



Université de Ghardaïa

N° d'ordre :
N° de série :

Faculté des Sciences de la Technologie
Département d'automatique et électromécanique

Mémoire présenté en vue de l'obtention du diplôme de

MASTER

Domaine : *Sciences et Technologies*

Filière : *Automatique*

Spécialité : *Automatique et systèmes*

Par : BAHAMIDA Aicha Imane

Thème

**L'utilisation des techniques de l'apprentissage
approfondi pour l'estimation des composantes solaires**

Soutenu publiquement le 30/09/2020

Devant le jury :

SEBA Hadj Yahia	MCB	Univ. Ghardaïa	Président
SADOUNI Radhwane	MCB	Univ. Ghardaïa	Examineur
BEN CHAABANE Acheur	MAA	Univ. Ghardaïa	Examineur
GUERMOUI Mawloud	Maitre de Recherche A	URAER. Ghardaïa	Encadrant
GAIRAA Kacem	Maitre de Recherche B	URAER. Ghardaïa	Co-Encadrant

Année universitaire 2019/2020

REMERCIEMENTS

De nos jours, toutes les études pour la réalisation de travaux se concrétisent grâce à une parfaite collaboration, un travail coopératif assure une bonne circulation de l'information et aussi un suivi sérieux et permanent de l'étude en question.

Dans ce contexte, je tiens à remercier vivement et à exprimer toute ma gratitude et ma reconnaissance en premier lieu à mon encadreur Dr GUERMOUI Mouloud, Maitre de Recherche A à l'Unité de Recherche Appliquée en Energies Renouvelables, Ghardaïa, pour avoir proposer et diriger ce travail et pour son support, ses conseils, sa confiance et sa motivation qui m'a apporté durant tout ce projet.

Je voudrais remercier vivement, mon Co-encadreur Dr GAIRAA Kacem, Maitre de Recherche B à l'Unité de Recherche Appliquée en Energies Renouvelables, Ghardaïa, pour sa disponibilité et surtout son aide et ses judicieux conseils, Je voudrais lui exprimer toute ma gratitude pour la grande confiance qu'il m'a accordée.

Je remercie également le Directeur de l'Unité de Recherche Appliquée en Energies Renouvelables (URAER, Ghardaïa), Dr DJAFER Djelloul, qui nous a permet de réaliser ce travail au sein de l'URAER.

Je tiens à saisir cette occasion pour adresser mes profonds remerciements à tous mes enseignants de l'Université Ghardaïa.

Je veux exprimer mes reconnaissances envers mes collègues qui m'ont apporté leur soutien moral et intellectuel tout au long de nos démarche.

Un grand merci à toute ma famille et à tous ceux qui me sont chers.

En ce jour solennel qui clôt le cycle de mes études qu'il me permet de dédier ce noble travail à la mémoire de ma mère K.L

RESUME

L'estimation précise des composantes du rayonnement solaire d'une location spécifique a été un des problèmes les plus importants des applications de l'énergie solaire. Dû au comportement non stationnaire des paramètres du rayonnement solaire et de la variété des conditions atmosphériques, les modèles de prévision autonomes sont insuffisants pour fournir une estimation précise dans certains cas. À cet égard, une nouvelle approche nommée CNN-SVM a été proposée pour la prévision multi-horaires du rayonnement solaire global.

Le modèle proposé est basé principalement sur l'utilisation du modèle de réseaux de neurones convolutifs (CVV-1D) comme fonction de sélection des entrées météorologiques et la machine à vecteur support (SVM) comme modèle de régression.

Les résultats obtenus montrent que le modèle combiné CNN-SVM donne une légère amélioration par rapport au modèle autonome en termes d'indicateurs statistiques.

Mots clés : Rayonnement solaire, Estimation des composantes du rayonnement solaire, réseaux de neurones convolutifs (CVV-1D), machine à vecteur support (SVM), Modèle combiné CNN et SVM.

LISTE DES FIGURES

Figure I.1	Structure de l'étoile soleil	04
Figure I.2	Analyse spectrale du rayonnement solaire	05
Figure I.3	Mouvement du globe terrestre	06
Figure I.4	La déclinaison solaire en fonction du jour	07
Figure I.5	Déclinaison solaire	07
Figure I.6	Les coordonnées horizontaux	09
Figure I.7	Variation de l'équation du temps en fonction du jour	11
Figure I.8	Variation du rayonnement hors atmosphère au cours de l'année	12
Figure I.9	Schéma explicatif de différentes couches atmosphériques	14
Figure I.10	Rayonnement diffus sur une surface inclinée	16
Figure I.11	Composition du rayonnement global sur un plan incliné	17
Figure II.1	Pyrhéliomètre Eppley : mesure du rayonnement solaire direct	20
Figure II.2	Pyranomètre équipé d'un pare-soleil : mesure du rayonnement diffus	21
Figure II.3	Instruments de mesure du rayonnement solaire	22
Figure II.4	Un spectroradiomètre	22
Figure II.5	L'héliographe de Campbell-Stokes	24
Figure II.6	Héliographe Cimel à fibre optique	24
Figure II.7	Réseau d'Héliographe (1983-2012)	26
Figure II.8	Réseau pyranométrique (ONM et CDER)	27
Figure II.9	Site de Ghardaïa en Algérie	29
Figure II.10	Unité de recherche appliquée en énergies renouvelables (URAER) Ghardaïa	30
Figure II.11	Station de mesure radiométrique de l'URAER (Ghardaïa)	31
Figure II.12	Exemple de mesures sur une journée au site de Ghardaïa	32
Figure II.13	Evolution de l'éclairement global	32
Figure II.14	Fréquence de distribution du rayonnement global	33
Figure II.15	Distribution journalière de l'indice de clarté	34
Figure II.16	Variation journalière de l'indice de clarté	34
Figure III.1	La séparation de deux ensembles de points par un hyperplan	39
Figure III.2	Vecteurs de support	40
Figure III.3	Séparateurs à vaste marge	40
Figure III.4	L'hyperplan correspondant à la fonction de décision d'un classifieur linéaire dans R^2	42
Figure III.5	Hyperplan optimal par la maximisation de la marge	43
Figure III.6	Maximisation de la marge pour obtenir l'hyperplan optimal	44
Figure III.7	Architecture d'une machine à vecteurs de support	44
Figure III.8	Cas linéairement séparable et non linéairement séparable	45
Figure III.9	Exemples d'hyperplans séparateurs dans R^2	45

LISTE DES FIGURES ET DES TABLEAUX

Figure III.10	Marge souple et Slack variable ξ	47
Figure III.11	Transformation d'un problème de séparation non linéaire en un problème de séparation linéaire dans un espace de re-description	47
Figure III.12	Un simple ConvNet	48
Figure III.13	Une image de (4x4) avec un filtre de (3x3)	53
Figure III.14	La première étape de convolution 2D	53
Figure III.15	Déplacement de filtre par colonne	54
Figure III.16	La dernière étape de convolution	54
Figure III.17	Convolution en 3D	55
Figure III.18	Max-pooling avec un filtre 2x2	55
Figure III.19	Modèle CNN-SVM	56
Figure IV.1	Variation du rayonnement global horaire à Ghardaïa entre 2013-2017	60
Figure IV.2	Comparaison des performances des trois modèles pour 12 heures en fonction de coefficient de corrélation	73
Figure IV.3	Comparaison des performances des trois modèles pour 12 heures en fonction de RMSE	74

LISTE DES TABLEAUX

Tableau I.1	Valeurs typiques de l'albédo du sol	15
Tableau I.2	La variance des constantes a_0 , a_1 , a_2 en fonction de type de climat	18
Tableau II.1	Potentiel solaire en Algérie	28
Tableau IV.1	Résultats de prédiction (cas global horaire)	61
Tableau IV.2	Résultats de prédiction (cas K_t horaire)	63
Tableau IV.3	Résultats du modèle SVM avec la meilleure combinaison	65
Tableau IV.4	Résultats avec 6 paramètres pour 12 heures	66
Tableau IV.5	Résultats du modèle CNN-1D	67
Tableau IV.6	Résultats du modèle CNN-1D pour 12 heures	68
Tableau IV.7	Résultats de prédiction du CNN-1D pour une heure	68
Tableau IV.8	Résultats du modèle CNN-1D avec 6 paramètres pour 12 heures	70
Tableau IV.9	Les résultats de combinaison de SVM-CNN dans 12 heures avec 5 entrées.	70
Tableau IV.10	Les résultats de combinaison de SVM-CNN pour 12 heures.	71
Tableau IV.11	Les résultats de combinaison de SVM et CNN avec différents filtres.	71
Tableau IV.12	Résultats de CNN-SVM avec des nouveaux nombres de filtres.	72

NOMENCLATURE

Symbole	Description	Unité
C	La vitesse de la lumière	m/s
λ	Longueur d'onde	Degrés
δ	Déclinaison du soleil	Degrés
N	Le nombre de jours	[-]
ω	L'angle horaire du soleil	Degrés
TSV	Le temps solaire vrai	Heure
ω_c	L'angle horaire du soleil à son coucher	Degrés
γ	Hauteur du soleil	Degrés
ϕ	Latitude du lieu	Degrés
z	La distance zénithale	Degrés
ψ	L'azimuth	Degrés
TU	Temps universel	Heure
TSM	Temps solaire moyen	Heure
TSV	Temps solaire vrai	Heure
ET	Equation de temps	Heure
I_0	La constante solaire	W/m ²
D_0	La distance moyenne terre-soleil	Km
R_s	Rayon du soleil	Km
σ	La constante de Stefan	W/m ² /K ⁴
G_{on}	Le rayonnement extraterrestre	W/m ²
D_n	Rayonnement direct à incidence normale	W/m ²
D_h	Rayonnement diffus horizontal	W/m ²
G_h	Rayonnement global horizontal	W/m ²
G_β	Le rayonnement solaire global sur une surface inclinée	W/m ²
R_b	Rapport concerne le rayonnement direct et l'inclinaison du capteur	[-]
R_d	Rapport concerne le rayonnement diffus et l'inclinaison du capteur	[-]
R_r	Rapport concerne le rayonnement réfléchi du sol et l'inclinaison du capteur	[-]
R_b	Un facteur géométrique	[-]
β	L'angle d'inclinaison du capteur	Degrés
ρ	L'albédo du sol	[-]
DNI	La composante normale directe	W/m ²
τ_{DNI}	La transmittance atmosphérique pour le rayonnement direct	W/m ²
μ	La fraction d'insolation	Wh/m ²
SS	La durée du jour mesurée	Heure
SS ₀	La durée maximale théorique	Heure

K_t	L'indice de clarté du ciel	W/m^2
X	L'espace des entrées	[-]
Y	L'espace des sorties	[-]
n	La taille de l'échantillon	[-]
S	L'échantillon d'apprentissage	[-]
f	Classifieur linéaire	[-]
ξ	Variable d'écart	[-]
r	Le coefficient de corrélation	(%)
RMSE	L'erreur quadratique moyenne	Wh/m^2
rRMSE	L'erreur relative moyenne	[-]
\hat{Y}	La valeur estimée	[-]
Y	La valeur mesurée	[-]
MABE	L'erreur de biais absolue moyenne	[-]

TABLE DES MATIERES

RESUME
NOMENCLATURE
LISTE DES FIGURES
LISTE DES TABLEAUX

INTRODUCTION GENERALE **01**

CHAPITRE I : LE RAYONNEMENT SOLAIRE

I.1.	Introduction	03
I.2.	Rayonnement solaire - aspect géométrique	03
I.2.1.	Le soleil et le rayonnement solaire	03
I.2.2.	Spectre solaire	04
I.2.3	Mouvement de la terre autour du soleil	04
I.2.4	Mouvement apparent du soleil	04
I.2.4.1	Les coordonnées équatoriales (Coordonnées horaires)	05
a)	Déclinaison δ	05
b)	Angle horaire ω	06
I.2.4.2	I.2.4.2 Les coordonnées horizontaux	06
a)	Hauteur du soleil (h_s)	06
b)	La distance zénithale z	07
c)	L'azimuth ψ	07
I.2.5	Le temps	08
I.2.5.1	Le temps universel (TU)	08
I.2.5.2	Le temps solaire moyen (TSM)	08
I.2.5.3	Le temps solaire vrai (TSV)	08
I.2.5.4	Equation du temps (ET)	08
I.2.6	La constante solaire I_0	09
I.2.7	Rayonnement extraterrestre	09
I.2.8	L'atmosphère terrestre	09
a)	La troposphère	10
b)	La stratosphère	11
c)	La mésosphère	11
d)	La thermosphère	12
I.3	Rayonnement solaire- aspect énergétique	13
I.3.1	Rayonnement solaire au sol	13
I.3.2	Rayonnement solaire sur une surface inclinée	13
I.3.3	Modèle de calcul du rayonnement solaire par ciel clair	14
I.3.3.1	Modèle Hottel et Liu & Jordan	14
I.4	Conclusion	14

CHAPITRE II : LES INSTRUMENTS DE MESURE DE RAYONNEMENT SOLAIRE

II.1	Introduction	19
II.2	Mesure du rayonnement solaire	19
II.2.1.	Rayonnement solaire direct	19
a)	Le pyréliomètre	19
II.2.1.2	Rayonnement solaire diffus	20
a)	Le pyranomètre	20
II.2.1.3	Rayonnement solaire global	21
II.2.1.4	Le spectroradiomètre	22
II.2.1.5	L'albédo du sol	23
II.2.1.6.	La durée d'isolation	23
a)	L'héliographe de Campbell-Stokes	23
II.2.1.7	La nébulosité	24
II.2.2	Paramètres dérivés de mesures	25
II.2.2.1	La fraction d'insolation et indice de clarté	25
II.3	Réseau de mesure radiométrique de l'Algérie	25
II.4	Gisement solaire en Algérie	27
II.5	Présentation du lieu de stage	28
II.5.1	Site de Ghardaïa	28
II.5.2	Aperçu sur l'unité de recherche appliquée en énergies renouvelables (URAER)	29
II.5.3	Données utilisées	30
II.5.4	Evolution du rayonnement solaire global	32
II.5.6	Indice de clarté	33
II.6	Conclusion	35

CHAPITRE III : LES METHODES DE REGRESSION SVM ET CNN

III.1	Introduction	36
III.2	Notions sur l'apprentissage statistique	36
III.2.1	Différents types d'apprentissage	37
III.2.1.1	Apprentissage supervisé	37
a)	Classification	37
b)	Régression	37
III.2.1.2	Apprentissage non supervisé	37
III.3	Les machines à vecteurs de supports	37
III.3.1	Notions de base	38
III.3.1.1	Hyperplan	38
III.3.1.2	Vecteurs de support	39
III.3.1.3	Marge	40
III.3.2	Définition	40
III.3.2.1	Séparatrice linéaire (Séparateur linéaire)	41
III.3.2.2	La notion de marge maximale	42

III.3.2.3	Fonction noyau	44
III.3.3	Machines à vecteurs supports linéaires	44
III.3.3.1	Cas des données séparables	44
III.3.3.2	Cas des données non Séparable	46
III.4	Le réseau de neurone convolutif (CNN)	48
III.4.1	L'architecture LeNet (années 1990)	48
III.4.2	Les différentes couches d'un CNN	49
III.4.2.1	La couche de convolution (CONV)	49
III.4.2.2	La couche de pooling (POOL)	49
III.4.2.3	La couche de correction (ReLU)	50
III.4.2.4	La couche entièrement connectée (FC)	50
III.4.2.5	Couche de perte (LOSS)	51
III.4.3	Choix des paramètres	51
III.4.3.1	Nombre de filtres	51
III.4.3.2	Forme de filtres	51
III.4.3.3	Conception du classificateur	52
a)	Training	52
b)	Test	52
III.4.4	Les types de convolutions	52
III.4.4.1	Convolution en 2D	53
III.4.4.2	Convolution en 3D	54
III.4.4.3	Convolution 1D	55
III.5	L'hybridation du CNN avec le modèle SVM	55
III.6	Performances et évaluation des prédicteurs	56
III.6.1	Racine de l'erreur moyenne quadratique (RMSE)	57
III.6.2	Erreur relative moyenne (rRMSE)	57
III.6.3	Coefficient de corrélation (r)	57
III.6.4	L'erreur de biais absolu moyenne (MABE)	58
III.7	Conclusion	58

CHAPITRE IV : RESULTATS ET DISCUSSION

IV.1	Introduction	59
IV.2	Prédiction du rayonnement global horaire	59
IV.3	Prédiction par le modèle SVM	60
IV.4	Résultats du modèle SVM	61
IV.5	Résultats du modèle CNN	66
IV.6	Résultats des modèles d'hybridation de SVM et CNN	70
IV.7	Comparaison des modèles	73
IV.8	Conclusion	74

CONCLUSION GENERALE	76
----------------------------	-----------

BIBLIOGRAPHIE

ANNEXE

INTRODUCTION GENERALE

INTRODUCTION GENERALE

Le monde connaît depuis ces dernières années un développement important économique résultant d'une croissance exceptionnelle des besoins énergétiques de l'humanité, Le développement industriel, l'augmentation du parc automobile et la multiplication des équipements domestiques ont provoqué une augmentation importante de la demande énergétique, mais cette croissance de la demande a été majoritairement visés par l'utilisation de sources d'énergies fossiles. Il est nécessaire de trouver de nouvelles sources d'énergie, tout en tenant compte des notions économiques et environnementales comme le soleil qui est une énergie renouvelable énorme et inépuisable, rayonne sur la terre avec une grande puissance dans toutes les longueurs d'ondes du spectre de la lumière visible. Les avantages des énergies renouvelables sont nombreux ; la durabilité, la disponibilité et la propreté mais l'inconvénient majeur reste le coût initial très élève pour l'installation du matériel de conversion [1].

Cette énergie renouvelable est une préoccupation mondiale et d'actualité surtout si elle est exploitée dans la conversion de l'énergie solaire en électricité, qui est l'un des enjeux de la politique énergétique de notre siècle.

Le gisement solaire en Algérie est important, il compte tenu des valeurs de la durée d'insolation enregistrées sur la quasi-totalité du territoire national dépassant les 2000 heures annuellement et pouvant atteindre même les 3900 heures sur les hauts plateaux et le Sahara, ce qui permet l'utilisation de l'énergie solaire dans différents domaines. La meilleure méthode pour connaître le gisement solaire sur un site donné est d'installer une bonne série de mesures de dispositifs du rayonnement au sol et la nécessité de connaître les approches de collecte, de traitement et d'analyser des données de ce rayonnement, afin d'évaluer correctement le potentiel d'énergie solaire de la région [2].

Les applications de l'énergie solaire nécessitent un développement des données de longue durée sur le rayonnement solaire. La qualité et la disponibilité de ces données sont obtenues dans des régions munies de stations radiométriques, ces dernières mesurent les paramètres météorologiques, la durée d'insolation et l'irradiation solaire ainsi le rayonnement solaire global, direct et diffus qui sont mesurés essentiellement sur un plan horizontal et incliné avec différents

angles d'inclinaison et angles azimutaux en utilisant les modèles statistiques pour l'estimation du rayonnement solaire.

Le but de ce travail est d'appliquer des méthodes de l'apprentissage approfondi pour la prédiction du rayonnement solaire horaire multi-horizon.

Cette mémoire qui décrit notre travail est architecturée en quatre chapitres :

Le premier chapitre est consacré à des notions de base du rayonnement solaire de point de vue géométriques et énergétiques ainsi le calcul du rayonnement par ciel clair présentée par le modèle Hottel et Liu & Jordan.

Le deuxième chapitre portera sur la présentation des instruments de mesure du rayonnement solaire ainsi une analyse de l'évolution du rayonnement solaire global au niveau de l'Unité de Recherche Appliquée en Energies Renouvelables (URAER) de Ghardaïa.

Dans le troisième chapitre, une approche théorique de l'apprentissage statistiques est présentée ainsi une description des différentes techniques d'apprentissage tel que la méthode de régression SVM (séparateur à vaste marge), les réseaux neurones convolutif CNN-1D. Ensuite nous présenterons un modèle d'hybridation entre SVM et CNN pour la prédiction du rayonnement solaire horaire.

Le dernier chapitre est dédié à la présentation des résultats obtenus à partir des données de mesure du rayonnement solaire au sol et l'application des deux méthodes CNN et SVM et la combinaison de ces deux derniers avec l'utilisation du programme MATLAB 2016bdans des conditions bien déterminées.

Finalement une conclusion générale contenant les interprétations et les perspectives envisagées terminera cette thèse.

CHAPITRE I

Le Rayonnement Solaire

I.1 Introduction

Le soleil est la source principale de tout rayonnement reçu par la terre, il lui procure l'énergie, la chaleur et la lumière qui sont nécessaires à toute forme de vie. La connaissance de la trajectoire du soleil et sa position en chaque point du ciel et la distance terre – soleil sont nécessaires quant à la détermination de la quantité du rayonnement extraterrestre [1].

Ce rayonnement solaire est affaibli lors de la traversée de l'atmosphère par absorption ou diffusion, suivant les conditions météorologiques et la latitude du lieu au niveau du sol. Dans le but de profiter au mieux de cette ressource énergétique, il est indispensable de connaître la quantité d'énergie solaire disponible à un endroit spécifique à chaque instant de la journée et de l'année. Pour cette raison, nous avons jugé utile de présenter dans ce chapitre le rayonnement solaire de point de vue géométriques et énergétiques [2].

I.2 Rayonnement solaire - aspect géométrique

I.2.1 Le soleil et le rayonnement solaire

Le Soleil est une étoile qui se trouve à une distance environ de 150 millions de kilomètres de la Terre, ce qui correspond à un voyage de 20 ans à peu près en avion. Le diamètre du soleil est de 1390000 km, soit environ 50 fois celui de la terre, son centre a une température de 10 à 15 millions de degré Celsius. Vu de notre planète, il se présente sous la forme d'un disque assez homogène et apparaît sous un angle d'un peu plus d'un demi-degré. Il faudrait donc 93 000 soleils côte à côte pour occuper l'étendue de la voûte céleste. Son rayon est 109 fois celui de la Terre (soit 696 000 km) et sa masse 333 000 fois celle de notre planète [3].

Il est composé à 80% d'hydrogène, 19% d'hélium et 1% d'un mélange de 100 éléments, soit pratiquement tous les éléments chimiques connus depuis que Langevin et Perrin, s'appuyant sur la théorie de la relativité d'Einstein, ont émis l'idée il y a une soixantaine d'années que c'est l'énergie de fusion nucléaire qui fournit au soleil sa puissance. Il est aujourd'hui admis que le soleil est une bombe thermonucléaire hydrogène –hélium transformant chaque seconde 564 millions de tonnes d'hydrogène en 560 millions de tonnes d'hélium ; la réaction se faisant dans son noyau à la température d'environ 25 millions de degrés Celsius. Ainsi, à chaque seconde, le soleil est allégé de 4 millions de tonnes dispersées sous forme de rayonnement [4].

Le soleil émet en permanence, dans toutes les directions de l'espace, des rayonnements électromagnétiques composés de « grains » de lumière appelés photons. Ces rayonnements se propagent à la vitesse de la lumière C ; il lui faut donc, en moyenne, 499 secondes, soit 8 minutes et 19 secondes, pour atteindre notre atmosphère. L'énergie de chaque photon est directement liée à la longueur d'onde λ [2].

Le rayonnement solaire désigne l'ensemble des ondes électromagnétiques émises par le Soleil. Il se compose donc de la lumière visible et d'ultraviolets, mais également d'ondes radio en plus de rayons cosmiques.

On distingue trois régions particulières dans le soleil : [5]

- **L'intérieur du soleil** : Cette région est divisée en trois zones ; le noyau, la zone radiative et la zone convective. Elle représente 40% de la masse du soleil et elle s'étend sur une épaisseur de 25104 km. Sa température atteint plusieurs millions de degrés et la pression un milliard d'atmosphère.

- **La photosphère** : C'est une couche opaque d'environ 300 km d'épaisseur. Sa température est de l'ordre de quelques millions de degrés, décroissant très rapidement dans l'épaisseur de la couche jusqu'à une température dite de surface de l'ordre de 4500°C.

- **La chromosphère et la couronne solaire** : C'est une région à faible densité où la matière est très diluée, elle représente l'atmosphère du soleil. Cette couche est caractérisée par un rayonnement émis très faible est très élevée (environ un million de degrés).

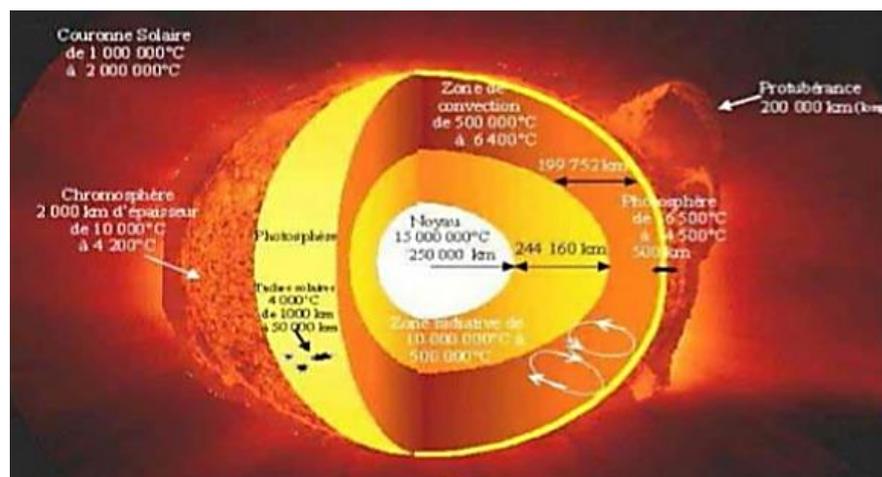


Figure I.1 : Structure de l'étoile soleil [5].

I.2.2 Spectre solaire

Le rayonnement électromagnétique est composé de photons. L'énergie de chaque photon est liée directement à la longueur d'onde λ :

Le spectre du rayonnement extraterrestre correspond environ à l'émission d'un corps noir porté à 5800° K . Une courbe standard, compilée selon les données recueillies par les satellites, est désignée sous le nom de AM0. Sa distribution en énergie est répartie en [4] :

- 6.4% dans la bande des ultraviolets (UV) ($0.20 < \lambda < 0.38 \text{ mm}$)
- 48.0% dans la bande visible ($0.38 < \lambda < 0.78 \text{ mm}$)
- 45.6% dans la bande des infrarouges (IR) ($0.78 < \lambda < 10 \text{ mm}$)

La figure suivante représente la répartition spectrale du rayonnement solaire au niveau du sol terrestre.

