People's Democratic Republic of Algeria
Ministry of Higher Education and Scientific Research
Ahmed Draïa Adrar University
Faculty of Material Sciences, Mathematics and Computer Science,
Material Sciences Department

Thesis

Presented with a view to obtaining the 3rd cycle Doctorate degree in

Physics

Option: Materials physics

By: DJELDJLI Halima

Study and forecast of solar radiation using hybrid models

Discuss publicly: 09 / 10 / 2024 In front of the jury :

President	BENATIALLAH Ali	Pr.	Univ Adrar
Examiner	KADDI Mohammed	MCA	Univ Adrar
Examiner	ARBAOUI Iliace	MCA	Univ Adrar
Examiner	MAOUEDJ Rachid	DR	URER MS. Adrar
Supervisor	BENATIALLAH Djelloul	MCA	Univ Adrar
Co-supervisor	TANOUGAST Camel	Pr.	Univ France

College year: 2023/2024

Acknowledgment

I am profoundly grateful to **Almighty Allah** for granting me the strength and opportunity to complete my Ph.D. Without God's guidance and blessings, this accomplishment would not have been feasible. I am deeply humbled and thankful for His providence throughout this journey.

I extend my sincerest appreciation to **Dr. BENATIALLAH Djelloul** for his invaluable guidance, support, and expertise as my thesis supervisor. His dedication and guidance have been crucial in forming this study.

Special thanks are also due to **Pr. TANOUGAST Camel** for his assistance and facilitation in completing this study.

I am grateful to the esteemed members of the committee, including **Pr. BENATIALLAH Ali, Dr. KADDI Mohammed, Dr. ARBAOUI Iliace**

And Dr. MAOUEDJ Rachid, for their willingness to evaluate and defend my Ph.D. thesis.

Your contributions and support have been invaluable to my research endeavor.

Lastly, I extend my thanks to father and uncle and my colleagues at the university, as well as the staff of laboratories LEESI and LDDI, for their encouragement, assistance, and the wealth of ideas and information they provided. Your steadfast encouragement and positivity have served as a continual source of motivation.

Djeldjli Halima

Dedication

To my cherished family,

With deep appreciation and love, this dissertation is dedicated to you. My journey has been shaped by your unfailing support, encouragement, and sacrifices. To my siblings, whose presence has been a source of joy and strength, and to my parents, whose unending love and wisdom have molded me into the person I am today.

To my father and uncle: I have always been guided by your knowledge and appreciate your unwavering belief in me. I want to thank my whole family for thier unending support.

This work is a reflection of all of your love, encouragement, and belief in me. I will always be appreciative of you being in my life.

With my deepest gratitude and affection.

Djeldjli Halima

الملـــخـــص

يلعب التنبؤ بالإشعاع الشمسي دورًا محوريًا في تصميم أنظمة الطاقة الشمسية وتحسينها، خاصة في المناطق التي تتميز بالمناخ الجاف مثل جنوب الجزائر. تبحث هذه الأطروحة في أداء نماذج التنبؤ المختلفة في تقدير الإشعاع الشمسي العالمي اليومي (G) عبر ست مدن في المنطقة: أدرار، ورقلة، بشار، الوادي، تيميمون، وتمنراست. تم تقييم أساليب متميزة - الشبكات العصبية الاصطناعية (ANN)، وشبكة (GA-ANN) القائمة على دمج الخوارزمية التحسين الجينية، وشبكة (FFA-ANN) القائمة على دمج خوارزمية التحسين اليراع - للتأكد من دقتها وموثوقيتها. تكشف الدراسة أن نماذج التنبؤ الهجينة، التي قمنا بتطويرها FFA-ANN وGA-ANN، تتفوق على نماذج ANN المستقلة من حيث دقة التنبؤ عبر جميع المواقع. في حين يقوم النموذج GA-ANN على تحسين اوزان الشبكة في التنبؤ بـ GSR ، يقوم النموذجFFA-ANN باختيار احسن مجموعة مدخلات مناسبة لكل موقع لضمان احسن أداء للنموذج، حيث حقق النموذجين قيمًا جيدة جدا في المؤشرات الإحصائية مثل معامل التحديد (R=0.9076 لـGA-ANN ،و R=0.9321 لـ FFA-ANN)، متوسط النسبة المئوية للخطأ المطلق (MAPE)، والجذر النسبي يعني خطأ مربع (rRMSE). والجدير بالذكر أن FFA-ANN يُظهر أداءً قويًا في تقدير GSR اليومي، مما يُظهر قدرته على التكيف مع الظروف المناخية المختلفة.علاوة على ذلك، يسلط البحث الضوء على أهمية معلمات المدخلات مثل الإشعاع الشمسي خارج الأرض (H_0)، والانحراف، ومتوسط درجة الحرارة (T_{avg})، والرطوبة النسبية (RH) في تعزيز القدرات التنبؤية لنموذج ANN في المناطق الجافة. تؤكد النتائج على إمكانات نماذج التنبؤ الهجينة في تحسين التنبؤ بالإشعاع الشمسي وتسهيل اتخاذ القرارات في تركيبات نظام الطاقة الشمسية وحجمها. وبشكل عام، تساهم هذه الأطروحة برؤى قيمة في تطوير وتطبيق نماذج التنبؤ الهجينة للتنبؤ بالإشعاع الشمسي في المناخات القاحلة. ومن خلال الاستفادة من تقنيات التحسين المتقدمة والتحليل الشامل للبيانات، تضع الدراسة الأساس لتعزيز كفاءة وفعالية أنظمة الطاقة الشمسية في جنوب الجزائر والمناطق المماثلة في جميع أنحاء العالم.

الكلمات المفتاحية: نموذج هجين، شبكة عصبية اصطناعية، خوارزمية اليراعة، إشعاع شمسي، ، خوارزمية جينيةGA .

Abstract

Solar irradiation prediction is essential to the planning and development of solar energy systems. especially in regions characterized by a dry climate such as southern Algeria. This thesis investigates the performance of different forecast models in predicting daily global solar radiation (GSR) across six cities in the region: Adrar, Ouargla, Béchar, El Oued, Timimoun, and Tamanrasset. Artificial neural networks (ANN), GA-ANN, and FFA-ANN—are evaluated for their accuracy and reliability. The study reveals that the hybrid prediction models we developed, FFA-ANN and GA-ANN, outperform standalone ANN models in terms of prediction accuracy across all sites. While the GA-ANN model optimizes the network weights when predicting GSR, the FFA-ANN model chooses the best combination set of inputs for each site to ensure the best performance of the model. The two models achieved very good statistical indicators results , the coefficient of determination (R=0.9076 for GA-ANN, and R=0.9321 for FFA-ANN), and the average absolute percentage error (MAPE), and relative root mean square error (rRMSE). Furthermore, the research highlights the significance of input parameters such as extraterrestrial solar irradiation (H_0), declination, average temperature (T_{avg}), and relative humidity (RH) in enhancing the predictive capabilities of the ANN model. The findings underscore the potential of hybrid forecasting models in optimizing solar radiation forecasting and facilitating informed decision-making in solar energy system installations and sizing. Overall, this thesis contributes valuable insights into the development and application of hybrid forecasting models for solar radiation forecasting in arid climates. By leveraging advanced optimization techniques and comprehensive data analysis, the study lays the foundation for enhancing the efficiency and effectiveness of solar energy systems in southern Algeria and similar regions worldwide.

Keywords: Hybrid model, Artificial neural network, Firefly algorithm, Solar radiation, Genetic algorithm.

Résumé

La prévision du rayonnement solaire joue un rôle central dans la conception et l'optimisation des systèmes d'énergie solaire, notamment dans les régions caractérisées par un climat sec comme le sud de l'Algérie. Cette thèse étudie les performances de différents modèles dans la prévision du rayonnement solaire global (GSR) quotidien dans six villes: Adrar, Ouargla, Béchar, El Oued, Timimoun et Tamanrasset. Les réseaux de neurones artificiels (ANN), GA-ANN et FFA-ANN – sont évaluées pour leur précision et fiabilité. L'étude révèle que les modèles de prédiction hybrides que nous avons développés, FFA-ANN et GA-ANN, surpassent les modèles ANN autonomes en termes de précision de prédiction sur tous les sites. Alors que le modèle GA-ANN optimise les pondérations du réseau lors de la prévision du GSR, le modèle FFA-ANN choisit la meilleure combinaison d'entrées pour chaque site afin de garantir les meilleures performances du modèle. Les deux modèles ont obtenu de très bons résultats en matière d'indicateurs statistiques, le coefficient de détermination (R=0.9076 pour GA-ANN, et R=0.9321 pour FFA-ANN), l'erreur absolue moyenne en pourcentage (MAPE) et l'erreur quadratique moyenne relative (rRMSE). En outre, la recherche met en évidence l'importance des paramètres d'entrée tels que l'irradiation solaire extraterrestre (H_0), la déclinaison, la température moyenne ($_{Tavg}$) et l'humidité relative (R_H) pour améliorer les capacités prédictives du modèle ANN. Les résultats soulignent le potentiel des modèles hybrides pour optimiser la prévision du rayonnement solaire et faciliter une prise de décision éclairée concernant les installations et le dimensionnement des systèmes d'énergie solaire. Dans l'ensemble, cette thèse apporte des informations précieuses sur le développement et l'application de modèles de prévision hybrides pour la prévision du rayonnement solaire dans les climats arides. En tirant parti de techniques d'optimisation avancées et d'une analyse complète des données, l'étude jette les bases de l'amélioration de l'efficience et de l'efficacité des systèmes d'énergie solaire dans le sud de l'Algérie et dans des régions similaires du monde entier.

Mots clés : Modèle hybride, Réseau de neurones artificiels, Algorithme Firefly, Rayonnement solaire, Algorithme génétique.

Table of Contents

Abstract iii

List of Figures vi

List of Tables ix

List of Abbreviations xi

List of Symbols xii

General Introduction 01

Chapter I: Solar Notions

I .1. Introduction 04

I .2. The issue of sustainable development and climate change 04

I .3. Global and national awareness of combating climate change 05

I .4. Renewable energies on a national scale 07

I .4.1. Solar Energy 10

I .5. Source Of Solar Irradiation 12

I .6. Solar Radiation and the Atmosphere 13

I .6.1. Earth's atmosphere composition and structure 13

I .6.2. Attenuation of solar radiation 16

I .7. Radiation on Ground 18

I .8. Flux of Irradiation 20

I .9. Solar Constant 20

I .10. Extraterrestrial radiation 22

I .11. Astronomy of Sun-Earth 22

I .11.1. Earth movement around the Sun 22

I .11.2. Solar Declination 23

I .11.3. Solar and Local Standard Time 24

I .11.4. Hour Angle 25

I .11.5. The Sun's Position 25

I .12. Clearness Index 27

I .13. Solar radiance (Tilted Plane) 28

Table of Contents

I .14. Conclusion ……. 30

I .15. References …….. 31

Chapter II: Data Sources And Processing

II.1 Introduction ……..... 33

II.2. Solar Radiation Measurements ……... 33

II.2.1 Direct Solar Radiation ……... 33

II.2.2 Diffuse Radiation …….. 34

II.2.3 Global Radiation ……...…… 35

II.2.4 Ground Albedo ……. 35

II.2.5 Sunshine Duration ……. 36

II.2.6 Cloud Cover …….. 37

II.3. Algerian Meteorological Network …….. 37

II.4. Remote Sensing Measurements …….. 39

II.4.1 Digital Image in Remote Sensing and its Characteristics ……. 41

II.4.2 Platforms and Sensors ……... 41

II.4.3 Characteristics of Satellites – Orbit …….. 42

II.4.4 Digital Image Processing …….. 43

II.5. EUMETSAT …… 44

II.6. History of METEOSAT Satellites …….. 44

II.7. Second Generation METEOSAT (MSG) …… 45

II.8. Data Preprocessing …….. 49

II.8.1. Data Preparation ……... 49

II.8.1.1. Data Cleaning ……. 50

II.8.1.2. Data Transformation …….. 51

II.8.1.3. Data Integration …….. 52

II.8.1.4. Data Normalization …… 52

II.8.1.5. Missing Data Imputation …… 52

II.8.1.6. Noise Identification …… 52

II.8.2. Data Reduction ……. 52

Table of Contents

II.8.2.1. Feature Selection …… 55

II.8.2.2. Instance Selection …….. 55

II.8.2.3. Discretization ……. 55

II.8.2.4. Feature Extraction/Instance Generation ……. 56

II.9. Conclusion …… 57

II.10. References …….. 58

Chapiter III : Prediction Models

III.1. Introduction ……... 60

III.2. Prediction Models …… 61

III.2.1. Persistence Model …… 61

III.2.2. Empirical Models ……. 61

III.2.2.1. Insolation-Based Models …… 61

III.2.2.2. Temperature-Based Models ……... 63

III.2.3. Time Series Models …….. 64

III.2.3.1. ARMA Model …… 64

III.2.3.2. ARIMA Model ……... 65

III.2.4. Machine learning …….. 65

III.2.4.1. Support Victor machine (SVM) …… 67

III.2.4.2. Artificial Neural Networks (ANNs) ……. 71

III.2.4.2.1. The Formal Neuron ……... 72

III.2.4.2.2. Types of Neural Networks …… 73

III.2.4.3. Gradient Boosting …… 77

III.2.4.4. Random Forest ……. 78

III.2.4.5. Linear Regression model …….. 79

III.3. Hybrid models ……. 81

III.4. Performance and Evaluation of Predictors …… 82

III.4.1. (R^2) ……. 82

III.4.2. (RMSE) …… 83

III.4.3. (MBE) …… 83

Table of Contents

III.4.4. (MAPE) …… 83
III.5. Conclusion ……. 84
III.6. References …….. 85

Chapter IV : Solar Radiation Forecasting Based on A New FFA-ANN Model

IV.1. Introduction …… 89
IV.2. Literature review of the models used …… 89
IV.3. Models and materials …… 91
IV.3.1. Areas of research and datasets …….. 91
IV.3.2. Methods and models ……. 93
IV.3.2.1. Artificial Neural Network (ANN) ……... 93
IV.3.2.2. ANN architecture …… 94
IV.3.2.3. The Support Victor machine (SVM) model …… 95
IV.3.2.4. Firefly algorithm (FFA)…… 96
IV.3.2.5. The Hybrid FFA-ANN model ……. 98
IV.3.3. Evaluation metrics ……. 100
IV.4. Findings and conversation …….. 100
IV.5. Conclusions …….. 107
IV.6. References …… 108

