007	1515	1362
Total des données éliminées	10147	3622	3399	4343	6150	3910
Total des données éliminées en (%)	31.84%	31.50%	39.59%	37.55%	54.48%	34.19%

12.3. Remplacement d'une longue série de données manquantes

Pour résoudre le problème des données manquantes au début de l'année 2013 dans les mesures de GHI à la ville de Fès, nous avons remarqué lors de notre analyse des données, une bonne corrélation entre l'irradiation horizontale globale (GHI) horaire et le rayonnement photo-synthétiquement actif (PAR) (Figure 1.20). Nous avons alors pensé qu'on peut utiliser cette corrélation pour remplacer les données manquantes du GHI à partir des mesures du PAR. Pour cela, nous avons d'abord éliminé, pour toutes les autres années, la même période manquante de l'année 2013 (3 Janvier – 23 Avril). Ensuite, nous avons tracé GHI en fonction du rayonnement photo-synthétiquement actif (PAR) pour la période restante de chaque année et nous avons déduit la relation entre ces deux composantes à travers laquelle nous avons calculé les données manquantes de GHI pour chaque année (Figure 1.21). Afin d'évaluer la performance de cette corrélation, nous avons comparé les données calculées de GHI avec celles mesurées pour la période 3 Janvier – 23 Avril, en traçant les valeurs de GHI calculées en fonction de celles mesurées (Figure 1.22). Selon cette figure, on remarque qu'on a obtenu un très bon coefficient de corrélation R pour chaque année.

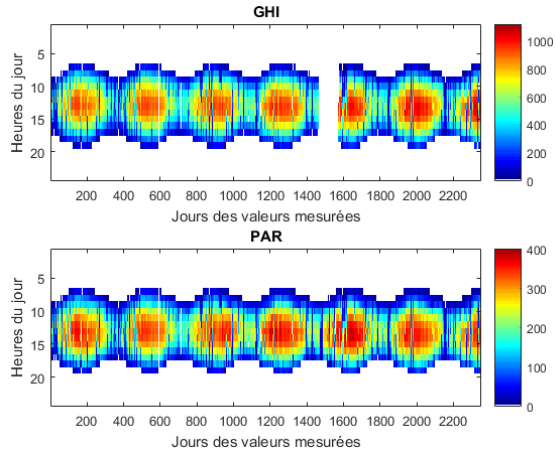


Figure 1.20. Les variations horaires de GHI, et PAR mesurés à Fès.

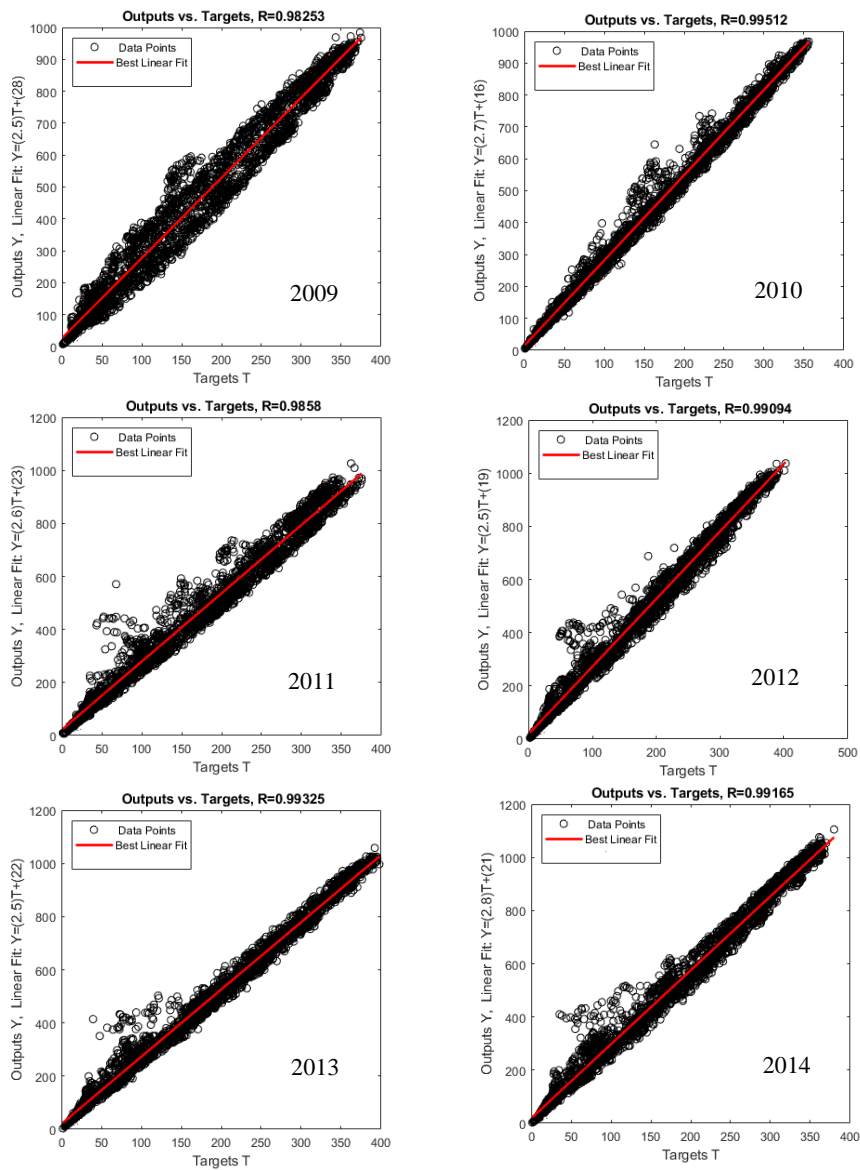


Figure 1.21. Régression linéaire entre GHI et PAR sans la période 3 Janvier – 23 Avril.

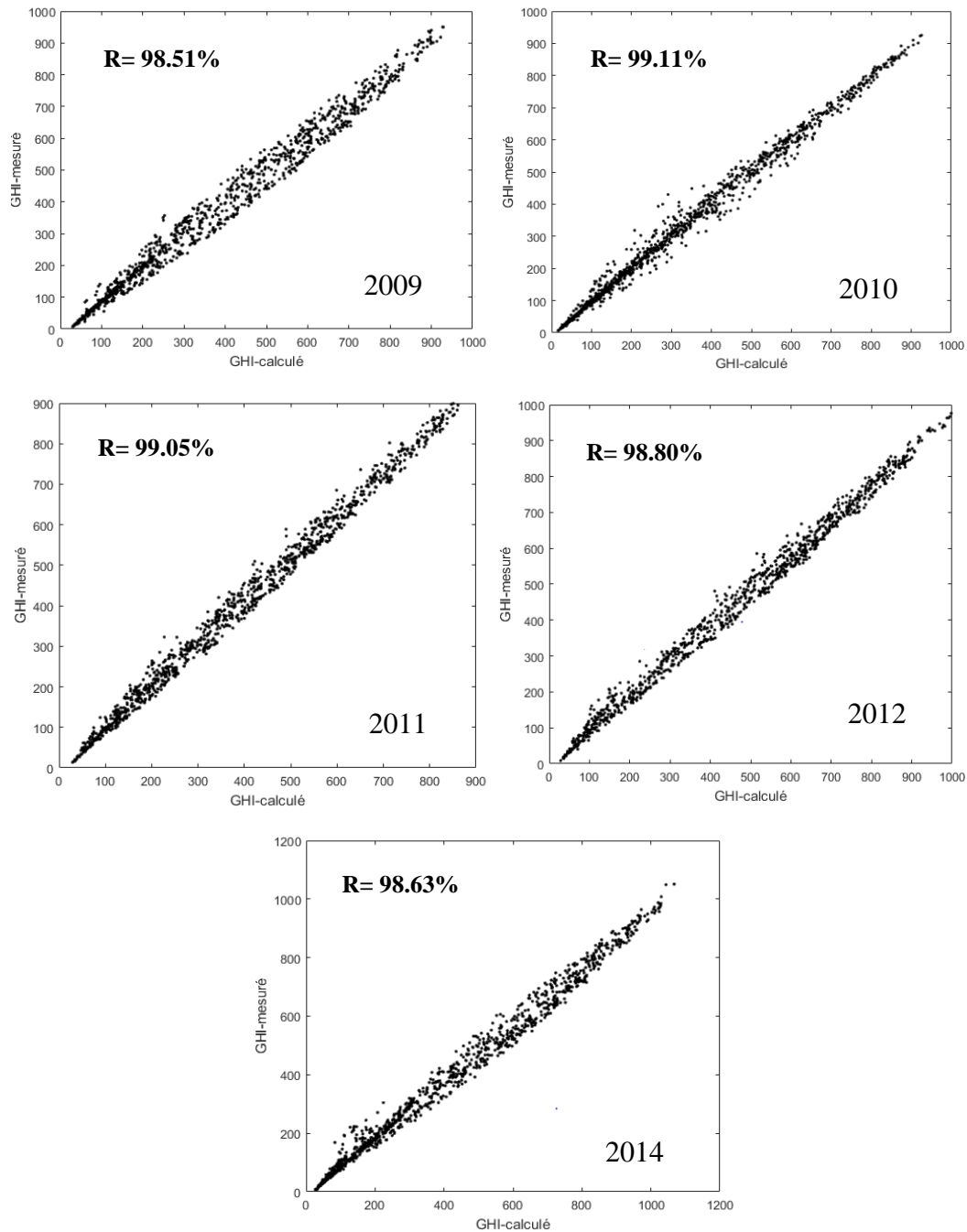


Figure 1.22. Corrélation entre les valeurs mesurées et calculées de GHI pour la période 3 Janvier – 23 Avril.

Le tableau 1.7 présente les indicateurs statistiques (Annexe D) calculés pour évaluer la performance de corrélation entre les valeurs calculées et mesurées du GHI pour la période manquante (3 Janvier – 23 Avril). Pour chaque année, les résultats montrent une excellente performance entre les valeurs de GHI mesurées et celles calculées ($R > 98\%$), ce qui prouve la possibilité de déduire les valeurs manquantes du GHI à partir des valeurs mesurées du PAR.

La figure 1.23 montre l'irradiation globale horaire reçue par un plan horizontal après correction des valeurs manquantes.

Tableau 1.7. Performances entre les valeurs calculées et mesurées du GHI pour la période 3 Janvier – 23 Avril.

Année	R (%)	nMAE	nRMSE
2009	98.513	0.067	0.083
2010	99.112	0.044	0.065
2011	99.056	0.050	0.063
2012	98.804	0.052	0.073
2014	98.635	0.062	0.080

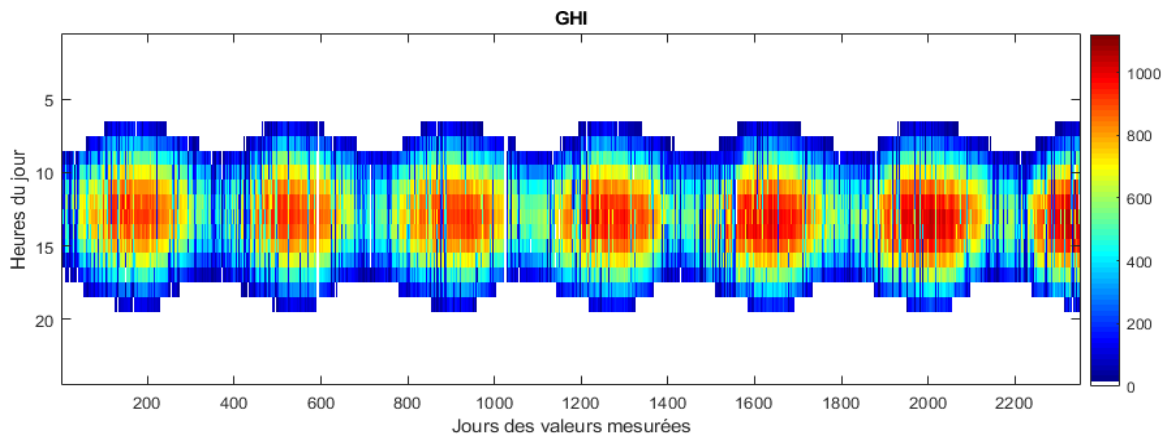


Figure 1.23. Variations horaires de GHI après correction des données manquantes au début de l'année 2013 à la ville de Fès.

13. Conclusion

Dans ce chapitre, nous avons présenté les principes fondamentaux du rayonnement solaire permettant de comprendre sa nature, sa composition et ses variations dans l'espace et dans le temps. Trois algorithmes (SPA, SG2 et Yallop) permettant de calculer, avec précision, la position solaire à chaque instant, ont été présentés. Ces algorithmes donnent approximativement les mêmes résultats. Dans nos calculs, nous avons adopté l'algorithme SG2.

Les données expérimentales provenant des six sites marocains utilisées dans le présent travail, ont été, ensuite, présentées et une procédure de contrôle de qualité leurs a été appliquée. Les résultats de ce contrôle de qualité ont montré que la précision des mesures des irradiances solaires varie d'une ville à une autre. Mais, il a été constaté que la précision des mesures à Erfoud, Fès et Zagora est meilleure à celle de Missouri, Oujda, TanTan.

Nous avons, à la fin, exploité la très bonne corrélation entre les mesures de la composante GHI et celle du PAR à la ville de Fès pour remplacer les données manquantes du GHI pour la période du 3 Janvier au 23 Avril 2013. Pour cette période, le pyranomètre de mesure du GHI souffrait d'un problème de connexion avec la chaîne d'acquisition des données.

Le prochain chapitre présentera les performances des modèles empiriques dans l'estimation des irradiances solaires journalières (GHI, DHI et DNI). Différents modèles empiriques proposés dans la littérature et des modèles que nous avons, nous-mêmes, proposés seront étudiés.

Chapitre 2

Estimation des irradiances solaires journalières
à l'aide des modèles empiriques

1. Introduction

Après la crise énergétique de 1973, plusieurs chercheurs, en particulier occidentaux, se sont beaucoup intéressés au développement de solutions alternatives pouvant assurer une indépendance vis-à-vis des énergies fossiles. Les énergies renouvelables occupaient une place de choix surtout avec les encouragements et subventions accordés par les Etats. Depuis, on a assisté à plusieurs alternances des hauts et des bas dans l'intérêt accordé aux sources à énergie renouvelable en particulier solaire et éolienne. Ces dernières années, et avec l'augmentation des prix des énergies fossiles, en général, et du pétrole, en particulier, plusieurs pays, dont le Maroc, ont complètement changé leur politique énergétique en intégrant, à grande échelle, les énergies renouvelables dans leurs mix énergétiques. Le problème majeur de l'exploitation de certaines sources à énergies renouvelables, comme le soleil et le vent, est leur intermittence. En effet, leur disponibilité dépend des aléas climatiques. La disponibilité du rayonnement solaire sur terre dépend de la qualité de l'atmosphère (turbidité), des passages nuageux en plus de l'alternance des jours et des nuits. La gestion des réseaux électriques classiques (actuels), en présence des centrales solaires, est délicate vu que les productions se font pendant le jour (période où la demande d'électricité n'est pas très grande), en plus de leur caractère imprévisible (passages nuageux). La migration vers des réseaux électriques intelligents (smart grids) est primordiale. Dans ces derniers, les prévisions des productions énergétiques est primordiale pour préparer des scénarios de gestion des réseaux électriques dont la vente du surplus à des prix convenables. Une autre solution à ce problème d'intermittence est le stockage de l'électricité en grandes quantités. Cette possibilité n'est pas encore possible, mais la recherche scientifique, dans ce domaine, est en plein développement.

La prévision des productions des centrales solaires passe nécessairement par la prédiction du rayonnement solaire. Pour cela, plusieurs méthodes ont été développées. Ces méthodes peuvent être classées en deux grandes familles. La première famille consiste en l'utilisation des modèles empiriques qui se basent sur un ensemble de relations mathématiques sous forme de relations linéaires ou de développements polynomiales pour exprimer la sortie désirée en fonction d'un ou plusieurs variables d'entrée tout en ajustant les paramètres de régression correspondant à chaque modèle sur la base des mesures expérimentales. La deuxième famille repose sur l'utilisation des techniques du Machine Learning qui font partie du grand domaine de l'intelligence artificielle et qui sont des méthodes d'apprentissage automatiques.

Dans ce chapitre, nous présentons, en premier lieu, un état de l'art des divers modèles empiriques utilisés pour l'estimation des irradiances solaires. Nous présentons, ensuite et d'une manière un peu plus détaillée, les modèles empiriques que nous avons testés dans le présent travail. Il s'agira des modèles de la littérature et des modèles que nous avons, nous-mêmes, proposés. Finalement, les résultats d'estimations des irradiances solaires journalières globale horizontale, diffuse ou directe normale seront présentés et analysés.

2. Modèles empiriques d'estimation des irradiances solaires globales journalières sur plan horizontal

Parmi les premières méthodes développées pour modéliser l'irradiation solaire globale journalière sur plan horizontal (GHI) au sol, on trouve les modèles empiriques qui consistent en des régressions linéaires, polynomiales ou complexes établies empiriquement à partir des mesures effectuées en des sites géographiques données. Le premier modèle dans cette catégorie est celui dérivé par Angström en 1924 (Angstrom, 1924), à partir des données mesurées à Stockholm:

$$\frac{H}{H_c} = 0.25 + 0.75 \sigma \quad (2-1)$$

où H est l'irradiation globale journalière sur plan horizontal, H_c est cette même irradiation mais par ciel clair et σ est la fraction d'insolation.

En exploitant les données enregistrées à l'observatoire du Mont Stramlo, à Canberra, en Australie, Prescott (Prescott, 1940) a modifié l'équation d'Ångström, en 1940, en introduisant l'irradiation solaire journalière extraterrestre H_0 au lieu de H_c :

$$\frac{H}{H_0} = 0.25 + 0.54 \sigma \quad (2-2)$$

Cette dernière équation est généralement connue sous le nom d'équation d'Ångström–Prescott. L'irradiation solaire journalières d'une surface horizontale au sommet de l'atmosphère H_0 [Wh/m²/jour] est donnée par l'expression suivante (Iqbal, 2012):

$$H_0 = \frac{24 \cdot I_0}{\pi} \left(1 + 0.033 \cos\left(\frac{360 \cdot DOY}{365}\right) \right) (\cos(\phi) \cos(\delta) \sin(\omega_s) + \frac{\pi \cdot \omega_s}{180} \sin(\phi) \sin(\delta)) \quad (2-3)$$

où I_0 est la constante solaire, DOY est le jour de l'année, ϕ est la latitude du lieu considéré, ω_s est l'angle horaire du coucher du soleil (en degrés) et δ est la déclinaison solaire.

La quantité :

$$K_t = \frac{H}{H_0} \quad (2-4)$$

représente ce qu'on appelle "indice de clarté".

L'équation d'Ångström–Prescott et ses variantes (quadratiques, cubiques, logarithmiques et exponentielles) ont été largement utilisées pour estimer la puissance du rayonnement solaire reçu au sol dans plusieurs sites sur terre.

Plusieurs auteurs ((Benson et al., 1984), (Ögelman et al., 1984), (Zabara, 1986), (Gopinathan, 1988), (Akinoğlu and Ecevit, 1990), (Akinoğlu, 1991), (Chelbi et al., 2015) ...) ont évalué la précision prédictive d'une multitude de corrélations empiriques basées sur la durée d'ensoleillement. Les résultats montrent que ces modèles sont, en général, capables d'estimer, avec une précision acceptable, le potentiel solaire disponible au niveau du sol. Ils sont, cependant, limités par le manque de disponibilité des enregistrements de la durée d'ensoleillement (Besharat et al., 2013). En effet, les données de la durée d'ensoleillement ne sont pas disponibles dans la plupart des stations météorologiques (Feng et al., 2017).

Par conséquent, plusieurs chercheurs ont essayé d'établir des relations empiriques entre le rayonnement solaire incident au sol et des variables couramment mesurées dans les stations météorologiques telles que la température ambiante, l'humidité relative de l'air, la vitesse et la direction du vent et les précipitations ((Besharat et al., 2013), (Despotovic et al., 2016), (Feng et al., 2020)). Dans le présent chapitre, on évaluera la précision de quelques modèles empiriques d'estimation de l'irradiation solaire globale journalière GHI, notée H , en utilisant des variables communément mesurées dans les stations météorologiques. On testera, aussi, quelques modèles d'estimation des irradiations solaires diffuses et directes journalières.

2.1. Modèles empiriques basés sur la température ou l'humidité relative de l'air

En raison de la grande disponibilité des données de la température dans les stations météorologiques, plusieurs chercheurs se sont intéressés à la modélisation du GHI à partir des seules mesures de la température de l'air. Le premier modèle dans cette catégorie a été proposé

par (Hargreaves and Samani, 1982). Ce modèle relie l'indice de clarté $\left(\frac{H}{H_0}\right)$ aux gradient journalier de la température de l'air $\Delta T(^{\circ}C)$:

$$\frac{H}{H_0} = a\Delta T^{0.5} \quad (2-5)$$

où $\Delta T = T_{max} - T_{min}$, avec T_{max} et T_{min} ($^{\circ}C$) les températures journalières maximale et minimale, respectivement, a est le coefficient empirique à calibrer. Hargreaves et al. (Hargreaves et al., 1985) modifient ce modèle pour l'écrire sous la forme:

$$\frac{H}{H_0} = a + b\Delta T^{0.5} \quad (2-6)$$

a et b sont des coefficients empiriques.

Quej et al. (Quej et al., 2016) ont évalué la précision et l'applicabilité de treize modèles empiriques dans six stations au Mexique (péninsule du Yucatán). Leurs résultats montrent que les modèles basés sur la température, les précipitations et l'humidité relative de l'air donnent de meilleurs résultats que les modèles utilisant uniquement des données de température. Toutefois, ils ont suggéré d'utiliser le modèle de Bristow et Campbell (Bristow and Campbell, 1984) si seules les données de température sont disponibles pour l'estimation du rayonnement solaire global. Ce modèle est donné par la relation:

$$\frac{H}{H_0} = a (1 - \exp(-b \Delta T^c)) \quad (2-7)$$

En utilisant des données à long terme provenant de 15 sites en Chine, Liu et al. (Liu et al., 2009) ont trouvé que le modèle de (Bristow and Campbell, 1984) avait des performances similaires à celles de la version modifiée du modèle de Hargreaves ((Hargreaves and Samani, 1982)).

Fan et al. (Fan et al., 2018a) ont testé 14 modèles empiriques existants et ont développé six nouveaux modèles pour l'estimation du GHI quotidien dans les régions subtropicales et tropicales humides de la Chine. Les données météorologiques de 1966 à 2015 provenant de 20 stations de mesure ont été utilisées. Pour les sites où seules les températures de l'air sont disponibles, ces chercheurs recommandent le modèle suivant :

$$H = H_0 (a + b\Delta T + c\Delta T^{0.25} + d\Delta T^{0.5}) + eT \quad (2-8)$$

D'autres modèles basés sur différents paramètres météorologiques ont, aussi, été proposés (voir section 2.2).

En général, il a été constaté que les modèles basés sur la température ont un bon potentiel pour estimer le rayonnement solaire global pour tous les endroits où les mesures des données de la durée d'ensoleillement ne sont pas disponible (Meenal et al., 2016).

Contrairement aux modèles empiriques basés sur la température, les modèles basés, exclusivement, sur l'humidité relative sont très rares ((Mubiru et al., 2007), (Korachagaon and Bapat, 2012), (Adeala et al., 2015), (Kolebaje et al., 2016)). Les mesures de l'humidité relative sont plus utilisées dans les modèles hybrides faisant intervenir plus d'un paramètre météorologique. Dans le présent travail, on testera les trois modèles suivants (Tableau 2.1):

- Modèle de (Mubiru et al., 2007):

$$\frac{H}{H_0} = a + bRH \quad (2-9)$$

- Modèle de (Kolebaje et al., 2016) :

$$\frac{H}{H_0} = a + bRH + cRH^d \quad (2-10)$$

- Modèle de (Korachagaon and Bapat, 2012):

$$\frac{H}{H_0} = a + bRH + cRH^2 + dRH^3 + eRH^4 + fRH^5 + gRH^6 + hRH^7 + iRH^8 + jRH^9 + kRH^{10} \quad (2-11)$$

RH est l'humidité relative et a, b, \dots, k sont les coefficients de régressions qui seront calibrés à l'aide des mesures.

2.2. Modèles empiriques basés sur des paramètres hybrides

Etant donné que le rayonnement solaire incident au sol varie régulièrement avec les conditions climatiques, plusieurs chercheurs ont pensé à construire des modèles hybrides basés sur des relations empiriques combinant deux ou plusieurs paramètres météorologiques tels que la température et l'humidité relative de l'air, la pression atmosphérique, la durée d'ensoleillement, la couverture nuageuse, les précipitations,... ((Lewis, 1983), (Garg and Garg, 1983), (Ododo, 1997), (Korachagaon and Bapat, 2012), (Chukwujindu, 2017), ...).

Dans leur étude, Fan et al (Fan et al., 2018a) ont constaté que le GHI quotidien peut être estimé avec une précision raisonnable en utilisant les modèles basés uniquement sur les données de la température de l'air. Toutefois, la précision de ces modèles est nettement améliorée si les précipitations et l'humidité relative quotidiennes sont également incluses. Ils ont noté que les performances de deux modèles incorporant une forme logarithmique des précipitations et la valeur réelle de l'humidité relative étaient supérieures à celles des autres modèles hybrides testés. Ces deux modèles sont donnés par les équations suivantes :

$$H = H_0 (a + b\Delta T + c\Delta T^{0.25} + d\Delta T^{0.5})[1 + e\varphi(P)] + fT \quad (2-12)$$

$$H = H_0 (a + b\Delta T + c\Delta T^{0.25} + d\Delta T^{0.5})[1 + e\varphi(P)] + fT + gRH \quad (2-13)$$

où P est la quantité réelle des précipitations journalières en (mm), $\varphi(P) = \ln(P + 1)$, \ln étant le logarithme népérien, a , b , c , d , e , f , et g sont des coefficients empiriques et RH est l'humidité relative de l'air. La forme logarithmique $\ln(P+1)$ a été utilisée au lieu de $\ln(P)$ car aucune solution ou des valeurs négatives sont obtenues pour la fonction logarithmique lorsque les précipitations quotidiennes sont nulles ou inférieures à 1 mm. Ces deux derniers modèles ne seront testés que pour la ville de Fès, pour laquelle on dispose des mesures des précipitations.

Dans la littérature, il existe plusieurs centaines de modèles empiriques pour l'estimation de GHI à partir des paramètres couramment mesurés dans les stations météorologiques. Pour tester le maximum de modèles qui peuvent être représentative sans alourdir notre tâche, nous nous sommes basés sur une étude effectuée par Chen et al., 2019. Dans cette étude, (Chen et al., 2019) ont examiné 294 modèles empiriques employant des variables météorologiques couramment mesurées et des facteurs géographiques ou astronomiques pour estimer le GHI au sol. Ces modèles ont été classés en 37 groupes en fonction de la nature des variables entrées utilisées. Les "trois meilleurs" modèles ayant les plus petits indicateurs d'erreur ont été identifiés pour chaque groupe.

Le tableau 2.1 regroupe les trois meilleurs modèles des catégories basées sur les paramètres communément mesurés dans les stations météorologiques et dont nous disposons dans nos bases de données. Ces modèles ont été testés pour estimer les GHI journalières dans six villes marocaines (Bounoua et al., 2021).

Tableau 2. 1. Modèles empiriques testés pour estimer l'irradiation global horizontale dans les six villes marocaines considérées.

Catégorie du modèle	ID du modèle	Equation	Référence
Température	T1	$\frac{H}{H_0} = a + (b + c\Delta T + d(\Delta T)^2 + e(\Delta T)^3)(\Delta T)^f$	(Hassan et al., 2016)
	T2	$\frac{H}{H_0} = (a + bT + cT^2 + dT^3)(\Delta T)^e$	(Hassan et al., 2016)
	T3	$\frac{H}{H_0} = a + (b + cT + dT^2)(\Delta T)^e$	(Hassan et al., 2016)
Température - Humidité relative	TR1	$H = a + H_0(bT_{max} + cT_{min} + dRH)$	(Li et al., 2013)
	TR2	$H = a + H_0(bT_{max} + cT_{min}) + dRH$	(Li et al., 2013)
	TR3	$\frac{H}{H_0} = a + bT_{max} + cT_{min} + d\Delta T + eRH$	(Korachagaon and Bapat, 2012)
Température - Pression	TPr1	$\frac{H}{H_0} = a + b(\Delta T)^{0.5} + cD$	(Chen et al., 2006)
	TPr2	$H = a + H_0(b + c(\Delta T)^{0.5}) + dD$	(Chen et al., 2006)
	TPr3	$H = a + H_0(b + c(\Delta T)^{0.5}) + d\log(D)$	(Chen et al., 2006)
Température - Humidité relative - Pression	TRPr1	$\frac{H}{H_0} = a + bT_{max} + cT_{min} + dT_{min} * T_{max} + eRH + fP_v$	(Chen and Li, 2013)
	TRPr2	$\frac{H}{H_0} = a + b(\Delta T)^{0.5} + cRH + dP_v$	(Chen and Li, 2013)
	TRPr3	$\frac{H}{H_0} = a + bT_{max} + cT_{min} + dRH + fP_v$	(Chen and Li, 2013)
Température - Facteurs géographiques	TG1	$\frac{H}{H_0} = a \left(1 + b \cos \left(\frac{cDOY\pi}{180} + d \right) \right) (1 - \exp(e(\Delta T)^f))$	(Quej et al., 2016)
	TG2	$H = (a(\Delta T)^b ICSKY^c - 2.4999)/0.8023$	(Mahmood and Hubbard, 2002)
	TG3	$\frac{H}{H_0} = a(1 + 2.7 * 10^{-5} Z)(\Delta T)^{0.5}$	(Annandale et al., 2002)
Température - Humidité relative - Facteurs géographiques	TRG1	$\frac{H}{H_0} = a \left(1 + b \cos \left(\frac{cDOY\pi}{180} + d \right) + eRH \right) \Delta T$	(Li et al., 2015)
	TRG2	$\frac{H}{H_0} = a \left(1 + b \cos \left(\frac{cDOY\pi}{180} + d \right) + eRH \right) (1 - \exp(f(\Delta T)^g))$	(Li et al., 2015)
	TRG3	$\frac{H}{H_0} = a + bW + c\delta$	(Garg and Garg, 1983)
Humidité relative	R1	$\frac{H}{H_0} = a + bRH + cRH^2 + dRH^3 + eRH^4 + fRH^5 + gRH^6 + hRH^7 + iRH^8 + jRH^9 + kRH^{10}$	(Korachagaon and Bapat, 2012)
	R2	$\frac{H}{H_0} = a + bRH + cRH^d$	(Kolebaje et al., 2016)
	R3	$\frac{H}{H_0} = a + bRH$	(Mubiru et al., 2007)
Pression	Pr	$H = a + bH_0 + cP_r$	(Skeiker, 2006)

Dans ce tableau, a, b, c, d, \dots, k sont les coefficients empiriques, T est la température moyenne journalière (°C), T_{max} et T_{min} sont les températures maximales et minimales journalières (°C), ΔT est le gradient de température journalier ($\Delta T = T_{max} - T_{min}$)(°C), RH est l'humidité relative moyenne journalière (%), Z est l'altitude du site en m, D est le déficit de pression de

vapeur de l'air défini comme étant la différence entre la pression de vapeur saturante (P_s) et la pression de vapeur réelle (P_v) en kPa pour une période de temps donnée :

$$D = V_{PD} = P_s - P_v \quad (2-14)$$

La pression de vapeur saturante peut être calculée à partir de la température de l'air (Allen, 1997):

$$P_s(T) = 0.6108 \exp \left[\frac{17.27 T}{T + 237.3} \right] \quad (2-15)$$

T étant la température de l'air [$^{\circ}\text{C}$].

La moyenne de la pression de vapeur saturante pour un jour, une semaine, une décennie ou un mois doit être calculée comme la moyenne entre la pression de vapeur saturante aux moyennes quotidiennes des températures maximales et minimales de l'air pour cette période (Allen, 1997):

$$P_{sj} = \frac{P_s(T_{max}) + P_s(T_{min})}{2} \quad (2-16)$$

où $P_s(T_{min})$ est la pression de vapeur saturante à la température minimale quotidienne [kPa], $P_s(T_{max})$ est la pression de vapeur saturante à la température maximale quotidienne [kPa].

La pression de vapeur réelle (P_v) peut être calculée à partir de l'humidité relative (Allen, 1997):

$$P_v = \frac{P_s(T_{max}) \frac{RH_{min}}{100} + P_s(T_{min}) \frac{RH_{max}}{100}}{2} \quad (2-17)$$

où RH_{max} et RH_{min} sont respectivement l'humidité relative maximale et minimale.

Dans le modèle TG2, le rayonnement solaire quotidien corrigé du ciel clair, ICSKY, est donné par la relation (Mahmood and Hubbard, 2002):

$$ICSKY = T_t * I_s \quad (2-18)$$

où,

$$T_t = 0.8 + 0.12 * [|182 - DOY| / 183]^{1.5} \quad (2-19)$$

est le coefficient de transmissivité, et,

$$I_S = 0.04188 * (A_c + B_c \sin[(2\pi (DOY + 10.5)/365) - \pi/2]) \quad (2-20)$$

est le rayonnement solaire pour un jour clair [MJ/m²/jour], avec

$$A_c = (0.29 \cos(\phi) + 0.52)[\sin(\phi)(46.355 LD - 574.3885) + 816.41 \cos(\phi)\sin(\pi LD/24)] \quad (2-21)$$

$$B_c = (0.29 \cos(\phi) + 0.52)[\sin(\phi)(574.3885 - 1.509 LD) - 29.59 \cos(\phi)\sin(\pi LD/24)] \quad (2-22)$$

où LD [heure] est la durée du plus long jour de l'année (DOY).

LD peut être estimé à partir de :

$$LD = 0.267 \sin^{-1}[0.5 + 0.007895/\cos(\phi) + 0.2168875 \tan(\phi)]^{0.5} \quad (2-23)$$

Dans le modèle TGR3, W [cm] est la vapeur d'eau atmosphérique précipitable par unité de volume d'air calculée à partir de :

$$W = 0.0049 * RH \left[\frac{\exp\left(26.23 - \frac{5416}{T_K}\right)}{T_K} \right] \quad (2-24)$$

T_K est la température moyenne de l'air en degré Kelvin (K).

En plus des modèles du tableau 2.1, nous avons aussi testé les modèles représentés par les équations (2.5) à (2.8) pour l'ensemble des six stations. Les modèles donnés par les équations (2.12) et (2.13) sont testés uniquement pour Fès, car, on ne dispose pas les données des précipitations pour les stations enerMENA. En plus de ces modèles, nous avons testé plusieurs modèles empiriques que nous avons nous-mêmes proposés basés uniquement sur l'humidité relative ou ceux dans lesquels on a combiné différents paramètres météorologiques couramment mesurés, pour améliorer la précision d'estimation journalière de GHI. Le modèle des paramètres hybrides MH3 est testé uniquement pour les stations d'enerMENA, vu l'indisponibilité des mesures de la pression pour la ville de Fès. Le tableau 2.2, ci-dessous, regroupe les modèles empiriques testés en plus des modèles du tableau 2.1. Dans le tableau 2.3, sont présentés les modèles testés pour la ville de Fès uniquement.

Tableau 2. 2. Modèles empiriques testés pour l'estimation journalière de GHI.

Modèle	Equation	Références
MT1	$H = H_0 a \Delta T^{0.5}$	(Hargreaves and Samani, 1982)
MT2	$H = H_0 (a + b \Delta T^{0.5})$	(Hargreaves and Samani, 1982)
MT3	$H = H_0 a (1 - \exp(-b \Delta T^c))$	(Bristow et Campbell, 1984)
MT4	$H = H_0 (a + b \Delta T + c \Delta T^{0.25} + d \Delta T^{0.5})$	(Fan et al., 2018a)
MT5	$H = H_0 (a + b \Delta T + c \Delta T^{0.25} + d \Delta T^{0.5}) + eT$	(Fan et al., 2018a)
MRH1	$H = H_0 (a + b RH_{max} + c RH_{min})$	Proposé
MH1	$H = H_0 (a + b \Delta T^{0.5} + c \Delta T + d \Delta RH^{0.5} + eT)$	Proposé
MH2	$H = H_0 \left(a + b \Delta T^{0.5} + c \Delta T + d \frac{T_{min}}{T_{max}} + e \frac{RH_{min}}{RH_{max}} \right)$	Proposé
MH3	$H = H_0 (a + b T + c RH + d \Delta T + e W_s + f P_r)$	Proposé

Tableau 2. 3. Modèles empiriques testés pour l'estimation journalière de GHI pour la ville de Fès uniquement.

Modèle	Equation	Références
MH4	$H = H_0 (a + b \Delta T + c \Delta T^{0.25} + d \Delta T^{0.5}) [1 + e \varphi(P)] + fT$	(Fan et al., 2018a)
MH5	$H = H_0 (a + b \Delta T + c \Delta T^{0.25} + d \Delta T^{0.5}) [1 + e \varphi(P)] + fT + gRH$	(Fan et al., 2018a)
MH6	$H = H_0 (a + b \Delta T^{0.25} + c \Delta T^{0.5} + dT) [1 + e \varphi(P)]$	Proposé

a, b, \dots, h sont les coefficients de régressions qui seront calibré à l'aide des mesures.

Dans les tableaux 2.2 et 2.3, les modèles MT représentent les modèles basés uniquement sur les mesures de la température ambiante, les MRH indiquent les modèles basés uniquement sur les mesures de l'humidité relative, et les MH représentent les modèles hybrides basés sur des combinaisons de deux ou plusieurs paramètres météorologiques.

3. Modèles empiriques d'estimation des irradiations solaires directes et diffuses journalières

Les modèles empiriques d'estimation des irradiations solaires directes normales (DNI) et diffuses horizontales (DHI) sont très peu nombreux et sont généralement basés sur des équations de régressions multiples faisant intervenir, en particulier, l'indice de clarté K_t .

Sur la base des données mesurées ou estimées du rayonnement global reçu par un plan horizontal, (Liu and Jordan, 1960) ont proposé un premier modèle qui permet de déterminer la corrélation entre l'indice de clarté et l'intensité instantanée du rayonnement diffus par temps

clair, les moyennes horaires et journalières à long terme du rayonnement diffus et les sommes journalières du rayonnement diffus pour des jours correspondants aux différents degrés de nébulosité. Par la suite, plusieurs modèles ont été proposés par de nombreux chercheurs pour estimer le rayonnement solaire diffus horizontal en fonction de l'indice de clarté et autres paramètres météorologiques dans un certain nombre d'emplacements géographiques à différentes échelles du temps: horaire ((Ihya et al., 2015), (Ruiz-Arias et al., 2010)), journalière (Fan et al., 2019a), (Yang et al., 2020), (Ihya et al., 2015), ...) et mensuelle journalière (Khorasanizadeh et al., 2016), (Jamil and Siddiqui, 2017), (Jamil and Akhtar, 2017). D'après le travail effectué par (Nwokolo and Ogbulezie, 2018), qui représente une large étude bibliographique sur toutes les catégories des modèles empiriques et précisément les modèles basés sur l'indice de clarté pour estimer le rayonnement diffus, il a été mentionné que les modèles basés sur l'indice de clarté (44.18%) sont souvent utilisés par les chercheurs africains en énergie solaire.

Dans le tableau 2.4, on a reporté les modèles que nous avons testé dans le présent travail. Ce tableau contient des modèles basés sur les paramètres hybrides (HBMs) et l'indice de clarté (CIBMs), qui sont très cités dans la littérature et ainsi que ceux que nous avons-nous-mêmes proposés dans le présent travail, pour améliorer la précision d'estimation des irradiations solaires.

Tableau 2. 4. Modèles empiriques d'estimation des irradiations diffuses horizontales et directes normales journalières.

Groupe	Composantes solaires	Modèle	Equation	Références
HBMs	DHI	DH1	$DHI = GHI (a + b K_t + c RH)$	(Nwokolo and Ogbulezie, 2018)
		DH2	$DHI = GHI (a + b K_t + c \delta)$	(Nwokolo and Ogbulezie, 2018)
		DH3	$DHI = GHI (a + b K_t + c H_0 + d DOY + e D)$	Proposé
	DNI	DNH1	$DNI = H_{0N} (a K_t^2 + b K_t + c DOY + d T + e)$	Proposé
CIBMs	DHI	DC1	$DHI = H_0 (a + b K_t)$	(Khorasanizadeh and Mohammadi, 2016)
		DC2	$DHI = GHI (a + b K_t)$	(Nwokolo and Ogbulezie, 2018)
	DNI	DNC1	$DNI = H_{0N} (a K_t^2 + b K_t + c)$	(Louche et al., 1991)

Pour un jour donné de l'année (DOY), l'irradiation solaire normale hors atmosphère (H_{0N}), est donnée par l'expression suivante :

$$H_{0N} = I_0 \left(1 + 0.033 \cos \left(\frac{360 DOY}{365} \right) \right) \quad (2-25)$$

4. Résultats et discussions

Cette section présente les performances des modèles empiriques retenus pour l'estimation des irradiations solaires journalières (GHI, DHI et DNI). Pour tous les modèles, chaque base de données a été divisée en deux sous-ensembles : le premier (ensemble de calibration) a été utilisé pour calibrer les paramètres du modèle (70% de données) et le second (ensemble de validation) a été utilisé pour valider le modèle calibré (30% de données). Les performances de chaque modèle ont été évaluées à l'aide des indicateurs statistiques présentés en Annexe D.

4.1. Performances des modèles empiriques d'estimation du GHI journalière

Les 22 modèles empiriques du tableau 2.1 ont été appliqués aux bases des données des 5 stations marocaines du réseau enerMENA (Erfoud, Missouri, Oujda, TanTan, Zagora) (Bounoua et al., 2021). Chaque modèle a été testé pour chaque station individuellement et pour une base de données formée par l'ensemble des données des cinq stations (Toutes les villes). Le tableau 2.5 montre les coefficients de régression obtenus.

Tableau 2. 5. Coefficients de régression des 22 modèles empiriques appliqués aux stations de mesures du réseau enerMENA.

Modèles Empiriques	Stations	Coefficients de régression											
		a	b	c	d	e	f	g	h	i	j	k	
Température	T1	Erfoud	-0.345640	0.065295	-0.003949	0.000079	0.721940	0.610740	-	-	-	-	-
		Missour	-1.292900	0.129640	-0.006459	0.000114	0.110740	1.253800	-	-	-	-	-
		Oujda	-1.132800	0.157470	-0.008971	0.000170	0.231040	1.115500	-	-	-	-	-
		TanTan	-12.593500	0.283550	-0.027111	0.000800	0.010373	12.580500	-	-	-	-	-
		Zagora	2.114100	0.008631	-0.001745	0.000042	0.177900	-2.513500	-	-	-	-	-
		Toutes les villes	0.453440	-0.026048	0.002277	-0.000062	0.264800	-0.049709	-	-	-	-	-
	T2	Erfoud	0.253660	-0.004317	0.000194	-0.000003	0.406940	-	-	-	-	-	-
		Missour	0.132680	0.003223	-0.000221	0.000003	0.597200	-	-	-	-	-	-
		Oujda	0.138920	0.002618	-0.000164	0.000003	0.603270	-	-	-	-	-	-
		TanTan	2.498600	-0.313960	0.015473	-0.000253	0.248960	-	-	-	-	-	-
		Zagora	0.261740	-0.007202	0.000298	-0.000004	0.458480	-	-	-	-	-	-
		Toutes les villes	0.184120	0.023499	-0.001019	0.000013	0.259540	-	-	-	-	-	-
	T3	Erfoud	-9.305500	9.287800	-0.001533	0.000023	0.027481	-	-	-	-	-	-
		Missour	-3.775300	3.606200	-0.000569	-0.000109	0.084415	-	-	-	-	-	-
		Oujda	10.382500	10.224000	0.000671	-0.000061	0.031548	-	-	-	-	-	-
TanTan		-41.339600	42.187900	-0.049726	0.001156	0.003930	-	-	-	-	-	-	
Zagora		-8.757800	8.701900	-0.002117	0.000010	0.033103	-	-	-	-	-	-	
Toutes les villes		4.451300	4.623700	0.012263	-0.000275	0.029824	-	-	-	-	-	-	
Température - Humidité relative	TR1	Erfoud	0.027097	-0.021413	0.029676	1428.368100	-	-	-	-	-	-	-
		Missour	0.032806	0.033990	0.122020	1216.897300	-	-	-	-	-	-	-
		Oujda	0.036860	-0.033060	0.033817	1088.818300	-	-	-	-	-	-	-
		TanTan	0.034734	-0.031464	0.227040	1368.693200	-	-	-	-	-	-	-
		Zagora	0.030140	-0.026231	-0.021899	1650.004200	-	-	-	-	-	-	-
		Toutes les villes	0.026079	-0.019733	0.196570	1161.579900	-	-	-	-	-	-	-
	TR2	Erfoud	0.024836	-0.019737	-929.387900	2179.820400	-	-	-	-	-	-	-
		Missour	0.032771	-0.033790	36.326600	1673.106700	-	-	-	-	-	-	-
		Oujda	0.035883	-0.032431	-698.370500	1799.958600	-	-	-	-	-	-	-
		TanTan	0.035707	-0.025299	1597.083700	618.841200	-	-	-	-	-	-	-
		Zagora	0.028881	-0.025271	-718.469800	2001.570200	-	-	-	-	-	-	-
		Toutes les villes	0.026380	-0.018428	1621.116900	921.025000	-	-	-	-	-	-	-
	TR3	Erfoud	0.581710	0.294820	-0.297880	-0.281210	-0.22568	-	-	-	-	-	-
		Missour	0.480460	0.316790	-0.324420	-0.292700	-0.16922	-	-	-	-	-	-
		Oujda	0.486870	0.311700	-0.316020	-0.282550	-0.23802	-	-	-	-	-	-
TanTan		0.531680	0.197720	-0.205910	-0.174700	0.08129	-	-	-	-	-	-	
Zagora		0.512740	0.296370	-0.299590	-0.277580	-0.15622	-	-	-	-	-	-	
Toutes les villes		0.436060	0.292690	-0.292680	-0.275870	0.005970	-	-	-	-	-	-	

Le tableau 2.6 représente les indicateurs statistiques des performances obtenus pour les 22 modèles empiriques considérés (Tableau 2.1), pour tous les cas étudiés. Les meilleurs résultats de chaque catégorie de modèles sont présentés en fond gris.

La première observation montre qu'aucun des 22 modèles empiriques ne s'est avéré être le meilleur pour toutes les stations. Les performances des quatre meilleurs modèles obtenus sont regroupées dans le tableau 2.7.

Tableau 2. 5. (suite) Coefficients de régression des 22 modèles empiriques appliqués aux stations de mesures du réseau enerMENA.

Modèles Empiriques	Stations	Coefficients de régression											
		a	b	c	d	e	f	g	h	i	j	k	
Température - Pression	TPr1	Erfoud	0.100660	0.147320	0.000674	-	-	-	-	-	-	-	-
		Missour	-0.126650	0.229800	-0.044216	-	-	-	-	-	-	-	-
		Oujda	-0.136310	0.234370	-0.013620	-	-	-	-	-	-	-	-
		TanTan	0.218010	0.184140	-0.093777	-	-	-	-	-	-	-	-
		Zagora	0.038380	0.178200	-0.007107	-	-	-	-	-	-	-	-
		Toutes les villes	0.275930	0.111480	-0.006362	-	-	-	-	-	-	-	-
	TPr2	Erfoud	-649.452500	0.242830	0.138590	-110.958300	-	-	-	-	-	-	-
		Missour	440.355800	-0.140410	0.219950	-425.244200	-	-	-	-	-	-	-
		Oujda	305.496400	-0.126860	0.213690	3.147300	-	-	-	-	-	-	-
		TanTan	1042.931600	0.031641	0.217240	-889.419600	-	-	-	-	-	-	-
		Zagora	-82.134200	0.051054	0.180190	-102.445400	-	-	-	-	-	-	-
		Toutes les villes	-201.071100	0.318530	0.107820	-89.367000	-	-	-	-	-	-	-
	TPr3	Erfoud	-680.393400	0.224780	0.139700	-185.942200	-	-	-	-	-	-	-
		Missour	193.785900	-0.123670	0.204730	-423.037900	-	-	-	-	-	-	-
		Oujda	672.995900	-0.049131	0.178000	339.853100	-	-	-	-	-	-	-
		TanTan	558.098500	0.168020	0.162940	96.278700	-	-	-	-	-	-	-
		Zagora	3.060800	0.022505	0.179740	-136.197500	-	-	-	-	-	-	-
		Toutes les villes	-159.723600	0.321500	0.100190	-34.985300	-	-	-	-	-	-	-
Température - Humidité relative - Pression	TRPr1	Erfoud	0.531440	0.014596	-0.015022	0.000009	-0.08038	-0.073684	-	-	-	-	-
		Missour	0.440950	0.026805	-0.022062	-0.000372	-0.27830	0.044525	-	-	-	-	-
		Oujda	0.637560	0.026307	-0.034525	-0.000157	-0.54484	0.149060	-	-	-	-	-
		TanTan	1.867200	-0.020424	-0.094957	0.001862	-0.69876	0.348300	-	-	-	-	-
		Zagora	0.468910	0.019261	-0.019294	-0.000004	0.020241	-0.090834	-	-	-	-	-
		Toutes les villes	0.453760	0.017891	-0.011914	-0.000216	-0.15753	0.056154	-	-	-	-	-
	TRPr2	Erfoud	0.282110	0.119000	-0.058225	-0.070383	-	-	-	-	-	-	-
		Missour	0.007629	0.188630	0.101120	-0.123040	-	-	-	-	-	-	-
		Oujda	0.038284	0.204680	-0.087892	-0.038867	-	-	-	-	-	-	-
		TanTan	0.170970	0.147810	0.198640	-0.044736	-	-	-	-	-	-	-
		Zagora	0.169040	0.150060	0.056284	-0.090889	-	-	-	-	-	-	-
		Toutes les villes	0.239640	0.112060	-0.009598	0.025418	-	-	-	-	-	-	-
	TRPr3	Erfoud	0.527030	0.014737	-0.014690	-0.077168	-0.07525	-	-	-	-	-	-
		Missour	0.547010	0.022923	-0.035189	-0.332300	0.094114	-	-	-	-	-	-
		Oujda	0.685340	0.024479	-0.039676	-0.561660	0.163670	-	-	-	-	-	-
		TanTan	1.217600	0.011803	-0.057402	-0.706390	0.356720	-	-	-	-	-	-
		Zagora	0.470800	0.019203	-0.019434	0.018937	-0.09020	-	-	-	-	-	-
		Toutes les villes	0.543080	0.014735	-0.019719	-0.210520	0.094247	-	-	-	-	-	-

D'après ces résultats, le modèle Température-Facteurs Géographiques TG1 (Quej et al., 2016) peut être recommandé pour calculer les irradiations solaires globales journalières. Ce modèle nécessite uniquement la connaissance des mesures du gradient de température journalier (ΔT) et par conséquent la connaissance des températures maximales et minimales journalières. Ces deux variables sont mesurées dans toutes les stations météorologiques.

Notons aussi, que malgré l'introduction de l'humidité relative de l'air dans le modèle TGR2, on n'a pas noté d'amélioration significative dans les résultats.

Tableau 2. 5. (suite) Coefficients de régression des 22 modèles empiriques appliqués aux stations de mesures du réseau enerMENA.

Modèles Empiriques	Stations	Coefficients de régression													
		a	b	c	d	e	f	g	h	i	j	k			
Température - Facteurs géographiques	TG1	Erfoud	0.69677	-0.0521	1.2587	0.7216	0.0130	2.1308	-	-	-	-	-	-	-
		Missour	0.78029	-0.0690	1.04300	2.192500	0.052016	1.3807	-	-	-	-	-	-	-
		Oujda	0.72400	0.02078	5.08240	0.856260	0.029693	1.839000	-	-	-	-	-	-	-
		TanTan	0.76600	0.32998	0.04295	1.877400	0.184080	1.778600	-	-	-	-	-	-	-
		Zagora	0.75401	-0.0445	1.34440	0.89192	0.04205	1.57490	-	-	-	-	-	-	-
	Toutes les villes	1.89430	-0.0269	1.1716	0.9836	0.1968	0.3079	-	-	-	-	-	-	-	
	TG2	Erfoud	0.40098	0.31932	1.02050	-	-	-	-	-	-	-	-	-	-
		Missour	0.54974	0.41658	0.82886	-	-	-	-	-	-	-	-	-	-
		Oujda	0.44491	0.47632	0.87332	-	-	-	-	-	-	-	-	-	-
		TanTan	0.78297	0.26853	0.89722	-	-	-	-	-	-	-	-	-	-
		Zagora	0.40855	0.40183	0.95135	-	-	-	-	-	-	-	-	-	-
	Toutes les villes	0.74798	0.20362	0.92734	-	-	-	-	-	-	-	-	-	-	
	TG3	Erfoud	0.17057	-	-	-	-	-	-	-	-	-	-	-	-
		Missour	0.16865	-	-	-	-	-	-	-	-	-	-	-	-
		Oujda	0.18531	-	-	-	-	-	-	-	-	-	-	-	-
TanTan		0.26001	-	-	-	-	-	-	-	-	-	-	-	-	
Zagora		0.17867	-	-	-	-	-	-	-	-	-	-	-	-	
Toutes les villes	0.18237	-	-	-	-	-	-	-	-	-	-	-	-		
Température - Humidité relative - Facteurs géographiques	TRG1	Erfoud	0.70880	0.03057	1.70180	1.9023	-0.03076	0.019540	1.941900	-	-	-	-	-	-
		Missour	1.87620	0.18765	0.37002	1.714400	-0.0209	0.0883	0.6787	-	-	-	-	-	-
		Oujda	0.92362	-0.0073	-0.6106	3.3339	-0.56476	2.544600	1.569400	-	-	-	-	-	-
		TanTan	0.74255	0.08151	0.59346	1.1695	-0.07810	0.481010	1.011500	-	-	-	-	-	-
		Zagora	0.75696	0.03045	1.86870	1.625100	0.080710	0.043296	1.522500	-	-	-	-	-	-
	Toutes les villes	1.41210	-0.0081	2.18070	2.008500	0.010577	0.262920	0.343790	-	-	-	-	-	-	
	TRG2	Erfoud	0.17414	-0.0578	1.30990	0.236370	0.021805	-	-	-	-	-	-	-	-
		Missour	0.16833	-0.0020	-2.0624	-0.48455	0.073353	-	-	-	-	-	-	-	-
		Oujda	0.11658	1.09160	-0.0040	5.47100	-0.22826	-	-	-	-	-	-	-	-
		TanTan	0.14827	0.04265	1.70810	3.495500	0.933160	-	-	-	-	-	-	-	-
		Zagora	0.17951	-0.0273	4.2898	3.465100	0.069106	-	-	-	-	-	-	-	-
	Toutes les villes	0.16325	-0.0107	4.84760	-2.34370	0.342550	-	-	-	-	-	-	-	-	
	TRG3	Erfoud	0.77600	0.00034	-8.9033	-	-	-	-	-	-	-	-	-	-
		Missour	0.69187	-0.0004	-2.3321	-	-	-	-	-	-	-	-	-	-
		Oujda	0.61111	-0.0003	0.37058	-	-	-	-	-	-	-	-	-	-
TanTan		0.66939	0.00010	-3.0453	-	-	-	-	-	-	-	-	-	-	
Zagora		0.81961	0.00050	-11.423	-	-	-	-	-	-	-	-	-	-	
Toutes les villes	0.71689	0.00002	-4.2757	-	-	-	-	-	-	-	-	-	-		
R1	Erfoud	4.6720	-148.35	2272.9	-18945.6	95731.2	-309044.7	651026.1	-890365.3	761422.3	-369627	77690	-	-	
	Missour	9.16680	-122.26	461.857	1768.796	-22353.4	91601.493	-208402.7	288229.56	-241594	113235.2	-22837	-	-	
	Oujda	2.5232	-59.81	837.37	-6490.93	30638.5	-92230.8	180431.9	-228094.7	179491.4	-79857.5	15332.1	-	-	
	TanTan	-287.31	6033.1	-54394	278743.5	-902055	1932380.6	-2783595	2670199	-163687	580500.9	-90649	-	-	
	Zagora	1.76480	-41.327	686.588	-6272.58	34627.34	-121062.6	273224.99	-396000.9	355337.8	-179454	38963.5	-	-	
Toutes les villes	0.94861	-6.9790	96.2220	-795.747	4048.334	-13058.58	27190.45	-36355.54	30038.83	-13904.2	2746.29	-	-		
Humidité relative	R2	Erfoud	0.69247	-0.1074	-0.3917	4.972100	-	-	-	-	-	-	-	-	-
		Missour	0.5043	1.1088	-1.5142	1.965500	-	-	-	-	-	-	-	-	-
		Oujda	-0.0340	-48.241	48.4634	0.962740	-	-	-	-	-	-	-	-	-
		TanTan	-0.6375	-88.767	89.8496	0.975120	-	-	-	-	-	-	-	-	-
		Zagora	0.72403	-0.1483	-1.1510	9.987300	-	-	-	-	-	-	-	-	-
Toutes les villes	0.73253	-0.1826	-0.6352	28.2416	-	-	-	-	-	-	-	-	-		
R3	Erfoud	0.72066	-0.2273	-	-	-	-	-	-	-	-	-	-	-	
	Missour	0.82892	-0.3852	-	-	-	-	-	-	-	-	-	-	-	
	Oujda	0.92200	-0.5186	-	-	-	-	-	-	-	-	-	-	-	
	TanTan	0.70721	-0.1585	-	-	-	-	-	-	-	-	-	-	-	
	Zagora	0.74368	-0.2431	-	-	-	-	-	-	-	-	-	-	-	
Toutes les villes	0.74712	-0.2256	-	-	-	-	-	-	-	-	-	-	-		
Pression	Pr	Erfoud	-2101.13	0.72115	1.62110	-	-	-	-	-	-	-	-	-	-
		Missour	-69802.3	0.76981	76.3534	-	-	-	-	-	-	-	-	-	-
		Oujda	-1099.30	0.76328	-0.0809	-	-	-	-	-	-	-	-	-	-
		TanTan	-62197.5	0.45436	62.3320	-	-	-	-	-	-	-	-	-	-
		Zagora	5694.43	0.70578	-6.3202	-	-	-	-	-	-	-	-	-	-
Toutes les villes	-511.163	0.69946	-0.0083	-	-	-	-	-	-	-	-	-	-	-	

Selon les résultats obtenus, on peut signaler aussi que le classement donné par (Chen et al., 2019) n'est pas respecté pour certaines catégories de modèles. On constate, par exemple, que le modèle TR3, classé au rang 3 dans la catégorie TR, a donné de meilleures performances que les modèles classés aux rangs 1 (TR1) et 2 (TR2). Cela peut signifier que les performances des modèles empiriques et, par conséquent, les influences des paramètres météorologiques sur le GHI peuvent varier d'un endroit géographique à l'autre.

Tableau 2. 6. Indicateurs statistiques de performances des 22 modèles empiriques pour l'estimation du GHI dans les six cas étudiés (Bounoua et al., 2021).