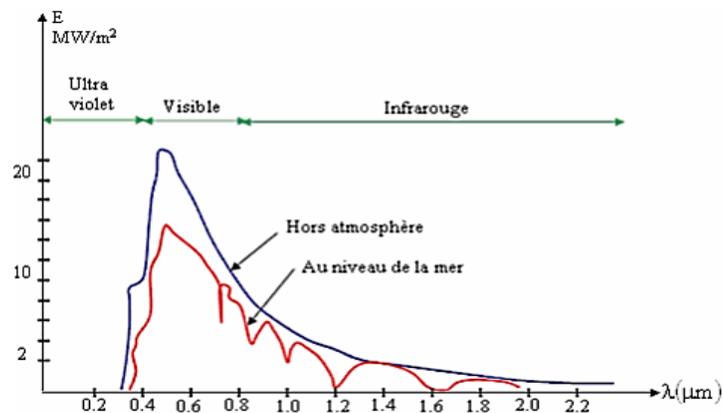


Figure I.2 : Analyse spectrale du rayonnement solaire [6]

I.2.3 Mouvement de la terre autour du soleil

La terre décrit autour du soleil une orbite elliptique quasi circulaire avec une période de 365,25 jours. Le plan de cette orbite est appelé plan de l'écliptique. C'est au solstice d'hiver (21 décembre) que la terre est la plus proche du soleil, et au solstice d'été (21 juin) quelle en est la plus éloignée [7].

La terre tourne sur elle-même avec une période de 24h. Son axe de rotation (l'axe des pôles) a une orientation fixe dans l'espace. Il fait un angle $d = 23^\circ 27'$ avec la normale au plan de l'écliptique

[8]. Le 21 mars et le 21 septembre sont appelés respectivement équinoxes de printemps et équinoxes d'automne. Aux équinoxes le jour et la nuit sont égaux [5].

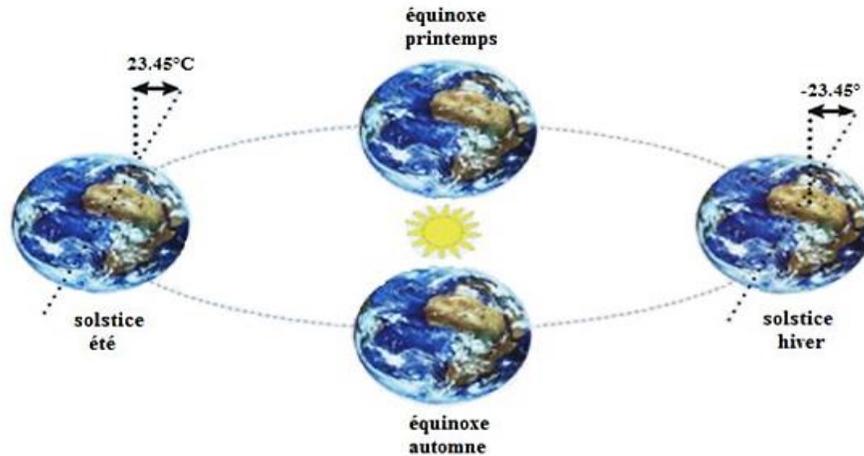


Figure I.3 : Mouvement du globe terrestre [2]

I.2.4 Mouvement apparent du soleil

La position du soleil dans la voûte céleste est repérée à tout instant de la journée par deux systèmes de coordonnées [9] :

I.2.4.1 Les coordonnées équatoriales (Coordonnées horaires)

Les coordonnées équatoriales sont indépendantes de la position de l'observateur sur la terre, mais elles sont liées à l'heure de l'observation. La position du soleil est exprimée par deux angles qui sont [9] :

a) Déclinaison δ

C'est l'angle formé par le vecteur Soleil-Terre avec sa projection sur le plan équatorial. Elle est due à l'inclinaison de l'axe des pôles terrestre par rapport au plan elliptique, ce qui est traduit par les différentes saisons. Elle varie au cours de l'année entre $-23,45^\circ$ et $+23,45^\circ$. Elle est nulle aux équinoxes de printemps (21 mars) et d'automne (23 septembre), et maximale aux solstices d'été (22 juin) et d'hiver (22 décembre).

La variation journalière de la déclinaison δ est d'environ de $0,5^\circ$. Elle est calculée par l'équation de Copernic approximative : [4]

$$\delta = 23.45 \sin \left[360 \frac{(284 + N)}{365} \right] \quad (I.01)$$

N : est le nombre de jours de l'année comptés à partir du 1er janvier, il varie de 1 à 365 ou 366 selon l'année.

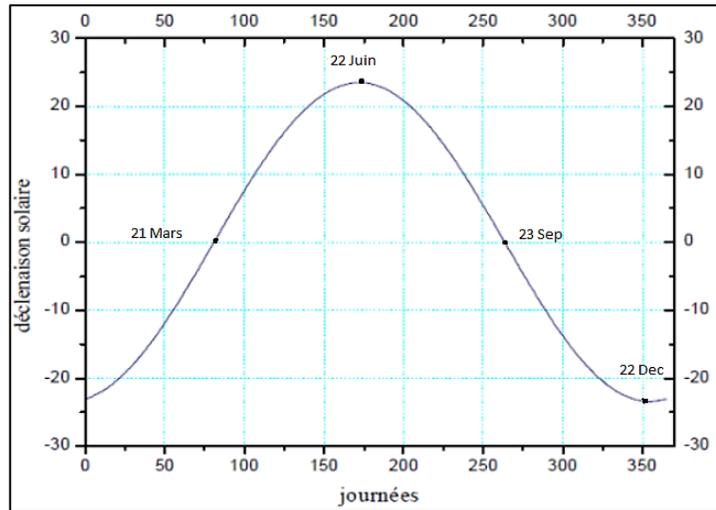


Figure I.4 : La déclinaison solaire en fonction du jour [2]

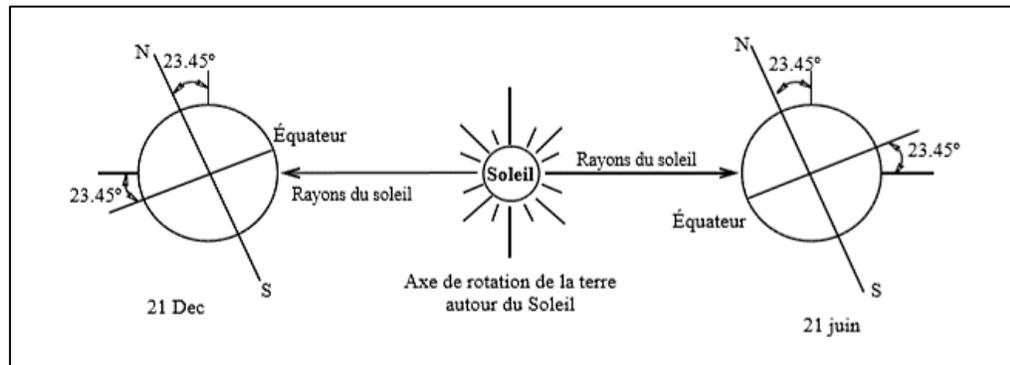


Figure I.5 : Déclinaison solaire [1]

b) Angle horaire ω

L'angle horaire d'un point sur la surface de la Terre est défini comme l'angle à travers laquelle la Terre tournerait pour amener le méridien du lieu directement. L'angle horaire à midi solaire est égal à zéro. À chaque heure qui s'écoule correspond une augmentation de l'angle horaire de 15° , soit encore 4 minutes de temps par degré de longitude. Les heures de l'après-midi sont considérées positives. L'angle horaire ω ($^\circ$) est donné par la formule suivante :

$$\omega = 15 (\text{TSV} - 12) \tag{I.02}$$

TSV est le temps solaire vrai en heure [2].

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 [4] :

$$\omega_c = \cos^{-1} (-\tan\varphi \tan\delta) \quad (\text{I. 03})$$

Pour l'angle horaire au lever du soleil est tout simplement l'opposé de l'angle horaire au coucher.

I.2.4.2 Les coordonnées horizontaux

Le soleil est repéré par les grandeurs suivantes :

a) Hauteur du soleil (hs)

C'est l'angle formé entre la direction du soleil et le plan horizontal, communément appelée hauteur ou élévation angulaire du soleil. γ varie de 0° (soleil à l'horizon) à $+90^\circ$ (vers le zénith) et de 0° à -90° (vers le nadir). La hauteur du soleil est bien représentée par une formule bien connue, qui avait été également adoptée par Capderou : [5,9]

$$\sin(\gamma) = [\sin(\varphi) \cdot \sin(\delta)] + [\cos(\varphi) \cdot \cos(\delta) \cdot \cos(\omega)] \quad (\text{I. 04})$$

Où

φ : Latitude du lieu

δ : La déclinaison du soleil

ω : L'angle horaire du soleil [4]

b) La distance zénithale z

C'est l'angle entre la direction du soleil et la verticale du lieu (zénith). L'angle z est complémentaire de γ [9].

$$\cos(z) = [\sin(\varphi) \cdot \sin(\delta)] + [\cos(\varphi) \cdot \cos(\delta) \cdot \cos(\omega)] \quad (\text{I. 05})$$

c) L'azimuth ψ

L'azimut est l'angle entre le plan vertical contenant le rayon solaire avec la direction du sud. Il se compte de 0° à 360° à partir du sud dans le sens rétrograde. L'azimut est compté positivement vers l'Ouest $\psi > 0$ et négativement vers l'est $\psi < 0$, à midi $\psi = 0$ [5,9].

L'azimut du soleil est représenté en fonction de la déclinaison solaire δ , latitude de lieu φ et l'angle horaire ω de la façon suivante :

$$\cos(\psi) = \frac{\cos \delta \sin \omega}{\cos \gamma} \quad (I.06)$$

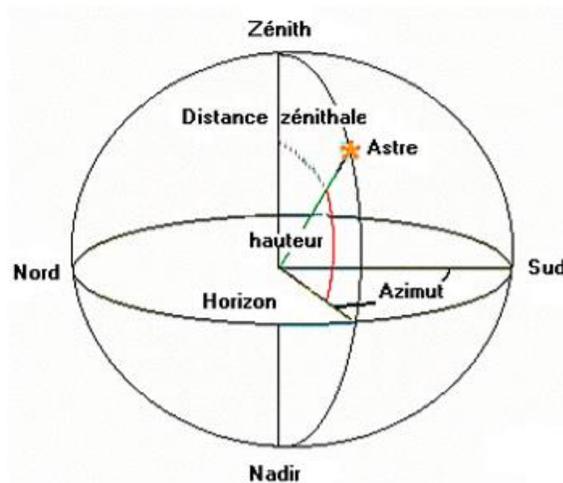


Figure I.6 : Les coordonnées horizontaux [8]

I.2.5 Le temps

La terre subite deux types de déplacement : la rotation de la terre autour de l'axe des pôles et sa révolution autour du soleil. Le mouvement de la terre sur elle-même définit la notion de journée solaire. Une rotation complète s'effectue en 24 heures. Elle amène ainsi la définition du temps, puisque chaque heure correspond à un écart angulaire de 15° . La rotation de la terre autour du soleil définit les saisons et amène à distinguer le temps solaire vrai [9].

I.2.5.1 Le temps universel (TU)

Le temps universel est le temps solaire moyen du méridien de Greenwich GMT (Greenwich Mean Time) qui est le méridien central du fuseau horaire. Il est exprimé par la formule suivante [9] :

$$TU = TSM + \frac{L}{15} \quad (I.07)$$

Avec L la longitude exprimée en degré et les temps en heure [5].

I.2.5.2 Le temps solaire moyen (TSM)

La rotation de la terre autour de l'axe des pôles introduit la notion de temps solaire moyen TSM. La journée solaire moyenne est le temps qui sépare, en moyenne, deux passages successifs du soleil au méridien d'un lieu, la journée solaire moyenne à une durée de : 1 jour = 24h 00m 00 s [5].

Il se déduit de l'équation suivante :

$$TSM = TU + \frac{\varphi}{15} \quad (I. 08)$$

Avec TSM en heure [9].

I.2.5.3 Le temps solaire vrai (TSV) :

Le temps solaire vrai est l'angle horaire du soleil en un lieu et à un instant donné qui est défini par les coordonnées angulaires vraies du soleil. C'est un concept qui montre à la fois le mouvement de rotation et le mouvement de translation de la terre. La variation de la durée du jour solaire vrai (intervalle séparant deux midi consécutifs) est comprise entre : 23h 59m 39s et 24h 00m 30s [5,9].

$$TSV = TSM + ET \quad (I. 09)$$

Où Et est l'équation du temps.

I.2.5.4 Equation du temps (ET)

La vitesse de la terre sur son orbite n'est pas constante au cours de l'année, la somme des inégalités du temps solaire vrai TSV porte le nom d'équation de temps ET avec : $ET = TSV - TSM$. La différence entre TSV et TSM varie selon la date, mais est nul en moyen [5].

Le temps Et varie de -14,5 minutes (du 10 au 15 Février) à +16,5 minutes (du 25 au 30 Octobre). L'équation du temps est donnée en fonction du jour de l'année N avec une bonne approximation du temps Et est donnée par l'équation suivante :

$$Et = 9.9 \sin [2 (0.986 N+100)] - 7.7 \sin (0.986 N-2) \quad (I. 10)$$

Où Et est exprimé en minutes et N est le numéro du jour dans l'année à partir du 1^{er} Janvier [10].

La figure ci-dessous montre la variation de l'équation du temps en fonction du jour de l'année.

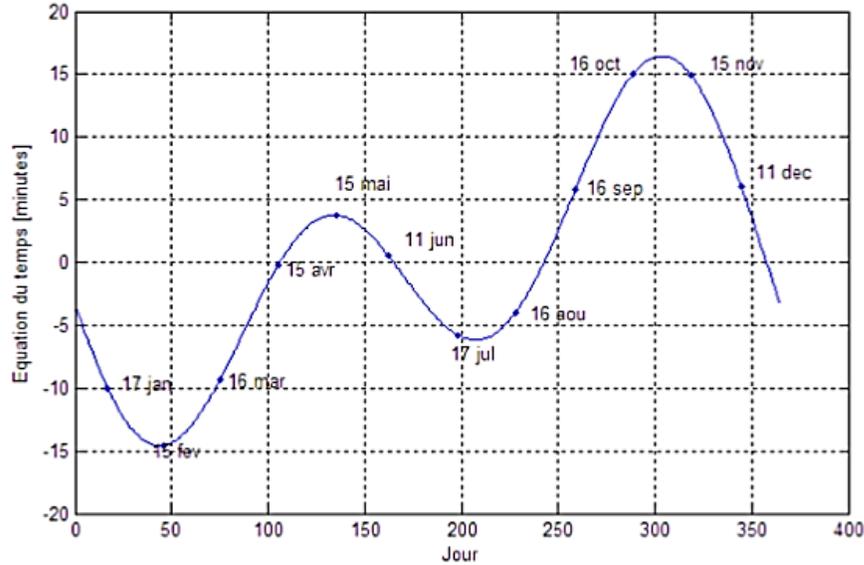


Figure I.7 : Variation de l'équation du temps en fonction du jour [5].

I.2.6 La constante solaire I_0

La constante solaire I_0 exprime le flux énergétique moyen reçu sur une surface unité placée orthogonalement aux rayons solaires au sommet de l'atmosphère terrestre. Cette valeur est soumise à des variations de quelques watts selon deux modes : une variabilité à long terme (période de 33 millions d'années) due au mouvement du système solaire dans la voie lactée et une variabilité à très court terme (un jour à quelques années) due à l'activité solaire [2].

La constante solaire peut être calculée à partir de la température de surface du soleil. Celle-ci est estimée à [1] : $T=5\ 780\ ^\circ\text{K}$

$$\text{Loi de Stefan} \quad \sigma T^4 \cdot 4 \pi R_s^2 = I_0 4 \pi D_0^2 \quad (\text{I. 11})$$

Avec R_s : rayon du soleil, D_0 : distance moyenne terre-soleil, σ : la constante de Stefan d'où :

$$I_0 = \sigma \left(\frac{R_s}{D_0} \right)^2 \cdot T^4 \quad (\text{I. 12})$$

La première valeur moyenne estimée de la constante solaire était 1353 W.m^{-2} . Cette valeur a été mise à jour en 1977 à 1377 W.m^{-2} et plus tard l'Organisation Météorologique Mondiale (OMM) a adopté une valeur de 1367 W.m^{-2} [11].

I.2.7 Rayonnement extraterrestre

Lorsque le Soleil est plus proche de la Terre (le 3 janvier), le rayonnement solaire hors atmosphère est d'environ 1400 W.m^{-2} ; et quand le Soleil est le plus éloigné (le 4 juillet) il est d'environ 1330 W.m^{-2} . Tout au long de l'année, le rayonnement hors atmosphère mesuré sur un plan normal au rayonnement sur la Nième journée de l'année « G_{on} » varie entre ses deux limites, comme est indiqué sur la figure (I.8), et peut être calculé par :

$$G_{on} = I_0 \left[1 + 0.033 \cos \left(\frac{360N}{365} \right) \right] \quad (I.13)$$

Où : N est le numéro du jour de l'année [11,12]

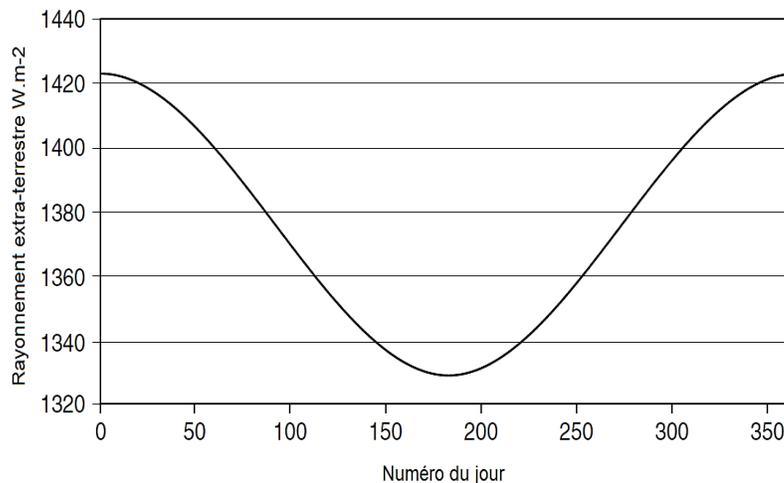


Figure I.8 : Variation du rayonnement hors atmosphère au cours de l'année [13]

I.2.8 L'atmosphère terrestre

La terre est entourée d'une mince couche gazeuse : l'atmosphère. Elle joue le rôle de bouclier protecteur pour toutes les espèces vivantes qui habitent à la surface du globe. En outre, elle les isole de l'espace glacé et les protège des rayons ultraviolets. L'organisation mondiale de la météorologie (OMM) définit quatre zones successives pour l'atmosphère à savoir la troposphère, la stratosphère, la mésosphère et la thermosphère. Ce sont les variations verticales de la température de l'air qui définissent la division de l'atmosphère en ces quatre grandes régions [14].

a) La troposphère

C'est la couche atmosphérique la plus proche du sol, dans laquelle nous vivons. Son épaisseur varie de 7 kilomètres de hauteur au-dessus des pôles, 18 kilomètres au-dessus de l'équateur et environ 13 kilomètres selon les saisons, dans la zone tempérée. La température diminue à mesure qu'on s'élève dans la troposphère, c'est-à-dire à mesure qu'on s'éloigne du sol, atteignant -56°C à la tropopause (zone séparant la troposphère de la stratosphère).

b) La stratosphère

C'est la couche atmosphérique au-dessus de la troposphère, dans la stratosphère on trouve la couche d'ozone. Sa formation il y a 3 milliards d'années a permis le développement de la vie sur notre planète. Car elle filtre une partie des rayons ultraviolets du soleil qui sont extrêmement nocifs pour tout être vivant. Cette absorption provoque un dégagement d'énergie sous forme de chaleur. C'est pourquoi la température augmente lorsqu'on s'élève dans la stratosphère.

c) La mésosphère

Elle se trouve juste au-dessus de la stratosphère. Dans cette couche, la température recommence à décroître avec l'altitude pour atteindre un minimum de -80 degrés Celsius à une altitude d'environ 80 kilomètres. Les poussières et particules qui proviennent de l'espace (les météores) s'enflamment lorsqu'elles entrent dans la mésosphère à cause de la friction de l'air. Ce phénomène nous apparaît sous la forme « d'étoiles filantes ».

d) La thermosphère

C'est la couche la plus haute. Dans cette couche, la température augmente avec l'altitude et peut atteindre environ 1500 degrés Celsius. La thermosphère atteint des milliers de kilomètres d'altitude et disparaît graduellement dans l'espace. La partie inférieure de la thermosphère réfléchit les ondes courtes (ondes radio), elle s'appelle l'ionosphère.

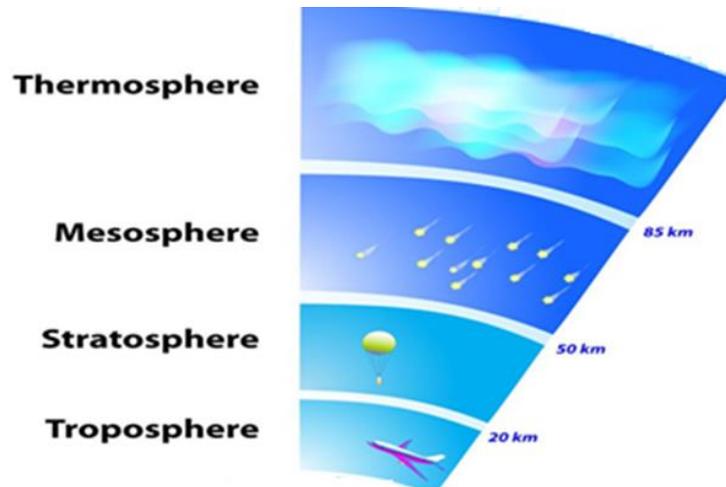


Figure I.9 : Schéma explicatif de différentes couches atmosphériques [15]

I.3 Rayonnement solaire- aspect énergétique

I.3.1 Rayonnement solaire au sol

Lorsque le rayonnement solaire pénètre la couche atmosphérique, il va subir des atténuations dues aux phénomènes d'absorption et de diffusion par les constitutions atmosphériques. En effet,

En effet, trois composantes fondamentales du rayonnement solaire atteignant le sol sont intéressantes pour l'évaluation et la prédiction de l'énergie solaire disponible en un site donné [16] :

1. Rayonnement direct à incidence normale (D_n) : c'est le rayonnement qui provient directement du disque solaire avec un angle d'ouverture de 5.7° , sans qu'il soit atténué par la turbidité atmosphérique.
2. Rayonnement diffus horizontal (D_h) : c'est le rayonnement qui a été diffusé par les nuages, les aérosols et d'autres constituants atmosphériques reçus sur une surface horizontale.
3. Rayonnement global horizontal (G_h) : c'est la somme algébrique des deux composantes précédentes (directe et diffuse).

I.3.2 Rayonnement solaire sur une surface inclinée

Le rayonnement solaire global sur une surface inclinée G_β est la somme des trois composantes : directe, diffuse et le réfléchi du sol. Ce dernier est directement proportionnel au rayonnement global. Le sol réfléchit une partie de l'éclairement qu'il reçoit en fonction de sa nature.

- sol cultivé 10 à 25%
- neige 60 à 90%
- sable 30 à 40%

Donc, le rayonnement global sur une surface inclinée est donné par la relation suivante :

$$G_{\beta} = D_h R_b + D_h R_d + G_h R_r \quad (I. 14)$$

Les rapports R sont des rapports qui tiennent compte de l'inclinaison du capteur, où R_b concerne le rayonnement direct, R_d le rayonnement diffus et R_r concerne le rayonnement réfléchi du sol. R_b est un facteur géométrique et il est fonction de l'angle d'incidence des rayons solaires sur la surface inclinée [17], tel que :

$$R_b = \frac{\cos(\theta)}{\cos(\theta_z)} \quad (I. 15)$$

$$R_b = \frac{(\sin \varphi \cos \beta - \cos \varphi \sin \beta \cos \psi) \sin \delta + (\cos \varphi \cos \beta + \sin \varphi \sin \beta \cos \psi) \cos \delta \cos \omega + \cos \delta \sin \beta \sin \psi \sin \omega}{\sin \varphi \sin \delta + \cos \varphi \cos \delta \cos \omega} \quad (I. 16)$$

Où : φ est la latitude du lieu, β est l'angle d'inclinaison du capteur, ψ l'azimut (orientation) du capteur, δ est la déclinaison du soleil et ω est l'angle horaire.

Le facteur R_r qui concerne le réfléchi du sol est fonction de l'albédo « ρ » qui est le rapport de l'énergie solaire réfléchie par une surface à l'énergie solaire incidente. Le tableau ci-dessous donne quelques valeurs de l'albédo en fonction de la nature du sol :

Tableau I.1 : Valeurs typiques de l'albédo du sol [17]

Nature du sol	Valeurs	Valeurs moyennes
Sable sec	0.1 à 0.4	0.25
Déserts	0.2 à 0.4	0.3
Prairies et herbages	0.2 à 0.3	0.25
Forêt	0.1 à 0.2	0.15
Neige fraîche	0.7 à 0.8	0.75
Neige ancienne	0.5 à 0.7	0.6
Habitat très dense	0.1 à 0.2	0.15

Habitat	0.15 à 0.25	0.2
---------	-------------	-----

Donc, le rayonnement réfléchi est donné par la relation suivante :

$$R_r = \frac{1}{2} \cdot \rho \cdot G_h \cdot (1 - \cos(\beta)) \quad (I. 17)$$

Le facteur R_d se diffère selon le traitement du rayonnement diffus. En effet, l'estimation du diffus sur une surface inclinée peuvent être classées en deux modèles [12] :

- Modèle isotrope, suppose que le rayonnement diffus est uniformément réparti sur la voûte céleste,
- Modèle anisotrope, suppose que l'anisotropie du rayonnement diffus se trouve dans la région du circumsolaire résultante du phénomène de diffusion et concentrée dans une partie du ciel, alors il est distribué d'une façon isotrope dans le reste de la voûte céleste, comme est illustré sur la figure suivante :

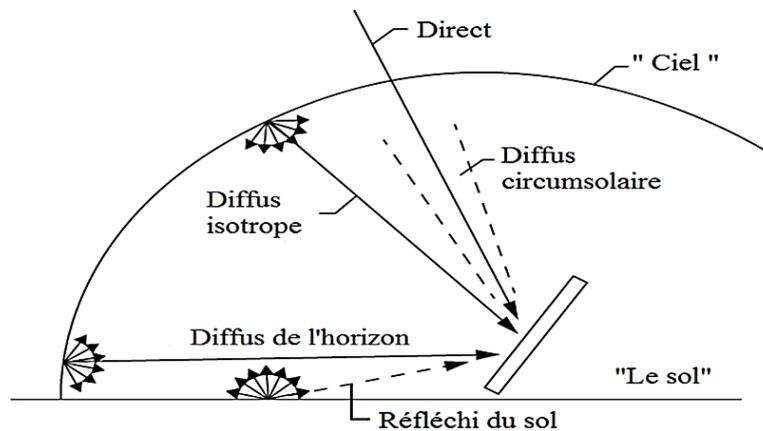


Figure I.10 : Rayonnement diffus sur une surface inclinée [12]

Cependant, dans la bibliographie, il existe plusieurs modèles estimant le diffus sur une surface inclinée selon l'isotropie ou l'anisotropie du rayonnement diffus.