Chapter V: Solar Radiation Estimation Based On A New GA-ANN model

V.1.Introduction ……... 112
V.2. Literature review of the models used …… 112
V.3. Material and Models …… 115
V.3. 1. Study Areas and Data Collection …… 115
V.3. 2. Data Preprocessing ……... 116
V.3..3. Artificial Neural Network (ANN) Model ……. 116
V.3..4 Adaptive Neuro Fuzzy Inference System (ANFIS) Model …… 118

Table of Contents

V.3. 5. Artificial Neural Network - Genetic Algorithm …… 119

V.4. Findings and discussions ……. 121

V.5. Conclusions …….. 132

V.6. Refrences …….. 133

General Conclusion and Perspective …….. 138

Chapter I: Solar Notions

Figure I.1: Distribution of the national RE program by technological sector …… 8
Figure I.2: Distribution of power generated from renewable sources (2012)…… 9
Figure I.3: Global irradiation on a horizontal plane…… 11
Figure I.4: Direct irradiation at normal incidence…… 11
Figure I.5: sun radiation's spectral distribution near the top of the atmosphere……. 13
Figure. I.6: Structure of the atmosphere. Where the strong line represents the variance in each atmospheric layer's average temperature…….…… ….. 15
Figure. I.7: Spectra of mean solar radiation incident……... 17
Figure. I.8: Main solar radiation components.…… 19
Figure I.9: Geometry of Sun-Earth……. 22
Figure.I-10: The earth movements……... 23
Figure I.11: The view of Earth's Axis from space……. 24
Figure I.12. The relationships between the terms of time…..…….. 25
Figure. I.13: azimuth (γ), solar altitude angle (αs), collector tilt (β), solar azimuth angle (γs)……. 26
Figure I.14. Position of the Sun…… 27

Chapter II: Data Sources And Processing

Figure II.1: (a) Diagram of a Pyrheliometer, (b) Pyrheliometer first-class type Hukseflux DR01…….. 34
Figure II.2 : Measurement of diffuse (a) pyranometer with shading ball mounted on a tracking system, (b) pyranometer with shading ring……...…….. 34
Figure II.3: Diagram of a pyranometer , and Eppley-type pyranometer (on the right)… 35
Figure II.4 : l‟Albédometer……. 35
Figure II.5: (a) Campbell-Stokes Heliograph; (b) Fiber Optic Heliograph……. 36
Figure II.6: Pyranometric Network (ONM and CDER)…….. 38
Figure II.7: Remote Sensing Process.…….. 29
Figure II.8: Digital Image…… 41
Figure II.9: Geostationary Satellites in Orbit……..…… 42
Figure II.10: Heliosynchronous Orbit Satellites.…….. 43

Figure II.11: METEOSAT Satellite Map…… 44
Figure II.12: Second Generation METEOSAT Satellite. ……. 45
Figure II.13: Sequence of three images captured by SEVIRI in the VIS 0.6 channel (C1) on June 19, 2004, at 06:00, 12:00, and 18:00 UTC ….…… 47
Figure II.14 Data preparation Forms …… 51
Figure II.15 : Data reduction forms…….. 54

Chapiter III : Prediction Models

Figure III.1. Machine learning techniques……... 66
Figure III. 2. Support vector maching …….. 70
Figure III-3: Correspondence between the biological neuron and the artificial neuron… 72
Figure III.4: Principle of the formal neuron……. 73
Figure III.5: Example of an MLP network with multiple layers ……. 74
Figure III.6: Example of a recurrent network …….. 75
Figure III.7: Typical Configuration of an RBF Network ……. 76
Figure III.8: The general structure of the gradient boosting algorithm…… 78
Figure III.9: The general structure of random forest model. …… 79
Figure III.10: simple linear regression structure ……. 80
Figure III.11: multiple linear regression structure…… 81

Chapter IV : Solar Radiation Forecasting Based On A New FFA-ANN Model

Figure IV.1. shows the locations of every site ……. 92
Figure IV.2. The model based on ANNs …….. 95
Figure IV.3. The FFA-ANN model ……. 99
Figure IV.4. The FFA-ANN and ANN's MAPE (%) and rRMSE (%). …… 103
Figure IV.5. Forcated and gathered values for the FFA-ANN approach.…….. 104
Figure IV.6. FFA-ANN models' scatter plot.…… 105

Chapter V: Solar Radiation Estimation Based On A New GA-ANN model

Figure V.1: Solar potential and sites locations ..……... 115
Figure V.2: Characteristics of our ANN …..…….. ……. 117
Figure V.3: Our ANFIS model.. …….. …….…. 119
Figure V.4: The steps of GA-ANN approach ….. …… 120
Figure V.5: Outcomes scatter plot…….……... 122
Figure V.6: rRMSE (%) and MAPE (%) of the models …….…... 123
Figure V.7: The overall result's values of error histogram …… 124
Figure V.8: The overall result's residuals vs. fitted Values ……... 125
Figure V.9: Error plot of models performance …… 126
Figure V.10: Residuals Quantile-Quantile of the approches Normality Assessment … 127
Figure V.11: Intercept tests and linear slope of the models ……..... 128

Chapter I: Solar Notions

Table I.1: The key events in the global awareness and response to climate change…… 06
Table I.2: The atmosphere's composition at the vicinity of the surface of the Earth ….. 16

Chapter II : Data Sources And Processing

Table II.1: Remote Sensing Process steps …….. 40
Table II.2. Characteristics and main objectives of the 12 image acquisition channels of SEVIRI …… 46

Chapter III : Prediction Models

Table III.1 : Constants of the Angstrom-Prescott Model for Various Locations …… 63
Table III.2 : Different kernel functions …… 70

Chapter IV : Solar Radiation Forecasting Based On A New FFA-ANN Model

Table IV.1. Literature review of the models used……... 90
Table IV.2. Coordinates of the location……... 92
Table IV.3. The Parameters used …… 93
Table IV.4. The characteristics of ANN…….. 94
Table IV.5. Parameters optimized for SVM ……... 96
Table IV.6. The parameters of the Firefly algorithm …….. 99
Table IV.7. Evaluation metrics ……... 100
Table IV.8. Statistical indices ……. 101
Table IV.9. Optimal combination of FFA-ANN …… 102
Table IV.10. Statistics assess errors across the different models …… 106

Chapter V: Solar Radiation Estimation Based On A New GA-ANN model

Table V.1: A Literature review of the models used …… 114
Table V.2: Training Details and Features of the Model …….. 117
Table V.3: The five layers of the ANFIS model …… 118
Table V.4: Parameters of ANN-GA method …….. 121
Table V.5: Model's overall performance …… ……. 121
Table V.6: Performance statistical indicators ……. 129
Table V.7: The performance of the models compared to other studies…….. 131

NOMENCLATURE

LIST OF ABREVIATIONS

CDER	Center Of Renewable Energy Development
CSP	Concentrated Solar Power
DM	Data Mining
DST	Daylight Saving Time
FIS	Fuzzy Inference Systems
LST	Local Solar Time
MV	Missing Values
ONM	National Meteorological Office
PV	Photovoltaic
RE	Renewable Energy
SODA	Simple Ocean Data Assimilation
TST	True Solar Time
VTA	Value Added Tax

NOMENCLATURE

LIST OF SYMBOLS

α_s	Solar Altitude Angle
β	Inclination, Collector Tilt
δ	Solar Declination
ε_0	Eccentricity Correction Factor
γ	Azimuth
γ_s	Azimuthal Angle
θ	Incidence Angle
θ_z	Zenith Angle
I_{norm}	Normalized Data
G_0 / G_{Ext} / H_0 / I_0	Extraterrestrial Solar Irradiance (Wh/m²)
Φ	Latitude Angle
Φ_{Sun}	Total Sun Radiation Flux
Ψ	Solar Azimuth
D	Diffuse Radiation
d_{re}	Reflected Radiation
E	Equation Of Time
E_{Sc}	Solar Constant
G	Global Radiation
G_h	Global Horizontal Irradiation
G_t	Solar Irradiance On A Slanted Surface
I_N	Normal Direct Radiation
I	Direct Radiation
K_t	Clearness Index
ρ	Albedo
S_d	Average Daily Insolation Duration
S_0	Maximum Insolation Duration
σ	Stefan-Boltzmann Constant
Y	The Year



General introduction



GENERAL INTRODUCTION

Solar power is a vital component of the global energy landscape since it is abundant and renewable. By using it, there is a way to lessen reliance on fossil fuels and lessen the negative environmental effects of using conventional energy sources. The smooth integration of solar energy into national electrical grids is severely hampered by its intermittent nature, which is defined by variations in sun radiation.

The fluctuations in solar irradiance can cause instability in energy generation, which can affect the dependability and efficiency of solar power plants. Additionally, the variability in solar radiation makes it difficult to match energy supply with demand, especially during peak consumption periods. These operational challenges are particularly felt by photovoltaic (PV) systems, which depend on consistent solar input for optimal performance.

Accurate solar irradiance predictions becomes essential to addressing these issues. Energy managers and grid operators may more effectively forecast variations in electricity generation and maximize the performance of photovoltaic systems by accurately forecasting solar radiation levels. Proactive decision-making, such as modifying energy production schedules, maximizing grid integration, and skillfully managing energy storage resources, is made possible by accurate projections.

Achieving high accuracy and reliability in solar radiation forecast models is still a challenge, even with major developments in the form of empirical, numerical, and machine learning-based approaches. Traditional forecasting techniques find it difficult to appropriately account for the complexities introduced by elements like terrain variations, cloud cover, and atmospheric variables.

Furthermore, areas with a variety of geographical and climatic circumstances have a greater need for improved forecasting. With their diverse geography and huge desert landscapes, nations like Algeria need customized forecasting methods that take into account local microclimates and weather patterns. For the purpose of optimizing resource allocation and maximizing energy output in these areas, the accuracy of solar energy projections is crucial.

The incorporation of sophisticated modeling tools, such as hybrid models that incorporate several forecasting methodologies, is part of the effort to improve solar radiation forecasting. To increase forecast accuracy and reliability, these hybrid models combine the best

features of several methodologies, including statistical analysis, physical modeling, and machine learning techniques.

Researchers hope to speed the shift to a sustainable energy future and enable the broad solar energy utilization by pushing the boundaries of solar energy predictions. Strong anticipating skills improve solar power projects' economic feasibility while also promoting global grid stability, energy security, and environmental sustainability.

Therefore, due to the limitation of the results of individual ML methods, models hybrids are developed based on a combination of two or more approaches. **This study's main goal** is to accurately anticipate fluctuations in solar irradiance through the use of hybrid models that use artificial intelligence techniques, metaheuristic algorithms, and other methods. In order to create hybrid forecasting models that can provide accurate predictions of solar radiation, this research will make use of a variety of input characteristics from various places around Algeria.

The thesis will be organized as follows in order to accomplish this goal and get deeper into the specifics of solar radiation forecasting:

Chapter I: Solar Notions:

An overview of solar radiation and its importance as a source of renewable energy is given in this chapter. It discusses the The sporadic character of solar power and the challenges it poses for energy management. It also examines the potential of solar power and the necessity of precise forecasting To be able to fully utilize its advantages.

Chapter II: Sources and Processing of Data Relating to Solar Radiation:

The second chapter discusses the different methods and tools for collecting and analyzing solar data, highlighting the importance of these data for efficient use of solar energy despite the challenges associated with the limited number of monitoring stations in Algeria.

Chapter III: Prediction Models:

This chapter explores a range of models for predicting solar radiation, including machine learning-based, deterministic, stochastic, and empirical methods ...; It goes over each model's advantages and disadvantages as well as how they affect forecasting solar energy. Also the chapter includes error indicators to access and compare these different models.

Chapter IV: FFA-ANN Model:

The creation and assessment of a hybrid forecasting methods for the prediction of solar irradiation are the topics of the fourth chapter. It talks about combining artificial intelligence method, and an optimization algorithm - Firefly Algorithm - to improve the forecasting process precision, this model is the hybrid FFA-ANN model, This chapter examines and evaluate hybrid model in various Algerian localactions along with two other methods (ANN and SVM) to determine the models efficiency.

Chapter V: Comparative Analysis and Evaluation a new GA-ANN model and ANFIS model:

A comparison of various solar radiation forecast models, including two hybrid methods created for this study, is given in the last chapter. By combining GA with ANN, the hybrid GA-ANN approach aims to improve the accuracy of solar forecasts, thereby promoting more efficient use of solar energy and environmental sustainability. The GA-ANN model was compared to the ANN and the hybrid ANFIS approachs, for three locations.

The study's main conclusions emphasize how crucial precise solar radiation prediction is to making the most optimal use of solar energy. It highlights the significance of ongoing developments in solar radiation prediction models for the shift to a sustainable energy future and explores possible directions for future research.

Chapter I:

Solar Notions

I.1. Introduction

Solar energy is the most plentiful and limitless source of energy in the globe. It's a renewable energy source. Direct electricity (produced by photovoltaic semiconductors) or solar thermal heat (used for power generation or heating) can be produced from sun radiation.

The choice to use solar energy has become essential to daily living. It is a great option for solar power applications, especially in desert regions, because of its favorable climate, remote locations, and high amounts of solar radiation. Transitioning from conventional energy systems to more abundant and environmentally friendly energy systems is now necessary.

This chapter provides an introduction of solar radiation, introducing us to several basic concepts (characteristic, nature, intensity, etc.).

I.2. The issue of sustainable development and climate change

The worldwide discussion surrounding climate change has evolved beyond a narrow focus on environmental issues to include a more comprehensive framework that links environmental objectives with those of social and economic development. Since it is known that climate change can have a negative influence on both economic growth and the achievement of the Sustainable Development Goals, managing climate change has become an essential component of both national and international strategies. In an international crisis

Table I.1 : The key events in the global awareness and response to climate change.

Year	Event
Late 1980s	The topic of sustainable development and climate change gained international attention at the same time. discussions between governments, economists, and scientists that also touch on public opinion over the use of fossil fuels, the overuse of natural resources, and measures to mitigate climate change without impeding economic growth.
1970s-1980s	Since the 1972 Stockholm Conference, which established the United Nations Environment Programme (UNEP), intergovernmental scientific conferences have concentrated on atmospheric and climate disturbances. Concerns about climate change were acknowledged by scientists at the first global climate conference held in Geneva in 1979. A crucial turning point in awareness came with the publication of the 1987 Brundtland Report, "Our Common Future."
1988	Global thought, spearheaded by the G7, established the Intergovernmental Panel on Climate Change (IPCC) in 1988 with the goal of producing recurring studies for decision-makers.
1992	The United Nations Framework Convention on Climate Change (UNFCCC) was adopted as a result of the United Nations Conference on Environment and Development (UNCED) in Rio de Janeiro. Stabilizing greenhouse gas concentrations was the aim in order to avoid harmful human interaction with the climate system.
1995 Onwards	Since 1995, countries around the world have gathered for the annual Conferences of the Parties (COP) to share ideas and talk about practical ways to address climate change. Important occasions include the Kyoto Protocol's creation at COP3 in Kyoto (1997), the historic Paris Agreement that replaced the Kyoto Protocol at COP15 in Copenhagen (2009), and COP21 in Paris (2015). The subject of the ongoing COP meetings, which included COP25 in Madrid, was the implementation of the Paris Agreement.

