



République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
UNIVERSITE MOHAMMED SEDDIK BENYAHIA



JIJEL
FACULTE DES SCIENCES ET DE LA TECHNOLOGIE
Département électronique
Laboratoire Energies Renouvelables

N° d'ordre :

Série :

Thèse Présentée

Mme. BELAÏD épouse BOUALIT SABRINA

En vue de l'obtention du diplôme de DOCTORAT en Science

Spécialité : Electronique

TITRE

**CONTRIBUTION A LA PREDICTION DE L'IRRADIATION
SOLAIRE EN MILLIEU SAHARIEN : APPLICATION AU SITE
DE GHARDAÏA.**

Soutenue le : 01/03/2017 Devant le jury composé de :

KEMIH KARIM	Professeur	UMSB Jijel	Président
MELLIT ADEL	Professeur	UMSB Jijel	Directeur de thèse
MALEK ALI	Directeur de recherche	CDER Alger	Examineur
AIT CHEIKH MED SALAH	Professeur	ENP Alger	Examineur

Dédicaces

À la mémoire de mon cher père.

*À ma mère ; Sans toi Mama, tout cela n'aurait pas connu le jour, que dieu
te garde en bonne santé et longue vie.*

À mon cher mari et nos adorables enfants :

Qu'ils sachent simplement qu'ils sont au fond de mon cœur

À chaque instant et en tout lieu.

REMERCIEMENTS

Avant tout, Mes remerciements vont à Dieu, le tout puissant, le miséricordieux, pour la volonté, la santé et surtout la persévérance qu'il m'a donné pour l'accomplissement de ce travail.

Je remercie vivement Monsieur Adel Mellit, Professeur à l'Université Mohammed Seddik Benyahia Jijel, de m'avoir confié et dirigé ce travail avec beaucoup d'efficacité, qu'il soit rassuré de mon estime. Grâce à son esprit pédagogue, il m'a fait prendre conscience du côté passionné et exaltant de la recherche scientifique.

J'en suis profondément honorée et j'exprime toute ma reconnaissance à Monsieur Kemih Karim, Professeur à l'Université Mohammed Seddik Benyahia Jijel, qui a bien voulu accepter la présidence du jury de cette thèse.

J'adresse mes respectueux remerciements à Monsieur Ait Cheikh Med Salah, Professeur à l'école nationale polytechnique d'Alger, pour avoir accepté de participer et d'honorer le jury de cette thèse.

Mes sincères remerciements sont également adressés à Monsieur Malek Ali, Directeur de recherche et président du conseil scientifique au centre de développement des énergies renouvelable Bouzaréah Alger, qui a accepté d'honorer par sa présence le jury de cette thèse.

Je remercie également Mr Djafer Djelloul, Directeur de notre unité de recherche appliquée en énergies renouvelables URAER, de m'avoir permis de suivre cette formation.

Je n'oublie pas de remercier mon cher mari et collègue, Dr. Boualit Abdelhamid. Rien de tout cela n'aurait été possible sans l'aide, le soutien, la patience qu'il m'a apporté tout au long de la préparation de cette thèse.

J'adresse mes remerciements les plus respectueux à mon amie et collègue Khalida Bousdira, Maître de recherche à l'URAER, pour son aide précieuse et soutien morale qu'elle m'ait apporté.

Merci à tous ceux qui ont contribué de près ou de loin à l'accomplissement de ce travail (Lynda, Karima, Nassima, Karim et Redha).

Résumé

Résumé : Des modèles prédictifs de l'irradiation solaire globale journalière (ISGJ) et de l'irradiation solaire globale moyenne journalière mensuelle (ISGM) sur une surface horizontale à Ghardaïa (Algérie) sont élaborés. Une technique basée sur l'apprentissage automatique appelée SVM (Machine à vecteurs support) est utilisée en exploitant des entrées simples, mesurables et disponibles in situ, telle la température. Les performances des modèles de l'ISGJ donnent un NRMSE qui varie de 13,163% à 13,305%, un MAPE de 10,403% à 10,503% et un coefficient de corrélation R supérieur à 89,4%. Pour le modèle prédictif de l'ISGM ; le NRMSE est de 7,442%, le MAPE de 8,940% et le R = 0,986. Ensuite, des études comparatives sont menées entre ces modèles SVM construits et ceux à base de réseaux de neurone (PMC perceptron multi couches), qui sont développés également pour le même site avec les mêmes inputs, ainsi que ceux publiés dans la littérature. Enfin, un modèle prédictif hybride est aussi développé en se basant sur le principe des séries temporelles, en combinant La technique SVM et le modèle SARIMA. Ainsi, le NRMSE du modèle hybride s'améliore de 0.38% par rapport à celui du modèle conventionnel SARIMA et le coefficient de corrélation atteint 87.4%.

Mots clés : Prévission – Irradiation solaire globale journalière - Irradiation solaire globale moyenne journalière mensuelle - SVM – Hybride – série temporelle.

Abstract: Predictive models of daily global solar radiation (ISGJ) and monthly average daily global solar radiation (ISGM) on a horizontal surface in Ghardaïa (Algeria) are developed. A technique based on automatic learning, which is the SVM (Support Vector Machine), is used by using measurable simple inputs that are available in situ, such as temperature. The performance of the ISGJ models gives an NRMSE which varies from 13.163% to 13.355%, a MAPE of 10.403% to 10.503%, and a correlation coefficient R greater than 89.4%. For the predictive model of the ISGM; The NRMSE is 7.442%, the MAPE is 8.940% and the R = 0.986. Then, comparative studies are carried out between these constructed SVM models and those based on neural networks (PMC perceptron multi-layer), developed also for the same site with the same inputs, and also those published in the literature. Finally, a hybrid predictive model based on the time series principle is also developed by combining the SVM technique and the SARIMA model. The NRMSE of the hybrid model improved by 0.38% compared to the conventional SARIMA model and the correlation coefficient reached 87.4%.

Keywords: Prediction – Daily global solar radiation – Mean monthly global solar radiation – SVM – Hybrid – time series.

ملخص: في هذه الأطروحة، نقترح نماذج لتنبؤ الإشعاع اليومي للطاقة الشمسية (ISGJ) ومتوسط الإشعاع اليومي الشمسي الشهري (ISGM) على سطح أفقي في غرداية (الجزائر). تستند هذه الطريقة المختارة في المقام الأول على تقنية التعلم الآلي، و التي طبقت حديثاً في مجال الأرصاد الجوية وخاصة الطاقة الشمسية، و هي SVM و ذلك باستخدام مدخلات بسيطة قابلة للقياس المتوفرة في الموقع، مثل درجة الحرارة. أداء نماذج التنبؤ الإشعاع اليومي تعطي NRMSE يتراوح من 13.163% إلى 13.305%، ومعامل الارتباط R أكبر من 89.4%. بالنسبة للنموذج التنبؤي الشهري، NRMSE هو 7.442% و R يساوي 0.986. كذلك أجريت دراسة مقارنة بين نماذج SVM المقترحة ونماذج الشبكات العصبية (متعدد الطبقات المستقبلات PMC) التي بنيت أيضاً لنفس الموقع مع نفس المدخلات وكذلك النماذج التي نشرت في المراجع. أيضاً، تم تطوير نموذج تنبؤي آخر هجين مبني على أساس مبدأ السلاسل الزمنية. وذلك بتكليف تقنية SVM إلى الجزء الغير الخطي لسلسلة الإشعاع الشمسي اليومي ونموذج SARIMA منهجية بوكس وجنكز) للجزء الخطي للسلسلة الزمنية. النموذج الهجين NRMSE حسن الاداء بنسبة 0.38% مقارنة بما كان عليه في النموذج التقليدي SARIMA ومعامل الارتباط وصل إلى 87.4%.

Sommaire

Tables des figures	1
Table des Tableaux	3
Nomenclature	4
INTRODUCTION GENERALE	6
CHAPITRE 1 : Etat de l'Art	11
1. Considérations fondamentales	12
1.1 Nature du rayonnement solaire atteignant le sol terrestre	12
1.2 Air Masse	14
1.3 La constante solaire	15
1.4 L'irradiation solaire globale hors atmosphère	16
1.5 La durée d'ensoleillement effective	16
1.6 La durée d'ensoleillement maximale	16
2. Etat de l'art sur les techniques prédictives de l'irradiation solaire	17
2.1 Modèles prédictifs à base de Réseaux de Neurones Artificiels	18
2.2 Modèles prédictifs à base de Support Vector Machine	22
2.3 Modèle prédictifs suivant la méthodologie de Box & Jenkins	25
2.4 Synthèse des performances des modèles prédictifs de l'irradiation solaire globale par les méthodes d'apprentissage statistique	26
2.4.1 Cas de l'irradiation solaire globale journalière ISGJ	26
2.4.2 Cas de l'irradiation solaire globale journalière moyenne mensuelle ISGM	28
2.5 Influence des paramètres d'entrées sur la prédiction de l'irradiation solaire	28
3. Présentation climatique du site étudié	33
3.1 L'irradiation solaire globale journalière hors atmosphère in situ (Ghardaïa)	33
3.2 Présentation de la station radiométrique installée sur le site de l'URAER	33
3.3 Le prétraitement des mesures effectuées	34
CHAPITRE 2 : Méthodologie	37
1. Les Support Vector Machine SVM	38
1.1 Principe générale des SVM	38
1.2 Régression par SVM	40
1.2.1 Formulation du problème dans le cas linéaire :	40
1.2.2 Formulation du problème dans le cas non linéaire	43
1.2.3 Quelques définitions	43
1.2.4 Propriétés des noyaux	44
1.2.5 Formulation duale et optimisation du problème de régression	44

Sommaire

2.	Les réseaux de neurones artificiels RNA	45
2.1	Historique	45
2.2	Le perceptron	46
2.3	Réseaux de neurones	46
2.4	Perceptron multicouche	47
2.5	Equation du réseau PMC	47
2.6	Apprentissage du réseau	48
3.	Modèles de prévision des séries temporelles	48
3.1	Introduction	48
3.2	Processus SARIMA (p,d,q)(P,D,Q)	49
3.3	Identification	50
3.4	Estimation	53
3.4.1	Système d'équations de Yule-Walker	53
3.4.2	Méthode du maximum de vraisemblance	54
3.5	Validation du modèle	55
3.5.1	La parcimonie :	55
3.5.2	Test sur les résidus	55
3.6	Prévision de la série temporelle via le modèle SARIMA	57
Chapitre3 : Prédiction de L'irradiation solaire globale par les techniques d'apprentissage statistiques		60
1.	Introduction	61
2.	Données et méthodologie	62
2.1	Base de données	62
2.2	Chois des entrées des modèles SVM	64
2.3	Méthodologie de construction des modèles SVM	65
3.	Résultats et discussion	68
3.1	Prédiction journalière	68
3.2	Prédiction mensuelle	72
4.	Etude comparative	75
4.1	Cas de l'irradiation solaire globale journalière ISGJ	75
4.1.1	Le modèle PMC développé pour ISGJ	75
4.1.2	Comparer les performances du modèle SVM proposé à ceux publiés dans la littérature	77
4.2	Cas de l'irradiation solaire globale moyenne mensuelle ISGM	78

Sommaire

4.2.1 Le modèle PMC développé pour ISGM	78
4.2.2 Comparer les performances du modèle SVM proposé à ceux publiés dans la littérature	80
5. Conclusion partielle	81
Chapitre4 : Modèle prédictif hybride de l'ISGJ	82
1. Introduction	83
2. La Méthodologie hybride SARIMA-SVM	83
3. Résultats et discussion	84
3.1 Modèle SARIMA (p,d,q)(P,D,Q)	84
3.2 Le modèle hybride SARIMA-SVM:	88
4. Conclusion partielle	90
CONCLUSION GENERALE	91
REFERENCES	95

Tables des figures

Figure 1 : Parcours du rayonnement solaire. a) Direct. b) Diffus. c) Réfléchis [12] 13

Figure 2 : Impact de la couverture nuageuse sur la série temporelle du rayonnement solaire global [11] 14

Figure 3 : L'angle et la distance du soleil par rapport à l'atmosphère..... 15

Figure 4 : Variation de la constante solaire durant l'année 15

Figure 5 : Type de réseau utilisé par Premalatha [52] 21

Figure 6 : l'irradiation solaire globale hors atmosphère G_0 par rapport au site de Ghardaïa 33

Figure 7 : Présentation de la station radiométrique installée sur le toit de L'URAER..... 34

Figure 8 : Exemple d'acquisition de données radiométriques in situ..... 35

Figure 9 : Le rayonnement solaire global horizontal pour différentes types de journée à Ghardaïa..... 36

Figure 10 : Hyperplan optimale et marge maximale 39

Figure 11 : Hyperplan et hyper-tube..... 40

Figure 12 : Fonctions de perte (a, b, c, d)..... 41

Figure 13 : Représentation de la marge douce et illustration de la fonction de pertes 42

Figure 14 : Régression dans le cas non linéaire par l'utilisation de noyaux 43

Figure 15 : Le perceptron 46

Figure 16 : Schéma de principe du réseau PMC 47

Figure 17 : Processus de modélisation et de prévision par la méthodologie Box & Jenkins ... 50

Figure 18 : Série du nombre de passager et sa prévision pour les 36 mois à compter de Juillet 1957 [12] 59

Figure 19 : les températures mesurées à L'URAER Ghardaïa..... 61

Figure 20 : Variation de l'Irradiation solaire globale journalière in situ..... 62

Figure 21 : Maximums, minimums et moyennes des ISGJ pour chacun des mois (2012, 2013 et 2014)..... 63

Figure 22 : Evolution de l'ISGJ mesurée ainsi que H_0 avec S_0 calculés durant la période 2012-2014..... 64

Figure 23 : Evolution de l'ISGJ par rapport à chacune des entrées 65

Figure 24 : Les étapes de construction des modèles SVM 66

Figure 25 : Les indicateurs de performance des modèles SVM développés pour la prédiction de l'ISGJ..... 69

Table des figures

Figure 26 : Tracés des valeurs entraînées et prédites par les modèles SVM ₂₅ .SVM ₂₈ , SVM ₃₉ et SVM ₄₂ en fonction des valeurs mesurées de l'ISGJ.	71
Figure 27 : Indicateurs de performances des différents modèles SVM développés pour l'ISGM.....	73
Figure 28 : Comparaison des ISGM mesurées avec celles entraînées et prédites.....	74
Figure 29 : Architecture du réseau PMC sélectionné (ISGJ)	75
Figure 30 : Comparaison des performances des deux modèles SVM et RNA développés.....	76
Figure 31 : Architecture du réseau PMC sélectionné (ISGM)	79
Figure 32 : Comparaison entre les mesures et valeurs prédites de l'ISGM par les deux modèles.....	80
Figure 33 : La série temporelle Kt et celle différenciée dKt	85
Figure 34 : Autocorrelogrammes simple et partiel de dKt	86
Figure 35 : ISGJ mesurées et estimées par le modèle SARIMA.....	87
Figure 36 : ISGJ mesurées et prédites par le modèle SARIMA.....	87
Figure 37 : ISGJ prédites par le modèle hybride SARIMA-SVM	89

Table des Tableaux

Tableau 1 : Comparaison des MAPE, publiés dans la littérature et dédiés à la prédiction de l'ISGM. [52] _____ 22

Tableau 2 : Résultats des performances de quelques modèles prédictifs de l'ISGJ [7] (1/2) 27

Tableau 3 : Résultats des performances de quelques modèles prédictifs de l'ISGM [7] _____ 29

Tableau 4 : Influence des variables d'entrées sur les performances prédictifs de l'irradiation solaire (1/3) _____ 30

Tableau 5: Exemples d'identification et d'estimation _____ 51

Tableau 6 : Propriétés des processus AR(p), MA(q) et ARMA (p,q) _____ 52

Tableau 7 : Les différentes combinaisons utilisées pour la construction des modèles SVM_ 67

Tableau8 : Résultats des performances des modèles sélectionnés (SVM₂₅, SVM₂₈, SVM₃₉ et SVM₄₂) _____ 70

Tableau 9: Hyper-paramètres des modèles SVM sélectionnés _____ 72

Tableau 10: Résultats des performances du modèle SVM₃₀ _____ 73

Tableau 11 : Hyper-paramètres obtenus pour le modèle SVM₃₀ _____ 74

Tableau 12: Comparaison des performances de prédiction de l'ISGJ par SVM et PMC ____ 77

Nomenclature

Nomenclature

Alt	: altitude (m)
ANFIS	: système d'inférence neuronal flou adaptative (Adaptive Neuro Fuzzy Inference System)
C	: paramètre de régularisation
D	: numéro du jour de l'année, commençant du 1 janvier (1-365)
DEC	: déclinaison du soleil (°)
DH ₀	: irradiation solaire globale journalière hors atmosphère (Wh/m ² /jour)
DT _{diff}	: maximum-minimum des températures journalières (°C)
DT _{max}	: température journalière maximale (°C)
DT _{min}	: température journalière minimale (°C)
DT _{moy}	: température journalière moyenne (°C)
E _v	: Evaporation
H ₀	: irradiation solaire globale hors atmosphère (Wh/m ²)
H _c	: irradiation solaire en jour clair (Wh/m ² /jour)
ISG	: irradiation solaire globale (Wh/m ²)
ISGJ	: irradiation solaire globale journalière (Wh/m ² /jour)
ISGM	: irradiation solaire globale moyenne journalière mensuelle (Wh/m ² /jour)
K _t	: indice de clarté
Lat	: latitude (°)
LLR	: régression linéaire locale (Local Linear Regression)
Lon	: longitude (°)
M	: mensuelle
MH ₀	: irradiation solaire globale hors atmosphère moyenne journalière mensuelle (Wh/m ² /jour)
MT _{diff}	: écart des moyennes mensuelles des températures maximales et minimales journalières (°C)
MT _{max}	: moyenne mensuelle des températures maximales journalières (°C)
MT _{min}	: moyenne mensuelle des températures minimales journalières (°C)
MT _{moy}	: moyenne mensuelle des températures moyennes journalières (°C)
N _u	: nébulosité
Pa	: pression atmosphérique (N/m ²)
PMC	: perceptron multi couche
P _r	: précipitation (mm)

Nomenclature

Pv	: pression de vapeur (bar)
R	: coefficient de corrélation
R ²	: coefficient de détermination
RBF	: fonction de base radiale (Radial Basis Function RBF)
Rh	: humidité relative (%)
RNA	: réseau de neurone artificiel
RNB	: réseau de neurone bayésien
RNBR	: réseau de neurone à fonction de base radiale
RNR	: réseau de neurone récurrent
RNRG	: réseau de Neurone à Régression Généralisée
S	: durée d'ensoleillement effective (Hours)
S ₀	: durée d'ensoleillement maximale (heure)
SARIMA	: autorégressif moyenne mobile intégré saisonnier (Seasonal Auto Regressive Integrated Moving Average)
SKT	: température de surface
SVM	: séparateurs à vaste marge (Support Vector Machine)
Tr	: température de rosée (°C)
W ₀	: angle horaire au couché du soleil (°)
WS	: vitesse du vent (m/s)
σ	: largeur de la bande passante du noyau RBF

INTRODUCTION GENERALE

INTRODUCTION GENERALE

Il est clair que l'Algérie par sa nouvelle politique énergétique tend à promouvoir le développement des énergies renouvelables particulièrement le solaire, de par sa situation géographique qui dispose d'un des gisements solaire les plus élevés au monde [1]. Mais l'exploitation de ce gisement énergétique solaire, soulève certains défis techniques en raison de son caractère intermittent et aléatoire, où son intégration dans les réseaux d'exploitation (électrique ou thermiques) pose des problèmes pour le maintien de l'équilibre production-consommation. Donc, une utilisation efficace de cette énergie solaire nécessite des informations de prévision fiable. Le défi actuel des chercheurs est de développer des modèles prédictifs qui permettent d'atteindre de bonnes performances en réduisant de manière significative le taux d'erreur par rapport aux différentes échelles de temps (minute, horaire, journalière ou mensuelle).

La prévision du comportement ou la réponse d'un quelconque phénomène que ce soit par les techniques d'apprentissage statistiques ou autres (empiriques, stochastiques,...), nécessite un ensemble de variables entrantes aux modèles ou ce qu'on appelle les inputs (entrées). Il a été remarqué que la majorité des modèles proposés dans la littérature dans le contexte de la prévision de l'énergie solaire sont principalement basé sur les réseaux de neurone artificiels ([2], [3], [4], [5],....) et utilisent plusieurs variables en entrée (durée d'ensoleillement, l'humidité, vitesse du vent, nébulosité, précipitation, température,...) ce qui est n'est pas toujours évident vu la non disponibilité des différents appareils nécessaires à la mesure de ces variables climatiques dans un site donnée.

L'objectif donc visé par ce travail est l'amélioration des performances des modèles prédictifs de l'énergie solaire d'une part et d'une autre part en exploitant seulement quelques variables mesurables simples et disponibles in situ, telle la température. Pour cela une nouvelle technique appelé Support Vector Machine ou SVM, qui est récemment appliqués dans le domaine de la météorologie et dans le solaire en particulier, est exploité. Les toutes premières investigations relatif à la prédiction de la composante solaire par les SVM n'ont été mené qu'en 2011 par Chen et al [6] en Chine en développant un modèle prédictif de l'irradiation solaire globale moyenne journalière mensuelle ISGM en combinant les températures à l'entrée du modèle.

On contribuera par ce travail à développer par cette technique SVM des modèles prédictifs de l'Irradiation solaire globale journalière et mensuelle à un pas de temps d'une journée et d'un mois respectivement. Ensuite, pour prouver l'efficacité de ces modèles, une étude comparative sera menée entre leurs performances, celles des modèles a base de RNA que nous développerons aussi, ainsi que celles des modèles déjà cités dans la littérature.

Tout en restant dans le contexte de la prédiction de l'irradiation solaire, mais cette fois ci un Modèle prédictif hybride SARIMA-SVM de l'ISGJ par le principe des séries temporelles est construit de sorte que la technique SVM est adapté à la partie non linéaire de la série après avoir procédé à la prédiction de la partie linéaire de la série temporelle par la méthodologie de Box & Jenkins (SARIMA)

Pour ce qui est de la région d'étude et l'application de nos modèles prédictifs, le choix s'est portée sur Ghardaïa (sud de l'Algérie) vu qu'elle est considérée comme une

plate-forme expérimentale pour des applications solaires, de par son fort potentiel solaire, donc favorable aux applications solaires (photovoltaïques ou thermique).

Ce manuscrit s'organise en 4 chapitres comme suit :

Le chapitre 1 est présenté en trois parties ; dans la première une présentation brève de quelques notions fondamentales de l'énergie solaire. Dans la seconde partie, nous exposons pour chacune des méthodes adoptées (les RNA, les SVM et le SARIMA) son état de l'art par rapport à la prédiction de l'énergie solaire. Ensuite nous apportons une synthèse des modèles prédictifs de l'irradiation solaire globale par les méthodes d'apprentissage statistique et ceux pour le cas journalier et mensuelle (cette partie a fait objet d'une publication [7]). Nous illustrons également une synthèse tirée de la littérature portant sur l'influence des paramètres d'entrées du modèle sur les performances de la prédiction de l'irradiation solaire. La troisième partie est dédiée à la présentation climatique du site étudiée.

Le chapitre 2 s'intéresse aux méthodes qui mettent en œuvre les prédicteurs déjà évoqués précédemment en apportant une description des algorithmes associés et le principe général de leurs utilisations.

Dans le chapitre 3, nous développons des modèles de prédiction de l'irradiation solaire globale journalière ISGJ et moyenne journalière mensuelle ISGM pour le site de Ghardaïa par les SVM en utilisant comme inputs aux modèles prévisionnels des paramètres simples et disponibles in situ à savoir ; la température ambiante mesurée, la durée d'ensoleillement calculée S_0 et l'irradiation solaire extra-terrestre H_0 . Aussi, une étude comparative est menée entre les modèles SVM prédictifs sélectionnés et les modèles de réseaux de neurones multi couches (PMC) qui sont aussi construits à partir

de la même base de données. Ensuite, ces modèles présentés sont comparés à d'autres modèles déjà publiés dans la littérature. (ce chapitre a fait objet d'une publication [7]).

Le chapitre 4 traitera la prédiction de l'ISGJ in situ via un modèle hybride par le principe des séries temporelles. Nous proposons donc un modèle hybride basé sur les SARIMA (Seasonal Auto-Regressif Integrated Moving Average) et les SVM (Support Vector machine). Car, d'après les investigations des chercheurs, les ARMA est parmi les techniques les plus compétitives pour la prédiction des séries temporelles [8]. (Ce chapitre a fait objet d'une communication internationale IRSEC2016).

Enfin, dans la conclusion générale nous présentons une synthèse des résultats de nos expérimentations qui met en évidence les apports de notre travail. Aussi, de nouvelles perspectives de recherche sont proposées.

CHAPITRE 1 : Etat de l'Art

1. Considérations fondamentales

Avant d'aborder le principe de la prédiction de l'irradiation solaire globale à la surface de la terre dans un site donnée, il est important d'exposer quelques faits théoriques qui permettent de comprendre les différents éléments qui composent le rayonnement solaire global. Mais pour plus de détails, on peut se référer aux ouvrages de [9].

Les termes *rayonnement solaire*, *irradiance*, *irradiation*, *radiance*, *éclairage*, *luminance* et *illumination* sont fréquemment rencontrés dans la littérature et une note sur leur utilisation doit être appropriée.

- *Le rayonnement solaire* (W/m^2) ou *la luminance* ($\text{candela}/\text{m}^2$ (cd/m^2)) désigne l'énergie provenant du soleil.
- *La luminance* est l'énergie contenue à l'intérieur de la partie visible du spectre de rayonnement solaire ($0,39$ à $0,78\mu\text{m}$).
- *Irradiance* (W/m^2) et *l'éclairage* (lm/m^2) se réfèrent à l'énergie incidente instantanée.
- Le terme *irradiation* (Wh/m^2 ou J/m^2) et *l'illumination* (lumière-heure ($\text{lm}\cdot\text{h}$) / m^2) se réfèrent à l'énergie incidente cumulée sur une surface dans une période de temps donnée.

1.1 Nature du rayonnement solaire atteignant le sol terrestre

Hors atmosphère, le flux énergétique moyen reçu par la terre est d'environ $1400\text{W}/\text{m}^2$ par jour. En traversant l'atmosphère, le rayonnement solaire incident subit plusieurs changements avant d'arriver au sol (Figure 1), car il est fortement influencé par l'atmosphère terrestre, par les phénomènes d'absorption et de réflexion [Ulgen et Hepbasli 2009]. La couche d'ozone absorbe la majeure partie du rayonnement ultraviolet, tandis que la vapeur d'eau absorbe le rayonnement infra-rouge. Ainsi, le flux énergétique reçu au sol dépend de nombreux facteurs, et son maximum ne dépasse guère $1000\text{W}/\text{m}^2$ [10].

De ce fait, la valeur de l'éclairage en un point du sol et à un instant donné dépend non seulement de la position du Soleil, mais aussi de l'état optique de l'atmosphère [Batlles et al.2000 ; Rehman 1999]. Ce phénomène permet de scinder en deux parties distinctes le phénomène du rayonnement global au sol [Oumbe 2009 ; Perez et al. 1990] : une composante que l'on nommera « déterministe » et qui ne dépend que de la distance entre le point de mesure et le centre du Soleil, et une composante que l'on nommera « stochastique » qui dépend de l'état de la couverture nuageuse (Figure 2), du taux d'ozone, du taux d'humidité, etc. [Hay et Davies 1978 ; Ineichen 2006 ; Bird et Hulstrum 198 ; Mellit et al. 2008]. La première composante peut se déterminer aisément avec des règles de mécanique céleste et de

physique énergétique [Iqbal 1983 et Jin et al. 2005]. Concernant la seconde, elle est beaucoup plus difficile à appréhender [Klucher 1979 ; Marty et Philipona 2000] du fait purement probabiliste des différents éléments qui la composent: occurrence de nuage, de concentration d'ozone, ...etc. [11].

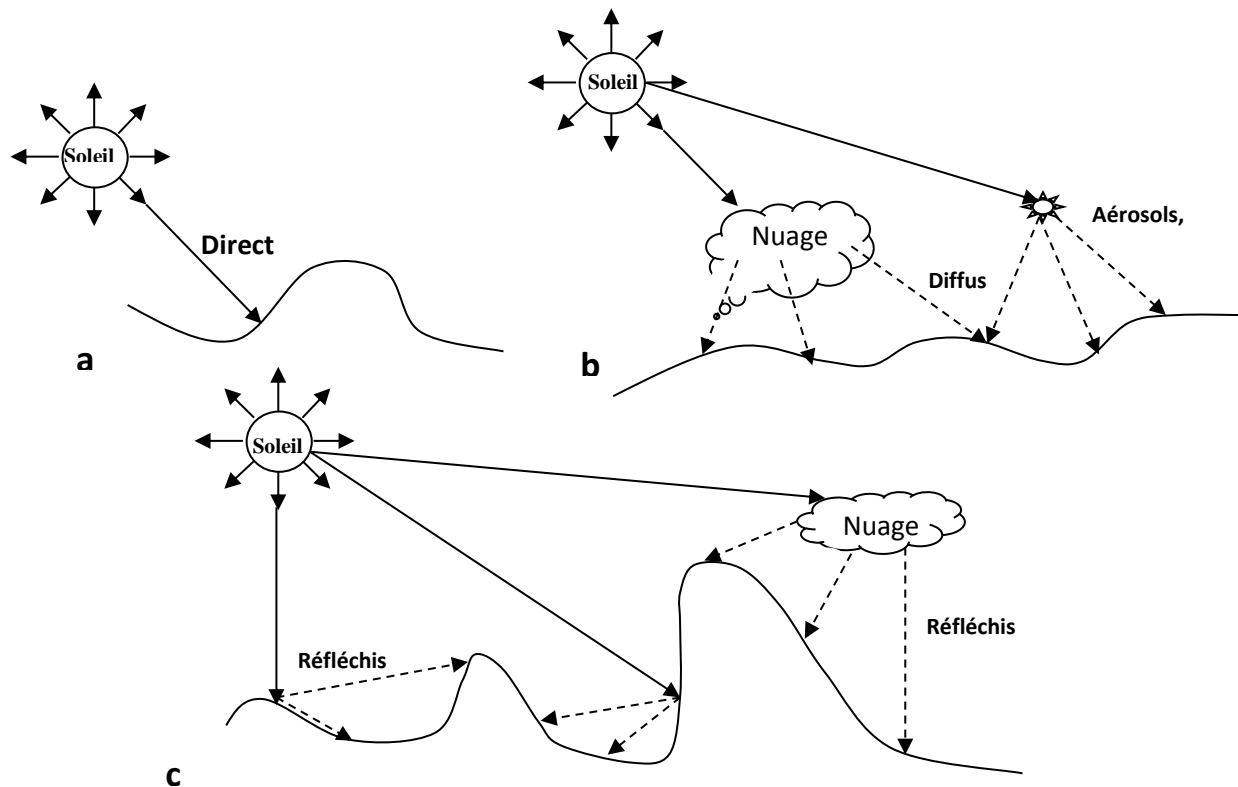


Figure 1 : Parcours du rayonnement solaire. a) Direct. b) Diffus. c) Réfléchi [12]

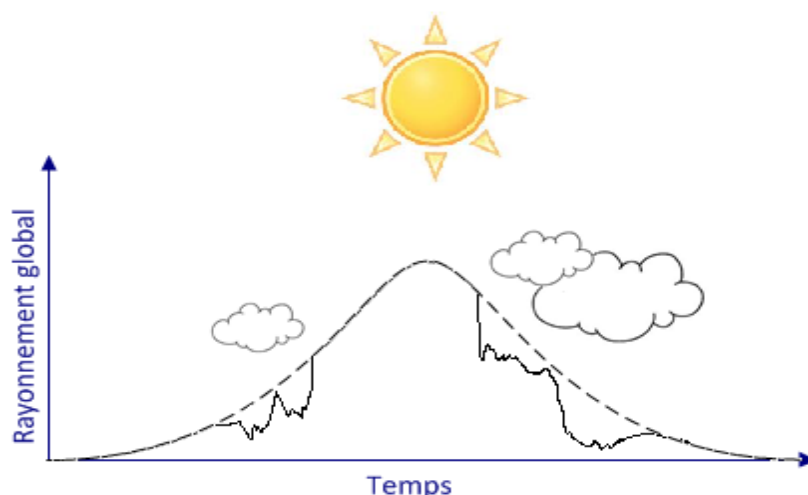


Figure 2 : Impact de la couverture nuageuse sur la série temporelle du rayonnement solaire global [11]

1.2 Air Masse

L'Air Masse est l'épaisseur de l'atmosphère que la lumière doit traverser et elle change d'heure en heure selon la position du soleil dans le ciel. Si le soleil est au zénith, la lumière traverserait une épaisseur de l'atmosphère définie par un AM1 pour atteindre la surface de la terre. L'après-midi ou le matin, le soleil est à un angle inférieur, et la distance que doit traverser la lumière pourrait être deux fois ou plus que celle au zénith (Figure 3).

L'équation mathématique définissant le nombre d'Air Masse est comme suit :

$$m = \frac{(p \cdot \exp(-Z/7.8))}{(1013 \cdot \sin(A))} \quad (1)$$

Avec :

Pa : est la pression atmosphérique (N/m²)

A(°) : élévation du soleil dans l'horizon (90° au zénith)

Z : altitude du site (Km)

Hors atmosphère, à haute altitude, on définit les conditions AM0.

Ce nombre d'Air Masse se sert à définir un spectre de référence pour calibrer dans le domaine pratique les cellules étalonnées destinées à qualifier les performances des dispositifs photovoltaïques.

