

Université Mohamed Boudiaf - M'sila

FACULTE DE TECHNOLOGIE

DEPARTEMENT D'ELECTRONIQUE



Numéro de série :

Numéro d'inscription :

Thèse

Présentée pour l'obtention du diplôme de

DOCTORAT EN SCIENCES

Spécialité : Electronique

Option : Communication Avancée

THEME

UTILISATION DES METHODES D'APPRENTISSAGE MACHINE POUR LA PREDICTION DES SIGNAUX ENERGETIQUES

Présenté Par :

ZEMOURI Nahed

Soutenue le : 04 /11 /2020

Devant le jury composé de :

<u>Nom & Prénom</u>	<u>Grade</u>	<u>Etablissement</u>	<u>Qualité</u>
MEZACHE Amar	Professeur	Univ. de M'sila	Président
BOUZGOU Hassen	Professeur	Univ. de Batna2	Encadreur
CHOUDER Aissa	Professeur	Univ. de M'sila	Examineur
MOUSS Leila Hayet	Professeur	Univ. de Batna 2	Examineur
TAIBI Soufiane	Professeur	Univ. de Batna 2	Invité

Année Universitaire : 2019/2020

Remerciements

C'est avec un grand plaisir que je tiens à exprimer tout d'abord ma gratitude et mes sincères remerciements à mon directeur de thèse, le Professeur **Hassen BOUZGOU** de l'université de Batna 2, pour m'avoir fait partager son expérience, pour la pertinence et la richesse de ses remarques et pour ses conseils ainsi que pour son temps consacré à suivre toutes les étapes de ma thèse.

Un très grand merci à **Christian A. Gueymard**, directeur et président du laboratoire Solar Consulting Services (Florida, USA), par ses aides techniques, ses conseils et son soutien qui m'a grandement aidé pour la publication des mes travaux.

Je tiens à remercier les membres de jury qui ont accepté de s'intéresser à mon travail. Je leur en suis très reconnaissante. Je tiens à remercier **Amar MEZACHE**, Professeur à l'université de M'sila, pour m'avoir fait l'honneur de présider ce jury de thèse. Je voudrais, exprimer mes remerciements aussi à **Aissa CHOUDER**, professeur à l'université de M'Sila, à **Leila Hayet MOUSS**, professeur à l'université de Université de Batna 2, et à **Soufiane TAIBI** professeur à l'université de Batna 2.

Un très grand merci à **Hatem MEZACHE** Maître Assistant à l'université de M'sila, pour ses aides techniques, ses conseils tout au long de l'élaboration de ce travail.

Je remercie enfin ma famille et mes amis pour m'avoir soutenue tout au long de ma thèse. A mes parents, Je n'aurais pu réussir sans eux, je tiens ici à les remercier.

Je ne saurais oublier de remercier toutes les personnes qui me sont chères, ma grande famille, en particulier mes frères Nacer, Omrani, ma sœur Nadia, son mari Nabil et leur fille Meriem.

Enfin, mon mari Abdelkarim, je ne le remercierais jamais assez pour tout ce qu'il a fait pour moi dans les pires moments de ma vie, j'ai toujours pu compter sur lui, je voulais qu'ils sachent à quel point son soutien a été d'une grande aide pour moi. Alors du fond du cœur... merci.

Je ne peux oublier mes trois enfants Ritadj, Abdelwaddod et Chahida El Rebeh.

Nahed

Table des matières

Glossaire des acronymes	IV
Table des Figures	VI
Liste des Tableaux	XI
Introduction générale	1
Chapitre I Energie Solaire	
I.1 Introduction.....	6
I.2 Le rayonnement solaire.....	6
I.2.1 La constante solaire.....	8
I.2.2 La masse d'air.....	8
I.2.3 Bilan énergétique du système terre / atmosphère.....	8
I.2.4 Nuages.....	10
I.2.5 Caractéristiques physiques.....	10
I.2.5.1 Trajectoire terre – soleil.....	10
I.2.5.2 Facteur de correction de la distance terre-soleil.....	11
I.2.6 La Position du soleil.....	12
I.2.6.1 Coordonnées équatoriales.....	12
I.2.6.2 Coordonnées Horizontales.....	12
I.3 Le rayonnement solaire au sol.....	17
I.3.1 Rayonnement solaire direct normal I_N	17
I.3.2 Rayonnement solaire diffus D	17
I.3.3 Rayonnement réfléchi d_{re}	18
I.3.4 Rayonnement solaire global G_h	18
I.3.4.1 Rayonnement global sur une surface horizontale G_h	18
I.3.4.2 Rayonnement global sur une surface inclinée.....	19
I.4 Rayonnement solaire et production photovoltaïque.....	19
I.4.1 L'effet photoélectrique.....	20
I.4.1.1 Génération de l'énergie électrique à partir du rayonnement solaire....	20
I.4.1.2 Systèmes photovoltaïques.....	21
I.4.1.3 l'énergie solaire thermique.....	22
I.5 Applications des prédictions d'irradiation et d'énergie photovoltaïque.....	23
I.6 Conclusion.....	24
Chapitre II Etat de l'art de la prédiction de l'irradiation solaire	
II.1 Introduction.....	25
II.2 Définition d'une série temporelle.....	25
II.3 Approche stochastique du signal.....	26
II.4 Fondamentaux pour la prévision du GHI.....	27
II.4.1 Modèles ciel clair.....	27
II.4.2 Indice de ciel clair K_c	28
II.4.3 Modèle de référence : la Persistance.....	28
II.5 Méthodes de prédiction des séries temporelles	29
II.6 Classification de la prédiction de rayonnement solaire.....	30
II.6.1 Classification de la prédiction de rayonnement solaire en fonction de l'horizon de prédiction.....	30
II.6.1.1 Prédications à court terme.....	31

II.6.1.2 Prédiction à moyen terme.....	31
II.6.1.3 Prédiction à long terme.....	31
II.6.2 Classification de la prédiction de rayonnement solaire basée sur la base de données historique.....	32
II.6.3 Classification de la prédiction de rayonnement solaire basée sur des méthodes de prédiction.....	33
II.6.3.1 Méthodes Régressives.....	33
II.6.3.2 Méthode de régression.....	34
II.6.3.3 Méthode de lissage exponentiel.....	34
II.6.3.4 Méthode physique.....	35
II.6.3.5 Méthode d'intelligence artificielle.....	35
II.6.3.6 Modèles hybrides.....	39
II.7 Prédiction d'ensemble	41
II.7.1 Prédiction d'ensemble compétitif.....	41
II.7.1.1 Diversité des données.....	42
II.7.1.2 Diversité de paramètres	43
II.7.2 Prédiction d'ensemble coopérative.....	43
II.7.2.1 Prétraitement.....	44
II.7.2.2 Post-traitement	45
II.8 Conclusion.....	45

Chapitre III Approche d'ensembles multi-modèles pour la prédiction de l'irradiation solaire

III.1 Introduction.....	47
III.2 Méthodologie proposée.....	48
III.2.1 Modélisation d'ensemble.....	49
III.2.2 Méthodes individuelles.....	50
III.2.2.1 Modèle autorégressifs.....	50
III.2.2.2 Modèles à base de réseaux de neurones artificiels (RNA).....	51
III.2.2.3 Modèles à noyau.....	55
III.2.3 Méthodes de combinaison.....	57
III. 2.3.1 Moyen simple, SA.....	58
III. 2.3.2 Méthode basée sur l'erreur, EB.....	58
III.2.3.3 Régression par moindres carrés, LSR.....	59
III.2.3.4 Méthode de pondération différentielle, DW.....	59
III.2.3.5 La méthode Outperformance, OP.....	60
III.2.4 Stratégie de classification.....	60
III.3 Sources de données.....	62
III.4 Estimation de l'erreur de prédiction.....	63
III.5 Résultats et discussion.....	64
III.5.1. Résultats obtenus avec des prédicteurs autorégressifs.....	64
III.5.2. Résultats obtenus avec les prédicteurs neuronaux.....	65
III.5.3. Résultats obtenus avec les prédicteurs SVM.....	68
III.5.4. Résultats obtenus avec un ensemble de prédictions hybride.....	68
III.5.5. Résultats obtenus avec la sélection des variables.....	70
III.5.6 Test d'hypothèse.....	75
III.6 Conclusion.....	76

Chapitre IV Estimation des paramètres du modèle autorégressif ARMA par les algorithmes génétiques

IV.1 Introduction.....	77
IV.2 Principe de fonctionnement des algorithmes génétiques.....	77
IV.3. Vocabulaire des algorithmes génétiques.....	78
IV.4 Structure de l’algorithme génétique.....	78
IV.4.1 Le codage	80
IV.4.1.1 Codage binaire.....	81
IV.4.1.2 Codage réel.....	81
IV.4.2 Création de la population.....	82
IV.4.3 Évaluation de la population.....	82
IV.5 Les mécanismes d’un AG.....	82
IV.5.1 La sélection.....	83
IV.5.2 Fonction d'adaptation.....	84
IV.5.3 Le croisement.....	84
IV.5.3.1 Croisement en un point.....	84
IV.5.3.2 Croisement deux points.....	84
IV.5.3.3 Croisement uniforme.....	84
IV.5.4 Mutation.....	85
IV.5.4.1 Mutation à valeurs réelles.....	85
IV.5.4.2 Mutation binaire.....	86
IV.5.5 Insertion	86
IV.5.6 Critère d’arrêt	86
IV.6 Influence des paramètres des AG	86
IV.6.1 La taille de la population N.....	86
IV.6.2 La longueur des individus S.....	87
IV.6.3 La probabilité de croisement p_c	87
IV.6.4 La probabilité de mutation p_m	87
IV.7 Optimisation des modèles ARMA.....	87
IV.7.1 Formulation d’estimation des paramètres de l’ordre d’ARMA à l’aide d’AG.....	88
IV.8 Résultats d'identification.....	89
IV.8.1 Interprétation des résultats.....	80
IV.8.2 Interprétation des courbes.....	95
IV.9 Conclusion.....	96
Conclusion générale	97
Bibliographie.....	99

Glossaire des acronymes

CIMO	Commission des Instruments et des Méthodes d'Observation
WMO	World Meteorological Organization
AU	Astronomical Unit
LT	Local Time
GMT	Greenwich Mean Time
TST	True Solar Time
UT	Universal Time
LST	Local Standard Time
PR	Performance Ratio
DC /AC	Direct Current / Alternative Current
PV	Photovoltaic
ESRA	European Solar Radiation Atlas
K_{cs}	Clear Sky index
AR	Autoregressive
MA	Moving Average
ARMA	Autoregressive Moving Average
ARIMA	Autoregressive Integrated Moving Average
SARIMA	Seasonal Autoregressive Integrated Moving Average
ARX	Autoregressive with exogenous variables
ARMAX	Autoregressive Moving Average with exogenous variables
VAR	Vector AR
ANN	Artificial Neural Networks
MLP	Multi Layer Perceptron
ELM	Extreme Learning Machines
RBF	Radial Basis Functions
ENN	Elman Neural Networks
kNN	k-Nearest Neighbors
SVM	Support Vector Machine
SVR	Support Vector Regression
RF	Random Forests
EWMA	Exponentially Weighted Moving Average

MLFFNN	Multilayer Feedforward Neural Network
GRNN	Generalized Regression Neural Network
ANFIS	Adaptive-Network-based Fuzzy Inference System
TDNN	Time Delay Neural Network
AdaBoost	Adaptive Boosting
TSI	Total Satellite Image
DWT	Discrete Wavelet Transform
MSE	Mean Squared Error
MAE	Mean Absolute Error
MBE	Mean Bias Error
RMSE	Root Mean Squared Error
NMSE	Normalized Mean Squared Error
SA	Simple Average
EB	Error-Based
LSR	Least-Squares Regression,
DW	Differential Weighting
OP	Outperformance
K – S	Kolmogorov – Smirnov
W-L-T	Win-Lose-Tie
GHI	Global Horizontal Irradiation
CS	Clear Sky

Liste des figures

Chapitre I

I.1	Spectre du rayonnement solaire reçu hors atmosphère	7
I.2	Bilan énergétique annuel du système Terre-Atmosphère.....	9
I.3	Classification des types de nuages.....	10
I.4	Le mouvement de la terre autour du soleil.....	11
I.5	Référentiel du plan équatorial.....	13
I.6	Coordonnées terrestres.....	13
I.7	Equation du temps en fonction du jour de l'année et de la déclinaison.....	14
I.8	Schéma représentant la variation annuelle de la déclinaison du soleil.....	15
I.9	Angle horaire ω du soleil.....	16
I.10	Système de coordonnées horizontales.....	17
I.11	Composantes principales du rayonnement solaire.....	19
I.12	L'effet photovoltaïque.....	20
I.13	Principe du raccordement d'un module <i>PV</i> au réseau électrique.....	22

Chapitre II

II.1	Représentation du rayonnement global horizontal (X_t pour les sites Ny-Alesund, Brasilia, Golden et Alicesprings.....	27
II.2	Présentation de l'indice K_c sur une journée de mesure.....	28
II.3	Modèle de persistance appliqué directement au GHI pour une journée ciel clair.....	29
II.4	Classification de la prédiction de GHI en fonction de l'horizon temporel.....	31
II.5	Classification de la prédiction de rayonnement solaire sur la base des données	

historiques	32
II.6 (a) Diagramme schématique d'une structure ANN composée d'une couche d'entrée, d'une couche cachée et d'une couche de sortie.(b) A modèle mathématique d'un ANN.....	37
II.7 Répartition des études par rapport à la technique utilisée.....	40
II.8 Schéma fonctionnel des méthodes de prédiction d'ensemble concurrentielles.....	42
Chapitre III	
III.1 Schéma fonctionnel du système de prédiction proposé.....	48
III.2 Architecture d'un réseau de neurones de type perceptron multicouche.....	53
III.3 Architecture d'un réseau de neurones à fonction radiale.....	53
III.4 Architecture du réseau Elman.....	54
III.5 Régression SVM linéaire avec ε -tube.....	57
III.6 Partition schématique d'un espace variable bidimensionnel pour un exemple composé de six modèles. Chaque région désigne l'algorithme de prévision individuel qui fournit la précision de prédiction optimale dans la zone correspondante de l'espace variable de la série temporelle.....	61
III.7 Schéma fonctionnel du processus d'apprentissage utilisant l'approche basée sur la classification.....	62
III.8 Série chronologique de données GHI sur deux années successives, des stations (Alice Springs, Brasilia, Ny-Ålesund, and Golden.....	63
III.9 Une petite fenêtre sur les séries temporelles mesurées et prédites par différents modèles autorégressifs sur le site d'Alice Springs.....	67
III.10 Une petite fenêtre sur les séries temporelles mesurées et prédites par différents modèles ANN sur le site de test d'Alice Springs	67
III.11 Une petite fenêtre sur les séries temporelles mesurées et prédites par différents modèles de SVM pour le site de test d'Alice Springs	71
III.12 Comparaison entre les modèles les plus performants pour chaque famille de prédiction sur le site de test d'Alice Springs.....	71
III.13 Diagramme de Taylor des valeurs GHI mesurées et prédites par différents modèles de combinaison pour le site de d'Alice Springs	74
III.14 Diagramme de Taylor des valeurs GHI mesurées et prédites par différents modèles de combinaison pour le site de Brasilia	74

III. 15	Diagramme de Taylor des valeurs GHI mesurées et prédites par différents modèles de combinaison pour le site de Ny-Ålesund	75
III. 16	Diagramme de Taylor des valeurs GHI mesurées et prédites par différents modèles de combinaison pour le site de Golden.....	75
Chapitre IV		
IV.1	Illustration d'un cycle dans un algorithme génétique.....	79
IV.2	Organigramme général de l'AG.....	80
IV.3	Structure d'un chromosome.....	81
IV.4	Schéma d'un exemple d'une roue de loterie.....	83
IV.5	Exemples d'opérations de croisement.....	85
IV.6	Mutation binaire.....	86
IV.7	Architecture de l'AG pour l'estimation du modèle ARMA.....	89
IV.8	Les étapes de construction du modèle ARMA.....	89
IV.9	Série chronologique de données GHI sur deux années successives, pour les sites Brasilia et Ny-Ålesund.....	91
IV.10	Critère AIC pour les sites : a) Brasilia,b)Ny-Ålesund.....	92
IV.11	ACF et PACF des données solaire du site Ny-Ålesund.....	93
IV.12	ACF et PACF des données solaire du site Brasilia	94
IV.13	Comparaison des différents modèles sur les deux stations de test.....	95

Liste des tableaux

Chapitre III

III.1	Informations sur les quatre stations de test utilisées dans cette étude. Les valeurs GHI moyennes et maximales mesurées sont données en W/m^2	62
III.2	Résultats statistiques obtenus à toutes les stations de test par différents modèles autorégressifs et leurs méthodes de combinaison. Dans chaque groupe de colonnes, les meilleurs résultats apparaissent en gras.....	65
III.3	Résultats des prédictions pour les quatre stations de test obtenues avec quatre modèles ANN et leurs méthodes de combinaison. Dans chaque groupe de colonnes, les meilleurs résultats apparaissent en gras.....	66
III.4	Résultats de prédiction obtenus avec trois modèles SVM et leurs méthodes de combinaison aux quatre stations de test. Dans chaque colonne, les meilleurs résultats apparaissent en gras.....	69
III.5	Résultats de prédictions obtenues avec le meilleur modèle de chaque famille de prédiction (SVM-RBF, ENN et ARIMA) et leurs méthodes de combinaison aux quatre stations de test. Dans chaque groupe de colonnes, les meilleurs résultats apparaissent en gras. Les modèles individuels incluent le modèle ciel claire (CS) [50] et Persistence (Persist.).....	60
III.6	Résultats des prédictions obtenus à l'aide du système à deux blocs (technique de sélection des caractéristiques et de régression). Dans chaque groupe de colonnes, les meilleurs résultats apparaissent en gras. SV: variables historiques sélectionnées.....	73
III.7	Résultats du test d'hypothèse comparant la stratégie de classification à d'autres stratégies de combinaison pour le site d'Alice Springs.....	76

Chapitre IV

IV.1	Vocabulaire des AG.....	78
IV.2	Résultats de la prédiction pour les deux sites de test obtenus avec les modèles ARMA, ARMA-GA et leur combinaison avec la méthode OP. Les meilleurs résultats sont en gras.....	90

Introduction générale

1. Contexte et objectifs

Les énergies renouvelables jouent un rôle crucial dans la société moderne. Les sources d'énergie tirent leur énergie des flux d'énergie existants, des processus naturels en développement, tels que le soleil, le vent, les eaux vives, les processus biologiques et les flux de chaleur géothermiques. Une définition commune des sources d'énergie renouvelables est que l'énergie renouvelable est capturée à partir d'une ressource énergétique qui est rapidement remplacée par un processus naturel tel que l'énergie générée par le vent ou le soleil [1]. À l'heure actuelle, les sources d'énergie les plus prometteuses sont l'énergie éolienne, l'énergie solaire et l'énergie géothermique [2, 3]. Les autres sources renouvelables comprennent les énergies hydroélectriques et océaniques, ainsi que la biomasse et l'éthanol en tant que carburants renouvelables. Malgré la diversité des sources d'énergie, la façon dont nous utilisons l'énergie (le produit final) répond en général à l'un des trois besoins à savoir :

- Production d'électricité.
- Génération de chaleur.
- Énergie pour le transport.

Les gouvernements du monde entier se tournent de plus en plus vers les énergies renouvelables pour mettre fin à leur dépendance aux combustibles fossiles. Dans ce contexte, l'énergie solaire photovoltaïque est l'une des technologies énergétiques alternatives la plus prometteuse du futur. L'histoire de l'énergie photovoltaïque a commencé en 1839 lorsqu'Edmond Becquerel, un physicien français, observe la production d'électricité par certains matériaux lorsqu'ils sont soumis à la lumière du soleil. Einstein étudie plus tard ce phénomène et publie en 1905 un article au sujet de l'effet photo-électrique pour lequel il reçoit le prix Nobel en 1921[4].

Cependant, l'énergie solaire est jugée trop coûteuse malgré les travaux de Robert Millikan et sa "cellule solaire" (1916). Elle est mise de côté pendant presque 40 ans. En 1960, la course à l'espace fait ressurgir la technologie solaire sur le devant de la scène : il s'agit d'une solution particulièrement adaptée à l'alimentation des premiers satellites. La recherche dans le domaine de l'énergie photovoltaïque bénéficie alors de fonds de la course à l'espace,

ce qui l'amène à devenir en 1970, avec la crise énergétique, une solution alternative prometteuse.

Dans le contexte énergétique actuel, l'énergie photovoltaïque est une solution viable et promue par de nombreux pays en reconversion énergétique.

La production d'énergie électrique par un système photovoltaïque est dépendante de nombreux facteurs. L'inclinaison des panneaux, leur technologie et leur température influent sur la production mais le rayonnement solaire reçu reste le facteur principal de variation. Dans le cadre de cette thèse, nous avons choisi de prédire la quantité du rayonnement solaire vu son importance dans beaucoup de problèmes liés à l'installation et l'utilisation de l'énergie solaire.

La prédiction du rayonnement solaire est un domaine de recherche très actif à l'heure actuelle. Nous distinguons quatre grandes familles de méthodes de prédiction selon les données utilisées et l'horizon de prédiction souhaité :

- *Méthodes statistiques* : elles s'appuient sur les séries temporelles d'un ou de plusieurs capteurs afin de construire des modèles statistiques de prévision. Ces méthodes supposent l'existence de séquences répétitives dans la série temporelle que l'on peut apprendre afin de les prévoir par la suite. On retrouve dans cette famille de nombreuses méthodes de régression (linéaires et non linéaires) existantes en traitement du signal (modèles analytiques, modèles autorégressifs, ...). L'horizon de prédiction dans ce type de méthodes est très dépendant de l'échantillonnage choisi pour la série temporelle.

- *Modèle de prédiction météorologique numérique* : les modèles numériques utilisent des algorithmes de résolution numérique des équations météorologiques assimilant les données météorologiques acquises par des stations de mesures réparties sur le globe. Elles fournissent ainsi des prévisions de nombreuses variables qualifiant l'état de l'atmosphère (température, pression, vitesse du vent,...etc.) pour chaque point de la grille spatiale de prédiction. Les prédictions nécessitent un temps de calcul important. La résolution temporelle des prévisions est de l'ordre de 3 à 6 heures. Ces modèles permettent une prévision de l'irradiation à long terme (plusieurs jours) via différents traitements statistiques des sorties obtenues.

- *Imagerie satellitaire* : les satellites météorologiques géostationnaires fournissent des images de l'atmosphère et de la terre sur l'ensemble du globe avec une résolution temporelle intra-horaire (15 minutes pour METEOSAT seconde génération). Ces

caractéristiques font de l'imagerie satellitaire un outil de premier ordre pour la description et le suivi de phénomènes météorologiques de moyenne et grande échelle. Ainsi, de nombreuses méthodes de prévision du l'irradiation profitent de ces images depuis 1980. L'horizon de prévision atteint est généralement de l'ordre de quelques heures.

- *Imagerie hémisphérique* : les images hémisphériques sont des images du ciel prises depuis le sol via une caméra possédant un angle de vision de 180°. Les travaux impliquant ces images ont émergé il y a une vingtaine d'années avec l'apparition de capteurs de plus en plus performants. Ces images permettent une description très fine de la couverture nuageuse avec une résolution temporelle de l'ordre de la seconde. Elles sont particulièrement adaptées à l'estimation et à la prédiction à très court-terme (jusqu'à 30 minutes).

2. Pourquoi prédire le rayonnement solaire ?

L'installation de tout système d'énergie solaire, dans un site donné, nécessite des études préalables. En effet, le dimensionnement et la simulation sont primordiaux pour assurer un fonctionnement optimal. Pour faire ce genre d'étude, des mesures fiables, sur des durées relativement longues, de certaines variables météorologiques, et spécialement, celles du rayonnement solaire sont indispensables.

Le manque de longues séries de données ou de séries de données de mauvaise qualité (discontinuités et non fiabilités) peut engendrer des erreurs lors de la conception de ces systèmes, ce qui a un impact négatif sur l'investissement. Malheureusement, les mesures du rayonnement solaire sont généralement inexactes et rares dans le monde entier [5]; en raison du prix élevé des appareils de mesure. Il n'y a qu'un nombre limité de stations de mesure de rayonnement solaire, c'est pourquoi il y a un manque de mesures d'irradiation solaire sur de grandes surfaces, ceci d'une part. D'autre part, quand ces données existent, il y a généralement des périodes de rupture dues à des pannes ou au mauvais suivi, puisque la majorité de ces stations appartiennent à des établissements non bénéficiaires économiquement de ces données.

Cependant, d'autres paramètres météorologiques tels que la température ambiante, l'humidité et la vitesse du vent sont relativement facile à mesurer dans un plus grand nombre de stations météorologiques avec un coût relativement faible par rapport à celui du rayonnement. D'autre part, l'optimisation d'un système solaire ou la simulation de ses performances nécessitent au moins, des données journalières, voire même, horaire de l'irradiation solaire. Par conséquent, il est souhaitable d'élaborer des relations entre les

données météorologiques disponibles et l'irradiation solaire et de développer des techniques précises (modèles) pour prédire cette dernière. Ces modèles sont des outils qui permettent la génération de longues séries de données de l'irradiation solaire à différents pas de temps. Ceci pourrait fournir des informations utiles aux décideurs en matière de sélection, de conception et de planification de nouvelles centrales solaires.

3. Objectifs de cette thèse

Dans la présente thèse, nous avons ciblé deux objectifs :

Le premier vise à trouver la meilleure performance de prédiction parmi un ensemble d'algorithmes de prédiction différents, en utilisant une nouvelle approche inspirée de la théorie d'apprentissage d'ensemble [6]. L'évaluation expérimentale de l'approche proposée sera réalisée à partir de données réelles.

La deuxième partie concerne l'utilisation des algorithmes génétiques pour l'optimisation des paramètres d'un modèle ARMA, ceci dans le but de trouver l'ordre optimal du modèle ARMA et d'améliorer l'estimation de ses paramètres, ensuite l'appliquer pour l'estimation de l'irradiation solaire globale (GHI) sur deux bases de données réelles.

4. Contributions de la thèse

Les principales contributions de la thèse sont:

- En ce qui concerne le choix du modèle de prédiction : une approche de fusion/combinaison basée sur la combinaison ou la classification des résultats de prédiction des différents algorithmes individuels est proposée, dans la quelles plusieurs familles de prédicteurs (autorégressives, neuronales et à noyaux) ont été testées et utilisées.
- Une méthode de prédiction incorporant un algorithme évolutionnaire, à savoir les algorithmes génétiques (AG) et la méthode de prédiction à moyenne mobile autorégressive (ARMA) a été proposée dans l'intérêt de tirer profit de leurs forces intrinsèques.

5. Organisation du manuscrit

Cette thèse présente l'ensemble des travaux de recherche et les différents résultats obtenus. Elle est répartie sur quatre chapitres.

Nous décrirons dans le premier chapitre le phénomène du rayonnement solaire global incident sur la surface de la Terre, en exposant les faits théoriques qui permettent de

comprendre les différents éléments qui le compose. Les différents phénomènes physiques qui sont sujets à ce rayonnement, la structure de l'atmosphère et son effet sur le rayonnement solaire. La connaissance de ces divers paramètres permet d'établir un bilan radiatif du système terre-atmosphère.

Dans le second chapitre, nous présentons l'état de l'art des méthodes de prévision, notamment des prédicteurs de natures différentes comme les réseaux de neurones, les méthodes auto-régressives, les machines à vecteurs de support (SVM), ...etc. De plus, ce chapitre présente une classification des modèles de prévision suivant les caractéristiques des données d'entrées et les horizons de prédiction.

Nous présenterons dans le chapitre trois, la méthodologie de prédiction multi-modèle adoptée dans la première partie.

Dans le quatrième chapitre, nous proposons une méthode de prédiction intégrant les algorithmes génétiques (AG) et la méthode de prédiction à moyenne mobile autorégressive (ARMA) afin de tirer profit de la force intrinsèque des deux modèles.

Nous terminons cette thèse par une conclusion générale, où nous résumons notre travail en mettant l'accent sur son originalité ainsi que sur les différentes perspectives ouvertes suite aux diverses études réalisées dans le cadre de cette thèse.

Chapitre I

Energie Solaire

I.1 Introduction

L'énergie solaire est la ressource énergétique la plus abondante sur terre. Elle est à l'origine de la majorité des autres énergies renouvelables. Le rayonnement solaire peut être utilisé pour produire soit directement de l'électricité à l'aide de semi-conducteurs photovoltaïque, soit de la chaleur solaire thermique pour le chauffage ou la production électrique. Cependant, ce rayonnement est sujet à plusieurs perturbations, qui varient avec la constitution verticale de l'atmosphère terrestre (nuage, aérosol, etc.), qui constitue une barrière empêchant ainsi l'irradiation solaire de passer.

Il serait tout d'abord intéressant de connaître quelques notions sur le rayonnement solaire (caractéristique, nature, intensité), avec ses différentes composantes qui le sont liées, la structure de l'atmosphère et ces effets sur le rayonnement solaire. Ainsi, la connaissance de ces divers paramètres permet d'établir un bilan radiatif du système terre- soleil et avec quelques avantages et inconvénients de cette énergie, ceci est présenté en détail dans ce chapitre.

I.2 Le rayonnement solaire

Le soleil est avant tout l'élément essentiel rendant la vie organique possible sur terre (chaleur, photosynthèse, et création des énergies fossiles). Le soleil est une sphère gazeuse chaude composée principalement d'Hydrogène (73.5%) et d'Hélium (24.9%). Son diamètre est de 1.39 millions de kilomètres, ce qui correspond à une surface totale égale à $6.08 \times 10^{18} \text{ m}^2$ et se trouve à une distance moyenne de 150 millions de kilomètres de la terre. Cette distance varie entre 152 millions de kilomètres (le premier juillet) et 147 millions de kilomètres (le premier janvier). Vu depuis la terre, le soleil présente un angle solide de $6.8 \times 10^{-5} \text{ rad}$ et une ouverture angulaire de 0.53° . Il émet du rayonnement dans des longueurs d'onde qui varient de 250 à 5000 nm. La terre reçoit une partie de ce rayonnement. Le flux énergétique reçu par la terre est appelé éclairement énergétique, abrégé en éclairement. La moyenne

annuelle de cet éclairement reçu sur un plan normal aux rayons du soleil situé au sommet de l'atmosphère est appelée constante solaire.

Le soleil peut être assimilé à un corps noir qui émet un rayonnement électromagnétique dont le spectre est présenté dans la figure I.1. Sur ce spectre, on note la différence entre le rayonnement arrivant au sommet de l'atmosphère et le rayonnement arrivant au niveau du sol. Cette différence est liée aux bandes d'absorption des gaz de l'atmosphère. La distribution de l'énergie émise par le soleil peut être divisée en : lumière visible 39.8%, ultraviolet (UV) 9% et infrarouge (IR) 51.2%.

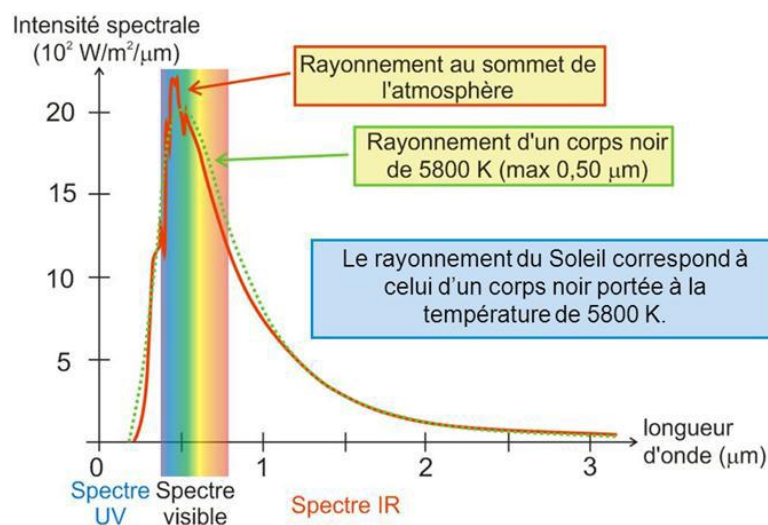


FIGURE I.1 Spectre du rayonnement solaire reçu hors atmosphère [7].

Le rayonnement solaire atteint le sol sous différentes formes à savoir, le rayonnement dispersé atteignant la surface de la terre désigné sous le nom du rayonnement diffus, le rayonnement direct provenant directement du disque solaire sans subir des modifications et également le rayonnement réfléti par la terre à différentes longueurs d'onde dues à la température de sa surface. Le rayonnement réfléchi par la surface terrestre est renvoyé dans une direction privilégiée (réflexion dite spéculaire) ou de manière diffuse. Le sol réfléchit plutôt le rayonnement de manière diffuse et anisotrope. La quantité d'énergie transportée par le rayonnement solaire à la surface terrestre dépend de certaines considérations géométriques dues au mouvement apparent du soleil et aux mouvements de la terre.

I.2.1 Constante solaire

L'énergie solaire reçue par unité de temps à la limite supérieure de l'atmosphère, sur une surface unité perpendiculaire aux rayons solaires et pour une distance Terre-Soleil égale à sa valeur moyenne, est appelée constante solaire. La constante solaire C a été estimée à 1367 W/m^2 par Claus Fröhlich et Christoph Wehrli du centre radiométrique mondial de Davos (Suisse), avec une incertitude de l'ordre de 1%. Cette valeur a été adoptée par la commission des instruments et des méthodes d'observation (CIMO) de l'Organisation météorologique mondiale (OMM) en octobre 1981. En première approximation, on peut calculer sa valeur en fonction du numéro du jour de l'année n par [8]:

$$G_0 = 1367(1 + 0.033 \cos(0.9856 n)) \text{ W/m}^2 \quad (\text{I.1})$$

Où n représente le numéro du jour de l'année compté à partir du 1^{er} janvier.

I.2.2 Masse d'air

Le rayonnement solaire traversant l'atmosphère au cours de la journée dépend de la position relative du soleil dans le ciel, pour tenir compte de cette position qui modifie l'épaisseur d'atmosphère traversée par les rayons lumineux, on définit un coefficient m appelé masse atmosphérique ou nombre masse d'air, défini par [9]:

$$m = \frac{P}{1013 \sin A} \exp\left(-\frac{Z}{7.8}\right) \quad (\text{I.2})$$

7.8 : L'épaisseur verticale moyenne standard de l'atmosphère exprimée en km ;

Z : L'épaisseur de l'atmosphère traversée par le rayonnement solaire exprimée en km ;

P : Pression atmosphérique N/ m^2 .

A : Élévation du soleil en degré.

I.2.3 Bilan énergétique du système terre / atmosphère

La moyenne annuelle de rayonnement arrivant au sommet de l'atmosphère est la constante solaire S (1367 W/m^2) [10]. Il s'agit de la quantité de rayonnement électromagnétique qui serait reçue sur un plan perpendiculaire aux rayons, à une distance de '1' unité astronomique (UA). En considérant la rotation de la Terre et l'angle avec lequel les rayons atteignent l'atmosphère selon l'endroit considéré, le rayonnement solaire moyen reçu est estimé à un quart de la constante solaire (340 W/m^2). C'est la valeur de référence

pour la figure I.2 qui présente un bilan énergétique global du rayonnement solaire sur Terre. En moyenne, seul 48% du rayonnement incident parvient au sol, l'atmosphère et les nuages retenant l'autre moitié soit par réflexion ou par absorption. L'étude de la couverture nuageuse et des paramètres de l'atmosphère joue donc un rôle capital pour l'étude du rayonnement global reçu au sol, ce qui constitue l'enjeu principal de plusieurs études.

L'éclairement solaire incident au sommet de l'atmosphère est estimé à environ 342 W/m² en moyenne annuelle. Une fraction du rayonnement qui arrive au sommet de l'atmosphère (environ 30%) est réfléchi et renvoyée vers l'espace par les différentes composantes de l'atmosphère (6 % par l'air, 20 % par les nuages) et par les surfaces terrestres (4%). L'albédo global terrestre exprime cette fraction. Lors de sa traversée de l'atmosphère, environ 19% du rayonnement est absorbé par un mélange de gaz et de particules en suspension (gaz, poussières et nuages). Seulement 53% de l'éclairement solaire incident parvient à la surface. L'énergie disponible en surface dépend en plus des fractions réfléchies et absorbées par l'atmosphère terrestre elle-même. L'atténuation du rayonnement à travers l'atmosphère est en fonction de la densité, de la taille et de la nature des molécules et des particules rencontrées. Elle varie aussi avec la longueur de la trajectoire des rayons solaires à travers l'atmosphère. Cette longueur est caractérisée par la 'masse atmosphérique', appelée aussi masse d'air optique.

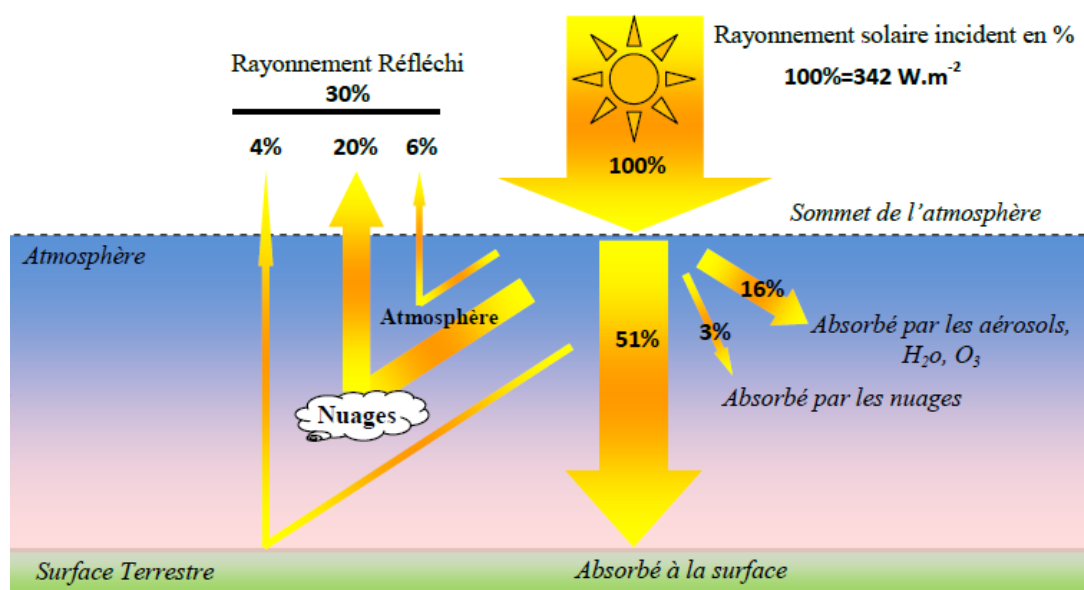


FIGURE I.2 Bilan énergétique annuel du système Terre-Atmosphère.