These events where only the start and the global awareness and response to climate change interests kept growing untill this day. Like many other nations, Algeria grew aware of the dangers of climate change at the national level when it ratified the Paris Agreement and the Kyoto Protocol, conforming to its climate-related obligations. Algeria is a developing nation and has never been responsible for the accumulation of greenhouse gases. Its existing responsibilities are rather limited due to its minimal greenhouse gas emissions [5].

Algeria hasn't avoided the effects of severe weather, though. Recognized for its semi-arid and arid environment, the nation is extremely vulnerable to global warming. There has

been a noticeable rise in extreme weather occurrences during the last 60 years. Thirty percent of animal species would perish in 2020 as a result of the expected shorter rainy season and a 1 to 1.5°C increase in temperature, according to experts from the Institute of Hydrometeorology for Training and Research. They also predict that global warming will cause an extra 3°C rise in temperature by 2050. Remarkably, the projected 2009 cost of damages resulting from climate change ranged from 1.3% to 4.3% of GDP.

Future efforts should concentrate on creating renewable energy sources, public transportation, insulation techniques for the construction industry, and putting in place monitoring and alert systems for heatwaves and droughts in cities that are susceptible, according to meteorological and environmental specialists in Algeria. Furthermore, the average yearly growth rate of energy consumption from fossil fuels was 7.7% in 2018, which was made worse by rising energy demand in comparison to a population that is estimated to be growing by 38,700,000 people [6]. As a result, the average annual consumption of oil equivalent per person increased.

In addition to the energy industry, Algeria has to deal with the requirements of its people in terms of housing, healthcare, and education within the framework of a financial crisis made worse by the drop in oil prices since 2014. Food security issues and the possibility of depleting natural resources exacerbate this. Algeria is thus potentially vulnerable to the effects of climate change, among other things, with consequences already being felt. These effects pose a particular threat to agriculture, water resources, and health. Thus, it is imperative that we step up our efforts to conserve the environment, fight pollution in all its manifestations, and instill in residents and decision-makers a respect for nature and the preservation of resources.

I.4. Renewable energies on a national scale

The nation's consumption of electricity has been steadily rising in the last few years. It is anticipated to increase by more than 6% a year, from roughly 64 TWh in 2015 to 83 TWh in 2020 and 150 TWh in 2030. 2013 witnessed a 4.3% increase in national electricity use from 2012 to 28% of total energy consumption (35% from natural gas and 30% from petroleum products). Remarkably, conventional sources account for 99.2% of the nation's electrical generation, with renewable energy providing just 0.1% and 0.7% of the total, respectively, for solar and hydropower [7].

Algeria has responded to this trend by showcasing its dedication to diversifying its energy sources and actively taking part in international initiatives aimed at lowering CO2 emissions. A goal of 40% of energy consumption from renewable sources by 2030 was established in 2011 as part of an ambitious program to develop and promote renewable energies [8]. Recently, this initiative was revised to include geothermal energy, biomass usage, cogeneration, and expanding solar thermal (CSP) by 2021.

By 2030, about 27% of the nation's electricity will be produced using renewable energy, provided the program is successfully implemented. According to the plans, more than 22 GW of renewable capacity will be installed, of which 10 GW will be used for export and 12 GW would be used to meet the country's electricity needs. The following is how this program is distributed among technology fields [8, 9]:

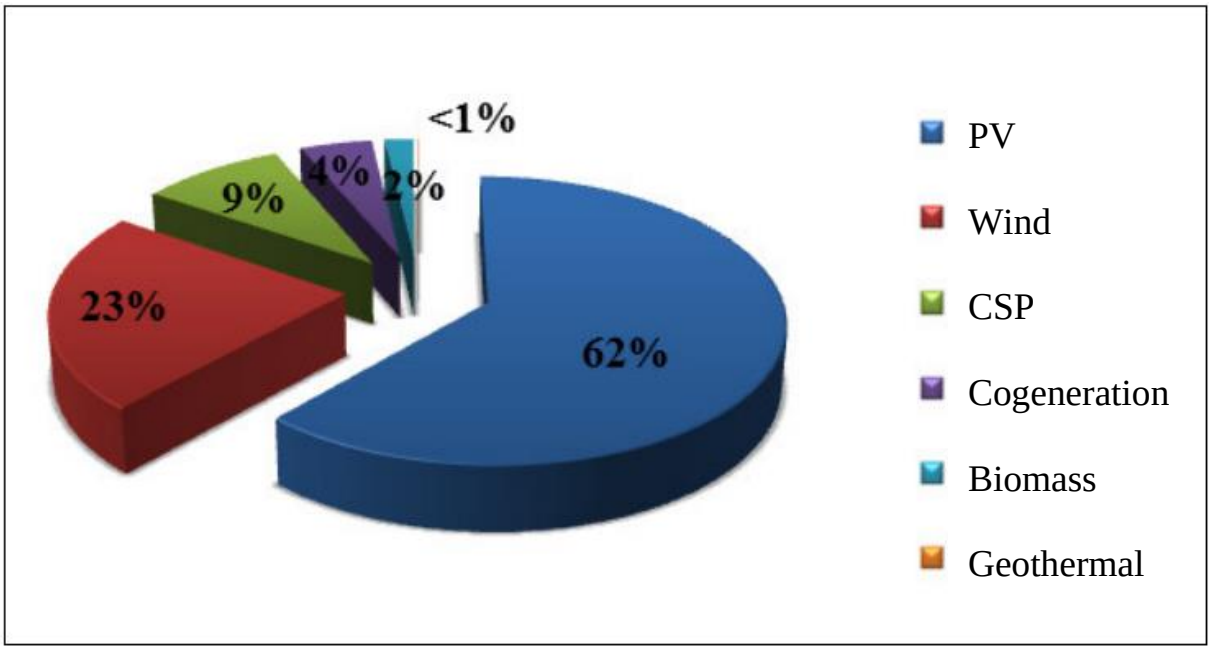


Figure I.1: Distribution of the national RE program by technological sector [8]

Consequently, the total is 22 GW, of which about 4.5 GW are expected to be completed by 2020. The program's results will have a big influence on industrialization, skill development, employment creation, and technical advancement, which will help the nation's economy grow, modernize, and protect the environment.

The percentage of electricity derived from renewable sources was projected to be 15% for solar and 85% for hydropower in 2012 [9]. This indicates that the pace of penetration of renewable energies in the country's electricity generation has increased.

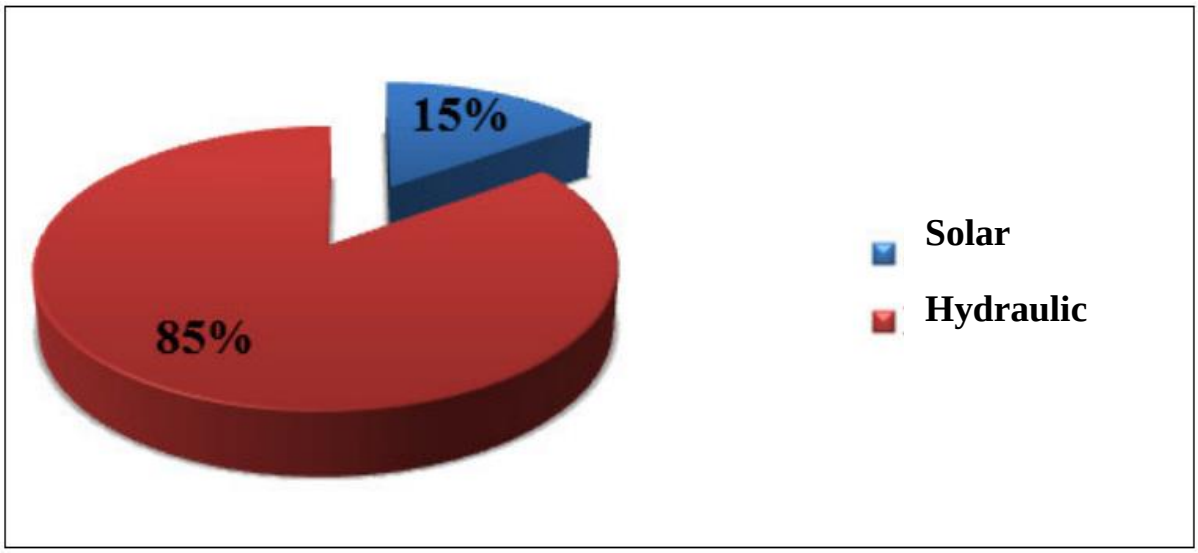


Figure I.2: Distribution of power generated from renewable sources (2012) [7]

Authorities envision a number of incentive measures to promote the implementation of this national plan for the growth of renewable energies and to entice businesses and industries interested in making investments in this area. The primary actions consist of [8]:

• Lowering of VAT and customs charges on equipment imports for the renewable energy sector (raw materials and semi-finished goods).

• The creation of the National Energy Management Fund (FNME), which would provide interest-free bank loans and financing for renewable energy projects.

• A law encouraging the application of sustainable energy;

• A statute pertaining to energy management.

Thus, sun, wind, and geothermal energy are Algeria's primary energy sources. Nevertheless, despite their unpredictability and sporadic nature, wind along solar power were anticipated of being crucial in meeting the nation's energy needs in the near future.

In addition to being essential for site qualification, an accurate characterization of these potentials is also required for the best possible energy conversion system design and methodology selection.

I.4.1. Solar Energy

The solar potential of Algeria is among the highest in the world because of its location; it has an estimated potential of 5 billion GWh/year and more than 2000 hours of annual insolation over almost the whole country, with 3900 hs throughout the Sahara and the upper plateaus [9, 10].

Over most of the country, the DGSR obtained on a horizontal ground of 1 m² is around 5 kWh; in the north, it is almost $1700kWh/m^2/year$, and in south, it is $2263kWh/m^2/yea$.

In northern regions, direct normal incidence (DNI) radiation is predicted to be $1800kWh/m^2/year$, whereas in southern regions, it might reach $2900kWh/m^2/year$.

This enormous potential shows that using this type of energy can be profitable in a number of industries, most notably the production of electricity. As a result, Algeria wants to use solar energy to provide 31.2% of its electricity by 2030 (9.2% for photovoltaic and 24% for concentrated solar power) [11].

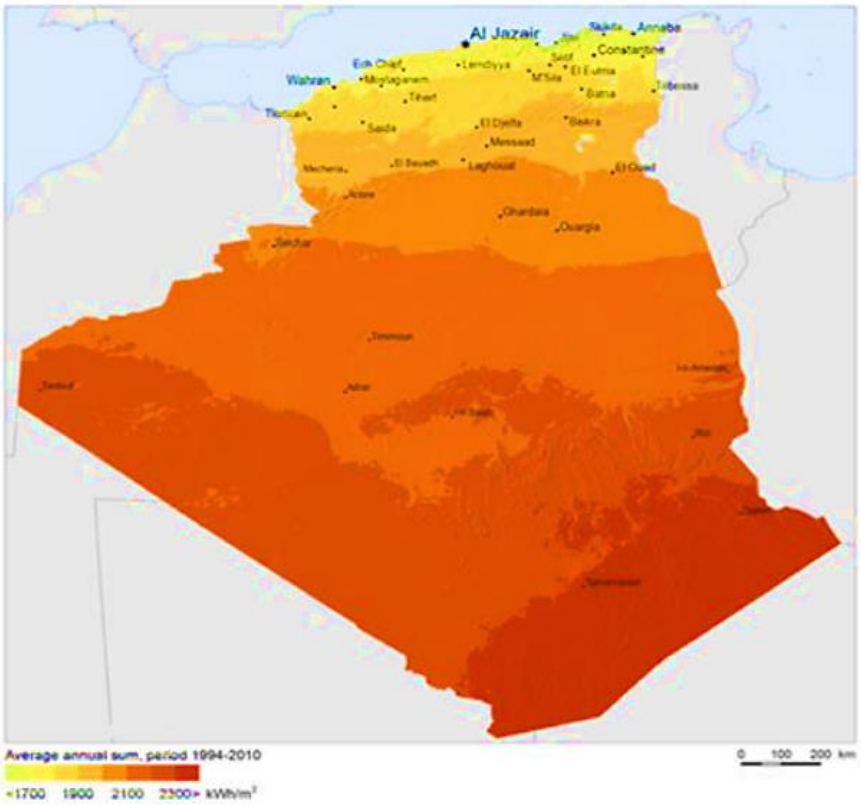


Figure I.3: Global irradiation on a horizontal plane [12]

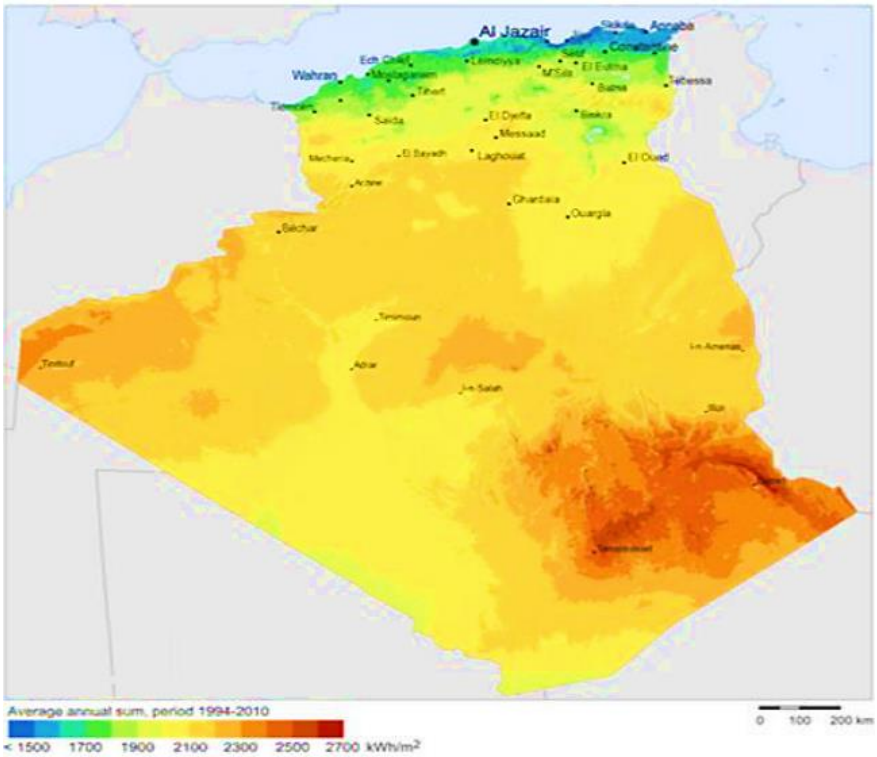


Figure I.4: Direct irradiation at normal incidence [12]

The advancements in solar energy systems have been around for a while. However, a number of applications pertaining to cathodic protections, public lighting, rural electrification, water pumps for irrigation in remote locations, and telemetry system power supply in hydrocarbon fields have been put into practice. PV plants are currently in the operational phase when it comes to producing power on a large scale. To evaluate the four current PV technologies, the Ghardaïa pilot plant, for instance, produces 1.1 MW of power.

