

**REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE.
MINISTRE DE L'ENSEIGNEMENT SUPERIEUR
ET DE LA RECHERCHE SCIENTIFIQUE.**



**Université Mohamed Seddik Ben Yahia de Jijel
Faculté des sciences et de la technologies
Département d'électronique**



PROJET DE FIN D'ETUDE

***En vue de l'obtention du Diplôme de MASTER ACADEMIQUE
Filière : Electronique***

OPTION : Electronique et Systèmes Embarqués

Thème

Prévision de l'irradiation solaire globale à l'aide
des méthodes d'apprentissage automatique

Présenté Par : Djebablia Hamza

Soutenu le 26/06/2023. Devant les Jurys composés de :

**Président : Bouaziz Fatiha
Examineur : Soukkou Ammar
Encadreur : Yacef Rima**

**MCA Univ JIJEL
MCA Univ JIJEL
MCB Univ JIJEL**

Promotion : 2022 /2023

Remerciement

Je tiens tout d'abord à remercier la personne qui a bien voulu se laisser convaincre par ma motivation et qui m'a permis de réaliser ce mémoire **Dr Yacef Rima**, qui m'a beaucoup aidé durant cette période, je tiens à lui exprimer, ici, ma reconnaissance et ma gratitude pour son aide, son support, et sa disponibilité.

Je remercie davantage et chaleureusement les membres du jury pour l'honneur qu'ils ont acceptée d'évaluer ce travail **Dr Soukkou Ammar, Dr Bouaziz Fatiha** qui ont acceptée de juger ce travail et d'y apporter leur caution. Je vous remercie vivement pour l'honneur que vous me faites en faisant partie du jury.

Enfin je remercie toute personne ayant contribué, de près ou de loin, à la réalisation de ce travail.

SOMMAIRE

Remerciements

Sommaire

Liste des figures

Liste des tableaux

Nomenclatures

Glossaire des acronymes

Introduction générale

Chapitre I: Contexte général et problématique

I.1 Introduction	3
I.2 irradiation solaire	3
I.2.1 Définition	3
I.2.2 Aspect géométrique	3
I.2.2.1 Mouvement apparent du soleil	4
I.2.2 .2 Coordonnées géographiques	5
I.2.2.3 Coordonnées horaires	6
I.2 .3 Aspect énergétique	8
I.2.3.1 Rayonnement solaire hors atmosphère et constante solaire	8
I.2.3.2 Les différentes composantes du rayonnement solaire	9
I.3 Les appareils de mesure des différentes composantes solaire	12
I.4 Conclusion	14

Chapitre II: Méthodes de l'intelligence artificielle

II.1 Introduction	15
II.2 L'intelligence artificielle	15
II.2.1 Définition	15
II.3 Apprentissage automatique	16
II.3.1 Définition	16
II.3.2 Les étapes de l'apprentissage automatique	17
II.3.2 .1 Collecte des données	17
II.3.2.2 Prétraitement des données	17

II.3.2. 3	Formation et test du modèle	18
II.3.2.4	L'évaluation	18
II.3.3	Types d'apprentissage automatique	18
II.3.3.1	Apprentissage supervisé	18
II.3.3.2	Apprentissage non supervisé	19
II.3.3.3	Apprentissage semi - supervisé	19
II. 3.3.4	Apprentissage par renforcement	19
II.3.3.5	Apprentissage par transfert	20
II.4	Quelques méthodes de l'apprentissage automatique	20
II.4.1	Les arbres de régression	20
II.4.1.1	Les arbres de régression simples	20
II.4.1.2	Les arbres de régression renforcés « boosting »	20
II.4.1.3	Les arbres de régression ensachés « bagging »	22
II.4.2	Régression linéaire	23
II.4.3	Machines à vecteurs supports pour la régression (MVS)	23
II.5	Conclusion	24
 Chapitre III: Résultats et discussion		
III.1	Introduction	26
III.2	Etat de l'art de la prévision de l'irradiation solaire globale	26
III.3	Les données expérimentales : Données de mesures, rayonnement et variables météorologique	29
III.4	Evaluation des modèles	30
III.5	Conclusion	32

Chapitre IV: Résultat et discussion

IV.1 Introduction	32
IV.2 Modélisation et prédiction de l'irradiation solaire journalière	32
IV.3 Conclusion	44
Références	48

Listes des figures

Figure I.1	Système Terre-Soleil	4
Figure I.2	Mouvement de la Terre au tour du soleil	5
Figure I.3	Longitude et latitude d'un lieu	6
Figure I.4	Déclinaison du soleil au cour d'une année	7
Figure I.5	Coordonnés horaires	8
Figure I.6	Composantes du rayonnement solaire global	11
Figure I.7	Héliographe Campbell-Stokes	13
Figure I.8	Le pyradiomètre : mesure du rayonnement solaire global et rayonnement diffus	13
Figure I.9	Pyrhéliomètre : mesure du rayonnement solaire directe	14
Figure II.1	Processus de l'apprentissage automatique	16
Figure II.2	L'hyperplan H optimale, vecteurs de supports et marge maximale	23
Figure IV.1	Distribution de la moyenne mensuelle effective de la durée d'insolation Pour le site d'AL-Médina	33
Figure IV.2	L'irradiation solaire globale mesurée et estimée pour la phase de test pour les modèles 1,2 ,3 de la régression linéaire	36
Figure IV.3	L'irradiation solaire globale mesurée et estimée pour la phase de test pour le modèle 4 de la régression linéaire	37
Figure IV.4	L'irradiation solaire globale mesurée et estimée pour la phase de test pour les modèles1, 2,3 de la régression par l'arbre de décision	38
Figure IV.5	L'irradiation solaire globale mesurée et estimée pour la phase de test pour la régression par l'arbre de décision Type « boosting - modèle 4»	39
Figure IV.6	L'irradiation solaire globale mesurée et estimée pour la phase de test pour la régression par l'arbre de décision Type « a bagged - modèle 5»	40
Figure IV.7	L'irradiation solaire globale mesurée et estimée pour la phase de test pour la régression par le machine a vecteur de support «modèle 1»	41
Figure IV.8	L'irradiation solaire globale mesurée et estimée pour la phase de test pour la régression par le machine a vecteur de support «modèle 2»	42
Figure IV.9	L'irradiation solaire globale mesurée et estimée pour la phase de test pour la régression par le machine a vecteur de support «modèle 3»	43

Liste des tableaux

Tableau III.1	Résumé des applications de l'intelligence artificielle pour l'estimation de l'irradiation solaire	27
Tableau III.2	Données météorologiques disponibles	30
Tableau IV.2	Indicateurs statistiques de précision pour les trois familles de modèles développés	

Nomenclature

ϕ	La latitude
λ	La longitude
δ	La déclinaison
ω	Angle horaire du soleil
TSV	Temps Solaire Vrai
TU	Temps Universel
TL	Temps Légal
Δt	Correction de l'équation du temps
ω_c	L'angle horaire du soleil à son couché
ω_l	L'angle horaire du soleil à son levé
Φ_z	L'angle zénithal
Dd	La Durée diurne
I_0	Constante solaire
h	Hauteur solaire
r	La distance Terre-Soleil
\bar{r}	La distance moyenne du Soleil
Y	Le nombre total de jour de l'année.
j	Le numéro du jour a compter du 1er janvier.
K_t	L'indice de clarté
G_0	L'irradiation solaire extraterrestre
G	L'irradiation solaire globale sur une surface horizontale
S	la durée d'ensoleillement

T	la température
RH	Humidité relative
G_{im}	L'irradiation solaire mesurée
G_{ie}	L'irradiation solaire estimée
R	Le coefficient de corrélation

Glossaire des acronymes

RMSE	Root Mean Square Error
MAE	Mean Absolute Error
MBE	Mean Bias Error
MPE	Mean Pourcentage Error
CART	Classification and Regression Tree
SVM	Machines à vecteurs supports pour la régression
SVR	Support Vector Regression
ANN	Artificial Neural Network
MAPE	Mean Absolute Percentage Error
ENS	Ensemble Learning
MSE	Mean Squared Error
FL	Feature Learning
ELM	Extreme Learning Machine
KRR	Kernel Ridge Regression
LASSO	Least Absolute Shrinkage and Selection Operator
RR	Ridge Regression
ARMA	Autoregressive Moving Average
RF	Random forest
MLR	Multiple Linear Regression
K-NN	k-Nearest Neighbor

L'irradiation solaire joue un rôle essentiel dans de nombreuses applications, telles que la production d'énergie solaire, la planification urbaine et l'agriculture de précision. Cependant, la prédiction précise de l'irradiation solaire est un défi complexe en raison de la variabilité des conditions météorologiques et des caractéristiques géographiques.

Dans ce mémoire, nous explorons l'utilisation des techniques d'apprentissage automatique, telles que la régression linéaire, les arbres de décision et les machines à vecteurs de support, pour modéliser et prédire l'irradiation solaire. Nous avons utilisé des données météorologiques exogènes pour la prédiction de l'irradiation solaire globale journalière [46].

la performance des modèles de prédiction de l'irradiation solaire. Ces métriques permettent de comparer les résultats des différentes techniques d'apprentissage automatique et d'identifier les approches les plus performantes.

Les résultats montrent que les méthodes d'apprentissage automatique peuvent fournir des prévisions précises de l'irradiation solaire, en prenant en compte les variations saisonnières, les conditions météorologiques et les caractéristiques géographiques spécifiques. Ces prévisions précises de l'irradiation solaire peuvent aider les professionnels de l'énergie solaire à optimiser la production d'énergie, les urbanistes à concevoir des bâtiments plus durables et les agriculteurs à améliorer leurs pratiques agricoles.

Dans le domaine de la prédiction de l'irradiation solaire en démontrant l'efficacité des méthodes d'apprentissage automatique. Il ouvre la voie à de futures recherches visant à améliorer les modèles de prédiction et à les intégrer dans des systèmes de gestion de l'énergie solaire plus larges.

Des études récentes à base des techniques de l'apprentissage automatiques portent sur l'estimation de l'radiation solaire [41] [32] [36] ont été élaborés au cours de ce mémoire .

Le mémoire présent comporte quatre chapitres :

Le premier chapitre de ce mémoire présente le contexte général de l'utilisation de l'énergie solaire et la problématique de la prédiction de l'irradiation solaire. Nous donnerons aperçut global sur quelques définitions générales des paramètres astronomiques et géographiques qui interviennent d'une manière direct dans le processus de prédiction de

l'irradiation solaire, ainsi on jette une lumière sur quelques notions préliminaires de l'irradiation solaire.

Dans le second chapitre, on se concentre sur les méthodes de l'intelligence artificielle utilisées dans la prédiction de l'irradiation solaire. Nous explorons les techniques telles que la régression linéaire, les machines à vecteurs supports, les arbres de décision, qui ont été appliquées avec succès dans d'autres domaines et qui présentent un potentiel prometteur dans la prédiction de l'irradiation solaire.

Ensuite, nous présenterons, dans le chapitre 3, les différentes techniques intelligentes existantes dans la littérature et destinées à l'estimation de l'irradiation solaire globale. Nous présenterons aussi dans ce chapitre les principaux paramètres d'erreur permettant de comparer les différentes méthodologies de prédiction. Sur la base des différents paramètres d'erreurs, une comparaison bibliographique des méthodes de l'apprentissage automatique d'estimations relatives à l'irradiation solaire globale sera également proposée.

Nous énoncerons dans le quatrième chapitre les résultats d'estimation obtenus pour les différents modèles développés. Nous analysons également les erreurs et les limitations de nos modèles. Tous les résultats liés aux estimateurs énoncés dans le chapitre II seront exposés pour l'horizon journalier.

Chapitre I

Contexte général

I.1 Introduction

L'énergie solaire est une source d'énergie renouvelable importante qui est utilisée pour produire de l'électricité, de la chaleur et d'autres formes d'énergie. Cette énergie provient du rayonnement solaire, qui est émis par le soleil sous forme d'ondes électromagnétiques. Ce rayonnement est composé de différentes longueurs d'ondes, qui sont classées en fonction de leur énergie et de leur longueur.

Le rayonnement solaire est essentiel pour la vie sur Terre, car il permet aux plantes de photo synthétiser et de produire de l'oxygène, ainsi qu'à de nombreux organismes de se réchauffer et de réguler leur température corporelle. Il est également utilisé pour produire de l'énergie à grande échelle, grâce à des technologies telles que les panneaux solaires photovoltaïques et les centrales solaires thermiques.

Dans ce chapitre, nous allons présenter le rayonnement solaire, sa composition et les aspects géométriques et énergétiques. Nous aborderons également les appareils de mesure des différentes composantes solaires.

I.2 L'irradiation solaire

I.2.1 Définition

L'irradiation solaire exprime la quantité d'énergie reçue, au niveau du sol, par unité de surface. Ses unités usuelles sont : J/m^2 , J/cm^2 , Wh/m^2 , ou bien kWh/m^2 . Elle dépend de nombreux facteurs, dont principalement la couverture nuageuse, la durée du jour, l'instant considéré dans la journée, l'orientation et l'inclinaison de la surface, la latitude du lieu, son degré de pollution et la hauteur angulaire du soleil au-dessus de l'horizon. La combinaison de tous ces paramètres produit la variabilité dans l'espace et dans le temps de l'irradiation solaire. Dans les régions du globe comprises entre 40° nord et sud, l'énergie globale annuelle reçue par une surface horizontale est comprise entre 1400 et 2500kWh/m^2 .

I.2.2 Aspect géométrique

Le soleil est une sphère gazeuse chaude composé principalement d'Hydrogène (73.5%) et d'Hélium (24.9%). Son diamètre est de 1.39 million de kilomètres, ce qui correspond à une surface totale égale à $6.08 \cdot 10^{18} \text{ m}^2$ et il se trouve à une distance moyenne de 1.496 millions de kilomètres de la terre, connue aussi sous le nom d'unité astronomique, UA12.

Cette distance varie de $\pm 1.7\%$ au cours de l'année, entre 152 millions de kilomètres (1 ou 2 juillet) et 147 millions de kilomètres (1 ou 2 janvier). Vu depuis la terre, le soleil présente un angle solide de $6.8 \cdot 10^{-5}$ rad et une ouverture angulaire de 0.53° (Figure I.1) [17].

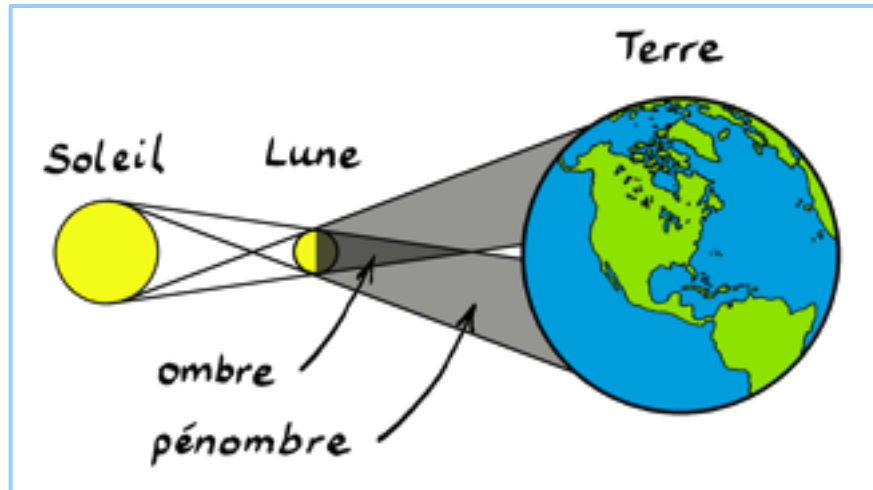


Figure I.1 Système Terre-Soleil

I.2.2.1 Mouvement apparent du soleil

Toute application solaire nécessite la connaissance du mouvement apparent du soleil pour un point donné de la surface terrestre, caractérisé par sa latitude, ϕ (positive pour l'hémisphère Nord) et sa longitude, ψ (définie par rapport au méridien de Greenwich) .

La terre est séparée par l'équateur en deux demi sphères, l'hémisphère Nord pour celle située du côté du pôle Nord, et l'hémisphère Sud pour celle qui est située du côté du pôle Sud.

D'autre part, elle est partagée d'Ouest en Est, par le méridien d'origine qui passe par Greenwich (près de Londres en Angleterre).

Pour un observateur terrestre, en considérant la déclinaison comme constante sur une journée, le soleil décrit un cercle autour de l'axe de rotation de la terre. Ce cercle est parallèle au plan de l'équateur, et sa hauteur apparente sur ce plan est donnée par la déclinaison.

La figure I.2 donne l'aspect de la trajectoire solaire diurne en zone intertropicale, en zone tempérée et en zone polaire. On remarque en particulier que:

- Au moment des équinoxes et pour toutes les latitudes le soleil se lève exactement à l'Est et se couche exactement à l'Ouest;
- Dans la zone intertropicale, la culmination solaire peut s'effectuer soit au Nord soit au Sud, selon la saison. Ceci ne manquera pas de poser des questions quant à l'orientation des modules solaires dans cette zone.

- Dans les calottes polaires, le soleil peut rester sous l'horizon pendant plusieurs jours, et six mois plus tard demeurer visible pendant plus d'un jour [17].