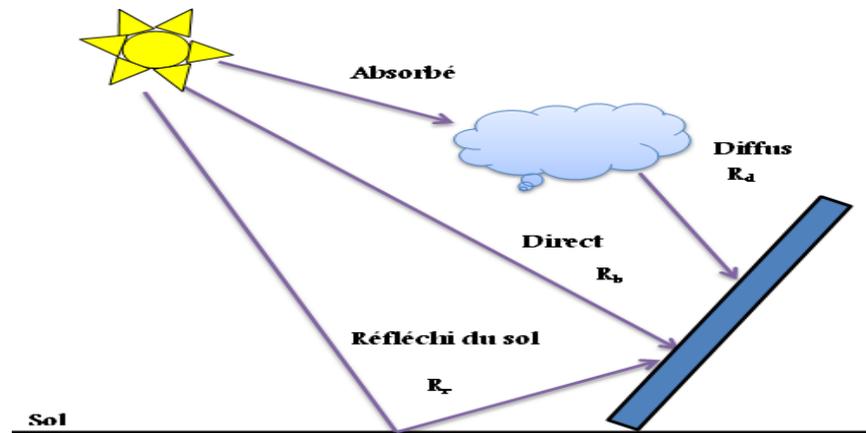


Figure I.11 : Composition du rayonnement global sur un plan incliné [16]

I.3.3 Modèle de calcul du rayonnement solaire par ciel clair

Dans la littérature, plusieurs modèles existent estimant la composante globale par ciel clair. Ces modèles prennent en compte les paramètres de base qui décrivent l'état de l'atmosphère tels que la profondeur optique des aérosols, la vapeur d'eau, la pression atmosphérique, etc. à titre indicatif, nous donnons le modèle Hottel et Liu & Jordan pour l'estimation du rayonnement global sur un plan horizontal.

I.3.3.1 Modèle Hottel et Liu & Jordan

Ce modèle est la combinaison du modèle Hottel [18] qui estime la transmission du rayonnement direct sous un ciel clair, et le modèle Liu & Jordan [19] qui évalue la transmission du rayonnement diffus. Dans ce modèle, la composante normale directe est donnée par la formule suivante :

$$DNI = I_0 \varepsilon \tau_{DNI} \quad (I.18)$$

La transmittance atmosphérique pour le rayonnement direct τ_{DNI} est illustrée par :

$$\tau_{DNI} = A_0 + A_1 \exp(-A_2 / \cos \theta_z) \quad (I.19)$$

Les constantes A_0 , A_1 et A_2 sont calculées à partir de a_0 , a_1 et a_2 pour une visibilité de 23 km et qui peuvent être évaluées pour des altitudes inférieures à 2,5 km comme suit :

$$A_0 = (0.4237 - 0.00821(6 - Z^2))a_0 \quad (I.20)$$

$$A_1 = (0.5055 - 0.00595(6.5 - Z^2))a_1 \quad (I.21)$$

$$A_2 = (0.2711 + 0.01858 (2.5 - Z^2)) a_2 \quad (I. 22)$$

Les constantes a_0 , a_1 , a_2 spécifiques à chaque type de climat sont reportées dans le tableau suivant [20] :

Tableau I.2 : La variance des constantes a_0 , a_1 , a_2 en fonction de type de climat [20]

Type de climat	a_0	a_1	a_2
Tropical	0.95	0.98	1.02
Estival des latitudes moyennes	0.97	0.99	1.02
Subarctique	0.99	0.99	1.01
Hivernal des latitudes moyennes	1.03	1.01	1.00

I.4 Conclusion

Dans ce chapitre, nous avons présenté les notions de base sur le rayonnement solaire, source principale de tous systèmes de conversion de l'énergie solaire, dès son émission par le soleil jusqu'à sa pénétration dans la surface terrestre. Nous avons vu que le rayonnement solaire global reçu sur une surface horizontale est composé de deux composantes : directe et diffus. En ajoutant la composante réfléchi, nous obtiendrons le rayonnement global sur une surface inclinée. Un modèle d'estimation du rayonnement global par ciel clair est également illustré.

Les divers instruments de mesure de rayonnement solaire ainsi que sa valorisation, par l'analyse des données mesurées au niveau de notre lieu de stage, feront l'objet du prochain chapitre.

CHAPITRE II

Les Instruments de Mesure de Rayonnement Solaire

II.1 Introduction

Ce chapitre est consacré à la présentation des instruments de mesure de différentes composantes de rayonnement solaire, ainsi qu'une description du réseau radiométrique existant en Algérie. Un cas d'étude du gisement solaire du site de Ghardaïa est abordé à travers l'analyse de l'évolution du rayonnement solaire global. Les mesures ainsi analysées sont élaborées au niveau de l'Unité de Recherche Appliquée en Énergies Renouvelables (URAER) de Ghardaïa.

II.2 Mesure du rayonnement solaire

La planification des grands projets solaires implique l'installation des stations et instruments radiométriques à hautes précisions, pour avoir des données d'une fiabilité accrue.

Les mesures du rayonnement solaire au sol sont essentiellement réalisées grâce aux pyranomètres pour le rayonnement global et le diffus, couvrant la gamme de longueurs d'onde s'étend de 0,3 à 3 μ m. Elles sont réalisées grâce aux pyréliomètres pour le rayonnement direct à incidence normale. Ces deux types d'appareil sont des capteurs conçus pour mesurer l'intensité du rayonnement électromagnétique émis directement par le soleil ou réfléchis par d'autres influences environnementales (atmosphère, nuages), ces mesures sont appelées éclaircissements dont l'unité est du (W/m²) [21].

La courbe typique de l'ensoleillement durant une journée est sous forme d'une cloche ; elle varie d'une valeur minimale au lever du soleil pour atteindre sa valeur maximale aux alentours du midi solaire avant de décroître pour avoir une autre fois sa valeur minimale au coucher du soleil. L'ensoleillement peut également exprimer la quantité d'énergie solaire captée sur un plan pendant un intervalle déterminé [22].

On distingue principalement selon sa provenance, sa constitution ou la méthode de mesure les paramètres suivants :

II.2.1 Rayonnement solaire directe

Ce type de rayonnement est essentiellement mesuré par un appareil appelé Pyréliomètre, dans le principe de fonctionnement est le suivant :

a) Le pyréliomètre

C'est un appareil dont la surface sensible est toujours disposée perpendiculairement aux rayons solaires. Cet instrument est en permanence orienté vers le Soleil de façon à ce qu'il ne mesure que le rayonnement provenant du seul disque solaire. Il comporte une ouverture réduite et une surface réceptrice qui doit être maintenue normale aux rayons du soleil grâce à un système de poursuite automatique. La surface sensible du capteur est un disque d'argent noirci placé à la base d'un tube muni d'un obturateur et d'un diaphragme limitant son angle d'ouverture à 5.7° (le diamètre apparent du soleil est de 0.5°). Pour pouvoir mesurer le rayonnement direct de façon continue [23,24].



Figure II.1 : Pyrhéliomètre Eppley : mesure du rayonnement solaire direct [3].

II.2.1.2 Rayonnement solaire diffus

La mesure du rayonnement solaire diffus est faite à l'aide d'un appareil appelé « Pyranomètre ».

a) Le pyranomètre

Cet instrument muni un dispositif ou un écran occultant totalement le rayonnement solaire direct. Selon le cas, ce dispositif peut être, soit une sphère ou un disque dirigé en permanence vers le soleil, soit une bande pare-soleil (anneau métallique cachant le disque solaire) parallèle au plan équatorial déplacée en fonction de la déclinaison solaire. Un simple déplacement axial du pare-

soleil tous les deux ou trois jours, permet d'ajuster la position de pyranomètre en fonction de la déclinaison solaire. Le pyranomètre de type **Kipp & Zonen** est le plus utilisé. Il est constitué d'une thermopile de Moll contenue dans un boîtier métallique fermé à sa partie supérieure par deux hémisphères en verre. Ces derniers limitent les échanges par convection ou par conduction moléculaire avec le milieu extérieur. Sa précision est de 3 à 4% et elle est due essentiellement aux limites de cet instrument imposée par la non linéarité de la réponse du pyranomètre. En outre, cet appareil est sensible aux variations de température c'est pour cela, un étalonnage fréquent de cet instrument est nécessaire (en moyenne tous les deux ans) [3,9].



Figure II.2 : Pyranomètre équipé d'un pare-soleil : mesure du rayonnement diffus [3]

II.2.1.3 Rayonnement solaire global

Le rayonnement global est mesuré par le même instrument que le rayonnement diffus : le pyranomètre. Cet instrument collecte le rayonnement solaire sur un plan horizontal ou incliné et absorbé par une surface noircie, en provenance d'un angle solide de 2π stéradians. Le rayonnement reçu est converti en chaleur par la surface noircie. La différence de température entre cette surface et le corps de l'instrument est proportionnelle à l'éclairement énergétique du rayonnement global ; elle est mesurée par une thermopile composée de plusieurs thermocouples montés en série avec

alternance des soudures chaudes et soudures froides. Les soudures chaudes sont en contact thermique avec la surface absorbante une « noircie » avec une capacité thermique faible. Les soudures froides sont en contact avec un matériau ayant une capacité thermique relativement élevée et reste à la température ambiante. Cet instrument doit être régulièrement étalonné et entretenu (poussière, condensation) [1,3,21].



Figure II.3 : Instruments de mesure du rayonnement solaire.

(a): Pyréliomètre (mesure du rayonnement direct), (b): Pyranomètre (mesure du rayonnement global et diffus). (Photo de l'Unité de Recherche Appliquée en Energies Renouvelables – URAER, Ghardaïa)

II.2.1.4 Le spectroradiomètre

On utilise un spectroradiomètre qui fournit la densité spectrale d'éclairement énergétique du rayonnement pour mesurer la répartition spectrale du rayonnement solaire [1].

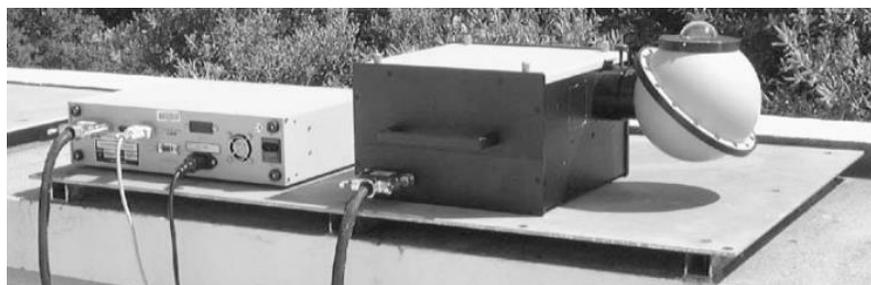


Figure II.4 : Un spectroradiomètre [3]

La figure ci-dessus représente un spectroradiomètre qui contient sur le côté droit un dispositif de captation du rayonnement solaire global (sphère intégrante) ; au centre, le

spectromètre proprement dit ; à gauche, l'électronique de commande et de numérisation, reliée à un microprocesseur. Cet instrument mesure le spectre du rayonnement solaire dans le domaine 250 nm - 800 nm [3].

II.2.1.5 L'albédo du sol

L'albédo du sol est le rapport du rayonnement solaire diffusé ou réfléchi par la surface au rayonnement global. Il est mesuré par un Albédomètre.

L'albédomètre est un instrument composé de deux pyranomètres montés en opposition, l'un captant le rayonnement solaire global et l'autre le rayonnement solaire réfléchi par la surface. En supposant que la parcelle de surface vue par le second capteur est suffisamment homogène et que le signal mesuré est équivalent à l'intégrale du flux réfléchi dans toutes les directions. Cet instrument doit être placé à grande distance du sol [21,22].

II.2.1.6 La durée d'insolation

La durée d'insolation ou d'ensoleillement exprime la longueur cumulée des périodes d'exposition directe au soleil entre le lever et le coucher du soleil pour un endroit donné du globe terrestre. Elle est mesurée à l'aide de L'héliographe [21]. En l'absence permanente de nuages, la durée d'insolation est pratiquement égale à la durée du jour. Les durées d'insolation mensuelle et annuelle sont obtenues par simple cumul des durées d'insolation des diverses journées. L'instrument destiné à mesurer ce paramètre est appelé un héliographe.

a) L'héliographe de Campbell-Stokes

Cet instrument employé dans la plupart des stations météorologiques et également pour les applications énergétiques (calculs de la production d'eau chaude ou d'électricité par panneaux ou moquettes solaires par exemple), ainsi que pour les écologues (la lumière étant la condition de la photosynthèse). Il a été adopté par l'Organisation météorologique mondiale « OMM » en 1962 comme instrument de référence, Il permet de mesurer avec une résolution de six minutes la durée d'ensoleillement quotidienne. Il mesure la durée pendant laquelle le rayonnement solaire direct a une puissance suffisante (en pratique plus de 120 W/m²) pour brûler ou décolorer une bande de papier changée chaque jour. La mesure de la longueur des parties brûlées permet de calculer la durée d'insolation [1,21].

L'évolution technologique de la fin du siècle dernier s'est répercutée par le remplacement de ces instruments "statiques" par des instruments "dynamiques" équipées de capteurs qui permettent d'émettre un signal régulier engendrant une impulsion dès que l'éclairement reçu par le capteur dépasse 120 W/m^2 . Il s'agit d'un système électronique capable de mesurer l'insolation solaire d'une façon autonome et automatique à partir d'un procédé utilisant des photodiodes [21].



Figure II.5 : L'héliographe de Campbell-Stokes [3].

Un héliographe plus récent mesure automatiquement l'insolation c'est un héliographe à fibre optique [1].



Figure II.6 : Héliographe Cimel à fibre optique [25].

II.2.1.7 La nébulosité

On dit la nébulosité ou la couverture nuageuse : C'est la fraction de la voûte céleste occultée par tout ou une partie des nuages à un instant et en un lieu donné. Elle est donnée par l'intermédiaire d'estimations visuelles du ciel relevée toutes les heures ou toutes les trois heures dans la plupart des stations professionnelles de l'ONM.

Ce paramètre est inclus dans les études du rayonnement solaire dans la mesure où il représente le principal facteur de son atténuation. Elle s'exprime en Octa : 0 correspond à un ciel clair, 8 à un ciel totalement couvert [22,26].

II.2.2 Paramètres dérivés de mesures

II.2.2.1 La fraction d'insolation et indice de clarté

La fraction d'insolation exprime le rapport entre la durée du jour mesurée et la durée maximale théorique. En générale elle est l'approche la plus utilisée pour l'estimation des moyennes mensuelles du rayonnement solaire global par application des modèles de régression.

Elle est exprimée par la relation suivante :

$$\sigma = \frac{SS}{SS_0} \quad (\text{II. 01})$$

Où : SS est la durée du jour et SS_0 est la durée du jour maximale théorique.

En plus de la fraction d'insolation, l'indice de clarté du ciel K_t est un autre paramètre qui exprime les variations liées à la couverture nuageuse et la clarté du site. Il est défini comme étant le rapport entre le rayonnement global sur une surface horizontale et le rayonnement extra-terrestre mesuré aussi sur une surface horizontale [21]. Sa description est donnée par la suite en bas (voir II.5.6).

II.3 Réseau de mesure radiométrique de l'Algérie

L'Office National de la Météorologie (O.N.M) dirige sur l'ensemble du territoire national un réseau météorologique constitué de quatre-vingt et une (81) stations mesurant la durée d'insolation. Parmi ces stations, on dispose de seules sept (07) stations équipées de pyranomètres et assurés donc la mesure des composantes globales et diffuses reçues sur le plan horizontal entre les années 1970 et 1980, alors que deux stations seulement sont équipées de pyréliomètres mesurant le rayonnement direct, sur lequel ils s'ajoutent quelques compagnes de mesures du rayonnement global effectuée par des universités ou des centres de recherche. En plus ces séries de mesures

réalisées comportent plusieurs données manquantes causées par des périodes plus ou moins longues de non fonctionnement des appareils de mesure [9,21,22].

Les informations recueillis en 2011 indiquaient que seules trois (03) stations à savoir Oran, Tamanrasset et Ksar Chellala assuraient d'une manière régulière la mesure du rayonnement solaire sur le plan horizontal et huit (08) autres stations automatiques mesurent le rayonnement depuis 2009 et qui sont : Alger (aéroport), Oran (Sénia), Ghardaïa, In amenas, Annaba, Tamanrasset, Constantine et Tlemcen [22].

Les données de la durée d'insolation ont été classées en fonction de leur qualité et de la durée de l'enregistrement en continu à étendre une période de 30 ans de mesures. Les années 1983 à 2012 ont été considérées comme périodes communes de référence. Le tableau (A.1) reprend la liste des coordonnées géographiques des stations météorologiques (Annexe A) [21].

La carte ci-dessous présente la distribution géographique des stations.

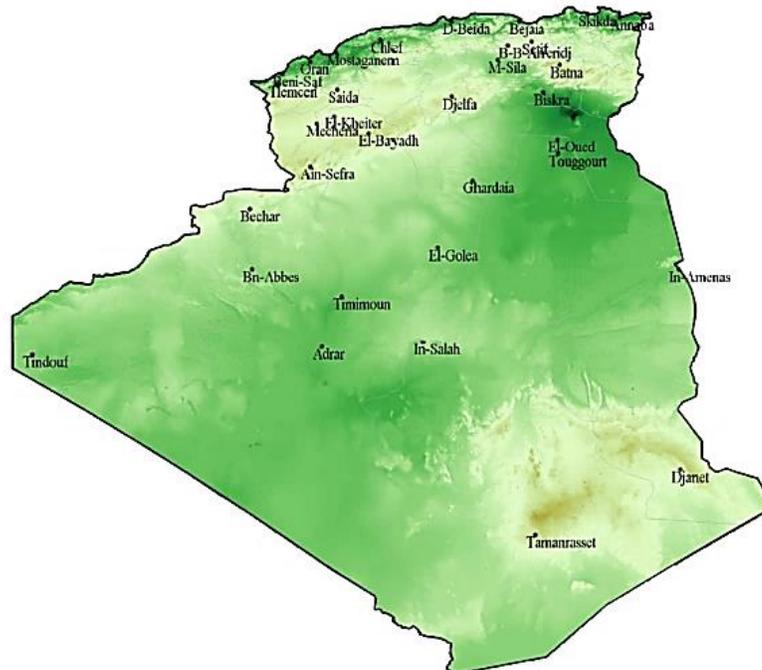


Figure II.7: Réseau d'Héliographe (1983-2012) [21].

On remarque une bonne répartition spatiale suffisante. Toutefois, certaines régions sont moins couvertes géographiquement de stations comme le centre, les régions du grand sud et les régions intérieures.

En ajoutant aux stations de l'ONM, le Centre de développement des énergies renouvelables (CDER) a installé trois stations de mesure radiométrique de hautes précisions à savoir : Bouzaréah, Adrar et Ghardaïa fournissent d'une manière régulière les différentes composantes solaires et certains paramètres météorologiques. La prise de mesure effectuée implique un entretien et un contrôle permanent des appareils de mesure et d'enregistrement. Les instruments sont régulièrement contrôlés et nettoyés et calibrés afin d'assurer une excellente fiabilité des mesures radiométriques [21].

La figure suivante montre la répartition géographique de l'ensemble des stations équipées de pyranomètres [22].

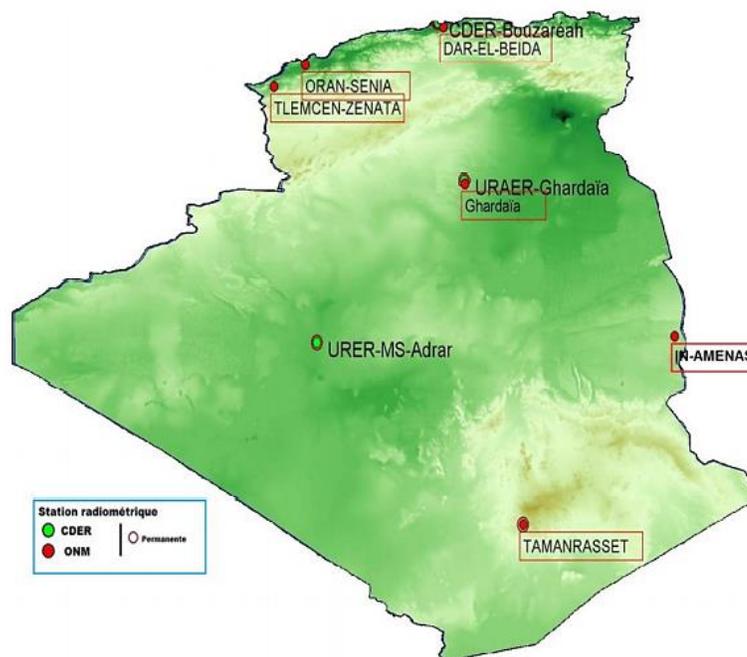


Figure II.8 : Réseau radiométrique (ONM et CDER) [22].

II.4 Gisement solaire en Algérie

L'Algérie dispose d'un des gisements solaires les plus élevés au monde par sa situation géographique, La durée d'insolation sur la quasi-totalité du territoire national dépasse les 2000 heures annuellement et peut atteindre les 3900 heures (hauts plateaux et Sahara). L'énergie reçue quotidiennement sur une surface horizontale de 1m^2 est de l'ordre de 5 KWh sur la majeure partie du territoire national, soit près de $1700\text{ kWh/m}^2/\text{an}$ au Nord et $2263\text{ kWh/m}^2/\text{an}$ au Sud du pays, Ce gisement solaire dépasse les 5 milliards de GWh [27].

Suite à une évaluation par satellites, l'Agence Spatiale Allemande (ASA) a conclu, que l'Algérie représente le potentiel solaire le plus important de tout le bassin méditerranéen, soit 169.000 TWh/an pour le solaire thermique, 13,9 TWh/an pour le solaire photovoltaïque. Le potentiel solaire Algérien est l'équivalent de 10 grands gisements de gaz naturel qui auraient été découverts à Hassi R'Mel. La répartition du potentiel solaire par région climatique au niveau du territoire Algérien est représentée dans le Tableau II.1 selon l'ensoleillement reçu annuellement [28]. Il est utilisé dans des domaines aussi variés que l'agriculture, la météorologie, les applications énergétiques et la sécurité publique [29].

Tableau II.1 : Potentiel solaire en Algérie [22]

Régions	Régions côtières	Hauts plateaux	Sahara
Superficie (%)	4	10	86
Durée moyenne d'ensoleillement (h/an)	2650	3000	3500
Energie moyenne reçue (kWh/m ² /an)	1700	1900	2650

II.5 Présentation du lieu de stage

II.5.1 Site de Ghardaïa

Le site de Ghardaïa se trouve au centre de l'Algérie avec une distance de 600 km d'Alger dont les coordonnées géographiques sont les suivantes : latitude de 32.39° Nord, longitude de 3.78° Est, et une altitude de 463 mètres par rapport au niveau de la mer.

Le site de Ghardaïa présentant un climat désertique, il est limité du côté Nord par la Wilaya de Laghouat (200 km), à l'Est par la Wilaya de Ouargla (200 km), au Sud-Ouest par la Wilaya d'Adrar (800 km), à l'Ouest par la Wilaya d'El-Bayad (350 km). Donc, Ghardaïa est un site idéal pour expérimenter les énergies renouvelables en particulier l'énergie solaire d'où la construction d'une unité de recherche scientifique en l'occurrence de l'Unité de Recherche Appliquée en Énergies Renouvelables (URAER), affiliée au Centre de Développement des Energies Renouvelables (CDER) situé à Bouzaréah, Alger. Cette unité est mise en service en 2002 par le ministère de l'Enseignement Supérieur et le Recherche Scientifique [30].

Le site de Ghardaïa a un caractère désertique où la pluviométrie est du type saharien avec une moyenne annuelle de 67.37 mm. Le nombre de jours de pluie par an en moyen est de 20 à 30 jours, les vitesses moyennes annuelles du vent sont de 3.1 à 4.4 m/s mesurées à 10m du sol.

La caractérisation de site se fait à travers l'analyse des paramètres météorologiques et radiométriques, comme : la précipitation, la température ambiante, le vent et l'humidité relative.

Les températures moyennes minimales mensuelles varient de 3.4 °C en Janvier à 29.5 °C en Juillet, Les températures moyennes maximales mensuelles varient de 14.9 °C en Janvier à 43.2° C en Juillet. Les températures moyennes mensuelles varient de 9 °C en Janvier à 36.5 °C en Juillet avec une moyenne annuelle de 22.6 °C [31].

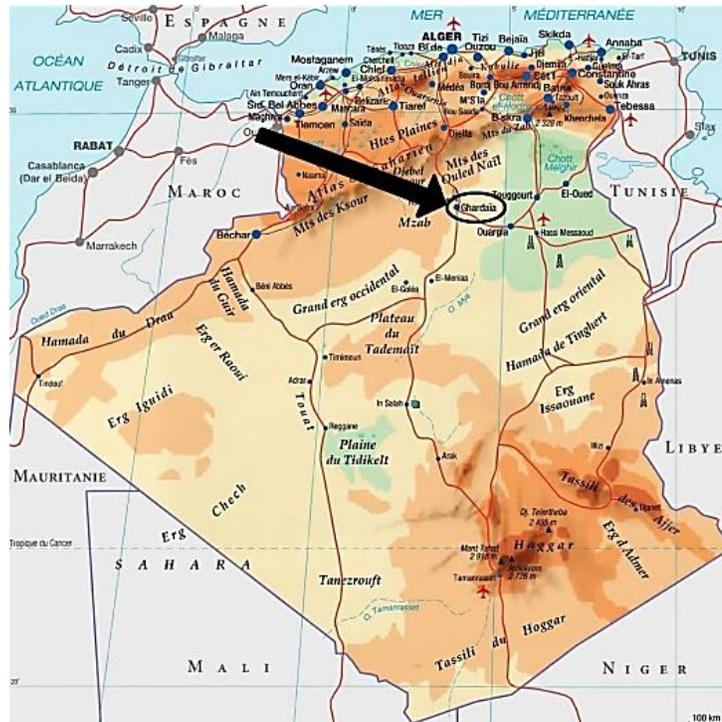


Figure II.9 : Localisation géographique du site de Ghardaïa [30].

II.5.2 Aperçu sur l'unité de recherche appliquée en énergies renouvelables (URAER)

L'Unité de Recherche Appliquée en Energies Renouvelables (URAER) a été inauguré en 1999, appartient au Centre de Développement des Energies Renouvelables (CDER) est située à 1 kilomètre de l'aéroport de Ghardaïa. L'URAER est une plateforme expérimentale dédiée aux différentes activités des énergies renouvelables.

Elle comporte deux divisions assurant la gestion de différentes activités. Ces divisions sont :

- La division Mini Centrales Solaires.
- La division Application des Energies Renouvelables dans les Milieux Arides et Semi Arides.