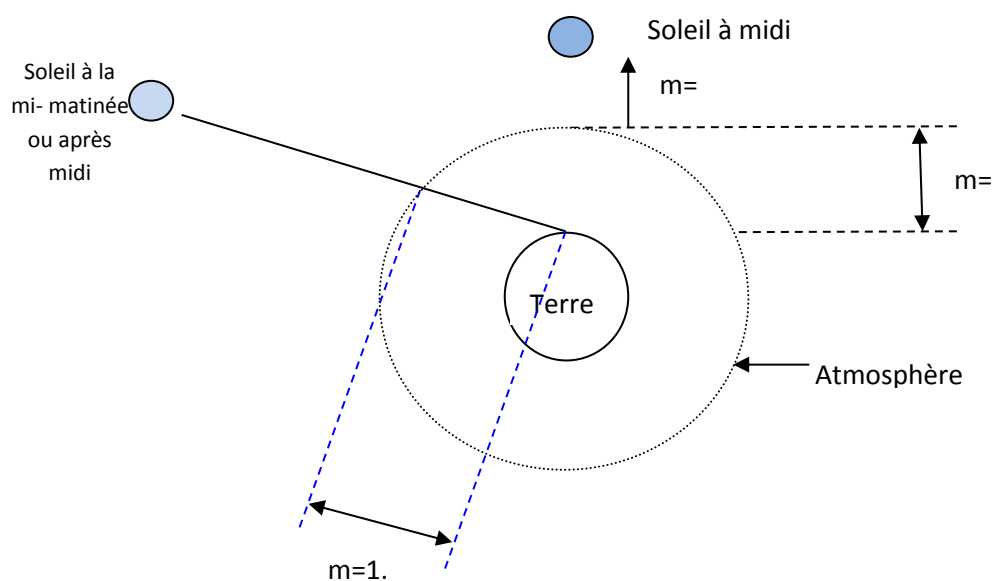


Figure 3 : L'angle et la distance du soleil par rapport à l'atmosphère

1.3 La constante solaire

La constante solaire est la densité de puissance du rayonnement sur l'atmosphère externe de la terre et exprimée en watts par mètre carré par minute (W/m^2). Iqbal (1986) fournit un calcul plus détaillée sur la constante solaire. Comme la distance entre le soleil et la terre change durant l'année, alors la valeur de la constante solaire change également pendant l'année (Figure 4). Le Centre Mondiale du Rayonnement a adopté une valeur de $1367 W/m^2$ avec une incertitude de 1%.

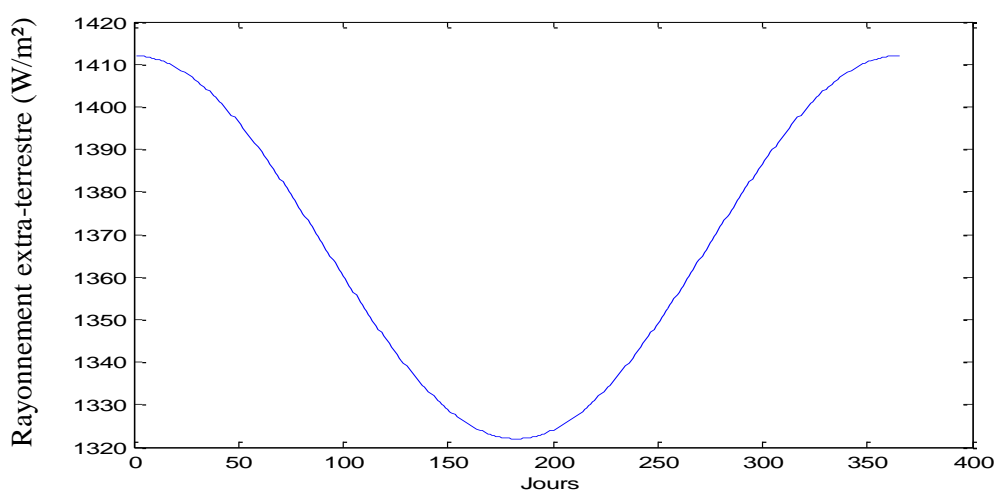


Figure 4 : Variation de la constante solaire durant l'année

1.4 L'irradiation solaire globale hors atmosphère

L'irradiation solaire globale hors atmosphère H_0 reçu sur un plan horizontal par rapport au site donné peut facilement se calculer à partir de la formule suivante [9]:

$$H_0 (Wh/m^2) = \left(\frac{24I_0}{\pi}\right) \left(1 + 0.033 \cos\left(\frac{360N}{365}\right)\right) \cdot \left(\cos Lat \cos DEC \sin w_0 + 2\pi w_0 \sin Lat \sin \frac{DEC}{360}\right) \quad (2)$$

Tel que ;

I_0 : est la constante solaire, $I_0 = 1367(W/m^2)$

N : est le numéro du jour à partir du 1^{er} Janvier

Lat : est la latitude du lieu

DEC : est la déclinaison du soleil

$$DEC = 23.45 \left[\sin\left(\frac{360(284+N)}{365}\right) \right] \quad (3)$$

w_0 : est l'angle horaire au coucher du soleil :

$$w_0 = \cos^{-1}(-\tan Lat \tan DEC) \quad (4)$$

NB : Tous les angles sont en radians si nous travaillons sur Excel.

1.5 La durée d'ensoleillement effective

La durée d'ensoleillement ou durée d'insolation est un indicateur climatique qui mesure le temps pendant lequel un endroit, un lieu est éclairé par le Soleil sur une période donnée durant laquelle le rayonnement solaire direct dépasse le seuil de 120 W/m².

La durée d'ensoleillement se mesure le plus souvent à l'aide d'un héliographe qui utilise une lentille optique en verre en forme de sphère.

1.6 La durée d'ensoleillement maximale

En absence permanente de nuages, la durée d'ensoleillement (d'insolation) est pratiquement égale à la durée du jour, qui s'appelle durée astronomique ou théorique du jour. Elle est définie par :

$$S_0 = \frac{2}{15} w_0 \quad (5)$$

Avec ;

S_0 ; Durée du jour (h)

w_0 : L'angle horaire au couché du soleil ($^\circ$)

2. Etat de l'art sur les techniques prédictives de l'irradiation solaire

Nombreuse sont les techniques utilisées dans la modélisation et la prédiction du rayonnement solaire (direct, globale, diffus). On peut classer ces techniques en grandes classes distinctes; la première classe comporte des modèles complexes de transferts radiatifs qui tiennent compte des interactions entre l'irradiation solaire et l'atmosphère terrestre, telles que la dispersion de Rayleigh, l'absorption radiative par l'ozone, par la vapeur d'eau ou par les aérosols [13]. Ces modèles ont été adoptés par certains auteurs; [14], [15], et [16]. Mais pour plus de connaissances sur ces modèles, on peut se référer aux travaux de [17], où ils résument une liste de modèles radiatifs existants dans la littérature, qui sont applicables à la prédiction des trois composantes du rayonnement solaire à différentes échelles de temps, et il a même appliqué ces modèles à la prédiction du rayonnement solaire dans 5 sites en Amérique.

La deuxième classe, se sont des modèles qu'on appelle empiriques qui sont en fonction de quelques paramètres météorologiques et/ou géographiques. Ces modèles découlent de l'équation originale d'Angström (1924) [9] définissant la relation existante entre l'indice de clarté (rapport du rayonnement solaire à la surface de terre et celui hors-atmosphère terrestre H/H_0) et le rapport d'insolation (la durée d'ensoleillement sur la durée d'ensoleillement maximale S/S_0). Plusieurs modèles résultent à partir de cette formulation d'Angström avec plus au moins de paramètres pour s'adapter à un site et un climat donnée. Par exemple, pour l'Espagne; Palomo (1989) , pour l'Italy ; Mustacchi et al (1979), pour Athènes ; Balouktsis et Tsalides (1986), pour Singapore ; Goh et Tan (1977), pour le Canada ; Graham et al (1987) et pour les USA ; Knight (1988) [18].

La troisième classe des modèles prédictifs du rayonnement ou de l'énergie solaire est basée sur le principe de l'apprentissage statistique. Ces techniques ont été développées ces dernières décennies et utilisées dans la modélisation ou l'estimation des phénomènes très complexes ou encore trop bruités tel le cas du rayonnement solaire et permettent de prédire au mieux leurs comportements à partir des séries d'observations.

La dernière classe s'agit des modèles basés sur le principe des séries temporelles. L'approche utilisée est basé sur l'utilisation des corrélations qui consiste à étudier les corrélations croisées de la série chronologique observée. Ces méthodes sont généralement paramétriques de type moyenne-mobiles (Moving Average MA), autorégressives (AR), une combinaison des deux (ARMA) ou autres (ARIMA, SARIMA, AR, MA, ARCH, GARCH, ARMAX,...).

Parmi ces différentes approches de prévision, nous illustrons dans ce qui suit l'état de l'art de la prédiction de l'énergie solaire par trois techniques qui font l'objet de notre travail; les Réseaux de Neurones Artificiels RNA, les Séparateurs à Vaste Marge SVM (Support Vector Machine) et les modèles de séries temporelles SARIMA (Seasonal AutoRegressive Integrated Moving Average)

2.1 Modèles prédictifs à base de Réseaux de Neurones Artificiels

A la fin des années 1990, des efforts ont été fourni pour introduire les techniques des réseaux de neurones artificiels RNA dans l'estimation de l'irradiation solaire [19]. Plusieurs approches étaient à l'investigation. Le problème a été abordé sous différents angles: la modélisation des approches employées est caractérisée par une grande variété de modèles de réseaux neuronaux avec différentes architectures et une diversité de paramètres d'entrée et de sortie. Ce qui est donc très intéressant de noter, est que dans moins de 20 ans d'applications de RNA à des études sur le rayonnement solaire, une grande variété d'approches d'entraînement neuronales ont été utilisés (différents algorithmes d'apprentissage, architectures,... etc.) et une multitude de variables d'entrée ont été utilisées que se soit météorologiques, géographique ou autres. Ce qui est commun dans presque toutes ces études est la validation de la méthodologie proposée avec des données indépendantes.

Différents modèles de RNA ont été appliqués à l'estimation ou à la prédiction des trois composantes du rayonnement solaire (globale, direct et diffus) à plusieurs échelles de temps. Horaire ([20], [21], [22] et [23]), journalier (ISGJ) ([24], [25] et [26]), et dans le cas de la moyenne journalière mensuelle (ISGM), on ne peut ne pas citer les toutes premières investigation menées par Mohandes et al ([27], [28]).

Ces travaux sont bien sur les toutes premières investigations qui ont été menées. Par la suite, d'autres efforts sont encore fournit mais cette fois ci en comparant à la fois les différentes architectures neuronales par les différents algorithmes d'apprentissage.

D'après une synthèse bibliographique publiée en 2009 par Mellit et al [29], qui est l'auteur de référence dans la manipulation des réseaux de neurones dans le domaine de la prédiction de l'irradiation solaire et de l'énergie PV (photovoltaïque), il montre que 79% des techniques de l'intelligence artificielle (IA) utilisées dans la prévision des données météorologiques sont basées sur une approche connexionniste (RNA) et plus particulièrement le PMC qui représente la grande majorité des travaux de recherche. D'ailleurs, en 2010, Mellit et al [30] développent un réseau MLP pour prévoir l'irradiation solaire globale à pas de temps de 24 heures en Italie afin de vérifier la capacité de généralisation des modèles prédictifs MLP. Ils se servent de la valeur présente de l'irradiation solaire ainsi que de la température comme entrées au modèle. Les performances sont telles que le coefficient de corrélation est entre 98 et 99% pour les journées claires et 94-96% pour des journées couvertes.

Dans ce qui suit, nous prendrons le soin d'identifier certains travaux dans le but d'exposer la méthodologie de l'utilisation des RNA ainsi que leurs résultats d'analyses par rapport à l'irradiation solaire en général :

En 2008, Bosch et al [5] ont choisi de tester si l'utilisation de l'information provenant d'un seul poste radiométrique et qui est facilement accessible pour fournir, par extrapolation, la répartition du rayonnement solaire pour la zone de topographie complexe étudiée, qui se situe dans le Sud Est de l'Espagne. Les entrées au réseau de neurone utilisé ont été sélectionnées en utilisant une technique de sélection des entrées, appelée méthodologie de détermination automatique de la pertinence (ARD Automatic Relevance Determination). Le modèle neuronal élaboré a permis d'atteindre des performances assez satisfaisantes.

Bilgili et al [31] ont exploité trois techniques différentes pour la prédiction de l'irradiation solaire globale en combinant plusieurs variables météorologiques telles que la durée d'ensoleillement, la température, la vitesse du vent, l'humidité ainsi que le jour et le mois de l'année. La première technique est basée sur la régression multiple linéaire (RML), la seconde est basée sur la régression multiple non linéaire (RMNL) et la dernière est celle des réseaux de neurones artificiels (RNA). Les résultats montrent que les RNA permettent d'obtenir de meilleures prédictions en se servant de toutes les variables d'entrées.

Il existe encore bien d'autres travaux basés toujours sur l'approche des RNA avec de bons résultats de performances [30, 32-48]. Mais avant de conclure sur ça, nous citerons quelques-uns des plus récents:

En 2016, Zou et al [49] ont développé un modèle basé sur les réseaux de neurones artificiels (RNA) et sur une technique d'interpolation pour estimer l'irradiation solaire globale (ISG) dans le Sud-Est de la Chine. Utilisant comme entrées plusieurs paramètres météorologiques telles que la durée d'ensoleillement, la température moyenne, la température maximale, la température minimale, l'humidité relative, les précipitations, la pression atmosphérique, la pression de la vapeur d'eau et la vitesse du vent. Le modèle élaboré a été comparé au modèle de Bristow-Campbell amélioré (IBC) et au modèle d'Angström-Prescott amélioré (IA-P) afin d'explorer le meilleur modèle à la prédiction de l'irradiation solaire globale. Le modèle ANN avec le réseau de 9-17-1 fournit une meilleure précision et il offre les performances suivantes ; $RMSE=1.34MJ/m^2$, $MBE= -0.11MJ/m^2$, and $R^2=0.91$ concernant la prédiction journalière ISGJ. Pour la prédiction des moyennes mensuelles journalières ISGM ; le RMSE est de $1.07MJ/m^2$, le MBE est de $0.62MJ/m^2$ et le coefficient de corrélation est supérieur à 0.95

Dans cette même année, Wu et al [50] proposent un nouveau modèle hybride à base de SOM-OPELM et de stratégies des séries temporelles pour prédire le rayonnement solaire global horizontal en Chine. Le modèle SOM-OPELM combine une technique d'apprentissage statique extrême OPELM (The Optimally Pruned Extrême Learning Machine) avec les Carte Auto-Organisatrice SOM(Self-organizing map). Tel que le SOM divise les données originales en groupes distincts et le OPELM établit le modèle de prédiction. Les trois stratégies de séries chronologiques ; à savoir la stratégie récursive, la stratégie direct-récurif (DirRec) et la stratégie DirRec MISMO (multi-input and multi-output with flexible horizon S) sont adoptées pour réaliser la prédiction multi-pas. L'auteur a même effectué une comparaison de son modèle avec d'autres méthodes et les résultats de simulation montrent que le modèle SOM-OPELM proposé avec la stratégie DirRec ou la stratégie MISMO surpasse les modèles suivants: récursive-BP, DirRec-BP, MISMO-BP, récursif-SOM-OPELM et ARIMA. Concernant les journées clair il atteint un coefficient de détermination de 0.96, pour les journées non clair le $R^2=0.83$ et le MAPE varie entre 29% et 43%. Comme inputs, il a utilisé 8 variables ;

La moyenne horaire de l'angle zénithal (deg).

La moyenne horaire de l'angle d'azimut (deg).

Le rayonnement hors atmosphère horaire sur une surface horizontale (Wh / m^2).

La moyenne horaire du rayonnement solaire direct (Wh / m^2).

La moyenne horaire du rayonnement solaire diffus (Wh / m^2).

La moyenne horaire du rayonnement solaire global d'un ciel clair (Wh / m^2).

La moyenne horaire du rayonnement solaire direct d'un ciel clair (Wh / m^2).

La moyenne horaire du rayonnement solaire diffus d'un ciel clair (Wh / m^2).

Toujours en 2016, Wang et al [51] développent trois types de réseaux de neurones artificiels ; PMC (Perceptron Multi Couche), RNRG (Réseau de Neurone à Régression Généralisée) et RNBR (Réseau de Neurone à fonction de Base Radiale) pour prédire ISGJ en utilisant comme entrées aux modèles plusieurs variables météorologiques ; la température de l'air, la pression atmosphérique, l'humidité relative, la pression de la vapeur d'eau et la durée d'ensoleillement, observées dans 12 stations de différentes zones climatiques en Chine au cours des années 1961-2014. Le modèle de Bristow-Campbell amélioré (IBC) a aussi été élaboré pour cette même base de données afin de prédire l'ISGJ pour chacun des sites étudiés. Les résultats ont montré que les performances des quatre modèles agissent différemment d'un site à un autre. Le RMSE varie de 2.19-4.58 MJ/m² et le R² est entre 0.52 et 0.92.

Enfin, Premalatha et al [52] ont comparé entre quatre modèles de RNA doté chacun d'un algorithme distinct (GD, LM, SCG, RP) pour prédire l'ISGM dans cinq différentes stations en Inde. Le modèle avec l'algorithme de Levenberg-Marquardt (LM) a donné de meilleurs résultats comparés aux autres avec un coefficient de détermination dépassant les 98% et un MAPE maximale de 4.24%. Les entrées utilisées ainsi que l'architecture du réseau sont représentées sur la Figure 5 suivante.

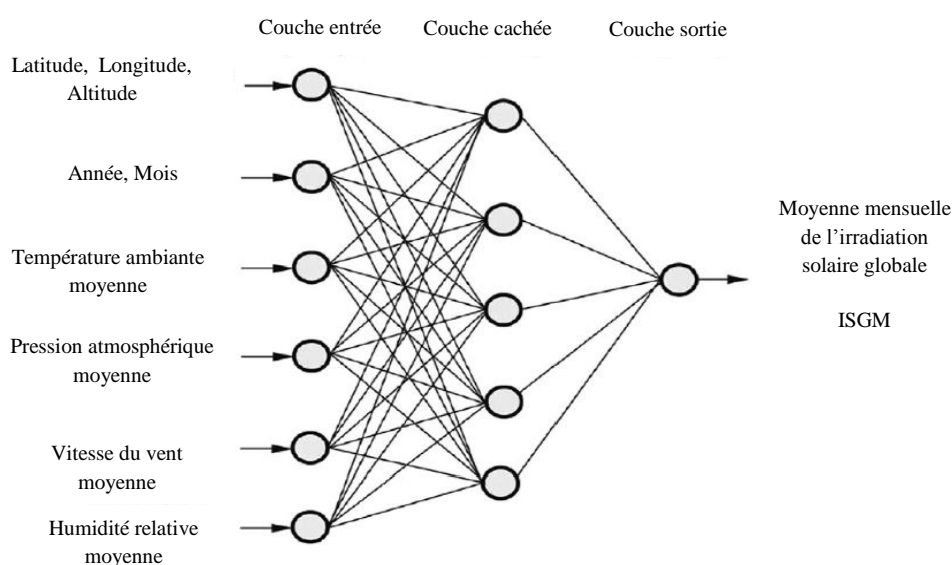


Figure 5 : Type de réseau utilisé par Premalatha [52]

Nous rapportons dans ce qui suit un tableau comparatif qui a été publiée par Premalatha [52] et qui résume quelques résultats obtenues par nos collègues concernant la prédiction de l'ISGM par les RNA (Tableau 1).

Tableau 1 : Comparaison des MAPE, publiés dans la littérature et dédiés à la prédiction de l' ISGM. [52]

Etude	site	MAPE (%)	Méthode/algorithme
Mohandes, Rehman, and Halawani (1998)	Kwash (Arabie Saoudite)	19.1	RNA/MLFF
Rehman and Mohandes (2008)	Abha (Arabie Saoudite)	11.8	RNA /MLFF
Alawi and Hinai (1998)	Majees (Nord Oman)	7.30	RNA /MLFF
Sözen, Arcaklioğlu, Özalp, and Caglar (2005)	Sirt (Turquie)	6.78	RNA /SCG
Sözen et al. (2004)	Mugla (Turquie)	6.73	RNA /LM
Premalatha (2016) [52]	Mumbai (Inde)	4.24	RNA /LM

MLFF; Multilayer Feed Forward
 GD; algorithme du Gradient descendant
 LM; algorithme de Levenberg-Marquardt

SCG ; algorithme du Gradient Conjugué
 RP ; résilient back propagation algorithme

2.2 Modèles prédictifs à base de Support Vector Machine

Les SVM (Support Vector Machine), souvent traduit par l'appellation de Séparateur à Vaste Marge, sont une classe d'algorithmes d'apprentissage initialement définis pour la discrimination (classification). Cette approche découle directement des travaux de Vapnik en théorie de l'apprentissage en 1995. Par la suite ces techniques d'apprentissage ont été généralisées pour des problèmes de régressions (prédiction).

Sauf que l'application des SVM dans la météorologie est très récente et les toutes premières investigations relatif à la prédiction de la composante solaire par les SVM ont été mené qu'en 2011 par Chen et al [6] et qui ont développé un modèle prédictif de l'irradiation solaire globale moyenne mensuelle ISGM en utilisant sept combinaisons de températures de l'air en entrée (la température maximale et minimale) tout en combinant également trois différentes fonctions noyaux qu'on peut utiliser dans la technique SVM à savoir ; la fonction linéaire, la fonction polynomiale et la fonction de base radiale (FBR) et il a été constaté que le modèle SVM avec le noyau polynomiale ainsi que la combinaison T_{max} et T_{min} comme variables d'entrées donne de meilleurs résultats avec un RMSE de 0,833MJ / m² et un RRMSE de 9,00%, aussi un plus élevé NSE (coefficient Nash-Sutcliffe) qui est de 0,999 et un coefficient de détermination $R^2= 0,969$. Par la suite ces mêmes auteurs [53] proposent d'appliquer les SVM pour estimer l'irradiation solaire globale journalière ISGJ en utilisant différentes combinaisons des entrée mais cette fois ci déduites a partir de la durée

d'ensoleillement, aussi cinq modèles empiriques à base de la durée d'ensoleillement sont évalués en exploitant des données météorologiques de trois différentes stations dans la province du Liaoning en Chine. Leur étude a montré que les modèles SVM sont plus performants que les modèles empiriques avec un RMSE $< 2,4 \text{ MJ / m}^2$ et un RRMSE $< 18\%$.

Toujours dans ce même contexte et vu que les travaux ne sont pas nombreux, on prendra le soin de les citer :

En 2013, Zeng et al [54], proposent de prédire la transmissivité atmosphérique à court terme (heure) pour en déduire le rayonnement solaire à partir de la latitude du site et du temps horaire. Comme entrées ils ont utilisé la représentation bidimensionnel 2D* de la transmissivité atmosphérique, la couverture nuageuse, l'humidité relative ainsi que la vitesse du vent. Aussi, ils ont mené une étude comparative avec d'autres techniques tels les modèles AR (modèles AutoRegressifs) et les modèles RNFBR (Réseau de Neurones à Fonction de Base Radiale) et il s'avère que les performances du modèle développé sont meilleurs avec un coefficient de corrélation de 0.974.

* Représentation unidimensionnelle 1D: le rayonnement solaire en fonction du temps t (heure)

* Représentation bidimensionnelle 2D: le rayonnement solaire en fonction de deux temps différents ; l'heure dans la journée (TOD time of day) et le jour dans l'année (DOY day of year)

En 2014, Ekici [55] a prédit, par les SVM, l'irradiation solaire globale journalière du jour J en fonction de cinq entrées du jour $J-1$ et qui sont ; le numéro du jour (à partir du 1 janvier), la température moyenne journalière, la température maximale journalière, la durée d'ensoleillement ainsi que l'irradiation solaire globale journalière. Son modèle est appliquée pour la Turquie et son coefficient de corrélation est de 99.294%.

En 2015, Mohammadi et al [56] ont combiné les SVM avec l'algorithme de la transformée en ondelettes (TO) pour prédire l'irradiation solaire global horizontal journalière (ISGJ) et moyenne mensuelle (ISGM) en Iran. Les différentes entrées utilisées sont; la durée d'ensoleillement relative (S / S_0) qui est le rapport entre la durée d'ensoleillement (S) et la durée maximale possible (S_0), la différence entre les températures ambiantes maximale et minimales ($T_{\max} - T_{\min}$), l'humidité relative (R_h), la pression de la vapeur d'eau (P_v), la température ambiante moyenne (T_{moy}) et l'irradiation solaire global hors atmosphère sur une

surface horizontale (H_0). Les performances du modèle sont tel que le MAPE = 7,000 %, le MABE = 0.841 MJ/m², et le $R^2 = 0,909$ pour le cas journalier. Concernant l'estimation des moyennes journalières mensuelles, les tests statistiques MAPE, le MABE, et le R^2 donnent respectivement 3,260%, 0,510 MJ/m² et 0,974.

Ensuite, ces même auteurs [57], ont contribué à la prédiction de l'ISGJ en Iran par les SVM avec deux fonction noyaux différentes ; polynomiale et radiale basique et qu'ils comparent aussi aux autres modèles de RNA et les ANFIS. Les entrées aux modèles sont ; la température minimale, la température maximale, la durée d'ensoleillement (S), la durée maximale possible d'ensoleillement (S_0), le numéro du jour de 1 à 365, l'irradiation solaire par journée clair et l'irradiation solaire global hors atmosphère sur une surface horizontale (H_0). Il en résulte que le meilleur modèles est le modèle SVM avec la fonction noyau de base radiale RBF (Radial Basic Function) avec un RMSE=3.2Mj/m² et un $R^2=0.90$.

Toujours en 2015, Shamshirband et al [58] ont développé des modèles prédictifs de l'ISGM reçu sur un plan horizontal, a base des techniques SVM, RNA et GP*. Comme entrée ils ont utilisé différentes combinaisons tels ; la différence de température maximale et minimale, la température moyenne, le rapport entre la durée d'ensoleillement et la durée d'ensoleillement maximale (S / S_0) ainsi que l'humidité relative. Les résultats obtenus ont montré que les modèles ayant en entrées tous les paramètres météorologiques pour chacune des technique utilisée apporte une meilleure prédiction (MAPE, MABE, RMSE, RRMSE, et le R^2 sont de 2.252%, 0.434MJ/m², 0.588MJ/m², 2.976% and 0.987, respectivement) suivit des modèles ayant en entrée la durée d'ensoleillement S avec la température.

* GP ; programmation génétique (Genetic programing)

En 2016, Urraca et al [59] ont construit des modèles prédictifs de l'irradiation solaire globale horaire dans un site en Espagne. Deux types d'ensembles d'entrées ont été sélectionnés comme inputs aux modèles. Le premier ensemble s'agit des caractéristiques climatiques du site tel ; la température de l'air, la pluviométrie, l'humidité relative, la vitesse du vent ainsi qu'une variable booléenne M qui prendra la valeur 0 si la pluviométrie n'est pas enregistré (0) ou M=1 dans le cas contraire. Le deuxième ensemble d'entrées dépend des caractéristiques solaire du site ; l'irradiation solaire globale hors atmosphère, l'angle solaire azimutale, l'angle d'élévation solaire, l'angle solaire horaire, le cosinus de l'angle zénithal solaire, l'irradiation solaire globale à ciel clair, l'indice de clarté, l'indice de clarté à un ciel

clair et l'irradiation solaire globale aux instants t , $t-1$ et $t-2$. En tous il y'a 20 inputs. Deux techniques basées sur le principe de l'intelligence artificielles ont été choisies à l'élaboration de ce travail. La première est la technique SVR (Support Vector Regression) et la deuxième est la méthode RFs* (Random Forests) introduite par Breiman en 2001. Ainsi, plusieurs combinaisons des entrées ont été testées à l'entrée de ces deux modèles. Les résultats montrent que le modèle SVR avec seulement 7 entrées donne de meilleures performances comparées aux autres avec un MAE de 58.74W/m^2 et un RMSE= 100.55W/m^2 . Aussi, le modèle RFs avec 11 entrées donne également de bonnes performances tel que le MAE= 62.64W/m^2 et le RMSE= 96.21W/m^2 .

*RFs: Random Forests, en français ; forêts aléatoires

2.3 Modèle prédictifs suivant la méthodologie de Box & Jenkins

L'étude des séries temporelles (ST) semble avoir atteint sa maturité au cours des années 70 où des développements significatifs sont apparus. En 1965, Cooley et Tukey ont bien contribué dans l'étude spectrale des séries en publiant leur article '*an algorithm for the machine calculation of complex Fourier series*', introduisant la Transformée de Fourier Rapide (Fast Fourier Transform FFT), qui est un algorithme permettant de calculer rapidement des périodogrammes [60].

A la même époque, en 1970, Box et Jenkins ont publié leur ouvrage '*Time series analysis, forecasting and control*', montrant que l'étude des séries temporelles à l'aide de processus de type ARMA (ARIMA, SARIMA, AR, MA, ARCH, GARCH, ARMAX,...) pouvait s'appliquer à de nombreux domaines tel que : la météorologie, l'astronomie, l'économétrie,..., et pouvait être facilement implémentée informatiquement. Ces modèles sont aussi appelés modèles classiques ou de référence, de par le nombre d'études les ayant utilisés surtout dans la finance et l'économétrie [11].

Ces modèles n'ont été repris dans la modélisation des variables climatiques que ces dernières décennies, notamment dans la prédiction du rayonnement solaire globale.

Les toutes premières investigations ont été menées en 1977 par Hoh et al [61] en proposant une modélisation par approche stochastiques de la série chronologique de l'irradiation solaire globale horaire et présentant ainsi un modèle prédictif appropriée pour un site à Singapour.

Ensuite, en 1998 ;Mora-Lopez et al [62] proposent une méthodologie pour générer une série de données de l'irradiation globale horaire basée sur l'utilisation d'un modèle

multiplicative ARMA(1,0)x(0,1) avec des composantes régulières et saisonnières qui changent d'un mois à l'autre dans dix différentes régions en Espagne.

En 1999, Craggs et al [63] ont modélisé le rayonnement solaire reçu sur un plan horizontale ainsi que verticale à l'aide des modèles SARIMA pour un pas de temps de 10 mn, et les résultats sont satisfaisant. Pour ce qui est du pas de temps horaire, une réduction considérable des performances est enregistrée par rapport aux deux orientations. Par conséquent, leurs travaux ont montré que ces modèles peuvent être utilisées pour prévoir les changements rapides et importants dans la production des panneaux photovoltaïques. Le même constat est effectué en 2009 par Reikard et al [64] en prouvant que les performances de la prédiction via les ARIMA dépendent de la résolution des données, ou du pas de temps.

Par la suite, Certain auteurs ont dédié une grande partie de leur recherche à la prédiction des séries temporelles du rayonnement global. C'est le cas de Voyant , qui en 2012 il combine les ARMA et les MLP pour développer un modèle prédictif de la série temporelle de l'irradiation solaire globale horaire avec un NRMSE=14.9% alors qu'avec le MLP le NRMSE était de 18.4% [65]. Ensuite en 2013, il développe encore la même approche mais cette fois ci en présentant un modèle pour chacune des saison(hiver, automne, printemps et été) afin d'améliorer la prédiction [8]. Encore, en 2014, Voyant et al [66] tentent de trouver une meilleure méthodologie permettant de prédire avec une grande précision la série temporelle de l'irradiation solaire global horaire en ajoutant aux prédicteurs développés (ARMA, MLP) des règles bayésiennes.

Récemment en 2015, [67] ont développé des modèles prédictifs de la série temporelle de l'irradiation solaire en Chine, en associant les ARIMA aux modèles GARCH (Generalized Autoregressive Conditional Heteroskedastic) qui sont des approches largement utilisées pour modéliser la moyenne et la volatilité des séries chronologiques. Ces modèles GARCH sont particulièrement appliqué dans la finance ou les décisions financières.

2.4 Synthèse des performances des modèles prédictifs de l'irradiation solaire globale par les méthodes d'apprentissage statistique

2.4.1 Cas de l'irradiation solaire globale journalière ISGJ

Pour avoir une idée générale sur les performances des techniques d'apprentissage statistique (SVM et RNA) par rapport à la prédiction de l'ISGJ, une synthèse de quelques résultats est illustrée dans le Tableau 2 [7].

Chapitre 1 : Etat de l'art

Tableau 2 : Résultats de performance de quelques modèles prédictifs de l'ISGJ [7] (1/2)

Modèles	Entrées	Régions	Auteurs	Performances des modèles prédictifs
PMC	D, T _{mean} , S, RH, VP, WS	Iran	[40]	MAPE = 5.210 –12.860%
RNBR	D, T _{mean} , S, RH, VP, WS	Iran	[40]	MAPE = 5.560 –12.390%
RNA	T _{max} , T _{min} , H ₀	South Iran	[41]	RMSE = 2.534 MJ/m ² NRMSE = 13.820 % R = 0.889
RNA	D, Lat, Lon, H ₀ , TCC, SKT, TCW, TCO	Spain	[68]	RMSE = 3.016 MJ/m ² NRMSE = 14.200%
LLR	T _{max} , T _{mean} , H ₀ , WS, P	Iran	[39]	NRMSE = 22.400 % R = 0.900
PMC	T _{max} , T _{mean} , H ₀ , WS, P	Iran	[39]	NRMSE = 28.100 % R = 0.830
RNA	D, Tmax	Saudi Arabia	[36]	MAPE = 10.300 %
RNA	D, T _{mean}	Saudi Arabia	[36]	MAPE = 11.800 %
RNA	D, T _{mean} , RH	Saudi Arabia	[36]	MAPE = 4.490 %
RNB	T _{mean} , RH, S, H ₀	(Al-Madinah) Saudi Arabia	[43]	MAPE = 5.939 % NRMSE = 9.318 % R = 0.972
RNA	D, S	(Al-Madinah) Saudi Arabia	[69]	R = 0.974
RNA	D, S, T _{mean}	(Al-Madinah) Saudi Arabia	[69]	R = 0.977
RNA	D, S, T _{mean} , RH	(Al-Madinah) Saudi Arabia	[69]	R = 0.975
RNA	D, T _{mean}	(Al-Madinah) Saudi Arabia	[69]	R = 0.892
RNA	D, RH	(Al-Madinah) Saudi Arabia	[69]	R = 0.870
RNA	D, T _{mean} , RH	(Al-Madinah) Saudi Arabia	[69]	R = 0.890
RNA	T _{mean} , T _{max} , T _{min} , S, RH, P, VP, WS, Ps	Chine	[49]	RMSE = 1.340 MJ/m ² NRMSE = 11.250% R = 0.951
PMC	T _{max} , T _{min} , S, RH, P, VP	12 Station Chine	[51]	RMSE = 1.94-3.27 MJ/m ² R = 0.854-0.959
RNBR	T _{max} , T _{min} , S, RH, P, VP	12 Station Chine	[51]	RMSE = 1.96-3.25 MJ/m ² R = 0.849-0.959

Tableau 2 : Résultats des performances de quelques modèles prédictifs de l'ISGJ [7] (2/2)

RNRG	T_{max} , T_{min} , S, RH, P, VP	12 Station Chine	[51]	RMSE = 2.00-3.29 MJ/m ² R = 0.849-0.949
PMC	T_{max} , T_{min} , T_{mean} , H_0	Ghardaïa (Algeria)	[7]	MAPE = 10.504 RMSE = 2.871 MJ/m ² NRMSE = 13.612% R = 0.889
SVM	T_{max} , T_{min} , T_{mean} , H_0	Ghardaïa (Algeria)	[7]	MAPE = 10.458 RMSE = 2.779 MJ/m ² NRMSE = 13.172% R = 0.896
Ps : précipitation		RNRG: Réseau de Neurone à Régression		
SVM : Séparateurs à Vaste Marge		Généralisée		
RNA : Réseau de Neurone Artificiel		RNBR: Réseau de Neurone à fonction de Base		
RNB : Réseau de Neurone Bayésien		Radiale		
PMC : Perceptron Multi Couche		RNR : Réseau de Neurone Récurent		

2.4.2 Cas de l'irradiation solaire globale journalière moyenne mensuelle ISGM

Le Tableau 3 récapitule les résultats de performances de quelques modèles basés sur les techniques de l'apprentissage statistique et qui permettent la prédiction de l'irradiation solaire globale journalière moyenne mensuelle ISGM.