In the extreme south and high plateaus, 23 projects totaling 400 MW are now being built. Indeed, as a component of the national initiative for renewable energy, 27 hybrid along 27 photovoltaic facilities are planned. Another method of using solar energy is to concentrate sun rays and produce electricity. The rays are focused using cylindrical-parabolic mirrors to heat a heat transfer fluid, which is often utilized to generate electricity.

Hassi R'mel's hybrid (solar-gas) power plant, which generates 150 MW, of which 25 MW is generated by solar energy, is one example of how this technology has been used. Since July 2011, this factory has been in operation. near the upcoming months, a second 20 MW CSP plant is supposed to be built near Beni Abbes, in the Province of Bechar. Reminder: by 2030, there will be six solar thermal plants scheduled.

I.5. Source Of Solar Irradiation

The electromagnetic power that the sun generates is known as solar irradiation, eventually reaching Earth in about eight minutes after it is released. Its definition is the quantity of energy delivered per unit area of the surface of the earth.

As a first approximation, the isotropic radiation emitted by the sun can be equated to the radiation of a black body, which is fully perfectly and emissive absorbent at $5800°K$. Solar irradiation is created in the sun's core by nuclear reactions, and its radiation towards the surface of the sun is rather steady. On the other hand, "varies substantially from one area to another and fluctuates throughout the year".

The sun's position in relation to the receiving ground has the biggest impact on the amount of solar radiation incident on surfaces of the planet. The wavelength range between 0.25 and 3μm accounts for 98 percent of the radiated energy. [13]

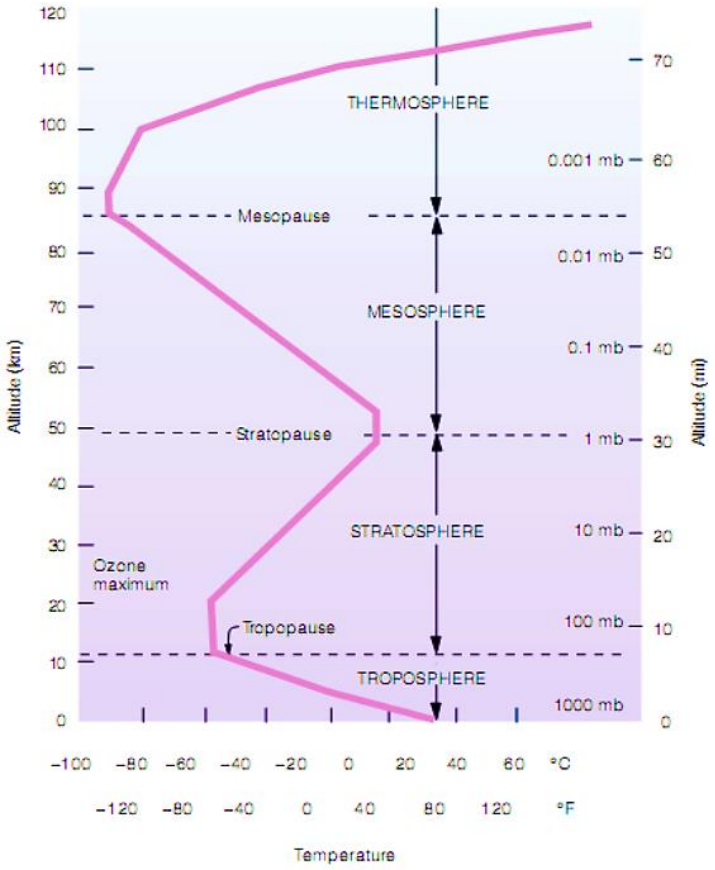


Figure. I.6: Structure of the atmosphere. Where the strong line represents the variance in each atmospheric layer's average temperature [15].

Aerosols and clouds are also present in the atmosphere. Aerosols make up a small fraction of the atmosphere's mass, but they have a significant impact on the behavior of solar radiation in the atmosphere via the different processes of reflection, diffusion, and absorption [17]. Clouds reflect around 20% of the solar energy that reaches the top of the atmosphere back into space.

The amount of direct illumination attenuated by clouds is determined by the type of cloud, its thickness, and the number of layers that comprise it. It is difficult to determine direct and diffuse radiation theoretically. Such a job need detailed information on the optical kinds and characteristics of clouds.

TABLE I.2: The atmosphere's composition at the vicinity of the surface of the Earth [15].

Permanent Gases			Variable Gases		
Gas	Symbol	Percent (by Volume)	Gas (and Particles)	Symbol	Percent (by Volume)
Nitrogen	N_2	78.08	Water vapor	H_2O	0 to 4
Oxygen	O_2	20.95	Carbon dioxide	CO_2	0.037
Argon	Ar	0.93	Methane	CH_4	0.00017
Neon	Ne	0.0018	Nitrous oxide	N_2O	0.00003
Helium	He	0.0005	Ozone	O_3	0.000004
Hydrogen	H_2	0.00006	Particles (dust, soot, etc.)		0.000001
Xenon	Xe	0.000009	Chlorofluorocarbons (CFCs)		0.00000002

I.6.2. Attenuation of solar radiation

As solar energy passes through the atmosphere, its intensity is greatly reduced by various atmospheric components such as molecules, aerosols, gases, cloud droplets, and ice crystals. Each of these elements has optical and physical characteristics of its own. Extinction-related physical phenomena are classified as follows:

- Radiation absorption by atmospheric gases such as O_2, O_3, H_2O, CO_2, and other pollutants.
- Radiation diffusion by particles, air molecules, and water vapor, which degrades during the diffraction, refraction, and reflection processes.

When passing through the 15 km thick atmosphere that roughly resembles the atmosphere, all wavelengths of solar energy do not respond in the same manner.

Schematically, we can point out that:

- In the ultraviolet region of the sun spectrum, wavelengths less than 0.3 micrometer are completely absorbed by ozone (O_3) near the top limit of the atmosphere.
- Visible radiation from 0.4 to 0.8 micrometer is relatively little absorbed in the case of a clear sky, but on the other hand its diffusion is all the more important as the wavelength is weak;
- Infrared radiation is absorbed by gases in the troposphere such as CO_2, H_2O, and O_3.

As shown in Figure I.7, these atmospheric components have an effect on the spectrum distribution of solar spectral radiation.

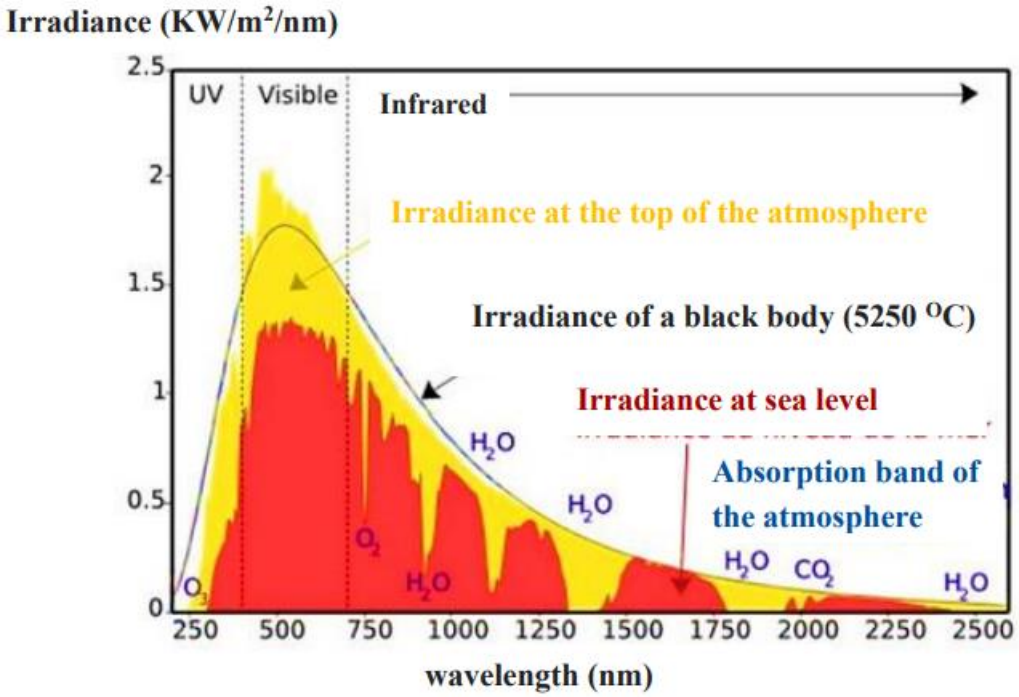


Figure. I.7: Spectra of mean solar radiation incident [18].

I.7. Radiation on Ground

Electromagnetic waves from the sun make their way to the top of the atmosphere. It is first separated into two sections:

- Before entering the atmosphere, the former is instantly reflected back to space.
- The second component dips into the atmosphere, where it is refracted and diffracted before being returned to space or to the earth.

Because of these processes, solar rays reaching the earth may be classified into three types: direct, diffuse, and reflected [19].

a) Normal direct radiation I_N

Direct solar radiation is radiation that arrives on a normal plane straight from the solar disk without any diffusion. It differs substantially from radiation reaching the upper limit of the atmosphere, due to absorption by the atmosphere's gaseous components.

b) Diffuse radiation D

The horizontal diffuse D_H is the radiation received on a horizontal surface coming from all the celestial vault except the solar disk. It is the radiation scattered by aerosols, water droplets, water vapor towards the ground.

c) Reflected radiation d_{re}

The reflected radiation is a component of the illumination by the ground, this component is only considered in the situation of an inclined surface, and it is mostly determined by the kind of surface.

Albedo is defined as the proportion of reflected energy compared to incident energy. This term is typically reserved for the ground or clouds, and it refers to the average value of their reflectance for the radiation under consideration and for all possible angles of incidence.

It is very dependent on the type of the soil and the state of the surface. The overall albedo of the Earth/Atmosphere system is around 0.3, which implies that 30% of solar energy is reflected and 70% is absorbed. [20]

d) Global radiation G_h

It is received on an area from the whole celestial vault and the solar disk.

❖ On a horizontal surface

The irradiation reaching the ground on a horizontal surface G_h is the sum of direct and diffuse radiation (Figure I.8), according to the following formula:

$$G_h = D_h + I_N \cos(\theta_S) \quad \text{(I.1)}$$

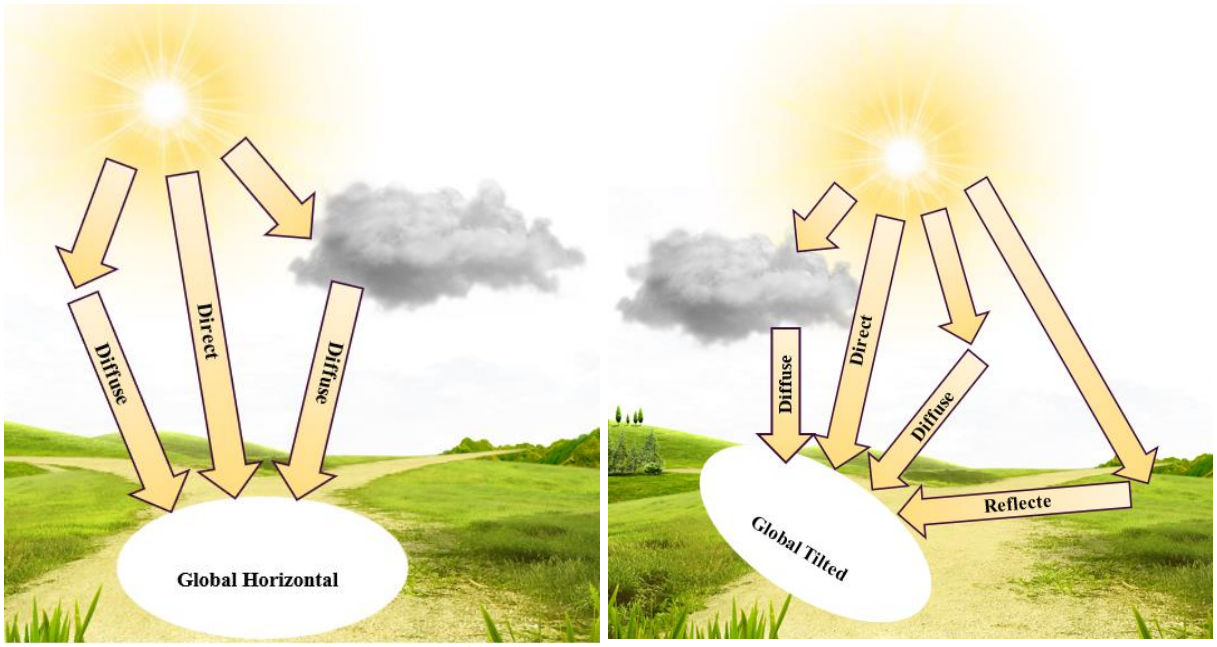


Figure. I.8: Main solar radiation components.

❖ On an inclined surface

The irradiation on an arbitrary surface with inclination β with respect to the horizontal and an azimuthal angle γ_s is the sum of the three components direct, diffuse and reflected by the ground of the solar radiation (Figure I.8) [13, 14].

$$G_\beta = D_\beta + I_{N_\beta} + d_{re} \quad \text{(I.2)}$$

Solar energy varies depending on:

- Place Latitude.
- Orientation and inclination of the collection plane.
- The season (maximum in summer, minimum in winter).

- The atmospheric situation (clear or cloudy sky).
- The sun's horizonal height, which reaches its greatest at solar noon.