Chaque division comporte des équipes de recherche veillant à développer les différentes applications liées aux énergies renouvelables. Elles assurent également l'encadrement des étudiants universitaires master et doctorat qui veulent finaliser leurs travaux de fin d'études.



Figure II.10 : Unité de recherche appliquée en énergies renouvelables (URAER) Ghardaïa [32].

II.5.3 Données utilisées

Nous avons choisi dans cette étude d'utiliser les données du site de Ghardaïa. Le choix de ce site est porté sur la disponibilité des données radiométriques recherchées. En fait, l'ensemble des données de rayonnement solaire utilisées dans ce mémoire ont été acquises par une station de mesure radiométrique localisée au niveau de l'unité de recherche appliquée en énergies renouvelables (URAER). Cette station est d'une haute précision et sert pour la mesure des trois composantes radiométriques. Elle est capable de suivre la trajectoire du soleil quotidiennement du lever au coucher, grâce à un système de poursuite. La station est pourvue des instruments suivants :

- ❖ Un pyrhéliomètre CHP1 (Kipp & Zonen) pour mesurer le rayonnement direct à incidence normale,

- ❖ Un pyranomètre de type CM11 (Kipp & Zonen), muni d'une boule sphérique pour mesurer le rayonnement diffus sur une surface horizontale,
- ❖ Un pyranomètre du type CM11 (Kipp & Zonen) pour la mesure de rayonnement global sur un plan horizontal.

Ces instruments sont reliés à une acquisition de données du type Campbell Scientific, modèle RS232 pour le stockage et la collecte des données radiométriques.

Ces mesures sont programmées à un pas de mesure d'une minute et de dix minutes.



Figure II.11 : Station de mesure radiométrique de l'URAER (Ghardaïa)

La figure ci-dessous montre un exemple de mesure des trois composantes solaires. Il est à signaler que le ciel est souvent clair à Ghardaïa toute au long de l'année.

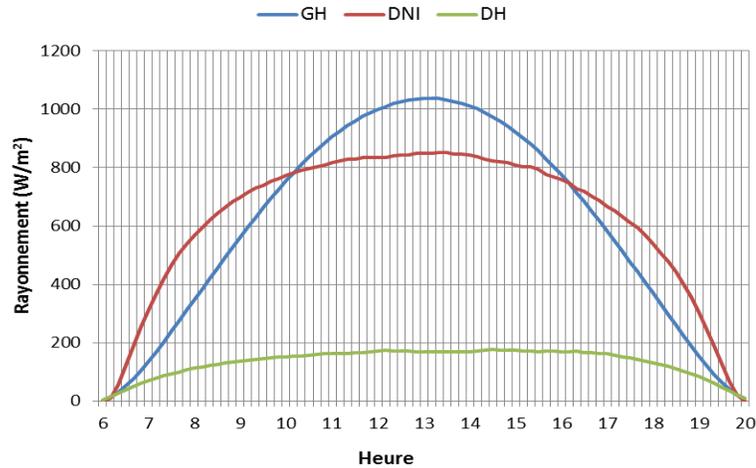


Figure II.12 : Exemple de mesures sur une journée au site de Ghardaïa.

Puisque notre étude s'est focalisée sur le rayonnement global, on va s'intéresser par la suite sur l'évaluation de ce paramètre.

II.5.4 Evolution du rayonnement solaire global

La figure (II.13) illustre l'évolution du rayonnement global sur une surface horizontale pendant une année de mesure. Les données indiquent que les valeurs des éclairagements sont élevées durant les mois d'été et relativement faibles pendant la saison hivernale. L'éclairage global maximum d'une valeur de 1117 W/m^2 est enregistré durant le mois de juin. Le global moyen instantané est important pendant la période allant du mois de mars au mois de septembre.

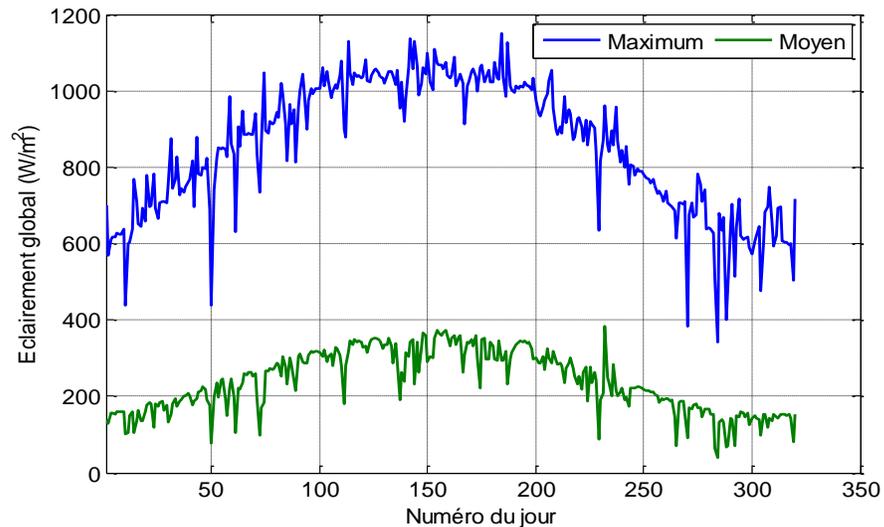


Figure II.13 : Evolution de l'éclairage global

La figure (II.14) montre les fréquences de distribution horaire du rayonnement global reçu sur un plan horizontal par classe d'énergie. On constate que, plus de 12% des données ont des valeurs aux alentours de 500-600 Wh/m² et 600-700 Wh/m², 10.5% des fréquences ayant des valeurs entre 700-800 Wh/m², 9.98% entre 900-1000 Wh/m² et seulement 3% des fréquences ont des valeurs entre 1000-1100 Wh/m². Cela peut être considéré comme point de fonctionnement de certaines applications solaires comme le photovoltaïque qui utilise le global comme paramètre d'entrée pour le dimensionnement.

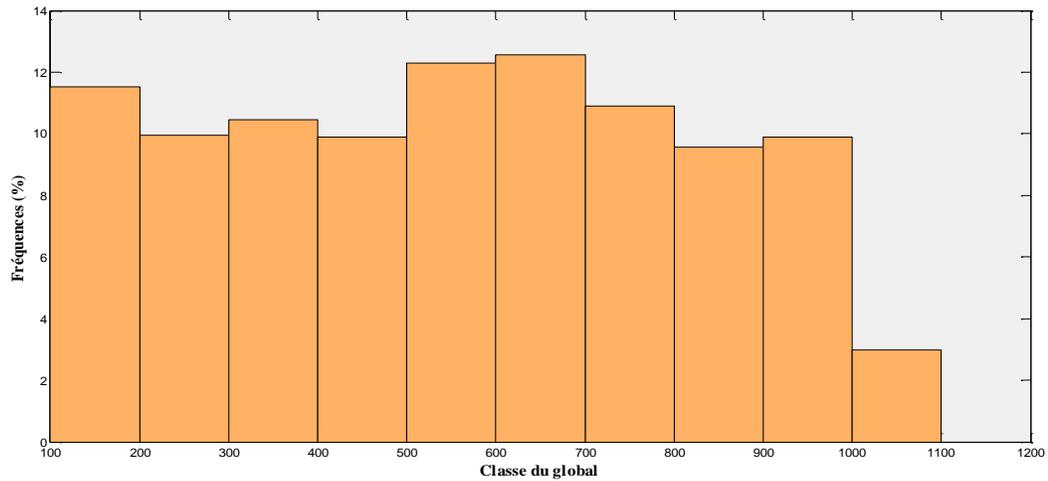


Figure II.14 : Fréquence de distribution du rayonnement global

II.5.6 Indice de clarté

L'indice de clarté (k_t) est un autre paramètre important qui est considéré dans la modélisation et dans l'estimation du rayonnement solaire. Il est défini comme étant le rapport entre le rayonnement solaire global reçu sur une surface horizontale (G_h) et le rayonnement extra-terrestre (G_{on}) ; il est donné par :

$$k_t = \frac{G_h}{G_{on}} \quad (II. 02)$$

L'indice de clarté est utilisé aussi pour traduire l'état du ciel, c'est-à-dire, un ciel couvert, partiellement couvert et un ciel clair. En réalité, il n'y a pas des valeurs exactes de k_t pour définir les conditions du ciel. Néanmoins, plusieurs chercheurs ont adopté des différentes valeurs de k_t pour classer l'état du ciel comme par exemple [11,13] :

- ❖ $k_t < 0.3$ pour un ciel couvert.
- ❖ $0.3 < k_t < 0.65$ pour un ciel partiellement couvert.
- ❖ $k_t > 0.65$ pour un ciel clair.

La distribution journalière de l'indice de clarté est illustrée dans la Figure (II.15). Pour le site de Ghardaïa, des journées à fort indice de clarté (0.7 à 0.8) ont été constatées avec plus de 75% des journées ayant un indice de clarté supérieur à 0.7. Le passage le plus fréquent caractérisant ce site est d'un type de temps moyennement ensoleillé à un beau temps. La variation journalière de l'indice de clarté est présentée dans la Figure (II.16). Les variations de k_t se situent entre 0.6 et 0.8, sauf pour quelques jours où des faibles de k_t valeurs lorsque les conditions de ciel réduisent considérablement le rayonnement solaire ; la valeur moyenne annuelle est de l'ordre de 0.67.

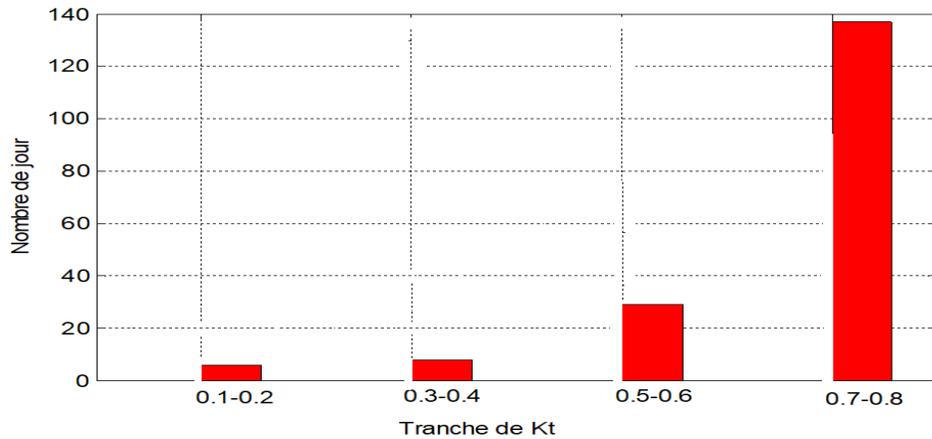


Figure II.15 : Distribution journalière de l'indice de clarté

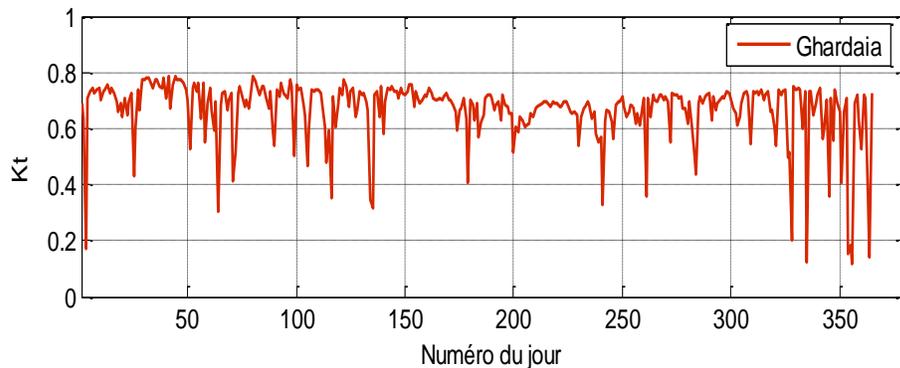


Figure II.16 : Variation journalière de l'indice de clarté

II.6 Conclusion

Dans ce chapitre, nous avons présenté les différents instruments de mesure du rayonnement solaire au sol et les différents paramètres connexes au rayonnement solaire. En la suite, une description du réseau de mesure radiométriques algérien a été illustrée. Pour notre étude, le choix s'est focalisé sur le site de Ghardaïa (URAER) à cause de la disponibilité des données. Une caractérisation de ce site de point de vue radiométrique a été présentée. Nous aborderons dans le prochain chapitre, les modèles de prévision du rayonnement solaire comme les machines à vecteur de support, le réseau neurone convolutif, etc.

CHAPITRE III

**Les Méthodes de Régression SVM
et CNN**

III.1 Introduction

Dans ce chapitre, nous commencerons par donner une approche théorique de l'apprentissage statistique ensuite nous présenterons une description des méthodes de régression, SVM (Séparateur à vaste marge) ; réseaux de neurones convolutif CNN-1D et la combinaison entre les deux modèles CNN-1D et SVM. L'apprentissage automatique a pour objectif de concevoir des algorithmes capables, à partir d'un nombre important d'exemples (les données correspondant à "l'expérience passée"), d'en assimiler la nature afin de pouvoir appliquer ce qu'ils ont ainsi appris aux cas futurs.

Nous intéresserons dans cette thèse d'analyser et appliquée des méthodes récentes de l'apprentissage profond telle que le modèle CNN-1D pour la prédiction du rayonnement solaire horaire multi-horizon. L'utilisation du modèle CNN-1D permet d'extraire les caractéristiques pertinentes dans les inputs. Dans notre cas, nous avons utilisé le modèle CNN pour générer des données pertinentes à partir des données météorologiques afin d'utiliser ces données comme entrée au modèle SVM pour la prédiction de la composante globale horaire du rayonnement solaire.

Dans ce chapitre, nous présenterons dans la première partie des notions de base sur les différentes techniques d'apprentissage est la théorie de L'SVM ainsi que le modèle CNN est dans la deuxième partie nous présenterons notre méthode proposée qui consiste de combiner les deux modèles CNN et SVM pour la prédiction du rayonnement solaire horaire.

III.2 Notions sur l'apprentissage statistique

L'apprentissage statistique, aussi appelé « machine learning » est un domaine à la jonction des statistiques et de l'intelligence artificielle qui a pour but la résolution automatique de problèmes complexes à partir des échantillons. La démarche de conception d'un modèle par apprentissage nécessite de postuler une fonction, dont les variables sont susceptibles d'avoir une influence sur la grandeur à modéliser. Cette fonction dépend des paramètres ajustables. L'apprentissage statistique consiste en l'ajustement de ces paramètres de telle manière que le modèle ainsi obtenu présente les qualités requises de généralisation [33].

D'après Mari & Napoli (1996) [34], une classification, c'est mettre en évidence, d'une part, des relations entre des objets et d'autre part les relations entre ces objets et leurs paramètres. Il s'agit de construire une partition de l'ensemble des objets en un ensemble de classes qui soient les plus homogènes possible. La classification, a donc deux objectifs à atteindre :

- Trouver un modèle capable de représenter la répartition des données (catégorisation).

- Définir de manière formelle l'appartenance à l'une ou l'autre des classes de toute nouvelle donnée (généralisation). En effet la classification a pour but de réduire l'espace de recherche dans une base de données lors du processus d'identification. En pratique, on peut rencontrer deux catégories de problèmes de classification, la classification supervisée et non-supervisée [35].

III.2.1 Différents types d'apprentissage

III.2.1.1 Apprentissage supervisé

Dans ce type d'apprentissage, on cherche à estimer une fonction $f^*(x)$ qui est la relation entre les objets et leurs classes et les objets utilisés comme données d'apprentissages sont accompagnés par la classe à laquelle ils appartiennent. On peut distinguer deux types d'apprentissage supervisé [36] :

a) Classification

La classification consiste donc à apprendre une fonction f_{class} de $X = \mathbb{R}^d$ dans $Y = \mathbb{N}$ qui associe à un vecteur sa classe. Si le nombre de classe est égal à 2, on parle alors de la classification binaire. X est l'espace des entrées (L'ensemble des objets à classifier) et Y l'espace des sorties (L'ensemble des catégories) [33].

b) Regression

Dans le cas de la régression, l'entrée n'est pas associée à une classe mais à une ou plusieurs quantités continues. Ainsi, l'entrée pourrait être les caractéristiques d'une personne (son âge, son sexe, son niveau d'études) et l'étiquette son revenu. La régression consiste donc à apprendre une fonction f_{regr} de $X = \mathbb{R}^d$ dans $Y = \mathbb{R}^k$ qui associe à un vecteur sa valeur associée [33].

III.2.1.2 Apprentissage non supervisé

Dans ce cas les données sont uniquement constituées d'entrées et les tâches à réaliser diffèrent de l'apprentissage supervisé. Ces tâches sont également effectuées par les humains. On cherche à regrouper les objets ayant des caractéristiques communes, les objets utilisés comme données d'apprentissage sont présentés sans leurs classes [33,36].

III.3 Les machines à vecteurs de supports

Depuis plus d'un demi-siècle, les chercheurs essaient de résoudre le problème de l'apprentissage automatique supervisé et plus particulièrement celui de la reconnaissance de formes avec la meilleure façon possible. En 1956, Frank Rosenblatt a construit la première machine d'apprentissage appelée Perceptron où il a décrit le modèle comme un algorithme capable d'être utilisé par des ordinateurs. Ce modèle reçoit en entrée une somme pondérée de signaux d'entrée plus un seuil appelé biais, et fournit une sortie qui dépend de l'entrée. Parmi les caractéristiques du Perceptron est sa capacité de généralisation sur d'autres données non utilisées dans l'apprentissage.

Malgré que les années 60 à 70 n'aient pas donné un résultat révolutionnaire, elles constituent néanmoins les années de la construction des fondements et bases de la théorie d'apprentissage statistique. Dans les années 80, il y a eu l'émergence des réseaux de neurones ou multicouches perceptrons inspirés par le système nerveux biologique. Ce type de méthodes est basé sur la minimisation du risque empirique [37].

Parmi les méthodes à noyaux, inspirées de la théorie statistique de l'apprentissage de Vladimir Vapnik, les SVM constituent la forme la plus connue. SVM est une méthode de classification binaire par apprentissage supervisé, elle fut introduite par Vapnik en 1995. Cette méthode est donc une alternative récente pour la classification. Cette méthode repose sur l'existence d'un classificateur linéaire dans un espace approprié. Puisque c'est un problème de classification à deux classes, cette méthode fait appel à un jeu de données d'apprentissage pour apprendre les paramètres du modèle. Elle est basée sur l'utilisation de fonction dites noyau (kernel) qui permettent une séparation optimale des données. Dans la présentation des principes de fonctionnements, nous schématiserons les données par des « points » dans un plan [38].

Les SVM ont été utilisés dans différents domaines d'expertise, tels que les diagnostics médicaux [39] [40] [41] [42], le traitement d'image [43], le traitement du signal [44], etc. Dans ce chapitre, nous présentons les aspects théoriques de la méthode SVM.

III.3.1 Notions de base

Dans cette section, nous donnons les notions utiles pour l'introduction des machines à vecteurs de support.

III.3.1.1 Hyperplan

Pour deux classes d'exemples donnés, le but de SVM est de trouver un classificateur qui va séparer les données et maximiser la distance entre ces deux classes [38]. On appelle hyperplan séparateur un hyperplan qui sépare les deux classes, en particulier il sépare leurs points d'apprentissage. Comme il n'est en générale pas possible d'en trouver un, on se contentera donc de chercher un hyperplan discriminant qui est une approximation au sens d'un critère à fixer (maximiser la distance entre ces deux classes) [45].

Dans le schéma qui suit, on détermine un hyperplan qui sépare les deux ensembles de points.

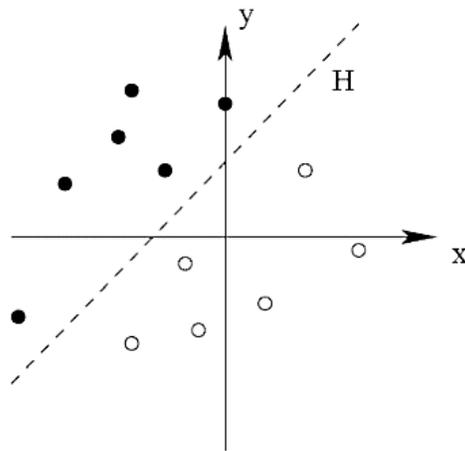


Figure III.1 : La séparation de deux ensembles de points par un hyperplan [38].

III.3.1.2 Vecteurs de support

Ce sont les points de la frontière entre les deux classes des données parmi l'ensemble total d'apprentissage, ces points sont appelés vecteurs de support [45].

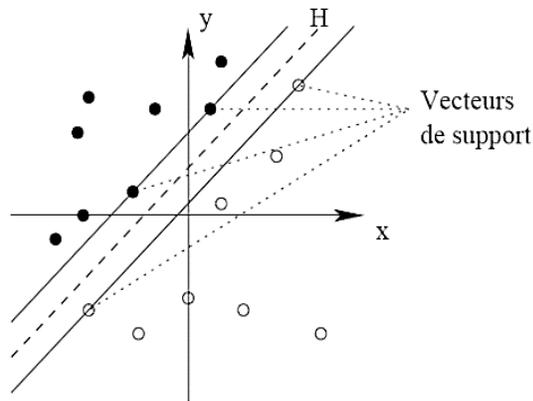


Figure III.2 : Vecteurs de support [38].

III.3.1.3 Marge

Il existe une infinité d'hyperplans capable de séparer parfaitement les deux classes d'exemples. Le principe des SVM est de choisir celui qui va maximiser la distance minimale entre l'hyperplan et les exemples d'apprentissage (i.e. la distance entre l'hyperplan et les vecteurs de support), cette distance est appelée la marge [45].

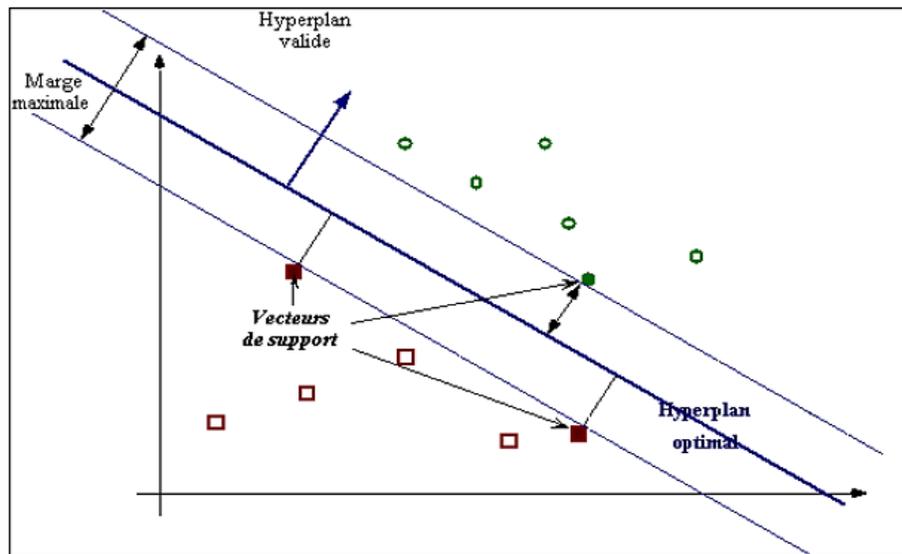


Figure III.3 : séparateurs à vaste marge [38].

III.3.2 Définition

Les SVM sont principalement conçues pour résoudre des problèmes comme :

- Des problèmes de discrimination, permettant de décider à quelle classe appartient un échantillon.

- Cependant, ils peuvent aussi résoudre des problèmes de régression visant à prédire la valeur numérique d'une variable c'est-à-dire de prédiction d'une variable continue en fonction d'autres variables, comme c'est le cas par exemple dans de la prédiction de consommation électrique en fonction de la période de l'année, de la température, etc.

La résolution de ces deux problèmes se base sur la construction d'une fonction h qui, à un vecteur d'entrée x , fait correspondre une sortie y ($y = h(x)$) [46].

III.3.2.1 Séparatrice linéaire (Séparateur linéaire)

Dans le cas de la discrimination biclasse, nous supposons que les données sont des couples (x_i, y_i) $1 \leq i \leq n \in X \times Y$, où X désigne l'espace des variables explicatives souvent pris dans \mathbb{R}^d , $Y = \{-1, +1\}$ et ' n ' étant la taille de l'échantillon. L'appartenance d'une observation x_i à une classe ou à une autre est matérialisée ici par la valeur -1 ou $+1$ de son étiquette y_i .

L'échantillon d'apprentissage S est ainsi une collection de réalisations i.i.d du couple aléatoire (x, y) dont la distribution P est fixe mais inconnue. Cet ensemble est souvent dénoté par :

$$S = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\} \subseteq (X \times Y)^n \tag{III. 01}$$

Une fonction séparatrice entre les classes c_1 et c_2 est une fonction de décision f définie de \mathbb{R}^d dans \mathbb{R} telle que toute observation x_i est affectée à la classe qui correspond au signe de $f(x_i)$ si :

- $f(x_i) \geq 0$, x_i est affecté à la classe positive (+1).

- $f(x_i) \leq 0$, x_i est affecté à la classe négative (-1).

Cette fonction peut être de nature variée. Si f est linéaire, on parle d'un séparateur linéaire ou bien d'un hyperplan séparateur, elle prend la forme générale suivante :

$$\begin{aligned} f(x) &= (w \cdot x) + b \quad \text{avec } (w, x) \in \mathbb{R}^d \times \mathbb{R} \\ &= \sum_{i=1}^n w_i x_i + b \end{aligned} \tag{III. 02}$$

Où w et b sont des paramètres, et $x \in \mathbb{R}$ est une variable.

La méthodologie d'apprentissage implique la recherche des paramètres w et b séparant le mieux possible les données d'apprentissage des classes c_1 et c_2 dans l'espace \mathbb{R}^d .

La fonction signe ($f(x)$) est appelée classifieur linéaire.

Géométriquement, cela revient à considérer un hyperplan qui est le lieu des points x satisfaisant $(w \cdot x) + b = 0$. En orientant l'hyperplan, la règle de décision correspond à observer de quel côté de l'hyperplan se trouve l'exemple x_i . La figure ci-dessous représente la situation dans \mathbb{R}^2 . On voit que le vecteur w définit la pente de l'hyperplan : w est perpendiculaire à l'hyperplan. Le terme b quant à lui permet de translater l'hyperplan parallèlement à lui-même [33].

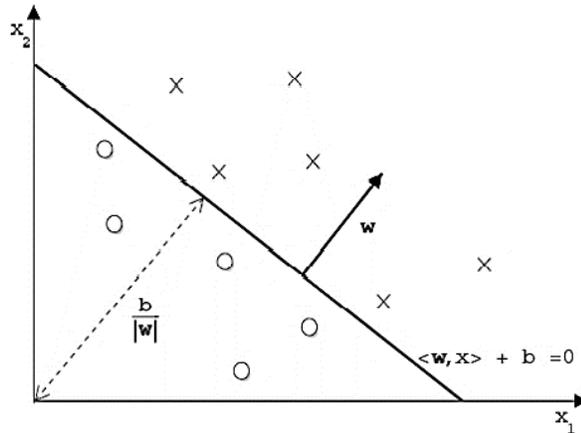


Figure III.4 : L'hyperplan correspondant à la fonction de décision d'un classifieur linéaire dans \mathbb{R}^2 [33].