2.5 Influence des paramètres d'entrées sur la prédiction de l'irradiation solaire

Suite à tous ce qui a été évoqué précédemment par rapport au contexte de la prédiction, il est bien clair que pour prévoir le comportement ou la réponse d'un quelconque phénomène que ce soit par les techniques de l'apprentissage statistique ou par les modèles empiriques, on a besoin de variables entrantes aux modèles ou ce qu'on appelle les inputs (entrées). Donc, pour bien illustrer l'influences de ces variables d'entrées sur la pertinence des modèles prédictifs et plus précisément ceux de l'irradiation solaire qui fait l'objet de notre étude, on a pris le soin de reprendre un résumé d'une étude faite sur le choix des paramètres d'entrée utilisées pour la prévision de l'irradiation solaire que Mohammadi et al [70] ont publié dans leurs article, auquel nous avons apporté quelques changements et mise à jour (Tableau 4) en ajoutant d'autres études réalisées après l'apparition de cette article.

D'après le Tableau 4, on remarque que la durée de l'ensoleillement (S) est un paramètre très utilisé et influence positivement dans la prédiction de l'irradiation solaire, quoi qu'il est à noter que sa mesure peut être faussée à cause du mauvais calibrage de l'appareil. Sachant que S est la somme des périodes durant lesquelles le rayonnement solaire direct dépasse les 120W/m² [71].

Chapitre 1 : Etat de l'art

Tableau 3 : Résultats des performances de quelques modèles prédictifs de l'ISGM [7]

Modèles	Entrées	Régions	Auteurs	Performances des modèles prédictifs
RNA	L, M, MP, MT, MVP, MRH, MWS, MS	Sultanat D' Oman.	[2]	MAPE = 5.45 - 7.30 % R = 0.950
CS-hard- ridge-RBF	12 Paramètres météorologiques		[72]	MAPE = 5.44 - 8.36 % RMSE = 280.93 – 387.21 Wh/m ²
RNBR	12 Paramètres météorologiques		[72]	MAPE = 9.98 - 17.05 % RMSE = 450.13 – 943.51 Wh/m ²
RNA -MTM	Lon, Lat, Alt, Kt	Algerie (4 locations)	[73]	NRMSE = 5 -8 % R= 0.900 – 0.920
PMC	Lon, Lat, Alt	Algerie	[34]	MAPE = 3.10 % R= 0. 950
RNBR	Lon, Lat, Alt	Algerie	[34]	MAPE = 3.30 % R=0. 960
RNR	Lon, Lat, Alt	Algerie	[34]	MAPE = 3.20 % R=0. 960
ANFIS	Lon, Lat, Alt	Algerie	[34]	MAPE = 2.20 % R= 0.970
PMC	Lon, Lat, Alt	10 cites en Arabie Saudite.	[74]	MAPE = 6.50 – 19.10 %
RNBR	Lon, Lat, Alt, MS	10 cites en Arabie Saudite.	[74]	MAPE = 4.90 – 15.50 %
RNA	Lon, Lat, Alt, MS, TCC, MTmax	Uganda	[35]	RMSE = 0.385 MJ/m ² R=0. 970
RNA	Lon, Lat, Alt, M, MS, MTmean	Siirt (Turquie)	[4]	MAPE = 6.78 % R= 0.998
RNA	Tmean, Tmax, Tmin, S, RH, P, RH, VP, WS, Ps	Chine	[49]	RMSE = 1.066 MJ/m ² NRMSE = 8.905 % R= 0.977
PMC	MT _{min} , MH ₀	Ghardaia (Algerie)	[7]	MAPE = 8.95 % RMSE = 1.596 Mj/m ² NRMSE = 7.510 % R= 970
SVM	MT _{min} , MH ₀	Ghardaia (Algerie)	[7]	MAPE = 8.94 % RMSE = 1.524 Mj/m ² NRMSE = 7.442 % R= 986

Chapitre 1 : Etat de l'art

Tableau 4 : Influence des variables d'entrées sur les performances prédictifs de l'irradiation solaire (1/3)

Modèles	Sites	Entrées considérées	Paramètres d'entrée pertinents	Sortie	Références
ANN	Chipre (Athalassa)	5 combinaisons : S, S ₀ , T _{max} et M	S, S ₀ et T _{max}	ISG	[75]
ANN-ARD	Espagne (12 stations dans le sud)	7 combinaisons : Lat, Long, Alt, l'inclinaison, angles d'azimut, H ₀ , K _t et D	Alt, D et K _t	ISG	[5]
ANN	Arabie Saoudite (Al-Medina)	6 combinaisons : S, T _{moy} , HR et D	S et T _{moy}	ISG	[38]
ANN	Iran (Dezful)	6 combinaisons : D, T _{moy} , HR, S, EV et WS	D, T _{moy} , HR, S et WS	ISG	[40]
ANN	Turquie (7 sites à Antalya)	6 combinaisons : Lat, Long, Alt, M, Nu, T _{moy} , HR, WS, S	lat, Long, Alt, M, Nu et S	ISG	[76]
BNN	Arabie Saoudite (Medina)	10 combinaisons : T _{moy} , HR, S et H ₀	S a été identifié comme étant le paramètre le plus pertinent, suivie de T _{moy}	ISG	[43]
ANN-MNLR	Turquie (31 stations)	10 combinaisons : Lat, Long, Alt, M, MT _{moy} , MT _{min} , MT _{max} , TS _{min} , TS _{max} , HR, WS, Pr, Pa, Pv, Nu et S	S, M, Nu, MT _{moy} , MT _{min} , MT _{max} , TS _{min} , TS _{max} , WS, Alt, Lat	ISG	[77]
GP-SA	Iran (Tehrani et Kerman)	T _{max} , T _{min} , HR, Pr et S	T _{max} , T _{min} influencent sur le modèle plus que les autres variables climatiques	ISG	[78]
HMGF	Inde (Gurgaon)	16 combinaisons : S, T _{moy} , HR, WS et Pa	S a été déterminé comme l'élément le plus important suivi par la T _{moy} , HR, Pa et WS	ISG	[79]
ANN	Iran (Téhéran)	5 combinaisons : T _{max} , T _{min} , T _{moy} , S, S ₀ , H ₀ , Pr, et D	T _{max} , T _{min} , S, S ₀ , H ₀ et D	ISG	[80]
NGA	Argentine (El Colmenar Tucumán)	Plusieurs combinaisons : T _{max} , T _{min} , T _{moy} , D, HR, Pa, Nu et S	Les auteurs n'ont pas spécifié	ISG	[81]
WEKA-ANN	Inde (26 sites à différents climatiques)	Différentes combinaisons : Lat, Long, Alt, T _{max} , T _{min} , T _{moy} , S	Alt, T _{max} , T _{min} et S	ISG	[82]

Chapitre 1 : Etat de l'art

Tableau 4 : Influence des variables d'entrées sur les performances prédictifs de l'irradiation solaire (Suite 2/3)

Modèles	Sites	Entrées considérées	Paramètres d'entrée pertinents	Sortie	Références
SVM-WT	Iran (Bandar Abass)	3 combinaisons : S, T_{diff} , HR, Pv, T_{moy} et H_0	S, T_{diff} , HR, Pv, T_{moy} et H_0	ISG	[56]
WEKA-ANN	Inde (26 locations dans l'état Himachal Pradesh)	Différentes combinaisons : T_{moy} , S, Lat, Long, Alt Kt, et H_0	T_{moy} , S et Alt	ISG	[83]
ANN-ARD	Le désert des USA	Différentes combinaisons : T_{moy} , HR, Pa, Tr, Pr, l'angle zénithal solaire, k_t , MAr et WS	K_t et MAr	ISD	[84]
ANFIS	Iran (Kerman)	Différentes combinaisons : H, S_0 , S, T_{max} , T_{min} , T_{moy} , HR, Pv, H_0 et l'angle de déclinaison solaire	Cas 1 entrée ; S Cas 2 entrées ; H et H_0 Cas 3 entrées ; H, H_0 et S	ISDF	[70]
ANN	Sud-Est de la Chine	Différentes combinaisons : S, T_{max} , T_{min} , T_{moy} , HR, Pr, Pa, Pv, WS	S, T_{max} , T_{min} , T_{moy} , HR, Pr, Pa, Pv, WS	ISG	[49]
SVR Et RFs*	Espagne	Différentes combinaisons; T_{moy} , Pr, HR, WS, H_0 , variable booléenne M (M=1 si Pr \neq 0, M=0 si Pr=0), l'angle solaire azimutale, l'angle d'élévation solaire, l'angle solaire horaire, le cosinus de l'angle zénithal solaire, Hc, Kt, Ktc et H aux instants t, t-1 et t-2	Cas du modèle SVR ; H (à t et t-2), Ktc (t), Hc (t+2), variable booléenne M (t), Pr (t) et T_{moy} (t) Cas du modèle RFs ; H (t), Ktc (t, t-1, t-2), Kt (t et t-1), Hc (t+2), l'angle solaire horaire (t+2), WS (t), T_{moy} (t) et HR	ISG	[59]
ANFIS	Iran (Isfahan, Kerman et Tabass)	Différentes combinaisons; S, S_0 , T_{max} , T_{min} , T_{moy} , HR, Pv, Pa et H_0	Cas 1 entrée : <ul style="list-style-type: none"> • S ou S_0 • Pv est le moins pertinent pour toutes les villes. Cas 2 entrées ; <ul style="list-style-type: none"> • site Isfahan ; S et H_0. • sites de Kerman et Tabass : S et S_0 cas 3 entrées ; <ul style="list-style-type: none"> • sites d'Ispahan et Kerman ; S_0, S et T_{max} • site de Tabass : S_0, S et H_0 	ISG	[85]

Chapitre 1 : Etat de l'art

Tableau 4 : Influence des variables d'entrées sur les performances prédictifs de l'irradiation solaire (Suite 3/3)

Modèles	Sites	Entrées considérées	Paramètres d'entrée pertinents	Sortie	Références
W-SVM	Australie (sites métropolitains et régionaux)	Différentes combinaisons; S, T_{max} , T_{min} , Pv, WS et Pr	S : variable très importante Certains site ; la combinaison de tous ces paramètres est optimale Autres sites ; S, T_{max} , T_{min} , et Pv	ISG	[86]
ANN	Zimbabwe	Différentes combinaisons; Lat, Long, Alt, HR, Pa, Kt et T_{moy}	Kt	ISG	[87]
SVM	Sud de l'Algérie (Ghardaïa)	42 combinaisons ; T_{max} , T_{min} , T_{moy} , T_{diff} , S_0 et H_0	Prédiction journalière ; T_{max} , T_{min} , T_{moy} et S_0 ou T_{max} , T_{min} , T_{moy} et H_0 Prédiction mensuelle ; T_{min} et H_0	ISG	[7]
ANN	Italie	Plusieurs combinaisons de 13 paramètres géographiques et climatologiques.	H_0 , S, NJP, Alt, Lat, Pr et période de temps	ISG	[88]
M: Mois		Pv : Pression de la Vapeur d'eau	WEKA: Waikato Environment for Knowledge Analysis	WEKA: Waikato Environment for Knowledge Analysis	
D: Numéro du jour de l'année		MT : Température Moyenne mensuelle	NJP : Nombre de Jour de Pluie	Hc : Irradiation solaire globale à ciel clair	
Ev : Evaporation		Tr : Température de Rosée	GP/SA : Genetic Programming Simulated Annealing	Ktc : Indice de clarté à ciel clair	
WS : Vitesse du Vent		MAR : Masse d'Air relative	ARD: Automatic Relevance Determination	HMGF: Hidden Markov-Generalized Fuzzy	
Nu : Nébulosité		NGA: Niching Genetic Algorithms	W-SVM: Wavelet Support Vector Machine		
Pr : Précipitation		MNLR: Multi-Nonlinear Regression			
Pa : Pression atmosphérique					

3. Présentation climatique du site étudié

3.1 L'irradiation solaire globale journalière hors atmosphère in situ (Ghardaïa)

La latitude du site de Ghardaïa est: $Lat = 32.23'N$

La Figure 6 représente l'évolution de l'irradiation solaire globale journalière hors atmosphère pour Ghardaïa.

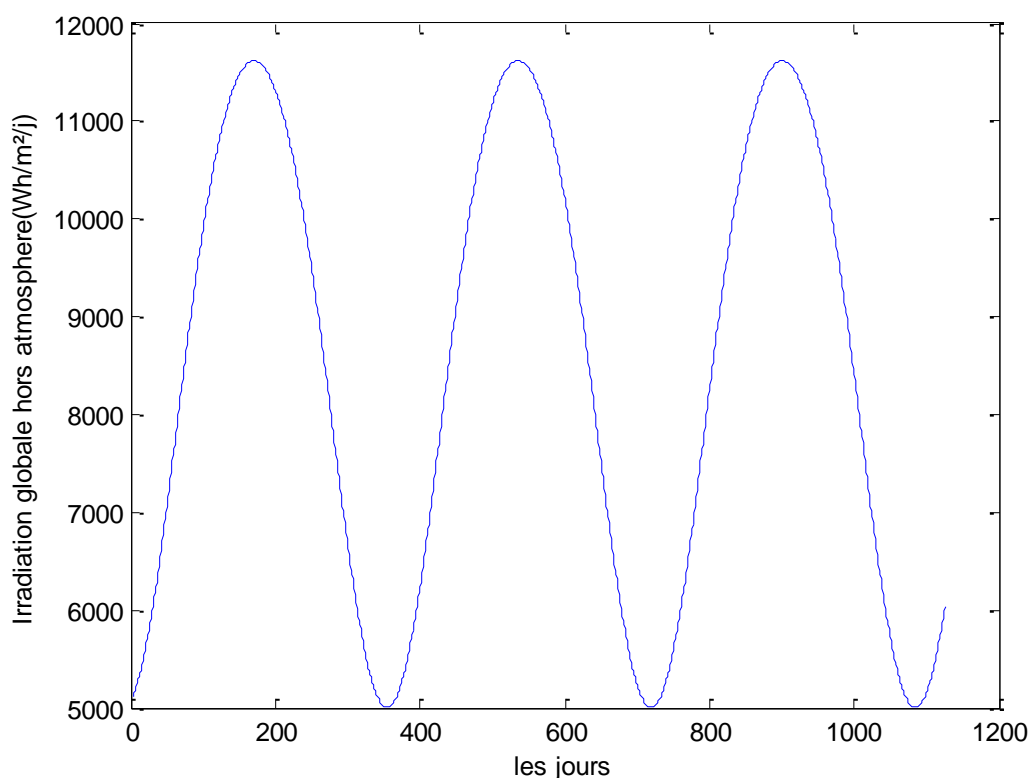


Figure 6 : l'irradiation solaire globale hors atmosphère G_0 par rapport au site de Ghardaïa

3.2 Présentation de la station radiométrique installée sur le site de l'URAER

C'est une station météorologique automatique conçue pour les réseaux d'observation météorologiques, supportant tous les capteurs nécessaires et la mesure du vent à 10 m. La technologie utilisée dans cette station garantit les performances exceptionnelles du système dans toutes les conditions climatiques même les plus rudes. Aussi, elle fonctionne sans intervention humaine avec de nombreux supports de télécommunications et permet une grande précision dans l'exploitation des données.

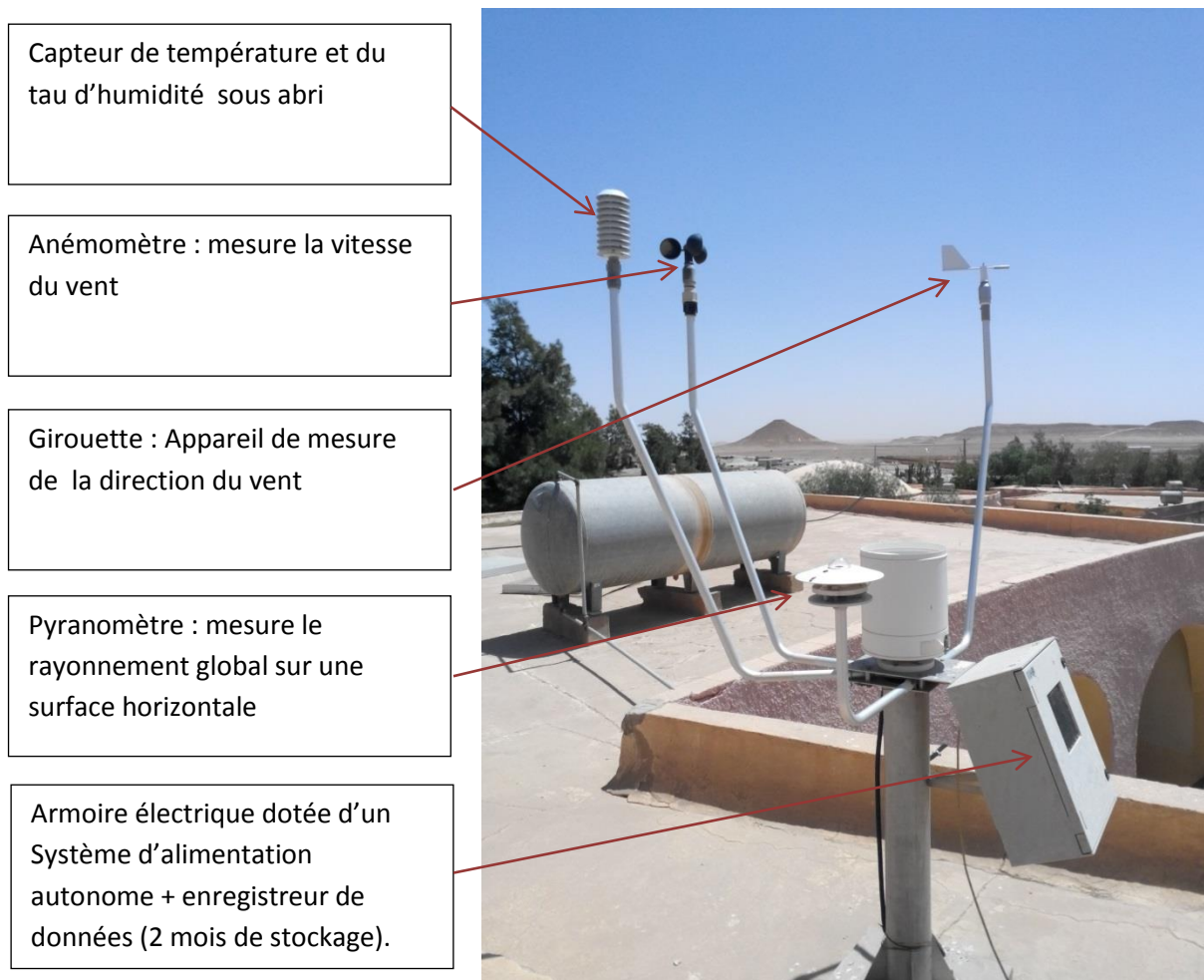


Figure 7 : Présentation de la station radiométrique installée sur le toit de L'URAER

3.3 Le prétraitement des mesures effectuées

Les mesures du rayonnement solaire global (W/m^2) au sol par rapport à une surface horizontale ont été prises à pas de temps de 1 heure, sans arrêt y compris la nuit.

La Figure 8 montre l'acquisition des données sur un fichier Excel.

Chapitre 1 : Etat de l'art

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Date	Heure	T	U	RR	RRX	IRRX	VT	VX	GIX	VM	GV	RG	P	PS	
2	09/02/2012	00:00	4,3	50,0	0,0	0	0	8	5	360	3	360	0	966,0	0	
3	09/02/2012	01:00	3,6	52,0	0,0	0	0	8	4	10	2	360	0	966,1	0	
4	09/02/2012	02:00	3,4	53,0	0,0	0	0	7	4	330	2	330	0	966,1	0	
5	09/02/2012	03:00	2,3	57,5	0,0	0	0	7	4	350	2	340	0	966,2	0	
6	09/02/2012	04:00	1,5	60,5	0,0	0	0	9	5	330	3	340	0	966,2	0	
7	09/02/2012	05:00	1,7	62,5	0,0	0	0	8	5	320	2	340	0	966,6	0	
8	09/02/2012	06:00	2,2	62,0	0,0	0	0	7	4	300	1	330	3	967,2	0	
9	09/02/2012	07:00	3,5	60,0	0,0	0	0	8	5	330	2	340	56	967,7	0	
10	09/02/2012	08:00	6,0	53,0	0,0	0	0	7	4	30	2	20	127	968,2	0	
11	09/02/2012	09:00	7,0	49,5	0,0	0	0	9	5	60	3	50	189	968,3	0	
12	09/02/2012	10:00	8,6	44,5	0,0	0	0	10	6	50	2	50	212	968,3	0	
13	09/02/2012	11:00	9,2	39,5	0,0	0	0	8	5	60	2	110	230	967,5	0	
14	09/02/2012	12:00	9,2	34,5	0,0	0	0	8	7	60	2	70	245	966,2	0	
15	09/02/2012	13:00	10,1	33,0	0,0	0	0	8	6	50	2	120	135	965,9	0	
16	09/02/2012	14:00	10,0	32,0	0,0	0	0	7	4	110	2	110	84	965,5	0	
17	09/02/2012	15:00	9,5	33,0	0,0	0	0	7	5	140	2	150	45	965,8	0	
18	09/02/2012	16:00	9,2	33,5	0,0	0	0	7	4	150	2	140	21	965,7	0	
19	09/02/2012	17:00	8,9	34,5	0,0	0	0	5	4	180	1	140	0	965,7	0	
20	09/02/2012	18:00	8,7	35,5	0,0	0	0	3	2	160	1	190	0	966,2	0	
21	09/02/2012	19:00	8,4	36,5	0,0	0	0	3	2	290	1	290	0	966,3	0	
22	09/02/2012	20:00	8,5	36,5	0,0	0	0	2	2	50	0	80	0	966,4	0	

Figure 8 : Exemple d'acquisition de données radiométriques in situ

Ensuite, un calcul intégral de ces données est effectué par rapport au temps et ceci du levé jusqu'au coucher du soleil afin d'avoir une énergie cumulée par rapport au temps (une journée) qui est l'irradiation solaire globale journalière estimée en Wh/m²/jour.

La Figure 9, présente un exemple de mesures prises du rayonnement solaire global durant trois journées du mois de Septembre avec des climats différents ; journée clair (bien ensoleillée) (a), journée partiellement nuageuse (b) et une journée perturbée (mauvais temps)(c).

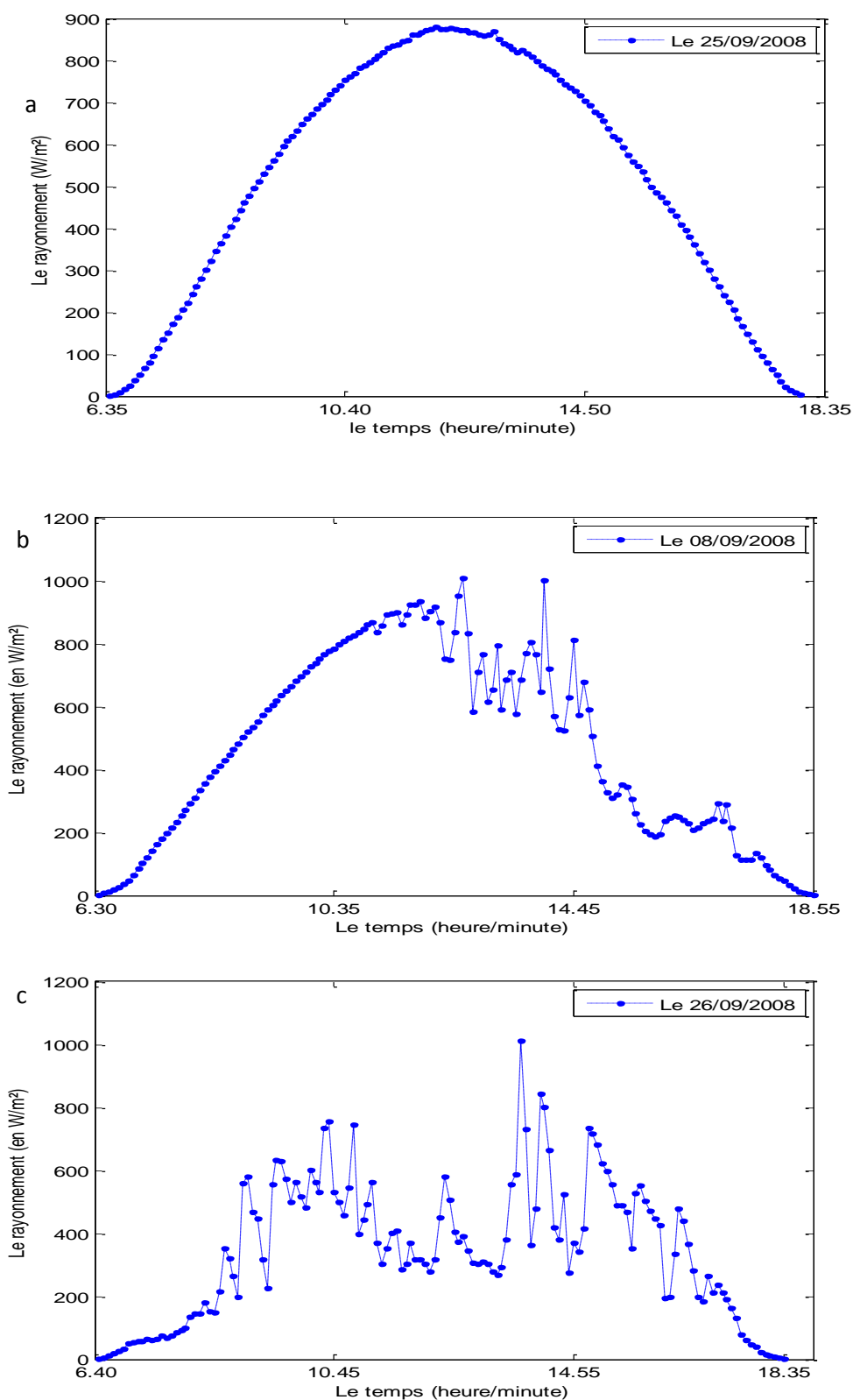


Figure 9 : Le rayonnement solaire global horizontal pour différentes types de journée à Ghardaïa

CHAPITRE 2 : Méthodologie

1. Les Support Vector Machine SVM

Les travaux de V. Vapnik, publiés en 1979, sur la théorie statistique de l'apprentissage (cf. Chapitre 1 : Modélisation) peuvent être considérés comme le point de départ de ce qui devint par la suite les machines à vecteurs supports (ou SVM pour Support Vector Machines). L'intérêt pour ce sujet ne prit de l'ampleur qu'au début des années 1990, en réponse à différents problèmes d'apprentissage supervisé [89].

Les machines à vecteurs supports ont été inventées en 1992, mais le terme « machines à vecteurs supports » n'est apparu qu'en 1995. Depuis lors, de nombreux développements ont été réalisés pour adapter cette méthode d'apprentissage à la régression, pour établir des liens avec les méthodes de régularisation, ou pour proposer des variantes, qui, comme les SVM, entrent dans la catégorie des méthodes de noyaux, et qui en reprennent les principes essentiels [89].

Dans de nombreuses applications, ces outils de modélisation se sont avérés très performants, surpassant parfois d'autres méthodes, notamment les réseaux de neurones, pour des problèmes de classification [90] ou de régression [91].

Vu que cette méthodologie SVM tient une part importante dans les travaux développés dans cette thèse, alors, nous avons décidé de présenter un peu plus en détail la technique de cette approche.

1.1 Principe générale des SVM

Pour expliquer le principe général de fonctionnement des SVM, on prend un problème de discrimination à deux classes ; cas linéairement séparable (Figure 10). L'idée est de chercher une frontière de décision entre les exemples de chaque classe. Mais, il est clair qu'il existe une multitude de classificateurs valide (Figure 10-a). Cependant, la propriété remarquable des SVM est que ce dernier doit être optimal. Ceci revient donc à trouver un classificateur qui va séparer à la fois les données et maximiser la distance entre ces deux classes (Figure 10-b). On appelle cette distance « marge » optimale, d'où l'appellation de Séparateurs à Vaste Marge et ce séparateur est appelé hyperplan.

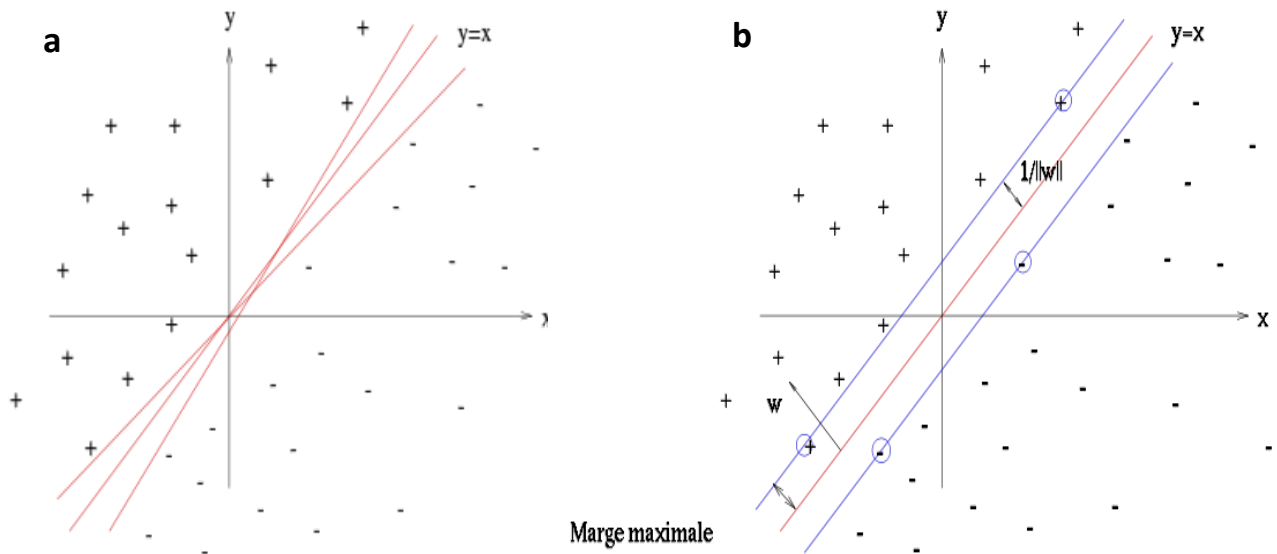


Figure 10 : Hyperplan optimale et marge maximale

Les échantillons entourés sont des vecteurs supports.

Lorsque, nous sommes devant un cas non linéairement séparable, l'idée était de faire une application de l'espace des entrées $x \in R^N$ vers un espace $\phi(x) \in R^{N'}$ de plus grande dimension $N' > N$, appelé espace des caractéristiques ou encore espace des représentations, où la tâche serait linéairement séparable ([COVER 1965]).

Soit :

$$x \rightarrow \phi = (x_1 \ x_2 \ \dots \ x_N \ x_1^2 \ x_1 x_2 \ \dots \ x_1 x_N \ x_2^2 \ x_2 x_3 \ \dots \ x_{n-1} x_N \ x_N^2) \quad (6)$$

La dimension du vecteur ϕ qui est $N' = N + N(N + 1)/2$.

Exemple : supposons que $N = 2$, et que l'on ait choisi l'espace des caractéristiques quadratiques défini par (1). On aura :

$$x \rightarrow \phi = (x_1 \ x_2 \ x_1^2 \ x_1 x_2 \ x_2^2)^T \quad (7)$$

D'où : $N' = 5$ et en tenant compte du seuil w_0 , les poids d'un perceptron dans cet nouvel espace caractéristique aura 6 composantes.

$$w \cdot \phi(x) = w_0 + w_1 x_1 + w_2 x_2 + w_{11} x_1^2 + w_{12} x_1 x_2 + w_{22} x_2^2 \quad (8)$$

Tel que ; les indices de chaque poids rappellent la forme du monôme qu'il pondère.

NB : Comme pour le perceptron, qui exige un poids supplémentaire noté b , nous incluons une composante constante $\phi_0 = x_0$ dans l'espace des caractéristiques pour pouvoir traiter le seuil w_0 comme un poids supplémentaire (que nous noterons par la suite b).

1.2 Régression par SVM

1.2.1 Formulation du problème dans le cas linéaire

La régression par SVM est un cas particulier de classification où les classes des exemples ne sont pas dénombrables c-à-d continues.

Soit un ensemble de données : $D = \{(x_i, y_i)\} \in \mathbb{R}^d * \mathbb{R}, i = 1:N$

L'objectif est de trouver une fonction f dans un espace de Hilbert H , par minimisation du risque structurel (SRM), établissant une relation entre les variables x et la grandeur à modéliser y ; $y = f(x)$ à partir de l'ensemble de mesures D . souvent f est considérée comme fonction linéaire :

$$\mathbf{f}(\mathbf{x}) = \langle \mathbf{w}, \mathbf{x} \rangle + \mathbf{b}, \quad \mathbf{w} \in \chi \text{ et } \mathbf{b} \in \mathbb{R} \quad (9)$$

Ou, $\langle ., . \rangle$ représente le produit scalaire dans χ , w est un vecteur et b un scalaire.