I.2.4 Nuages

Les nuages sont constitués de petites gouttes d'eau et de cristaux de glace en suspension dans l'air. La plupart des nuages sont situés dans la troposphère. Ils se forment par condensation habituellement au niveau des courants ascendants. Selon leur altitude et leur composition, ils présentent diverses formes et propriétés. L'Organisation Météorologique Mondiale (WMO) propose une classification des nuages en 10 principales classes (figure I.3)

L'impact des nuages sur le rayonnement perçu au sol est difficile à décrire de façon quantitative. En effet, les nuages d'un même type peuvent avoir un effet d'atténuation différent en fonction de leurs propriétés macroscopiques et microscopiques (géométrie, épaisseur optique, quantité d'eau liquide, etc.). Cependant, les nuages ont généralement un effet réducteur sur le rayonnement reçu au sol puisqu'ils réfléchissent une partie de la lumière incidente vers l'espace.

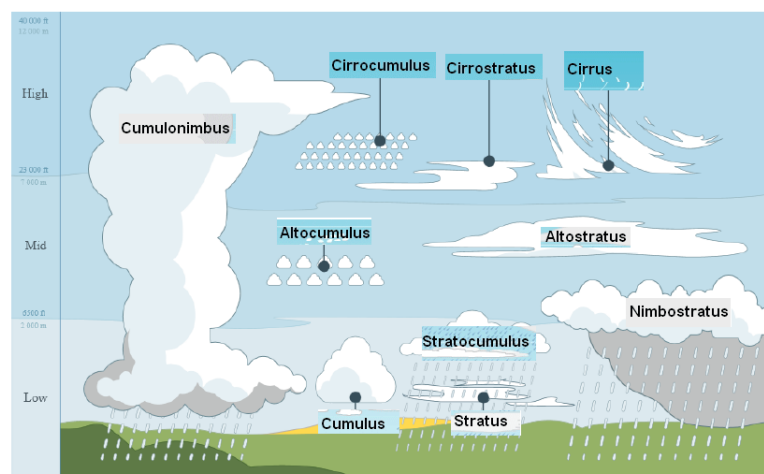


FIGURE I.3 Classification des types de nuages [11].

I.2.5 Caractéristiques physiques

I.2.5.1 Trajectoire terre – soleil

La terre crée autour du soleil une orbite elliptique quasi-circulaire avec une période de 365,25 jours. Le plan de cette orbite est appelé plan de l'écliptique comme indiqué sur la figure I.4. C'est au solstice d'hiver (21 décembre) que la terre est très près du soleil avec 147 millions de km de distance et au solstice d'été qu'elle en est la plus éloignée : avec 153 millions de km, et la distance moyenne étant de 149,5 millions de km, la terre coupe le petit axe de son orbite aux équinoxes de printemps (21 Mars) et d'automne (21 septembre). Elle tourne sur elle-même avec une période de 24 heures. Son axe de rotation (l'axe des pôles) à

une orientation fixe dans l'espace, il fait un angle avec la normale du plan de l'écliptique. L'axe de rotation de la Terre sur elle-même est incliné par rapport au plan de l'écliptique céleste. On appelle déclinaison δ , l'angle formé par l'axe Terre -Soleil avec le plan équatorial. La déclinaison varie de $+23^{\circ}27'$ solstice d'été (21 juin), à $-23^{\circ}27'$ Solstice d'hiver (21 décembre), elle s'annule deux fois par an les 21 mars et 23 septembre (ce sont les équinoxes) et elle est responsable des saisons. Les parties diurnes et nocturnes de la journée ont alors la même durée aux équinoxes, la déclinaison passe par 0° ($-23^{\circ}27' < \delta < +23^{\circ}27'$) [10,12].

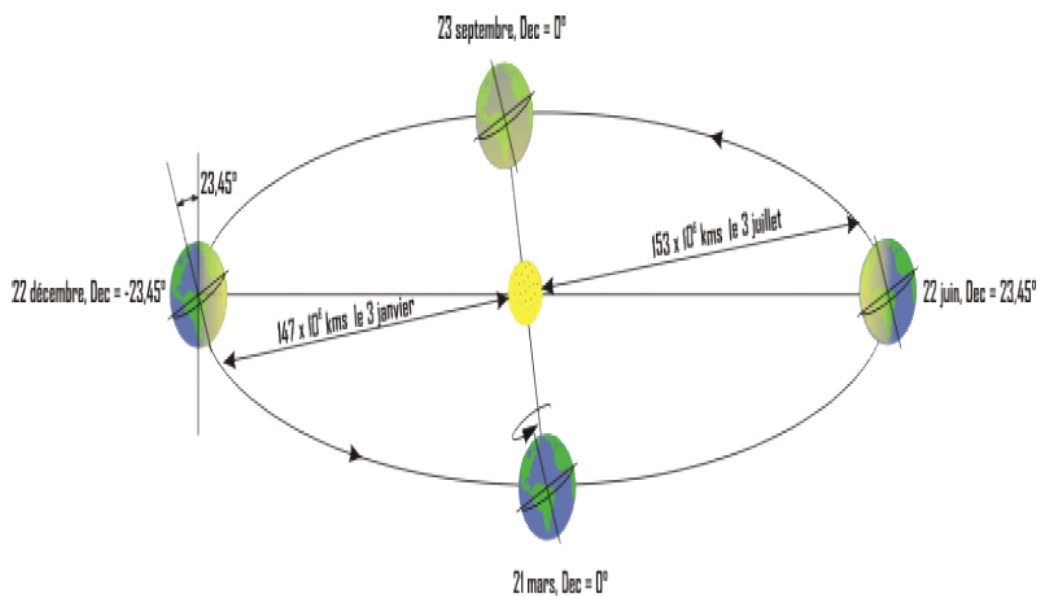


FIGURE I.4 Le mouvement de la terre autour du soleil [8].

I.2.5.2 Facteur de correction de la distance terre-soleil

La distance entre la terre et le soleil varie sensiblement en raison de la trajectoire elliptique de la terre autour du soleil. Cette distance peut alors atteindre 153 millions de Km au maximum le 3 juillet, et 147 millions de Km au minimum le 3 janvier. Ce facteur dépend du numéro du jour de l'année j , la correction de la distance terre-soleil (c_{t-s}) peut être calculée à l'aide de l'équation suivante [13] :

$$C_{t-s} = 1 + 0.033 \cos \left[\frac{360n}{365} \right] \quad (I.3)$$

avec

n : Le numéro du jour de l'année à partir du 1^{er} Janvier.

I.2.6. La Position du soleil

Pour un lieu donné, la position du soleil est repérée à chaque instant de la journée et de l'année par deux référentiels différents, grâce à deux plans, plan horizontal du lieu et plan équatorial de la terre. Pour repérer un site donné à la surface terrestre, on définit les grandeurs suivantes :

I.2.6.1 Coordonnées équatoriales

Dans ce référentiel, le soleil est localisé au moyen de deux angles, la déclinaison solaire δ donnée par l'équation (4) et l'angle horaire Ω défini par [14]:

$$\Omega = 14(TSV - 12)$$

$$TSV = TL - GMT + \left(\frac{Et+4\lambda}{60}\right) \quad (I.4)$$

$$Et = 9.87 \sin 2N - 7.35 \cos N - 1.5 \sin N \quad (I.5)$$

$$N = \frac{360}{365}(n - 81) \quad (I.6)$$

GMT : décalage horaire par rapport au méridien de Greenwich ;

TSV : temps solaire vrai ;

TL : temps légal (donné par une montre) ;

Et : l'équation de correction du temps ;

λ : Longitude du lieu ;

L'angle horaire Ω varie de -180° à $+180^\circ$, prend la valeur nulle à midi *TSV*, négative durant la période avant midi *TSV* et positive après.

I.2.6.2 Coordonnées horizontales

- **Latitude et longitude**

La latitude φ permet de repérer la distance angulaire d'un point quelconque par rapport à l'équateur. Elle varie de 0° à l'équateur à 90° au pôle nord. L'angle de latitude est l'angle entre une ligne tracée d'un point sur la surface de la terre avec le centre de cette dernière, et le cercle équatorial de la terre. L'intersection du cercle équatorial avec la surface de la terre forme l'équateur et est indiquée en tant que latitude de 0° , le pôle nord par la latitude $+90^\circ$ et le pôle sud par la latitude -90° .

- **Equation du temps**

Tous les points à longitude constante font l'expérience de midi et de toute autre heure à la même heure. L'heure locale (LT), également appelée heure solaire, est une mesure de la position du Soleil par rapport à une localité. À midi heure locale, le soleil traverse sa position la plus haute dans le ciel. La figure I.7 représente graphiquement l'équation du temps en fonction du jour Julien et de sa déclinaison. Le temps universel (TU) peut être défini comme l'heure locale au méridien zéro. Pour éviter toute confusion due à des heures locales infinies, des fuseaux horaires ont été introduits sous le concept d'heure standard. L'Heure Standard (SDT) a été proposée par Sandford Fleming en 1879[15]; cela consistait à diviser le monde en 24 fuseaux horaires, chacun couvrant exactement 15° , car la Terre tourne de 15° par heure. A cause des considérations politiques, ils ont maintenant augmenté le nombre de fuseaux horaires standard à 39. L'heure locale standard (LST) est la même heure dans tout le fuseau horaire. De plus, l'horloge est généralement décalée d'une heure entre avril et octobre pour mieux utiliser la lumière du soleil, prétendument pour économiser de l'énergie. La figure I.7 illustre l'équation du temps en fonction du jour de l'année et de la déclinaison δ .

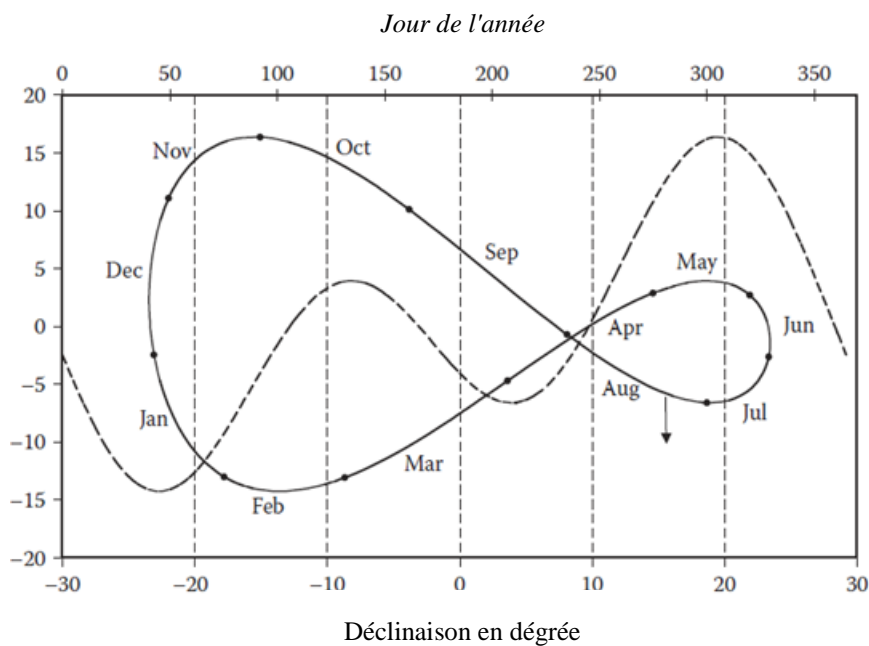


FIGURE 1.7 Equation du temps en fonction du jour de l'année et de la déclinaison.

La relation entre l'heure solaire et l'heure standard doit être connue pour décrire la position du soleil. Pour la plupart des endroits où les zones standards avancent d'une heure à l'autre, l'ajustement de l'heure solaire pour la longitude peut être effectué en soustrayant la longitude de l'observateur (λ_{local}) à la longitude du méridien standard (λ_{STD}) du fuseau

horaire de l'observateur et en la multipliant par *4minutes*. Le soleil prend pour se déplacer 1° à travers le ciel. L'équation 1.7 estime la différence de temps en minutes entre l'heure solaire et l'heure standard plus une correction due à l'irrégularité de la durée naturelle d'un jour. Cette irrégularité est provoquée par l'orbite non circulaire de la Terre qui tourne autour du Soleil et par l'inclinaison de l'axe nord-sud par rapport au Soleil:

$$LT - STD = 4(\lambda_{STD} - \lambda_{local}) + E_t \quad (I.7)$$

où E_t est appelée l'équation du temps en fonction de l'angle quotidien Γ donné par l'équation I.8.

$$E_t = (0.000075 + 0.001868 \cos \Gamma - 0.032077 \sin \Gamma - 0.014615 \cos 2\Gamma - 0.04089 \sin 2\Gamma - 0.04089 \sin 2\Gamma)(229.18) \quad (I.8)$$

- **La déclinaison et l'angle horaire**

La déclinaison (δ) est l'angle que forme la direction du soleil et le plan équatorial. La déclinaison varie de façon sinusoïdale au cours de l'année. Plusieurs expressions ont été développées pour évaluer la déclinaison, la plus simple est celle utilisée par Gueymard[16].

$$\delta = 23.45 \sin\left[\frac{360}{365}(n + 284)\right] \quad (I.9)$$

où : n est le numéro du jour de l'année.

La figure I.8 montre le changement de la déclinaison solaire en fonction du jour de l'année.

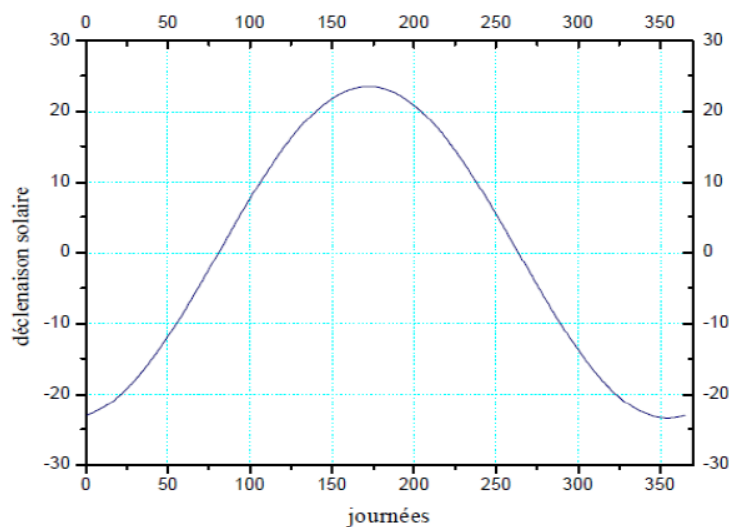


FIGURE I.8 Schéma représentant la variation annuelle de la déclinaison du soleil.

L'angle horaire ω mesure le mouvement du soleil par rapport à midi qui est l'instant où le soleil passe au plan méridien du lieu zénith. Cet angle est formé entre la projection du soleil sur le plan équatorial à un moment donné et la projection du Soleil sur ce même plan au midi vrai. L'angle horaire est donné par la relation suivante [17,18]:

$$\omega = 15(T_{SV} - 12) \quad (I.10)$$

Où T_{SV} représente le temps solaire vrai en heure basé sur la rotation de la terre autour de son axe polaire et de sa révolution autour du soleil. La durée du jour solaire varie durant l'année du fait que : la terre balaye des surfaces différentes dans sa rotation autour du soleil et l'axe de la terre est incliné par rapport au plan écliptique.

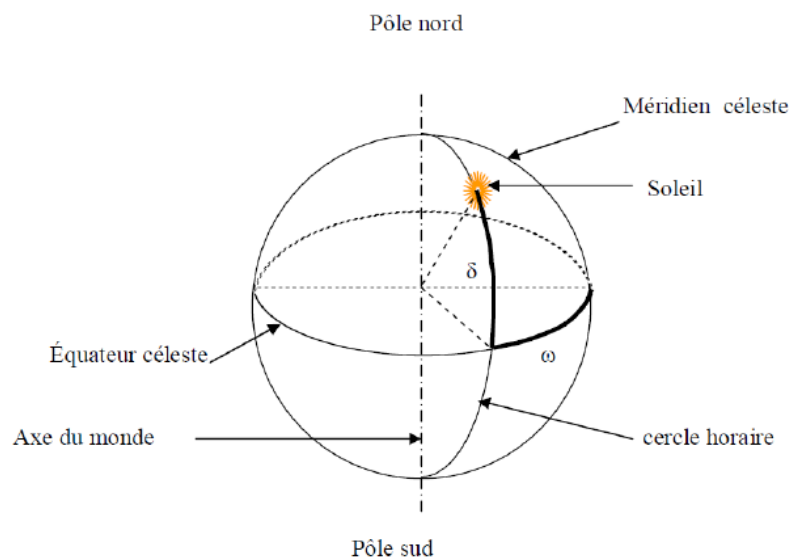


FIGURE I.9. Angle horaire ω du soleil.

- **Azimet A_z**

C'est l'angle horizontal de la direction du soleil avec la direction du sud. La connaissance de l'azimet ' A_z ' permet de calculer l'angle d'incidence des rayons sur une surface non horizontale.

- **Hauteur angulaire**

C'est l'angle vertical de la direction du soleil avec le plan horizontal, parfois on parle de l'angle zénithal qui est le complément de h , tel que $A_z = 90^\circ - h$. La hauteur angulaire du soleil h mesure la distance angulaire du soleil à partir de l'horizon.

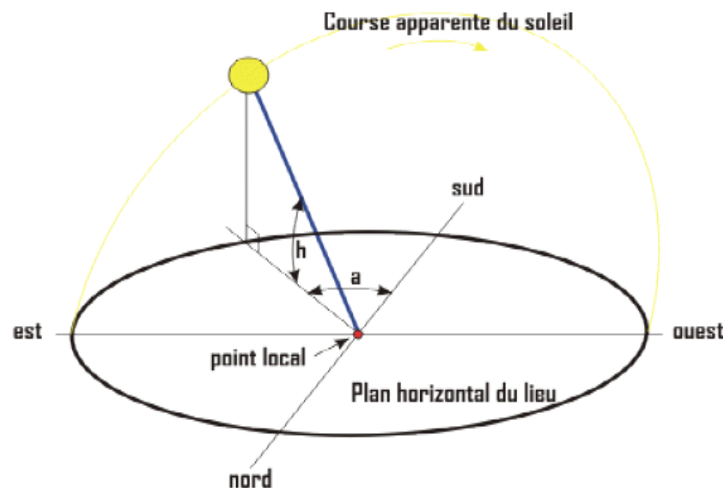


FIGURE I.10 Système de coordonnées horizontales.

I.3 Le rayonnement solaire au sol

L'irradiation solaire arrive au sommet de l'atmosphère sous forme d'ondes électromagnétiques. Elle se subdivise premièrement en deux parties: La première est directement réfléchi vers l'espace avant de pénétrer dans l'atmosphère et la deuxième partie s'enfonce dans l'atmosphère, subit des réfractions et des diffractions pour être à son tour soit renvoyée vers l'espace soit vers le sol. Ces phénomènes font que l'on peut regrouper les rayons solaires parvenant au sol en trois composantes principales, qui sont le direct, le diffus et le réfléchi [19].

I.3.1 Le rayonnement solaire direct normal I_N

C'est le rayonnement solaire qui se forme de rayons parallèles provenant du soleil sans avoir été dispersé par l'atmosphère. Le rayonnement direct est le rayonnement incident sur un plan donné en provenance d'un petit angle solide centré sur le disque solaire, il parvient en ligne droite et par temps clair. C'est ce type de rayonnement qui est utilisé dans les systèmes à concentration (four solaire, centrales thermo-solaires et photovoltaïques).

I.3.2 Le rayonnement solaire diffus D

C'est une partie du rayonnement du soleil qui a subi de multiples dispersions par les molécules et les *aérosols* de l'atmosphère qui sont définis comme des éléments fins solides et/ou liquides en suspension dans l'atmosphère, ils peuvent être attribués aux activités humaines (anthropiques) ou à des sources naturelles. Les principales sources naturelles d'aérosols sont les océans, les déserts et les volcans. C'est un rayonnement qui provient de

l'atmosphère, il contribue à l'éclairement mais, il ne peut pas être concentré (à l'inverse du rayonnement direct). Le rayonnement diffus, produit par des processus complexes, est beaucoup plus difficile à analyser que le rayonnement direct.

I.3.3 Rayonnement solaire réfléchi d_{re}

Le rayonnement solaire qui arrive sur le sol est en partie réfléchi. La réflexion du rayonnement solaire par le sol se fait de façon sélective (couleur des corps). Par temps clair, le problème est compliqué par la présence des zones d'ombre. Pour caractériser de façon globale les propriétés réfléchives du sol, on utilise son *albédo*, c'est-à-dire le rapport du rayonnement réfléchi au rayonnement incident sur toutes les fréquences et tous les angles d'incidences.

La fraction d'énergie réfléchie par rapport à l'énergie incidente est appelée *Albédo*. Ce terme étant généralement réservé au sol ou aux nuages, est une valeur moyenne de leur réflectance pour le rayonnement considéré et pour tous les angles d'incidences possibles. Il dépend fortement de la nature du sol et de l'état de la surface. L'albédo total du système Terre/Atmosphère vaut environ 0.3 ce qui signifie que 30% du rayonnement solaire est réfléchi et 70% est absorbé. L'albédo des forêts est proche de 0.15, celui des surfaces désertiques vaut environ 0.35, tandis que celui des surfaces englacées s'élève à 0.6 [8].

I.3.4 Le rayonnement solaire global G_h

C'est le rayonnement reçu sur une surface provenant de toute la voûte céleste et du disque solaire

I.3.4.1 Rayonnement global sur une surface horizontale G_h

Le rayonnement global parvenant au sol sur une surface horizontale G_h est la somme du rayonnement direct et du diffus (figure I.11), selon la formule suivante :

$$G_h = D_h + I_N \cos(\theta_s) \quad (I.11)$$

Où θ_s est l'angle zénithal.

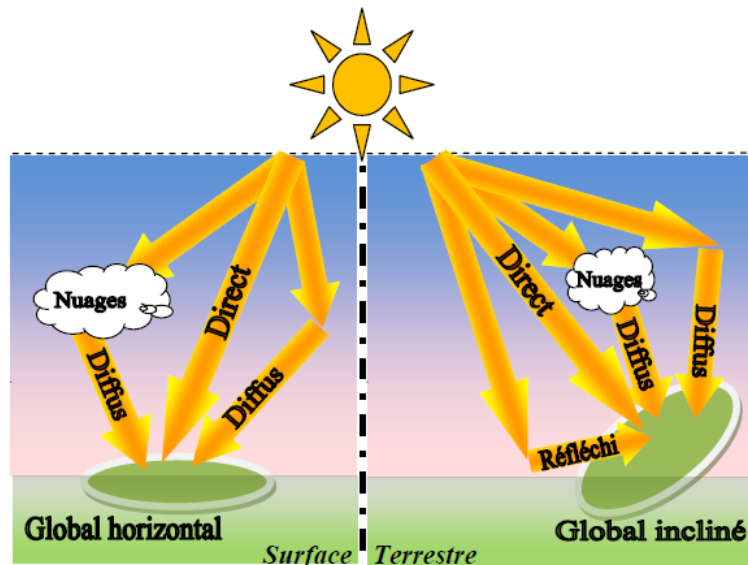


FIGURE I.11 Composantes principales du rayonnement solaire.

I.3.4.2 Le rayonnement global sur une surface inclinée

Le rayonnement solaire global sur une surface arbitraire ayant une inclinaison β par rapport à l'horizontale et un angle azimutal γ est la somme des trois composantes directe, diffuse et réfléchi par le sol du rayonnement solaire (figure I.11)[8, 20].

$$G_{\beta} = D_{\beta} + I_{N_{\beta}} + d_{re} \quad (\text{I.12})$$

L'énergie solaire varie en fonction de :

- La latitude du lieu ;
- L'orientation et l'inclinaison du plan de collecte ;
- La saison (maximale en été, minimale en hiver) ;
- La situation atmosphérique (ciel clair ou couvert) ;
- La hauteur du soleil sur l'horizon (maximale à midi solaire).

I.4 Rayonnement solaire et production photovoltaïque

Le terme « photovoltaïque » désigne le processus physique qui consiste à transformer l'énergie lumineuse en énergie électrique par le transfert de l'énergie des photons aux électrons d'un matériau. Le rayonnement global est induit par un apport d'énergie à la surface terrestre. Cependant, outre l'aspect calorifique que ce dernier représente spontanément, il est possible à partir de cellules spécialement créées, de convertir cette énergie surfacique en énergie électrique. Ce processus passe par l'utilisation d'un module photovoltaïque intégré

dans un système photovoltaïque complet. Il est associé à un système de stockage (batterie, H₂, volant d'inertie, etc.) afin de stocker l'énergie produite.

I.4.1 L'effet photoélectrique

I.4.1.1 Génération de l'énergie électrique à partir du rayonnement solaire

L'effet photoélectrique a été initialement découvert par le physicien allemand Heinrich Rudolf Hertz (1886). Il réalisa expérimentalement qu'un matériau métallique exposé à la lumière pouvait émettre des électrons. Cette découverte fut appelée « effet photoélectrique ». Malheureusement, Hertz ne fut pas en mesure d'expliquer théoriquement le phénomène, car certaines caractéristiques de cet effet n'étaient pas formalisables avec la théorie classique de l'électromagnétisme de l'époque. En effet, ces dernières sont réalisées en associant deux matériaux semi-conducteurs de types différents, l'un est dopé *N* et l'autre est dopé *P*, comme le montre la figure II.12. De nombreuses technologies de cellules photovoltaïques existent. Les plus fréquentes sont celles en Silicium cristallin et les couches minces [21].

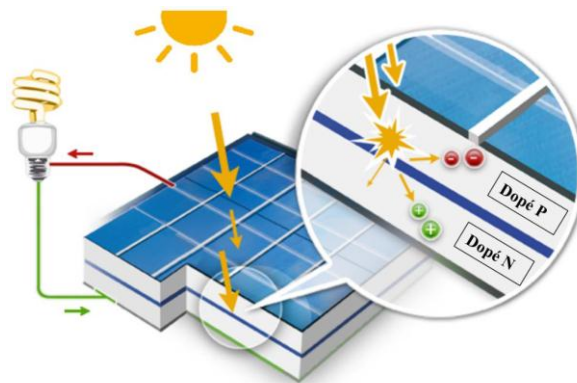


FIGURE I.12 L'effet photovoltaïque [22].

Quand un photon de lumière, d'énergie suffisante, heurte un atome de la surface du matériau photovoltaïque, il excite un électron et l'arrache de sa structure moléculaire, créant ainsi une paire « électron-trou ». Les électrons et les trous sont des charges de signes opposés, et c'est cette opposition qui génère une différence de potentiel électrique (l'effet photovoltaïque). L'association de deux types de matériaux pour la création de la jonction permet de récupérer les charges avant la recombinaison des électrons dans le matériau (redevient alors neutre). La présence de la jonction PN permet ainsi le maintien d'une circulation d'un courant électrique continu jusqu'à ses bornes. Ce courant est alors recueilli par des fils métalliques très fins

connectés les uns aux autres et acheminé à la cellule suivante. Le courant s'additionne en passant d'une cellule à l'autre jusqu'aux bornes de connexion du panneau, et il peut ensuite s'additionner à celui des autres panneaux raccordés en « champs ».

I.4.1.2 Les systèmes photovoltaïques

Dans des conditions d'ensoleillement standards (1000W/m^2 ; 25°C), la puissance maximale délivrée par une cellule en silicium de 150 cm^2 est d'environ $2,3\text{W}$ sous une tension de $0,5\text{V}$. Une cellule photovoltaïque élémentaire constitue donc un générateur électrique de faible puissance, insuffisant en tant que tel pour la plupart des applications domestiques ou industrielles [23]. Les générateurs photovoltaïques sont, de ce fait, réalisés par association, en série et/ou en parallèle, d'un grand nombre de cellules élémentaires. Outre cette juxtaposition de cellules, le courant généré par les cellules photovoltaïques produit uniquement du courant continu à partir du rayonnement solaire.

L'utilisation de ce courant continu diffère d'une installation à l'autre, selon le but de celle-ci. On distingue principalement deux types d'utilisation selon si l'installation photovoltaïque est connectée ou non à un réseau de distribution d'électricité. Les installations non connectées peuvent directement consommer l'électricité produite. C'est le cas des sites non raccordés au réseau électrique (en montagne, sur des îles ou des voiliers, un satellite, ... etc.) qui peuvent en outre être équipés de batteries, d'accumulateurs pour disposer d'électricité au cours de périodes sans lumière (la nuit notamment).

Des installations photovoltaïques peuvent également être connectées à un réseau de distribution. Sur les grands réseaux de distribution (Amérique du Nord, Europe, Japon, etc.) des installations photovoltaïques produisent de l'électricité et l'injectent dans le réseau. Pour ce faire, ces installations transforment le courant continu en courant alternatif aux caractéristiques du réseau (fréquence de 50 Hz en Europe ou 60 Hz en Amérique du Nord par exemple). Elles n'ont pas besoin d'installation de stockage (batteries) car l'électricité est consommée à l'instant où elle est produite par les consommateurs les plus proches sur le réseau. Ce raccordement nécessite l'utilisation d'un onduleur, c'est-à-dire un dispositif qui a pour fonction de transformer un courant électrique continu en courant alternatif, identique à celui circulant sur le réseau. Il existe des onduleurs issus de différente technologie, cependant, leurs principes et objectifs sont les mêmes : faire passer l'énergie du *DC* (courant continu) à l'*AC* (courant alternatif).

Cette transition entraîne une modification des performances globales du système. Il convient d'introduire le facteur PR (performance ratio) qui dépend du choix des composants comme l'onduleur, la température, l'énergie perdue...etc., et peut modifier considérablement la production d'électricité au cours du temps. Le PR est calculé par le rapport entre le rendement mesuré et celui théorique. La figure 1.13 détaille le principe d'utilisation d'une source renouvelable d'énergie de type solaire raccordée au réseau électrique.

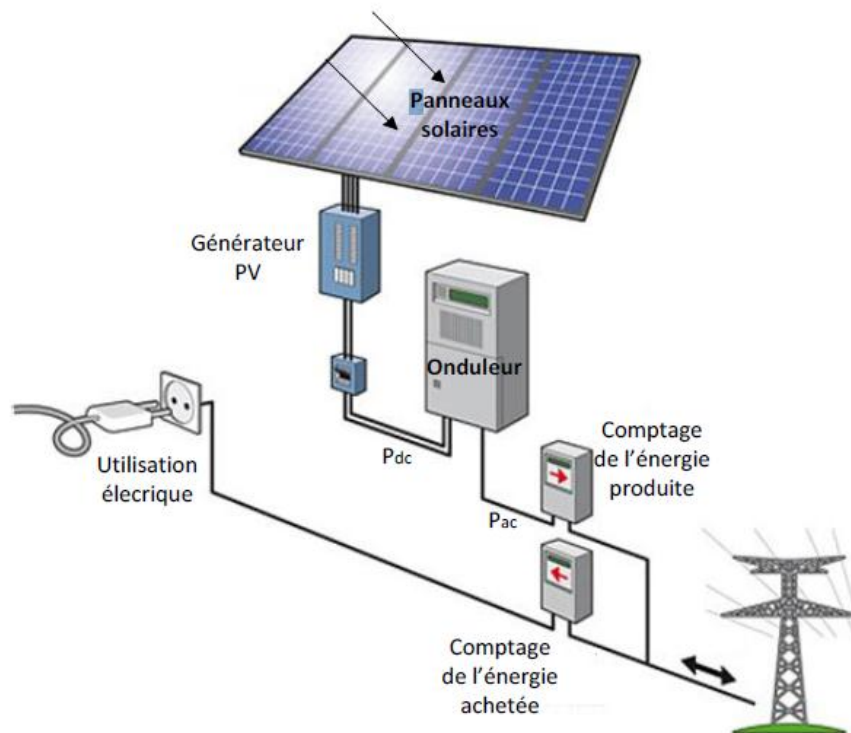


FIGURE 1.13 Principe du raccordement d'un module PV au réseau électrique.

I.4.1.3 L'énergie solaire thermique

Une grande partie du rayonnement solaire n'est pas convertie en électricité et accroît leur température entraînant ainsi une baisse de leur rendement électrique. Les capteurs solaires PVT hybrides sont des systèmes utilisant des modules photovoltaïques (PV) comme absorbeur thermique. Par récupération d'une partie de la chaleur dissipée par les modules PV à l'aide d'un fluide caloporteur, ces capteurs solaires hybrides permettent la production simultanée d'énergies thermique et électrique. Ils associent ainsi un capteur solaire photovoltaïque intégré au bâti et un capteur solaire thermique [24].

I.5 Applications des prédictions d'irradiation et d'énergie photovoltaïque

Une utilisation efficace de la production d'énergie fluctuante des systèmes photovoltaïques nécessite des informations prévisionnelles fiables pour les stratégies de gestion et d'exploitation. Cette prédiction est nécessaire pour l'intégration au réseau des systèmes PV ainsi que pour les petits réseaux et les systèmes autonomes.

L'application la plus importante des prédictions d'énergie photovoltaïque est la prise en charge d'une intégration économique de grandes quantités d'énergie solaire dans le système d'alimentation en électricité. La contribution de la production d'énergie renouvelable en particulier des systèmes PV, des centrales thermiques solaires et des convertisseurs d'énergie éolienne à l'alimentation électrique est en augmentation constante. Les entreprises de services publics et les gestionnaires de réseau de transport doivent faire face à la fluctuation des apports de ces différentes sources d'énergie renouvelables.

L'un des principaux objectifs de la prévision de la puissance photovoltaïque consiste à optimiser la production et la consommation la valeur de l'énergie renouvelable produite sur le réseau. D'un point de vue technique, cet objectif est atteint en réduisant le besoin d'équilibrer l'énergie produite et l'énergie de réserve. Le besoin d'informations prévisionnelles sur la production prévue d'énergie solaire et éolienne augmente avec la quantité d'énergie installée. Aujourd'hui, les systèmes de prévision de l'énergie solaire sont largement utilisés au niveau opérationnel et ont montré leur fort impact économique et leurs avantages pour l'intégration de l'énergie solaire dans le réseau électrique [25,26]. En conséquence, la prévision des rendements solaires devient de plus en plus importante pour les services publics qui doivent intégrer des quantités croissantes d'énergie solaire, en particulier pour les pays où la législation encourage le déploiement de centrales solaires.

Il existe également une variété d'applications de prévisions d'irradiation non liées à la photovoltaïque. L'utilisation de prévisions d'irradiation directe pour optimiser les stratégies d'exploitation des centrales thermiques solaires est aussi important [27]. En outre, à très court terme (quelques minutes), les prévisions d'éclairement énergétique direct pourraient être utiles pour le contrôle des récepteurs dans les centrales solaires thermiques. D'autres exemples incluent l'utilisation des prévisions météorologiques et de l'énergie solaire pour le contrôle du chauffage, de la ventilation et du refroidissement des bâtiments, l'utilisation des prévisions d'irradiation pour améliorer la gestion des réseaux de chauffage urbain intégrant le chauffage solaire de l'eau et l'utilisation de prévisions en agriculture, par exemple pour la récolte[28].

I.6 Conclusion

Le rayonnement solaire peut être utilisé pour produire soit directement de l'électricité à l'aide de semi-conducteurs photovoltaïque, soit de la chaleur solaire thermique pour le chauffage ou la production électrique. En exploitant le rayonnement solaire, considéré comme la ressource principale, qui est très considérable à l'échelle planétaire, et plus spécialement dans notre pays, ce type d'énergie offre une véritable alternative à la consommation des ressources fossiles classiques avec un faible impact environnemental et un fort potentiel de réduction des coûts.

L'aspect intermittent de cette source d'énergie montre l'importance et la nécessité de connaître le potentiel solaire avant de procéder à l'installation de n'importe quel système solaire. Le meilleur moyen d'identifier ce potentiel sur un site donné est de disposer d'une bonne série de mesures de différents paramètres de rayonnement au sol, ce qui n'est pas disponible dans plusieurs lieux et localités. Pour remédier au manque d'informations et mesures dans plusieurs situations, on fait recours aux modèles d'estimation et de prédiction de l'énergie solaire incidente sur le site en question.

Dans le prochain chapitre, nous présentons un état de l'art des méthodes de prédiction avec une petite définition sur des fondamentaux pour la prévision du GHI.

Chapitre II

Prédiction de l'irradiation solaire :

Etat-de-l'art

II.1 Introduction

Les séries temporelles, appelées aussi séries chronologiques ou même chroniques, occupent une place importante dans tous les domaines de l'observation ou de la collecte de données. Le terme série est employé pour évoquer des objets (des nombres ou des mots) classés dans un certain ordre. L'ordre utilisé est le temps, plus précisément, on utilise une mesure de temps exprimée en : années, mois, jours, minutes ou n'importe quelle autre unité de mesure. En d'autres termes, les séries temporelles associent des objets divers à des marques temporelles successives plus ou moins équidistantes. La série est dite temporelle, parce qu'elle indexe l'objet ou la valeur par le temps.

Dans ce chapitre, nous commencerons par expliquer ce que sont les séries temporelles, Nous dégagerons ensuite les différentes méthodologies de prédiction existantes dans la littérature et plus particulièrement celles liées au rayonnement global d'irradiation solaire (GHI) en détaillant les principaux modèles utilisables. L'état de l'art de la prédiction des séries temporelles va nous servir pour comparer les résultats de notre approche de prédiction avec la littérature.