III.3.2.2 La notion de marge maximale

Selon Vapnik et Kotz (1982), il existe un unique hyperplan optimal permettant de résoudre ce problème. Cet hyperplan optimal, est défini comme étant l'hyperplan qui maximise la marge entre les échantillons et l'hyperplan séparateur. Ce choix est justifié théoriquement par le fait que la capacité des classes d'hyperplans séparateurs diminue lorsque leur marge augmente. La marge est la plus petite distance entre les échantillons d'apprentissage et l'hyperplan séparateur qui satisfait la condition de séparabilité. Ces échantillons sont appelés vecteurs supports [45].

- **Pourquoi maximiser la marge?**

Intuitivement, le fait d'avoir une marge plus large procure plus de sécurité lorsque l'on classe un nouvel exemple. De plus, si l'on trouve le classificateur qui se comporte le mieux vis-à-vis des données d'apprentissage, il est clair qu'il sera aussi celui qui permettra au mieux de classer les nouveaux exemples. Dans le schéma qui suit, la partie droite nous montre qu'avec un hyperplan optimal, un nouvel exemple reste bien classé alors qu'il tombe dans la marge. On constate sur la partie gauche qu'avec une plus petite marge, l'exemple se voit mal classé [38,45].

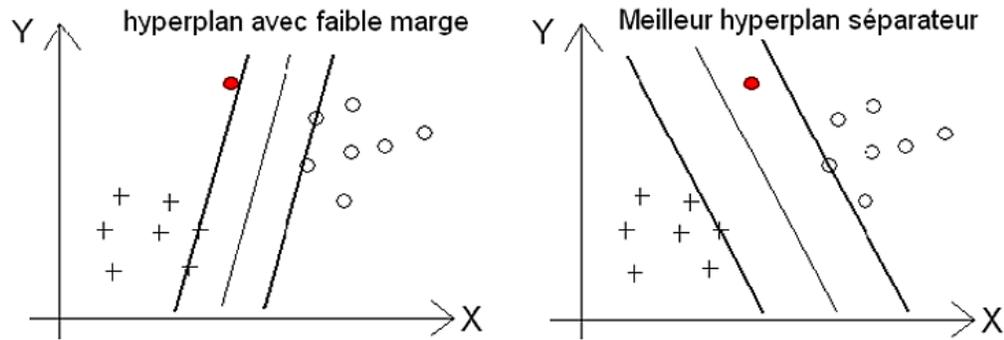


Figure III.5 : Hyperplan optimal par la maximisation de la marge [38].

En général, la classification d'un nouvel exemple inconnu est donnée par sa position par rapport à l'hyperplan optimal. Dans le schéma suivant, le nouvel élément sera classé dans la catégorie des « + ».

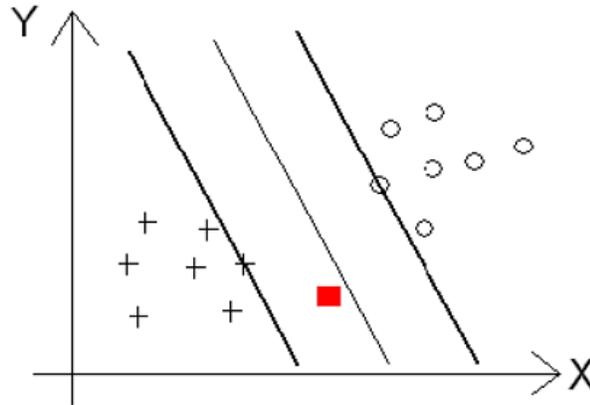


Figure III.6 : Maximisation de la marge pour obtenir l'hyperplan optimal [45].

III.3.2.3 Fonction noyau

Les fonctions du noyau sont utilisées pour cartographier efficacement les données d'entrée qui peuvent ne pas être séparées linéairement en un espace de grande dimension, où des méthodes linéaires peuvent alors être appliquées [45].

Afin de résoudre le problème de données non linéaires, la fonction noyau joue le rôle central de liaison des vecteurs d'entrées à l'espace de caractéristiques de grande dimension [35].

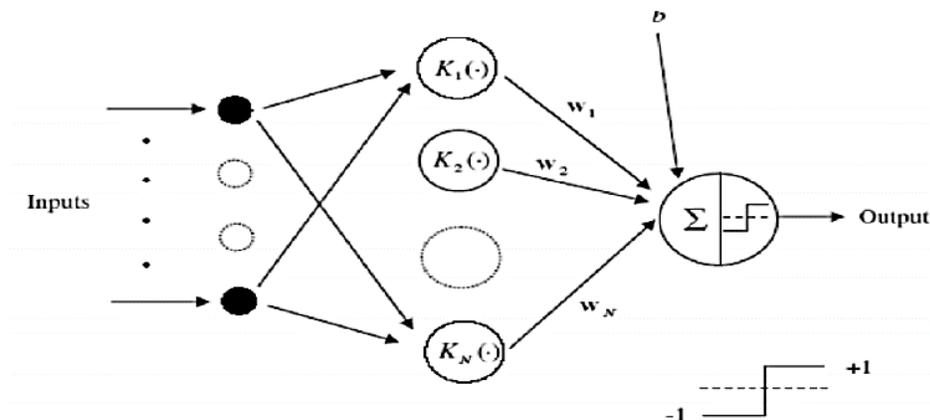


Figure III.7 : Architecture d'une machine à vecteurs de support [36].

III.3.3 Machines à vecteurs supports linéaires

III.3.3.1 Cas des données séparables

Parmi les modèles des SVM, on constate les cas linéairement séparables et les cas non linéairement séparable. Les premiers sont les plus simple de SVM car ils permettent de trouver facilement le classificateur linéaire. Dans la plupart des problèmes réels il n'y a pas de séparation linéaire possible entre les données, le classificateur de marge maximale ne peut pas être utilisé car il fonctionne seulement si les classes de données d'apprentissage sont linéairement séparables [38].

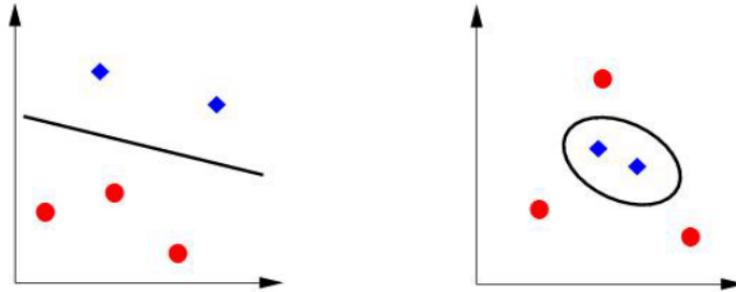


Figure III.8 : Cas linéairement séparable et non linéairement séparable [38].

Dans le cas de données séparables, il existe une infinité d'hyperplans permettant de séparer les deux classes, comme l'illustre la figure suivante [33].

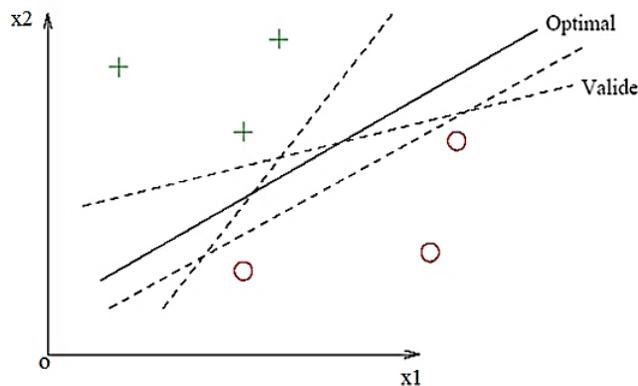


Figure III.9 : Exemples d'hyperplans séparateurs dans R^2 [33].

La séparabilité des données implique que la contrainte suivante est remplie pour chaque exemple.

$$\{(w \cdot x_i) + b \leq -1 \text{ si } y_i = -1\} \quad (\text{III.03})$$

$$\{(w \cdot x_i) + b \geq 1 \text{ si } y_i = 1\} \quad (\text{III. 04})$$

Ce qui équivalent à :

$$y_i(w \cdot x_i + b) \geq 1 \text{ pour } i = 1, \dots, n \quad (\text{III. 05})$$

Nous n'allons plus nous contenter d'en trouver un hyperplan séparateur, mais nous allons en plus chercher parmi ceux-ci l'hyperplan optimal. Selon la théorie de Vapnik, l'hyperplan optimal est celui qui maximise la marge $M(H)$ (voir figure III.3) [33].

III.3.3.2 Cas des données non Séparable

Nous considérons ici le cas où des exemples sont mal classés par l'hyperplan optimal. Cela peut résulter du bruit dans les données. Pour résoudre ce problème, Courte et Vapnik en 1995 ont introduit la notion de « marge souple » (soft margin) qui correspond toujours à la recherche d'un hyperplan de marge optimale, mais avec une règle d'exception qui autorise que quelques exemples soient à une distance plus faible de l'hyperplan que la marge correspondante [33].

L'idée de la marge souple est de rechercher l'hyperplan de marge optimale pénalisée par l'importance des variables ressorts. Le terme de marge souple vient du fait que l'on peut considérer que les exemples pour les quels $\xi_i > 0$, ont une marge géométrique réduite de $d(1 - \xi_i)$. Le terme de pénalisation est de la forme $C \sum \xi_i$ avec C une constante qui permet de définir l'importance de la pénalisation avec $\xi_i = \max(1 - y_i \cdot f(x_i), 0)$ un indice mesurant l'importance de la pénétration de l'exemple x_i dans la zone définie par l'hyperplan H de marge géométrique 'd' (voir figure 9). Cette variable est appelée variable d'écart (Slack variable) et si $\xi_i > 1$, l'exemple n'est pas du bon côté de l'hyperplan relativement à sa classe [33].

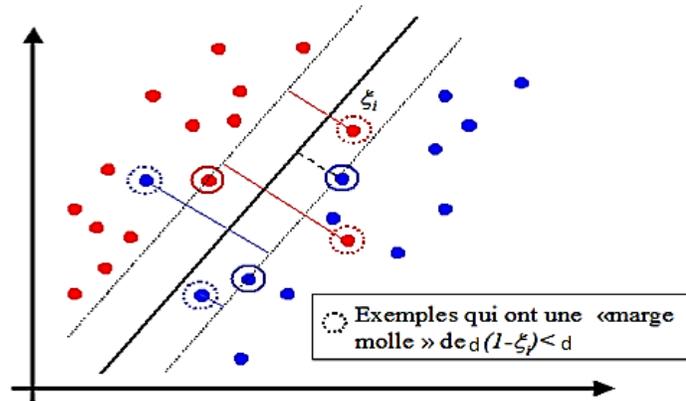


Figure III.10 : Marge souple et Slack variable ξ [38].

Pour vaincre les inconvénients des cas non linéairement séparable, l'idée des SVM est de changer l'espace des données. La transformation non linéaire des données peut permettre une séparation linéaire des exemples dans un nouvel espace. Ce changement de dimension appelée « espace de re-description ». En effet, plus la dimension de l'espace de re-description est grande, plus la probabilité de pouvoir trouver un hyperplan séparateur entre les exemples est élevée. Ceci est illustré par le schéma suivant [38] :

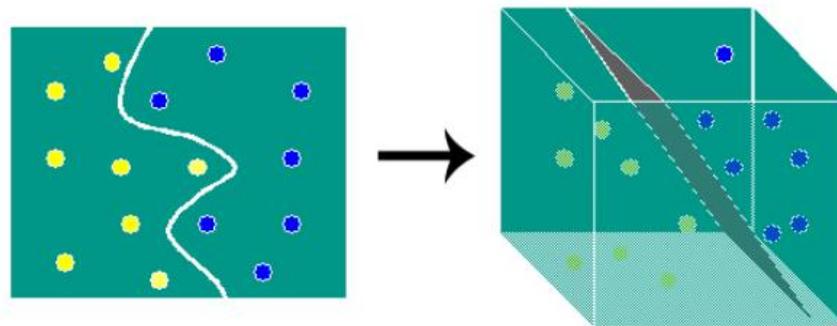


Figure III.11 : transformation d'un problème de séparation non linéaire en un problème de séparation linéaire dans un espace de re-description [33].

Cette transformation non linéaire est réalisée via une fonction noyau. En pratique, quelques familles de fonctions noyau paramétrables sont connues et il revient à l'utilisateur de SVM d'effectuer des tests pour déterminer celle qui convient le mieux pour son application. On peut citer les exemples de noyaux suivants : polynomiale, gaussien, sigmoïde et laplacien [38].

III.4 Le réseau de neurone convolutif (CNN)

Les réseaux de neurones convolutifs (Convolutional Neural Network ou ConvNets ou CNNs) sont une catégorie de réseaux neuronaux qui se sont connues très efficaces dans des domaines tels que la reconnaissance d'image et la classification. ConvNets ont réussi à identifier les visages, les objets et les panneaux de signalisation en dehors de la vision de puissance dans les voitures autonomes et les robots. Le terme "convolutional" vient de l'opération de convolution de matrices utilisée dans le traitement des images. Deux nouveaux types de couche ont été ajoutés dans le réseau : la couche convolution (convolutional layer) et la couche de mise en commun (pool layer) [47].

III.4.1 L'architecture LeNet (années 1990)

LeNet a été l'un des tout premiers réseaux neuronaux convolutionnels qui ont aidé à propulser le domaine de l'apprentissage profond. Cette œuvre pionnière de Yann LeCun a été nommée LeNet5 d'après de nombreuses itérations réussies précédentes depuis l'année 1988. À cette époque, l'architecture LeNet était principalement utilisée pour des tâches de reconnaissance de caractères telles que la lecture de codes postaux, de chiffres, etc.

La figure ci-dessous montre une intuition de la façon dont l'architecture LeNet apprend à reconnaître les images. Il y a eu plusieurs nouvelles architectures proposées au cours des dernières années qui sont des améliorations sur le LeNet, mais ils utilisent tous les concepts principaux du LeNet.

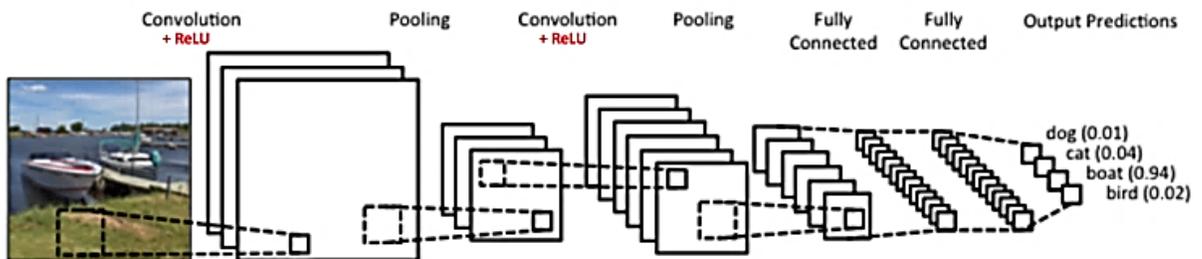


Figure III.12 : Un simple ConvNet [47]

Le réseau neuronal convolutionnel de la (figure III.14) est similaire en architecture à l'original LeNet et classe une image d'entrée en quatre catégories : chien, chat, bateau ou oiseau (le LeNet original a été utilisé principalement pour des tâches de reconnaissance de caractères).

Comme il ressort de la figure ci-dessus, lors de la réception d'une image de bateau comme entrée, le réseau attribue correctement la probabilité la plus élevée pour le bateau (0,94) parmi les quatre catégories [48].

III.4.2 Les différentes couches d'un CNN

III.4.2.1 La couche de convolution (CONV)

La couche de convolution est la composante clé des réseaux de neurones convolutifs et constitue au moins leur première couche, elle traite les données d'un champ récepteur. Dans le but de repérer la présence d'un ensemble de fonctionnalités dans les images reçues en entrée. On réalise un filtrage par convolution, le principe est de faire "glisser" une fenêtre représentant la fonctionnalité sur l'image puis calculer le produit de convolution entre la fonctionnalité et chaque portion de l'image balayée.

La couche de convolution reçoit donc en entrée plusieurs images, et calcule la convolution de chacune d'entre elles avec chaque filtre. Les filtres correspondent exactement aux fonctionnalités que l'on souhaite retrouver dans les images.

On obtient pour chaque paire (image, filtre) une carte d'activation ou (feature map), qui nous indique où se situent les fonctionnalités dans l'image : plus la valeur est élevée, plus l'endroit correspondant dans l'image ressemble à la fonctionnalité [49].

Trois paramètres permettent de dimensionner le volume de la couche de convolution : la profondeur, le pas et la marge [50].

III.4.2.2 La couche de pooling (POOL)

Un deuxième outil très puissant utilisé par les CNNs s'appelle le Pooling, cette couche permet de compresser l'information en réduisant la taille de l'image intermédiaire (souvent par sous-échantillonnage). Le Pooling est une méthode permettant de prendre une large image et d'en réduire la taille tout en préservant les informations les plus importantes qu'elle contient. Les mathématiques derrière la notion de pooling ne sont pas très complexes. En effet, il suffit de faire glisser une petite fenêtre pas à pas sur toutes les parties de l'image et de prendre la valeur maximum de cette fenêtre à chaque pas. En pratique, on utilise souvent une fenêtre de 2 ou 3 pixels de côté et une valeur de 2 pixels pour ce qui est de la valeur d'un pas. Après avoir procédé au pooling, l'image n'a plus qu'un quart du nombre de ses pixels de départ. Parce qu'il garde à chaque pas la valeur maximale contenue

dans la fenêtre, il préserve les meilleures caractéristiques de cette fenêtre. Cela signifie qu'il ne se préoccupe pas vraiment d'où a été extraite la caractéristique dans la fenêtre. Le résultat est que le CNN peut trouver si une caractéristique est dans une image, sans se soucier de l'endroit où elle se trouve. Cela aide notamment à résoudre le problème lié au fait que les ordinateurs soient hyper-littéraires. La couche de pooling permet de réduire le nombre de paramètres et de calculs dans le réseau. On améliore ainsi l'efficacité du réseau et on évite le sur-apprentissage [49,51].

Au final, une couche de pooling est simplement un traitement de pooling sur une image ou une collection d'images. L'output aura le même nombre d'images mais chaque image aura un nombre inférieur de pixels. Cela permettra ainsi de diminuer la charge de calculs. Par exemple, en transformant une image de 8 mégapixels en une image de 2 mégapixels, ce qui nous rendra la vie beaucoup plus facile pour le reste des opérations à effectuer par la suite [51].

III.4.2.3 La couche de correction (ReLU)

Un élément important dans l'ensemble du processus est l'Unité linéaire rectifiée ou ReLU (Rectified Linear Units). Les mathématiques derrière ce concept sont assez simples encore une fois : chaque fois qu'il y a une valeur négative dans un pixel, on la remplace par un 0. Ainsi, on permet au CNN de rester en bonne santé (mathématiquement parlant) en empêchant les valeurs apprises de rester coincer autour de 0 ou d'explorer vers l'infinie. ReLU désigne la fonction réelle non-linéaire définie par $\text{ReLU}(x)=\max(0,x)$ [51].

La couche de correction ReLU remplace donc toutes les valeurs négatives reçues en entrées par des zéros. Elle joue le rôle de fonction d'activation [49].

III.4.2.4 La couche entièrement connectée (FC)

Les couches entièrement connectées qui sont de type perceptron, prennent les images filtrées de haut niveau et les traduisent en votes. Ce type de couche reçoit un vecteur en entrée et produit un nouveau vecteur en sortie. Pour cela, elle applique une combinaison linéaire puis éventuellement une fonction d'activation aux valeurs reçues en entrée [49,51].

La dernière couche fully-connected permet de classifier l'image en entrée du réseau : elle renvoie un vecteur de taille M, où M est le nombre de classes dans notre problème de classification d'images. Chaque élément du vecteur indique la probabilité pour l'image en entrée d'appartenir à une classe.

Par exemple, si le problème consiste à distinguer les chats des chiens, le vecteur final sera de taille 2 : le premier élément (respectivement, le deuxième) donne la probabilité d'appartenir à la classe "chat" (respectivement "chien"). Ainsi, le vecteur [0.90.1] signifie que l'image à 90% de chances représente un chat. Chaque valeur du tableau en entrée "vote" en faveur d'une classe. Les votes n'ont pas tous la même importance : la couche leur accorde des poids qui dépendent de l'élément du tableau et de la classe [49].

Pour calculer les probabilités, la couche fully-connected multiplie donc chaque élément en entrée par un poids, fait la somme, puis applique une fonction d'activation (logistique si $N=2$, softmax si $N>2$), Ce traitement revient à multiplier le vecteur en entrée par la matrice contenant les poids. Le fait que chaque valeur en entrée soit connectée avec toutes les valeurs en sortie explique le terme entièrement connecté.

III.4.2.5 Couche de perte (LOSS)

Elle est normalement la dernière couche dans le réseau. Diverses fonctions de perte adaptées à différentes tâches peuvent y être utilisées. La fonction « Softmax sur les classes de sortie [50].

III.4.3 Choix des paramètres

Les CNNs utilisent plus de paramètres qu'un perceptron multicouches (MLP Multi Layer Perceptron) standard. Même si les règles habituelles pour les taux d'apprentissage et des constantes de régularisation s'appliquent toujours, il faut prendre en considération les notions de nombre de filtres, leur forme et la forme du max pooling.

III.4.3.1 Nombre de filtres

Comme la taille des images intermédiaires diminue avec la profondeur du traitement, les couches proches de l'entrée ont tendance à avoir moins de filtres tandis que les couches plus proches de la sortie peuvent en avoir davantage. Pour égaliser le calcul à chaque couche, le produit du nombre de caractéristiques et le nombre de pixels traités est généralement choisi pour être à peu près constant à travers les couches.

III.4.3.2 Forme de filtres

Ils sont généralement choisis en fonction de l'ensemble de données. Les meilleurs résultats sur les images de MNIST (28x28) sont habituellement dans la gamme de 5x5 sur la première couche, tandis que les ensembles de données d'images naturelles (souvent avec des centaines de

pixels dans chaque dimension) ont tendance à utiliser de plus grands filtres de première couche. Le défi est donc de trouver le bon niveau de granularité de manière à créer des abstractions à l'échelle appropriée et adaptée à chaque cas [50].

III.4.3.3 Conception du classificateur : Il y a deux grands processus qui englobent la conception d'un classificateur :

a. Training : c'est le processus le plus important parce qu'on va créer notre modèle grâce à des configurations précises.

- La Dataset : c'est une base de données d'images répertoriées en classes. Par exemple, si on prend une dataset d'animaux domestiques, on peut trouver des classes comme 'chien', 'chat', 'coq', 'oiseau' ... etc.

- Labels classes : c'est un fichier texte qui portera les noms des classes de notre base. Par exemple on trouvera un fichier comportera 'la classe chien', 'la classe chat' ...etc.

- Cnn et paramètres : c'est notre algorithme de création d'un réseau neurones convolutionnels qui sera configuré avec des paramètres, par exemple : nombre de filtre, nombre de couche, nombre d'époques ...etc.

On va exécuter la base sur notre algorithme CNN qui sera paramétrée pour générer un modèle puis on va utiliser ce modèle pour le Test.

b. Test : Dans le processus Test on retrouve :

- L'image : c'est l'entrée pour les tests. Ce sont plusieurs images ou une seule.
- Le modèle : c'est un fichier généré dans notre training.
- L'affichage de la classification : son nom résume son travail, on va afficher le résultat sorti du modèle qui est le nom d'une classe [52].

III.4.4 Les types de convolutions

Une convolution est la façon dont l'entrée est modifiée par un filtre. Dans les réseaux convolutionnels, plusieurs filtres sont pris pour couper l'image et les cartographier un par un et apprendre différentes parties d'une image d'entrée [53].

Notez qu'une image est en 2 dimensions avec la largeur et la hauteur. Si l'image est colorée, il est considéré comme ayant une dimension de plus pour la couleur RGB. Pour cette raison, les convolutions 2D sont généralement utilisées pour les images en noir et blanc, tandis que les convolutions 3D sont utilisées pour les images colorées [53,54].

III.4.4.1 Convolution en 2D

Commençons par une image d'entrée (4 x 4) sans rembourrage et nous utilisons un filtre de convolution (3 x 3) pour obtenir une image de sortie.

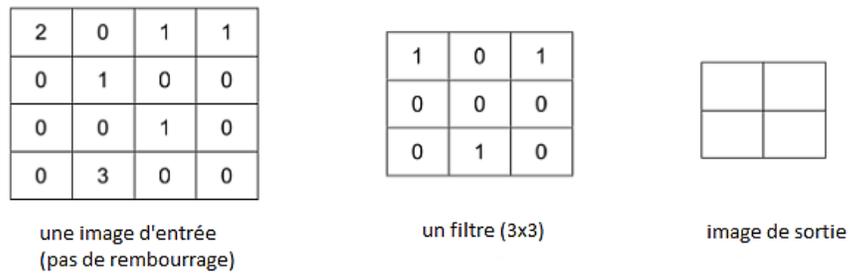


Figure III.13 : Une image de (4x4) avec un filtre de (3x3) [53]

La première étape consiste à multiplier la région jaune dans l'image d'entrée avec un filtre. Chaque élément est multiplié par un élément dans l'emplacement correspondant. Ensuite, vous summez tous les résultats, qui sont une valeur de sortie [48,53].

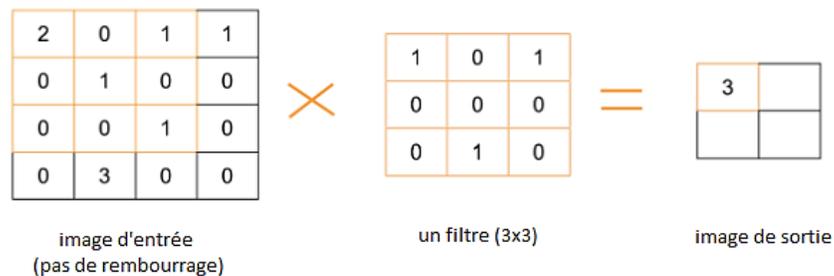


Figure III.14 : la première étape de convolution 2D [53]

Mathématiquement, c'est $(2 * 1) + (0 * 0) + (1 * 1) + (0 * 0) + (1 * 0) + (0 * 0) + (0 * 0) + (0 * 1) + (1 * 0) = 3$

Ensuite, nous répétons la même étape en déplaçant le filtre par une colonne. Et nous obtenons la deuxième et la troisième sortie. Notez que nous avons déplacé le filtre d'une seule colonne. La taille de l'étape lorsque le filtre glisse sur l'image est appelée foulée. Ici, la foulée est 1. Une taille de foulée supérieure à 1 réduira toujours l'image. Si la taille est 1, la taille de l'image restera la même [48,53].