Le problème revient donc à trouver un l'hyperplan caractérisé par w^* et b^* qui minimisent l'écart global entre f et les y_i :

$$(\mathbf{w}^*, \mathbf{b}^*) = \operatorname{argmin}_{\mathbf{w}, \mathbf{b}} \sum_{i=1}^N (y_i - \langle \mathbf{w}, \mathbf{x}_i \rangle - \mathbf{b})^2 \quad (10)$$

Comme pour la classification, on propose de modéliser la fonction de régression par un hyperplan qui se situe au centre d'un hyper-tube de largeur 2ε contenant la majorité des exemples d'entraînement (Figure 11).

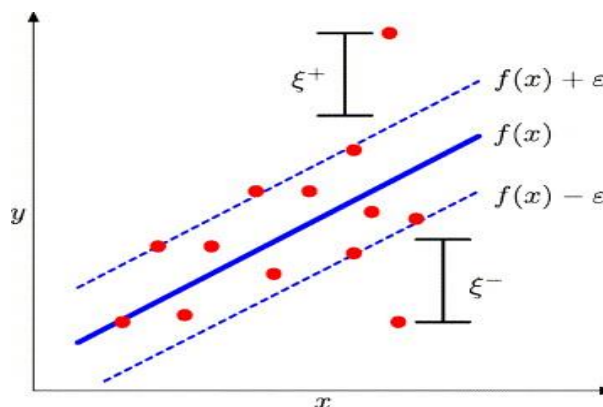
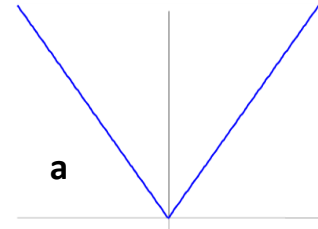


Figure 11 : Hyperplan et hyper-tube

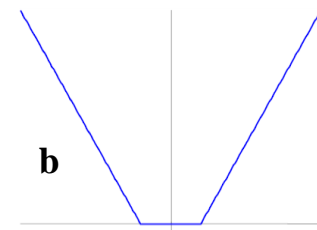
Cet hyper-tube est caractérisé par une fonction qu'on appellera la fonction de perte qui peut prendre plusieurs formes et sont présenté sur les figures suivantes (Figure 12).

❖ **La fonction de perte ε -laplace**



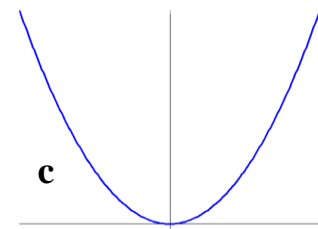
❖ **La fonction de perte ε -insensible :**

$$L_{\varepsilon}(y) = \begin{cases} 0 & \text{for } |f(x) - y| < \varepsilon \\ |f(x) - y| - \varepsilon & \text{otherwise} \end{cases}$$



❖ **La fonction de perte quadratique :**

$$L_{\text{quad}}(f(x) - y) = (f(x) - y)^2$$



❖ **La fonction de perte de Huber:**

$$L_{\text{huber}}(f(x) - y) = \begin{cases} \frac{1}{2}(f(x) - y)^2 & \text{for } |f(x) - y| < \mu \\ \mu|f(x) - y| - \frac{\mu^2}{2} & \text{otherwise} \end{cases}$$

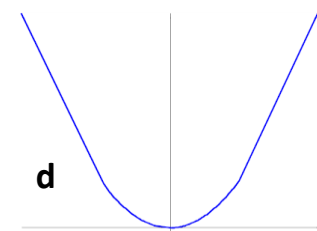


Figure 12 : Fonctions de perte (a, b, c, d)

Chapitre 2 : Méthodologie

Formellement, nous pouvons écrire ce problème comme un problème d'optimisation convexe présenté ci-dessous dans sa forme primale :

$$\begin{aligned} & \text{Minimiser} && L(w) = \frac{1}{2} \|w\|^2 \\ & \text{Sous les contraintes} && \begin{cases} \mathbf{y}_i - \langle \mathbf{w}, \mathbf{x}_i \rangle - \mathbf{b} \leq \varepsilon \\ \mathbf{y}_i - \langle \mathbf{w}, \mathbf{x}_i \rangle - \mathbf{b} \geq -\varepsilon \end{cases} \end{aligned} \quad (11)$$

Ceci veut dire qu'il existe une fonction f qui approxime tous les couples de données (x_i, y_i) avec une précision ε , en d'autres termes, que le problème d'optimisation convexe possède une solution.

Mais, bien souvent, ceci n'est pas le cas et nous devons donc nous résoudre à autoriser quelques erreurs en introduisant de nouvelles variables appelées variables ressorts (ou variables de relâchement) qu'on notera ξ_i et ξ_i^* , pour gérer les contraintes irréalisables du problème d'optimisation convexe, ce qui est appelé *la marge douce* (soft margin) (Figure 13).

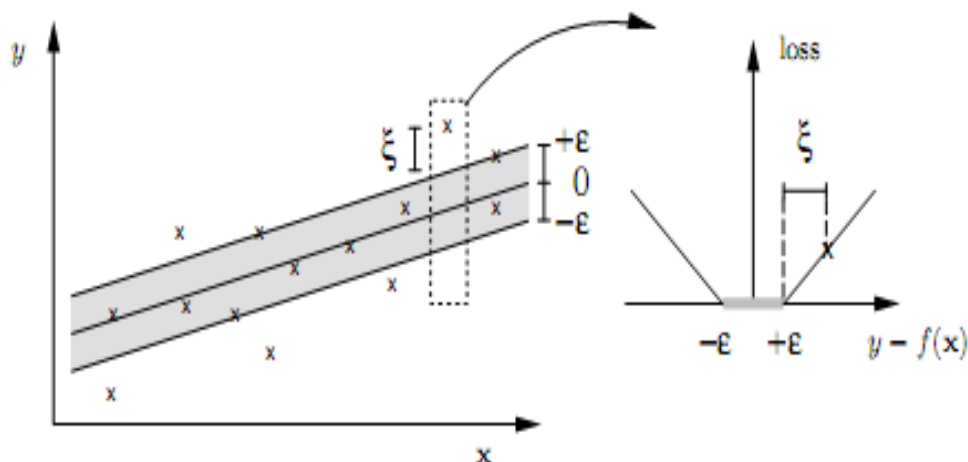


Figure 13 : Représentation de la marge douce et illustration de la fonction de perte

En introduisant le principe de la marge douce, la nouvelle formulation du problème d'optimisation convexe sous contraintes donne la formulation primale suivante :

La formulation primale :

$$\min_{w \in \mathcal{X}, \xi \in \mathbb{R}^n, b \in \mathbb{R}} L(w, \xi, \xi^*) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \quad (12)$$

Sous les contraintes
$$\begin{cases} y_i - \langle w, x_i \rangle - b \leq \varepsilon + \xi_i \\ y_i - \langle w, x_i \rangle - b \geq -\varepsilon - \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases} \quad \forall i \in \{1, \dots, n\} \quad (13)$$

C représente une constante déterminante du compromis entre les deux objectifs opposés : la minimisation de l'erreur et la maximalisation de la marge.

1.2.2 Formulation du problème dans le cas non linéaire

Dans le cas des données non-linéaires, les entrées sont projetées dans un espace de plus grande dimension appelé espace caractéristique (feature space) en utilisant une fonction de projection K prédéterminée (fonction noyau) (Cortes et Vapnik, 1995; Scholkopf et al, 1999; Scholkopf et al, 2000) de telle manière que le problème devienne un problème de régression linéaire dans ce nouvel espace (Figure 14).

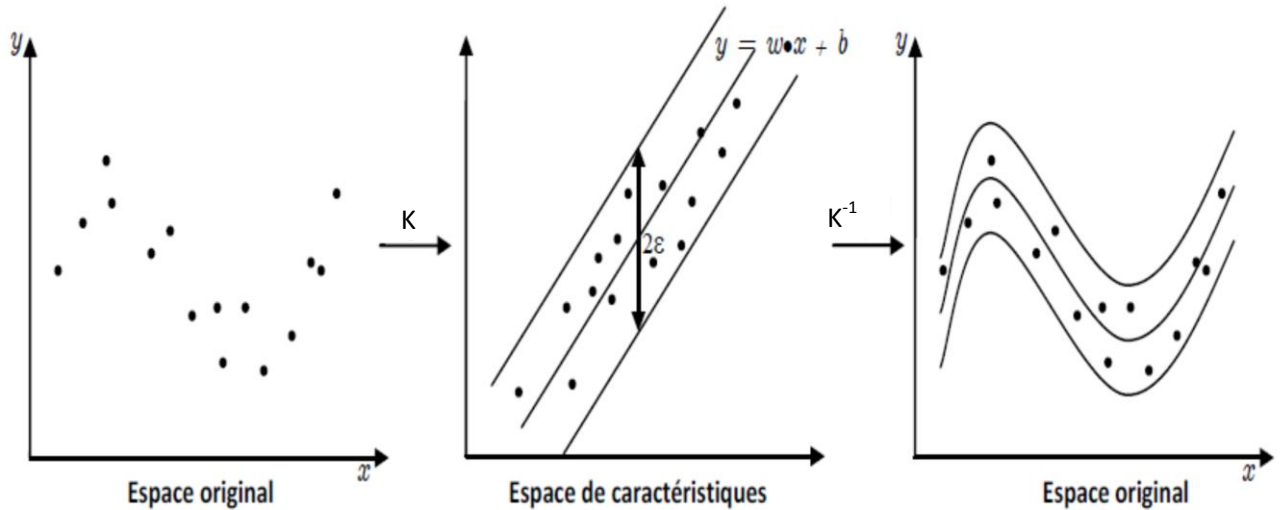


Figure 14 : Régression dans le cas non linéaire par l'utilisation de noyaux

1.2.3 Quelques définitions

- **Espaces de Hilbert à noyaux reproduisant**

C'est l'espace dans lesquels s'effectue de manière implicite la régression non linéaire par les méthodes de noyaux

- **Définition1**

- On définit un espace préhilbertien H comme un espace vectoriel muni d'un produit scalaire $\langle \cdot, \cdot \rangle_H$.
- Un espace de Hilbert est un espace préhilbertien complet pour la norme $\|\cdot\|_H$ induite par le produit scalaire via $\|h\|_H^2 = \langle h, h \rangle_H$ pour tout h dans H .

- **Définition2**

On appelle « Espace de Hilbert à Noyau Reproductant » ou RKHS (Reproducing Kernel Hilbert Space) un espace de Hilbert H qui possède les propriétés suivantes :

- étant donné un espace X , H contient toutes les fonctions telles que

$$\forall \mathbf{x} \in X \quad \mathbf{K}_x: \mathbf{t} \rightarrow \mathbf{K}(\mathbf{x}, \mathbf{t}) \quad (14)$$

- il existe une fonction K_x telle que, pour tout \mathbf{x} de X et pour tout $f(\cdot)$ de H , on ait :

$$\mathbf{f}(\mathbf{x}) = \langle \mathbf{f}, \mathbf{k}_x \rangle_H \quad (15)$$

Où $\langle f, g \rangle_H$ désigne le produit scalaire défini dans H . k_x est le noyau reproductant de H ; s'il existe (donc si H est un RKHS), il est unique.

1.2.4 Propriétés des noyaux

- Un noyau reproductant vérifie la *propriété de reproduction* :

$$\langle \mathbf{k}_x, \mathbf{k}_y \rangle_H = \mathbf{K}(x, y) \quad (16)$$

- Un noyau reproductant existe si et seulement si $f(x)$ est continue.
- Tout noyau reproductant est un noyau défini positif : étant donnés N vecteurs $\{x_i, i = 1 \dots N\}$, la matrice d'élément général $k_{ij} = \mathbf{K}(x_i, x_j)$ est semi-définie positive.
- Toute combinaison linéaire de fonctions noyaux est une fonction noyau.
- Tout produit de fonctions noyaux est une fonction noyau.

1.2.5 Formulation duale et optimisation du problème de régression

Afin de résoudre le problème de minimisation de l'équation (12) sous les contraintes (13), La méthode des multiplicateurs de Lagrange est une solution vu qu'elle regroupe la fonction à optimiser et la contrainte en une seule fonction. De ce fait, nous obtenons une nouvelle formulation au problème appelé duale. Nous nous passerons des différentes étapes de démonstration (pour les détails se référer aux travaux de Drucker [92]) et

Chapitre 2 : Méthodologie

nous résumons les résultats de la formulation duale du problème non linéaire sous ses contraintes dans l'équation suivante (Eq 17):

Avec α_i et α_i^* les multiplicateurs de Lagrange.

Le principe donc revient à optimiser le Lagrangien $L(\alpha_i, \alpha_i^*)$ suivant ;

$$\begin{aligned} \text{minimiser} \quad L(\alpha_i, \alpha_i^*) = & - \sum_{i=1}^n y_i (\alpha_i - \alpha_i^*) + \varepsilon \sum_{i=1}^n y_i (\alpha_i + \alpha_i^*) \\ & + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) K(x_i, x_j) \end{aligned} \quad (17)$$

Sous les contraintes ;

$$\begin{aligned} 0 \leq \alpha_i, \alpha_i^* \leq C, \quad i = 1, \dots, n \\ \sum_{i=1}^n (\alpha_i - \alpha_i^*) = 0 \end{aligned} \quad (18)$$

Calcul de w :

$$w = \sum_{i=1}^n x_i (\alpha_i - \alpha_i^*) \quad (19)$$

Calcul de b^* ;

$$b^* = y - \sum_{i=1}^n (\alpha_i - \alpha_i^*) K(x, x_i) - \varepsilon \quad \text{Pour} \quad 0 \leq \alpha_i \leq C \quad (20)$$

Enfin, la fonction de régression s'écrit ;

$$f(x, \alpha_i, \alpha_i^*) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) K(x, x_i) + b^* \quad (21)$$

2. Les réseaux de neurones artificiels RNA

2.1 Historique

L'appellation d'Intelligence Artificielle (IA), a pour objectif de simuler les comportements du cerveau humain. Les premières tentatives de modélisation du cerveau sont anciennes et précèdent même l'ère informatique. C'est en 1943 que Mc Culloch (neurophysiologiste) et Pitts (logicien) qui proposèrent les premières notions du neurone formel. Ce concept fut ensuite mis en réseau avec une couche d'entrée et une couche de sortie, par Rosenblatt en 1957 au laboratoire d'aéronautique de l'université de Cornell, pour inventer le perceptron ! [93]

2.2 Le perceptron

L'entrée du perceptron est une couche qui reçoit des informations. La couche de sortie fait une somme linéaire de la couche d'entrée pondérée par des poids. On rajoute une fonction de sortie f qui suivra la sortie.

La Figure 15, représente un neurone de la couche de sortie, où x_i est la valeur de sortie de la i^e cellule d'entrée, w_{ij} est l'intensité de la connexion entre la i^e cellule d'entrée et la j^e cellule de sortie, a_j est l'activation de la j^e cellule de sortie, f est la fonction de transfert et o_j est la règle de décision. Le perceptron est non linéaire à cause de la fonction de transfert qui simule le seuil d'activation du neurone.

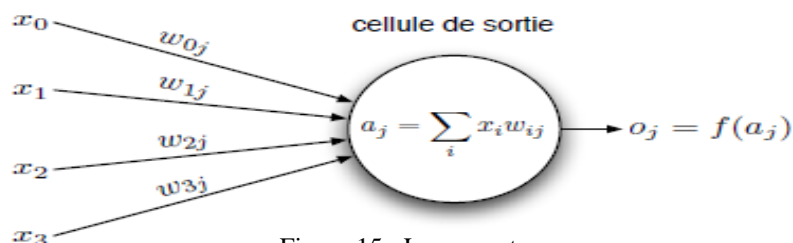


Figure 15 : Le perceptron

2.3 Réseaux de neurones

Un réseau neuronal est l'association, en un graphe plus ou moins complexe, d'objets élémentaires (les neurones formels).

Les principaux réseaux se distinguent par :

- l'organisation du graphe ou son architecture
- son niveau de complexité (le nombre de neurones, présence ou non de boucles de rétroaction dans le réseau)
- le type des neurones (leurs fonctions de transition ou d'activation)
- l'objectif visé : apprentissage supervisé ou non, optimisation, systèmes dynamiques.

Les principaux types de fonction d'activation :

- linéaire, f est la fonction identité,
- sigmoïde $f(x) = 1/(1 + e^x)$
- seuil $f(x) = 1$

- radiale $f(x) = \sqrt{\frac{1}{2\pi}} e^{-x^2/2}$

2.4 Perceptron multicouche

Le perceptron multi couche PMC est très utilisé dans les problèmes de prédiction des phénomènes météorologiques et plus particulièrement le solaire ([49], [76], [68], [3], [73],...) de par sa simplicité et le résultat de ses performances. Cependant, ce type de réseau sera adopté pour l'élaboration de notre travail.

Le perceptron multicouche (PMC) est un réseau composé de couches de neurones successives totalement connectées. La fonction sigmoïde est principalement utilisée comme fonction d'activation. Les neurones d'une même couche ne sont pas connectés entre eux mais sont connectés à la couche ou aux couches suivantes (réseaux à propagation vers l'avant feed-forward). Le principe de l'apprentissage repose sur la minimisation d'une fonction coût par un algorithme adaptatif de type gradient. Souvent la fonction de coût choisit est l'erreur quadratique moyenne.

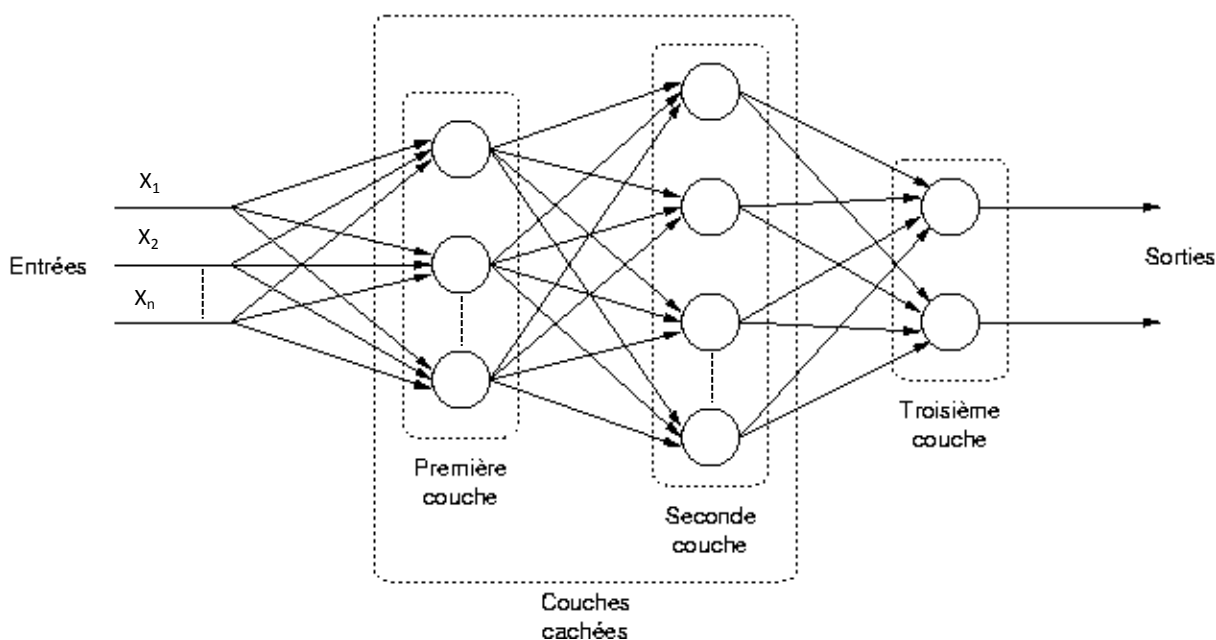


Figure 16 : Schéma de principe du réseau PMC

2.5 Equation du réseau PMC

Soit le réseau PMC à L couches, p entrées et q sorties. La couche d'entrée contient les composantes du vecteur d'entrée $(x_i)_{0 \leq i \leq p}$. Les calculs se font couche par couche de l'entrée vers la sortie. La sortie du j^e neurone de la couche l ($0 \leq l \leq L$) est donnée par :

$$S_j^l = \sum_{i=0}^{N^{l-1}} w_{ji}^l u_i^{l-1} \quad (22)$$

$$u_j^l = f(S_j^l)$$

Ou :

S_j : est la somme pondérée à l'entrée du j^e neurone dans la couche(l).

N^l : est le nombre de neurones de la couche(l).

u_j^l : est la sortie du j^e neurone dans la couche(l).

w_{ji}^l : est le poids qui relie le i^e neurone de la couche ($l - 1$) au j^e neurone dans la couche (l).

L : Nombre de couches dans le réseau. La couche 0 correspond aux nœuds d'entrée.

$f(.)$: est la fonction d'activation, elle peut être linéaire ou non pour les neurones de la couche de sortie, mais toujours non linéaire pour les neurones appartenant à une couche cachée.

On note que le premier élément de chaque vecteur $u^l (l = 0, 1, \dots, L - 1)$ est mis à un.

Le produit $(w_{j0}^l * u_0^{l-1})$ représente la valeur du seuil interne du j^e neurone dans la couche l .

2.6 Apprentissage du réseau

L'algorithme de Levenberg-Marquardt est très utilisé pour l'entraînement des réseaux de neurones de petite et moyenne taille. Il est le plus rapide et assure la meilleure convergence vers un minimum de l'erreur quadratique quand le nombre de poids du réseau est inférieur à quelques centaines [94]. Cet algorithme est dit du second ordre car il prend en considération la dérivée seconde de la fonction coût. Les étapes de calcul et d'implémentation de l'algorithme peuvent être trouvées dans plusieurs travaux ([73], [30], [40], ...). Pour plus de détails, on peut se référer aux ouvrages originaux ([95], [96], [97]).

3. Modèles de prévision des séries temporelles

3.1 Introduction

Dans cette section, nous aborderons plus particulièrement les processus SARIMA qui n'est rien d'autre qu'une généralisation des processus ARIMA contenant une partie saisonnière et sont particulièrement utiles pour décrire le comportement des séries temporelles univariées. Pour de plus amples détails sur la modélisation et la prédiction des séries temporelles sous toutes ses formes, se référer à l'ouvrage de [98] et aux cours de [60].

3.2 Processus SARIMA (p,d,q)(P,D,Q)

Soit l'équation caractéristique du modèle SARIMA[60] :

$$(1 - L^s)\varphi(L)\varnothing(L^s)(1 - L)^d X_t = \theta(L)\vartheta(L^s)\varepsilon_t + \varphi_0 \tag{23}$$

Avec :

- $\varphi(L) = 1 - \varphi_1 L - \varphi_2 L^2 - \dots - \varphi_p L^p$
- $\theta(L) = 1 - \theta_1 L - \theta_2 L^2 - \dots - \theta_q L^q$
- $\varnothing(L^s) = 1 - \varnothing_1 L^s - \varnothing_2 L^{2s} - \dots$
- $\vartheta(L^s) = 1 - \vartheta_1 L^s - \vartheta_2 L^{2s} - \dots$

Les polynômes $\varphi(L)$ et $\theta(L)$ caractérisent respectivement les parties AR (autorégressives) et MA (moving average) du modèle. Et les polynômes $\varnothing(L^s)$ $\vartheta(L^s)$ caractérisent respectivement les parties AR et MA saisonnières.

- p est le degré du polynôme AR
- q est le degré du polynôme MA
- d est le degré de différenciation, en général le maximum est de 3. Si $d = 1$ alors, $(1 - L)^1 X = X_t - X_{t-1}$
- $(1 - L)^d$: sert à différencier la série chronologique afin de la rendre stationnaire.
- L est l'opérateur retard.

$$\nabla = (1 - L)$$

$$\nabla_{365} = (1 - L^{365})$$

- ε_t est un bruit blanc : $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$

Remarquons que ces processus ont à peu près la même définition que les processus ARMA, la seule différence est qu'on permet au polynôme autorégressif d'avoir la racine unité.

Trois étapes sont nécessaires à la construction de ces modèles:

- une phase d'identification,
- une phase d'estimation,
- une phase de validation.

À l'issue de ces trois phases, une fois déterminé le meilleur modèle ARMA, il est utilisé pour la prédiction.

3.3 Identification

Il n'y a pas de règles rigoureuses qui nous permettent de connaître les degrés p et q des polynômes du modèle ARMA (p,q) de la série stationnaire. D'après Box et Jenkins [98], Les moments d'ordre deux définis par la fonction d'autocorrélation simple (FAC) et la fonction d'autocorrélation partielle (FACP) de la série temporelle étudiée, peuvent nous renseigner sur les degrés maximums des polynômes.

Cependant, après différentes combinaisons ; plusieurs modèles candidats seront retenus et éliminés successivement durant les étapes deux et trois du cycle (estimation, validation) jusqu'à obtention du bon (Figure 17).

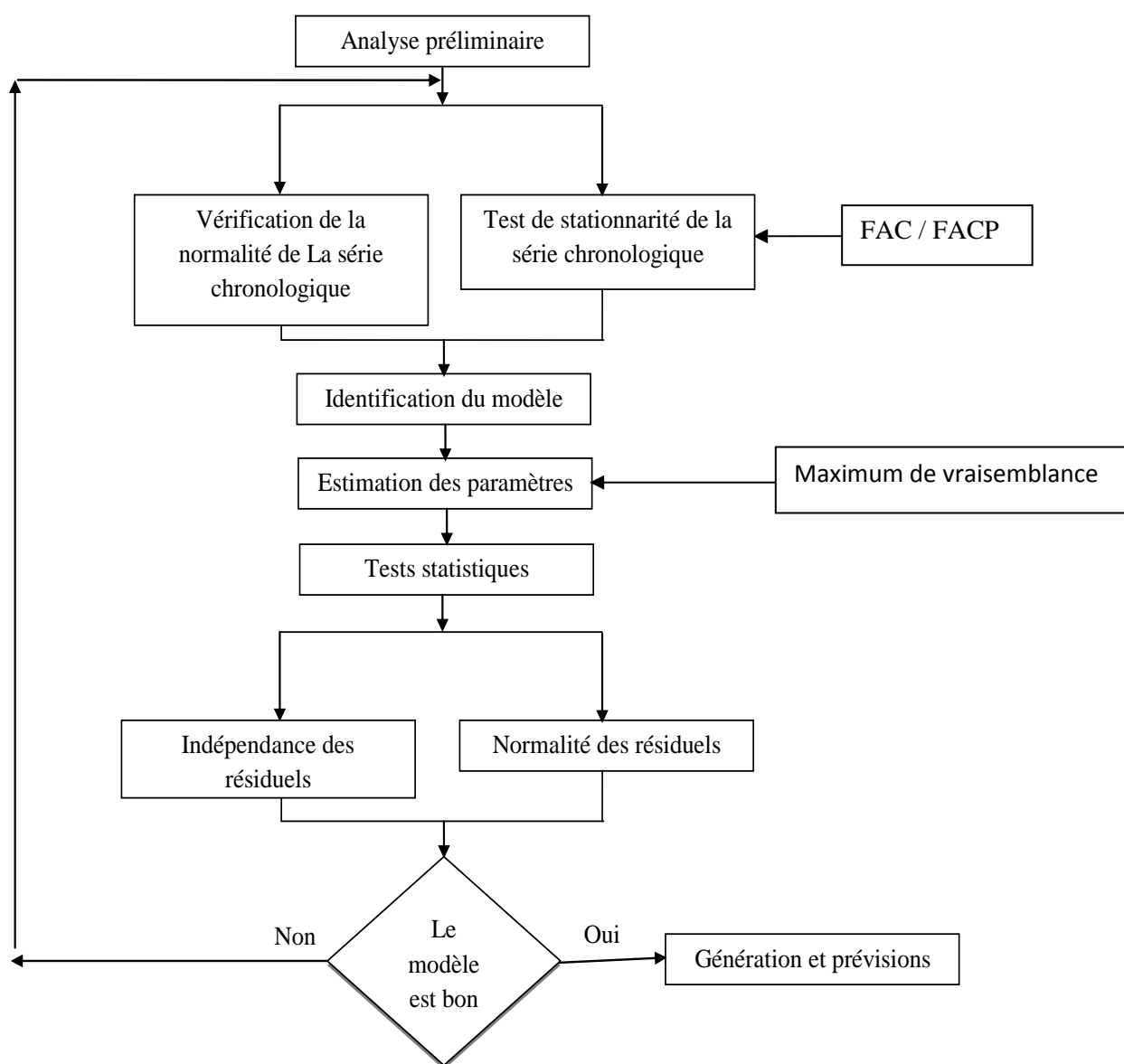


Figure 17 : Processus de modélisation et de prévision par la méthodologie Box & Jenkins

Chapitre 2 : Méthodologie

Plusieurs critères sont nécessaires pour identifier les modèles. Parmi ces critères, on cite quelques-uns des plus simples :

- Pour un modèle AR(p) : la FAC de la série stationnaire ou devenu stationnaire décroît exponentiellement ou sinusoidalement et la FACP s'annule après un retard $k=p$.
- Pour un modèle MA(q) : la FACP de la série stationnaire ou devenu stationnaire décroît exponentiellement ou sinusoidalement et la FAC s'annule après un retard $k=q$.

Dans le tableau 5, on illustre quelques cas d'identification tels les modèles AR(1), AR(2), MA(1), MA(2) et ARMA(1,1), ainsi que les estimations préliminaires de leurs coefficients [99].

Rappelons que :

ρ_k est le coefficient d'autocorrélation de la série stationnaire

ϕ_{kk} est le coefficient d'autocorrélation partiel de la série stationnaire

Tableau 5: Exemples d'identification et d'estimation

ordre	(1,d,0)	(0,d,1)
✓ Comportement de ρ_k	Décroît exponentiellement	$\rho_1 \neq 0$
✓ Comportement de ϕ_{kk}	$\phi_{11} \neq 0$	Décroît exponentiellement
✓ Estimations préliminaire des paramètres du modèle	$\varphi_1 = \rho_1$	$\rho_1 = \frac{-\theta_1}{1+\theta_1^2}$
✓ Région d'admissibilité	$-1 < \varphi_1 < 1$	$-1 < \theta_1 < 1$
ordre	(2,d,0)	(0,d,2)
✓ Comportement de ρ_k	Une sinusoïde ou exponentiel atténuée	toujours $\neq 0$
✓ Comportement de ϕ_{kk}	φ_1 et $\varphi_2 \neq 0$	Une sinusoïde ou exponentiel atténuée
✓ Estimations préliminaire des paramètres du modèle	$\begin{cases} \varphi_1 = \frac{\rho_1(1-\rho_2)}{1-\rho_1^2} \\ \varphi_2 = \frac{\rho_2-\rho_1^2}{1-\rho_1^2} \end{cases}$	$\begin{cases} \rho_1 = \frac{-\theta_1(1-\theta_2)}{1+\theta_1^2+\theta_2^2} \\ \rho_2 = \frac{-\theta_2}{1+\theta_1^2+\theta_2^2} \end{cases}$
✓ Région d'admissibilité	$\begin{cases} -1 < \varphi_2 < 1 \\ \varphi_2 + \varphi_1 < 1 \\ \varphi_2 - \varphi_1 < 1 \end{cases}$	$\begin{cases} -1 < \theta_2 < 1 \\ \theta_2 + \theta_1 < 1 \\ \theta_2 - \theta_1 < 1 \end{cases}$

Chapitre 2 : Méthodologie

ordre

(1,d,1)

✓ Comportement de ρ_k	décroissance exponentielle après le premier retard	
✓ Comportement de ϕ_{kk}	dominé par une décroissance exponentielle après le premier retard	
✓ Estimations préliminaire des paramètres du modèle	$\rho_1 = \frac{(1-\theta_1\phi_1)(\phi_1-\theta_1)}{1+\theta_1^2-2\phi_1\theta_1}$	$\rho_2 = \rho_1\phi_1$
✓ Région d'admissibilité	$-1 < \phi_1 < 1$	$-1 < \theta_1 < 1$

Tableau 6 : Propriétés des processus AR(p), MA(q) et ARMA (p,q)

	AR(p)	MA(q)	ARMA(p,q)
✓ Modèle en fonction de précédents X_t	$\varphi(B)X_t = \varepsilon_t$	$\theta^{-1}(B)X_t = \varepsilon_t$	$\theta^{-1}(B)\varphi(B)X_t = \varepsilon_t$
✓ Modèle en fonction de précédents ε_t	$X_t = \varphi^{-1}(B)\varepsilon_t$	$X_t = \theta(B)\varepsilon_t$	$X_t = \varphi^{-1}(B)\theta(B)\varepsilon_t$
✓ Conditions de stationnarité	Les racines de $\varphi(B) = 0$ se trouvent en dehors du cercle unité	Toujours stationnaire	Les racines de $\varphi(B) = 0$ se trouvent en dehors du cercle unité
✓ Conditions d'inversibilité	Toujours inversible	Les racines de $\theta(B) = 0$ se trouvent en dehors du cercle unité	Les racines de $\theta(B) = 0$ se trouvent en dehors du cercle unité
✓ Fonction d'autocorrélation	Infinie (exponentielle et/ou onde sinusoidale atténuée) diminue	finie S'annule	Infinie (exponentielle et/ou onde sinusoidale atténuée après q-p retards) diminue
✓ Fonction d'autocorrélation partielle	finie S'annule	Infinie (dominée par une exponentielle et/ou onde sinusoidale atténuée) diminue	Infinie (dominée par une exponentielle amortie et/ou onde sinusoidale atténuée après p-q retards) diminue

Pour un processus ARMA (p,q) stationnaire d'ordres élevés, la FAC se comporte comme celle d'un processus AR(p) après (q - p) retards, et la FAP se comporte comme celle d'un processus MA(q) après (p - q) retards.

Le Tableau 6 résume les propriétés des différents processus AR(p), MA(q) et ARMA(p,q). [99]

3.4 Estimation

On suppose que les ordres des paramètres p et q sont connus avant d'effectuer leurs estimation.