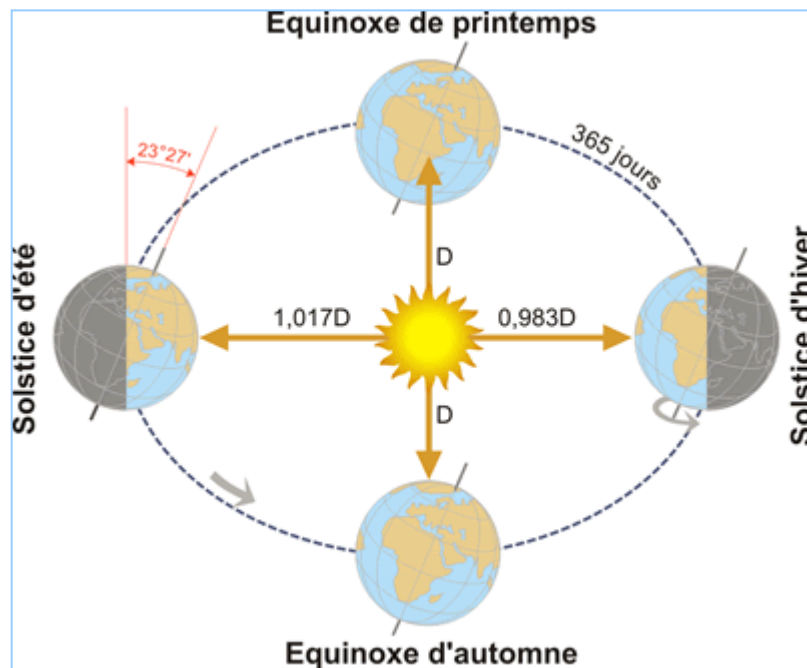


Figure I.2 Movement de la Terre au tour du soleil

I.2.2.2 Coordonnées géographiques

- **Longitude λ**

La longitude λ d'un lieu correspond à l'angle que fait le plan méridien passant par ce lieu avec un plan méridien retenu comme origine. On choisit pour méridien origine (0°) le plan passant par l'observatoire de Greenwich. Par convention on affecte du signe (+) les méridiens situés à l'est de ce méridien (-) les méridiens situés à l'ouest.

- **Latitude ϕ**

La latitude ϕ d'un lieu correspond à l'angle avec le plan équatorial que fait le rayon joignant le centre de la terre à ce lieu. L'équateur terrestre est donc caractérisé par une latitude égale à 0° , le pôle nord par la latitude $+90^\circ$ et le pôle sud par la latitude -90° .

Cette convention de signe affecte le signe (+) à tous les lieux de l'hémisphère nord et le signe (-) tous les lieux de l'hémisphère sud.

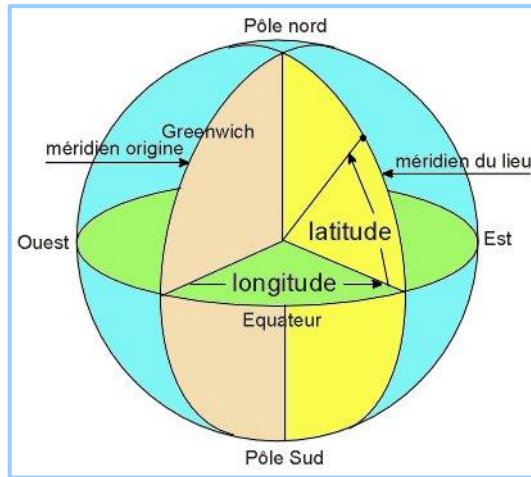


Figure I.3 Longitude et latitude d'un lieu

I.2.2.3 Coordonnées horaires

Le mouvement du Soleil est repéré par rapport au plan équatorial de la Terre à l'aide de deux angles (δ , ω) qui sont définies par la figure ci-dessous (Figure I.4).

a) La déclinaison (δ)

La déclinaison (δ) est l'angle formé par la direction du Soleil avec le plan équatorial. Elle varie de $+23^{\circ},45$ en degrés décimaux au solstice d'été (22 juin) à $-23,45^{\circ}$ au solstice d'hiver (22 décembre), en passant par la valeur 0 aux équinoxes (21 mars et 23 septembre). Cette déclinaison est due à l'inclinaison de l'axe des pôles terrestres par rapport au plan écliptique ce qui nous donne les différentes saisons (la Terre est plus proche du Soleil en hiver. La valeur de la déclinaison peut être calculée par la relation [6] :

$$\delta = 23.45 \sin \left[\frac{284 + j}{365} \times 360 \right] [\text{Degré}] \quad (\text{I.4})$$

Ou J: le numéro du jour de l'année compté à partir du 1^{er} janvier, c'est-à-dire il varie de 1 à 365.

La variation de la déclinaison du soleil au cours de l'année est représentée par la Figure I.4.

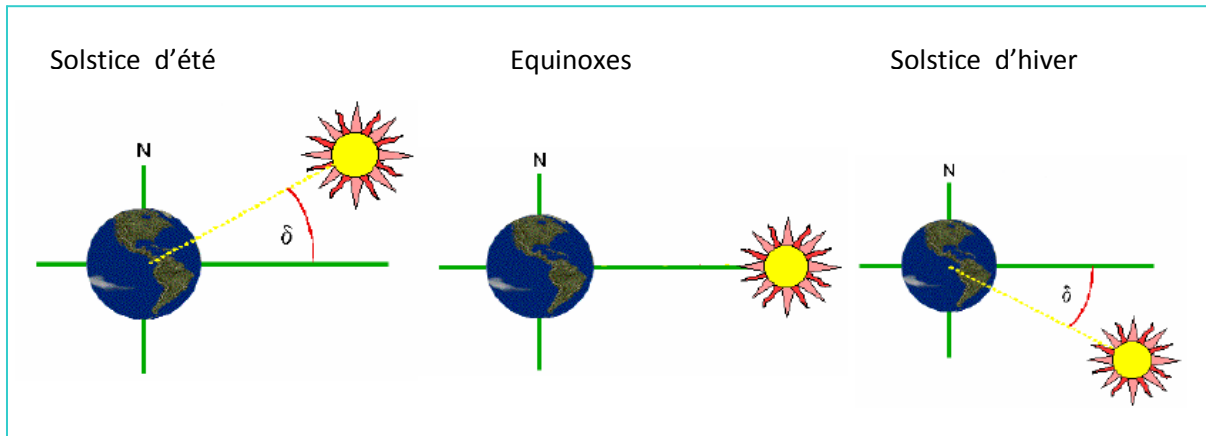


Figure I.4 Déclinaison du soleil au cours d'une année.

b) Angle horaire du soleil (ω)

C'est l'angle que font les projections de la direction du soleil avec la direction du méridien du lieu, l'angle horaire du Soleil varie à chaque instant de la journée selon la relation :

$$\omega = 15(TSV - 12) \quad (I.1)$$

avec,

$$TSV = TL - TU + \left(\frac{\Delta t + 4\lambda}{60} \right) = TL - \left(TU - \frac{\lambda}{15} \right) + \frac{\Delta t}{60} \quad (I.2)$$

$$\Delta t = 9.87 \cdot \sin 2 \left(\frac{360}{365} (j - 81) \right) - 7.35 \cdot \cos \left(\frac{360}{365} (j - 81) \right) - 1.5 \cdot \sin \left(\frac{360}{365} (j - 81) \right) \quad (I.3)$$

Ou TSV: Temps Solaire Vrai : temps repéré de façon que le soleil se trouve au zénith à midi,

TU : Temps Universel : est le décalage horaire par rapport au méridien de Greenwich,

TL: Temps Légal : temps donné par une montre,

Δt : Correction de l'équation du temps,

λ : longitude du lieu,

L'angle horaire varie de -180° à $+180^\circ$.

- $\omega > 0$ après midi
- $\omega < 0$ le matin
- $\omega = 0$ midi TSV

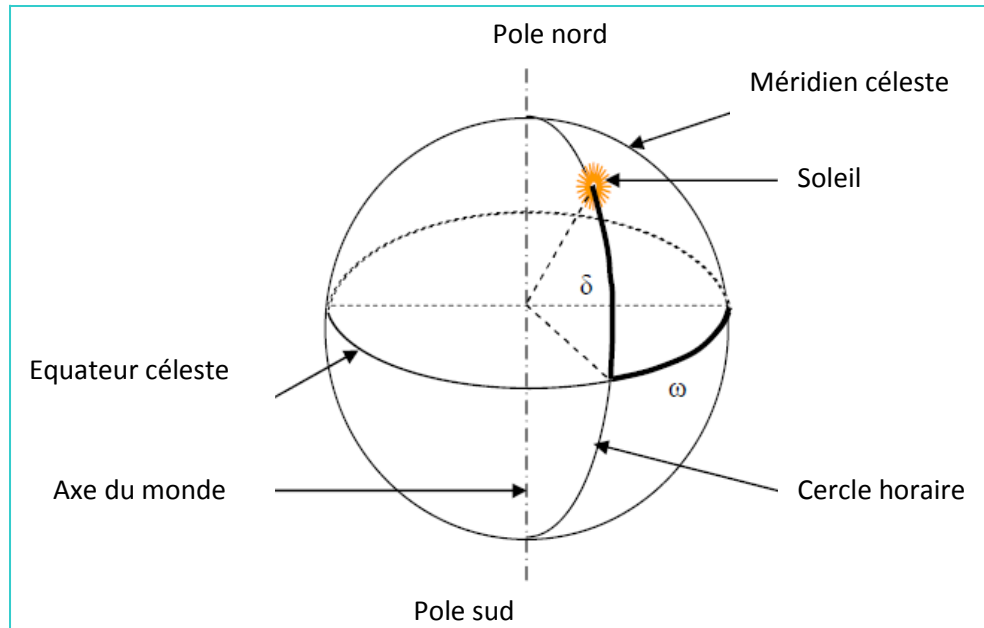


Figure I.5 Coordonnées horaires.

On peut également définir l'angle horaire du Soleil à son couché ω_c et à son levé ω_L . L'angle horaire du Soleil à son coucher ω_c est l'angle horaire solaire correspondant à l'heure où le soleil se couche, il est donné par l'équation suivante :

$$\omega_c = \arccos(-\tan(\varphi) \cdot \tan(\delta)) \quad (\text{I.4})$$

$$\omega_L = -\omega_c \quad (\text{I.5})$$

Est-on déduire la Durée diurne D_d :

$$D_d = \omega_c - \omega_L = 2\omega_c = 2 \arccos(-\tan(\varphi) \cdot \tan(\delta)) \quad (\text{I.6})$$

I.2.3 Aspect énergétique

I.2.3.1 Rayonnement solaire hors atmosphère et constante solaire

L'énergie solaire reçue par unité de temps à la limite supérieure de l'atmosphère, sur une surface unité perpendiculaire aux rayons solaires et pour une distance Terre-Soleil égale à sa valeur moyenne, est appelée la **constante solaire** (I_0). La constante solaire a été estimée à 1367 W/m^2 d'après des résultats de mesures faites entre 1969 et 1980. Cette valeur de la constante solaire a été adoptée par la Commission des instruments et des méthodes d'observation (CIMO) de l'Organisation météorologique mondiale (OMM) en octobre 1981 [12].

Les mesures récentes faites à l'aide de radiomètres embarqués sur satellite ont permis de bien mettre en évidence les faibles variations de la constante solaire liées à la rotation du Soleil sur lui-même et au cycle d'activité de onze ans du Soleil.

Le rayonnement solaire hors atmosphère ne dépend que de la constante solaire (I_0), de la hauteur solaire (h) et de la distance Terre-Soleil (r).

Le rayonnement solaire hors atmosphère sur une surface perpendiculaire à la direction de propagation du rayonnement solaire pour chaque variation de la distance Terre-Soleil est donné par l'expression suivante [6]:

$$G_0 = \left(\frac{\bar{r}}{r} \right)^2 \cdot I_0 \cdot \cos(\theta_z) \quad (\text{I.7})$$

$$\frac{\bar{r}}{r} = \frac{1}{1 - 0.033 \cdot \cos \frac{2\pi \cdot j}{Y}} \quad (\text{I.8})$$

\bar{r} : est la distance moyenne du Soleil égale à $1.49 \cdot 10^{11}$ m.

r : est la distance actuelle du Soleil.

Y : est le nombre total de jour de l'année.

j : le numéro du jour à compter du 1^{er} janvier.

θ_z : est l'angle zénithale : l'angle formé entre la normale du plan et les rayons du soleil. C'est l'angle d'incidence du rayonnement solaire direct sur un plan horizontal.

$$\cos(\theta_z) = \cos(\Phi) \cos(\delta) \cos(\omega) + \sin(\Phi) \sin(\delta) \quad (\text{I.9})$$

Le rayonnement solaire total hors atmosphère entre le couché et le levé du Soleil est calculé comme suit :

$$G_0 = \int_{\omega_L}^{\omega_s} \left(\frac{\bar{r}}{r} \right)^2 \cdot I_0 \cdot \cos \theta_z \cdot d\omega \quad (\text{I.10})$$

La valeur réelle du rayonnement extraterrestre (hors atmosphère) varie de $\pm 3.3\%$ au cours de l'année du fait de l'ellipticité de l'orbite terrestre.

I.2.3.2 Les différentes composantes du rayonnement solaire

L'ensoleillement correspond à l'intensité du rayonnement solaire reçu sur un plan à un moment donné. Il s'exprime habituellement en watts par mètre carré (W/m^2). L'ensoleillement varie de zéro, au lever du Soleil, à sa valeur maximale, typiquement au midi solaire.

L'ensoleillement peut également exprimer la quantité d'énergie solaire captée sur un plan pendant un intervalle déterminé. Il s'exprime habituellement en kilowatt heure par mètre

carré (kWh/m^2), en « heure de soleil maximum », en mégajoule par mètre carré (MJ/m^2) ou en calorie par centimètre carré (cal/cm^2) pour l'intervalle déterminé – une journée ou une heure, par exemple.

$$\begin{aligned} 1\text{kWh} / \text{m}^2 / \text{jour} &= 1 \text{ heure de soleil maximum } (1000\text{W/m}^2/\text{jour}) \\ &= 3,6 \text{ MJ/m}^2 / \text{jour} \\ &= 86 \text{ Cal/cm}^2/\text{jour} \end{aligned}$$

Les valeurs quotidiennes d'ensoleillement sont habituellement présentées sous forme de moyennes mensuelles pour différentes latitudes et inclinaisons à des endroits précis (afin de rendre compte de l'effet de conditions météorologiques différentes). Le plus souvent, on exprime l'ensoleillement en « heures de soleil maximum », c'est-à-dire par le nombre équivalent d'heures par jour où l'éclairement est en moyenne de 1000 W/m^2 . Les heures de soleil maximum sont un indice utile pour le dimensionnement des systèmes photovoltaïques, car ces systèmes sont habituellement évalués en watt crête (W_c), c'est-à-dire sous un éclairement de 1000 W/m^2 .

Lorsque le rayonnement solaire se propage dans l'atmosphère, il interagit avec les constituants gazeux de celle-ci et avec toutes les particules présentes en suspension (aérosols, gouttelettes d'eau et cristaux de glace). Les particules dont on parle ici ont des dimensions variant du centième de μm à quelques centaines de μm .

En traversant l'atmosphère, le rayonnement solaire est absorbé et diffusé. Au sol, on distingue plusieurs :

- **Réfléchi** : par la surface terrestre, c'est-à-dire renvoyé dans une direction privilégiée (réflexion dite spéculaire) ou de manière diffuse. Le sol réfléchit plutôt le rayonnement de manière diffuse et anisotrope.
- **Diffusé** : c'est-à-dire renvoyé dans toutes les directions. Le phénomène de diffusion se produit dans un milieu contenant de fines particules ou des molécules et dépend fortement de la taille des particules considérées. Par exemple, l'influence des molécules est plus intense pour les courtes longueurs d'onde (bleu) que pour les grandes (rouge). C'est la raison pour laquelle la voûte céleste apparaît en général bleue et le Soleil couchant rougeâtre (les rayonnements violet et bleu ayant été diffusés). Les molécules diffusent la lumière dans toutes les directions, cependant, deux directions sont privilégiées : la diffusion avant et la diffusion arrière. Pour les particules les plus grosses (cas des gouttelettes de nuages), la diffusion se fait majoritairement en avant.
- **Absorbé** : par les composants gazeux de l'atmosphère. Cette absorption est dite sélective, car elle s'opère pour des valeurs de longueur d'onde bien précises. Elle est due

essentiellement à la vapeur d'eau, à l'ozone, au dioxyde de carbone et à un degré moindre, à l'oxygène.

- **Direct** : On appelle rayonnement solaire direct celui qui arrive au sol sans avoir subi de diffusion. Il s'éloigne de façon notable du rayonnement atteignant la limite supérieure de l'atmosphère, en particulier du fait de l'absorption par les constituants gazeux de l'atmosphère. Dans certaines bandes de longueur d'onde, le rayonnement est atténué ou même annulé.
- **Albédo** : est la partie réfléchi par le sol. Il dépend de l'environnement du site, il faudra en tenir compte pour évaluer le rayonnement sur plans inclinés.
- **Global** : est le rayonnement solaire parvenant au niveau du sol sur une surface horizontale, soit directement, soit après diffusion. Il s'exprime simplement comme la somme du rayonnement diffus et du rayonnement direct pondéré par le cosinus de l'angle zénithal du Soleil.