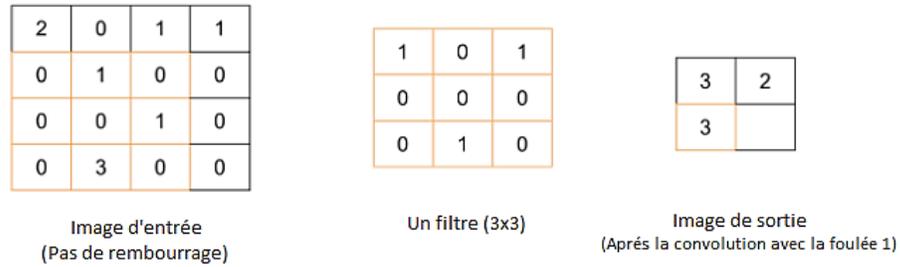


Figure III.15 : déplacement de filtre par colonne [53]

Enfin, nous obtenons la sortie finale.

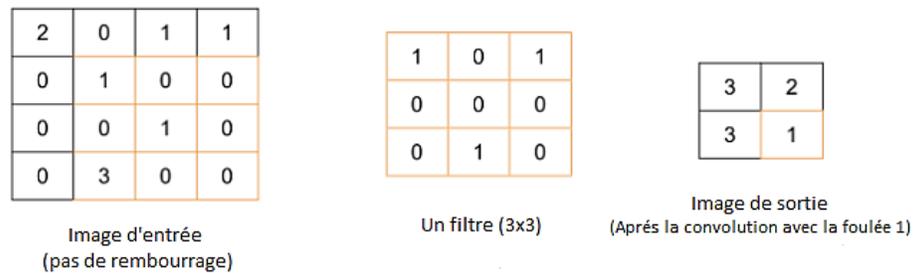


Figure III.16 : La dernière étape de convolution [53]

Nous voyons que la taille de l'image de sortie est plus petite que celle de l'image d'entrée. En fait, c'est vrai dans la plupart des cas.

III.4.4.2 Convolution en 3D

Convolution en 3D est tout comme la 2D, sauf que nous faisons le travail 2D 3 fois, parce qu'il y a 3 canaux de couleur.

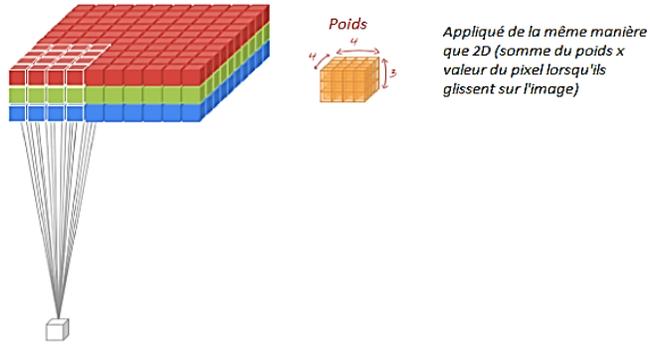


Figure III.17 : Convolution en 3D [53]

III.4.4.3 Convolution 1D

La convolution 1D est couverte ici, parce qu'elle est généralement sous-expliquée, mais elle a des avantages remarquables. Elle est très efficace pour dériver des entités à partir d'un segment de longueur fixe de l'ensemble de données global, ou il n'est pas si important que l'entité se trouve dans le segment. Elle fonctionne bien pour l'analyse d'une série chronologique de données de capteurs et données de signal sur une période de longueur fixe comme l'enregistrement audio. La différence entre CNN 1D et un CNN 2D réside dans la structure des données d'entrées et dans la manière où le filtre appelé noyau de convolution se déplace sur les données en plus les réseaux 1D permettent d'utiliser des filtres de plus grande taille, par exemple un filtre de taille 7 ou 9 ne contient que 7 ou 9 vecteurs de caractéristiques, alors que dans un CNN 2D un filtre de taille 7 contiendra 49 vecteurs ce qui en fait une sélection très large [55].

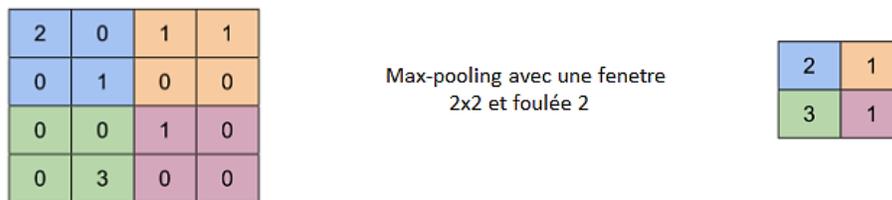


Figure III.18: Max-pooling avec un filtre 2x2 [53]

III.5 L'hybridation du CNN avec le modèle SVM

Un réseau de neurones convolutifs permet d'extraire des caractéristiques optimisées pour un problème donné sur ces différentes couches et plus la couche est profonde, plus elle a été convoluée

et sous-échantillonnée. La profondeur de la couche d'extraction définit donc une résolution de caractéristiques. Nous souhaitons de bénéficier par les avantages des deux modèles SVM et CNN pour améliorer la performance de prédiction du composant global horaire multi-horizon. Nous allons maintenant connecter le réseau de SVM aux différentes couches du CNN. Le réseau SVM reçoit en entrée des nouvelles entrées de caractéristiques du modèle CNN, ces entrées sont divisées en deux catégories la première est appelée les données d'entraînement à pour objectif d'entraîner le modèle SVM est estimé ces paramètres est la deuxième appelée ensemble de teste pour tester la performance du modèle par les critères statistiques.

Le modèle CNN-SVM qui consiste à combiner le CNN et le modèle SVM avec l'utilisation toujours du K_t comme sortie. Le principe de fonctionnement du modèle CNN-SVM est le suivant :

- Après la collecte et le traitement des données, ces derniers sont introduits dans le modèle CNN afin de générer de nouvelles données par utilisation des différents filtres (nombre de filtre, longueur).
- Une fois le meilleur modèle CNN a été sélectionné, les données générées sont utilisées par le modèle SVM pour la prédiction du rayonnement solaire.

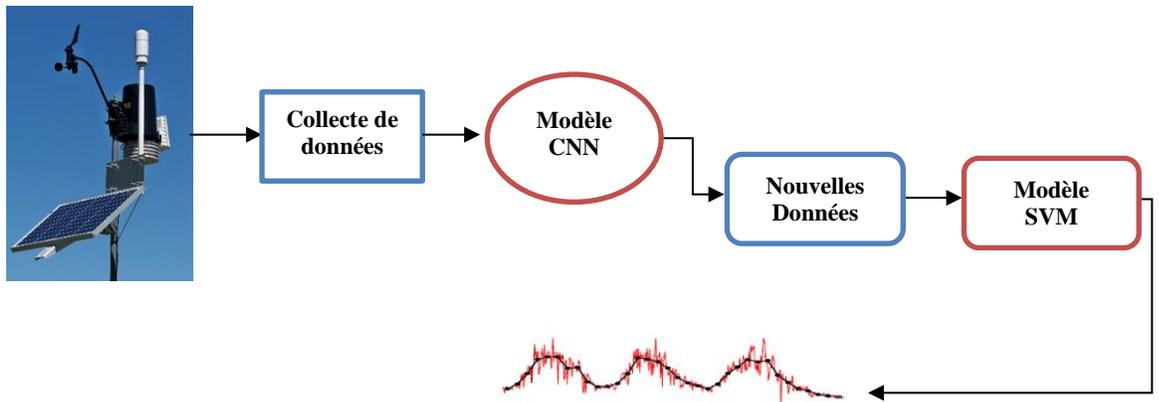


Figure III.19 : Modèle CNN-SVM

III.6 Performances et évaluation des prédicteurs

Dans ce travail, la performance des modèles proposées est évaluée en comparant les valeurs mesurées avec les valeurs estimée à l'aide de différents indicateurs statistiques comme le

coefficient de corrélation (r), l'erreur quadratique moyenne (RMSE), l'erreur relative moyenne (rRMSE). Ces outils sont exposés dans ce qui suit :

III.6.1 Racine de l'erreur moyenne quadratique (RMSE)

Le RMSE représente la différence entre les valeurs prédites estimées par le modèle et les valeurs mesurées. En fait, RMSE identifie la précision du modèle calculée par [56] :

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{Y} - Y)^2} \quad (\text{III. 06})$$

Où \hat{Y} est la valeur estimée et Y est la valeur mesurée [57].

III.6.2 Erreur relative moyenne (rRMSE)

Le rRMSE est calculé en divisant le RMSE par la moyenne des données mesurées comme suit [57] :

$$\text{rRMSE} = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{Y} - Y)^2}}{\frac{1}{N} \sum_{i=1}^n Y} \times 100 \quad (\text{III. 07})$$

Où N est le nombre de données et les gammes de rRMse définissent les performances du modèle comme suit [57] :

Excellent si : $\text{rRMSE} < 10\%$,

Bien si : $10\% < \text{rRMSE} < 20\%$,

Juste si : $20\% < \text{rRMSE} < 30\%$,

Mauvais si : $\text{rRMSE} > 30\%$.

III.6.3 Coefficient de corrélation (r)

Le coefficient de corrélation [58] peut être utilisé pour déterminer la relation linéaire entre les valeurs mesurées et les valeurs estimées, qui peuvent être calculées à partir de l'équation suivante [56,59] :

$$r = \frac{\sum_{i=1}^n (\hat{Y} - \bar{\hat{Y}}) \cdot (Y - \bar{Y})}{\sum_{i=1}^n (\hat{Y} - \bar{\hat{Y}}) \cdot \sum_{i=1}^n (Y - \bar{Y})} \quad (\text{III. 08})$$

III.6.4 L'erreur de biais absolu moyenne (MABE)

Il donne la valeur absolue moyenne de l'erreur de biais [60].

$$\text{MABE} = \frac{1}{n} \sum_{i=1}^n |\hat{Y} - Y| \quad (\text{III. 09})$$

Les entrées sont montrées dans l'annexe B.

III.7 Conclusion

Dans le présent chapitre, nous avons revu les différentes approches et fondements permettant la prédiction et la prévision du rayonnement solaire. Le modèle SVM est le modèle le plus exploité dans la famille des modèles de l'intelligence artificielle.

Récemment, de nouvelles techniques ont été développées à partir de Deep Learning, qui ont influencé plusieurs études sur le traitement du signal spécialement la prédiction des séries temporelles. Pour cela, on a développé un nouveau modèle basé sur l'apprentissage profond appelé CNN-1D pour la prédiction du rayonnement global horaire.

En outre, les outils statistiques permettant l'évaluation des performances des prédicteurs ont été aussi synthétisés. Le prochain chapitre va détailler la méthodologie de combinaison et d'hybridation entre des approches différentes dont le souhait d'améliorer la précision du prédicteur développé.

CHAPITRE IV

Résultats et Discussion

IV.1 Introduction

Dans cette partie de travail, nous allons aborder la prédiction du rayonnement global horaire multi-horizon reçu sur un plan horizontal. Pour cela, le site de Ghardaïa ou les données sont disponibles à notre possession sont utilisés pour validée notre approche. Dans le chapitre précédent, les différents modèles de prédiction ont été revus. Par la suit, nous détaillons en premier lieu, les démarches à suivre pour la prédiction du rayonnement global par ces modèles, avant de présenter tous les résultats obtenus par l'application du modèle SVM et CNN et l'hybridation entre ces deux derniers pour la prédiction du rayonnement solaire sur le site de Ghardaïa. Une comparaison des résultats obtenus aura lieu par la suit.

IV.2 Prédiction du rayonnement global horaire

Les données de rayonnement global horizontal utilisées dans cette étude ont été collectées sur le site de Ghardaïa. Les données instantanées du rayonnement global horizontal (G_h) sont mesurées chaque dix minutes, les mesures relatives aux nuits sont éliminées pour plus de lisibilité. Ensuite, une intégration des données instantanées a été faite afin d'obtenir les données horaires.

Dans cette étude, nous avons travaillé sur les données de cinq ans (2013-2017) pour l'établissement des modèles. Figure IV.1 montre l'évolution du rayonnement global horaire pour notre site. On peut noter que le comportement du rayonnement global ressemble à celui des signaux périodiques bruité par conséquent, la série de données du rayonnement global a besoin d'être stationnaires.

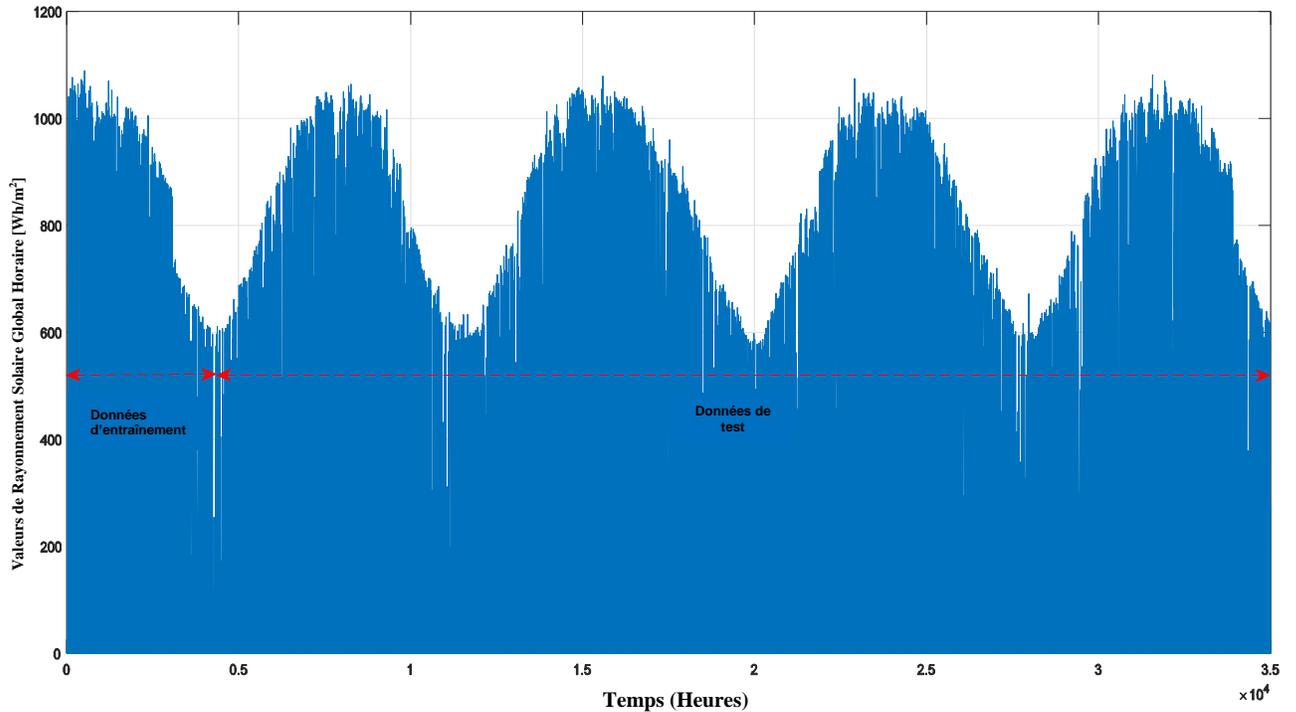


Figure IV.1 : Variation du rayonnement global horaire à Ghardaïa entre 2013-2017.

Cependant, la prédiction du rayonnement solaire à la surface de la Terre est perturbée par la non-stationnarité du signal et la périodicité due aux phénomènes atmosphériques. Effectivement, un prétraitement des données d'entrée peut considérablement contribuer à améliorer l'exactitude des prévisions. Il existe donc, plusieurs façons de traiter une série non stationnaire, comme le cas du rayonnement solaire, pour la mettre dans une forme appropriée. Dans notre cas, nous avons utilisé l'indice de clarté comme moyen de stationnarisation, dans le but d'éliminer la saisonnalité de la ressource, c'est-à-dire la composante déterministe. Par la suite on montre les résultats de prédiction avec et sans stationnarisation du signal pour voir l'effet de la saisonnalité sur la performance des modèles.

IV.3 Prédiction par le modèle SVM

Après avoir stationnariser la série du rayonnement global par le passage par l'indice de clarté K_t , nous expliquerons notre démarche pour la prédiction du rayonnement global. Cependant, l'implémentation du modèle SVM passe par la division de la base de données en trois ensembles, ensemble d'entraînement pour entraîner le modèle, ensemble de validation pour sélectionner le meilleur modèle et ensuite l'ensemble de teste pour tester la performance de prédiction du modèle.

IV.4 Résultats du modèle SVM

Comme on a déjà mentionné pour voir l'effet de la saisonnalité du signal nous avons utilisé le signal global horaire comme sortie du modèle dans la première partie des expériences et dans la deuxième partie on a utilisé l'indice de clarté K_t comme sortie du modèle SVM. Dans toutes les expériences plusieurs combinaisons des données météorologiques sont testées afin de choisir la meilleure combinaison.

Les résultats sont reportés au Tableau IV.1 et Tableau IV.2. Ces tableaux montrent que les valeurs t-statistiques de tous les paramètres des deux modèles ARMA obtenus sont supérieures à 2. Ceci confirme que ces paramètres sont significativement différents de zéro à 5 %.

Pour la prédiction du rayonnement solaire on peut utiliser le cas global et l'indice de clarté K_t comme illustré aux deux tableaux suivants, ensuite on va choisir le meilleur résultat entre ces deux derniers :

Tableau IV.1 : Résultats de prédiction (cas global horaire)

Modèle SVM	Indicateur statistiques				
	Données Météorologiques	RMSE (Wh/m ²)	rRMSE	MABE	r (%)
T_{moy}	298,5976	0,6202	256,3398	0,2071	0.0326
T_{min}	300,0384	0,6232	256,8941	1,1928	0.0294
T_{max}	299,9559	0,6230	256,5613	0,1971	0.0355
$T_{min} - T_{max}$	287,9797	0,5981	240,3231	0,3334	0.0851
$T_{moy} - T_{min}$	290,2909	0,6029	246,5495	0,2990	0.0846
$T_{moy} - T_{max}$	307,6647	0,6390	243,7297	0,2504	-0.0172
$T_{moy} - T_{min} - T_{max}$	312,8263	0,6497	240,7868	0,2525	-0.1131
HR_{moy}	299,6391	0,6223	255,1746	0,1832	0.0250
HR_{min}	299,1250	0,6213	254,8200	0,1895	0.0290
HR_{max}	300,0385	0,6232	255,4842	0,1784	0.0228
$HR_{min} - HR_{max}$	289,5478	0,6014	242,0565	0,3171	0.0891
$HR_{moy} - HR_{min}$	290,9519	0,6043	244,9739	0,2936	0.0750
$HR_{moy} - HR_{max}$	294,3076	0,6113	246,6965	0,2744	0.0564
$HR_{moy} - HR_{min} - HR_{max}$	291,7555	0,6060	244,7166	0,2968	0.0734
Pr_{moy}	303,5850	0,6305	258,9600	0,1670	-0.0177
Pr_{min}	306,5040	0,6366	261,2230	0,1430	-5.9304
Pr_{max}	306,3783	0,6363	260,8941	0,1529	-0.0155

$Pr_{\min} - Pr_{\max}$	3.8177e+03	7.9291	883.2936	0.0500	-70.9703
$Pr_{\text{moy}} - Pr_{\min}$	2.4066e+04	49.9828	3.6372e+03	- 0.0078	-1.9868e+03
$Pr_{\text{moy}} - Pr_{\max}$	2.3341e+03	4.8477	642.9276	0.0202	-92.3729
$Pr_{\text{moy}} - Pr_{\min} - Pr_{\max}$	3.5022e+03	7.2738	864.1417	0.0500	-203.9882
$T_{\text{moy}} - HR_{\text{moy}} - Pr_{\text{moy}}$	324.7916	0.6746	264.5571	0.1904	-0.1724
$T_{\text{moy}} - HR_{\text{moy}}$	306.3711	0.6363	258.4717	0.1700	-0.0078
$T_{\text{moy}} - Pr_{\text{moy}}$	300.5083	0.6241	251.3266	0.2560	0.0183
$HR_{\text{moy}} - Pr_{\text{moy}}$	312.3668	0.6488	259.6041	0.2044	-0.0360
$T_{\min} - HR_{\min} - Pr_{\min}$	323.3025	0.6715	264.6361	0.1939	-0.1507
$T_{\min} - HR_{\min}$	302.7401	0.6288	257.7270	0.1736	0.0046
$T_{\min} - Pr_{\min}$	300.8587	0.6249	251.8550	0.2522	0.0208
$HR_{\min} - Pr_{\min}$	312.7349	0.6495	259.9607	0.2060	-0.0642
$T_{\max} - HR_{\max} - Pr_{\max}$	326.5114	0.6781	265.0940	0.1942	-0.1717
$T_{\max} - HR_{\max}$	301.8501	0.6269	255.6747	0.2081	-0.0070
$T_{\max} - Pr_{\max}$	300.1950	0.6235	251.5247	0.2502	0.0149
$HR_{\max} - Pr_{\max}$	305.5943	0.6347	255.2803	0.2194	-0.0146
$T_{\text{moy}} - HR_{\min} - Pr_{\max}$	323.1628	0.6712	263.9372	0.1962	-0.1336
$T_{\min} - HR_{\text{moy}} - Pr_{\text{moy}}$	323.0213	0.6709	264.0944	0.1894	-0.1326
$T_{\max} - HR_{\text{moy}} - Pr_{\text{moy}}$	325.9509	0.6770	265.2479	0.1897	-0.1532
$T_{\text{moy}} - HR_{\min} - Pr_{\min}$	326.6493	0.6784	266.2981	0.1927	-0.1582
$T_{\min} - HR_{\text{moy}} - Pr_{\min}$	323.7025	0.6723	264.6637	0.1868	-0.1374
$T_{\text{moy}} - HR_{\text{moy}} - Pr_{\min}$	321.4676	0.6677	262.8176	0.1904	-0.1217
$T_{\text{moy}} - HR_{\max} - Pr_{\max}$	325.7070	0.6765	264.8586	0.1931	-0.1515
$T_{\text{moy}} - HR_{\min} - Pr_{\text{moy}}$	325.6554	0.6764	265.5011	0.1954	-0.1511
$T_{\min} - HR_{\max} - Pr_{\max}$	321.7224	0.6682	263.0896	0.1924	-0.1235
$T_{\max} - HR_{\max} - Pr_{\min}$	329.2398	0.6838	266.6980	0.1913	-0.1766
$T_{\text{moy}} - HR_{\max} - Pr_{\text{moy}}$	325.6130	0.6763	264.5725	0.1954	-0.1508
Dur_{inso}	248.0622	0.5152	196.8229	0.5821	0.3321
$Dur_{\text{inso}} + T_{\min}$	247.1035	0.5132	196.8775	0.5828	0.3372
$Dur_{\text{inso}} + T_{\max}$	244.5416	0.5079	194.9105	0.5933	0.3509
$Dur_{\text{inso}} + T_{\text{moy}}$	246.2259	0.5114	196.2558	0.5863	0.3419
$Dur_{\text{inso}} + T_{\text{moy}} + T_{\min}$	233.0323	0.4840	182.9213	0.6420	0.4106
$Dur_{\text{inso}} + T_{\text{moy}} + T_{\max}$	237.3482	0.4930	184.3234	0.6290	0.3885
$Dur_{\text{inso}} + T_{\min} + T_{\max}$	234.0668	0.4861	181.9453	0.6401	0.4053
$Dur_{\text{inso}} + T_{\text{moy}} + T_{\min} + T_{\max}$	237.2784	0.4928	182.7217	0.6306	0.3889
$Dur_{\text{inso}} + T_{\text{moy}} + T_{\min} + Ev_{\max} + Dec + SS$	164.7283	0.3421	115.0181	0.8617	0.7055

$Dur_{inso} + HR_{moy}$	247.0051	0.5130	194.3793	0.5860	0.3377
$Dur_{inso} + HR_{min}$	246.6621	0.5123	194.2619	0.5874	0.3396
$Dur_{inso} + HR_{max}$	247.1681	0.5134	194.2299	0.5853	0.3369
$Dur_{inso} + HR_{moy} + HR_{min}$	240.3516	0.4992	188.8258	0.6151	0.3729
$Dur_{inso} + HR_{moy} + HR_{max}$	241.1802	0.5009	190.2975	0.6122	0.3686
$Dur_{inso} + HR_{min} + HR_{max}$	238.9342	0.4963	188.6184	0.6204	0.3803
$Dur_{inso} + HR_{moy} + HR_{min} + HR_{max}$	239.7652	0.4980	189.3621	0.6173	0.3760
$Dur_{inso} + HR_{min} + HR_{max} + Ev_{max} + Dec + SS$	129.1688	0.2683	94.5156	0.9065	0.8189
$Dur_{inso} + Pr_{moy}$	251.6216	0.5226	201.4954	0.5673	0.3128
$Dur_{inso} + Pr_{min}$	250.1401	0.5195	200.4141	0.5724	0.3208
$Dur_{inso} + Pr_{max}$	254.0189	0.5276	203.1527	0.5594	0.2996
$Dur_{inso} + Pr_{moy} + Pr_{min}$	274.9640	0.5711	209.0236	0.4945	0.1793
$Dur_{inso} + Pr_{moy} + Pr_{max}$	264.0640	0.5484	206.4369	0.5238	0.2431
$Dur_{inso} + Pr_{min} + Pr_{max}$	251.2952	0.5219	201.4420	0.5725	0.3145
$Dur_{inso} + Pr_{moy} + Pr_{min} + Pr_{max}$	631.6364	1.3119	296.0826	0.2491	-3.3306
$Dur_{inso} + Pr_{min} + Ev_{max} + Dec + SS$	317.9996	0.6605	169.0368	0.5875	-0.0977

Tableau IV.2 : Résultats de prédiction (cas K_t horaire)

Modèle SVM	Indicateur statistiques				
	Données Météorologiques	RMSE (Wh/m ²)	rRMSE	MABE	r (%)
T_{moy}	113.9247	0.2367	87.3765	0.9452	0.8590
T_{min}	114.6681	0.2383	87.4266	0.9433	0.8584
T_{max}	113.7505	0.2364	87.1983	0.9452	0.8594
$T_{min} - T_{max}$	114.3390	0.2376	83.2418	0.9374	0.8586
$T_{moy} - T_{min}$	114.1187	0.2371	83.5135	0.9382	0.8620
$T_{moy} - T_{max}$	113.8111	0.2365	83.7261	0.9386	0.7928
$T_{moy} - T_{min} - T_{max}$	113.8140	0.2365	83.2806	0.9382	0.8596
HR_{moy}	112.2368	0.2332	85.5906	0.9477	0.8632
HR_{min}	111.9678	0.2327	85.4786	0.9481	0.8638
HR_{max}	111.4310	0.2316	85.1558	0.9487	0.8651
$HR_{min} - HR_{max}$	111.2057	0.2311	83.0236	0.9451	0.8656
$HR_{moy} - HR_{min}$	111.2656	0.2312	83.6561	0.9462	0.8667
$HR_{moy} - HR_{max}$	112.0283	0.2328	83.8430	0.9446	0.8643
$HR_{moy} - HR_{min} - HR_{max}$	111.3630	0.2314	83.2625	0.9450	0.8652
Pr_{moy}	114.1532	0.2372	84.1708	0.9390	0.8584
Pr_{min}	114.0048	0.2369	83.8766	0.9390	0.8588