Modèles empiriques	Stations	Calibration			Validation			
		R (%)	nMAE (%)	nRMSE (%)	R (%)	nMAE (%)	nRMSE (%)	
Température	T1	Erfoud	93.09	7.92	10.76	89.95	9.11	12.51
		Missour	89.54	11.89	15.05	85.38	13.19	17.48
		Oujda	93.36	10.34	14.89	92.78	10.95	15.02
		TanTan	72.12	17.43	21.98	71.13	18.27	22.81
		Zagora	92.39	7.87	11.01	93.07	7.39	10.18
		Toutes les villes	86.34	12.78	17.26	86.01	12.96	17.43
	T2	Erfoud	91.29	9.06	12.13	89.54	9.84	12.69
		Missour	90.15	11.20	14.55	87.53	12.24	15.84
		Oujda	91.97	12.02	16.29	91.07	12.81	16.56
		TanTan	63.52	19.79	24.47	66.54	19.59	24.28
		Zagora	91.95	8.18	11.31	92.73	7.78	10.48
		Toutes les villes	86.60	12.60	17.10	86.45	12.79	17.17
	T3	Erfoud	91.59	8.92	11.90	89.67	9.70	12.62
		Missour	90.42	10.98	14.35	87.89	11.98	15.66
		Oujda	92.70	11.21	15.59	91.91	12.10	15.85
		TanTan	65.29	19.38	23.99	66.69	19.53	24.23
		Zagora	92.32	8.04	11.05	93.03	7.61	10.26
		Toutes les villes	86.50	12.65	17.17	86.42	12.76	17.19
Température - Humidité relative	TR1	Erfoud	87.59	11.18	14.21	86.02	11.51	14.60
		Missour	87.33	12.91	16.37	84.91	13.45	17.39
		Oujda	90.69	13.17	17.47	88.42	14.38	18.76
		TanTan	56.93	20.65	26.04	58.69	20.99	26.35
		Zagora	89.50	9.85	12.80	90.12	9.58	12.09
		Toutes les villes	84.10	14.09	18.51	83.17	14.48	18.97
	TR2	Erfoud	87.73	11.25	14.13	86.40	11.36	14.40
		Missour	86.94	13.17	16.60	84.03	13.94	17.87
		Oujda	90.75	13.12	17.41	88.48	14.18	18.71
		TanTan	55.87	20.67	26.27	56.58	21.23	26.80
		Zagora	89.58	9.90	12.75	90.09	9.65	12.11
		Toutes les villes	82.72	14.63	19.22	81.60	15.11	19.75
	TR3	Erfoud	91.52	8.98	11.92	90.30	9.50	12.24
		Missour	89.94	11.44	14.69	87.73	12.41	15.78
		Oujda	91.55	12.71	16.75	90.65	13.42	17.09
		TanTan	58.21	20.71	25.82	63.43	20.30	25.11
		Zagora	91.76	8.47	11.42	92.61	8.02	10.56
		Toutes les villes	86.07	13.04	17.42	85.49	13.38	17.73
Température - Pression	TPr1	Erfoud	91.02	9.21	12.24	89.37	9.93	12.80
		Missour	89.96	11.20	14.68	87.61	12.11	15.91
		Oujda	92.12	11.93	16.22	91.12	12.93	16.70
		TanTan	63.10	19.60	24.68	64.92	19.60	24.67
		Zagora	91.72	8.37	11.44	92.53	7.88	10.59
		Toutes les villes	86.30	12.79	17.30	86.10	13.01	17.38
	TPr2	Erfoud	91.30	8.81	12.01	89.73	9.67	12.65
		Missour	90.00	11.17	14.65	87.69	12.06	15.81
		Oujda	92.19	11.92	16.06	90.86	12.97	16.71
		TanTan	64.85	19.30	24.11	63.36	20.00	25.12
		Zagora	91.77	8.32	11.40	92.62	7.74	10.51
		Toutes les villes	86.38	12.77	17.23	86.15	12.99	17.35
	TPr3	Erfoud	91.22	8.92	12.07	89.57	9.76	12.75
		Missour	89.24	11.89	15.16	86.74	12.71	16.37
		Oujda	92.47	11.74	15.78	90.84	12.78	16.74
		TanTan	62.75	19.96	24.66	63.61	20.38	25.08
		Zagora	91.66	8.39	11.47	92.60	7.84	10.52
		All cities	86.27	12.89	17.30	85.92	13.14	17.48
Température - Humidité relative - Pression	TRPr1	Erfoud	91.55	9.05	11.89	90.31	9.51	12.24
		Missour	90.44	11.11	14.34	88.02	12.12	15.60
		Oujda	92.05	12.35	16.25	91.65	12.68	16.06
		TanTan	62.09	20.09	24.84	63.19	20.08	25.13
		Zagora	91.85	8.47	11.35	92.96	7.90	10.33
		Toutes les villes	86.45	12.86	17.20	85.82	13.19	17.54
	TRPr2	Erfoud	91.90	8.85	11.65	90.47	9.38	12.15
		Missour	89.86	11.64	14.76	87.56	12.42	15.97
		Oujda	92.30	11.95	16.06	91.22	12.96	16.69
		TanTan	61.53	20.05	25.02	64.93	19.88	24.69
		Zagora	92.29	8.23	11.05	93.27	7.69	10.10
		Toutes les villes	86.21	12.96	17.34	85.83	13.19	17.54
	TRPr3	Erfoud	91.55	9.04	11.88	90.32	9.50	12.23
		Missour	90.14	11.27	14.56	87.71	12.42	15.78
		Oujda	91.99	12.38	16.31	91.49	12.85	16.22
		TanTan	61.49	20.20	25.00	63.60	20.16	25.01
		Zagora	91.85	8.48	11.35	92.96	7.90	10.33
		Toutes les villes	86.27	13.00	17.30	85.68	13.28	17.62

Tableau 2. 6. (Suite) Indicateurs statistiques de performances des 22 modèles empiriques pour l'estimation du GHI dans les six cas étudiés (Bounoua et al., 2021).

Modèles empiriques	Stations	Calibration			Validation			
		R (%)	nMAE (%)	nRMSE (%)	R (%)	nMAE (%)	nRMSE (%)	
Température – Facteurs géographiques	TG1	Erfoud	93.71	7.42	10.28	90.69	8.65	12.10
		Missour	90.35	11.08	14.40	87.29	12.22	16.07
		Oujda	93.40	10.37	14.86	92.73	11.03	15.09
		TanTan	72.51	17.27	21.96	72.38	17.94	22.39
		Zagora	92.88	7.47	10.63	93.46	6.96	9.89
		Toutes les villes	86.37	12.65	17.24	86.34	12.83	17.23
	TG2	Erfoud	91.68	8.83	11.77	89.84	9.78	12.57
		Missour	89.15	12.22	15.23	85.79	12.92	16.86
		Oujda	92.46	11.65	15.80	91.22	12.60	16.35
		TanTan	61.85	20.50	24.89	62.58	20.99	25.35
		Zagora	91.84	8.39	11.35	92.83	7.96	10.36
		Toutes les villes	85.32	13.22	17.84	85.36	13.32	17.79
	TG3	Erfoud	90.64	9.58	12.45	88.82	10.39	13.18
		Missour	88.93	12.39	15.58	85.37	13.76	17.71
		Oujda	92.04	12.01	16.23	90.80	13.04	16.82
		TanTan	61.02	20.40	25.55	58.53	21.55	26.96
		Zagora	91.62	8.59	11.54	92.47	8.02	10.62
		Toutes les villes	83.06	14.65	20.44	83.14	14.86	20.47
Température – Humidité relative – Facteurs géographiques	TRG1	Erfoud	93.35	7.77	10.57	90.69	8.83	12.15
		Missour	89.81	11.71	14.78	86.82	12.74	16.32
		Oujda	87.37	15.76	20.17	86.67	15.64	19.92
		TanTan	69.37	18.49	22.96	70.29	18.78	23.40
		Zagora	92.78	7.62	10.72	93.36	7.15	9.96
		Toutes les villes	86.20	12.85	17.35	86.12	13.00	17.37
	TRG2	Erfoud	91.57	9.08	11.87	89.22	10.34	13.16
		Missour	88.89	12.30	15.53	85.39	13.54	17.55
		Oujda	92.18	11.99	16.14	90.92	13.19	16.91
		TanTan	63.47	19.58	24.86	62.00	20.35	25.84
		Zagora	91.88	8.27	11.33	92.55	7.78	10.54
		Toutes les villes	83.37	14.05	19.24	83.73	14.15	19.08
	TRG3	Erfoud	90.14	9.69	13.26	89.29	9.68	13.01
		Missour	83.23	16.06	19.30	78.51	16.38	20.33
		Oujda	83.02	19.04	24.05	83.15	18.04	22.76
		TanTan	50.29	22.57	27.93	61.34	20.72	25.63
		Zagora	88.98	9.37	13.17	91.38	8.25	11.49
		Toutes les villes	82.43	15.08	19.79	81.74	15.27	20.01
Humidité relative	R1	Erfoud	89.26	9.60	13.28	87.90	10.25	13.64
		Missour	86.70	13.51	16.76	80.44	14.79	19.51
		Oujda	90.72	12.77	17.45	90.22	12.73	17.16
		TanTan	64.62	19.36	24.52	67.61	19.44	24.01
		Zagora	87.02	10.20	14.20	89.04	9.24	12.69
		Toutes les villes	85.10	13.60	17.97	84.33	13.98	18.36
	R2	Erfoud	88.94	9.79	13.47	88.43	10.15	13.35
		Missour	86.47	13.74	16.89	80.76	14.61	19.29
		Oujda	90.07	13.25	18.01	89.72	13.08	17.58
		TanTan	59.04	20.76	25.88	65.74	19.75	24.40
		Zagora	86.89	10.21	14.26	88.99	9.32	12.73
		Toutes les villes	84.53	13.94	18.28	83.81	14.28	18.63
	R3	Erfoud	88.68	9.92	13.64	88.20	10.39	13.54
		Missour	84.54	14.63	17.95	79.46	15.52	19.92
		Oujda	87.34	15.78	20.19	86.66	15.67	19.92
		TanTan	51.65	22.51	27.73	61.90	20.53	25.49
		Zagora	86.46	10.42	14.54	88.81	9.60	12.89
		Toutes les villes	83.62	14.31	18.76	82.47	14.76	19.33
Pression	Pr	Erfoud	87.10	10.43	14.47	86.71	10.89	14.21
		Missour	83.82	14.62	18.33	78.95	15.69	20.17
		Oujda	83.14	17.54	23.03	83.47	17.08	21.93
		TanTan	52.36	21.63	26.98	62.97	20.79	25.67
		Zagora	83.91	11.10	15.61	87.29	9.82	13.57
		Toutes les villes	79.39	15.74	20.80	78.34	16.09	21.24

Pour la ville de Tan-Tan, on constate que tous les modèles empiriques testés donnent de faibles performances ($22\% < nRMSE < 27\%$) par rapport aux autres stations. Ceci sera discuté en détails dans la section comparaison et analyse des performances du Chapitre 3 (section 6.3.3).

Tableau 2. 7. Performances des quatre meilleurs modèles obtenus.

Modèle	Stations	Calibration			Validation		
		R (%)	nMAE (%)	nRMSE (%)	R (%)	nMAE (%)	nRMSE (%)
T1 (Hassan et al., 2016) $\frac{H}{H_0} = a + (b + c\Delta T + d(\Delta T)^2 + e(\Delta T)^3)(\Delta T)^f$	Erfoud	93.09	7.92	10.76	89.95	9.11	12.51
	Missour	89.54	11.89	15.05	85.38	13.19	17.48
	Oujda	93.36	10.34	14.89	92.78	10.95	15.02
	TanTan	72.12	17.43	21.98	71.13	18.27	22.81
	Zagora	92.39	7.87	11.01	93.07	7.39	10.18
	Toutes les villes	86.34	12.78	17.26	86.01	12.96	17.43
TG1 (Quej et al., 2016) $\frac{H}{H_0} = a \left(1 + b \cos \left(\frac{cDOY\pi}{180} + d \right) \right) (1 - \exp(e(\Delta T)^f))$	Erfoud	93.71	7.42	10.28	90.69	8.65	12.10
	Missour	90.35	11.08	14.40	87.29	12.22	16.07
	Oujda	93.40	10.37	14.86	92.73	11.03	15.09
	TanTan	72.51	17.27	21.96	72.38	17.94	22.39
	Zagora	92.88	7.47	10.63	93.46	6.96	9.89
	Toutes les villes	86.37	12.65	17.24	86.34	12.83	17.23
T2 (Hassan et al., 2016) $\frac{H}{H_0} = (a + bT + cT^2 + dT^3)(\Delta T)^e$	Erfoud	91.29	9.06	12.13	89.54	9.84	12.69
	Missour	90.15	11.20	14.55	87.53	12.24	15.84
	Oujda	91.97	12.02	16.29	91.07	12.81	16.56
	TanTan	63.52	19.79	24.47	66.54	19.59	24.28
	Zagora	91.95	8.18	11.31	92.73	7.78	10.48
	Toutes les villes	86.60	12.60	17.10	86.45	12.79	17.17
TRPr1 (Chen and Li, 2013) $\frac{H}{H_0} = a + bT_{max} + cT_{min} + dT_{min} * T_{max} + eRH + fP_v$	Erfoud	91.55	9.05	11.89	90.31	9.51	12.24
	Missour	90.44	11.11	14.34	88.02	12.12	15.60
	Oujda	92.05	12.35	16.25	91.65	12.68	16.06
	TanTan	62.09	20.09	24.84	63.19	20.08	25.13
	Zagora	91.85	8.47	11.35	92.96	7.90	10.33
	Toutes les villes	86.45	12.86	17.20	85.82	13.19	17.54

La figure 2.1 montre le graphique en radar des valeurs de nRMSE correspondants à chaque modèle empirique testé pour l'ensemble de données de validation. On remarque que les meilleurs résultats sont obtenus pour les stations de Zagora puis d'Erfoud avec des nRMSEs variant, respectivement, de 9.89% (modèle TG1) à 13.57% (modèle Pr), et de 12.1% (modèle TG1) à 14.6% (modèle TR1). Pour Missour, Oujda et la base de données de « toutes les villes », des performances intermédiaires sont obtenues avec des nRMSEs variant de 15.6% (modèle TRPr1) à 20.33% (modèle TRG3), de 15.02% (modèle T1) à 22.76% (TRG3), et de 17.17% (modèle T2) à 21.24% (modèle Pr) pour les trois cas respectivement.

Les modèles empiriques ayant les meilleures performances (T1, TG1, T2 et TRPr1) ont été appliqué à la base de données de la ville de Fès. Les indicateurs statistiques obtenus par ces quatre modèles sont présentés dans le tableau 2.8. La meilleure performance est obtenue par le modèle TG1 avec des valeurs de nRMSE, nMAE, et R égales à 15.01%, 11.24%, 93.16%, respectivement pour l'ensemble de validation.

Tableau 2. 8. Performances des quatre meilleurs modèles empiriques sur la ville de Fès.

Modèles empiriques	Coefficients de régression						Calibration			Validation		
	a	b	c	d	e	f	R (%)	nMAE (%)	nRMSE (%)	R (%)	nMAE (%)	nRMSE (%)
TG1	0.733	0.032	1.293	4.875	0.095	1.256	91.60	12.03	16.50	93.16	11.24	15.01
T1	-1.258	0.140	0.0078	0.00015	0.234	1.446	91.52	11.95	16.62	93.06	11.26	15.08
T2	0.201	-0.002	0.0001	-2.799	0.501	-	90.98	12.92	17.08	92.24	12.28	15.95
TRPr1	0.022	0.200	-0.027	-	-	-	90.83	13.02	17.22	92.08	12.36	16.08

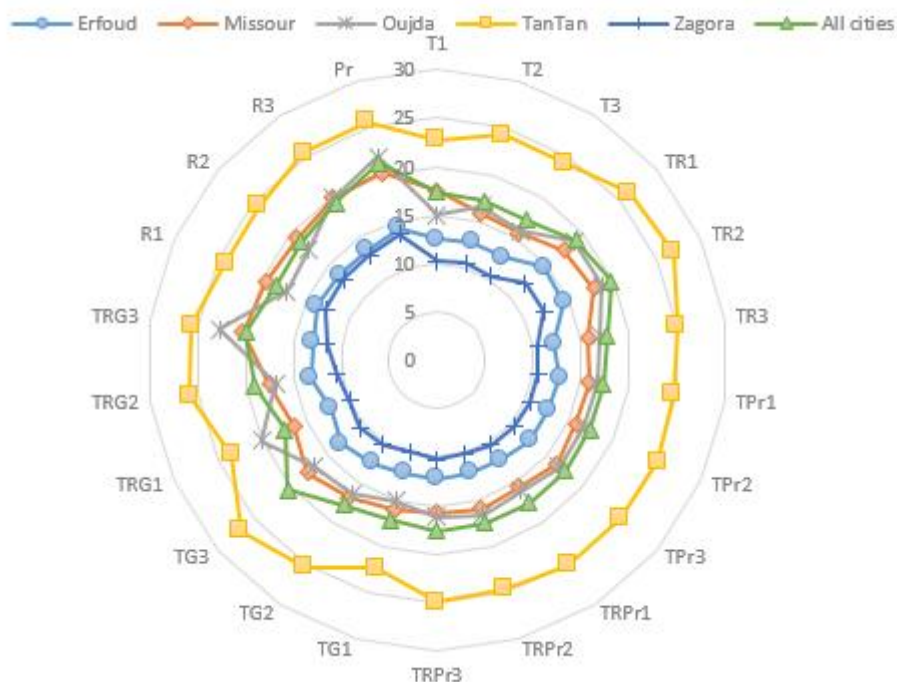


Figure 2. 1. Graphique en Radar des valeurs de nRMSEs de chaque modèle empirique pour l'ensemble de données de validation.

Les figures 2.2 et 2.3 montrent les diagrammes de dispersion des valeurs quotidiennes mesurées et estimées de GHI pour les ensembles de données de validation en utilisant les modèles ayant les meilleures performances. La valeur de R obtenue pour chacun de ces modèles empiriques est présentée dans la légende de chaque graphique.

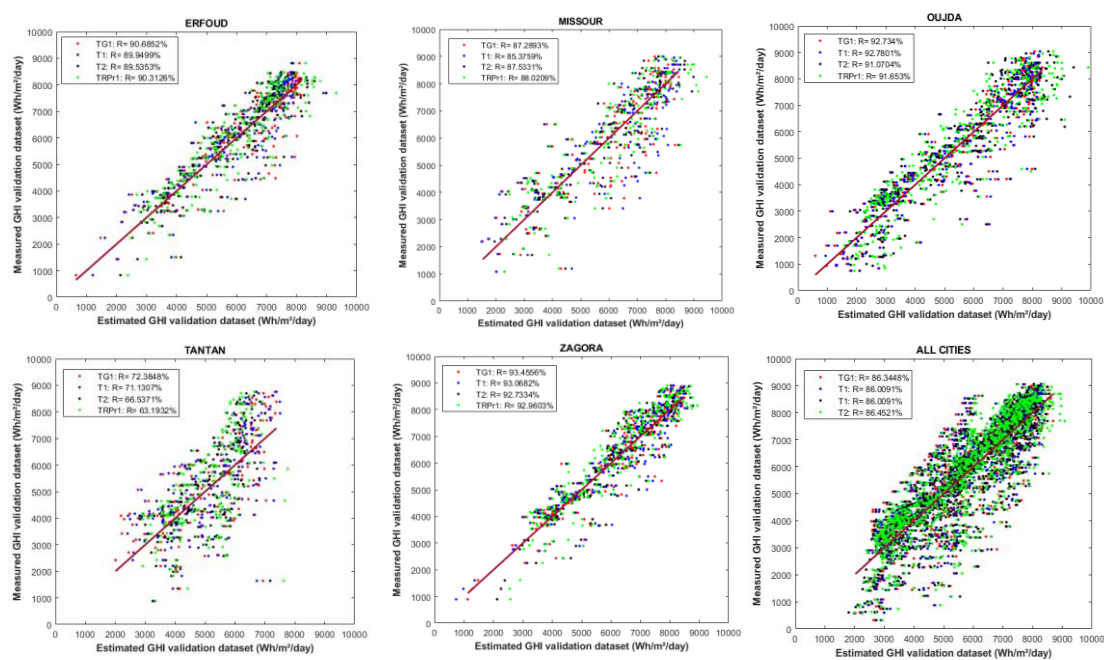


Figure 2. 2. Nuages de points des valeurs quotidiennes mesurées et estimées de GHI en utilisant les modèles les plus performants pour les ensembles de données de validation.

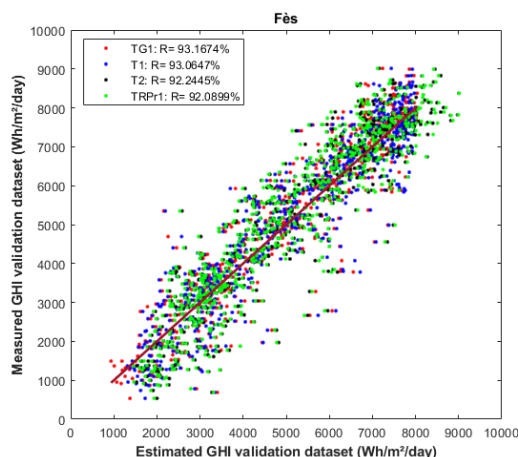


Figure 2. 3. Nuages de points des valeurs quotidiennes mesurées et estimées de GHI en utilisant les modèles les plus performants pour les ensembles de données de validation à la ville de Fès.

Tableau 2. 9. Résultats des modèles empiriques du tableau 2.2 pour les 6 stations marocaines considérées.

Modèle	Stations	Calibration			Validation			Coefficients empiriques				
		R (%)	nMAE (%)	nRMSE (%)	R (%)	nMAE (%)	nRMSE (%)	a	b	c	d	e
Modèle MT1 Hargreaves and Samani (1982) $\frac{H}{H_0} = a\Delta T^{0.5}$	Erfoud	90.64	9.58	12.45	88.82	10.39	13.18	0.1745				
	Missour	88.93	12.39	15.58	85.37	13.76	17.71	0.1737				
	Oujda	92.04	12.01	16.23	90.80	13.04	16.82	0.1884				
	TanTan	61.02	20.41	24.55	58.53	21.55	26.96	0.2605				
	Zagora	91.62	8.59	11.54	92.47	8.02	10.62	0.1824				
	Fès	90.59	13.32	17.61	91.37	12.96	16.91	0.1823				
Modèle MT2 Hargreaves and Samani (1985) $\frac{H}{H_0} = a + b\Delta T^{0.5}$	Erfoud	91.03	9.21	12.24	89.37	9.94	12.79	0.1004	0.1479			
	Missour	88.91	12.37	15.54	85.40	13.15	17.45	0.0245	0.1673			
	Oujda	92.19	11.96	16.21	90.88	13.23	17.07	-0.1055	0.2195			
	TanTan	61.72	20.13	25.00	65.07	19.82	24.63	0.2775	0.1374			
	Zagora	91.60	8.46	11.53	92.52	7.94	10.56	0.0420	0.1714			
	Fès	90.60	13.25	17.50	91.34	12.92	16.86	0.0384	0.1711			
Modèle MT3 (Bristow and Campbell, 1984) $\frac{H}{H_0} = a(1 - \exp(-b\Delta T^c))$	Erfoud	92.95	7.99	10.86	90.16	9.03	12.38	0.6988	0.0149	2.0580		
	Missour	89.54	11.88	15.05	85.40	13.15	17.45	0.7405	0.0506	1.4653		
	Oujda	93.34	10.39	14.92	92.71	11.04	15.10	0.7244	0.0307	1.8239		
	TanTan	71.86	17.49	22.14	71.73	18.04	22.60	0.6542	0.2011	1.7139		
	Zagora	92.33	7.97	11.03	93.05	7.48	10.19	0.7988	0.0788	1.2299		
	Fès	91.56	11.97	16.57	92.72	11.50	15.48	0.7239	0.1011	1.2643		
Modèle MT4 (Fan et al., 2018a) $\frac{H}{H_0} = a + b\Delta T + c\Delta T^{0.25} + d\Delta T^{0.5}$	Erfoud	92.71	8.11	11.04	90.28	8.98	12.28	-0.8819	-0.119	-0.5426	1.1414	
	Missour	88.74	12.44	15.6	85.17	13.67	17.6	0.1478	0.0065	-0.0105	0.1152	
	Oujda	92.80	11.21	15.46	91.75	12.15	15.95	-0.2341	-0.0441	-0.2338	0.5371	
	TanTan	70.54	17.86	22.5	69.35	18.49	23.38	-0.216	-0.1698	-0.5769	1.1391	
	Zagora	91.54	8.36	11.56	92.55	7.88	10.54	0.0759	0.0001	0.013	0.1555	
	Fès	91.50	12.03	16.62	92.62	11.58	15.28	-0.3703	-0.0727	-0.308	0.7125	
Modèle MT5 (Fan et al., 2018a) $H = H_0(a + b\Delta T + c\Delta T^{0.25} + d\Delta T^{0.5}) + eT$	Erfoud	91.02	9.16	12.28	89.47	9.85	12.73	0.0566	0.0076	0.1993	0.0281	-0.1177
	Missour	88.91	12.32	15.47	85.28	13.53	17.51	-0.0312	-0.0005	0.0954	0.1353	-0.0481
	Oujda	93.02	10.90	15.25	92.25	11.64	15.51	-2.2663	-0.0201	2.3302	-0.3331	-2.6906
	TanTan	62.17	20.03	24.89	66.19	19.41	24.29	0.3077	-0.0194	-0.1146	0.2458	-0.3270
	Zagora	91.65	8.30	11.49	92.62	7.82	10.48	-0.0335	-0.0043	0.0738	0.1700	0.0132
	Fès	90.70	13.08	17.34	91.43	12.77	16.73	0.0057	-0.0049	0.0258	0.1840	0.0099

Tableau 2. 9. (Suite) Résultats des modèles empiriques du tableau 2.2 pour les 6 stations marocaines considérées.

Modèle	Stations	Calibration			Validation			Coefficients empiriques							
		R (%)	nMAE (%)	nRMSE (%)	R (%)	nMAE (%)	nRMSE (%)	a	b	c	d	e	f		
Modèle MRH1 (Proposé) $H = H_0 (a + b RH_{max} + c RH_{min})$	Erfoud	90.63	9.70	12.55	87.95	10.40	13.55	0.00070	1.11 e-06	-6.33e-06					
	Missour	89.24	13.32	16.87	87.22	14.75	17.10	0.00069	1.58 e-06	-6.60 e-06					
	Oujda	90.64	14.56	17.12	89.92	15.71	18.27	0.00077	1.49 e-06	-7.34 e-06					
	TanTan	60.19	20.80	23.09	58.04	21.26	24.38	0.00054	4.63e-06	-5.62e-06					
	Zagora	89.12	10.50	13.38	89.90	9.66	12.10	0.0007	1.58e-06	-8.32e-06					
	Fès	90.35	14.82	18.02	89.21	15.23	18.57	0.82783	-0.00552	-0.0159					
Modèle MH1 (Proposé) $\frac{H}{H_0} = a + b\Delta T^{0.5} + c\Delta T + d\Delta H^{0.5} + eT$	Erfoud	93.35	7.84	10.66	90.68	8.7	12.01	-1.2915	1.0100	-0.1196	-0.0143	-0.0026			
	Missour	90.65	10.95	14.2	87.99	11.85	15.55	-0.7523	0.6467	-0.0610	-0.0043	-0.0060			
	Oujda	93.21	10.68	15.03	92.61	11.30	15.11	-0.8929	0.7867	-0.0890	-0.0091	-0.0016			
	TanTan	71.38	17.58	22.18	71.12	18.17	22.82	-0.4712	0.8664	-0.1580	0.0147	-0.0064			
	Zagora	92.87	7.62	10.69	93.47	7.16	9.94	-1.0053	0.8111	-0.0903	-0.0012	-0.0023			
	Fès	91.71	11.96	16.38	92.89	11.43	15.32	-0.5349	0.5429	-0.0585	0.0086	-0.0024			
Modèle MH2 (Proposé) $\frac{H}{H_0} = a + b\Delta T^{0.5} + c\Delta T + d\frac{T_{min}}{T_{max}} + e\frac{RH_{min}}{RH_{max}}$	Erfoud	92.59	7.82	11.46	90.58	8.29	12.29	-0.5227	0.616	-0.0725	-0.0697	-0.1243			
	Missour	90.69	11.07	14.17	87.58	12.18	15.86	-0.389	0.5363	-0.0574	-0.2394	-0.1256			
	Oujda	93.13	10.68	15.11	92.54	11.25	15.17	-0.7539	0.7239	-0.0855	-0.0474	-0.0753			
	TanTan	70.64	17.79	22.42	69.64	18.55	23.28	-0.0131	0.7763	-0.1514	-0.4578	0.0038			
	Zagora	93.05	7.46	10.53	93.37	7.04	9.97	-0.7421	0.7322	-0.0847	-0.1221	-0.1178			
	Fès	91.69	11.91	16.4	92.89	11.35	15.31	-0.2041	0.4997	-0.0608	-0.1454	-0.1629			
Modèle MH3 (Proposé) $H = H_0 (a + bT + cRH + d\Delta T + eWs + fP_s)$	Erfoud	91.24	10.11	12.23	91.51	8.44	10.05	0.00016	-4.13 e-06	-3.00 e-06	1.30 e-05	1.63 e-06	5.37 e-07		
	Missour	89.53	11.66	14.91	89.32	11.42	14.88	-0.0027	-8.71 e-06	-1.73 e-06	2.41 e-05	-4.24e-06	3.64 e-06		
	Oujda	91.42	13.72	16.93	91.85	12.86	15.29	0.00064	-3.80e-06	-2.71e-06	2.40e-05	-5.90e-06	-7.54 e-08		
	TanTan	62.64	22.40	25.31	63.49	20.38	23.61	-0.0103	2.06e-06	7.7 e-08	2.49 e-05	-2.37e-06	1.06 e-05		
	Zagora	91.92	8.02	11.28	92.85	7.85	10.30	0.0029	-4.18e-06	-2.72e-06	1.82e-05	-4.55e-06	-2.55e-06		

Dans le tableau 2.9, nous avons reporté les résultats obtenus en testant les modèles empiriques du tableau 2.2 pour les 6 stations marocaines étudiées dans le présent travail.

Les modèles de Hargreaves et Samani (Hargreaves and Samani, 1982) et Hargreaves et al. (Hargreaves et al., 1985) donnent à peu près les mêmes performances avec une légère amélioration pour le dernier en particulier pour le coefficient de corrélation et le nRMSE. Aussi, les modèles MT3, MT4, MH1 et MH2 donnent des performances voisines. Le modèle MH1 donne les meilleures performances pour les stations d'Erfoud, Missour, Zagora et Fès alors que le modèle MT3 excelle pour les stations d'Oujda et TanTan. Le modèle MT3 (Bristow and Campbell, 1984) donne des performances très proches de celles obtenues par le modèle TG1 de Quej et al., 2016 (tableau 2.1). Ces deux modèles ont la même formulation mathématique sauf au niveau du coefficient multiplicatif qui est constant pour le modèle MT3 alors qu'il est variable en fonction du jour de l'année pour le modèle TG1. Ces deux modèles ne nécessitent que la connaissance des données des températures maximale et minimale journalières. Ces deux paramètres sont mesurés dans toutes les stations météorologiques.

Pour le modèle MRH1 basé sur les valeurs extrêmes de l'humidité relative (proposé), la plus faible valeur de nRMSE est celle trouvée à Zagora (12.10%), alors que, la valeur la plus élevée de nRMSE est celle donnée par TanTan (24.38%).

Dans le tableau 2.10, on présente les résultats des modèles appliqués à la station de Fès uniquement.

Tableau 2. 10. Résultats des modèles empiriques appliqués à la ville de Fès uniquement.

Modèle	Calibration			Validation			Coefficients empiriques						
	R (%)	nMAE (%)	nRMSE (%)	R (%)	nMAE (%)	nRMSE (%)	a	b	c	d	e	f	g
Modèle MH4 (Fan et al., 2018a) $\frac{H}{H_0} = (a + b\Delta T + c\Delta T^{0.25} + d\Delta T^{0.5})[1 + e\ln(1 + P)] + fT$	91.19	12.66	16.89	91.83	12.39	16.35	-0.1131	-0.0085	0.1675	0.156	-0.1459	0.0884	-
Modèle MH5 (Fan et al., 2018a) $\frac{H}{H_0} = (a + b\Delta T + c\Delta T^{0.25} + d\Delta T^{0.5})[1 + e\ln(1 + P)] + fT + gRH$	92.11	11.52	16.06	92.84	11.28	15.36	-0.5565	-0.0565	0.1781	0.451	-0.5337	-0.3462	-0.1129
Modèle MH6 (Proposé) $\frac{H}{H_0} = [a + b\Delta T^{0.25} + c\Delta T^{0.5} + dT] * [1 + e\ln(1 + P)]$	92.24	11.53	15.87	93.20	11.2	15.01	-2.4133	2.9068	-0.6610	-0.0026	-0.3258	-	-

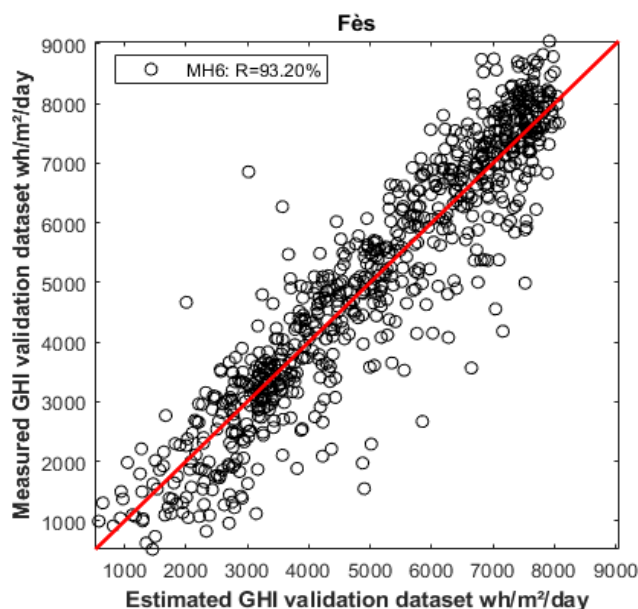


Figure 2. 4. Nuages de points des valeurs quotidiennes mesurées et estimées de GHI en utilisant le modèle proposé MH6 pour les ensembles de données de validation à la ville de Fès.

On remarque qu'en introduisant le terme des précipitations P (modèle MH4), on obtient une légère amélioration des résultats par rapport au modèle MT5. Le modèle proposé (MH6) donne

des performances légèrement meilleures que tous les modèles testés pour la ville de Fès, y compris le modèle TG1. La figure 2.4 montre le nuage de points de comparaison des GH quotidiennes mesurées et estimées en utilisant le MH6.

4.2. Performances des modèles empiriques d'estimation journalière de DHI et DNI

La mesure du rayonnement solaire direct et/ou diffus est plus délicate que la mesure du rayonnement solaire global. En effet, les instruments utilisés sont équipés de systèmes de poursuite du soleil ou de bandes pare-soleil qui nécessitent un entretien et un suivi très minutieux. Sinon, les bases de données seront entachées d'erreurs et donc inutilisables pour la modélisation ou la gestion des centrales solaires. Pour assurer une comparaison non erronée entre les modèles, on a choisi d'utiliser les données de la ville de Zagora pour le choix des meilleurs modèles avant de les tester sur les autres sites.

Le tableau 2.11 montre les coefficients de régression ajustés et les indicateurs de performances statistiques obtenus pour les modèles empiriques d'estimations des irradiations solaires diffuses d'un plan horizontal (Tableau 2.4). On remarque que le modèle DH3, qu'on a proposé, est plus approprié que les autres modèles avec la plus faible valeur du nRMSE évalué à 24.83% et la plus grande valeur de R égale à 89.41. Le modèle DH3 a amélioré les précisions des modèles DH1 et DH2 de 7.29% et 16.88%, respectivement.

Tableau 2. 11. Indicateurs de performances statistiques obtenus pour l'estimation des irradiations solaires diffuses journalières (DHI) à la ville de Zagora.

Modèle	Coefficients de régression						Calibration					Validation				
	a	b	c	d	e	f	R	MAE	RMSE	MBE	nRMSE (%)	R	MAE	RMSE	MBE	nRMSE (%)
DH1	1.6224	-1.7982	-0.0026	-	-	-	0.8825	0.4063	0.5109	-0.011	26.6651	0.8972	0.3759	0.4779	0.0674	26.7917
DH2	1.4617	-1.6658	-0.0001	-	-	-	0.8880	0.4311	0.5485	-0.051	29.3045	0.8795	0.4422	0.5638	-0.048	29.8807
DH3	1.4767	-1.7721	3.12e-06	-0.0001	0.0229	-	0.8913	0.3774	0.4880	0.001	25.5746	0.8941	0.3578	0.4476	0.018	24.8368
DC1	0.0005	-0.0005	-	-	-	-	0.7668	0.5610	0.7116	-0.060	37.7679	0.7736	0.5376	0.6778	-0.029	36.4712
DC2	1.4949	-1.7141	-	-	-	-	0.8841	0.4357	0.5592	-0.053	29.8317	0.8834	0.4135	0.5226	-0.057	27.7802

Pour les modèles basés sur l'indice de clarté (Tableau 2.11), il est clair que le modèle DC2 est plus précis que le modèle DC1. Le modèle DC2 donne des performances acceptables, avec des valeurs de R et nRMSE égales à 88.34% et 27.78%, respectivement, pour l'ensemble de validation ; en réduisant la valeur du nRMSE obtenue par le modèle DC1 de 23.82%.

Le tableau 2.12 montre les performances obtenues par les modèles empiriques d'estimation journalière des irradiations solaires directes d'un plan normal (Tableau 2.4). Le modèle empirique que nous avons proposé pour estimer le DNI quotidien (DNH1) a donné de bonnes

performances avec des valeurs de nRMSE et de R égales à 13.40% et 94.75%, respectivement, pour l'ensemble de validation.

Tableau 2. 12. Indicateurs statistiques obtenus par les modèles empiriques d'estimation journalière des irradiations solaires directes d'un plan normal (DNI) à la ville de Zagora.

Modèle	Coefficients de régression					Calibration				Validation					
	a	b	c	d	e	R	MAE	RMSE	MBE	nRMSE (%)	R	MAE	RMSE	MBE	nRMSE (%)
DNH1	2.4896	-1.4283	0.0002	0.0032	0.2357	0.9475	0.6750	0.8414	0.0058	12.3056	0.9488	0.7021	0.8873	0.0010	13.4011
DNC1	2.6512	-1.6502	0.2694	-	-	0.9307	0.7759	0.9864	0.0414	14.6122	0.9349	0.7774	0.9644	0.0097	14.1341

La figure 2.5 montre les diagrammes de dispersion des valeurs quotidiennes mesurées et estimées de DHI et DNI obtenues en utilisant les modèles les plus performants dans chaque catégorie pour les ensembles de données de validation à Zagora. La valeur de R obtenue pour chacun de ces modèles empiriques est présentée dans la légende de chaque graphique.

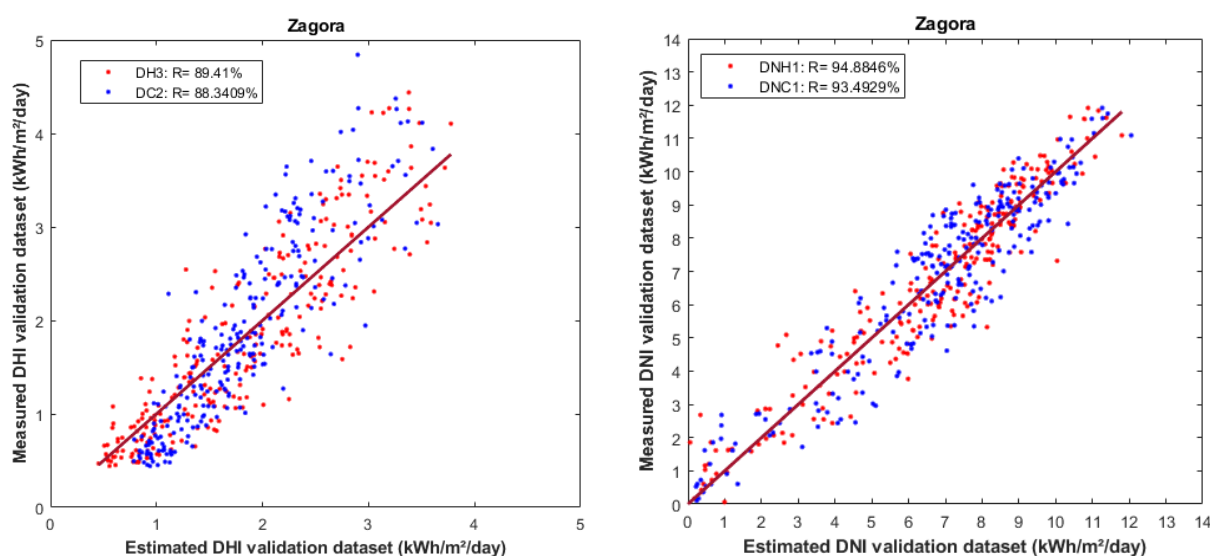


Figure 2. 5. Diagrammes de dispersion des valeurs quotidiennes mesurées et estimées de DHI et DNI en utilisant les modèles les plus performants dans chaque catégorie pour les ensembles de données de validation à Zagora.

Les meilleurs modèles ainsi obtenus, (DH3 et DNH1), ont été testé par les autres sites d'étude. Le tableau 2.13 montre les performances obtenues pour chaque station. Les résultats de ce tableau montrent que les meilleures performances d'estimation journalière du DHI (DH3) et DNI (DNH1), sont obtenues à TanTan et Fès, avec des nRMSEs de validation égales à 18.95% et 11.14%, respectivement.

Tableau 2. 13. Performances statistiques des modèles DH3 et DNHI appliqués aux autres sites.

Site	Modèle	Coefficients de régression						Calibration				Validation						
		a	b	c	d	e	f	R	MAE	RMSE	MBE	nRMSE (%)	R	MAE	RMSE	MBE	nRMSE (%)	
Erfoud	DH3	1.5237	-1.886	3.39 e-06	-	0.00017	0.024	-	0.8688	0.3933	0.5024	0.001	26.5085	0.8772	0.3673	0.4712	0.023	25.1369
	DNHI	2.7206	-1.573	0.00018	-0.003	0.2677	-	0.9449	0.6790	0.8498	0.006	13.5430	0.9364	0.7509	0.9123	-0.008	14.3019	
Missour	DH3	1.3002	-1.548	5.08 e-06	-6.29 e-05	0.0185	-	0.8859	0.3256	0.4286	-0.000	23.6063	0.8719	0.3576	0.4766	-0.042	25.6727	
	DNHI	1.8855	-0.900	0.00010	-0.003	0.1651	-	0.9648	0.5960	0.7778	0.003	12.2596	0.9649	0.6265	0.8094	0.040	13.3232	
Oujda	DH3	1.2121	-1.513	1.13 e-05	-1.73 e-05	0.0141	-	0.8954	0.3144	0.4182	0.003	24.0971	0.9080	0.2851	0.3868	0.002	22.2365	
	DNHI	1.8703	-0.820	7.22 e-05	-0.003	0.1344	-	0.9713	0.5495	0.7029	0.012	12.1983	0.9672	0.5622	0.7630	0.003	13.4629	
TanTan	DH3	1.3796	-1.697	1.54 e-05	-8.29 e-05	0.0164	-	0.9154	0.3598	0.4515	0.000	18.7701	0.9090	0.3623	0.4684	-0.008	18.9550	
	DNHI	2.4325	-1.447	0.0001	0.0048	0.294	-	0.9644	0.5698	0.7372	0.023	18.1948	0.9628	0.5962	0.7575	-0.046	18.4214	
Fès	DH3	1.0511	-	3.51 e-06	0.0120	0.0120	-	0.8552	0.2037	0.2636	-0.005	23.7453	0.8669	0.2050	0.2667	-0.005	22.8053	
	DNHI	0.991	0.015	0.00010	-0.0016	0.0123	-	0.9676	0.5380	0.6881	-0.001	11.3687	0.9665	0.5359	0.6831	0.000	11.1420	

En se basant sur l'indice de clarté (Tableau 2. 14), la précision obtenue par le modèle DC2 pour estimer le DHI quotidien à TanTan est supérieure à celles des autres sites (nRMSE=18.84%). Ainsi, la plus faible valeur de nRMSE obtenue par le modèle DNC1 pour estimer le DNI quotidien est celle de Fès (11.77%).

Tableau 2. 14. Indicateurs statistiques obtenus par les modèles CIBMs pour l'estimation des DHI et DNI aux autres sites.

Site	Modèle	Coefficients de régression				Calibration				Validation				
		a	b	c	R	MAE	RMSE	MBE	nRMSE (%)	R	MAE	RMSE	MBE	nRMSE (%)
Erfoud	DC2	1.5003	-1.7833	-	0.8794	0.4281	0.5449	-0.0512	29.6568	0.8350	0.4520	0.5644	-0.0985	28.0783
	DNC1	2.698	-1.5897	0.24244	0.9229	0.8500	1.0319	0.0536	16.5146	0.9182	0.7749	0.9770	0.0608	15.1688
Missour	DC2	1.3745	-1.5672	-	0.8801	0.3571	0.4642	-0.0335	26.1921	0.8558	0.3557	0.4665	-0.0487	23.8201
	DNC1	2.0774	-1.1456	0.19609	0.9575	0.6652	0.8556	0.0264	13.4290	0.9615	0.6677	0.8524	0.0140	14.1769
Oujda	DC2	1.259	-1.4187	-	0.9131	0.3372	0.4492	-0.0457	25.8856	0.8977	0.3451	0.4523	-0.0558	26.0066
	DNC1	1.8093	-0.8095	0.10654	0.9625	0.6247	0.8144	0.0385	14.2030	0.9653	0.6115	0.7730	0.0698	13.4810
TanTan	DC2	1.536	-1.7503	-	0.9207	0.3767	0.4776	-0.0352	20.1107	0.9149	0.3582	0.4792	-0.0754	18.8434
	DNC1	2.3655	-1.3619	0.19797	0.9616	0.5887	0.7658	0.0302	19.3395	0.9604	0.6097	0.7841	0.0574	18.1198
Fès	DC2	1.0082	-1.2358	-	0.8500	0.2104	0.2747	-0.0030	24.2152	0.8663	0.1985	0.2544	-0.0169	22.8670
	DNC1	1.0854	-0.0870	0.0266	0.9649	0.0006	0.0007	0.0000	11.7143	0.9651	0.0006	0.0007	0.0000	11.7724

5. Conclusion

Ce chapitre a été consacré aux modèles empiriques largement utilisés dans la modélisation des différentes composantes du rayonnement solaire incident au sol. On s'est intéressé, en particulier, aux modèles basés sur des variables météorologiques couramment mesurées et des facteurs géographiques et/ou astronomiques bien connus.

Les modèles d'estimation des irradiations solaires globales sur plan horizontal sont les plus répandus. Nous avons testé des modèles sélectionnés dans la littérature et ceux que nous avons-nous-mêmes proposé. L'application de ces modèles aux 6 sites marocains considérés dans la présente étude, a montré que plusieurs modèles permettent l'estimation des valeurs journalières du GHI avec des précisions acceptables. Toutefois, nous pouvons recommander le modèle TG1 $\left(\frac{H}{H_0} = a \left(1 + b \cos\left(\frac{cDOY\pi}{180} + d\right)\right) \left(1 - \exp(e(\Delta T)^f)\right)\right)$ pour l'ensemble des stations étant donné qu'il a été le meilleur modèle dans plusieurs stations et que ses performances avoisinent celles des modèles donnant les meilleures performances pour les autres villes. Ce modèle ne nécessite que la connaissance du gradient de température journalier $\Delta T = T_{max} - T_{min}$. T_{max} et T_{min} sont mesurées dans toutes les stations météorologiques.

Pour les composantes solaires directe sur plan normal (DNI) et diffuse sur plan horizontal (DHI), nous avons testé des modèles basés sur des paramètres hybrides (HBM) et des modèles basés sur l'indice de clarté (CIBM). Nous avons aussi proposé deux modèles empiriques HBM l'un pour l'estimation du DHI et l'autre pour l'estimation du DNI. La comparaison des performances des différents modèles a été effectuée sur la base de données du site de Zagora. Cette base a été jugée comme étant la moins entachée d'erreurs lors de nos tests de contrôle de qualité des données. La comparaison entre les différents modèles a révélé que les deux modèles que nous avons proposés donnent les meilleures performances.

Dans les chapitres qui suivent, les performances des modèles empiriques étudiés dans le présent chapitre seront comparées à celles de quelques modèles d'apprentissage automatique issus de l'intelligence artificielle.

Chapitre 3

Estimation des irradiances solaires journalières
en utilisant les modèles d'apprentissage
automatique

1. Introduction

Dans le chapitre précédent, l'estimation des différentes composantes solaires journalières en utilisant les modèles empiriques, première famille des estimateurs connue, a été évaluée. Les résultats obtenus ont montré que les performances des modèles empiriques restent limitées étant donné que ces modèles n'arrivent pas à résoudre significativement les problèmes complexes non linéaires. Dans le présent chapitre, on s'intéressera aux techniques du Machine Learning qui font partie du grand domaine de l'intelligence artificielle et qui sont des méthodes d'apprentissage automatique. Ces méthodes permettent aux modèles d'adapter leurs paramètres aux nouvelles situations ou à l'évolution du système sans avoir besoin d'être reprogrammés explicitement. Un modèle d'apprentissage automatique apprend les relations entre les entrées et les sorties à partir des exemples (expériences) et sa performance s'améliore au fur et à mesure que la base de données d'entraînement s'élargit avec le temps.

En premier lieu, une large étude bibliographique sur les différentes techniques d'apprentissage automatique sera présentée en discutant et comparant l'efficacité de chaque modèle d'après les travaux bibliographiques recensés. Ensuite, nous allons décrire la méthodologie suivie, en détaillant chaque modèle d'apprentissage automatique développé durant notre travail. Trois grandes catégories de techniques seront considérées: les Réseaux de Neurones Artificiels (RNAs), les Méthodes d'Ensemble (ME) et la Régression par Machine à Vecteur de Support (SVM). Enfin, les résultats et les performances des différents modèles seront présentés, analysés et discutés.

2. Etat de l'art des modèles d'apprentissage automatique (Machine learning)

Les modèles d'apprentissage automatique font partie des méthodes d'intelligence artificielle, avec une grande capacité à résoudre des problèmes non linéaires complexes en apprenant par l'exemple les relations entre les entrées et sorties d'un système. Dans plusieurs domaines, l'apprentissage automatique est devenu particulièrement important pour résoudre les problèmes de classification, de prédiction, d'optimisation, de reconnaissance des formes, etc. Deux types de techniques d'apprentissage sont utilisés dans les modèles d'apprentissage automatique : l'apprentissage supervisé qui développe un modèle prédictif basé sur des données d'entrée et de sortie connues pour prédire les sorties futures, et l'apprentissage non supervisé, qui regroupe et interprète les données en se basant uniquement sur les données d'entrée. De nos jours, les

modèles d'apprentissage automatique sont de plus en plus utilisés pour prévoir le rayonnement solaire à différents horizons temporels (Fan et al., 2019b). Parmi ces modèles, les Réseaux de Neurones Artificiels (RNAs) ont acquis une grande popularité au cours des trente dernières années. Basés sur la structure et les fonctions du cerveau humain, les RNAs ont été appliqués, à grande échelle, pour résoudre différents problèmes complexes et non linéaires. Les réseaux Perceptron MultiCouches (PMC), également connus sous le nom de réseaux de "rétro-propagation" ou "feedforward", sont les RNAs les plus couramment utilisés pour modéliser et prédire les différentes composantes du rayonnement solaire. Selon Rehman et Mohandes (Rehman and Mohandes, 2008), il a été démontré que les RNAs sont capables d'estimer le rayonnement solaire global en utilisant uniquement les données disponibles de la température et d'humidité relative. Pour plus d'informations, certaines études ont été menées ((Behrang et al., 2010),(Sharifi et al., 2016) , (Feng et al., 2019)) pour estimer le rayonnement solaire global quotidien à l'aide de techniques de réseaux de neurones artificiels (RNAs) basées sur la température, l'humidité relative et les paramètres hybrides.

Si les capacités des RNAs ont été largement explorées, certaines nouvelles méthodes d'apprentissage automatique n'ont pas été, quant à elles, suffisamment utilisées dans la littérature pour estimer ou prédire le rayonnement solaire à l'aide des variables météorologiques couramment mesurées. Il s'agit, notamment, de la régression par machine à vecteur de support (Support Vector Machine Regression (SVMR)) et les méthodes d'ensemble tels que, les forêts aléatoires (Random forests (RF)), le Bagging (Bag) et le Boosting (Boost). Selon l'aperçu de la prévision du rayonnement solaire à l'aide de méthodes d'apprentissage automatique présenté par (Voyant et al., 2017), il a été remarqué que les SVMs et la méthode d'ensemble des forêts aléatoires (RF) donnent des résultats très prometteurs. Néanmoins, leurs performances dépendent de plusieurs facteurs (diversité du jeu de données, pas de temps, ...) et ont globalement une erreur de prédiction équivalente. Certains auteurs ont suggéré d'utiliser des modèles hybrides pour améliorer la précision de la prédiction ((Mohammadi et al., 2015), (Salcedo-Sanz et al., 2018)). De même, une étude bibliographique sur la prévision de l'énergie solaire et éolienne en appliquant les modèles de machine à vecteur de support (SVM) a été présenté dans (Zendehboudi et al., 2018), et il a été conclu que l'approche SVM est rapide, simple à utiliser, fiable et fournit des résultats précis par rapport aux autres approches de modélisation.

L'un des premiers travaux traitant le problème de la prédiction du rayonnement solaire à l'aide des RNAs est celui d'Elizondo et al. (Elizondo et al., 1994). Leurs résultats montrent qu'un Perceptron MultiCouches (PMC) avec une seule couche cachée de 15 neurones peut être utilisé pour estimer l'irradiation solaire globale horizontale (GHI) quotidienne lorsque seules les mesures quotidiennes de la température de l'air et les précipitations maximales et minimales sont disponibles. Benghanem et al. (Benghanem et al., 2009) ont testé six architectures du PMC en utilisant différentes combinaisons de quatre entrées : jour de l'année, durée d'ensoleillement, température, et humidité relative de l'air. Les meilleurs résultats ont été obtenus avec une architecture de 5 neurones cachés et la combinaison des entrées formée par le jour de l'année, la durée d'ensoleillement et la température de l'air. Maamar et al. (Maamar et al., 2014) montrent qu'un PMC avec 6 neurones cachés est bien capable d'estimer le GHI quotidien à partir de la température, de l'humidité relative et de la vitesse du vent. Les résultats des études de Kumar et Kaur (Kumar and Kaur, 2016) ont révélé que l'ensemble de données d'entrée contenant les températures moyennes, minimales et maximales, l'humidité relative, les précipitations, la vitesse du vent et le rayonnement extraterrestre, en utilisant un PMC avec 5 neurones dans la couche cachée, sont les meilleurs en termes de la qualité et l'efficacité de calcul pour estimer le GHI de la ville de Varanasi en Inde. Ihya et al. (Ihya et al., 2015) ont développé et testé des modèles PMCs pour prédire la fraction solaire diffuse horaire et quotidienne à la ville de Fès, au Maroc. Les résultats montrent que l'utilisation de l'indice de clarté quotidien et la durée du jour comme entrées, en employant un PMC avec 10 neurones cachés, sont plus appropriés.

Une considération importante et fondamentale dans la construction des modèles RNAs est le choix des variables d'entrée pertinentes. De nombreux travaux ont accordé une attention particulière à ce problème. A cet égard, Yadav et al. (Yadav et al., 2015) ont comparé la précision de trois modèles développés à l'aide des RNAs (fitting tool (nftool), Radial Basis Function Neural Network (RBFNN) et Generalized Regression Neural Network (GRNN)). Ils ont constaté que la température maximale, la température minimale, l'altitude, l'heure d'ensoleillement sont les paramètres d'entrée les plus pertinents pour la prédiction du rayonnement solaire dans 76 endroits du nord-ouest de l'Inde. Alsina et al. (Alsina et al., 2016) ont utilisé 17 sites pour la phase d'entraînement et 28 sites pour la phase de test pour prédire la moyenne mensuelle quotidienne du rayonnement solaire global en Italie, à l'aide d'un réseau RNA. Le choix de la meilleure combinaison composée de sept paramètres d'entrée (l'irradiation au sommet de l'atmosphère, la durée du jour, le nombre de jours pluvieux, la moyenne des

précipitations, la latitude, et l'altitude) a été effectué à l'aide de la méthode de détermination automatique de la pertinence (ARD). Pour un horizon temporel allant jusqu'à 6 jours, Marquez et Coimbra (Marquez and Coimbra, 2011) ont développé un modèle de réseau de neurones artificiel (RNA) pour prévoir une irradiation solaire à moyen terme en adoptant comme entrées les variables météorologiques prédites à partir de la base de données de prévision du Service météorologique national (US National Weather Service (NWS)). Dans ce travail, les entrées les plus appropriées sont sélectionnées à l'aide d'un ensemble de critères en combinant une version du test Gamma avec un algorithme génétique. Ils ont constaté qu'en utilisant les ensembles d'entrées pertinents identifiés par le test Gamma, les modèles de prévision développés améliorent les RMSEs de GHI de 10 à 15 % par rapport au modèle de référence. En outre, la précision du modèle proposé diminue avec l'augmentation de l'horizon de prévision, mais moins fortement pendant l'été que pendant les mois d'hiver. Mellit et Pavan, (Mellit and Pavan, 2010), ont proposé une prévision de 24 heures de l'irradiation solaire en utilisant un modèle de réseau de neurones artificiel (RNA) de type Perceptron MultiCouche (PMC), en utilisant comme entrées les valeurs actuelles de la moyenne quotidienne d'irradiation solaire et de la température de l'air. Le modèle proposé montre de bonnes performances avec un coefficient de corrélation qui se situe entre 98 et 99% pour les jours ensoleillés et entre 94 et 96% pour les jours nuageux.

Un autre type de modèles parmi la famille des réseaux de neurones artificielles (RNAs) est celui des réseaux de neurones non linéaires autorégressifs avec des entrées exogènes (NARX) qui ont été souvent utilisés pour prédire l'irradiation solaire à différents horizons temporels. Al-Sbou et Alawasa, (Al-Sbou and Alawasa, 2017), ont utilisé les modèles NARX pour prédire le rayonnement solaire à la ville de Mutah en Jordanie. Ils ont développé sept modèles basés sur sept combinaisons de trois entrées météorologiques qui sont la température, l'humidité et la vitesse du vent. Leurs résultats ont montré que le modèle NARX utilisant les trois variables comme entrées était le modèle le plus approprié pour les prédictions et les investigations. Sansa et Bellaaj (Sansa and Bellaaj, 2018) ont prouvé l'efficacité du modèle NARX pour prédire le rayonnement solaire, en particulier pour les faibles fluctuations du rayonnement solaire. Boussaada et al. (Boussaada et al., 2018) ont proposé un modèle de réseau de neurones de type NARX pour prédire l'irradiation solaire directe journalière. La phase d'apprentissage du réseau de neurones est effectuée périodiquement et les meilleurs résultats ont été obtenus pour un ensemble de données de 10 jours avec une moyenne mobile de 30 min sur des intervalles d'une heure.

Une étude bibliographique présentant des travaux de recherche sur la prévision de la vitesse et de la puissance du vent et sur la prévision de l'irradiation solaire en utilisant des méthodes d'ensemble a été effectuée par Ren et al. (Ren et al., 2015). Ces méthodes d'ensemble ont été classées en deux principales catégories, la prévision d'ensemble compétitive qui a été classée en fonction de la diversité des données et des paramètres et la prévision d'ensemble coopérative qui a été divisée en fonction du prétraitement et du post-traitement. Ainsi, le potentiel des trois méthodes de prévision d'ensemble a été évalué par les auteurs pour prédire l'irradiation solaire et la vitesse du vent. Pour un horizon de prévision plus long, les meilleures performances ont été obtenues par la méthode de prévision d'ensemble compétitive (Bagging-BackPropagation). Hassan et al. (Muhammed A Hassan et al., 2017) ont proposé une étude pour explorer le potentiel des méthodes d'ensemble basées sur les arbres de décision dans la modélisation du rayonnement solaire. Pour cela, trois modèles (Gradient Boosting, Bagging et Random Forests (RF)) ont été développés pour estimer les valeurs journalières et horaires des trois composantes solaires (GHI, DHI et DNI). En général, les résultats ont montré que les techniques proposées sont très fiables et précises, malgré leur relative simplicité. Cependant, les deux algorithmes Bagging et RF ont donné les meilleures estimations par rapport au modèle du Gradient Boosting. En considérant la précision de prédiction, la stabilité du modèle et l'efficacité de calcul, Fan et al. (Fan et al., 2018b) ont confirmé que la nouvelle méthode d'ensemble appelée, Boosting à gradient extrême (XGBoost) est fortement recommandée pour estimer le GHI quotidien en utilisant les données de la température et de la précipitation communément disponibles avec une excellente performance obtenue pour les climats subtropicaux humides.

Dans le travail de Torres-Barrán et al. (Torres-Barrán et al., 2019), les méthodes d'ensemble telles que la régression par forêts aléatoires (RFR), la régression par boosting à gradient (GBR) et le boosting à gradient extrême (XGB) ont été appliquées aux deux problèmes de prédiction de l'énergie éolienne et du rayonnement solaire. La principale conclusion tirée de cette étude est que ces méthodes d'ensemble sont considérées parmi les candidats compétitifs des modèles d'apprentissage automatique (Machine-learning) pour les prévisions opérationnelles du vent ou du soleil. Karasu et Altan (Karasu and Altan, 2019) ont proposé un modèle de reconnaissance de haute performance pour les séries temporelles du rayonnement solaire basé sur la méthode d'ensemble des forêts aléatoires (RF) avec une approche de sélection des caractéristiques. Finalement, les résultats du modèle proposé pour estimer le rayonnement solaire en sélectionnant les caractéristiques pertinentes ont montré de bonnes performances avec des

valeurs de MSE et R^2 égales à 735.04 et 0.9963, respectivement. Afin de fournir une prédiction fiable à court terme de l'irradiation solaire basée sur divers facteurs météorologiques, Lee et al. (Lee et al., 2020) ont évalué la performance de prédiction de quatre méthodes d'ensemble: Boosted Trees, Bagged Trees, Random Forests et Generalized Random Forests. Les résultats trouvés montrent que les modèles d'ensemble ont une prédiction cohérente et fiable lorsqu'ils sont appliqués à des données provenant de différents endroits.