3.4.1 Système d'équations de Yule-Walker

Cette technique est utilisé dans le cas d'un processus AR(p), on retrouve donc les estimateurs des paramètres $\varphi_1, \varphi_2, \dots, \varphi_p$ et σ_ε^2 du modèle par la résolution des équations de Yule-Walker qui sont déduites directement de la fonction d'autocorrélation.[98]

$$\rho_k = \begin{cases} \sum_{j=1}^p \varphi_j \rho_{k-j} & \text{si } k > 0 \\ \sum_{j=1}^p \varphi_j \rho_{-j} + \sigma_\varepsilon^2 & \text{si } k = 0 \end{cases} \quad (24)$$

L'écriture matricielle est la suivante :

$$\begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \vdots \\ \rho_p \end{bmatrix} = \begin{bmatrix} \rho_0 & \rho_{-1} & \dots & \rho_{1-p} \\ \rho_1 & \rho_0 & \dots & \rho_{2-p} \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{p-1} & \rho_{p-2} & \dots & \rho_0 \end{bmatrix} \begin{bmatrix} \varphi_1 \\ \varphi_2 \\ \vdots \\ \vdots \\ \varphi_p \end{bmatrix} \quad (25)$$

Tel que, les ρ_i ($i = 0; p \in \mathbb{N}$) sont les estimateurs des coefficients de corrélation de la série stationnaire X_t .

$$\rho_k = \frac{\text{cov}(X_t, X_{t-k})}{\sigma_{X_t} \sigma_{X_{t-k}}} = \frac{\sum_{t=k+1}^n (X_t - \bar{X})(X_{t-k} - \bar{X})}{\sqrt{\sum_{t=k+1}^n (X_t - \bar{X})^2} \sqrt{\sum_{t=k+1}^n (X_{t-k} - \bar{X})^2}} \quad (26)$$

Le signal étant réel et $\rho_k = \rho_{-k}$; on retrouve alors une matrice d'autocorrélation qui est symétrique et appelé aussi de Toeplitz.

Il suffit donc de résoudre les p équations à p inconnues pour pouvoir trouver les coefficients $\varphi_1, \varphi_2, \dots, \varphi_p$ et pour la variance σ_ε^2 on pose $k = 0$.

3.4.2 Méthode du maximum de vraisemblance

C'est la méthode la plus couramment utilisée pour estimer les différents paramètres $(\varphi_1, \dots, \varphi_p, \theta_1, \dots, \theta_q, \sigma_\varepsilon^2)$ du modèle ARMA (p,q) dont la représentation est la suivante :

$$\varphi(L)X_t = \theta(L)\varepsilon_t + \varphi_0 \quad (27)$$

D'où :

$$\varepsilon_t = \sum_{i=1}^p \varphi_i X_{t-i} - \sum_{i=1}^q \theta_i \varepsilon_{t-i} - \varphi_0 \quad (28)$$

Le principe de la vraisemblance revient à déterminer la valeur des paramètres $(\varphi_0, \varphi_1, \dots, \varphi_p, \theta_1, \dots, \theta_q, \sigma_\varepsilon^2)$ fonction des observations (X_1, X_2, \dots, X_T) ; qui assure la plus grande probabilité d'apparition de ces observations (X_1, X_2, \dots, X_T) .

Pour commencer ; on fait l'hypothèse de la normalité des résidus ε_t avant de spécifier une forme fonctionnelle à la vraisemblance du modèle tel que ; $\varepsilon_t \sim (0, \sigma_\varepsilon^2)$ dont la fonction de distribution est :

$$f(\varepsilon) = \frac{1}{\sigma_\varepsilon \sqrt{2\pi}} e^{-\frac{\varepsilon_t^2}{2\sigma_\varepsilon^2}} \quad (29)$$

Et la fonction vraisemblance associée au vecteur de réalisation (X_1, X_2, \dots, X_T) est :

$$\begin{aligned} f(X_1, X_2, \dots, X_T; \varphi_1, \dots, \varphi_p, \theta_1, \dots, \theta_q, \sigma_\varepsilon^2) &= \prod_{t=1}^T \frac{1}{\sigma_\varepsilon \sqrt{2\pi}} e^{-\frac{(X_t - E(X_t))^2}{2\sigma_\varepsilon^2}} \\ &= \left(\frac{1}{\sigma_\varepsilon \sqrt{2\pi}} \right)^T \exp \left[\sum_{t=1}^T -\frac{(X_t - E(X_t))^2}{2\sigma_\varepsilon^2} \right] \end{aligned} \quad (30)$$

Afin de maximiser la fonction de vraisemblance, on doit d'abord considérer son logarithme pour faciliter les calculs :

$$\begin{aligned} \text{Ln} \left(f(X_1, X_2, \dots, X_T; \varphi_1, \dots, \varphi_p, \theta_1, \dots, \theta_q, \sigma_\varepsilon^2) \right) &= T \text{Ln} \left(\frac{1}{\sigma_\varepsilon \sqrt{2\pi}} \right) - \left[\sum_{t=1}^T \frac{(\varepsilon_t)^2}{2\sigma_\varepsilon^2} \right] \\ &= -\frac{T}{2} \text{Ln}(\sigma_\varepsilon^2) - \frac{T}{2} \text{Ln}(2\pi) - \left[\sum_{t=1}^T \frac{(\varepsilon_t)^2}{2\sigma_\varepsilon^2} \right] \end{aligned} \quad (31)$$

Enfin, pour maximiser la fonction vraisemblance, cela revient à minimiser le terme d'erreur : $(\varepsilon_t)^2$

3.5 Validation du modèle

Après avoir estimé les paramètres du modèle, on doit s'assurer de sa validité avant de procéder à la prévision. Par conséquent, plusieurs critères sont à vérifier à travers des tests.

3.5.1 La parcimonie :

D'après Box & Jenkins, l'une des règles de sélection d'un modèle est la parcimonie qui consiste à déterminer et estimer le minimum de paramètres possible. Dans ce processus de sélection d'un modèle ARMA (p,q) ; une augmentation de p ou de q fait diminuer la somme des résidus au carré mais fait aussi diminuer le nombre de degré de liberté ce qui va dans le sens opposé de la parcimonie. Pour cela on doit faire un arbitrage entre ces deux facteurs à l'aide de deux critères de comparaison ; AIC (Akaike information criterion)(Akaike,1974) et BIC (Baisian information criterion)(Schwarz,1978) qui sont les premiers critères apparaissant dans la littérature et largement diffusés et appliqués. Le principe est d'avoir le minimum possible pour un meilleur modèle [12].

a) Critère d'information d'Akaike (AIC)

La formulation mathématique d'AIC est définie comme suit [99] :

$$AIC(p + q) = N \ln(\sigma_{\varepsilon}^2) + 2(p + q) \quad (32)$$

Où p et q sont le nombre de paramètres d'AR et de MA à estimer. Le modèle qui donne le minimum AIC est choisi comme modèle parcimonieux, de même pour le BIC.

b) Critère d'information Bayésien (BIC)

La formulation mathématique d'AIC est définie comme suit [99] :

$$BIC(p + q) = N \ln \sigma_{\varepsilon}^2 + (p + q) \ln N \quad (33)$$

N : le nombre d'observation

σ_{ε}^2 : Variance des résidus

3.5.2 Test sur les résidus

a) Test d'autocorrélations des résidus

La fonction d'autocorrélation des résidus (RACF) du modèle à choisir pour la prévision devrait obéir à celle d'un bruit blanc. Pour cela, deux méthodes seront utilisées pour vérifier cela:

Chapitre 2 : Méthodologie

• La première est de dessiner le corrélogramme en traçant $r_k(\varepsilon)$ en fonction du retard k : Si une partie du RACF est significativement différente de zéro, ceci peut signifier que le modèle actuel est insatisfaisant.

• La seconde est la statistique $Q(r)$ qui est la plus récente et suggérée par Ljung et Box (1978). Le principe est de vérifier l'hypothèse d'adéquation du modèle en choisissant un niveau de signification ensuite en comparant la valeur statistique de χ^2 calculé à χ^2 tabulée de la valeur critique.

Si la valeur calculée de χ^2 est plus petite que la valeur critique χ^2 tabulée, le modèle est adéquat sur la base des données disponibles.

La statistique de $Q(r)$ est construite comme suit: [100]

$$Q(r) = T' (T' + 2) \sum_{k=1}^m (T' - k)^{-1} r_k(\varepsilon)^2 \xrightarrow{T \rightarrow \infty} \chi^2 (m - p - q) \quad (34)$$

Tel que :

$$r_k(\varepsilon) = \frac{\sum_{i=k+1}^T \varepsilon_i \varepsilon_{i-k}}{\sum_{i=1}^T \varepsilon_i^2} \quad (35)$$

$Q(r)$: statistique de Ljung–Box au retard m

$r_k(\varepsilon)$: ACF de ε_k au retard k

m : un nombre présélectionné des autocorrélations

$N' = N - d$: le nombre d'observations restantes après d différenciations afin de stationnariser la série

b) Test de normalité des résidus

Pour vérifier si le processus des résidus $\{\varepsilon_t, t \in \mathbb{Z}\}$ est un bruit blanc gaussien, plusieurs tests peuvent être utilisés, mais le test le plus courant est celui de Jaque et Bera. Ce dernier est fondé sur la notion de skewness S_k (moment d'ordre 3 et asymétrie) et de Kurtosis K_u (moment d'ordre 4 et queue de distribution).

La statistique du test de Jaque et Bera est comme suit :

$$S = \frac{T}{6} S_k + \frac{T}{24} (K_u - 3)^2 \xrightarrow{T \rightarrow \infty} \chi^2(2) \quad (36)$$

Si : $S \geq \chi_{1-\alpha}^2(2)$, On rejette l'hypothèse de normalité des résidus au seuil de α %

3.6 Prédiction de la série temporelle via le modèle SARIMA

Soit le processus SARIMA (p,d,q)(P,D,Q) suivant :

$$(1 - L)^d(1 - L^s)X_t\varphi(L)\vartheta(L^s) = \theta(L)\vartheta(L^s)\varepsilon_t \quad (37)$$

Ou :

$$\nabla^d \nabla_s X_t \varphi(L) \vartheta(L^s) = \theta(L) \vartheta(L^s) \varepsilon_t \quad (38)$$

Tel que :

$$\begin{cases} \nabla^d = (1 - L)^d \\ \nabla_s = (1 - L^s) \end{cases} \quad (39)$$

Après l'identification et l'estimation des paramètres du processus stationnaire $W_t = \nabla^d \nabla_s X_t$, on revient à notre série originale X_t pour faire la prédiction à l'horizon h.

- **Etape de la prédiction**

Afin de mieux comprendre les étapes de la prédiction, on s'appuiera sur un exemple illustré par Box et Jenkins [98] et il s'agit d'un signal périodique (nombre des passager dans l'avion) dont la périodicité est de 12 mois ainsi, la représentation par le modèle SARIMA du processus est la suivante :

NB ; nous avons pris le soin d'entreprendre le développement de l'algorithme de calculs en utilisant Excel et Matlab.

Soit :

$$\nabla \nabla_{12} X_t = (1 - \theta L)(1 - \vartheta L^{12}) \varepsilon_t \quad (40)$$

Avec $\begin{cases} \hat{\theta} = 0.4 \\ \hat{\vartheta} = 0.6 \end{cases}$

Et $\varepsilon_t \sim (0, \sigma_\varepsilon^2)$, $\sigma_\varepsilon^2 = 1.34 * 10^{-3}$

Après développement de l'équation (121), on aura :

$$X_t = X_{t-1} + X_{t-12} - X_{t-13} + \varepsilon_t - 0.4\varepsilon_{t-1} - 0.6\varepsilon_{t-12} + 0.24\varepsilon_{t-13} \quad (41)$$

La prédiction à l'horizon « h » selon la méthodologie de Box & Jenkins s'écrit comme suit :

$$\begin{aligned}\widehat{X}_t^*(h) = & X_{t+h-1} + X_{t+h-12} - X_{t+h-13} + \varepsilon_{t+h} - 0.4\varepsilon_{t+h-1} - 0.6\varepsilon_{t+h-12} \\ & + 0.24\varepsilon_{t+h-13} \quad (42)\end{aligned}$$

Tel que :

$$\begin{aligned}X_{t+j} = & \begin{cases} X_{t+j} & j \leq 0 \\ \widehat{X}_t(j) & j > 0 \end{cases} \\ \varepsilon_{t+j} = & \begin{cases} \varepsilon_{t+j} & j \leq 0 \\ 0 & j > 0 \end{cases}\end{aligned}$$

Ceci revient à dire que ; nous remplacerons les valeurs inconnues de Y_{t+j} par leurs prévisions $\widehat{Y}_t(j)$ et les résidus inconnu par des zéros. Bien sur, après la première étape de prévision, on doit calculer les nouveaux résidus (mettre a jour) à chaque fois, comme suit :

$$\varepsilon_t = X_t - \widehat{X}_{t-1}(1) \quad (43)$$

Par exemple, pour prévoir la valeur de $\widehat{X}_t(3) = X_{t+3}$, c.à.d. la troisième valeur après l'origine t, on procédera de la manière suivante :

$$X_{t+3} = X_{t+2} + X_{t-9} - X_{t-10} + \varepsilon_{t+3} - 0.4\varepsilon_{t+2} - 0.6\varepsilon_{t-9} + 0.24\varepsilon_{t-10} \quad (44)$$

Ainsi la prévision de X_{t+3}^* est :

$$\begin{aligned}X_{t+3}^* = \widehat{X}_t^*(3) = & \widehat{X}_t^*(2) + X_{t-9} - X_{t-10} - 0.6\{X_{t-9} - \widehat{X}_{t-10}^*(1)\} \\ & + 0.24\{X_{t-10} - \widehat{X}_{t-11}^*(1)\} \quad (45)\end{aligned}$$

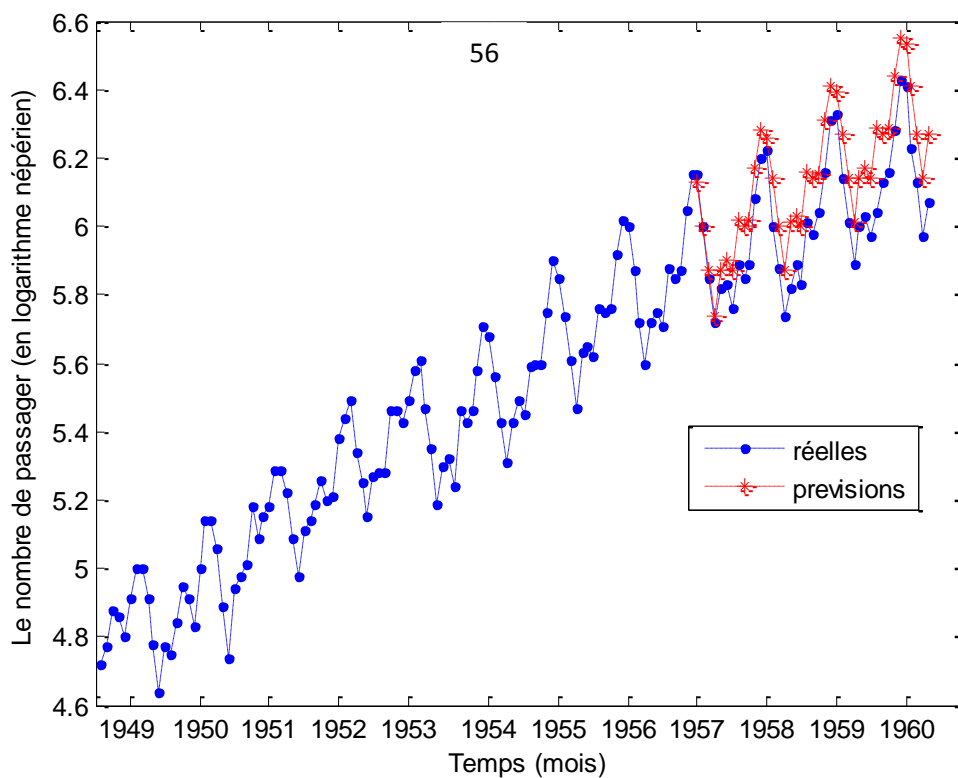
$$\widehat{X}_t^*(3) = \widehat{X}_t^*(2) + 0.4X_{t-9} - 0.76X_{t-10} + 0.6\widehat{X}_{t-10}^*(1) - 0.24\widehat{X}_{t-11}^*(1) \quad (46)$$

Pour le calcul de $\widehat{X}_{t-10}^*(1)$ et $\widehat{X}_{t-11}^*(1)$, on se réfère aussi à l'équation (42).

NB : comme l'ordre de l'équation est égal à 13, alors il nous faut 13 x 2 valeurs de la série temporelle pour faire la prévision.

Chapitre 2 : Méthodologie

La Figure 18, montre les résultats de la prévision de la série citée en exemple dont nous avons entrepris les calculs.



NB : En règle générale, la prévision d'une série temporelle, nécessite un nombre d'observation supérieur au moins à deux fois l'ordre de l'équation caractéristique du modèle SARIMA.

*Chapitre 3 : Prédiction de
L'irradiation solaire globale par les
techniques d'apprentissage
statistiques*

1. Introduction

Dans ce chapitre, nous développerons des modèles de prédiction de l'irradiation solaire globale journalière et moyenne journalière mensuelle dans un climat aride (Ghardaïa) par les SVM en utilisant comme inputs au modèle prévisionnel des paramètres simples et disponibles in situ à savoir ; la température ambiante mesurée, la durée d'ensoleillement calculée S_0 et l'irradiation solaire extra-terrestre H_0 .

Une description de la base de données utilisée est présentée, ainsi que les étapes de construction des différents modèles SVM, que se soit pour le cas de la prédiction de l'irradiation solaire globale journalière ou moyenne journalière mensuelle (ISGJ et ISGM respectivement) reçus sur un plan horizontale. Ensuite, Les résultats obtenus sont analysés et discutés pour chacun des cas. Enfin une étude comparative est menée entre les modèles SVM prédictifs sélectionnés et les modèles de réseaux de neurones artificiels (RNA) qui sont aussi construits à partir de la même base de données. Aussi, ces modèles présentés sont comparés à d'autres modèles publiés dans la littérature.

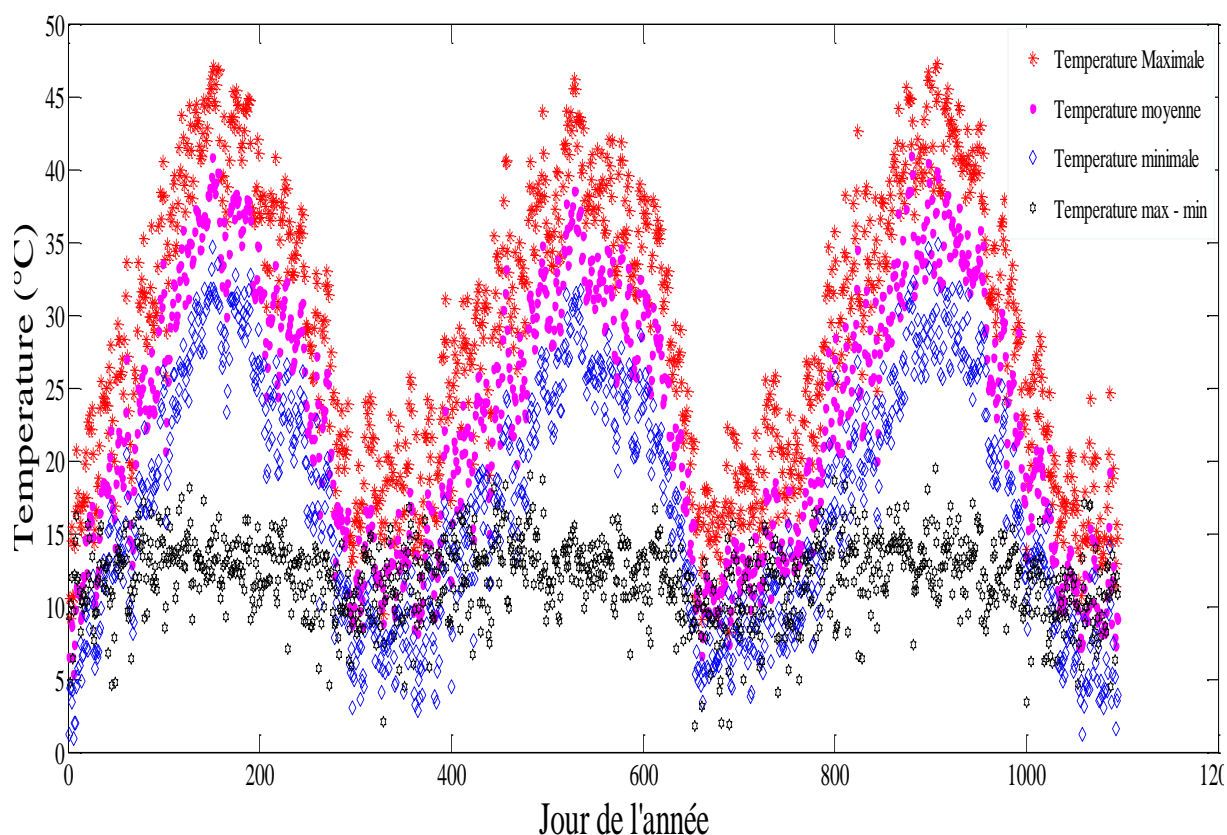


Figure 19 : Les températures mesurées à L'URAER Ghardaïa

2. Données et méthodologie

2.1 Base de données

L'ensemble des données météorologiques mesurées in situ et exploitées pour l'élaboration des modèles prédictifs sont : les températures ($^{\circ}\text{C}$) (voir Figure 19) ainsi que l'irradiation solaire globale journalière reçu sur un plan horizontale ($\text{Wh}/\text{m}^2/\text{J}$) (voir Figure 20).

La Figure 19, présente les différentes températures mesurées sur le site de Ghardaïa tel les températures ; maximale (T_{max}), minimale (T_{min}), moyenne (T_{moy}) et maximale-minimale (T_{diff}).

La Figure 20, présente les variations le l'irradiation solaire globale journalières tout au long des trois années considérées (2012, 2013, et 2014) pour le développement des modèles prédictifs.

Comme on peut le constater, l'ISGJ mesurées est comprise entre $2500 \text{ Wh}/\text{m}^2/\text{J}$ et $8000 \text{ Wh}/\text{m}^2/\text{J}$. L'énergie cumulée par an dépasse les $2\text{MWh}/\text{m}^2/\text{an}$.

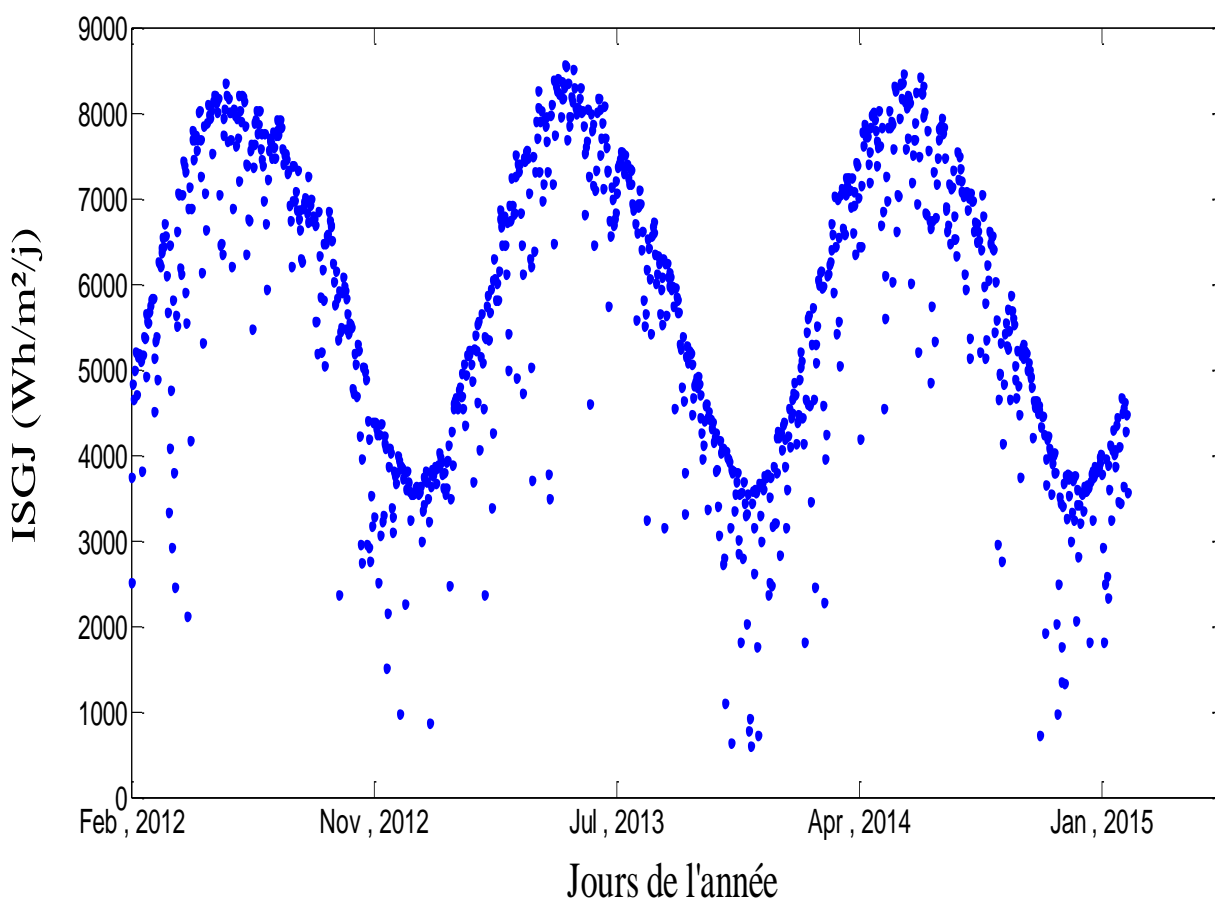


Figure 20 : Variation de l'Irradiation solaire globale journalière in situ

Pour chacun des mois des années considérées, nous avons pris le soin de calculer le minimum, le maximum ainsi que la moyenne de l'ISGJ (Figure 21).

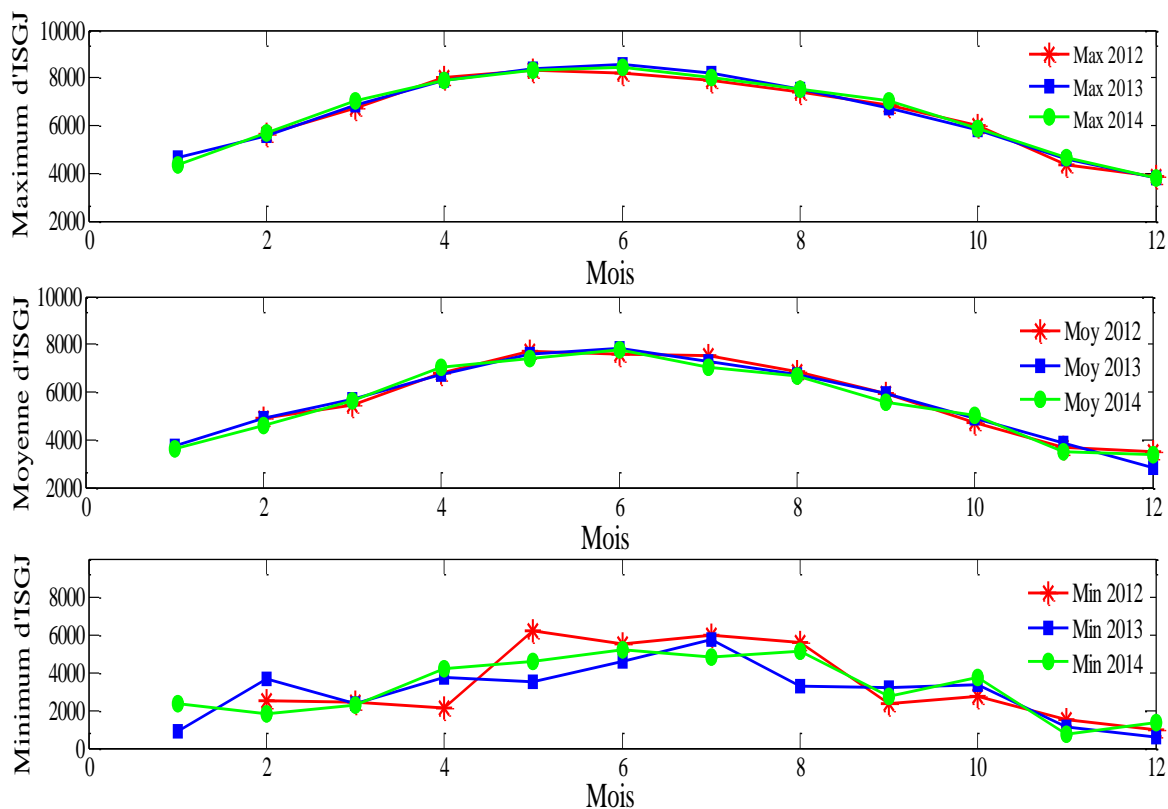


Figure 21 : Maximums, minimums et moyennes des ISGJ pour chacun des mois (2012, 2013 et 2014)

Durant l'année, l'ISGJ varie d'un minimum de 592Wh/m²/J(en Décembre) à un maximum de 8565Wh/m²/J (en Juin) et la moyenne journalière mensuelle sur une surface horizontale est de 5551Wh/ m²/J

On constate également que les valeurs moyennes et maximales de l'ISGM pour les trois années sont pratiquement confondues, ce qui est la caractéristique principale de ce site. En ce qui concerne la valeur minimale de l'ISGM, une différence a été observée pour presque tous les mois, et cela peut être expliqué soit par une rare journée de mauvais temps ou par un autre phénomène très imprévisible qui est le vent de sable.

2.2 Choix des entrées des modèles SVM

Pour la prédiction à une longueur d'avance de l'ISGJ et l'ISGM, nous avons choisit d'élaborer des modèles SVM (Support Vector Machine) en exploitant respectivement des données journalières et moyennes journalières mensuelles.

Les entrées (attributs, inputs) utilisées sont: T_{max} , T_{min} , T_{moy} , $(T_{max} - T_{min})$, l'irradiation solaire hors atmosphère (H_0) et la durée d'ensoleillement maximale (S_0).

On pose $T_{max} - T_{min} = T_{diff}$

La Figure 22, illustre l'évolution journalière de H_0 , S_0 calculés ainsi que de l'ISGJ mesurées pour les années 2012, 2013 et 2014.

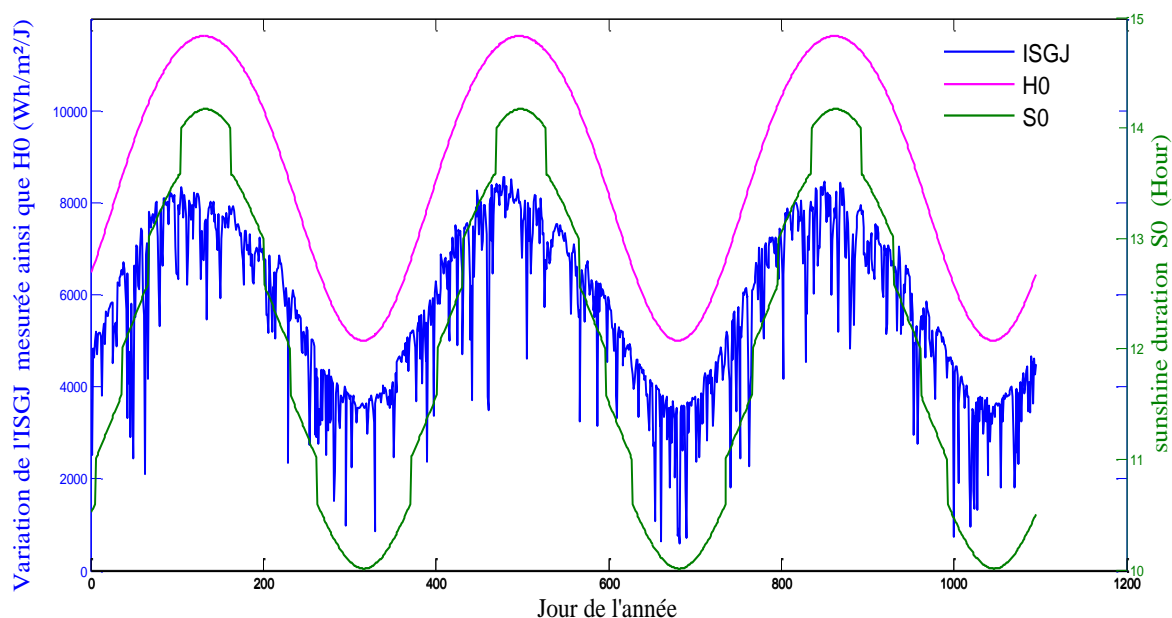


Figure 22 : Evolution de l'ISGJ mesurée ainsi que H_0 avec S_0 calculés durant la période 2012-2014

Afin de justifier le choix de ces inputs (T_{min} , T_{max} , T_{diff} , T_{moy} , S_0 et H_0), nous traçons, les variations de l'ISGJ en fonction de chacune de ces entrées tout en calculant leurs coefficients de corrélation (Figure 23). Aussi la P-value (P) exprimée en probabilité est calculée et sert à déterminer l'importance de chaque corrélation. Plus la valeur de P est petite (<0.01) mieux c'est [101].

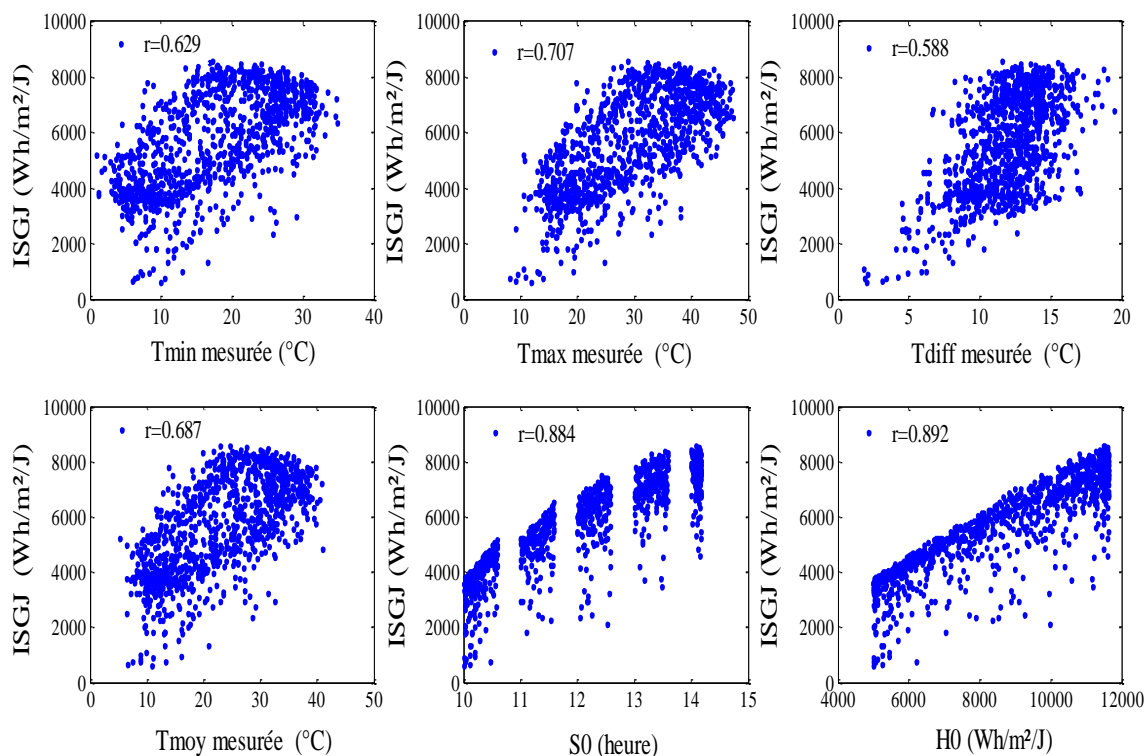


Figure 23 : Evolution de l'ISGJ par rapport à chacune des entrées

De cette Figure 23, on constate que les coefficients des corrélations sont compris entre 0.58 et 0.89 (avec $p < 0.001$). Par conséquent, ces entrées peuvent être candidats pour la prédiction de l'ISGJ et l'ISGM. Différentes combinaisons de ces inputs ont été réalisées (Tableau 7) pour la construction des modèles SVM prédictifs.