Pour les besoins de divers modèles, nous définissons également l'indice de clarté K_t comme le rapport du rayonnement au sol normalisé à l'extraterrestre (hors atmosphère). Cette grandeur est une mesure de l'atténuation dans l'atmosphère, et peut être définie pour chaque composante globale, diffuse et directe.

$$K_t = \frac{G}{G_0} \quad (\text{I.11})$$

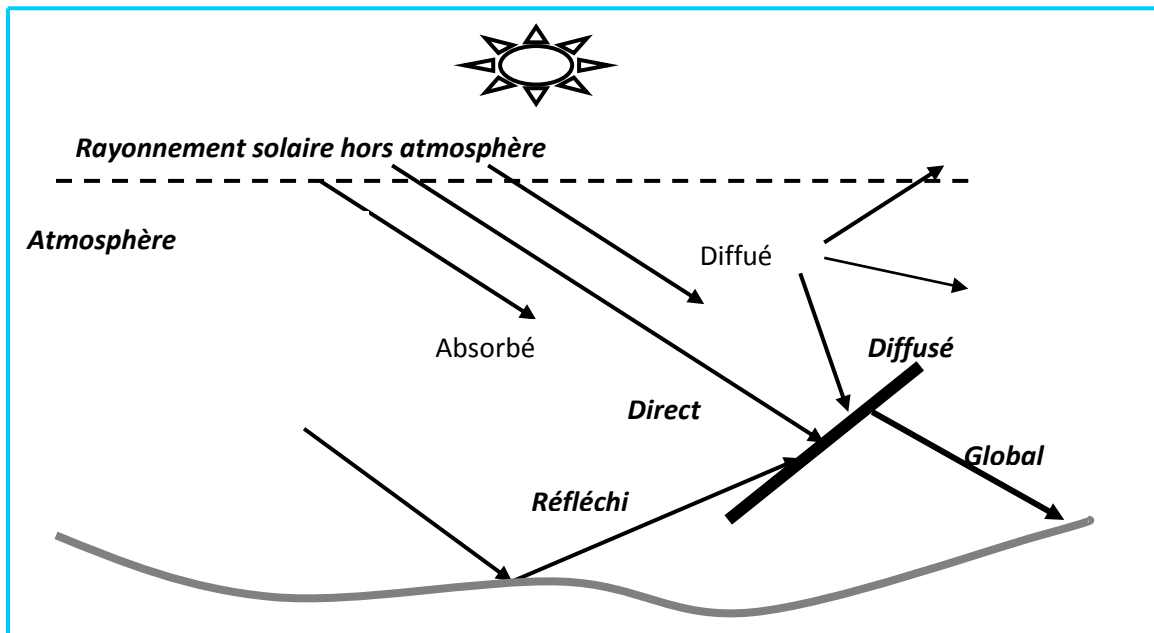


Figure I.6 Composantes du rayonnement solaire global.

Lieu géographique, orientation, saison, heure de la journée,..., jouent directement sur le rayonnement instantané. C'est ce qui déroute bien souvent les utilisateurs pour définir leurs systèmes solaires. L'énergie délivrée par un panneau solaire est hautement variable ! .

L'éclairement instantané n'est d'ailleurs pas très utile, on se servira plutôt de valeurs globales intégrées sur une journée pour chiffrer l'énergie récupérable [30].

I.3 Les appareils de mesure des différentes composantes solaire

L'utilisation du rayonnement solaire comme source d'énergie pose donc un problème bien particulier. En effet, le rayonnement solaire n'est pas toujours disponible, en outre on ne peut l'emmagasiner ni le transporter. Le concepteur d'un système qui emploie le rayonnement solaire comme source d'énergie doit donc déterminer la quantité d'énergie solaire disponible à l'endroit visé et le moment où cette énergie est disponible.

Plusieurs appareils de mesure existent depuis des années et servent à la mesure de différentes composantes du rayonnement solaire. Néanmoins ces appareils ne peuvent pas être disponibles sur n'importe quelle site pour leurs onéreux coûts.

De ces appareils de mesure on peut citer :

- **L'héliographe**

L'héliographe est un instrument qui permet de mesurer la durée de l'ensoleillement sur un point de la surface de la planète. Plus précisément, il enregistre la durée pendant laquelle le rayonnement solaire est d'une intensité suffisante pour produire des ombres distinctes.

Un héliographe est constitué d'une boule de cristal de 10 cm de diamètre, qui, exposée au soleil, fait effet de lentille (ce dernier est représenté sur la figure I.7). Cette lentille concentre le rayonnement de manière à brûler une feuille de papier rigide spécial placée sur un support incurvé (équivalent à la bande équatoriale d'un cadran éponyme) .

Le papier brûle au fur et à mesure du déplacement du point focal, tant que le soleil brille. Cette bande de papier (héliogramme) est graduée en heures, ou plus finement, elle permet de déterminer la durée d'ensoleillement.

Le déplacement du soleil fait changer le point de convergence des rayons sur le carton, ce qui produit une ligne brûlée sur celui-ci. Si un nuage cache les rayons du soleil, il y aura sur le carton un endroit, correspondant à cette heure, qui ne sera pas brûlé.

L'emplacement de l'appareil, doit être dégagé de tout obstacle matériel pouvant porter ombre sur l'héliographe.

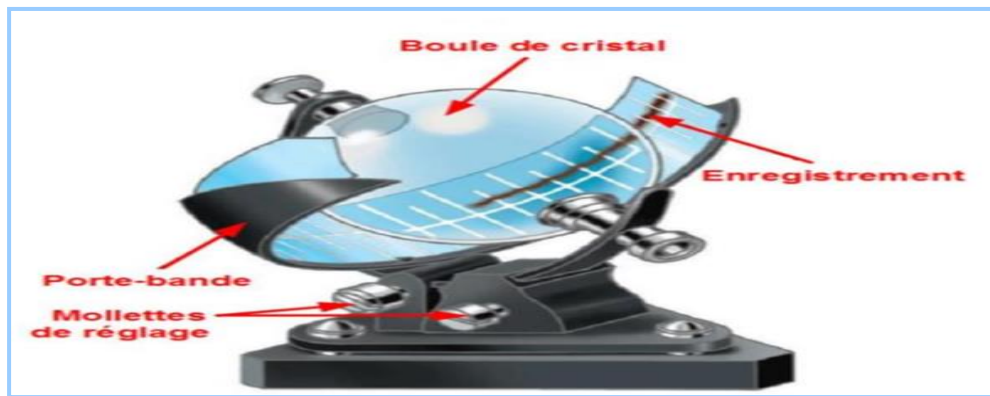


Figure I.7 Héliographe Campbell-Stokes.

- **Pyranomètre**

Il mesure le rayonnement qui provient de la voûte céleste (diffusé à travers les nuages) donc l'éclairement énergétique global (représenté par la figure I.8).

L'élément sensible du pyranomètre est une thermopile constituée de 64 thermocouples imprimés sur un support de céramique et disposés en étoile, les soudures froides sont placées sur le bord du substrat et en contact thermique avec le boîtier, les soudures chaudes sont placées près du centre.

La thermopile exposée au rayonnement solaire délivre une différence de potentiel proportionnel au flux incident ($0 - 15 \mu\text{V} / \text{W} \cdot \text{m}^{-2}$).

Le pyradiomètre mesure également le rayonnement diffus c'est-à-dire celui qui provient du sol terrestre échauffé par les rayons solaire. Il a donc un intérêt tout particulier dans l'évaluation de l'augmentation de l'effet de serre. Tous les ans, il doit être étalonné.



Figure I.8 Le pyradiomètre

- **Pyrhéliomètre**

Le pyrhéliomètre mesure les irradiances directes. Il comporte une ouverture réduite et une surface réceptrice qui doit être maintenue normale aux rayons du soleil par un système automatique.

L'élément sensible (thermopile) du pyrhéliomètre est constitué de thermocouples montés en série, exposés au rayonnement solaire et dont les soudures froides sont maintenues à la température de l'air par conduction. Il délivre une différence de potentiel proportionnelle au flux incident. La géométrie de l'appareil limite le rayonnement incident au rayonnement solaire direct et au rayonnement provenant du ciel circumsolaire.

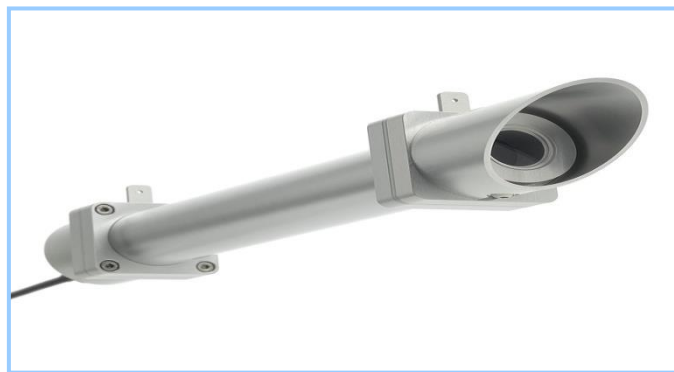


Figure I.9 Pyrhéliomètre.

I.4 Conclusion

En conclusion de ce chapitre, nous avons exploré en détail le rayonnement solaire, en mettant l'accent sur sa composition, les aspects géométriques et énergétiques qui le caractérisent, ainsi que les différents appareils de mesure utilisés pour quantifier ses différentes composantes.

Nous avons appris que le rayonnement solaire est une source d'énergie essentielle provenant du soleil, qui se compose de plusieurs composantes. Chacune de ces composantes joue un rôle spécifique dans l'irradiation solaire totale atteignant la surface terrestre.

Chapitre II

Méthodes de l'intelligence artificielle

II.1 Introduction

L'intelligence artificielle (IA) est un domaine de l'informatique qui vise à créer des systèmes informatiques capables d'imiter, de comprendre et d'effectuer des tâches qui nécessitent normalement l'intelligence humaine. L'IA cherche à développer des programmes et des machines qui peuvent raisonner, apprendre, percevoir, planifier, comprendre le langage naturel et résoudre des problèmes complexes.

Dans ce chapitre, nous explorerons en profondeur les différentes méthodes d'intelligence artificielle qui ont révolutionné le domaine. Nous examinerons les approches classiques, mais aussi les techniques de pointe qui ont émergé récemment, et qui repoussent les limites de ce qui était auparavant considéré comme possible.

Nous commencerons par les bases de l'apprentissage automatique, une discipline fondamentale de l'IA, qui permet aux machines d'apprendre à partir de données et de prendre des décisions en se basant sur ces connaissances acquises. Nous discuterons les algorithmes d'apprentissage supervisé et non supervisé, ainsi que les techniques d'apprentissage par renforcement qui permettent à une machine d'apprendre à partir de l'interaction avec son environnement.

Nous aborderons également quelques méthodes de l'apprentissage automatique telles que les arbres de décision, la régression linéaire ainsi que les machines à vecteurs supports.

II.2 L'intelligence artificielle

II.2.1 Définition

L'intelligence artificielle (IA) se réfère à la capacité des machines et des systèmes informatiques à réaliser des tâches qui nécessitent normalement l'intelligence humaine. L'IA vise à reproduire, voire à dépasser, les capacités cognitives humaines telles que le raisonnement, l'apprentissage, la résolution de problèmes, la perception sensorielle, la reconnaissance de la parole et du langage, ainsi que l'interaction sociale.

Cette dernière, utilise diverses méthodes et techniques, notamment l'apprentissage automatique (machine learning), les réseaux de neurones artificiels, le traitement du langage naturel, la vision par ordinateur, l'exploration de données (data mining), la logique floue et d'autres approches de modélisation et d'analyse des données.

II.3 Apprentissage automatique

II.3.1 Définition

L'apprentissage automatique (ou « machine learning » en Anglais) est un champ d'études de l'intelligence artificielle qui se fonde sur des approches mathématiques et statistiques pour donner aux ordinateurs la capacité d'apprendre à partir de données, c'est-à-dire d'améliorer leurs performances à résoudre des tâches sans être explicitement programmés pour chacune.

Plus largement, il concerne la conception, l'analyse, l'optimisation, le développement et l'implémentation de telles méthodes.

L'apprentissage automatique comporte généralement deux phases. La première consiste à estimer un modèle à partir de données, appelées observations, qui sont disponibles et en nombre fini, lors de la phase de conception du système. L'estimation du modèle consiste à résoudre une tâche pratique, telle que traduire un discours, estimer une densité de probabilité, reconnaître la présence d'un chat dans une photographie ou participer à la conduite d'un véhicule autonome.

Cette phase dite « d'apprentissage » ou « d'entraînement » est généralement réalisée préalablement à l'utilisation pratique du modèle. La seconde phase correspond à la mise en production : le modèle étant déterminé, de nouvelles données peuvent alors être soumises afin d'obtenir le résultat correspondant à la tâche souhaitée. En pratique, certains systèmes peuvent poursuivre leur apprentissage une fois en production, pour peu qu'ils aient un moyen d'obtenir un retour sur la qualité des résultats produits.

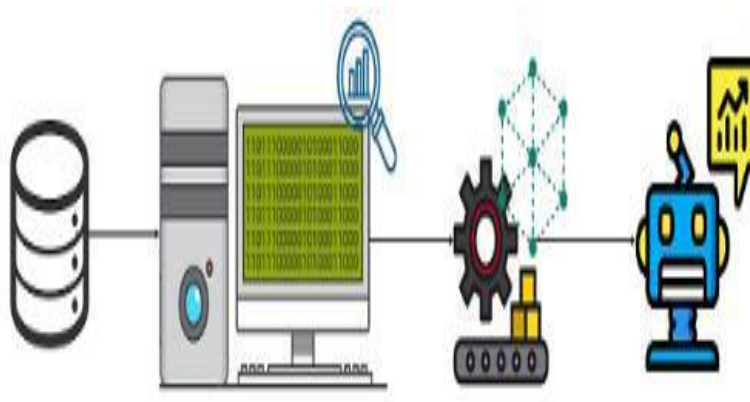


Figure II.1 Processus de l'apprentissage automatique.

II.3.2 Les étapes de l'apprentissage automatique

Afin de développer et gérer un modèle d'apprentissage automatique adapté à une mise en production, il faut passer par les étapes suivantes :

- Collecte des données.
- Prétraitement des données.
- Formation et test du modèle.
- Évaluation.

II.3.2.1 Collecte des données

La première étape consiste à la collecte de données, celle-ci est très importante car la qualité et la quantité des données collectées déterminent la qualité du modèle à venir.

L'ensemble de données (Data set en anglais) peut être collecté à partir de diverses sources telles qu'un fichier, une base de données, un capteur et de nombreuses autres sources similaires.

Cependant les données collectées ne peuvent pas être utilisées directement, elles peuvent avoir beaucoup de champs manquants, des valeurs extrêmement grandes ou non organisées, ou bien être bruyantes.

II.3.2.2 Prétraitement des données

Le prétraitement des données est un processus de nettoyage des données brutes en données propres.

Voici quelques techniques de prétraitement de base :

- Conversion des données : Comme les modèles d'apprentissage automatique ne peuvent gérer que des fonctionnalités numériques, les données catégorielles et ordinales doivent donc être converties en fonctionnalités numériques.
- Ignorer les valeurs manquantes : A la détection des données manquantes, la ligne ou la colonne des contenants peuvent être supprimées selon les besoins. Cette méthode est connue pour être efficace, mais elle ne devrait pas être appliquée excessivement.
- Remplissage des valeurs manquantes : Les données manquantes dans l'ensemble de données peuvent être remplacées manuellement par la valeur moyenne, médiane ou la plus haute fréquence utilisée.
- Détection des valeurs aberrantes : Certaines données d'erreurs qui pourraient être présentes dans l'ensemble de données s'écartent considérablement des autres

observations de l'ensemble de données. Par exemple : Poids humain = 800 kg, en raison d'une faute de frappe sur l'extra 0. Celles-ci doivent être supprimées ou remplacées.

II.3.2.3 Formation et test du modèle

Pour l'entraînement d'un modèle, il faut initialement diviser le modèle en trois sections qui sont : Données de formation «Training Set en anglais », Données de validation «Validation Set en anglais » et Données de test «Testing Set en anglais ».

Par exemple, on entraîne le classificateur à l'aide d'un ensemble de données d'apprentissage, et on ajuste les paramètres à l'aide d'un ensemble de validation, puis on teste les performances du classificateur sur un « ensemble de données de test ». Un point important à noter est que pendant la formation du classificateur, seul l'ensemble d'entraînement et / ou de validation est disponible. L'ensemble de données de test ne doit pas être utilisé pendant l'entraînement du classificateur. L'ensemble de test ne sera disponible que lors du test du classificateur.

II.3.2.4 L'évaluation

Elle est une partie intégrante du processus d'élaboration du modèle. Elle aide à trouver le meilleur modèle qui représente les données et à quel point le modèle choisi fonctionnera à l'avenir.

II.3.3 Types d'apprentissage automatique

Pour donner à la machine la capacité d'apprendre, on utilise différentes méthodes d'apprentissage automatique

II.3.3.1 Apprentissage supervisé

Cette approche consiste à faire apprendre une machine à travers des exemples d'entrées qui sont labellisés avec la sortie souhaitée afin que l'algorithme puisse prédire les valeurs des sorties en fonction des entrées.

Plus formellement, étant donné un ensemble de données D , décrit par un ensemble de caractéristiques X , un algorithme d'apprentissage supervisé va trouver une fonction f en entrée x et la variable à prédire y .

La fonction décrivant la relation entre x et y s'appelle un modèle de prédiction .D'une manière générale, la machine peut apprendre une relation $f: x \rightarrow y$ qui relie x à y en ayant

analysé des millions d'exemples d'associations $x \rightarrow y$ et l'utiliser pour catégoriser les données non triées par la suite.

On distingue deux problèmes d'apprentissage supervisé la classification et la régression que nous décrivons dans ce qui suit :

a) Classification

La classification consiste à prédire la valeur ou bien la classe d'une variable d'entrée discrète.

Par exemple, la classification peut servir prédire si un email est un spam ou non selon le nombre de liens présent dans l'email.

b) Régression

La régression consiste à prédire la valeur ou bien la classe d'une variable d'entrée continue.

Par exemple, en utilisant la régression on peut prédire le prix d'appartement selon sa surface habitable.

II.3.3.2 Apprentissage non supervisé

On parle d'apprentissage non supervisé quand on ne dispose pas d'étiquettes. L'algorithme doit apprendre tout seul à trouver le point commun entre les données d'entrée et les regrouper dans des clusters ou classes.

On prend comme exemple la reconnaissance d'images. L'algorithme va pouvoir regrouper des images de voiture dans un cluster et des immeubles dans un autre cluster ainsi de suite, sans la structure ou le concept de l'objet ne soit prédéfini qui est dans notre cas une voiture et un immeuble. Il saura seul que le point en commun entre les images de voitures est la roue, et qu'un bâtiment a des forme plus droite.

II.3.3.3 Apprentissage semi - supervisé

L'apprentissage semi-supervisé est une combinaison entre l'apprentissage supervisé (donnée labellisé) et l'apprentissage non supervisé (donnée non labellisé).

Cette approche a pour but de prédire les labels des données non labellisé et améliorer les performances.