I.8. Flux of Irradiation

radiant flux that the sun emits is described by radiation laws. The sun is assumed to be a black body emitter with a surface temperature of 5780K [21].

According to physics laws, A blackbody's (M) total radiative flux density at a particular temperature (T) is expressed in terms of fallows.

$$M = \sigma * T^4 \quad \text{(I.3)}$$

With $\sigma = 5.67 * 10 - 8\, Wm^{-2}\, K^{-4}$ is the Stefan-Boltzmann constant with $T = 5780\, K$.

$$\mathrm{M_{Sun}} = 5.67 * 10^{-8} * 57804\, Wm^{-2} \quad \text{(I.4)}$$

$$\mathrm{M_{Sun}} = 6.33 * 10^{7}\, Wm^{-2} \quad \text{(I.5)}$$

Surface area of the sun, A_{Sun} is calculated as:

$$A_{Sun} = 4\pi * {r_{Sun}}^2 \quad \text{(I.6)}$$

With $r_{Sun} = 6.96 * 10^8 m$ Total sun radiation flux,

$$\Phi_{Sun} = M_{Sun} * A_{Sun} = 3.85 * 1026\, W \quad \text{(I.7)}$$

I.9. Solar Constant

The solar constant represents the amount of solar radiation received per unit area on a surface oriented perpendicular to the sun's rays at the average distance between the Earth and the sun, measured outside the Earth's atmosphere. Based on the principles of energy conservation, the total radiant flux emitted by the sun (Φ_{Sun}) is equivalent to the flux passing through any spherical surface centered on the sun.

This is particularly relevant for a sphere with a radius equal to the average distance from the sun to the Earth (r_{ES}), as depicted in Figure I.9.

$$4\pi * r_{Sun}^2 * M_{Sun} = 4\pi * r_{ES}^2 * E_{Sc} \quad \text{(I.8)}$$

Equation 1.9 may be used to calculate the solar constant, which is the radiation flux passing through a unit area at the average distance between the Earth and the sun.

$$E_{Sc} = \left(\frac{r_{Sun}}{R_{ES}}\right)^2 M_{Sun} = 1367\ Wm^{-2} \pm 4\ Wm^{-2} \quad \text{(I.9)}$$

As a result, solar flux received by the cross sectional area of the earth disk as seen by the sun may be computed using E_{Sc}.

$$\Phi_{Earth} = A_{Earth} * E_{Sc} = \pi r^2 * E_{Sc} \quad \text{(I.10)}$$

With r being the mean earth radius of $6371\ km$.

And so,

$$\Phi_{Earth} = 3.142 * (6.371 * 10^6\ m)^2 * 1367\ Wm^{-2}$$

$$= 1.75 * 10^{17}\ W \quad \text{(I.11)}$$

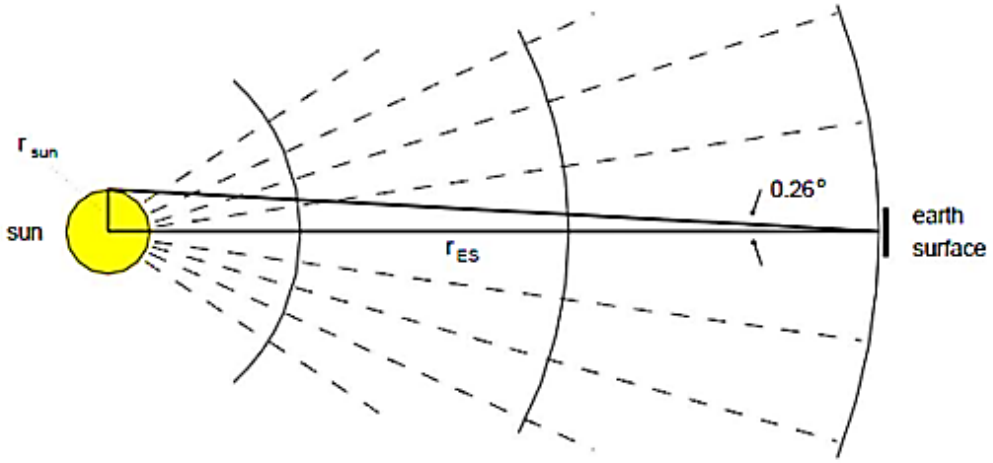


Figure I.9: Geometry of Sun-Earth [21]

I.10. Extraterrestrial radiation

Due to the Earth's elliptical orbit, the intensity of solar radiation at the top of the atmosphere varies throughout the year. Consequently, the distance between the Earth and the sun (r_{ES}) changes, introducing some eccentricity. This variation can be adjusted using an eccentricity correction factor (ε_0), which was empirically calculated by Spencer in 1971.

$$\varepsilon_0 = 1.00011 + 0.034221\ cos\ d + 0.00128\ sin\ d + 0.000719\ cos\ 2d + 0.000077\ sin\ 2d \quad \text{(I.12)}$$

With $d = 2\pi * \frac{(n-1)}{365}$ (I.13)

Equation I.12's simple approach commonly used,

$$\varepsilon_0 \approx 1 + 0.033 * cos\left(\frac{360\,n}{365}\right) \quad \text{(I.14)}$$

And the Extraterrestrial irradiance could be;

$$G_{Ext} = \varepsilon_0 * E_{Sc} \approx E_{Sc} * \left[1 + 0.033 * cos\left(\frac{360\,n}{365}\right)\right] \quad \text{(I.15)}$$

I.11. Astronomy of Sun-Earth

I.11.1. Earth movement around the Sun:

The Earth's orbit around the sun is elliptical, with the sun at one of the focal points. This orbital plane is known as the ecliptic. [22]

The Earth travels along this path at a constant average speed of 29.8 km/s (or 107,280 km/h), covering about 1° per day since it completes a full orbit in one year (360°/365 days). The plane of Earth's orbit around the sun is inclined at an angle of 23° 27' relative to the equatorial plane.

As the Earth orbits the sun, its axial tilt remains parallel to itself, making the annual journey similar to a geometric translation along a circular path. [23]

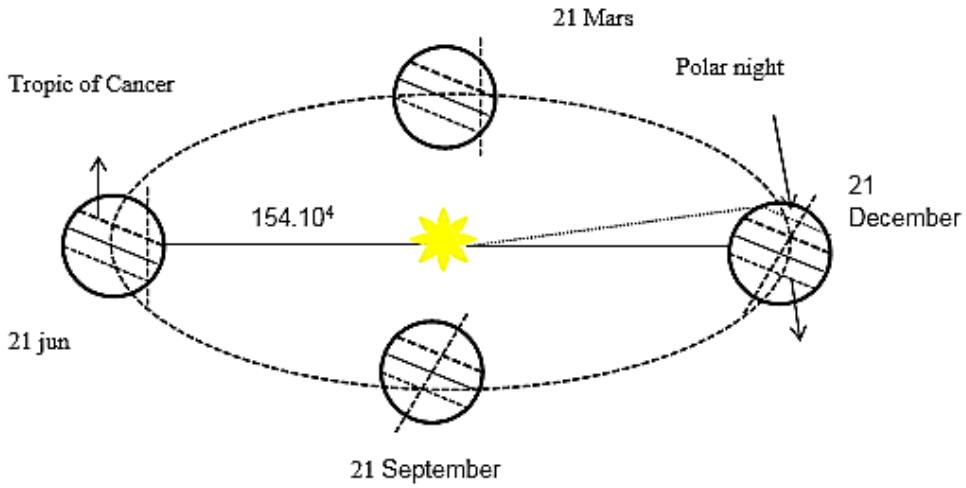


Figure.I-10: The earth movements.

- During the winter solstice on December 21, the sun's rays strike the Earth at a declination angle of -23° 27', marking the minimum declination value.
- On the spring equinox (March 21) and autumn equinox (September 22), the sun's rays are aligned with the equator, resulting in a declination angle of 0°, which corresponds to equal lengths of day and night.
- On the summer solstice, June 23, the Earth's position is opposite to that of December 21, and the sun's rays strike the northern hemisphere at a maximum declination angle of 23° 27'. [24]

I.11.2. Declination

The Earth travels along an elliptical plane with its axis tilted approximately 23.45° relative to its orbit around the sun (see Figure I.11).

On June 21, during the northern hemisphere's summer, the Earth's axis tilts towards the sun. Conversely, on December 22, during the northern hemisphere's winter, the axis tilts away from the sun. Solar declination is the angle between the Earth's equatorial plane and its orbital plane around the sun.

The solar declination for day n can be expressed as:

$$\delta \approx 23.45 * \ sin\left(\frac{360(348+n)}{365}\right) \quad \text{(I.16)}$$

with n taking values of 1 to 365.

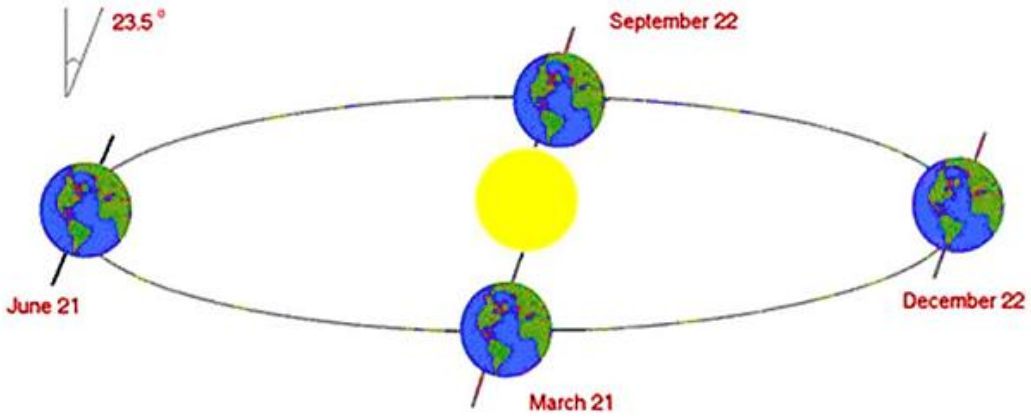


Figure I.11: The view of Earth's Axis from space [25]

I.11.3. Local Standard and Solar Time

Solar time is an essential parameter for calculating the variance in solar irradiation for different places.

As a result, solar time varies based on longitude on the Plant's surface. The relationship between local and solar standard time considered essential for determining the sun's position. A day is represented by the interval between the lowest and highest solar altitudes, which marks midday in true solar time. This period averages 24 hours.

Due to the Earth's elliptical orbit and its axis not being perpendicular to the ecliptic plane, the length of the day varies throughout the year. Local time, however, consistently runs for exactly 24 hours each day.

The difference in minutes between solar time and local mean time can be expressed using the equation of time, E, in 1.17.

$$E \approx 9.87 * sin2\,\Phi - 7.53 * cos\,\Phi - 1.5 * cos\,\Phi \quad \text{(I.17)}$$

The time zone determines the standard time. Solar time differs from standard time because of differences in how long a solar day is and the difference between the local longitude and the standard longitude of the relevant time zone (See Figure.I.12).

$$\boldsymbol{True\ Solar\ Time = Local\ Time + E} \quad \text{(I.18)}$$

Local time is determined by local longitude, L_l, and standard time is determined by standard longitude, L_s. True Solar Time (TST) may be derived from Local Time (LST),

$$\boldsymbol{TST = LST - DST + 4(L_s - L_l) + E} \quad \text{(I.19)}$$

In this case, DST = 0 outside of daylight saving time and = 1 hour during it.

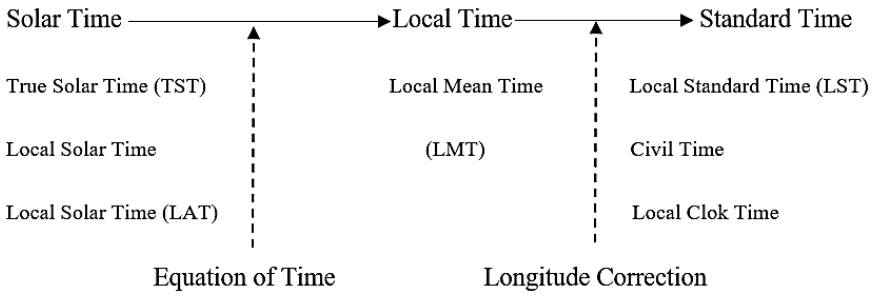


Figure I.12. The relationships between the terms of time [21]

I.11.4. Hour Angle

TST's measurement unit is hours. The hour angle (ω) is defined to facilitate computations in Cos and Sine equations.

The sun's angular displacement from the local meridian due to the earth's rotation is known as the hour angle. Radians are used to measure it. Hour angle and time are related to each other by

$$\omega = \pi * (12 - TST)/12 \quad \text{(I.20)}$$

The angle of the sun with respect to a line that is perpendicular to the surface of the earth is known as the zenith angle (θ). Furthermore, the azimuthal angle (γ_s) indicates the sun's position with respect to the north-south axis.

I.11.5. The Sun's Position

According to [26], the position of the sun with respect to the surface is determined by certain angles, which also explain the geometric relationship between any mobile surface with respect to the earth at any given time—regardless of whether the surface is stationary or moving [26]. The angles are illustrated in Figure.I.13.