II.2 Définition d'une série temporelle

Une série temporelle est définie comme étant une suite de mesures ou d'observations au cours du temps représentant un phénomène [29, 30,31]. Il est aujourd'hui, de plus en plus fréquent de parler de « prédiction ». Pour utiliser le formalisme des séries temporelles, il est nécessaire, au préalable de donner certaines définitions. Ainsi, la valeur courante en t de la chronique X est notée X_t où, le temps, est compris entre 1 et n, avec n le nombre total d'observations. On appelle h le nombre de points ou de valeurs à prédire de la série temporelle. La prédiction de la série temporelle de $(n + 1)$ à $(n + h)$, connaissant l'historique de X_1 à X_n , porte le nom d'horizon de prédiction (horizon 1, ..., horizon h). Aussi pour un

horizon 1 (cas le plus simple), le formalisme générale la prédiction sera représenté par l'équation II-1 où ε représente l'erreur entre la prédiction et la mesure, f_n le modèle à estimer et t le paramètre temporel qui prend les $(n - p)$ valeurs suivantes : $n, n - 1, \dots, p + 1, p$. Où n est le nombre d'observations et p le nombre de paramètres du modèle, on suppose que $n \gg p$.

$$X_{t+1} = f_n(X_t, X_{t-1}, \dots, X_{t-p+1}) + \varepsilon(t + 1) \quad (\text{II.1})$$

où ε est l'erreur entre la valeur prédite et la valeur mesurée.

Il est toujours utile en première analyse de représenter l'évolution temporelle d'un phénomène (profil et allure de la série) à l'aide d'un graphique ayant en ordonnée la valeur du phénomène x_t et en abscisse le temps t . Ainsi sur la figure II-1, on peut se rendre compte que le phénomène de rayonnement global (GHI) X_t est un ensemble de signaux périodiques avec un bruit (lié à la couverture nuageuse) qui semble plus important durant les mois d'hiver que durant ceux d'été.

II.3 Approche stochastique du signal

Les techniques traditionnelles d'analyse des séries temporelles procèdent souvent par décomposition et recombinaison. L'approche de décomposition suppose que la structure de chaque chronique peut être scindée en éléments simples (modélisables), et donc plus facilement prévisibles, pour ensuite être reconstituée pour donner la prédiction de la chronique. Les premières études [30,32] sur les chroniques ont amené à considérer de façon standard trois grandes composantes de séries temporelles:

- **la tendance** : décrit le mouvement sur le long terme (extra-annuel).
- **la composante saisonnière** : est une composante cyclique relativement régulière de période intra-annuelle. Il existe différents types de saisonnalités pour le rayonnement global on peut parler de périodicité intrinsèque « rigide », ou déterministe car elle est bien marquée et répétitive.
- **la composante résiduelle (bruit ou résidu)** : correspond à des fluctuations irrégulières, en général de faible intensité mais de nature aléatoire. C'est une composante qui existe par défaut, elle regroupe ce que les autres composantes n'ont pu intégrer.

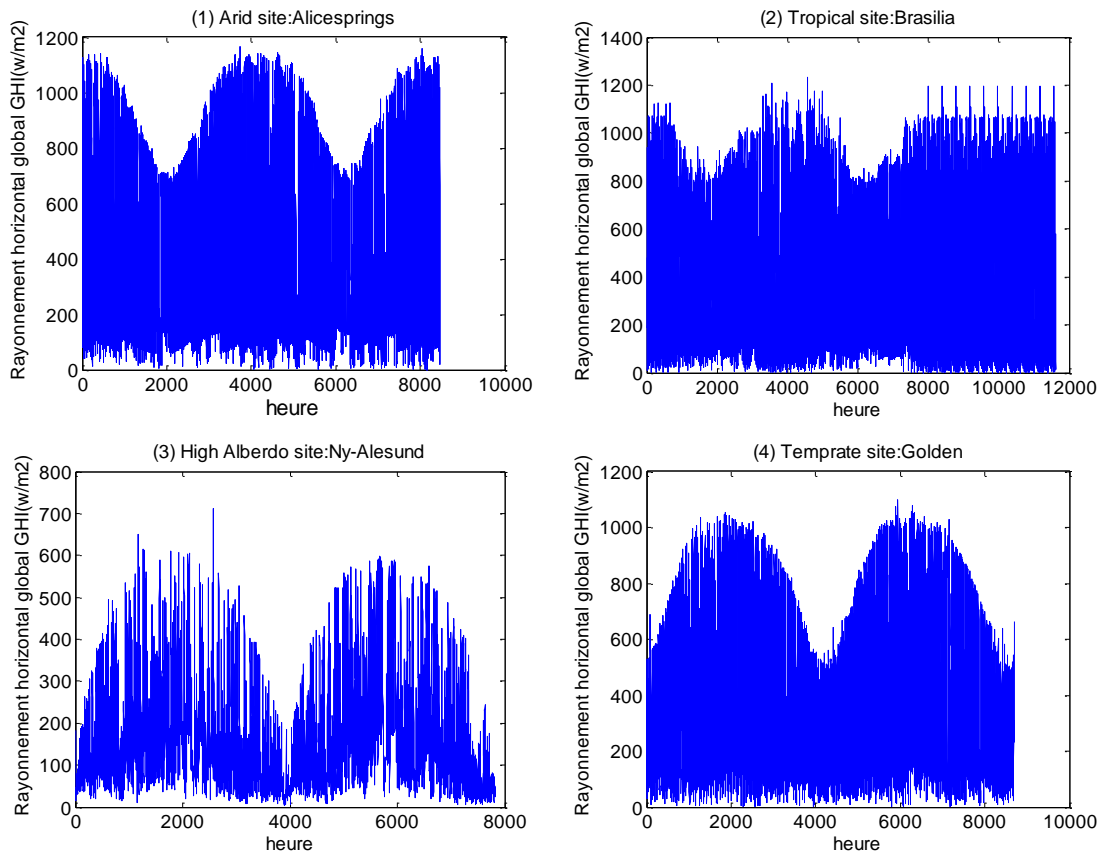


FIGURE II.1 Exemples de rayonnement global horizontal pour les sites Alice Springs, Brasilia, Ny-Alesund et Golden.

II.4 Fondamentaux pour la prévision du GHI

Dans cette section, quelques concepts de base sur l'irradiation solaire et la production d'énergie solaire sont expliqués, ce qui facilitera la compréhension des parties restantes du chapitre.

II.4.1 Modèles ciel clair

L'irradiation solaire est principalement influencée par la présence de nuages, qui rend difficile la prédiction d'irradiation. Cependant, il est possible d'estimer l'éclairement énergétique par temps clair, c'est-à-dire en l'absence de nuages. Cette valeur peut être utilisée pour calculer des indices solaires, normaliser des mesures et obtenir la production d'une centrale solaire dans des conditions stationnaires.

Généralement, les modèles de ciel clair sont alimentés avec des variables météorologiques et une géométrie solaire, en utilisant des modèles de transfert radiatif pour établir les connexions entre les entrées.

Il existe un grand nombre de modèles de ciel clair, qui diffèrent les uns des autres principalement par les entrées nécessaires à chaque modèle. Certains des modèles de ciel clair

les plus largement utilisés sont le modèle Solis [33], le modèle de l'atlas européen du rayonnement solaire ESRA (European Solar Radiation Atlas) [34], le modèle d'Ineichen [35] et le modèle d'évaluation de référence sur la transmission solaire[36].

Une description détaillée de ces modèles et d'autres modèles de ciel clair sont présentés dans les références [37] et [38]. Certains modèles ne nécessitent qu'une entrée (ESRA) alors que d'autres nécessitent un grand nombre d'entre eux (Solis). Comme détaillé dans [39], le choix d'un modèle à ciel clair pour un lieu déterminé dépend de la disponibilité et de la qualité des données d'entrée, qui constituent le principal facteur limitant.

En pratique, le modèle ciel clair présente les cycles journaliers (jour/nuit) et annuels comme illustré sur la figure II.2.

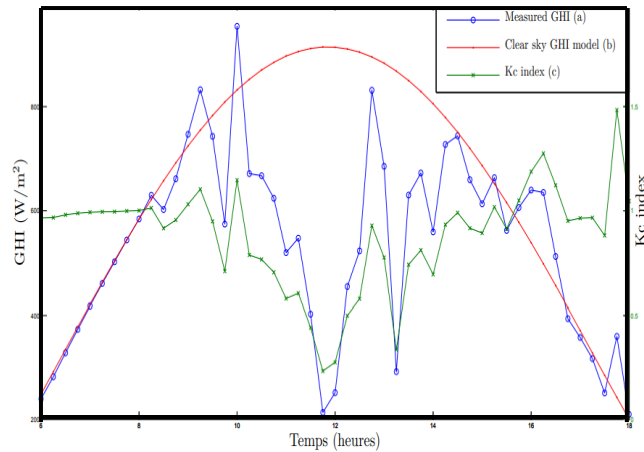


FIGURE II.2 Présentation des courbes du modèle clair ciel, GHI_m et l'indice k_c . sur une journée de mesure [38].

II.4.2 Indice de ciel clair K_c

Il existe deux paramètres, l'indice de ciel clair k_c et l'indice de clarté k_t , qui sont largement utilisés pour classifier les conditions météorologiques et pour calculer des modèles de *persistance intelligents*. Ils sont obtenus de manière similaire, mais différents par la variable de normalisation. L'indice de ciel clair est le rapport entre l'irradiation solaire mesurée et l'irradiation solaire modélisée du ciel clair I_c . Cet indice est donc généralement préféré au GHI dans la plupart des études sur la prévision ou l'estimation locale du GHI.

$$K_c = \frac{GHI}{GHI_{\text{ciel clair}}} \quad (\text{II.2})$$

II.4.3 Modèle de référence : la Persistance

Le modèle de persistance est reconnu comme la référence des méthodes de prédiction du GHI pour sa simplicité et son universalité. Il s'appuie sur l'hypothèse de persistance de la

mesure d'un instant à un autre de proche en proche. La prédiction d'une variable temporelle $X(t)$ par persistance peut être formulée par l'équation suivante :

$$\hat{X}(t + 1) = \hat{X}(t) \quad (\text{II.3})$$

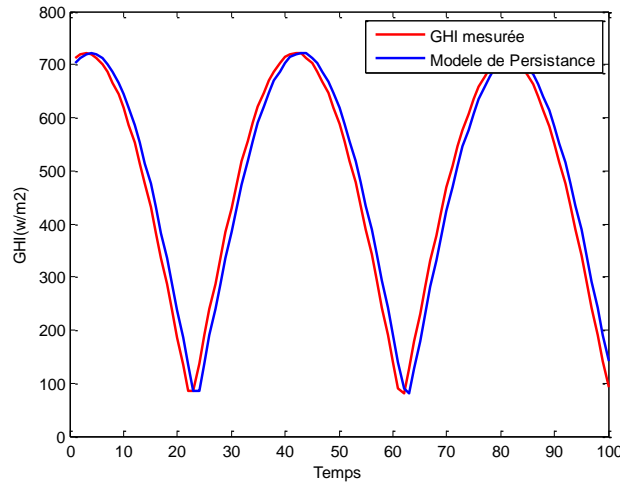


FIGURE II.3 Modèle persistance appliqué directement au GHI pour une journée ciel clair.

L'application de ce modèle à la prédiction du GHI nécessite l'utilisation de l'index K_c . En effet, en appliquant le modèle directement au GHI on obtiendrait une prédiction toujours en décalage avec la mesure à cause de la variation journalière connue (figure II.3). La prédiction de GHI par persistance s'écrit donc avec les équations suivantes :

$$\begin{aligned} \hat{GHI}(t+1) &= \hat{K}_c(t+1) \times GHI_{\text{cielClair}}(t+1) \\ \hat{K}_c(t+1) &= \hat{K}_c(t) \end{aligned} \quad (\text{II.4})$$

Cette méthode de référence a l'avantage d'être très performante dès que les conditions météorologiques sont assez stables. Les scores de persistance sont d'ailleurs très importants pour évaluer et comparer des méthodes de prédiction développées et testées sur des jeux de données différents. Les scores du modèle de persistance donnent d'une part une référence pour la prévision mais aussi une information sur la difficulté de celle-ci, ou du moins sur la variabilité relative des jeux de données. Un bon score du modèle de persistance indique un site où les conditions météorologiques sont relativement stables et vice versa [40].

II.5 Les méthodes de prédiction des séries temporelles

La prédiction des séries temporelles est un problème qui recouvre de nombreux domaines d'application. Les études menées dans la finance et l'économétrie ont permis de dégager de nombreux modèles plus ou moins sophistiqués. Ces derniers ont été repris dans le

cadre d'autres thématiques, dont la prédiction du rayonnement solaire par modélisation des séries temporelles.

En ce qui concerne l'aspect temporel des prédictions, il est important d'introduire trois concepts : horizon de prédiction h , résolution de prédiction r et intervalle de prédiction f_i .

L'horizon de prédiction est la durée entre le temps présent t et le temps effectif des prédictions. La résolution de prédiction décrit *la fréquence* à laquelle les prédictions sont émises et l'intervalle de prédiction indique l'intervalle de temps des prédictions.

Il existe de nombreux modèles permettant de faire une prédiction de séries temporelles. Il est possible de les rassembler en quatre grands groupes.

- **Les modèles de type « naïf »** qui sont primordiaux pour vérifier la pertinence des modèles complexes. On peut citer la persistance, la moyenne ou les k -plus proches voisins ;
- **Les modèles à probabilités conditionnelles** rarement mentionnés dans la littérature en ce qui concerne le rayonnement global. On peut citer les chaînes de Markov et les prédictions basées sur les inférences Bayésiennes ;
- **Les modèles de type connexionnistes (réseau de neurones)** et plus particulièrement le Perceptron Multi-Couche (*PMC*), qui est un type de réseaux de neurones à fort potentiel prédictif et le plus souvent utilisé.
- **Les modèles de référence**, de par le nombre d'études les ayant utilisés, ils sont issus de la grande famille des modèles autorégressifs ;

Plusieurs modèles de prédiction du rayonnement solaire existent dans la littérature. Ils diffèrent par : les entrées disponibles, leur classification, l'horizon et la méthode de prédiction utilisée.

II.6 Classification de la prédiction de rayonnement solaire

Les chercheurs ont classé la prédiction du rayonnement solaire dans différentes catégories en fonction de différents facteurs qui sont liés à l'horizon de prédiction, données historiques de l'irradiation solaire et autres modèles de données météorologiques.

II.6.1 Classification selon l'horizon de prédiction

Le but et la précision d'un modèle de prédiction dépend de l'horizon de prédiction. Lipperheide et al. [41] ont analysé les performances de l'énergie photovoltaïque sur différents horizons de prédiction, tels que 20s, 40s, 60s,.... 180s. L'erreur de prédiction (RMSE) du modèle de prédiction proposé est comprise entre 3,2 et 15,5% pour les horizons de prédiction

compris entre 20 et 180 s. Lonij et al. [42] ont conçu un modèle de prédiction de la puissance PV où les erreurs changent avec les horizons de prédiction allant de 15 min à 90 min. La précision de la prédiction varie avec le changement d'horizon de prédiction dans le même modèle avec les mêmes paramètres. Par conséquent, l'horizon de prédiction doit être pris en compte avant de concevoir le modèle de prédiction approprié. Il n'existe aucun critère bien défini pour classer le modèle de prédiction en fonction de l'horizon de prédiction. Néanmoins, selon la plupart des rapports de recherche, la prédiction de l'énergie solaire peut être divisée en trois catégories en fonction de l'horizon temporel, comme illustré à la figure II.4.

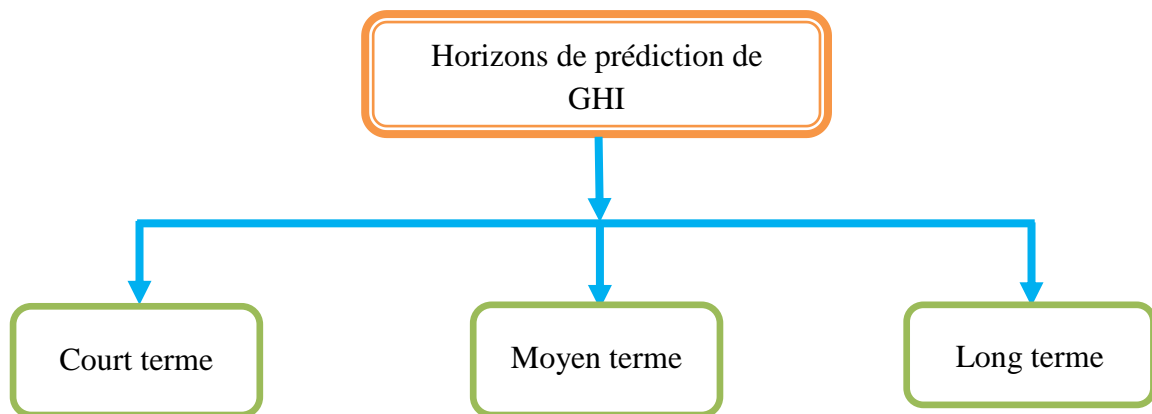


FIGURE II.4 Classification de la prédiction de GHI en fonction de l'horizon temporel.

II.6.1.1 Prédiction à court terme

La prédiction du rayonnement solaire réalisée pendant une heure, plusieurs heures, un jour ou jusqu'à sept jours sont appelées prédictions à court terme. La prédiction à court terme de l'énergie solaire garantit l'engagement, la planification et la répartition de l'énergie électrique des unités. Ce type de modèle de prédiction est utile pour concevoir un système de gestion de l'énergie intégré au système photovoltaïque. La prédiction à court terme améliore également la sécurité du fonctionnement du réseau.

II.6.1.2 Prédiction à moyen terme

La prédiction de rayonnement solaire à moyen terme est effectuée sur plus d'une semaine à un mois. Ce type de prédiction aide à la planification du système d'alimentation et le calendrier de maintenance en prévoyant la disponibilité de l'énergie électrique à l'avenir.

II.6.1.3 Prédiction à long terme

La prédiction à long terme du rayonnement solaire se fait à partir d'un mois à un an. Ce type de prédiction de rayonnement solaire est utile pour la planification de la production,

du transport et de la distribution de l'électricité en dehors des enchères d'énergie et de la sécurisation des opérations.

Cependant, certains chercheurs ont divisé l'horizon prévisionnel de l'énergie solaire en quatre catégories [43]. La quatrième catégorie est appelée "horizon de prédiction à très court terme". La prédiction à très court terme est prise en compte pour quelques secondes, une minute ou plusieurs minutes (<1 h) de prédiction. Ce type de prédiction a été fait pour le lissage de la puissance, répartition de l'électricité en temps réel et réserves optimales.

II.6.2 Classification de la prédiction de rayonnement solaire sur la base des données historiques

Les méthodes de prédiction de rayonnement solaire peuvent être classées en quatre types basés sur l'utilisation de données historiques de l'irradiation solaire et les variables météorologiques associées. Ces modèles sont (a) la persistance, (b) les méthodes statistiques, (c) l'apprentissage automatique et (d) les méthodes hybrides, comme indiqué à la figure II.5 avec les sous-catégories.

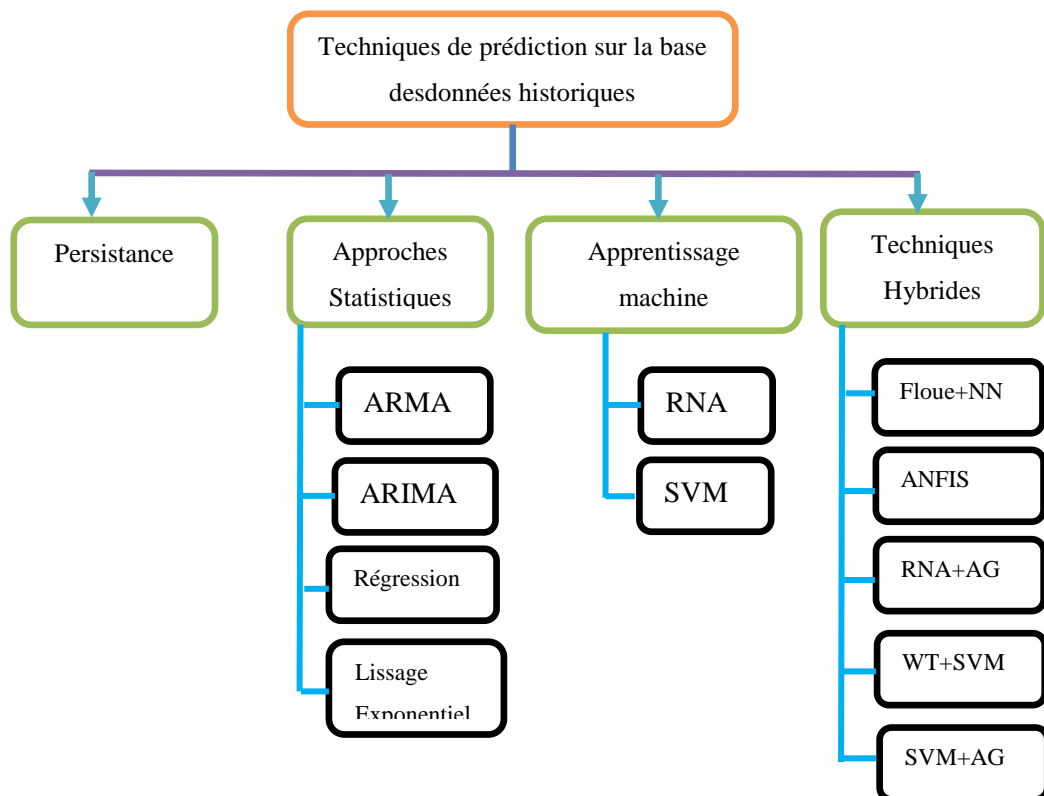


FIGURE II.5. Classification des techniques de prédiction utilisant des données historiques [43].

Dans le modèle de persistance, la valeur de l'irradiation solaire (IS) estimée à l'instant IS_{t+1} est égale à la valeur réelle précédente IS_t . Dans cette méthode, seules les données

historiques sont nécessaires pour prévoir l'IS. Généralement, ce modèle est utilisé comme modèle de référence (benchmark). Dans les méthodes statistiques, l'IS est prédite via l'analyse statistique des différentes variables d'entrée, par conséquent, les données chronologiques historiques sont utilisées dans ces méthodes.

Normalement, les méthodes précédentes sont adoptées pour la prédiction à court terme. Les données historiques récentes devraient être utilisées dans ces méthodes pour augmenter la précision du modèle. L'exigence de la taille de la série de données d'entrée dans ce modèle est moindre par rapport à la méthode d'apprentissage automatique. Inversement, dans l'apprentissage automatique, une grande quantité de données est nécessaire pour prévoir correctement l'IS avec précision. Les modèles d'apprentissage automatique sont des modèles intelligents, ils peuvent modéliser des données linéaires, non-linéaires et non-stationnaires. La combinaison de deux techniques ou plus est utilisée pour concevoir un modèle de prévision appelé modèle hybride. Le modèle hybride affiche de meilleurs résultats qu'un modèle unique pour différents problèmes de prévision en combinant les avantages de chaque technique.

II.6.3 Classification des méthodes de prédiction du rayonnement solaire

De nombreux chercheurs ont étudié les différentes méthodes de prédiction de rayonnement solaire. Ils ont proposé et développé plusieurs techniques et modèles pour prévoir la prédiction de rayonnement solaire. Parmi les méthodes de prédiction utilisées dans la littérature, certaines sont basées sur des modèles linéaires tels que la régression linéaire (Linear Regression LR), la moyenne mobile autorégressive (Auto-Regressive Moving Average ARMA) et le processus autorégressif (Auto-Regressive AR) [44, 45]. Cependant, le comportement non linéaire du rayonnement solaire a suscité les chercheurs à proposer plusieurs autres modèles : ceux basés sur le modèle numérique météorologique, les ondelettes, les modèles flous, les systèmes d'inférence neuro-flou adaptatifs (Adaptive Neural Fuzzy Inference Systems ANFIS), les forêts aléatoires (Random Forests RF), les k-plus proches voisins (k-Nearest Neighbors kNN) ainsi que les Réseaux de Neurones Artificiels (RNA) [46].

II.6.3.1 Processus autorégressifs

Ces techniques permettent d'estimer la relation entre une variable dépendante (irradiation solaire) et des variables indépendantes, appelées des fois *prédicteurs*. Selon la manière dont les séries chronologiques sont traitées (linéaire ou non linéaire et stationnaire ou non-stationnaire), une autre classification apparaît. Les séries temporelles stationnaires sont

des séries chronologiques fluctuant autour d'une moyenne statique alors que les séries non stationnaires ne montrent pas une telle moyenne.

Nous trouvons ici les modèles autorégressifs (AR), qui modélisent la sortie sous forme d'une combinaison linéaire des valeurs retardées des prédicteurs; les modèles simples de moyenne mobile (MA) utilisés lorsque les données présentent une variance constante sur une position d'équilibre autour de la moyenne, pour laquelle la moyenne des données historiques est utilisée comme prédiction; les modèles double MA utilisés lorsqu'il y a une tendance; les modèles à moyenne mobile autorégressive (ARMA) qui tiennent compte à la fois des valeurs antérieures décalées et des erreurs; Les modèles AR eXogenous (ARX), qui ajoutent des données exogènes à un modèle AR, et les modèles de moyenne mobile autorégressive à variables eXogenes (ARMAX), qui introduisent des variables externes. Dans l'analyse des séries chronologiques (c.-à-d. à partir de la prédiction numérique du temps), pour traiter l'aspect probabiliste, il existe certaines adaptations, telles que le vecteur AR (VAR) ou le vecteur ARX (VARX) [47,48].

II.6.3.2 Méthodes de régression

La méthode de régression est une méthode statistique utilisée pour établir une relation entre les variables explicatives et dépendantes. Dans ce modèle, la variable dépendante est prédite en connaissant les variables explicatives. Dans le cas de la prédiction de rayonnement solaire, le RS prédit est considéré comme une variable dépendante et les variables météorologiques sont considérés comme des variables explicatives. Oudjana et al. [49] ont développé un modèle pour prédire la production d'énergie photovoltaïque, comprenant deux modèles de régression différents : les régressions simples et les régressions linéaires multiples. Le modèle de régression utilisant l'irradiation solaire et la température en entrée a donné de meilleurs résultats par rapport au cas où l'un ou l'autre était considéré en entrée. Un modèle mathématique et plusieurs variables explicatives sont nécessaires pour concevoir un modèle de prévision basé sur la régression, ce qui constitue la faiblesse de cette méthode.

II.6.3.3 Méthodes de lissage exponentiel

La méthode de lissage exponentiel, suggérée pour la première fois par Brown [50], est connue sous le nom de méthode de lissage exponentiel simple. Cette méthode a été développée par Holt [51], connue sous le nom de méthode de Holt; Winter a également modifié cette méthode sous le nom de Holt-Winter méthode [52]. Dans la méthode de lissage exponentiel, un ensemble de pondérations des données historiques au lieu des pondérations

égales est imposé aux données passées. Cependant, les poids des données passées diminuent de manière exponentielle des points de données les plus récents aux plus distants. Le lissage exponentiel simple est une méthode simple, également connue sous le nom de MA pondérée de manière exponentielle (EWMA). La forme du modèle EWMA est écrite comme suit:

$$\widehat{X}_{t+1} = \alpha X_t + (1 - \alpha)\widehat{X}_t = \widehat{X}_t + \alpha(X_t - \widehat{X}_t) \quad (\text{II.5})$$

où α est la constante de lissage et peut prendre toute valeur entre 0 et 1. Cette méthode nécessite une prédiction initiale qui doit être estimée ou supposée. Par conséquent, l'ensemble initial de prédictions \widehat{X}_t est souvent considéré comme $\widehat{X}_t = X_t$. Le modèle de prédiction EWMA indique que la valeur de prédiction à la période $t + 1$ est égale à la somme des dernières valeurs prédites \widehat{X}_t plus un terme d'ajustement de l'erreur de prédiction $\alpha (X_t - \widehat{X}_t)$.

II.6.3.4 Méthodes physiques

La méthode de prédiction physique consiste en un ensemble de méthodes mathématiques. L'équation qui décrit l'état physique et le mouvement dynamique de l'atmosphère [53]. Les modèles physiques sont conçus en fonction des caractéristiques du site, telles que la localisation, les différentes variables météorologiques et les données historiques d'orientation. Ces modèles sont considérablement simples (lorsque basés uniquement sur l'irradiance solaire globale) ou compliqués (s'ils incluent des paramètres supplémentaires) [54]. La précision du modèle de prédiction physique est supérieure lorsque les conditions météorologiques sont stables [55]. Cependant, les performances de prédiction sont largement affectées par les fortes variations des variables météorologiques. Différentes études ont proposé [54,56], la méthode de prédiction physique combinée à différentes méthodes d'intelligence artificielle et statistiques afin de concevoir un modèle de prédiction hybride peut avoir généralement une meilleure précision de la prédiction. Cependant, ces modèles sont très sensibles à la prédiction météorologique.

II.6.3.5 Méthodes d'intelligence artificielle

- **Réseau de neurones artificiels (RNA)**

Le RNA est la méthode la plus efficace et populaire parmi les chercheurs depuis 1980. Cette méthode a été utilisée dans différentes applications de prédiction, y compris la prédiction du rayonnement solaire avec un niveau de réussite parmi les plus élevés. Les RNA sont largement utilisés pour prévoir le rayonnement solaire dans la plupart des recherches, en raison de la non-linéarité des données météorologiques. Le RNA est plus approprié que les

méthodes statistiques lorsqu'un lien non linéaire existe entre les données sans hypothèse préalable.

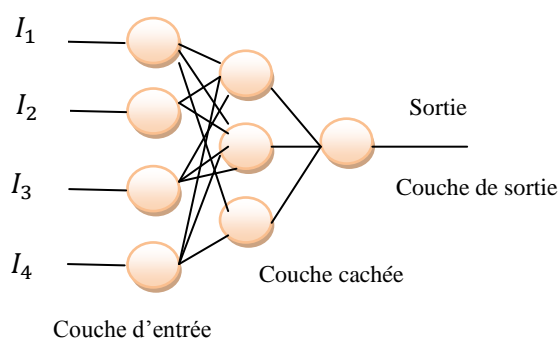
Les principaux composants d'un RNA sont l'entrée, la/les couches cachées, les couches d'entrée et de sortie, les neurones et les connexions. La couche d'entrée reçoit les différentes informations d'entrée. La couche cachée, qui peut consister en un seul ou plusieurs couches, analyse les informations d'entrée. La couche de sortie reçoit les résultats analysés et fournit la sortie (valeur prédite). La connexion fait un lien entre les neurones des différentes couches avec mise à jour des poids. La figure II.6 (a),(b) montre l'architecture de base d'un RNA et un modèle schématique de la procédure de traitement dans une cellule de neurone, respectivement.

La figure indique que la cellule neuronale a deux parties. La première partie est la «fonction de combinaison» qui produit une valeur en résumant toutes les entrées. La deuxième partie est la "fonction d'activation". La sortie du réseau est générée en additionnant les entrées pondérées en utilisant la fonction d'activation. Par conséquent, la fonction d'activation du réseau agit comme une fonction de compression pour transférer l'entrée sous la forme de sortie. La formule mathématique de base du RNA exprimée par [57] :

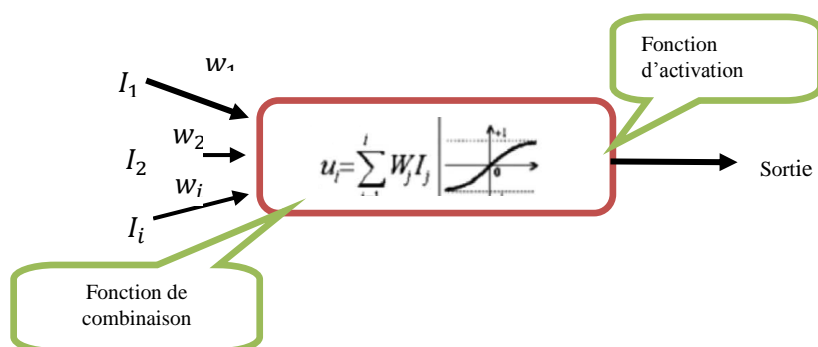
$$U_N = b + \sum_{j=1}^N (W_j \times I_j) \quad (\text{II.6})$$

Où U_N , b , N , W_j et I_j sont la sortie du réseau, poids de biais, nombre d'entrée, poids des connexions et entrée réseau, respectivement. Dans la plupart des cas, le RNA a plusieurs entrées et une ou plusieurs sorties.

L'apprentissage et les tests sont les deux opérations de base d'un RNA. À la phase d'apprentissage, le réseau est formé à l'aide d'une base de données *d'apprentissage* via l'algorithme d'apprentissage. L'algorithme d'apprentissage du RNA tente de mapper les relations d'entrées et de sortie(s) en mettant à jour les valeurs de poids synaptique. La sortie générée par le réseau est comparée à la valeur souhaitée, l'erreur de la sortie est ensuite calculée. Par conséquent, les poids et les biais (agissant en tant que compensateurs), les valeurs de RNA sont mises à jour en fonction de l'erreur. Ce cycle continue jusqu'à ce que la sortie souhaitée soit atteinte. Ainsi, le réseau fournit la sortie finale en fonction de la base de données utilisée et des valeurs de pondération du modèle. Cependant, la sortie du réseau varie en fonction du changement d'architecture, de la fonction d'activation et des entrées



(a)



(b)

FIGURE II.6 (a) Diagramme schématisé d'une structure RNA composée d'une couche d'entrée, une couche cachée et une couche de sortie.(b)modèle mathématique du neurone artificiel [59].

Dans la plupart des cas, les RNA ont une seule couche et sont utilisés pour résoudre les différents problèmes. Cependant, un nombre important de problèmes complexes qui sont liés au type de données peuvent utiliser un RNA monocouche. Dans plusieurs cas, des relations complexes d'entrée et de sortie existent entre différentes variables. Afin de surmonter ce problème complexe, les RNA ont été modifiés en plusieurs types qui suivent différentes architectures et différentes procédures de mappage entrée-sortie. Parmi les plus utilisés sont le perceptron multicouche (MLP), RNA à fonction de base radiale (RBF), RN récurrent (RNN), RN à régression générale (GRNN) et systèmes adaptatif d'interface neuro-floue (ANFIS).

Le RNA donne de bons résultats en apprenant différentes relations complexes et structures de calcul. Par conséquent, il est considéré comme un bon outil pour la prédiction des données chronologiques [60]. Yona et al. ont signalé que des erreurs dans les résultats de prédiction sont considérablement minimisées par RNA par rapport à FFNN [61]. GRNN est un réseau probabiliste qui fait la régression plutôt que des tâches de classification [62]. Avec

une forte capacité non-linéaire de cartographie, GRNN peut résoudre efficacement des problèmes non linéaires [63]. Cependant, la croissance de cette méthode nécessite une grande puissance de calcul. La sélection du paramètre de réseau σ est importante pour l'utilisation de GRNN [64]. Les chercheurs ont introduit un réseau modifié de BP pour remédier à ces limitations [65]. Parmi les méthodes liées à RNA, le réseau de neurones à propagation arrière (BPNN), qui a été largement utilisé en raison de son excellente fonction de cartographie non linéaire, convient à la résolution des problèmes de régression complexes [66].

ANFIS est un type de réseau multicouche adaptatif anticipé ;il est appliqué à la prédiction non linéaire, dans laquelle des échantillons antérieurs sont utilisés pour prévoir l'échantillon à l'avance [67]. Le réglage des fonctions d'appartenance est nécessaire pour améliorer les performances d'ANFIS [68]. Parmi les systèmes flous, l'ANFIS est le plus largement utilisé car il est moins coûteux en calcul et produit des résultats aussi robustes que les modèles statistiques [69].

- ***Machines à vecteurs de support (SVM).***

Il s'agit d'une méthode de modélisation supervisée, introduite pour la première fois par [70], puis développée par [71] pour être utilisée dans les problèmes de classification. Lorsqu'elles sont appliquées à des problèmes de régression, elles sont appelées machines de régression vectorielle (Support Vector Regression ou SVR). Elles se distinguent par leur forte capacité de généralisation et leur capacité à traiter des problèmes non linéaires. Elles fonctionnent comme une régression linéaire multiple utilisant des prédicteurs transformés tout en conservant une faible complexité et un bon ajustement des données. Trois paramètres principaux dominent les performances de la technique et doivent être ajustés : la précision, le paramètre de coût, qui concerne le compromis entre précision et complexité, et γ , qui régule la fonction du noyau, utilisée pour transformer les prédicteurs en un espace de fonctions de plus grande dimension. Les SVM / SVR ont montré un grand potentiel dans plusieurs études [72,73].

- ***k- plus proches voisins (k-NN)***

C'est l'une des méthodes les plus simples d'apprentissage machine. Il repose sur un algorithme de reconnaissance de formes, qui compare l'état actuel à des échantillons d'apprentissage dans un espace de fonctions. Les distances euclidiennes sont ainsi calculées et les k premiers voisins les plus proches sont sélectionnés pour les prédictions.

- **Forêts Aléatoires (RF)**

Développées dans un premier temps par Breiman (2001), elles consistent en un ensemble d'arbres de décision / régression, dont les résultats montrent la prédiction moyenne des arbres individuels. Normalement, en se concentrant sur les arbres de régression, elles se caractérisent par le sur-apprentissage des données de formation, ce qui donne un biais faible mais une variance élevée. Il est courant de cultiver plusieurs arbres pour chaque étude de cas en faisant la moyenne des résultats de plusieurs arbres, il est possible de réduire la variance au prix d'une légère augmentation du biais. Néanmoins, la performance globale du modèle s'améliore. Les arbres de régression simples sont très sensibles au bruit de la base de données d'apprentissage. Le simple calcul de la moyenne des résultats de plusieurs arbres pour résoudre ce problème ne fonctionne pas, car s'ils étaient entraînés avec la même base de données, ils seraient très corrélés. Ce problème est résolu avec la méthode bagging, qui consiste à faire construire des arbres, chacun avec un échantillon initialisé (*bag*) de l'ensemble d'apprentissage. Les échantillons laissés en dehors du modèle sont utilisés pour évaluer les performances de chaque modèle.