II.3.3.4 Apprentissage par renforcement

L'algorithme va apprendre à se comporter à partir d'un environnement réel ou simulateur, en interagissant avec ce dernier afin de construire son propre data set (données).

II.3.3.5 Apprentissage par transfert

L'apprentissage automatique permet de sauvegarder toutes les données et connaissances déjà traité ; de ce fait, on peut résoudre une tâche à partir d'une autre tâche similaire déjà résolu.

II.4 Quelques méthodes de l'apprentissage automatique

II.4.1 Les arbres de régression

Les arbres de décision sont une catégorie d'arbres qui emploient une représentation hiérarchique de la structure des données qualitatives sous forme des séquences de décisions (tests) en vue de la prédiction d'un résultat ou d'une classe. Chaque individu (ou observation), qui doit être attribué(e) à une classe, est décrit(e) par un ensemble de variables qui sont testées dans les nœuds de l'arbre. Les tests s'effectuent dans les nœuds internes et les décisions sont prise dans les nœuds feuilles.

Exemple, les arbres de décision peuvent répartir une population d'individus en groupes homogènes selon un ensemble de variables descriptives (ex. âge, temps passé sur un site Web, etc.) et en fonction d'un objectif fixé (variable de sortie, par exemple : chiffre d'affaires, probabilité de cliquer sur une publicité, etc.)

II.4.1.1 Les arbres de régression simples

Les arbres de régression classiques ont été la première déclinaison des arbres de décision utilisés pour traiter des problèmes de régression sur les données. Ce sont [Hastie et Tibshirani 1986] qui ont proposés le premier formalisme de ce modèle régressif, la formule mathématique est :

$$\hat{x}(t + h) = \sum_{i=1}^{t-1} k_i \times H(x(t - i)) \quad (\text{II.1})$$

où k_i est un facteur, H est une fonction qui retourne 1 si la donnée est utilisée ou 0 si elle ne l'est pas.

Une fois que l'arbre est construit, un modèle de régression est appliqué à chaque nœud. Lors de la phase d'apprentissage, un processus itératif, on cherche à minimiser l'erreur entre la valeur à l'instant t et la même valeur prédite.

II.4.1.2 Les arbres de régression renforcés « boosting »

L'arrivée de la méthode de boosting dans le monde de l'apprentissage automatique découle de l'introduction de l'algorithme Adaboost [20], un algorithme de classification binaire.

L'idée de base est de combiner les réponses de plusieurs prédicteurs basiques (utilisables sur n'importe quel modèle de machine learning) pour obtenir une réponse plus précise obtenant ainsi un prédicteur plus complexe.

Le classifieur de base, le plus souvent utilisé comme prédicteur basique par l'algorithme, est un arbre simple de classification à une seule division.

L'algorithme crée des arbres successifs à partir de différentes pondérations des données à chaque étape, les données sont classifiées par l'état actuel et ces classifications sont utilisées comme poids pour générer les arbres suivants.

Les données mal classifiées sont pondérées de manière plus grande que les données correctement classifiées, Les données difficiles à classifier reçoivent des poids toujours plus élevés, cela augmente les chances de les classer correctement.

Le résultat final de la classification est obtenu par la majorité pondérée de la classification à travers la génération des arbres du modèle. Cette méthode d'amélioration, d'abord développée pour les arbres de classification, a été adaptée aux arbres de régression.

C'est la première méthode qui utilise la technique des ensembles. En effet la famille des modèles à base d'arbres de régression a vu naître un nouveau sous-groupe de modèles, les modèles dits « d'ensemble learning ». Les arbres dont la réponse améliore la prévision sont pondérés et la prévision finale est obtenue par une combinaison linéaire des arbres [17].

La méthode de boosting peut être décrite par l'algorithme suivant pour la régression :

Soit x_0 la variable à prévoir, et $z = \{(x_1/y_1), \dots, (x_n/y_n)\}$ un échantillon.

Initialiser p par la distribution uniforme $P = \{p_i = 1/n ; i = 1, \dots, n\}$.

For $m = 1$ à M do

Tirer avec remise dans z un échantillon z_m^* suivant la distribution P .

Estimer \hat{f}_m sur l'échantillon z_m^*

Calculer à partir de l'échantillon initial z :

$l_m(i) = l(y_i, \hat{f}_m(x_i))$ $i = 1, \dots, n$. Ou l est la fonction de perte.

$\hat{\epsilon}_m = \sum_{i=1}^n p_i l_m(i)$ $w_i = g(l_m(i))p_i$. (g : est continue non décroissante).

Calculer les nouvelles probabilités : $p_i \leftarrow \frac{w_i}{\sum_{i=1}^n w_i}$

End For

II.4.1.3 Les arbres de régression ensachés « bagging »

Le terme de bagging est l'acronyme de « bootstrap aggregating ». Il s'agit d'un autre niveau d'amélioration des modèles de prévision et notamment des arbres de régression.

La description de cette méthode a été réalisée par Breiman (1996). Cette méthode consiste à générer des arbres de régression sur des échantillons du jeu de données, mais contrairement à la méthode de boosting, les arbres successifs ne dépendent pas des réponses des arbres précédents.

En effet, chaque arbre est construit en utilisant un échantillon d'amorce, dit « Bootstrap, » du jeu de données. Un Bootstrap est défini comme un nouvel échantillon créé à partir du jeu de données originel. Chaque échantillon est construit par tirage avec remise, c'est ce que l'on appelle le ré-échantillonnage. L'hypothèse gaussienne sur la distribution des échantillons n'est pas nécessaire. Lorsqu'on a généré les arbres et que l'on dispose des réponses correspondantes, un simple vote majoritaire est utilisé pour réaliser la prévision. Cette méthode d'ensemble donne des améliorations significatives dans le traitement de données concernant des phénomènes complexes et dont les relations entrées/sorties sont difficiles à appréhender.

Tout comme pour la méthode du bagging, il est possible de représenter cette méthode par un algorithme :

Soit x_0 la variable à prévoir, et $z = \{(x_1/y_1), \dots, (x_n/y_n)\}$ un échantillon.

For $b = 1$ à B do

Tirer dans z un échantillon z_b^*

Estimer $\hat{f}_{z_b}(x_0)$ sur l'échantillon bootstrap.

End For .

Calculer l'estimation moyenne $\hat{f}_B(x_0) = \frac{1}{B} \sum_{b=1}^B \hat{f}_{z_b}(x_0)$ le résultat du vote.

Le nombre d'arbres générés est déterminé par une méthode heuristique, le test est effectué sur un nombre croissant d'arbres. On calcule l'erreur sur la prévision pour déterminer k .

II.4.2 Régression linéaire

La régression linéaire est une méthode fondamentale de l'apprentissage automatique et de l'analyse statistique, elle permet de modéliser la relation entre une variable dépendante continue et une ou plusieurs variables indépendantes continues ou catégorielles. Le modèle de régression linéaire est basé sur l'hypothèse que la variable dépendante peut être exprimée comme une combinaison linéaire des variables indépendantes, pondérée par des coefficients.

Une fois le modèle de régression linéaire construit, il peut être utilisé pour effectuer des prédictions sur de nouvelles données en utilisant les valeurs des variables indépendantes.

De plus, l'analyse des coefficients permet de comprendre l'impact relatif de chaque variable indépendante sur la variable dépendante.

En résumé, la régression linéaire est une technique puissante pour modéliser et analyser les relations linéaires entre des variables, elle constitue une pierre angulaire de l'analyse statistique et de l'apprentissage automatique, offrant une compréhension et des prédictions précises dans de nombreux domaines.

II.4.3 Machines à vecteurs supports pour la régression (MVS)

Les machines à vecteurs de support sont un ensemble de techniques d'apprentissage supervisé destinées à traiter des problèmes de discrimination ou de régression. Ce sont des classifieurs linéaires généralisés.

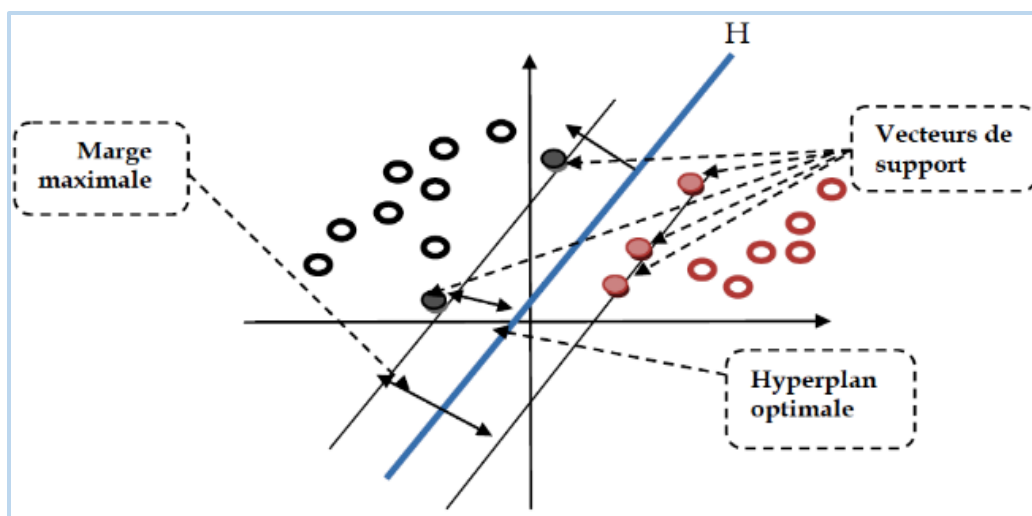


Figure II.2 L'hyperplan H optimale, vecteurs de supports et marge maximale.

La figure II.2 présente l'hyperplan H optimale, vecteurs de supports et la marge maximale.

Le développement de ces modèles est directement issu des travaux de Valdimir Vapnik dans les années 80 [Vapnik, 1986]. Les « support vector regression » (SVR) sont issus de l'adaptation des machines à vecteur de support utilisés pour la régression sur les données. Cette méthode a été appliquée avec succès à la prévision de séries temporelles. Le formalisme des SVR est similaire à celui des processus Gaussiens. La formule mathématique des SVR appliquée à la prévision de séries temporelles [Lauret *et al.*, 2015] est comme suit :

$$\hat{x}(t+h) = \sum_{i=1}^n \alpha_i \cdot k_{rbf}(x_i, x_{test}) + b \quad (II.2)$$

Dans laquelle, x_i est le $i^{\text{ème}}$ vecteur d'entrées pour l'apprentissage et x_{test} le vecteur de test. La fonction de base radiale est donnée par :

$$k_{rbf}(x_p, x_q) = \exp \left[\frac{-(x_p - x_q)^2}{2\sigma^2} \right] \quad (II.3)$$

Le paramètre b (paramètre de biais) est déduit de l'équation précédente, l'hyperparamètre σ également, il définit notamment la complexité du modèle.

Concernant les SVR, les coefficients α_i sont liés à la différence de deux multiplicateurs de Lagrange qui sont déduits par la résolution d'un problème de programmation quadratique. Contrairement aux réseaux de neurones artificiels, sujets aux problèmes de minimums locaux, pour les SVR le problème est strictement convexe et ne possède donc qu'une seule et unique solution. En outre, il faut souligner (contrairement aux processus Gaussiens) que tous les modèles d'entraînement ne participent pas à la relation précédente. En effet, un choix pratique d'une fonction de coût (fonction ϵ de Vapnik) dans le problème quadratique, permet d'obtenir une solution particulière construite à partir d'une régression. A noter alors que certains des coefficients α_i seront non-nuls.

II.5 Conclusion

Dans le présent chapitre, nous avons mis un œil sur l'apprentissage automatique, également connu sous le nom de machine Learning, qui est un domaine de l'intelligence artificielle qui se concentre sur le développement de techniques permettant aux ordinateurs d'apprendre à partir de données sans être explicitement programmés. Nous avons également présenté les différents types d'apprentissage automatique tout en passant par trois familles de modèles de l'apprentissage

automatique à savoir : les arbres de régression, la régression linéaire et aussi les machines à vecteurs supports.

Ces derniers modèles font l'objet du quatrième chapitre dont ils vont être développés et comparés.

Chapitre III

Méthodologie et modèles de prévision

III.1 Introduction

L'irradiation solaire dépend de différents facteurs géographiques et météorologiques. A cet égard, plusieurs modèles empiriques et intelligents ont été développés pour calculer l'irradiation solaire globale à l'aide de divers paramètres climatiques [44] [10] [18] [19] [9] [1] [39] [22]

Dans ce chapitre, nous allons présenter quelques modèles proposés dans la littérature pour l'estimation de l'irradiation solaire globale en fonction de quelques paramètres météorologiques et géographiques. En plus les modèles estimant l'irradiation solaire globale en fonction du taux d'insolation, et de la température seront plus particulièrement détaillés dans le quatrième chapitre parce qu'ils constituent la base de la méthodologie retenue et développée au cours de notre travail.

Nous présenterons aussi dans ce qui suit, les principaux critères de performances permettant de comparer ces différentes méthodologies de prédiction. Sur la base de différents indicateurs de précision, une comparaison bibliographique des méthodes d'estimations relatives à l'irradiation solaire globale sera également proposée.

III.2 Etat de l'art de la prévision de l'irradiation solaire globale

Dans les dernières années, les techniques intelligentes (réseau de neurones, logique floue, réseau neuro-flou,..., les méthodes d'apprentissage automatiques ... etc.) sont largement utilisées pour résoudre des problèmes réels dans le cas où les méthodes classiques sont insuffisantes. La prédiction de l'irradiation solaire est l'un des domaines où les techniques intelligentes ont été utilisées avec succès, et en particulier les réseaux de neurones artificiels et les mathodes de l'apprentissage automatiques [45] [22] [1] [39].

Plusieurs applications de l'intelligence artificielle pour la modélisation, et la prédiction de l'irradiation solaire mensuelle, journalière et horaire existent dans la littérature.

Le Tableau III.1 résume quelques travaux existants dans la littérature et qui peuvent nous servir d'exemple par la suite pour construire notre méthodologie de prédiction.

Nous nous concentrons dans cette section à faire une mise au point sur les derniers travaux cités dans la littérature dans lesquels il est possible de retrouver une grande partie de tous ce qui révèle du couplage des techniques de l'IA avec l'irradiation solaire.

Références	Année	Localisation	Horizon	Méthode utilisée	Indice de performance
Liu J et al.	2015	Minqin,Gansu	1jour d'avance	ANN	MAPE :7.65%
Gigoni L et al.	2015	Italy	1 jour d'avance	ENS	nMAE :1.27–4.04
Zhang Yet al.	2015	SanDiego, Braedstrup and Catania, Italy	1 jour d'avance	k-NN	nMAE: 7.4, 6.38 et 7.74
Baharin K.A et al .	2016	Melaka, Malaysia	12 h d'avance Min	SVR	RMSE: 4.29%–6.85%
Pierro M et al.	2016	Bolzano,Italy	1 Jour d'avance	MME	RMSE :10.5%
Leva S et al.	2017	Milano, Italy	1 jour d'avance	ANN	MAE < 15%
Liu F et al.	2017	Queensland, Australia.	1h d'avance	FL	MAE =9.77%
Leva S et al.	2017	Milano, Italy	1h	ANN	MAE < 15%
Pierro M et al.	2017	Italy	Jusqu'à 48	ANN	RMSE: 5%–7% Pour 1-4h RMSE: 7%–7.5% pour 1–2 jour
Liu W et al.	2018	Beijing,China	1h d'avance	SVM et ANN	MRE :11.61%
Al-Dahidi et al.	2018	Amman,jordan	24 d'avance	ELM-ANN	MAE :1.08%
Yao X et al.	2019	China	1h d'avance	ESN	MAPE:–0.00195%
Han Y et al.	2019	China	Quelques heures	ELM	MAE :2.13%
Kesh B Pun.	2020	Wichita, Kansas, USA	/	SVR KRR LASSO RR ARMA	RMSE: 0.0943% RMSE: 0.1106% RMSE: 0.1107% RMSE: 0.1106% RMSE: 0.0984%
Bouguerra O et al.	2020	HI-SEAS Weather Station	/	RF	MAE: 9.64
Chahboun S et al.	2021	/	/	MLR SVR RF	MAE: 0.58 /R ² : 0.87 MAE: 0.36 /R ² : 0.92 MAE: 0.25 /R ² : 0.96

Tableau III.1 Résumé des applications de l'intelligence artificielle pour l'estimation de l'irradiation solaire.

D'autres méthodes de l'intelligence artificielle et en particulier les méthodes de l'apprentissage automatique sont plus détaillées dans ce qui suit :

Dans l'article [32], un modèle de prévision basé sur un réseau de neurones à perceptron multicouche (MLP) a été entraîné en utilisant des prévisions météorologiques et des données historiques. Le modèle a donné de meilleurs résultats pendant les journées ensoleillées que pendant les journées partiellement nuageuses. L'erreur moyenne absolue (MAE) normalisée était inférieure à 15% pour tous les cas étudiés.

Dans le travail de [41], une nouvelle méthode d'agrégation a été développée pour estimer la puissance produite par une installation photovoltaïque (PV) en Italie. La méthode utilise des données provenant de satellites et de modèles de prévision numérique du temps (NWP) pour estimer la production solaire à l'échelle régionale. La méthode a été appliquée à la production d'électricité de 1985 petites installations photovoltaïques installées dans la région du Tyrol du Sud, en Italie (la superficie totale couverte était de 800 km²). L'erreur quadratique moyenne (RMSE) était dans la plage de 5-7% pour un horizon temporel de 4 heures et dans la plage de 7-7,5% pour l'estimation d'une journée.

La méthode utilise des données provenant de satellites et de modèles de prévision numérique du temps (NWP) pour estimer la génération solaire à l'échelle régionale. La méthode a été appliquée à la production d'électricité de 1985 petites centrales photovoltaïques installées dans la région du Tyrol du Sud, en Italie (la superficie totale couverte était de 800 km²). L'erreur quadratique moyenne (RMSE) se situait dans la fourchette de 5 à 7 % pour une période de prévision de 4 heures, et dans la fourchette de 7 à 7,5 % pour l'estimation sur 1 jour. Les méthodes SVM (Support Vector Machines) et MLP (Multilayer Perceptron) ont été utilisées pour la prévision à très court terme d'une petite centrale photovoltaïque installée à Beijing, en Chine [36].