Certains auteurs ont dédié une grande partie de leurs travaux pour estimer le rayonnement solaire en utilisant la régression par machine à vecteur de support. C'est le cas de Ji-Long Chen qui est l'auteur de référence de ce type de modélisation. L'objectif principal de Chen et al. (Chen et al., 2011) a été d'examiner la faisabilité des SVMs dans l'estimation du rayonnement solaire mensuel, en utilisant 7 différentes combinaisons d'entrées basées seulement sur les données mesurées de la température de l'air (les températures de l'air mensuelles à long terme, les températures maximales et minimales). Ainsi, une analyse comparative des modèles SVMs avec plusieurs modèles empiriques basés sur la température a également été effectuée et il a été montré que la méthodologie SVM peut être une alternative prometteuse par rapport aux approches traditionnelles pour prédire le rayonnement solaire lorsque les enregistrements de la température de l'air sont disponibles. Selon Chen et al. (Chen et al., 2013), l'utilisation des données de la durée d'ensoleillement offre une alternative importante pour l'estimation du rayonnement solaire en l'absence de ses mesures. Pour cela, sept modèles SVM utilisant différents variables d'entrée et cinq modèles empiriques basés sur l'ensoleillement sont évalués pour estimer le rayonnement solaire quotidien à l'aide des données météorologiques de trois stations de la province de Liaoning en Chine. Les résultats trouvés montrent qu'il est préférable d'utiliser le modèle nouvellement développé, SVM1, en raison de sa plus grande précision et de la simplicité de sa variable d'entrée (fraction d'ensoleillement) aussi sa performance a été meilleure en hiver qu'en été. De plus, SVM1 est capable de bien estimer le rayonnement solaire quotidien en utilisant les données des stations voisines. Dans le même sens, un total de 20 modèles SVMs utilisant différentes combinaisons de la fraction d'ensoleillement, des températures maximale et minimale de l'air, d'humidité relative et de la pression de vapeur d'eau atmosphérique comme variables d'entrée ont été explorés par (Chen and Li, 2014) à l'aide des données météorologiques mesurées en 15 stations en Chine. Les résultats indiquent que le modèle SVM2, utilisant la fraction d'ensoleillement et le gradient journalier de la température de l'air comme variables d'entrées, est nettement plus performant que les modèles empiriques,

avec une précision moyenne supérieure de 26%. Aussi, le modèle SVM19 utilisant la température maximale, la température minimale et la pression de vapeur d'eau atmosphérique surpasse de manière significative les modèles empiriques basés sur la température avec une précision moyenne supérieure de 18%. L'étude effectuée dans ((Chen et al., 2015)), a montré que le rayonnement solaire sur un site (site d'estimation) peut être bien estimé par le modèle SVM développé sur un autre site (site source) et sa précision dépend de la distance et de la différence de température entre les deux sites, ainsi que de l'altitude du site source.

Belaid et Mellit (Belaid and Mellit, 2016) ont appliqué la machine à vecteur de support (SVM) pour une prédiction à un pas en avant du rayonnement solaire global quotidien et mensuel sur une surface horizontale à Ghardaïa (Algérie), en utilisant différentes combinaisons des données de la température ambiante mesurée, de la durée maximale d'ensoleillement calculée et du rayonnement solaire extraterrestre calculé. Le principal avantage trouvé est que les modèles SVM proposés ne nécessitent que quelques paramètres simples pour obtenir une bonne précision. Dans ((Bakhashwain, 2016)), le modèle SVM a été entraîné en utilisant la température et l'humidité relative de l'air mesurées pour prédire le rayonnement solaire global (GSR) à Sharurha, une ville située dans le sud-ouest de l'Arabie Saoudite. Les résultats obtenus montrent que la méthode SVM est capable de prédire précisément le GSR à partir des valeurs mesurées de ces deux paramètres météorologiques.

3. Comparaison bibliographique des performances prédictives des modèles d'apprentissage automatique

Fan et al. (Fan et al., 2019b) ont examiné et évalué les performances de 12 modèles empiriques et de 12 algorithmes d'apprentissage automatique pour estimer le GHI quotidien à partir de la durée d'ensoleillement dans les différentes zones climatiques de la Chine. Leurs résultats ont révélé que les modèles d'apprentissage automatique ont généralement surpassé les modèles empiriques en termes de précision des prédictions, mais aucun modèle d'apprentissage automatique n'a été dominant dans toutes les stations des différentes zones climatiques étudiées. En plus dans (Fan et al., 2018b),. Il a été constaté que les modèles SVM et XGBoost (Boosting de Gradient Extrême) étaient plus performants que les modèles empiriques étudiés, et ont recommandé le modèle XGBoost comme un modèle d'apprentissage automatique prometteur pour l'estimation du GHI. Torres-Barrán et al. (Torres-Barrán et al., 2019) ont également montré que les deux méthodes d'ensemble, la régression boostée par gradient (GBR : Gradient

Boosted Regression) et la méthode d'ensemble XGBoost peuvent améliorer les performances obtenues par le SVR dans la modélisation du rayonnement solaire.

Dans le travail d'Ahmad et al. (Ahmad et al., 2015), un réseau de neurones récurrent autorégressif non linéaire avec des entrées exogènes (NARX) a été comparé avec un réseau de neurones artificiel (RNA) basé le Perceptron MultiCouche (PMC), une approche statistique utilisant une moyenne mobile autorégressive (ARMA) et une approche de référence de persistance, pour fournir une prévision du rayonnement solaire global horaire à un jour d'avance à la Nouvelle-Zélande. Leurs résultats ont démontré que l'approche NARX a une grande capacité de généralisation pour produire des prévisions précises de l'irradiation global à un certain nombre de lieux différents. Voyant et al. (Voyant et al., 2017) stipulent qu'il existe plusieurs méthodes pour estimer le rayonnement solaire, celles qui sont souvent utilisées (RNA, ARIMA (AutoRegressive Integrated Moving Average)), méthodes naïves), d'autres qui commencent à être utilisées plus couramment (SVM, SVR, k-mean), et d'autres qui sont rarement utilisées (Boosting, arbre de régression, forêts aléatoires, etc.). De plus, il a été mentionné que RNA et ARIMA sont équivalents en termes de précision de prédiction, mais la simplicité du RNA en tant qu'approximation non linéaire généralisée le rend plus désirable que l'ARIMA traditionnel. Aussi, la précision de ces méthodes est généralement liée à la qualité des données d'entraînement et il est nécessaire que les SVMs, les arbres de régression et les forêts aléatoires soient utilisés dans les années à venir, étant donné leurs résultats positifs qui sont très prometteurs.

Ayodele et al. (Ayodele et al., 2019) ont développé un algorithme de k-moyenne (k-means) hybridé avec la régression par vecteur de support (SVR) pour prédire le rayonnement solaire global quotidien. L'objectif de l'utilisation des k-moyennes était de regrouper les données saisonnières en groupes distincts indépendants des saisons. Leurs résultats montrent que les performances de l'algorithme proposé étaient excellentes par rapport aux modèles RNA, ARMA et Angstrom-Prescott. (Olatomiwa et al., 2015), (Olatomiwa et al., 2015), proposent une nouvelle méthode hybride nommée SVM-FFA pour prédire le rayonnement solaire global en couplant la machine à vecteur de support (SVM) avec l'algorithme Firefly (FFA). Les résultats obtenus montrent que le modèle SVM-FFA est plus efficace que les modèles RNAs (Artificial Neural Networks) et GP (Genetic Programming). Hassan et al. (Muhammed A Hassan et al., 2017) ont exploré le potentiel des méthodes d'ensemble basées sur des arbres de

régression (Gradient Boosting, Bagging et Random Forests (RF)) pour estimer les composantes globales, diffuses et directes normales du rayonnement solaire à l'échelle quotidienne et horaire. Ces modèles ont été testés sur cinq sites de la région MENA et comparés à un Perceptron MultiCouche (PMC), à la régression par vecteur de support (SVR) et aux arbres de décision (DT). Dans leur procédure de calcul, ils ont mentionné que les observations quotidiennes du ciel couvert sont éliminées avant d'appliquer les méthodes d'apprentissage automatique, ce qui explique les bonnes performances obtenues dans leur étude. Les résultats obtenus ont montré que les modèles d'ensemble sont plus stables que les arbres de décision et les PMCs et ont des performances proches des modèles SVRs optimisés, avec des coûts de calcul nettement inférieurs. Les auteurs affirment que les modèles suggérés pourraient être utilisés pour générer les données synthétiques du rayonnement à utiliser dans la simulation, la conception ou l'évaluation des performances des différents systèmes d'énergie solaire. Benali et al. (Benali et al., 2019) ont également comparé trois méthodes de prévision, la Persistance intelligente (Smart Persistence (SP)), le réseau de neurones artificiel (Artificial Neural Network (ANN)) et les forêts aléatoires (Random Forests (RF)) pour prévoir les trois composantes de l'irradiation solaire (GHI, DHI, et DNI) mesurées à Odeillo, en France, pour des horizons temporels allant de h+1 à h+6. Leurs résultats montrent que la méthode RF est la plus efficace et prévoit les trois composantes avec les valeurs de nRMSE les plus faibles comparées à ANN et SP. Aussi, la précision de la méthode RF augmente avec l'horizon de prévision. D'ailleurs, l'étude effectuée par Liu et Sun (Liu and Sun, 2019) a confirmé que la meilleure performance de la prévision de la production d'énergie photovoltaïque (PV) est obtenue en utilisant l'analyse en composantes principales et l'algorithme de regroupement K-moyennes combinés à l'algorithme des forêts aléatoires (RF) optimisé par l'évolution différentielle Grey Wolf Optimizer.

Onze outils statistiques et d'apprentissage automatique sont analysés et appliqués par Fouilloy et al. (Fouilloy et al., 2018) pour prévoir l'irradiation solaire horaire pour un horizon temporel de 1 à 6 h. Les données solaires collectées sur trois sites présentant une variabilité météorologique faible, moyenne et élevée sont considérées. Il a été constaté que pour une faible variabilité, la moyenne mobile autorégressive (ARMA) et le perceptron multicouche (PMC) sont les plus efficaces. Pour une variabilité moyenne, l'ARMA et l'arbre de régression Bagging (Bag) sont les meilleurs prédicteurs, tandis que pour une variabilité élevée, seules des méthodes plus complexes telles que l'arbre de régression Bagging et l'approche des forêts aléatoires peuvent être utilisées pour garantir des résultats satisfaisants. Le tableau 3.1 présente quelques

travaux réalisés par certains chercheurs pour prédire le rayonnement solaire global journalier et horaire dans différentes localisations géographiques en précisant la performance de prédiction obtenue par chaque modèle d'apprentissage automatique (Machine-Learning) utilisé. Plus de détails sur la prédiction du rayonnement solaire global à l'aide de modèles d'apprentissage automatique ont été présentés dans un article qui a été publié récemment par Zhou et al. (Zhou et al., 2021). Cette étude fournit un examen complet et systématique sur tous les aspects importants entourant les modèles d'apprentissage automatique, y compris les paramètres d'entrée, la sélection des caractéristiques et le développement du modèle, en se basant sur 232 articles concernant les modèles d'apprentissage automatique pour la prédiction du rayonnement solaire global. Cette comparaison bibliographique nous a permis de conclure que la modélisation ou la prédiction des séries temporelles du rayonnement solaire est une tâche difficile et complexe. Malgré le grand nombre de modèles établis, il est généralement impossible de trouver un modèle universel pouvant être appliqué à tous les problèmes d'estimation de l'irradiation solaire, en raison de la diversité des paramètres météorologiques et des emplacements géographiques. En outre, les performances des modèles de prévision sont toujours controversées dans diverses études et il y a un manque de comparaison systématique entre eux.

Sur la base de l'étude bibliographique ci-dessus, les méthodes d'ensemble semblent sous-exploitées dans la modélisation des différentes composantes solaires (GHI, DHI et DNI) malgré leur très bon potentiel. En effet, la plupart des chercheurs se sont principalement concentrés sur l'application des réseaux neuronaux artificiels (RNAs) et des modèles empiriques pour estimer ou prédire le rayonnement solaire. Par conséquent, la proposition de cette thèse est d'explorer le potentiel des différentes techniques, mais en mettant en exergue les nouvelles méthodes émergentes et, en particulier, les méthodes d'ensemble basées sur les arbres de régression, à savoir, le Boosting, le Bagging et les forêts aléatoires (Random Forests), ainsi que la régression par machine à vecteur de support (SVMR). Nous procéderons à l'évaluation de ces modèles et à la comparaison de leurs performances avec celles des réseaux de neurones artificiels (RNAs) et les modèles empiriques présentés dans le chapitre précédent.

Chapitre 3 : Estimation des irradiations solaires journalières en utilisant les modèles d'apprentissage automatique.

Tableau 3. 1. Performances des modèles d'apprentissage automatique pour la prédiction de l'irradiation solaire globale journalière et horaire dans la littérature.

Réfs	Composante solaire	Localisation	Modèle	Performances
(Sözen and Arcaklioglu, 2005)	Irradiation solaire globale journalière	12 villes en Turquie	RNA (PMC)	MAPE<=3.832% R ² > 0.998 RMSE <=1.89 % MPE=2.2077% RMSE=0.049515 MBE=1.2514 R=94.65%
(Benghanem et al., 2009)	Irradiation solaire globale journalière	Al-Madinah (Arabie Saoudite)	RNA (PMC)	MAPE=5.21% R ² =99.57% R=0.97 R ² = 0.89
(Behrang et al., 2010)	Irradiation solaire globale journalière	Dezful (Iran)	RNA (PMC)	RMSE=2.53 MJ m ⁻² d-1 RE=13.83% MAE=1.00 MJ/m ² day MAPE=9.23% R=0.9751 R ² =0.9508
(Rahimkhoob, 2010)	Irradiation solaire globale journalière	Ahwaz (Iran)	RNA (PMC)	MBE=-0.49 MJ m ⁻² RMSE= 14.20% R= 0.93
(Bilgili and Ozgoren, 2011)	Irradiation solaire globale journalière	Adana (Turquie)	RNA	MBE=-0.317 MJ/m ² RMSE=0.998 MJ/m ² R ² =0.977
(Linares-Rodríguez et al., 2011)	Irradiation solaire globale journalière	Espagne	RNA (PMC)	MBE=-0.459 MJ/m ² RMSE=1.056 MJ/m ² R ² =0.975
		Le Caire (Égypte)	Boosting	MBE=-0.274 MJ/m ² RMSE=1.040 MJ/m ² R ² =0.975
			Bagging	MBE=-0.294 MJ/m ² RMSE=1.093 MJ/m ² R ² =0.973
			Random Forests	MBE=-0.272 MJ/m ² RMSE=1.214 MJ/m ² R ² =0.966
			RNA (PMC)	MBE=0.000 MJ/m ² RMSE=0.896 MJ/m ² R ² =0.983 MJ/m ²
			SVR	MBE=-0.235 MJ/m ² RMSE=0.809 MJ/m ² R ² =0.986
		Ma'an (Jordanie)	Boosting	MBE=-0.074 MJ/m ² RMSE=1.018 MJ/m ² R ² =0.978
			Bagging	MBE=-0.039 MJ/m ² RMSE=0.947 MJ/m ² R ² =0.981
			Random Forests	MBE=0.006 MJ/m ² RMSE=0.978 MJ/m ² R ² =0.980
			RNA (PMC)	MBE=0.045 MJ/m ² RMSE=1.518 MJ/m ² R ² =0.960
			SVR	MBE=0.055 MJ/m ² RMSE=1.483 MJ/m ² R ² =0.962
(Muhammed A Hassan et al., 2017)	Irradiation solaire globale journalière	Tataouine (Tunisie)	Boosting	MBE=0.159 MJ/m ² RMSE=1.878 MJ/m ² R ² =0.939
			Bagging	MBE=0.261 MJ/m ² RMSE=1.779 MJ/m ² R ² =0.945
			Random Forests	MBE=0.311 MJ/m ² RMSE=1.838 MJ/m ² R ² =0.942
			RNA (PMC)	MBE=0.288 MJ/m ² RMSE=0.996 MJ/m ² R ² =0.973
			SVR	MBE=0.078 MJ/m ² RMSE=0.882 MJ/m ² R ² =0.979
		Ghardaia (Algérie)	Boosting	MBE=0.304 MJ/m ² RMSE=1.220 MJ/m ² R ² =0.959
			Bagging	MBE=0.258 MJ/m ² RMSE=0.986 MJ/m ² R ² =0.973
			Random Forests	MBE=0.286 MJ/m ² RMSE=1.100 MJ/m ² R ² =0.967
			RNA (PMC)	MBE=-0.493 MJ/m ² RMSE=1.403 MJ/m ² R ² =0.895
			SVR	MBE=-0.489 MJ/m ² RMSE=1.209 MJ/m ² R ² =0.922
		Tan-Tan (Maroc)	Boosting	MBE=-0.261 MJ/m ² RMSE=1.337 MJ/m ² R ² =0.905
			Bagging	MBE=-0.387 MJ/m ² RMSE=1.348 MJ/m ² R ² =0.903
			Random Forests	MBE=-0.247 MJ/m ² RMSE=1.569 MJ/m ² R ² =0.869

Chapitre 3 : Estimation des irradiations solaires journalières en utilisant les modèles d'apprentissage automatique.

Tableau 3. 1. (Suite) Performances des modèles d'apprentissage automatique pour la prédiction de l'irradiation solaire globale journalière et horaire dans la littérature.

Réfs	Composante solaire	Localisation	Modèle	Performances	
(Fan et al., 2019b)	Irradiation solaire globale journalière	Chine	PMC	-0.284 < MBE < 0.612 MJ m ⁻² d ⁻¹ 0.129 < nRMSE < 0.200 MJ m ⁻² d ⁻¹ 0.880 < R ² < 0.934	
			SVM	-0.437 < MBE < 0.637 MJ m ⁻² d ⁻¹ 0.131 < nRMSE < 0.197 MJ m ⁻² d ⁻¹ 0.879 < R ² < 0.933	
			RF	-0.364 < MBE < 0.744 MJ m ⁻² d ⁻¹ 0.144 < nRMSE < 0.213 MJ m ⁻² d ⁻¹ 0.839 < R ² < 0.919	
			Boosting	-0.371 < MBE < 0.733 MJ m ⁻² d ⁻¹ 0.132 < nRMSE < 0.197 MJ m ⁻² d ⁻¹ 0.879 < R ² < 0.932	
(Mellit and Pavan, 2010)	Irradiation solaire globale horaire	Italie	RNA (PMC)	R=94.00% RMSE=67.00% MBE=32.00% R ² =0.9242	
(Shaddel et al., 2016)	Irradiation solaire globale horaire avec un plan incliné de 45°	Iran	RNA	RMSE=0.055 MAE=0.0284 MBE=-3.666	
			RNA (PMC)	RMSE=126.605 R ² =0.797 MBE=5.249	
			SVR	RMSE=88.752 R ² =0.900	
			Le Caire (Égypte)	Boosting	MBE=-10.197 RMSE=97.510 R ² =0.879
				Bagging	MBE=-0.527 RMSE=92.156 R ² =0.892
			Random Forests	MBE=-3.675 RMSE=90.435 R ² =0.896	
			RNA (PMC)	MBE=-19158 RMSE=105.037 R ² =0.870	
			SVR	MBE=6.611 RMSE=95.576 R ² =0.892	
		Ma'an (Jordanie)	Boosting	MBE=-1.925 RMSE=98.395 R ² =0.8886	
			Bagging	MBE=2.230 RMSE=102.422 R ² =0.876	
			Random Forests	MBE=-1.045 RMSE=100.272 R ² =0.881	
			RNA (PMC)	MBE=-4.702 RMSE=133.144 R ² =0.790	
			SVR	MBE=-24.320 RMSE=121.422 R ² =0.825	
			Tataouine (Tunisie)	Boosting	MBE=-0.496 RMSE=123.120 R ² =0.821
				Bagging	M MBE=14.041 RMSE=130.116 R ² =0.800
				Random Forests	MBE=-8.431 RMSE=127.840 R ² =0.807
RNA (PMC)	MBE=-3.666 RMSE=126.605 R ² =0.797				
SVR	MBE=-21.114 RMSE=102.447 R ² =0.866				
Ghardaia (Algérie)	Boosting	MBE=-0.597 RMSE=103.083 R ² =0.864			
	Bagging	MBE=7.506 RMSE=112.860 R ² =0.837			
	Random Forests	MBE=-0.945 RMSE=109.588 R ² =0.847			
	RNA (PMC)	MBE=-23.541 RMSE=189.998 R ² =0.623			
	SVR	MBE=-2.499 RMSE=181.161 R ² =0.657			
	Tan-Tan (Maroc)	Boosting	MBE=-25.652 RMSE=173.167 R ² =0.687		
		Bagging	MBE=-6.863 RMSE=179.768 R ² =0.662		
		Random Forests	MBE=-16.635 RMSE=179.024 R ² =0.665		
RNA		R = 0.9887			
(Khosravi et al., 2018)	Irradiation solaire globale horaire	Iran	RNA	RMSE = 41.08 Wh/m ² MSE = 1688.188Wh/m ²	

4. Réseaux de Neurones Artificiels (RNAs)

Les réseaux de neurones artificiels (RNAs) sont des modèles inspirés des réseaux de neurones biologiques. En 1943, les premiers travaux sur les réseaux de neurones ont été effectués par deux chercheurs américains, Warren McCulloch et Walter Pitts. Ces travaux ont donné naissance au premier modèle mathématique simplifié du neurone biologique communément appelé : neurone formel. Similaire aux neurones biologiques, les neurones artificiels visent à reproduire les mêmes raisonnements intelligents du cerveau mais d'une manière artificielle. Ainsi, en se basant sur le fonctionnement d'un cerveau humain qui est la machine de traitement d'information la plus sophistiquée, les réseaux de neurones artificiels s'adaptent de la même façon qu'un cerveau humain aux nouvelles situations, selon un processus d'adaptation rapide, automatique et selon l'apprentissage reçu (Boussaada, 2018).

Les RNAs constituent une partie importante de l'intelligence artificielle et ont une grande capacité à résoudre des problèmes non linéaires complexes en apprenant (à partir de quelques exemples) la relation entre les entrées et les sorties d'un processus ((Al-Sbou and Alawasa, 2017), (Boussaada et al., 2018)). Les modèles RNAs sont utilisés dans plusieurs applications comme la classification, la prédiction, l'optimisation, la reconnaissance de formes, etc. Deux catégories de RNAs se distinguent dans le domaine de la prédiction : les réseaux bouclés et les réseaux non bouclés. Les réseaux de neurones bouclés se caractérisent par des connexions récurrentes qui permettent d'exploiter les sorties précédentes comme entrées pour la prédiction de la sortie présente. Dans ce type de réseaux, la notion du temps est prise en considération (Dreyfus et al., 2002). Plus généralement, ces réseaux sont utilisés pour effectuer des tâches de modélisation des systèmes dynamiques, de commande de processus ou de filtrage. Les RNAs non bouclés sont constitués, quant à eux, d'un ensemble de neurones connectés entre eux et où l'information circule, d'une manière directe (feedforward), depuis les entrées vers les sorties sans retour en arrière. Ce type de réseaux est souvent utilisé dans les problèmes de classification ou d'estimation non linéaires.

Dans un RNA, l'apprentissage est une phase très importante permettant d'ajuster itérativement les poids du réseau pour obtenir un compromis satisfaisant entre la précision d'approximation sur l'ensemble d'apprentissage et celle sur l'ensemble de validation. On peut distinguer trois types d'apprentissage :

- Apprentissage supervisé : les entrées et les sorties sont connues et présentées au réseau pour ajuster les poids des différentes connexions dans le sens de minimiser l'erreur ou l'écart entre la sortie calculée et la valeur cible (mesurée) correspondante.
- Apprentissage non supervisé : les sorties désirées ne sont pas connues et les poids du réseau sont ajustés de façon à réaliser une catégorisation (clustering) des entrées.
- Apprentissage par renforcement : les sorties désirées ne sont pas fournies au réseau, mais une information sur la qualité de la performance est fournie pour trouver la meilleure solution qui s'approche du but visé.

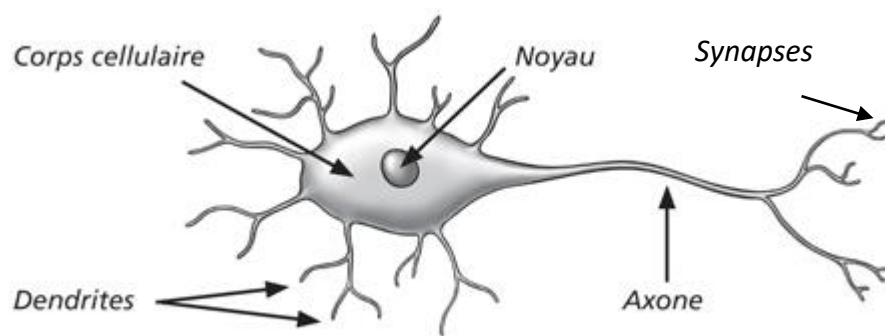


Figure 3. 1. Représentation schématique d'un neurone biologique.

Pour bien comprendre le principe des RNAs, il est important de définir leur élément fonctionnel connu par le neurone formel. Ce dernier représente une modélisation mathématique du neurone biologique (Figure 3.1).

La Figure 3.2 présente l'analogie entre le neurone formel et biologique qui peut être résumée comme suit :

- Les signaux d'entrée et la sortie du neurone formel ou artificiel représentent respectivement les dendrites et l'axone du neurone biologique.
- Les synapses du neurone biologique sont modélisées par des poids synaptiques (coefficients numériques) associées aux entrées du neurone formel et ajustées pendant la phase d'apprentissage.
- Le corps cellulaire du neurone biologique désigne la fonction de transfert, appelée aussi fonction d'activation du neurone artificiel

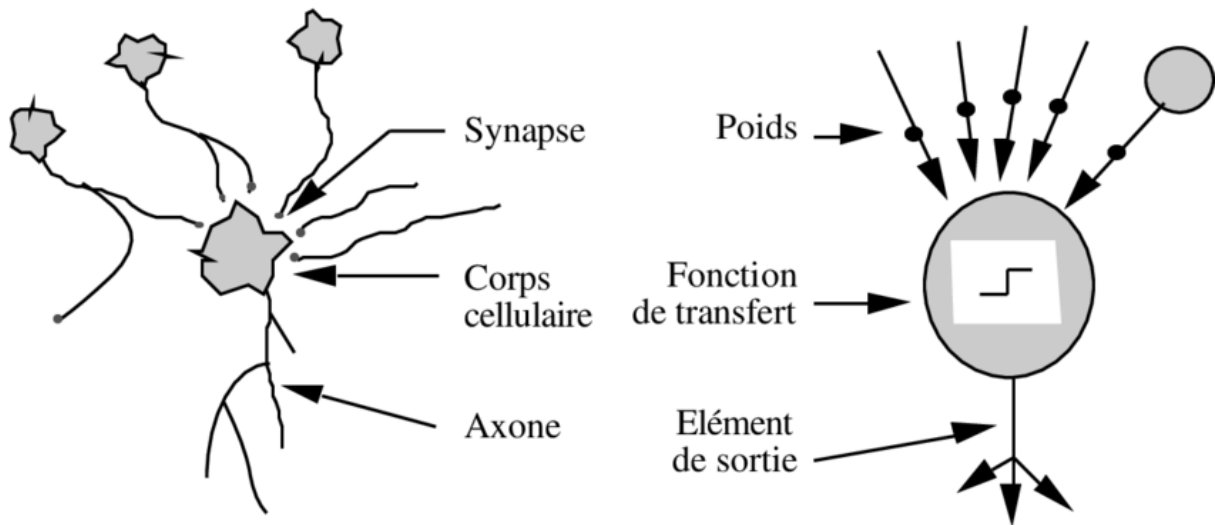


Figure 3. 2. Analogie entre le neurone biologique et le neurone formel.

La procédure générale suivie par le neurone formel pour calculer une sortie est définie par les principales étapes suivantes :

- Multiplier les entrées de la couche d'entrée (X_i), séparément par des poids (W_i), chaque entrée est associée à un poids synaptique d'une valeur numérique notée W_1 pour l'entrée 1 jusqu'à W_i pour l'entrée i .
- Faire, ensuite, la somme de ces produits en ajoutant un biais (b) pour ajuster indépendamment les entrées.
- Transformer le résultat de cette somme à travers une fonction d'activation non linéaire (F) pour produire la sortie (Y) qui est définie par l'équation 1:

$$Y = F(X_1W_1 + X_2W_2 + X_3W_3 + \dots + X_iW_i + b) \quad (3-1)$$

Dans cette thèse, seuls deux types de RNAs à apprentissage supervisé sont utilisés : le modèle Perceptron Multicouches (PMC) et le modèle autorégressif non linéaire avec des entrées exogènes (NARX).

4.1. Perceptron Multicouches (PMC)

Le modèle Perceptron Multicouches (PMC) est un RNA direct ayant une architecture composée de plusieurs couches de neurones où l'information circule directement depuis la couche d'entrée vers la couche de sortie.

Dans notre travail, nous avons utilisé un PCM, entièrement connecté de M entrées, une seule couche cachée de N neurones avec la tangente hyperbolique comme fonction de transfert et un seul neurone linéaire de sortie. Le choix de cette architecture est motivé par le fait qu'il a été démontré que les PMCs directs avec une seule couche cachée sont des estimateurs universels capables de résoudre tout problème de régression ((Gallant and White, 1988), (Hornik et al., 1989), (Pinkus, 1999), (Kainen et al., 2013)). La structure du modèle PMC choisie est présentée dans la Figure 3.3.

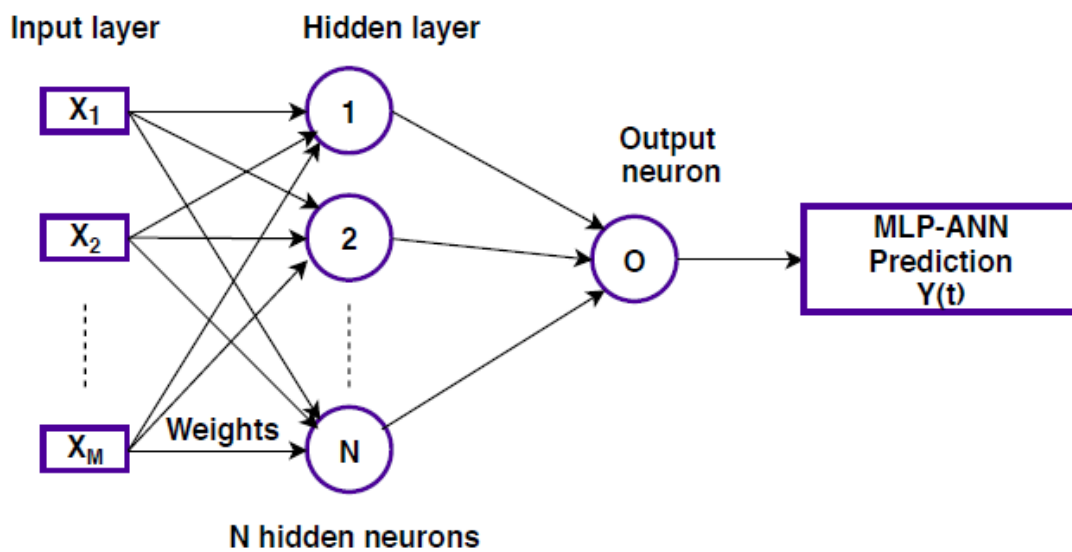


Figure 3. 3. Architecture du PMC.

Dans cette architecture, la relation entre les entrées et la sortie peut être écrite sous la forme suivante :

$$Y = W_{oh} \tanh(W_{hi}X + b_h) + b_o \quad (3-2)$$

où W_{oh} est un vecteur de N éléments représentant les poids des connexions entre les neurones cachés (h : hidden) et le neurone de sortie (o : output), W_{hi} est une matrice NxM dont les éléments sont les poids des connexions entre les entrées (i : input) et les neurones cachés, X est le vecteur colonne des M entrées, b_h est un vecteur colonne représentant les N biais des neurones cachés et b_o est le biais du neurone de sortie. Les poids et les biais sont initialisés aléatoirement et un processus d'apprentissage supervisé utilisant un ensemble de mesures entrées-sorties est utilisé pour les ajuster en minimisant une fonction coût (généralement l'erreur-quadratique entre les sorties calculées et celles mesurées).

L'algorithme choisi pour effectuer l'apprentissage supervisé du PMC est celui de la rétro-propagation du gradient (Back-propagation Algorithm) basé sur la méthode d'optimisation de Levenberg-Marquardt, souvent utilisée pour obtenir une solution numérique dans tout problème de minimisation de fonctions non linéaire à plusieurs variables.

Dans notre étude, nous avons utilisé la fonction tangente hyperbolique ($f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$) et la fonction linéaire ($f(x) = x$) comme fonctions d'activation des neurones cachés et du neurone de sortie, respectivement.

En plus du choix des fonctions de transfert, la construction d'un modèle PMC efficace nécessite un choix judicieux des variables d'entrée pertinentes et un nombre de neurones dans la couche cachée juste suffisant pour éviter le sur-ajustement ou le sous-ajustement. Cependant, il n'existe pas de méthode universelle pour effectuer ces choix. La procédure suivie pour déterminer les entrées et le nombre de neurones cachés optimaux pour modéliser ou prédire l'irradiation solaire, sera expliquée dans les paragraphes ci-dessous (Section 6.3.1.1 et 6.3.1.2)

4.2. Modèles Non linéaires AutoRégressifs avec des entrées eXogènes (NARX)

Etant donné que le modèle Non linéaire AutoRégressif avec entrées eXogènes (NARX) est un bon prédicteur des séries temporelles ((Pisoni et al., 2009), (Ruiz et al., 2016)), nous l'avons adopté parmi les réseaux de neurones dynamiques récurrents pour prédire l'irradiation solaire. Le modèle NARX représente une généralisation non linéaire du modèle autorégressif exogène (ARX), qui est un modèle standard dans l'identification des systèmes linéaires de type boîte noire (Ferreira et al., 2012). Un réseau de neurones NARX est un modèle récurrent possédant des connexions récurrentes permettant de relier la valeur actuelle d'une série chronologique aux valeurs passées de la même série et aux valeurs actuelles et passées des entrées exogènes. Ce type de modèles peut être utilisé dans diverses applications telles que la modélisation d'une grande variété de systèmes dynamiques non linéaires (Thiaw, 2008).

Le concept du réseau de neurones NARX est le même que celui d'un modèle Perceptron Multicouches (PMC) auquel des sorties et, éventuellement, des entrées retardées sont ajoutées (Figure 3.4) et dont l'apprentissage se fait par l'algorithme de rétro-propagation du gradient. Le degré de récurrence dépend de l'horizon temporel considéré.

Afin d'obtenir de bonnes prestations du modèle NARX dans la prédiction des séries temporelles non linéaires, il est intéressant d'utiliser sa capacité de mémoire en employant les valeurs passées des séries chronologiques prédites ou réelles (Boussaada et al., 2018). Le modèle NARX possède deux différents types d'architecture qui sont présentées sur la figure 3.4: une architecture en boucle ouverte (série-parallèle) et une architecture en boucle fermée (parallèle). Les équations qui caractérisent le modèle NARX en boucle ouverte ou fermée sont présentées ci-dessous :

- Architecture série parallèle (boucle ouverte) :

$$\hat{y}(t) = f(x(t), x(t-1), \dots, x(t-d), \dots, y(t-1), \dots, y(t-d)) \quad (3-3)$$

- Architecture parallèle (boucle fermée) :

$$\hat{y}(t) = f(x(t), x(t-1), \dots, x(t-d), \hat{y}(t), \hat{y}(t-1), \dots, \hat{y}(t-d)) \quad (3-4)$$

Dans l'architecture série parallèle (boucle ouverte), la valeur de la sortie calculée à l'instant t ($\hat{y}(t)$) est prédite en fonction des :

- Valeurs actuelles $x(t)$ et précédentes $x(t-1), \dots, x(t-d)$ des entrées.
- Valeurs précédentes mesurées $y(t-1), \dots, y(t-d)$ de la sortie.

Dans l'architecture parallèle (boucle fermée), la valeur de la sortie calculée $\hat{y}(t)$ est prédite en fonction des :

- Valeurs actuelles $x(t)$ et précédentes $x(t-1), \dots, x(t-d)$ des entrées.
- Valeurs précédentes prédites $\hat{y}(t-1), \dots, \hat{y}(t-d)$ de la sortie.

Comme pour le réseau PMC, l'algorithme d'entraînement sélectionné pour le réseau de neurones NARX est celui de la rétro-propagation basée sur la méthode de minimisation Levenberg-Marquardt. Aussi, la fonction d'activation de la couche cachée choisie est une tangente hyperbolique, tandis qu'une fonction linéaire est utilisée pour l'unité de sortie.

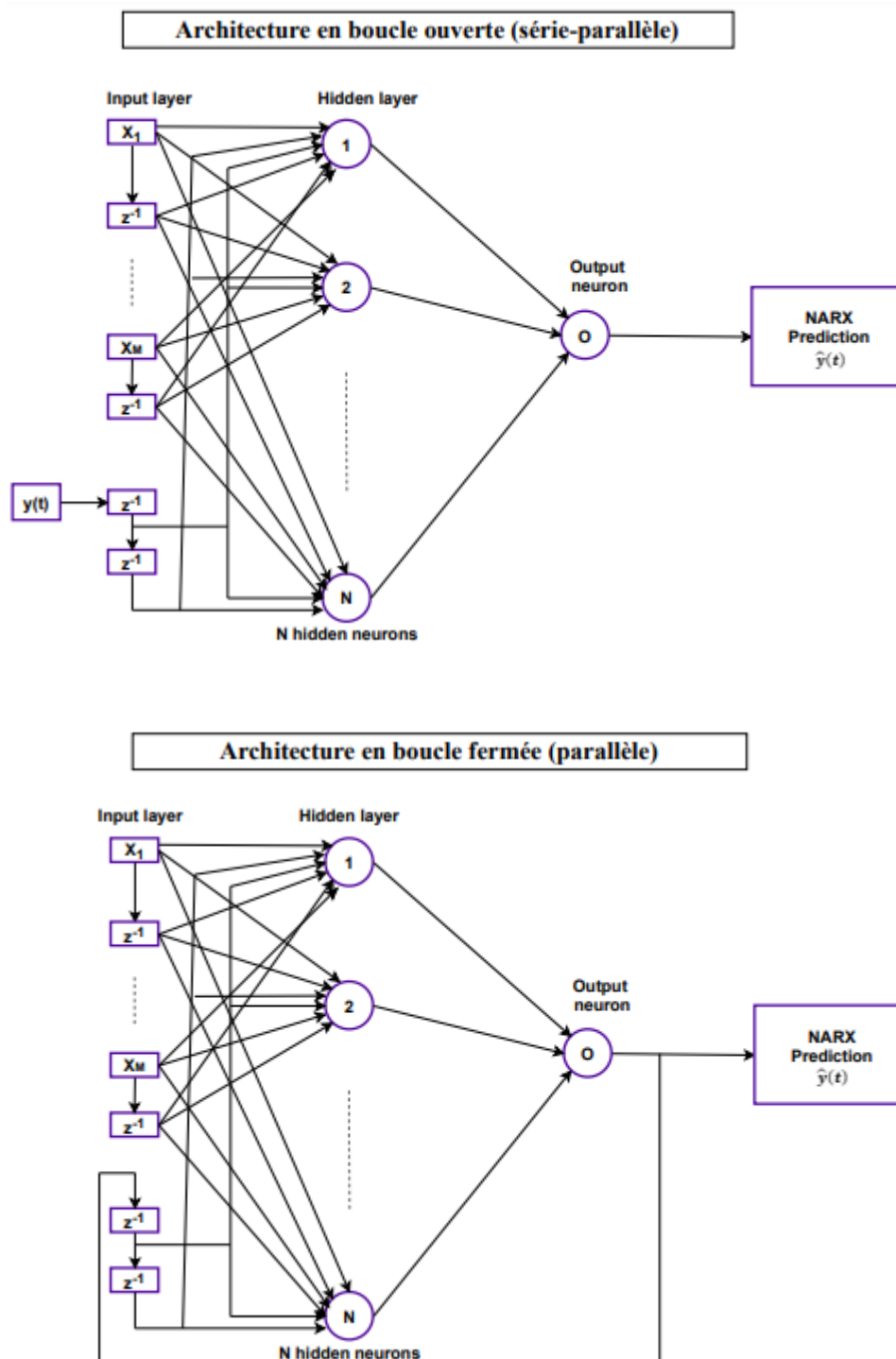


Figure 3. 4. Architectures du réseau de neurones NARX.

Généralement, la procédure typique suivie pour créer le réseau de neurones NARX consiste à utiliser, dans un premier temps, l'architecture série-parallèle (boucle ouverte) durant la phase d'entraînement, vue que ce type d'architecture présente deux principaux avantages. Le premier

avantage est que l'utilisation des valeurs réelles de la sortie en entrées du réseau récurrent peut contribuer à l'amélioration de la précision de la prédiction. Le deuxième avantage est que le réseau résultant à une architecture analogue à celle du PCM et peut donc être entraîné à l'aide de la méthode de rétro-propagation du gradient. Une fois notre réseau NARX est entraîné, il peut être transformé, par la suite, en une architecture parallèle (boucle fermée) pour générer une prédiction anticipée de plusieurs valeurs de la cible considérée.

5. Méthodes d'ensemble

Le principe des méthodes d'ensembles se base sur la simple idée que « l'union crée la force ». Si on réunit un ensemble de prédicteurs, chacun est individuellement différent des autres et donne des résultats relativement bons, le prédicteur agrégé donnera tout naturellement des résultats significativement bons par rapport aux résultats obtenus par chaque prédicteur individuellement (Marsland, 2015). Dans le cas des êtres humains, travailler en équipe et réunir un ensemble de cerveau humain ayant le même objectif donnera normalement des résultats meilleurs que de travailler de manière individuelle.

Contrairement aux approches d'apprentissage ordinaires, qui tentent de construire un seul apprenant à partir des données d'entraînement, les méthodes d'ensemble construisent un ensemble d'apprenants et les combinent (Zhou, 2012). Et, le fait de réunir un ensemble d'apprenants élargit l'horizon des solutions. Certains apprenants (learners) apprennent mieux certaines choses et les autres apprennent mieux d'autres, donc après agrégation on peut s'attendre à obtenir un prédicteur qui donnera des prédictions plus proches de la réalité. La figure 3.5 illustre l'architecture commune des méthodes d'ensemble.

Dans notre approche, deux types d'algorithmes des méthodes d'ensemble sont considérés: les algorithmes basés sur la construction adaptative (Boosting) ou aléatoire (Bagging et Random forest (forêts aléatoires)). Pour ces trois méthodes, les phases d'entraînement et de validation sont effectuées en utilisant plusieurs combinaisons des entrées, et leurs performances sont examinées sur la base de l'ensemble de données de validation, en exécutant le processus de chaque méthode d'ensemble plusieurs fois.

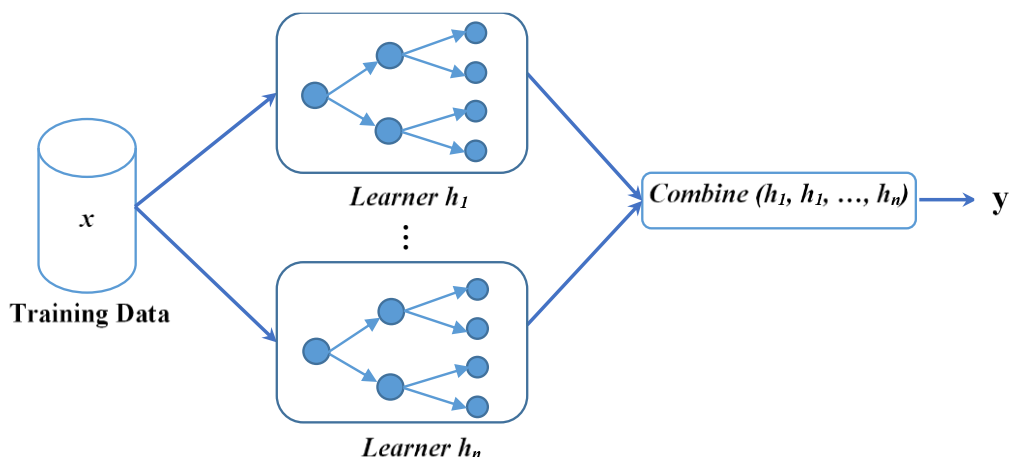


Figure 3. 5. Architecture des méthodes d'ensemble.

5.1. Boosting

Le Boosting est une méthode d'ensemble séquentielle qui a été introduite en 1996 par Freund et Schapire (Freund and Schapire, 1996). La principale caractéristique de cet algorithme est sa capacité à convertir les apprenants faibles (weak learners) en apprenants forts : a) en associant séquentiellement les sorties des apprenants faibles pendant la phase d'entraînement, et b) en attribuant, à chaque itération, des poids plus élevés aux échantillons mal prédits. Il s'agit d'une méthode d'apprentissage renforcée permettant de réduire la fonction de perte (loss function) qui est simplement une mesure de l'erreur d'ajustement. La minimisation de cette fonction s'effectue en utilisant une procédure itérative qui consiste à ajouter un nouveau prédicteur basé sur les résidus des anciens.

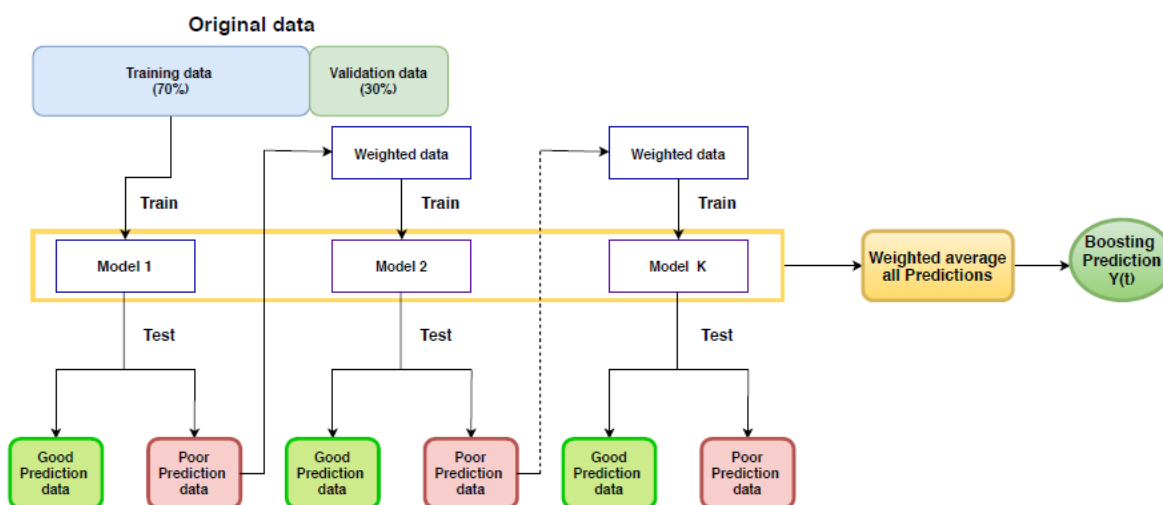


Figure 3. 6. Diagramme du processus du Boosting.

Le ré-échantillonnage dépend de la performance du prédicteur généré par l'échantillon précédent. L'échantillon suivant est choisi de manière à se concentrer sur les faiblesses du prédicteur précédent, c'est-à-dire sur les valeurs prédites erronées. L'objectif est de ré-entraîner la partie la plus complexe de la base de données. Le diagramme présentant le processus du Boosting est montré dans la figure 3.6.

Tableau 3. 2. Procédure générale du Boosting pour la régression (Gey and Poggi, 2006).

Soit x_0 à prévoir ,

$Z = \{(x_1, y_1), \dots, (x_n, y_n)\}$, un échantillon d'entraînement de taille n , et

K le nombre d'itérations

Initialiser p par la distribution uniforme D , $p_1 = D$ sur $\{1, \dots, n\}$.

Pour k allant de 1 à K :

- Tirer aléatoirement de Z avec remplacement, en fonction de p_k , un échantillon Z_k de taille n .
- Estimer \hat{f}_k sur l'échantillon Z_k .
- Calculer à partir de l'échantillon initial Z : $i=1, \dots, n$

$$l_k(i) = ((Y_i - \hat{f}_k(X_i))^2 \text{ et } \varepsilon_{p_k}^{(boost)} = \sum_{i=1}^n p_k(i) l_k(i)$$

$$\beta_k = \frac{\varepsilon_{p_k}^{(boost)}}{\max_{1 \leq i \leq n} l_k(i) - \varepsilon_{p_k}^{(boost)}} \text{ et } d_k(i) = \frac{l_k(i)}{\max_{1 \leq i \leq n} l_k(i)}$$

$$\text{Si } \varepsilon_{p_k}^{(boost)} < 0.5 \max_{1 \leq i \leq n} l_k(i) \text{ , } \omega_{k+1}(i) = \beta_k^{1-d_k(i)} p_k(i),$$

$$\text{Sinon } \omega_{k+1} = p_1,$$

Fin si

- Calculer les nouvelles probabilités : $p_{k+1}(i) = \frac{\omega_{k+1}(i)}{\sum_{j=1}^n \omega_{k+1}(j)}$

Fin pour

Calculer $\hat{f}(x_0)$ moyenne ou médiane des prévisions $\hat{f}_k(x_0)$ pondérées par des coefficients $\log\left(\frac{1}{\beta_k}\right)$

Le tableau 3.2 résume de manière générale la procédure suivie par la méthode Boosting pour résoudre un problème de régression. Cet algorithme génère K prédicteurs à partir d'un ensemble d'échantillons bootstrap Z_k , dont le tirage se fait aléatoirement et dépend des probabilités p mises à jour à chaque itération. A l'itération k , le calcul de la distribution p_{k+1} permettra

d'augmenter la probabilité d'apparition des observations de l'échantillon Z_k mal prédites par \hat{f}_k dans le prochain échantillon Z_{k+1} , pour que le prochain prédicteur \hat{f}_{k+1} se concentrera davantage sur ces observations (les données présentant le plus grand nombre d'erreurs). Ainsi, la mise à jour des probabilités est effectuée en fonction du paramètre β_k qui est considéré comme un indicateur de la performance de \hat{f}_k sur l'échantillon Z_k , c'est-à-dire, plus la valeur de β_k est petite, plus l'estimation de \hat{f}_k est performante. Et puisque, $d_k(i)$ représente la performance individuelle de \hat{f}_k sur l'échantillon Z , la multiplication de $p_k(i)$ par $\beta_k^{1-d_k(i)}$ fusionne dans p_{k+1} la performance globale de \hat{f}_k sur Z_k et aussi les performances individuelles sur Z . À la fin, l'estimation finale est obtenue en calculant la moyenne ou la médiane de toutes les prévisions pondérées par les coefficients correspondants à la performance respective de chacune de ces prévisions (Gey and Poggi, 2006).

5.2. Bagging

Leo BREIMAN a développé l'algorithme du Bagging en 1996 (Breiman, 1996). Cette méthode est basée sur deux étapes principales : le bootstrap et l'agrégation. La première étape consiste à collecter, de manière aléatoire et indépendante, des échantillons de l'ensemble d'apprentissage et à construire des prédicteurs. L'échantillonnage est effectué de manière répétitive, avec remplacement à partir de l'échantillon initial, d'où le nom de bootstrap. La deuxième étape est l'agrégation, qui consiste à généraliser les résultats en combinant les différents prédicteurs construits. Le processus de cet algorithme est expliqué sous-forme d'un diagramme dans la figure 3.7.

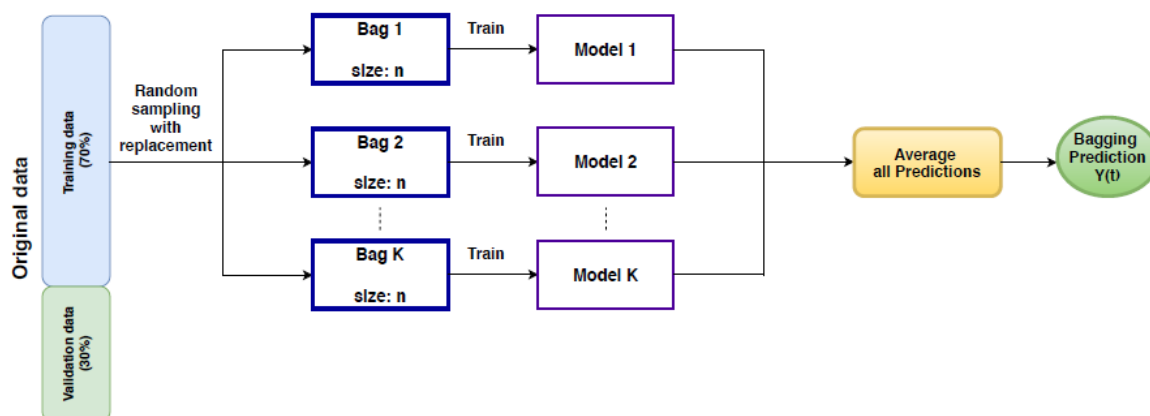


Figure 3. 7. Diagramme du processus du Bagging.

Pour l'ensemble de données d'entraînement, un tirage au sort est effectué en fonction du nombre d'observations afin d'obtenir K répliques d'échantillons bootstrap. Ensuite, une estimation est effectuée pour chaque échantillon bootstrap Z_k^* . A la fin, l'estimation finale de l'ensemble est obtenue en suivant la stratégie populaire adoptée par le Bagging pour résoudre les problèmes de régression qui consiste à agréger les sorties de l'apprenant de base en calculant la moyenne des estimations obtenues à partir de chaque échantillon. L'algorithme de régression présentant la procédure générale du Bagging est donné par le tableau 3.3.

Tableau 3. 3. Procédure générale du Bagging pour la régression (Agrégation de modèles, 2018).

Soit x_0 à prévoir et

$Z = \{(x_1, y_1), \dots, (x_n, y_n)\}$, un échantillon d'entraînement de taille n

K le nombre d'itérations

Pour k allant de 1 à K :

- Tirer un échantillon bootstrap Z_k^* .
- Estimer $\hat{f}_{Z_k}(x_0)$ sur l'échantillon bootstrap.

Fin pour

Déterminer l'estimation finale en calculant la moyenne des estimations de chaque échantillon :

$$\hat{f}_K(x_0) = \frac{1}{K} \sum_{k=1}^K \hat{f}_{Z_k}(x_0).$$

5.3. Forêts aléatoires (Random Forests)

Les forêts aléatoires sont la combinaison de plusieurs arbres aléatoires construits sur la base de la méthode Bagging. Elle a été proposée par Breiman (Breiman, 2001), et représente l'agrégation ou l'association d'un certain nombre de prédicteurs, chacun développé différemment, à partir d'un échantillon arbitraire d'arbre de décision. Un arbre de régression est développé pour chaque échantillon avec une sélection aléatoire des variables d'entrée et une sélection arbitraire d'un sous-ensemble de prédicteurs. Dans cette méthode, la performance de la prédiction dépend du nombre d'arbres sélectionnés à l'origine dans la forêt, c'est-à-dire que plus le nombre d'arbres est élevé, plus l'erreur de généralisation tend vers une limite. La différence d'erreur trouvée entre les arbres de décisions permet de les compléter afin de construire des arbres efficaces individuellement. Un diagramme du processus de la méthode des forêts aléatoires est illustré à la figure 3.8.

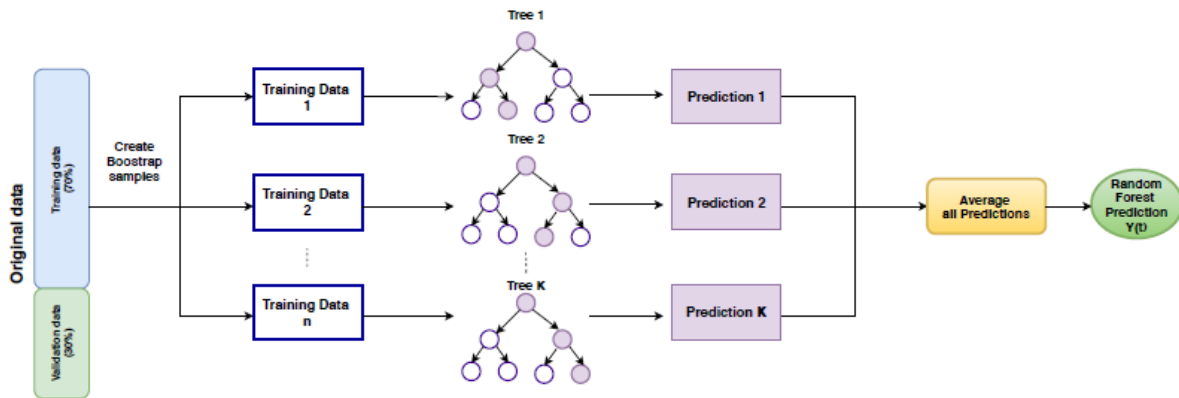


Figure 3. 8. Diagramme du processus des forêts aléatoires (Random Forests).

Pour le modèle des forêts aléatoires, le tirage de K répliques d'échantillons bootstrap est effectué de la même manière que dans l'algorithme de la méthode Bagging pour l'ensemble de données d'entraînement. La différence entre ces deux modèles est que dans les forêts aléatoires, un arbre de régression est développé pour chaque échantillon avec une sélection aléatoire des variables d'entrée et un tirage arbitraire d'un sous-ensemble de prédicteurs pour trouver la division optimale sur chaque nœud. L'estimation finale de l'ensemble est obtenue en calculant l'estimation moyenne de tous les arbres développés. La procédure générale des forêts aléatoires pour résoudre un problème de régression est présentée dans le tableau 3.4.

Tableau 3. 4. Procédure générale des forêts aléatoires pour la régression (Agrégation de modèles, 2018).

Soit x_0 à prévoir et $Z = \{(x_1, y_1), \dots, (x_n, y_n)\}$, un échantillon d'entraînement de taille n

Pour k allant de 1 à K :

- Tirer un échantillon bootstrap Z_k^* .
- Développer un arbre de régression pour cet échantillon avec randomisation des variables : la recherche de chaque division optimale est précédée d'un tirage aléatoire d'un sous-ensemble de m prédicteurs.
- Estimer $\hat{f}_{Z_k}(x_0)$ pour l'arbre développée.

Fin pour

Calculer l'estimation moyenne $\hat{f}_K(x_0) = \frac{1}{K} \sum_{k=1}^K \hat{f}_{Z_k}(x_0)$.

6. Régression par machine à vecteur de support (SVM)

La machine à vecteur de support (SVM), inventée par Vladimir Vapnik et ses collaborateurs (Vapnik, 2013), se caractérise par l'utilisation de noyaux (kernel), la parcimonie de la solution, l'absence de minima locaux et le contrôle de la capacité obtenue en agissant sur la marge ou le nombre de vecteurs de support. La machine à vecteur de support est une technique non paramétrique basée sur le principe de la minimisation du risque structurel, en réduisant l'erreur de généralisation de la limite supérieure (Muhammed A. Hassan et al., 2017).

La machine à vecteur de support peut être appliquée dans le cas de problèmes de régression (Awad and Khanna, 2015), en utilisant les mêmes caractéristiques pour caractériser l'algorithme de la marge maximale (une fonction non linéaire qui est adaptée par un mappage linéaire d'apprentissage automatique dans un espace de caractéristiques à haute dimension induit par un noyau). La capacité du système est contrôlée par les paramètres qui ne dépendent pas de la dimension de l'espace des caractéristiques. Pareillement à l'approche de classification, la motivation de cet algorithme consiste à rechercher et optimiser les limites de généralisation qui sont données par la régression. Ces limites sont utilisées pour définir la fonction de perte qui ignore les erreurs qui sont situées à une distance donnée de la valeur réelle.

La régression SVM utilise une entrée x qui est cartographiée dans un espace de caractéristiques à m dimensions. Ensuite, un modèle linéaire est réalisé dans cet espace caractéristique. La notation mathématique pour le modèle linéaire $f(x)$ est donnée par :

$$f(x) = \sum_{j=1}^m w_j g_j(x) + b \quad (3-5)$$

où w_j est le vecteur de poids, $g_j(x), j = 1, \dots, m$, désigne les caractéristiques de l'espace de haute dimension qui mappent le vecteur d'espace d'entrée x , et b est le terme du biais. Les équations détaillées du modèle SVMR sont disponibles dans les travaux originaux Vapnik ((Vapnik, 1995),(Vapnik, 1998)).

La Figure 3.9 montre un exemple d'une régression linéaire unidimensionnelle. Ce problème d'approximation de la fonction est formulé par le SVR comme étant un problème d'optimisation qui cherche à étreindre le plus possible le tube centré autour de la surface, tout en minimisant l'erreur de prédiction (la distance entre la sortie souhaitée et prédite). Pour cela, le SVR adopte une fonction de perte insensible ϵ pour pénaliser les prédictions qui sont éloignées

des sorties souhaitées. La valeur de la fonction de perte ε détermine la largeur du tube, ce qui explique que plus la valeur de ε est petite, plus la marge d'erreur entre la prédiction et l'observation est faible. Ainsi, cela permettra également d'affecter le nombre de vecteurs de support et, par conséquent, la finesse de la solution. Selon la figure 3.9, il est clairement visible que plus la valeur de ε est petite, plus la limite du tube est déplacée vers l'intérieur (Awad and Khanna, 2015). Plusieurs fonctions de perte peuvent être adoptées pour minimiser l'erreur de prédiction, notamment, la fonction linéaire, quadratique, ou de Huber, comme le montre la figure 3.10. L'asymétrie de ces trois fonctions de perte permet de limiter la sous-estimation ou la surestimation et leur convexité permet de garantir que le problème d'optimisation a une solution unique qui peut être trouvée en un nombre fini d'étapes (Awad and Khanna, 2015). De plus, le choix de la fonction de perte est déterminé en fonction de plusieurs paramètres telles que la distribution du bruit affectant les échantillons de données (Huber, 1964), la spécificité du modèle recherchée et la complexité du calcul d'entraînement (Awad and Khanna, 2015).