Néanmoins, les arbres peuvent toujours être corrélés, dans la mesure où certaines variables sont des prédicteurs très puissants pour la sortie, elles seront sélectionnées dans la plupart des arbres en sac. Les forêts aléatoires traitent ce problème avec la mise en sac des entités, qui consiste à sélectionner un sous-ensemble aléatoire d'entités au niveau de chaque nœud et à réduire ainsi la corrélation.

La Figure.II.7 montre la répartition des études analysées concernant la technique utilisée. Comme on le voit, l'approche la plus courante parmi les articles analysés est l'utilisation de techniques statistiques, en particulier les RNA, qui représentent 24% des études.

II.6.3.6 Modèles hybrides

Certains modèles peuvent omettre certaines informations en raison de la façon dont chaque technique transforme les données. Ainsi, il est également courant de combiner des techniques pour renforcer leurs atouts afin d'améliorer la précision, ce qui est appelé modèle hybride, modèle mixte, modèle combiné ou modèle d'ensemble. Les modèles peuvent être mélangés de plusieurs manières, telles que bagging, boosting, voting ou stacking[76,77].

Deux approches peuvent être suivies, soit en combinant deux techniques statistiques ou plus (hybride statistique), soit en associant une technique statistique à un modèle de performance (hybride physique). Plusieurs travaux peuvent être trouvés en utilisant la

première approche. Bouzerdoum et al. [78] ont associé SARIMA à SVM dans leurs prévisions avec un horizon d'une heure, alors que Ramsami et Oree[79] ont utilisé ARIMA avec ANN, Vaz et al. [80] ont appliqué une technique d'ensemble de RNA et d'ARX non linéaire (NARX). Hossain et al. [81], ont mis au point une méthodologie pour sélectionner la meilleure combinaison de techniques de régression afin de construire un modèle d'ensemble, ils ont montré qu'ils ont des performances meilleures que les techniques individuelles.

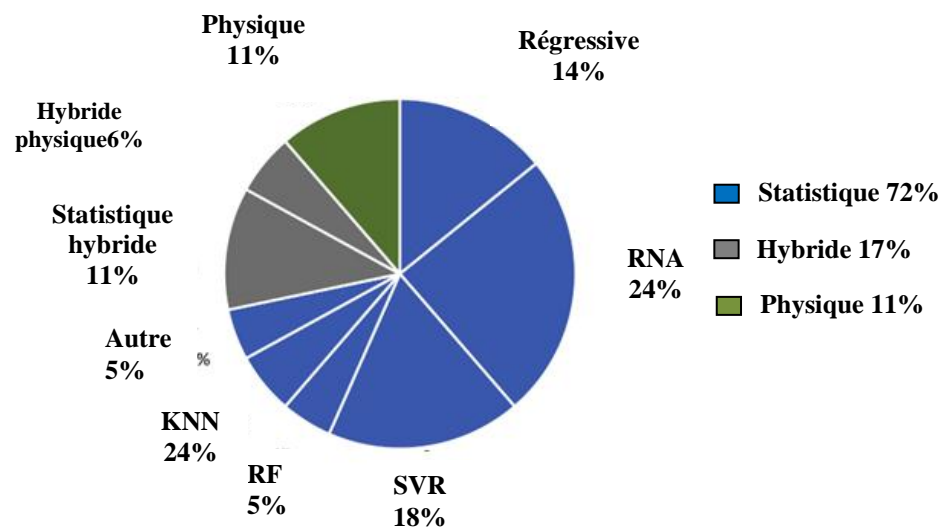


FIGURE II.7 Répartition des études par rapport à la technique utilisée [43].

Certaines études, cependant, ont construit un modèle boîte noire (black box) pour prédire mais ils ont inclus dans sa structure des expressions physiques. C'est le cas de la méthode du réseau de neurones artificiels hybride physique (PHANN), qui combinait un réseau RNA avec un modèle ciel clair. Ils ont également combiné certaines expressions physiques, telles que le calcul de l'indice de clarté et la température, avec les techniques NAR et RNA [82].

Dans [83], un modèle hybride composé d'un modèle ARMA et d'un réseau de neurones à retard temporel (Time Delay Neural Network TDNN) est examiné. Ce modèle est utilisé pour la prédiction du rayonnement solaire à court terme (une heure). Ce modèle hybride a le potentiel d'exploiter les avantages de ces deux techniques. Le modèle ARMA s'est avéré adapté au problème linéaire et le TDNN est efficace pour le problème non-linéaire. Dans le cas du rayonnement solaire contenant à la fois des caractéristiques linéaires et non linéaires, la précision de ce modèle hybride est assez satisfaisante. Cependant, ces différents modèles

fonctionnent très bien quand le temps est clair, mais lorsque les conditions météorologiques sont mauvaises, la précision des modèles de prévision diminue considérablement.

Cependant, la complexité informatique est accrue dans un modèle hybride en raison de l'utilisation de deux techniques ou plus. La performance d'un modèle hybride dépend de la performance d'un modèle individuel. La précision du modèle hybride est affectée par le choix d'une seule technique susceptible d'être peu performante, ce qui constitue une limitation du modèle hybride.

II.7 Méthode d'ensemble

La méthode d'ensemble est populaire en statistiques et en apprentissage machine, elle utilise plusieurs prédicteurs pour obtenir une décision agrégée qui sera meilleure que n'importe lequel des prédicteurs de base. Selon [84], il existe deux types de méthodes d'ensemble : compétitif et coopératif. Semblable à la classification, la prédiction d'ensemble peut être catégorisée en prédiction d'ensemble concurrentielle et coopérative. Un Ensemble compétitif de prédiction consiste à former différents prédicteurs individuels soit avec différents jeux de données ou bien avec le même ensemble de données, mais avec des paramètres différents, puis la prédiction est obtenue en faisant la moyenne (ou autre approche) de la décision de tous les prédicteurs individuels (prédicteurs de base). D'autre part, la prédiction d'ensemble coopérative consiste à diviser la tâche de prédiction en plusieurs sous-tâches et à sélectionner les prédicteurs appropriés pour chaque sous-tâche en fonction des caractéristiques des sous-tâches. La décision finale est la somme de tous les résultats des prédicteurs de base.

II.7.1 Ensemble de prédiction compétitif

L'approche de prédiction compétitive globale utilise plusieurs prédicteurs construits avec des conditions initiales légèrement différentes ou des paramètres différents, permettant de construire des modèles de prédiction individuels afin de former un modèle de prédiction d'ensemble (ensemble multi-modèle). Les résultats de prédiction de tous les modèles ou des modèles sélectionnés après l'élagage est agrégé par moyenne ou autres méthode de combinaison. Le niveau de confiance peut être mesuré par les variations de la dispersion des résultats individuels utilisés lors du processus de combinaison [85].

La diversité est un élément clé de la prédiction d'ensemble compétitive. Si les prédicteurs de base rendent des décisions similaires, il y aura moins d'amélioration. En prédiction concurrentielle (compétitive), si les sous-tâches sont similaires, les sorties des prédicteurs de base seront similaires et l'amélioration des performances d'ensemble sera

marginale. Pour la classification, la diversité peut être encore catégorisée comme diversité de données, diversité de paramètres et diversité de noyau [86,87]. Un schéma fonctionnel de la prédiction d'ensemble concurrentiel est montré dans la figure II.6. Il existe deux variantes basées sur la diversité des données et la diversité de paramètres.

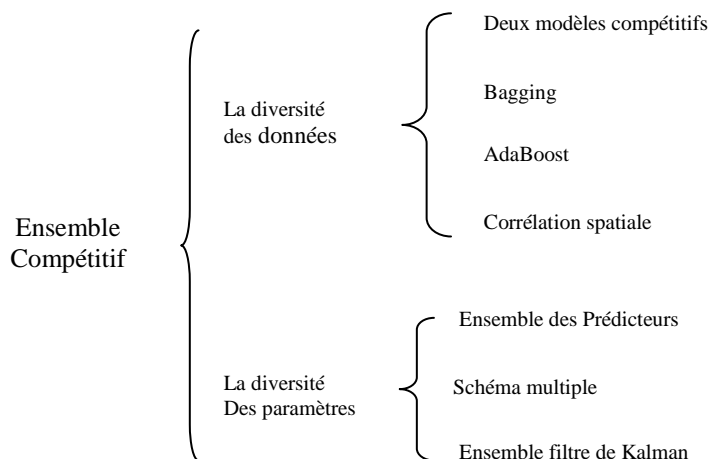


FIGURE II.8 Schéma fonctionnel des méthodes de prédiction d'ensemble concurrentielles.

II.7.1.1 Diversité des données

Pour la prédiction d'ensemble compétitive basée sur la diversité des données, plusieurs ensembles de données d'entrée sont introduits dans le système de prédiction. Il existe deux variantes, comme indiqué dans les équations. (II.7) et (II.8). Éq. (II.7) représente l'approche de prédiction qui applique N prédicteurs. $f_1(.) \dots f_N(.)$ pour N ensembles de données d'entrée $X_1 \dots X_N$ et la prédiction finale est une moyenne pondérée. Eq(II.8) représente une autre approche de prédiction qui utilise un seul prédicteur $f(.)$ pour N ensembles de données d'entrée.

$$\hat{y}(t+h) = \frac{1}{n} \sum_{i=1}^N w_i f_i(X_i(t)) \quad (\text{II.7})$$

$$\hat{y}(t+h) = f(X_1(t), \dots, X_N(t)) \quad (\text{II.8})$$

Où \hat{y} est la valeur prédite et h est l'horizon de prédiction.

Une étude sur la prédiction de l'irradiation solaire à Fontana, en Californie en 2009-2010 a été reportée [88]. Un modèle concurrentiel à deux méthodes de prédiction d'ensemble a été proposé : le premier modèle prédit l'irradiation solaire en utilisant une régression non linéaire des données météorologiques et le deuxième modèle prédit l'irradiation solaire basé sur la reconnaissance des formes. Les sorties de ces deux modèles ont été combinées pour

l'horizon de prédiction de 1h-3h. Les auteurs ont comparé leurs résultats avec les résultats précédemment rapportés et ont déclaré que la méthode de prédiction d'ensemble concurrentiel à deux modèles présente une erreur plus petite que les erreurs des méthodes précédemment rapportées, les auteurs ont également proposé d'utiliser une méthode d'apprentissage adaptatif pour mettre à jour les modèles sur une base journalière.

Le bagging est une méthode populaire de la théorie d'ensemble [89]. Elle comporte deux phases: 1) la phase de bootstrap, consiste à échantillonner la base de données d'origine avec remplacement pour obtenir N bases de données. 2) la phase d'agrégation pour combiner les sorties des N bases, les prédicteurs formés par chaque jeu de donnée sont mémorisés. Dans [90], un modèle RNA avec bagging a été introduit pour la prédiction à court terme de l'irradiation solaire.

Le modèle RNA avec bagging se compose de trois types différents des RNA: MLP, fonction de base radiale (RBFNN) et réseau de neurones récurrentes (RNN) avec des données historiques. Les données utilisées provenaient de l'Agence Météorologique Japonaise pendant la période 2007–2008. Les résultats ont montré que les mesures d'erreur de RNA avec bagging étaient plus petites par rapport aux modèles individuels.

Pour la prédiction de l'irradiation solaire, la corrélation spatiale est largement utilisée. Cette méthode est appelée totale satellite image (TSI), elle est habituellement utilisée dans les prédictions à court terme [91]. Les images du ciel sont prises par des satellites avec différentes coordonnées spatiales et temporelles. Dans [92], un prédicteur d'irradiation solaire par corrélation spatiale basé sur le RNA a été mis en œuvre avec 17 bases de données géographiques et météorologiques en Turquie de 2000 à 2002.

II.7.1.2 Diversité de paramètres

La diversité des paramètres par rapport aux variables... $\theta_1 \dots \theta_N$ d'un prédicteur pour produire N prédicteurs $f_1(.) \dots f_N(.)$ et chaque prédicteur apprendra du même base de donnée x. La valeur prédite \hat{y} est obtenue par la moyenne des performances de tous les prédicteurs comme il est montré en équation II.9.

$$\hat{y}(t + h) = \frac{1}{N} \sum_{i=1}^N f_i(X(t), \theta_i) \quad (\text{II.9})$$

Ou \hat{y} est la valeur prédite et h est l'horizon de prédiction.

II.7.2 Prédiction d'ensemble coopérative

La prédiction coopérative par ensembles divise une tâche de prédiction en plusieurs sous-tâches et résout chaque sous-tâche individuellement. Il existe deux variantes: l'une est en cours de prétraitement et l'autre en post-traitement.

II.7.2.1 Prétraitement

Le prétraitement consiste à diviser la base de données d'entrée en plusieurs sous-ensembles de données. Chaque sous-ensemble de données est modélisé et prédit par un prédicteur. Généralement, les prédicteurs sont identiques pour tous les sous-ensembles de données. La prédiction finale est une somme de toutes les sorties des prédicteurs.

Une méthode de décomposition en série chronologique appelée décomposition en ondelettes a été rapportée dans plusieurs publications [93,94]. La théorie des ondelettes consiste à étudier les séries temporelles dans le domaine fréquentiel ainsi que dans le domaine temporel. La décomposition en ondelettes consiste à décomposer la série temporelle en un ensemble de sous-séries basées sur une ondelette mère qui peut être prédite avec plus de précision. Il existe deux types de décompositions: continue et discrète. Pour des applications pratiques, la transformée en ondelettes discrète (DWT) est généralement utilisée pour la décomposition. Les équations clés de la prédiction basée sur la décomposition en ondelettes sont présentées dans l'équation II.10.

$$\begin{aligned} \{X_{D_i}(t), X_{A_n}(t)\} &= \text{DWT}(X(t)) \\ \hat{y}_{D_i}(t+h) &= f(X_{D_i}(t)) \\ \hat{y}_{A_j}(t+h) &= f(X_{A_j}(t)) \\ \hat{y}(t+h) &= \sum_{i=1}^n \hat{y}_{D_i}(t+h) + \hat{y}_{A_j}(t+h) \end{aligned} \quad (\text{II.10})$$

Où X est la base de données d'origine, X_{D_i} est la $i^{\text{ème}}$ composante détaillée, X_{A_j} est la $j^{\text{ème}}$ composante approximative, h est l'horizon de prédiction, \hat{y} est la valeur prédite et $f(\cdot)$ est le prédicteur. Dans [95], une ondelette Daubechies de nombre 6 a été utilisée pour la décomposition en ondelettes discrètes. Après décomposition, il y a trois décompositions détaillées ($X_{D_i}(t)$, $i = 1,2,3$) et une décomposition approximative $X_{A_3}(t)$. ARIMA a été choisi pour être le prédicteur pour chaque sous-série. Les résultats de prévision ont été agrégés pour obtenir le résultat final. Le modèle d'ondelettes-ARIMA présenté a été évalué dans le cas de la prévision de la vitesse du vent de 3,5 et 10 h, et la performance était meilleure que la méthode ARIMA conventionnelle.

Cao and Cao [96] ont présenté un réseau de propagation rétro-ondulante à répétition (RBPN) pour la prédiction de l'irradiation solaire. Le RBPN est un RNA dynamique qui permet de restituer tout ou partie des sorties. Les auteurs ont utilisé la méthode cut-and-trial pour déterminer le nombre de neurones cachés d'un RBPN à 3 couches.

II.7.2.2 Post-traitement

Une série chronologique peut avoir plus d'une caractéristique et chaque caractéristique est adaptée à une méthode particulière. Par exemple, le modèle ARIMA convient à la modélisation de séries chronologiques linéaires et Les RNA sont plus préférables pour la modélisation de séries chronologiques non linéaires. La prédiction de la série chronologique consécutivement par deux prédicteurs ou plus est considérée comme une prévision d'ensemble coopérative basée sur le post-traitement.

Plusieurs modèles coopératifs de prédiction d'ensemble basés sur le post-traitement ont été décrits dans la littérature, tels que ARIMA – GARCH [97], ARIMA – RNA, ARIMA–SVM [98] et SVR-SVC [99].

Dans [100], un modèle de réseau neuronal à retard temporel (TDNN) ARMA a été décrit. Les auteurs ont tout d'abord, utilisé plusieurs méthodes statistiques pour générer une série temporelle d'irradiation solaire stationnaire, puis ils ont utilisé ARMA pour prédire la partie linéaire de la série temporelle et le réseau TDNN pour prédire la partie non linéaire. Ensuite, ils ont combiné les deux résultats pour avoir la prédiction finale. Les auteurs ont comparé le modèle hybride ARMA–TDNN à un seul modèle ARMA et TDNN et les résultats ont montré une RMSE normalisée (NRMSE) plus petite dans le cas du modèle hybride.

II.8 Conclusion

Comme indiqué dans le présent chapitre, de nombreuses méthodes et types de méthodes sont disponibles. Il y a beaucoup de méthodes d'estimation du rayonnement solaire, certaines sont souvent utilisées (réseaux de neurones, processus autorégressifs,...), d'autres commencent à être utilisées (SVM, GP, LSTM...) et d'autres rarement (boosting, arbres de décision, forêts aléatoires, ...etc). En conclusion, on peut dire que les méthodes basées sur les RNA et ARIMA sont équivalentes en termes de qualité de prédiction dans certaines conditions de variabilité, mais la flexibilité des RNA en tant que modèles non linéaires universels les rend plus préférables que les processus autorégressifs classique.

En règle générale, la précision de ces méthodes dépend de la qualité des données d'apprentissage. En fait, en considérant les articles publiés, ces méthodes produisent des statistiques d'erreur très proches.

A partir de la littérature vue dans le présent chapitre, la méthodologie des ensembles a montré toujours une meilleure performance par rapport aux prédicteurs simples (uniques), ce qui montre l'intérêt à les utiliser dans cette thèse.

Chapitre III

Approche d'ensembles multi-modèles pour la prédiction de l'irradiation solaire

III.1 Introduction

La prédiction de l'irradiation solaire est devenue une étape clé pour prévoir correctement les performances des systèmes de conversion de l'énergie solaire et pour garantir la stabilité et la sécurité des réseaux électriques. Les compagnies d'électricité doivent contrôler les systèmes de production en se basant sur la connaissance préalable de la charge et de la capacité de production attendues au cours des prochaines minutes, heures et jours, car certains groupes électrogènes peuvent être démarrés rapidement, tandis que d'autres ont besoin de beaucoup de temps avant d'être prêts. De plus, le prix de la production d'électricité se négocie sur un marché en évolution constante et rapide. Cela explique pourquoi des prédictions précises de la production et de la consommation d'électricité sont nécessaires pour réduire les coûts de l'électricité [100].

La procédure normale de prédiction de la puissance électrique produite par un système de conversion solaire est basée sur le rayonnement solaire disponible, ainsi que sur plusieurs aspects concernant la performance du système et les problèmes environnementaux connexes au moment de l'intérêt. En utilisant des générateurs basés sur les technologies de conversion photovoltaïque (PV), l'électricité produite est, en principe au moins, linéairement liée à l'irradiation solaire incidente. Ainsi, à l'aide d'un outil de simulation d'énergie photovoltaïque approprié, la prévision de la production d'électricité est directement liée à celle de l'éclairement énergétique incident. Cette dernière est cependant très variable dans le temps pour des raisons météorologiques, ce qui rend difficile sa prévision exacte. D'un point de vue plus général, la prédiction de l'irradiation solaire avec une bonne précision devrait augmenter l'efficacité et la valeur de la production d'énergie de tout type de système d'énergie solaire.

La littérature que nous venons d'examiner dans le chapitre trois montre que de nombreuses méthodes stochastiques et de l'IA sont possibles et peuvent être combinées

efficacement. Bien que les méthodes individuelles soient largement utilisées, leur «force intrinsèque» n'est pas forcément la meilleure option possible. La nouvelle méthodologie présentée dans ce chapitre, tente d'améliorer les prédictions en sélectionnant de manière dynamique le prédicteur le plus approprié pour chaque série temporelle de GHI en utilisant plusieurs approches de combinaison et de classification [101].

On peut considérer qu'une méthode fiable serait une méthode qui diminue le risque de modèle (lorsque il est impliqué dans l'ensemble) et cherche à améliorer la précision des prédictions en tirant profit des différentes forces de divers modèles potentiellement efficaces, tout en compensant leurs limites. Le résultat de la nouvelle approche proposée dans cette étude est la prédiction produite par le modèle d'apprentissage distingué par la performance maximale dans la région de *l'espace des variables*. Ce dernier est également appelé «espace d'observations» car chaque valeur historique de la série GHI (dimension / coordonnée) correspond à une observation du GHI historique enregistré. Dans ce chapitre, divers types d'ensembles sont évalués à l'aide de plusieurs modèles statistiques et / ou modèles à base des techniques de l'IA [102].

III.2 Méthodologie proposée

Considérons un ensemble de N modèles de prédiction $f_i(X)$ ($i = 1, 2, \dots, N$) formés indépendamment sur les données d'apprentissage disponibles. Le problème schématisé sur la figure III.1, consiste à définir une approche de combinaison $\theta\{\}$ de telle sorte que les prédictions obtenues après la combinaison des différents modèles individuels pour un point de données particulier (c'est-à-dire non inclus dans l'ensemble d'apprentissage) puissent être décrites par (III.1):

$$\hat{F}(X) = \theta\{f_1(X), f_2(X), \dots, f_N(X)\} \quad (\text{III.1})$$

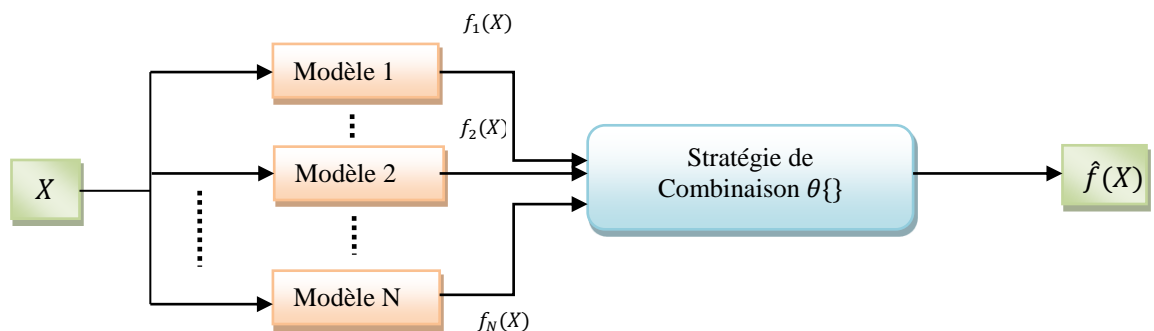


FIGURE III.1 Schéma fonctionnel du système de prédiction proposé.

III.2.1 Modélisation d'ensemble

Les méthodes d'ensemble sont des algorithmes d'apprentissage qui construisent un ensemble de modèles, puis classifient / prédisent les nouveaux points de données en prenant une combinaison pondérée de leurs prédictions [103,104].

La construction d'un ensemble de différents modèles de prédiction peut être faite à l'aide du même modèle de prédiction avec différents paramètres de réglage (tuning) ou selon différentes méthodes d'apprentissage.

Dans une autre perspective, les méthodes d'ensemble sont des méta-algorithmes qui combinent de nombreuses méthodes d'apprentissage automatique en un seul modèle dans le but de réduire la variance (bagging [89] ou le biais (boosting) [105] ou d'améliorer les prédictions (stacking). Les méthodes existantes peuvent être divisées en deux groupes:

- ***Méthodes d'ensemble séquentielles***

Dans lequel les apprenants de base sont créés consécutivement (par exemple, AdaBoost) [106]. L'objectif ici est d'exploiter la dépendance entre les différents apprenants de base. Les performances générales peuvent être améliorées en analysant les mauvais exemples (mauvaises sorties) avec des poids de grandes valeurs, comme dans les méthodes de renforcement classiques.

- ***Méthodes d'ensemble parallèles***

Les apprenants de base sont créés en parallèle (par exemple, Random Forest) [107]. Le but ici est d'exploiter l'indépendance potentielle entre tous les apprenants de base, car l'erreur peut être fortement réduite en faisant la moyenne.

Habituellement, un seul algorithme d'apprentissage de base peut être utilisé pour produire des modèles de base homogènes à partir de modèles du même type, conduisant à des ensembles homogènes.

Alternativement, des modèles hétérogènes, c'est-à-dire des modèles différents qui conduisent à des ensembles hétérogènes, dont l'objectif est d'avoir un modèle global plus précis que chacun de ses membres individuels, les modèles individuels de l'ensemble doivent être aussi précis que possible et aussi diversifiés que possible [104]. Une question centrale dans la théorie des ensembles est que, pour améliorer le processus de combinaison, les modèles individuels doivent être différents. Sinon, les performances ne seraient pas améliorées si les modèles simples dupliqués étaient combinés.

Dans cette thèse, nous ne considérons que la seconde approche, dans laquelle la modélisation *hétérogène* est utilisée. Cette considération est confortée par l'intérêt porté à

l'examen de l'efficacité de l'approche pour la prédiction de GHI à court terme (de 1 minute à 1 heure à l'avance). A des taux de rafraîchissement aussi élevés, les variables exogènes qui pourraient être utilisées pour prédire le GHI avec un certain degré de fiabilité (par exemple, fraction nuageuse, précipitation ou profondeur optique d'aérosol) ne sont généralement pas disponibles à partir d'observations locales ou de données modélisées et ré-analysées (dont la résolution spatiale est quand même trop large).

Comme mentionné ci-dessus, dans ce qui suit, plusieurs approches de combinaison existent et sont conçues selon deux stratégies différentes. Dans la première, appelée «combinaison», la prédiction finale est obtenue après la combinaison des prédictions obtenues par les différents modèles uniques. L'autre stratégie, dite «sélection», consiste en une nouvelle méthode (détaillée ci-après), qui obtient la prédiction finale en sélectionnant le résultat (prédiction) du «meilleur modèle individuel» à l'aide d'une méthodologie qui définit une partition appropriée de l'espace des variables. Dans la présente étude, trois familles de modèles sont utilisées pour construire l'ensemble optimal, à savoir : (i) les modèles autorégressifs ; ii) modèles à base de Réseaux de Neurones Artificiels (RNA); et (iii) les modèles à noyaux.

III.2.2 Méthodes individuelles

III.2.2.1 Modèles autorégressifs

Les modèles ARIMA [108] constituent l'origine de la plupart des techniques de prédiction statistiques. Un modèle ARIMA (p, d, q), généralement appelé modèle de Box-Jenkins, suppose que toute valeur future d'une série temporelle peut être exprimée sous forme de fonction linéaire de certaines observations passées, en plus de certains termes de bruit blanc. Il utilise une procédure de différenciation courante d'ordre d pour rendre la série temporelle stationnaire. Il se décompose en différents modèles, comme suit:

- ***Processus autorégressif AR***

Un processus stochastique Y_t avec des valeurs réelles est considéré comme un processus AR(p) d'ordre p si:

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} + \dots + a_p y_{t-p} + \varepsilon_t. \quad (\text{III.2})$$

La valeur AR (p) à l'instant t est ensuite estimée à l'aide des p valeurs passées et d'un terme aléatoire ε_t .

- **Moyen mobile MA**

Un processus stochastique Y_t est dit processus MA d'ordre q , noté MA (q), s'il existe q coefficients b_1, b_2, \dots, b_q , et un bruit blanc ε_t tel que :

$$y_t = b_0\varepsilon_t + b_1\varepsilon_{t-1} + \dots + b_q\varepsilon_{t-q}. \quad (\text{III.3})$$

La valeur du processus MA (q) au temps t est donc régressée sur ses propres erreurs passées.

- **Autorégressif à moyenne mobile, ARMA**

Les processus AR (p) sont des cas particuliers d'ARMA. Un processus stochastique Y_t est considéré comme un processus ARMA d'ordre p, q si:

$$y_t = \varepsilon_t + (a_1y_{t-1} + a_2y_{t-2} + \dots + a_p y_{t-p}) + (b_0\varepsilon_t + b_1\varepsilon_{t-1} + \dots + b_q\varepsilon_{t-q}) \quad (\text{III.4})$$

Ceci peut aussi être formulé comme $\phi(z)Y_t = \theta(z)\varepsilon_t$, ou $\phi(z) = 1 + a_1z + a_2z^2 + \dots + a_pz^p$ et $\theta(z) = 1 + b_1z + b_2z^2 + \dots + b_qz^q$ sont les termes polynomiaux des composantes AR et MA du processus ARMA(p, q). y_t , et z est l'opérateur de rétrogradation.

- **Autorégressif à moyenne mobile intégrée, ARIMA**

ARIMA est une extension de la classe de modèles ARMA dans laquelle une opération de différenciation est ajoutée. Un processus y_t est considéré comme un ARIMA(p, d, q) si $(1 - z)^d Y_t$ est une ARMA (p, q) causal. Le modèle ARIMA peut être formalisé comme suit:

$$\phi(z)(1 - z)^d Y_t = \theta(z)\varepsilon_t. \quad (\text{II.5})$$

- **Autorégressif à moyenne mobile saisonnière intégrée, SARIMA**

Le modèle SARIMA peut être décrit comme suit:

$$\phi_p(z)\Phi_P(z^S)(1 - z)^d(1 - z^S)^D Y_t = \theta_q(z)\Theta_Q(z^S)\varepsilon_t \quad (\text{III.6})$$

Où $\phi_p(z), \Phi_P(z^S), \theta_q(z)$ et $\Theta_Q(z^S)$ sont les termes polynomiaux d'ordres p, P, q et Q , respectivement, d et D sont les ordres de la différenciation non saisonnière et saisonnière, respectivement.

III.2.2.2 Modèles à base de réseaux de neurones artificiels (RNA)

Les RNA ont été créés pour surmonter les limitations rencontrées avec des méthodes plus traditionnelles lorsqu'on essaie de résoudre des problèmes complexes. Les RNA apprennent de certains exemples en construisant une relation d'entrées-sorties permettant d'effectuer des prédictions. Les RNA contiennent une interconnexion de plusieurs neurones.

Dans la littérature, plusieurs architectures de réseaux avec différents algorithmes d'apprentissage ont été proposées [109]. Un algorithme d'apprentissage est essentiellement une procédure dont la fonction est de réguler les poids d'un réseau donné. Pour un ensemble de données d'apprentissage donné, cette opération peut être réalisée en minimisant une fonction objective entre la fonction cible souhaitée et estimée. La présente étude se concentre sur quatre types de RNA, à savoir: MLP (Perceptron Multicouches), RBF (Réseaux à Fonction de Base Radial), ELM (Extreme Learning Machines) et Réseau de neurones Elman (ENN).

- **Réseaux de neurones de perceptron multicouches PMC**

Un PMC est composé de plusieurs unités interconnectées (neurones) structurées en couches et fonctionnant en parallèle avec un flux d'informations en retour, comme illustré dans la figure III.2. Les signaux circulent successivement dans les différentes couches, de la couche d'entrée à la couche de sortie. Les couches entre les couches d'entrée et de sortie sont appelées couches cachées. Dans chaque couche, chaque unité calcule un produit scalaire entre le vecteur des poids (pondération) et le vecteur de sortie issu de la couche précédente. Ensuite, une fonction de transfert est appliquée au résultat précédent pour créer une entrée pour la couche suivante [111]. Généralement, la fonction de transfert pour les couches cachées est la sigmoïde:

$$f(x) = \frac{1}{1+e^{-x}} \quad (\text{III.6})$$

- **Réseaux de neurones à fonctions radiales, RBF**

Un réseau neuronal RBF est composé de trois couches (une couche entrée, une couche cachée et une couche de sortie). Comme illustré sur la figure III.3, les neurones d'entrée (c'est-à-dire les variables d'entrée) transmettent uniquement les variables d'entrée X_j . Chaque neurone de la couche cachée est caractérisé par une fonction noyau φ_j (généralement une fonction gaussienne) décrite par une largeur σ_j et un centre c_j , tels que :

$$\varphi_j(\|y - c_j\|) = \exp\left(-\frac{1}{2}\left(\frac{\|y - c_j\|}{\sigma_j}\right)^2\right). \quad (\text{III.7})$$

La couche de sortie contient un seul neurone fournissant la valeur prédite. La fonction de sortie peut être écrite comme suit:

$$f(y) = \sum_{j=1}^m \lambda_j \varphi_j(\|y - c_j\|) \quad (\text{III.8})$$

où λ_j et m sont respectivement le poids et le nombre de fonctions RBF [104,108].

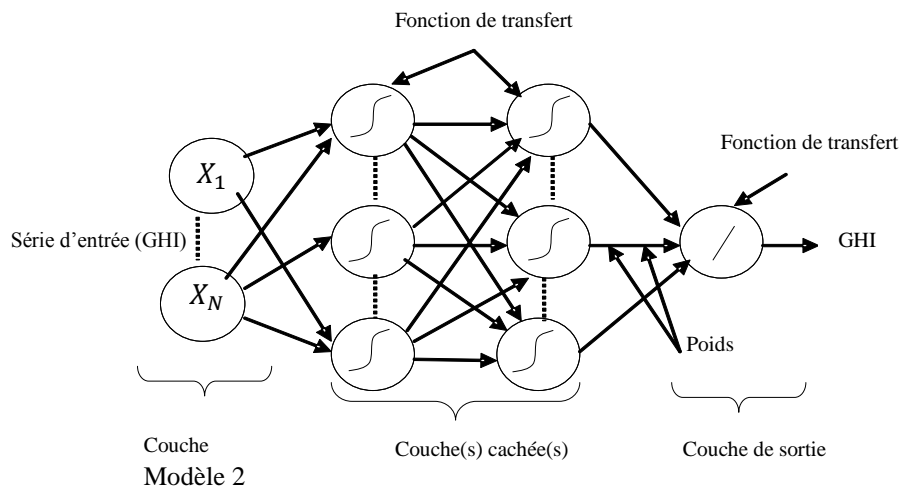


FIGURE III.2 Architecture d'un réseau de neurones de type perceptron multicouches.

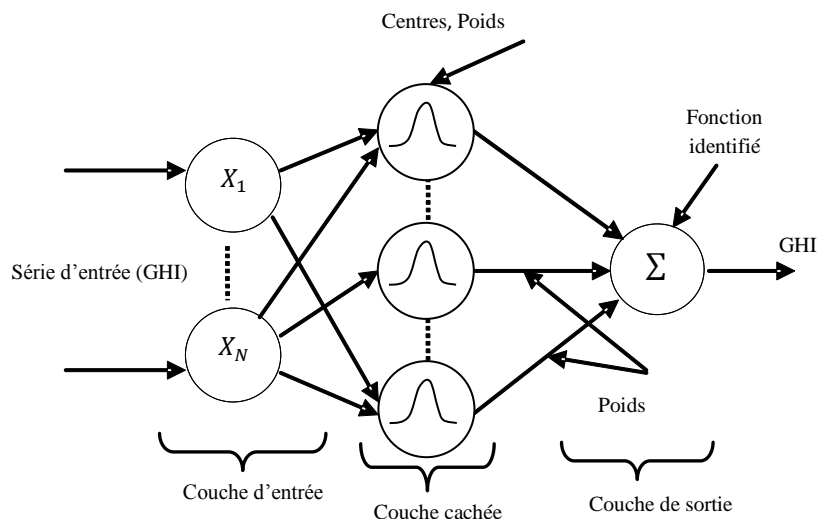


FIGURE III.3 Architecture d'un réseau de neurones à fonction radiale

- **Réseaux de Neurones Elman ENN**

Un ENN est un type de réseau de neurones dynamique, basé essentiellement sur la structure de réseaux de neurones artificiels classiques. Il stocke un état interne pour élaborer la fonction de ses caractéristiques dynamiques et fournit au système la possibilité de s'adapter à toute caractéristique variable dans le temps. Un ENN peut donc refléter directement les caractéristiques dynamiques d'un système, ce qui est une caractéristique importante d'un outil de prédiction avancé. La structure ENN est illustrée sur la figure III.4.

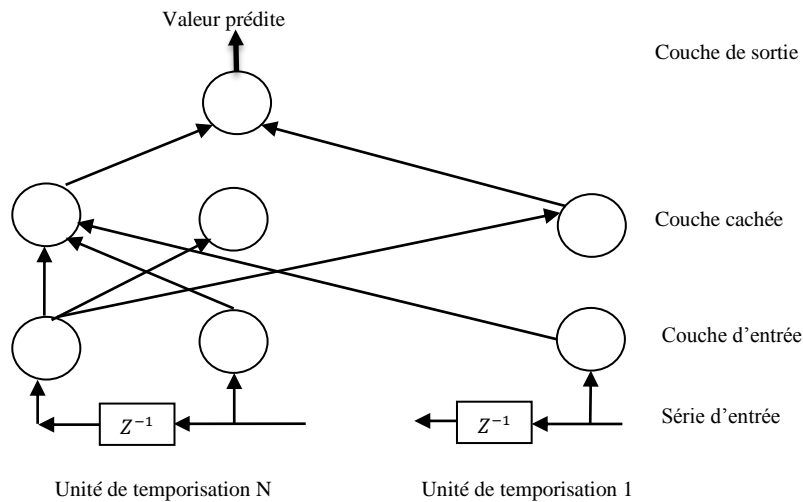


FIGURE III.4 Architecture du réseau Elman.

Généralement, les ENN sont divisés en quatre couches: la couche d'entrée, la couche cachée, couche de réalisation (*undertake*) et la couche de sortie. Les nœuds de la couche d'entrée agissent simplement comme des unités de transmission de signal. La couche de sortie réagit comme une fonction de sommation, tandis que la couche de réalisation (ou couche d'état) est utilisée pour mémoriser les valeurs de sortie précédentes des unités de la couche cachée qui devront être renvoyées à l'entrée du réseau.