2.3 Méthodologie de construction des modèles SVM

Les étapes classiques de prévision sont principalement réalisées en deux phases: une phase d'apprentissage et une phase de validation. L'estimateur reçoit en entrée les données d'apprentissage utilisées pour générer un modèle. Par la suite, il reçoit des données de test qui serviront à prédire de nouvelles valeurs sur la base du modèle généré précédemment. Une fois le processus terminé, les données de test et les résultats de prévision sont comparés afin d'estimer le taux d'erreur (voir Figure 24).

Deux ans de données (2012 et 2013) sont choisis pour la phase d'apprentissage, et une année (2014) est utilisée pour la phase de test. Les données sont normalisées entre -1 et +1 en utilisant la formule suivante (N est le nombre d'observation) :

$$q_{\text{normalisé}} = \frac{2}{\max(q_i)_{i=1..N} - \min(q_i)_{i=1..N}} \left(q - \frac{\max(q_i)_{i=1..N} + \min(q_i)_{i=1..N}}{2} \right)$$

D'après [102], les SVM sont efficaces pour résoudre les problèmes non linéaires même avec une base de données d'apprentissage restreinte.

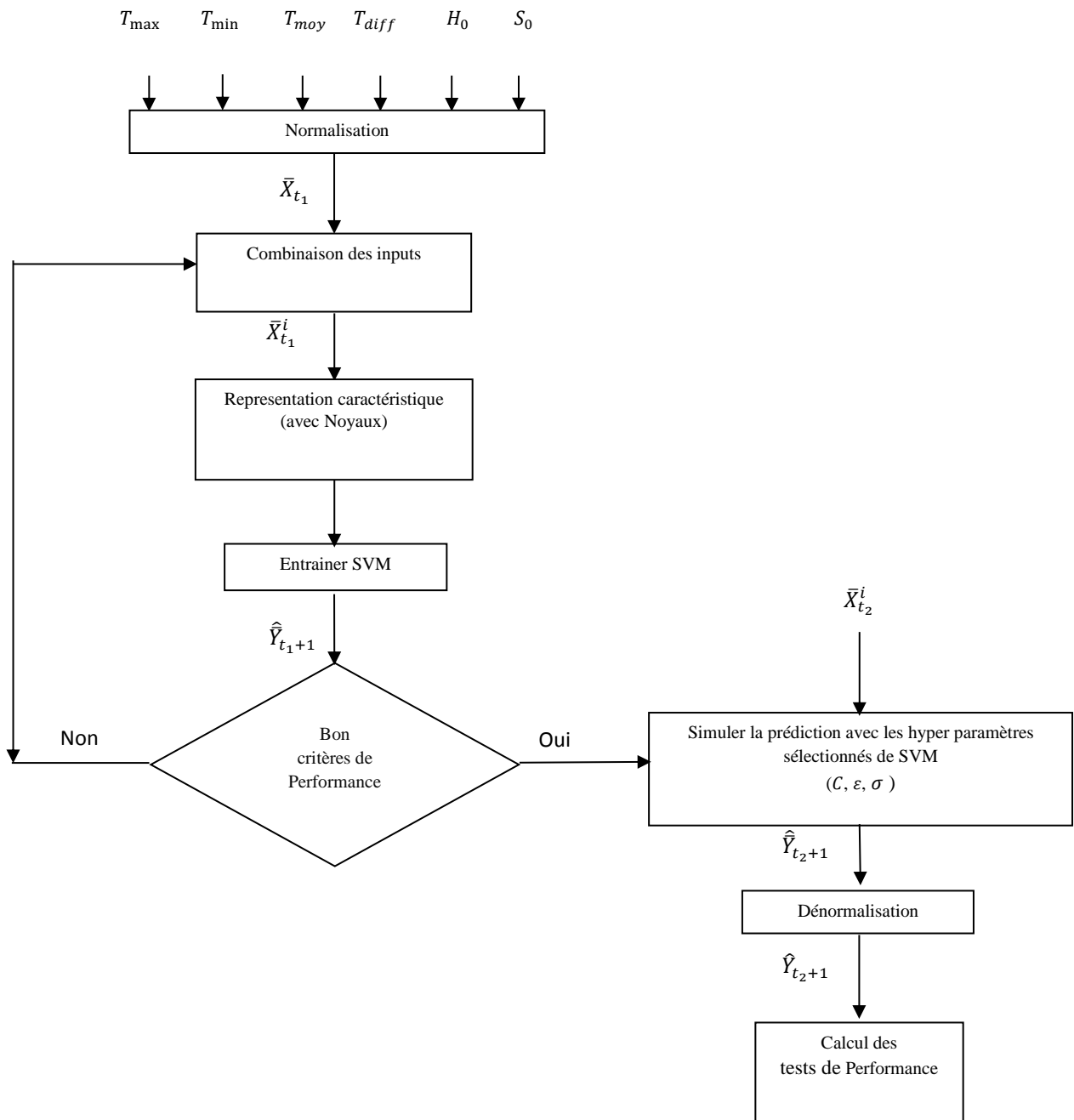


Figure 24 : Les étapes de construction des modèles SVM

Chapitre 3 : Prédiction de l'irradiation solaire globale par les techniques d'apprentissage statistiques

Avec ;

$\bar{X}_{t_1}^i$: Données d'entrée pour l'entraînement (2ans)

$\bar{X}_{t_2}^i$: Données d'entrée pour le test (1 an)

\hat{Y}_{t_1+1} : l'ISG estimée par SVM pour un pas de temps en avance

\hat{Y}_{t_2+1} : L'ISG prédite par SVM à un pas de temps en avance

Tableau 7 : Les différentes combinaisons utilisées pour la construction des modèles SVM

Modèle	Inputs	Modèle	Inputs	Modèle	Inputs
SVM ₁	DT _{max}	SVM ₁₅	DT _{max} , S ₀	SVM ₂₉	DT _{max} , H ₀
SVM ₂	DT _{min}	SVM ₁₆	DT _{min} , S ₀	SVM ₃₀	DT _{min} , H ₀
SVM ₃	DT _{moy}	SVM ₁₇	DT _{moy} , S ₀	SVM ₃₁	DT _{moy} , H ₀
SVM ₄	DT _{diff}	SVM ₁₈	DT _{diff} , S ₀	SVM ₃₂	DT _{diff} , H ₀
SVM ₅	DT _{max} , DT _{min}	SVM ₁₉	DT _{max} , DT _{min} , S ₀	SVM ₃₃	DT _{max} , DT _{min} , H ₀
SVM ₆	DT _{max} , DT _{moy}	SVM ₂₀	DT _{max} , DT _{moy} , S ₀	SVM ₃₄	DT _{max} , DT _{moy} , H ₀
SVM ₇	DT _{max} , DT _{diff}	SVM ₂₁	DT _{max} , DT _{diff} , S ₀	SVM ₃₅	DT _{max} , DT _{diff} , H ₀
SVM ₈	DT _{min} , DT _{moy}	SVM ₂₂	DT _{min} , DT _{moy} , S ₀	SVM ₃₆	DT _{min} , DT _{moy} , H ₀
SVM ₉	DT _{min} , DT _{diff}	SVM ₂₃	DT _{min} , DT _{diff} , S ₀	SVM ₃₇	DT _{min} , DT _{diff} , H ₀
SVM ₁₀	DT _{moy} , DT _{diff}	SVM ₂₄	DT _{moy} , DT _{diff} , S ₀	SVM ₃₈	DT _{moy} , DT _{diff} , H ₀
SVM ₁₁	DT _{max} , DT _{min} , DT _{moy}	SVM ₂₅	DT _{max} , DT _{min} , DT _{moy} , S ₀	SVM ₃₉	DT _{max} , DT _{min} , DT _{moy} , H ₀
SVM ₁₂	DT _{max} , DT _{min} , DT _{diff}	SVM ₂₆	DT _{max} , DT _{min} , DT _{diff} , S ₀	SVM ₄₀	DT _{max} , DT _{min} , DT _{diff} , H ₀
SVM ₁₃	DT _{max} , DT _{moy} , DT _{diff}	SVM ₂₇	DT _{max} , DT _{moy} , DT _{diff} , S ₀	SVM ₄₁	DT _{max} , DT _{moy} , DT _{diff} , H ₀
SVM ₁₄	DT _{min} , DT _{moy} , DT _{diff}	SVM ₂₈	DT _{min} , DT _{moy} , DT _{diff} , S ₀	SVM ₄₂	DT _{min} , DT _{moy} , DT _{diff} , H ₀

Le Tableau 7 rapporte les différentes combinaisons possibles qui sont utilisées comme entrées aux modèles SVM développés. La lettre D, précédant les noms des inputs, désigne les données journalières. Par ailleurs, afin d'éviter la répétition du tableau pour le cas de l'ISGM, la lettre D sera remplacée par M pour désigner les moyennes journalières mensuelles. Cependant, 42 modèles prédictifs SVM sont développés pour chacun des cas (ISGJ et ISGM).

3. Résultats et discussion

3.1 Prédiction journalière

Afin d'analyser les résultats obtenus par chacun des modèles élaborés (SVM_1, \dots, SVM_{42}), quelques tests de performances sont calculés: Erreur quadratique moyenne normalisée, (NRMSE), erreur quadratique moyenne (RMSE), pourcentage d'erreur moyenne absolue (MAPE), biais d'erreur moyenne (MBE) ainsi que le coefficient de corrélation (R).

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{q_i - \hat{q}_i}{q_i} \right| * 100\%$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (q_i - \hat{q}_i)^2}$$

$$NRMSE = \frac{RMSE}{\frac{1}{N} \sum_{i=1}^N q_i} * 100$$

$$MBE = \frac{1}{N} \sum_{i=1}^N (q_i - \hat{q}_i)$$

q_i représente les mesure et \hat{q}_i représente les valeurs calculées via le modèle.

D'après [57] Différentes rangés du NRMSE peuvent définir la qualité de la prédiction ou l'efficacité du modèle:

Excellent pour $NRMSE < 10 \%$

Bon pour $10 \% < NRMSE < 20 \%$

Faible pour $20 \% < NRMSE < 30 \%$

Très faible pour $NRMSE > 30 \%$

La Figure 25 résume les résultats des performances des 42 modèles développés pour le cas de l'ISGJ.

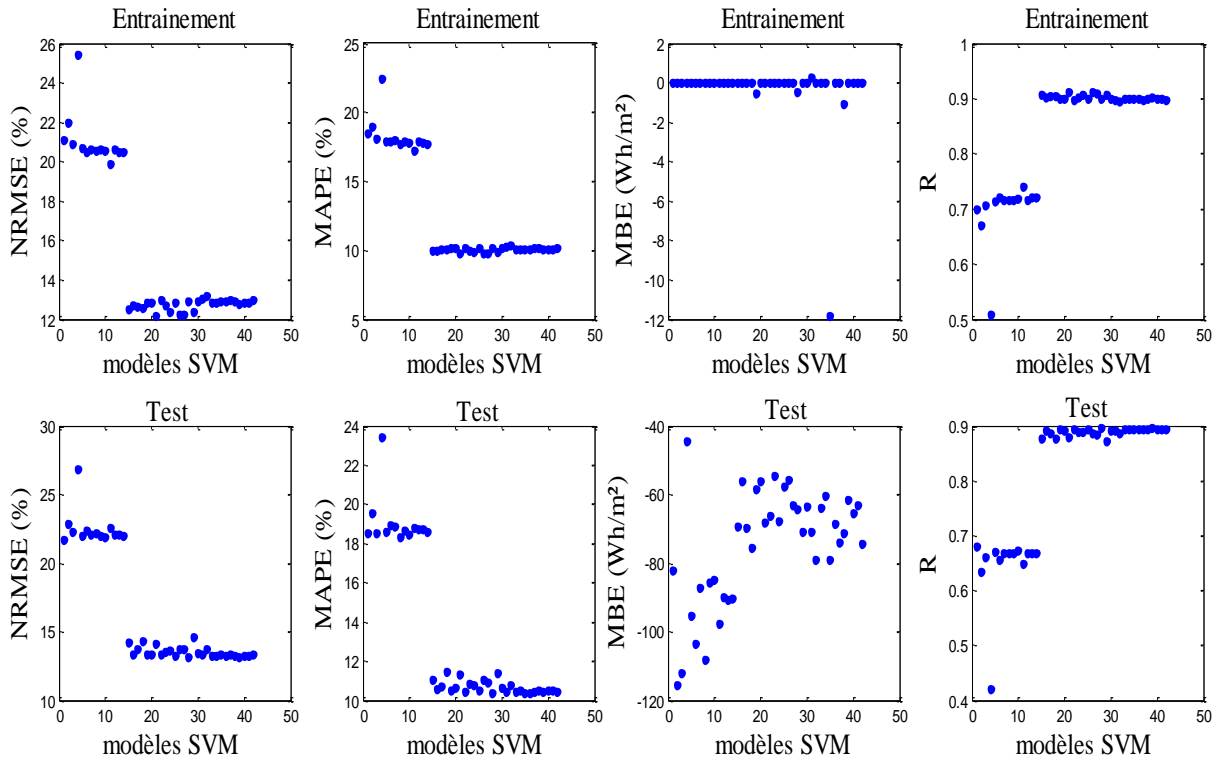


Figure 25 : Les indicateurs de performance des modèles SVM développés pour la prédiction de l'ISGJ

A partir de la Figure 25, on remarque d'ores et déjà qu'il y a deux groupes de modèles qui sont visiblement distingués par les valeurs de leurs performances (NRMSE, MAPE, MBE et R). Le premier groupe se compose de modèles ne comportant que des combinaisons de températures dans les ses entrées (SVM₁..SVM₁₄), ces modèles ne sont pas efficaces, puisque la valeur inférieure de NRMSE est de 20,095% et la plus élevée valeur de R est égale à 0.740. Un deuxième groupe de modèles utilise comme entrées les températures ainsi que H_0 ou S_0 (SVM₁₅...SVM₄₂), celui-ci donne de meilleures performances avec une valeur inférieure de NRMSE (13.163%) et la valeur plus élevée de $R = 0,896$ ($p < 0,001$). Par conséquent, il en résulte que la durée d'ensoleillement maximale S_0 ainsi que le rayonnement solaire hors atmosphère H_0 améliorent considérablement les performances des modèles, car la relation entre l'ISG et S_0 ou H_0 est relativement bonne (d'après la Figure 23).

Parmi ce second groupe, quatre modèles sont sélectionnés, pour la prédiction de l'ISGJ, pour leurs meilleures performances, qui sont résumés dans le Tableau 8. On note que ces

Chapitre 3 : Prédiction de l'irradiation solaire globale par les techniques d'apprentissage statistiques

quatre modèles considèrent les différentes températures mesurées DT_{max} , DT_{min} , DT_{moy} en plus de H_0 ou S_0 .

Tableau 8 : Résultats des performances des modèles sélectionnés **SVM₂₅**, **SVM₂₈**, **SVM₃₉** et **SVM₄₂**

Modèles	Inputs		RMSE	NRMSE	MAPE	MBE	R
			(Mj/m ²)	(%)	(%)	(Mj/m ²)	
SVM ₂₅	DT_{max} , DT_{min} , DT_{moy} , DS_0	Entrainment	2.727	12.740	10.181	0.105	0.900
		prediction	2.798	13.266	10.503	-0.207	0.894
SVM ₂₈	DT_{min} , DT_{moy} , DT_{diff} , DS_0	Entrainment	2.746	12.829	10.293	0.115	0.900
		prediction	2.777	13.163	10.403	-0.232	0.896
SVM ₃₉	DT_{max} , DT_{min} , DT_{moy} , DH_0	Entrainment	2.727	12.742	10.137	0.101	0.901
		prediction	2.779	13.172	10.458	-0.221	0.896
SVM ₄₂	DT_{min} , DT_{moy} , DT_{diff} , DH_0	Entrainment	2.755	12.875	10.058	0.069	0.898
		prediction	2.807	13.305	10.440	-0.267	0.894

Le Tableau 8 illustre que lors de l'étape de prédiction, les valeurs du MAPE sont inférieures à 11%, ce qui est très acceptable [57], les valeurs du RMSE sont comprises entre 2,777 et 2,807 MJ /m²/j et le coefficient de corrélation R est compris entre 0,894 et 0,896 avec $p < 0,001$ qui signifie que par ces modèles, la prédiction de l'ISGJ à un pas de temps d'une journée est réalisée avec un intervalle de confiance supérieur à 95%. Aussi, le MBE varie de - 0,267 MJ/m²/j à -0.207 MJ/m²/jour, montrant que ces modèles ont tendance à surestimer légèrement l'ISGJ in situ (Figure 26).

Les hyper-paramètres obtenus lors de l'entraînement des modèles SVM sélectionnés pour la prédiction sont résumés dans le Tableau 9. 'C' est le paramètre de régularisation et ' σ ' est la largeur de la bande passante du noyau RBF utilisé.

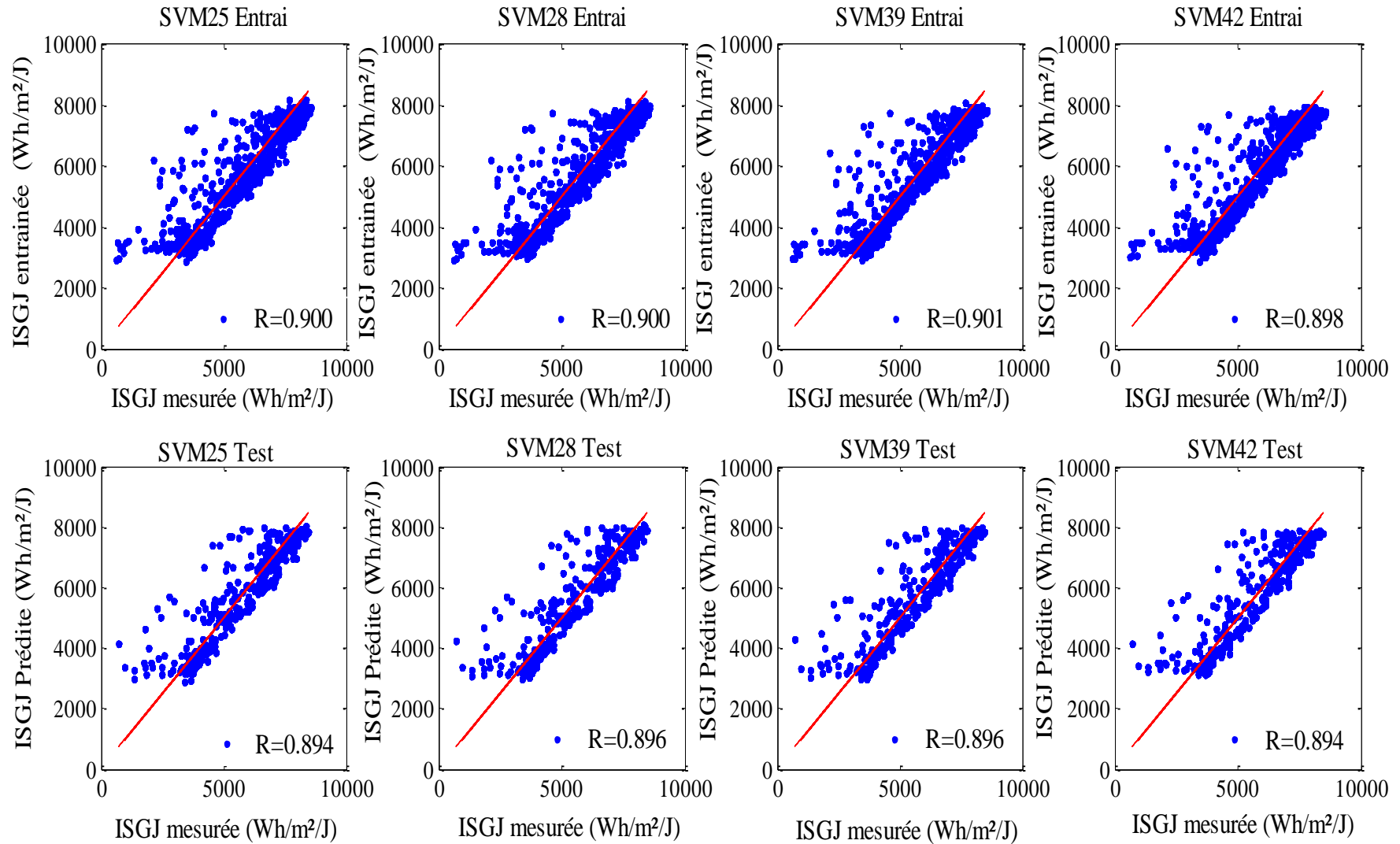


Figure 26 : Tracés des valeurs entraînées et prédites par les modèles SVM₂₅, SVM₂₈, SVM₃₉ et SVM₄₂ en fonction des valeurs mesurées de l'ISGJ.

Tableau 9: Hyper-paramètres des modèles SVM sélectionnés

SVM model	C	σ	SVM model	C	σ
SVM ₂₅	11643	2.774	SVM ₃₉	2172500	8.666
SVM ₂₈	3042	2.274	SVM ₄₂	967100	12.364

La Figure 26 illustre les graphes associés aux ISGJ entraînées ainsi que celles prédites par les modèles SVM₂₅, SVM₂₈, SVM₃₉ et SVM₄₂ en fonction des valeurs mesurées de l'ISGJ in situ. Le tracé de la première bissectrice $y=x$ est ajouté également pour chacun des graphes afin de montrer les bonnes corrélations qu'il y a entre les résultats de l'entraînement avec la prédiction et les mesures.

3.2 Prédiction mensuelle

Après construction des différents modèles prédictifs de l'ISGM (SVM₁..SVM₄₂) à partir des combinaisons des entrées considérées précédemment (T_{max} , T_{min} , T_{moy} , T_{diff} , H_0 et S_0), les tests statistiques calculés pour chacun des modèles sont calculés et présentés sur la Figure 27.

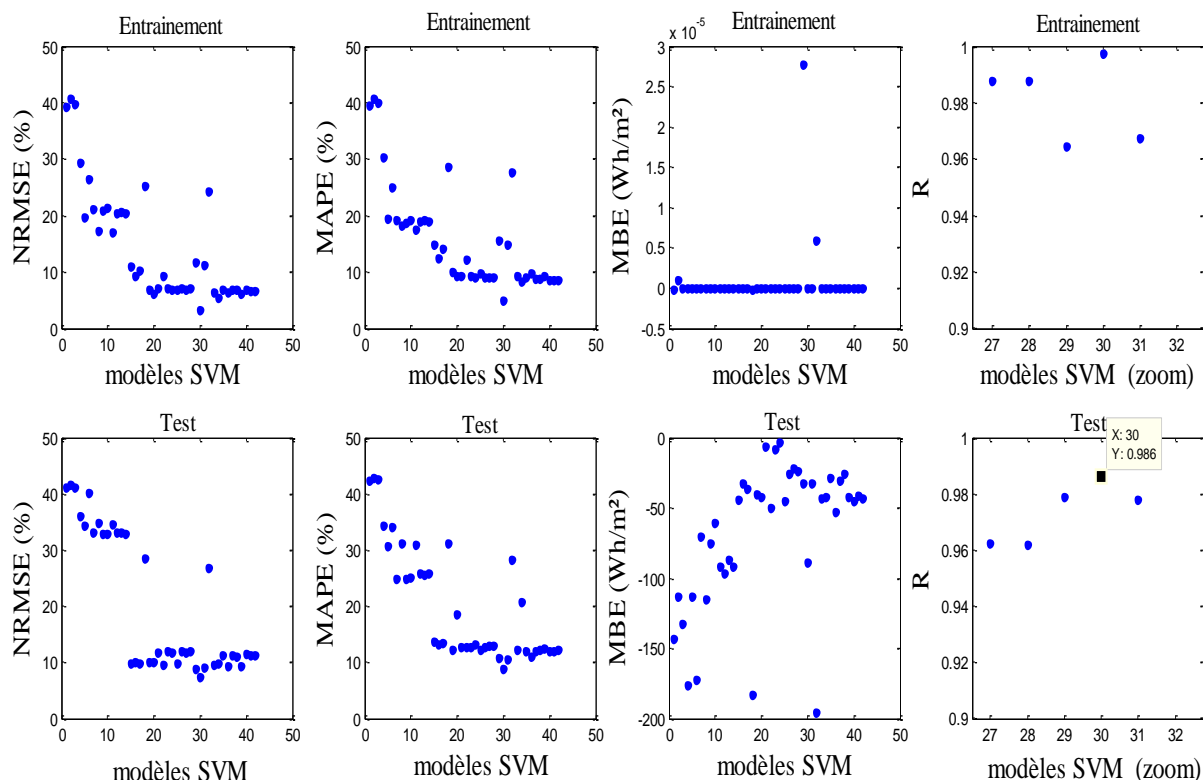


Figure 27 : Indicateurs de performances des différents modèles SVM développés pour l'ISGM

D'après la Figure 27, on observe clairement que le modèle SVM30, dont les entrées sont la température minimale moyenne mensuelle MT_{\min} et l'irradiation solaire hors atmosphère moyenne mensuelle MH_0 , se distingue par rapport aux autres modèles SVM développés. Ce modèle SVM₃₀ présente de meilleures performances dans les deux phases ; d'apprentissage et de prédiction. Les tests statistiques du modèle ainsi sélectionné sont présentés sur le Tableau 10.

Tableau 10: Résultats des performances du modèle SVM₃₀

Modèle	Inputs	RMSE	NRMSE	MAPE	MBE	R	
		(MJ/m ²)	(%)	(%)	(MJ/m ²)		
SVM ₃₀	MH ₀ , MT _{min}	Entraînement	0.673	3.305	4.842	0	0.997
		Test	1.524	7.442	8.940	-0.069	0.986

Les résultats affichés sur le Tableau 10 montrent que le modèle SVM₃₀ peut prédire avec précision l'ISGM, les performances de ce derniers sont tel que le coefficient de corrélation R

est de 0,986 (avec $p < 0,001$), le MAPE est de 8.940% qui est pratiquement excellent (d'après [57]) et le NRMSE=7.442%.

Les hyper-paramètres obtenus lors de l'entraînement du modèle SVM₃₀ sélectionné sont résumés dans le Tableau 11.

Tableau 11 : Hyper-paramètres obtenus pour le modèle SVM₃₀

SVM model	C	σ
SVM ₂₅	147.19	0.256

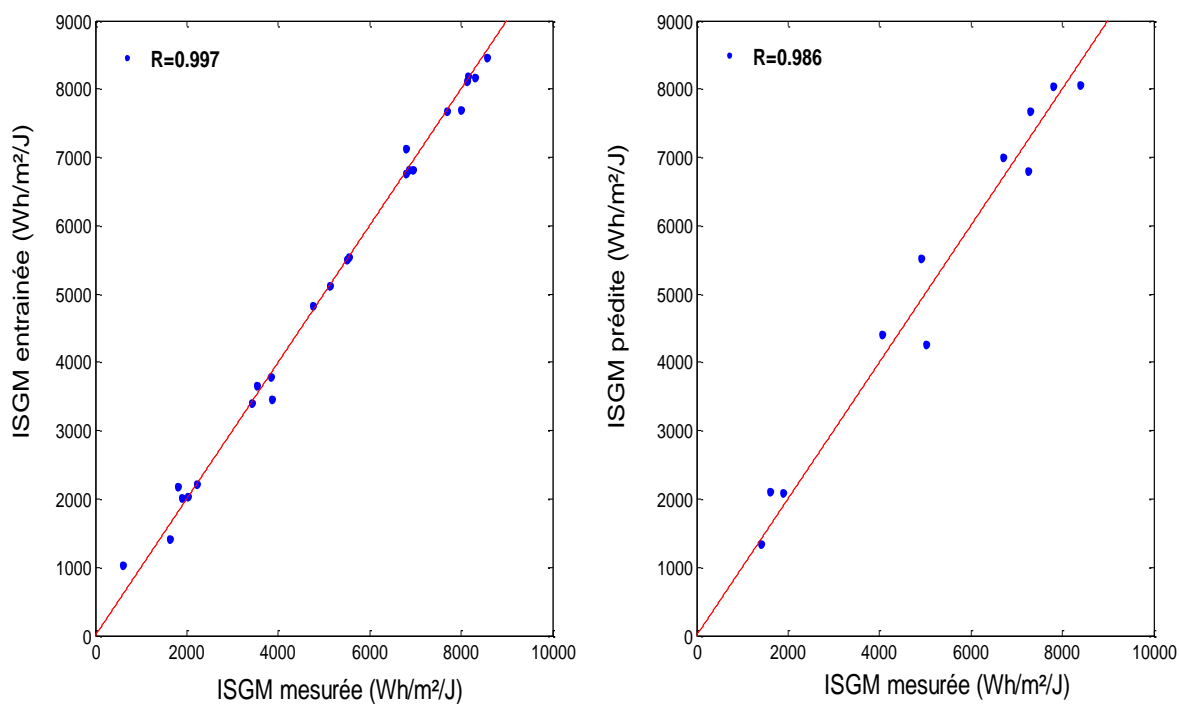


Figure 28 : Comparaison des ISGM mesurées avec celles entraînées et prédites

La Figure 28 illustre les graphes associés aux ISGM entraînées ainsi que celles prédites par le modèle SVM₃₀ en fonction des valeurs mesurées de l'ISGM in situ. Les valeurs des résultats obtenus à l'entraînement sont pratiquement collés à la première bissectrice ($y=x$) et celles prédites sont aussi parfaitement corrélées vu que le coefficient de corrélation dépasse les 98% ($p < 0,001$). Aussi, il est à noter que le modèle prédictif de l'ISGM est obtenu avec seulement un paramètre simple à mesurer (température minimale MT_{min}) et un paramètre

calculable (H_0). La température minimale mensuelle MT_{\min} est un paramètre important pour la prédiction de l'ISGM dans cette région aride.

4. Etude comparative

4.1 Cas de l'irradiation solaire globale journalière ISGJ

Les réseaux neuronaux artificiels (RNA) sont une méthode classique la plus commune basée sur des algorithmes d'apprentissage et qui ont prouvé leur exploit dans plusieurs domaines d'application, en particulier dans la prédiction de l'énergie solaire [38, 103]. Par conséquence, nous avons choisi de construire des modèles à base de perceptron multi couche (PMC) en utilisant des entrées qui seront les mêmes que celles utilisées dans les modèles SVM déjà sélectionnés (SVM_{25} , SVM_{28} , SVM_{39} and SVM_{42}), afin qu'on puisse faire une comparaison entre les performances de prédiction de l'ISGJ par les modèles RNA et celles des modèles SVM.

4.1.1 Le modèle PMC développé pour ISGJ

Comme il n'y a pas de règles rigoureuses sur le choix du nombre de couches ou de neurones dans chacune de ces couches à construire pour apprendre un quelconque phénomène, il a fallu tester plusieurs architectures. La Figure 29 illustre l'architecture du réseau MLP adopté (4-10-10-1) qui permet d'approcher les meilleurs résultats de prédiction de l'ISG.

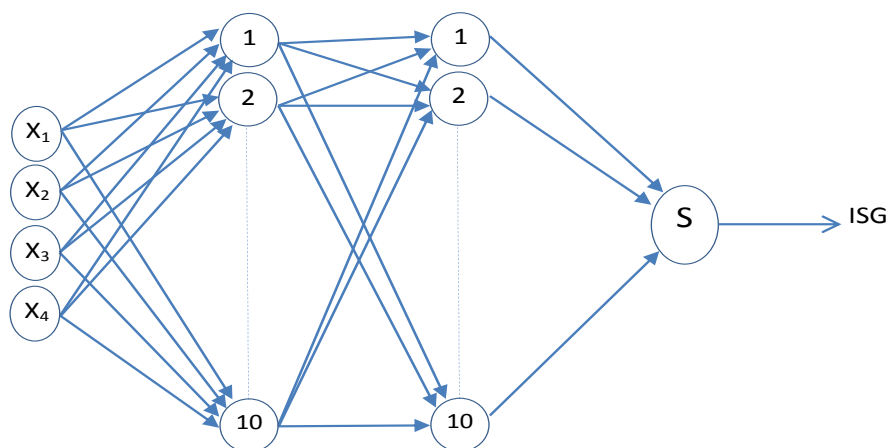


Figure 29 : Architecture du réseau PMC sélectionné (ISGJ)

Après avoir développé les modèles PMC (RNA), les résultats de leurs coefficients de corrélation ainsi que ceux des modèles SVM sont représentés simultanément sur la Figure 30.