Pr_{max}	114.1634	0.2372	84.1175	0.9389	0.8584
$Pr_{min} - Pr_{max}$	613.3369	1.2746	172.0633	0.3621	-3.0869
$Pr_{moy} - Pr_{min}$	3.8439e+03	7.9880	596.9513	0.0354	-159.5247
$Pr_{moy} - Pr_{max}$	1.4710e+03	3.0568	287.0531	0.1323	-26.6802
$Pr_{moy} - Pr_{min} - Pr_{max}$	809.9191	1.6831	199.0130	0.2740	-4.2515
$T_{moy} - HR_{moy} - Pr_{moy}$	120.3223	0.2500	84.8066	0.9220	0.8420
$T_{moy} - HR_{moy}$	117.0300	0.2432	87.1855	0.9389	0.8514
$T_{moy} - Pr_{moy}$	116.2575	0.2416	83.0957	0.9311	0.8540
$HR_{moy} - Pr_{moy}$	115.6686	0.2404	84.2349	0.9337	0.8543
$T_{min} - HR_{min} - Pr_{min}$	119.9184	0.2492	84.7646	0.9228	0.8467
$T_{min} - HR_{min}$	115.3160	0.2396	86.1937	0.9411	0.8558
$T_{min} - Pr_{min}$	116.3842	0.2419	83.2940	0.9310	0.8519
$HR_{min} - Pr_{min}$	116.1351	0.2413	84.5385	0.9331	0.8534
$T_{max} - HR_{max} - Pr_{max}$	120.5507	0.2505	85.0680	0.9228	0.8420
$T_{max} - HR_{max}$	118.8859	0.2471	88.7604	0.9372	0.8465
$T_{max} - Pr_{max}$	115.2275	0.2395	82.5993	0.9327	0.8556
$HR_{max} - Pr_{max}$	113.9873	0.2369	83.3552	0.9366	0.8590
$T_{moy} - HR_{min} - Pr_{max}$	119.6196	0.2486	84.6068	0.9235	0.8445
$T_{min} - HR_{moy} - Pr_{moy}$	120.3996	0.2502	84.8479	0.9220	0.8425
$T_{max} - HR_{moy} - Pr_{moy}$	120.5409	0.2505	84.9530	0.9218	0.8421
$T_{moy} - HR_{min} - Pr_{min}$	120.4032	0.2502	85.0738	0.9218	0.8425
$T_{min} - HR_{moy} - Pr_{min}$	120.5684	0.2506	84.9799	0.9216	0.8421
$T_{moy} - HR_{moy} - Pr_{min}$	120.4817	0.2504	84.9178	0.9216	0.8423
$T_{moy} - HR_{max} - Pr_{max}$	121.0220	0.2515	85.3887	0.9216	0.8409
$T_{moy} - HR_{min} - Pr_{moy}$	119.1230	0.2475	84.2372	0.9243	0.8458
$T_{min} - HR_{max} - Pr_{max}$	119.6188	0.2486	84.6163	0.9242	0.8445
$T_{max} - HR_{max} - Pr_{min}$	121.3158	0.2521	85.4038	0.9210	0.8401
$T_{moy} - HR_{max} - Pr_{moy}$	121.2966	0.2521	85.4872	0.9209	0.8402
Dur_{inso}	73.3202	0.1524	53.6616	0.9759	0.9416
$Dur_{inso} + T_{min}$	75.1059	0.1561	54.6193	0.9740	0.9387
$Dur_{inso} + T_{max}$	75.0379	0.1559	54.6902	0.9742	0.9388
$Dur_{inso} + T_{moy}$	75.1377	0.1561	54.7354	0.9741	0.9387
$Dur_{inso} + T_{moy} + T_{min}$	77.8300	0.1617	55.6793	0.9695	0.9342
$Dur_{inso} + T_{moy} + T_{max}$	78.1033	0.1623	55.4712	0.9696	0.9337
$Dur_{inso} + T_{min} + T_{max}$	77.3445	0.1607	55.6014	0.9699	0.9350
$Dur_{inso} + T_{moy} + T_{min} + T_{max}$	77.2439	0.1605	55.8511	0.9702	0.9352

$Dur_{inso} + T_{max} + Ev_{max} + Dec + SS$	80.9908	0.1683	55.2555	0.9638	0.9287
$Dur_{inso} + HR_{moy}$	74.1199	0.1540	53.9514	0.9752	0.9403
$Dur_{inso} + HR_{min}$	74.2741	0.1543	54.0767	0.9751	0.9401
$Dur_{inso} + HR_{max}$	74.3101	0.1544	54.0584	0.9750	0.9400
$Dur_{inso} + HR_{moy} + HR_{min}$	75.4243	0.1567	54.1967	0.9728	0.9382
$Dur_{inso} + HR_{moy} + HR_{max}$	75.4844	0.1569	54.6459	0.9729	0.9381
$Dur_{inso} + HR_{min} + HR_{max}$	75.8369	0.1576	54.6296	0.9723	0.9375
$Dur_{inso} + HR_{moy} + HR_{min} + HR_{max}$	76.1729	0.1583	55.0861	0.9722	0.9370
$Dur_{inso} + HR_{moy} + Ev_{max} + Dec + SS$	76.1429	0.1582	51.2402	0.9680	0.9370
$Dur_{inso} + Pr_{moy}$	100.4024	0.2086	61.9208	0.9439	0.8905
$Dur_{inso} + Pr_{min}$	221.8780	0.4611	77.7168	0.7795	0.4652
$Dur_{inso} + Pr_{max}$	614.5989	1.2772	122.4479	0.4242	-3.1038
$Dur_{inso} + Pr_{moy} + Pr_{min}$	86.9837	0.1808	56.9635	0.9601	0.9178
$Dur_{inso} + Pr_{moy} + Pr_{max}$	90.2397	0.1875	59.3649	0.9558	0.9115
$Dur_{inso} + Pr_{min} + Pr_{max}$	652.3698	1.3557	129.3998	0.3914	-3.6237
$Dur_{inso} + Pr_{moy} + Pr_{min} + Pr_{max}$	858.5571	1.7842	154.2975	0.3442	-7.0082
$Dur_{inso} + Pr_{moy} + Pr_{min} + Ev_{max} + Dec + SS$	475.5879	0.9883	107.0172	0.5471	-1.4573

A partir de ces deux tableaux, on peut remarquer que le K_t donne des meilleures performances que le global. Donc, tous nos essais seront basés sur le K_t comme sortie des modèles et les meilleurs paramètres (inputs) qu'on va les utilisées sont 5 inputs (Dur_{inso} HR_{moy} Ev_{max} Dec SS) pour la prédiction de 12 heures qui viennent.

Tableau IV.3 : Résultats du modèle SVM avec la meilleure combinaison.

Horizon	Indicateur statistiques				
Heures	RMSE (Wh/m ²)	rRMSE	MABE	r (%)	Nech
1h	76.1429	0.1582	51.2402	0.9680	0.9370
2h	96.6835	0.2009	65.5120	0.9489	0.8985
3h	110.5968	0.2298	76.4048	0.9354	0.8671
4h	117.0794	0.2433	77.8032	0.9236	0.8511
5h	107.3947	0.2232	74.9918	0.9382	0.8747
6h	98.0704	0.2038	65.4872	0.9465	0.8955
7h	101.4878	0.2109	66.4871	0.9425	0.8881
8h	103.2507	0.2145	69.3841	0.9410	0.8842
9h	109.6289	0.2278	75.0954	0.9339	0.8694

10h	115.0228	0.2390	81.1647	0.9312	0.8563
11h	113.3390	0.2355	79.8835	0.9316	0.8604
12h	113.0380	0.2348	78.8365	0.9299	0.8612

De la même façon nous ajouterons pour notre programme un autre paramètre qui est le global de l'heur précédente comme entrée dans SVM sans utilisation du cas global, dans le but d'améliorer la performance des résultats, le tableau suivant représente les résultats obtenus :

Tableau IV.4 : Résultats avec 6 paramètres de 1h à 12 heures.

Horizon	Indicateur statistiques			
Heures	RMSE (Wh/m²)	rRMSE	MABE	r (%)
1h	71.1809	0.1480	44.9521	0.9723
2h	97.3571	0.2025	63.7812	0.9490
3h	114.5850	0.2383	75.6190	0.9300
4h	117.4083	0.2442	78.2051	0.9258
5h	111.4980	0.2318	80.2035	0.9365
6h	101.8186	0.2117	69.8769	0.9428
7h	108.5746	0.2257	72.0284	0.9344
8h	106.8742	0.2222	72.8899	0.9371
9h	110.4943	0.2297	76.8161	0.9346
10h	125.8888	0.2618	83.0358	0.9103
11h	112.0251	0.2329	73.9409	0.9305
12h	112.8916	0.2347	74.2433	0.9285

Comme on peut le voir, l'ajout de la composante global précédente améliore la performance du modèle SVM.

IV.5 Résultats du modèle CNN

Un réseau CNN-1D basé sur différent filtres a été optimisé pour la prédiction de l'indice de clarté (K_t). Cependant, l'optimisation du CNN est effectuée de la manière suivante (1) choix du nombre des couches, (2) choix du nombre de filtre dans chaque couche et la taille adéquate pour chaque filtre. Plusieurs essais sont effectués afin de sélectionnées l'architecture optimal du modèle CNN.

Donc, après plusieurs essais, un réseau avec deux couches, la première avec 5 filtres de taille 2 et la deuxième avec 3 filtres de taille 1 est retenue comme meilleure architecture pour la prédiction du K_t au site de Ghardaïa (Voir Tableau IV.5).

Tableau IV.5 : Résultats du modèle CNN-1D

Modèle CNN-D1								
Couche + Taille + Nbr.Filtres		mse1	nmse1	t_2	RMSE	rRMSE	MABE	r(%)
1 ^{er} Couche	2eme Couche							
1 filtre, taille 2	1 filtre, taille 1	0.0153	0.4976	259.6398	83.1469	0.1724	57.1011	0.9630
1 filtre, taille 2	2 filtres, taille 1	0.0154	0.4997	415.1262	82.5149	0.1711	55.4836	0.9630
1 filtre, taille 2	3 filtres, taille 1	0.0153	0.4969	447.6726	83.0275	0.1722	56.8300	0.9630
1 filtre, taille 2	4 filtres, taille 1	0.0154	0.5012	532.8255	82.2410	0.1706	55.0164	0.9632
1 filtre, taille 2	5 filtres, taille 1	0.0153	0.4967	616.3259	83.1669	0.1725	57.3431	0.9631
2 filtres, taille 2	1 filtre, taille 1	0.0153	0.4973	372.1146	83.4676	0.1731	58.0845	0.9632
2 filtres, taille 2	2 filtres, taille 1	0.0144	0.4673	490.5373	81.8336	0.1697	54.2834	0.9634
2 filtres, taille 2	3 filtres, taille 1	0.0151	0.4919	605.7531	81.6305	0.1693	55.3112	0.9640
2 filtres, taille 2	4 filtres, taille 1	0.0146	0.4746	706.1912	81.2328	0.1685	54.3641	0.9639
2 filtres, taille 2	5 filtres, taille 1	0.0152	0.4934	904.4897	82.6838	0.1715	57.1410	0.9636
3 filtres, taille 2	1 filtres, taille 1	0.0152	0.4950	504.5325	81.8637	0.1698	54.8253	0.9635
3 filtres, taille 2	2 filtres, taille 1	0.0138	0.4508	631.8706	78.7466	0.1633	53.0938	0.9663
3 filtres, taille 2	3 filtres, taille 1	0.0143	0.4642	769.3771	81.4235	0.1689	55.7014	0.9642
3 filtres, taille 2	4 filtres, taille 1	0.0144	0.4692	906.3607	80.2952	0.1665	55.3593	0.9653
3 filtres, taille 2	5 filtres, taille 1	0.0145	0.4718	1.0715e+03	80.6400	0.1672	55.6435	0.9651
4 filtres, taille 2	1 filtre, taille 1	0.0152	0.4937	642.7975	82.0194	0.1701	54.9144	0.9634
4 filtres, taille 2	2 filtres, taille 1	0.0137	0.4457	825.4033	77.6033	0.1609	53.6525	0.9678
4 filtres, taille 2	3 filtres, taille 1	0.0139	0.4510	1.0012e+03	79.9622	0.1658	53.6003	0.9651
4 filtres, taille 2	4 filtres, taille 1	0.0140	0.4557	1.1383e+03	78.3073	0.1624	52.8786	0.9666
4 filtres, taille 2	5 filtres, taille 1	0.0137	0.4476	1.3311e+03	78.0431	0.1619	52.0381	0.9667
5 filtres, taille 2	1 filtre, taille 1	0.0153	0.4983	711.7470	81.5773	0.1692	54.3250	0.9637
5 filtres, taille 2	2 filtres, taille 1	0.0130	0.4222	941.3463	77.2757	0.1603	51.0620	0.9674
5 filtres, taille 2	3 filtres, taille 1	0.0132	0.4287	1.1439e+03	76.0114	0.1576	51.5134	0.9688
5 filtres, taille 2	4 filtres, taille 1	0.0129	0.4186	1.4408e+03	77.4185	0.1606	51.7359	0.9674
5 filtres, taille 2	5 filtres, taille 1	0.0133	0.4325	1.5691e+03	77.2276	0.1602	53.6475	0.9687

Après avoir sélectionné la structure optimale du modèle CNN-1D, l'étape suivante consiste à utiliser le meilleur ensemble de données pour la prédiction de 12 heures qui viennent de la composante globale avec l'utilisation toujours du K_t comme sortie du modèle.

Le tableau IV.6 montre les résultats obtenus par le modèle CNN avec 5 entrées pour 12 heures après la sélection de la meilleure structure du CNN-1D.

Tableau IV.6 : Résultats du modèle CNN-1D de 1h à 12 heures.

CNN – 1D							
Heures	mse1	nmse1	t2	RMSE	rRMSE	MABE	r(%)
1h	0.0132	0.4287	1.1439e+03	76.0114	0.1576	51.5134	0.9688
2h	0.0209	0.6814	1.0045e+03	103.9633	0.2156	72.4257	0.9426
3h	0.0198	0.6464	1.0471e+03	112.8691	0.2341	76.4900	0.9353
4h	0.0294	0.9572	1.0275e+03	116.6602	0.2420	89.1222	0.9445
5h	0.0243	0.7921	3.8254e+03	108.6790	0.2254	76.9865	0.9385
6h	0.0204	0.6629	1.0549e+03	101.9704	0.2115	70.9807	0.9445
7h	0.0183	0.5965	1.0314e+03	101.0778	0.2096	67.2879	0.9440
8h	0.0205	0.6675	1.0464e+03	105.2867	0.2183	74.2113	0.9413
9h	0.0237	0.7732	1.0387e+03	108.2438	0.2244	76.5932	0.9389
10h	0.0267	0.8722	1.0411e+03	112.9359	0.2341	83.1873	0.9393
11h	0.0283	0.9251	1.0364e+03	115.5264	0.2395	88.8922	0.9455
12h	0.0250	0.8180	1.0844e+03	115.0082	0.2384	86.1692	0.9360

Dans la deuxième partie des expériences, on ajoute la composante globale horaire précédente aux ensembles des inputs. Les résultats de prédiction sont illustrés au tableau IV.7.

Tableau IV.7: Résultats de prédiction du CNN-1D pour une heure.

CNN -1D (Dur _{inso} + HR _{moy} + Ev _{max} + Dec + SS + Kt)								
Couche + Taille + Nbr.Filtres		mse1	nmse1	t2	RMSE	rRMSE	MABE	r(%)
1 ^{er} Couche	2eme Couche							
1 filtre, taille 3	1 filtre, taille 1	0.0153	0.4980	256.9787	83.2835	0.1727	56.4877	0.9625
1 filtre, taille 3	2 filtres, taille 1	0.0152	0.4964	352.7492	83.8804	0.1740	57.9341	0.9625
1 filtre, taille 3	3 filtres, taille 1	0.0153	0.4967	419.7824	83.5519	0.1733	57.0446	0.9624
1 filtre, taille 3	4 filtres, taille 1	0.0153	0.4993	500.3376	82.8829	0.1719	55.4950	0.9626
1 filtre, taille 3	5 filtres, taille 1	0.0153	0.4981	587.3117	83.0704	0.1723	55.8688	0.9625
2 filtres, taille 3	1 filtre, taille 1	0.0127	0.4139	379.1652	72.1050	0.1495	48.9530	0.9718
2 filtres, taille 3	2 filtres, taille 1	0.0126	0.4105	466.9033	71.4696	0.1482	48.0271	0.9723
2 filtres, taille 3	3 filtres, taille 1	0.0124	0.4021	591.7264	71.1172	0.1475	47.7307	0.9726
2 filtres, taille 3	4 filtres, taille 1	0.0123	0.4020	742.1796	71.6522	0.1486	48.3210	0.9723
2 filtres, taille 3	5 filtres, taille 1	0.0125	0.4059	806.5554	72.8596	0.1511	50.3686	0.9720
3 filtres, taille 3	1 filtre, taille 1	0.0131	0.4261	462.6957	70.6052	0.1464	45.7318	0.9734
3 filtres, taille 3	2 filtres, taille 1	0.0123	0.4003	594.4331	70.1582	0.1455	46.6418	0.9733
3 filtres, taille 3	3 filtres, taille 1	0.0121	0.3941	756.9661	71.7676	0.1488	49.1762	0.9727
3 filtres, taille 3	4 filtres, taille 1	0.0120	0.3896	872.3127	71.0634	0.1474	48.0124	0.9729
3 filtres, taille 3	5 filtres, taille 1	0.0122	0.3959	1.0509e+03	70.4012	0.1460	47.2534	0.9732
4 filtres, taille 3	1 filtre, taille 1	0.0127	0.4119	729.7427	70.3076	0.1458	47.1288	0.9732

4 filtres, taille 3	2 filtres, taille 1	0.0117	0.3825	924.3254	68.6653	0.1424	45.2030	0.9744
4 filtres, taille 3	3 filtres, taille 1	0.0120	0.3907	868.4294	70.5981	0.1464	46.9647	0.9730
4 filtres, taille 3	4 filtres, taille 1	0.0120	0.3915	1.0228e+03	70.6413	0.1465	47.2743	0.9730
4 filtres, taille 3	5 filtres, taille 1	0.0121	0.3949	1.1933e+03	70.2777	0.1458	46.1368	0.9731
5 filtres, taille 3	1 filtre, taille 1	0.0126	0.4097	657.9535	71.5669	0.1484	49.6933	0.9731
5 filtres, taille 3	2 filtres, taille 1	0.0125	0.4074	850.3978	70.6229	0.1465	47.6335	0.9730
5 filtres, taille 3	3 filtres, taille 1	0.0117	0.3811	1.0519e+03	71.2370	0.1477	46.5456	0.9725
5 filtres, taille 3	4 filtres, taille 1	0.0120	0.3900	1.2127e+03	72.1530	0.1496	49.5824	0.9726
5 filtres, taille 3	5 filtres, taille 1	0.0120	0.3903	1.3978e+03	71.1304	0.1475	47.5931	0.9727
1 filtre, taille 1	1 filtre, taille 2	0.0162	0.5281	268.4654	89.4568	0.1855	61.5773	0.9569
1 filtre, taille 1	2 filtres, taille 2	0.0161	0.5226	380.7592	90.9423	0.1886	63.7766	0.9561
1 filtre, taille 1	3 filtres, taille 2	0.0161	0.5254	434.2345	89.2260	0.1850	60.3676	0.9564
1 filtre, taille 1	4 filtres, taille 2	0.0162	0.5271	514.3436	89.1143	0.1848	60.3791	0.9566
1 filtre, taille 1	5 filtres, taille 2	0.0162	0.5279	591.9766	88.9872	0.1846	59.9175	0.9566
2 filtres, taille 1	1 filtre, taille 2	0.0160	0.5206	382.4037	91.8234	0.1904	65.7952	0.9572
2 filtres, taille 1	2 filtres, taille 2	0.0155	0.5052	501.1461	90.4759	0.1876	63.6450	0.9564
2 filtres, taille 1	3 filtres, taille 2	0.0153	0.4990	607.8874	88.8308	0.1842	62.0692	0.9574
2 filtres, taille 1	4 filtres, taille 2	0.0151	0.4908	683.1979	88.6999	0.1840	62.7224	0.9583
2 filtres, taille 1	5 filtres, taille 2	0.0158	0.5128	850.7112	89.5938	0.1858	62.2107	0.9568
3 filtres, taille 1	1 filtre, taille 2	0.0161	0.5244	484.0457	87.9241	0.1823	60.3865	0.9582
3 filtres, taille 1	2 filtres, taille 2	0.0158	0.5135	619.1328	88.2707	0.1831	61.3833	0.9580
3 filtres, taille 1	3 filtres, taille 2	0.0147	0.4774	764.0651	87.9272	0.1824	61.3314	0.9582
3 filtres, taille 1	4 filtres, taille 2	0.0153	0.4991	903.7043	89.3483	0.1853	62.5357	0.9571
3 filtres, taille 1	5 filtres, taille 2	0.0148	0.4807	9.6491e+03	88.0500	0.1826	60.9864	0.9581
4 filtres, taille 1	1 filtre, taille 2	0.0161	0.5232	575.6381	87.7616	0.1820	60.4525	0.9585
4 filtres, taille 1	2 filtres, taille 2	0.0151	0.4917	762.4732	88.3816	0.1833	62.2690	0.9586
4 filtres, taille 1	3 filtres, taille 2	0.0150	0.4867	961.2262	88.0012	0.1825	61.2388	0.9580
4 filtres, taille 1	4 filtres, taille 2	0.0148	0.4806	1.1921e+03	89.2975	0.1852	63.7998	0.9586
4 filtres, taille 1	5 filtres, taille	0.0148	0.4826	1.4679e+03	89.1538	0.1849	63.6573	0.9588
5 filtres, taille 1	1 filtre, taille 2	0.0161	0.5254	691.1983	87.8326	0.1822	60.0803	0.9581
5 filtres, taille 1	2 filtres, taille 2	0.0151	0.4927	917.3142	87.4816	0.1814	60.8972	0.9588
5 filtres, taille 1	3 filtres, taille 2	0.0153	0.4980	1.1297e+03	88.9739	0.1845	62.1415	0.9575
5 filtres, taille 1	4 filtres, taille 2	0.0144	0.4699	1.3339e+03	88.2432	0.1830	62.4640	0.9593
5 filtres, taille 1	5 filtres, taille 2	0.0146	0.4739	1.5310e+03	86.8394	0.1801	59.0899	0.9588

D'après le tableau IV.7 on peut voir que l'architecture optimale du modèle contenait 4 filtres de taille 3 dans la première couche et 2 filtres de taille 1 dans la deuxième couche. L'étape suivante consiste à utiliser cette architecture avec l'ensemble sélectionnées des données météorologiques pour la prédiction des 12 heures qui viennent. Les résultats sont montrés dans le tableau suivant (voir tableau IV.8).

Tableau IV.8 : Résultats du modèle CNN-1D avec 6 paramètres de 1h à 12 heures.

Modèle CNN – 1D							
Heures	mse1	nmse1	t2	RMSE	rRMSE	MABE	r(%)
1h	0.0117	0.3825	924.3254	68.6653	0.1424	45.2030	0.9744
2h	0.0176	0.5721	724.0762	97.7612	0.2028	67.7759	0.9502
3h	0.0207	0.6737	727.7382	117.0341	0.2427	79.2535	0.9278
4h	0.0234	0.7610	824.0405	119.2929	0.2474	84.8843	0.9279
5h	0.0236	0.7692	730.0495	105.8437	0.2195	75.3481	0.9416
6h	0.0204	0.6656	722.3677	99.6332	0.2066	67.4186	0.9462
7h	0.0183	0.5960	704.2616	100.4661	0.2084	68.7611	0.9456
8h	0.0199	0.6486	743.8005	103.2367	0.2141	71.6445	0.9423
9h	0.0237	0.7739	774.2537	112.7966	0.2339	81.3161	0.9364
10h	0.0228	0.7469	735.4549	116.5397	0.2416	82.7272	0.9303
11h	0.0238	0.7790	728.8889	114.2151	0.2368	83.2729	0.9357
12h	0.0223	0.7310	773.7272	110.3159	0.2287	79.6907	0.9367

D’après les résultats obtenus on peut voir que l’ajout du globale de l’heure précédente améliore la précision de prédiction du modèle CNN-1D en termes des indicateurs statistiques.

IV.6 Résultats des modèles d’hybridation de SVM et CNN

Dans cette partie nous présenterons les résultats de prédiction du modèle proposé (CNN-1D-SVM) qui consiste à utiliser les données sélectionnées comme entrée au CNN, ensuite les nouvelles données résultant du modèle CNN sont utilisées comme des entrées au modèle SVM pour la prédiction du K_t . Le tableau IV.9 montre les résultats de prédiction du modèle hybrides CNN-SVM en termes d’indicateurs statistiques.

Tableau IV.9 : Les résultats de combinaison de SVM-CNN de 1h à 12 heures avec 5 entrées.

Modèle CNN-SVM				
Heures	RMSE (Wh/m²)	rRMSE	MABE	r (%)
1h	80.4448	0.1673	54.6120	0.9650
2h	103.6925	0.2156	69.8253	0.9423
3h	112.4974	0.2339	74.7545	0.9334
4h	122.8874	0.2555	88.9367	0.9270
5h	114.6974	0.2384	79.9699	0.9306
6h	104.1537	0.2165	70.8152	0.9412

7h	102.3072	0.2126	68.4394	0.9426
8h	105.3156	0.2189	72.9602	0.9398
9h	109.4545	0.2275	76.9669	0.9373
10h	116.0717	0.2413	80.9742	0.9306
11h	117.3956	0.2440	83.7608	0.9317
12h	111.9910	0.2328	80.1619	0.9351

Dans la deuxième partie des essais, on ajoute les composantes globales horaires précédentes ou ensemble des données d'entrées pour tester l'influence de cette dernière sur la prédiction du K_t . Le tableau IV.10 présente les résultats obtenus par la combinaison du CNN et SVM avec 6 entrées.

Tableau IV.10 : les résultats de combinaison de SVM-CNN de 1h à 12 heures.