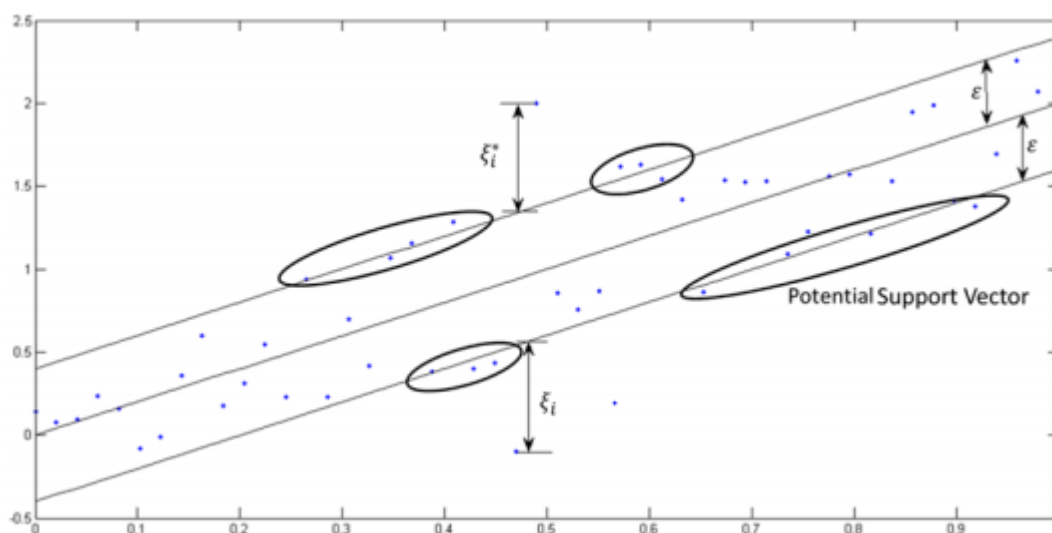


Figure 3. 9. Régression linéaire unidimensionnelle SVR (Awad and Khanna, 2015).

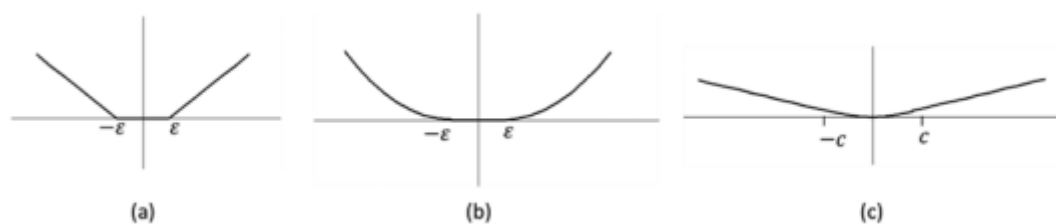


Figure 3. 10. Types de fonctions de perte : a) linéaire, b) quadratique, c) Huber (Awad and Khanna, 2015).

7. Résultats et discussions

Plusieurs tests ont été effectués pour estimer les irradiations solaires journalières aux différents sites d'étude, en utilisant les modèles d'apprentissage automatique présentés ci-dessus. Les résultats obtenus seront présentés et discutés brièvement dans la présente section, tout en analysant et comparant les performances des modèles d'apprentissage automatique avec celles des modèles empiriques présentés dans le chapitre 2. La comparaison des performances sera basée sur les indicateurs statistiques suivants : coefficient de corrélation (R), coefficient de détermination (R^2), racine de l'erreur quadratique moyenne normalisée ($nRMSE$), pourcentage de l'erreur absolue moyenne ($MAPE$) et l'erreur absolue moyenne normalisée ($nMAE$) (Annexe D).

7.1. Performances des réseaux de neurones artificiels pour l'estimation de GHI sur la base de données de la ville de Fès

Dans la présente étude, nous nous intéressons à l'estimation de l'irradiation solaire globale quotidienne reçue par une surface horizontale (GHI) à la ville de Fès, à partir de quelques variables météorologiques couramment mesurées (température, humidité relative de l'air, vitesse du vent) (Bounoua and Mechaqrane, 2018). Nous testons deux types de réseaux de neurones artificiels (RNAs) : le Perceptron MultiCouche (PMC) et le réseau de neurones non linéaire autorégressif avec entrées exogènes (NARX). Dans notre cas, nous avons utilisé un PMC à propagation directe (feedforward) avec M entrées, une couche cachée avec N neurones tangents hyperboliques et un neurone linéaire de sortie (Section 3.1). Aussi, l'architecture du modèle NARX est la même que celle du PMC à laquelle des sorties retardées sont ajoutées (Section 3.2). Le degré de récurrence dépend de l'horizon temporel considéré. Ces deux architectures ont été optimisées en choisissant la meilleure combinaison des paramètres d'entrée et en optimisant le nombre de neurones cachés. De plus, la performance de chaque modèle (PMC et NARX) est évaluée à l'aide des trois indicateurs statistiques R^2 , $nRMSE$ et $MAPE$ (Annexe D).

Pour établir un modèle RNA supervisé, notre base de données a été divisée en trois ensembles : les données de la période entre 2009 et 2011 sont utilisées pour l'ensemble d'entraînement, les données de 2012 à 2013 sont utilisées pour l'ensemble de test et les données de 2014 est utilisé pour l'ensemble de validation. Pour les deux modèles (PMC et NARX), nous avons commencé par utiliser une architecture avec 5 neurones cachés. Ce nombre sera optimisé ultérieurement.

Chapitre 3 : Estimation des irradiations solaires journalières en utilisant les modèles d'apprentissage automatique.

En outre, et puisque l'initialisation des poids et du biais des RNAs se fait de manière aléatoire, nous prenons les meilleurs résultats sur cinq exécutions.

7.1.1. Performances du modèle PMC

Le modèle PMC est testé pour estimer les valeurs journalières de GHI en utilisant comme variables d'entrées différentes combinaisons des paramètres météorologiques couramment mesurés, tels que la température moyenne de l'air ambiant T ses valeurs extrêmes T_{max} et T_{min} , le gradient journalier de la température $\Delta T = T_{max} - T_{min}$, l'humidité relative de l'air RH , ses valeurs extrêmes RH_{max} et RH_{min} , le gradient journalier de l'humidité relative $\Delta H = RH_{max} - RH_{min}$, la vitesse moyenne journalière du vent W_s , le jour de l'année, DOY et l'irradiation solaire journalière hors atmosphère G_{toa} . Plusieurs combinaisons ont été testées. Le tableau 3.5 montre les résultats obtenus pour certaines d'entre elles.

Tableau 3. 5. Performances du PMC pour l'estimation du GHI quotidienne à Fès (Bounoua and Mechaqrane, 2018).

DOY	G_{toa}	T	T_{max}	T_{min}	RH	RH_{max}	RH_{min}	W_s	ΔT	ΔH	$\sqrt{\Delta T}$	$\sqrt{\Delta H}$	R^2 (%) Train Validation Test	MAPE (%) Train Validation Test	nRMSE (%) Train Validation Test
X	X												94.96	30.84	23.78
													92.93	33.74	28.14
													95.15	26.02	23.28
X	X	X											95.56	28.63	22.44
													93.52	30.91	27.03
													95.36	24.55	22.88
X	X		X										96.70	22.86	19.45
													95.03	25.18	23.79
													96.35	20.25	20.39
X				X			X						97.67	17.84	16.42
													96.54	22.60	19.91
													97.98	15.89	15.10
X			X	X		X	X						98.28	16.28	14.14
													96.71	20.07	19.63
													98.14	14.52	14.58
	X							X	X				98.19	15.88	14.47
													96.33	19.83	20.68
													98.01	14.49	15.03
	X		X	X		X	X					X	98.34	15.84	13.89
													96.98	19.99	18.71
													98.28	14.07	13.98
X	X		X	X		X	X					X	98.35	15.53	13.81
													96.98	19.46	18.69
													98.17	13.95	14.40
X	X							X					98.14	16.43	14.66
													96.25	19.86	20.97
													97.91	14.48	15.43

Chapitre 3 : Estimation des irradiations solaires journalières en utilisant les modèles d'apprentissage automatique.

En considérant les données de l'ensemble de validation, les meilleurs résultats sont obtenus à partir de la combinaison constituée par DOY , G_{toa} , T , T_{max} , T_{min} , RH_{max} , RH_{min} et $\sqrt{\Delta T}$ comme entrées (ligne en fond gris dans le tableau 3.5).

7.1.2. Performances du modèle NARX

Pour le modèle NARX, nous avons utilisé les mêmes entrées que dans le modèle PMC auxquelles sont ajoutées certaines entrées et des valeurs de sortie retardées. Le tableau 3.6 montre les résultats obtenus (Bounoua and Mechaqrane, 2018).

A partir de ces résultats, on constate que nous pouvons obtenir l'estimation du GHI quotidien à différents horizons du temps, avec approximativement la même précision (lignes en fond gris dans le tableau 3.6). Dans cette partie, le degré de récurrence dépend de trois horizons temporels, un horizon d'un jour, deux jours ou trois jours. Nous notons que les meilleurs résultats pour chaque horizon x sont obtenus en utilisant la même meilleure combinaison du PMC avec des valeurs récurrentes de $GHI-x$, T_{max-x} et T_{min-x} .

Tableau 3. 6. Performances du modèle NARX pour l'estimation du GHI à différents horizons temporels (jour j+1, j+2 et j+3).

G_{toa}	T_{max}	T_{min}	RH_{max}	RH_{min}	$\sqrt{\Delta T}$	DOY_{j-1}	GHI_{-1}	G_{toa-1}	T_{max-1}	T_{min-1}	D_{j-2}	GHI_{-2}	G_{toa-2}	T_{max-2}	T_{min-2}	D_{j-3}	GHI_{-3}	G_{toa-3}	T_{max-3}	T_{min-3}	R^2 (%)	MAPE (%)	nRMS E (%)
																					Train Validation Test	Train Validation Test	Train Validation Test
X	X	X	X	X	X		X		X												98.43	15.20	13.49
																					96.87	18.13	19.06
																					98.23	13.35	14.16
X	X	X	X	X	X		X			X											98.37	14.99	13.77
																					96.94	17.58	18.95
																					98.16	13.59	14.46
X	X	X	X	X	X		X		X	X											98.34	15.09	13.84
																					97.47	16.58	17.19
																					98.48	13.11	13.17
X	X	X	X	X	X							X		X							98.35	15.14	13.82
																					96.84	18.58	19.15
																					98.14	14.03	14.54
X	X	X	X	X	X							X			X						98.36	15.10	13.81
																					96.59	18.92	19.95
																					98.09	13.50	14.75
X	X	X	X	X	X							X		X	X						98.36	15.26	13.78
																					96.97	19.10	18.72
																					98.43	13.69	13.31
X	X	X	X	X	X												X		X		98.38	14.15	13.66
																					95.95	20.23	21.56
																					97.95	14.12	15.20
X	X	X	X	X	X												X		X	X	98.38	15.04	13.66
																					96.95	19.90	18.69
																					98.25	13.77	14.09
X	X	X	X	X	X		X					X					X				98.38	14.69	13.65
																					96.49	19.38	20.07
																					98.18	13.79	14.35

7.1.3. Optimisation du nombre de neurones cachés

Il est bien connu qu'un RNA donne les meilleurs résultats lorsque son architecture est optimale. Par conséquent, il est toujours nécessaire d'éviter le sous-ajustement (En anglais Underfitting : faible nombre de neurones et faible complexité) et le sur-ajustement (En anglais Overfitting : nombre élevé de neurones) qui conduisent tous les deux à de mauvaises performances. Nous devons donc trouver l'architecture juste suffisante pour bien représenter notre problème.

Dans cette section, nous présentons les résultats obtenus après l'optimisation des architectures des deux modèles PMC et NARX. Ces architectures sont optimisées en choisissant le nombre de neurones dans la couche cachée de chaque modèle. Pour cela, une approche constructive est utilisée permettant de tester les performances des différentes architectures avec un nombre de neurones cachés allant de 1 à 20 neurones. Le tableau 3.7 présente les meilleures architectures obtenues et les performances correspondantes.

Tableau 3. 7. Meilleures architectures et leurs performances (Bounoua and Mechaqrane, 2018).

Modèle	Nombre de neurones cachés	R^2 (%)	MAPE(%)	nRMSE(%)
		Train Validation Test	Train Validation Test	Train Validation Test
PMC	11	98.35	15.58	13.82
		97.14	19.50	18.19
		98.45	13.49	13.24
		98.47	14.51	13.35
NARX _{d+1}	17	97.25	17.49	17.89
		98.55	12.96	12.80
		98.49	14.90	13.25
NARX _{d+2}	14	96.76	19.03	19.59
		98.47	13.19	13.55
		98.38	14.30	13.70
NARX _{d+3}	7	97.12	18.42	18.24
		98.35	13.13	13.73

Les figures 3.11 à 3.14 représentent, pour toutes les bases de données et pour tous les modèles testés, les régressions correspondant aux meilleures performances obtenues. La valeur R au-dessus de chaque graphique représente le coefficient de corrélation. Nous pouvons voir que pour tous les modèles, le $nRMSE$ est inférieur à 20% et le coefficient de détermination est supérieur à 97% (Bounoua and Mechaqrane, 2018).

Chapitre 3 : Estimation des irradiations solaires journalières en utilisant les modèles d'apprentissage automatique.

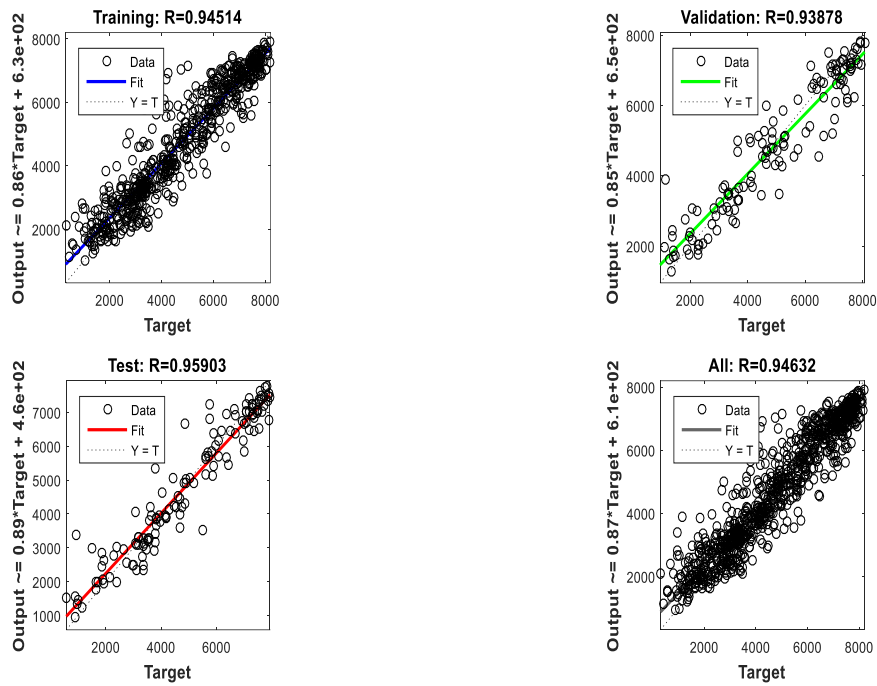


Figure 3. 11. Régressions obtenues par le meilleur modèle PMC.

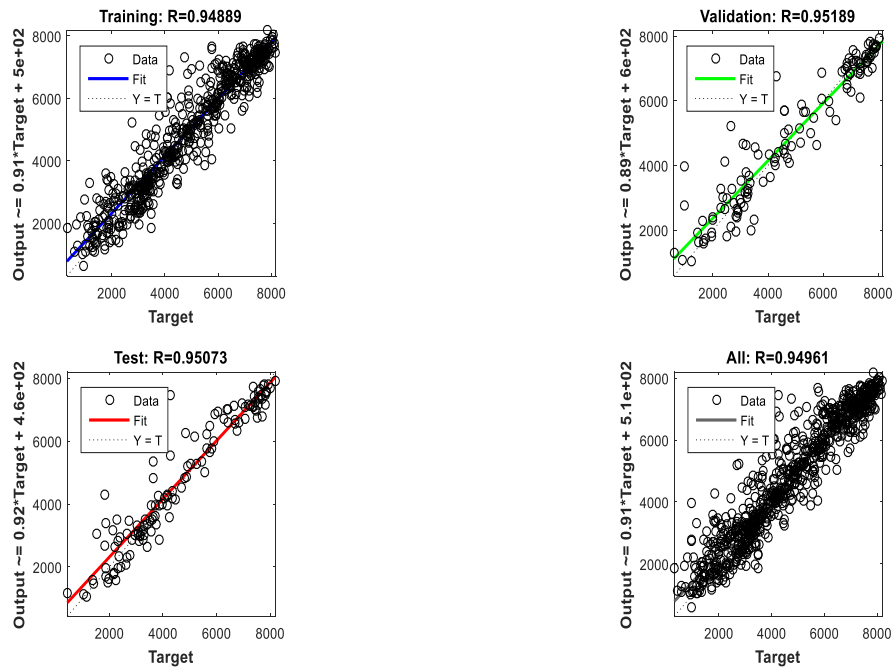


Figure 3. 12. Régressions obtenues par le meilleur modèle NARX avec une récurrence d'un jour.

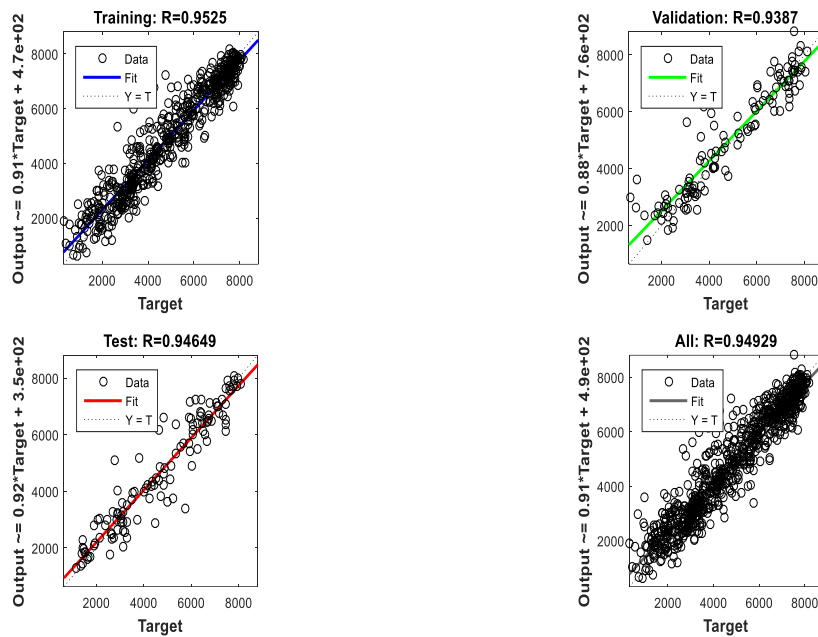


Figure 3.13. Régressions obtenues par le meilleur modèle NARX avec une récurrence de deux jours.

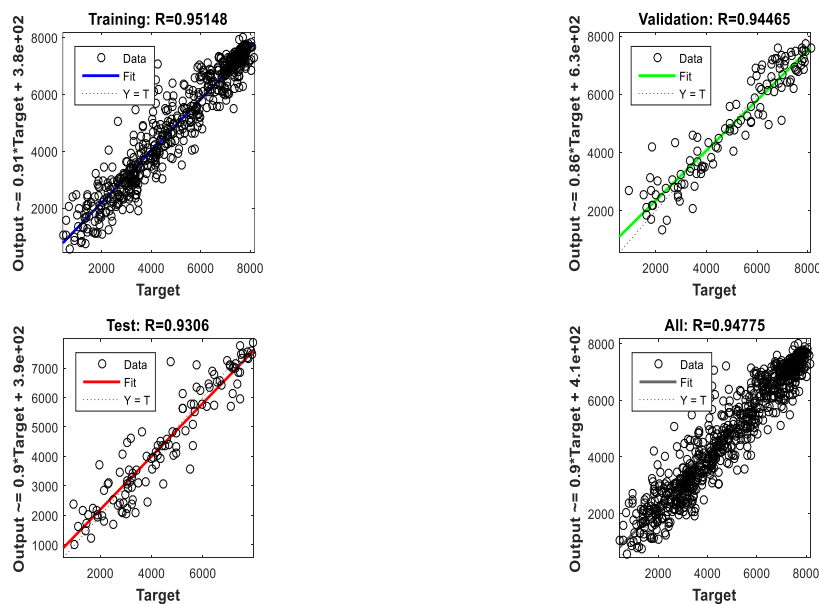


Figure 3.14. Régressions obtenues par le meilleur modèle NARX avec une récurrence de trois jours.

7.2. Performances des meilleures architectures obtenues des deux modèles PMC et NARX sur les bases de données enerMENA.

Les meilleures architectures optimisées des deux modèles PMC et NARX obtenues précédemment, ont été testées sur les autres sites d'enerMENA. Les performances statistiques obtenues pour chaque station sont présentées dans le tableau 3.8, pour l'ensemble de validation.

Chapitre 3 : Estimation des irradiations solaires journalières en utilisant les modèles d'apprentissage automatique.

Ces résultats montrent que la meilleure précision du modèle PMC est obtenue à Zagora avec des valeurs de R et $nRMSE$ égales 98.68% et 11.84%, respectivement pour l'ensemble de validation. De même, la meilleure performance d'estimation journalière de GHI à l'aide du modèle NARX est celle trouvée à Zagora avec des valeurs de R et $nRMSE$ égales respectivement à 98.87% et 10.40%, pour un horizon d'un jour, à 98.48% et 12.89%, pour un horizon de deux jours et à 98.73% et 11.56%, pour un horizon de trois jours, pour l'ensemble de validation. Par contre, les plus faibles performances d'estimation journalière de GHI obtenues par les deux modèles PMC et NARX ont été constatées à TanTan ($20\% < nRMSE < 25\%$). Ceci est peut-être dû au fait que les données collectées en ce site présentent une variabilité météorologique élevée (Ceci sera expliqué brièvement dans la section suivante).

Tableau 3. 8. Performances statistiques des meilleures architectures des deux modèles PMC et NARX appliqués aux sites enerMENA, pour l'ensemble de validation.

Site	Modèle	R^2	MAPE	nRMSE
		(%) Validation	(%) Validation	(%) Validation
Erfoud	PMC (11 neurones)	98.67	11.26	12.16
	NARX _{d+1} (17 neurones)	98.87	9.27	10.89
	NARX _{d+2} (14 neurones)	98.06	13.36	14.86
	NARX _{d+3} (7 neurones)	97.98	13.21	14.82
Missour	PMC (11 neurones)	97.79	16.83	16.02
	NARX _{d+1} (17 neurones)	97.75	15.63	15.47
	NARX _{d+2} (14 neurones)	97.40	16.22	16.78
	NARX _{d+3} (7 neurones)	97.50	15.98	16.17
Oujda	PMC (11 neurones)	97.28	16.91	17.70
	NARX _{d+1} (17 neurones)	97.42	16.55	17.29
	NARX _{d+2} (14 neurones)	96.41	17.65	19.52
	NARX _{d+3} (7 neurones)	96.55	16.87	18.49
TanTan	PMC (11 neurones)	94.95	21.41	24.61
	NARX _{d+1} (17 neurones)	95.16	19.44	22.18
	NARX _{d+2} (14 neurones)	94.55	21.38	24.06
	NARX _{d+3} (7 neurones)	94.98	21.88	23.12
Zagora	PMC (11 neurones)	98.68	11.14	11.84
	NARX _{d+1} (17 neurones)	98.87	9.99	10.40
	NARX _{d+2} (14 neurones)	98.48	13.28	12.89
	NARX _{d+3} (7 neurones)	98.73	11.03	11.56

7.3. Exploration du potentiel des méthodes d'ensemble dans l'estimation de GHI journalière.

L'objectif principal dans cette section est d'évaluer le potentiel de trois méthodes d'ensemble basées sur les arbres de régression (Bagging, Boosting et Random Forests) dans l'estimation du GHI quotidien aux cinq sites marocains (Erfoud, Missour, Oujda, TanTan, et Zagora). Aussi, une sixième base de données (Toutes les villes) a été construite en combinant les cinq ensembles de données des cinq stations. Ceci a été pensé dans le but d'élaborer des modèles moyens pouvant représenter l'ensemble de la zone géographique représentée par les cinq sites enerMENA et se trouvent dans un croissant représentant la zone la plus ensoleillée du Maroc (Figure 1.10 du Chapitre 1). Les performances de ces méthodes d'ensemble sont comparées à celles du modèle PMC et des 22 modèles empiriques testés (Tableau 2.1 du Chapitre 2). Une telle comparaison, effectuée pour différents climats, permettrait de tirer des conclusions plus réalistes et générales sur les performances des différents modèles. Tous les modèles testés sont basés, uniquement, sur des entrées constituées par des paramètres astronomiques facilement calculables et des variables météorologiques couramment mesurées dans les stations météorologiques (Bounoua et al., 2021).

7.3.1. Procédure des modèles développés

Les performances de chaque modèle sont évaluées pour chaque station (Erfoud, Missour, Oujda, TanTan et Zagora), puis pour l'ensembles des stations (tous les ensembles de données combinés en un seul). Pour tous les modèles, chaque base de données a été divisée en deux ensembles : l'ensemble d'entraînement et l'ensemble de validation. L'ensemble d'entraînement (70% des données) est utilisé pour ajuster les modèles, tandis que l'ensemble de validation (30% des données) est utilisé pour évaluer leur crédibilité. Le modèle le plus performant pour chaque station a été sélectionné sur la base de trois indicateurs statistiques R^2 , $nRMSE$ et $nMAE$.

Dans cette étude, un modèle PMC est considéré. Pour les trois méthodes d'ensemble (Boosting, Bagging, Random Forests), les phases d'entraînement et de validation sont effectuées en utilisant la meilleure combinaison d'entrées trouvée pour le modèle PMC. La validation de ces modèles développés est effectuée sur l'ensemble de données de validation en exécutant le processus de chaque méthode d'ensemble cinq fois. Les principaux paramètres des méthodes d'ensemble sont fixés comme suit : le nombre d'arbres choisi (nombre d'apprenants); dépend de la méthode (Boosting, Bagging, ou Random Forests) et la base de données utilisées et se

situé principalement entre 40 et 400, le nombre minimum d'observations par feuille d'arbre (MinLeafSize) est fixé à 5 et le taux d'apprentissage est égal à 0.1.

7.3.1.1. Sélection des entrées pertinentes en utilisant le modèle PMC

Les entrées pertinentes sont celles qui ont l'impact le plus significatif sur les fluctuations de la variable de sortie. Dans le domaine de l'estimation du rayonnement solaire, de nombreux chercheurs ont accordé une attention particulière au choix des variables d'entrée pertinentes dans la construction des modèles RNAs. Benghanem et al (Benghanem et al., 2009) ont développé six modèles pour estimer les valeurs du rayonnement solaire en sélectionnant les paramètres d'entrée en fonction de leur corrélation avec le rayonnement solaire global. Ils ont trouvé que le modèle combinant la durée d'ensoleillement et la température comme entrées conduit aux meilleurs résultats par rapport aux autres modèles RNAs et de régression conventionnels testés. (Behrang et al., 2010) ont appliqué les techniques PMC et RBF (Radial Basis Function) pour prédire le rayonnement solaire global horizontal quotidien à la ville de Dezful en Iran, en considérant six combinaisons différentes de l'ensemble des variables {jour de l'année, température moyenne quotidienne de l'air, humidité relative, heures d'ensoleillement, évaporation, vitesse du vent}. Leurs résultats indiquent qu'en utilisant le jour de l'année, la température moyenne quotidienne de l'air, l'humidité relative, les heures d'ensoleillement et la vitesse du vent, le réseau PMC était meilleur que les autres modèles RNAs. Dans le travail de (Yadav et al., 2014), sept variables d'entrée pertinentes sont identifiées à l'aide du logiciel Waikato, un environnement pour l'analyse des connaissances (WEKA : Waikato Environment for Knowledge Analysis) pour prédire le rayonnement solaire en 26 localités indiennes. (Marquez and Coimbra, 2011) ont développé des modèles de prévision pour les irradiations horaires horizontales globales (GHI) et directes normales (DNI) en utilisant les modèles RNAs pour un horizon allant jusqu'à 6 jours. La contribution importante de leur étude est le développement d'un ensemble de critères pour la sélection des entrées pertinentes. Les variables géométriques/temporelles solaires (cosinus de l'angle zénithal solaire et l'angle horaire normalisé) se sont avérées être d'une importance critique, tandis que les variables météorologiques les plus pertinentes comprennent la couverture nuageuse du ciel, la probabilité de précipitation et les températures maximales et minimales. Dans (Yadav et al., 2015), les précisions de prédiction des modèles RNA, RBF et les réseaux de neurones à régression généralisée (GRNN : Generalized Regression Neural Network) ont été comparés. Leurs résultats ont montré que les températures maximale et minimale, l'altitude et la durée

d'ensoleillement sont les paramètres les plus pertinents pour prédire le rayonnement solaire pour 76 emplacements dans le nord-ouest de l'Inde. (Alsina et al., 2016) se sont intéressés à la prédiction de la moyenne mensuelle du rayonnement solaire global quotidien en Italie à l'aide d'un RNA. Ils ont utilisé 17 emplacements pour la phase d'entraînement et 28 emplacements pour la phase de test. Pour chaque emplacement, 13 paramètres d'entrée sont considérés, y compris les coordonnées géographiques et les valeurs mensuelles des paramètres climatologiques les plus fréquemment adoptés. Une méthode automatique est utilisée pour indiquer les entrées les plus pertinentes. La meilleure configuration du réseau RNA comprend sept paramètres, à savoir le rayonnement solaire hors atmosphère (TOA), la durée du jour, le nombre de jours pluvieux et la pluviométrie moyenne, la latitude et l'altitude. (Kheradmanda et al., 2016) ont proposé un modèle RNA intégré pour prédire l'indice de clarté en utilisant des facteurs environnementaux et météorologiques mensuels collectés pendant 30 ans en 19 villes iraniennes. La sélection des variables d'entrée a été faite en considérant les plus corrélées entre elles. Ils ont constaté que l'erreur quadratique moyenne minimale est obtenue en utilisant la température, l'humidité relative, l'ensoleillement et les précipitations comme entrées. (Mohammadi et al., 2016) ont utilisé le processus ANFIS pour sélectionner les variables d'entrée parmi plusieurs combinaisons testées en considérant les données de trois villes situées au centre et au sud d'Iran. Selon cette étude, il a été conclu que la sélection appropriée des paramètres d'entrée joue un rôle très important dans la précision de prédiction de l'irradiation solaire globale. Cependant, les effets climatiques et les microclimats de chaque région rendent impossible l'adoption d'un modèle universel avec une combinaison idéale de différents paramètres d'entrée.

L'objectif principal de notre travail est de construire un modèle PMC capable d'estimer le GHI quotidien à partir de certains paramètres astronomiques et variables météorologiques couramment mesurées (Bounoua et al., 2021). Le tableau 3.9 présente toutes les variables météorologiques et astronomiques considérées. Différentes combinaisons des variables d'entrée ont été testées à l'aide d'un modèle PMC à architecture fixe (cinq neurones cachés). Pour chaque cas, la précision est évaluée sur un ensemble de cinq exécutions. Les résultats ont révélé que les entrées les plus pertinentes sont : la température moyenne quotidienne de l'air, l'humidité relative moyenne quotidienne, la direction et la vitesse du vent, le déficit de pression de vapeur d'air quotidien (équation 2-22 du chapitre 2), l'irradiation solaire horizontale quotidienne hors

atmosphère et le gradient de la température quotidien (T , RH , W_d , W_s , D , H_0 , ΔT). Notons que, dans les cinq stations considérées, les entrées pertinentes obtenues sont les mêmes.

7.3.1.2. Optimisation du nombre de neurones cachés du modèle PMC

Une fois les entrées pertinentes sélectionnées, il est crucial d'optimiser le nombre de neurones dans la couche cachée. Pour cela, une approche constructive est utilisée qui consiste en une construction itérative de la couche cachée. Le nombre de neurones dans la couche cachée est augmenté de façon itérative de 1 à 20. Les performances de chaque architecture ont été évaluées sur cinq exécutions. Les résultats montrent que, pour toutes les stations considérées, le modèle PMC est plus performant lorsqu'une couche cachée de 10 neurones est utilisée (Bounoua et al., 2021).

Tableau 3. 9. Données météorologiques et astronomiques considérées (Bounoua et al., 2021).

Nombre	Abréviation	Paramètre d'entrée
1	DOY	Jour de l'année
2	T	Température moyenne quotidienne de l'air ($^{\circ}C$)
3	RH	Humidité relative moyenne quotidienne (%)
4	W_s	Vitesse moyenne quotidienne du vent (m/s)
5	W_d	Direction moyenne quotidienne du vent
6	D	Déficit de pression de vapeur d'air quotidien (kPa)
7	H_0	Irradiation solaire horizontale quotidienne hors atmosphère (Wh/m^2)
8	T_{max}	Température maximale quotidienne de l'air ($^{\circ}C$)
9	T_{min}	Température minimale quotidienne de l'air ($^{\circ}C$)
10	$\Delta T = (T_{max} - T_{min})$	Gradient de température quotidien ($^{\circ}C$)
11	$\sqrt{\Delta T}$	Gradient de température quadratique ($^{\circ}C$)
13	$T_r (T_{min}/T_{max})$	Rapport de température

7.3.2. Performances obtenues par les méthodes d'ensemble et le modèle PMC pour l'estimation journalière de GHI

Le tableau 3.10 représente les résultats de performances d'estimation du GHI journalier en appliquant les trois méthodes d'ensemble et le modèle PMC. Les meilleures performances obtenues sont mentionnées par un fond gris. On peut remarquer que tous les modèles développés donnent de bonnes performances pour les ensembles de données d'entraînement et de validation: les nRMSE sont dans la gamme de fiabilité ($0 < nRMSE < 20\%$). Les erreurs statistiques obtenues diffèrent d'un modèle à l'autre. Les valeurs du coefficient de corrélation R varient entre 87.53 et 96.20% pour les forêts aléatoires (Random Forests), 86.02 et 95.09% pour le Bagging, 83.44 et 94.69% pour le Boosting et 81.73 et 95.14% pour le PMC. Parmi les modèles d'apprentissage automatique, l'algorithme Random Forests donne les meilleures performances sur toutes les stations. Le nRMSE sur les ensembles de données de validation varie entre 7.85 et 15.33% pour le Random Forests, 8.22 et 16.16% pour le Bagging, 8.88 et

Chapitre 3 : Estimation des irradiations solaires journalières en utilisant les modèles d'apprentissage automatique.

18% pour le Boosting et 8.22 et 18.23% pour le PMC. Ces résultats montrent que les modèles d'apprentissage automatique sont nettement plus performants que les modèles empiriques (Bounoua et al., 2021).

Les meilleures performances sont obtenues à Zagora, avec des valeurs du nRMSE égales à 7.85 % (RF), 8.22 % (Bagging), 8.22 % (PMC) et 8.88 % (Boosting) pour les ensembles de données de validation. De même à Erfoud, les modèles PMC et RF ont des précisions d'estimation proches en termes des trois indicateurs statistiques pour l'ensemble de validation, avec des valeurs du nRMSE égales à 8.53% et 9.18%, des valeurs de R égales à 94.89% et 94.35% et des valeurs du nMAE égales à 6.25% et 6.94% respectivement. De plus, tous les modèles développés ont montré de bonnes performances à Missouri et Oujda avec des valeurs du nRMSE égales à 12.30% et 11.23% (RF), 12.41% et 12.68% (Bagging), 13.05% et 13.31% (Boosting) et 12.61% et 13.16% (PMC), respectivement, pour l'ensemble de validation. Les résultats obtenus à Tan-Tan montrent, encore une fois, des performances inférieures à celles trouvées dans les autres stations, avec des valeurs du nRMSE de 15.33% (RF), 16.16% (Bagging), 18% (Boosting) et 18.23% (ANN), pour l'ensemble de validation (Bounoua et al., 2021).

Tableau 3. 10. Performances d'estimation du GHI quotidien des quatre méthodes d'apprentissage automatique pour tous les cas étudiés (ensembles de données d'entraînement et de validation) (Bounoua et al., 2021).

Station	Modèle	Structure	Entraînement			Validation		
			R (%)	nMAE (%)	nRMSE (%)	R (%)	nMAE (%)	nRMSE (%)
Erfoud	PMC	10 neurones	94.70	6.95	9.56	94.35	6.94	9.18
	Boosting	338 apprenants	95.66	6.21	8.58	94.68	6.99	9.41
	Bagging	42 apprenants	97.11	5.04	7.23	94.77	6.64	9.22
	RF	132 apprenants	97.07	5.11	7.36	94.89	6.25	8.53
Missour	PMC	10 neurones	91.96	9.65	13.12	92.51	9.52	12.61
	Boosting	342 apprenants	94.02	8.86	11.50	91.08	9.80	13.05
	Bagging	53 apprenants	95.99	7.23	9.76	92.04	9.65	12.41
	RF	149 apprenants	95.90	7.24	9.80	93.63	9.86	12.30
Oujda	PMC	10 neurones	95.04	9.50	12.86	94.01	9.71	13.16
	Boosting	311 apprenants	94.61	9.91	13.36	94.15	10.62	13.31
	Bagging	43 apprenants	97.46	6.50	9.43	94.46	9.30	12.68
	RF	133 apprenants	97.42	6.54	9.47	96.19	8.84	11.23
Tan-Tan	PMC	10 neurones	82.39	13.59	17.85	81.73	13.63	18.23
	Boosting	345 apprenants	84.06	13.37	17.13	83.44	13.86	18.00
	Bagging	47 apprenants	93.34	9.06	12.08	86.02	12.06	16.16
	RF	143 apprenants	93.32	8.95	12.15	87.53	11.81	15.33
Zagora	PMC	10 neurones	94.97	6.52	9.06	95.14	6.04	8.22
	Boosting	340 apprenants	95.44	6.36	8.57	94.69	6.61	8.88
	Bagging	44 apprenants	97.65	4.57	6.44	95.09	5.88	8.22
	RF	125 apprenants	97.49	4.69	6.61	96.20	5.84	7.85
Toutes les villes	PMC	10 neurones	91.75	9.88	13.83	91.00	9.87	13.38
	Boosting	340 apprenants	90.80	10.74	14.45	90.76	10.86	14.09
	Bagging	44 apprenants	96.67	6.38	9.01	92.56	9.64	12.93
	RF	125 apprenants	96.69	6.49	9.06	92.54	9.29	12.63

Chapitre 3 : Estimation des irradiations solaires journalières en utilisant les modèles d'apprentissage automatique.

Malgré les différences de climat et de positions géographiques entre les stations, les cinq bases de données ont été rassemblées (Erfoud, Missouri, Oujda, TanTan, et Zagora) afin d'évaluer les performances d'estimation du GHI journalier. Ce qui permettra de généraliser nos modèles à chaque site du croissant géographique reliant toutes les stations d'Oujda à Tan-Tan (Voir figure 1.10 du Chapitre 1). Cette zone géographique comprend une partie des régions les plus ensoleillées du Maroc. Les résultats obtenus montrent une bonne précision, avec des valeurs du nRMSE égales à 12.63% pour le RF, 12.93% pour le Bagging, 13.38% pour le PMC et 14.09% pour le Boosting, pour l'ensemble de validation. En général, les performances obtenues sont de l'ordre de celles obtenues pour les stations d'Oujda et de Missouri, malgré la prise en compte des données de la station de Tan-Tan (résultats les moins performants). Sans les données de Tan-Tan, les valeurs du nRMSE deviennent 11.20% (RF), 11.63% (Bagging), 12.07% (PMC) et 12.43% (Boosting) (Bounoua et al., 2021).

La figure 3.15 représente les diagrammes de dispersion des quatre méthodes d'apprentissage automatique développées pour estimer le GHI quotidien pour chaque ensemble de données de validation.

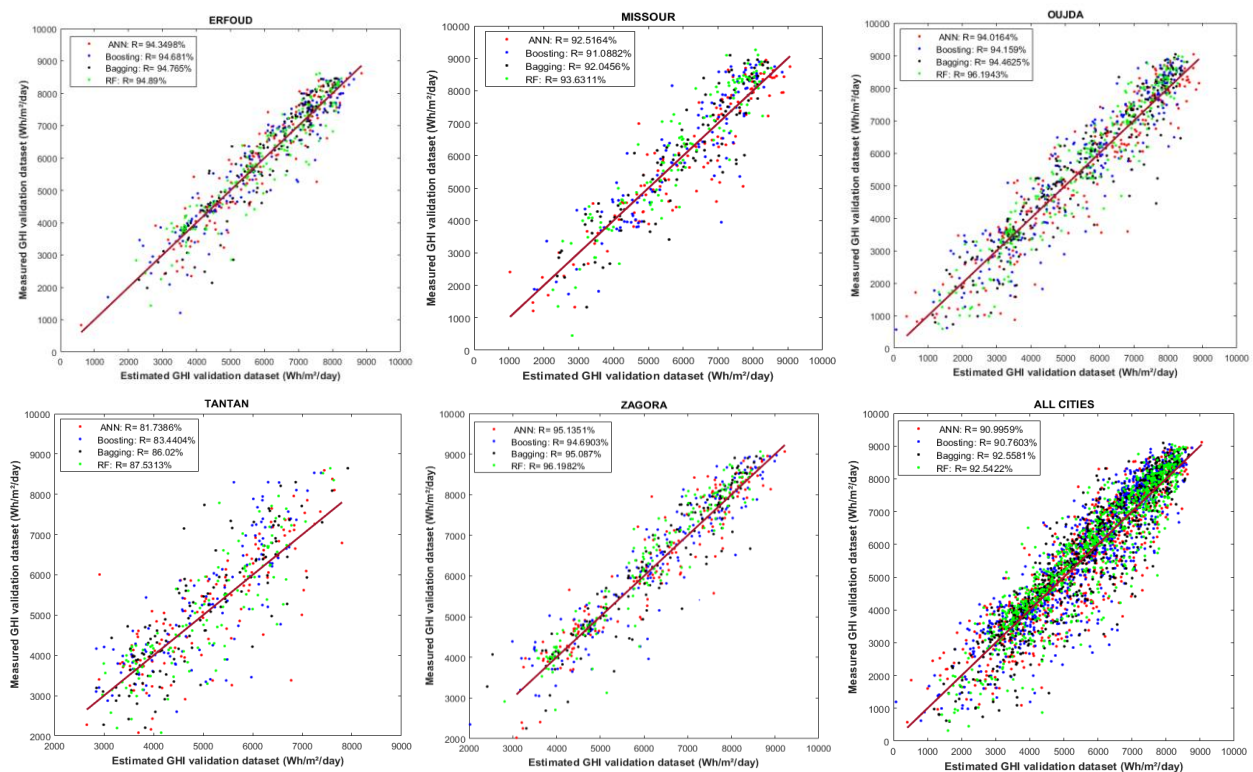


Figure 3. 15. Nuages de points des valeurs quotidiennes mesurées et estimées de GHI en utilisant les quatre modèles d'apprentissage automatique pour les ensembles de données de validation.

7.3.3. Comparaison et analyse des performances

A partir de l'analyse de tous les résultats de validation représentés dans la partie précédente, il est possible de conclure que les méthodes d'ensemble donnent des performances assez similaires à celles obtenues par le modèle PMC basé sur les réseaux de neurones artificiels (RNA). Ceci prouve que les méthodes d'ensemble (Boosting, Bagging et Random Forests) peuvent concurrencer le modèle PMC pour estimer le GHI quotidien. Au contraire, les modèles empiriques ont montré de faibles performances par rapport aux modèles d'apprentissage automatique (Chapitre 2, Section 4.1).

Dans le cas de Tan-Tan, le nRMSE sur l'ensemble de données de validation est de 18.23 % pour le PMC, 18 % pour le Boosting, 16.16 % pour le Bagging, 15.33 % pour le RF et 22.39 % pour le modèle empirique TG1 (Chapitre 2, Section 5.1). La valeur de R est de 81.73 % pour le PMC, 83.84 % pour le Boosting, 86.02 % pour le Bagging, 87.53 % pour le RF et de 72.51 % pour le modèle empirique TG1.

Selon les travaux de (Schüler et al., 2016), le climat de Tan-Tan est caractérisé par des nuages et des brouillards fréquents, et une humidité relative élevée de 83% en moyenne sur toute l'année. La figure 3.16 présente les variations de l'humidité relative maximale observée quotidiennement à Tan-Tan. On peut voir que pendant plusieurs jours, l'humidité relative maximale de l'air atteint 100%. Dans ces conditions, l'humidité de l'air peut se condenser et de petites gouttelettes d'eau apparaissent sur les instruments de mesure. Cela entraînerait des inexactitudes dans les mesures du rayonnement solaire incident, à moins que les instruments ne soient nettoyés (entretien). Cependant, le nombre d'événements de nettoyage évalués dans cette station n'était que de 84 pour la période du 26 juillet 2014 au 1er juin 2015 (Schüler et al., 2016). Ce qui explique les performances relativement faibles des différents à Tan-Tan.

D'un autre côté, il a été constaté à partir des signaux de mesure que les stations RSI (Erfoud et Zagora) sont généralement plus propres que les capteurs thermiques (Oujda, Missour et Tan-Tan) (Schüler et al., 2016). Cela peut justifier les meilleures performances obtenues à Erfoud et Zagora par rapport aux autres stations. D'après nos résultats, le modèle des forêts aléatoires (RF) est plus efficace pour estimer le GHI quotidien. Ce modèle est le plus robuste même dans le cas de données bruitées (cas de Tan-Tan), tandis que le PMC est le plus sensible parmi les modèles d'apprentissage automatique.

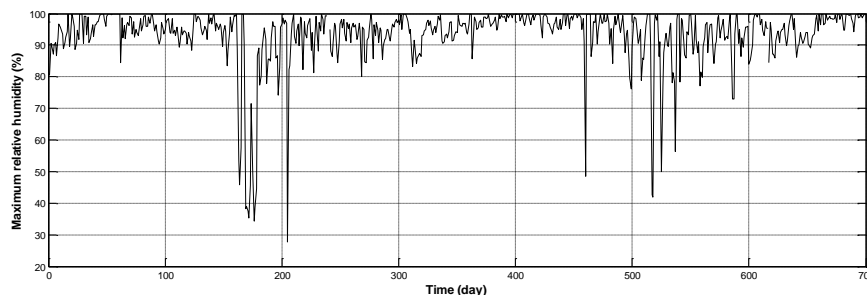


Figure 3. 16. Humidité relative de l'air maximale journalière mesurée à Tan-Tan.

La figure 3.17 montre le graphique radar des valeurs du nRMSE obtenues sur les ensembles de données de validation, pour tous les modèles d'apprentissage automatique testés. Le modèle RF donne les meilleures performances avec une excellente précision pour Zagora et Erfoud ($nRMSE < 10\%$) et une bonne précision pour tous les autres cas (Voir tableau 3.10). La variabilité de la précision peut être attribuée aux différences de climat et à la qualité des mesures (Bounoua et al., 2021).

Notons que les modèles Boosting donnent des valeurs de nRMSE légèrement inférieures à celles du modèle PMC (Missour (Boosting : 13.05%, PMC : 12.61%), Oujda (Boosting : 13.31%, PMC : 13.16%) et Tan-Tan (Boosting : 18%, PMC : 18.23%)), pour l'ensemble de validation. Alors que les modèles Bagging montrent des valeurs de nRMSE légèrement meilleures que le modèle PMC (Missour (Bagging : 12.41%, PMC : 12.61%), Oujda (Bagging : 12.68%, PMC : 13.16%), Tan-Tan (Bagging : 16.16%, PMC: 18.23%)), et pour toutes les villes (Bagging : 12.93%, PMC : 13.38%)), pour l'ensemble de validation. Cependant, le Bagging et le PMC ont les mêmes valeurs de nRMSE pour la station de Zagora (Bagging : 8.22%, PMC : 8.22%) pour l'ensemble de validation. Pour Erfoud, les valeurs de nRMSE obtenues par le modèle PMC sont meilleures que celles des modèles Boosting et Bagging (PMC : 9.18%, Bagging : 9.22%, et Boosting : 9.41%) (Bounoua et al., 2021).

Le modèle RF donne des nRMSE meilleures de 2.46% à 15.91% par rapport au PMC, de 5.75% à 15.63% par rapport au Boosting et de 0.89% à 11.43% par rapport au Bagging. D'après nos résultats, on peut dire que, pour l'estimation des irradiances solaires globales journalières, le modèle RF donne les meilleures performances, suivi du modèle Bagging, du modèle PMC et du modèle Boosting.

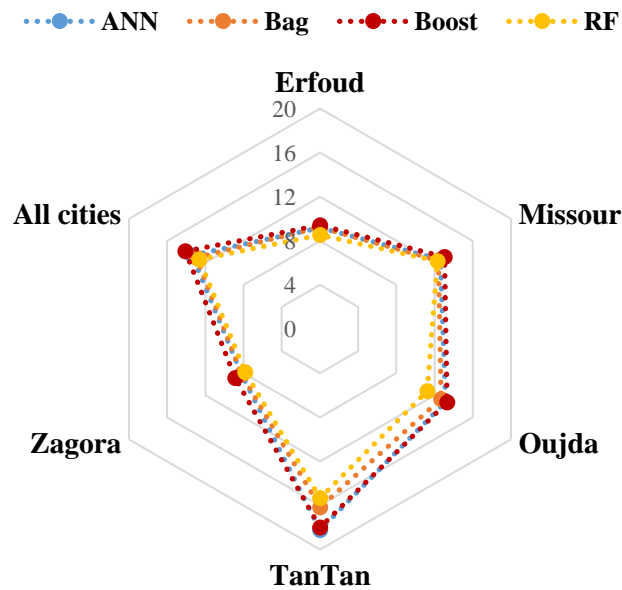


Figure 3. 17. Graphique Radar des valeurs de nRMSE obtenues pour les estimations journalières de GHI pour les ensembles de données de validation et pour tous les modèles testés.

7.4. Performances des méthodes d'ensemble pour l'estimation journalière de GHI à la ville de Fès

Le tableau 3.11 représente les résultats d'estimation de GHI journalière à la ville de Fès, en utilisant la meilleure combinaison d'entrées sélectionnées précédemment (Section 6.3.1.1). D'après ces résultats, on constate que la meilleure performance pour l'ensemble de validation est obtenue encore une fois par le modèle RF (nRMSE = 12.5%), suivi du modèle PMC (nRMSE=13.27%), du modèle Bagging (nRMSE=13.87%) et du modèle Boosting (nRMSE=14.68%).

Tableau 3. 11. Performances des méthodes d'apprentissage automatique dans l'estimation du GHI quotidienne pour la ville de Fès (ensembles de données d'entraînement et de validation).

Station	Modèle	Structure	Entrainement			Validation		
			R (%)	nMAE (%)	nRMSE (%)	R (%)	nMAE (%)	nRMSE (%)
Fès	PMC	10 neurones	94.17	9.85	13.85	94.66	9.66	13.27
	Boosting	326 apprenants	94.33	10.07	13.65	93.41	10.71	14.68
	Bagging	54 apprenants	97.66	6.21	8.97	94.14	9.63	13.87
	RF	143 apprenants	97.37	6.62	9.55	95.29	9.27	12.65

La figure 3.18 montre les nuages de points des valeurs quotidiennes mesurées et estimées de GHI à la ville de Fès, en utilisant les quatre modèles d'apprentissage automatique pour les ensembles de données d'entraînement et de validation.

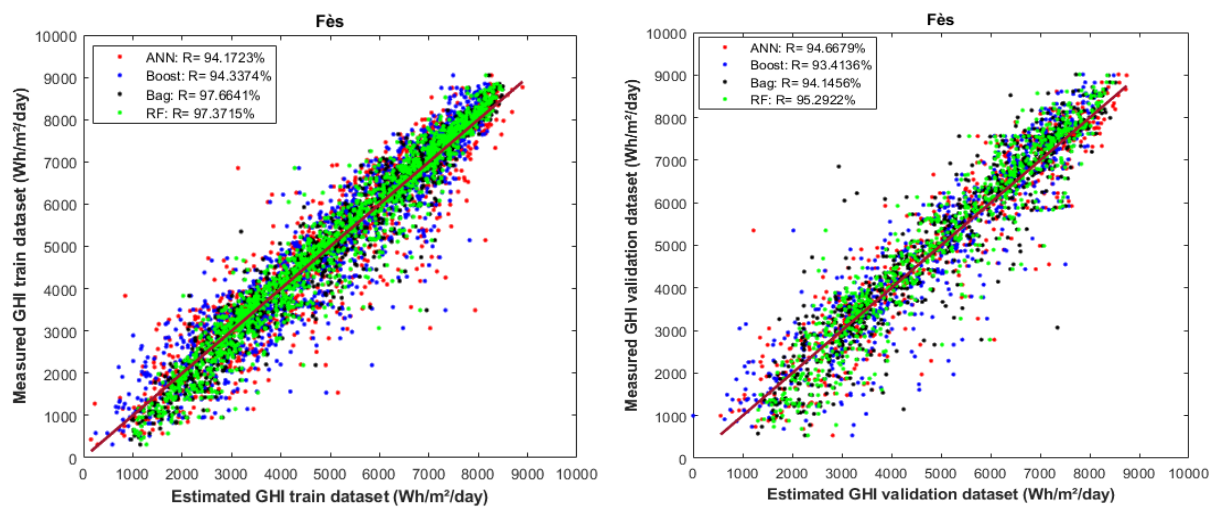


Figure 3. 18. Nuages de points des valeurs quotidiennes mesurées et estimées de GHI à la ville de Fès, en utilisant les quatre modèles d'apprentissage automatique pour les ensembles de données d'entraînement et de validation.

7.5. Catégorisation des modèles d'apprentissage automatique pour l'estimation journalière des composantes solaires à la ville de Zagora

Dans cette étude (Bounoua and Mechaqrane, 2021a), plusieurs méthodes d'apprentissage automatique, dont le PMC, la régression par machine à vecteur de support (SVM) et le RF, sont examinées pour estimer les trois composantes de l'irradiation solaire journalières au sol: l'irradiations horizontale globale (GHI), l'irradiations horizontale diffuse (DHI) et l'irradiation normale directe (DNI) à la ville de Zagora. Ces modèles sont regroupés en quatre catégories d'entrées : les modèles basés sur la température (GTBMs), les modèles basés sur l'humidité relative (RHBMs) pour GHI, les modèles basés sur les paramètres hybrides (HBM) pour GHI, DHI et DNI et les modèles basés sur l'indice de clarté (CIBMs) pour DHI et DNI.

L'objectif principal de cette partie est de définir la méthode d'apprentissage automatique la plus précise pour estimer les trois composantes solaires en fonction des quatre catégories d'entrées considérées. Tous ces modèles d'apprentissage automatique catégorisés ont été développés et optimisés. De plus, le degré d'amélioration d'estimation de ces modèles est évalué en les comparant aux modèles empiriques cités dans la littérature ou nouvellement suggérés (Chapitre 2, Section 4) (Bounoua and Mechaqrane, 2021a).

7.5.1. Processus de catégorisation et évaluation des modèles

Dans cette section, chaque modèle développé est présenté par un nom spécial, qui fait référence à son type, à la composante solaire à estimer, au groupe d'entrées correspondant, et à un numéro pour indiquer son ordre dans chaque catégorie. La première lettre du nom du modèle commence par "P", "S", ou "R" pour les modèles PMC, SVM, ou RF, respectivement. Suivie de "G", "D", ou "DN", pour GHI, DHI, ou DNI, respectivement. Puis de "T", "RH", "H", ou "C", pour les modèles basés sur la température (TBM), l'humidité (RHBM), paramètres hybrides (HBM), ou l'indice de clarté (CIBM), respectivement. Ensuite, un numéro pour désigner s'il s'agit d'un 1er, 2e, 3e, 4e ou 5e modèle. Le tableau 3.12 présente les détails concernant le groupe, l'abréviation, les entrées et les sorties de chaque modèle d'apprentissage automatique.

Tableau 3. 12. Détails des modèles d'apprentissage automatique développés.

Groupe	Abréviation	Modèle			Entrées	Sortie
Modèles basés sur la température	TBM	PGT1	SGT1	RGT1	H_0, T	GHI
		PGT2	SGT2	RGT2	H_0, T_{max}, T_{min}	GHI
		PGT3	SGT3	RGT3	$H_0, \Delta T$	GHI
		PGT4	SGT4	RGT4	H_0, T_r	GHI
Modèles basés sur l'humidité relative	RHBM	PGRH1	SGRH1	RGRH1	H_0, RH	GHI
		PGRH2	SGRH2	RGRH2	H_0, RH_{max}, RH_{min}	GHI
		PGH1	SGH1	RGH1	H_0, T, RH	GHI
		PGH2	SGH2	RGH2	H_0, T_r, RH, T	GHI
		PGH3	SGH3	RGH3	$H_0, T_r, RH, \Delta T$	GHI
		PGH4	SGH4	RGH4	$H_0, T, RH, \Delta T, W_s, P_r$	GHI
Modèles basés sur des paramètres hybrides	HBM	PGH5	SGH5	RGH5	$H_0, \Delta T, T_{max}, T_{min}, RH$	GHI
		PDH1	SDH1	RDH1	GHI, K_t, RH	DHI
		PDH2	SDH2	RDH2	GHI, K_t, δ	DHI
		PDH3	SDH3	RDH3	GHI, K_t, H_0, DOY, D	DHI
		PDNH1	SDNH1	RDNH1	H_{0N}, K_t, DOY, T	DNI
		PDC1	SDC1	RDC1	H_0, K_t	DHI
Modèles basés sur l'indice de clarté	CIBM	PDC2	SDC2	RDC2	GHI, K_t	DHI
		PDNC1	SDNC1	RDNC1	H_{0N}, K_t	DNI

Pour tous les modèles développés, 850 observations sont soumises à une validation croisée en utilisant la méthode de partition aléatoire non stratifiée « Holdout ». Cette partition divise les observations en un ensemble d'entraînement et de validation pour garantir la performance de chaque modèle et généraliser son potentiel d'estimation. Pour cette procédure de validation croisée, 70% des observations sont utilisées pour la phase de d'entraînement et les 30% restants pour la phase de validation. De plus, cinq indicateurs statistiques sont utilisés pour évaluer la précision d'estimation de chaque modèle développé, le coefficient de corrélation (R), l'erreur absolue moyenne (MAE), l'erreur quadratique moyenne (RMSE), l'erreur de biais moyenne

(MBE) et l'erreur quadratique moyenne normalisée (nRMSE) (Annexe D). En se basant principalement sur les valeurs de nRMSE pour sélectionner les modèles les plus performants.

7.5.2. Performances des modèles d'apprentissage automatique catégorisés

Les résultats des modèles d'estimation du GHI basés sur la température (GTBM) sont présentés dans le tableau 3.13 tout en indiquant les paramètres ou la structure de chaque modèle utilisé (nombre de neurones cachés pour le modèle PMC, type de la fonction noyau pour le modèle SVM et le nombre d'apprenants pour le modèle RF). Les résultats obtenus montrent que les modèles d'apprentissage automatique (PMC, SVR et RF) surpassent le meilleur modèle empirique MT3 (nRMSE = 10.19%) (Chapitre 2, Section 4.1) dans l'estimation du GHI à partir des données de la température. La meilleure performance est obtenue par le modèle SGT3 qui est un modèle SVM utilisant comme entrées l'irradiation solaire journalière d'un plan horizontal hors atmosphère H_0 et le gradient de température journalier $\Delta T = T_{max} - T_{min}$. Le nRMSE obtenu sur les données de validation est de 9.09% ce qui représente une amélioration de 10.79% par rapport au modèle empirique MT3. D'autre part, les modèles PMC et RF (PGT3 et RGT3) donnent aussi des performances voisines de celles du modèle SVM avec des nRMSEs de l'ordre de 9.46% et 9.56% respectivement, c'est-à-dire une amélioration par rapport au modèle empirique MT3 de l'ordre de 7.26% et 6.28%, respectivement. La figure 3.19 montre le tracé des séries temporelles du GHI quotidien mesuré et estimé par le modèle SGT3 pour les ensembles d'entraînement et de validation (Bounoua and Mechaqrane, 2021a).

Tableau 3. 13. Indicateurs statistiques des modèles TBMs pour estimer les valeurs journalières de GHI. Les lignes en surbrillance représentent les meilleures performances obtenues pour chaque méthode utilisée. En revanche, les chiffres en gras indiquent le meilleur modèle trouvé dans l'ensemble du groupe.