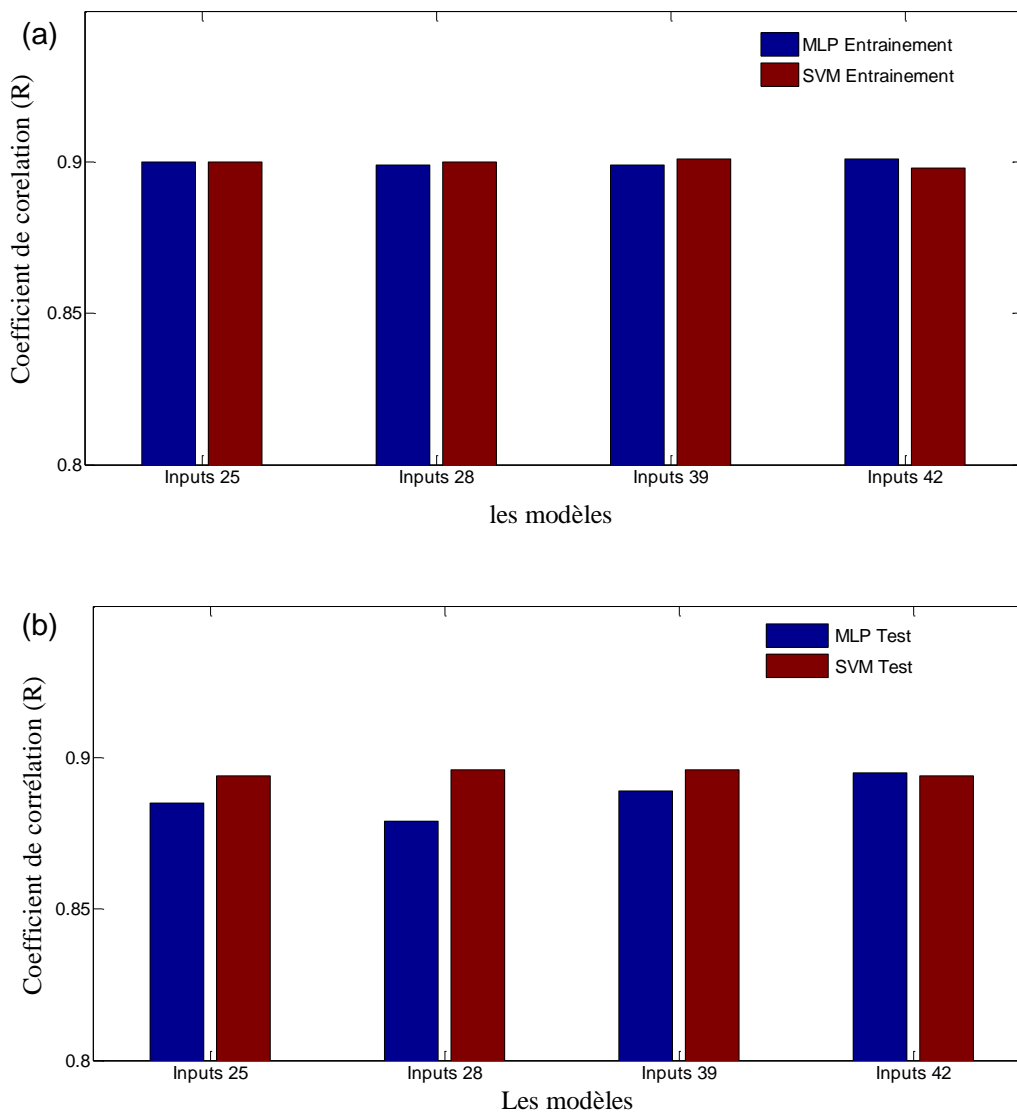


Figure 30 : Comparaison des performances des deux modèles SVM et RNA développés
(a) Entraînement, (b) Prédiction

D'après ces graphes, on voit que les modèles PMC et SVM fournissent pratiquement les mêmes résultats de performances durant l'étape d'entraînement, tandis qu'à l'étape de test, les modèles SVM améliorent légèrement les performances de la prédiction. Ces dernières sont affichées sur le Tableau 12.

Tableau 12: Comparaison des performances de prédiction de l'ISGJ par SVM et PMC

Inputs	Modèles	RMSE (Mj/m ²)	NRMSE (%)	MAPE (%)	MBE (Mj/m ²)	R
DT _{max} , DT _{min} , DT _{mean} , DS ₀	SVM ₂₅	2.798	13.266	10.503	-0.207	0.894
	MLP ₂₅	2.925	13.869	10.937	-0.320	0.885
DT _{min} , DT _{mean} , DT _{diff} , DS ₀	SVM ₂₈	2.777	13.163	10.403	-0.232	0.896
	MLP ₂₈	3.047	14.443	10.531	-0.673	0.879
DT _{max} , DT _{min} , DT _{mean} , DH ₀	SVM ₃₉	2.779	13.172	10.458	-0.221	0.896
	MLP ₃₉	2.871	13.612	10.504	-0.373	0.889
DT _{min} , DT _{mean} , DT _{diff} , DH ₀	SVM ₄₂	2.807	13.305	10.440	-0.267	0.894
	MLP ₄₂	2.788	13.216	10.281	-0.292	0.895

A partir du Tableau 12, on peut constater que les résultats obtenus par les modèles SVM sont dans la même gamme de performances par rapport à ceux obtenues par les modèles MLP. Pour les modèles de prédiction SVM, le coefficient de corrélation est compris entre 0,894 et 0,896 et le NRMSE varie de 13,163% à 13,305%. Les modèles de prédiction MLP présentent relativement les mêmes performances avec R allant de 0,879 à 0,895 et un NRMSE de 13,216% à 14,443%. Toutefois, il convient de noter que les modèles de réseaux de neurones (PMC) ne garantissent pas la convergence vers une solution optimale contrairement au SVM.

4.1.2 Comparer les performances du modèle SVM proposé à ceux publiés dans la littérature

Afin de prouver l'efficacité des SVM dans la prédiction de l'énergie solaire (ISGJ), ainsi que sa compétitivité par rapports aux autres techniques d'apprentissage statistique, nous allons donc nous référer au Tableau 2 du Chapitre 1, ou nous résumons les résultats des performances de certains modèles prédictifs de l'ISGJ publiés dans la littérature y compris les modèles SVM que nous venons de développer dans de cette section.

Commentaire

A partir de ce Tableau 2, on remarque que les modèles utilisant la durée d'ensoleillement S parmi leurs inputs donnent des résultats de performance meilleurs(modèles : [40], [43] , [69]). Mais ce qu'il faut savoir, c'est que S est en parfaite corrélation avec le rayonnement

solaire global. D'après [38], la durée d'ensoleillement permet d'atteindre des résultats de performances très élevés. Par exemple, Dans [69], le coefficient de corrélation obtenue est de 0,94 pour des données recueillies en Arabie Saoudite

Par contre, le modèle SVM développé dans ce travail concurrence les modèles qui n'utilisent pas la durée d'ensoleillement (S) en entrée. Exemple le modèle ANN, développé pour prédire l'irradiation solaire en Espagne [68] et emploie différentes entrées (D , Lat , Lon , Ho , TCC , SKT , TCW , TCO), a des performances ($RMSE = 3.016MJ / m^2$ et $NRMSE = 14,20\%$) inférieures à celles obtenus par notre modèle SVM ($RMSE = 2.779MJ / m^2$ et $NRMSE = 13,172\%$). De même, le modèle construit dans [39], qui est basé sur la méthode de régression linéaire locale (RLL) utilisant en entrées ; la vitesse du vent (WS), la pression atmosphérique (Pa), les températures et le rayonnement extraterrestre (H_0) pour prédire l'ISGJ en Iran, a une coefficient de corrélation ($R = 0,900$) qui est pratiquement le même que notre modèle SVM ($R = 0,896$).

Cependant, Si nous prenons le modèle prédictif ANN développé par [41], dont les entrées utilisées sont pratiquement les mêmes que celle de notre modèle SVM (T_{max} , T_{min} , T_{moy} et H_0), on remarquera qu'ils ont pratiquement les mêmes performances.

Comme conclusion, on peut dire que ces modèles SVM peuvent être une alternative prometteuse pour prédire l'irradiation solaire globale journalière dans ce climat aride en utilisant que des températures de l'air comme inputs mesurables.

4.2 Cas de l'irradiation solaire globale moyenne mensuelle ISGM

Afin de réaliser une étude comparative concernant les modèles prédictifs de l'ISGM, un modèle a base de RNA en utilisant l'architecture perceptron multi couche (PMC) est construit. Les entrées utilisées dans le modèle SVM sélectionné (la température minimum mensuel MT_{min} et l'irradiation solaire globale moyenne mensuel MH_0) sont les mêmes à considérer pour ce modèle PMC.

4.2.1 Le modèle PMC développé pour ISGM

La Figure 31, illustre l'architecture du réseau MLP adopté (2-10-10-1) qui permet d'approcher les meilleurs résultats de prédiction de l'ISGM.

Les critères de performances sont déterminées pour chaque modèle (SVM_{30} et PMC_{30}) et résumées dans le Tableau 13.

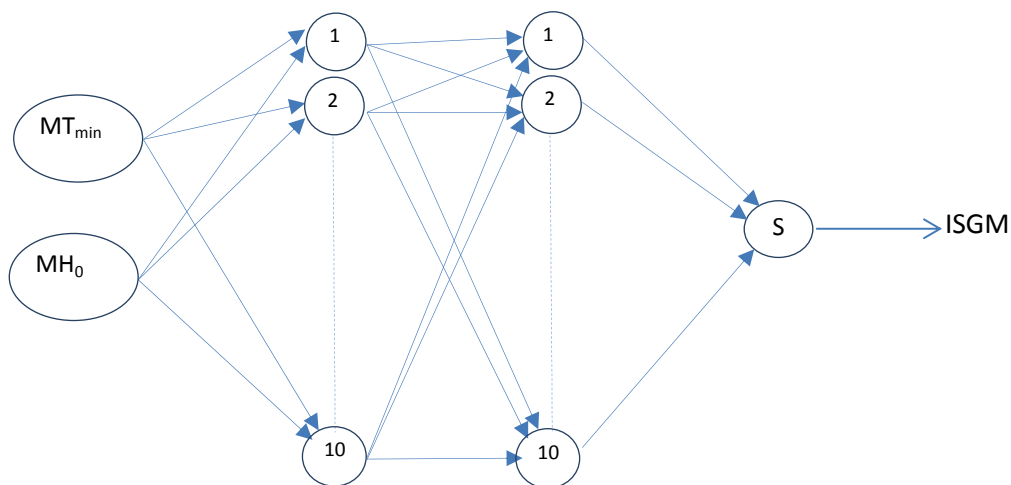


Figure 31 : Architecture du réseau PMC sélectionné (ISGM)

Tableau 13 : Comparaison entre les performances de la prédiction des modèles PMC_{30} et SVM_{30}

Inputs	Modèles		RMSE	NRMSE	MAPE	MBE	R
			(MJ/m ²)	(%)	(%)	(MJ/m ²)	
MT_{min} , MH_0	SVM_{30}	Entraînement	0.673	3.305	4.842	0	0.997
		Prediction	1.524	7.442	8.940	-0.069	0.986
	PMC_{30}	Entraînement	1.052	4.972	4.620	-0.057	0.984
		Prediction	1.596	7.510	8.950	0.099	0.970

D'après les résultats affichés dans le Tableau 13 les performances du modèle SVM_{30} sont meilleurs que celles données par le modèle PMC_{30} . Le coefficient de corrélation $R=0,986$ pour le SVM_{30} , tandis que pour le PMC_{30} $R =0,970$. La Figure 32 illustre les tracés représentatifs de l'ISGM mesurées et prédites par les deux modèles SVM_{30} et PMC_{30} .

Les déductions énoncées précédemment sont fortement soutenue par la Figure 32. Les valeurs de l'ISGM obtenues à l'entraînement ainsi qu'à la prédiction par le modèle SVM_{30} suivent correctement la tendance de la courbe des mesures. Alors que, pour le modèle PMC_{30} , des écarts significatifs sont observés entre les valeurs mesurées et entraînées (Octobre à Février), et aussi entre les valeurs mesurées et prédites (novembre à Janvier). Donc, il est clairement démontré que le modèle SVM_{30} est plus performant que le modèle

PMC₃₀ pour la prédiction de l'ISGM dans cette région qui est caractérisée par un climat aride.

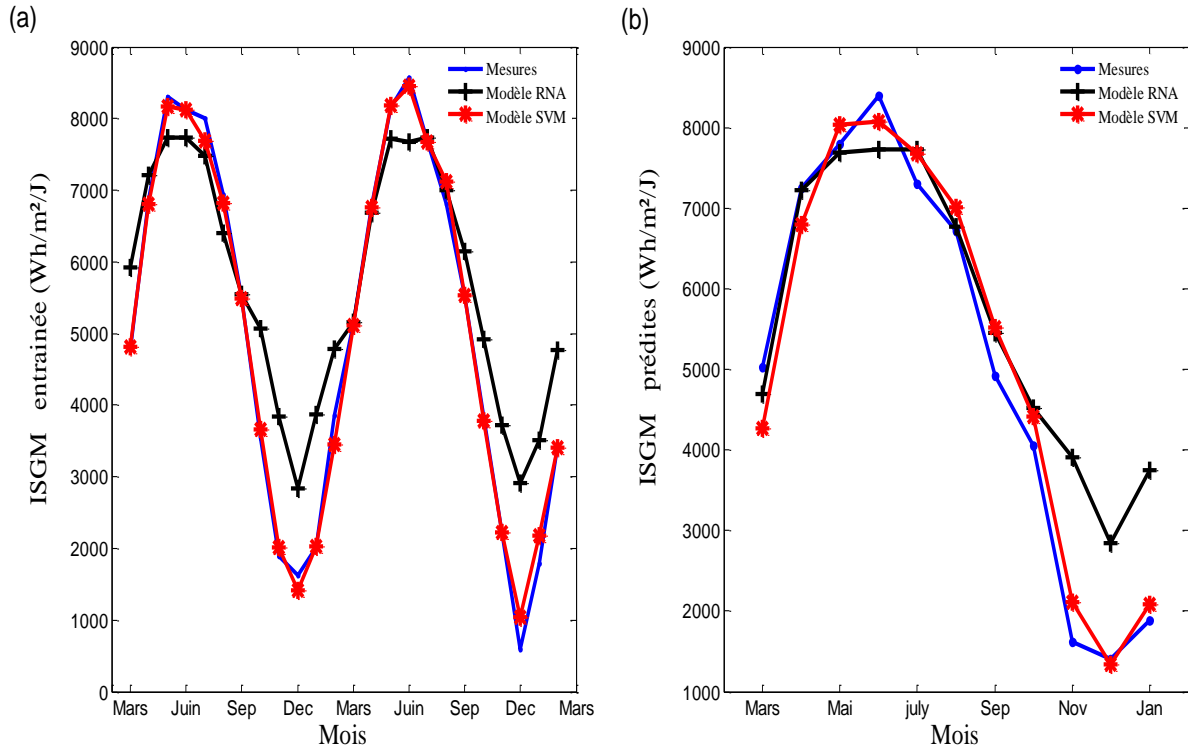


Figure 32 : Comparaison entre les mesures et valeurs prédites de l'ISGM par les deux modèles

(SVM₃₀ and MLP₃₀): a) à l'entraînement b) à la prédiction

4.2.2 Comparer les performances du modèle SVM proposé à ceux publiés dans la littérature

Afin de prouver l'efficacité des modèles SVM dans la prédiction de l'ISGM, une étude comparative est menée entre le modèle SVM₃₀ et d'autres modèles développés et publiés dans la littérature. Les différents tests de performances résultant par chacun de ces modèles sont affichés sur Le Tableau 3 du Chapitre 1.

Commentaires

En se référant au Tableau 3, il peut être clairement remarqué que tous les modèles mentionnés ont des performances inférieures par rapport à notre modèle SVM₃₀ développé, à l'exception du modèle de Sözen et al [4] vu qu'il utilise parmi ses entrées un autre paramètre

mesurable qui est la durée d'ensoleillement (MS) et aboutit à un coefficient de corrélation $R = 0,997$. Mais, notre modèle proposé SVM_{30} réalise tous de même de très bonnes performances ($R = 0,99$ et $0,986$ à l'entraînement et à la prédiction respectivement) avec un seul paramètre mesurable (température minimale quotidienne moyenne mensuelle MT_{min}) et un paramètre calculable (irradiation solaire globale hors atmosphère MH_0)

5. Conclusion partielle

Dans ce chapitre, des modèles prédictifs de l'ISGJ et l'ISGM à Ghardaïa à un pas de temps (Jour et mois respectivement) à base de techniques SVM ont été développés en utilisant des paramètres d'entrée simple : Les températures ($T_{max}, T_{min}, T_{moy}, (T_{max} - T_{min})$), ainsi que deux autres paramètres calculables : la durée d'ensoleillement maximale S_0 et l'irradiation solaire hors atmosphère (extraterrestre) H_0 .

Les résultats de la prédiction journalière (ISGJ) ont montré une amélioration significative des performances par l'introduction de DH_0 ou DS_0 au coté des températures mesurées en entrée. Ainsi, quatre modèles SVM ont été sélectionnés pour la prédiction de l'ISGJ. Les tests statistiques indiquent que les valeurs du NRMSE varient de 13,163% à 13,305%, le MAPE de 10,403% à 10,503% et R varie de 0,894 à 0,896 ($p < 0,001$).

Pour la prévision de l'irradiation solaire globale mensuelle (ISGM), les résultats ont montré clairement qu'avec seulement la température minimale (MT_{min}) et l'irradiation solaire extraterrestre MH_0 comme entrées, une très bonne prédiction peut être atteinte avec un $NRMSE = 7,442\%$, $MAPE = 8,940\%$ et un coefficient de corrélation R proche de 0,986 ($p < 0,001$).

L'efficacité de tous ces modèles SVM proposés a été prouvée par une étude comparative avec des modèles construits à base de réseaux de neurones artificiels (Perceptron Multi Couche) également développés in situ et ceux publiés dans la littérature.

**Chapitre4 : *Modèle prédictif hybride
de l'ISGJ***

1. Introduction

Toujours dans le même contexte de la prédiction de l'irradiation solaire globale, plusieurs efforts sont fournis par la communauté scientifique dans l'optimisation des résultats prédictifs en utilisant différentes techniques ; les modèles statistiques [104], [105], les processus stochastiques [106], [107], [108] ou les techniques d'intelligence artificielle (réseaux de neurones, logique floue, algorithmes génétiques, SVM, ...etc.) ([7], [30], [54], [44], [43]). Par la suite, ces dernières décennies, il y a eu l'apparition des modèles hybrides pour l'objectif d'amélioration de la prédiction d'un grand nombre de phénomènes climatiques, en particulier celle de l'ISGJ exemple ; [56] combine les Support Vector Machine (SVM) avec l'algorithme de la transformée en ondelette (Wavelet Transform WT), [65] développe un modèle hybride ARMA/RNA et [109] propose de combiner les ARMA avec TDNN (Time Delay Neural Network)...etc.

Dans ce chapitre, nous prendrons le soin de proposer un modèle hybride dans le but d'améliorer la prédiction de l'ISGJ in situ mais cette fois-ci, nous l'aborderons par le principe des séries temporelles.

Donc, nous proposons un modèle hybride basé sur la méthodologie de Box & Jenkins (SARIMA ; Seasonal Auto-Regressif Integrated Moving Average) ainsi que les SVM (Support Vector machine). Car, d'après les investigations des chercheurs, les ARMA est parmi les techniques les plus compétitives pour la prédiction des séries temporelles[8].

2. La Méthodologie hybride SARIMA-SVM

Comme le modèle SARIMA représente la partie linéaire du problème et le modèle SVM peut représenter la partie non linéaire. Alors, le modèle hybride à proposer dans cette étude peut modéliser les deux modèles linéaires et non linéaires pour une meilleure performance de prédiction.

La série temporelle à étudier dans ce travail qu'on notera Z_t peut être considérée comme une combinaison d'une structure d'autocorrélation linéaire L_t et une composante non linéaire N_t .

$$Z_t = L_t + N_t \quad (47)$$

Chapitre 4 : Modèle prédictif hybride de l'ISGJ

Chacun des polynômes L_t et N_t sera estimé à partir de l'ensemble de données à considérer.

Premièrement, on modélise la série temporelle Z_t avec le modèle SARIMA. Les résidus résultants après l'estimation de Z_t représenteront la partie non linéaire

ε_t représente le résidu au temps "t" ;

$$\varepsilon_t = Z_t - \widehat{L}_t \quad (48)$$

\widehat{L}_t est la prévision par le modèle SARIMA à instant t.

La partie non linéaire peut être trouvée en modélisant le résidu ε_t en utilisant les SVM.

$$\varepsilon_t = f(\varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-n}) + e_t = N_t \quad (49)$$

ou f est une fonction non linéaire déterminée par SVM et e_t est l'erreur.

Par conséquent, la prévision combinée sera;

$$\widehat{Z}_t = \widehat{L}_t + \widehat{N}_t \quad (50)$$

Ou:

\widehat{L}_t : La prévision par SARIMA

\widehat{N}_t : La prévision par SVM

3. Résultats et discussion

3.1 Modèle SARIMA (p,d,q)(P,D,Q)

Nous avons considéré une série temporelle de l'irradiation solaire globale journalière (ISGJ) mesurée sur une surface horizontale à Ghardaïa (le sud de l'Algérie) du 1 Avril 2012 au 1 Avril 2016 (quatre années) ; la série temporelle est prise telle qu'elle sans aucune modification apportée aux données mesurées.

A partir de ces données, nous déduisons la série temporelle de l'indice de clarté K_t qui est le rapport entre l'ISGJ et celle hors atmosphère (Equation 51).

$$K_t = \frac{ISGJ}{DH_0} \quad (51)$$

DH_0 : Irradiation globale journalière hors atmosphère.

La première étape de la modélisation SARIMA consiste à identifier l'ordre des polynômes caractéristique du modèle.

Chapitre 4 : Modèle prédictif hybride de l'ISGJ

En traçant la série temporelle K_t , on remarque déjà qu'elle présente une périodicité de 365 jours (voir Figure 33a). Alors on doit d'abord rendre la série stationnaire en procédant à la différentiation par rapport à la saisonnalité (Figure 33b).

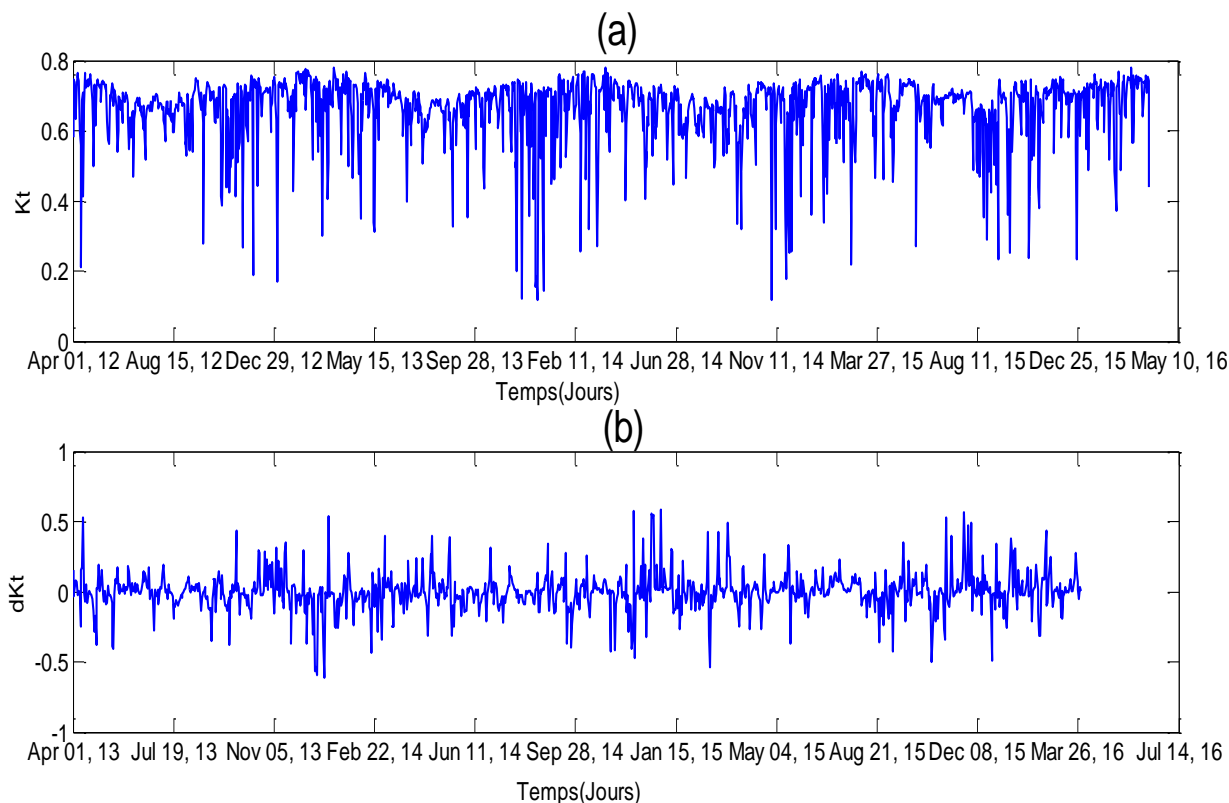


Figure 33 : La série temporelle K_t et celle différenciée dK_t

Ensuite, le tracé des autocorrélogrammes de dK_t (Figure 34) permettra de déterminer l'ordre des polynômes ($\varphi(L)$ et $\theta(L)$). On voit que la série temporelle dK_t admet un paramètre AR et un paramètre MA

Enfin, l'équation caractéristique de la série temporelle K_t par le modèle SARIMA est SARIMA $(1,0,1)(0,1,0)_{365}$ (Equation 52).

$$K_t(1 - L^{365})(1 - \varphi_1 L) = (1 - \theta_1 L)\varepsilon_t \quad (52)$$

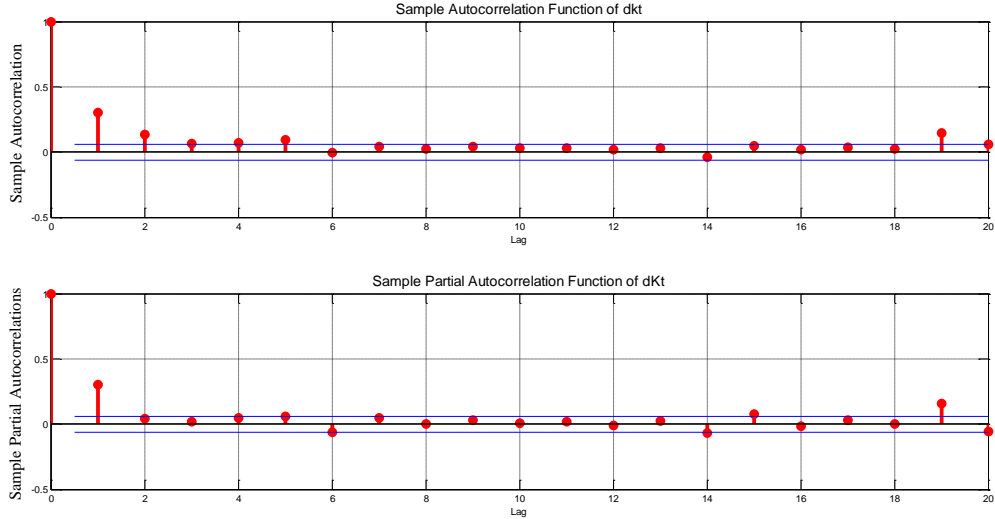


Figure 34 : Autocorrelogrammes simple et partiel de dK_t

Après avoir identifié les paramètres du modèle SARIMA de la série temporelle K_t , nous passerons à l'étape de l'estimation de chacun de ces paramètres AR et MA par la méthode du maximum de vraisemblance. On a choisit trois années pour l'estimation et gardé une année pour les prévisions.

La prévision à l'horizon h ($h \geq 1$) suivant la méthodologie de Box & Jenkins est comme suit:

$$\hat{K}_t^*(h) = \varphi_1 K_{t+h-1} + K_{t+h-365} - \varphi_1 K_{t+h-366} + \varepsilon_{t+h} - \theta_1 \varepsilon_{t+h-1} \quad (53)$$

Such as ;

$$K_{t+j} = \begin{cases} K_{t+j} & j \leq 0 \\ \hat{K}_t(j) & j > 0 \end{cases} ; \quad (54)$$

$$\varepsilon_{t+j} = \begin{cases} \varepsilon_{t+j} & j \leq 0 \\ 0 & j > 0 \end{cases} ; \quad (55)$$

$$\varepsilon_t = K_t - \hat{K}_{t-1}(1) \quad (56)$$

Avec $\hat{K}_t(j)$ l'estimation de la prévision.

Après avoir estimé et prédit K_t par le modèle SARIMA(1,0,1)(0,1,0), on revient à notre série temporelle de l'ISGJ par un simple calcul via l'équation 51.

Chapitre 4 : Modèle prédictif hybride de l'ISGJ

Les tracés de l'estimation et de la prédiction de l'ISGJ par SARIMA sont montrés sur les figures suivantes.

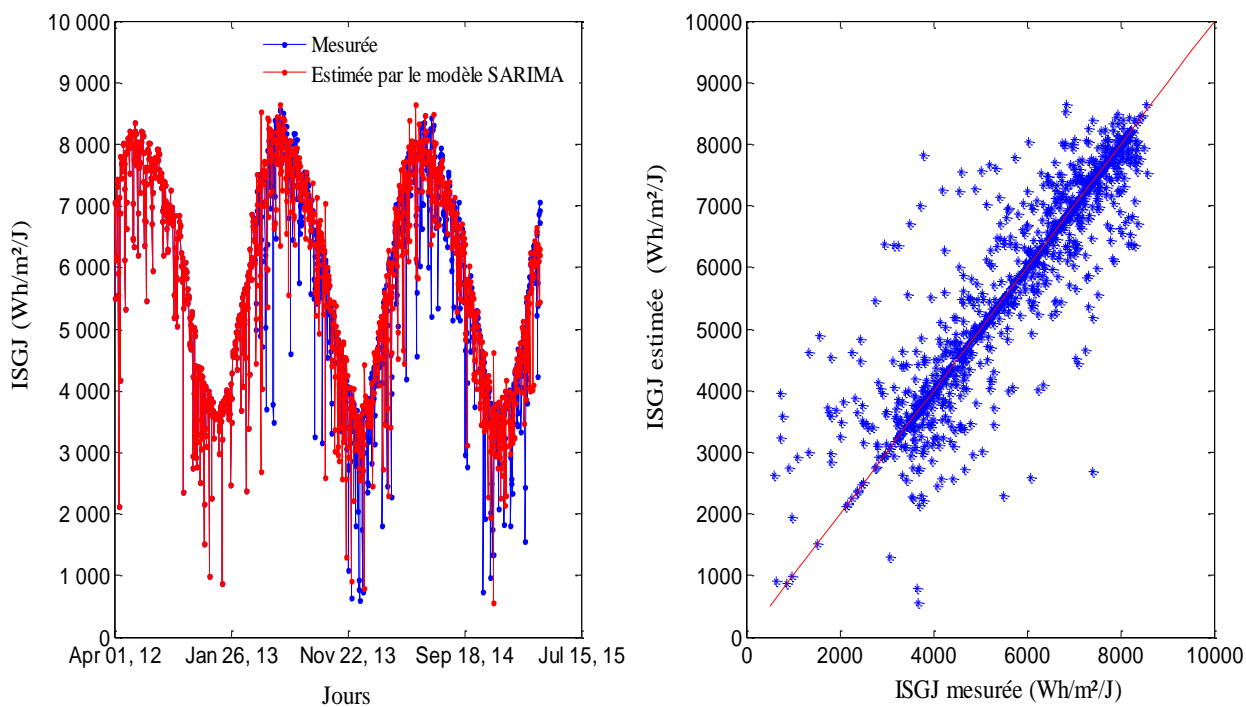


Figure 35 : ISGJ mesurées et estimées par le modèle SARIMA

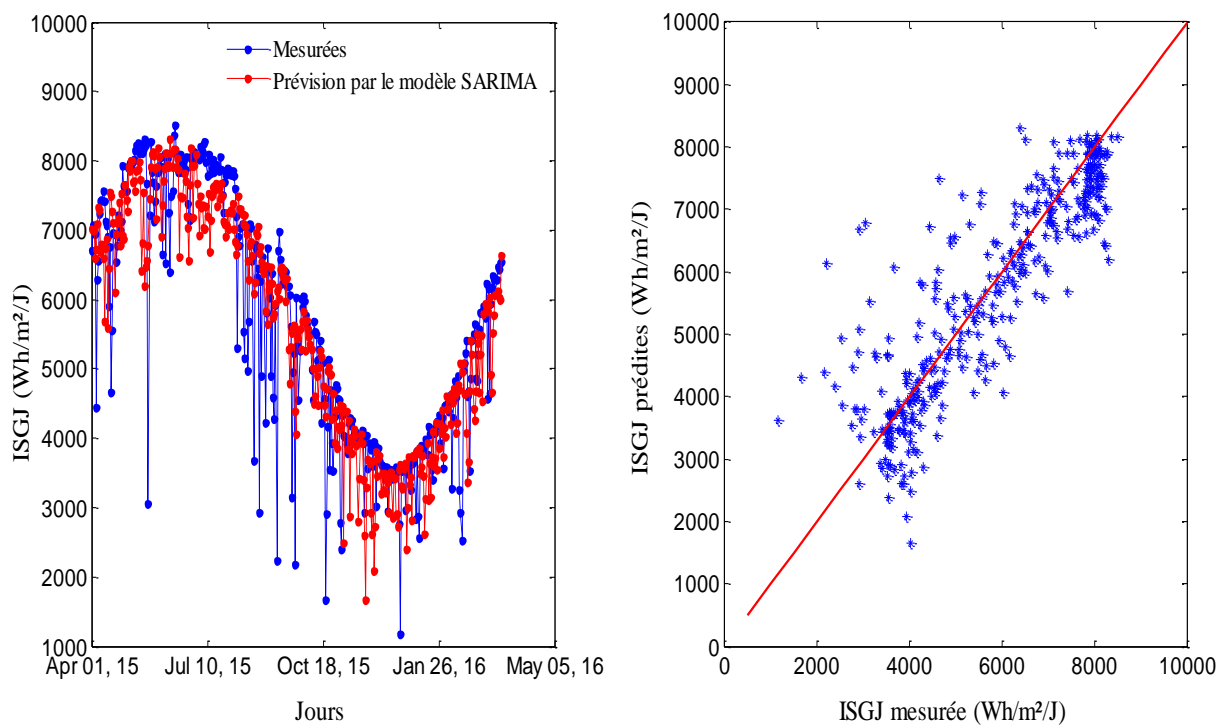


Figure 36 : ISGJ mesurées et prédites par le modèle SARIMA

Chapitre 4 : Modèle prédictif hybride de l'ISGJ

Les graphes de gauches des figures 35 et 36 représentent respectivement l'ISGJ estimées et prédites par le modèle SARIMA in situ.