Modèle CNN-SVM				
Heures	RMSE (Wh/m²)	rRMSE	MABE	r (%)
1h	73.8974	0.1537	48.9435	0.9705
2h	97.0781	0.2019	67.4652	0.9512
3h	117.3727	0.2440	79.2288	0.9286
4h	121.4773	0.2526	84.7978	0.9264
5h	108.4388	0.2254	74.8596	0.9376
6h	99.6111	0.2071	66.3558	0.9457
7h	99.9715	0.2078	67.0030	0.9451
8h	103.5433	0.2152	72.4359	0.9421
9h	110.1845	0.2290	77.1114	0.9359
10h	112.5070	0.2339	77.3742	0.9339
11h	112.8414	0.2346	80.0958	0.9347
12h	112.9312	0.2347	79.5522	0.9329

Dans cette partie d'expérience nous essayerons d'ajouter d'autres combinaisons des filtres et garder toujours la même taille des filtres afin d'améliorer la performance de l'hybridation de SVM et CNN.

Tableau IV.11 : Les résultats de combinaison de SVM et CNN avec différents filtres.

Modèle CNN - SVM					
Couche + Taille + Nbr.Filtres		RMSE	rRMSE	MABE	r(%)
1^{er} Couche	2eme Couche				
6 filtres, taille 3	1 filtre, taille 1	69.8536	0.1453	46.3850	0.9736

6 filtres, taille 3	2 filtres, taille 1	74.3597	0.1546	48.9082	0.9702
6 filtres, taille 3	3 filtres, taille 1	70.8746	0.1474	46.2116	0.9729
6 filtres, taille 3	4 filtres, taille 1	72.4586	0.1507	47.6872	0.9717
6 filtres, taille 3	5 filtres, taille 1	75.4831	0.1570	49.6098	0.9693
6 filtres, taille 3	6 filtres, taille 1	73.1625	0.1521	49.0995	0.9713
1 filtre, taille 3	6 filtres, taille 1	81.8336	0.1702	54.8768	0.9634
2 filtres, taille 3	6 filtres, taille 1	72.9992	0.1518	48.9266	0.9714
3 filtres, taille 3	6 filtres, taille 1	78.2365	0.1627	52.5793	0.9672
4 filtres, taille 3	6 filtres, taille 1	76.0075	0.1581	46.6517	0.9686
5 filtres, taille 3	6 filtres, taille 1	73.7326	0.1533	47.7641	0.9706
7 filtres, taille 3	1 filtre, taille 1	70.4676	0.1465	47.3611	0.9733
7 filtres, taille 3	2 filtres, taille 1	70.5573	0.1467	46.4027	0.9730
7 filtres, taille 3	3 filtres, taille 1	77.4868	0.1611	45.9573	0.9673
7 filtres, taille 3	4 filtres, taille 1	74.0797	0.1540	46.3412	0.9703

D'après les résultats obtenus on remarque que la performance du modèle CNN-SVM est liée directement au choix optimal des filtres du modèle CNN. Dans notre cas on voit bien que la combinaison proposée atteint sa meilleure performance avec l'utilisation de 6 filtres de taille 3 dans la première couche et 1 filtre de taille 1 dans la deuxième couche.

Par la suite, nous exécutons le modèle proposé pour la prédiction de 12 heures qui viennent. Les résultats obtenus sont montrés dans le tableau.

Tableau IV.12 : Résultats de CNN-SVM avec des nouveaux nombres de filtres.

Modèle CNN-SVM				
Heures	RMSE (Wh/m²)	rRMSE	MABE	r (%)
1h	69.8536	0.1453	46.3850	0.9736
2h	94.8758	0.1973	65.4322	0.9524
3h	110.2898	0.2293	76.4176	0.9395
4h	125.6860	0.2613	84.5048	0.9188
5h	103.9196	0.2160	74.8935	0.9458
6h	100.4269	0.2087	70.0070	0.9459
7h	104.7104	0.2176	69.1670	0.9397
8h	104.0523	0.2163	71.8613	0.9412
9h	109.6184	0.2278	77.4358	0.9376
10h	114.8982	0.2388	85.9999	0.9407
11h	113.6158	0.2362	81.6441	0.9344

12h	111.0483	0.2308	80.8017	0.9371
-----	----------	--------	---------	--------

IV.7 Comparaison des modèles

Dans cette partie, nous comparons graphiquement les performances de prédiction des trois modèles en termes de coefficient de corrélation (r) et l'erreur de prédiction (RMSE), voire figure 4.1 et 4.2 respectivement. Pour cela, nous avons pris les meilleurs résultats de chaque modèle et les comparons ensemble pour 12 heures qui viennent. Comme on peut le voir sur les deux figures 4.1 et 4.2, nous pouvons observer que les trois modèles ont révélé des résultats significatifs. Cependant, la combinaison des deux modèles (CNN-SVM) proposé présente une légère amélioration par rapport aux modèles uniques dans la plupart des cas. Une autre remarque importante est que le modèle SVM donne les performances les plus faibles en termes de coefficient de corrélation pour toute l'horizon de prédiction qui confirme la performance des modèles de l'apprentissage profond par rapport aux modèles conventionnel.

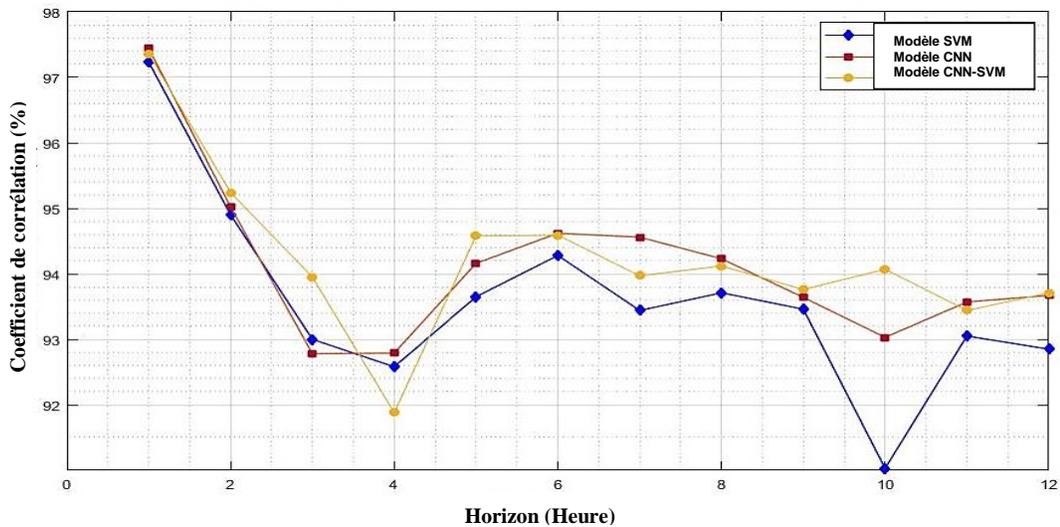


Figure IV.2 : Comparaison des performances des trois modèles de 1h à 12 heures en fonction de coefficient de corrélation.

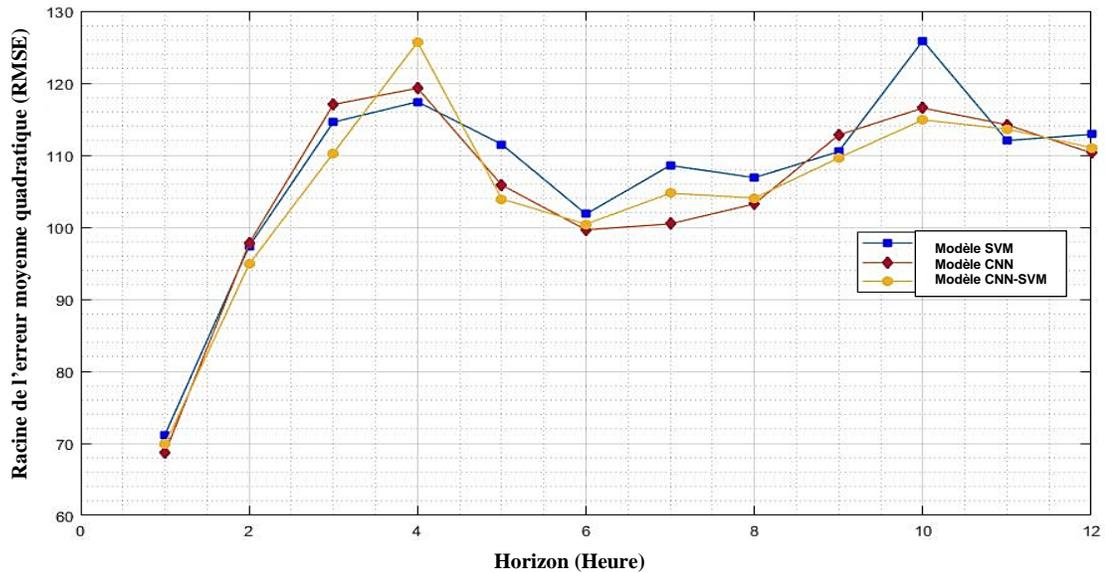


Figure IV.3 : Comparaison des performances des trois modèles de 1h à 12 heures en fonction de RMSE.

Donc, sur la lumière de cette comparaison statistique, nous pouvons constater que le modèle hybride CNN-SVM a présenté une légère amélioration à la prédiction, si en le comparant avec chacune des deux approches.

IV.8 Conclusion

Dans ce chapitre, une nouvelle approche hybride CNN-SVM pour la prédiction du rayonnement horaire multi-horizon sur plan horizontal, a été proposée. L'approche proposée bénéficie de l'avantage des deux modèles : précision du modèle SVM et nouvelles données pertinentes produites par le modèle CNN. Les tests effectués ont montré l'effet appréciable des paramètres d'entrée sur la précision des trois modèles. Il est à noter que le modèle CNN a prouvé sa précision par rapport au modèle SVM, qui confirme toujours la performance des modèles de Deep Learning par rapport aux modèles conventionnel de l'intelligence artificiel.

Le test statistique des modèles a démontré que l'approche combinée donne une légère amélioration dans la plupart des cas de prédiction (heurs de prédiction). En effet, la comparaison

entre les valeurs observées et celles prédites par les différents modèles, a suggéré que le modèle hybride est plus performant.

Conclusion Générale

CONCLUSION GENERALE

L'énergie est considérée comme le facteur le plus important dans le développement économique des nations. La demande en énergie dans le monde est prévue d'augmenter rapidement au cours des prochaines années avec l'augmentation de la population humaine, l'urbanisation et la modernisation. Parmi les diverses sources d'énergies possibles est l'énergie solaire, une bonne connaissance des puissances instantanées du rayonnement solaire à un endroit géographique donné est d'une importance majeure pour le développement des dispositifs d'énergies solaires et pour l'évaluation de leurs performances, notamment si on considère que l'Algérie est un pays de soleil par excellence où le nombre des heures d'insolation s'élève à presque 3300 h/an. Cependant, pour beaucoup de pays en voie de développement ; les mesures du rayonnement solaire ne sont pas facilement disponibles. Il est donc, important d'élaborer des modèles d'estimation de ce paramètre sur la base des données astronomiques et météorologiques.

L'approche que nous avons développée dans ce mémoire est conçue pour l'estimation des composantes solaire atteignant le sol par l'utilisation des techniques de l'apprentissage approfondi : la méthode de régression SVM et CNN. Après une étude bibliographique sur le rayonnement solaire de point de vue géométrique et énergétique, les instruments de mesure de différentes composantes avec une étude sur le gisement solaire du site de Ghardaïa à l'URAER, les valeurs instantanées du rayonnement diffus horizontal D_h , global horizontal G_h , direct normal D_n , global incliné G_β et le rayonnement extraterrestre G_{on} avec l'utilisation de différentes grandeurs tels que la déclinaison du soleil δ , l'angle horaire ω , la hauteur du soleil γ , l'azimut solaire ψ et l'indice de clarté K_t , nous avons commencé par l'étude détaillée de notre problématique et à la programmation des deux modèles par Matlab.

La méthode utilisée pour mettre en œuvre notre application est de développer un nouveau modèle basé sur l'apprentissage profond appelé CNN-1D pour la prédiction du rayonnement global horaire par la combinaison et d'hybridation entre les deux approches CNN et SVM dont le but d'améliorer la précision du prédicteur développé.

L'approche proposée pour la prédiction du rayonnement horaire multi-horizon montre que le modèle CNN est mieux précise que le modèle SVM et cette combinaison donne une légère amélioration de nos résultats et une performance mieux que CNN.

BIBLIOGRAPHIE

BIBLIOGRAPHIE

- [1] Mouhous-chaouchi S., Etude statistique du rayonnement solaire sur un plan incliné. Mémoire de Magister, Université Abou Bekr Belkaid, Tlemcen 2012.
- [2] Kherous N., Hamzaoui J., Etude et réalisation d'un concentrateur solaire parabolique. Mémoire de Master, Université Mouloud Mammeri, Tizi-Ouzou 2015.
- [3] Bessemoulin P., Oliviéri J., Le rayonnement solaire est sa composante ultraviolette. Physique atmosphérique, La Météorologie 8° série – n°31 – Septembre France 2000.
- [4] Aouali K., Influence des facteurs météorologiques sur le rayonnement solaire à Bejaïa. Mémoire de Magister, Université de Abderrahmane Mira, Bejaïa 2011.
- [5] Baroud H., Djekaoua H., Estimation de l'ensoleillement par deux modèles semi-empiriques dans la région de Ghardaïa. Mémoire de Master, Université de Ghardaïa, Ghardaïa 2018.
- [6] Yettou F., Conception et réalisation d'un système de cuisson solaire destiné au site saharien (Ghardaia, Algérie). Thèse de Doctorat, Université de Hadj Lakhdar, Batna 2015.
- [7] Ricaud A., Gisement solaire et transferts énergétiques. Mémoire de Master, Université de CERGY-PONTOISE, France 2011.
- [8] Zeghib I., Etude et réalisation d'un concentrateur solaire parabolique. Mémoire de Magister, Université Mentouri, Constantine 2005.
- [9] Meziani F., Détermination du gisement solaire par traitement d'images MSG. Mémoire de Magister, Université Mouloud Mammeri, Tizi-Ouzou.
- [10] Attaf Y., Détermination du rayonnement solaire globale par imagerie météosat. Thèse de Magister en Automatique Industrielle, Université Mouloud Mammeri, Tizi-Ouzou 2003.
- [11] Badesco V., Modeling solar radiation at the earth surface, recent advances. Edition Springer, 2008.
- [12] Iqbal M., An introduction to solar radiation. Academic Press Inc, 1983.
- [13] Duffie JP., Beckman WA., Solar engineering of thermal process. 2nd ed, New York: John-Wiley, 1991.

- [14] Nia M., Etude comparative des méthodes d'estimation du rayonnement solaire. Mémoire de Magister, Université Ferhat Abbas, Sétif 2010.
- [15] www.rts.ch
- [16] Gairaa K., Prédiction du rayonnement solaire en Algérie à partir des modèles hybrides. Thèse de Doctorat, Université Ibn Khaldoun, Tiaret 2018.
- [17] Gairaa K., Khellaf A., Chellali F., Benkaciali S., Bakelli Y., Bezari S., Maximisation and optimization of the total solar radiation reaching solar collector surfaces, *Progress in Clean Energy: Novel systems and Applications 2*, pp.873-886,2015.
- [18] Liu, B. Y. H., and R. C. Jordan. The interrelationship of Direct, Diffuse and Characteristic Distribution Total Solar Radiation. *Solar Energy* 4.3 (1960): 1-19.
- [19] Kasten F., The Linke turbidity factor based on improved values of the integral Rayleigh optical thickness. *Solar energy* 56.3 (1996): 239-244.
- [20] Gairaa K., Benkaciali S., Gurmoui M., Clear-sky models evaluation of two sites over Algeria for PV forecasting purpose. *Eur. Phys. J. Plus* 134, 534 (2019).
- [21] Bouchouicha K., Modélisation multispectrale des images satellitaire-Application : quantification du bilan d'énergie Sol-atmosphère. Thèse de Doctorat, Université des sciences et de la technologie Mohamed Boudiaf, Oran 2017.
- [22] Benatallah D., Détermination du gisement solaire par imagerie satellitaire avec intégration dans un système d'information géographique pour le sud d'Algérie. Thèse de Doctorat, Université Africaine Ahmed Draia, Adrar 2019.
- [23] ISO-9060 Standard & Pyranometer Measurement Accuracy (2014). Consulté en Avril 2014. https://ases.conference-services.net/resources/252/2859/pres/SOLAR2012_0829_presentation.pdf
- [24] Geuder, N., F. Trieb, C. Schillings, R. Meyer, V. Quaschnig, (2003). Comparison of different methods for measuring solar irradiation data. In proceedings of the 3rd International Conference on Experiences with Automatic Weather Stations, 19th-21st of February 2003, Torremolinos, Spain. 9 pp.
- [25] www.futura-sciences.com
- [26] McArthur L. (1998): Baseline Surface Radiation Network (BSRN) Operations Manual, (Version 1.0), WMO/TD No. 879. Geneva, Switzerland : World Climate Research Program, WMO, 1998.
- [27] Portail Ministère de l'Énergie, www.energy.gov.dz, Consulté en Février 2018.

- [28] Mefti A., Bouroubi M.Y., Mimouni H., (2002), Evaluation du potentiel énergétique solaire, Bulletin des Energies Renouvelables, N° 2, pp.12, Décembre 2002.
- [29] Mefti A., Bouroubi M.Y., Khellaf A., (1999), « Analyse Critique du Modèle de l'Atlas de l'Algérie », Revue des Energies Renouvelables, 2(2), 69 - 85.
- [30] Yakhelef Y., Mehdaoui A., Contribution à la détermination expérimentale des performances thermiques d'un cuiseur solaire parabolique. Mémoire de Master, Université El-Hadj Lakhdar, Batna 2019.
- [31] Merdaoui Z., Caractérisation radiométrique des sites de Bouzaréah et de Ghardaïa. Mémoire de Magister, département d'électronique, Université de Saad Dahleb, Blida 2007.
- [32] Chikh M.E., Benramdane A.Y., Analyse des profils thermiques de deux cuiseurs solaires expérimentés sous climat Aride du sud Algérien (Etude expérimental-Centre de recherche Ghardaïa). Université Aboubaker Belkaid, Tlemcen 2019.
- [33] Mahdjane K., Détection d'anomalies sur des données biologiques par SVM. Mémoire de Magister, Université Mouloud Mammeri, Tizi-Ouzou 2012.
- [34] Tezmol A., Sari-Sarraf H., Mitra S., Long R. Customized Hough Transform for Robust Segmentation of the Cervical Vertebrae from X-Ray Images". SSIAl2002, Proc. IEEE 1537, pp.224-228, 2002.
- [35] Mohamed Salem L., Prédiction de la fréquence de Résonance d'une antenne microbande avec une ouverture dans le plan de masse. Mémoire de Master, Université de Mohamed Boudiaf, M'sila 2016.
- [36] Benyahia Belaidi S., Application de la classe des méthodes d'apprentissage statistique SVM (support vector machine) pour la reconnaissance des formes dans les images. Mémoire de Magister, Université Abou Bakr Bel-Kaid, Tlemcen 2012.
- [37] Louati N., Diagnostic de défauts par les Machines à Vecteurs Supports : application à différents systèmes multivariables non linéaires. Thèse de Doctorat, Université Claude Bernard Lyon 1, Lyon 2012.
- [38] Mohamadally H., Fomani B., SVM : Machines à Vecteurs de Support ou Séparateurs à Vastes Marges. France 2006.
- [39] Zidelmal Z., Amirou A., Djeddi M., Djouaher N., Application des SVMs basés sur l'algorithme SMO pour la détection d'anomalies cardiaques. 4th International Conference: Sciences of Electronic, Technologies of Information and Telecommunications. IEEE. Tunisia. March 25-29, 2007.

- [40] Abibullaev B., Kang W., S. Hyun Lee S., Jinung, An Classification of Cardiac Arrhythmias using Biorthogonal Wavelets and Support Vector Machines. International Journal of Advancements in Computing Technology Volume 2, Number 2, June, 2010.
- [41] Sudhir D, Ashok. A, Amol. P: Neural Network Aided Breast Cancer Detection and Diagnosis Using Support Vector Machine. Proceedings of the 7th WSEAS International Conference on Neural Networks, Cavtat, Croatia, (pp158-163), June 12-14, 2006.
- [42] Majumder S.K, Ghosh N, Gupta P.K., Support vector machine for optical diagnosis of cancer. Journal of Biomedical Optics 10(2), 024034, March/April ,2005.
- [43] Bertozzi M., Broggi A., Del Rose M., Felisa M., Rakotomamonjy A., Suard F., A Pedestrian Detector Using Histograms of Oriented Gradients and a Support Vector Machine Classifier. Proceedings of the 2007 IEEE Intelligent Transportation Systems Conference Seattle, WA, USA, Sept. 30 - Oct. 3, 2007.
- [44] Ramona M., Classification automatique de flux radiophoniques par Machines à Vecteurs de Support. Thèse de Docteur, l'Ecole Télécom ParisTech, 2010.
- [45] Ouakkaf Belgasmi H., Reconnaissance automatique des expressions faciales par support vector machine. Mémoire de Master, Université Larbi Ben M'hidi Oum El Bouaghi 2017.
- [46] P.V.M.J. Jones, Robust Real Time Face Detection, international journal of computer vision kluwer academie publishers, 2004.
- [47] www.ujjwalkarn.me
- [48] www.cs231n.github.io
- [49] www.openclassrooms.com
- [50] Malki N., Classification automatique des textes par Les réseaux de neurones à convolution. Université Larbi Ben M'Hidi, Oum El-Bouaghi 2019.
- [51] www.medium.com
- [52] Oulmi M., Kaloun S., Classification d'objet avec le Deep learning. Mémoire de Master, Université Akli Mohand Oulhadj, Bouira 2018.
- [53] www.towardsdatascience.com
- [54] www.coursera.org
- [55] www.missinglink.ai

- [56] Rabehi A., Guermoui M., Khelifi R., Mekhalfi ML., (2020). Decomposing global solar radiation into its diffuse and direct normal radiation. *International Journal of Ambient Energy*, 41(7), 738-743.
- [57] Khelifi R., Guermoui M., Rabehi A., Lalmi L., Multi-step-ahead forecasting of daily solar radiation components in the Saharan climate. *International Journal of Ambient Energy*, 2018. <https://www.tandfonline.com/doi/full/10.1080/01430750.2018.1490349>
- [58] Abdul Majeed Muzathik, Wan Mohd Norsani Wan Nik, Khalid Samo, Mohd. Zamri Ibrahim, 'Hourly Global Solar Radiation Estimates on a Horizontal Plane', *Journal of Physical Science*, Vol. 21, N°2, pp. 51 - 66, 2010.
- [59] Lantri F., Bachari NI., Belbachir H., Estimation et cartographie des différentes composantes de rayonnement solaire au sol à partir des données météorologiques. *Revue des Energies Renouvelables* Vol. 20 N°1. 111 – 130 (2017).
- [60] Guermoui M., Rabehi A., Soft Computing for Solar Radiation Potential Assessment in Algeria. *International Journal of Ambient Energy*, 2018. <https://www.tandfonline.com/doi/full/10.1080/01430750.2018.1517686>

ANNEXES

Annexe A

Mesure du rayonnement - Norme ISO 9060

Tableau. A.1: Spécifications iso 9060 pyranomètre [21].

Specifications ISO 9060	Pyranomètre Secondary	Standard First Class	Second Class
Temps de réponse :	< 15s	< 30s	< 60s
Zero-offset: Offset-A:	+ 7 W/m²	+ 7 W/m²	+ 7 W/m²
Offset-B:	± 2 W/m²	± 2 W/m²	± 2 W/m²
Non-stabilité :	± 0.8%	± 1.5%	± 3%
Non-linéarité : ± 0.5% ± 1% ± 3%			
Réponse directionnelle (pour le direct) : Directe ≈ 1000 W/m²	± 10 W/m²	± 20 W/m²	± 20 W/m²
Sélectivité spectrale :	± 3%	± 5%	± 10%
Réponse en température : (de -10 à +40°C)	2%	4%	8%
Réponse en inclinaison : Directe ≈ 1000 W/m²	± 0.5%	± 2%	± 5%

Annexe B

Les entrées de programme MATLAB sont comme suit :

Température de l'air (°C)

T_{moy} : Température moyenne.

T_{min} : Température minimum.

T_{max} : Température maximum.

Pression atmosphérique (kPa)

Pr_{moy} : Pression moyenne.

Pr_{min} : Pression minimum.

Pr_{max} : Pression maximum.

Humidité relative (%)

HR_{moy} : Humidité relative moyenne.

HR_{min} : Humidité relative minimum.

HR_{max} : Humidité relative maximum.

H_{max} : Hauteur maximum du soleil.

GHI : Horaire précédent GHI.

K_t : L'indice de clarté.

Dur_{inso} : Durée d'insolation.

Ev_{max} : Elévation maximale.

Dec : Déclinaison.

SS : Durée des jours.

ABSTRACT

Accurate estimation of solar radiation components of a specific location has been one of the most important issues of solar energy applications. Due to the non-stationary behaviour of solar radiation parameters and variety of atmosphere conditions, stand-alone forecasting models are insufficient for providing accurate estimation in some cases.

In this respect, a new approach named CNN-SVM was proposed for multi-hourly global solar radiation forecasting. The proposed model is based mainly on the use of convolution neural network model (CNN-1D) as feature selection of meteorological input and support vector machine (SVM) as regression model.

The achieved results show the combined model CNN-SVM provides slight improvement against stand-alone model in terms of statistical indicators.

Key words: Solar radiation, Estimation of solar radiation components, Convolution neural network model (CNN-1D), Support vector machine (SVM), The combined model CNN-SVM.

ملخص

يعد التقدير الدقيق لمكونات الإشعاع الشمسي لموقع معين أحد أهم مشاكل استخدامات تطبيقات الطاقة الشمسية، ونظرا لعدم ثبات معاملات الإشعاع الشمسي وتنوع ظروف الملائمة للغلاف الجوي، فإن نماذج التنبؤ المستقلة غير كافية لتوفير تقدير دقيق في بعض الحالات. في هذا الصدد تم اقتراح مقارنة جديدة تحت اسم CNN-SVM للتنبؤ بالإشعاع الشمسي الكلي لعدة ساعات، حيث تعتمد هذه المقارنة بشكل أساسي على استخدام نموذج الشبكة العصبية الالتفافية (CNN-1D) كنموذج لاختبار مدخلات الطقس وجهاز متجه الدعم (SVM) كنموذج انحدار. تظهر النتائج المحققة أن النموذج المشترك CNN-SVM يعطي تحسنا طفيفا عن النموذج المستقل من حيث المؤشرات الإحصائية.

الكلمات المفتاحية: الإشعاع الشمسي، تقدير مكونات الإشعاع الشمسي، نموذج الشبكة العصبية الالتفافية (CNN-1D)، جهاز متجه الدعم (SVM)، النموذج المشترك CNN-SVM.