Modèle	Structure	Entraînement					Validation				
		R	MAE	RMSE	MBE	nRMSE (%)	R	MAE	RMSE	MBE	nRMSE (%)
PGT1	5 neurones	0.8672	0.6005	0.8860	0.0018	14.3927	0.8641	0.6281	0.8466	-0.0273	13.3436
PGT2	18 neurones	0.9421	0.4196	0.5951	-0.0409	9.5827	0.9346	0.4524	0.6405	0.0280	10.4079
PGT3	10 neurones	0.9397	0.4399	0.6153	0.0047	9.9663	0.9356	0.4215	0.5902	0.0200	9.4578
PGT4	17 neurones	0.9293	0.4501	0.6475	-0.0200	10.4880	0.9289	0.4740	0.6674	-0.0154	10.6952
SGT1	RBF kernel	0.8605	0.5221	0.9431	0.2840	15.3379	0.8869	0.5154	0.8381	0.2206	13.3061
SGT2	RBF kernel	0.9456	0.3824	0.5916	0.1103	9.6120	0.9334	0.4176	0.6168	0.0784	9.8151
SGT3	RBF kernel	0.9319	0.4340	0.6518	0.1002	10.6004	0.9443	0.4081	0.5727	0.0251	9.0929
SGT4	RBF kernel	0.9273	0.4370	0.6804	0.1118	11.1327	0.9221	0.4145	0.6482	0.0259	10.1540
RGT1	126 apprenants	0.9193	0.4782	0.7058	0.0046	11.3899	0.8915	0.5440	0.7821	-0.0458	12.6429
RGT2	134 apprenants	0.9543	0.3787	0.5421	0.0051	8.7767	0.9346	0.4778	0.6601	-0.0766	10.5918
RGT3	114 apprenants	0.9606	0.3517	0.4982	0.0037	8.0898	0.9413	0.4531	0.5998	-0.1211	9.5575
RGT4	131 apprenants	0.9518	0.3616	0.5453	0.0055	8.8159	0.9292	0.4622	0.6642	-0.0178	10.6907

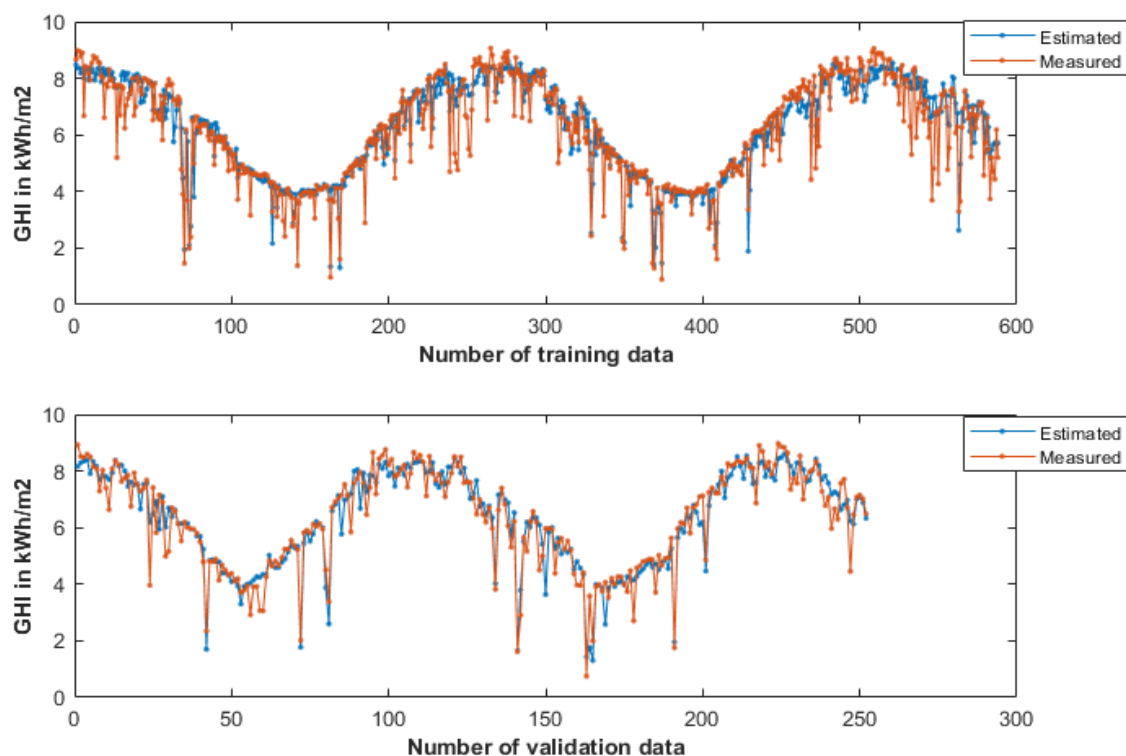


Figure 3. 19. Tracé des séries temporelles du GHI mesuré et estimé quotidiennement par le modèle SGT3, pour les ensembles d'entraînement et de validation.

Pour les modèles basés sur l'humidité relative RHBM (tableau 3.14), le meilleur modèle pour estimer le GHI quotidien est le PMC (PGRH2) utilisant comme entrées H_0 , RH_{\max} et RH_{\min} . Sur l'ensemble de validation, le PGRH2 donne des valeurs de R et du nRMSE égales à 91.13% et 11.41%, respectivement. Il réduit, ainsi, la valeur du nRMSE du meilleur modèle empirique MRH1 (qu'on a proposé (chapitre 2 section 4.1)) (nRMSE = 12.10%) de 5.70%. Les modèles SGRH2 (SVM) et RGRH2 (RF) présentent une précision presque égale à celle du modèle PGRH2 (PMC), avec des valeurs respectives du nRMSE égales à 11.88 % et 11.71% et représentent une légère amélioration par rapport au modèle empirique MRH1 (1.81% pour le modèle SGRH2 et 3.22% pour le modèle RGRH2) (Bounoua and Mechaqrane, 2021a).

Tableau 3. 14. Indicateurs statistiques des modèles RHBM pour estimer les valeurs journalières de GHI. Les lignes en surbrillance représentent les meilleures performances obtenues pour chaque méthode utilisée. En revanche, les chiffres en gras indiquent le meilleur modèle trouvé dans l'ensemble du groupe.

Modèle	Structure	Entraînement					Validation				
		R	MAE	RMSE	MBE	nRMSE (%)	R	MAE	RMSE	MBE	nRMSE (%)
PGRH1	6 neurones	0.8873	0.5923	0.8109	0.0135	13.1075	0.9023	0.5461	0.7643	-0.0396	12.3072
PGRH2	18 neurones	0.9106	0.5288	0.7427	0.0162	11.9080	0.9113	0.5075	0.6953	-0.0114	11.4124
SGRH1	RBF kernel	0.8953	0.5364	0.8240	0.2010	13.4104	0.8867	0.5353	0.8083	0.2120	12.8144
SGRH2	RBF kernel	0.9112	0.4790	0.7496	0.1754	12.1109	0.9110	0.5037	0.7373	0.1210	11.8845
RGRH1	138 apprenants	0.9344	0.4448	0.6331	0.0015	10.2102	0.8957	0.5582	0.7868	0.0511	12.7362
RGRH2	125 apprenants	0.9444	0.4224	0.5969	-0.0019	9.6973	0.9054	0.5188	0.7361	-0.0208	11.7174

Chapitre 3 : Estimation des irradiations solaires journalières en utilisant les modèles d'apprentissage automatique.

Le tableau 3.15 montre les performances obtenues par les modèles basés sur les paramètres hybrides (HBM). D'après ces résultats, le meilleur modèle pour estimer le GHI quotidien est le modèle RGH4 (RF) ($H_0, T, RH, \Delta T, W_s, P_r$), avec des valeurs de R et nRMSE égales à 95.44% et 8.36% respectivement, pour l'ensemble de validation. Ce modèle réduit le nRMSE du modèle empirique MH3 (nRMSE= 10.30%) de 18.83% (Chapitre 2, Section 4.1). Aussi, les modèles PGH4 (PMC) et SGH4 (SVM) ont minimisé le nRMSE de validation du modèle empirique MH3 de 11.74% et 11.06%, respectivement (Chapitre 2, Section 4.1).

Pour estimer le DHI quotidien en utilisant les paramètres hybrides (Tableau 3.16), les modèles RF sont plus appropriés que les modèles empiriques (Chapitre 2, Section 4.2), le SVM et le PMC, avec la plus faible valeur du nRMSE égale à 18.55% et la plus grande valeur de R égale à 95.75% obtenus par le modèle RDH3. En comparant cette valeur (nRMSE =18.55%) avec celle du meilleur nouveau modèle empirique DH3 (qu'on a proposé) (Chapitre 2, Section 4.2), on constate une diminution relative de 25.29%. Les modèles PMC et SVM donnent une légère amélioration par rapport au modèle empirique DH3. L'erreur d'estimation du modèle DH3 (nRMSE = 24.83%) est réduite de 21.58% et 14.74% en utilisant PDH3 (PMC), SDH3 (SVM), respectivement.

Tableau 3. 15. Indicateurs statistiques des modèles HBMs pour estimer les valeurs journalières de GHI. Les lignes en surbrillance représentent les meilleures performances obtenues pour chaque méthode utilisée. En revanche, les chiffres en gras indiquent le meilleur modèle trouvé dans l'ensemble du groupe.

Modèle	Structure	Entraînement					Validation				
		R	MAE	RMSE	MBE	nRMSE (%)	R	MAE	RMSE	MBE	nRMSE (%)
PGH1	9 neurones	0.9104	0.5027	0.7399	0.0153	12.0920	0.9045	0.5145	0.7197	-0.0184	11.3007
PGH2	15 neurones	0.9501	0.4055	0.5581	-0.0444	9.0299	0.9368	0.4166	0.6113	-0.0622	9.8211
PGH3	6 neurones	0.9489	0.3941	0.5576	-0.0022	9.0454	0.9360	0.4237	0.6160	-0.0219	9.8360
PGH4	10 neurones	0.9418	0.4184	0.6010	0.0194	9.7759	0.9418	0.4177	0.5733	-0.0382	9.0991
PGH5	26 neurones	0.9423	0.4165	0.5879	-0.0335	9.5222	0.9449	0.4490	0.5957	-0.0843	9.5464
SGH1	RBF kernel	0.9104	0.4441	0.7646	0.1919	12.2846	0.9123	0.4556	0.7092	0.1093	11.5822
SGH2	RBF kernel	0.9426	0.3647	0.6018	0.1176	9.7523	0.9378	0.3905	0.6190	0.0916	9.9066
SGH3	RBF kernel	0.9475	0.3581	0.5856	0.1135	9.4654	0.9357	0.3862	0.6052	0.1000	9.7468
SGH4	RBF kernel	0.9478	0.3474	0.5643	0.1057	9.2075	0.9476	0.3807	0.5818	-0.0059	9.1682
SGH5	RBF apprenants	0.9507	0.3374	0.5554	0.1031	8.9758	0.9402	0.3831	0.6064	0.0698	9.7703
RGH1	123 l apprenants	0.9424	0.4279	0.6152	0.0054	10.0075	0.9202	0.4923	0.6568	-0.0997	10.4209
RGH2	131 apprenants	0.9679	0.3172	0.4530	0.0024	7.3597	0.9500	0.4088	0.5627	-0.0960	8.9538
RGH3	127 apprenants	0.9718	0.3049	0.4298	0.0027	7.0080	0.9468	0.4140	0.5607	-0.1210	8.8504
RGH4	130 apprenants	0.9747	0.2950	0.4132	0.0033	6.7206	0.9544	0.3987	0.5265	-0.1131	8.3600
RGH5	118 apprenants	0.9688	0.3242	0.4480	0.0038	7.2649	0.9491	0.4079	0.5725	-0.0666	9.1496

Pour l'estimation journalière du DNI (Tableau 3.17), le modèle RDNH1 (RF) a donné la plus faible valeur du nRMSE et la plus grande valeur de R (10.26% et 96.26%, respectivement), pour l'ensemble de validation. Cette performance est suivie par celle du modèle PDNH1(PMC) (nRMSE=10.60%) et celle du modèle SDNH1 (SVM) (nRMSE=12.56%). La performance du

Chapitre 3 : Estimation des irradiations solaires journalières en utilisant les modèles d'apprentissage automatique.

nouveau modèle empirique DNH1 proposé (Chapitre 2, Section 4.2) a été améliorée de 23.43 %, 20.89 % et 6.26 % en utilisant respectivement RF (RDNH1), PMC (MDNH1) et SVM (SDNH1) (Bounoua and Mechaqrane, 2021a).

Tableau 3. 16. Indicateurs statistiques des modèles HBM pour estimer les valeurs journalières de DHI. Les lignes en surbrillance représentent les meilleures performances obtenues pour chaque méthode utilisée. En revanche, les chiffres en gras indiquent le meilleur modèle trouvé dans l'ensemble du groupe.

Modèle	Structure	Entraînement					Validation				
		R	MAE	RMSE	MBE	nRMSE (%)	R	MAE	RMSE	MBE	nRMSE (%)
PDH1	18 neurones	0.9340	0.0560	0.0729	-0.0004	22.8688	0.9372	0.0543	0.0689	0.0031	20.8056
PDH2	5 neurones	0.9143	0.0633	0.0804	-0.0002	24.8384	0.9275	0.0629	0.0787	0.0047	24.6426
PDH3	9 neurones	0.9412	0.0502	0.0664	0.0003	20.7625	0.9569	0.0486	0.0648	0.0016	19.4736
SDH1	RBF kernel	0.9330	0.0624	0.0757	0.0156	24.1625	0.9356	0.0629	0.0772	0.0006	21.5252
SDH2	RBF kernel	0.9153	0.0687	0.0826	0.0111	26.0457	0.9234	0.0676	0.0831	0.0107	24.2214
SDH3	RBF kernel	0.9448	0.0630	0.0742	0.0143	22.7612	0.9386	0.0530	0.0654	0.0134	21.1751
RDH1	142 apprenants	0.9605	0.0437	0.0574	-0.0001	17.9453	0.9456	0.0514	0.0643	-0.0028	19.3155
RDH2	138 apprenants	0.9542	0.0472	0.0613	0.0001	19.5044	0.9248	0.0629	0.0788	-0.0192	22.2255
RDH3	132 apprenants	0.9710	0.0367	0.0500	0.0000	15.2561	0.9575	0.0422	0.0559	0.0039	18.5504

Tableau 3. 17. Indicateurs statistiques des modèles HBM pour estimer les valeurs journalières de DNI. Les chiffres en gras indiquent le meilleur modèle trouvé dans l'ensemble du groupe.

Modèle	Structure	Entraînement					Validation				
		R	MAE	RMSE	MBE	nRMSE (%)	R	MAE	RMSE	MBE	nRMSE (%)
PDNH1	13 neurones	0.9668	0.0338	0.0424	0.0011	10.3611	0.9552	0.0359	0.0447	0.0028	10.6087
SDNH1	RBF kernel	0.9529	0.0508	0.0594	-0.0038	14.4675	0.9534	0.0442	0.0523	-0.0034	12.5618
RDNH1	127 apprenants	0.9781	0.0296	0.0374	0.0001	9.1652	0.9626	0.0351	0.0439	-0.0024	10.2649

Les séries temporelles des trois composantes solaires (GHI, DHI, et DNI) mesurées et estimées correspondant aux meilleurs modèles basés sur les paramètres hybrides (RGH4, RDH3 et RDNH1) pour les 120 jours de l'ensemble de validation sont présentées dans la figure 3.20.

Le tableau 3.18 montre la précision des modèles basés sur l'indice de clarté pour estimer le DHI et le DNI quotidiens. Pour le DHI, il est clair que le modèle RDC2 (RF) est plus précis que tous les modèles empiriques (Chapitre 2, Section 4.2), SVM et PMC utilisés dans cette catégorie d'entrées. Le modèle RDC2 donne des valeurs de R et du nRMSE égales à 93.61% et 20.44%, respectivement, pour l'ensemble de validation. Ce modèle (RDC2) a réduit le nRMSE de validation du meilleur modèle empirique basé sur l'indice de clarté (DC2 (nRMSE=27.78%)) de 26.42%. Le modèle SDC2 (SVM) présente une performance proche du modèle RDC2 en donnant un nRMSE de validation égale à 20.99% et en minimisant le nRMSE de validation du modèle empirique DC2 de 24.44%. D'autre part, le modèle PDC2 (PMC) améliore le nRMSE de validation du modèle empirique DC2 de 22.24%.

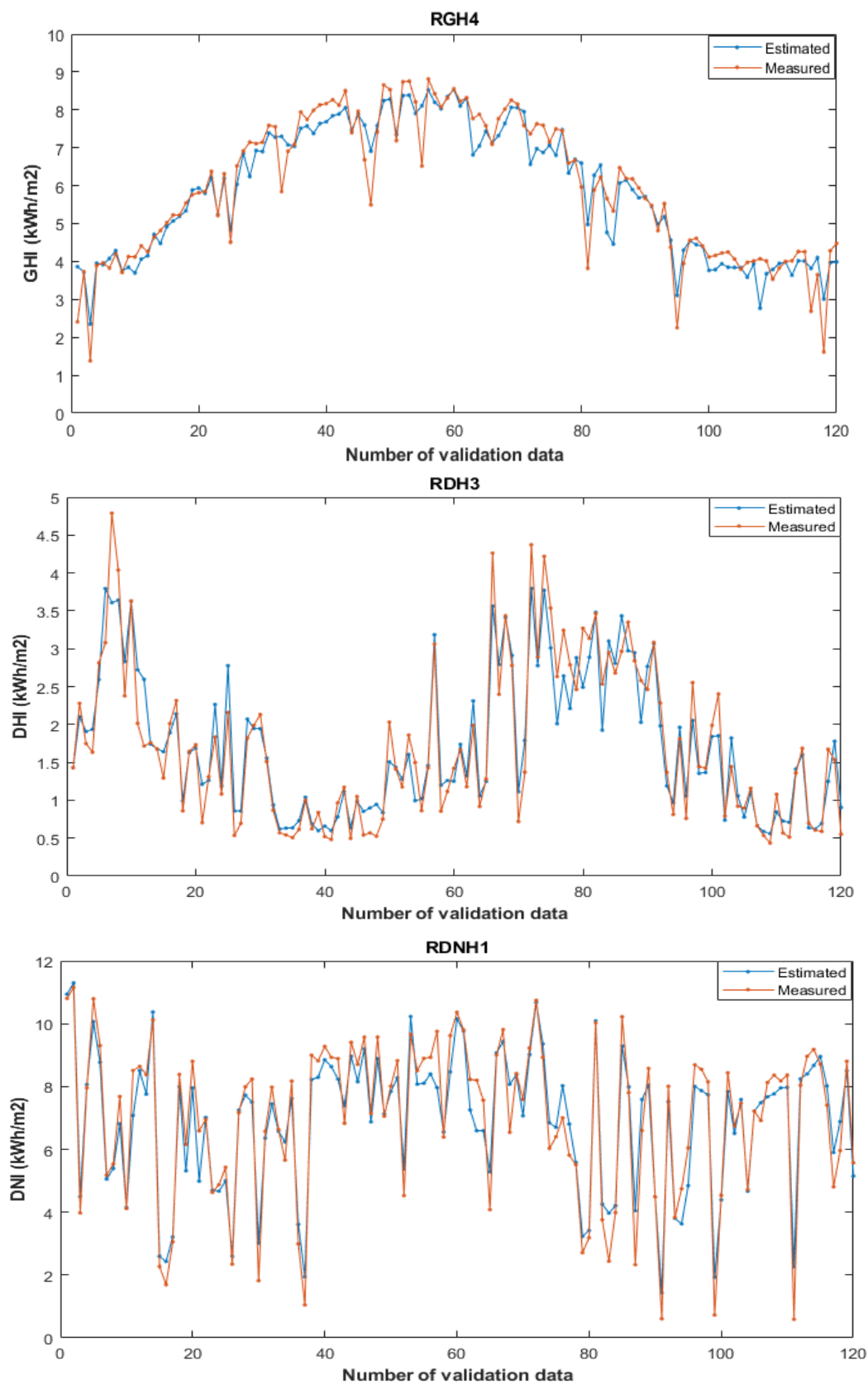


Figure 3. 20. Séries temporelles des trois composantes solaires (GHI, DHI et DNI) mesurées et estimées à partir des meilleurs modèles basés sur les paramètres hybrides (RGH4, RDH3 et RDNH1) pour les 120 jours de l'ensemble de validation.

Chapitre 3 : Estimation des irradiations solaires journalières en utilisant les modèles d'apprentissage automatique.

Pour estimer le DNI quotidien à l'aide de l'indice de clarté (Tableau 3.19), il a été constaté que le modèle PDNC1(PMC) (nRMSE= 12.43%) surpasse le modèle empirique DNC1 (nRMSE= 14.13%) en diminuant son nRMSE de validation de 12.03% (Chapitre 2 section 4.2). La performance du modèle PMC (PDNC1) est suivie par celle du modèle RDNC1 (RF) (nRMSE= 12.94%) et celle du modèle SDNC1 (SVM) (nRMSE= 13.61%), en minimisant le nRMSE de validation du modèle empirique DNC1 par 8.42% et 3.68%, respectivement.

La figure 3.21 présente les séries temporelles de DHI et DNI mesurés et estimés correspondantes aux meilleurs modèles basés sur l'indice de clarté (RDC2 et PDNC1) pour les 120 jours de l'ensemble de validation.

Tableau 3. 18. Indicateurs statistiques des modèles CIBMs pour estimer les valeurs journalières de DHI. Les lignes en surbrillance représentent les meilleures performances obtenues pour chaque méthode utilisée. En revanche, les chiffres en gras indiquent le meilleur modèle trouvé dans l'ensemble du groupe.

Modèle	Structure	Entraînement					Validation				
		R	MAE	RMSE	MBE	nRMSE (%)	R	MAE	RMSE	MBE	nRMSE (%)
PDC1	13 neurones	0.9078	0.3228	0.4370	-0.0048	23.2061	0.9148	0.3328	0.4283	0.0196	23.4488
PDC2	12 neurones	0.9095	0.3263	0.4421	-0.0157	23.4081	0.9166	0.3055	0.3896	0.0170	21.6000
SDC1	RBF kernel	0.9098	0.3115	0.4353	-0.0225	23.2322	0.9223	0.3160	0.4058	-0.0448	21.7810
SDC2	RBF kernel	0.9082	0.3162	0.4372	-0.0220	23.2522	0.9326	0.2936	0.3859	-0.0559	20.9964
RDC1	122 apprenants	0.9479	0.2513	0.3387	-0.0017	18.0363	0.9270	0.2990	0.4015	-0.0254	21.7295
RDC2	131 apprenants	0.9508	0.2408	0.3246	0.0001	17.5889	0.9361	0.2960	0.4041	-0.0691	20.4488

Tableau 3. 19. Indicateurs statistiques des modèles CIBMs pour estimer les valeurs journalières de DNI. Les lignes en surbrillance représentent les meilleures performances obtenues pour chaque méthode utilisée. En revanche, les chiffres en gras indiquent le meilleur modèle trouvé dans l'ensemble du groupe.

Modèle	Structure	Entraînement					Validation				
		R	MAE	RMSE	MBE	nRMSE (%)	R	MAE	RMSE	MBE	nRMSE (%)
PDNC1	24 neurones	0.9380	0.0445	0.0565	0.0003	13.9323	0.9412	0.0426	0.0542	0.0002	12.4306
SDNC1	RBF kernel	0.9352	0.0531	0.0636	-0.0003	15.5141	0.9292	0.0476	0.0570	-0.0042	13.6126
RDNC1	125 apprenants	0.9580	0.0367	0.0472	0.0001	11.5591	0.9360	0.0433	0.0549	0.0011	12.9439

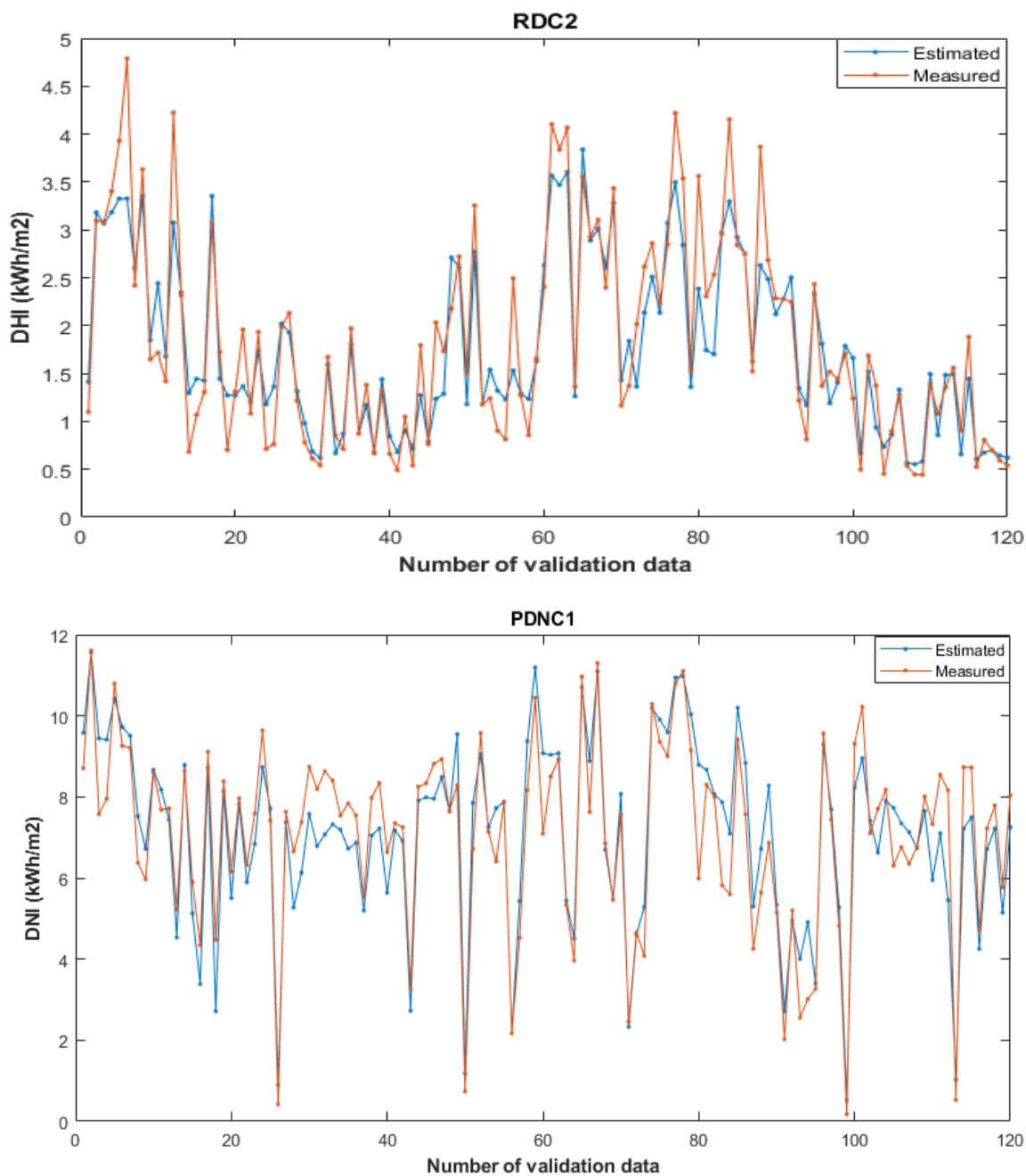


Figure 3. 21. Séries temporelles de DHI et DNI mesurées et estimées à partir des meilleurs modèles basés sur l'indice de clarté (RDC2 et PDNC1) pour les 120 jours de l'ensemble de validation.

7.5.3. Etude comparative des modèles catégorisés

Selon les résultats de la section précédente, il est clair que le choix du meilleur modèle pour estimer les GHI, DHI, et DNI quotidiens dépend de la catégorie de variables d'entrées utilisée. D'après nos résultats, le RF est le modèle plus approprié pour estimer le GHI, le DHI et le DNI quotidiens en utilisant les variables d'entrées basées sur les paramètres hybrides (HBMs), ou le

DHI en utilisant les variables d'entrées basées sur l'indice de clarté (CIBMs). Le PMC est le modèle optimal pour estimer le GHI en utilisant les variables d'entrées basées sur l'humidité relative (RHBMs) ou le DNI en utilisant les variables d'entrées basées sur l'indice de clarté (CIBMs). Le modèle SVM s'est avéré, quant à lui, plus approprié pour l'estimation du GHI quotidien en employant les variables d'entrées basées sur la température (GTBMs).

Aussi, il a été constaté que les modèles empiriques sont moins efficaces que les méthodes d'apprentissage automatique concernant toutes les catégories de variables d'entrées considérées. Un graphique montrant la meilleure méthode d'apprentissage automatique à utiliser pour estimer chaque composante solaire en fonction de chaque catégorie d'entrées est illustré à la figure 3.22.

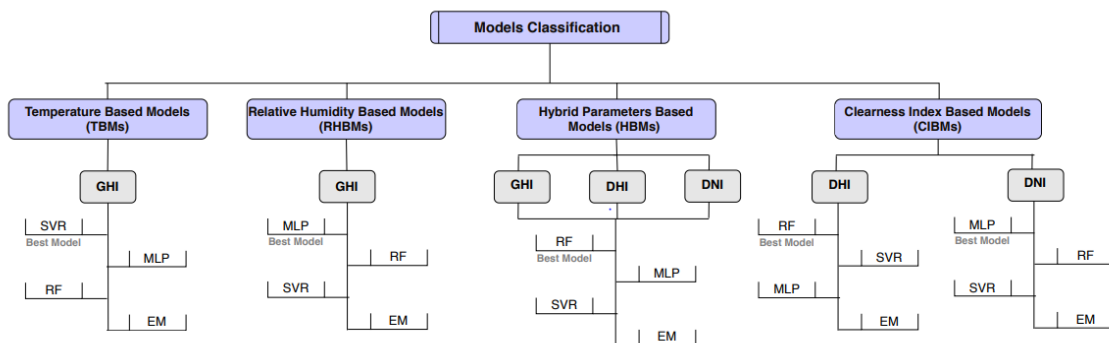


Figure 3. 22. La meilleure méthode d'apprentissage automatique par catégories d'entrées pour estimer le GHI, le DHI et le DNI quotidiens.

De plus, l'utilisation de modèles basés sur des paramètres hybrides (HBMs) avec une variété de données météorologiques donne les meilleures performances (plus faibles valeurs du nRMSE) pour l'estimation journalière de GHI, DHI, et DNI. Pour le GHI, le modèle basé sur les paramètres hybrides RGH4 améliore le nRMSE de validation du modèle basé sur la température SGT3 et du modèle basé sur l'humidité relative PGRH2 de 8.03% et 26.73%, respectivement. En ce qui concerne le DHI, le modèle basé sur les paramètres hybrides RDH3 réduit le nRMSE de validation du modèle basé sur l'indice de clarté RDC2 de 9.24 %. Tandis que pour le DNI, le modèle basé sur les paramètres hybrides RDNH1 minimise le nRMSE de validation du modèle PDNC1 de 17.45%.

Chapitre 3 : Estimation des irradiations solaires journalières en utilisant les modèles d'apprentissage automatique.

La figure 3.23 montre les diagrammes de dispersion des valeurs quotidiennes mesurées et estimées de GHI, DHI et DNI obtenus en utilisant le modèle le plus précis trouvé dans chaque catégorie d'entrées.

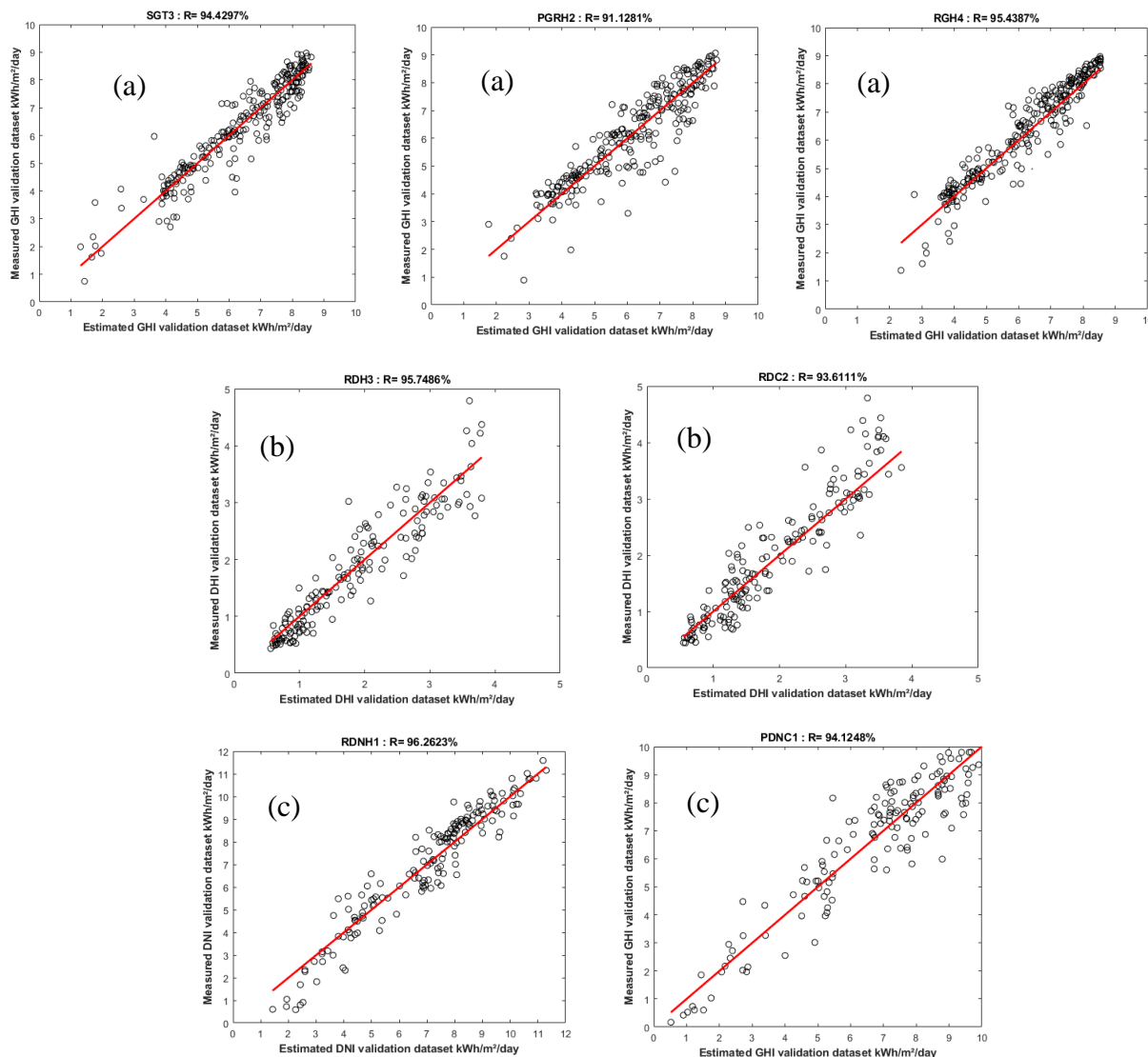


Figure 3. 23. Nuages de points des valeurs quotidiennes mesurées et estimées de GHI (a), DHI (b) et DNI (c) à partir du modèle le plus précis trouvé dans chaque catégorie d'entrées.

En général, les estimations des modèles basés sur la température (TBMs), sur l'humidité relative (RHBMs), ou sur l'indice de clarté (CIBMs) sont aussi crédibles et peuvent être considérés pour examiner le potentiel du rayonnement solaire. Pour estimer le GHI, les meilleurs modèles sélectionnés qui sont basés sur la température SGT3 et l'humidité relative MGRH2 ont fourni un bon coefficient de corrélation R égal à 94.43 % et 91.13 %, respectivement. De même, les meilleurs modèles sélectionnés en se basant que sur l'indice de clarté ont également donné des

Chapitre 3 : Estimation des irradiations solaires journalières en utilisant les modèles d'apprentissage automatique.

valeurs satisfaisantes de R, égales 93.61 % pour l'estimation du DHI (RDC2), et 94.12% pour l'estimation du DNI (MDNC1).

7.6. Performances des meilleurs modèles catégorisés obtenus sur la base de données des autres villes

Dans cette section, les meilleurs modèles d'apprentissage automatique obtenus précédemment dans chaque catégorie des entrées ont été testés sur les autres sites d'étude pour estimer les trois composantes solaires journalières (GHI, DHI, et DNI). Les tableaux 3.20, 3.21, et 3.22 présentent les performances obtenues pour chaque station, respectivement. En général, les résultats obtenus montrent que les modèles RF (forêts aléatoires) basés sur les paramètres hybrides surpassent tous les autres modèles dans l'estimation journalière des trois composantes solaires. Pour le GHI (Tableau 3.20), les meilleures performances sont obtenues par le modèle RGH4 avec des nRMSEs de validation égaux à 9.55%, 14.26%, 12.91%, 16.71%, pour Erfoud, Missouri, Oujda, et TanTan, respectivement (le modèle RGH4 n'a pas été testé pour Fès, vu l'indisponibilité des mesures de la pression).

Tableau 3. 20. Indicateurs statistiques obtenus par les meilleurs modèles d'apprentissage automatique catégorisés pour l'estimation journalière de GHI aux autres sites d'étude. Les chiffres en gras indiquent la meilleure performance obtenue parmi ces modèles.

Site	Modèle	Structure	Entraînement					Validation				
			R	MAE	RMSE	MBE	nRMSE (%)	R	MAE	RMSE	MBE	nRMSE (%)
Erfoud	SGT3	RBF kernel	0.9349	0.4717	0.6758	0.1201	11.4738	0.9334	0.4476	0.6212	0.0400	10.4332
	MGRH2	28 neurones	0.9078	0.5124	0.7245	0.0242	12.2478	0.9182	0.5117	0.7047	0.0718	11.9526
Missour	RGH4	122 apprenants	0.9659	0.3357	0.4725	-0.0018	8.0303	0.9403	0.4193	0.5702	-0.0210	9.5534
	SGT3	RBF kernel	0.8944	0.6528	0.8819	0.1184	15.1152	0.8828	0.7064	0.9510	0.0439	15.5635
Oujda	MGRH2	13 neurones	0.8785	0.7217	0.9336	-0.0050	15.6520	0.8814	0.7494	0.9628	0.0742	16.5816
	RGH4	132 apprenants	0.9556	0.4622	0.6112	-0.0017	10.5235	0.8911	0.6845	0.8809	-0.0602	14.2696
TanTan	SGT3	RBF kernel	0.9311	0.5460	0.8093	0.0994	14.9199	0.9399	0.5331	0.7555	0.0382	14.2471
	MGRH2	6 neurones	0.9189	0.6309	0.8723	-0.0016	16.5673	0.9168	0.6290	0.8590	-0.0346	15.1405
Fès	RGH4	127 apprenants	0.9735	0.3668	0.5196	0.0028	9.8393	0.9428	0.5327	0.7280	-0.0193	12.9129
	SGT3	RBF kernel	0.7938	0.7212	1.0050	0.1405	19.8245	0.8199	0.6797	0.9184	0.0601	17.7903
Fès	MGRH2	9 neurones	0.7906	0.7719	0.9962	-0.0930	19.4079	0.8031	0.7652	0.9817	-0.0385	19.5756
	RGH4	118 apprenants	0.9260	0.4763	0.6471	0.0044	12.7303	0.8733	0.6532	0.8576	-0.0197	16.7150
Fès	SGT3	RBF kernel	0.9341	0.5026	0.7601	0.0885	14.9176	0.9463	0.4765	0.6705	0.0896	13.2257
	MGRH2	12 neurones	0.9215	0.5633	0.8183	-0.0172	16.2062	0.9280	0.5540	0.7691	-0.0599	14.8565

Tableau 3. 21. Indicateurs statistiques obtenus par les meilleurs modèles d'apprentissage automatique catégorisés pour l'estimation journalière de DHI aux autres sites d'étude. Les chiffres en gras indiquent la meilleure performance obtenue parmi ces modèles.

Site	Modèle	Structure	Entraînement					Validation				
			R	MAE	RMSE	MBE	nRMSE (%)	R	MAE	RMSE	MBE	nRMSE (%)
Erfoud	RDH3	142 apprenants	0.9617	0.0412	0.0573	0.0001	17.2147	0.9415	0.0574	0.0730	-0.0020	20.6965
	RDC2	125 apprenants	0.9368	0.2715	0.3625	-0.0004	19.3838	0.9008	0.3203	0.4172	0.0361	22.4812
Missour	RDH3	138 apprenants	0.9665	0.0431	0.0584	0.0001	16.6305	0.9549	0.0547	0.0743	-0.0057	21.3400
	RDC2	121 apprenants	0.9430	0.2435	0.3260	-0.0008	17.6119	0.8989	0.2921	0.3785	0.0743	22.5014
Oujda	RDH3	130 apprenants	0.9780	0.0351	0.0503	0.0002	13.3504	0.9649	0.0473	0.0663	-0.0015	17.1611
	RDC2	145 apprenants	0.9556	0.1917	0.2711	-0.0002	15.5678	0.9361	0.2395	0.3322	0.0344	19.5648
TanTan	RDH3	123 apprenants	0.9789	0.0420	0.0560	0.0006	10.6681	0.9730	0.0551	0.0695	-0.0169	13.2053
	RDC2	135 apprenants	0.9557	0.2352	0.3224	-0.0025	13.4895	0.9468	0.2834	0.3750	0.0039	14.6035
Fès	RDH3	140 apprenants	0.9813	0.0298	0.0398	0.0002	14.3774	0.9723	0.0405	0.0511	-0.0008	17.6224
	RDC2	126 apprenants	0.9833	0.0286	0.0385	-0.0003	13.7090	0.9666	0.0406	0.0524	-0.0035	18.6685

Chapitre 3 : Estimation des irradiations solaires journalières en utilisant les modèles d'apprentissage automatique.

Concernant l'estimation journalière de DHI (Tableau 3.21), le modèle RDH3 a donné de bonnes performances dans toutes les stations, en marquant une meilleure précision à TanTan (nRMSE=13.20%). De même, le modèle RDNH1 a donné une très bonne précision d'estimation journalière du DNI dans chaque station (Tableau 3.22). Notons que la meilleure performance obtenue est celle à Oujda avec la plus faible valeur du nRMSE égale à 10.40% (Bounoua and Mechaqrane, 2022).

Tableau 3. 22. Indicateurs statistiques obtenus par les meilleurs modèles d'apprentissage automatique catégorisés pour l'estimation journalière de DNI aux autres sites d'étude. Les chiffres en gras indiquent la meilleure performance obtenue parmi ces modèles.

Site	Modèle	Structure	Entraînement					Validation				
			R	MAE	RMSE	MBE	nRMSE (%)	R	MAE	RMSE	MBE	nRMSE (%)
Erfoud	RDNH1	136 apprenants	0.9756	0.0208	0.0286	0.0001	8.5730	0.9530	0.0300	0.0388	-0.0003	11.3570
	MDNC1	13 neurones	0.9292	0.0471	0.0583	-0.0010	15.3142	0.9441	0.0438	0.0537	-0.0088	13.8477
Missour	RDNH1	128 apprenants	0.9844	0.0199	0.0280	0.0001	7.4990	0.9625	0.0278	0.0392	-0.0031	10.8640
	MDNC1	17 neurones	0.9671	0.0355	0.0460	-0.0002	12.3010	0.9613	0.0382	0.0483	-0.0005	12.6493
Oujda	RDNH1	130 apprenants	0.9863	0.0142	0.0214	0.0001	6.7450	0.9676	0.0214	0.0293	0.0002	10.4070
	MDNC1	8 neurones	0.9659	0.0347	0.0450	-0.0004	12.9624	0.9734	0.0318	0.0409	0.0011	12.0584
TanTan	RDNH1	134 apprenants	0.9745	0.0165	0.0235	0.0002	9.954	0.9510	0.0264	0.0344	0.0002	12.6480
	MDNC1	10 neurones	0.9661	0.0344	0.0446	-0.0008	18.0814	0.9634	0.0382	0.0470	-0.0062	16.9991
Fès	RDNH1	121 apprenants	0.9830	0.0251	0.0326	-0.000	8.2342	0.9721	0.0332	0.0415	0.0008	10.5306
	MDNC1	7 neurones	0.9691	0.0330	0.0427	0.0014	10.8908	0.9629	0.0342	0.0445	0.0033	11.1024

8. Conclusion

Nous nous sommes intéressés dans ce chapitre à l'estimation journalière des composantes solaires au sol (globale et diffuse sur plan horizontal (GHI et DHI) et directe sur plan normal (DNI)) à l'aide des modèles d'apprentissage automatique: réseaux de neurones artificiels (ANN et NARX), méthodes d'ensemble (Boosting, Bagging, Forêts aléatoires), régression par machine à vecteur de support (SVM). Seules les variables couramment mesurées dans les stations météorologiques (Température, Humidité relative de l'air, Vitesse du vent, ...) et quelques paramètres astronomiques facilement calculables sont considérés

En premier lieu, nous avons proposé deux architectures des réseaux de neurones artificiels (PMC et NARX) pour estimer le GHI à la ville de Fès. Les meilleures architectures optimisées des deux modèles PMC et NARX obtenues ont été, ensuite, testées sur les autres sites d'enerMENA. Les résultats obtenus montrent que les valeurs de nRMSE sont inférieures à 20% pour toutes les stations, à l'exception de la ville de TanTan qui se caractérise par une variabilité météorologique élevée.

En deuxième lieu, nous avons exploré le potentiel des trois méthodes d'ensemble basées sur des arbres de régression (Bagging, Boosting et Forêts aléatoires) dans l'estimation du GHI

quotidien dans les cinq sites d'enerMENA. Aussi, une sixième base de données a été construite en combinant les données des cinq stations enerMENA. Les performances de ces méthodes d'ensemble ont été comparées à celles du modèle PMC et des 22 modèles empiriques étudiés dans le deuxième chapitre. Une fois la combinaison des entrées pertinentes-déterminée, les trois méthodes d'ensemble ainsi que le modèle PMC ont été testés sur les données de la ville de Fès. Les résultats obtenus montrent que tous les modèles d'apprentissage automatique testés donnent de bonnes performances sur les ensembles de données d'entraînement et de validation ($nRMSE < 20\%$). Parmi ces modèles, le modèle des forêts aléatoires (Random Forests (RF)) s'est avéré être le plus performant pour toutes les stations, donnant un degré de confiance élevé à Zagora, avec des indicateurs statistiques de validation R, nMAE et nRMSE de 96.20%, 5.84 % et 7.85 %, respectivement, pour une estimation quotidienne de GHI. Il a été, aussi, constaté que le modèle RF est le plus robuste dans le cas de données bruitées (TanTan), alors que le modèle PMC est le plus sensible. A la fin, cette étude nous a permis de remarquer que les méthodes d'ensemble donnent de bonnes performances et un haut degré de stabilité et peuvent concurrencer les réseaux neuronaux artificiels, qui ont été largement utilisés dans le domaine de l'estimation et de la prédiction solaire. On peut donc conclure que les méthodes Bagging, Boosting et Random Forests sont les plus prometteuses.

Enfin, nous avons proposé un travail qui vise à tester les performances des méthodes d'apprentissage automatique (PMC, RF, et SVM) dans l'estimation des composantes solaires (GHI, DHI, DNI) lorsque seulement des mesures restreintes sont disponibles. Cette étude a concerné uniquement le site de Zagora dont la base de données est de meilleure qualité. Les différents modèles ont été catégorisés selon le type des entrées, en n'utilisant que des entrées couramment mesurées : La température (TBMs: Temperature Based Models) et l'humidité relative (RHBMs: Relative Humidity Based Models) pour l'estimation journalière de GHI, les paramètres hybrides (HBMs: Hybrid Based Models) pour l'estimation journalière de GHI, DHI et DNI, et l'indice de clarté (CIBMs: Clarity Index Based Models) pour l'estimation journalière de DHI et DNI. L'idée principale est d'évaluer les performances de chaque méthode d'apprentissage automatique lorsque seulement quelques variables météorologiques sont mesurées. Ces modèles ont été également comparés aux modèles empiriques cités dans la littérature ou proposés dans le présent travail (Chapitre 2). Les résultats montrent que les modèles d'apprentissage automatique sont plus performants que les modèles empiriques dans toutes les catégories considérées.

A partir de cette analyse, nous avons pu conclure que le modèle RF peut être utilisé pour l'estimation de l'ensemble des composantes solaires (GHI, DHI, DNI), en utilisant comme variables d'entrées les paramètres hybrides suivants :

- Pour le GHI: l'irradiation solaire hors atmosphère d'un plan horizontal (H_0), le gradient de la température (ΔT), la vitesse du vent (W_S) et la pression (P_r). Les performances obtenues sont: $R=95.44\%$ et $nRMSE=8.36\%$.
- Pour le DHI: l'irradiation globale d'un plan horizontal (GHI), l'indice de clarté (K_t), le l'irradiation solaire hors atmosphère d'un plan horizontal H_0 , le jour de l'année (DOY) et le déficit de pression de vapeur de l'air D (Equation 2.14 du chapitre 2). Les performances obtenues sont: $R=95.75\%$ et $nRMSE=18.55\%$.
- Pour le DNI: l'irradiation solaire hors atmosphère d'un plan normal (H_{0N}), l'indice de clarté (K_t), le jour de l'année (DOY) et la température journalière moyenne (T). Les performances obtenues sont: $R=96.26\%$ et $nRMSE=10.26\%$.

Toutefois, si seuls les mesures des températures sont disponibles, on peut utiliser la variante GT3 des modèles SVM, PMC et RF pour estimer le GHI en utilisant uniquement les mesures du gradient de température journalier (ΔT) et l'irradiation solaire journalière d'un plan horizontal hors atmosphère H_0 . Le coefficient de corrélation obtenu est de l'ordre de 94% pour l'ensemble des trois modèles, alors que le nRMSE est de 9.09%, 9.46% et 9.56% pour les trois modèles respectivement.

Si seules les mesures des valeurs extrêmes RH_{max} et RH_{min} de l'humidité relative sont disponibles, les modèles PMC, RF et SVM peuvent être utilisés indifféremment étant donné que leurs performances sont très proches ($nRMSE = 11.41\%$, 11.72% et 11.88% , respectivement).

Le prochain chapitre présentera la prédiction à l'avance du rayonnement global horizontal (GHI) pour des horizons temporelles horaires (1h) et sub-horaires (10min) en employant la nouvelle technique d'apprentissage profond (LSTM) ainsi que les modèles d'apprentissage automatique.

Chapitre 4

Prédiction à différents horizons temporels de
l'irradiation solaire globale sur plan horizontal

1. Introduction

Une connaissance précise des composantes globales, diffuses et directes du rayonnement solaire incident au sol est cruciale pour déterminer les régions ayant un potentiel de production d'énergie solaire, pour évaluer la consommation énergétique des bâtiments, pour subventionner des études éco-physiologiques, pour estimer l'évapotranspiration des cultures (production végétale tirée de l'exploitation de la terre) et pour soutenir la planification urbaine (Marques Filho et al., 2016). Toutefois, l'intégration de l'électricité solaire dans le mix énergétique entraîne de sérieux problèmes pour la maintenance de l'équilibre des réseaux électriques. En effet, les centrales solaires peuvent produire de l'électricité lorsque la demande est faible. Par conséquent, la prédiction des potentiels de production électrique des centrales solaires peut permettre aux gestionnaires des réseaux de préparer des scénarios de décision à l'avance. Le recours aux modèles de prédiction à l'avance des composantes solaires est une technique alternative très pratique pour bien exploiter cette quantité d'énergie renouvelable dont l'intermittence pénalise péniblement son intégration aux réseaux électriques.

Dans les chapitres 2 et 3, deux familles de modèles d'estimateurs ont été développées et testées pour estimer les trois composantes solaires journalières dont les résultats ont montré la supériorité des modèles d'apprentissage automatique par rapport aux modèles empiriques. En revanche, la prédiction à l'avance du rayonnement solaire à différents horizons temporels impose l'intervention d'un sous domaine de l'intelligence artificiel et du machine learning qui commence à se démarquer actuellement et connu par le nom : Deep Learning (apprentissage profond). Cette technique d'apprentissage est utilisée dans diverses applications telles que la prédiction des séries temporelles, position et expression du visage dans une image, la reconnaissance faciale, la segmentation des images, De plus, elle nécessite des bases de données considérablement volumineuses par rapport aux autres techniques du Machine Learning pour garantir de bons résultats.

Dans ce chapitre, une étude bibliographique sur les différentes techniques de l'apprentissage profond utilisées dans la prédiction des irradiances solaires sera présentée en premier lieu. Ensuite, la nouvelle approche de l'apprentissage profond connue par les réseaux de neurones à longue mémoire à court terme (LSTM: Long-Short Term Memory networks) est décrite, tout en montrant les différentes étapes suivies par la couche LSTM pour effectuer ses calculs de base. Nous procédons, enfin, à la présentation des résultats de deux études effectuées pour prédire à l'avance l'irradiation solaire globale horizontale (GHI) à différents horizons

temporels. La première étude a pour objectif de prédire le GHI horaire à court ($h+1$) et à long ($h+48$) termes en appliquant deux types de réseaux de neurones artificiels (RNAs). La deuxième étude consiste à prédire les GHI horaires et sub-horaires à l'aide du nouveau modèle d'apprentissage profond (LSTM), tout en comparant ses performances à celles de deux méthodes d'apprentissage automatique. Dans cette étude, deux scénarios de prédiction annuel et saisonnier ont été proposés pour des horizons temporels allant de $h+1$ à $h+6$ et de $10\text{min}+10$ à $10\text{min}+60$.

2. Modèles d'apprentissage profond (Deep Learning)

Le Deep Learning est un système qui fonctionne à partir de plusieurs couches de réseaux neuronaux en combinant différents algorithmes inspirés du fonctionnement du cerveau humain. Cette approche est capable de résoudre des problèmes complexes en se basant sur des données non structurées. Un système d'apprentissage profond permet d'identifier lui-même les caractéristiques discriminantes des données en cherchant dans chaque couche, un nouveau critère spécifique qui sert à bien prédire une valeur ou classifier un objet à la fin du processus.

En analysant un certain nombre de travaux effectués dans la littérature, nous avons remarqué qu'il existe deux types de modèles d'apprentissage profond qui ont été utilisés pour estimer ou prédire le rayonnement solaire: les réseaux de neurones à longue mémoire à court terme (LSTM) et les réseaux de neurones bidirectionnels à longue mémoire à court terme (Bidirectional Long-Short Term Memory networks: BiLSTM) qui représentent l'architecture d'un réseau neuronal artificiel récurrent (RNN). La différence entre ces deux techniques est que la sortie de la structure LSTM à l'instant t est liée seulement à l'instant $t-d$ (avec $d \geq 1$) (l'entraînement du réseau n'est basé que sur les informations du passé). Par contre dans la structure de BiLSTM, toutes les informations du passé et du futur sont utilisées pour l'entraînement, c'est-à-dire la sortie de la structure BiLSTM à l'instant t est liée à la fois à l'instant $t-d$ et à $t+d$ (avec $d \geq 1$) (Li et al., 2020). Un nouveau modèle basé sur les LSTM pour prédire l'irradiation solaire horaire à un jour à l'avance en utilisant les données de prévision météorologique a été proposé par Qing et Niu (Qing and Niu, 2018). Ainsi, en exploitant l'ensemble de données recueillies sur l'île de Santiago, au Cap-Vert, les résultats expérimentaux ont montré que l'algorithme proposé présente moins de sur-ajustement et une meilleure capacité de généralisation et donne les meilleures performances. De même, Srivastava et Lessmann (Srivastava and Lessmann, 2018) ont effectué une étude comparative pour évaluer les

performances des LSTM pour prévoir l'irradiation solaire globale horizontale à l'horizon d'un jour en utilisant les données satellitaires. Cette étude a prouvé que la méthode LSTM surpasse un grand nombre de méthodes alternatives et qu'elle peut être considérée comme une technique prometteuse parmi les nouvelles techniques de prévision. Dans (Kaba et al., 2018), le modèle d'apprentissage profond (DL) utilisé a donné des résultats très précis pour l'estimation de l'irradiation globale quotidienne. Les résultats trouvés sont généralement meilleurs ou comparables à ceux de nombreuses études existantes dans la littérature. Par conséquent, il a été conclu que la méthode est une bonne alternative et peut être appliquée avec succès à des régions similaires.

Pour prédire l'irradiation solaire horaire à un jour à l'avance en se basant uniquement sur des données météorologiques largement disponibles, Husein et Chung (Husein and Chung, 2019) ont développé un modèle de prévision à l'aide d'un réseau de neurones récurrent profond, à longue mémoire à court terme (LSTM-RNN). Six expériences ont été effectuées en utilisant des mesures provenant de différentes stations météorologiques de climats distincts. Leurs résultats ont montré que l'approche proposée est la plus précise et entraîne une augmentation de 2% des économies d'énergie annuelles. En considérant que le modèle de prévision de l'irradiation solaire horaire est la superposition d'un modèle de prévision de l'irradiation moyenne quotidienne et d'un modèle de prévision de l'amplitude de l'irradiation, Li et al. (Li et al., 2021) ont construit deux sous-modèles en utilisant un réseau BiLSTM profond. Ils ont trouvé que la méthode proposée est la plus performante pour 84%, 72% et 68% par rapport au nombre total de stations utilisées, en termes d'erreur quadratique moyenne, d'erreur absolue moyenne et du coefficient de détermination, respectivement.

En proposant une étude comparative sur les réseaux LSTM dans la prévision à un jour à l'avance de l'irradiation solaire globale horizontale, Srivastava et Lessmann (Srivastava and Lessmann, 2018) ont montré que le LSTM surpasse un grand nombre de méthodes alternatives avec une marge substantielle de 52.2% par rapport au modèle de persistance. Aussi, en comparant le modèle LSTM avec l'algorithme de persistance, la régression linéaire des moindres carrés et les réseaux de neurones multicouches à anticipation utilisant l'algorithme de rétro-propagation (BPNN), les résultats expérimentaux de Qing et Niu (Qing and Niu, 2018) ont confirmé la supériorité du LSTM par rapport à ces algorithmes concurrents pour la prédiction d'une seule sortie. De plus, les analyses comparatives effectuées par Husein et al. (Husein and Chung, 2019)

ont montré que l'approche proposée en utilisant un réseau de neurones récurrent à longue mémoire à court terme (LSTM-RNN) est plus précise que le réseau de neurones à anticipation (FFNN) et atteint une précision de 60.31 W/m^2 en termes de RMSE, pour une prévision d'un jour en avance de l'irradiation solaire.

3. Réseaux de neurones à longue mémoire à court terme (LSTM)

Un réseau LSTM est une approche d'apprentissage profond, inventée en 1997 par Hochreiter et Schmidhuber (Hochreiter and Schmidhuber, 1997). Cette technique est basée sur l'architecture des réseaux de neurones récurrents (RNN) et elle est actuellement appliquée dans diverses applications, telles que les prédictions de séries temporelles, le contrôle de robots, la reconnaissance vocale, Les LSTM-NN sont capables d'apprendre les dépendances à long terme dans les données. Le modèle LSTM possède une architecture spéciale appelée cellule de mémoire, basée sur différentes portes (gates) qui sont responsables de la sélection des données à conserver ou à éliminer. De plus, le modèle LSTM a une grande capacité à mémoriser plus de données que le RNN grâce au contrôle du gradient (Haider et al., 2019).

Comme le montre la figure 4.1, la structure de la cellule de mémoire du LSTM comprend trois portes pour ajuster son état de cellule S_t : une porte d'entrée i_t (input gate), une porte de sortie o_t (output gate) et une porte d'oubli f_t (forget gate). A chaque pas de temps t , chacune des trois portes est présentée en fonction de l'entrée courante x_t et des valeurs précédentes de la sortie de la cellule de mémoire h_{t-1} . Chaque porte fonctionne comme un filtre pour accomplir un objectif bien défini (Fischer and Krauss, 2018) :

- La porte d'entrée détermine les informations (nouvelles entrées) à ajouter à l'état de la cellule de mémoire.
- La porte de sortie définit les informations à utiliser comme sortie de l'état de la cellule de mémoire (sortie de la cellule).
- La porte d'oubli spécifie les informations à supprimer ou à conserver pour une référence future à partir de l'état de la cellule de mémoire.

Les calculs de base effectués par la couche LSTM à chaque pas de temps t pour mettre à jour sa cellule mémoire sont décrits en quatre étapes distinctes ((Fischer and Krauss, 2018), (Ghimire et al., 2019)):

- **Étape 1** : La couche LSTM détermine l'information qui doit être rejetée à partir des états précédents de ses cellules S_{t-1} . Cela permet de calculer les valeurs d'activation des portes d'oubli f_t au pas du temps t en fonction de l'entrée courante x_t et des valeurs précédentes des sorties h_{t-1} des cellules mémoire ainsi que des termes de biais b_f des portes d'oubli.

$$f_t = \sigma(W_{f,x} x_t + W_{f,h} h_{t-1} + b_f) \quad (4-1)$$

Dans l'équation (4-1), σ présente la fonction sigmoïde logistique qui met à l'échelle toutes les valeurs d'activation dans une plage comprise entre 0 (oubli complet) et 1 (souvenir complet) ; $W_{f,x}$ et $W_{f,h}$ sont les matrices des poids.