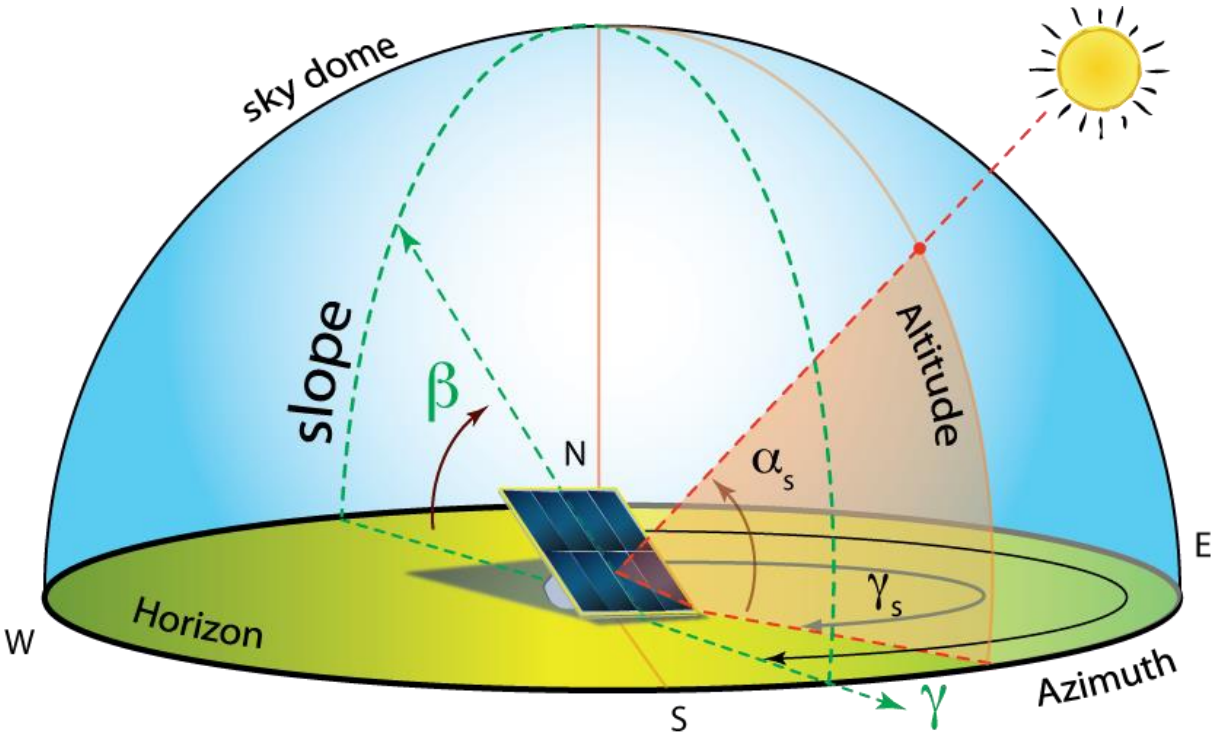


Figure. I.13: azimuth (γ), solar altitude angle (α_s), collector tilt (β), solar azimuth angle (γ_s) [27].

Angle of solar altitude (also known as elevation angle):

As shown in Figure.I.13, solar altitude is the angle formed by the line of incoming sun rays and the horizontal surface. This solar altitude is represented by

$$\alpha_S = sin^{-1}\left[sin(\Phi)\, sin(\delta) + cos(\Phi)\, cos(\delta)\, cos(\omega)\right] \quad \text{(I.21)}$$

where $\boldsymbol{\alpha_s}$ is the solar altitude angle, Φ the latitude angle, the declination angle δ, and the hour angle ω. To compute the irradiance normal to a given tilted plane, the sun's location with respect to that plane must be known.

The location of the Sun in relation to the sky hemisphere is characterized by solar height (elevation above the horizon) and solar azimuth (See Figure I.14).

The height of the Sun is determined by

$$\text{Sin } h = \sin\delta * \sin\phi + \cos\delta * \cos\phi \quad \text{(I.22)}$$

The solar azimuth may be calculated using,

$$\cos\Psi = \frac{\sin\alpha * \sin\Phi - \sin\delta}{\cos\alpha * \cos\Phi} \quad \text{(I.23)}$$

The angle produced between the normal to the receiving surface and the incoming solar beam radiation is known as the incidence angle (θ) (see to Figure I.14).

Geographical location (latitude), solar declination (year), hourly angle (time of day), and slanted surface orientation all affect this angle.

The plane's azimuth angle, or the deviation of the plane's normal from the local meridian, and slope angle, or the angle between the collector plane and the horizontal surface, define the plane's orientation.

The incidence angle (θ) can be found using

$$\cos\theta = \sin\delta\,(\sin\Phi\cos\beta + \cos\Phi\sin\beta\cos\gamma) + \cos\delta\cos\omega\,(\cos\Phi\cos\beta - \sin\Phi\sin\beta\cos\gamma) - \cos\delta\sin\beta\sin\gamma\sin\omega \quad \text{(I.24)}$$

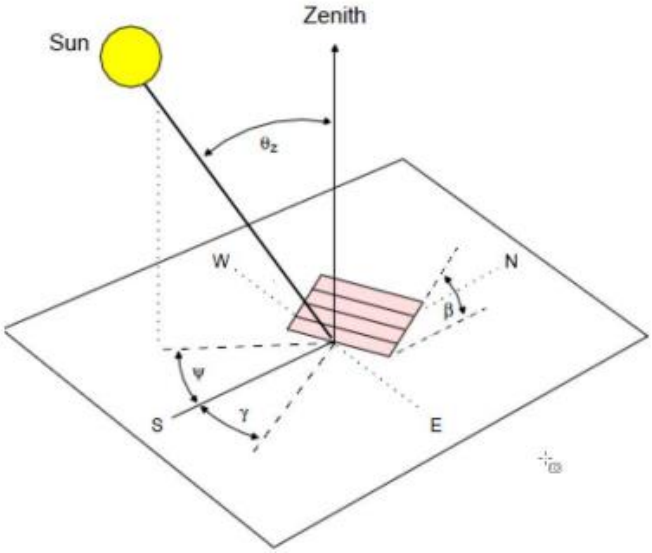


Figure I.14. Position of the Sun [21]

I.12. Clearness Index

Aerosols, clouds, water vapor, air molecules, and so on absorb or disperse a portion of the solar energy that penetrates the earth's atmosphere.

Solar radiation that reaches the earth's surface directly is referred to as direct solar radiation. [21].

Diffuse solar radiation is the radiation that has been dispersed from the direct beam [21].

As a result, the total of the direct and diffuse components of sunlight on a horizontal surface is the global solar irradiance.

Clouds and turbidity in the atmosphere reduce the amount of incoming radiation. Such a decline can be quantified using the metric clearness index, which normalizes global irradiance to extraterrestrial irradiance.

Because of this, the clearness index is a dimensionless figure that shows the percentage of extraterrestrial radiation that is reduced as a result of air molecules, aerosols, clouds, and water vapor absorbing and scattering light. The definition of the hourly clearness index, or k_t, is

$$k_t = \frac{\text{Global Horizontal radiance}(\text{Wm}^2)}{\text{Extraterrestrial radiance } (\text{Wm}^2)} \quad \text{(I.25)}$$

Also daily clearness index k_t maybe defined as,

$$k_t = \frac{\text{Global Horizontal daily sun radiance}(\text{Wm}^{-2})}{\text{Extraterrestrial daily sun radiance } (\text{Wm}^{-2})} \quad \text{(I.26)}$$

I.13. Solar radiance (Tilted Plane)

The horizontal plane is used extensively for irradiance measurement.

However, for solar energy research and applications, Since nonhorizontal irradiance data is relevant to solar energy conversion systems, it is required for slanted surfaces.

As a result, data records of horizontal radiation must be simulated to obtain approximations of the irradiance on inclined surfaces.

And this by conversion models on inclined planes must first be constructed.

It is common practice to divide solar radiation into three components when assessing irradiance on a slanted surface (G_t): Sky Diffuse (G_{dt}), Direct Beam (G_{bt}), and Ground Reflected Irradiance (G_{rt}).

$$G_t = G_{bt} + G_{dt} + G_{rt} \quad \text{(I.27)}$$

The definition of the global slanted irradiance's direct radiation component is as follows:

$$G_{bt} = G_{\mathrm{h}} \frac{\cos\theta}{\cos\theta_Z} \quad \text{(I.28)}$$

Where G_{h} is the global horizontal radiation and θ_z, θ are the incidence zenith angle and angle, respectively.

Under the supposition that the ground reflects irradiation isotropically, It is possible to provide the global slanted irradiance component of reflected radiation.

$$G_{rt} = G_h \rho \frac{(1-\cos\beta)}{2} \quad \text{(I.29)}$$

Where $\boldsymbol{\rho = G_r / G_h}$ is the surface's albedo, or reflectance.

The model developed by Liu and Jordan, 1963, represents the diffuse radiation component of global tilted irradiance.

Over the sky dome, the diffuse sky radiation is anticipated to be distributed evenly (isotropically).

The diffused horizontal radiation and the view factor from the surface to the sky are therefore used to calculate the diffuse radiation on a tilted surface (G_{dt}).

$$G_{dt} = G_d \frac{(1+\cos\beta)}{2} \quad \text{(I.30)}$$

Where G_d is global diffuse horizontal radiance and β is plane's slope.

Finally, the equation I.27 may be used to calculate global irradiance on a tilted plane.

$$G_t = G_h \frac{\cos\theta}{\cos\theta_Z} + G_d \frac{(1+\cos\beta)}{2} + G_h \rho \frac{(1-\cos\beta)}{2} \quad \text{(I.31)}$$

Empirical models that link the diffuse fraction to the clearness index can also be used to break down global radiation into its diffuse and global components. These approaches have been presented by authors in the papers [28], [29], [30], and [31], but the model suggested in [29] is still widely applied.

I.14. Conclusion

Photovoltaic semiconductors can directly generate electricity from solar radiation, or solar thermal heat can be used for power generation or heating. With little negative effects on the environment and significant potential for cost savings, this form of energy provides a viable substitute for the use of fossil fuels by harnessing solar radiation, which is thought to be the primary resource and is highly significant on a planetary scale.

This energy source's randomness highlights how crucial it is to understand the sun's potential before installing any solar systems. A comprehensive set of ground-based measurements of several radiation characteristics is the most effective method to determine this potential at a specific location.

This is the reason that we will outline the many approaches and methodologies for forecasting as well as measuring solar radiation in the upcoming chapters, along with a description of each approach.

I.15. References

1. Commission Economique pour l'Afrique (2015) note conceptuelle, Développement Durable et Changement Climatique: comment se positionne l'Afrique du Nord, p. 16.
2. Veyret, y. (2017). Le développement Durable,Paris, édition sedes, p. 22. https://doi.org/10.4000/developpementdurable.2959
3. OECD (2008), "Promoting Sustainable Consumption: Good Practices in OECD Countries" OECD insights, p. 141, www.oecd.org/dataoecd/1/59/40317373.pdf
4. Nations Unies (1992). CCNUCC, Art2, p. 5. https://www.ifdd.francophonie.org/wp-content/uploads/2021/09/647_Resume_decideurs_CdP21-climat_IFDD-2.pdf
5. CPDN-Algérie (2015). Contribution prévue déterminée au niveau national, p. 3. https://unfccc.int/sites/default/files/NDC/2022-06/Alg%C3%A9rie%20-INDC-%2003%20septembre%202015.pdf
6. l' Agence Nationale de la Promotion et de la Rationalisation de l'Utilisation de l'énergie (2019) La situation énergétique régionale, p. 19.
7. Observ'ER. (2015). La production d'électricité d'origine renouvelable dans le monde. http://www.energies-renouvelables.org/observ-er/html/inventaire/pdf/15e-inventaire-Chap01-Fr.pdf
8. Programme des énergies renouvelables et de l'efficacité énergétique (mars 2011). https://www.cder.dz/vlib/bulletin/pdf/bulletin_020_19.pdf
9. www.energy.gov.dz .Accessed on 21/03/2022
10. www.sktm.dz . Accessed on 06/01/2022
11. Othenio H., Awange J., (2016). Energy resources in Africa; Distribution, opportunities and challenges. Edition Springer.
12. www.solargis.info
13. Bessemoulin P. and Oliviéri J., (2000) ''Le rayonnement solaire et sa composante ultraviolette, La Météorologie 8e série - n° 31 – septembre, 2000.
14. Codron F., (2012). Transfert radiatif bilan énergétique, Note du cours, Université Pierre et Marie Curie-Paris 6, France, 2012.
15. Ahrens D. C, (2000). Essentials of meteorology, an invitation to the atmosphere, 3Ième Ed, 2000.
16. CEA C., (2009). Terre et environnement, Memo A « Voyage au centre de la Terre et aux confins de l'atmosphère, Hiver 2008-2009, N°57, PP21-24, 2009.

17. Demoor G, (2007). transfert radiatif en météorologie dynamique, Cours et manuels, N°17, Météo-France, 2007.
18. Wikipedia. (2024). Bilan radiatif de la Terre, lien : https://fr.wikipedia.org/wiki/Bilan_radiatif_de_la_Terre. Accédé en mars 2024.
19. Tadili R. et M. N. Bargach, (2005). Une méthode d'estimation du rayonnement solaire global reçu par une surface inclinée, La Météorologie - n° 50, pp. 46-50, Septembre 2005.
20. Oumbe A., (2009). Exploitation des nouvelles capacités d'observation de la terre pour évaluer le rayonnement solaire au sol, PhD thesis, MINES ParisTech, 2009.
21. Heinemann D., (2002). Energy Meteorology, Postgraduate program in Renewable Energy», University of Oldenburg, Germany.
22. Bernard J, (2004). Energie solaire calculs et optimisation », Ellipse Edition Marketing.
23. Daoudi A., Hadadji M., (2009). Calcule du rayonnent solaire pour une surface inclinée (cas d'un site situé à Ouargla), mémoire d'ingénieur d'état, Université Kasdi Merbah Ouargla.
24. Raud R., (2007). Capteur solaire thermique concentrateur a conduite manuelle », Association SOLEIL & VAPEUR.
25. Vignola F., Solar Radiation Monitoring Laboratory, University of Oregon, (2010). Available: http://solardat.uoregon.edu/SolarRadiationBasics.html#Fig1. Accessed 2023.
26. Benford, F. and Bock, J.E., (1938) A time analysis of sunshine. General electric Company, Research laboratory.
27. Brownson, J. (2016) Tools for Time and Space Relationships. Available at: https://www.e-education.psu.edu/eme810/node/576
28. Orgill J.F. and Hollands K.G., (1977) Correlation equation for hourly diffuse radiation on a horizontal surface. Solar Energy, 19, 357-359. Master Thesis-Renewable Energy-2014 72
29. Collares-Pereira M. and Rabl A. (1979) The average distribution of solar radiation correlations between diffuse and hemispherical and between daily and hourly insolation values. Solar Energy, 22, 155-164.
30. Erbs D.G., Klein S.A. and Duffie J.A., (1982) Estimation of the diffuse radiation fraction for hourly, daily and monthly-average global radiation. Solar Energy, 28, 293-304.
31. Reindl D.T., Beckman W.A. and Duffie J.A., (1990) Diffuse fraction correlations. Solar Energy, 45 1-7

Chapter II: Sources And Processing of Data Relating To Solar Radiation

II.1.Introduction

Understanding solar potential is crucial before sizing and constructing any solar capture system because solar energy is unpredictable and discontinuous, and because its availability changes over time in response to the demands of energy exploitation. The most effective method for evaluating solar potential at a specific location is to take a thorough set of measurements of ground radiation. Therefore, in order to effectively assess the region's potential for solar energy, it is imperative to understand the procedures for gathering, interpreting, and evaluating data on this radiation. A crucial component of energy planning is a precise and trustworthy evaluation of solar resources.

Although Algeria has a relatively small network of measurement stations and hence low territorial coverage, this chapter clarifies the significance of radiometric measurements, which is a contribution to our current work. This work highlights the necessity of optimizing the sizing of any proposed installation and simulating the operation of anticipated solar energy systems.