Dans l'article [34], un modèle spécialement conçu pour les conditions environnementales particulières de brouillard a été développé pour prévoir la production d'une installation photovoltaïque à Beijing, en Chine. Les données d'entrée de ce modèle comprennent les mesures sur site de la température de l'air, de l'humidité relative et des indices d'aérosols.

Dans l'article [5], un algorithme ELM (Extreme Learning Machine) a été développé pour entraîner un réseau de neurones MLP qui prévoit la production d'une installation photovoltaïque installée à Amman, en Jordanie, avec une anticipation de 24 heures. L'ELM se

révèle plus performant que l'algorithme classique de rétro propagation du gradient (BP) en termes de précision. La technique a montré la plus faible MAE (erreur moyenne absolue) de 1,08% en juin, tandis que la plus grande MAE de 18,83% correspondait à février et mars.

Dans l'article [47], un modèle basé sur un réseau d'états récurrents à réservoirs multiples (MR-ESN) a été proposé. L'algorithme quasi-Newton a été utilisé pour optimiser les paramètres du réservoir. Les résultats ont montré un MAPE (erreur absolue moyenne en pourcentage) très proche de zéro (0,00195%). Par rapport à une prévision d'une heure à l'avance, le modèle a donné de meilleurs résultats que d'autres techniques telles que les réseaux de neurones à rétro propagation du gradient (BPNNs), la régression par vecteurs de support (SVR-ANN) et la transformation en ondelettes (WT).

Dans l'article [24], un modèle multi-modèle basé sur l'ELM a été proposé pour prévoir la production d'une installation photovoltaïque de 250 kWp installée à Beijing, en Chine. En ce qui concerne l'exactitude du multi-modèle, la MAE était de 2,13% au printemps et de 1,7% en été, tandis que pour un modèle unique annuel, la MAE était de 2,43% au printemps et de 1,81% en été. Le multi-modèle conçu tient compte des fluctuations de la production d'électricité afin d'améliorer la précision des prévisions.

III.3 Les données expérimentales : Données de mesures, rayonnement et variables météorologiques

Pour valider nos approches, nous avons effectué des prédictions de rayonnement global sur le site de l'Arabie saoudite. Nous avons essayé de construire un outil de prédiction adaptable. Les données que nous allons présenter sont des mesures météorologiques spécifiques pour la localité d'étude dites exogènes, c'est-à-dire de nature différente que la prédiction (le rayonnement global).

La région étudiée durant ce mémoire est l'Arabie saoudite: le site d'Al Médina (Latitude 24°28' N / Longitude 39°36' E / Altitude 594 m) . Cette région présente un climat désertique, caractérisé par une aridité constante, des températures souvent très chaudes en journée. Les mesures du rayonnement solaire considèrent des données journalières.

Le site d'AL Médina comporte une base de données qui a été gracieusement mise à notre disposition par le laboratoire national des énergies renouvelables (NREL).

La base de données comporte des mesures journalières du rayonnement solaire, la durée d'ensoleillement (S) en h, la température (T) en °C et l'humidité relative (RH) en %.

Tableau III.2 Données météorologiques disponibles.

Site	Variable exogène utilisées	Base	Période couverte
AL Medina	La durée d'insolation, la température et l'humidité.	Journalière	01/01/1998 31/12/2002

III.4 Evaluation des modèles

Dans le cadre d'estimation de l'irradiation solaire globale ou de la puissance PV, il existe de nombreuses méthodes pour vérifier si un estimateur ou prédicteur est efficace [3] [28] [37]. Durant ce mémoire, nous effectuons un test statistique pour la comparaison de toutes les approches présentées ou développées. Les outils que nous avons utilisés sont décrits ci-après :

- **le coefficient de détermination (R^2)** est un indicateur qui permet de juger la qualité d'une régression linéaire, simple ou multiple. D'une valeur comprise entre 0 et 1, il mesure l'adéquation entre le modèle et les données observées. Dans le cadre d'une régression linéaire simple, c'est le carré du coefficient de corrélation. Ce coefficient s'applique aussi bien à une régression simple qu'à une régression multiple.

$$R^2 = 1 - \left(\frac{\sum_{i=1}^N |G_{im} - G_{ie}|^2}{\sum_{i=1}^N G_{im}} \right) \quad (\text{III.1})$$

ou, N est le nombre de données, et notons que G_{im} et G_{ie} la $i^{\text{ème}}$ valeur mesurée et estimée de l'irradiation solaire.

Il est possible d'utiliser le coefficient de détermination ajusté car ce dernier tient compte du nombre de variables. Cependant, dans nos travaux comme le nombre de variables était très largement inférieur au nombre d'observation, il n'y avait pas grand intérêt à manipuler cette grandeur.

- **Le coefficient de corrélation (R)** est un coefficient statistique permettant de mettre en évidence, une liaison entre deux types de séries de données statistiques. Le coefficient de corrélation est compris entre 1 et -1 et on considère généralement que si sa valeur absolue est

supérieure à 0,95, il y a une liaison forte entre les deux séries. Si la valeur est proche de 1, les séries évoluent dans le même sens, si elle est proche de - 1, elles évoluent en sens opposé.

$$R = \frac{\sum_{i=1}^N (G_{im} - \bar{G}_{im})(G_{ie} - \bar{G}_{ie})}{\sqrt{\sum_{i=1}^N (G_{im} - \bar{G}_{im})^2} \sqrt{\sum_{i=1}^N (G_{ie} - \bar{G}_{ie})^2}} \quad (\text{III.2})$$

Avec : \bar{G}_{im} et \bar{G}_{ie} la $i^{\text{ème}}$ moyenne mesurée et estimée de l'irradiation solaire.

- **L'erreur quadratique moyenne (Root Mean Square Error; RMSE)**, est une mesure de l'amplitude moyenne de l'erreur et utilisée pour manipuler une erreur basée sur la même échelle que la mesure, donnée par :

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (G_{im} - G_{ie})^2}{N}} \quad (\text{III.3})$$

- **L'erreur moyenne absolue (Mean Absolute Error; MAE)**, est une quantité souvent utilisée pour mesurer l'écart des prédictions et des mesures. Sa formule est donnée par :

$$MAE = \frac{1}{N} \sum_{i=1}^N |G_{im} - G_{ie}| \quad (\text{III.4})$$

- **Le pourcentage moyen (Mean Pourcentage Error; MPE)**, est donné par :

$$MPE = \left(\frac{1}{N} \sum_{i=1}^N \frac{|G_{im} - G_{ie}|}{G_{im}} \right) \times 100 \quad (\text{III.5})$$

- **Le biais moyen (Mean Bias Error; MBE)**, est défini comme étant l'écart algébrique moyen entre la simulation et la mesure, sa formule est donnée comme suit:

$$MBE = \frac{\sum_{i=1}^N (G_{im} - G_{ie})}{N} \quad (\text{III.6})$$

Ce paramètre se rapproche plus de l'erreur sur la prédiction du cumul de la ressource que de l'erreur commise à horizon. En effet, une valeur positive de MBE signifie que sur le long terme, le cumul des données de prédiction sous-estime le cumul des mesures réelles alors qu'une valeur négative signifiera l'inverse.

III.5 Conclusion

Dans ce chapitre, nous avons mis une lumière sur quelques méthodes d'approches intelligentes existantes dans la littérature et destinées à l'estimation de l'irradiation solaire globale.

A travers l'analyse bibliographique dressée dans ce chapitre, une tendance semble se dégager: la synthèse de tous les travaux destinés à l'estimation de l'irradiation solaire qui se considère comme une série temporelle est assez délicate, car en règles générales, les approches, les estimations d'erreurs et les spécificités géographiques peuvent être très différentes.

La simplicité d'implémentation de ces approches nous, a poussé à penser à de nouvelles approches basées sur l'apprentissage automatique pour l'estimation de l'irradiation solaire.

Enfin, les critères de performances utilisés tout au long de ce mémoire sont aussi présentés dans ce chapitre .

Chapitre VI

Interprétation des résultats

IV.1 Introduction

Les méthodologies présentées dans les deux derniers chapitres précédents nous ont permis d'entreprendre une série de manipulations en vue de la prédiction à horizon de l'irradiation solaire globale.

Dans le chapitre III, nous avons défini les trois grandes familles de modèle de l'intelligence artificielle existants dans la littérature et qui régissent la dynamique solaire et plus particulièrement le rayonnement global au sol.

Ce chapitre fut aussi l'occasion de présenter le principe des méthodes de l'apprentissage automatique proposées dans ce mémoire et qui seront développées, testées et comparées au cours de ce quatrième chapitre.

Le chapitre III, a présenté les différents critères de performance qui seront par la suite employés pour tester la fiabilité des modèles proposés.

La prochaine section de ce chapitre est dédiée à l'étude et l'estimation de l'irradiation solaire journalière dont plusieurs approches intelligentes seront développées et testées par le biais d'outils de comparaison et d'évaluation d'erreurs.

IV.2 Modélisation et prédiction de l'irradiation solaire journalière:

La prédiction de l'irradiation solaire dans notre cas d'étude est basée sur l'utilisation de paramètres météorologiques et géographiques « dites variables exogènes » qui fait intervenir comme paramètres météorologiques et géographiques : la température, l'humidité relative, la durée d'insolation et l'irradiation solaire extraterrestre, dont trois familles de modèles seront développées et testés :

- ✓ La première famille de modèle est basée sur l'utilisation de la régression linéaire de l'apprentissage automatique, cette dernière est composée à son tour de 4 modèles de la régression linéaire à savoir : la régression linéaire simple, à interaction, pas à pas et robuste.
- ✓ La deuxième famille de modèles est basée sur l'utilisation des arbres de régression cette dernière est composée à son tour de 5 modèles à savoir : Fine tree ,médiu m tree, Coarse tree, arbre de décision à boosting de gradient et Arbre de décision à bagging.
- ✓ La troisième famille de modèles est basée sur l'utilisation des arbres de régression cette dernière est composée à son tour de 5 modèles à savoir : Fine tree ,médiu m tree, Coarse tree, arbre de décision à boosting de gradient et Arbre de décision à bagging.

Les trois familles de modèles seront testées pour une base de données comportant 5 années de mesures.

Les mesures journalières de l'irradiation solaire globale, la durée d'insolation, l'humidité relative et la température ont été récupérées auprès du laboratoire national des énergies renouvelables (National Laboratory of Renewable Energy ; NREL) de la station d'AL-Médina pour les années 1998-2002.

Deux étapes essentielles sont nécessaires pour le phénomène de modélisation : l'étape de développement « d'apprentissage » et l'étape d'évaluation « de test ». Pour cette raison la base de données est divisée en deux sous ensembles de données, l'un est utilisé pour le développement des modèles, et l'autre pour l'évaluation de ces derniers. Dans notre contexte d'étude les quatre premières années de mesures (1988-2001) seront utilisées pour le développement des modèles quand à une année de mesure (2001-2002) sera utilisée pour l'évaluation et le test des performances.

La Figure IV.1 présente la distribution de la moyenne mensuelle effective de la durée d'ensoleillement pour une période de scinque années pour le site d'AL-Médina. Elle est de plus de 8 heures par jour pendant les mois d'hiver et de plus de 10 heures par jour pendant les mois de juin à juillet. Le site est donc très favorable à l'utilisation de l'énergie solaire.

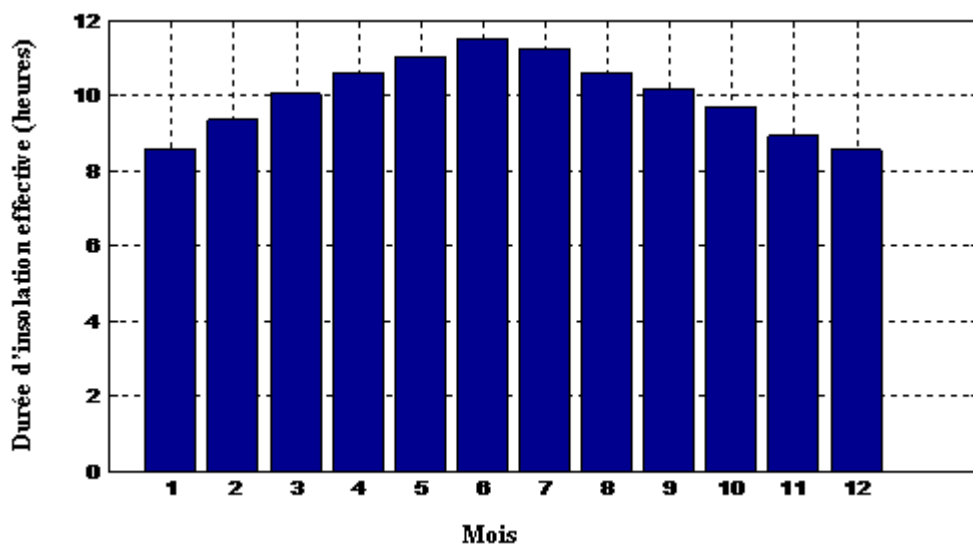


Figure IV.1 Distribution de la moyenne mensuelle effective de la durée d'insolation pour le site d'AL-Médina.

Les modèles de l'apprentissage automatique ont été développés sous MATLAB 18, en utilisant les mesures respectives de l'irradiation solaire globale, la durée d'insolation, l'humidité relative et la température.

Pour évaluer les résultats obtenus par les différents modèles, nous avons calculé les écarts absolu (MAE), et quadratique (RMSE) ainsi que le coefficient de détermination R^2 .

Le Tableau IV.2 résume les différents modèles développés ainsi que les indicateurs statistiques de précision MAE, RMSE et R^2 pour les 4 premières années d'apprentissage.

Modèles	Type de modèle	Critères de performances (Apprentissage)		
		RMASE %	MAE%	R^2
Régression linéaire (RL)	Modèle 1: Linaire	5.01	3.48	0.88
	Modèle 2: a interactions	4.44	3.09	0.91
	Modèle 3: Pas a pas	4.42	3.09	0.91
	Modèle 4: Robuste	5.48	3.18	0.86
Régression par l'arbre de décision (AD)	Modèle 1: Fine tree	4.56	3.14	0.90
	Modèle 2: Medium tree	4.24	2.93	0.91
	Modèle 3: Coarse tree	4.72	3.16	0.89
	Modèle 4: Arbre de décision a boosting de gradient	4.79	3.87	0.89
	Modèle 5: Arbre de décision a bagged	4.48	3.13	0.90
Régression par machine a vecteurs de supports (MVS)	Modèle 1: Linéaire MVS	5.27	3.20	0.87
	Modèle 2: Quadratique MVS	3.95	2.67	0.92
	Modèle 3: Cubique MVS	3.89	2.64	0.93

Tableau IV.2 Indicateurs statistiques de précision pour les trois familles de modèles développés.

Il est à noter que les valeurs estimées par les différents modèles sont en bonnes concordances avec les valeurs mesurées, cependant les meilleurs résultats sont obtenus par les modèles :

- ✓ De la régression linéaire type pas a pas (Modèle 3) avec un RMSE de 4.42%, un MAE de 3.09 % et un R^2 de 0.91.

- ✓ De l'arbre de décision type médium tree (**modèle 2**) avec un RMSE de 4.24%, un MAE de 2.93 % et un R^2 de 0.91.
- ✓ De l'arbre de décision type bagging (**modèle 5**) avec un RMSE de 3.89%, un MAE de 3.13 % et un R^2 de 0.90.

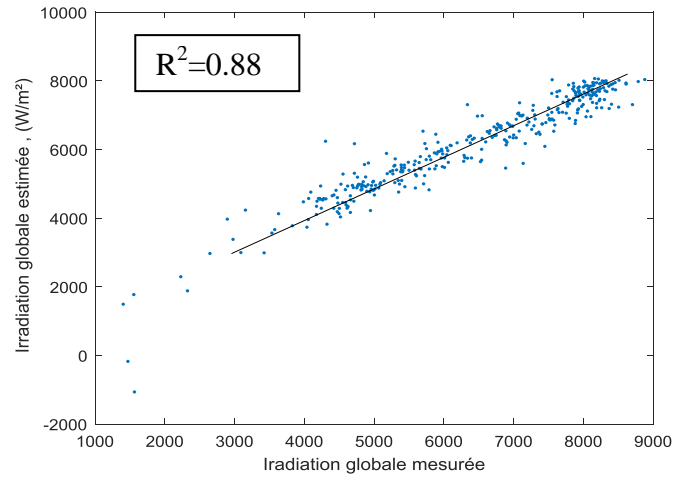
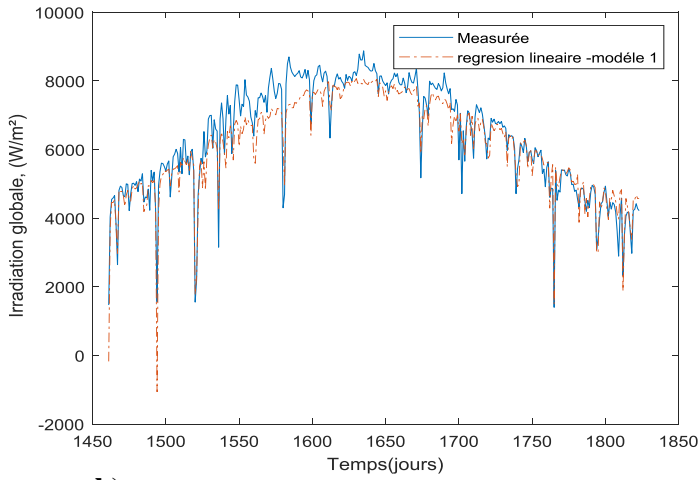
La troisième partie du Tableau IV.2 reflète la performance des différents types d'estimateurs étudiés par machine à vecteurs supports (MVS). Les différentes erreurs d'estimation montrent que le (MVS-cubique) est le meilleur modèle dans notre étude comparativement aux autres estimateurs de même type SVM (linéaire et quadratique).