La capacité de stockage de l'ENN le rend plus sensible aux données historiques, ce qui permet au retour d'information interne d'améliorer l'aptitude du réseau à gérer les variations dynamiques des données. De plus, un ENN peut approximer n'importe quelle fonction non linéaire avec un degré de précision très acceptable et peut éviter l'impact du bruit externe. Les équations mathématiques dans l'espace d'état non linéaire d'un ENN peuvent s'écrire comme suit :

$$y(x) = g(\omega^3 x(x)) \quad (\text{III.9})$$

$$x(k) = f(\omega^1 x_c(k) + \omega^2 (u(k-1))) \quad (\text{III.10})$$

$$x_c(k) = x(k-1) \quad (\text{III.11})$$

où

y est le vecteur de sortie de dimension égale à M ;

x est le vecteur de nœud de la couche cachée avec M dimensions;

u est le vecteur d'entrée avec r dimensions;

x est le vecteur d'état de retour avec N dimensions;

ω^3 sont les poids reliant la couche cachée à la couche de sortie;

g est la fonction de transfert du noeud de sortie;

f est la fonction de transfert des nœuds des couche cachées.

Généralement, les ENNs utilisent également l'algorithme de rétro-propagation pour déterminer les poids du réseau.

- **Extreme Learning Machines, ELM**

Le modèle ELM a été proposé par Huang et al. [110]. Ce type de réseau neuronal à une seule couche cachée peut être utilisé pour les problèmes de régression et de classification. Supposons un ELM avec Q neurones cachés et une fonction d'activation $g(x)$ qui est conçu pour apprendre N différents échantillons (X_i, t_i) , ou $X_i = [x_{i1}, x_{i2}, \dots, x_{in}]^T$ et $t_i = [t_{i1}, t_{i2}, \dots, t_{im}]^T$. Les poids d'entrée et les biais cachés sont choisis aléatoirement. Par conséquent, le système non linéaire peut être converti en un système linéaire comme suit:

$$H\beta = T \quad (\text{III.12})$$

Où $H = \{h_{ij}\}$ ($i = 1, 2$) est la matrice de sortie de la couche cachée unique, $h_{ij} = g(W_j * Y_i + b_j)$ désigne la sortie du j_{th} neurone caché par rapport à X_i ; $W_j = [w_{j1}, w_{j2}, \dots, w_{jn}]^T$. et b_j est le biais du neurone caché j_{th} ; $W_j * Y_i$ est le produit intérieur de W_j et X_i ; $\beta = [\beta_1, \beta_2, \dots, \beta_K]^T$ est la matrice des poids de sortie et $\beta_j = [\beta_{j1}, \beta_{j2}, \dots, \beta_{jm}]^T$ ($j=1, \dots, K$) indique le vecteur de poids reliant le neurone caché j_{th} au neurone de sortie. Enfin, $T = [t_1, t_2, \dots, t_N]^T$ est la matrice de sortie (cibles) (c'est-à-dire la sortie souhaitée de GHI). Par conséquent, les poids de sorties sont déterminés en recherchant la solution des moindres carrés du système linéaire décrit par l'équation (III.13), tel que:

$$\hat{\beta} = H^+T \quad (\text{III.13})$$

où H^+ est l'inverse généralisé de Moore-Penrose de la matrice H . Habituellement, les ELM ont tendance à donner de bonnes performances de généralisation avec une vitesse d'apprentissage extrêmement rapide d'où vient leur nom [111,112].

III.2.2.3 Les modèle à noyau

- **Machine à vecteurs de support, SVM**

Les SVM sont des techniques non paramétriques supervisées qui ne nécessitent donc aucune hypothèse sur la distribution des données. Le but d'une approche de régression SVM est de définir une fonction *lisse* ayant (au maximum) ε déviation par rapport aux échantillons d'apprentissage, qui sont dans le cas présent issus d'une série temporelle de rayonnement

solaire [113,114]. Ceci peut être réalisé en projetant la série originale d'entrée de dimension d dans un espace variable transformé, c'est-à-dire $\Gamma(x) \in \mathbb{R}^{d'}$ ($d' > d$) et en fixant une estimation linéaire comme suit:

$$f(x) = \omega^* \cdot \Gamma(x) + b^* \quad (\text{III.14})$$

La fonction linéaire dans l'espace variable de dimension supérieure est obtenue en optimisant une fonction de coût, dont l'objectif est de minimiser deux critères (norme euclidienne et minimisation d'erreur), elle est exprimée comme suit:

$$\Psi(\omega, \xi) = \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^N (\xi_i + \xi_i^*). \quad (\text{III.15})$$

Où C est un paramètre de régularisation. Cette fonction est soumise aux contraintes ci-dessous:

$$\begin{cases} y_i - (\omega \cdot \Gamma(x_i) + b) \leq \varepsilon + \xi_i \\ (\omega \cdot \Gamma(x_i) + b) - y_i \leq \varepsilon + \xi_i^* \end{cases}, \quad \xi_i, \xi_i^* \geq 0 \text{ et } i \in [1, N] \quad (\text{III.16})$$

où ξ_i et ξ_i^* sont des variables *slack* liées à des échantillons qui ne correspondent pas au tube de déviation ε . La formulation de la fonction d'erreur est conçue pour traiter une fonction de perte ε -insensible $|\gamma|_\varepsilon$ définie comme suit :

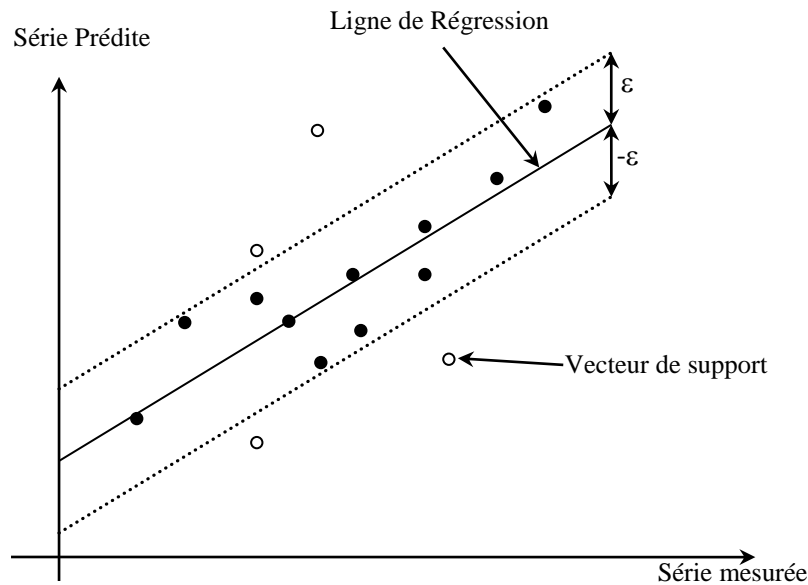
$$\begin{cases} 0, \text{ si } |\tau| \leq \varepsilon \\ |\tau| - \varepsilon, \text{ autrement} \end{cases} \quad (\text{III.17})$$

où τ est représenté l'écart par rapport à la sortie souhaitée. Cela revient à dire que les différences entre les valeurs mesurées et les valeurs prédites sont acceptées à l'intérieur du tube (erreur inférieure à ε), alors qu'une pénalité linéaire est imposée aux points situés à l'extérieur du tube insensible. Un exemple de SVM linéaire est présenté dans la figure III.5.

Le résultat du problème d'optimisation est une fonction représentée dans l'espace variable (original) d'entrée comme suit :

$$f(x) = \sum_{i \in S} (\alpha_i - \alpha_i^*) K(x_i, x) + b^* \quad (\text{III.18})$$

Où $K(\cdot)$ est une fonction noyau S est le sous-ensemble d'indices $i = (1, 2, \dots, N)$ correspondant au multiplicateurs de Lagrange non nuls α_i ou α_i^* . En fonction de leur importance dans la détermination d'une solution, les multiplicateurs de Lagrange évaluent chaque point d'apprentissage. Les points d'apprentissage avec des poids non nuls sont appelés des vecteurs de support.

FIGURE III.5 Régression SVM linéaire avec ϵ -tube.

III.2.3 Méthodes de combinaison

Une stratégie de combinaison tente d'améliorer les performances de prédiction globales en combinant différentes approches. Elle tente d'augmenter la précision des prédictions par le biais d'une solution pratique, mais aux dépens d'une complexité de calcul supplémentaire. Par conséquent, il convient de bien comprendre les propriétés de tous les modèles d'ensemble et de faire preuve de prudence dans l'utilisation des informations accessibles dans le but de développer des méthodes de combinaison efficaces [115]. Après avoir généré un ensemble d'apprenants de base, au lieu d'essayer de déterminer le meilleur modèle, les méthodes d'ensemble sélectionnent une combinaison pour obtenir la meilleure capacité de généralisation, ou la méthode de combinaison joue un rôle important.

Dietterich [116] a attribué le bénéfice obtenu avec le processus de combinaison aux trois facteurs fondamentaux suivants:

- **Statistique:** habituellement, l'espace des hypothèses est trop grand pour permettre la découverte de petits ensembles de données de formation, où d'autres hypothèses différentes peuvent donner la même précision sur les données de formation. Si le modèle d'apprentissage sélectionne un espace parmi ces hypothèses, il existe un risque qu'une hypothèse mal choisie ne prédit pas correctement les données futures. Par conséquent, en combinant des hypothèses, le risque de choisir une hypothèse incorrecte peut être réduit.

- **Coût de calcul:** plusieurs modèles d'apprentissage permettent de créer différents types de recherche locale pouvant être bloqués dans les optima locaux. Bien que les données de formation soient suffisantes, il reste très difficile de déterminer la meilleure hypothèse. La combinaison de recherches locales lancées à partir de plusieurs points de départ distincts peut fournir une meilleure estimation de la véritable hypothèse inconnue. Dans ce cas, la combinaison d'hypothèses réduit le risque de choisir un mauvais minimum local.

- **Représentationnel:** dans plusieurs problèmes d'apprentissage automatique, la vraie hypothèse inconnue ne peut pas être caractérisée. Dans ce cas, l'utilisation d'une combinaison d'hypothèses permet une expansion de l'espace représentable des fonctions. Par conséquent, le modèle d'apprentissage peut constituer une estimation plus précise de la véritable hypothèse inconnue.

Les trois classes de facteurs que nous venons de décrire, expliquent généralement pourquoi les techniques d'apprentissage classiques ont tendance à échouer [104]. Dans ce qui suit, on fournit un bref résumé des techniques utilisées dans cette étude pour la combinaison linéaire de plusieurs prédictions.

III. 2.3.1 Moyen simple, SA

La méthode de combinaison la plus simple est la moyenne simple, où toutes les prédictions individuelles ont le même poids ($w_j = 1/n$). Plusieurs études ont révélé que curieusement, la moyenne simple améliore souvent la précision globale des prédictions, même lorsque certaines prédictions individuelles apparaissent comme des valeurs aberrantes (outliers) ou des valeurs extrêmes [117, 118].

III. 2.3.2 Méthode basée sur l'erreur, EB

Dans cette méthode, la série chronologique est divisée en deux sous-ensembles complémentaires, à savoir l'ensemble d'apprentissage et celui de validation. Les modèles individuels sont formés sur les données de formation et leurs erreurs de prédiction résultantes (en utilisant la donnée de validation) sont enregistrées. Le poids de la combinaison pour chaque prédiction individuelle est ensuite calculé comme étant inversement proportionnel à l'erreur de prédiction du modèle correspondant, c'est-à-dire

$$W_j = \frac{e_j^{-1}}{\sum_{j=1}^n e_j^{-1}} \quad (\text{III.19})$$

III.2.3.3 Régression par moindres carrés, LSR

Dans cette méthode, les poids d'une combinaison linéaire sont attribués en minimisant la somme des erreurs carrées (SSE) entre les données prédites et les données mesurées. La formulation classique de la combinaison linéaire peut être exprimée comme suit:

$$\hat{Y} = UW \quad (\text{III.20})$$

Où

$$U = \left[\hat{y}_1^{(1)} \hat{y}_1^{(2)} \dots \hat{y}_1^{(n)}; \hat{y}_2^{(1)} \hat{y}_2^{(2)} \dots \hat{y}_2^{(n)}; \dots; \hat{y}_M^{(1)} \hat{y}_M^{(2)} \dots \hat{y}_M^{(n)} \right]^T$$

$$W = [w_1, w_2, \dots, w_n]^T \quad (\text{III.21})$$

Et où \hat{Y} et Y sont les séries prédites et mesurées, respectivement, et W est lié avec les poids des modèles individuels. Les SSE de prédictions peuvent alors être exprimées comme suit:

$$\begin{aligned} \text{SSE} &= \sum_{t=1}^M (y_t - \hat{y}_t)^2 \quad (\text{III.22}) \\ &= (Y - UW)^T (Y - UW) \\ &= Y^T Y - 2W^T U^T Y + W^T U^T U W. \end{aligned}$$

En minimisant les SSE par rapport à W , le vecteur de pondération est calculé comme suit:
 $W = (U^T U)^{-1} U^T Y$.

III.2.3.4 Méthode de pondération différentielle, DW

Similairement à la méthode LSR, les poids d'une combinaison linéaire de prédictions sont obtenus en minimisant la variance de l'erreur de prédiction combinée [103]. En fait, cette approche nécessite la connaissance de la matrice de covariance des erreurs de prédiction, inconnue dans la pratique. Les poids sont calculés avec :

$$W_i = \frac{\sum_{s=t-v}^{t-1} (e_s^{(i)})^2}{\sum_{j=1}^n \sum_{s=t-v}^{t-1} (e_s^{(j)})^2} \quad (\text{III.23})$$

Où n est le nombre de modèles individuels, t est la période de prédiction, W_i est la pondération attribuée à la deuxième méthode en fonction de la période précédente des données $(t-1)$, v est une constante et $e_t^{(i)}$ est le pourcentage d'erreur de prédiction à l'instant t , défini comme suit:

$$e_t^{(i)} = \frac{y_t - \hat{y}_t^{(i)}}{y_t} \quad \forall i = 1, 2, \dots, n. \quad (\text{III.24})$$

III.2.3.5 la méthode Outperformance, OP

Proposé par Lemke et Gabrys [119], OP adopte un cadre bayésien de probabilités subjectives d'attribution des poids appropriés à l'ensemble des prévisions individuelles. Le poids d'un modèle de prévision individuel est déterminé à partir du nombre de fois où il a dépassé les autres modèles. Pour simplifier, la combinaison de deux modèles individuels F_1 et F_2 est réalisée comme suit

$$\hat{y}_k = w\hat{y}_k^{(1)} + (1 - w)\hat{y}_k^{(2)}, \forall k = 1, 2, \dots, N. \quad (\text{III.25})$$

Dans le cas présent, on suppose que les performances des deux modèles sont évaluées à l'aide d'une erreur absolue et enregistrées M fois. Avec ce scénario, il est raisonnable d'accepter que le modèle avec l'erreur absolue la plus petite surpasse les autres modèles. Si le nombre de fois F_1 bat F_2 est une fraction k de M , on peut supposer que k suit une distribution bêta $B(k|a_1, a_2)$ ($0 \leq k \leq 1; a_1, a_2 > 0$) [119]. Pour chaque réalisation de prédiction i , la variable de Bernoulli δ_i est exprimée:

$$\delta_i = \begin{cases} 1, & \text{si } F_1 \text{ bat } F_2 \\ 0, & \text{sinon} \end{cases} \quad (\text{III.26})$$

La structure générale peut être étendue aux modèles, en supposant que les fractions de surperformance k_i ($i = 1, 2, \dots, n$) des modèles individuels suivent la distribution de Dirichlet à n paramètres, ce qui est équivalent à la distribution bêta.

III.2.4 Stratégie de classification

Cette stratégie est motivée par l'idée que divers prédicteurs à hautes performances peuvent être connectés à différentes zones de l'espace des variables en entrée de la série temporelle. Le meilleur modèle individuel est celui qui minimise l'erreur dans différentes parties de l'espace des variables. L'approche consiste ici, à créer une partition de l'espace variable, où chaque point de l'espace variable (série temporelle unique) est lié au prédicteur de l'ensemble qui obtient l'erreur de prédiction la plus petite. Comme illustré sur la figure III.6, l'étape d'apprentissage comprend l'identification (parmi l'ensemble des modèles de prédiction individuels existants) du meilleur prédicteur individuel $\hat{P}(X)$ pour chaque série S de l'espace variable. La division optimale de l'espace variable en un ensemble de régions est acquise en analysant les données d'apprentissage S_i ($i = 1, 2, \dots, M$), chacune étant attribuée à un seul prédicteur. Par conséquent, l'objectif de l'élément de sélection est de rechercher le meilleur moyen possible d'obtenir la partition la plus appropriée de l'espace variable, de telle sorte que :

$$\hat{P}(X) = \operatorname{argmin}_{j=1,2,\dots,N} \{|f_j(S_i) - y_i|\} \quad (\text{III.27})$$

Néanmoins, cela ne modifierait pas la partition obtenue, car le résultat de l'opérateur «argmin» n'est pas influencé par des fonctions monotones dans la phase de prévision. Chaque série inconnue $S \in \mathfrak{R}^d$ est introduite en tant qu'entrée dans le modèle de prévision. Sa sortie est alors la prédiction $\hat{P}(X) \in \{1, 2, \dots, N\}$ du prédicteur le plus précis pour la série considérée. La valeur prédite $\hat{F}(X)$ fournie par divers prédicteurs peut être exprimée comme suit:

$$F(X) = f_{\hat{P}(X)}(X). \quad (\text{III.28})$$

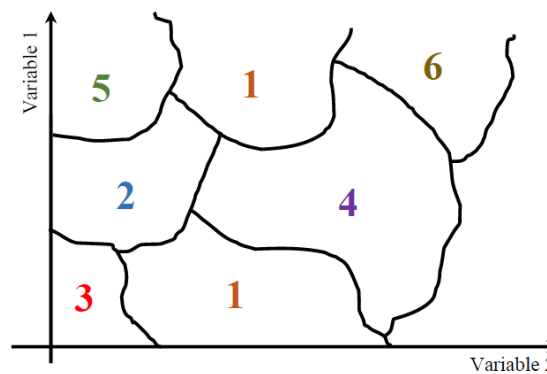


FIGURE III.6 Partition schématique d'un espace de variables bidimensionnel pour un exemple composé de six modèles. Chaque région désigne l'algorithme de prévision individuel qui fournit la précision de prédiction optimale dans la zone correspondante de l'espace variable de la série temporelle.

D'un point de vue opérationnel, l'approche de classification peut être utilisée en fonction de l'utilisation des informations contenues dans les enregistrements GHI accessibles. Cela implique deux parties: l'apprentissage et le test. Comme l'illustre la figure III.7, la phase d'apprentissage consiste à déterminer le modèle le plus probable pour chaque point de l'espace des variables. La meilleure partition de l'espace variable (constitué d'un ensemble de zones) est obtenue par l'analyse de la série de formations, chacune étant attribuée à un seul modèle de prédiction. Le but de l'approche de sélection est de construire la partition optimale de la meilleure façon possible. Ici, l'optimalité peut être formulée en termes d'erreur de prévision. Dans le cas présent, le critère de l'erreur minimale absolue (MAE) est sélectionné pour cette tâche. Dans la phase de test, chaque série (pour laquelle la valeur future de GHI est inconnue) est fournie en entrée au classifieur, qui fournit la valeur prévue du modèle le plus précis possible pour cette série [120 ,121].

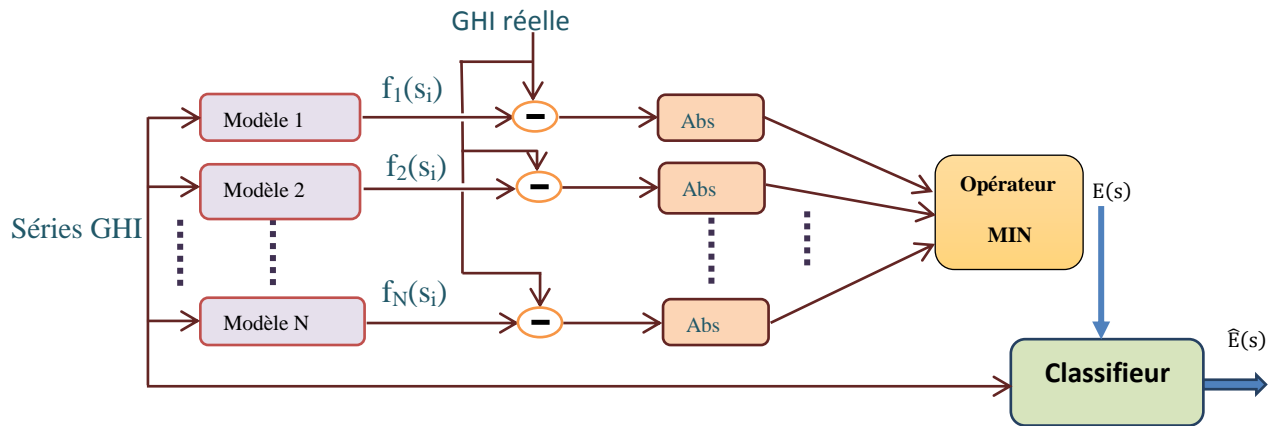


FIGURE III.7 Schéma fonctionnel du processus d'apprentissage utilisant l'approche basée sur la classification.

III.3 Sources de données

Une base de données de haute qualité de mesures d'irradiation solaire en 1 minute provenant de quatre stations radiométriques qui observent les composants d'irradiation (y compris GHI) avec des radiomètres à thermopile, a été examinée [122]. Cette dernière référence fournit des informations détaillées sur ces stations et sur le processus de *contrôle de qualité* de leurs données. Par conséquent, le tableau III.2 donne une brève description. Les quatre stations illustrent respectivement des types de climat radiatif distincts: Aride, Tempéré, Tropical et Albédo élevé qui favorisent différents régimes de nuages et différents schémas saisonniers, comme le montre la figure III.8. Pour ce premier exemple d'application selon cette méthode, les données initiales sur une minute ont été compilées en données horaires pour prédire la prochaine période horaire, ce qui est un horizon commun à court terme dans la pratique. Les développements futurs examineront d'autres applications, telles que l'horizon de 15 minutes, qui est également très important dans la répartition de la distribution d'électricité et les négociations des prix.

TABLEAU III.1. Informations sur les quatre stations de test utilisées dans cette étude. Les valeurs GHI moyennes et maximales mesurées sont données en W/m^2 .

Site	Source	Altitude [m]	Longitude [°]	Latitude [°]	Max_GHI [W/m^2]	Moy_GHI [W/m^2]	Période	Climat
Alice Springs	BSRN	547	133.88	-23.79	1260.7	543.2	2009–2010	Arid
Brasilia	BSRN	1023	-47.71	-15.60	1335.1	475.6	2006–2007	Tropical
Golden	NREL	1829	-05.18	39.74	1209.3	459.4	2007–2008	Tempéré
Ny-Ålesund	BSRN	11	11.93	78.92	899.0	184.3	2010–2011	Albédo élevé

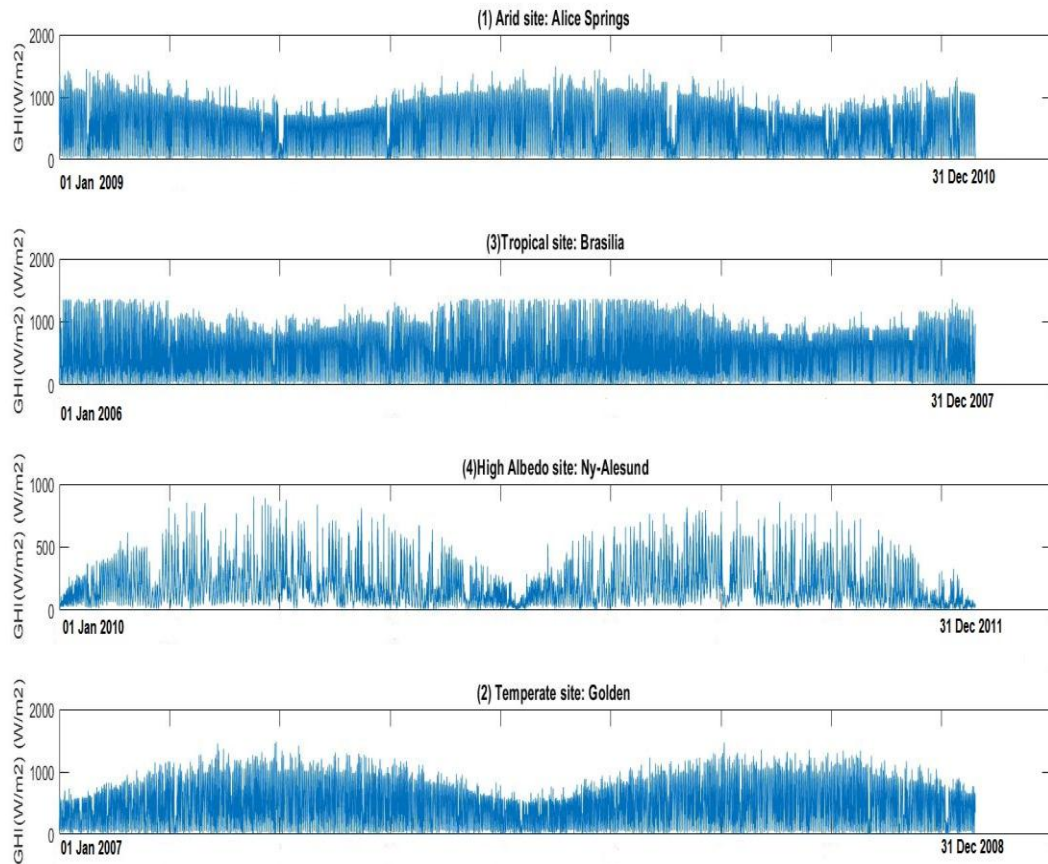


FIGURE III.8 Série chronologique de données GHI sur deux années successives, des stations (Alice Springs, Brasilia, Ny-Ålesund, and Golden, du haut jusqu'en bas).

III.4 Estimation de l'erreur de prédiction

Différentes métriques ont été utilisées dans la littérature pour évaluer la précision des modèles de rayonnement solaire et des prévisions. Dans la présente étude, les indicateurs statistiques suivants sont utilisés.

- **Coefficient de détermination (R^2):**

$$R^2 = 1 - \frac{\text{var}(y_t - \hat{y}_t)}{\text{var}(y_t)} \quad (\text{III.29})$$

- **Erreur moyenne absolue en pourcentage (MAPE):**

$$\text{MAPE} = \frac{1}{M} \sum_{t=1}^M \frac{|y_t - \hat{y}_t|}{|y_t|} \times 100 \quad (\text{III.30})$$

• **Erreur quadratique moyenne normalisée (NMSE):**

$$\text{NMSE} = \frac{\frac{1}{M} \sum_{t=1}^M (y_t - \hat{y}_t)^2}{\text{var}(y)} \quad (\text{III.31})$$

où $\text{var}(y)$ est la variance des valeurs mesurées.

• **Racine de l'erreur quadratique moyenne (RMSE):**

$$\text{RMSE} = \sqrt{\frac{1}{M} \sum_{t=1}^M (y_t - \hat{y}_t)^2} \quad (\text{III.32})$$

où y_t et \hat{y}_t sont respectivement les valeurs mesurées et prédites et M est le nombre de séries prédites.

• **Compétence de prévision (FS):**

FS peut être exprimé comme le rapport de la performance du modèle à celle du modèle de persistance, en termes de RMSE:

$$\text{FS} = 1 - \frac{\text{RMSE}_m}{\text{RMSE}_p} \quad (\text{III.33})$$

III.5 Résultats et discussion

Pour les expériences rapportées ici, la série chronologique GHI comprend deux années de données sur chacune des quatre stations de test. La série GHI, définie dans un espace des variables à 50 dimensions (c.-à-d. Les derniers points de l'heure à tout moment, où les périodes de nuit sont exclues), est subdivisée en trois ensembles : deux ensembles d'apprentissage (formation et validation, environ 25% de l'ensemble des données pour chacun d'eux) et un ensemble de test (environ la moitié de l'ensemble de données). L'ensemble d'apprentissage est utilisé pour apprendre les différents modèles individuels de l'ensemble, alors que l'ensemble de validation est nécessaire pour former les différents schémas de combinaison / classification supervisés proposés dans cette étude (SA, LSR, EB, DWS, OP et Class.).

III.5.1. Résultats obtenus avec des prédictors autorégressifs

Les prédictors autorégressifs sont la classe de modèles la plus générale pour la prédiction de séries chronologiques. Dans cette thèse, cinq modèles sont utilisés (AR, MA, ARMA, ARIMA et SARIMA). On constate que le meilleur modèle est ARIMA sur tous les

sites, probablement parce qu'il tire profit de deux modèles de base (AR et MA) avec la procédure mathématique AR. appelée «opération de différenciation» appliquée pour rendre la série immobile. En ce qui concerne les méthodes de combinaison, la méthode de classification donne les meilleurs résultats sur tous les sites (voir tableau III.2 et figure III. 9).

TABLEAU III.2. Résultats statistiques obtenus à toutes les stations de test par différents modèles autorégressifs et leurs méthodes de combinaison. Dans chaque groupe de colonnes, les meilleurs résultats apparaissent en gras.

Site/Méthode	Modèles Individuels				Méthodes de Combinaison					
	ARMA	ARIMA	SARIMA	Persist	SA	LSR	EB	DWS	OP	Class
Alice Springs										
FS	0.09	0.11	-0.07	0.04	0.09	0.22	0.11	0.03	0.11	0.38
R ²	0.79	0.80	0.71	0.77	0.79	0.82	0.72	0.76	0.80	0.90
MABE	128.3	115.8	134.8	139.7	120.0	99.4	111.3	121.7	110.8	71.9
MAPE	25.0	22.6	26.3	27.4	23.4	19.7	21.8	23.7	21.6	14.0
RMSE	159.1	155.4	187.8	168.0	158.8	134.2	155.1	168.9	155.2	108.1
Brasillia										
FS	0.05	0.06	0.05	0.14	0.19	0.24	0.19	0.20	0.19	0.43
R ²	0.72	0.73	0.72	0.77	0.80	0.83	0.79	0.80	0.80	0.90
MABE	128.9	113.6	123.0	121.0	97.6	95.3	97.6	102.8	97.4	62.2
MAPE	25.5	22.4	24.3	24.0	19.3	18.8	19.3	20.3	19.2	12.3
RMSE	162.6	160.5	163.5	147.2	138.5	129.3	137.9	136.6	138.5	97.1
Ny-Ålesund										
FS	0.06	0.04	-0.06	0.08	-0.03	0.08	-0.01	-0.17	0.006	0.27
R ²	0.60	0.57	0.47	0.61	0.52	0.63	0.47	0.37	0.55	0.76
MABE	67.6	60.5	71.5	66.3	71.9	65.2	70.7	82.2	69.2	44.7
MAPE	32.7	28.7	34.0	32.0	34.8	31.5	34.2	39.8	33.4	21.6
RMSE	87.5	83.1	92.6	85.5	96.1	85.2	94.2	109.0	92.4	67.7
Golden										
FS	0.10	0.13	-0.02	0.20	0.11	0.19	0.12	-0.01	0.11	0.40
R ²	0.76	0.77	0.69	0.81	0.76	0.81	0.76	0.69	0.76	0.89
MABE	102.1	93.9	106.6	94.6	92.2	89.7	92.2	103.8	92.1	57.5
MAPE	23.9	22.0	24.9	22.1	21.6	21.01	21.6	24.3	21.6	13.4
RMSE	132.7	128.4	150.8	117.7	130.8	119.1	129.6	149.2	130.5	87.6

III.5.2. Résultats obtenus avec les prédicteurs neuronaux

Les réseaux de neurone sont devenus une technologie du monde réel, avec des applications efficaces dans différents domaines. La majorité de ces applications traitent des problèmes de reconnaissance de formes, principalement de classification, de régression et de regroupement. Ils exploitent des architectures de réseaux telles que le perceptron multicouche (MLP) ou le réseau à fonction de base radiale [123]. Dans la présente étude, plusieurs architectures sont étudiées, à savoir: MLP, RBF, ELM et ENN. Ici, ENN semble avoir les meilleurs résultats pour tous les indicateurs statistiques (voir tableau III.3 et figure III.10), ce qui peut s'expliquer par le fait qu'un ENN révèle le comportement dynamique temporel d'une séquence temporelle. Contrairement aux réseaux de neurones à anticipation directe

(feedforward), les ENN peuvent utiliser leur état interne (mémoire) pour traiter diverses séquences de mesures de GHI. Cela les rend plus appropriés pour ce genre de tâches.

TABLEAU III.3. Résultats des prédictions pour les quatre stations de test obtenues avec quatre modèles RNA et leurs méthodes de combinaison. Dans chaque groupe de colonnes, les meilleurs résultats apparaissent en gras.

Site/ Méthode	Modèles Individuels					Méthodes de Combinaison					
	MLP	RBF	ELM	ENN	Persist	SA	LSR	EB	DWS	OP	Class
Alice Springs											
FS	0.32	0.28	0.30	0.34	0.13	0.34	0.35	0.34	0.34	0.35	0.49
R ²	0.89	0.87	0.87	0.89	0.81	0.89	0.89	0.89	0.89	0.89	0.93
MABE	80.6	86.9	85.0	75.4	129.7	76.7	74.8	76.5	76.8	76.1	50.7
MAPE	15.7	16.9	16.5	14.7	25.3	14.9	14.6	14.9	14.9	14.8	9.8
RMSE	118.5	125.6	122.5	114.6	151.9	114.3	113.0	114.2	114.2	113.9	88.1
Brasilia											
FS	0.36	0.32	0.34	0.38	0.12	0.38	0.42	0.39	0.35	0.39	0.54
R ²	0.88	0.87	0.87	0.88	0.76	0.89	0.89	0.89	0.87	0.89	0.93
MABE	80.4	89.8	86.9	80.2	128.8	79.9	65.5	79.6	85.4	79.6	50.6
MAPE	15.9	17.7	17.2	15.9	25.5	15.8	12.9	15.7	16.9	15.7	10.0
RMSE	108.4	115.4	112.7	106.1	150.6	104.9	99.1	104.8	110.9	104.8	78.8
Ny-Ålesund											
FS	0.14	0.12	0.13	0.15	0.12	0.16	0.20	0.16	0.12	0.31	0.37
R ²	0.69	0.66	0.67	0.69	0.62	0.70	0.71	0.69	0.66	0.79	0.83
MABE	61.4	60.8	59.6	57.7	59.2	57.2	55.2	57.2	60.7	41.6	48.4
MAPE	29.7	29.4	28.8	27.9	29.6	27.7	26.7	27.7	29.3	20.1	23.4
RMSE	79.5	81.5	80.0	78.2	80.2	77.4	73.7	77.3	81.5	63.6	58.1
Golden											
FS	0.34	0.29	0.29	0.34	0.20	0.35	0.38	0.35	0.35	0.35	0.51
R ²	0.877	0.85	0.85	0.876	0.81	0.87	0.88	0.87	0.88	0.88	0.93
MABE	73.3	84.2	81.1	71.5	101.5	74.3	63.6	73.8	72.7	72.4	47.1
MAPE	17.1	19.7	19.0	16.7	23.7	17.4	14.9	17.3	17.0	16.9	11.0
RMSE	97.3	104.3	104.9	96.7	117.9	95.9	90.8	95.8	95.9	94.8	71.5

III.5.3. Résultats obtenus avec les prédicteurs SVM

La sélection de ce type de prédicteur vient de l'intérêt porté à une évaluation globale et large de l'efficacité de la technique SVM lorsqu'elle est utilisée dans les problèmes de prédiction de GHI. Plus précisément, trois différents types de SVM sont considérés : un SVM linéaire (SVM-Linear; égal à un SVM sans transformation avec noyau), un SVM avec un noyau polynomial (SVM-Polynomial) et un SVM avec un noyau gaussien (SVM-RBF).

L'ajout de ce dernier type de SVM permet d'évaluer l'effet de la transformation noyau sur la méthode de prévision SVM. De plus, des indications précieuses sont ensuite obtenues sur les modèles de prédiction à choisir pour mettre en œuvre les diverses stratégies de combinaison et évaluer l'efficacité de l'ensemble. Comme indiqué dans le tableau III.4, les performances du SVM-RBF sont optimales sur tous les sites de test, comme l'indiquent toutes les mesures de performance. Cela peut être lié à la flexibilité du noyau RBF pour modéliser

les phénomènes non linéaires (voir figure III.11). En ce qui concerne les méthodes de combinaison, la stratégie basée sur la classification surpasse largement les autres.

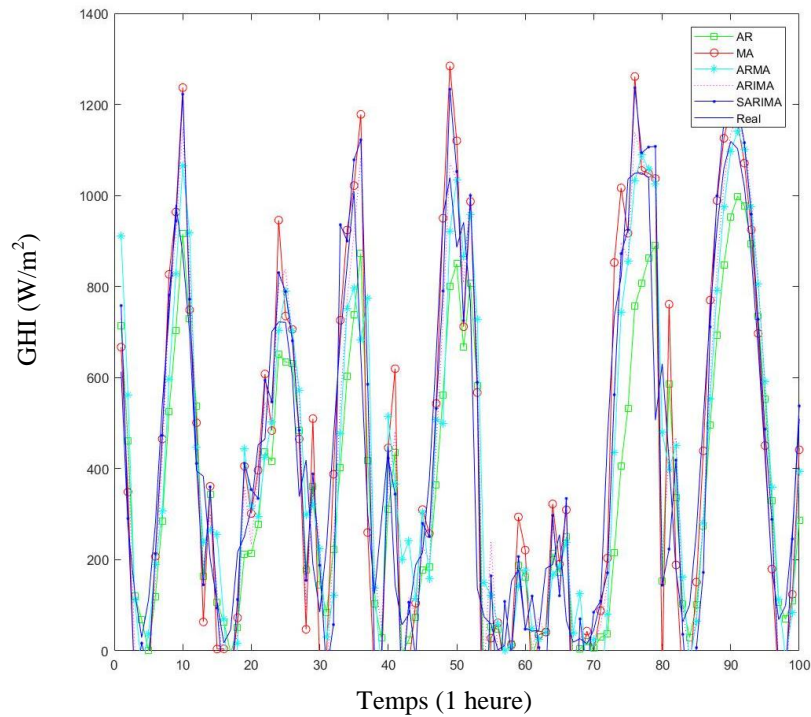


FIGURE III.9 : Une petite fenêtre sur les séries temporelles mesurées et prédites par différents modèles autorégressifs sur le site d'Alice Springs.