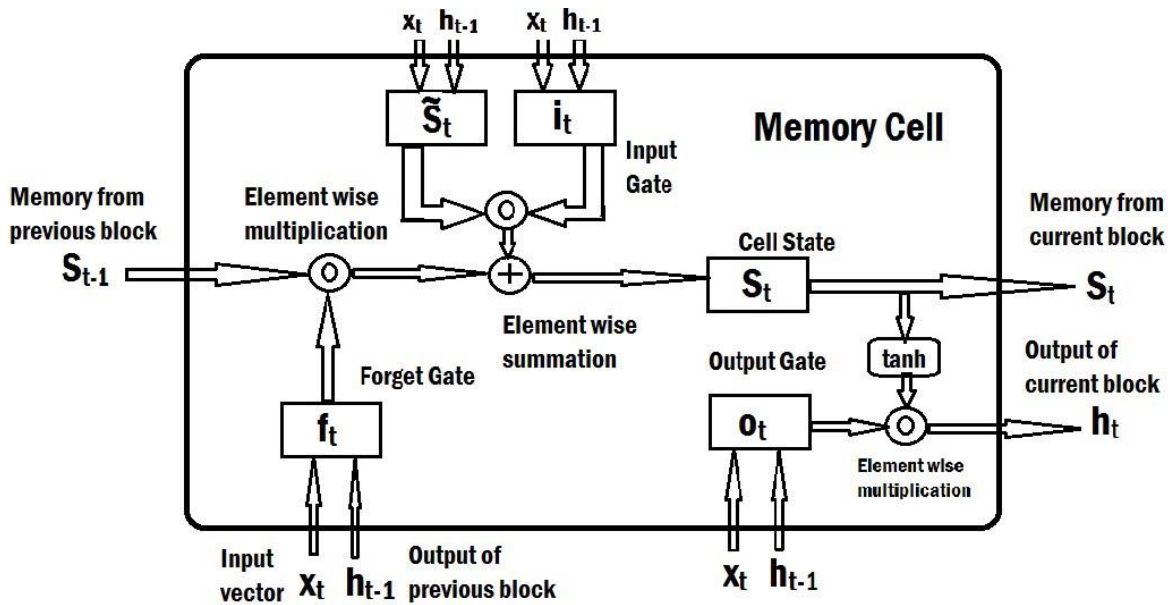


Figure 4. 1. Structure d'une cellule à longue mémoire à court terme (Amey laddad, 2019).

- **Étape 2** : Les informations déterminées par la couche LSTM sont ajoutées aux états de la cellule du réseau S_t . Par conséquent, un nouveau candidat d'état de cellule \tilde{S}_t est généré et les valeurs d'activation i_t des portes d'entrée sont calculées.

$$\tilde{S}_t = \tanh(W_{\tilde{s},x} x_t + W_{\tilde{s},h} h_{t-1} + b_{\tilde{s}}) \quad (4-2)$$

$$i_t = \sigma(W_{i,x} x_t + W_{i,h} h_{t-1} + b_i) \quad (4-3)$$

- **Étape 3 :** Mise à jour du nouvel état de la cellule S_t , en combinant l'état de la cellule précédent S_{t-1} affecté par la porte d'oubli f_t et le nouveau candidat d'état de cellule \tilde{S}_t mis à l'échelle par la porte d'entrée i_t .

$$S_t = f_t * S_{t-1} + i_t * \tilde{S}_t \quad (4-4)$$

- **Étape 4 :** La sortie souhaitée h_t des cellules de mémoire est déterminée en fonction de deux étapes principales. La première étape consiste à construire une nouvelle porte appelée "porte de sortie" o_t pour décider quelles parties de l'état de la cellule doivent se trouver dans la sortie. La deuxième étape consiste à filtrer l'état de la cellule S_t activée par la fonction tangente hyperbolique (\tanh) en la multipliant par o_t .

$$o_t = \sigma(W_{o,x} x_t + W_{o,h} h_{t-1} + b_o) \quad (4-5)$$

$$h_t = o_t * \tanh(S_t) \quad (4-6)$$

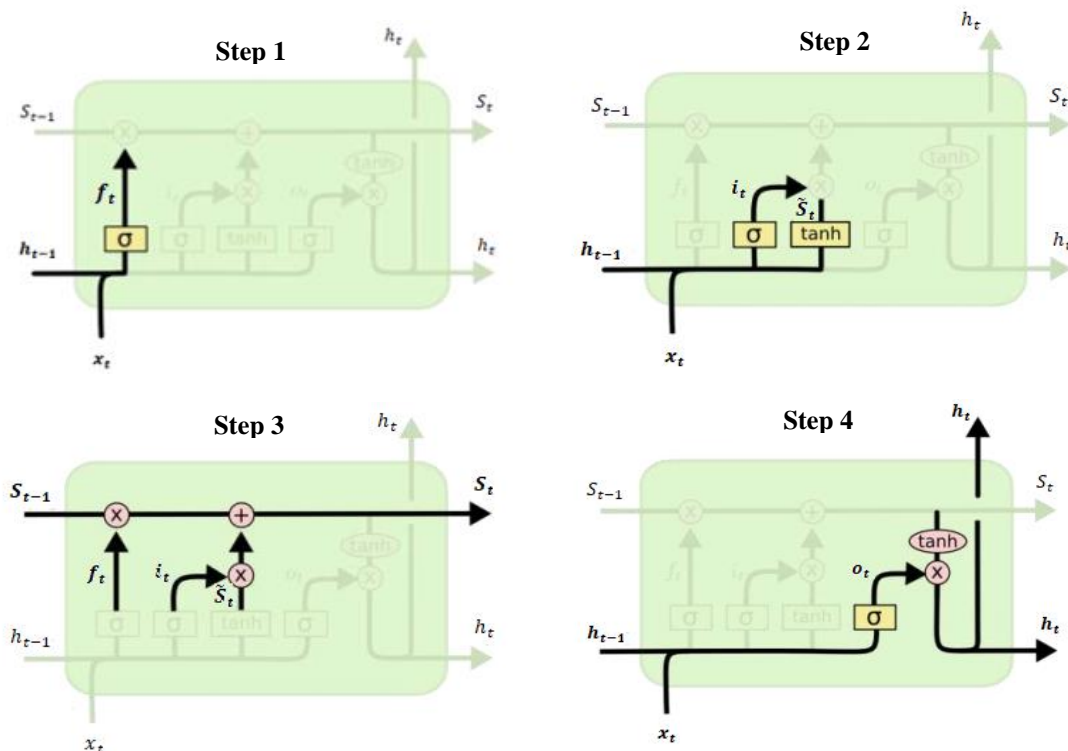


Figure 4. 2. Procédure des calculs de base effectués par la couche LSTM étape par étape (Jeff Hu, 2017).

La figure 4.2 illustre, étape par étape, la procédure des calculs de base effectués par la couche LSTM.

4. Résultats et discussions

Deux études ont été effectuées pour prédire à l'avance l'irradiation solaire globale horizontale (GHI) pour des horizons temporels horaires et sub-horaires. Une étude pour prédire la GHI à court ($h+1$) et à long ($h+48$) termes à l'aide des réseaux de neurones artificiels et une étude pour prédire cette même GHI en utilisant la nouvelle approche d'apprentissage profond basée sur les réseaux LSTM pour des horizons temporels allant de $h+1$ à $h+6$ et de $10\text{min}+10$ à $10\text{min}+60$. Les performances de cette nouvelle technique seront ensuite comparées à celles d'un réseau de neurones PMC (Perceptron Multi-Couches) et à la méthode des forêts aléatoires (Random Forests). Toutes les performances des modèles développés seront évaluées à l'aide des indicateurs statistiques présentés dans l'Annexe D.

4.1. Prédiction de l'irradiation solaire globale horizontale à court et à long termes à la ville de Fès

Dans cette partie, nous nous sommes intéressés à la prédiction de la valeur horaire du GHI une heure à l'avance ($h+1$) (court terme) et à deux jours à l'avance (prédiction de la valeur horaire du GHI 48 heures à l'avance $h+48$) (long terme). Deux types de réseaux de neurones artificiels (RNAs) seront testés: le Perceptron multicouche (PMC) et le réseau de neurones autorégressif à entrées exogènes (NARX). Dans la littérature, les études basées sur les techniques de réseaux de neurones de type PMC ont été plus utilisées que les méthodes NARX pour prédire la GHI à différents horizons temporels. Notre objectif ici est de prédire la GHI horaire aux horizons $h+1$ et $h+48$ à la ville de Fès en utilisant que des paramètres météorologiques communément mesurés, tout en comparant les performances du modèle NARX à celles du modèle PMC (Bounoua and Mechaqrane, 2020).

Dans cette étude, six années de données horaires (2009-2014) ont été utilisées pour développer les modèles permettant de prédire la GHI horaire à court terme ($h+1$). Les modèles développés sont ensuite utilisés pour prédire la GHI à long terme ($h+48$) pour l'année 2015 (les modèles PMC et NARX développés pour prédire GHI à $h+1$ seront évalués et testés pour prédire les valeurs horaires de GHI des deux jours de l'année 2015 (1 et 2 Janvier 2015)). Pour établir ces modèles, 70% des données sont présentées au réseau pendant l'entraînement et le réseau est ajusté en fonction de l'erreur obtenue. Par la suite, 15% des données sont utilisées pour mesurer

la généralisation du réseau et pour arrêter l'entraînement lorsque la généralisation cesse de s'améliorer. Enfin, les 15% restants des données fournissent une mesure indépendante de la performance du réseau pendant et après l'entraînement, avec aucun effet sur la phase d'entraînement (Bounoua and Mechaqrane, 2020).

4.1.1. Sélection de la meilleure combinaison de variables d'entrée

La précision d'un modèle de prédiction dépend du choix des entrées pertinentes qui ont un impact significatif sur les fluctuations des variables de sortie. Pour cela, nous avons effectué une analyse statistique pour tester la corrélation de chaque variable d'entrée avec le GHI en calculant le coefficient de corrélation. Le tableau 1 montre le degré de corrélation trouvé pour chaque variable d'entrée. Selon ce tableau, cinq entrées météorologiques (T, RH, TOA, h et SD) ont une valeur de régression acceptable par rapport aux autres variables d'entrée (W_s , W_d et P).

Afin de tester l'effet du regroupement des variables météorologiques sur la précision de la prévision du GHI, nous avons proposé d'examiner la possibilité de combiner entre les différents paramètres d'entrée du tableau 4.1 (avec bon ou mauvais degré de corrélation avec le GHI). Les architectures des deux réseaux de neurones (NARX et PMC) ont été entraînées, testées et validées en utilisant différentes combinaisons entre les variables météorologiques présentées dans le tableau 4.1.

Tableau 4. 1. Corrélation du GHI horaire avec les variables météorologiques.

Variables d'entrées	R (%)	Variables d'entrées	R (%)
Température (T)	43.23	Précipitation (P)	13.21
Humidité Relative (RH)	47.43	Irradiation solaire hors atmosphère (TOA)	84.93
Vitesse du vent (W_s)	11.37	Altitude solaire (h)	79.78
Direction du vent (W_d)	23.49	Durée d'ensoleillement (SD)	44.23

Les performances des modèles PMC et NARX obtenues pour prédire les valeurs horaires de GHI ($h+1$) en utilisant les dix combinaisons significatives ($nRMSE < 30\%$) des huit variables météorologiques sont présentées dans le tableau 4.2 et 4.3, respectivement. Ces résultats montrent que la meilleure combinaison obtenue est combinant les données horaires de la température (T), de l'humidité relative (RH), de l'irradiation solaire au sommet de l'atmosphère

Chapitre 4 : Prédiction à différents horizons temporels de l'irradiation solaire globale sur plan horizontal.

(TOA), de l'altitude solaire (h), de la durée d'ensoleillement (SD) et de la vitesse du vent (W_s) avec un degré de récurrence d'une heure (mesurés à une heure d'avant), ainsi les valeurs horaires de GHI mesurées à deux heures d'avant (deux valeurs retardées de la sortie). Les indicateurs statistiques obtenus sont $R = 96.37\%$, $nMAE = 9.81\%$, $nRMSE = 16.07\%$, et $R = 92.21\%$, $nMAE = 16.50\%$, $nRMSE = 23.81\%$, pour les modèles NARX et PMC, respectivement (Bounoua and Mechaqrane, 2020).

Tableau 4. 2. Indicateurs statistiques obtenus par le modèle PMC en testant différentes combinaisons d'entrées pour l'ensemble de données de validation sur 5 exécutions.

Combinaisons d'entrées	PMC		
	R (%)	nMAE (%)	nRMSE (%)
$T_{t-1}, RH_{t-1}, W_{s_{t-1}}, W_{d_{t-1}}, P_{t-1}, TOA_{t-1}, h_{t-1}, SD_{t-1}, GHI_{t-1}, GHI_{t-2}$	91.68	17.53	24.82
$T_{t-1}, RH_{t-1}, W_{s_{t-1}}, TOA_{t-1}, GHI_{t-1}, GHI_{t-2}$	90.98	18.21	26.21
$T_{t-1}, RH_{t-1}, TOA_{t-1}, h_{t-1}, SD_{t-1}, GHI_{t-1}, GHI_{t-2}$	92.01	16.82	24.15
$T_{t-1}, RH_{t-1}, TOA_{t-1}, h_{t-1}, P_{t-1}, GHI_{t-1}, GHI_{t-2}$	90.19	18.54	27.13
$T_{t-1}, RH_{t-1}, W_{s_{t-1}}, W_{d_{t-1}}, TOA_{t-1}, GHI_{t-1}, GHI_{t-2}$	90.074	19.71	26.62
$T_{t-1}, RH_{t-1}, W_{s_{t-1}}, W_{d_{t-1}}, TOA_{t-1}, SD_{t-1}, GHI_{t-1}, GHI_{t-2}$	91.62	17.61	24.81
$T_{t-1}, RH_{t-1}, TOA_{t-1}, SD_{t-1}, GHI_{t-1}, GHI_{t-2}$	91.31	17.83	25.62
$T_{t-1}, RH_{t-1}, TOA_{t-1}, h_{t-1}, SD_{t-1}, W_{s_{t-1}}, GHI_{t-1}, GHI_{t-2}$	92.21	16.50	23.81
$T_{t-1}, RH_{t-1}, TOA_{t-1}, h_{t-1}, SD_{t-1}, P_{t-1}, GHI_{t-1}, GHI_{t-2}$	91.79	17.13	24.52
$T_{t-1}, TOA_{t-1}, h_{t-1}, SD_{t-1}, GHI_{t-1}, GHI_{t-2}$	90.93	18.25	25.98

En revanche, il faut noter que le deuxième modèle NARX (Tableau 4.3) n'utilisant que la température (T_{t-1}), l'humidité relative (RH_{t-1}), la vitesse du vent ($W_{s_{t-1}}$), l'irradiation solaire au sommet de l'atmosphère (TOA_{t-1}), et les deux valeurs précédentes de GHI (GHI_{t-1} et GHI_{t-2}) donne une très bonne précision de prédiction en absence des valeurs mesurées de la durée d'ensoleillement (SD), avec une valeur de $nRMSE$ égale à 16.67%. Notons que la valeur de $nRMSE$ obtenue par le meilleur modèle NARX ($T_{t-1}, RH_{t-1}, TOA_{t-1}, h_{t-1}, SD_{t-1}, W_{s_{t-1}}, GHI_{t-1}, GHI_{t-2}$) est de 16.07% qui représente une amélioration que de 3.59% par rapport au deuxième modèle NARX ($T_{t-1}, RH_{t-1}, W_{s_{t-1}}, TOA_{t-1}, GHI_{t-1}, GHI_{t-2}$). Ceci indique qu'en ne se basant que sur les paramètres météorologiques couramment mesurés, le deuxième modèle est le plus approprié.

Chapitre 4 : Prédiction à différents horizons temporels de l'irradiation solaire globale sur plan horizontal.

Tableau 4. 3. Indicateurs statistiques obtenus par le modèle NARX en testant différentes combinaisons d'entrées pour l'ensemble de données de validation sur 5 exécutions.

Combinaisons d'entrées	NARX		
	R (%)	nMAE (%)	nRMSE (%)
$T_{t-1}, RH_{t-1}, W_{S_{t-1}}, W_{d_{t-1}}, P_{t-1}, TOA_{t-1}, h_{t-1}, SD_{t-1}, GHI_{t-1}, GHI_{t-2}$	95.82	11.23	17.23
$T_{t-1}, RH_{t-1}, W_{S_{t-1}}, TOA_{t-1}, GHI_{t-1}, GHI_{t-2}$	96.15	10.61	16.67
$T_{t-1}, RH_{t-1}, TOA_{t-1}, h_{t-1}, SD_{t-1}, GHI_{t-1}, GHI_{t-2}$	96.03	10.82	16.52
$T_{t-1}, RH_{t-1}, TOA_{t-1}, h_{t-1}, P_{t-1}, GHI_{t-1}, GHI_{t-2}$	95.79	10.94	17.35
$T_{t-1}, RH_{t-1}, W_{S_{t-1}}, W_{d_{t-1}}, TOA_{t-1}, GHI_{t-1}, GHI_{t-2}$	95.73	11.15	17.84
$T_{t-1}, RH_{t-1}, W_{S_{t-1}}, W_{d_{t-1}}, TOA_{t-1}, SD_{t-1}, GHI_{t-1}, GHI_{t-2}$	96.09	10.72	16.71
$T_{t-1}, RH_{t-1}, TOA_{t-1}, SD_{t-1}, GHI_{t-1}, GHI_{t-2}$	95.85	10.91	17.10
$T_{t-1}, RH_{t-1}, TOA_{t-1}, h_{t-1}, SD_{t-1}, W_{S_{t-1}}, GHI_{t-1}, GHI_{t-2}$	96.37	09.81	16.07
$T_{t-1}, RH_{t-1}, TOA_{t-1}, h_{t-1}, SD_{t-1}, RF_{t-1}, GHI_{t-1}, GHI_{t-2}$	96.20	10.26	16.91
$T_{t-1}, TOA_{t-1}, h_{t-1}, SD_{t-1}, GHI_{t-1}, GHI_{t-2}$	96.16	10.36	17.13

4.1.2. Sélection du nombre de neurones dans la couche cachée

Dans cette partie, les meilleures combinaisons d'entrées sélectionnées précédemment sont utilisées avec différents nombres de neurones dans la couche cachée afin de trouver la meilleure architecture pour les deux modèles NARX et PMC. Une approche constructive est utilisée, en augmentant itérativement le nombre de neurones dans la couche cachée de 1 à 30. Pour chaque architecture, les performances ont été évaluées sur 5 exécutions. Les résultats de validation montrent que les meilleures performances des deux modèles NARX et PMC sont obtenues avec 12 neurones cachés avec, un R égal à 96.57% et 92.96%, un nMAE égal à 9.51% et 15.90%, un nRMSE égal à 15.73% et 22.75%, respectivement.

La figure 4.3 montre, pour l'ensemble de données de validation, les régressions correspondant aux meilleures performances obtenues pour les modèles NARX et PMC. La valeur de R au-dessus de chaque graphique représente le coefficient de corrélation.

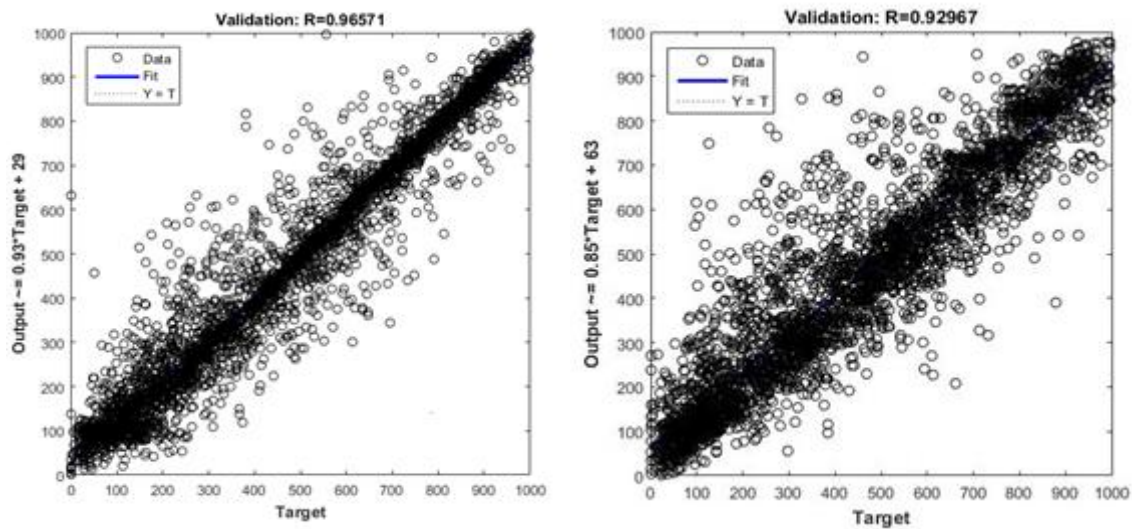


Figure 4. 3. Régressions entre les valeurs mesurées et prédites par les modèles NARX (à gauche) et PMC (à droite) pour les ensembles de données de validation.

La figure 4.4 montre les séries temporelles prédites et mesurées du GHI à l'horizon h+1, correspondantes aux meilleures performances obtenues par les modèles NARX et PMC pour une semaine sélectionnée aléatoirement. On note que le modèle NARX a une très bonne précision dans la prédiction des valeurs futures du GHI, comparé au modèle PMC, qui a, toutefois, une précision acceptable.

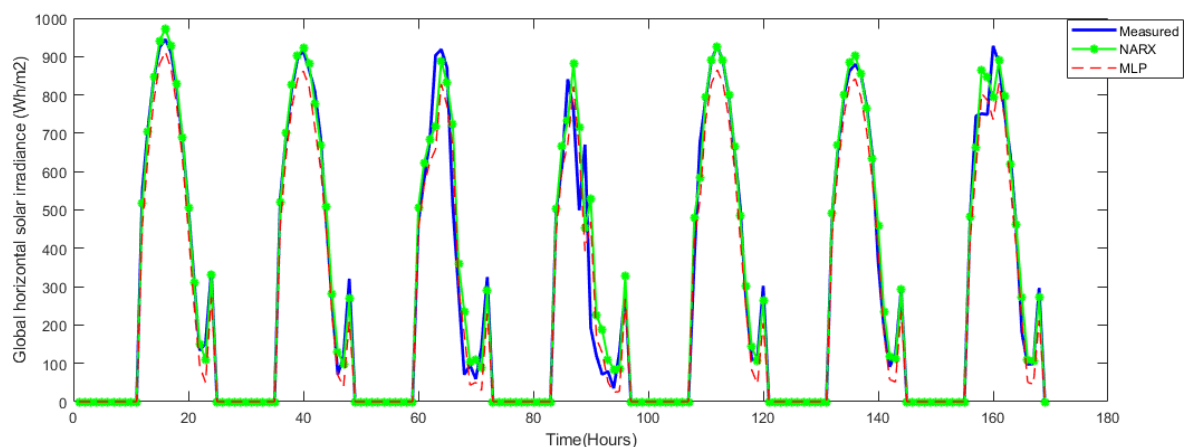


Figure 4. 4. Valeurs horaires prédites et mesurées de GHI pour une semaine choisie aléatoirement en utilisant les modèles NARX et PMC (MLP).

Pour explorer la capacité de prédiction des modèles développés à long terme (h+48), les modèles NARX et PMC sont utilisés pour prédire la GHI horaire pour deux jours à l'avance durant l'année 2015. D'après la figure 4.5, on peut noter que les prédictions à l'aide du modèle NARX sont meilleures que celles du modèle PMC.

On peut conclure que le modèle NARX est plus performant pour les prévisions à court et à long terme du GHI, avec des valeurs du nRMSE nettement meilleures que celles du modèle PMC (Bounoua and Mechaqrane, 2020).

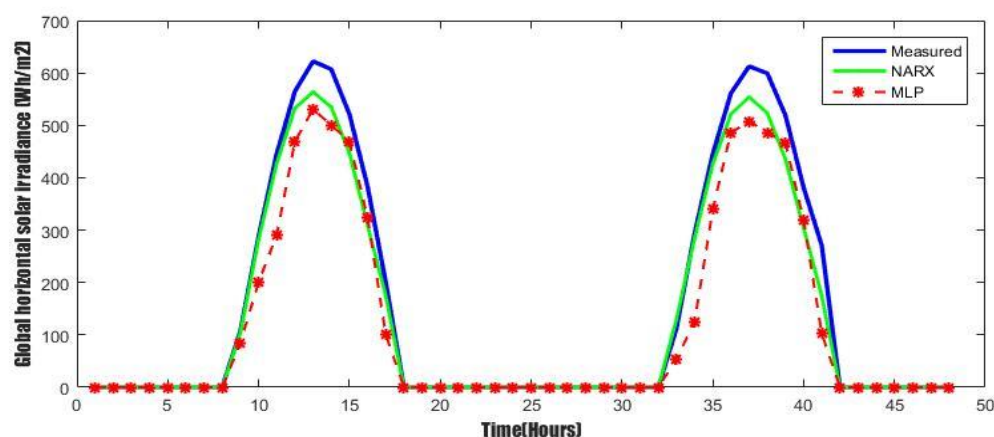


Figure 4. 5. Prédiction horaire de l'irradiation globale horizontale pour deux jours à l'avance ($h+48$) par rapport aux données mesurées du 1 et 2 janvier 2015 au site de Fès.

4.2. Performances du modèle NARX sur les bases de données enerMENA

En utilisant la meilleure combinaison d'entrées (T , RH , TOA , h , SD , W_s) et la meilleure architecture (12 neurones dans la couche cachée) du modèle NARX trouvées précédemment, nous avons testé l'efficacité de ce modèle sur les cinq bases de données d'enerMENA afin de prédire la GHI à court terme ($h+1$). Le tableau 4.4 montre les performances d'entraînement et de validation obtenues pour chaque site. D'après les résultats de validation, on constate que la meilleure précision est obtenue à Zagora (en gras) avec les plus faibles valeurs des nRMSE (10.86%), nMAE (6.31%), et la plus grande valeur de R (97.93%). La plus faible performance est celle observée à TanTan avec des valeurs de nRMSE et de R égales à 15.87% et 96.85%, respectivement.

Pour l'ensemble de données d'entraînement et de validation, les régressions obtenues à la ville de Zagora en utilisant le modèle NARX pour prédire la GHI à l'horizon $h+1$ sont présentées dans la figure 4.6. La figure 4.7 montre les séries temporelles prédites et mesurées à la ville de Zagora pour une semaine sélectionnée aléatoirement. Cette figure montre un très bon accord entre les valeurs horaires mesurées et prédites.

Tableau 4. 4. Indicateurs statistiques obtenus en testant la meilleure combinaison d'entées et la meilleure architecture du modèle NARX ($h+1$) sur les sites enerMENA pour l'ensemble de données d'entraînement et de validation pour 5 exécutions.

Site	Entrainement			Validation		
	R (%)	nMAE (%)	nRMSE (%)	R (%)	nMAE (%)	nRMSE (%)
Erfoud	98.30	6.07	9.84	97.82	6.60	10.96
Missour	97.04	9.09	14.25	96.78	9.90	15.41
Oujda	97.09	9.05	14.67	97.15	8.67	14.04
TanTan	97.14	10.05	15.03	96.85	10.59	15.87
Zagora	97.97	6.66	10.87	97.93	6.31	10.86

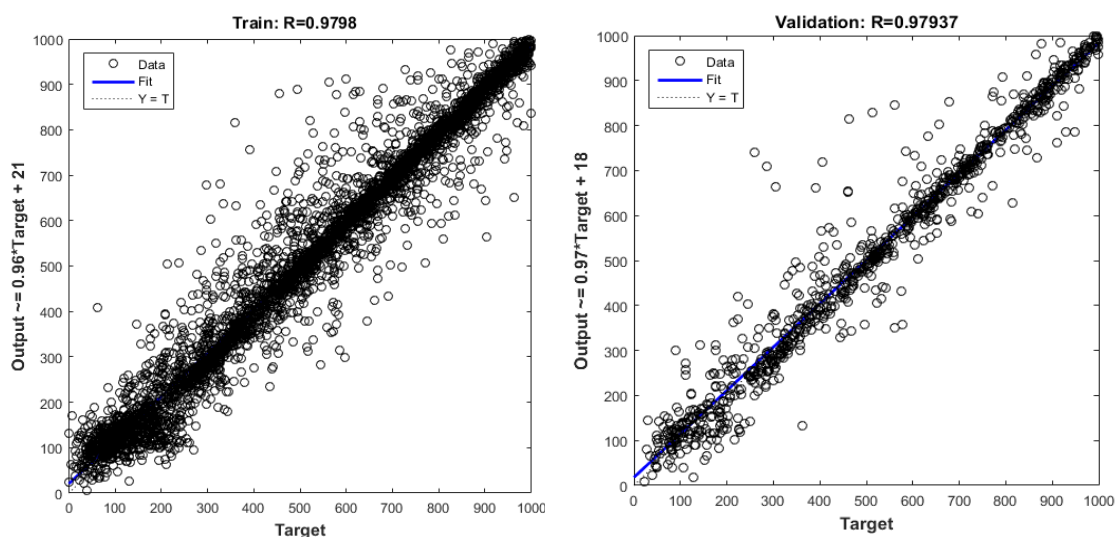


Figure 4. 6. Régressions obtenues pour la ville de Zagora en utilisant le modèle NARX pour prédire l'irradiation globale horaire à l'horizon $h+1$ pour l'ensemble de données d'entraînement et de validation.

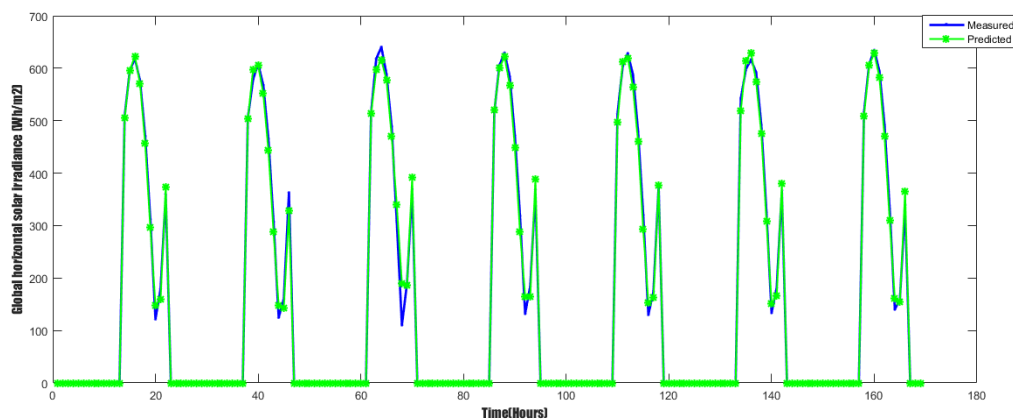


Figure 4. 7. Valeurs horaires prédites et mesurées de GHI pour une semaine choisie aléatoirement en utilisant le modèle NARX à Zagora ($h+1$).

Chapitre 4 : Prédiction à différents horizons temporels de l'irradiation solaire globale sur plan horizontal.

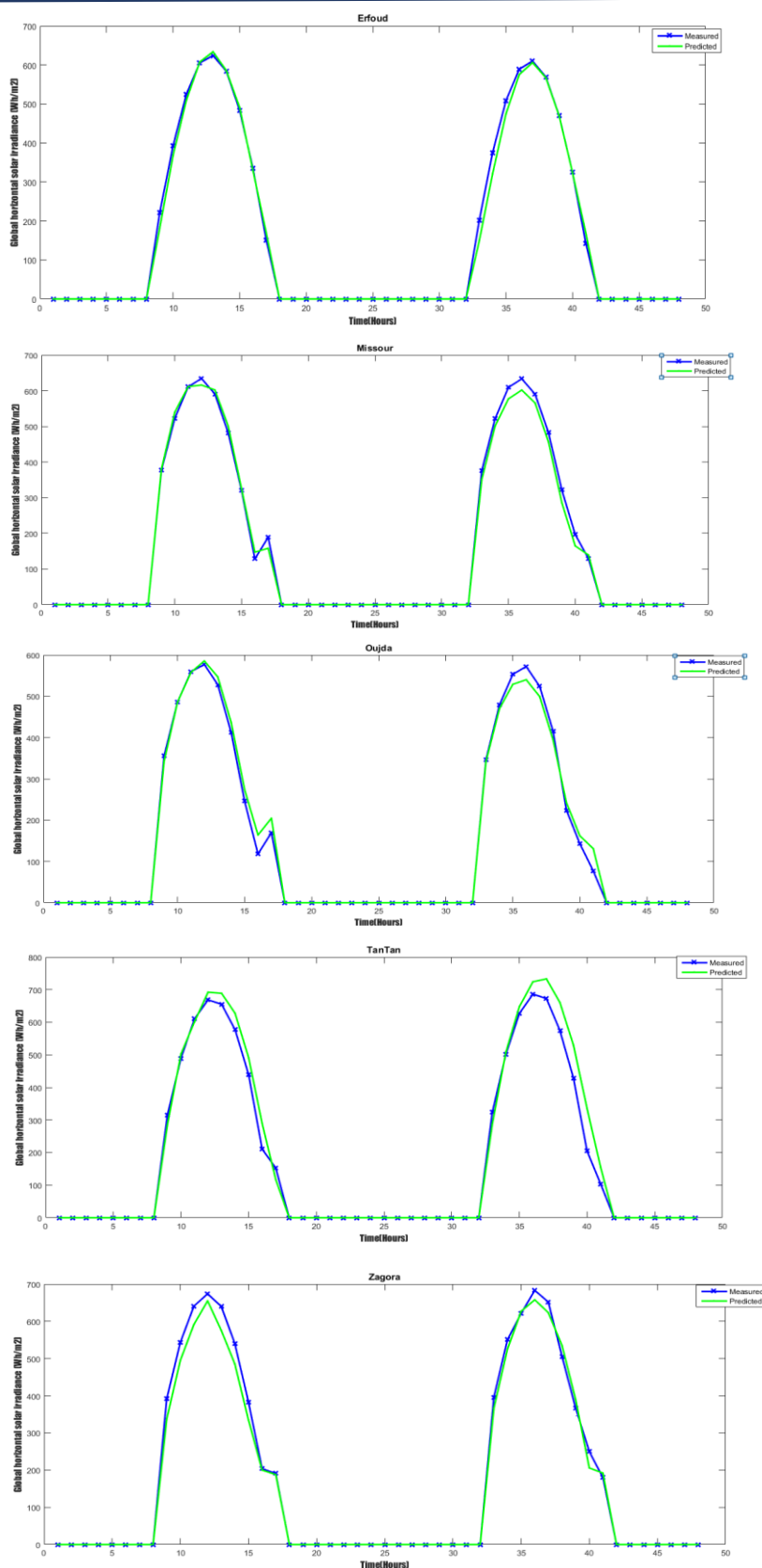


Figure 4. 8. Prévisions horaires de GHI pour deux jours à l'avance ($h+48$) par rapport aux données mesurées du 1 et 2 janvier 2015 pour chaque site enerMENA.

L'efficacité du modèle NARX a été aussi évaluée pour la prédiction à long terme de l'irradiation global horaire aux différents sites enerMENA. Pour cela, les modèles NARX développés précédemment pour chaque base de données ont été utilisés et testés pour prédire deux jours à l'avance (h+48) la GHI au début de l'année 2015. Les prévisions horaires obtenues pour chaque base de données sont présentées dans la figure 4.8. Pour toutes les stations, on remarque que les valeurs horaires prédites suivent précisément les valeurs horaires mesurées.

4.3. Prédiction à l'avance des GHI horaires et sub-horaires à l'aide de la nouvelle approche d'apprentissage profond

Les travaux de recherche sur les techniques d'apprentissage profond pour prévoir le rayonnement solaire à court et à très court terme (en particulier les horizons sub-horaires) sont encore rares et il est nécessaire de comparer leurs précisions à celle des méthodes d'apprentissage automatique. Dans ce sens, nous proposons une nouvelle approche d'apprentissage profond, à savoir les réseaux de neurones à longue mémoire à court terme (LSTM) pour la prévision horaire et sub-horaire de la GHI à la ville d'Erfoud, capitale de la région du Tafilalet au Maroc. L'objectif de ce travail est d'explorer le potentiel des modèles LSTMs dans la prédiction simultanée de GHI à des horizons à court terme, de 1 à 6 heures et à très court terme de 10 à 60 minutes, en se basant uniquement sur des données historiques endogènes (valeurs mesurées précédentes de GHI). Dans cette étude, deux scénarios ont été proposés, le premier scénario appelé « Performances annuelles » dans lequel les phases d'entraînement et de validation sont effectuées en utilisant la base de données entière de la ville considérée, quelles que soient les conditions météorologiques tout au long de l'année. Le second scénario nommé « Performances saisonnières » dans lequel les phases d'entraînement et de validation sont effectuées en utilisant, pour chaque ville, quatre bases de données partitionnées en fonction de chaque saison de l'année, afin de distinguer les impacts des conditions météorologiques caractérisant chaque saison. Enfin, une étude comparative est menée avec les deux techniques populaires d'apprentissage automatique, à savoir le réseau de neurones artificiel à perceptron multicouches (PMC) et la méthode d'ensemble des forêts aléatoires (Random Forest (RF)), afin d'évaluer les performances des modèles LSTMs développés pour chaque horizon temporel. Pour les deux scénarios proposés, cette comparaison est présentée comme l'écart de performances en termes de trois indicateurs statistiques (Annexe D). On donnera, aussi, une comparaison illustrative des tracés des séries temporelles des GHI horaire et sub-horaire mesurées et prédites pour l'ensemble de validation, pour les modèles PMC, RF

et LSTM et pour des horizons temporels allant de 60min (10min+60) à 6h (h+6) (Bounoua and Mechaqrane, 2021b).

4.3.1. Sélection des données d'entrée

Dans ce travail, les valeurs futures des GHI horaires et sub horaires sont prédites simultanément à différents horizons temporels, de h+1 à h+6 et de 10min+10 à 10min+60, en utilisant les valeurs mesurées antérieures de GHI, comme expliqué mathématiquement par l'équation suivante :

$$\hat{Y}_{t+h} = f(Y_t, Y_{t-1}, Y_{t-2}, \dots, Y_{t-n}) \quad (4-7)$$

\hat{Y}_{t+h} , présente les valeurs prévisionnelles simultanées de GHI à un horizon temporel (t+h), basées sur les données observées de Y mesurées aux instants (t, t-1, t-2, ..., t-n). Par exemple, pour un horizon temporel h+1, on prévoit 1 heure à l'avance le GHI, pour h+2, on prévoit simultanément 1 et 2h à l'avance le GHI, ... et pour h+6, on prévoit simultanément 1, 2, 3, 4, 5, et 6h à l'avance le GHI. De même, pour l'horizon sub-horaire 10min+10, on prévoit 10 minutes à l'avance le GHI, pour 10min+20, on prévoit simultanément 10 et 20 minutes à l'avance le GHI, ... et pour 10min+60, on prévoit simultanément 10, 20, 30, 40, 50 et 60 minutes à l'avance le GHI.

Pour sélectionner le nombre pertinent de valeurs historiques endogènes à utiliser comme entrées pour prédire les valeurs futures de GHI, une analyse de la fonction d'autocorrélation (ACF) est utilisée ((Alsharif et al., 2019), (Madugu et al., 2019)). Cette fonction donne les valeurs d'autocorrélation de toute série avec ses valeurs décalées. Elle décrit, tout simplement, à quel point la valeur actuelle de la série est liée à ses valeurs passées, en calculant le degré de corrélation entre Y_{t+h} et Y_{t-i} avec i variant de 0 à n.

La figure 4.9-a présente les coefficients de corrélation de 10 délais des entrées endogènes avec les valeurs futures du GHI horaire. On peut observer que les 5 premiers délais sont positivement corrélés avec le GHI futur, tandis que les 5 autres délais sont négativement corrélés avec le GHI futur. Cependant, seuls les 3 premiers délais ont un bon accord par rapport aux autres délais (coefficient de corrélation supérieur à 0.6).

La figure 4.9-b montre les coefficients de corrélation des 10 délais des entrées endogènes avec les valeurs futures du GHI sub-horaire. On remarque que les valeurs futures du GHI sub-horaire

sont corrélées avec un grand nombre de leurs valeurs passées par rapport aux observations horaires de la figure 4.9-a. Cela explique que la variabilité entre les mesures sub-horaires (toutes les 10 minutes) est relativement faible par rapport aux observations horaires. Néanmoins, il suffit d'utiliser seulement 6 délais avec un coefficient de corrélation supérieur à 0.9 pour prédire le GHI sub-horaire.

Sur la base des résultats de l'analyse ACF, il est recommandé d'utiliser 3 entrées et 6 entrées (ou délais) pour prédire le GHI horaire et sub-horaire, respectivement.

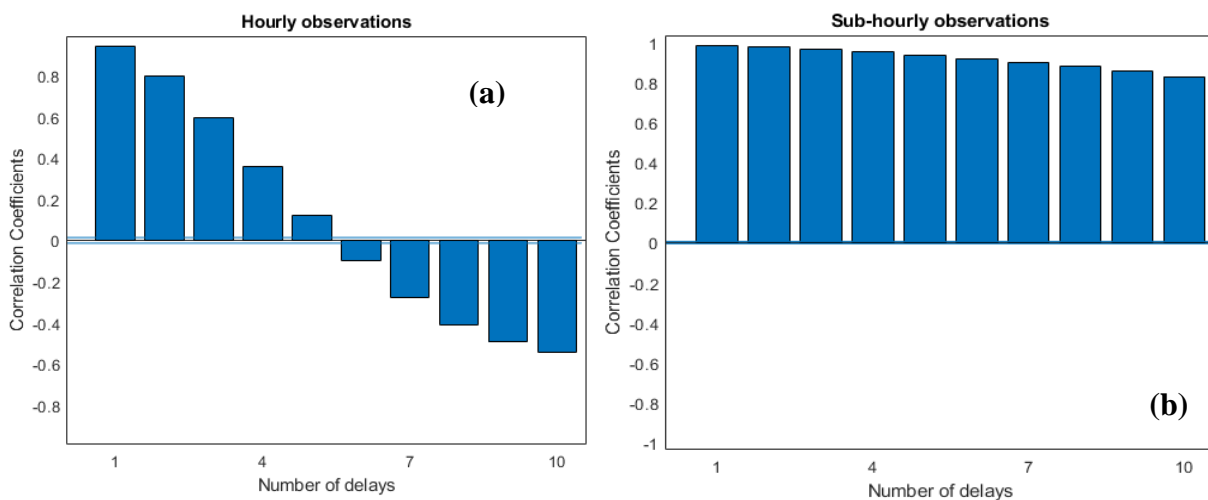


Figure 4. 9. Corrélation des valeurs passées des entrées endogènes avec les valeurs futures horaires (a) et sub-horaires (b) de GHI.

4.3.2. Paramétrage des modèles et évaluation de la précision

Dans cette étude (Bounoua and Mechaqrane, 2021b), nous avons comparé les performances du modèle LSTM proposé pour prédire à l'avance les valeurs horaires et sub-horaires de GHI avec celles des deux modèles d'apprentissage automatique, à savoir le réseau de neurone artificiel à Perceptron MultiCouches (PMC) et la méthode d'ensemble des forêts aléatoires (RF). Les paramètres optimisés de chaque algorithme sont décrits ci-dessous :

- Pour le présent PMC, le nombre optimal de neurones cachés a été choisi et fixé à 20 après avoir effectué plusieurs tests (de 1 à 50) et les performances de chaque architecture ont été évaluées sur cinq exécutions.

Chapitre 4 : Prédiction à différents horizons temporels de l'irradiation solaire globale sur plan horizontal.

- Dans ce travail, les principaux paramètres du modèle RF sont fixés comme suit : le nombre d'arbres dans la forêt (nombre d'apprenants) dépend de chaque modèle RF utilisé et se situe principalement entre 100 et 150. Le nombre minimum d'observations par une feuille d'arbre (MinLeafSize) est égal à 5 et le nombre de prédicteurs à sélectionner au hasard pour chaque division est choisi égal à 2.
- L'architecture du modèle d'apprentissage profond LSTM est fixée à trois couches, à savoir une couche des entrées décalées par N pas de temps sélectionnés par l'analyse ACF dans la section précédente, une couche cachée, dans laquelle le nombre de neurones est fixé à 50 après un certain nombre de tests et une couche de sortie avec un neurone linéaire. L'erreur quadratique moyenne (RMSE) est choisie comme fonction de perte de la couche de régression. En outre, le solveur de réseau d'entraînement appelé Adam est sélectionné comme optimiseur et le nombre maximal d'itérations a été fixé à 250. Les hyper-paramètres du modèle LSTM que nous avons optimisés sont résumés dans le tableau 4.5.

Tableau 4. 5. Optimisation des hyper-paramètres du modèle LSTM.

Hyper paramètre	Plage de recherche	Valeur
Solveur d'optimisation	adam, sgd, rmsprop	Adam
Nombre de couches	1, 2, 3, 4, 5	1
neurones cachés/couche	10-100	50
Nombre d'itérations	100-500	250
Taux d'apprentissage	0.1, 0.01, 0.005, 0.001, 0.0005	0.005

Avant d'appliquer tous les modèles précédents, 20197 observations horaires et 120816 observations sub-horaires ont été réparties à l'aide d'une méthode de partition de base (sans utiliser la validation croisée). En effet, les modèles LSTM, PMC et RF sont exécutés 10 fois sur les mêmes données avec une initialisation aléatoire des poids, afin d'assurer une comparaison équitable. Cette partition permet de diviser les observations en deux ensembles, un ensemble d'entraînement et un ensemble de validation. Pour cela, 70% des observations sont utilisées pour la phase d'entraînement et les 30% restantes pour la phase de validation. Aussi, toutes les observations sont normalisées entre [-1, 1]. La précision de prédiction de tous les

modèles développés est évaluée à l'aide de trois indicateurs statistiques : le coefficient de corrélation (R), l'erreur absolue moyenne normalisée (nMAE) et l'erreur quadratique moyenne normalisée (nRMSE) (Annexe D).

4.3.3. Prévisions horaires du GHI

4.3.3.1. Scénario 1 : Performances annuelles

Pour chaque horizon temporel, les performances annuelles de l'ensemble de validation pour la prédiction horaire du GHI en appliquant les deux modèles d'apprentissage automatique (PMC et RF) et le modèle d'apprentissage profond (LSTM) sont présentés dans le tableau 4.6. Les meilleures performances obtenues sont indiquées en chiffres gras pour chaque horizon et chaque indicateur statistique. D'après ces résultats, on peut observer que tous les modèles développés (PMC, RF et LSTM) donnent une bonne précision pour l'ensemble de données de validation. Comme le montre la figure 4.10, les nRMSEs obtenus se situent dans une marge de fiabilité ($10 < nRMSE < 23\%$) pour chaque horizon temporel (de h+1 à h+6). Les erreurs statistiques obtenues diffèrent d'un modèle à l'autre. Les valeurs de R varient entre 96.40% et 89.06% pour le PMC, entre 96.55% et 90.33% pour le RF et entre 97.35% et 94.10% pour le LSTM, pour les horizons temporels allant de h+1 à h+6. Néanmoins, il est clair que le meilleur modèle pour prédire à l'avance le GHI horaire est le LSTM pour chaque horizon temporel (meilleures erreurs statistiques). La figure 4.11 présente les diagrammes de dispersion des valeurs mesurées et prédites de GHI de l'ensemble de validation pour h+1 (a), h+2 (b), h+3 (c), h+4 (d), h+5 (e) et h+6 (f) en utilisant les modèles LSTMs. La valeur de R obtenue pour chaque horizon temporel est présentée au-dessus de chaque graphique.

Tableau 4. 6. Performances annuelles des modèles pour la prédiction à l'avance des GHI horaires pour l'ensemble de validation. Les chiffres en gras représentent le meilleur prédicteur pour chaque horizon et chaque indicateur statistique.

Modèle	Indicateurs statistiques (%)	Horizons temporels					
		h+1	h+2	h+3	h+4	h+5	h+6
PMC	R	96.40	94.52	93.01	91.40	90.52	89.06
	nMAE	8.69	11.38	12.80	14.65	15.19	16.79
	nRMSE	13.58	16.47	18.43	20.34	21.21	22.70
RF	R	96.55	95.15	93.72	92.41	91.32	90.33
	nMAE	8.69	10.73	12.37	13.68	14.61	15.45
	nRMSE	13.33	15.70	17.63	19.21	20.44	21.51
LSTM	R	97.35	96.36	95.79	95.14	94.46	94.10
	nMAE	7.30	8.94	9.46	10.12	10.86	11.47
	nRMSE	11.63	13.50	14.66	15.83	16.89	17.35

D'après le tableau 4.6, le modèle LSTM surpasse les modèles PMC et RF dans la prédiction de l'irradiation solaire globale horizontale pour tous les horizons horaires, en notant que la plus faible valeur du nRMSE est obtenue pour h+1 (nRMSE = 11.63%). Pour h+1, le modèle LSTM a réduit les valeurs du nRMSE des modèles PMC et RF de 14.35% et 12.75%, respectivement. Aussi, les performances des modèles PMC et RF diminuent significativement avec l'horizon temporel par rapport aux modèles LSTM. Nous notons que l'écart entre les nRMSE des deux modèles d'apprentissage automatique (PMC et RF) et du modèle d'apprentissage profond (LSTM) augmente avec l'horizon temporel de 14.35% et 12.75% à h+1 à 23.56% et 19.33% à h+6 respectivement pour les modèles PMC et RF.

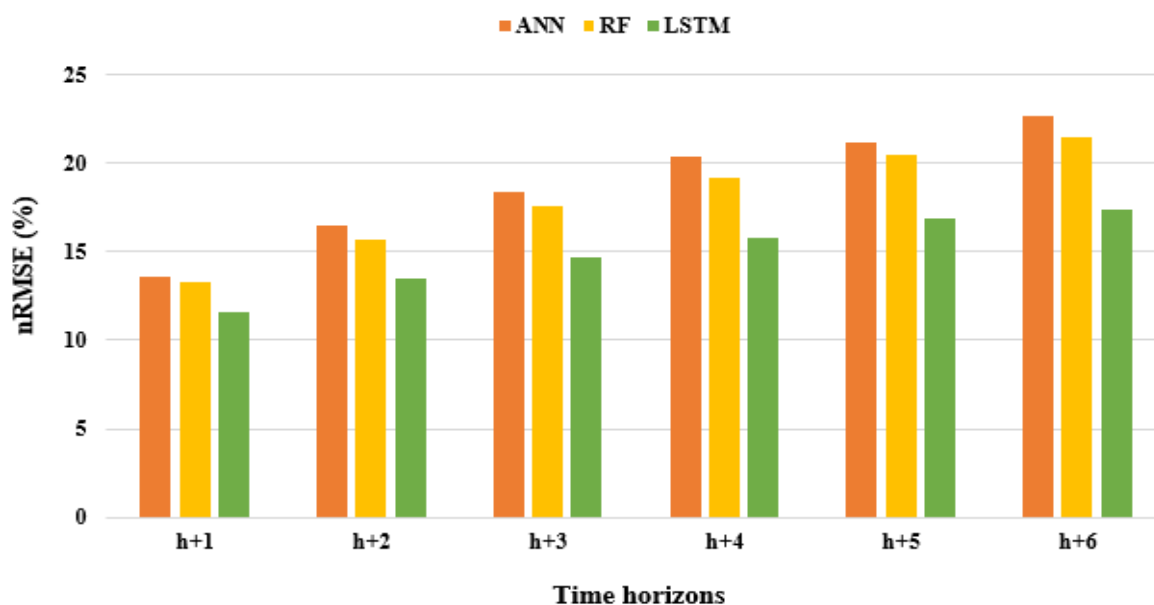


Figure 4. 10. Comparaison entre les performances annuelles des modèles de prédiction à l'avance du GHI en termes de nRMSE pour les différents horizons horaires.

Pour l'ensemble de validation, la figure 4.12 montre le tracé des séries temporelles de 77 heures (7 jours) mesurées et prédites de GHI correspondantes aux modèles PMC, RF et LSTM pour un horizon temporel allant jusqu'à 6h. On peut noter que tous les modèles développés donnent de bonnes performances sur les ensembles de données de validation avec une performance significativement meilleure du modèle LSTM qui suit de près les valeurs mesurées par rapport aux autres modèles (PMC et RF).

Chapitre 4 : Prédiction à différents horizons temporels de l'irradiation solaire globale sur plan horizontal.

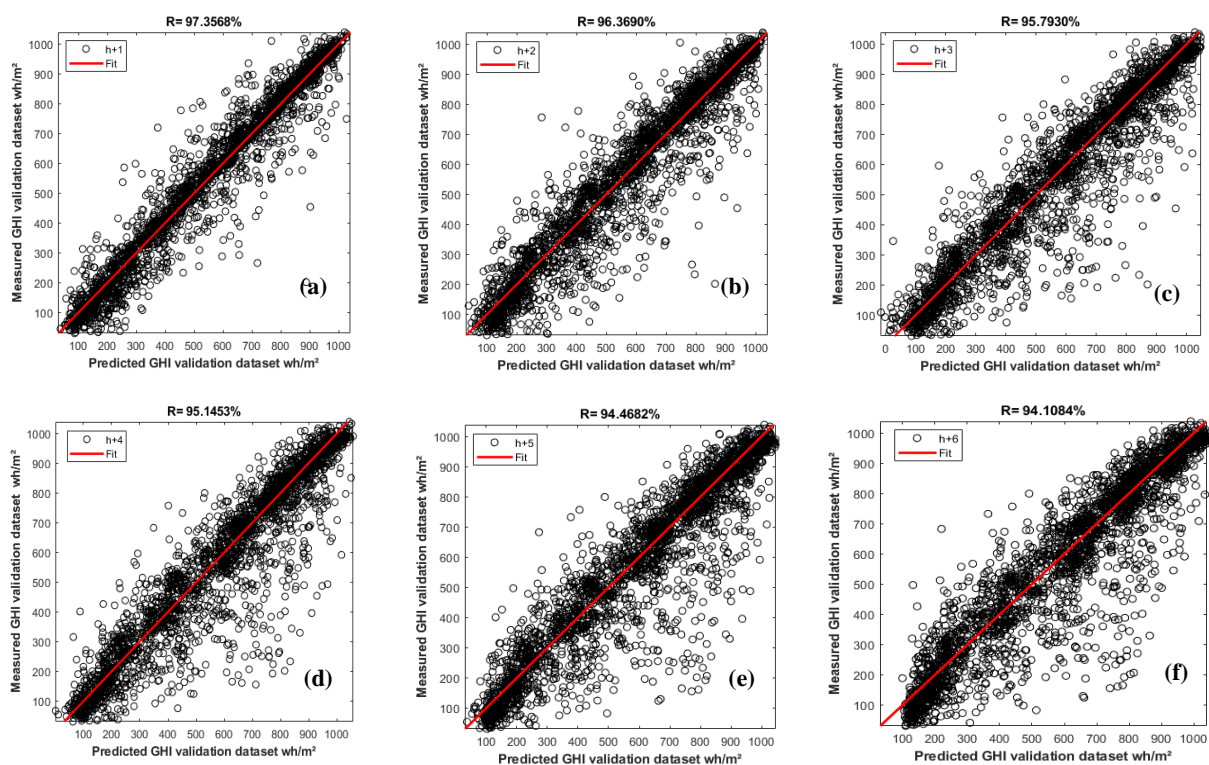


Figure 4. 11. Nuages de points des valeurs mesurées et prédites de GHI pour $h+1$ (a), $h+2$ (b), $h+3$ (c), $h+4$ (d), $h+5$ (e) et $h+6$ (f) en utilisant les modèles LSTMs, pour l'ensemble de validation.

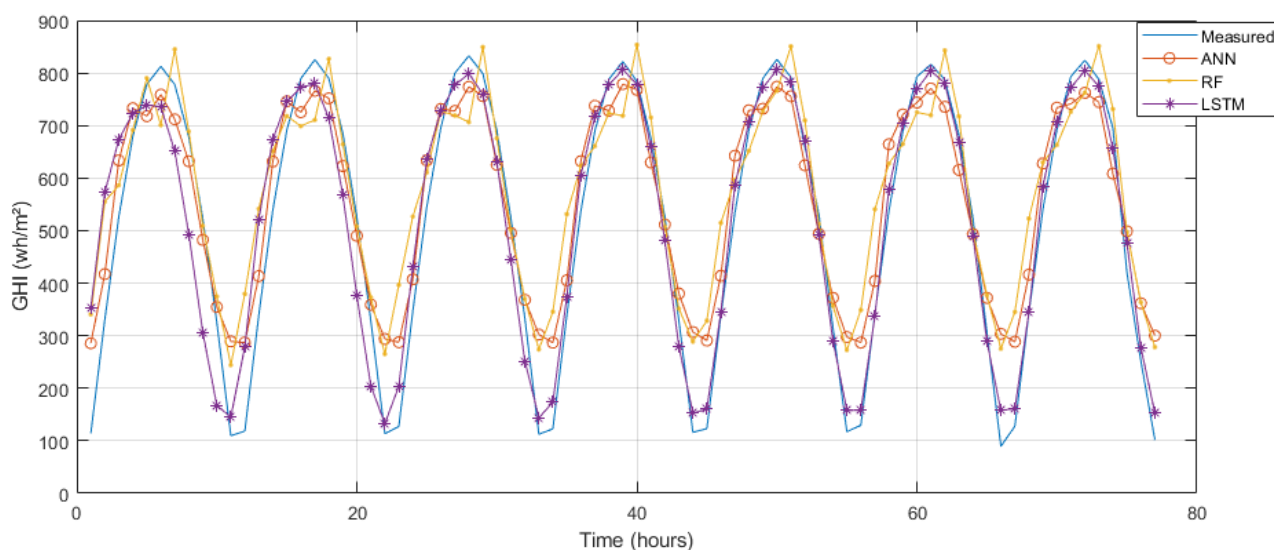


Figure 4. 12. Tracé des séries temporelles de 77 heures (7 jours) mesurées et prédites de GHI correspondantes aux modèles ANN, RF et LSTM pour l'ensemble de validation ($h+6$).

4.3.3.2. Scénario 2 : Performances saisonnières

Pour évaluer l'impact de la variabilité des conditions météorologiques sur la précision des prédictions, on a jugé nécessaire d'examiner les performances des modèles développés saison par saison. Dans ce but, le tableau 4.7 résume les performances saisonnières de la prédiction à l'avance des irradiations globales horizontales horaires pour la phase de validation. Les chiffres en gras présentent le meilleur prédicteur pour chaque horizon et chaque indicateur statistique. En général, le LSTM est le modèle le plus performant quel que soit l'horizon temporel. L'écart en terme de nRMSE entre PMC et LSTM passe de 4.93 % à h+1 à 22.13% à h+6 pour l'hiver, de 10.34% à h+1 à 18.77% à h+6 pour le printemps, de 12.32% à h+1 à 18.84% à h+6 pour l'été et de 8.11% à h+1 à 16.63% à h+6 pour l'automne. De même, le LSTM réduit le nRMSE du modèle RF de 7.02% à h+1 à 24.01% à h+6 pour l'hiver, de 7.33% à h+1 à 14.84% à h+6 pour le printemps, de 8.52% à h+1 à 13.83% à h+6 pour l'été et de 9.46% à h+1 à 16.42% à h+6 pour l'automne.

La figure 4.13 montre que les meilleures performances sont obtenues au printemps avec des valeurs du nRMSE variant entre 9.86 et 16.88% pour le PMC, 9.54 et 16.10% pour le RF et 8.84 et 13.71% pour le LSTM, pour l'horizon temporel allant de h+1 à h+6. Suivi par les résultats d'hiver et d'été avec des valeurs du nRMSE variant respectivement entre 11.55 et 22.72%, 14.93 et 21.49% pour le PMC, 11.81 et 23.28%, 14.31 et 20.24 pour le RF, et 10.98 et 17.69%, 13.09 et 17,44% pour le LSTM, pour l'horizon temporel allant de h+1 à h+6. Les plus faibles performances sont obtenues à l'automne avec des valeurs du nRMSE variant entre 21.56 et 36.61% pour le PMC, 21.88 et 36.52% pour le RF et 19.81 et 30.52% pour le LSTM, pour l'horizon temporel allant de h+1 à h+6.

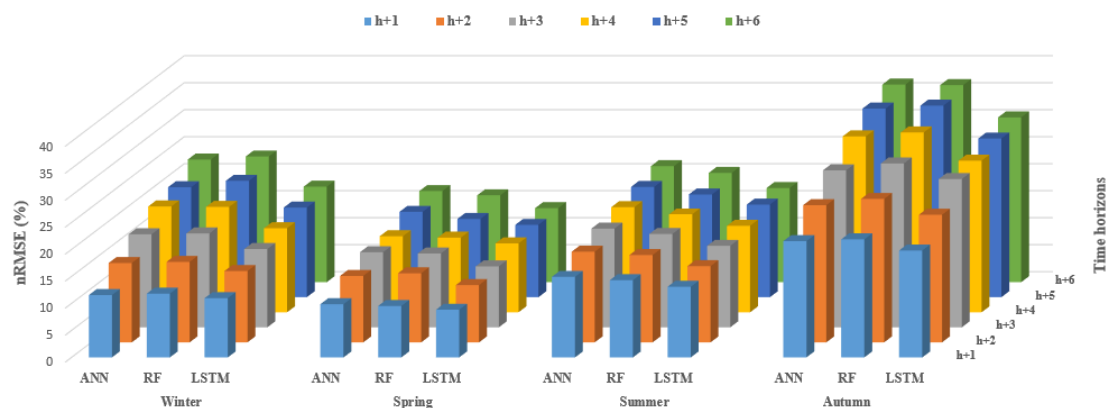


Figure 4.13. Comparaison entre les performances saisonnières des modèles de prédiction à l'avance de GHI en termes du nRMSE pour les différents horizons horaires.