II.2.Solar Radiation Measurements

Sunshine corresponds to the measurement of solar radiation received by a surface during a given period, representing the intensity of solar radiation received on a plane at a specific moment. It is usually expressed in watts per square meter (W/m^2). Sunshine varies from zero at sunrise and sunset to its maximum value, typically at solar noon. Sunshine can also express the amount of solar energy captured on a plane during a specific interval.

The parameters are mainly distinguished based on their origin, composition, or the measurement method, as follows [1]:

II.2.1. Direct Solar Radiation

The measurement of direct solar radiation is carried out using a pyrheliometer (Figure II.1), an apparatus with a sensitive surface always positioned perpendicular to solar rays. Direct solar radiation is the sunlight that reaches the ground without undergoing any diffusion. It significantly differs from the radiation reaching the upper limit of the atmosphere, particularly due to absorption by the gaseous constituents of the atmosphere.

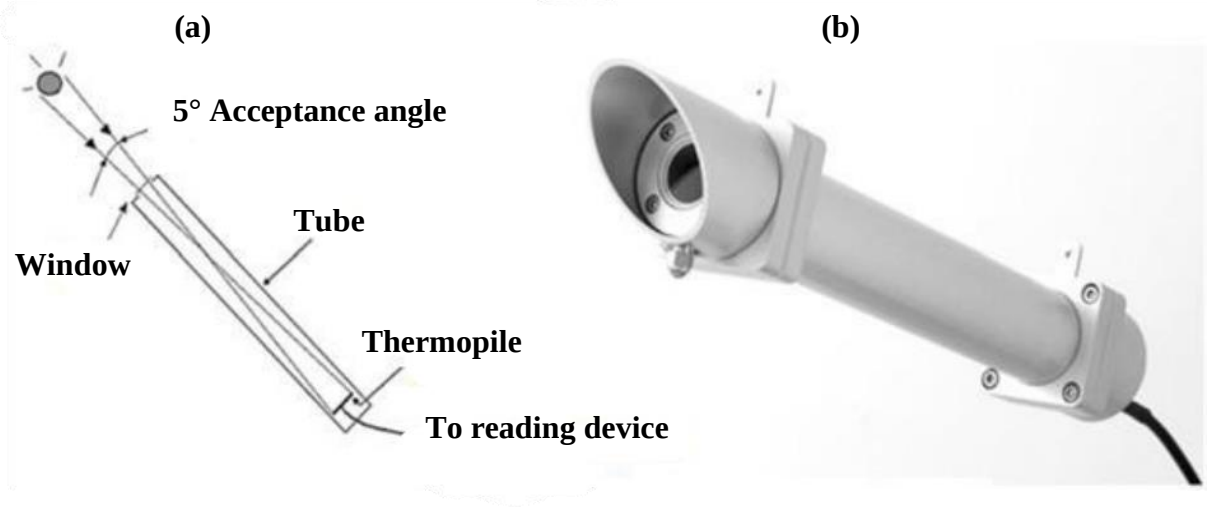


Figure II.1: (a) Diagram of a Pyrheliometer, (b) Pyrheliometer first-class type Hukseflux DR01 [2]

II.2.2. Diffuse Radiation

This type of radiation is measured by a device called a "Pyranometer," equipped with a sunshield (a metal ring that hides the solar disk). This radiation comes from the celestial vault, excluding the solar disk. In other words, it represents the portion of solar radiation scattered by solid or liquid particles suspended in the atmosphere.

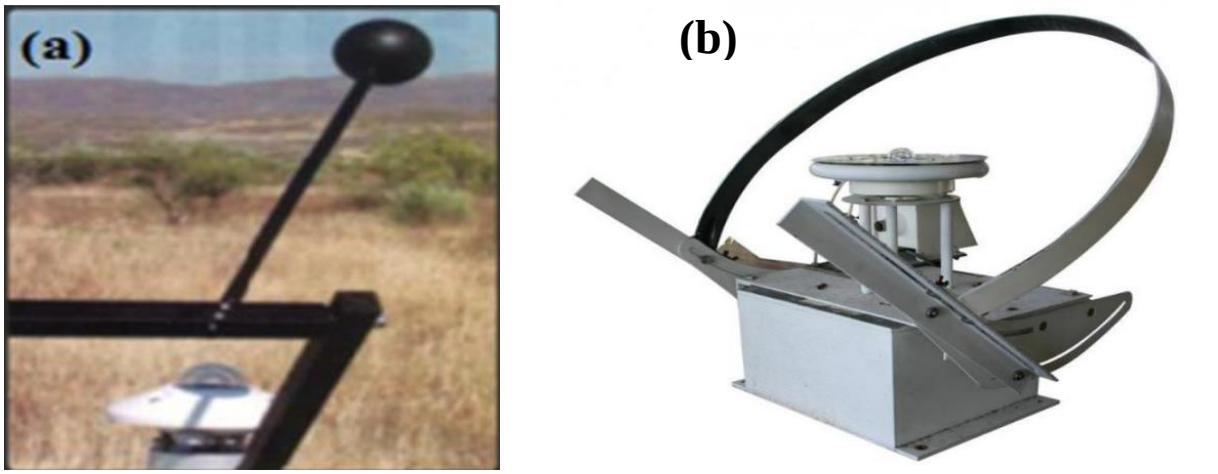


Figure II.2 : Measurement of diffuse (a) pyranometer with shading ball mounted on a tracking system, (b) pyranometer with shading ring [2]

II.2.3. Global Radiation

The pyranometer (Figure II.3) is an instrument for measuring solar irradiance (both direct and diffuse) reaching the Earth's surface in the wavelength range of 0.3 to 3 μm. With this instrument, it is possible to measure not only the global radiation but also the reflected solar radiation. Global radiation is also defined as the sum of direct and diffuse radiation, according to the formula: $G = I \sin \alpha_s + D$; Where α_s is the solar altitude.

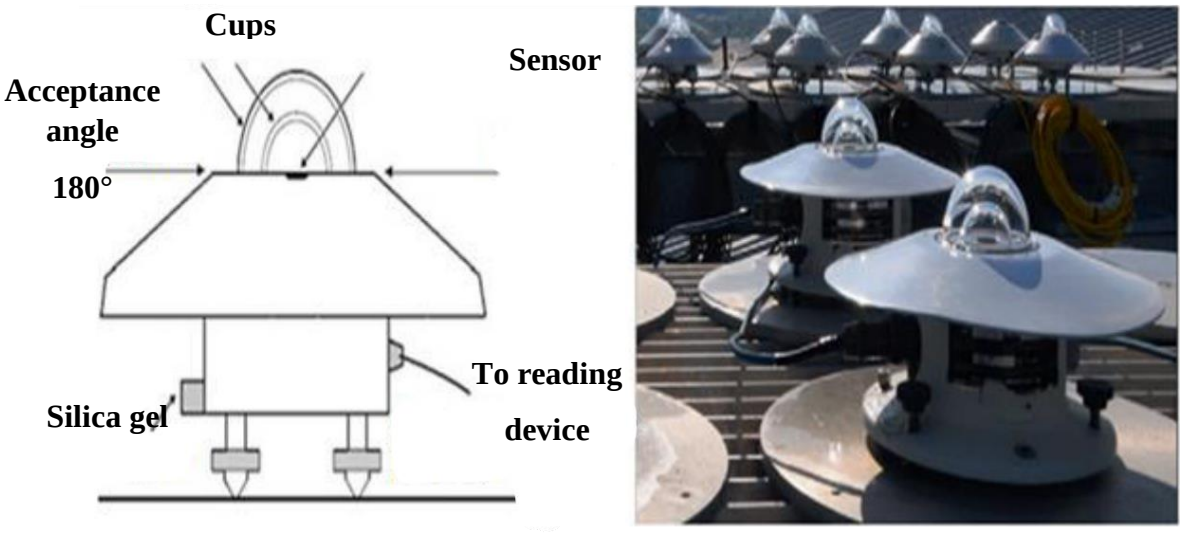


Figure II.3: Diagram of a pyranometer , and Eppley-type pyranometer (on the right) [2]

II.2.4. Ground Albedo

The albedo for solar radiation is defined as the ratio of the radiation diffused or reflected by the ground to the global radiation. It is measured by an Albedometer (Figure II.4): two pyranometers mounted in opposition, one facing the ground and the other facing the sky.



Figure II.4 : l‟‟Albédometer

II.2.5. Sunshine Duration

Sunshine duration is most commonly measured using a Campbell-Stokes heliograph, which employs a glass optical lens in the shape of a crystal sphere to focus the sun's rays onto a paper that burns as it moves (Figure II.5). Thus, only moments when the sun is clearly visible are recorded; this is referred to as actual or effective sunshine duration and depends on whether the rising sun is visible from the observation point or obscured by clouds.

This optical instrument records the duration during which solar radiation is of sufficient intensity to produce distinct shadows, aligning with the definition of this climatic indicator. Various models of these instruments exist. Sunshine duration corresponds to the number of hours in a day, between sunrise and sunset, during which the sun is clearly visible.

In the absence of a heliograph, it is possible to estimate the theoretical duration of the day by calculating the relative astronomical movement of the sun and the Earth; that is, the duration there would be if clouds did not obscure the sun. This duration is calculated based on the latitude of the site and the apparent declination, which itself depends on the time of the year considered [3].



Figure II.5: (a) Campbell-Stokes Heliograph; (b) Fiber Optic Heliograph [2, 4]

Several factors, including the following, influence the measured value's uncertainty:

- Resolution (the smallest change in radiation levels that the device is capable of detecting),
- The response's non-linearity (sensitivity variations correlated with incident radiation intensity),
- Directional response deviation (response in azimuth and cosine),
- The instrument's time constant, or the amount of time needed to get to 95% of the final value,
- Variations in sensitivity brought on by changes in weather-related factors (such as wind, pressure, humidity, and temperature),
- The ratio of the electrical output signal to the applied radiation exposure, or the long-term variation of sensitivity. [4].

A well-characterized instrument should be aware of all these uncertainties. But some instruments are more effective than others under specific climate and radiation conditions; therefore, instruments should be chosen based on their intended application.

II.2.6.Cloud Cover

Cloud cover, or cloudiness, is the fraction of the sky obscured at a specific location and time by all or part of the clouds. It is a measurable quantity typically assessed by the observer in octas: 0 corresponds to a clear sky, and 8 indicates a completely overcast sky. This parameter is included in studies of solar radiation as it represents the primary factor in its attenuation. It is provided every hour or every three hours in most professional meteorological stations of the National Meteorological Office (ONM) [5].

II.3. Algerian Meteorological Network

The National Meteorological Office (O.N.M.) uses its network of eighty-one (81) meteorological stations to record sunshine duration in order to estimate solar radiation in Algeria. Between 1970 and 1989, only seven of these sites measured global and diffuse elements received on the horizontal plane. Additionally, these measurement series include missing data due to varying periods of non-functionality of the measuring instruments, with the geographical coordinates of the meteorological stations provided. Information collected in 2011 indicated that only three (03) stations, namely Oran, Tamanrasset, and Ksar Chellala,

regularly measured solar radiation on the horizontal plane. Eight (08) other automatic stations have been measuring radiation since 2009, namely: Algiers (airport), Tlemcen,Oran (Sénia), Ghardaïa, In Amenas, Annaba, Constantine, and Tamanrasset [6].

The geographical distribution of all stations equipped with pyranometers is shown in Figure II.6.

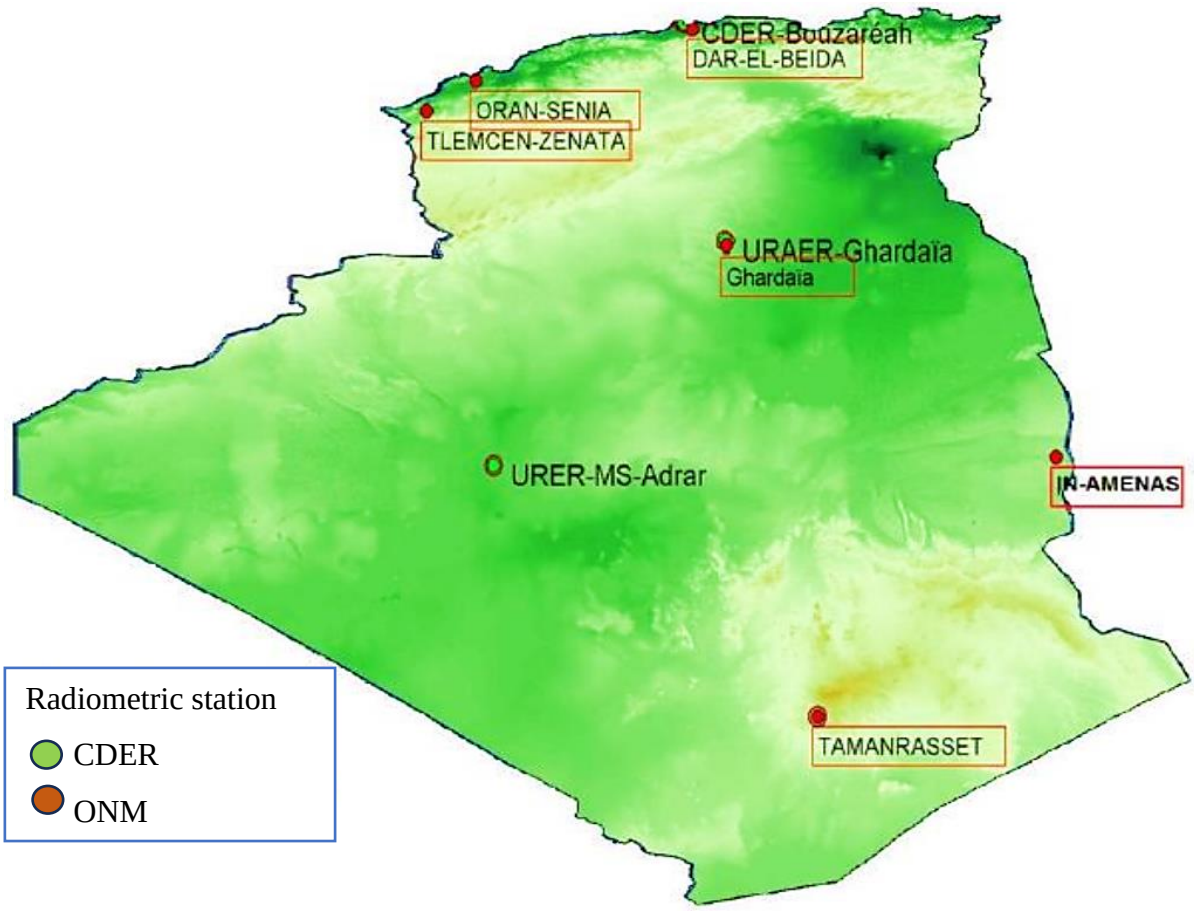


Figure II.6: Pyranometric Network (ONM and CDER)

II.4. Remote Sensing Measurements

Remote sensing is the method that makes it possible to obtain information about the surface of the planet without making physical contact by taking pictures [7]. The complete process of detecting, registering, and processing electromagnetic radiation energy that is reflected or emitted, as well as its subsequent processing, analysis, and application, is referred to as remote sensing.