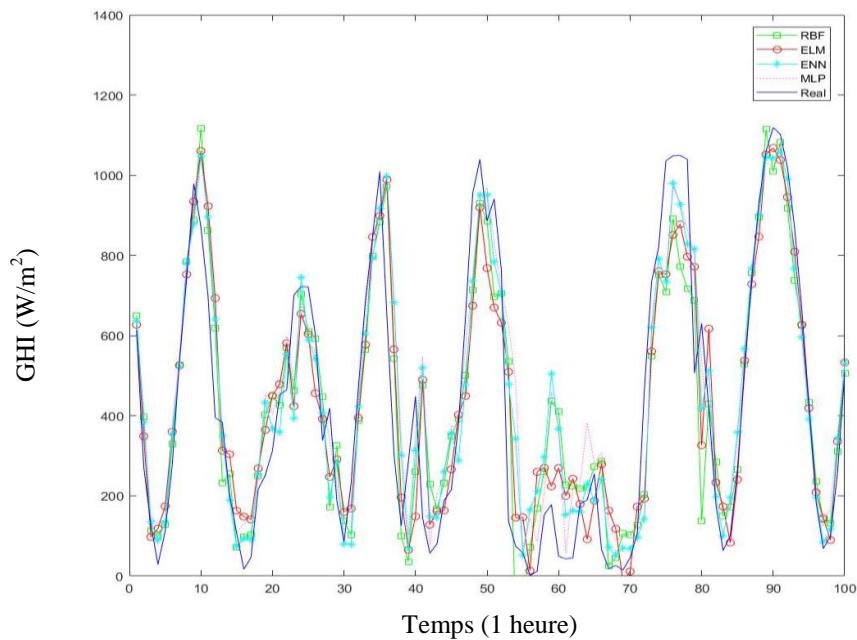


FIGURE III.10 Une petite fenêtre sur les séries temporelles mesurées et prédites par différents modèles d'RNA sur le site de test d'Alice Springs.

TABLEAU III.4. Résultats de prédiction obtenus avec trois modèles à noyau SVM et leurs méthodes de combinaison aux quatre stations de test. Dans chaque colonne, les meilleurs résultats apparaissent en gras.

Site/ Méthode	Modèles Individuels				Méthodes de Combinaison					
	SVM- RBF	SVM- Poly	SVM- Lin.	Persist.	SA	LSR	EB	DWS	OP	Class
Alice Springs										
FS	0.35	0.27	0.27	0.14	0.33	0.35	0.34	0.35	0.34	0.48
R ²	0.89	0.86	0.86	0.81	0.89	0.89	0.89	0.89	0.89	0.93
MABE	75.8	86.2	89.5	127.3	78.6	76.05	78.3	75.8	77.9	73.2
MAPE	14.7	16.8	17.4	24.8	15.3	14.8	15.2	14.7	15.1	10.7
RMSE	113.6	126.6	126.9	149.2	115.8	113.5	115.67	113.6	115.3	91.0
Brasilia										
FS	0.39	0.38	0.37	0.16	0.39	0.42	0.39	0.39	0.39	0.48
R ²	0.89	0.88	0.88	0.78	0.89	0.89	0.89	0.89	0.89	0.92
MABE	80.5	81.4	84.1	123.5	81.1	66.7	81.1	80.5	81.3	63.4
MAPE	15.9	16.1	16.6	24.5	16.0	13.2	16.0	15.9	16.1	12.5
RMSE	105.0	106.3	108.3	144.3	104.5	99.0	104.5	104.8	104.5	88.6
Ny-Ålesund										
FS	0.15	0.13	0.07	0.11	0.14	0.18	0.14	0.15	0.15	0.26
R ²	0.68	0.66	0.63	0.40	0.68	0.70	0.68	0.68	0.68	0.76
MABE	57.7	59.6	65.2	78.2	58.4	56.1	58.3	57.5	58.2	45.1
MAPE	27.9	28.8	31.5	33.1	28.2	27.1	28.2	27.8	28.1	21.8
RMSE	78.2	80.5	85.8	99.5	79.3	75.7	79.	78.2	79.0	68.1
Golden										
FS	0.34	0.33	0.31	0.28	0.34	0.35	0.34	0.34	0.34	0.45
R ²	0.87	0.86	0.86	0.84	0.87	0.87	0.87	0.87	0.87	0.91
MABE	72.3	75.4	78.4	92.6	73.2	70.1	73.1	72.3	73.2	56.6
MAPE	16.9	17.6	18.3	21.6	17.1	16.4	17.1	16.9	17.1	13.2
RMSE	97.2	98.7	101.1	106.3	96.3	95.4	96.3	97.2	96.3	80.0

III.5.4. Résultats obtenus avec un ensemble de prédictions hybride

Dans ce cas, pour chaque famille de prédictions, le plus performant est sélectionné. Le meilleur modèle individuel de l'ensemble semble être SVM-RBF. On pouvait s'y attendre dans une certaine mesure, en raison de la meilleure capacité de généralisation des SVM aux problèmes d'apprentissage. En ce qui concerne les stratégies de combinaison, comme pour les combinaisons précédentes, la stratégie basée sur la classification surpasse largement les autres (voir le tableau III.5 et la figure III.12).

TABLEAU III.5. Résultats de prédictions obtenus avec le meilleur modèle de chaque famille de prédiction (SVM-RBF, ENN et ARIMA) et leurs méthodes de combinaison aux quatre stations de test. Dans chaque groupe de colonnes, les meilleurs résultats apparaissent en gras. Les modèles individuels incluent le modèle ciel clair (CS) [124] et le Persistence (Persist).

Site/Method	Modèles Individuels				Méthodes de Combinaison						
	ARIMA	ENN	SVM-RBF	CS	Persist.	SA	LSR	EB	DWS	OP	Class
Alice Springs											
FS	0.11	0.34	0.35	0.15	0.17	0.35	0.36	0.36	0.35	0.36	0.51
R ²	0.80	0.89	0.89	0.58	0.82	0.89	0.90	0.89	0.89	0.89	0.94
MABE	115.8	75.4	75.8	140.1	121.0	75.9	71.4	73.9	75.7	73.4	49.3
MAPE	22.6	14.7	14.8	27.3	23.6	14.8	13.9	14.4	14.7	14.3	9.6
RMSE	155.4	114.6	113.6	260.0	145.4	113.6	110.8	112.1	113.5	111.4	84.5
Brasillia											
FS	0.06	0.38	0.39	0.26	0.11	0.40	0.42	0.41	0.40	0.39	0.59
R ²	0.73	0.88	0.89	0.71	0.75	0.89	0.89	0.88	0.89	0.88	0.94
MABE	113.6	80.2	80.5	91.5	131.4	73.2	65.6	73.1	77.0	73.4	46.0
MAPE	22.4	15.8	15.9	18.1	26.0	14.5	12.9	14.4	15.2	14.5	9.1
RMSE	160.5	106.1	105.0	186.3	152.5	102.8	99.2	100.4	101.8	104.1	70.3
Ny-Ålesund											
FS	0.04	0.15	0.15	-0.01	0.10	0.20	0.20	0.14	0.17	0.13	0.38
R ²	0.57	0.69	0.68	-1.29	0.70	0.50	0.74	0.41	0.70	0.66	0.83
MABE	60.5	57.8	57.7	155.6	58.8	78.3	26.4	61.0	56.6	62.8	39.3
MAPE	28.7	27.9	27.9	117.4	28.4	37.8	54.7	29.5	27.3	30.4	19.0
RMSE	83.1	78.2	78.2	187.9	75.2	98.1	71.1	79.6	76.8	80.8	56.8
Golden											
FS	0.13	0.34	0.34	0.20	0.17	0.36	0.37	0.36	0.35	0.36	0.53
R ²	0.77	0.87	0.87	0.60	0.79	0.88	0.88	0.87	0.87	0.88	0.93
MABE	93.9	71.5	72.3	106.0	102.5	67.0	65.0	66.8	68.8	66.9	43.5
MAPE	22.0	16.7	16.9	24.8	23.9	15.7	15.2	15.6	16.1	15.6	10.1
RMSE	128.4	96.7	97.2	197.0	121.5	94.2	92.6	93.3	95.5	93.8	68.2

III.5.5. Résultats obtenus avec la sélection des variables

Une autre façon d'étudier les performances de la méthodologie de prédiction proposée consiste à utiliser une étape de traitement additif avant d'introduire la série temporelle dans les différents prédicteurs. Ici, la méthode de sélection de caractéristiques décrite dans [112] est utilisée et le meilleur ensemble de variables historiques sélectionnées est obtenu à l'aide du principe de sélection de modèle [125]. Pour chaque cas, le meilleur ensemble est obtenu en fonction des performances de la technique de prédiction dans l'ensemble de validation. Le tableau III.6 montre une légère amélioration des performances par rapport aux résultats du tableau III.5 (sans sélection de caractéristiques). Une exception importante est celle du modèle autorégressif (ARIMA), où les performances de prédiction dégradent.

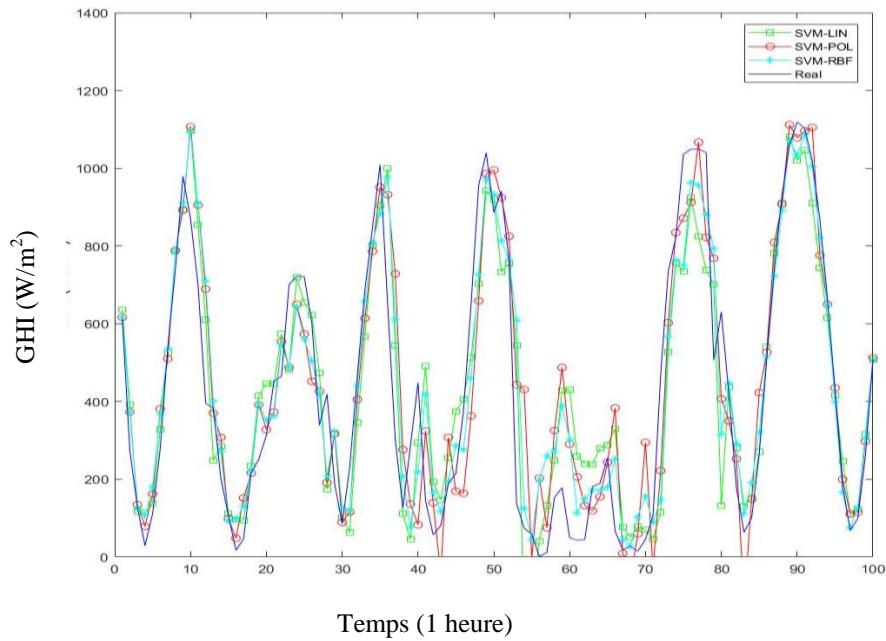


FIGURE III.11. Une petite fenêtre sur les séries temporelles mesurées et prédites par différents modèles de SVM pour le site de test d'Alice Springs.

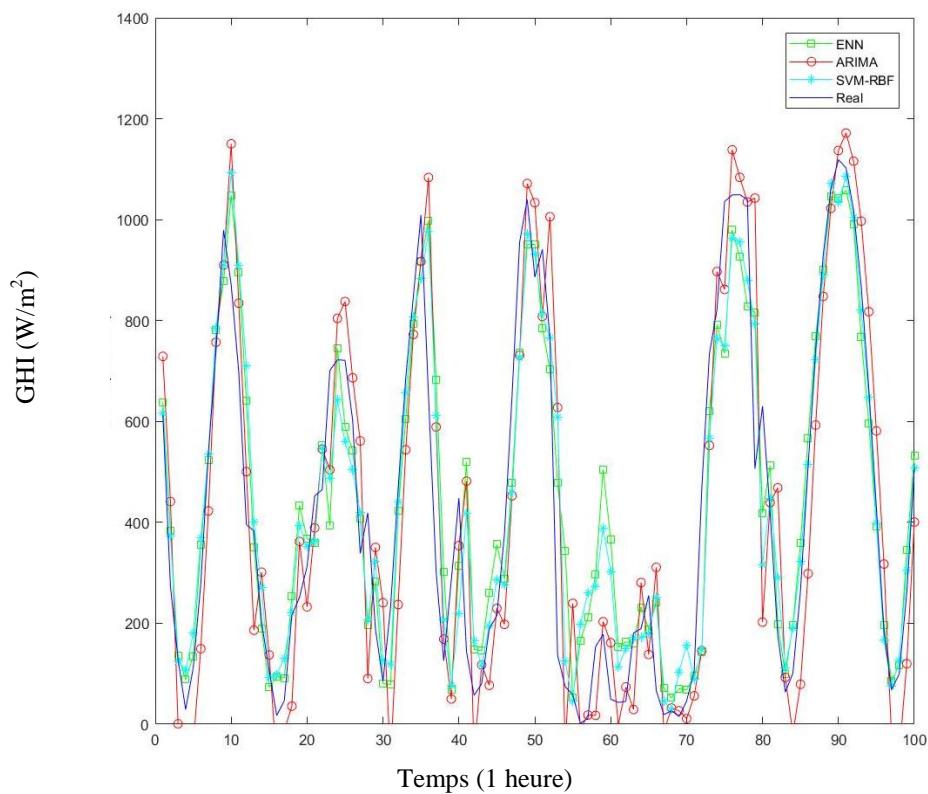


FIGURE III.12. Comparaison entre les modèles les plus performants pour chaque famille de prédiction sur le site de test d'Alice Springs.

Cela peut s'expliquer par le fait que les modèles ARIMA sont basés sur la modélisation polynomiale et nécessitent davantage de variables historiques pour comprendre le comportement de la série chronologique du signal de rayonnement solaire.

Néanmoins, lorsqu'on utilise différentes techniques de combinaison, le système conserve le même niveau de performance malgré les mauvais résultats de la technique ARIMA. Cela confirme l'utilité de la stratégie de combinaison lorsque les techniques d'entrée sont suffisamment diversifiées (différentes familles de prédiction).

Un autre moyen d'analyse pour comparer les performances des différents modèles consiste à utiliser le diagramme de Taylor [126], qui est une illustration graphique de la connexion statistique entre deux variables aléatoires. Il caractérise trois statistiques reliant deux variables en même temps, à savoir : la corrélation, l'écart type et la différence quadratique moyenne. Ces graphiques sont particulièrement adaptés pour évaluer plusieurs aspects des modèles complexes ou pour juger la compétence relative des différents modèles

Dans l'espace bidimensionnel d'un diagramme de Taylor, chaque point peut décrire la combinaison de trois statistiques énoncées ci-dessus, elles sont liées entre elles par :

$$E'^2 = \sigma_t^2 + \sigma_m^2 - 2\sigma_t\sigma_m R \quad (\text{III.34})$$

où E' est la différence quadratique moyenne centrée entre la valeur mesurée et celle prédite, R est le coefficient de corrélation entre la série mesurée et la série prédite et σ_m et σ_f sont les variances des champs de test et de modèle, respectivement. La différence RMS centrée entre les séries temporelles GHI prédites et mesurées est relative à la distance entre la prévision et le point de l'axe des x identifié comme "Mesuré". De même, l'écart type de la série prédite est proportionnel à la distance radiale entre la prédiction et l'origine.

En utilisant les mesures réelles et les modèles de prédiction, les diagrammes de Taylor pour les différents sites sont montrés aux figures III.13-16. Ces diagrammes indiquent que toutes les valeurs de corrélation entre les séries chronologiques mesurées et prévues se situent dans l'intervalle [0,95–0,99] pour Alice Springs, Ny-Ålesund et Brasilia et dans l'intervalle [0,90–0,95] pour le site de Golden. Ces diagrammes indiquent également que la stratégie de classification surpasse tous les autres modèles de combinaison. Il est à noter que le diagramme de Taylor ne fournit pas d'informations

sur les biais généraux, mais décrit simplement l'erreur du modèle centrée.

TABLEAU III.6. Résultats des prédictions obtenus à l'aide du système à deux blocs (technique de sélection des caractéristiques et de régression). Dans chaque groupe de colonnes, les meilleurs résultats apparaissent en gras. SV: variables historiques sélectionnées.

Site/Méthode	Modèles Individuels			Méthodes de combinaison					
	ARIMA	ENN	SVM_ RBF	SA	LSR	EB	DWS	OP	Class
Alice Springs	SV=15	SV=12	SV=19						
FS	-1.62	0.35	0.35	-0.09	0.37	0.28	0.35	0.31	0.48
R ²	-0.72	0.89	0.89	0.70	0.90	0.88	0.89	0.88	0.93
MABE	386.5	77.5	75.5	159.2	14.4	91.0	75.6	89.4	57.1
MAPE	75.3	15.1	14.7	31.0	74.1	17.7	14.7	17.4	11.1
RMSE	460.4	113.0	113.2	191.6	109.3	124.9	112.9	120.1	91.1
Brasilia	SV=12	SV=11	SV=16						
FS	-1.31	0.38	0.38	-0.001	0.43	0.31	-0.02	0.30	0.50
R ²	-0.65	0.895	0.89	0.69	0.90	0.85	0.68	0.85	0.92
MABE	334.1	81.4	81.3	144.8	65.5	94.5	146.8	97.3	62.1
MAPE	66.1	16.1	16.0	28.6	12.9	18.7	29.0	19.2	12.2
RMSE	397.4	105.5	105.3	172.3	96.7	118.3	175.7	119.6	85.7
Ny-Ålesund	SV=17	SV=14	SV=11						
FS	-1.83	-0.20	-0.15	-0.17	0.12	-0.04	-0.15	-0.04	0.23
R ²	-0.66	0.78	0.78	0.75	0.84	0.75	0.80	0.82	0.89
MABE	113.5	43.1	39.9	47.3	29.6	40.0	41.7	40.3	26.9
MAPE	85.7	32.5	30.1	35.6	22.4	30.2	31.4	30.4	20.3
RMSE	145.1	61.5	58.9	60.3	44.6	53.6	59.1	53.5	38.9
Golden	SV=17	SV=17	SV=13						
FS	-0.60	0.36	0.33	0.17	0.39	0.30	-0.39	0.31	0.53
R ²	0.23	0.88	0.87	0.80	0.89	0.85	0.41	0.86	0.93
MABE	193.2	70.8	78.2	99.7	14.8	81.9	169.7	81.6	49.6
MAPE	45.2	16.6	18.3	23.3	63.3	19.1	39.7	19.1	11.6
RMSE	236.6	94.0	98.9	121.3	89.6	102.4	206.7	101.5	68.4

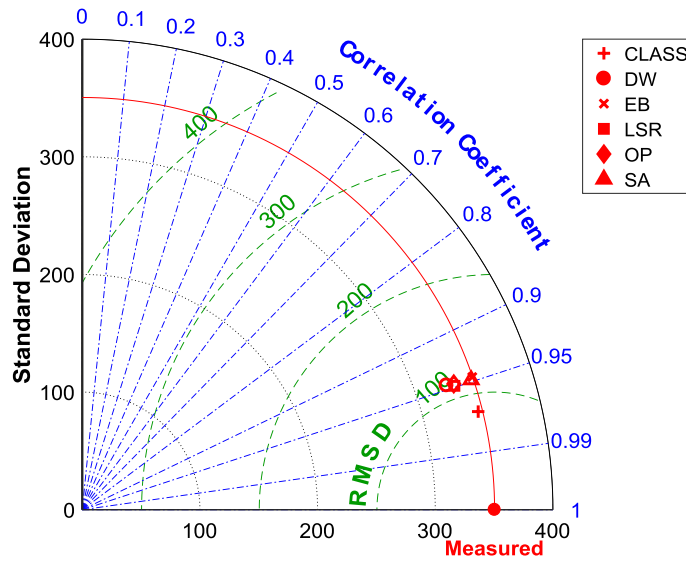


FIGURE III. 13. Diagramme de Taylor des valeurs de GHI mesurées et prédites par différents modèles de combinaison pour le site de Alice Springs.

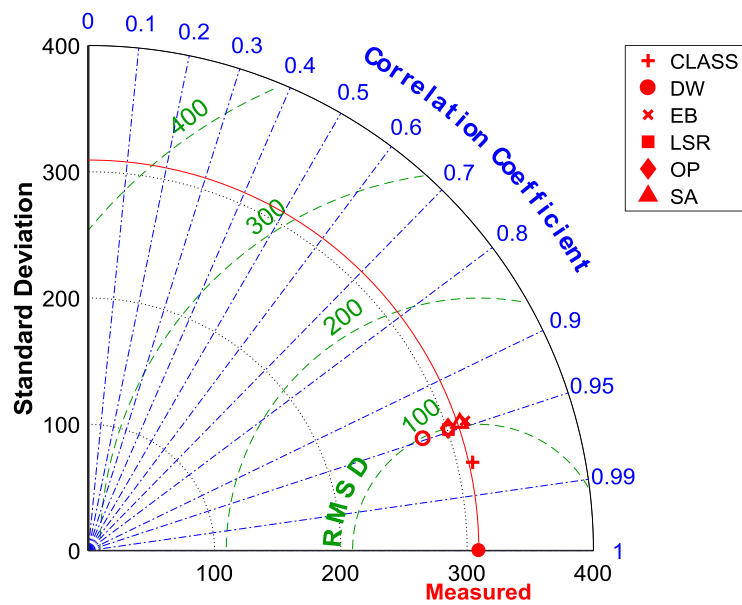


FIGURE III. 14. Diagramme de Taylor des valeurs GHI mesurées et prédites par différents modèles de combinaison pour le site de Brasilia.

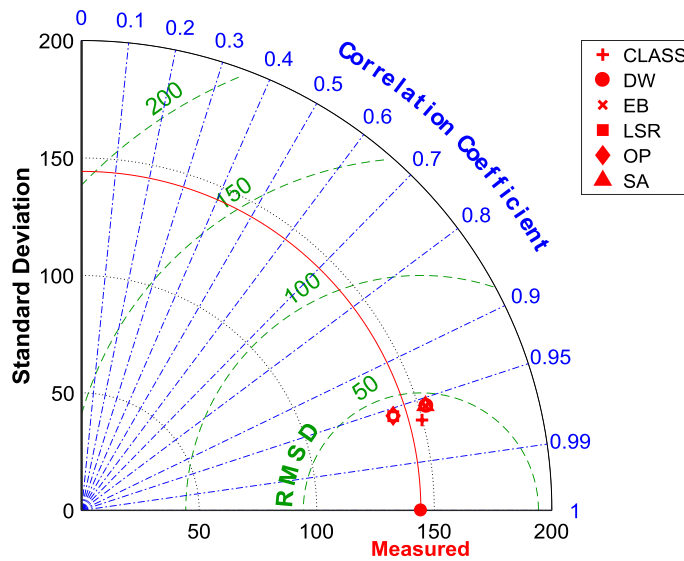


FIGURE III. 15. Diagramme de Taylor des valeurs GHI mesurées et prédites par différents modèles de combinaison pour le site de Ny-Ålesund.

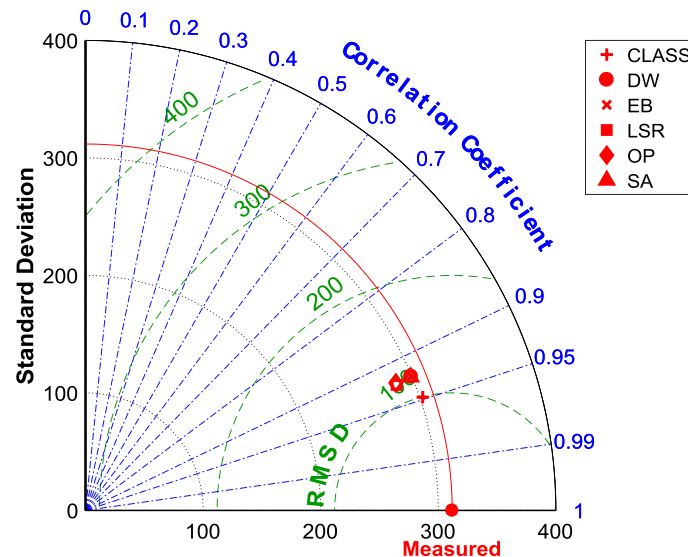


FIGURE III. 16. Diagramme de Taylor des valeurs GHI mesurées et prédites par différents modèles de combinaison pour le site de Golden.

II.5.6 Test d'hypothèse

Le test d'hypothèse statistique utilisé dans cette étude est principalement basé sur le test t apparié et le test de Fisher signé. Ils sont utiles pour déterminer si une approche de combinaison particulière des meilleurs modèles individuels, à savoir (ARIMA) du modèle autorégressif, (MLP) des réseaux neuronaux artificiels et (SVM-RBF) des machines à vecteurs de support, est plus performante que la meilleure stratégie de combinaison. Cette dernière est la stratégie de classification [127,128]. Pour effectuer les tests statistiques, l'ensemble de données complet est subdivisé en M sous-ensembles de sorte que, chaque

modèle soit formé en sous-division i et testé en sous-division $i + 1$. Pour obtenir un groupe de 51 sous-ensembles différents, dans la présente étude, K est fixé à 50. Le test de *Kolmogorov–Smirnov* ($K - S$) peut ensuite être exécuté pour confirmer la normalité des données. Si en effet la distribution des données est normale, un test t est exécuté pour comparer les performances des diverses stratégies combinées et découvrir les meilleurs. Sinon Le test non paramétrique de Fisher signé est effectué pour obtenir la même comparaison.

Le tableau III.6 compare les différentes stratégies de fusion et la méthode de classification proposée sur la base de test d'Alice Springs. Le niveau de signification, α , est fixé à 0,05, ce qui correspond à un niveau de confiance de 95%. Les résultats du test $K-S$ et ceux du t -test ou du test de Fisher sont également présentés. De plus, pour chaque cas, les valeurs *Win-Lose-Tie* ($W-L-T$) sont fournies. Ici, la valeur probabiliste p est la probabilité que, lorsque l'hypothèse nulle est vraie, la performance statistique de la méthode de classification proposée est identique ou supérieure à celle d'autres scénarios. En d'autres termes, une valeur p inférieure indique une amélioration des performances du modèle proposé. D'après le tableau III.7, il est évident que l'approche de classification surpasse statistiquement toutes les autres stratégies de combinaison, comme en témoignent les valeurs de $W-L-T$.

TABLEAU III.7 Résultats du test d'hypothèse comparant la stratégie de classification à d'autres stratégies de combinaison pour le site d'Alice Springs.

Class. vs. autres méthodes de combinaison	W-L-T	Test de Fisher p	Test K-S p	Test t p
LSR	27-22-1	—	0.840	0.900
OP	26-22-2	—	0.150	0.910
EB	44-6-0	—	0.004	0.020
DWS	36-14-0	—	0.030	0.650
SA	47-0-3	—	0.190	0.710

III.6. Conclusion

L'analyse et la prédiction des séries temporelles constituent un domaine de recherche dynamique, qui revêt une importance capitale dans de nombreux domaines pratiques. L'amélioration de l'exactitude des prédictions de séries chronologiques est une tâche qui retient de plus en plus l'attention des chercheurs au cours des dernières décennies. Dans ce chapitre, une nouvelle méthodologie a été proposée pour prévoir l'irradiation solaire globale horizontale (GHI) de l'heure qui suit, sur la base d'un ensemble multi modèles. La méthodologie vise à exploiter de la diversité entre un ensemble de modèles de prédiction

variés, dans le but d'améliorer la précision du processus de prédiction. Plusieurs stratégies ont été décrites et mises en œuvre. Les différentes stratégies varient les unes des autres par rapport à : (i) la procédure de combinaison ; et (ii) la technique utilisée pour exploiter les connaissances antérieures accessibles de la série chronologique horaire de GHI.

Les divers ensembles multi modèles utilisés dans les expériences actuelles reposent sur différentes familles de prédictions, principalement : (1) modèles autorégressifs ; (2) modèles neuronaux ; et (3) modèles de machines à vecteurs de support. Les méthodologies proposées ici peuvent améliorer les performances de la prédiction de GHI, en combinant différentes techniques, simplement avec la condition de diversité, c'est-à-dire en utilisant différentes familles de prédicteurs.

Afin de réaliser une évaluation cohérente et complète de l'efficacité de la méthodologie proposée, différentes expériences (avec différentes architectures) ont été considérées. Les expériences ont simulé plusieurs architectures, dans lesquelles les membres de l'ensemble étaient soit basés sur la même famille de prédiction, soit sur le meilleur modèle de chaque famille de prédiction inclus dans l'ensemble.

L'évaluation expérimentale réalisée pour évaluer l'efficacité du système multi modèle a clairement montré qu'il pouvait fournir une précision et des compétences en prédiction meilleures que le meilleur modèle unique introduit dans l'ensemble, avec des performances plus élevées lors de l'utilisation de la stratégie de classification. Ce résultat est particulièrement encourageant car les compétences améliorées de la méthode d'ensemble proposée sont maintenues sur les quatre sites de test utilisés ici, lesquels sont représentatifs de divers climats et régimes de nuages. Les développements ultérieurs de la méthode permettront d'évaluer ses compétences en matière de prédiction à plus court terme, par exemple pour l'horizon 15 min.

Chapitre IV

Estimation des paramètres du modèle autorégressif ARMA par les algorithmes génétiques

IV.1 Introduction

L'Algorithme Génétique est une approche d'optimisation évolutive basée sur des algorithmes de recherche aléatoires, développée par Holland en 1975 [129]. L'AG est une méthode d'optimisation combinatoire très répandue en raison de sa robustesse pour les problèmes complexes et non linéaires. Elle présente de nombreux avantages par rapport aux autres méthodes d'optimisation classiques. Afin d'obtenir de meilleures solutions, différents étapes de calcul telles que la sélection, la mutation et le croisement ont été implémentés dans l'algorithme. L'algorithme peut facilement converger vers une bonne solution plus rapidement que les autres approches classiques.

Les étapes de base du principe de fonctionnement de l'AG sont les suivantes : 1) initialisation d'une population de solutions possibles, puis 2) application des opérateurs de sélection, de croisement et de mutation, respectivement. La fonction d'évaluation est calculée pour chaque solution candidate. Après élimination de la mauvaise solution de la population existante, une nouvelle population est créée à nouveau à l'aide des opérateurs d'AG et le mécanisme de travail est maintenu jusqu'à ce que le critère soit satisfait [130].

Dans ce chapitre, nous proposons une méthode de prédiction intégrant les algorithmes génétiques et la méthode de prédiction à moyenne mobile autorégressive (ARMA) afin de tirer profit de la force intrinsèque des deux modèles à savoir ARMA et les algorithmes génétiques. L'approche est proposée pour prédire l'irradiation solaire globale (Global Horizontal Irradiance GHI) dans deux sites de test qui sont (Brasilia et Ny-Alesund).

IV.2 Principe de fonctionnement des algorithmes génétiques

Les algorithmes génétiques manipulent un ensemble de points dans l'espace de recherche, appelé population d'individus. Chaque individu ou chromosome représente une solution possible du problème posé. Il est constitué d'éléments, appelés gènes, dont les valeurs sont appelées allèles. Les algorithmes génétiques font évoluer cette population

d'individus par générations successives, en utilisant des opérateurs inspirés de la théorie de l'évolution qui sont : la sélection, le croisement et la mutation. De génération en génération, la force des individus de la population augmente et après un certain nombre d'itérations, la population est entièrement constituée d'individus tous forts, soit de solutions quasi-optimales du problème posé [131].

IV.3. Vocabulaire des algorithmes génétiques

Dû au fait que les AG s'appuient à la fois sur la génétique naturelle et l'intelligence artificielle, les terminologies utilisées pour les AG dans la littérature sont un mélange du naturel et de l'artificiel. Les termes usuels appliqués dans les AG sont résumés dans le tableau ci-dessous.

TABLEAU IV.1 Vocabulaire des AG [132].

Algorithmes Génétiques	Significations
Chromosome (individu)	Solution
Gènes (variable)	Partie de la solution
Lieu	Position du gène
Allèles	Valeur du gène
Phénotype	Solution décodée
Génotype	Solution encodée

IV.4 Structure de l'algorithme génétique

L'implémentation d'un AG est spécifique au problème à résoudre. Pour l'utiliser, il faut disposer des cinq éléments suivants [133]:

- Un principe de codage des éléments de la population, qui consiste à associer à chacun des points de l'espace d'état une structure de données, la qualité de ce codage des données conditionne le succès des algorithmes génétiques ; bien que le codage binaire a été très utilisé à l'origine, les codages réels sont désormais largement exploités, notamment dans les domaines applicatifs pour l'optimisation de problèmes à variables réelles.
- un mécanisme de génération de la population initiale qui doit être capable de produire une population d'individus non homogène servant de base pour les générations futures; le choix de la population initiale est important, car il influence la rapidité de la convergence vers l'optimum global; dans le cas où l'on ne dispose que de peu

d'informations sur le problème à résoudre, il est essentiel que la population initiale soit répartie sur tout le domaine de recherche.

- une fonction à optimiser : celle-ci retourne une valeur appelée *fitness* ou fonction d'évaluation de l'individu.
- des opérateurs permettant de diversifier la population au cours des générations et d'explorer l'espace d'état; l'opérateur de croisement recompose les gènes d'individus existant dans la population alors que l'opérateur de mutation garantit l'exploration de l'espace d'état.
- des paramètres de dimensionnement, représentés par la taille de la population, le nombre total de générations, ou le critère d'arrêt, ainsi que les probabilités d'application des opérateurs de croisement et de mutation.

Un algorithme génétique est défini par un cycle de population et fait intervenir trois facteurs importants : fitness, croisement, mutation (figure VI.1). Un cycle représente le passage d'une population à la génération suivante soit l'évolution génétique d'une population.

L'organigramme fonctionnel de la figure VI.2 illustre la structure de l'algorithme génétique. Les diverses phases et les mécanismes associés à chacune d'entre elles seront présentés dans les sections suivantes.

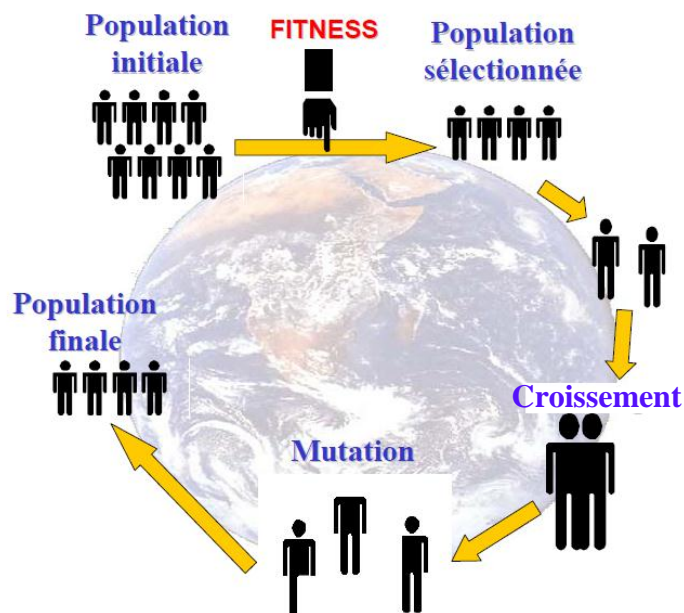


FIGURE VI.1 Illustration d'un cycle dans un algorithme génétique.

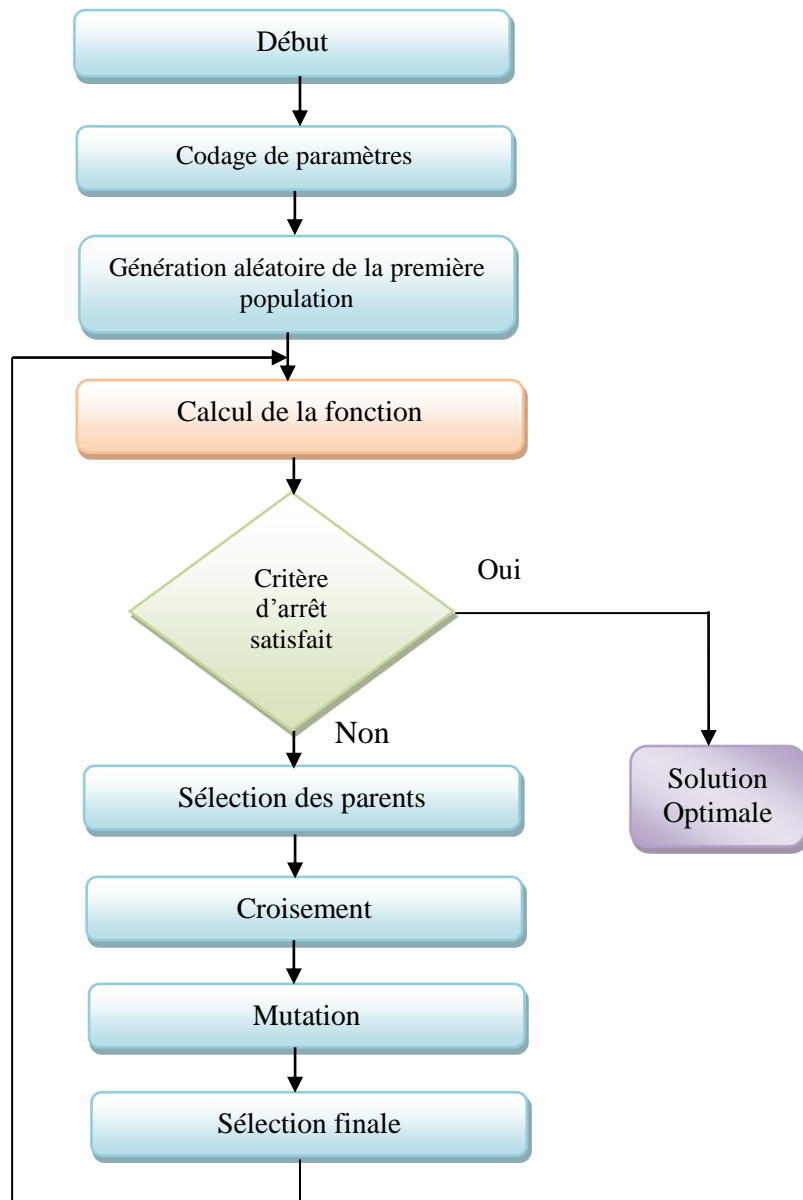


FIGURE IV.2 Organigramme général de l'AG [134].