Chapitre 4 : Prédiction à différents horizons temporels de l'irradiation solaire globale sur plan horizontal.

Pour l'ensemble de validation, la figure 4.14 montre les variabilités horaires mesurées et prédites (7 jours) de GHI correspondantes aux modèles PMC, RF et LSTM pour chaque saison et pour un horizon temporel allant jusqu'à 6h (h+6). D'après cette figure, on peut conclure que l'irradiation solaire en automne est plus difficile à prévoir que les autres saisons, en raison de la grande variabilité météorologique observée pendant cette saison.

Tableau 4. 7. Performances saisonnières de la prédiction à l'avance des GHI pour l'ensemble de validation. Les chiffres en gras représentent le meilleur prédicteur pour chaque horizon et chaque indicateur statistique.

Modèle	Indicateurs statistiques (%)	Horizons temporels					
		h+1	h+2	h+3	h+4	h+5	h+6
a) Hiver							
PMC	R	97.07	95.30	93.33	91.20	90.43	88.00
	nMAE	8.34	10.06	12.77	15.02	14.85	17.09
	nRMSE	11.55	14.70	17.24	19.61	20.38	22.72
RF	R	97.04	95.54	93.52	91.53	89.40	87.34
	nMAE	8.59	11.07	13.22	14.94	16.47	17.61
	nRMSE	11.81	14.92	17.40	19.54	21.57	23.28
LSTM	R	97.21	95.91	95.04	94.53	93.80	93.19
	nMAE	7.38	9.00	10.18	16.95	11.95	11.90
	nRMSE	10.98	13.21	14.51	15.58	16.62	17.69
b) Printemps							
PMC	R	98.00	96.73	95.80	95.56	94.41	93.70
	nMAE	6.37	8.37	10.16	9.89	11.55	12.75
	nRMSE	9.86	12.32	13.91	14.09	15.82	16.88
RF	R	98.10	96.41	95.85	95.83	95.39	94.23
	nMAE	6.21	8.32	9.46	9.79	10.43	11.64
	nRMSE	9.54	12.81	13.68	13.83	14.51	16.10
LSTM	R	98.34	97.64	97.30	96.42	96.04	96.03
	nMAE	5.49	7.27	7.65	9.19	9.79	10.03
	nRMSE	8.84	10.62	11.30	12.76	13.40	13.71
c) Été							
PMC	R	95.86	94.63	93.57	92.65	91.82	91.02
	nMAE	9.51	11.28	12.70	13.45	14.82	15.77
	nRMSE	14.93	16.80	18.29	19.48	20.44	21.49
RF	R	96.19	95.03	94.28	93.64	92.97	92.35
	nMAE	9.19	10.73	11.77	12.70	13.53	14.02
	nRMSE	14.31	16.19	17.29	18.17	19.04	20.24
LSTM	R	96.84	96.27	95.78	95.18	94.46	94.25
	nMAE	8.27	9.61	10.90	11.51	12.45	12.87
	nRMSE	13.09	14.12	15.10	16.03	17.16	17.44
d) Automne							
PMC	R	92.20	88.51	84.59	79.77	76.79	74.58
	nMAE	14.93	18.33	21.24	23.53	25.95	26.80
	nRMSE	21.56	25.42	29.08	32.60	34.95	36.61
RF	R	91.98	87.38	82.75	78.81	75.70	74.78
	nMAE	15.40	19.04	22.16	24.42	26.12	26.40
	nRMSE	21.88	26.58	30.36	33.32	35.50	36.52
LSTM	R	93.41	90.07	86.18	85.04	84.65	84.02
	nMAE	13.69	16.74	20.01	21.20	22.35	23.51
	nRMSE	19.81	23.65	27.48	28.10	29.40	30.52

Chapitre 4 : Prédiction à différents horizons temporels de l'irradiation solaire globale sur plan horizontal.

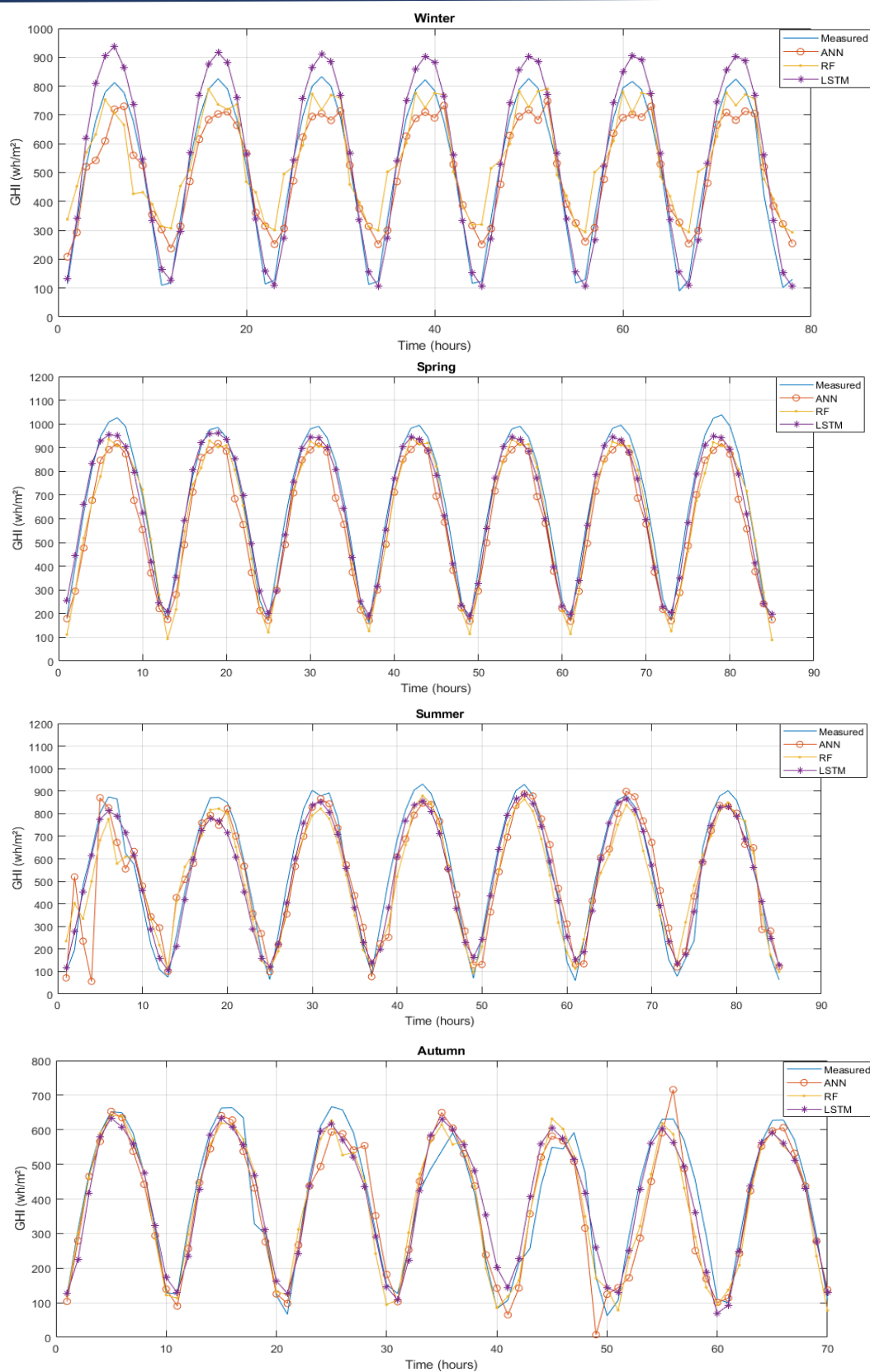


Figure 4. 14. Variabilités horaires mesurées et prédites (7 jours) de GHI correspondantes aux modèles PMC, RF et LSTM pour chaque saison et pour un horizon temporel allant jusqu'à 6h ($h+6$).

4.3.4. Prévisions sub-horaires du GHI

4.3.4.1. Scénario 1 : Performances annuelles

Pour l'ensemble de validation, le tableau 4.8 résume les résultats de la prédiction à l'avance des GHI sub-horaires à l'aide des deux modèles d'apprentissage automatique (PMC et RF) et du modèle d'apprentissage profond (LSTM). Pour chaque horizon temporel, les chiffres en gras présentent les meilleures erreurs statistiques obtenues. En analysant ces résultats, on peut voir que les meilleurs résultats sont obtenus par le modèle LSTM malgré que quelques valeurs du nMAE obtenues par le modèle RF sont très légèrement meilleures.

D'après le tableau 4.8, les performances de tous les modèles (PMC, RF et LSTM) en termes de valeurs du nRMSE sont très fiables ($11\% < nRMSE < 17\%$). Comme le montre la figure 4.15, les valeurs du nRMSE varient entre 11.73% et 16.79% pour le PMC, 11.58% et 16.28% pour le RF et 11.28% et 14.88% pour le LSTM, pour un horizon temporel allant de 10min+10 à 10min+60.

Tableau 4. 8. Performances annuelles pour la prédiction à l'avance des GHI sub-horaires pour l'ensemble de validation. Les chiffres en gras représentent le meilleur prédicteur pour chaque horizon et chaque indicateur statistique.

Modèle	Indicateurs statistiques (%)	Horizons temporels					
		10min+10	10min+20	10min+30	10min+40	10min+50	10min+60
PMC	R	96.72	95.80	95.09	94.18	93.68	92.90
	nMAE	5.80	6.84	7.61	8.42	8.96	9.51
	nRMSE	11.73	13.14	14.16	15.28	15.91	16.79
RF	R	96.79	95.91	95.20	94.56	93.95	93.36
	nMAE	5.58	6.53	7.26	7.89	8.50	9.05
	nRMSE	11.58	12.98	14.00	14.84	15.60	16.28
LSTM	R	96.96	96.22	95.83	95.42	94.97	94.56
	nMAE	6.00	6.89	7.56	8.10	8.53	8.97
	nRMSE	11.28	12.51	13.10	13.74	14.33	14.88

Encore une fois, ces résultats montrent la supériorité du LSTM dans la prévision à l'avance du GHI pour chaque horizon sub-horaire. En particulier, lorsque l'horizon temporel augmente, le modèle LSTM se distingue plus des modèles PMC et RF. En effet, les améliorations des performances en termes des valeurs du nRMSE entre les deux modèles d'apprentissage automatique (PMC et RF) et le LSTM augmentent avec l'horizon temporel de 3.83% et 2.59% à 10min+10 à 11.37% et 8.59% à 10min+60 pour le PMC et le RF respectivement.

Chapitre 4 : Prédiction à différents horizons temporels de l'irradiation solaire globale sur plan horizontal.

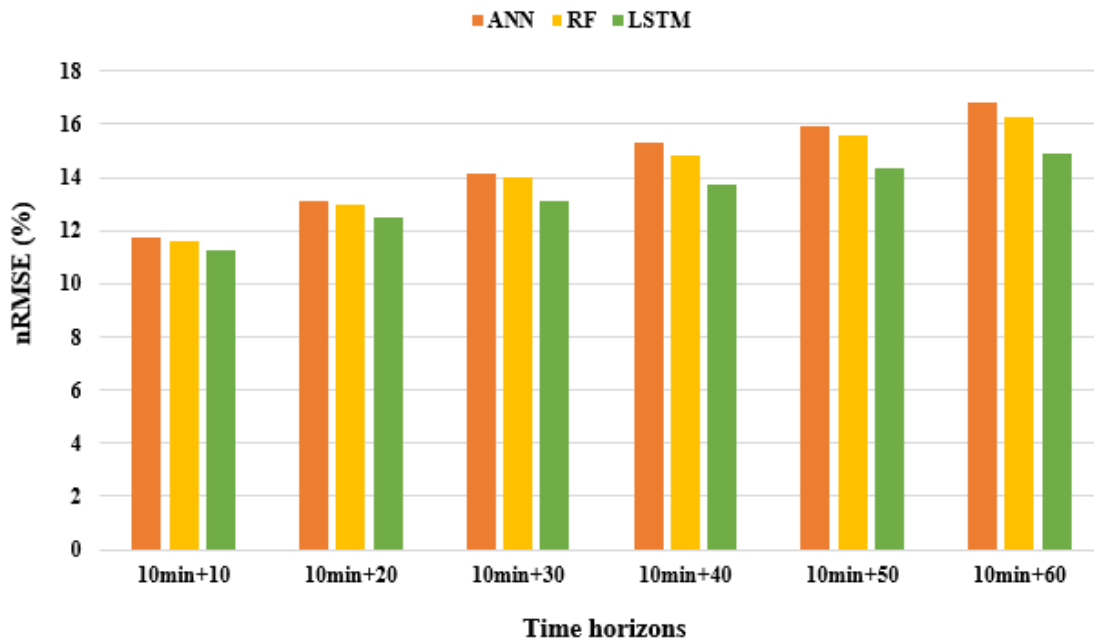


Figure 4. 15. Comparaison entre les performances annuelles des modèles de prédiction à l'avance du GHI en termes du nRMSE pour les différents horizons sub-horaires.

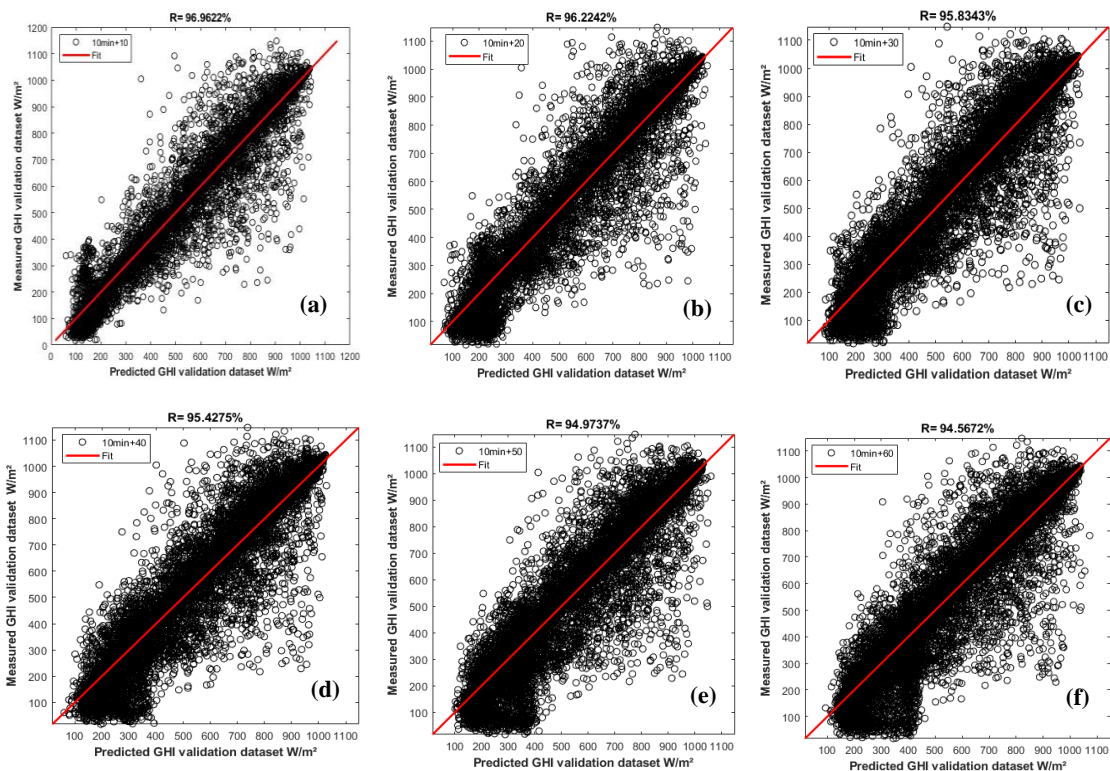


Figure 4. 16. Nuages de points des valeurs mesurées et prédites de GHI pour 10min+10 (a), 10min+20 (b), 10min+30 (c), 10min+40 (d), 10min+50 (e) et 10min+60 (f) en utilisant le modèle LSTM, pour l'ensemble de validation.

La figure 4.16 illustre les diagrammes de dispersion des valeurs mesurées et prédites de GHI pour 10min+10 (a), 10min+20 (b), 10min+30 (c), 10min+40 (d), 10min+50 (e) et 10min+60 (f) en utilisant le modèle LSTM, pour l'ensemble de validation. On constate que les valeurs du coefficient de corrélation (R) du modèle LSTM diminuent avec l'horizon temporel, en passant de 96.96 % à 10min+10 à 94.56 % à 10min+60.

La figure 4.17 montre les séries temporelles de 500 valeurs sub-horaires mesurées et prédites de GHI correspondantes aux modèles PMC, RF et LSTM pour un horizon temporel de 10min+60. On peut observer que le modèle LSTM suit de près les variations de l'irradiation globale quels que soient les changements météorologiques par rapport aux modèles PMC et RF. En revanche, les modèles PMC et RF sont très précis pour des conditions météorologiques stables.

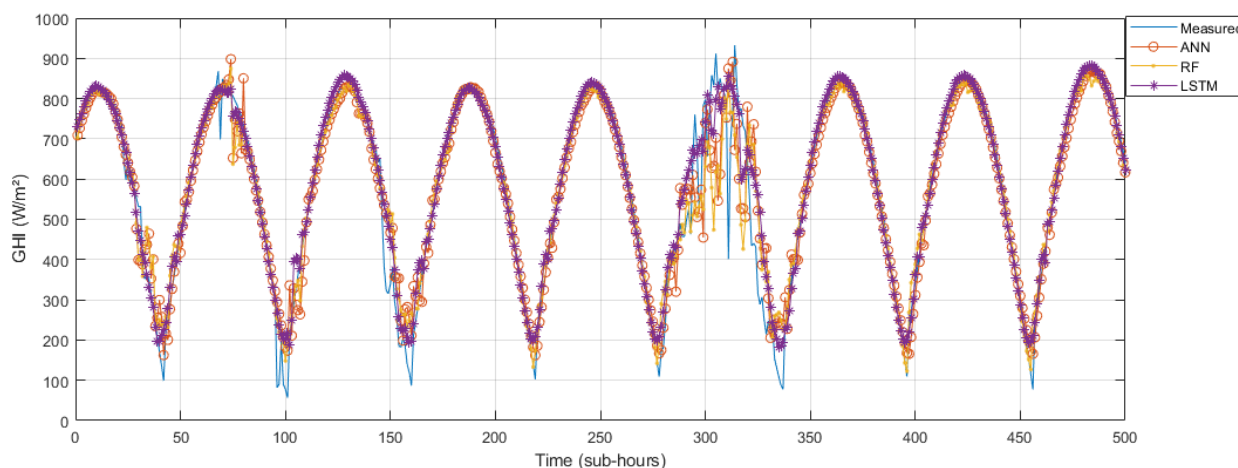


Figure 4. 17. *Séries temporelles de 500 valeurs sub-horaires mesurées et prédites de GHI correspondantes aux modèles PMC, RF et LSTM pour l'ensemble de validation (10min+60).*

4.3.4.2. Scénario 2 : Performances saisonnières

Les performances saisonnières de la prévision à l'avance des GHI sub-horaires pour la phase de validation sont présentées dans le tableau 4.9. Les chiffres en gras indiquent les meilleurs indicateurs statistiques obtenus pour chaque horizon temporel.

Comme pour les performances annuelles, les meilleurs résultats sont obtenus par le modèle LSTM malgré que certaines valeurs du nMAE obtenues par le modèle RF sont légèrement meilleures. Aussi, il a été remarqué que pour les prédictions sub-horaires du GHI, que ce soit en performance annuelle ou saisonnière, la précision diminue lorsque l'horizon temporel augmente. En effet, les valeurs du nRMSE des modèles LSTM passent de 10.10% pour

Chapitre 4 : Prédiction à différents horizons temporels de l'irradiation solaire globale sur plan horizontal.

10min+10 à 13.88% pour 10min+60 pour l'hiver, de 9.13% pour 10min+10 à 11.51% pour 10min+60 pour le printemps, de 13.21% pour 10min+10 à 17.69% pour 10min+60 pour l'été et de 15.96% pour 10min+10 à 23.83% pour 10min+60 pour l'automne.

Tableau 4. 9. Performances saisonnières de la prédiction à l'avance des GHI sub-horaires pour l'ensemble de validation. Les chiffres en gras représentent le meilleur prédicteur pour chaque horizon et chaque indicateur statistique.

Modèle	Indicateurs statistiques (%)	Horizons temporels					
		10min+10	10min+20	10min+30	10min+40	10min+50	10min+60
a) Hiver							
PMC	R	96.62	95.69	94.33	94.03	93.27	92.48
	nMAE	5.60	6.56	6.94	7.77	8.55	9.05
	nRMSE	10.83	12.20	13.76	14.10	14.92	15.76
RF	R	96.82	95.96	95.28	94.62	94.11	93.61
	nMAE	4.75	5.38	6.17	6.77	7.30	7.79
	nRMSE	10.54	11.80	12.69	13.48	14.10	14.66
LSTM	R	97.08	96.32	95.56	95.09	94.61	94.29
	nMAE	5.11	5.47	6.37	7.19	7.83	7.92
	nRMSE	10.10	11.24	12.26	12.87	13.46	13.88
b) Printemps							
PMC	R	97.55	96.92	96.52	96.17	95.78	95.32
	nMAE	4.44	5.35	5.57	5.89	6.59	6.87
	nRMSE	9.61	10.70	11.33	11.85	12.43	13.07
RF	R	97.70	97.16	96.72	96.36	96.04	95.71
	nMAE	4.12	4.72	5.23	5.63	6.04	6.41
	nRMSE	9.33	10.30	11.02	11.58	12.07	12.54
LSTM	R	97.80	97.33	97.06	96.85	96.64	96.45
	nMAE	4.13	5.41	5.30	5.64	6.43	6.80
	nRMSE	9.13	10.02	10.46	10.83	11.20	11.51
c) Été							
PMC	R	95.80	94.52	93.63	92.72	91.95	91.05
	nMAE	7.31	8.17	9.09	9.76	10.48	11.04
	nRMSE	13.59	15.39	16.50	17.55	18.37	19.33
RF	R	95.97	94.94	94.13	93.38	92.65	91.98
	nMAE	6.78	7.85	8.64	9.15	9.91	10.47
	nRMSE	13.33	14.80	15.87	16.77	17.59	18.32
LSTM	R	96.04	95.04	94.44	93.97	93.33	92.68
	nMAE	7.06	8.23	8.67	9.17	10.11	10.97
	nRMSE	13.21	14.70	15.53	16.13	16.92	17.69
d) Automne							
PMC	R	94.08	92.49	90.60	88.64	87.64	86.02
	nMAE	8.85	10.21	12.58	13.12	14.21	15.39
	nRMSE	16.58	18.51	20.84	22.53	23.33	24.46
RF	R	94.21	92.21	90.53	89.08	87.75	86.50
	nMAE	8.53	10.37	11.79	12.98	13.98	14.88
	nRMSE	16.42	18.87	20.67	22.11	23.32	24.39
LSTM	R	94.48	92.68	90.88	89.42	88.40	87.46
	nMAE	8.80	10.77	12.08	13.04	14.10	14.89
	nRMSE	15.96	18.19	20.33	21.85	22.77	22.80

Chapitre 4 : Prédiction à différents horizons temporels de l'irradiation solaire globale sur plan horizontal.

Ces résultats permettent de conclure que les modèles LSTM sont plus performants que les modèles PMC et RF pour tous les horizons temporels et toutes les saisons. Ainsi, les modèles LSTM réduisent les valeurs du nRMSE des modèles PMC et RF, respectivement, de 6.74% et 4.17% à 10min+10 et de 11.92% et 5.32% à 10min+60 pour l'hiver; de 4.99% et 2.14% à 10min+10 et de 11.93% et 8.21% à 10min+60 pour le printemps; de 2.79% et 0.90% à 10min+10 et de 8.48% et 3.43% à 10min+60 pour l'été et de 3.73% et 2.80% à 10min+10 et de 6.78% et 6.51% à 10min+60 pour l'automne.

La figure 4.18 montre le diagramme à barres des valeurs nRMSE pour les modèles PMC, RF et LSTM en fonction de chaque saison. Cette figure révèle que les valeurs du nRMSE obtenues par tous les modèles au printemps et en hiver sont meilleures qu'en été et en automne. Au printemps et en hiver, les valeurs du nRMSE varient respectivement entre 9.61%, 10.83% et 13.07%, 15.76% pour PMC, 9.33%, 10.54% et 12.54%, 14.66% pour RF et 9.13%, 10.10% et 11.51%, 13.88% pour LSTM, de 10min+10 à 10min+60. Pour l'été, les valeurs du nRMSE varient entre 13.59% et 19.33% pour PMC, 13.33% et 18.32% pour RF, 13.21% et 17.69% pour LSTM, de 10min+10 à 10min+60, respectivement. Les plus faibles précisions sont observées en automne, avec des valeurs du nRMSE variant entre 16.58% et 24.46% pour PMC, 16.42% et 24.39% pour RF, 15.96% et 22.80% pour LSTM, de 10min+10 à 10min+60, respectivement.

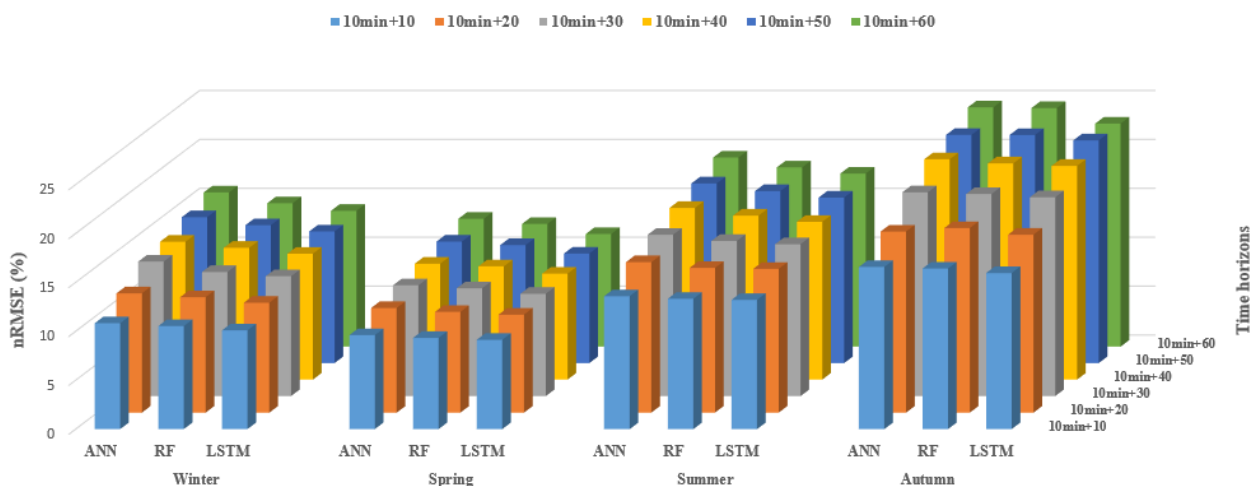


Figure 4.18. Comparaison entre les performances saisonnières des modèles de prédiction à l'avance de GHI en termes du nRMSE pour les différents horizons sub-horaires.

Chapitre 4 : Prédiction à différents horizons temporels de l'irradiation solaire globale sur plan horizontal.

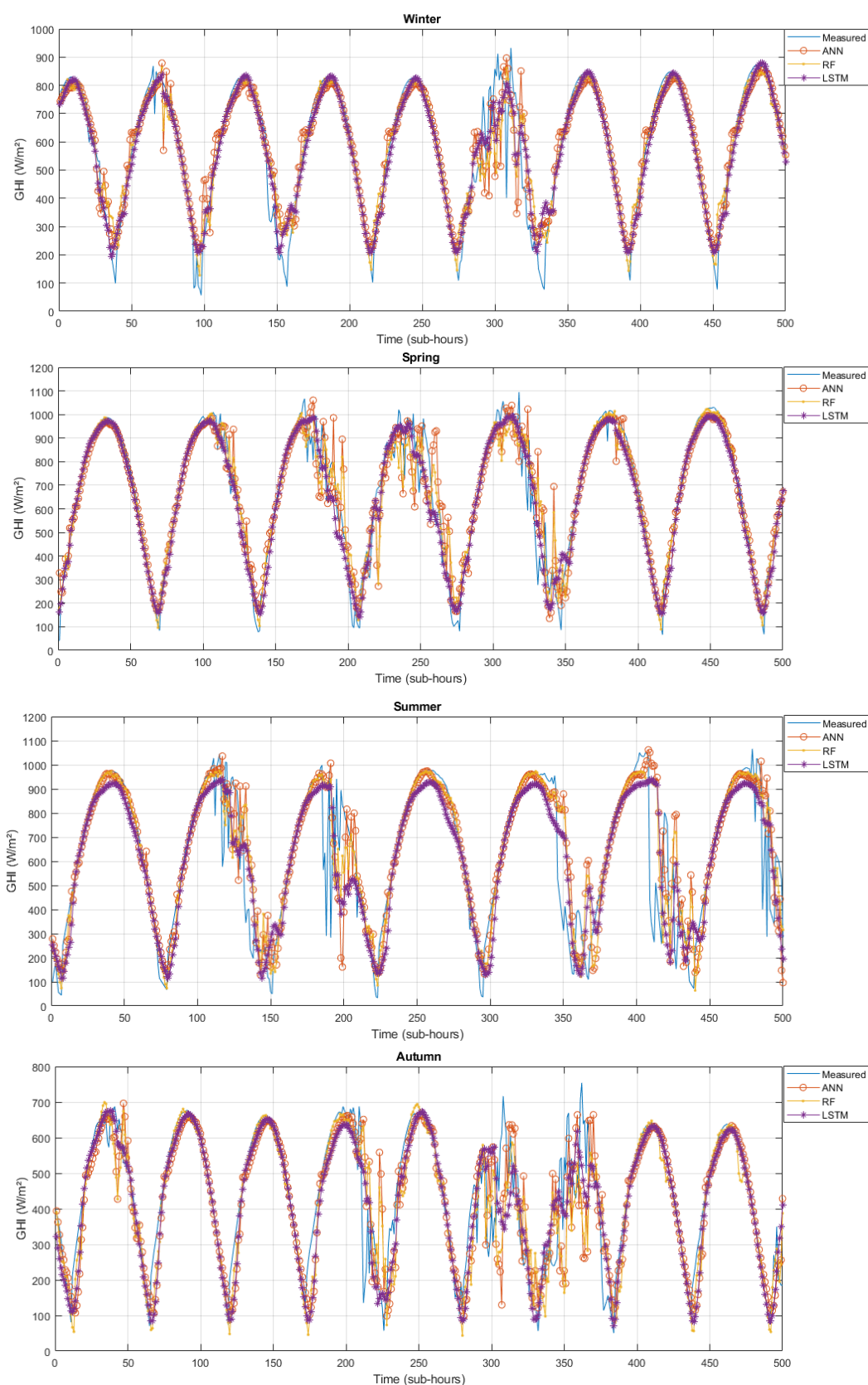


Figure 4. 19. Séries temporelles de 500 valeurs sub-horaires mesurées et prédites de GHI correspondantes aux modèles PMC, RF et LSTM pour chaque saison et pour un horizon temporel 10min+60.

La figure 4.19 illustre les séries temporelles par saison de 500 valeurs sub-horaires mesurées et prédites de GHI correspondantes aux modèles PMC, RF et LSTM pour l'horizon temporel 10min+60. Une fois de plus, l'automne reste la saison la plus difficile pour prédire le GHI. En effet, cette période de l'année est caractérisée par des variabilités importantes des conditions météorologiques par rapport aux autres saisons. Par conséquent, tous les modèles développés ont du mal à suivre ces variabilités. On note, toutefois, une performance de prédiction du modèle LSTM nettement meilleure par rapport aux autres modèles.

4.4. Performances du modèle d'apprentissage profond LSTM sur les bases de données des autres villes

L'étude effectuée précédemment pour prédire à l'avance les GHI horaires et sub-horaires à Erfoud, nous a révélé que le modèle d'apprentissage profond LSTM est le plus performant par rapport aux deux modèles d'apprentissage automatique PMC et RF pour des horizons temporels allant de h+1 à h+6 et de 10min+10 à 10min+60. Par conséquent, nous avons évalué la précision du modèle LSTM pour la prédiction à l'avance du GHI horaire sur tous les sites d'étude restants et pour la prédiction à l'avance du GHI sub-horaire, uniquement sur les bases de données enerMENA, vu l'indisponibilité des mesures sub-horaires (10min) à la ville de Fès.

4.4.1. Prédiction horaire du GHI

4.4.1.1. Scénario 1 : Performances annuelles

Pour chaque site d'étude, les performances annuelles pour la prédiction à l'avance du GHI à l'aide du modèle LSTM pour l'ensemble de validation sont présentées dans le tableau 4.10. Les résultats montrent que la meilleure précision est obtenue à Zagora pour tous les horizons horaires, avec les plus faibles valeurs du nRMSE qui varient entre 11.22% et 17.29% pour des horizons temporels allant de h+1 à h+6.

4.4.1.2. Scénario 2: Performances saisonnières

Le tableau 4.11 présente les performances saisonnières du modèle LSTM dans la prédiction à l'avance des GHI horaires aux différents sites. Ces résultats montrent que les meilleures performances de prédiction sont obtenues à la ville de Zagora avec des valeurs de nRMSE variant entre 8.01% et 13.44% au Printemps, entre 12.05% et 20.25% en Hiver et entre 17.86% et 29.60% en Automne, pour un horizon temporel allant de h+1 à h+6.

Chapitre 4 : Prédiction à différents horizons temporels de l'irradiation solaire globale sur plan horizontal.

Tableau 4. 10. Performances annuelles pour la prédiction à l'avance du GHI à l'aide du modèle LSTM pour l'ensemble de validation. Les chiffres en gras représentent la meilleure précision obtenue pour chaque horizon.

Site	Indicateurs statistiques (%)	Horizons temporels					
		h+1	h+2	h+3	h+4	h+5	h+6
Missour	R	95.45	93.50	91.24	89.68	89.45	88.51
	nMAE	11.33	13.36	15.34	17.21	16.97	18.84
	nRMSE	16.74	19.76	22.91	25.10	26.02	26.33
Oujda	R	96.49	95.01	93.32	91.00	89.61	88.35
	nMAE	9.37	11.11	13.05	15.07	16.16	17.15
	nRMSE	14.33	17.05	19.29	22.07	23.70	25.06
TanTan	R	96.06	93.62	92.00	90.28	88.83	87.91
	nMAE	11.34	14.64	16.64	18.51	20.04	21.02
	nRMSE	16.55	20.56	22.82	25.02	26.80	27.85
Zagora	R	97.53	96.46	95.79	94.95	94.29	94.08
	nMAE	6.72	8.75	9.47	10.77	11.66	11.24
	nRMSE	11.22	13.27	14.48	15.80	16.82	17.29
Fès	R	95.85	94.47	92.79	91.64	90.57	89.51
	nMAE	11.19	13.10	15.46	16.61	17.67	19.16
	nRMSE	17.39	19.78	22.41	24.00	25.44	26.86

Tableau 4. 11. Performances saisonnières de la prédiction à l'avance des GHI à l'aide du modèle LSTM pour l'ensemble de validation. Les chiffres en gras représentent la meilleure précision obtenue pour chaque horizon temporel et chaque saison.

Site	Indicateurs statistiques (%)	Horizons temporels					
		h+1	h+2	h+3	h+4	h+5	h+6
a) Hiver							
Missour	R	94.61	91.96	90.07	89.16	86.74	85.82
	nMAE	11.52	13.47	15.27	17.71	18.22	18.38
	nRMSE	17.23	20.62	23.01	25.24	26.61	27.42
Oujda	R	93.91	91.17	88.96	87.42	83.81	84.31
	nMAE	13.58	16.26	18.55	20.06	23.05	23.53
	nRMSE	19.49	22.68	25.22	26.70	29.96	29.94
TanTan	R	95.17	91.33	86.74	86.46	86.44	85.35
	nMAE	13.25	17.65	21.78	21.65	20.49	21.41
	nRMSE	17.41	22.26	27.42	27.40	28.66	29.32
Zagora	R	96.71	93.70	92.28	92.17	90.35	90.14
	nMAE	8.21	11.98	13.34	12.94	13.91	14.70
	nRMSE	12.05	16.16	17.62	18.05	20.05	20.25
Fès	R	94.22	92.06	90.11	89.55	85.87	85.11
	nMAE	12.76	15.37	17.89	19.24	21.60	22.85
	nRMSE	18.65	21.40	24.62	25.33	27.25	28.50
b) Printemps							
Missour	R	95.86	92.73	91.38	89.89	89.04	88.47
	nMAE	12.50	14.79	15.91	18.06	19.21	19.37
	nRMSE	17.87	20.80	22.65	24.27	25.38	25.93
Oujda	R	95.98	94.75	93.30	89.91	88.77	87.71
	nMAE	9.89	11.24	12.83	14.79	17.95	18.61
	nRMSE	14.76	16.56	18.54	20.59	23.77	24.76
TanTan	R	96.97	93.82	92.51	91.03	90.35	89.25
	nMAE	11.20	14.80	16.53	18.43	20.93	21.66
	nRMSE	15.37	17.82	19.27	20.15	22.31	23.87
Zagora	R	98.59	98.31	97.62	97.38	96.53	95.99
	nMAE	4.80	6.09	7.26	8.01	8.49	9.75
	nRMSE	8.01	9.02	10.44	11.07	12.44	13.44
Fès	R	95.42	93.54	91.56	90.20	89.65	89.21
	nMAE	11.33	13.55	15.46	17.65	18.51	18.88
	nRMSE	17.30	20.20	22.35	23.87	24.70	24.95

Chapitre 4 : Prédiction à différents horizons temporels de l'irradiation solaire globale sur plan horizontal.

En Eté, les plus faibles valeurs de nRMSE ont été obtenues à la ville de Fès et varient entre 12.02% et 18.04% pour h+1 et h+6, respectivement. Ces résultats ont confirmé encore une fois que l'automne est la saison la plus difficile pour prédire les valeurs horaires de GHI.

Tableau 4.11. (Suite) Performances saisonnières de la prédiction à l'avance des GHI à l'aide du modèle LSTM pour l'ensemble de validation. Les chiffres en gras représentent la meilleure précision obtenue pour chaque horizon temporel et chaque saison.

Site	Indicateurs statistiques (%)	Horizons temporels					
		h+1	h+2	h+3	h+4	h+5	h+6
c) Eté							
Missour	R	94.55	92.19	89.56	88.98	88.31	87.58
	nMAE	11.63	14.84	17.89	18.66	18.94	20.12
	nRMSE	18.82	22.47	25.71	26.05	26.76	27.60
Oujda	R	95.18	92.86	90.26	89.45	89.12	89.02
	nMAE	12.38	14.58	16.80	17.35	17.84	18.06
	nRMSE	17.81	21.09	24.40	25.35	25.87	26.41
TanTan	R	95.63	93.45	90.76	90.19	89.20	88.63
	nMAE	12.29	15.19	18.06	19.11	20.33	21.12
	nRMSE	17.59	20.90	24.38	25.24	26.40	27.20
Zagora	R	96.30	95.00	94.15	93.65	93.20	92.54
	nMAE	9.53	10.37	11.78	12.26	13.15	14.39
	nRMSE	14.61	15.83	16.48	17.41	17.89	18.87
Fès	R	97.38	96.39	95.71	94.88	94.46	93.86
	nMAE	6.77	8.44	9.13	10.05	10.64	11.40
	nRMSE	12.02	14.00	15.14	16.47	17.14	18.04
d) Automne							
Missour	R	91.48	89.55	85.30	84.76	84.13	83.95
	nMAE	18.07	20.02	22.16	23.40	23.96	24.36
	nRMSE	24.61	26.76	28.66	29.52	30.05	30.64
Oujda	R	92.65	89.89	85.45	84.80	84.35	84.10
	nMAE	17.58	19.20	21.13	22.36	22.85	23.15
	nRMSE	23.32	25.40	26.95	27.69	28.40	29.33
TanTan	R	91.22	88.75	84.95	84.12	83.92	83.13
	nMAE	18.40	20.67	21.85	23.55	24.30	25.02
	nRMSE	25.36	27.30	28.49	29.65	29.89	30.80
Zagora	R	94.01	90.45	88.02	86.68	85.10	84.81
	nMAE	11.80	16.35	17.55	18.65	20.78	21.85
	nRMSE	17.86	23.50	26.26	27.36	28.17	29.33
Fès	R	92.86	90.43	86.46	85.62	84.44	84.05
	nMAE	15.30	17.68	19.65	20.48	21.50	22.26
	nRMSE	21.84	24.78	27.39	28.56	28.89	29.75

4.4.2. Prédiction sub-horaire du GHI

4.4.2.1. Scénario 1 : Performances annuelles

Le tableau 4.12 montre les performances annuelles pour la prédiction sub-horaire de GHI dans les autres sites enerMENA, en utilisant le modèle LSTM pour un horizon temporel allant de 10min+10 à 10min+60. En général, les résultats trouvés indiquent une très bonne précision de la prédiction en avance du GHI sub-horaire pour toutes les stations. Néanmoins, la meilleure performance est celle obtenue à Zagora avec des valeurs de R et de nRMSE variant entre

97.19% et 95.02%, 10.71% et 14.03%, pour 10min+10 et 10min+60, respectivement. Aussi, on constate, qu'en terme du nRMSE, les performances des prédictions sub-horaires sont meilleures que les performances des prédictions horaires pour l'ensemble des sites considérés.

Tableau 4. 12. Performances annuelles pour la prédiction à l'avance de GHI à l'aide du modèle LSTM pour l'ensemble de validation. Les chiffres en gras représentent la meilleure précision obtenue pour chaque horizon.

Site	Indicateurs statistiques (%)	Horizons temporels					
		10min+10	10min+20	10min+30	10min+40	10min+50	10min+60
Missour	R	96.96	95.77	94.75	93.86	93.21	92.22
	nMAE	7.34	9.06	10.36	12.14	12.26	13.52
	nRMSE	12.90	15.03	16.63	18.03	18.79	20.03
Oujda	R	97.12	96.15	95.49	94.97	94.29	93.86
	nMAE	6.51	7.93	8.77	9.56	10.46	11.05
	nRMSE	12.02	13.78	14.83	15.63	16.62	17.21
TanTan	R	97.14	96.56	95.98	95.26	94.70	94.28
	nMAE	6.98	8.38	9.26	10.18	11.07	11.52
	nRMSE	12.19	14.13	15.21	16.44	17.35	17.98
Zagora	R	97.19	96.65	96.08	95.61	95.37	95.02
	nMAE	5.53	6.08	7.12	7.58	7.79	8.31
	nRMSE	10.71	11.78	12.57	13.23	13.55	14.03

4.4.2.2. Scénario 2: Performances saisonnières

Les performances saisonnières du modèle LSTM pour la prédiction à l'avance du GHI sub-horaire aux différents sites enerMENA sont présentées dans le tableau 4.13, pour un horizon temporel allant de 10min+10 à 10min+60. De même que pour les performances saisonnières (Tableau 4.11), les meilleures précisions de prédiction sont obtenues à Zagora avec les plus faibles valeurs de nRMSE variant entre 8.98% et 14.58% en Hiver, entre 8.78% et 11.20% au printemps et entre 14.50% et 20.73% en Automne pour 10min+10 et 10min+60, respectivement. Par contre, les meilleures précisions de prédiction en Été sont obtenues à Oujda avec un nRMSE qui varie entre 11.73% et 18.01% pour 10min+10 et 10min+60, respectivement.

En général, cette différence de précision entre les performances obtenues revient au niveau de variabilités météorologiques qui perturbent les mesures solaires dans chaque site d'étude. Les sites d'Erfoud et Zagora se caractérisent par un climat plus stable (moins de variabilités météorologiques) par rapport aux autres villes. Aussi, il a été remarqué que les mesures horaires et sub-horaires des données solaires ou météorologiques effectuées en ces deux stations (Erfoud et Zagora) sont plus précises à celles des autres stations (Missour, Oujda, et TanTan).

Chapitre 4 : Prédiction à différents horizons temporels de l'irradiation solaire globale sur plan horizontal.

Tableau 4. 13. Performances saisonnières de la prédiction à l'avance des GHI à l'aide du modèle LSTM pour l'ensemble de validation. Les chiffres en gras représentent la meilleure précision obtenue pour chaque horizon temporel et chaque saison.

Site	Indicateurs statistiques (%)	Horizons temporels					
		10min+10	10min+20	10min+30	10min+40	10min+50	10min+60
a) Hiver							
Missour	R	97.31	96.32	95.15	93.82	92.74	90.78
	nMAE	7.08	8.53	10.14	11.28	12.90	14.84
	nRMSE	11.92	13.84	15.79	17.51	19.04	21.23
Oujda	R	95.31	93.92	92.81	91.68	91.07	90.23
	nMAE	9.51	11.13	12.57	13.88	14.69	15.57
	nRMSE	16.37	18.43	19.89	21.26	21.99	22.90
TanTan	R	96.53	95.52	94.05	93.81	92.68	91.00
	nMAE	8.50	9.95	11.53	11.85	13.26	14.61
	nRMSE	13.19	14.84	16.77	17.07	18.49	20.21
Zagora	R	97.55	96.42	95.89	94.77	93.76	93.19
	nMAE	4.89	6.41	6.68	7.77	8.99	9.81
	nRMSE	8.98	10.87	11.49	12.84	13.96	14.58
b) Printemps							
Missour	R	96.45	95.19	94.12	93.09	92.02	91.33
	nMAE	7.33	8.92	10.32	11.46	12.54	13.71
	nRMSE	14.00	16.09	17.65	19.02	20.34	21.12
Oujda	R	96.77	95.90	95.18	94.59	93.83	93.61
	nMAE	7.75	8.36	9.48	10.12	11.37	11.34
	nRMSE	12.93	14.29	15.44	16.37	17.34	17.60
TanTan	R	97.91	97.83	97.06	96.73	96.45	96.06
	nMAE	6.07	7.54	8.53	9.57	10.44	10.98
	nRMSE	10.26	11.76	13.02	14.06	15.08	15.83
Zagora	R	98.45	97.95	97.41	96.95	96.67	96.39
	nMAE	3.70	4.20	4.74	5.58	5.27	5.83
	nRMSE	8.78	9.56	10.11	10.72	10.75	11.20
c) Été							
Missour	R	96.66	95.36	94.20	93.10	92.03	91.53
	nMAE	6.91	8.55	9.75	10.98	12.10	12.67
	nRMSE	13.62	15.76	17.63	19.08	20.43	21.12
Oujda	R	97.47	96.47	95.73	94.56	94.39	93.68
	nMAE	6.39	7.81	8.78	10.18	10.45	11.35
	nRMSE	11.73	13.68	14.96	16.78	17.07	18.01
TanTan	R	97.64	96.63	95.49	94.77	94.06	93.12
	nMAE	6.88	8.20	9.96	10.94	11.85	13.00
	nRMSE	11.89	14.02	16.05	17.17	18.32	19.73
Zagora	R	95.91	94.86	94.06	93.39	92.98	92.11
	nMAE	7.31	8.72	9.94	10.37	11.17	12.09
	nRMSE	13.95	15.50	16.60	17.52	17.98	19.02
d) Automne							
Missour	R	94.26	91.83	90.35	88.49	86.75	85.44
	nMAE	11.60	13.66	15.23	17.13	18.82	19.88
	nRMSE	18.86	22.15	23.84	25.80	27.55	28.72
Oujda	R	96.30	94.91	93.92	93.32	92.14	91.61
	nMAE	9.64	11.97	13.32	14.27	15.72	17.04
	nRMSE	15.90	18.41	20.00	20.91	22.62	23.45
TanTan	R	94.75	92.88	91.58	90.44	89.16	88.18
	nMAE	9.37	11.20	12.49	13.58	14.57	15.39
	nRMSE	15.97	18.27	19.76	20.96	22.23	23.13
Zagora	R	96.78	95.10	94.28	93.88	92.55	92.07
	nMAE	8.02	9.51	10.73	11.79	12.16	14.13
	nRMSE	14.50	16.49	17.74	19.00	19.10	20.73

5. Conclusion

Ce chapitre a été consacré à la prédiction à l'avance de l'irradiation solaire globale horizontale (GHI) à l'aide des modèles d'apprentissage automatique (Machine learning) et profond (Deep learning) pour différents horizons temporels horaires et sub-horaires. Nous avons d'abord présenté une étude bibliographique sur les différentes techniques d'apprentissage profond pour la prédiction du rayonnement solaire. Puis, une description détaillée du nouveau modèle d'apprentissage profond nommé réseaux de neurones à longue mémoire à court terme (LSTM) a été proposée. Enfin, les résultats des prédictions à l'avance des GHI horaires et sub-horaires ont été présentés, discutés et analysés.

En général, les résultats trouvés ont montré que les différents modèles d'apprentissage automatiques (PMC, NARX, RF) ou le modèle LSTM d'apprentissage profond peuvent être utilisés pour la prédiction à l'avance des GHI horaires ou sub-horaires. Toutefois, le modèle LSTM est plus performant surtout lorsque l'horizon temporel de prédiction augmente. Les améliorations des performances apportées par le modèle LSTM augmentent avec l'horizon temporel et varient, en termes des valeurs du nRMSE, de 3.83% et 2.59% à 11.37% et 8.59% pour les horizons 10min+10 et 10min+60 par rapport aux modèles PMC et RF, respectivement. Ces améliorations varient de 14.35% et 12.75% à 23.56% et 19.33% pour les horizons h+1 et h+6 par rapport aux modèles PMC et RF, respectivement.

Pour les prédictions saisonnières, il a été constaté que les plus faibles performances ont été obtenues pour la saison d'automne. Ceci peut être expliqué par le fait que cette saison est caractérisée par des fluctuations considérables dans les conditions météorologiques. Pour cette saison, le modèle LSTM est le plus performant.

Enfin, on peut conclure que le modèle LSTM est le plus prometteur pour prévoir le GHI à des horizons temporels à très court, à court et à long termes avec de bonnes performances et un haut degré de stabilité par rapport aux modèles RNAs qui ont été largement utilisés dans le domaine de la prévision solaire.

Conclusion générale

L'étude de la viabilité économique et le dimensionnement de toute installation solaire nécessitent la disponibilité des mesures fiables du rayonnement solaire à différentes résolutions temporelles. Or, le réseau de stations de mesure du rayonnement solaire au sol est très faible autour du monde. Une voie de recherche d'intérêt majeur consiste à développer des modèles de prédiction précis en cherchant des relations entre les variables météorologiques communément mesurées et les différentes composantes de l'irradiation solaire.

L'objectif principal de notre thèse consiste à chercher et développer des prédicteurs précis de la ressource solaire afin de déterminer la quantité d'énergie solaire disponible au niveau du sol et faciliter, ainsi, la conception des installations solaire et leur intégration aux réseaux électriques. A cet égard, nous avons abordé deux axes de recherche en relation avec cette problématique. Le premier axe concerne l'estimation des irradiances solaires journalières globales horizontales (GHI), diffuses horizontales (DHI) et normales directes (DNI) à l'aide des modèles d'apprentissage automatique (Machine learning) en se basant uniquement sur les paramètres couramment mesurés dans les stations météorologiques (température, humidité relative de l'air, vitesse du vent, ...). Les performances de ces modèles ont été comparées à celles des modèles empiriques. Le deuxième axe de notre travail de recherche concerne la prédiction à l'avance des GHI horaires et sub-horaires pour des horizons temporels allant de $h+1$ à $h+6$ et de $10\text{min}+10$ à $10\text{min}+60$, en n'utilisant comme entrées que les variables endogènes (valeurs historiques de GHI) et en employant les nouvelles techniques d'apprentissage automatique et profond (Deep learning).

Avant d'entamer l'estimation ou la prédiction des différentes composantes solaires, il a fallu d'abord examiner la qualité des données disponibles. Pour cela, un état de l'art détaillé sur les différentes techniques de contrôle de qualité existantes a été présenté dans le premier chapitre. Les données des irradiances solaires mesurées aux six stations météorologiques considérées dans le présent travail (Fès, Erfoud, Missouri, Oujda, Zagora, TanTa), ont été, alors, examinées à l'aide d'une procédure de contrôle de qualité ce qui nous a permis d'éliminer les mesures les plus ou moins douteuses. Les résultats de ce contrôle ont révélé que la qualité des mesures des irradiances solaires varie d'une ville à l'autre. En effet, pour nos bases de données, il a été

constaté que la qualité des mesures à Erfoud, Fès, et Zagora est meilleure à celle de Missour, Oujda, TanTan.

Le deuxième chapitre a été consacré à l'estimation des irradiances solaires journalières GHI, DHI et DNI en utilisant des modèles empiriques. Pour ce faire, plusieurs modèles empiriques, largement utilisés dans la littérature, ou que nous avons, nous-mêmes, proposés, ont été testés. Les performances de 22 modèles empiriques, sélectionnés dans la littérature, ont été, ainsi, évaluées pour estimer les GHI journalière aux 6 sites marocains considérés. Les résultats trouvés nous ont permis de recommander le modèle TG1 qui ne nécessite que la connaissance (mesure) du gradient de température journalier $\Delta T = T_{max} - T_{min}$. Les valeurs maximales (T_{max}) et minimales (T_{min}) de la température de l'air sont mesurées dans toutes les stations météorologiques. Pour les composantes DNI et DHI journalières, nous avons testé des modèles basés sur des paramètres hybrides (HBM) et des modèles basés sur l'indice de clarté (CIBM). La comparaison des performances des différents modèles a montré que les deux modèles basés sur les paramètres hybrides (DH3 et DNH1) que nous avons proposés donnent les meilleurs résultats.

Dans le troisième chapitre, des modèles d'apprentissage automatique, tels que, les réseaux de neurones artificiels (ANN et NARX), les méthodes d'ensemble (Boosting, Bagging, Forêts aléatoires) et la régression par machine à vecteur de support (SVM) ont été proposés pour estimer les GHI, DHI et DNI journalières, en n'utilisant que les variables couramment mesurées dans les stations météorologiques (température, humidité relative de l'air, vitesse du vent, ...) et quelques paramètres astronomiques facilement calculables. Dans un premier temps, deux architectures des réseaux de neurones artificiels (PMC et NARX) ont été proposées pour estimer le GHI quotidien aux six sites marocains. Les résultats obtenus montrent une bonne performance d'estimation pour toutes les stations ($10\% < nRMSE < 20\%$), à l'exception de la ville de TanTan qui se caractérise par une variabilité météorologique élevée (brouillards fréquents). Ensuite, les potentiels des trois méthodes d'ensemble basées sur des arbres de régression (Bagging, Boosting et Forêts aléatoires) ont été explorés pour l'estimation du GHI quotidien dans les cinq sites d'enerMENA. Les performances de ces méthodes ont été comparées à celles du modèle PMC et des 22 modèles empiriques testés dans le deuxième chapitre. Après optimisation de tous les paramètres, nous avons constaté que l'ensemble des modèles d'apprentissage automatique utilisés donnent de bonnes performances ($7\% < nRMSE < 20\%$). Aussi, le modèle des forêts aléatoires (Random Forests (RF)) s'est avéré être le plus

robuste dans toutes les stations, en particulier, lorsqu'il s'agit d'une base de données bruitée (TanTan) ($R= 87.53\%$ et $nRMSE=15.33\%$). Plus généralement, cette étude nous a permis de conclure que les méthodes d'ensemble sont les plus prometteuses et donnent un haut degré de stabilité par rapport aux réseaux neuronaux artificiels, qui ont été largement utilisés dans le domaine de l'estimation et de la prédiction du rayonnement solaire. A la fin de ce chapitre, les performances des méthodes d'apprentissage automatique (PMC, RF et SVM) ont été évaluées pour l'estimation des trois composantes solaires (GHI, DHI, DNI) lorsque seulement des mesures restreintes sont disponibles. Ces modèles ont été catégorisés selon le type des entrées utilisées : La température (TBMs: Temperature Based Models) et l'humidité relative (RHBMs: Relative Humidity Based Models) pour l'estimation journalière de GHI, les paramètres hybrides (HBMs: Hybrid Based Models) pour les estimations journalières de GHI, DHI et DNI et l'indice de clarté (CIBMs: Clarity Index Based Models) pour les estimations journalières de DHI et DNI. Les résultats ont montré que le modèle RF est le plus performant pour l'estimation de l'ensemble des composantes solaires (GHI, DHI, DNI), en utilisant comme variables d'entrées les paramètres hybrides. Néanmoins, la variante GT3 des modèles SVM, PMC et RF peut être utilisée pour estimer le GHI en prenant en compte uniquement les mesures du gradient de température journalier (ΔT) et l'irradiation solaire journalière d'un plan horizontal hors atmosphère G_{TOA} . De plus, les modèles PMC, RF et SVM peuvent être utilisés indifféremment en utilisant les valeurs extrêmes de l'humidité relative comme entrées.

En ce qui concerne la prédiction en avance du rayonnement solaire, en particulier l'irradiation globale horizontale à partir des variables exogènes (données météorologiques) et endogènes (valeurs historiques de GHI), un ensemble de modèles d'apprentissage automatique et profond ont été développés dans le dernier chapitre. Dans un premier temps, deux modèles RNAs (PMC et NARX) ont été développés pour prédire à court ($h+1$) et à long termes ($h+48$) le GHI horaire, en utilisant les paramètres météorologiques couramment mesurés. Les performances obtenues ont confirmé que le modèle NARX est capable de prédire efficacement le GHI à court ($h+1$) et à long termes ($h+48$), avec la plus faible valeur du $nRMSE$ qui est égale à 15.73% . De plus, en testant les performances du modèle NARX proposé sur les bases de données d'enerMENA, il a été conclu que ce modèle (NARX) est efficace pour la prédiction du GHI horaire aux différents sites. Ensuite, une nouvelle approche d'apprentissage profond basée sur les réseaux de neurones LSTM (Long-Short Term Memory networks) a été utilisée pour prédire en avance les GHI horaires et sub-horaires. Pour les horizons temporels de $h+1$ à $h+6$ et de $10min+10$ à $10min+60$, le modèle LSTM proposé prévoit simultanément les valeurs horaires et sub-horaires

de GHI, en n'utilisant que les valeurs historiques de GHI (variables endogènes). Deux scénarios ont été proposés, un scénario des performances annuelles et un scénario des performances saisonnières. En comparant les performances du modèles LSTM aux deux modèles populaires d'apprentissage automatique (PMC et RF), il a été constaté que le LSTM est plus performant pour la prévision horaire et sub-horaire du GHI. En particulier, une grande amélioration en terme des valeurs du nRMSE a été observée lorsque l'horizon temporel de prévision augmente (LSTM :17.35%, RF : 21.51%, PMC : 22.70% pour h+6, et LSTM :14.88%, RF : 16.28%, PMC : 16.79% pour 10min+60). L'étude saisonnière a montré que la prédiction du GHI en automne est moins fiable que pendant l'hiver, le printemps et l'été en raison de la grande variabilité météorologique observée pendant cette période. Néanmoins, le modèle LSTM est plus robuste (nRMSE =30.52% pour h+6 et nRMSE= 22.80% pour 10min+60) à cette variabilité par rapport aux modèles PMC (nRMSE=36.61% pour h+6 et nRMSE=24.46% pour 10min+60) et RF (nRMSE= 36.52% pour h+6 et nRMSE= 24.39% pour 10min+60). Cette étude nous a permis de conclure que le modèle LSTM est un outil prometteur pour la prédiction en avance de GHI avec un haut degré de fiabilité et de stabilité.

En conclusion, tous les modèles développés durant cette thèse fournissent, avec de bonnes précisions, les séries temporelles des irradiances solaires qui sont très utiles pour la conception des systèmes solaires. Toutefois, les modèles basés sur les techniques d'apprentissage automatique donnent de meilleures performances que les modèles empiriques. Aussi, le modèle d'apprentissage profond LSTM est le modèle le plus performant pour la prédiction à l'avance des irradiances solaires globales horizontales horaires et sub-horaires.

En terme de perspectives, ce travail peut ouvrir de nouveaux horizons de recherche :

- Tester l'efficacité des modèles développés dans d'autres régions marocaines avec des climats différents.
- Combiner les algorithmes bagging et boosting aux réseaux de neurones artificiels afin d'améliorer encore la prédiction du GHI.
- Exploiter les différents modèles pour l'élaboration d'une carte du gisement solaire marocain pour déterminer le potentiel solaire des zones géographiques dans lesquelles l'installation des centrales solaires est plus rentable.
- Intégrer les modèles de prédiction élaborés dans les stratégies de commande des centrales solaires et de gestion du mix énergétique.