En effet, c'est le modèle qui a abouti aux plus faibles erreurs (une RMSE estimée à 389%, , MAE égale à 2.64%) et au plus grand coefficient de corrélation estimé à 0.93, les MVS de type cubique et quadratique sont acceptables avec un coefficient de corrélation qui varie entre 0.92 et 0.93 respectivement. On peut conclure que, dans ce contexte le choix du noyau s'avère primordial.

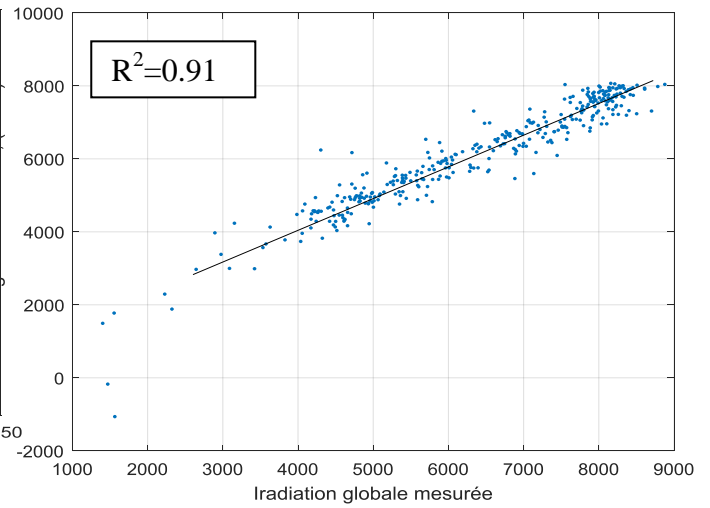
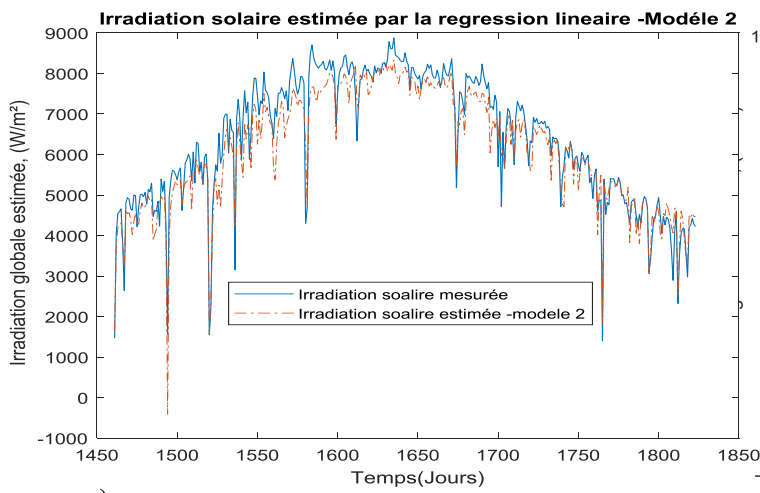
La Figure IV.2 a), b) et c) présente les trois modèles développés : Modèle 1, Modèle 2 et modèle 3 de la régression linéaire . Une bonne concordance entre l'irradiation solaire globale estimée et mesurée est très remarquable pour les trois différents modèles avec un coefficient de détermination égale à 0.88 ,0.91 et 0.91 respectivement .

La Figure IV.3, présente le modèle de la régression linéaire. Il est clair qu'une bonne concordance entre l'irradiation solaire globale estimée et mesurée est très remarquable pour ce modèle de la régression linéaire type robuste avec un coefficient de détermination acceptable égale à 0.86.

a)



b)



c)

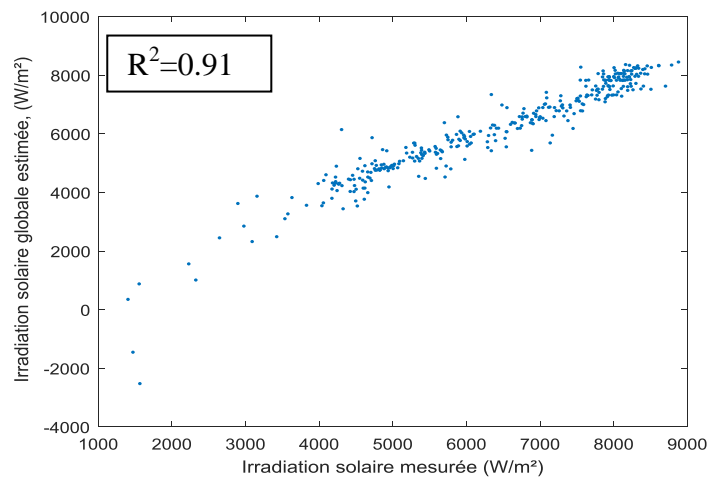
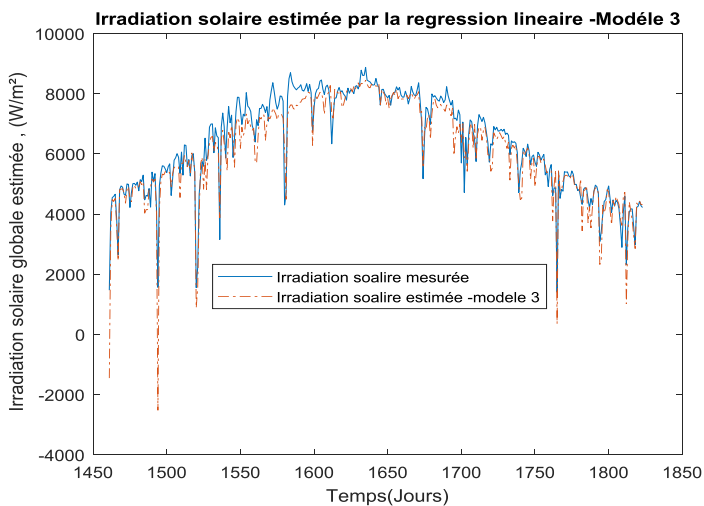


Figure IV.2 L'irradiation solaire globale mesurée et estimée pour la phase de test pour les modèles de la régression linéaire a) « *modèle 1* » b) « *modèle 2* » et c) « *modèle 3* ».

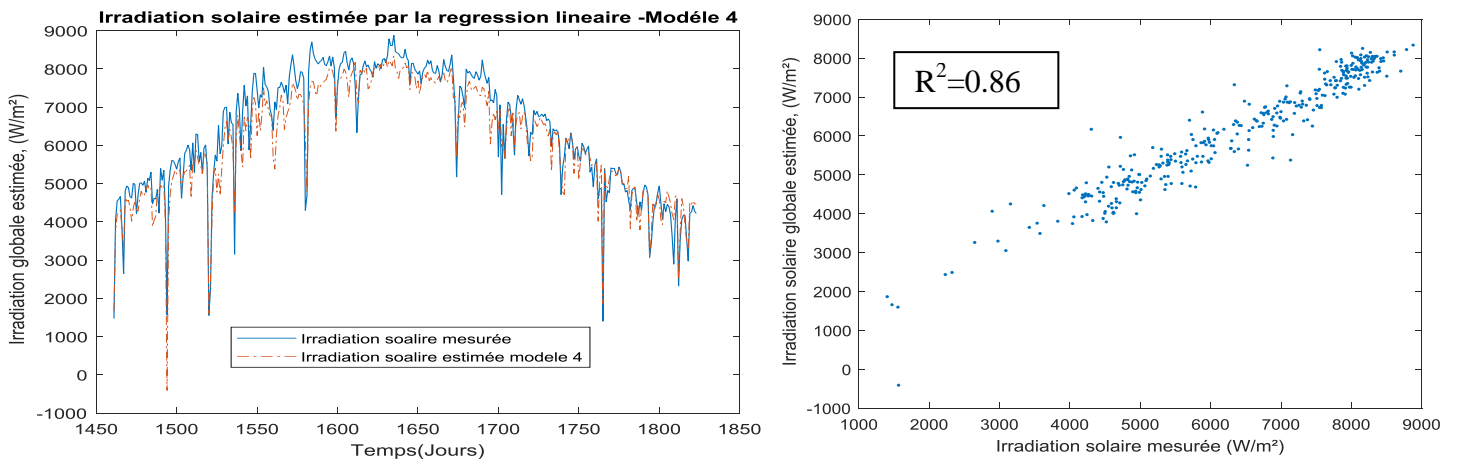


Figure IV.3 L'irradiation solaire globale mesurée et estimée pour la phase de test pour le modèle 4 de la régression linéaire.

La Figure IV.4 a), b) et c) présente les trois modèles développés : Modèle 1, Modèle 2 et modèle 3 de la régression par l'arbre de décision. Une bonne concordance entre l'irradiation solaire globale estimée et mesurée est très remarquable pour les trois différents modèles avec un coefficient de détermination égale à 0.90, 0.91 et 0.89 respectivement. Un coefficient de détermination très proche entre les trois modèles proposés.

La Figure IV.5, présente le modèle 4 (à boosting de gradient) de la régression par l'arbre de décision. Il est clair qu'une bonne concordance entre l'irradiation solaire globale estimée et mesurée est très remarquable avec un coefficient de détermination acceptable égale à 0.89.

La Figure IV.6, présente le modèle 4 (à bagged) de la régression par l'arbre de décision. Il est clair qu'une bonne concordance entre l'irradiation solaire globale estimée et mesurée est très remarquable avec un coefficient de détermination acceptable égale à 0.90.

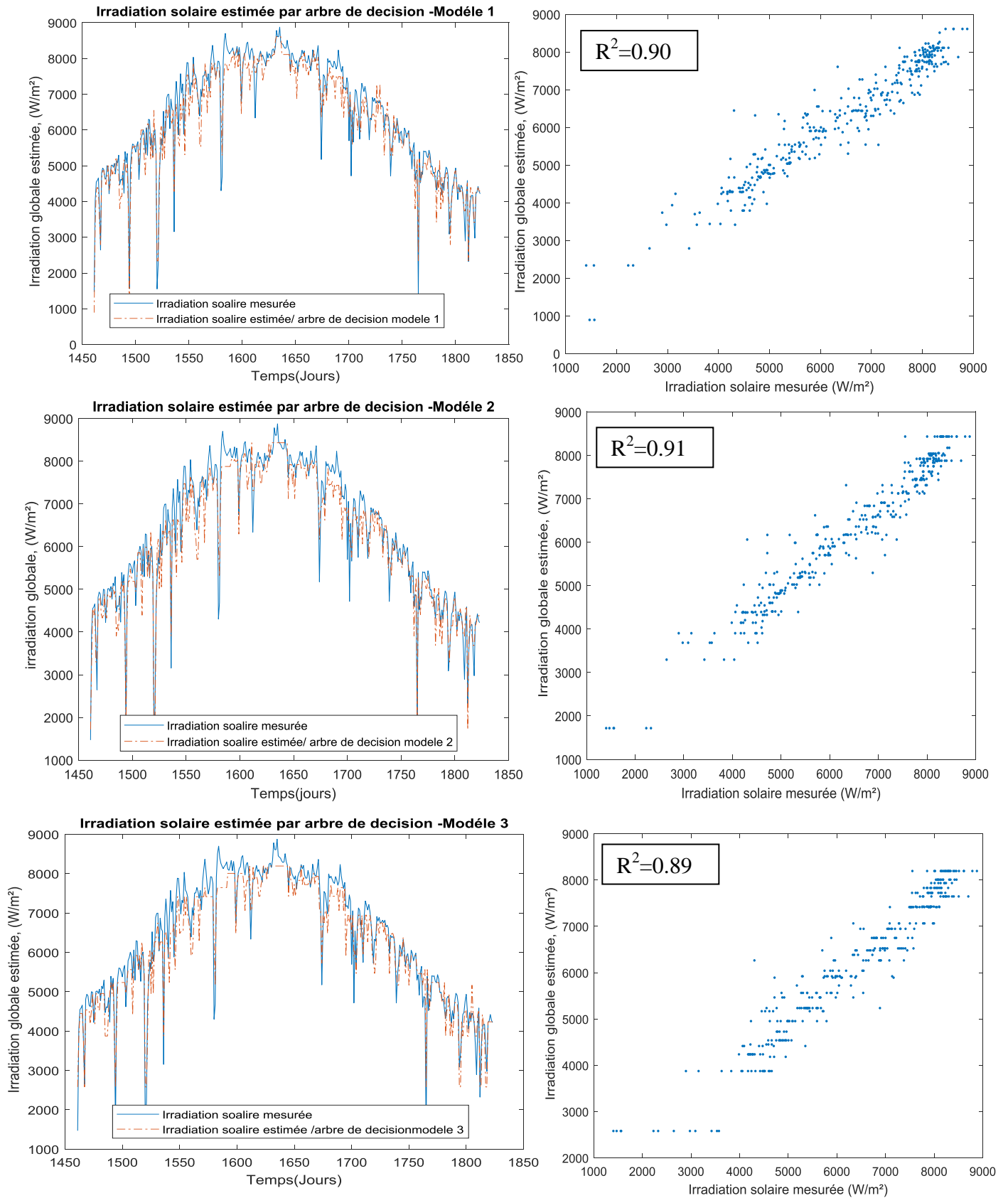


Figure IV.4 L'irradiation solaire globale mesurée et estimée pour la phase de test pour les modèles de la régression par l'arbre de décision a) « *modèle 1* », b) « *modèle 2* » et c) « *modèle 3* ».

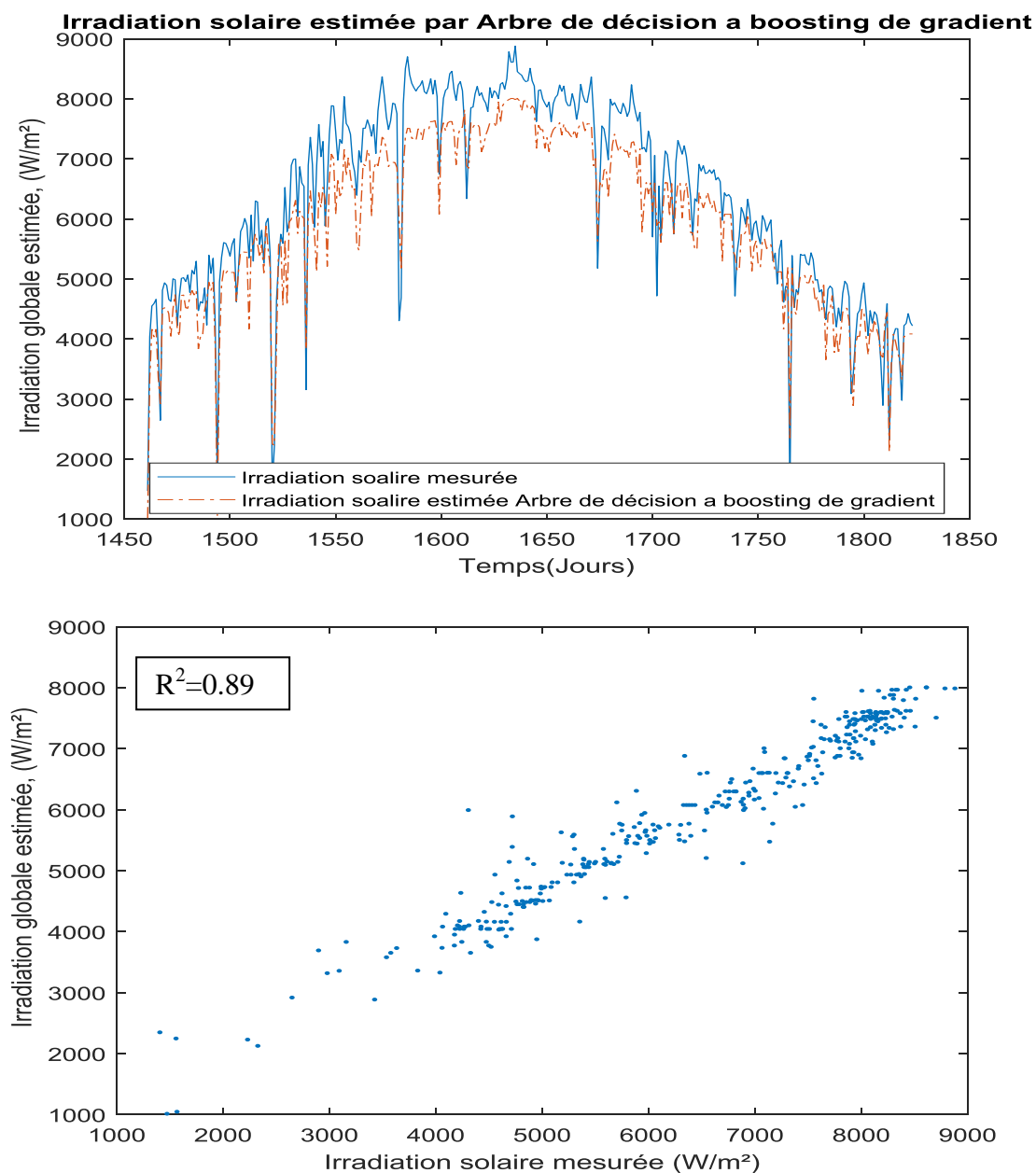


Figure IV.5 L'irradiation solaire globale mesurée et estimée pour la phase de test pour la régression par l'arbre de décision Type « boosting de gradient - modèle 4 ».