IV.4.1 Le codage

Dans une première phase, il faut représenter les différents états possibles de la variable dont on cherche la valeur optimale sous une forme utilisable par un AG : c'est le codage. Chaque paramètre d'une solution est assimilé à un gène, toutes les valeurs qu'il peut prendre sont les allèles de ce gène. Un chromosome est une suite de gènes, on peut choisir de regrouper les paramètres similaires dans un même chromosome et chaque gène sera repérable par sa position. Chaque individu est représenté par un ensemble de chromosomes.

Il existe principalement deux types de codage : le codage binaire et le codage réel :

IV.4.1.1 Codage binaire

Chaque gène dispose du même alphabet binaire $\{0, 1\}$. Chaque paramètre x_i situé dans un intervalle $\{x_{i\min}, x_{i\max}\}$, est associé à une chaîne binaire $b_0, b_1, \dots, b_{L_{x_i}-1}$ définie sur L_{x_i} bits. A chaque chaîne correspond une valeur entière naturelle :

$$g_i = \sum_{i=0}^{L_{x_i}-1} 2^i b_i \quad (\text{IV.1})$$

Cette chaîne doit être décodée pour pouvoir calculer la valeur de la fonction coût (fonction d'évaluation) qui lui est associée. Le paramètre réel x_i de l'espace de recherche relatif à g_i est obtenu par interpolation linéaire :

$$x_i = x_{i\min} + \frac{x_{i\max} - x_{i\min}}{2^{L_{x_i}-1}} \cdot g_i \quad (\text{IV.2})$$

IV.4.1.2 Codage réel

Cela peut-être utile notamment dans le cas où l'on recherche le maximum d'une fonction réelle. La représentation des gènes par des nombres réels arrive naturellement dans le cas d'optimisation de paramètres avec des variables comprises dans des domaines continus. La taille des chromosomes est la même que la taille du vecteur qui sera la solution du problème. Suite à cette approche, chaque gène représente une variable du problème.

Dans le codage binaire, l'utilisation de domaines vastes amène à une perte de précision des solutions.

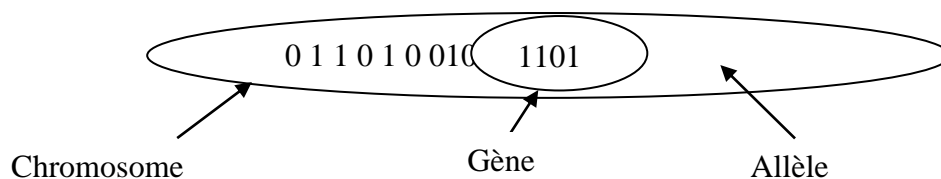


FIGURE IV.3 Structure d'un chromosome.

Autre avantage de l'utilisation des nombres réels est leur capacité à exploiter la gradualité (les changements survenus au niveau des variables impliquent des changements au niveau de la fonction) des fonctions qui ont des variables continues. Dans ce cas, un avantage très important est la capacité de réglage local de solutions, qui se produit beaucoup plus rapidement que dans le cas du codage binaire des algorithmes génétiques.

IV.4.2 Création de la population

La population initiale d'un AG représente la population de départ de l'algorithme. Elle doit contenir des chromosomes qui sont bien répartis dans l'espace des solutions pour fournir à l'AG un matériel génétique varié. Lors de la genèse de la population, l'ensemble dans lequel est effectuée la recherche de solutions doit être de dimension finie. La qualité et le choix du codage des chromosomes de la population est déterminant et conditionne le succès de la recherche pour la meilleure solution. Le choix de la taille de la population est aussi important que le codage, car une population trop grande entraînera une recherche exhaustive, tandis qu'une population trop petite ne permettra pas à l'algorithme d'atteindre une solution optimale.

IV.4.3 Évaluation de la population

Il s'agit d'un processus au cours duquel, les chromosomes de la population sont évalués par la fonction sélective (*fitness function*). C'est une fonction qui permet une évaluation de la population selon le critère de performance choisi. Elle permet ainsi de choisir les candidats les plus performants. Le critère de performance est lié à la nature du problème. Dans certains cas, la fonction sélective peut être établie comme une fonction de l'erreur. La plupart des algorithmes ont pour action de maximiser la valeur de la fonction sélective. Etant donné une fonction f réelle à une ou plusieurs variables, le problème d'optimisation sur l'espace de recherche E s'écrit de la manière suivante :

$$\max_{x \in E} f(x) \quad (\text{VI.5})$$

Dans beaucoup de problèmes, l'objectif est exprimé sous forme de minimisation d'une fonction coût h :

$$\min_{x \in E} h(x) \quad (\text{VI.6})$$

Le passage du problème de minimisation à un problème de maximisation est obtenu par transformation de la fonction h selon la relation suivante :

$$f(x) = \frac{1}{(1+h(x))} \quad (\text{VI.7})$$

IV.5 Les mécanismes d'un AG

A partir d'une première population d'individus créée aléatoirement, les AG génèrent de nouveaux individus plus performants que leurs prédécesseurs en effectuant des opérations génétiques. Les AG utilisent des outils tels que la reproduction, le croisement et la mutation.

Ces outils sont basés sur des processus aléatoires. La reproduction est une version artificielle de la sélection naturelle, c'est un processus dans lequel chaque individu est copié en fonction des valeurs de la fonction d'évaluation. Le croisement est l'opérateur le plus dominant dans un AG, il permet à deux chaînes d'échanger des portions de leurs structures produisant ainsi de nouvelles chaînes. La mutation est un opérateur local qui est appliqué avec une très faible probabilité. Dans ce qui suit, nous allons discuter le rôle de ces paramètres.

IV.5.1 La sélection

Cet opérateur est peut-être le plus important puisqu'il permet aux individus d'une population de survivre, de se reproduire ou de mourir. En règle générale, la probabilité de survie d'un individu sera directement reliée à son efficacité relative au sein de la population.

Il existe plusieurs méthodes pour la reproduction. La méthode la plus connue et utilisée est la roue de loterie biaisée (*roulette wheel*) de [135]. Selon cette méthode, chaque chromosome sera dupliqué dans une nouvelle population proportionnellement à sa valeur d'adaptation. L'inconvénient majeur de cette méthode repose sur le fait qu'un individu n'étant pas le meilleur, peut tout de même dominer la sélection. Elle peut aussi engendrer une perte de diversité par la domination d'un super individu. Un autre inconvénient c'est sa faible performance vers la fin quand l'ensemble des individus se ressemblent.

Brièvement, il existe d'autres méthodes, la plus connue étant celle du tournoi (*tournament selection*) : on tire deux individus aléatoirement dans la population et on reproduit le meilleur des deux dans la nouvelle population. On refait cette procédure jusqu'à ce que la nouvelle population soit complète. Cette méthode donne de bons résultats. Toutefois, aussi important que soit la phase de sélection, elle ne crée pas de nouveaux individus dans la population. Ceci est le rôle des opérateurs de croisement et de mutation.

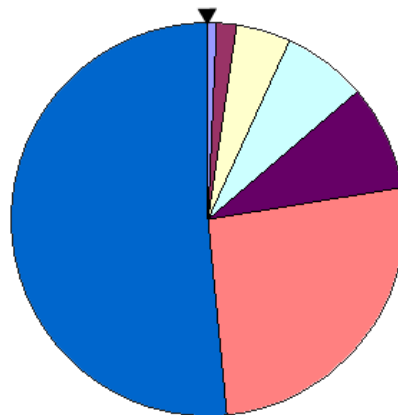


FIGURE IV.4 Schéma d'un exemple d'une roue de loterie.

IV.5.2 Fonction d'adaptation

Ce sont les principes de survie et de reproduction décrits par Charles Darwin [136], où une population ayant une grande variété engendrera, de génération en génération, une contenance des individus dont le génotype se traduit par une meilleure adaptation. Pour cela, il nous faut une fonction dite d'adaptation qui va être dans notre cas la fonction objectif elle-même [137].

IV.5.3 Croisement

L'opérateur de croisement permet la création de nouveaux individus selon un processus très simple. Il permet donc l'échange d'information entre les chromosomes (individus). Tout d'abord, deux individus, qui forment alors un couple, sont tirés au sein de la nouvelle population issue de la sélection. Puis un (potentiellement plusieurs) site de croisement est tiré aléatoirement.

Il existe différents types de croisements pour un algorithme génétique classique. Sur la figure IV.4, les trois principaux types de croisement sont présentés (en un point, en deux points et uniforme). Cette opération est contrôlée par une probabilité p_c souvent supérieure à 0.5 [138].

IV.5.3.1 Croisement en un point

Dans ce type de recombinaison, un point de croisement est choisi au hasard. À partir de ce point de coupure, les données sont échangées entre les individus afin de créer des enfants (voir figure IV.5.a).

IV.5.3.2 Croisement en deux points

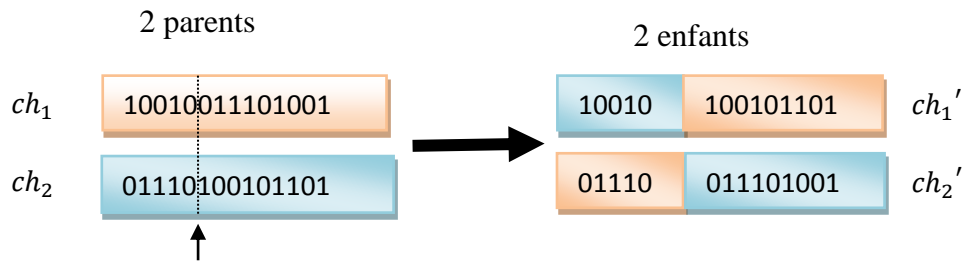
Dans ce cas, on choisit au hasard deux points de coupure à partir desquels on procède à l'échange de données, afin de créer des enfants comme le montre la figure IV.4.

IV.5.3.3 Croisement uniforme

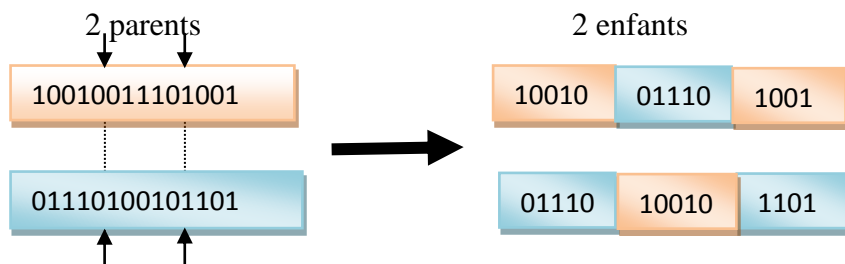
La mise en œuvre de ce procédé est très simple, elle consiste à définir de manière aléatoire un "masque", c'est-à-dire une chaîne de bits de même longueur que les chromosomes des parents sur lesquels il sera appliqué. Ce masque est destiné à savoir, pour chaque *locus* (position sur le chromosome), de quel parent le premier fils devra hériter le gène.

Si face à un *locus*, le masque présente un 0, le fils héritera le gène du parent n°1, s'il présente un 1 il en héritera du parent n°2. La création du fils n°2 se fait de manière symétrique: si pour un gène donné, le masque indique que le fils n°1 devra recevoir celui du

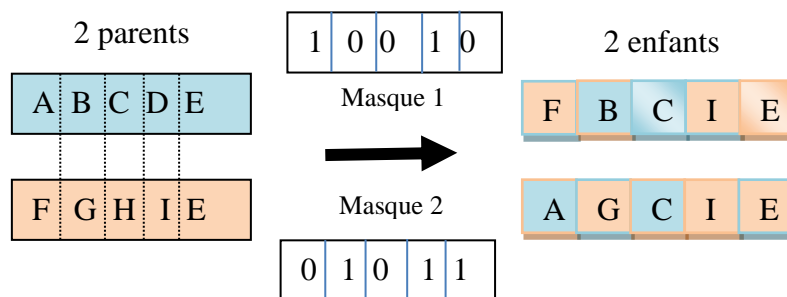
parent n° 1 alors le fils n° 2 le recevra du parent n°2, et si le fils n°1 le reçoit du parent n°2 alors le fils 2 le recevra du parent n°1, comme illustrée sur la figure IV.4.c [139].



(a) croisement en un point.



(b) Croisement en deux points.



(c) Croisement uniforme.

FIGURE IV.5 Exemples d'opérations de croisement [140].

IV.5.4 Mutation

Durant la mutation, des individus sont altérés de façon aléatoire. Ces altérations sont la plupart du temps petites. Selon la représentation des variables, différentes méthodes de mutation peuvent être employées. On a ainsi les méthodes pour des variables à valeurs réelles et les méthodes pour des variables binaires.

IV.5.4.1 Mutation à valeurs réelles

Ici la mutation est réalisée en rajoutant aux individus des réels de très petites valeurs créées de manière aléatoire.

IV.5.4.2 Mutation binaire

La mutation est définie ici comme étant l'inversion d'un bit dans un chromosome (voir figure IV.6). Le choix du bit à inverser est fait de manière aléatoire. Les mutations empêchent la recherche de la solution de stagner, en apportant une légère variation. Elles permettent aussi d'assurer une recherche globale plutôt que locale.

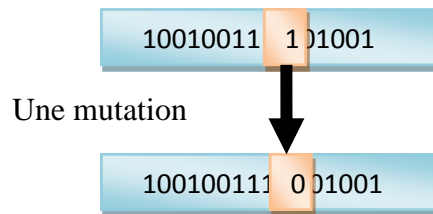


FIGURE IV.6 Mutation binaire.

IV.5.5 Insertion

La population issue des opérations génétiques sera insérée dans l'ancienne population d'une manière à garder les individus ayant la fonction fitness la plus grande.

IV.5.6 Critère d'arrêt

Les opérations génétiques précédentes seront exécutées autant de fois que nécessaire, et à chaque fois on obtient une nouvelle population ayant des individus de plus en plus adaptés. Le critère d'arrêt peut être choisi par plusieurs façons, soit par le test de la fonction coût, soit par le nombre de générations, soit par le test de changements dans la population. Si le nombre de générations est fixé, on obtient un temps d'exécution fixe mais le programme peut ne pas s'arrêter à la précision souhaitée. Si l'on connaît la valeur maximale de la fonction fitness, le programme s'arrêtera une fois il atteint cette valeur, on obtient la précision souhaitée, mais cela peut exiger un temps de calcul non contrôlé.

IV.6 Influence des paramètres des AG

Les opérations de l'algorithme génétique sont guidées par un certain nombre de paramètres fixés à l'avance. La valeur de ces paramètres influe sur la réussite d'un algorithme génétique. Ces paramètres sont les suivants :

IV.6.1 La taille de la population N

Si N est trop grand le temps de calcul de l'algorithme peut s'avérer très important, et si N est trop petit, il peut converger trop rapidement vers un mauvais chromosome (pas de diversité).

IV.6.2 La longueur des individus S

La longueur de chaque individu représente la précision des solutions. Plus S est grand, plus on gagne de précision, et on perd dans le temps de calcul et la taille mémoire nécessaire.

IV.6.3 La probabilité de croisement p_c

Plus elle est élevée, plus la population subit de changements importants. Les valeurs généralement admises sont comprises entre 0,5 et 0,9.

IV.6.4 La probabilité de mutation p_m

Ce taux est généralement faible, puisque un taux élevé risque de conduire à une solution sous optimale. Nous avons examiné les principales caractéristiques des algorithmes génétiques qui les rendent des outils simples et efficaces pour explorer un espace de performance donné.

Dans ce qui suit, les AG seront utilisés pour estimer les paramètres d'un modèle ARMA et améliorer ainsi sa performance pour la prédiction de l'irradiation solaire globale (GHI).

IV.7 Optimisation des modèles ARMA

Les modèles Box-Jenkins (modèles Autoregressifs) sont des méthodes de prédiction très populaires en raison de leurs bonnes capacités de prédiction pour différents types de données. [141], ont mis au point un nouveau modèle appelé ARMA (modèle de moyenne mobile autorégressif) en 1994, qui est appliqué aux problèmes de prédictions dans le cas des séries temporelles stationnaires.

Le modèle ARMA est une combinaison des modèles AR et MA. Dans le cas des séries non stationnaires, les données doivent d'abord être transformées en une forme stationnaire en introduisant l'opérateur de différentiation. C'est la première condition pour construire un modèle à moyenne mobile autorégressif (ARMA). Ici, la valeur future d'une variable est supposée être une fonction linéaire de plusieurs facteurs, plus précisément, les observations passées et les erreurs aléatoires. Généralement, un modèle ARMA est défini par l'équation suivante comme il a été détaillé dans le chapitre III:

$$y_t = \varepsilon_t + (a_1 y_{t-1} + a_2 y_{t-2} + \dots + a_p y_{t-p}) + (b_0 \varepsilon_t + b_1 \varepsilon_{t-1} + \dots + b_q \varepsilon_{t-q}). \quad (\text{IV.8})$$

où y_t est la valeur réelle, ε_t est une erreur aléatoire à la période t . a_i ($i = 1, 2, \dots, p$) et b_j ($j =$

0,1,2, ... q) sont les paramètres du modèle. L'ordre d'un modèle ARMA est généralement désigné par la notation ARMA (p, q), où p est l'ordre de la partie autorégressive (AR), q est l'ordre du processus de moyenne mobile (MA).

Le choix du meilleur modèle ARMA peut être effectué en trois étapes principales: premièrement, l'étape d'identification qui choisit le modèle le mieux adapté à la fonction d'auto-corrélation (ACF) et la fonction d'auto-corrélation partielle (PACF) de la série chronologique. Si les ACF de la série temporelle disparaissent rapidement, nous considérons la série temporelle comme étant stationnaire. Inversement, si la courbe ACF décline très lentement, la série chronologique peut être considérée comme non stationnaire.

La figure IV.10 illustre la sélection du meilleur modèle ARMA basé sur les comportements d'ACF et de PACF. La deuxième étape consiste à estimer les paramètres du modèle et enfin appliquer la méthode de Box-Jenkins pour vérifier la validité statistique du modèle estimé.

En règle générale, cela ne suffit pas pour déterminer l'ordre du modèle en utilisant uniquement le PACF. Ainsi, nous utilisons différents critères pour choisir l'ordre optimal. Ici, nous utilisons uniquement le critère d'information Akaike (AIC) [142], considéré comme le plus utilisé, qui est défini comme suit:

$$AIC = \log(v) + 2m/N \quad (VI.11)$$

Où $m = p + q$ et v représente la fonction de vraisemblance. Les résultats des calculs de l'AIC sont illustrés sur la figure IV.10.

IV.7.1 Formulation d'estimation des paramètres de l'ordre d'ARMA à l'aide d'AG

L'algorithme génétique est un outil de recherche puissant pour résoudre les problèmes d'optimisation de manière optimale. Il estime l'ordre et les paramètres du modèle ARMA comme illustré dans la figure IV.7. Pour commencer la procédure d'estimation, il faut d'abord définir une fonction objective et effectuer la recherche conformément à cette fonction objective.

Comme notre objectif est de produire un modèle ARMA optimisé, la fonction objective vise à réduire la différence entre les modèles réels et optimisés. Par conséquent, la fonction objective, appelée fonction de fitness, peut être définie comme [143]:

$$\text{Fitness} = \frac{1}{1+RMSE} \quad (IV.9)$$

Où

$$RMSE = \sqrt{\frac{1}{N} \sum_0^N (y_t - \hat{y}_t)^2} \tag{IV.10}$$

Ici, RMSE est l'erreur quadratique moyenne, y_t est la sortie réelle, \hat{y}_t est la sortie estimée et N est le nombre d'observations. Pour le modèle ARMA donné dans l'équation (1), le signal solaire est généralement enregistré avec un bruit blanc additif gaussien.

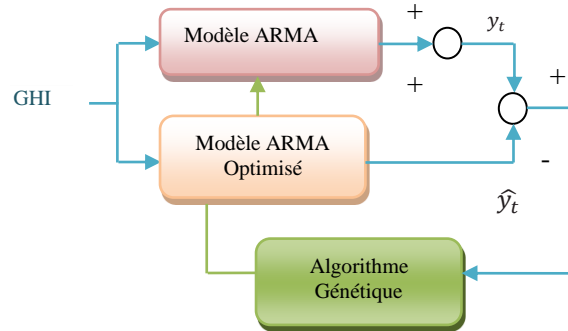


FIGURE IV.7 Architecture de l'AG pour l'estimation du modèle ARMA [139].

Les étapes de construction des modèles ARMA sont illustrées dans la figure VI.8.

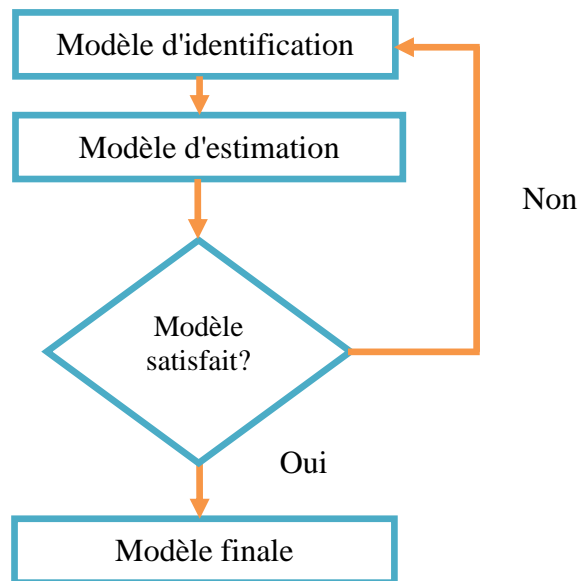


FIGURE IV.8 Les étapes de construction du modèle ARMA.

IV.8 Résultats d'identification

Les données d'entrée de l'algorithme proposé sont divisées en deux parties: les paramètres liés à l'AG et les paramètres relatifs au modèle ARMA. Les paramètres liés à l'AG inclus dans notre simulation sont les suivants:

La taille de la population, $N_p=80$, la probabilité de croisement, $P_c=0,65$, la probabilité

de mutation, $P_m=0,3$ pour le site de Ny_Ålesund. En ce qui concerne le site de Brasilia, la taille de la population a été fixée à $N_p = 65$, la probabilité de croisement $P_c = 0,7$, la probabilité de mutation $P_m = 0,3$. Ici, l'analyse met l'accent sur le RMSE moyen et le nombre moyen de générations nécessaires pour la convergence.

En raison de la nature stochastique des AG, différentes exécutions sont effectuées avec un nombre arbitraire de générations aléatoires.

L'analyse liée au modèle ARMA concerne la détection d'ordre et l'optimisation des paramètres. L'analyse d'optimisation des paramètres est axée sur la convergence des paramètres du modèle ARMA. Dans l'algorithme proposé, la première étape de l'opération consiste à détecter le véritable ordre. Pour examiner l'efficacité de la méthode, différents tests ont été effectués pour détecter l'ordre réel avant d'estimer les paramètres opérationnels.

Le tableau IV.2 donne les résultats de prévision des modèles, selon les critères d'erreurs suivant: MAPE, RMSE, NMSE et R^2 , le modèle AG-ARMA fonctionne mieux que l'ARMA classique.

TABLEAU IV.2 Résultats de la prédiction pour les deux sites de tests obtenus avec les modèles ARMA, AG-ARMA et leur combinaison avec la méthode OP. Les meilleurs résultats sont en gras.

<i>Méthodes</i>	<i>Sites</i>	<i>R²</i>	<i>RMSE</i>	<i>MAPE</i>	<i>NMSE</i>
ARMA	Ny_Ålesund	85.57	47.46	19.56	0.13
	Brasilia	60.44	211.19	35.36	0.42
AG_ARMA	Ny_Ålesund	85.90	39.140	18.47	0.12
	Brasilia	61.92	199.71	33.94	0.38
OP	Ny_Ålesund	85.92	38.24	18.15	0.11
	Brasilia	62.35	198.99	33.65	0.37

IV.8.1 Interprétation des résultats

D'après les résultats expérimentaux obtenus, il est clair que le modèle ARMA estimé par AG donne de meilleurs résultats par rapport au modèle ARMA classique, ceci est en raison de sa structure dynamique, adaptative et objective. Séparément, les modèles ARMA n'ont pas été en mesure de montrer l'efficacité des prévisions GHI en raison de structures traditionnelles et de jugements subjectifs sur la détermination des paramètres ARMA. Les modèles combinés AG-ARMA répondent à ces lacunes en termes d'obtention de meilleures solutions et de prévisions cohérentes. Comme le montre la figure IV.9.

De plus, nous pouvons améliorer davantage les résultats de prévision par la

proposition d'une autre méthode basée sur la combinaison des sorties des deux modèles obtenus. Nous pouvons ainsi clairement observer que, pour chaque série temporelle de GHI, les méthodes de combinaison génèrent les plus petites erreurs, et par conséquent, les meilleures précisions en termes de mesures d'erreur. Nous trouvons donc que les méthodes combinées ont une précision de prédiction globalement supérieure à celle des méthodes individuelles basées sur la prédiction de GHI par ARMA et AG-ARMA.

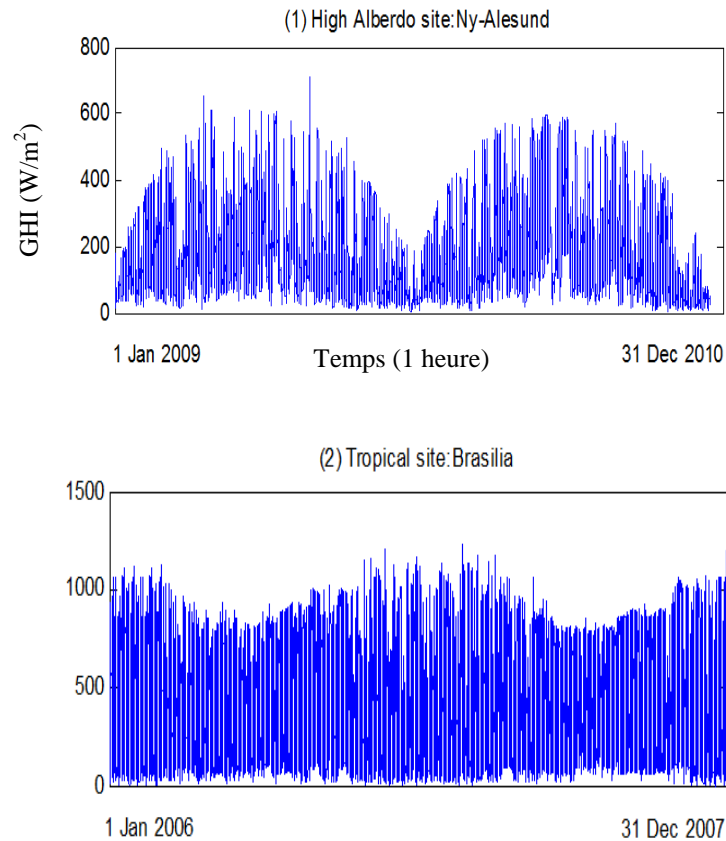
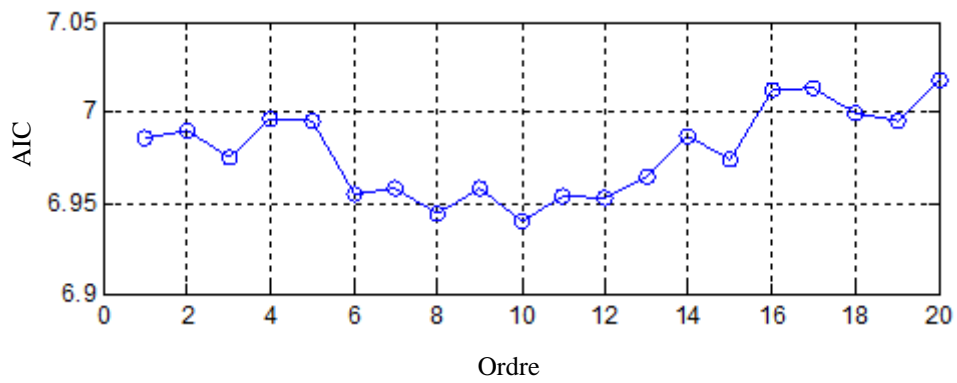
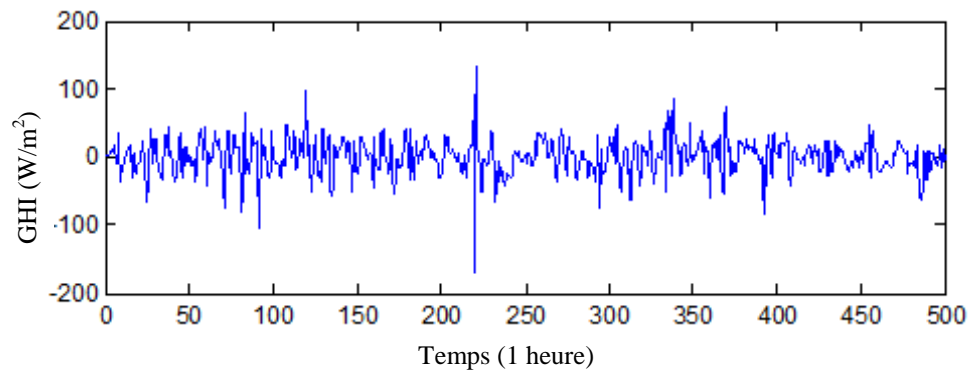
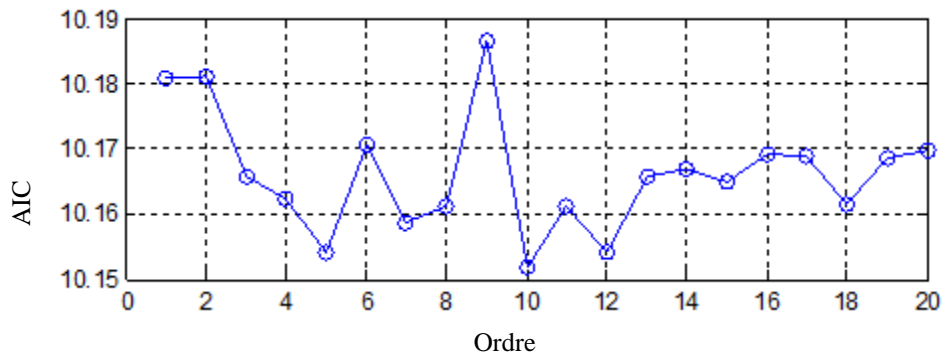
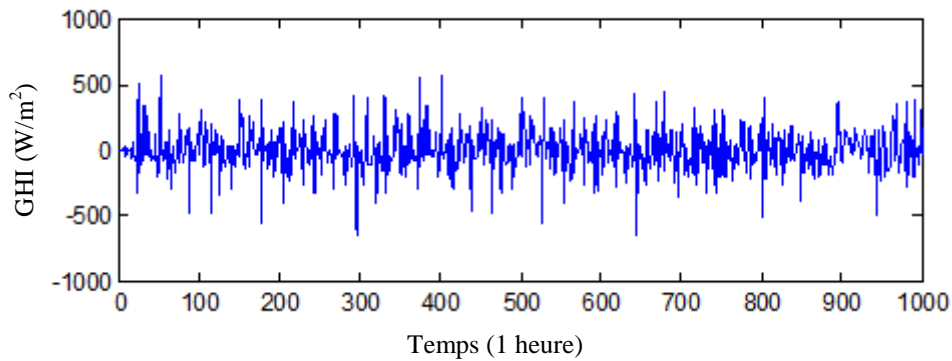


FIGURE IV.9 Série chronologique de données GHI sur deux années successives, pour les sites Brasilia et Ny-Ålesund.



(a)



(b)

FIGURE IV.10 Critère AIC pour les sites : (a) Brasilia, (b) Ny-Ålesund.

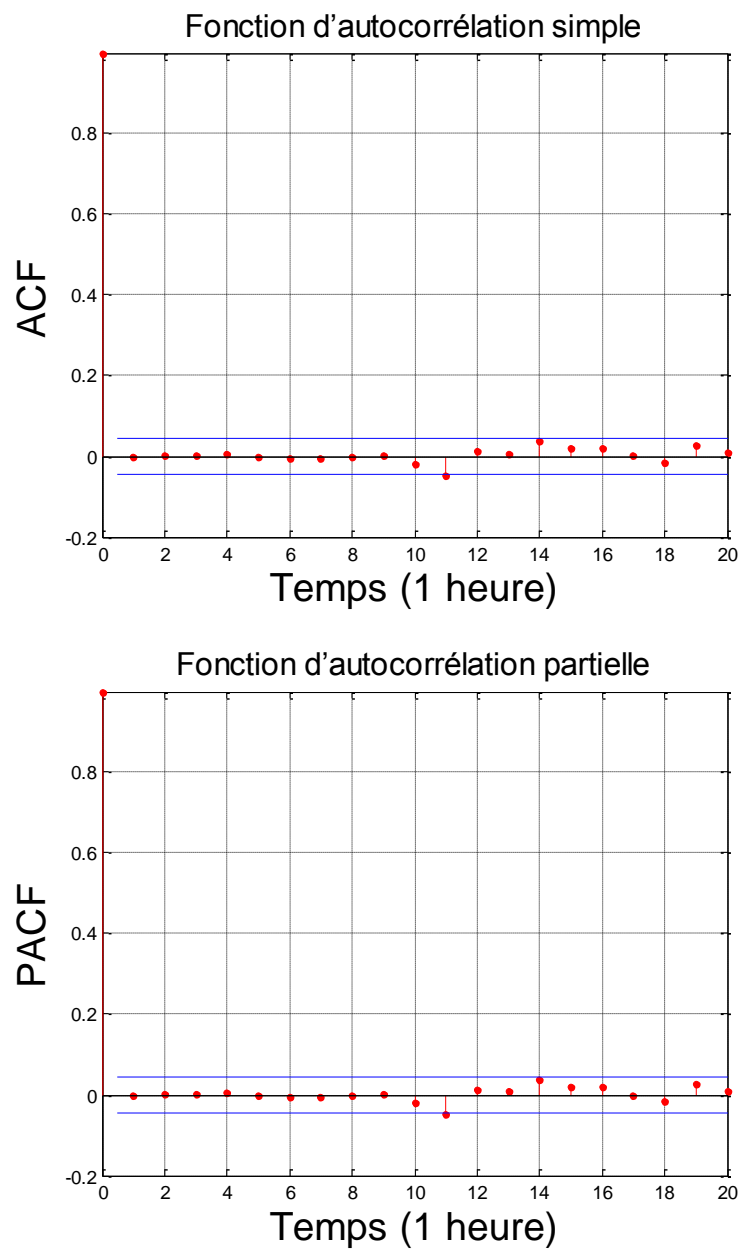


FIGURE IV.11 ACF et PACF des données solaire du site Ny-Ålesund.