Les tracés de droite des Figure 35 et 36 montre respectivement l'ISGJ estimé et prédites en fonction de celles mesurées par le modèle SARIMA. La droite de la première bissectrice est présentée sur chacun de ces graphes afin d'illustrer les bonne corrélations entre les valeurs estimé ou prédites et celle mesurées. Les résultats de performance du modèle SARIMA sont montrés sur le Tableau 14.

Tableau 14 : Performances du modèle SARIMA(1,0,1)(0,1,0)₃₆₅.

	RMSE (Wh/m²)	NRMSE (%)	MAPE (%)	MBE (Wh/m²)	R
Entraînement	772.164	13.033	11.092	-25.086	0.900
Prédiction	889.768	14.913	14.139	57.772	0.867

À partir de ce Tableau 14, on voit que le MAPE est au alentour de 14%, qui est une valeur acceptable, le RMSE = 889.768 Wh/m²/J, le coefficient de corrélation R = 0.867 (p<0.001) qui signifie que par ce modèle SARIMA(1,0,1)(0,1,0)₃₆₅, l'ISGJ peut être prédites avec un intervalle de confiance supérieur à 95%. Le MBE = 57.772 Wh/m²/J, montrant que le modèle tend à sous-estimer légèrement l'ISGJ dans cette région aride.

3.2 Le modèle hybride SARIMA-SVM:

Après avoir estimé l'indice de clarté K_t par le modèle SARIMA, on procède au calcul des résidus :

$$\varepsilon_t = K_t - \widehat{K}_t \quad (57)$$

Par la suite, nous appliquerons le principe des SVM pour prédire $\varepsilon_t = f(\varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-n})$.

Une fois la prédiction de ε_t obtenue, on calcule le nouvel indice de clarté issu du modèle hybride ;

$$\widehat{K}_{t/\text{Hybrid}} = \widehat{K}_{t/\text{SARIMA}} + \varepsilon_{t/\text{SVM}} \quad (58)$$

Par la même formule que l'Equation 51, nous obtenons l'ISGJ prédite par le modèle hybride SARIMA_SVM.

Chapitre 4 : Modèle prédictif hybride de l'ISGJ

Les tracés de l'irradiation solaire globale journalière mesurées et prédites sont présentés sur la Figure 37 et les résultats de performances de la prévision sont affichés sur le Tableau

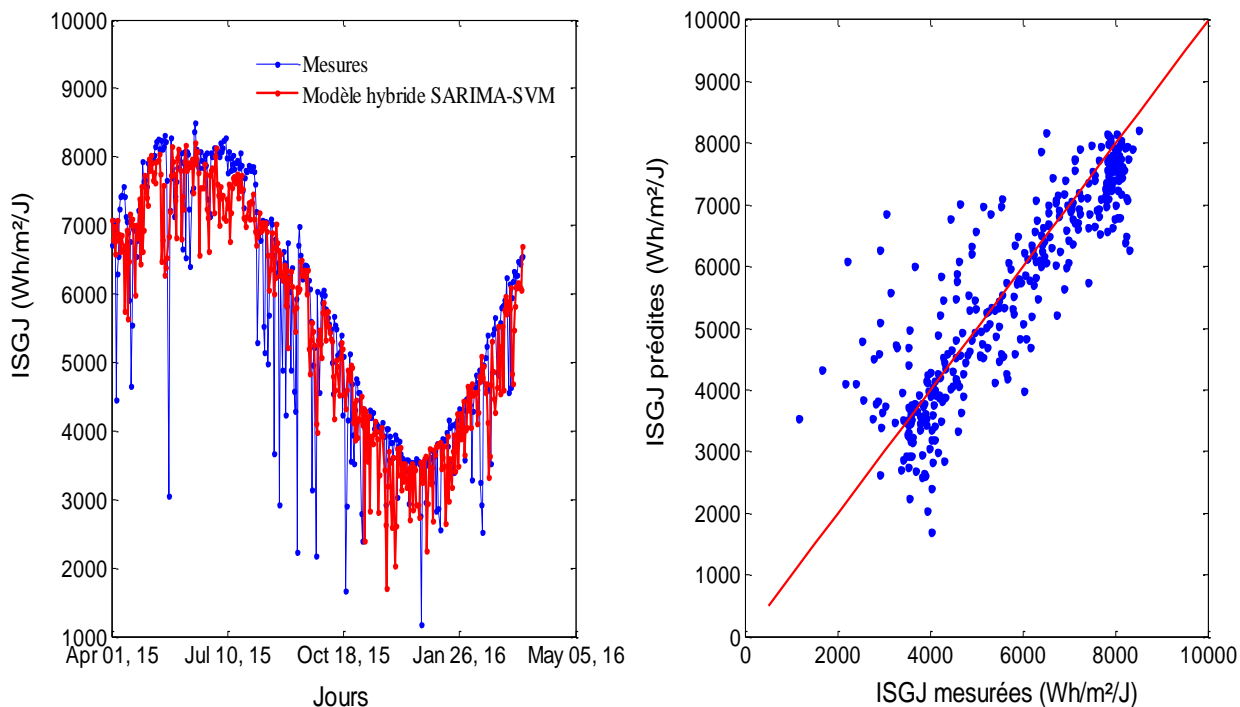


Figure 37 : ISGJ prédites par le modèle hybride SARIMA-SVM

15.

Tableau 15 : comparaison entre le modèle SARIMA et le modèle hybride

Test	RMSE (Wh/m ²)	NRMSE (%)	MAPE (%)	R
SARIMA	889.768	14.913	14.139	0.867
Hybride	866.814	14.529	13.820	0.874

Le Tableau 15 affiche les résultats de tests statistiques obtenus à la prévision de l'ISGJ par chacun des modèles SARIMA et Hybride.

On remarque qu'il y a une légère amélioration des résultats de la prédiction de l'ISGJ par le modèle Hybride SARIMA-SVM : le MAPE est de 13.820% pour le modèle Hybride alors qu'il était de 14.139% pour le modèle SARIMA, le NRMSE passe de 14.913% à 14,529% et le coefficient de corrélation a atteint 87.4% avec le modèle Hybride alors qu'il était de 86.7% concernant le modèle SARIMA.

On peut dire qu'en combinant les SARIMA avec les SVM on contribue à l'amélioration de la prédiction de la série temporelle de l'irradiation solaire globale.

4. Conclusion partielle

Le principe de la prédiction des séries temporelles a été abordé dans ce chapitre.

Le modèle conventionnelle basé sur la méthodologie de Box & Jenkins a été développé et a permis la prédiction de la série temporelle de l'ISGJ à partir de celle de l'indice de clarté in situ. Les performances sont telles que le coefficient de corrélation est de 0.867 et le NRMSE est de 14.913%.

Ces performances ont pu être améliorées légèrement par l'introduction des SVM dans la partie non linéaire du problème. Par conséquent le NRMSE s'améliore de 0.38% et le coefficient de corrélation est de 87.4%.

Certes, le modèle hybride SARIMA_SVM proposé a apporté une légère amélioration à la prédiction de l'ISGJ in situ, mais il faut noter que ce phénomène étudié est très aléatoire, alors, même si le modèle améliore légèrement les performances prédictifs, ça reste toujours bénéfique.

CONCLUSION GENERALE

CONCLUSION GENERALE

Dans cette thèse, deux différentes approches sont utilisées pour la prédiction de l'irradiation solaire globale (ISG) dans un site caractérisé par un climat aride (Ghardaïa) et favorable aux installations solaires.

La première approche consiste à développer des modèles prédictifs de l'ISGJ (irradiation solaire globale journalière) et de l'ISGM (irradiation solaire globale moyenne journalière mensuelle) à un pas d'avance (Jour et mois respectivement) via une technique d'apprentissage statistique nouvellement appliquée dans le domaine de la météorologie qui est la technique SVM (Support Vector Machine).

Ces modèles SVM élaborés utilisent un seul paramètre d'entrée mesurable qui est la température (T_{\max} , T_{\min} , T_{moy} , $(T_{\max} - T_{\min})$), avec deux autres paramètres calculables à partir de formules théoriques: la durée d'ensoleillement maximale S_0 et l'irradiation solaire globale hors atmosphère (extraterrestre) H_0 .

Le choix de la température est soutenu par le fait qu'elle est en bonne corrélation avec le rayonnement solaire ($p < 0,001$), aussi par la simplicité de sa collecte, sa disponibilité ainsi que le faible coût des dispositifs utilisés pour sa mesure.

Les résultats de la prédiction journalière (ISGJ) ont montré une amélioration significative des performances par l'introduction de DH_0 ou DS_0 au côté des températures mesurées en entrée. Les tests statistiques indiquent que les valeurs du NRMSE varient de 13.163% à 13,305%, le MAPE de 10.403% à 10,503% et le coefficient de corrélation R varie de 0,894 à 0,896 ($p < 0,001$).

Pour les modèles élaborés pour la prévision de l'irradiation solaire globale moyenne journalière mensuelle (ISGM), les résultats ont montré clairement qu'avec seulement la température minimale (MT_{\min}) et l'irradiation solaire hors atmosphère (MH_0) comme entrées, une très bonne prédiction peut être atteinte avec un NRMSE = 7.442%, MAPE = 8.940% et un coefficient de corrélation R proche de 0,986 ($p < 0,001$).

L'efficacité de tous ces modèles SVM a été prouvée par des études comparatives avec les modèles de RNA (PMC) développés in situ ainsi que ceux déjà publiés dans la littérature.

La deuxième approche consiste à appliquer un modèle hybride en associant deux techniques complètement différentes, l'une basée sur le principe des séries temporelles (SARIMA) et l'autre sur l'apprentissage statistique (SVM) afin de prédire à un pas de temps d'une journée l'irradiation solaire globale journalière ISGJ in situ.

Ainsi, le modèle hybride (SARIMA-SVM) a permis d'améliorer légèrement les performances de la prédiction par rapport au modèle conventionnelle qui est basé sur la méthodologie de Box & Jenkins (SARIMA). Le NRMSE s'améliore de 0.38% et le coefficient de corrélation atteint 87.4%.

Bien que le modèle hybride SARIMA_SVM proposé a apporté une légère amélioration à la prédiction de l'ISGJ in situ, mais ça reste toujours bénéfique vu que ce phénomène étudié est très aléatoire et la série d'observations est considérée sans aucune modification. Cependant, cette technique hybride pourrait être une solution pour d'autres phénomènes plus ou moins aléatoires.

Perspectives

Comme perspectives, nous essayerons d'adapter ces modèles SVM développés à d'autres sites avoisinant tout en utilisant les mêmes entrées et en s'inspirant des travaux menés récemment par [110], où ils prouva que le rayonnement solaire sur un site (site d'estimation) pourrait bien être estimée par un modèle SVM développé sur un autre site (site source), à condition qu'il y ait des conditions climatiques et topographiques semblables.

REFERENCES

- [1] <http://www.energy.gov.dz>.
- [2] S. M. Al-Alawi and H. A. Al-Hinai, "An ANN-based approach for predicting global radiation in locations with no direct measurement instrumentation," *Renewable Energy*, vol. 14, pp. 199-204, 1998.
- [3] S. R. M. Mohandes, T. O. Halawani, " Estimation of global solar radiation using artificial neural networks," *Renewable Energy*, vol. Vol. 14, pp. pp. 179-184, 1998.
- [4] A. Sözen, E. Arcaklıoğlu, M. Özalp, and N. Çağlar, "Forecasting based on neural network approach of solar potential in Turkey," *Renewable Energy*, vol. 30, pp. 1075-1090, 2005.
- [5] J. L. Bosch, G. López, and F. J. Batlles, "Daily solar irradiation estimation over a mountainous area using artificial neural networks," *Renewable Energy*, vol. 33, pp. 1622-1628, 2008.
- [6] J.-L. Chen, H.-B. Liu, W. Wu, and D.-T. Xie, "Estimation of monthly solar radiation from measured temperatures using support vector machines – A case study," *Renewable Energy*, vol. 36, pp. 413-420, 2011.
- [7] S. Belaid and A. Mellit, "Prediction of daily and mean monthly global solar radiation using support vector machine in an arid climate," *Energy Conversion and Management*, vol. 118, pp. 105-118, 2016.
- [8] C. Voyant, M. Muselli, C. Paoli, and M.-L. Nivet, "Hybrid methodology for hourly global radiation forecasting in Mediterranean area," *Renewable Energy*, vol. 53, pp. 1-11, 2013.
- [9] T. Muneer, *Solar Radiation and Dayligh*: Elsevier Ltd, 2004.
- [10] G. S. Garg HP, "Prediction of global solar radiation from bright sunshine hours and other meteorological data," *Energy Conversion Management* 1983.
- [11] c. voyant, "Prédiction de séries temporelles de rayonnement solaire global et de production d'énergie photovoltaïque à partir de réseaux de neurones artificiels," ECOLE DOCTORALE ENVIRONNEMENT ET SOCIETE, UNIVERSITE DE CORSE-PASCAL PAOLI, 2011.
- [12] S. Belaid, "Modelisation des données de l'irradiation solaire cas du site Ghardaia," Magister, Faculté de la Technologie Département d'Electronique, Université Abderrahmane Mira de Bejaia, 2010.
- [13] G. N. K. Dahmani , R. Dizène et C. Paoli, "Etat de l'art sur les réseaux de neurones artificiels appliqués à l'estimation du rayonnement solaire," *Revue des Energies Renouvelables* vol. Vol. 15 N°4, 2012.
- [14] J. A. Davies and D. C. McKay, "Estimating solar irradiance and components," *Solar Energy*, vol. 29, pp. 55-64, 1982.
- [15] C. Gueymard, "Critical analysis and performance assessment of clear sky solar irradiance models using theoretical and measured data," *Solar Energy*, vol. 51, pp. 121-138, 1993.
- [16] M. Iqbal, "An Introduction to Solar Radiation," *Academic Press; Canada*, 1983.
- [17] C. A. Gueymard, "Clear-sky irradiance predictions for solar resource mapping and large-scale applications: Improved validation methodology and detailed performance analysis of 18 broadband radiative models," *Solar Energy*, vol. 86, pp. 2145-2169, 2012.
- [18] R. F. M. A. B. Brahim, "A global solar radiation model for the design of solar energy systems," *Asian journal of scientific research*, 2008.
- [19] V. Badescu, *Modeling Solar Radiation at the Earth's Surface* Springer, 2008.
- [20] R. J. Hontoria L, Zufiria P, Aguilera J, "Improved generation of hourly solar irradiation artificial series using neural networks," in *Engineering Applications of Neural Networks, EANN99 Conference*, Warsaw, Poland, 1999, pp. 87-92.
- [21] C. A. Sfetos A, "Univariate and multivariate forecasting of hourly solar radiation with artificial intelligence techniques," *Sol Energy*, 2000.
- [22] J. J. Dorvlo ASS, Al-Lawati A, "Development of a neural network to predict daily solar radiation," *Agr Forest Meteorol*, 2002.

Annexes

- [23] A. F. Elminir HK, Elsayed TS, "Estimation of solar radiation components incident on Helwan site using neural networks," *Sol Energy*, 2005.
- [24] H. G. Elizondo D, McClendon R, "Development of a neural network to predict daily solar radiation," *Agr Forest Meteorol*, 1994.
- [25] O. S. Kemmoku Y, Nakagawa S, Sakakibara T, "Daily insolation forecasting using a multistage neural network," *Sol Energy*, 1999.
- [26] B. Mellit A, Kalogirou SA "An adaptive wavelet-network model for forecasting daily total solar radiation," *Appl Energ*, 2006.
- [27] R. S. Mohandes AM, Halawani TO, "Estimation of global solar radiation using artificial neural networks," *Renew Energ*, 1998.
- [28] B. A. Mohandes M, Kassas M, Rehman S, Halawani TO, "Use of radial basis functions for estimating monthly mean daily solar radiation," *Sol Energy*, 2000.
- [29] A. Mellit, S. A. Kalogirou, L. Hontoria, and S. Shaari, "Artificial intelligence techniques for sizing photovoltaic systems: A review," *Renewable and Sustainable Energy Reviews*, vol. 13, pp. 406-419, 2009.
- [30] A. Mellit and A. M. Pavan, "A 24-h forecast of solar irradiance using artificial neural network: Application for performance prediction of a grid-connected PV plant at Trieste, Italy," *Solar Energy*, vol. 84, pp. 807-821, 2010.
- [31] M. Bilgili and M. Ozgoren, "Daily total global solar radiation modeling from several meteorological data," *Meteorology and Atmospheric Physics*, vol. 112, pp. 125-138, 2011.
- [32] J. Cao and X. Lin, "Study of hourly and daily solar irradiation forecast using diagonal recurrent wavelet neural networks," *Energy Conversion and Management*, vol. 49, pp. 1396-1406, 2008.
- [33] M. Chaabene and M. Ben Ammar, "Neuro-fuzzy dynamic model with Kalman filter to forecast irradiance and temperature for solar energy systems," *Renewable Energy*, vol. 33, pp. 1435-1443, 2008.
- [34] A. Mellit, S. A. Kalogirou, S. Shaari, H. Salhi, and A. Hadj Arab, "Methodology for predicting sequences of mean monthly clearness index and daily solar radiation data in remote areas: Application for sizing a stand-alone PV system," *Renewable Energy*, vol. 33, pp. 1570-1590, 2008.
- [35] J. Mubiru and E. J. K. B. Banda, "Estimation of monthly average daily global solar irradiation using artificial neural networks," *Solar Energy*, vol. 82, pp. 181-187, 2008.
- [36] S. Rehman and M. Mohandes, "Artificial neural network estimation of global solar radiation using air temperature and relative humidity," *Energy Policy*, vol. 36, pp. 571-576, 2008.
- [37] P. L. Zervas, H. Sarimveis, J. A. Palyvos, and N. C. G. Markatos, "Prediction of daily global solar irradiance on horizontal surfaces based on neural-network techniques," *Renewable Energy*, vol. 33, pp. 1796-1803, 2008.
- [38] M. Benganem, A. Mellit, and S. N. Alamri, "ANN-based modelling and estimation of daily global solar radiation data: A case study," *Energy Conversion and Management*, vol. 50, pp. 1644-1655, 2009.
- [39] A. Moghaddamnia, R. Remesan, M. H. Kashani, M. Mohammadi, D. Han, and J. Piri, "Comparison of LLR, MLP, Elman, NNARX and ANFIS Models—with a case study in solar radiation estimation," *Journal of Atmospheric and Solar-Terrestrial Physics*, vol. 71, pp. 975-982, 2009.
- [40] M. A. Behrang, E. Assareh, A. Ghanbarzadeh, and A. R. Noghrehabadi, "The potential of different artificial neural network (ANN) techniques in daily global solar radiation modeling based on meteorological data," *Solar Energy*, vol. 84, pp. 1468-1480, 2010.
- [41] A. Rahimikhoob, "Estimating global solar radiation using artificial neural network and air temperature data in a semi-arid environment," *Renewable Energy*, vol. 35, pp. 2131-2135, 2010.

- [42] C. Voyant, M. Muselli, C. Paoli, and M.-L. Nivet, "Optimization of an artificial neural network dedicated to the multivariate forecasting of daily global radiation," *Energy*, vol. 36, pp. 348-359, 2011.
- [43] R. Yacef, M. Benghanem, and A. Mellit, "Prediction of daily global solar irradiation data using Bayesian neural network: A comparative study," *Renewable Energy*, vol. 48, pp. 146-154, 2012.
- [44] S. X. Chen, H. B. Gooi, and M. Q. Wang, "Solar radiation forecast based on fuzzy logic and neural networks," *Renewable Energy*, vol. 60, pp. 195-201, 2013.
- [45] S. A. Kalogirou, E. Mathioulakis, and V. Belessiotis, "Artificial neural networks for the performance prediction of large solar systems," *Renewable Energy*, vol. 63, pp. 90-97, 2014.
- [46] J. Wu, C. K. Chan, Y. Zhang, B. Y. Xiong, and Q. H. Zhang, "Prediction of solar radiation with genetic approach combing multi-model framework," *Renewable Energy*, vol. 66, pp. 132-139, 2014.
- [47] H. Citakoglu, "Comparison of artificial intelligence techniques via empirical equations for prediction of solar radiation," *Computers and Electronics in Agriculture*, vol. 118, pp. 28-37, 2015.
- [48] L. Mazonra Aguiar, B. Pereira, M. David, F. Díaz, and P. Lauret, "Use of satellite data to improve solar radiation forecasting with Bayesian Artificial Neural Networks," *Solar Energy*, vol. 122, pp. 1309-1324, 2015.
- [49] L. Zou, L. Wang, A. Lin, H. Zhu, Y. Peng, and Z. Zhao, "Estimation of global solar radiation using an artificial neural network based on an interpolation technique in southeast China," *Journal of Atmospheric and Solar-Terrestrial Physics*, vol. 146, pp. 110-122, 2016.
- [50] Y. Wu and J. Wang, "A novel hybrid model based on artificial neural networks for solar radiation prediction," *Renewable Energy*, vol. 89, pp. 268-284, 2016.
- [51] L. Wang, O. Kisi, M. Zounemat-Kermani, G. A. Salazar, Z. Zhu, and W. Gong, "Solar radiation prediction using different techniques: model evaluation and comparison," *Renewable and Sustainable Energy Reviews*, vol. 61, pp. 384-397, 2016.
- [52] N. Premalatha and A. Valan Arasu, "Prediction of solar radiation for solar systems by using ANN models with different back propagation algorithms," *Journal of Applied Research and Technology*, vol. 14, pp. 206-214, 2016.
- [53] J.-L. Chen, G.-S. Li, and S.-J. Wu, "Assessing the potential of support vector machine for estimating daily solar radiation using sunshine duration," *Energy Conversion and Management*, vol. 75, pp. 311-318, 2013.
- [54] J. Zeng and W. Qiao, "Short-term solar power prediction using a support vector machine," *Renewable Energy*, vol. 52, pp. 118-127, 2013.
- [55] B. B. Ekici, "A least squares support vector machine model for prediction of the next day solar insolation for effective use of PV systems," *Measurement*, vol. 50, pp. 255-262, 2014.
- [56] K. Mohammadi, S. Shamshirband, C. W. Tong, M. Arif, D. Petković, and S. Ch, "A new hybrid support vector machine-wavelet transform approach for estimation of horizontal global solar radiation," *Energy Conversion and Management*, vol. 92, pp. 162-171, 2015.
- [57] K. Mohammadi, S. Shamshirband, M. H. Anisi, K. A. Alam, and D. Petković, "Support vector regression based prediction of global solar radiation on a horizontal surface," *Energy Conversion and Management*, vol. 91, pp. 433-441, 2015.
- [58] S. Shamshirband, K. Mohammadi, P. L. Yee, D. Petković, and A. Mostafaeipour, "A comparative evaluation for identifying the suitability of extreme learning machine to predict horizontal global solar radiation," *Renewable and Sustainable Energy Reviews*, vol. 52, pp. 1031-1042, 2015.
- [59] R. Urraca, J. Antonanzas, M. Alia-Martinez, F. J. Martinez-de-Pison, and F. Antonanzas-Torres, "Smart baseline models for solar irradiation forecasting," *Energy Conversion and Management*, vol. 108, pp. 539-548, 2016.

- [60] Arthur Charpentier, "Cours de séries temporelles théorie et applications," ed. www.math.univ-montp2.fr.
- [61] T. N. Goh and K. J. Tan, "Stochastic modeling and forecasting of solar radiation data," *Solar Energy*, vol. 19, pp. 755-757, 1977.
- [62] L. Mora-López and M. Sidrach-de-Cardona, "Multiplicative ARMA models to generate hourly series of global irradiation," *Solar Energy*, vol. 63, pp. 283-291, 1998.
- [63] C. Craggs, E. Conway, and N. M. Pearsall, "Stochastic modelling of solar irradiance on horizontal and vertical planes at a northerly location," *Renewable Energy*, vol. 18, pp. 445-463, 1999.
- [64] G. Reikard, "Predicting solar radiation at high resolutions: A comparison of time series forecasts," *Solar Energy*, vol. 83, pp. 342-349, 2009.
- [65] C. Voyant, M. Muselli, C. Paoli, and M.-L. Nivet, "Numerical weather prediction (NWP) and hybrid ARMA/ANN model to predict global radiation," *Energy*, vol. 39, pp. 341-355, 2012.
- [66] C. Voyant, C. Darras, M. Muselli, C. Paoli, M.-L. Nivet, and P. Poggi, "Bayesian rules and stochastic models for high accuracy prediction of solar radiation," *Applied Energy*, vol. 114, pp. 218-226, 2014.
- [67] H. Sun, D. Yan, N. Zhao, and J. Zhou, "Empirical investigation on modeling solar radiation series with ARMA-GARCH models," *Energy Conversion and Management*, vol. 92, pp. 385-395, 2015.
- [68] A. Linares-Rodríguez, J. A. Ruiz-Arias, D. Pozo-Vázquez, and J. Tovar-Pescador, "Generation of synthetic daily global solar radiation data based on ERA-Interim reanalysis and artificial neural networks," *Energy*, vol. 36, pp. 5356-5365, 2011.
- [69] M. Benghanem and A. Mellit, "Radial Basis Function Network-based prediction of global solar radiation data: Application for sizing of a stand-alone photovoltaic system at Al-Madinah, Saudi Arabia," *Energy*, vol. 35, pp. 3751-3762, 2010.
- [70] K. Mohammadi, S. Shamshirband, D. Petković, and H. Khorasanizadeh, "Determining the most important variables for diffuse solar radiation prediction using adaptive neuro-fuzzy methodology; case study: City of Kerman, Iran," *Renewable and Sustainable Energy Reviews*, vol. 53, pp. 1570-1579, 2016.
- [71] S. Kalogirou, "Solarenergyengineering:processesandsystems," *1sted.Elsevier Inc*, 2009.
- [72] H. Jiang, Y. Dong, J. Wang, and Y. Li, "Intelligent optimization models based on hard-ridge penalty and RBF for forecasting global solar radiation," *Energy Conversion and Management*, vol. 95, pp. 42-58, 2015.
- [73] A. Mellit, M. Benghanem, A. H. Arab, and A. Guessoum, "A simplified model for generating sequences of global solar radiation data for isolated sites: Using artificial neural network and a library of Markov transition matrices approach," *Solar Energy*, vol. 79, pp. 469-482, 2005.
- [74] A. B. M. Mohandes, M. Kassas, S. Rehman and T. O. Halawani, "Use of radial basis functions for estimating monthly mean daily solar radiation," *Solar Energy*, vol. Vol. 68, pp. pp. 161-168, 2000.
- [75] F. S. Tymvios, C. P. Jacovides, S. C. Michaelides, and C. Scouteli, "Comparative study of Ångström's and artificial neural networks' methodologies in estimating global solar radiation," *Solar Energy*, vol. 78, pp. 752-762, 2005.
- [76] A. Koca, H. F. Oztog, Y. Varol, and G. O. Koca, "Estimation of solar radiation using artificial neural networks with different input parameters for Mediterranean region of Anatolia in Turkey," *Expert Systems with Applications*, vol. 38, pp. 8756-8762, 2011.
- [77] M. Ozgoren, M. Bilgili, and B. Sahin, "Estimation of global solar radiation using ANN over Turkey," *Expert Systems with Applications*, vol. 39, pp. 5043-5051, 2012.
- [78] E. S. Mostafavi, S. S. Ramiyani, R. Sarvar, H. I. Moud, and S. M. Mousavi, "A hybrid computational approach to estimate solar global radiation: An empirical evidence from Iran," *Energy*, vol. 49, pp. 204-210, 2013.

Annexes

- [79] S. Bhardwaj, V. Sharma, S. Srivastava, O. S. Sastry, B. Bandyopadhyay, S. S. Chandel, and J. R. P. Gupta, "Estimation of solar radiation using a combination of Hidden Markov Model and generalized Fuzzy model," *Solar Energy*, vol. 93, pp. 43-54, 2013.
- [80] Z. Ramedani, M. Omid, and A. Keyhani, "Modeling Solar Energy Potential in a Tehran Province Using Artificial Neural Networks," *International Journal of Green Energy*, vol. 10, pp. 427-441, 2013/04/21 2013.
- [81] A. Will, J. Bustos, M. Bocco, J. Gotay, and C. Lamelas, "On the use of niching genetic algorithms for variable selection in solar radiation estimation," *Renewable Energy*, vol. 50, pp. 168-176, 2013.
- [82] A. K. Yadav, H. Malik, and S. S. Chandel, "Selection of most relevant input parameters using WEKA for artificial neural network based solar radiation prediction models," *Renewable and Sustainable Energy Reviews*, vol. 31, pp. 509-519, 2014.
- [83] A. K. Yadav and S. S. Chandel, "Solar energy potential assessment of western Himalayan Indian state of Himachal Pradesh using J48 algorithm of WEKA in ANN based prediction model," *Renewable Energy*, vol. 75, pp. 675-693, 2015.
- [84] G. López, F. J. Batlles, and J. Tovar-Pescador, "Selection of input parameters to model direct solar irradiance by using artificial neural networks," *Energy*, vol. 30, pp. 1675-1684, 2005.
- [85] K. Mohammadi, S. Shamshirband, A. Kamsin, P. C. Lai, and Z. Mansor, "Identifying the most significant input parameters for predicting global solar radiation using an ANFIS selection procedure," *Renewable and Sustainable Energy Reviews*, vol. 63, pp. 423-434, 2016.
- [86] R. C. Deo, X. Wen, and F. Qi, "A wavelet-coupled support vector machine model for forecasting global incident solar radiation using limited meteorological dataset," *Applied Energy*, vol. 168, pp. 568-593, 2016.
- [87] K. Chiteka and C. C. Enweremadu, "Prediction of global horizontal solar irradiance in Zimbabwe using artificial neural networks," *Journal of Cleaner Production*, vol. 135, pp. 701-711, 2016.
- [88] E. F. Alsina, M. Bortolini, M. Gamberi, and A. Regattieri, "Artificial neural network optimisation for monthly average daily global solar radiation prediction," *Energy Conversion and Management*, vol. 120, pp. 320-329, 2016.
- [89] M. Lucea, "Modélisation dynamique par réseaux de neurones et machines à vecteurs supports : contribution à la maîtrise des émissions polluantes de véhicules automobiles," L'UNIVERSITE Pierre Marie Curie PARIS 6, 2006.
- [90] K. S. B. Schölkopf, C. Burges, F. Girosi, P. Niyogi, T. Poggio, V. Vapnik, "Comparing support vector machines with Gaussian kernels to radial basis function classifiers" *IEEE Transaction on Signal Processing*, vol. Vol. 45, pp. pp. 2758-2765, 1997.
- [91] C. J. C. B. H. Drucker, L. Kaufman, A. Smola, V. Vapnik, "Support Vector Regression Machines," *In Advances in Neural Information Processing Systems*, vol. Vol. 9, pp. pp. 155-161, 1997.
- [92] H. B. Drucker, Christopher J. C.; Kaufman, Linda; Smola, Alexander J.; and Vapnik, Vladimir N, "Support Vector Regression Machines," *in Advances in Neural Information Processing Systems 9, NIPS 1996, 155-161, MIT Press., 1997.*
- [93] <https://fr.wikipedia.org/wiki/Perceptron>.
- [94] H. B. D. M.T.Hagan, M.Beale, "Neural Network Toolbox TM6, User's Guide," *The Math Works*, 1992-2009.
- [95] D. E. Rumelhart, Hinton, G.E., Williams, J.R, "Learning internal representation by error propagation Parallel Distributed Processing," *vol. 1. MIT Press, MA, pp. 318-362, 1986.*
- [96] S. Haykin, "Neural Networks, a Comprehensive Foundation," *Macmillan, New York, 1994.*
- [97] M. Riedmiller, Braum, H, "direct adaptive method for faster Back-propagation learning: The RPROP algorithm.," *Proc. IEEE, Trans. Sys. Man. Cybe., pp. 581-591., 1993.*
- [98] G. E. P. Box Jenkins, Jenkins, G. M., & Reinsel, G. C, "Time Series Analysis: Forecasting and Control

Annexes

Control," *John Wiley and Sons, Inc*, 2008.

- [99] H. Akaike, "A Bayesian analysis of the minimum AIC procedure.," *Annals of the Institute of Statistics and Mathematics* 1978.
- [100] B. G. Ljung GM, "On a measure of lack of fit in time series models," *Biometrika*, 1978.
- [101] M. Gaultier, *Analyse, probabilités et méthode statistique*: Vuibert, 1984.
- [102] H.-x. Zhao and F. Magoulès, "A review on the prediction of building energy consumption," *Renewable and Sustainable Energy Reviews*, vol. 16, pp. 3586-3592, 2012.
- [103] A. Mellit, H. Eleuch, M. Benghanem, C. Elaoun, and A. M. Pavan, "An adaptive model for predicting of global, direct and diffuse hourly solar irradiance," *Energy Conversion and Management*, vol. 51, pp. 771-782, 2010.
- [104] V. Prema and K. U. Rao, "Development of statistical time series models for solar power prediction," *Renewable Energy*, vol. 83, pp. 100-109, 2015.
- [105] S. safi, A. Zeroual, and M. Hassani, "Prediction of global daily solar radiation using higher order statistics," *Renewable Energy*, vol. 27, 2002.
- [106] J. Hassan, "ARIMA and regression models for prediction of daily and monthly clearness index," *Renewable Energy*, vol. 68, pp. 421-427, 2014.
- [107] F. O. Hocaoglu, "Stochastic approach for daily solar radiation modeling," *Solar Energy*, vol. 85, pp. 278-287, 2011.
- [108] A. Zeroual, M. Ankrim, and A. J. Wilkinson, "Stochastic modelling of daily global solar radiation measured in Marrakesh, Morocco," *Renewable Energy*, vol. 6, pp. 787-793, 1995.
- [109] W. Ji and K. C. Chee, "Prediction of hourly solar radiation using a novel hybrid model of ARMA and TDNN," *Solar Energy*, vol. 85, pp. 808-817, 2011.
- [110] J.-L. Chen, G.-S. Li, B.-B. Xiao, Z.-F. Wen, M.-Q. Lv, C.-D. Chen, Y. Jiang, X.-X. Wang, and S.-J. Wu, "Assessing the transferability of support vector machine model for estimation of global solar radiation from air temperature," *Energy Conversion and Management*, vol. 89, pp. 318-329, 2015.