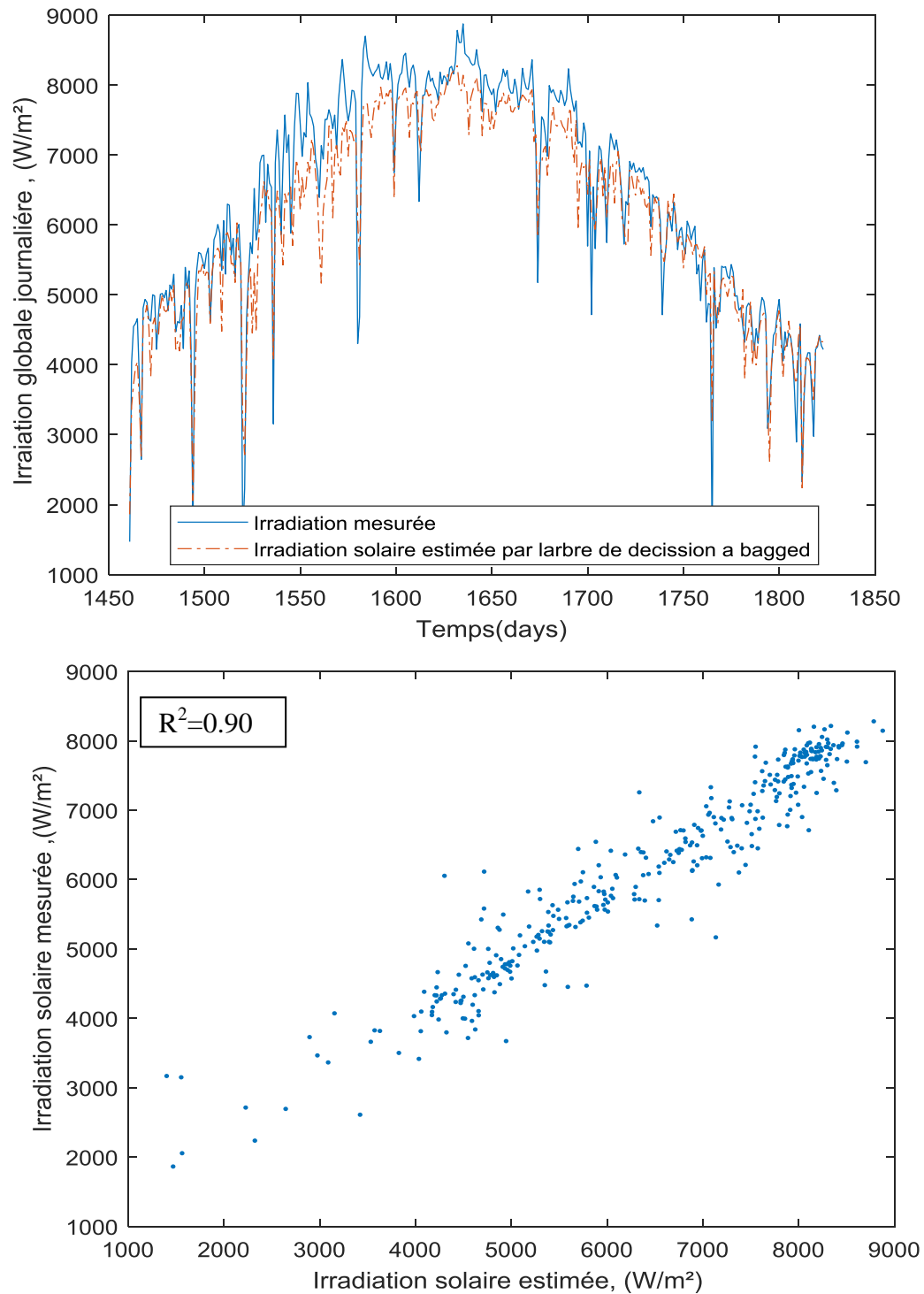


Figure IV.6 L'irradiation solaire globale mesurée et estimée pour la phase de test pour la régression par l'arbre de décision Type « a bagged - modèle 5 ».

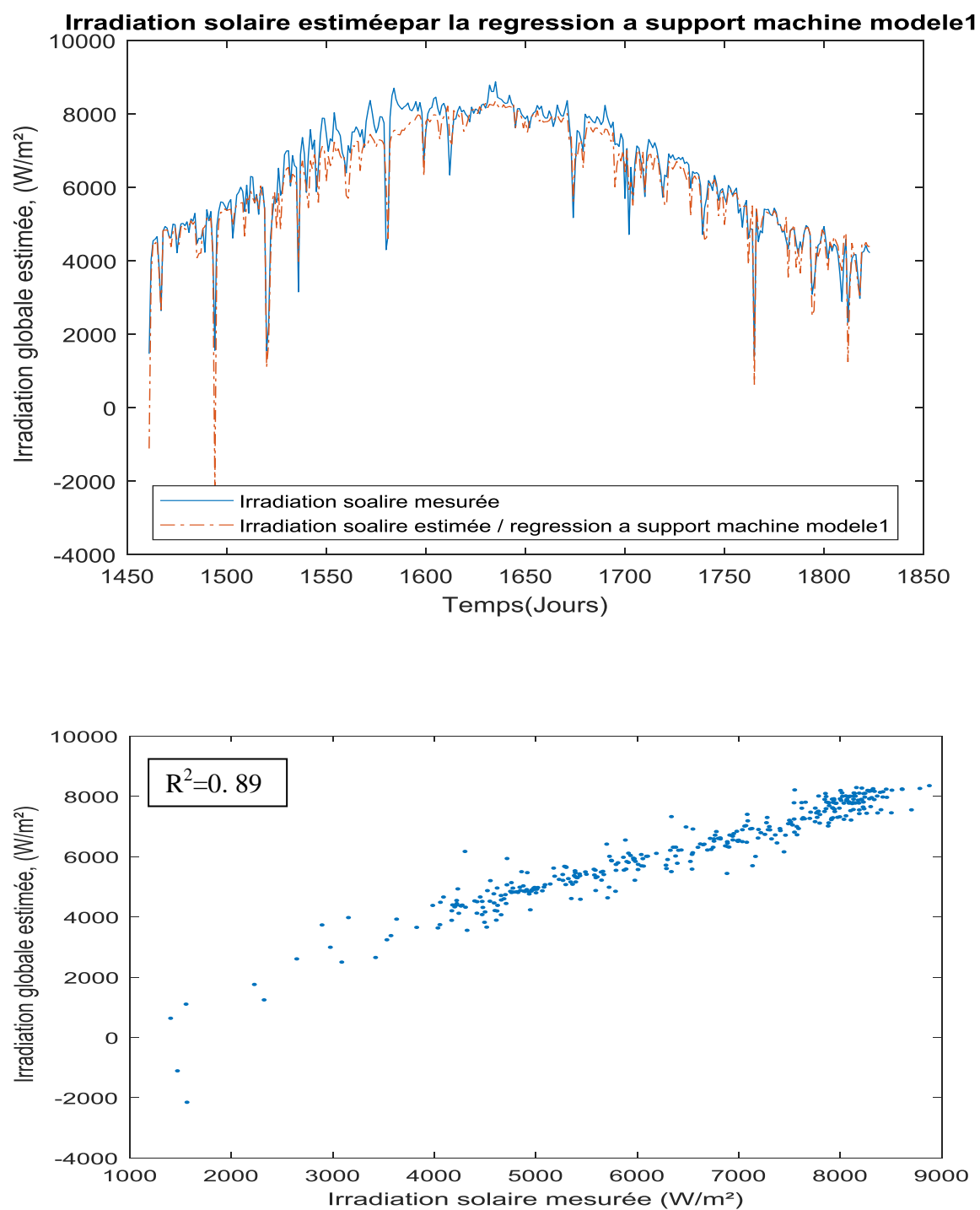


Figure IV.7 L'irradiation solaire globale mesurée et estimée pour la phase de test pour la régression par le machine a vecteur de support «**modèle 1**».

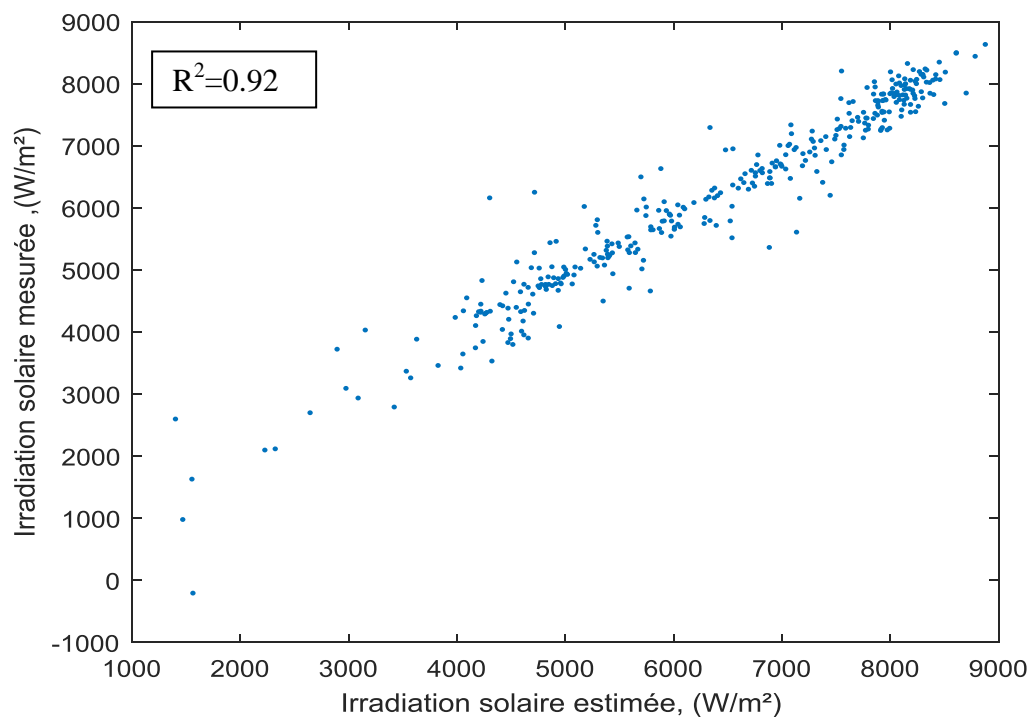
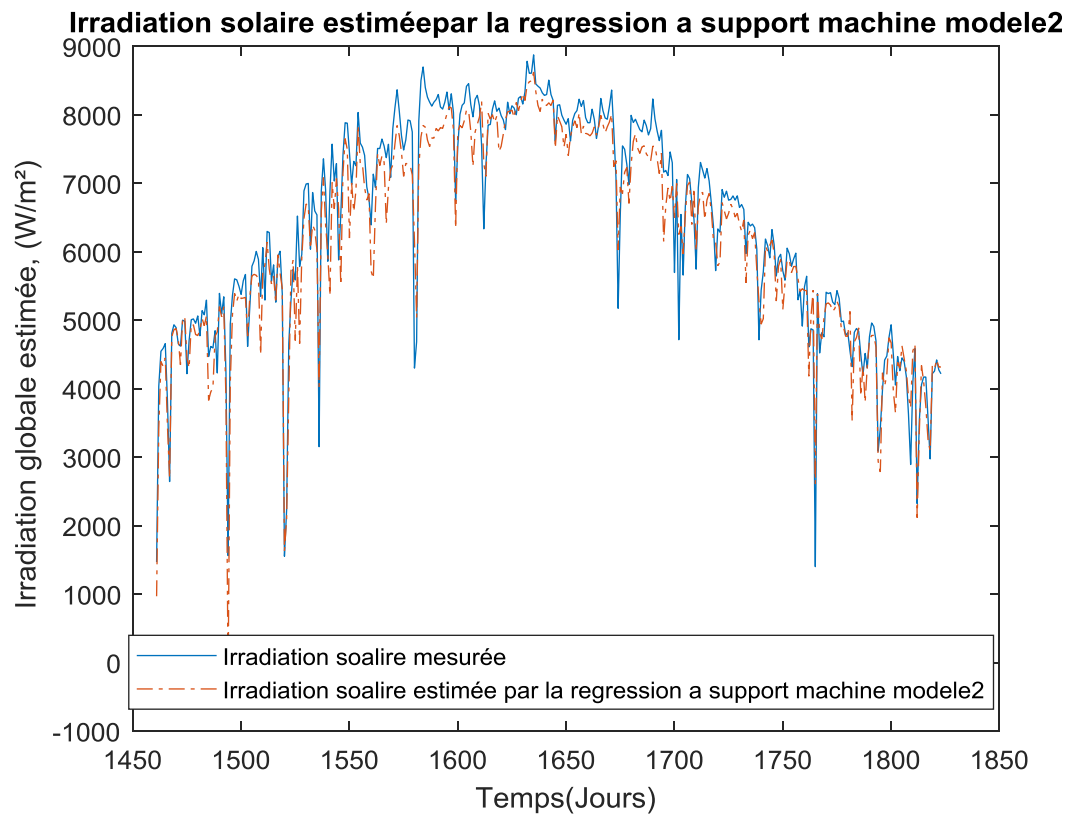


Figure IV.8 L'irradiation solaire globale mesurée et estimée pour la phase de test pour la régression par le machine a vecteur de support «**modèle 2**».

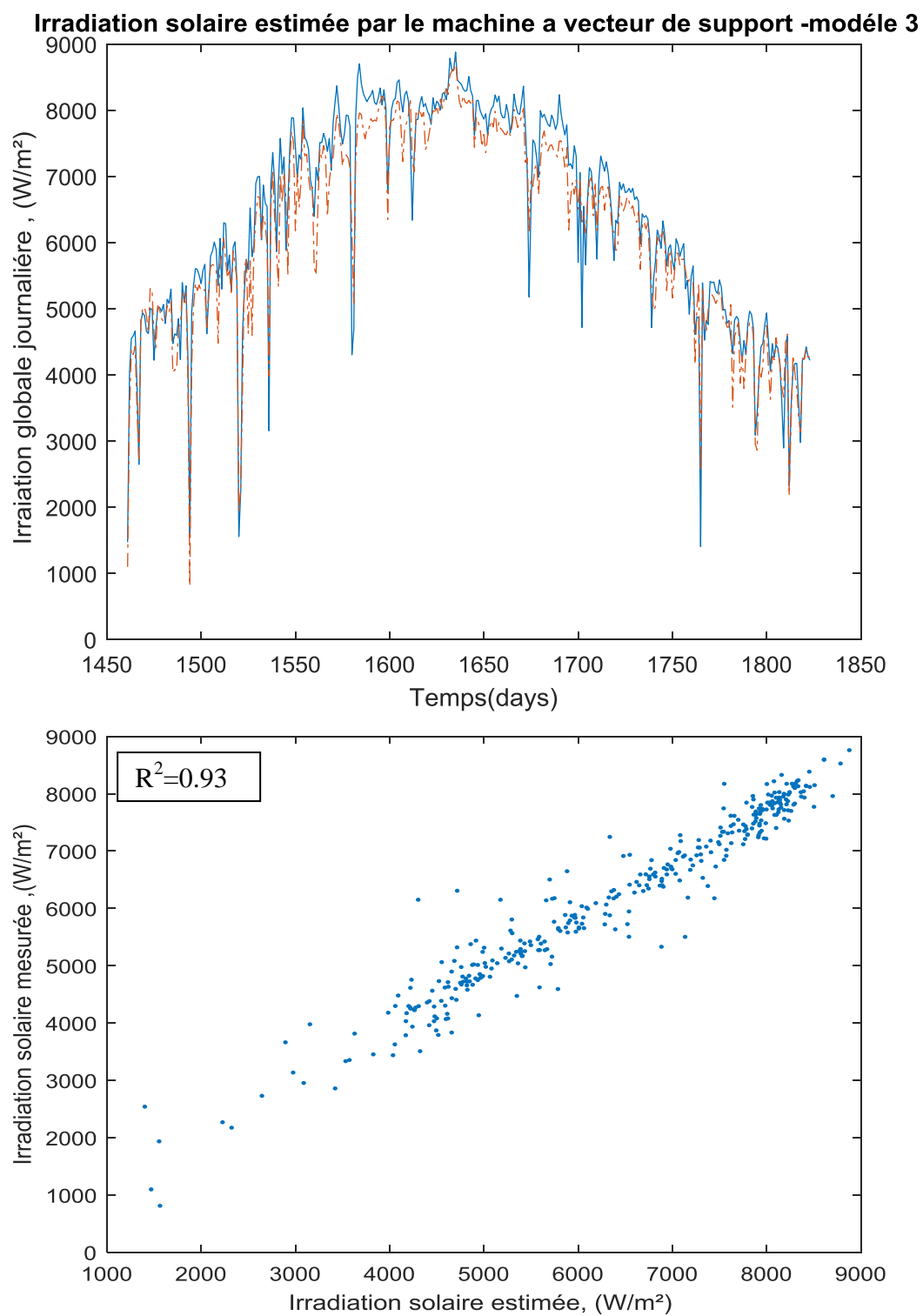


Figure IV.9 L'irradiation solaire globale mesurée et estimée pour la phase de test pour la régression par le machine a vecteur de support «**modèle 3**».

La Figure IV.7, présente le modèle 1 de la régression par la machine vectrice de support (MVS). Il est clair qu'une bonne concordance entre l'irradiation solaire globale estimée et mesurée est très remarquable avec un coefficient de détermination acceptable égale à 0.87.

La Figure IV.8, présente le modèle 2 de la régression par la machine vectrice de support (MVS). Il est clair qu'une bonne concordance entre l'irradiation solaire globale estimée et mesurée est très remarquable avec un coefficient de détermination acceptable égale à 0.92.

La Figure IV.9, présente le modèle 3 de la régression par la machine vectrice de support (MVS). Il est clair qu'une bonne concordance entre l'irradiation solaire globale estimée et mesurée est très remarquable par rapport aux modèles de la même famille avec un coefficient de détermination très acceptable égale à 0.93.

IV.3 Conclusion

Dans ce chapitre, nous venons de présenter les modèles et les résultats obtenus dans le cas de la prédiction de l'irradiation globale horaire et journalière par différents modèles de l'apprentissage automatique présentés dans le chapitre III. La méthodologie de prédiction la plus sophistiquée, est certes assez lourde à mettre en place, mais donne des résultats largement supérieurs à ceux issus des autres techniques testées. Les outils de validation des processus de modélisation ont été mis au point. Leurs études ont révélé qu'une très bonne concordance entre l'irradiation solaire mesurée et estimée pour les trois grandes familles de modèles proposés (LR, AD et MVS), néanmoins il est très clair que les méthodes à noyaux (MSV) donnent une meilleure estimation par rapport à d'autres modèles proposés.