Références

- A. Redouane et al, 2018. “Business Opportunities Report for Morocco’s Renewable Energy Sector”, Ministry of Foreign affairs, Commissioned by Netherlands Enterprise Agency, 2018, <https://www.rijksoverheid.nl/binaries/rijksoverheid/documenten/rapporten/2018/06/01/business-opportunities-report-for-morocco%E2%80%99s-renewable-energy-sector/business-opportunities-report-for-morocco%E2%80%99s-renewable-energy-sector.pdf> (Consulté le 08/05/2019).
- Adeala, A.A., Huan, Z., Enweremadu, C.C., 2015. Evaluation of global solar radiation using multiple weather parameters as predictors for South Africa provinces. *Thermal Science* 19, 495–509.
- Agrégation de modèles, 2018. <https://www.math.univ-toulouse.fr/~besse/Wikistat/pdf/st-m-app-agreg.pdf> (consulté le 14/06/2021).
- Ahmad, A., Anderson, T.N., Lie, T.T., 2015. Hourly global solar irradiation forecasting for New Zealand. *Solar Energy* 122, 1398–1408.
- Akinoğlu, B.G., 1991. A review of sunshine-based models used to estimate monthly average global solar radiation. *Renewable Energy* 1, 479–497.
- Akinoğlu, B.G., Ecevit, A., 1990. A further comparison and discussion of sunshine-based models to estimate global solar radiation. *Energy* 15, 865–872.
- Allen, R.G., 1997. Self-calibrating method for estimating solar radiation from air temperature. *Journal of Hydrologic engineering* 2, 56–67.
- Al-Sbou, Y.A., Alawasa, K.M., 2017. Nonlinear autoregressive recurrent neural network model for solar radiation prediction. *International Journal of Applied Engineering Research* 12, 4518–4527.
- Alsharif, M.H., Younes, M.K., Kim, J., 2019. Time series ARIMA model for prediction of daily and monthly average global solar radiation: The case study of Seoul, South Korea. *Symmetry* 11, 240.
- Alsina, E.F., Bortolini, M., Gamberi, M., Regattieri, A., 2016. Artificial neural network optimisation for monthly average daily global solar radiation prediction. *Energy conversion and management* 120, 320–329.
- Amey laddad, 2019. Basic understanding of LSTM “ <https://blog.goodaudience.com/basic-understanding-of-lstm-539f3b013f1e>”. (Accessed June 30, 2021).
- Angstrom, A., 1924. Solar and terrestrial radiation. Report to the international commission for solar research on actinometric investigations of solar and atmospheric radiation. *Quarterly Journal of the Royal Meteorological Society* 50, 121–126.
- Annandale, J., Jovanovic, N., Benade, N., Allen, R., 2002. Software for missing data error analysis of Penman-Monteith reference evapotranspiration. *Irrigation Science* 21, 57–67.
- Awad, M., Khanna, R., 2015. *Efficient learning machines: theories, concepts, and applications for engineers and system designers*. Springer nature.
- Ayodele, T.R., Ogunjuyigbe, A.S.O., Amedu, A., Munda, J.L., 2019. Prediction of global solar irradiation using hybridized k-means and support vector regression algorithms. *Renewable Energy Focus* 29, 78–93.
- Bakhashwain, J.M., 2016. Prediction of global solar radiation using support vector machines. *International Journal of Green Energy* 13, 1467–1472.

- Behrang, M., Assareh, E., Ghanbarzadeh, A., Noghrehabadi, A., 2010. The potential of different artificial neural network (ANN) techniques in daily global solar radiation modeling based on meteorological data. *Solar Energy* 84, 1468–1480.
- Belaid, S., Mellit, A., 2016. Prediction of daily and mean monthly global solar radiation using support vector machine in an arid climate. *Energy Conversion and Management* 118, 105–118.
- Benali, L., Notton, G., Foulloy, A., Voyant, C., Dizene, R., 2019. Solar radiation forecasting using artificial neural network and random forest methods: Application to normal beam, horizontal diffuse and global components. *Renewable energy* 132, 871–884.
- Benghanem, M., Mellit, A., Alamri, S., 2009. ANN-based modelling and estimation of daily global solar radiation data: A case study. *Energy conversion and management* 50, 1644–1655.
- Benson, R.B., Paris, M.V., Sherry, J.E., Justus, C.G., 1984. Estimation of daily and monthly direct, diffuse and global solar radiation from sunshine duration measurements. *Solar energy* 32, 523–535.
- Besharat, F., Dehghan, A.A., Faghieh, A.R., 2013. Empirical models for estimating global solar radiation: A review and case study. *Renewable and Sustainable Energy Reviews* 21, 798–821.
- Bessemoulin, P., Oliviéri, J., 2000. Le rayonnement solaire et sa composante ultraviolette. *La météorologie*.
- Bilgili, M., Ozgoren, M., 2011. Daily total global solar radiation modeling from several meteorological data. *Meteorology and Atmospheric Physics* 112, 125–138.
- Blanc, P., Wald, L., 2012. The SG2 algorithm for a fast and accurate computation of the position of the Sun for multi-decadal time period. *Solar Energy* 86, 3072–3083.
- Blanco-Muriel, M., Alarcón-Padilla, D.C., López-Moratalla, T., Lara-Coira, M., 2001. Computing the solar vector. *Solar energy* 70, 431–441.
- Bounoua, Z., Chahidi, L.O., Mechaqrane, A., 2021. Estimation of daily global solar radiation using empirical and machine-learning methods: a case study of five Moroccan locations. *Sustainable Materials and Technologies* 28, e00261.
- Bounoua, Z., Marzouq, M., Mechaqrane, A., 2018. Assessment of a quality control procedure of hourly solar irradiances at Fez city, Morocco, in: *IOP Conference Series: Earth and Environmental Science*. IOP Publishing, p. 012010.
- Bounoua, Z., Mechaqrane, A., 2022. Estimation of Daily Direct Normal Solar Irradiation Using Machine-Learning Methods, in: *WITS 2020*. Springer, pp. 627–638.
- Bounoua, Z., Mechaqrane, A., 2021a. Categorized machine-learning models to estimate daily solar radiation in Zagora City, Southern Morocco (Under Review).
- Bounoua, Z., Mechaqrane, A., 2021b. Hourly and Sub-Hourly Ahead Global Horizontal Solar Irradiation Forecasting using a Long Short-Term Memory Neural Networks Deep Learning Approach: A Case Study (Under Review).
- Bounoua, Z., Mechaqrane, A., 2020. Prediction of Short-Term and Long-Term Hourly Global Horizontal Solar Irradiation Using Artificial Neural Networks Techniques in Fez City, Morocco, in: *International Conference on Electronic Engineering and Renewable Energy*. Springer, pp. 683–690.
- Bounoua, Z., Mechaqrane, A., 2018. Prediction of daily global horizontal solar irradiation using artificial neural networks and commonly measured meteorological parameters, in: *AIP Conference Proceedings*. AIP Publishing LLC, p. 020024.
- Boussaada, Z., 2018. Contribution à la Gestion et à la Commande des Micro-Réseaux par les Systèmes Multi Agents.

- Boussaada, Z., Curea, O., Remaci, A., Camblong, H., Mrabet Bellaaj, N., 2018. A nonlinear autoregressive exogenous (NARX) neural network model for the prediction of the daily direct solar radiation. *Energies* 11, 620.
- Breiman, L., 2001. Random forests. *Machine learning* 45, 5–32.
- Breiman, L., 1996. Bagging predictors. *Machine learning* 24, 123–140.
- Bretagnon, P., 1982. Theorie du mouvement de l'ensemble des planetes (VSOP82). *Astronomy and astrophysics* 114, 278.
- Bretagnon, P., Francou, G., 1988. Planetary theories in rectangular and spherical variables-VSOP 87 solutions. *Astronomy and astrophysics* 202, 309–315.
- Bristow, K.L., Campbell, G.S., 1984. On the relationship between incoming solar radiation and daily maximum and minimum temperature. *Agricultural and forest meteorology* 31, 159–166.
- Brochure of the SP Lite2 Silicon Pyranometer - Kipp & Zonen, 2014. “<https://www.kippzonen.com/Download/592/Brochure-Pyranometers-SP-Lite2-French?ShowInfo=true>” (consulté le 14/10/2021) [WWW Document].
- Calça, M.V.C., Raniero, M.R., Fernando, D.M.Z., Rodrigues, S.A., Dal Pai, A., 2019. Outliers Detection in a Quality Control Procedure for Measurements of Solar Radiation. *IEEE Latin America Transactions* 17, 1815–1822.
- Chelbi, M., Gagnon, Y., Waewsak, J., 2015. Solar radiation mapping using sunshine duration-based models and interpolation techniques: Application to Tunisia. *Energy Conversion and Management* 101, 203–215.
- Chen, J., Li, G., 2013. Estimation of monthly average daily solar radiation from measured meteorological data in Yangtze River Basin in China. *International journal of climatology* 33, 487–498.
- Chen, J.-L., He, L., Yang, H., Ma, M., Chen, Q., Wu, S.-J., Xiao, Z., 2019. Empirical models for estimating monthly global solar radiation: A most comprehensive review and comparative case study in China. *Renewable and Sustainable Energy Reviews* 108, 91–111.
- Chen, J.-L., Li, G.-S., 2014. Evaluation of support vector machine for estimation of solar radiation from measured meteorological variables. *Theoretical and applied climatology* 115, 627–638.
- Chen, J.-L., Li, G.-S., Wu, S.-J., 2013. Assessing the potential of support vector machine for estimating daily solar radiation using sunshine duration. *Energy conversion and management* 75, 311–318.
- Chen, J.-L., Li, G.-S., Xiao, B.-B., Wen, Z.-F., Lv, M.-Q., Chen, C.-D., Jiang, Y., Wang, X.-X., Wu, S.-J., 2015. Assessing the transferability of support vector machine model for estimation of global solar radiation from air temperature. *Energy Conversion and Management* 89, 318–329.
- Chen, J.-L., Liu, H.-B., Wu, W., Xie, D.-T., 2011. Estimation of monthly solar radiation from measured temperatures using support vector machines—a case study. *Renewable Energy* 36, 413–420.
- Chen, R., Kang, E., Lu, S., Yang, J., Ji, X., Zhang, Z., Zhang, J., 2006. New methods to estimate global radiation based on meteorological data in China. *Energy conversion and management* 47, 2991–2998.
- Chen, Z., Yang, Y., 2004. Assessing forecast accuracy measures. *Preprint Series 2010*, 2004–10.
- Chukwujindu, N.S., 2017. A comprehensive review of empirical models for estimating global solar radiation in Africa. *Renewable and sustainable energy reviews* 78, 955–995.

- Despotovic, M., Nedic, V., Despotovic, D., Cvetanovic, S., 2016. Evaluation of empirical models for predicting monthly mean horizontal diffuse solar radiation. *Renewable and Sustainable Energy Reviews* 56, 246–260.
- Dornelles, K., Roriz, M., Roriz, V., Caram, R., 2011. Thermal performance of white solar-reflective paints for cool roofs and the influence on the thermal comfort and building energy use in hot climates, in: *Solar World Congress*. ISES. International Solar Energy Society.
- Doucouré, B., 2015. Proposition, intégration dans un système de gestion de réseau intelligent et validation expérimentale d'une méthode de prédiction pour un système d'énergies renouvelables (PhD Thesis). Université du Québec à Trois-Rivières.
- Dreyfus, G., Martinez, J.M., Samuelides, M., Gordon, M.B., Badran, F., Thiria, S., Héroult, L., 2002. *Réseaux de neurones*. Eyrolles Paris.
- Elizondo, D., Hoogenboom, G., McClendon, R.W., 1994. Development of a neural network model to predict daily solar radiation. *Agricultural and Forest Meteorology* 71, 115–132.
- Espinar, B., Wald, L., Blanc, P., Hoyer-Klick, C., Homscheidt, M.S., Wanderer, T., 2011. Project ENDORSE-Excerpt of the report on the harmonization and qualification of meteorological data: Procedures for quality check of meteorological data (PhD Thesis). Mines ParisTech.
- Fan, J., Chen, B., Wu, L., Zhang, F., Lu, X., Xiang, Y., 2018a. Evaluation and development of temperature-based empirical models for estimating daily global solar radiation in humid regions. *Energy* 144, 903–914.
- Fan, J., Wang, X., Wu, L., Zhou, H., Zhang, F., Yu, X., Lu, X., Xiang, Y., 2018b. Comparison of Support Vector Machine and Extreme Gradient Boosting for predicting daily global solar radiation using temperature and precipitation in humid subtropical climates: A case study in China. *Energy Conversion and Management* 164, 102–111.
- Fan, J., Wu, L., Zhang, F., Cai, H., Ma, X., Bai, H., 2019a. Evaluation and development of empirical models for estimating daily and monthly mean daily diffuse horizontal solar radiation for different climatic regions of China. *Renewable and Sustainable Energy Reviews* 105, 168–186.
- Fan, J., Wu, L., Zhang, F., Cai, H., Zeng, W., Wang, X., Zou, H., 2019b. Empirical and machine learning models for predicting daily global solar radiation from sunshine duration: A review and case study in China. *Renewable and Sustainable Energy Reviews* 100, 186–212.
- Feng, Y., Cui, N., Zhang, Q., Zhao, L., Gong, D., 2017. Comparison of artificial intelligence and empirical models for estimation of daily diffuse solar radiation in North China Plain. *International Journal of Hydrogen Energy* 42, 14418–14428.
- Feng, Y., Gong, D., Jiang, S., Zhao, L., Cui, N., 2020. National-scale development and calibration of empirical models for predicting daily global solar radiation in China. *Energy Conversion and Management* 203, 112236.
- Feng, Y., Gong, D., Zhang, Q., Jiang, S., Zhao, L., Cui, N., 2019. Evaluation of temperature-based machine learning and empirical models for predicting daily global solar radiation. *Energy Conversion and Management* 198, 111780.
- Ferreira, A.A., Ludermit, T.B., De Aquino, R.R., 2012. Comparing recurrent networks for time-series forecasting, in: *The 2012 International Joint Conference on Neural Networks (IJCNN)*. IEEE, pp. 1–8.
- Fischer, T., Krauss, C., 2018. Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research* 270, 654–669.

- Fouilloy, A., Voyant, C., Notton, G., Motte, F., Paoli, C., Nivet, M.-L., Guillot, E., Duchaud, J.-L., 2018. Solar irradiation prediction with machine learning: Forecasting models selection method depending on weather variability. *Energy* 165, 620–629.
- Freund, Y., Schapire, R.E., 1996. Experiments with a new boosting algorithm. Presented at the icml, Citeseer, pp. 148–156.
- Fröhlich, C., 1977. World radiometric reference. WMO/CIMO Final Rep., WMO Rep 490, 97–110.
- Gallant, A.R., White, H., 1988. There exists a neural network that does not make avoidable mistakes. Presented at the Proc. of the International Conference on Neural Networks, San Diego.
- Garg, H., Garg, S., 1983. Prediction of global solar radiation from bright sunshine hours and other meteorological data. *Energy Conversion and Management* 23, 113–118.
- Geiger, M., Diabaté, L., Ménard, L., Wald, L., 2002. A web service for controlling the quality of measurements of global solar irradiation. *Solar energy* 73, 475–480.
- Gey, S., Poggi, J.-M., 2006. Boosting and instability for regression trees. *Computational statistics & data analysis* 50, 533–550.
- Ghimire, S., Deo, R.C., Raj, N., Mi, J., 2019. Deep solar radiation forecasting with convolutional neural network and long short-term memory network algorithms. *Applied Energy* 253, 113541.
- Gopinathan, K.K., 1988. A simple method for predicting global solar radiation on a horizontal surface. *Solar & wind technology* 5, 581–583.
- Haider, S.A., Naqvi, S.R., Akram, T., Umar, G.A., Shahzad, A., Sial, M.R., Khaliq, S., Kamran, M., 2019. LSTM neural network based forecasting model for wheat production in Pakistan. *Agronomy* 9, 72.
- Hargreaves, G.H., Samani, Z.A., 1982. Estimating potential evapotranspiration. *Journal of the irrigation and Drainage Division* 108, 225–230.
- Hargreaves, G.L., Hargreaves, G.H., Riley, J.P., 1985. Irrigation water requirements for Senegal River basin. *Journal of Irrigation and Drainage Engineering* 111, 265–275.
- Hassan, G.E., Youssef, M.E., Mohamed, Z.E., Ali, M.A., Hanafy, A.A., 2016. New temperature-based models for predicting global solar radiation. *Applied energy* 179, 437–450.
- Hassan, Muhammed A, Khalil, A., Kaseb, S., Kassem, M., 2017. Exploring the potential of tree-based ensemble methods in solar radiation modeling. *Applied Energy* 203, 897–916.
- Hassan, Muhammed A., Khalil, A., Kaseb, S., Kassem, M.A., 2017. Potential of four different machine-learning algorithms in modeling daily global solar radiation. *Renewable energy* 111, 52–62.
- Hay, J.E., 1993. Solar radiation data: validation and quality control. *Renewable energy* 3, 349–355.
- Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. *Neural computation* 9, 1735–1780.
- Hornik, K., Stinchcombe, M., White, H., 1989. Multilayer feedforward networks are universal approximators. *Neural networks* 2, 359–366.
- Huber, P.J., 1964. Robust estimation of a location parameter: *Annals Mathematics Statistics*, 35.
- Husein, M., Chung, I.-Y., 2019. Day-ahead solar irradiance forecasting for microgrids using a long short-term memory recurrent neural network: A deep learning approach. *Energies* 12, 1856.

- Ihya, B., Mechaqrane, A., Tadili, R., Bargach, M., 2015. Prediction of hourly and daily diffuse solar fraction in the city of Fez (Morocco). *Theoretical and applied climatology* 120, 737–749.
- Iqbal, M., 2012. *An introduction to solar radiation*. Elsevier.
- Jamil, B., Akhtar, N., 2017. Comparative analysis of diffuse solar radiation models based on sky-clearness index and sunshine period for humid-subtropical climatic region of India: A case study. *Renewable and Sustainable Energy Reviews* 78, 329–355.
- Jamil, B., Siddiqui, A.T., 2017. Generalized models for estimation of diffuse solar radiation based on clearness index and sunshine duration in India: applicability under different climatic zones. *Journal of Atmospheric and Solar-Terrestrial Physics* 157, 16–34.
- Jeff Hu, 2017. RNN & LSTM Neural Network for Sequential Data
[“https://fr.slideshare.net/YaochiehHu/rnn-lstm-neural-network-for-sequential-data”](https://fr.slideshare.net/YaochiehHu/rnn-lstm-neural-network-for-sequential-data)
 (consulté le 30/06/2021).
- Journée, M., Bertrand, C., 2011. Quality control of solar radiation data within the RMIB solar measurements network. *Solar Energy* 85, 72–86.
- Kaba, K., Sarigül, M., Avci, M., Kandirmaz, H.M., 2018. Estimation of daily global solar radiation using deep learning model. *Energy* 162, 126–135.
- Kainen, P.C., Kůrková, V., Sanguineti, M., 2013. Approximating multivariable functions by feedforward neural nets, in: *Handbook on Neural Information Processing*. Springer, pp. 143–181.
- Kalogirou, S.A., 2013. *Solar energy engineering: processes and systems*. Academic Press.
- Karasu, S., Altan, A., 2019. Recognition model for solar radiation time series based on random forest with feature selection approach, in: *2019 11th International Conference on Electrical and Electronics Engineering (ELECO)*. IEEE, pp. 8–11.
- Kheradmanda, S., Nematollahi, O., Ayoobia, A.R., 2016. Clearness index predicting using an integrated artificial neural network (ANN) approach. *Renewable and Sustainable Energy Reviews* 58, 1357–1365.
- Khorasanizadeh, H., Mohammadi, K., 2016. Diffuse solar radiation on a horizontal surface: Reviewing and categorizing the empirical models. *Renewable and Sustainable Energy Reviews* 53, 338–362.
- Khorasanizadeh, H., Mohammadi, K., Goudarzi, N., 2016. Prediction of horizontal diffuse solar radiation using clearness index based empirical models; A case study. *International Journal of Hydrogen Energy* 41, 21888–21898.
- Khosravi, A., Koury, R.N.N., Machado, L., Pabon, J.J.G., 2018. Prediction of hourly solar radiation in Abu Musa Island using machine learning algorithms. *Journal of Cleaner Production* 176, 63–75.
- Kolebaje, O., Ikusika, A., Akinyemi, P., 2016. Estimating solar radiation in Ikeja and Port Harcourt via correlation with relative humidity and temperature. *International Journal of Energy Production and Management* 1, 253–262.
- Korachagaon, I., Bapat, V., 2012. General formula for the estimation of global solar radiation on earth’s surface around the globe. *Renewable energy* 41, 394–400.
- Kumar, S., Kaur, T., 2016. Development of ANN based model for solar potential assessment using various meteorological parameters. *Energy Procedia* 90, 587–592.
- Laiti, L., Andreis, D., Zottele, F., Giovannini, L., Panziera, L., Toller, G., Zardi, D., 2014. A solar atlas for the Trentino region in the Alps: quality control of surface radiation data. *Energy Procedia* 59, 336–343.
- Lee, J., Wang, W., Harrou, F., Sun, Y., 2020. Reliable solar irradiance prediction using ensemble learning-based models: A comparative study. *Energy Conversion and Management* 208, 112582.

- Lemos, L.F., Starke, A.R., Boland, J., Cardemil, J.M., Machado, R.D., Colle, S., 2017. Assessment of solar radiation components in Brazil using the BRL model. *Renewable Energy* 108, 569–580.
- Lewis, G., 1983. Estimates of irradiance over Zimbabwe. *Sol. Energy;(United Kingdom)* 31.
- Li, C., Zhang, Y., Ren, X., 2020. Modeling hourly soil temperature using deep BiLSTM neural network. *Algorithms* 13, 173.
- Li, C., Zhang, Y., Zhao, G., Ren, Y., 2021. Hourly solar irradiance prediction using deep BiLSTM network. *Earth Science Informatics* 14, 299–309.
- Li, H., Cao, F., Bu, X., Zhao, L., 2015. Models for calculating daily global solar radiation from air temperature in humid regions—A case study. *Environmental Progress & Sustainable Energy* 34, 595–599.
- Li, M.-F., Tang, X.-P., Wu, W., Liu, H.-B., 2013. General models for estimating daily global solar radiation for different solar radiation zones in mainland China. *Energy conversion and management* 70, 139–148.
- Linares-Rodríguez, A., Ruiz-Arias, J.A., Pozo-Vázquez, D., Tovar-Pescador, J., 2011. Generation of synthetic daily global solar radiation data based on ERA-Interim reanalysis and artificial neural networks. *Energy* 36, 5356–5365.
- Liu, B., Jordan, R., 1961. Daily insolation on surfaces tilted towards equator. *ASHRAE J.:(United States)* 10.
- Liu, B.Y., Jordan, R.C., 1960. The interrelationship and characteristic distribution of direct, diffuse and total solar radiation. *Solar energy* 4, 1–19.
- Liu, D., Sun, K., 2019. Random forest solar power forecast based on classification optimization. *Energy* 187, 115940.
- Liu, X., Mei, X., Li, Y., Wang, Q., Jensen, J.R., Zhang, Y., Porter, J.R., 2009. Evaluation of temperature-based global solar radiation models in China. *Agricultural and forest meteorology* 149, 1433–1446.
- Long, C.N., Dutton, E.G., 2010. BSRN Global Network recommended QC tests, V2. x.
- Long, C.N., Shi, Y., 2008. An automated quality assessment and control algorithm for surface radiation measurements. *The Open Atmospheric Science Journal* 2.
- Louche, A., Notton, G., Poggi, P., Simonnot, G., 1991. Correlations for direct normal and global horizontal irradiation on a French Mediterranean site. *Solar Energy* 46, 261–266.
- M. Amara, 2015. The integration of Morocco into the Mediterranean energy market constitutes one of the major axes of its energy strategy, Ministry of Energy, Mines, Water and Environment, Kingdom of Morocco, 2015. [Online]. Available: <http://www.mem.gov.ma/SitePages/Activites2015/Act11Dec15.aspx> [Consulté le 15/01/2019].
- Maamar, L., Salah, H., Nawal, C., Omar, N., 2014. Predicting global solar radiation for North Algeria. *algorithms* 3, 4.
- Madugu, I.S., Olufeagba, B.J., Adediran, Y.A., Abdulkadir, F., Abdulkarim, A., Inaku, J.U., Lawan, A.U., 2019. A novel model for solar radiation prediction. *Telkomnika* 17, 3100–3109.
- Mahmood, R., Hubbard, K.G., 2002. Effect of time of temperature observation and estimation of daily solar radiation for the Northern Great Plains, USA. *Agronomy Journal* 94, 723–733.
- Marques Filho, E.P., Oliveira, A.P., Vita, W.A., Mesquita, F.L., Codato, G., Escobedo, J.F., Cassol, M., França, J.R.A., 2016. Global, diffuse and direct solar radiation at the surface in the city of Rio de Janeiro: Observational characterization and empirical modeling. *Renewable Energy* 91, 64–74.

- Marquez, R., Coimbra, C.F., 2011. Forecasting of global and direct solar irradiance using stochastic learning methods, ground experiments and the NWS database. *Solar Energy* 85, 746–756.
- Marsland, S., 2015. *Machine learning: an algorithmic perspective*. CRC press.
- Meenal, R., Boazina, P.G., Selvakumar, A.I., 2016. Temperature based radiation models for the estimation of global solar radiation at horizontal surface in India. *Indian Journal of Science and Technology* 9, 1–6.
- Meeus, J., 1998. *Astronomical algorithms*, *Astronomical algorithms by J. Meeus*. Richmond, VA: Willmann-Bell.
- Mellit, A., Pavan, A.M., 2010. 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* 84, 807–821.
- Michalsky, J.J., 1988. The astronomical almanac's algorithm for approximate solar position (1950–2050). *Solar energy* 40, 227–235.
- Mirás-Avalos, J.M., Rodríguez-Gómez, B.A., del Carmen Meizoso-López, M., Sande-Fouz, P., González-García, M.Á., Paz-González, A., 2012. Data quality assessment and monthly stability of ground solar radiation in Galicia (NW Spain). *Solar energy* 86, 3499–3511.
- Mohammadi, K., Shamshirband, S., Kamsin, A., Lai, P., Mansor, Z., 2016. Identifying the most significant input parameters for predicting global solar radiation using an ANFIS selection procedure. *Renewable and Sustainable Energy Reviews* 63, 423–434.
- Mohammadi, K., Shamshirband, S., Tong, C.W., Arif, M., Petković, D., Ch, S., 2015. A new hybrid support vector machine–wavelet transform approach for estimation of horizontal global solar radiation. *Energy Conversion and Management* 92, 162–171.
- Moradi, I., 2009. Quality control of global solar radiation using sunshine duration hours. *Energy* 34, 1–6.
- Moreno-Tejera, S., Ramírez-Santigosa, L., Silva-Pérez, M., 2015. A proposed methodology for quick assessment of timestamp and quality control results of solar radiation data. *Renewable Energy* 78, 531–537.
- Mubiru, J., Banda, E., D'Ujanga, F., Senyonga, T., 2007. Assessing the performance of global solar radiation empirical formulations in Kampala, Uganda. *Theoretical and Applied Climatology* 87, 179–184.
- Muneer, T., 2007. *Solar radiation and daylight models*. Routledge.
- Muneer, T., Younes, S., Munawwar, S., 2007. Discourses on solar radiation modeling. *Renewable and Sustainable Energy Reviews* 11, 551–602.
- Nwokolo, S.C., Ogbulezie, J.C., 2018. A qualitative review of empirical models for estimating diffuse solar radiation from experimental data in Africa. *Renewable and Sustainable Energy Reviews* 92, 353–393.
- Ododo, J.C., 1997. Prediction of solar radiation using only maximum temperature and relative humidity: South-east and north-east Nigeria. *Energy Conversion and Management* 38, 1807–1814.
- Ögelman, H., Ecevit, A., Tasdemiroğlu, E., 1984. A new method for estimating solar radiation from bright sunshine data. *Solar Energy* 33, 619–625.
- Ohmura, A., Dutton, E.G., Forgan, B., Fröhlich, C., Gilgen, H., Hegner, H., Heimo, A., König-Langlo, G., McArthur, B., Müller, G., 1998. *Baseline Surface Radiation Network (BSRN/WCRP): New precision radiometry for climate research*. *Bulletin of the American Meteorological Society* 79, 2115–2136.
- Olatomiwa, L., Mekhilef, S., Shamshirband, S., Mohammadi, K., Petković, D., Sudheer, C., 2015. A support vector machine–firefly algorithm-based model for global solar radiation prediction. *Solar Energy* 115, 632–644.

- Page, J.K., Albuissou, M., Wald, L., 2001. The European Solar Radiation Atlas: a valuable digital tool. *Solar Energy* 71, 81–83. [https://doi.org/10.1016/S0038-092X\(00\)00157-2](https://doi.org/10.1016/S0038-092X(00)00157-2)
- Pashiardis, S., Kalogirou, S.A., 2016. Quality control of solar shortwave and terrestrial longwave radiation for surface radiation measurements at two sites in Cyprus. *Renewable Energy* 96, 1015–1033.
- Perez-Astudillo, D., Bachour, D., Martin-Pomares, L., 2018. Improved quality control protocols on solar radiation measurements. *Solar Energy* 169, 425–433. <https://doi.org/10.1016/j.solener.2018.05.028>
- Pinkus, A., 1999. Approximation theory of the MLP model in neural networks. *Acta numerica* 8, 143–195.
- Pisoni, E., Farina, M., Carnevale, C., Piroddi, L., 2009. Forecasting peak air pollution levels using NARX models. *Engineering Applications of Artificial Intelligence* 22, 593–602.
- Prescott, J., 1940. Evaporation from a water surface in relation to solar radiation. *Trans. Roy. Soc. S. Aust.* 46, 114–118.
- Principe de la mesure du rayonnement solaire incident, 2021. “<http://www.precis-mecanique.com/specialites/rayonnement-solaire-energie-duree/principe-de-la-mesure-du-rayonnement-solaire-incident.html>” (consulté le 14/10/2021) [WWW Document].
- Qing, X., Niu, Y., 2018. Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM. *Energy* 148, 461–468.
- Quej, V.H., Almorox, J., Ibrakhimov, M., Saito, L., 2016. Empirical models for estimating daily global solar radiation in Yucatán Peninsula, Mexico. *Energy conversion and management* 110, 448–456.
- Rahimikhoob, A., 2010. Estimating global solar radiation using artificial neural network and air temperature data in a semi-arid environment. *Renewable Energy* 35, 2131–2135.
- Reda, I., Andreas, A., 2004. Solar position algorithm for solar radiation applications. *Solar energy* 76, 577–589.
- Rehman, S., Mohandes, M., 2008. Artificial neural network estimation of global solar radiation using air temperature and relative humidity. *Energy Policy* 36, 571–576.
- Ren, Y., Suganthan, P.N., Srikanth, N., 2015. Ensemble methods for wind and solar power forecasting—A state-of-the-art review. *Renewable and Sustainable Energy Reviews* 50, 82–91.
- REN21, 2018. http://www.ren21.net/wp-content/uploads/2018/06/17-8652_GSR2018_FullReport_web_final_.pdf (consulté le 08/05/2019).
- Ridley, B., Boland, J., Lauret, P., 2010. Modelling of diffuse solar fraction with multiple predictors. *Renewable Energy* 35, 478–483.
- Ruiz, L.G.B., Cuéllar, M.P., Calvo-Flores, M.D., Jiménez, M.D.C.P., 2016. An application of non-linear autoregressive neural networks to predict energy consumption in public buildings. *Energies* 9, 684.
- Ruiz-Arias, J., Alsamamra, H., Tovar-Pescador, J., Pozo-Vázquez, D., 2010. Proposal of a regressive model for the hourly diffuse solar radiation under all sky conditions. *Energy Conversion and Management* 51, 881–893.
- Salcedo-Sanz, S., Deo, R.C., Cornejo-Bueno, L., Camacho-Gómez, C., Ghimire, S., 2018. An efficient neuro-evolutionary hybrid modelling mechanism for the estimation of daily global solar radiation in the Sunshine State of Australia. *Applied Energy* 209, 79–94.
- Sansa, I., Bellaaj, N.M., 2018. Solar Radiation Prediction Using NARX Model, in: *Advanced Applications for Artificial Neural Networks*. IntechOpen.
- Scharmer, K., Page, J.K., Wald, L., Albuissou, M., Czeplak, G., Bourges, B., Aguiar, R., Lund, H., Joukoff, A., Terzenbach, U., 2000. The European Solar Radiation Atlas Vol. 1: Fundamentals and maps. Presses de l’Ecole des Mines de Paris.

- Schüler, D., Wilbert, S., Geuder, N., Affolter, R., Wolfertstetter, F., Prah, C., Röger, M., Schroedter-Homscheidt, M., Abdellatif, G., Guizani, A.A., 2016. The enerMENA meteorological network–Solar radiation measurements in the MENA region. Presented at the AIP Conference Proceedings, AIP Publishing LLC, p. 150008.
- Seidlitz, H.K., Thiel, S., Krins, A., Mayer, H., 2001. Solar radiation at the Earth's surface. *Comprehensive series in photosciences* 3, 705–738.
- Shaddel, M., Javan, D.S., Baghernia, P., 2016. Estimation of hourly global solar irradiation on tilted absorbers from horizontal one using Artificial Neural Network for case study of Mashhad. *Renewable and Sustainable Energy Reviews* 53, 59–67.
- Sharifi, S.S., Rezaverdinejad, V., Nourani, V., 2016. Estimation of daily global solar radiation using wavelet regression, ANN, GEP and empirical models: A comparative study of selected temperature-based approaches. *Journal of Atmospheric and Solar-Terrestrial Physics* 149, 131–145.
- Skeiker, K., 2006. Correlation of global solar radiation with common geographical and meteorological parameters for Damascus province, Syria. *Energy conversion and management* 47, 331–345.
- Sözen, A., Arcaklioğlu, E., 2005. Solar potential in Turkey. *Applied Energy* 80, 35–45.
- Srivastava, S., Lessmann, S., 2018. A comparative study of LSTM neural networks in forecasting day-ahead global horizontal irradiance with satellite data. *Solar Energy* 162, 232–247.
- Stafford, B., Davis, M., Chambers, J., Martínez, M., Sanchez, D., 2009. Tracker accuracy: field experience, analysis, and correlation with meteorological conditions, in: 2009 34th IEEE Photovoltaic Specialists Conference (PVSC). IEEE, pp. 002256–002259.
- Thermopile Pyranometer calibration by ITC India Pvt Ltd, 2019. . Electrical Safety Testing Laboratory. URL <https://www.itcindia.org/thermopile-pyranometer-calibration/> (accessed 5.19.21).
- Thiaw, L., 2008. Identification de systemes dynamiques non linéaires par réseaux de neurones et multimodeles (PhD Thesis).
- Toledo, C., Gracia Amillo, A.M., Bardizza, G., Abad, J., Urbina, A., 2020. Evaluation of solar radiation transposition models for passive energy management and building integrated photovoltaics. *Energies* 13, 702.
- Torres-Barrán, A., Alonso, Á., Dorronsoro, J.R., 2019. Regression tree ensembles for wind energy and solar radiation prediction. *Neurocomputing* 326, 151–160.
- Trouillet, F., Vallée, J.-M., 2008. Mesure de la constante solaire "http://aces.ens-lyon.fr/acces/thematiques/paleo/systemclim/effet-de-serre/data/constante_solaire_flo.pdf "(consulté le 14/10/2021).
- Vapnik, V., 2013. *The nature of statistical learning theory*. Springer science & business media.
- Vapnik, V.N., 1998. *Statistical Learning Theory*, New York: Wiley.
- Vapnik, V.N., 1995. *The nature of statistical learning. Theory*.
- Voyant, C., 2011. Prédiction de séries temporelles de rayonnement solaire global et de production d'énergie photovoltaïque à partir de réseaux de neurones artificiels (PhD Thesis). Université Pascal Paoli.
- Voyant, C., Notton, G., Kalogirou, S., Nivet, M.-L., Paoli, C., Motte, F., Fouilloy, A., 2017. Machine learning methods for solar radiation forecasting: A review. *Renewable Energy* 105, 569–582.
- Voyant, C., Soubdhan, T., Lauret, P., David, M., Muselli, M., 2015. Statistical parameters as a means to a priori assess the accuracy of solar forecasting models. *Energy* 90, 671–679.
- Yadav, A.K., Malik, H., Chandel, S., 2015. Application of rapid miner in ANN based prediction of solar radiation for assessment of solar energy resource potential of 76

- sites in Northwestern India. *Renewable and Sustainable Energy Reviews* 52, 1093–1106.
- Yadav, A.K., Malik, H., Chandel, S.S., 2014. Selection of most relevant input parameters using WEKA for artificial neural network based solar radiation prediction models. *Renewable and Sustainable Energy Reviews* 31, 509–519.
- Yang, L., Cao, Q., Yu, Y., Liu, Y., 2020. Comparison of daily diffuse radiation models in regions of China without solar radiation measurement. *Energy* 191, 116571.
- Younes, S., Claywell, R., Muneer, T., 2005. Quality control of solar radiation data: Present status and proposed new approaches. *Energy* 30, 1533–1549.
- Zabara, K., 1986. Estimation of the global solar radiation in Greece. *Solar & wind technology* 3, 267–272.
- Zendehboudi, A., Baseer, M.A., Saidur, R., 2018. Application of support vector machine models for forecasting solar and wind energy resources: A review. *Journal of cleaner production* 199, 272–285.
- Zhou, Y., Liu, Y., Wang, D., Liu, X., Wang, Y., 2021. A review on global solar radiation prediction with machine learning models in a comprehensive perspective. *Energy Conversion and Management* 235, 113960.
- Zhou, Z.-H., 2012. *Ensemble methods: foundations and algorithms*. CRC press.
- Zobaa, A.F., Bansal, R.C., 2011. *Handbook of renewable energy technology*. World Scientific.

Annexes

Annexe A : Algorithme SPA (Solar Position Algorithm)

Intitulé de l'équation	Equation
Le temps terrestre (TT)	$TT = TAI + 233.148$ Avec TAI est le temps atomique international
Le temps universel (UT)	$UT = TT - \Delta T$ Avec ΔT est la différence entre le TT et le temps de la rotation de la terre. $UT = UTC - \Delta UT$ Avec ΔUT est une valeur positive ou négative, qui est ajouté au temps universel coordonnée (UTC) pour régler le taux de rotation irrégulier de la terre.
La date (D)	$D = jour + \frac{heure}{24} + \frac{minute}{24 * 60} + \frac{second}{24 * 60 * 60}$
Le jour Julien (JD)	$JD = int(365.25 * annee + 4716)) + (30.6001 * (mois + 1)) + D + B - 1524.5$ - Si le mois est inférieur à 2 => mois=mois+12 et annee =annee-1. - Si le mois est supérieur à 2 => mois=mois et annee =annee. - B est égal à 0, pour le calendrier julien JD inférieur ou égal à 2299160. - B est égal à (2-A+int(A/4)), pour le calendrier grégorien (JD > 2299160). - A =int (annee/100).
Le jour julien des Ephémérides (JDE)	$JDE = JD + \frac{\Delta T}{86400}$
Le siècle (JC)	$JC = \frac{JD - 2451545}{36525}$
Ephémérides siècle (JCE)	$JCE = \frac{JDE - 2454545}{36525}$
Ephémérides millénaire (JME)	$JME = \frac{JCE}{10}$
Le terme $L0i$	$L0i = Ai * \cos(Bi + Ci * JEM)$ Avec Ai, Bi et Ci sont les valeurs déterminées à partir des termes de rotation de la Terre
Le terme $L0$	$L0 = \sum L0i$
La longitude héliocentrique de la terre (L)	$L = \frac{L0 + L1 * JME + L2 * JME^2 + L3 * JME^3 + L4 * JME^4 + L5 * JME^5}{10^8}$
La fraction décimale de la division (F)	$F = L/360$ -Si L est positif => L=360*F -Si L est négatif => L=360-360*F
La longitude géocentrique (en degré)	$\theta = L + 180$
La latitude géocentrique (en degré)	$\beta = -B$
L'allongement moyen de la lune du soleil (en degré)	$X0 = 297.85036 + 44527.112480 * JCE - 0.0019142 * JCE^2 + \frac{JCE^3}{189474}$
L'anomalie moyenne du soleil (Terre) (en degré)	$X1 = 357.52772 + 35999 * JCE.050340 * JCE^2 - \frac{JCE^3}{300000}$

Intitulé de l'équation	Equation
L'anomalie moyenne de la lune (degré)	$X2 = 134.96298 + 477198.867398 * JCE + 0.008672 * JCE^2 + \frac{JCE^3}{56250}$
L'argument de la lune sur la latitude (en degré)	$X3 = 93.27191 + 483202.017538 * JCE - 0.0036825 * JCE^2 + \frac{JCE^3}{327270}$
La longitude du nœud ascendant de l'orbite moyenne de la lune sur l'écliptique (en degré)	$X4 = 125.04452 - 1934.136261 * JCE + 0.0020708 * JCE^2 + \frac{JCE^3}{450000}$
Le terme $\Delta\psi_i$	$\Delta\psi_i = (a_i + b_i * JCE) * \sin(\sum (X_j * Y_{i,j}))$
Le terme $\Delta\epsilon_i$	$\Delta\epsilon_i = (c_i + d_i * JCE) * \cos(\sum (X_j * Y_{i,j}))$
La nutation en longitude (en degré)	$\Delta\psi = \sum_{i=0}^{i=n} \frac{\Delta\psi_i}{36000000}$
La nutation en obliquité (en degré)	$\Delta\epsilon = \sum_{i=0}^{i=n} \frac{\Delta\epsilon_i}{36000000}$
L'obliquité moyenne de l'écliptique (en degré)	$\epsilon_0 = 84381.448 - 4680.93 * U - 1.55 * U^2 + 1999.25 * U^3 - 51.38 * U^4 - 249.67 * U^5 - 39.05 * U^6 + 7.12 * U^7 + 27.87 * U^8 + 5.79 * U^9 + 2.45 * U^{10}$ avec $U = \frac{JEM}{100}$
L'obliquité réelle de l'écliptique (en degré)	$\epsilon = \frac{\epsilon_0}{3600} + \Delta\epsilon$
La correction d'aberration (en degré)	$\Delta\tau = -\frac{20.4898}{3600 * R}$
La longitude apparente du soleil (en degré)	$\lambda = \theta + \Delta\psi + \Delta\tau$
Le temps sidéral moyen à Greenwich (en degré)	$V_0 = 280.46061837 + 360.98564736629 * (JD - 241545) + 0.000387933 * JC^2 - \frac{JC^3}{38710000}$
Le temps apparent de la sidère à Greenwich (en degré)	$V = V_0 + \Delta\psi * \cos\epsilon$
L'ascension droit du soleil (en radian)	$\alpha = \arctan 2\left(\frac{\sin\lambda \cos\epsilon - \tan\beta \sin\epsilon}{\cos\lambda}\right)$
La déclinaison géocentrique du soleil (en degré)	$\delta = \arcsin(\sin\beta + \cos\beta \sin\epsilon \sin\lambda)$ Avec δ est positif ou négatif si le soleil se trouve au nord ou au sud de l'équateur céleste, respectivement.
L'angle horaire locale de l'observateur (en degré)	$H = v + \sigma - \alpha$ Avec α est la longitude géographique de l'observateur, positif ou négatif pour l'est ou l'ouest de Greenwich, respectivement.
La parallaxe horizontale équatoriale du soleil (en degré)	$\xi = \frac{8.794}{3600 * R}$
Le terme u (en degré)	$u = \arctan(0.99664719 * \tan\phi)$ Avec ϕ est la latitude géographique de l'observateur, positif ou négatif au nord ou au sud de l'équateur, respectivement.
Le terme x	$x = \cos u + \frac{E}{6378140} * \cos\phi = \rho * \cos\phi'$ Avec E est l'élévateur de l'observateur, ρ est la distance de l'observateur au centre de la terre et ϕ' est la latitude géocentrique de l'observateur.
Le terme y	$y = 0.99664719 * \sin u + \frac{E}{6378140} * \sin\phi$

Intitulé de l'équation	Equation
La parallaxe dans l'ascension droite du soleil (en degré)	$\Delta\alpha = \arctan2\left(\frac{-x * \sin\xi * \sin H}{\cos\delta - x * \sin\xi * \cos H}\right)$
L'ascension droite au soleil topocentrique (en degré)	$\alpha' = \alpha + \Delta\alpha$
La déclinaison du soleil topocentrique (en degré)	$\delta' = \arctan2\left(\frac{(\sin\delta - y * \sin\xi) * \cos\Delta\alpha}{\cos\delta - y * \sin\xi * \sin\xi * \cos H}\right)$
L'angle horaire de locale topocentrique (en degré)	$H' = H - \Delta\alpha$
L'angle d'élévation topocentrique sans correction de réfraction atmosphérique (en degré)	$e_0 = \arcsin(\sin\phi * \sin\delta' + \cos\phi * \cos\delta' * \cos H)$
La correction de la réfraction atmosphérique (en degré)	$\Delta e = \frac{P}{1010} * \frac{283}{273 + T} * \frac{1.02}{60 * \tan\left(e_0 + \frac{10.3}{e_0 + 5.11}\right)}$ <p>Avec P est la pression moyenne du local (en mbar) et T est la température annuelle moyenne du local (en °C)</p>
L'angle d'élévation topocentrique (en degré)	$e = e_0 + \Delta e$
L'angle zénith topocentrique (en degré)	$\theta = 90 - e$
L'angle de l'azimut des astronomes topocentriques (en degré)	$\Gamma = \arctan2\left(\frac{\sin H'}{\cos H' * \sin\phi - \tan\delta' * \cos\phi}\right)$
L'angle d'azimut topocentrique pour les navigateurs et les utilisateurs de rayonnement solaire (en degré)	$\psi = \Gamma + 180$
L'angle d'incidence pour une surface orientée dans n'importe quelle direction (en degré)	$I = \arccos(\cos\theta * \cos\omega * \sin\theta * \cos(\Gamma - \sigma))$ <p>Avec ω est la pente de la surface mesurée à partir du plan horizontal, σ est l'angle de rotation de l'azimut de surface</p>

Annexe B: Algorithme SG2 (Solar Geometry 2)

Intitulé de l'équation	Equation
1. Coordonnées héliocentriques	
Rayon héliocentrique du soleil	$R = R_{SG2} = \rho^R \cos(2\pi f^R j_{TT}^* - \phi^R) + a^R j_{TT}^* + b^R$ <p style="text-align: center;">Avec : $j_{TT}^* = j_{TT} - 2444239.5$</p> $j_{TT} = j_{UT} + \frac{\Delta_{TT}}{86400}$ $a^R = 0 ; b^R = 1.000140 ; f^R = 1/365.254902 ;$ $\rho^R = 0.016704 ; \phi^R = -3.091159$
Longitude héliocentrique du soleil	$L = L_{SG2} = \sum_{k=1}^{10} \rho_k^L \cos(2\pi f_k^L j_{TT}^* - \phi_k^L) + a^L j_{TT}^* + b^L$ <p>Avec : $a^L = 1/58.130101 ; b^L = 1.742145$</p> <p>Pour</p> <p>k=1 : $f_k^L = \frac{1}{365.261278} ; \rho_k^L = 3.401508 \times 10^{-2} ; \phi_k^L = 1.600780$ k=2 : $f_k^L = \frac{1}{182.632412} ; \rho_k^L = 3.486440 \times 10^{-4} ; \phi_k^L = 1.662976$ k=3 : $f_k^L = \frac{1}{29.530634} ; \rho_k^L = 3.136227 \times 10^{-5} ; \phi_k^L = -1.195905$ k=4 : $f_k^L = \frac{1}{399.529850} ; \rho_k^L = 3.578979 \times 10^{-5} ; \phi_k^L = -1.042052$ k=5 : $f_k^L = \frac{1}{291.956812} ; \rho_k^L = 2.676185 \times 10^{-5} ; \phi_k^L = 2.012613$ k=6 : $f_k^L = \frac{1}{583.598201} ; \rho_k^L = 2.333925 \times 10^{-5} ; \phi_k^L = -2.867714$ k=7 : $f_k^L = \frac{1}{4652.629372} ; \rho_k^L = 1.221214 \times 10^{-5} ; \phi_k^L = 1.225038$ k=8 : $f_k^L = \frac{1}{1450.236684} ; \rho_k^L = 1.217941 \times 10^{-5} ; \phi_k^L = -0.828601$ k=9 : $f_k^L = \frac{1}{199.459709} ; \rho_k^L = 1.343914 \times 10^{-5} ; \phi_k^L = -3.108253$ k=10 : $f_k^L = \frac{1}{365.355291} ; \rho_k^L = 8.499475 \times 10^{-4} ; \phi_k^L = -2.353709$</p>
2. Paramètres géocentriques	
Correction de l'aberration stellaire	$\Delta\tau = \Delta\tau_{SG2} = -9.933735 \times 10^{-5} \text{ rad}$
Parallaxe horizontale équatoriale du soleil	$\xi = \xi_{SG2} = 4.263521 \times 10^{-5} \text{ rad}$
Nutation dans la longitude géocentrique du soleil	$\Delta\psi = \Delta\psi_{SG2} = \rho^{\Delta\psi} \cos(2\pi f^{\Delta\psi} j_{TT}^* - \phi^{\Delta\psi})$ <p>Avec : $f^{\Delta\psi} = \frac{1}{6791.164405} ; \rho^{\Delta\psi} = 8.329092 \times 10^{-5} ; \phi^{\Delta\psi} = -2.052757$</p>
Obliquité terrestre réelle	$\varepsilon = \varepsilon_{SG2} = \rho^\varepsilon \cos(2\pi f^\varepsilon j_{TT}^* - \phi^\varepsilon) + a^\varepsilon j_{TT}^* + b^\varepsilon$ <p>Avec : $a^\varepsilon = -6.216374 \times 10^{-9} ; b^\varepsilon = 4.091383 \times 10^{-1}$ $f^\varepsilon = \frac{1}{6791.164405} ; \rho^\varepsilon = 4.45618 \times 10^{-5} ; \phi^\varepsilon = 2.660352$</p>
Longitude géocentrique du Soleil apparent	$\Theta_a = L + \pi + \Delta\psi + \Delta\tau$
Déclinaison géocentrique solaire	$\delta^g = \text{asin}(\sin(\Theta_a) \sin(\varepsilon))$
Ascension droite géocentrique du Soleil	$r_a^g = \text{atan2}(\sin(\Theta_a) \cos(\varepsilon), \cos(\Theta_a))$
3. Temps sidéral	
Temps sidéral moyen	$v_0 = v_{0-SG2} = 6.3000388 j_{TT}^* + 1.742079$

Intitulé de l'équation	Equation
4. Paramètres topo-centriques	
Effets de parallaxe dans l'ascension droite topo-centrique du soleil	$\Delta r_{\alpha} = \Delta r_{\alpha-SG2} = -x \frac{\sin \omega^g}{\cos \delta^g} \xi$
Déclinaison topo-centrique du soleil	$\delta = \delta_{SG2} = \delta^g + (x \cos \omega^g \sin \delta^g - y \cos \delta^g) \xi$
Angle horaire topo-centrique du soleil	$\omega = v_0 + \Delta\psi \cos \varepsilon - r_{\alpha}^g + \lambda - \Delta r_{\alpha}$
Azimut topo-centrique du soleil	$\alpha_s = atan2(\sin \omega, \cos \omega \sin \varphi - \tan \delta \cos \varphi) + \pi$
Angle d'élévation topo-centrique du soleil sans correction de la réfraction atmosphérique	$\gamma_{s0} = asin(\sin \varphi \sin \delta + \cos \varphi \cos \delta \cos \omega)$
	Pour : $\gamma_{s0} \geq -0.01 \text{ rad}$
	$\Delta\gamma_s = \frac{P}{283} \Delta\gamma_{S-SG2}^+ = \frac{2.96706 \times 10^{-4}}{1010273 + T \tan(\gamma_{s0} + 0.0031376(\gamma_{s0} + 0.089186))^{-1}}$
Terme de correction de la réfraction atmosphérique	
	Pour : $\gamma_{s0} < -0.01 \text{ rad}$
	$\Delta\gamma_s = \Delta\gamma_{S-SG2}^+ = -\frac{P}{1010273 + T} \frac{1.005516 \times 10^{-4}}{\tan \gamma_{s0}}$
Angle d'élévation topo-centrique du soleil	$\gamma_s = \gamma_{s0} + \Delta\gamma_s$

Annexe C: Algorithme Yallop

Intitulé de l'équation	Equation
Heure en décimal et le temps universel	$heure = heure + \frac{minute}{60} + \frac{seconde}{3600}$ $UT = heure + \frac{longitude}{15} + TZ$ <p>Avec TZ: Time zone (http://www.timeanddate.com/worldclock/)</p> <p>Si le mois est supérieur à 2 ou égal 1 alors mois=mois-3 ; année=année Sinon année=année-1 et mois = mois+12</p> $temps = \frac{\frac{UT}{24} + jour + int(30.6 * mois + 0.5) + int(365.25 * (année - 1967)) - 8707.5}{36525}$ $G = 357.528 + 35999.05 * temps$ $C = 1.915 \sin G + 0.020 \sin 2G$ $L = 280.460 + 36000.770 * temps + C$ $\alpha = L - 2.466 \sin 2L + 0.053 \sin 4L$
Angle horaire de Greenwich	$\omega GHA = 15 * UT - 180 - C + L - \alpha$ <p>Si $\omega < 0$ alors $x = \text{ceil}(-\omega GHA / 360)$; $\omega GHA = \omega GHA + 360 * x$</p> <p>Si $\omega > 0$ alors $x = \text{ceil}(\omega GHA / 360)$; $\omega GHA = \omega GHA - 360 * x$</p>
Obliquité de l'écliptique	$\epsilon = 23.4393 - 0.013 * temps$
Déclinaison du soleil	$\delta = \arctan(\tan \epsilon * \sin \alpha)$
Equation du temps	$EoT = (L - C - \alpha) * \frac{60}{15}$
Temps solaire vrai	$TSV = heure + \frac{EoT}{60} + \frac{longitude}{15} - TZ$
Angle horaire du lieu géographique	$\omega = 15 * (TSV - 12)$
Hauteur du soleil en degré (°)	$H = \arcsin(\sin(latitude) * \sin(\delta) + \cos(latitude) * \cos(\delta) * \cos(\omega))$

Intitulé de l'équation	Equation
Angle d'azimut du soleil en degré (°)	$A_z = \arctan2(\sin(\omega GHA), \cos(\omega GHA) * \sin(latitude) - \tan(\delta) * \cos(latitude))$
Angle de zénith degré (°)	$sunZenith = 90 - H$
Angle d'incidence en degré (°)	$I = \arccosd(\cosd(sunZenith) * \cosd(surface_{slope}) - \sind(surface_{slope}) * \sind(sunZenith) * \cosd(A_z - surface_{azimuth}))$
Temps du lever et du coucher	$TSV_{lever} = 12 - \frac{\frac{180}{\pi} * \arccos(-\tan(\delta) * \tan(latitude))}{15}$ $TSV_{coucher} = 12 + \frac{\frac{180}{\pi} * \arccos(-\tan(\delta) * \tan(latitude))}{15}$ $TU_{lever} = TSV_{lever} - \frac{EoT}{60} - \frac{longitude}{15}$ $TU_{coucher} = TSV_{coucher} - \frac{EoT}{60} - \frac{longitude}{15}$

Annexe D: Evaluation de la performance d'estimation et de prédiction

La performance d'un modèle d'estimation ou de prédiction des séries temporelles du rayonnement solaire peut être évalué en utilisant un ensemble de métriques qui ont été proposées par un certain nombre d'auteurs dans la littérature ((Chen and Yang, 2004),(Voyant et al., 2015)). Dans le présent travail, les différents indicateurs statistiques utilisés pour évaluer la précision d'estimation de chaque modèle développé sont décrits ci-dessous :

- a) **Le coefficient de détermination (R^2)** : est un indicateur permettant d'évaluer la qualité d'une régression linéaire simple. Plus sa valeur est proche de 1, plus l'équation de régression est adaptée pour décrire la distribution des points autour de la droite de régression. Alors que sa valeur proche de 0 signifie que le modèle utilisé n'explique absolument pas la distribution des points autour de la droite de régression.

$$R^2 = \frac{\sum_{i=1}^n (y_{mi} - y_{ei})^2}{\sum_{i=1}^n (y_{ei} - \bar{y}_m)^2} \quad (D-1)$$

y_{mi} et y_{ei} sont la $i^{\text{ème}}$ valeur mesurée et la $i^{\text{ème}}$ valeur estimée, respectivement. \bar{y}_m et \bar{y}_e sont la valeur moyenne de n mesures de y_{mi} ($\bar{y}_m = \frac{1}{n} \sum_{i=1}^n y_{mi}$) et la valeur moyenne de n estimations de y_{ei} ($\bar{y}_e = \frac{1}{n} \sum_{i=1}^n y_{ei}$), respectivement.

- b) **Le coefficient de corrélation (R)** : est un indicateur qui permet de mesurer la qualité de la relation linéaire entre deux variables. C'est un paramètre sans unité et compris entre -1 et 1. Plus la valeur de ce coefficient est proche de 0, plus la relation linéaire entre les variables est faible. Par contre, plus sa valeur est proche de 1 ou -1, plus la relation linéaire positive ou négative entre les variables est forte. La corrélation positive ($R > 0$) indique que les valeurs des deux variables augmentent ensemble, alors que, la corrélation négative ($R < 0$) indique que les valeurs d'une variable augmentent et que les valeurs de l'autre variable diminuent.

$$R = \frac{\sum_{i=1}^n (y_{ei} - \bar{y}_e)(y_{mi} - \bar{y}_m)}{\sqrt{\sum_{i=1}^n (y_{ei} - \bar{y}_e)^2 \sum_{i=1}^n (y_{mi} - \bar{y}_m)^2}} \quad (D-2)$$

- c) **La racine carrée de l'erreur quadratique moyenne (Root Mean square error : RMSE)** : est un indicateur qui indique à quel point les données sont concentrées autour de la ligne du meilleur ajustement. En d'autres termes, le RMSE mesure l'éloignement des points de données par rapport à la ligne de régression. Le RMSE est plus sensible aux erreurs de prévision importantes, et donc adaptée aux applications où les petites erreurs sont plus tolérables que les grandes (Benali et al., 2019).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_{ei} - y_{mi})^2} \quad (D-3)$$

RMSE représente un facteur de fiabilité qui est largement utilisé pour évaluer la performance d'un modèle pour prédire les variabilités temporelles de la ressource solaire. Mais, ce facteur ne peut pas être utilisé pour comparer les résultats des études qui sont effectuées sur des sites géographiquement différents. Dans ce cas, il est essentiel de normaliser le RMSE (Voyant, 2011). Il existe deux méthodes pour calculer le RMSE normalisé. Soit en utilisant la valeur moyenne du rayonnement ou les valeurs maximales et minimales du rayonnement. Dans notre travail, la valeur moyenne de l'irradiation \bar{y}_m a été utilisée comme référence pour normaliser le RMSE comme suit :

$$nRMSE = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (y_{ei} - y_{mi})^2}}{\bar{y}_m} \quad (D-4)$$

- d) **L'erreur moyenne absolue (Mean Absolute Error : notée MAE)** : est un facteur qui est souvent utilisé pour mesurer l'écart entre les valeurs mesurées et les valeurs prédites. La diminution de la valeur du MAE signifie une amélioration moyenne de la prédiction.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_{ei} - y_{mi}| \quad (D-5)$$

Comme dans le cas du RMSE, il est possible de normaliser la valeur de MAE en divisant par la valeur moyenne de l'irradiation que l'on nomme nMAE :

$$nMAE = \frac{\frac{1}{n} \sum_{i=1}^n |y_{ei} - y_{mi}|}{\bar{y}_m} \quad (D-6)$$

e) **Le biais moyen (Mean Bias Error : MBE)** : est une erreur qui représente la différence algébrique moyenne entre les données mesurées et prédites. Une valeur positive du MBE indique que l'estimation, en moyenne, surestime la mesure réelle, alors qu'une valeur négative signifie que l'estimation, en moyenne, sous-estime la mesure réelle. Aussi, une valeur nulle du MBE ne signifie pas forcément une très bonne estimation (Voyant, 2011). Les formules du biais moyen (MBE) et sa normalisation (nMBE) sont données par :

$$MBE = \frac{1}{n} \sum_{i=1}^n (y_{ei} - y_{mi}) \quad (D-7)$$

$$nMBE = \frac{\frac{1}{n} \sum_{i=1}^n (y_{ei} - y_{mi})}{\bar{y}_m} \quad (D-8)$$

f) **Erreur moyenne absolue en pourcentage (Mean absolute percentage error MAPE)** : est un indicateur qui mesure la moyenne des écarts en valeur absolue par rapport aux valeurs observées.

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_{ei} - y_{mi}}{y_{mi}} \right| \times 100 \quad (D-9)$$

Le MAE et le RMSE normalisés sont utilisés pour éviter la dépendance à l'échelle.

Plus R est proche de 1, et plus nMAE et nRMSE sont proches de 0, meilleure est la prédiction. Le nMAE donne une meilleure idée sur la précision de la prédiction. Le nRMSE indique la dispersion des données ou la variabilité de la précision de la prédiction. Un bon modèle est celui pour lequel nMAE et nRMSE sont les plus faibles possibles. Selon ((Li et al., 2013); (Despotovic et al., 2016)), la précision du modèle peut être divisée en quatre classes, comme le montre le tableau D.1.

Tableau D.1
Différents états de précision d'un modèle.

Classes	Valeurs de nRMSE (%)	Etat de précision
1	nRMSE < 10%	Excellent
2	10% < nRMSE < 20%	Bon
3	20% < nRMSE < 30%	Acceptable
4	nRMSE > 30%	Mauvais