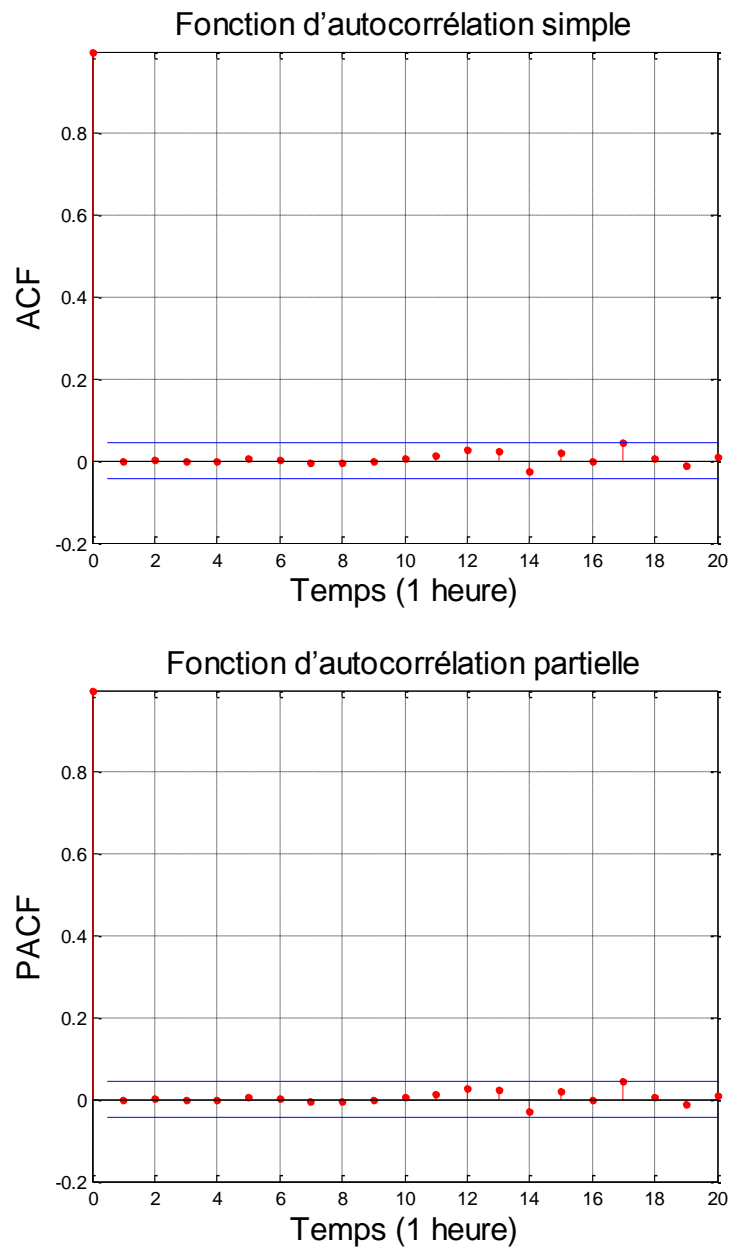


FIGURE IV.12 ACF et PACF des données solaire du site Brasilia

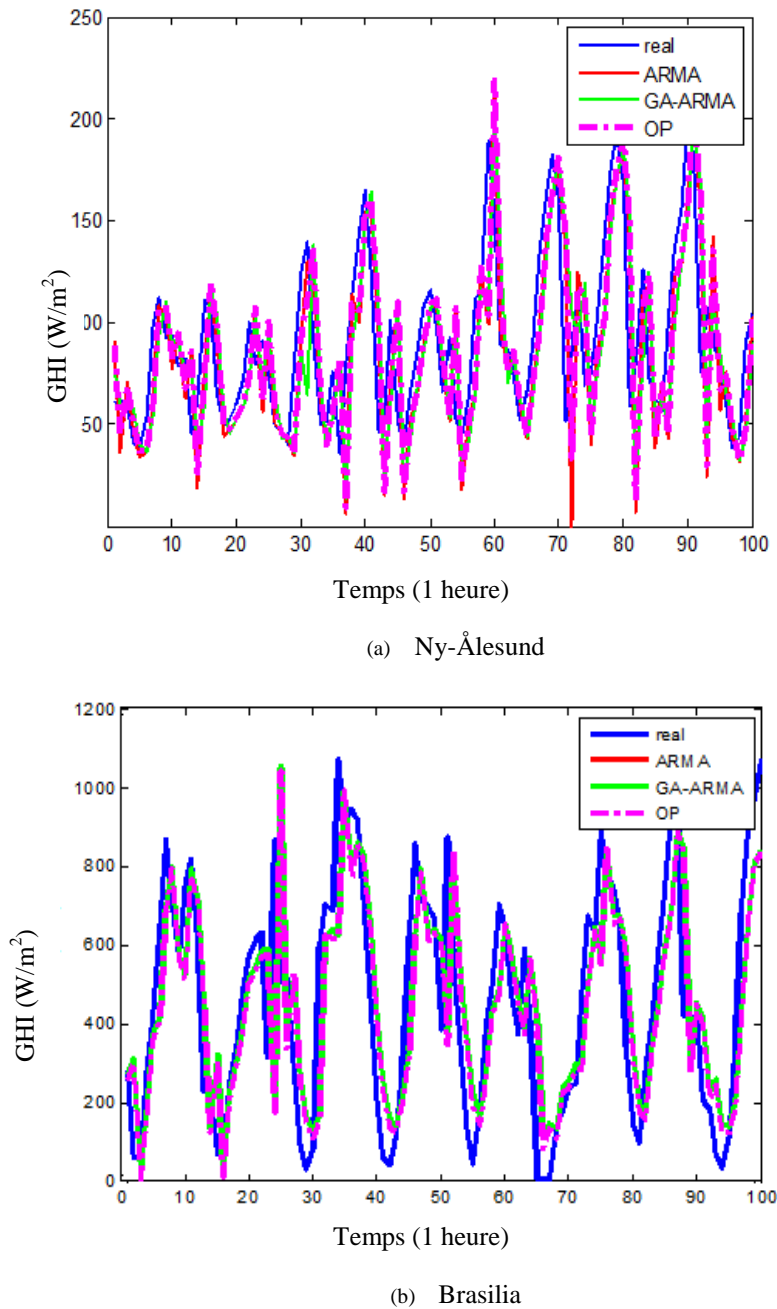


FIGURE IV.13 Comparaison des différents modèles sur les deux stations de test.

IV.8.2 Interprétations des courbes

Les tests effectués sur les deux sites (Brasilia et Ny-Alesund), montrent que les algorithmes génétiques constituent un outil puissant pour l'amélioration des résultats via l'optimisation des paramètres du modèle ARMA. Les critères utilisés, dans notre cas, pour juger la performance des AG sont : RMSE, NRMSE, R^2 et MAPE.

Les algorithmes génétiques ont donné la possibilité de déterminer l'ordre et le nombre des paramètres à inclure dans un modèle ARMA automatiquement par la sélection du meilleur chromosome adéquat pour le signal étudié, ainsi que l'estimation supplémentaire de la valeur des paramètres par ajustement.

Enfin nous déduisons que les AG promettent d'être un moyen efficace et robuste au service d'autres algorithmes de prédiction classiques.

VI.9. Conclusion

Dans ce chapitre, une approche AG pour l'optimisation des paramètres du modèle ARMA est introduite comme nouvelle méthodologie de prédiction pour les séries temporelles solaires de GHI. L'algorithme proposé joue un rôle efficace dans l'optimisation de la technique autorégressive ARMA. La fonction fitness calcule uniquement le RMSE, ce qui signifie qu'il n'y a pas de calculs complexes (dérivant des équations pour le processus d'estimation) dans le processus d'évaluation. L'algorithme minimise l'erreur dans l'estimation des paramètres et fournit de meilleurs résultats dans la détection de l'ordre du modèle par rapport à l'ARMA classique. L'algorithme proposé peut tout d'abord, détecter le vrai ordre dans un premier temps, puis estimer les paramètres en fonction de la valeur d'ordre obtenue dans un deuxième temps.

Il a été constaté que la technique de choix de l'ordre du modèle AG-ARMA et l'estimation des paramètres peuvent donner de bons résultats par rapport aux méthodes ARMA classiques. En outre, l'ajout d'une autre étape de calcul en combinant les deux modèles de prédiction peut améliorer davantage les performances de prédiction.

Conclusion Générale

1. Contributions

Cette thèse explore le problème de la prédiction du rayonnement solaire par deux méthodologies différentes et indépendantes utilisées pour différents types de données de GHI.

Comme nous l'avons vu dans l'introduction, l'intégration d'énergies renouvelables intermittentes à un réseau électrique est très délicate, et particulièrement dans le cas de l'énergie solaire. Une possibilité en vue de cette intégration, réside dans la caractérisation de ces différentes ressources.

Cette étude s'inscrit donc, dans la recherche de meilleures techniques de prédiction de la ressource solaire afin de quantifier l'énergie disponible pour en faciliter la gestion au sein d'un réseau. Nous avons vu dans le chapitre 2 que cette démarche essentielle est tributaire de différents paramètres comme, notamment, l'horizon considéré, la nature des données, la sélection des variables à utiliser, le calcul de l'erreur,...etc. Nous avons essayé de tester différentes techniques de prédiction, notamment, les méthodes autorégressive, les méthodes neuronale, les machines à vecteur de support et les méthodes de combinaison. Toutes ces méthodes sont conçues et testées sur différents sites avec un climat différent dans chaque cas.

Pour la première partie, le grand succès des systèmes d'ensembles, nous a permis d'étendre l'application du système d'ensembles pour résoudre le problème de la prédiction du rayonnement solaire, rencontré dans le domaine d'énergies renouvelables et de météorologie. La création de prédictors précis à partir d'un ensemble d'exemples est extrêmement importante pour différents problèmes d'apprentissage automatique. La combinaison de prédictors a été proposée comme un moyen très efficace afin d'améliorer les performances de généralisation et de précision. Leur compréhension conduit à de nouvelles stratégies, à une meilleure identification des stratégies existantes et à une caractérisation des régions où elles fonctionnent le mieux.

Dans ce contexte, une nouvelle approche a été développée pour traiter ce type de problèmes, elle propose de combiner plusieurs techniques de prévision avec différentes

familles (statistiques, neuronales et à noyau) pour résoudre le problème de la prédiction de GHI.

L'évaluation expérimentale des méthodologies a été réalisée sur des bases de données réelles avec des types de climat radiatif distincts : Aride, Tempéré, Tropical, et High-Albedo, qui favorisent différents régimes de nuages et différents schémas saisonniers. Les résultats obtenus montrent la performance exceptionnelle des stratégies de combinaison par rapport aux prédicteurs individuels, ce qui a été confirmée par la stratégie de classification.

Dans la deuxième partie de cette thèse, nous avons proposé une autre méthode de prédiction intégrant les algorithmes génétiques AG et la méthode de prédiction à moyenne mobile autorégressive ARMA afin de tirer profit de la force intrinsèque des deux modèles. L'approche présentée est proposée pour prédire l'irradiation solaire globale dans le cas de l'horizon de prédiction d'une heure, la méthodologie a été appliquée sur deux sites de test qui sont (Brasilia et Ny-Alesund), où nous avons trouvé de meilleurs résultats comparativement avec ARMA sans estimation ou optimisation.

2. Perspectives

Le problème de la prévision du rayonnement solaire a été discuté et analysé de manière approfondie dans cette thèse, notamment par la proposition de deux méthodologies différentes.

Cette thèse constitue une synthèse et une extension utile de la littérature actuelle en matière de prédiction du rayonnement solaire. Nous espérons que les chercheurs trouveront une utilité dans les méthodes proposées et découvriront de nouvelles orientations stimulantes basées sur les lignes directrices définies dans ce qui suit :

- 1- Exploration d'autres techniques de prédiction comme l'apprentissage profond (Deep learning);
- 2- Estimer et prédire la production d'énergie produite dans le cas des centrales solaires pour bien gérer son intégration dans les réseaux électriques intelligents ;
- 3- Utiliser et explorer d'autres moyens de prétraitement des séries solaires brutes telles que : les algorithmes de décomposition, les algorithmes de sélection de caractéristiques...etc.

Bibliographie

- [1] Jannot, Yves. "Cours thermique solaire."
URL : <http://www.thermique55.com/principal/thermiquesolaire.pdf>. Accédé le 20/11/2020
- [2] Mondol, Jayanta Deb, Yigzaw G. Yohanis, and Brian Norton. "Solar radiation modelling for the simulation of photovoltaic systems." *Renewable Energy* 33.5 (2008): 1109-1120.
- [3] M. Salmi, H. Bouzgou, and H. Boursas, "A comparison study of solar energy application to arabic countries." 3rd Conference of Basic Science, 2009. University of Aljabal Algharbi, Lybia, 2009.
- [4] Albert. Einstein, "The photoelectric effect." *Ann. Phys* 17.132 (1905): 4.
- [5] M. A Atwater, and J. T. Ball. "A numerical solar radiation model based on standard meteorological observations." *Solar Energy*, 21.3 (1978): 163-170.
- [6] L. Yong, and X. Yao. "Ensemble learning via negative correlation." *Neural networks* 12.10 (1999): 1399-1404.
- [7] Gueymard, Christian A. "REST2: High-performance solar radiation model for cloudless-sky irradiance, illuminance, and photosynthetically active radiation– Validation with a benchmark dataset." *Solar Energy* 82.3 (2008): 272-285.
- [8] Oumbe, Armel. *Exploitation des nouvelles capacités d'observation de la Terre pour évaluer le rayonnement solaire incident au sol*. Diss. Thèse de doctorat. Ecole Nationale Supérieure des Mines de Paris, 2009.
- [9] Malardel, Sylvie. *Fondamentaux de météorologie*. Cépadues Editions, 2009.
- [10] Kopp, Greg, and Judith L. Lean. "A new, lower value of total solar irradiance: Evidence and climate significance." *GeophysicalResearchLetters* 38.1 (2011).
- [11] Jannot, Yves. "Cours thermique solaire." *Mars* (2011).
- [12] <http://nicolaselie.free.fr/tpe/html/mouvements.html>. Consulté le 01 juin 2020.
- [13] Mohammed, Bencherif. "Modélisation des systèmes énergétiques photovoltaïques et éoliens: intégration dans un système hybride à basse tension." Mémoire de Doctorat, Université Abou-BekrBelkaid Tlemcen (2014)
- [14] Fleming, Sandford. *Report in Reference to the Canadian Pacific Railway*. Ottawa, MacLean, 1879.

Bibliographie

- [15] Gueymard, Christian. "Critical analysis and performance assessment of clear sky solar irradiance models using theoretical and measured data." *Solar Energy* 51.2 (1993): 121-138.
- [16] Ch.Chauliaguet, P. Baratcabal, J. P. Batelier, "L'énergie solaire dans le bâtiment" Ed industrie Paris (1975).
- [17] De Brichambaut, Christian Perrin, and Guy Lambole. Le rayonnement solaire au sol et ses mesures. Thermique et industrie, 1968.
- [18] Tadili, R., and M. N. Bargach. "Une méthode d'estimation du rayonnement solaire global reçu par une surface inclinée: Application aux sites marocains." *La météorologie* (2005).
- [19] Bessemoulin, Pierre, and Jean Oliviéri. "Le rayonnement solaire et sa composante ultraviolette." *La météorologie* (2000).
- [20] F. Codron, "Transfert radiatif bilan énergétique", Note du cours, Université Pierre et Marie Curie-Paris 6, France, 2012.
- [21] Mathieu, Henry, and Hervé Fanet. *Physique des semiconducteurs et des composants électroniques*. Vol. 5. Paris:Dunod, 2001.
- [22] <http://www.photovoltaique.info/L-effet-photovoltaique.html>, dernier accès Juin 2019.
- [23] Mondol, Jayanta Deb, Yigzaw G. Yohanis, and Brian Norton. "Solar radiation modelling for the simulation of photovoltaic systems." *Renewable Energy* 33.5 (2008): 1109-1120.
- [24] Enteria, Napoleon, and Aliakbar Akbarzadeh, eds. *Solar energy sciences and engineering applications*. CRC Press, 2013.
- [25] Lange, M. "A mature market? The history of short-term prediction services. POWWOW Best-Practises Workshop, Delft, 2006." (2006).
- [26] Landberg, Lars, et al. "Short-term prediction—an overview." *Wind Energy: An International Journal for Progress and Applications in Wind Power Conversion Technology* 6.3 (2003): 273-280.
- [27] Wittmann, Michael, et al. "Case studies on the use of solar irradiance forecast for optimized operation strategies of solar thermal power plants." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 1.1 (2008): 18-27.
- [28] Traunmüller, Wolfgang, and Gerald Steinmaurer. "Solar irradiance forecasting, benchmarking of different techniques and applications of energy meteorology." *Proceedings of the EuroSun 2010 conference*. 2010.

Bibliographie

- [29] Bourbonnais, Régis, and Michel Terraza. *Analyse des séries temporelles en économie*. FeniXX, 1998.
- [30] Bourbonnais, Régis, and Michel Terraza. *Analyse des séries temporelles: Applications à l'économie et à la gestion*. 2008.
- [31] Simon, Geoffroy. *Méthodes non linéaires pour séries temporelles: prédiction par Double Quantification Vectorielle et sélection du délai en hautes dimensions*. Presses univ. de Louvain, 2007.
- [32] Hamilton, James D., and Raul Susmel. "Autoregressive conditional heteroskedasticity and changes in regime." *Journal of econometrics* 64.1-2 (1994): 307-333.
- [33] Mueller, R. W., et al. "Rethinking satellite-based solar irradiance modelling: The SOLIS clear-sky module." *Remotesensing of Environment* 91.2 (2004): 160-174.
- [34] Rigollier, Christelle, Olivier Bauer, and Lucien Wald. "On the clear sky model of the ESRA—European Solar Radiation Atlas—with respect to the Heliosat method." *Solar energy* 68.1 (2000): 33-48.
- [35] Ineichen, Pierre, and Richard Perez. "A new airmass independent formulation for the Linke turbidity coefficient." *Solar Energy* 73.3 (2002): 151-157.
- [36] Gueymard, Christian A. "REST2: High-performance solar radiation model for cloudless-sky irradiance, illuminance, and photosynthetically active radiation—Validation with a benchmark dataset." *Solar Energy* 82.3 (2008): 272-285.
- [37] Inman, Rich H., Hugo TC Pedro, and Carlos FM Coimbra. "Solar forecasting methods for renewable energy integration." *Progress in energy and combustion science* 39.6 (2013): 535-576.
- [38] Badescu, Viorel, et al. "Accuracy analysis for fifty-four clear-sky solar radiation models using routine hourly global irradiance measurements in Romania." *Renewable Energy* 55 (2013): 85-103.
- [39] Ineichen, Pierre. "Comparison of eight clear sky broadband models against 16 independent data banks." *Solar Energy* 80.4 (2006): 468-478.
- [40] Marquez, Ricardo, and Carlos FM Coimbra. "Proposed metric for evaluation of solar forecasting models." *Journal of solarenergy engineering* 135.1 (2013).
- [41] Lipperheide, Moritz, J. L. Bosch, and Jan Kleissl. "Embedded nowcasting method using cloud speed persistence for a photovoltaic power plant." *Solar Energy* 112 (2015): 232-238.
- [42] Lonij, Vincent PA, et al. "Intra-hour forecasts of solar power production using measurements from a network of irradiance sensors." *Solar Energy* 97 (2013): 58-66.

Bibliographie

- [43] Amral, N., C. S. Ozveren, and D. King. "Short term load forecasting using multiple linear regression." *2007 42nd International universities power engineering conference*. IEEE, 2007.
- [44] Zhang, Wen-Yu, et al. "Short term wind speed forecasting for wind farms using an improved autoregression method." *2011 International Conference of Information Technology, Computer Engineering and Management Sciences*. Vol. 1. IEEE, 2011.
- [45] Wang, Zhenyu, et al. "Hourly solar radiation forecasting using a volterra-least squares support vector machine model combined with signal decomposition." *Energies* 11.1 (2018): 68.
- [46] C.Voyant, "Prédiction de séries temporelles de rayonnement solaire global et de production d'énergie photovoltaïque à partir de réseaux de neurones artificiels," vol. 6134, p. 257, 2011.
- [47] Li, Yanting, Yan Su, and Lianjie Shu. "An ARMAX model for forecasting the power output of a grid connected photovoltaic system." *RenewableEnergy* 66 (2014): 78-89.
- [48] Bacher, Peder, Henrik Madsen, and Henrik Aalborg Nielsen. "Online short-term solar power forecasting." *Solar energy* 83.10 (2009): 1772-1783.
- [49] Oudjana, S. Hamid, A. Hellal, and I. HadjMahamed. "Short term photovoltaic power generation forecasting using neural network." *2012 11th International Conference on Environment and Electrical Engineering*. IEEE, 2012.
- [50] Brown, Robert G. "Exponential smoothing for predicting demand." *Operations Research*. Vol. 5. No. 1. 1957.
- [51] Chatfield, Chris. "The Holt-winters forecasting procedure." *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 27.3 (1978): 264-279..
- [52] Tikunov, Denis, and Toshikazu Nishimura. "Traffic prediction for mobile network using Holt-Winter's exponential smoothing." *2007 15th International Conference on Software, Telecommunications and Computer Networks*. IEEE, 2007.
- [53] Monteiro, Claudio, et al. "Short-term forecasting models for photovoltaic plants: Analytical versus soft-computing techniques." *Mathematical problems in engineering* 2013 (2013).
- [54] Dolara, Alberto, et al. "A physical hybrid artificial neural network for short term forecasting of PV plant power output." *Energies* 8.2 (2015): 1138-1153.
- [55] Soman, Saurabh S., et al. "A review of wind power and wind speed forecasting methods with different time horizons." *North American Power Symposium 2010*. IEEE, 2010.

Bibliographie

- [56] Gandelli, Alessandro, et al. "Hybrid model analysis and validation for PV energy production forecasting." *2014 international joint conference on neural networks (IJCNN)*. IEEE, 2014.
- [57] Hossain, Md Sazzad, et al. "Artificial neural networks for vibration based inverse parametric identifications: A review." *Applied Soft Computing* 52 (2017): 203-219.
- [58] Mellit, Adel, and Soteris A. Kalogirou. "Artificial intelligence techniques for photovoltaic applications: A review." *Progress in energy and combustion science* 34.5 (2008): 574-632.
- [59] Aminzadeh, Fred, and Paul De Groot. *Neural networks and other soft computing techniques with applications in the oil industry*. Eage Publications, 2006.
- [60] Serra, Pedro Henrique Cardeal. "Short-Term Forecasting of Photovoltaic Power Plants.", Master thesis, University of Porto(2014).
- [61] A .Yona, T.Senju, AY.Saber, T.Funabashi, H.Sekine, and C. Kim ,“ Application of neural network to one-day-ahead 24h generating power forecasting for photovoltaic system ,” In: Proceedings of international conference on intelligent systems applications to power systems, ISAP 2007. IEEE; 2007. p. 1–6, 2007 .
- [62] Senju, Tomonobu. "Application of recurrent neural network to short-term-ahead generating power forecasting for wind power generator." *Proc. of International Conference on Electrical Engineering 2006, Jul. 9-13*. 2006.
- [63] Ibrić, Svetlana, et al. "Artificial neural networks in evaluation and optimization of modified release solid dosage forms." *Pharmaceutics* 4.4 (2012): 531-550.
- [64] AlHakeem, Donna, et al. "A new strategy to quantify uncertainties of wavelet-GRNN-PSO based solar PV power forecasts using bootstrap confidence intervals." *2015 IEEE Power & Energy Society General Meeting*. IEEE, 2015.
- [65] Bizzarri, Federico, et al. "Model of photovoltaic power plants for performance analysis and production forecast." *IEEE transactions on sustainable energy* 4.2 (2012): 278-285.
- [66] Smolensky, Paul, Michael C. Mozer, and David E. Rumelhart, eds. *Mathematical perspectives on neural networks*. Psychology Press, 2013.
- [67] Yun, Zhang, et al. "RBF neural network and ANFIS-based short-term load forecasting approach in real-time price environment." *IEEE Transactions on power systems* 23.3 (2008): 853-858.
- [68] Shamshirband, Shahaboddin, et al. "Developing an ANFIS-PSO model to predict mercury emissions in combustion flue gases." *Mathematics* 7.10 (2019): 965.
- [69] Bayoumi, Ehab HE, Mohammed A. Awadallah, and Hisham M. Soliman. "Robust control of switched reluctance motor drives using Kharitonov theorem and swarm intelligence." *International Journal of Modelling, Identification and Control* 7.3

Bibliographie

- (2009): 234-245.
- [70] Vapnik, Vladimir. "Pattern recognition using generalized portrait method." *Automation and remote control* 24 (1963): 774-780.
- [71] Vapnik, Vladimir. "Support-vector networks." *Machine learning* 20 (1995): 273-297.
- [72] Rana, Mashud, Irena Koprinska, and Vassilios G. Agelidis. "2D-interval forecasts for solar power production." *Solar Energy* 122 (2015): 191-203.
- [73] da Silva Fonseca Jr, Joao Gari, et al. "Use of support vector regression and numerically predicted cloudiness to forecast power output of a photovoltaic power plant in Kitakyushu, Japan." *Progress in photovoltaics: Research and applications* 20.7 (2012): 874-882.
- [74] Pavlov, YuryLeonidovich. "Random forests." *Probabilistic methods in discrete mathematics (Petrozavodsk, 1996)* (1997): 11-18.
- [75] Breiman, Leo. "Manual on setting up, using, and understanding random forests v3. 1." *Statistics Department University of California Berkeley, CA, USA* 1 (2002): 58.
- [76] Zamo, M., et al. "A benchmark of statistical regression methods for short-term forecasting of photovoltaic electricity production, part I: Deterministic forecast of hourly production." *Solar Energy* 105 (2014): 792-803.
- [77] Antonanzas, Javier, et al. "Review of photovoltaic power forecasting." *Solar Energy* 136 (2016): 78-111.
- [78] Bouzerdoum, Moufida, Adel Mellit, and A. MassiPavan. "A hybrid model (SARIMA–SVM) for short-term power forecasting of a small-scale grid-connected photovoltaic plant." *Solar Energy* 98 (2013): 226-235.
- [79] Ramsami, Pamela, and VishwamitraOree. "A hybrid method for forecasting the energy output of photovoltaic systems." *Energy Conversion and Management* 95 (2015): 406-413.
- [80] Nazari-pouya, Hamidreza, et al. "Univariate time series prediction of solar power using a hybrid wavelet-ARMA-NARX prediction method." *2016 IEEE/PES Transmission and Distribution Conference and Exposition (T&D)*. IEEE, 2016.
- [81] Hossain, Monowar, et al. "Application of extreme learning machine for short term output power forecasting of three grid-connected PV systems." *Journal of Cleaner Production* 167 (2017): 395-405.
- [82] Gandelli, Alessandro, et al. "Hybrid model analysis and validation for PV energy production forecasting." *2014 International joint conference on neural networks (IJCNN)*. IEEE, 2014.

Bibliographie

- [83] Ji, Wu, and Keong Chan Chee. "Prediction of hourly solar radiation using a novel hybrid model of ARMA and TDNN." *Solar Energy* 85.5 (2011): 808-817.
- [84] Opitz, David, and Richard Maclin. "Popular ensemble methods: An empirical study." *Journal of artificial intelligence research* 11 (1999): 169-198.
- [85] Peter Manousos, « Ensemble Prediction Systems » [archive], WeatherPrediction Center, 19 juillet 2009 (consulté Juin 2020).
- [86] Ye, Ren, and Ponnuthurai N. Suganthan. "A kernel-ensemble bagging support vector machine." *2012 12th International Conference on Intelligent Systems Design and Applications (ISDA)*. IEEE, 2012.
- [87] Kuncheva, Ludmila I., and Christopher J. Whitaker. "Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy." *Machine learning* 51.2 (2003): 181-207.
- [88] Hall, James, and Jeffrey Hall. "Forecasting solar radiation for the Los Angeles basin—phase II report." *American Solar Energy Society annual SOLAR 2011 Conference Proceedings*. 2011.
- [89] Breiman, Leo. "Bagging predictors." *Machine learning* 24.2 (1996): 123-140.
- [90] Chaouachi, Aymen, et al. "Neural network ensemble-based solar power generation short-term forecasting." *World Academy of Science, Engineering and Technology* 54 (2009): 54-59.
- [91] Sözen, Adnan, ErolArcaklioğlu, and Mehmet Özalp. "Estimation of solar potential in Turkey by artificial neural networks using meteorological and geographical data." *Energy Conversion and Management* 45.18-19 (2004): 3033-3052.
- [92] Liu, Hui, et al. "A hybrid statistical method to predict wind speed and wind power." *Renewable energy* 35.8 (2010): 1857-1861.
- [93] Y.Ren, PN.Suganthan, "EMD-Knn models for wind speed forecasting,". *J Power Energy Eng*, 2(4):176–85, 2014.
- [94] Ren, Ye, P. N. Suganthan, and N. Srikanth. "Ensemble methods for wind and solar power forecasting—A state-of-the-art review." *Renewable and Sustainable Energy Reviews* 50 (2015): 82-91.
- [95] Cao, JCSH Cao, and S. H. Cao. "Study of forecasting solar irradiance using neural networks with preprocessing sample data by wavelet analysis." *Energy* 31.15 (2006): 3435-3445.
- [96] Lojowska, Alicja, et al. "Advantages of ARMA-GARCH wind speed time series modeling." *2010 IEEE 11th International Conference on Probabilistic Methods Applied to Power Systems*. IEEE, 2010.

Bibliographie

- [97] Shi, Jing, JinmeiGuo, and Songtao Zheng. "Evaluation of hybrid forecasting approaches for wind speed and power generation time series." *Renewable and SustainableEnergyReviews* 16.5 (2012): 3471-3480.
- [98] Ji, Wu, and Keong Chan Chee. "Prediction of hourly solar radiation using a novel hybrid model of ARMA and TDNN." *Solar Energy* 85.5 (2011): 808-817.
- [99] Ji, Guo-Rui, Pu Han, and Yong-JieZhai. "Wind speed forecasting based on support vector machine with forecasting error estimation." *2007 International Conference on Machine Learning and Cybernetics*. Vol. 5. IEEE, 2007.
- [100] Praene, Jean Philippe, et al. "Renewable energy: Progressing towards a net zero energy island, the case of Reunion Island." *Renewable and Sustainable Energy Reviews* 16.1 (2012): 426-442.
- [101] Zemouri, Nahed, Hassen Bouzgou, and Christian A. Gueymard. "Multimodel ensemble approach for hourly global solar irradiation forecasting." *The European Physical Journal Plus* 134.12 (2019): 594.
- [102] Diagne, Maimouna, et al. "Review of solar irradiance forecasting methods and a proposition for small-scale insular grids." *Renewable and Sustainable Energy Reviews* 27 (2013): 65-76.
- [103] Clemen, Robert T. "Combining forecasts: A review and annotated bibliography." *International journal of forecasting* 5.4 (1989): 559-583.
- [104] Zhou, Zhi-Hua. *Ensemble methods: foundations and algorithms*. CRC press, 2012.
- [105] De'Ath, Glenn. "Boosted trees for ecological modeling and prediction." *Ecology* 88.1 (2007): 243-251.
- [106] Liu, Hui, et al. "Comparison of four Adaboost algorithm based artificial neural networks in wind speed predictions." *Energy Conversion and Management* 92 (2015): 67-81.
- [107] Mondol, Jayanta Deb, Yigzaw G. Yohanis, and Brian Norton. "Solar radiation modelling for the simulation of photovoltaic systems." *RenewableEnergy* 33.5 (2008): 1109-1120.
- [108] Wijitcharoen, A., et al. "ICA-DEODA: An independent feature extraction model for stock index forecasting." *2016 International Computer Science and Engineering Conference (ICSEC)*. IEEE, 2016.
- [109] Wang, Lipo P., and Chunru R. Wan. "Comments on" The extreme learning machine." *IEEE Transactions on Neural Networks* 19.8 (2008): 1494-1495.
- [110] Huang, Guang-Bin, Qin-Yu Zhu, and Chee-Kheong Siew. "Extreme learning machine: theory and applications." *Neurocomputing* 70.1-3 (2006): 489-501.
- [111] Zemouri, Nahed, and Hassen Bouzgou. "Ensemble of support vector methods to

Bibliographie

- estimate global solar radiation in Algeria." *International Conference in Artificial Intelligence in Renewable Energetic Systems*. Springer, Cham, 2017.
- [112] Bouzgou, Hassen, and Christian A. Gueymard. "Minimum redundancy–Maximum relevance with extreme learning machines for global solar radiation forecasting: Toward an optimized dimensionality reduction for solar time series." *Solar Energy* 158 (2017): 595-609.
- [113] Vapnik, Vladimir, and Vladimir Vapnik. "Statistical learning theory Wiley." *New York* 1 (1998): 624.
- [114] A.Schölkopf, and K-R.Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural computation* 10.5: 1299-1319, 1998.
- [115] De Menezes, Lilian M., Derek W. Bunn, and James W. Taylor. "Review of guidelines for the use of combined forecasts." *European Journal of Operational Research* 120.1 (2000): 190-204.
- [116] Dietterich, Thomas G. "Approximate statistical tests for comparing supervised classification learning algorithms." *Neural computation* 10.7 (1998): 1895-1923.
- [117] Winkler, Robert L. "Combining forecasts: A philosophical basis and some current issues." *International Journal of Forecasting* 5.4 (1989): 605-609.
- [118] Armstrong, Michael. *A handbook of human resource management practice*. Kogan Page Publishers, 2006.
- [179] Lemke, Christiane, and Bogdan Gabrys. "Meta-learning for time series forecasting and forecast combination." *Neurocomputing* 73.10-12 (2010): 2006-2016.
- [120] Bouzgou, Hassen, and Nabil Benoudjit. "Multiple architecture system for wind speed prediction." *Applied Energy* 88.7 (2011): 2463-2471.
- [121] Bruzzone, Lorenzo, and Farid Melgani. "Robust multiple estimator systems for the analysis of biophysical parameters from remotely sensed data." *IEEE Transactions on Geoscience and Remote Sensing* 43.1 (2005): 159-174.
- [122] Gueymard, Christian A., and Jose A. Ruiz-Arias. "Extensive worldwide validation and climate sensitivity analysis of direct irradiance predictions from 1-min global irradiance." *Solar Energy* 128 (2016): 1-30.
- [123] Turkoglu, I., and A. Arslan. "An intelligent pattern recognition system based on neural network and wavelet decomposition for interpretation of heart sounds." *2001 Conference Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. Vol. 2. IEEE, 2001.
- [124] Perez, Richard, et al. "A new operational model for satellite-derived irradiances: description and validation." *Solar Energy* 73.5 (2002): 307-317.

Bibliographie

- [125] H. Bouzgou, "Advanced Methods for the Processing and Analysis of Multidimensional Signals: Application to Wind Speed," Doctorate Thesis. Université de Batna , 2012.
- [126] Taylor, Karl E. "Summarizing multiple aspects of model performance in a single diagram." *Journal of Geophysical Research: Atmospheres* 106.D7 (2001): 7183-7192.
- [127] Bouzgou, Hassen, and Christian A. Gueymard. "Fast short-term global solar irradiance forecasting with wrapper mutual information." *Renewable Energy* 133 (2019): 1055-1065.
- [128] Davis, Gary E., and Todd W. Anderson. "Population estimates of four kelp forest fishes and an evaluation of three in situ assessment techniques." *Bulletin of Marine Science* 44.3 (1989): 1138-1151.
- [129] Holland, John Henry. *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*. MIT press, 1992.
- [130] Nur'afifah, Binti Rusdi. *Parameter Estimation of Box-Jenkins Model Using Genetic Algorithm*. Master thesis. Universiti Teknologi Malaysia, 2013.
- [131] Nesrine, Talbi. "Conception des Systèmes d'Inférence Floue par des Approches Hybrides: Application pour la Commande et la Modélisation des Systèmes Non linéaires." *Mémoire de Master, Université Setif* (2014).
- [132] Eke, Didier Félix. *Méthode d'identification en virgule fixe d'un modèle non linéaire basé sur les algorithmes génétiques*. Doctorat thesis. Université du Québec à Trois-Rivières, 2008.
- [133] Durand, Nicolas, Jean-Marc Alliot, and Joseph Noailles. "Algorithmes génétiques: un croisement pour les problèmes partiellement séparables." *Proceedings of the Journées Evolution Artificielle Francophones*. 1994.
- [134] Yu, Feng, and Xiaozhong Xu. "A short-term load forecasting model of natural gas based on optimized genetic algorithm and improved BP neural network." *Applied Energy* 134 (2014): 102-113.
- [135] Hummeltenberg, Wilhelm, Decision Engineering. Springer Gabler, Wiesbaden, 175-199. 2014
- [136] Goldberg, D. E. "Genetic Algorithms in Search, Optimization, and Machine Learning. (1989) Addison Wesley." *Reading, Mass* (1989).
- [137] Es-Sadek, Mohamed Zeriab. *Contribution à l'optimisation globale: approche déterministe et stochastique et application*. Doctorate Thesis. Rouen, INSA, 2009.
- [138] Pal, Sankar K., Sanghamitra Bandyopadhyay, and C. A. Murthy. "Genetic

Bibliographie

- algorithms for generation of class boundaries." *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 28.6 (1998): 816-828.
- [139] Eke, Didier Félix. *Méthode d'identification en virgule fixe d'un modèle non linéaire basé sur les algorithmes génétiques*. Doctorate thesis. Université du Québec à Trois-Rivières, 2008.
- [140] Mertz, Christopher George. *Utilization of Genetic Algorithms and Constrained Multivariable Function Minimization to Estimate Load Model Parameters from Disturbance Data*. Doctorate thesis. Virginia Tech, 2013.
- [141] Box, George EP, et al. "Time series analysis: forecasting and control. John Wiley & Sons." *Hoboken, NJ* (2008).
- [142] Akaike, Hirotugu. "Statistical predictor identification." *Annals of the institute of Statistical Mathematics* 22.1 (1970): 203-217.
- [143] Abo-Hammour, Zaer S., Othman MK Alsmadi, and Adnan M. Al-Smadi. "A Novel Technique for ARMA Modelling with Order and Parameter Estimation Using Genetic Algorithms." *International Conference on Networked Digital Technologies*. Springer, Berlin, Heidelberg, 2010.

Résumé

Dans cette thèse, nous proposons d'étudier le problème de la prédiction de l'irradiation solaire par deux approches : dans la première, une méthodologie inspirée de la théorie d'apprentissage d'ensembles est proposée, où différentes familles de prédicteurs ont été examinées et testées, à savoir : techniques neuronales, techniques à noyau et techniques autorégressives. Différentes configurations ont été intégrées dans l'ensemble avec différentes méthodes de combinaison. Dans ce contexte une combinaison basée sur la technique de classification est proposée.

Pour la deuxième, nous proposons une nouvelle architecture de prédiction hybride basée sur l'optimisation d'une technique autorégressive.

Les deux approches proposées sont validées sur la base de jeux de données réels, où les résultats montrent la supériorité des approches proposées.

***Mots-clés:** séries temporelles, apprentissage machine, énergie solaire, prédiction, modèles statistiques, modèles intelligents, combinaison, optimisation.*

Abstract

In this thesis, we propose to study the problem of predicting solar irradiation by two approaches: in the first, a methodology inspired by ensemble learning theory is proposed, where different families of predictors have been examined and tested, namely: neural techniques, kernel techniques and autoregressive techniques. Different configurations have been incorporated into the ensemble with different combination methods. In this context, a combination based on the classification technique is proposed.

For the second, we propose a new hybrid prediction architecture based on the optimization of an autoregressive technique.

The two proposed approaches are validated on the basis of real data sets, where the results show the superiority of the proposed approaches.

Keywords: *time series, machine learning, solar energy, prediction, statistical models, intelligent models, combination, optimization.*

ملخص

في هذه الأطروحة، نقتح دراسة مشكلة التنبؤ بالإشعاع الشمسي من خلال نهجين: في الأول ، تم اقتراح منهجية مستوحاة من نظرية التعلم الجماعي ، حيث تم فحص واختبار عائلات مختلفة من التقنيات ، وهي: التقنيات العصبية ، تقنيات النواة وتقنيات الانحدار الذاتي . تم دمج تكوينات مختلفة في المجموعة بطرق تجميع مختلفة . في هذا السياق ، تم اقتراح مجموعة تعتمد على تقنية التصنيف .

بالنسبة للثاني ، نقتح بنية تنبؤ هجينة جديدة تعتمد على تحسين تقنية الانحدار التلقائي . يتم التحقق من صحة النهجين المقترحين على أساس مجموعات من البيانات الحقيقية ، حيث تظهر النتائج تفوق النهج المقترحة .

الكلمات المفتاحية: السلاسل الزمنية، التعلم الآلي، الطاقة الشمسية، التنبؤ، النماذج الإحصائية، النماذج الذكية، الجمع، التحسين.