Le choix du type de MSV reste toujours primordial pour la prédiction de l'irradiation solaire (le MSV cubique donne une meilleure estimation rapport au MSV quadratique) avec un coefficient de détermination égale 0.93.

Ce travail de mémoire de master s'inscrit dans la cadre de la modélisation et la prédiction de l'irradiation solaire globale. Les résultats obtenus dans le contexte de ce travail ont ouvert des perspectives novatrices sur la prédiction de l'irradiation solaire à partir de données météorologiques et géographiques. Avant de présenter quelques perspectives, nous allons tirer les principales conclusions de cette étude.

Pour valider nos approches proposées, nous avons effectué des prédictions de l'irradiation solaire globale en fonction de trois principaux paramètres météorologiques à savoir : la température, l'humidité relative et la durée d'insolation pour le site de l'Arabie saoudite (Al -Medinah).

Notre méthodologie d'étude est basée sur l'étude de l'horizon journalier, pour la prédiction de l'irradiation solaire dont des modèles intelligents basés essentiellement sur l'apprentissage automatique ont été développés à savoir : les arbres de régression, la régression linéaire et le machine à vecteurs supports dont plusieurs paramètres météorologiques et géographiques ont été intervenus.

Afin de pouvoir arriver à réaliser l'objectif principale de notre travail, qui est l'obtention des modèles qui rassemble à la fois la simplicité de mise en oeuvre et l'efficacité en terme d'estimation, nous avons effectué une comparaison de plusieurs modèles appartenant à la même famille , commençant par la régression linéaire dont quatre modèles ont été développés et testés ,la deuxième famille de modèles et les arbres de décision dont cinq modèles ont été développés et testés ,et enfin la famille de machine à vecteurs de supports dont trois modèles ont été développés et testés .

Tous les modèles développés ont été testés à l'aide des différents critères de performances présentés par différentes erreurs RMSE et MAE et le coefficient de détermination R^2 et comparés pour les deux phases d'apprentissage et de test.

Il est à noter que les valeurs estimées par les différents modèles sont en bonnes concordances avec les valeurs mesurées, cependant les meilleurs résultats sont obtenus par les modèles :

- ✓ De la régression linéaire type pas à pas (Modèle 3) avec un RMSE de 4.42%, un MAE de 3.09 % et un R^2 de 0.91.
- ✓ De l'arbre de décision type médium tree (modèle 2) avec un RMSE de 4.24%, un MAE de 2.93 % et un R^2 de 0.91.
- ✓ De l'arbre de décision type bagged (**modèle 5**) avec un RMSE de 3.89%, un MAE de 3.13 % et un R^2 de 0.90.

- ✓ Pour les trois estimateurs étudiés par machine à vecteurs de supports (MVS). Les différentes erreurs d'estimation montrent que le (MVS-cubique) est le meilleur modèle dans notre étude comparativement aux autres estimateurs de même type SVM (linéaire et quadratique). En effet, c'est le modèle qui a abouti aux plus faibles erreurs (une RMSE estimée à 389%, , MAE égale à 2.64%) et au plus grand coefficient de corrélation estimé à 0.93, les MVS de type cubique et quadratique donne une estimation très acceptables avec un coefficient de corrélation qui vari entre 0.92 et 0.93 respectivement. On peut conclure que, dans ce contexte le choix du noyau s'avère primordial.

A l'issus de ce travail de mémoire, plusieurs perspectives se dégagent :

- Plusieurs méthodes basées sur l'apprentissage automatiques et profond sont intéressantes a développées telles que les forets aléatoires, algorithme K-NN (k-Nearest Neighbors), le naïve Bayes, les réseaux neuronaux convolutifs (CNN), les réseaux neuronaux récurrents (RNN), les réseaux de mémoire à long et court terme (LSTM) ... ect.
- De même développé un estimateur intelligent pour la région du sud ,et exceptionnellement le sarah algérienne qui présente l'un des plus grand gisement solaire au monde , et donc élargissement de régions d'étude et de base de données de même climat que celui d'Al -Medinah offre l'opportunité d'avoir un estimateur régional intelligent de l'irradiation solaire globale destiné au climat désertiques et le développé sur carte FPGA.
- Prédiction de la puissance photovoltaïque PV à partir de l'irradiation solaire prédite (de l'irradiation solaire à la production photovoltaïque).

Bibliographie

- [1] **Ağbulut U, Gürel A.E , Biçen Y** .Prediction of daily global solar radiation using different machine learning algorithms: evaluation and comparison .Renew. Sustain. Energy Rev., Vol 135 ,2021.
- [2] **Aguilar R, Fraidenraich N** . Analysis of a new relationship between monthly global solar irradiation and sunshine hours from a database of Brazil.Renewable Energy Vol.30, 957, 2005.
- [3] **Ahlburg D** .Error measures and the choice of a forecast method. International Journal of forecasting n°6;8(1), page 99-100, 1992.
- [4] **Alan Turing**. Computing machinery and intelligence, Mind (en), Oxford University Press, vol. 59, no 236, page 433–460, octobre 1950
- [5] **Al-Dahidi S, Ayadi O, Adeeb J, Alrbai M, Qawasmeh B**. Extreme learning machines for solar photovoltaic power predictions. Energies .vol. 11, page 2725. 2019
- [6] **Badescu V**. Modeling Solar Radiation at the Earth's Surface: recent advances. Springer, 2008.
- [7] **Baharin KA, Abdul Rahman H, Hassan MY, Gan CK**. Short-term forecasting of solar photovoltaic output power for tropical climate using ground-based measurement data. J Renew Sustain Energy ,vol.8. 2016.
- [8] **Bayrakç H.C, Demircan C, Keçebaş A**.The development of empirical models for estimating global solar radiation on horizontal surface: a case study, Renew. Sustain. Energy Rev. Vol.81 2771–2782, 2018.
- [9] **Belaid S, Mellit A, Boualit H , Zaiani M** . Hourly global solar forecasting models based on supervised machine learning algorithm and time series principle. International journal of ambient energy .Janvier 2020.
- [10] **Benghanem M , Joraid AA**.A multiple correlation between different solar parameters in Medina, Saudi Arabia. Renewable Energy Vol.32, page 2424-2435, 2007.
- [11] **Benghanem M ,Mellit M**. A simplified calibrated model for estimating daily global solar radiation in Madinah, Saudi Arabia. Springer-Verlag Wien, DOI 10.1007/s00704-013-0884-2,2013.
- [12] **Bessemoulin P, Oliviéri J**. Le rayonnement solaire et sa composante ultraviolette. La Météorologie 8^e série, n° 31, 2000.

- [13] **Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J.** Classification and regression trees, 1984.
- [14] **Bhandari B.** Hybrid Convolutional Neural Network-Multilayer Perceptron Model for Solar Radiation Prediction. Cognitive Computation, Springer 2022.
- [15] **Daniel Crevier et Nathalie Bukcek** (trad. de l'anglais), À la recherche de l'intelligence artificielle, Paris, Flammarion, 438 p. (ISBN 2-08-081428-1), 1997.
- [16] **De'ath, ,** Boosted Trees for Ecological Modeling and Prediction. Ecology Vol.88, page 243-251,2007.
- [17] **Drif M .** Contribution au développement de méthodes d'évaluation des pertes d'énergie inhérentes aux générateurs photovoltaïques partiellement ombragés.Thèse de doctorat, université de biskra, septembre 2022.
- [18] **EL-Sebaili AA, Trabea AA.** Estimation of global solar radiation on horizontal surfaces over Egypt, Egypt J. Solids, Vol.28, p 163- 175, 2005.
- [19] **Feng Y, Gong D, Zhang Q, Jiang S, Zhao L, Cui N.** Evaluation of temperature-based machine learning and empirical models for predicting daily global solar radiation, Energy Convers and Manag. Vol. 198 ,2019.
- [20] **Freund, Y., Schapire, R.E.,** A Short Introduction to Boosting Vol.14, page 771 - 780,1999
- [21] **Gigoni L, Betti A, Crisostomi E, Franco A, Tucci M, Bizzarri F, Mucci D .**Day-ahead hourly forecasting of power generation from photovoltaic plants. IEEE Trans Sustain Energy Vol. 9, page 831–842,2015.
- [22] **Guermoui M, Melgani F, Gairaa K, Mekhalfi M.L .** A comprehensive review of hybrid models for solar radiation forecasting J. Clean. Prod.,vol 258 ,2020.
- [23] **Guermoui M, Melgani F, Danilo C .**Multi-step ahead forecasting of daily global and direct solar radiation: a review and case study of Ghardaia region J. Clean. Prod., Vol. 201, page 716-734, 2018.
- [24] **Han Y, Wang N, Ma M, Zhou H, Dai S, Zhu H** A PV power interval forecasting based on seasonal model and nonparametric estimation algorithm. Sol Energy vol 184, page 515–526 2019.
- [25] **Hastie, T., Tibshirani, R.,** Generalized additive models. Stat. Sci. 1, vol. 297, page 318.1986
- [26] **Judith Hurwitz and Daniel Kirsch,** Machine Learning, IBM Limited Edition, 2018.

- [27] **Inman RH, Pedro HTC, Coimbra CFM.** Solar forecasting methods for renewable energy integration. *Progress in Energy and Combustion Science*, article in press, page 1-42, 2013.
- [28] **Kalogirou S.** Artificial neural networks in renewable energy systems applications: a review. *Renewable and Sustainable Energy Reviews*, n°12; 5(4), page 373-401, 2001.
- [29] **Kesh B Pun.** Short term forecasting of solar power with machine learning and time series techniques. PhD thesis. Wichita State University, 2020.
- [30] **Labouret A, Cumunel P, Braun JP, Faraggi B.** Cellules solaires Les bases de l'énergie photovoltaïque. Edition Technique et Scientifique Française « ETSF » ,5^{ème} édition, paris, 2010.
- [31] **Lauret P, Voyant C, Soubdhan T, David, M, Poggi P.** A benchmarking of machine learning techniques for solar radiation forecasting in an insular context. *Sol. Energy* ,Vol.112, page 446-457,2015.
- [32] **Leva S, Dolara A, Grimaccia F, Mussetta M, Ogliari E.** Analysis and validation of 24 hours ahead neural network forecasting of photovoltaic output power. *Math Comput Simul* vol.131, page 88–100,2017.
- [33] **Li H, Ma W, Lian Y, Wang X.** Estimating daily global solar radiation by day of year in China. *Applied Energy* Vol.87, p 3011-3017, 2010.
- [34] **Liu F, Li R, Li Y, Yan R, Saha T.** Takagi-Sugeno fuzzy model-based approach considering multiple weather factors for the photovoltaic power short-term forecasting. *IET Renew Power Gener* vol 11, page 1281–1287,2017.
- [35] **Liu J, Fang W, Zhang X, Yang C.** An improved photovoltaic power forecasting model with the assistance of aerosol index data. *IEEE Trans Sustain Energy* vol 6, page 434–442,2015.
- [36] **Liu W, Liu C, Lin Y, Ma L, Xiong F, Li J.** Ultra-short-term forecast of photovoltaic Output power under fog and haze weather. *Energies* vol11, page 528,2018.
- [37] **Mellit A, Kalogirou SA, Shaari S, Salhi H, Hadj Arab A.** 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,n°7,page 1570-1590, 2008.
- [38] **Menges HO, Ertekin C, Sonmete MH.** Evaluation of solar radiation models for Konya, Turkey, *Energy Conversion and Management*, Vol. 47, page 3149-3173, 2006.

- [39] **Obiora C.N, Hasan A.N, Ali A, Alajarmeh N.** Forecasting hourly solar radiation using artificial intelligence techniques, *IEEE Can. J. Electr. Comput. Eng.* Vol 44, page 497–508,2021.
- [40] **Oussama Bouguerra and Oussama Benslimane.** Solar radiation prediction using machine learning. Thèse de doctorat, 2020.
- [41] **Pierro M, De FeliceM,Maggioni E, Moser D, Perotto A, Spada F,CornaroC .** Data-driven upscaling methods for regional photovoltaic power estimation and forecast using satellite and numerical weather prediction data. *Sol Energy* vol 158,page 1026–1038,2017.
- [42] **Souhaila Chahboun and Mohamed Maaroufi.** “Performance Comparison of Support Vector Regression, Random Forest and Multiple Linear Regression to Forecast the Power of Photovoltaic Panels”. In: 2021 9th International Renewable and Sustainable Energy Conference (IRSEC). IEEE, page 1–4,2021.
- [43] **Stuart J. Russell et Peter Norvig,** Intelligence artificielle 3^{ème} édition. Artificial Intelligence : A Modern Approach ,Pearson Education France, page 1198,2018.
- [44] **Tiba C, de Troncoso, A., Salcedo-Sanz, S., Casanova-Mateo, C., Riquelme, J.C., Prieto, L.** Local 2015.
- [45] **Wang L, Kisi O, Zounemat-Kermani M, Salazar G.A, Zhu Z, Gong W.** Solar radiation prediction using different techniques: model evaluation and comparison *Renew. Sustain. Energy Rev.*, Vol ,61 page 384-397, 2016.
- [46] **Yacef R, Mellit A, Belaid S, Şen Z.**New combined models for estimating daily global solar radiation from measured air temperature in semi-arid climates: application in Ghardaïa, Algeria, *Energy Convers. Manag.* Vol.79, page 606–615,2014.
- [47] **Yao X,Wang Z, Zhang H .**A novel photovoltaic power forecasting model based on echo state network. *Neurocomputing* vol 325, page 182–189,2019.
- [48] **Zhang Y, Beaudin M, Taheri R, Zareipour H, Wood D .**Day-ahead power output forecasting for small-scale solar photovoltaic electricity generators. *IEEE Trans Smart Grid* Vol.6, page 2253–2262,2015.

Résumé:

Ce mémoire de fin d'études propose une approche novatrice pour la prévision de l'irradiation solaire en utilisant des méthodes d'apprentissage automatique. L'irradiation solaire joue un rôle essentiel dans de nombreuses applications, telles que la production d'énergie solaire, la planification urbaine et l'agriculture de précision. Cependant, la prédiction précise de l'irradiation solaire est un défi complexe en raison de la variabilité des conditions météorologiques et des caractéristiques géographiques. Dans ce mémoire, nous explorons l'utilisation de techniques d'apprentissage automatique, telles que la régression linéaire, les arbres de décision et les machines à vecteurs de support, pour modéliser et prédire l'irradiation solaire. Nous avons utilisé des données météorologiques exogènes pour la prédiction de l'irradiation solaire globale journalière. Les outils de validation des processus de modélisation ont été mis au point. Leurs études ont révélé qu'une très bonne concordance entre l'irradiation solaire mesurée et estimée pour les trois grandes familles de modèles proposés (LR, AD et MVS), néanmoins il est très clair que les méthodes à noyau (MSV) donnent une meilleure estimation par rapport à d'autres modèles proposés.

Le choix du type de MSV reste toujours primordial pour la prédiction de l'irradiation solaire (le MSV cubique donne une meilleure estimation rapport au MSV quadratique) avec un coefficient de détermination égale 0.93.

Mots clés : Irradiation solaire, apprentissage automatique, machine à vecteur de support, la régression linéaire, arbre de décision.

Abstract:

This dissertation proposes an innovative approach for solar irradiation forecasting using machine learning methods. Solar irradiation plays a crucial role in various applications, such as solar energy production, urban planning, and precision agriculture. However, accurately predicting solar irradiation is a complex challenge due to the variability of weather conditions and geographical features. In this study, we explore the use of machine learning techniques, such as linear regression, decision trees, and support vector machines, to model and predict solar irradiation. We utilized exogenous weather data for daily global solar irradiation prediction. The modeling process validation tools were developed, and their analyses revealed a strong agreement between the measured and estimated solar irradiation for the three main proposed model families (LR, DT, and SVM). However, it is evident that the kernel-based methods (SVM) provide better estimation compared to other proposed models. The choice of SVM type remains crucial for solar irradiation prediction (cubic SVM performs better than quadratic SVM) with a coefficient of determination equal to 0.93.

Key words : Solar irradiation, machine Learning, support vector machine, linear regression.

ملخص :

يقترح هذا البحث نهجاً مبتكراً لتوقع الإشعاع الشمسي باستخدام أساليب التعلم الآلي. يلعب الإشعاع الشمسي دوراً حاسماً في تطبيقات متعددة، مثل إنتاج الطاقة الشمسية والتخطيط الحضري والزراعة المتقدمة. ومع ذلك، فإن التنبؤ بالإشعاع الشمسي بدقة هو تحدي معقد بسبب تغيرات ظروف الطقس والملاحة الجغرافية في هذه الدراسة، نستكشف استخدام تقنيات التعلم الآلي، مثل الانحدار الخطي وأشجار القرار وآلة الدعم النوعية، لنمذجة وتوقع الإشعاع الشمسي. قمنا باستخدام بيانات جوية خارجية لتوقع الإشعاع الشمسي العالمي اليومي. تم تطوير أدوات التحقق من عملية النمذجة، وأظهر تحليلها توافقاً قوياً بين الإشعاع الشمسي المقاس والمقدر للعائلات الثلاثة الرئيسية المقترحة من النماذج (الانحدار الخطي وأشجار القرار وآلة الدعم النوعية). ومع ذلك، فإنه واضح أن أساليب النواة (آلة الدعم النوعية) توفر تقديراً أفضل مقارنة بالنماذج الأخرى المقترحة يبقى اختيار نوع آلة الدعم النوعية أمراً حاسماً لتوقع الإشعاع الشمسي (آلة الدعم النوعية الكوبية أفضل من آلة الدعم النوعية التربيعية) مع معامل تحديد يساوي 0.93.

كلمات مفتاحية: الإشعاع الشمسي، التعلم الآلي، آلة الدعم النوعية، الانحدار الخطي، أشجار القرار