Targets and incident energy interact in the majority of remote sensing scenarios. The seven steps listed below are involved in remote sensing using image technology. It is important to remember, nevertheless, that non-imaging sensors and released energy can also be used in distant sensing.Remote sensing goes through seven steps:

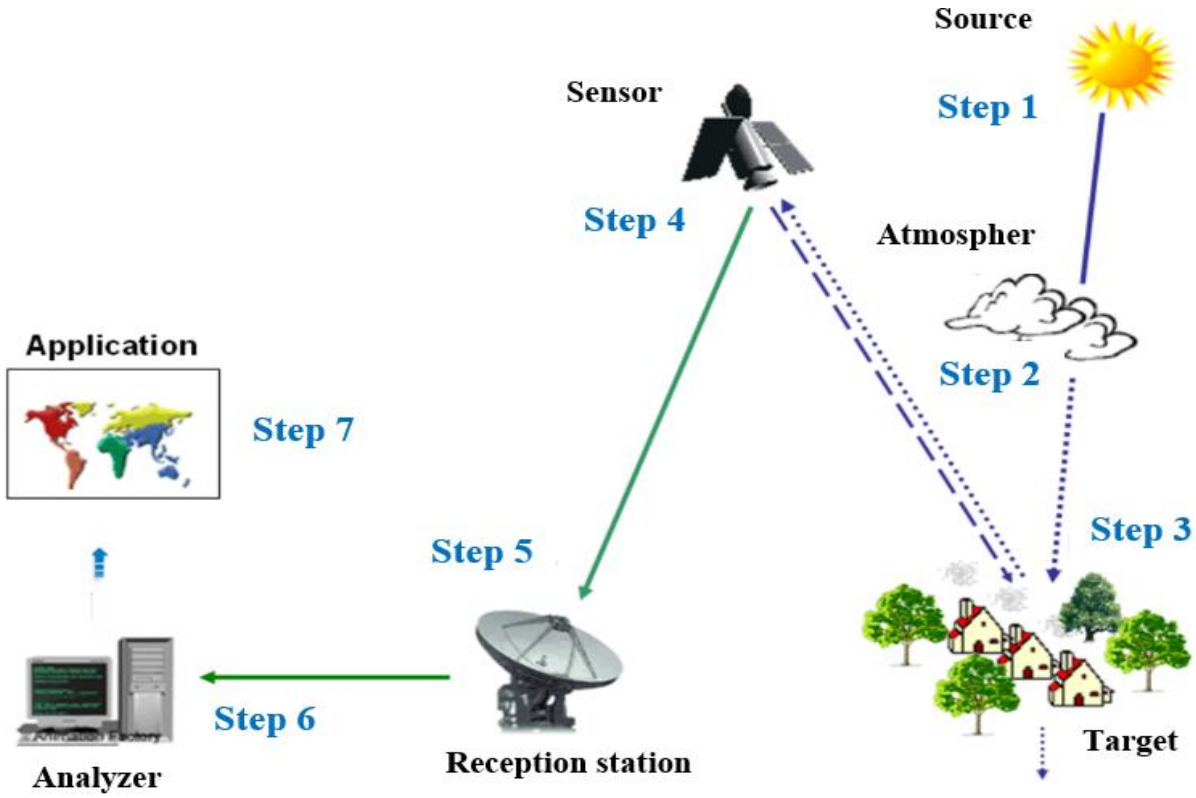


Figure II.7: Remote Sensing Process.

Table II.1: Remote Sensing Process steps.

Step	Description
1	**Energy or Illumination Source:** At the origin of any remote sensing operation, there is inevitably an energy source to illuminate the target. Most often, the energy source is the sun. However, the satellite itself can be an energy source; this is the case for radar remote sensing.
2	**Radiation and Atmosphere:** During its "outbound" journey between the energy source and the target, radiation interacts with the atmosphere. A second interaction occurs during the "return" journey between the target and the sensor.
3	**Interaction with the Target:** Once reaching the target, the energy interacts with its surface. This interaction depends on the characteristics of the radiation and the properties of the surface. Each geographic object emits or reflects radiation in various frequencies of the electromagnetic spectrum. This characteristic is called spectral behavior.
4	**Energy Archiving by the Sensor:** Once the energy scattered or emitted by the target, it needs to be remotely captured by a sensor not in direct contact with the target but embedded on a satellite or aircraft, for example, and finally recorded in digital format.
5	**Transmission, Reception, and Processing:** This information recorded by the sensor is transmitted, often through electronic means, to a reception station generally located on the ground where the information is transformed into images (digital or photographic).
6	**Interpretation and Analysis:** A visual and/or digital interpretation of the processed image is essential to extract the desired information about the target.
7	**Application:** The final step of the process involves using the extracted information from the image to better understand the target, i.e., the studied portion of space (Foggaras, AEP network, flooded area, a forest, etc.), to help solve a specific problem.

II.4.1. Digital Image in Remote Sensing and its Characteristics

It is a planar representation of a terrestrial landscape. At every point (x, y) on the image plane, one can assign a value G(x, y) representing the radiometry of the corresponding ground point. A remote sensing image is a two-dimensional geometric matrix.

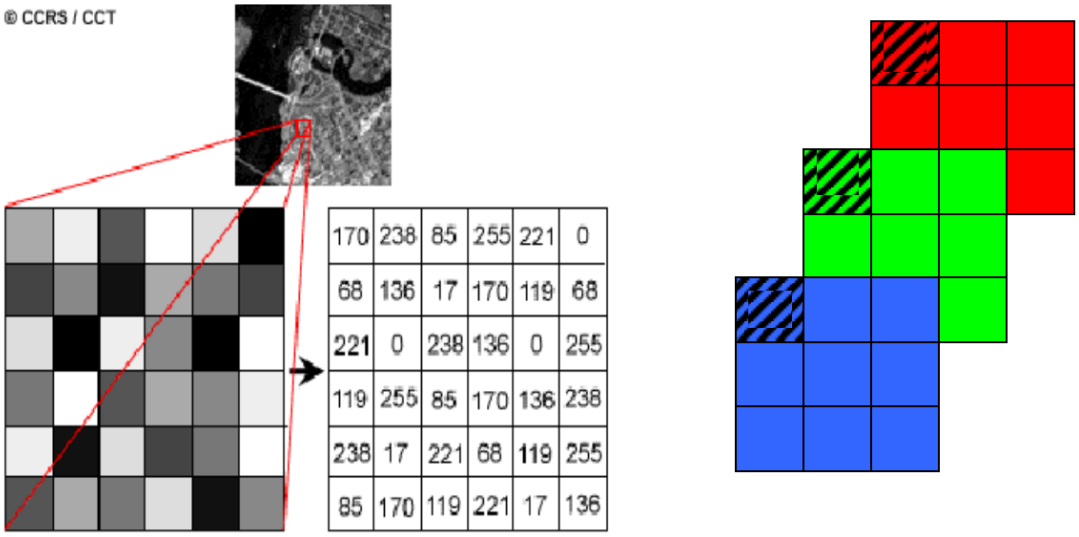


Figure II.8: Digital Image

The pixel is the smallest perceivable area by a sensor. At any given pixel, 3 values are associated, allowing these values to be in digital count (CN). Each digital count is encoded in 8 bits, and the grayscale levels are stored between 0 and 255.

II.4.2. Platforms and Sensors

In this section, we'll look more closely at how a sensor recorded this energy by analyzing the features of the remote sensing platform, the sensors, and the data that was collected. Ground sensors are frequently employed to capture comprehensive surface data. After that, this data is contrasted with information gathered from satellites or aircraft. Ground sensors are frequently mounted on tall buildings, cranes, scaffolds, scales, etc.

Remote sensing is sometimes conducted in space from satellites, but it is also occasionally conducted from the space shuttle. Space platforms are able to cover the Earth's surface repeatedly and continuously because of their orbit. Cost is frequently a deciding issue when selecting between several platforms.

II.4.3. Characteristics of Satellites - Orbit

Although ground-based or airborne platforms are used, satellites provide the majority of remote sensing data today. An orbit is the route a satellite takes around the Earth. The mission objectives and the capability of the satellite's sensors are taken into consideration when determining the satellite's orbit. The altitude (the height of the satellite above the surface of the Earth), direction, and rotation of the satellite with respect to the Earth all influence the choice of orbit. Based on these factors, two types of satellites can be defined [7]:

A. Geostationary Orbit Satellites

These are extremely high-altitude satellites that continuously monitor the same area of the planet. From their height of about 36,000 kilometers, geostationary satellites appear to be stationary because they travel at the same speed as the Earth. The satellite can continuously monitor and collect data about a particular area because to its orbital structure. Satellites for weather monitoring and communication are situated in these orbits. Certain weather satellites can monitor conditions and clouds throughout the whole hemisphere of the planet because of their great altitude.

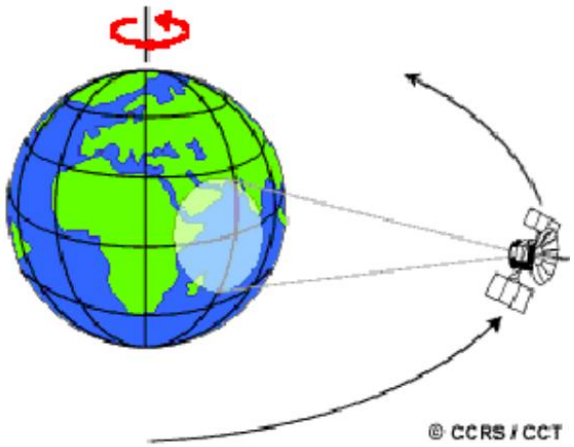


Figure II.9: Geostationary Satellites in Orbit

B. Heliosynchronous Orbit Satellites

These are satellites that cover the entire planet at the same local solar time every single time. The Sun's location in the sky at a certain latitude during a specific season when the satellite travels over that area will never change.

These space platforms follow an orbit practically from north to south or vice versa. Because of this arrangement and the Earth's west-east rotation, the satellites have observed nearly the whole surface of the planet during a predetermined amount of time. The orbit's inclination with respect to a line that passes across the Earth's North and South Poles gives rise to the term "quasi-polar orbit."

Similar sun illumination conditions are ensured by this orbital characteristic while gathering data for a given season over several years or for a single region over multiple days. This is a crucial consideration when comparing two consecutive photographs or creating a mosaic with nearby images since it eliminates the need to adjust the images for solar illumination.

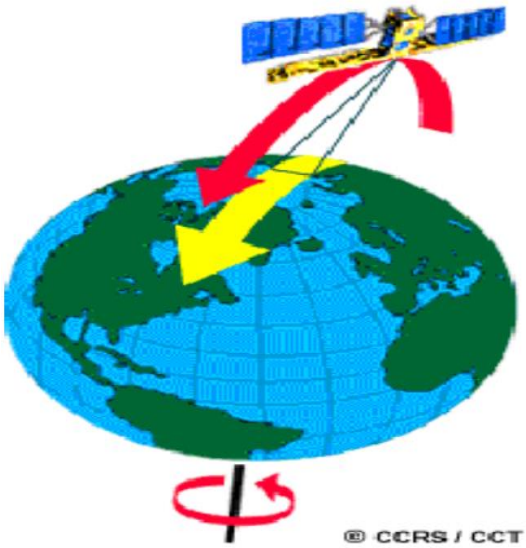


Figure II.10: Heliosynchronous Orbit Satellites.

II.4.4. Digital Image Processing

Digital image processing necessarily requires a computer system (or image analysis system) as well as the equipment as well as tools for data processing. A number of commercial software programs have been created expressly to process and analyze of remote sensing images. We can group commonly available image processing functions into four groups:

- Image enhancement
- Preprocessing

- Image classification and analysis
- Image transformation

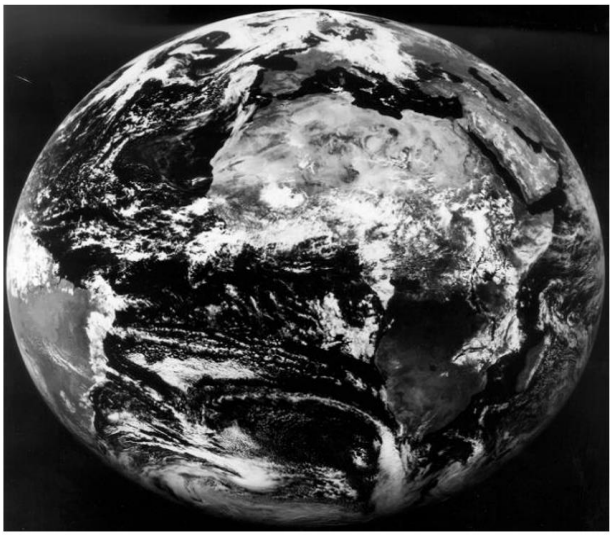


Figure II.11: METEOSAT Satellite Map

II.5. EUMETSAT

EUMETSAT is a European intergovernmental organization [8] active in the space domain, responsible for establishing and operating operational meteorological satellite systems.

EUMETSAT currently has more than 20 European member states, including Germany, Austria, Belgium, Denmark, Spain, Finland, France, etc. The organization has also signed several privileged cooperation agreements with around ten other countries, such as Bulgaria, Croatia, Hungary, Latvia, etc. These countries have access to data products 24 hours a day, every day of the year, notably via EUMETCast, the digital data dissemination system.

II.6. History of METEOSAT Satellites

The METEOSAT group of European satellites is part of operational geostationary satellites, providing observation of the entire Earth, excluding the poles. METEOSAT 1 was launched in 1977 for an experimental task but achieved such success that the service became operational with the launch of METEOSAT 2 in 1981. Four additional satellites were launched in 1988, 1989, 1991, and 1993. A seventh satellite of the same design was launched in 1997 under the METEOSAT program.

II.7. Second Generation METEOSAT (MSG)

The first satellite in the series, named Second Generation METEOSAT (MSG), was launched on August 28, 2002 [9], after a development phase conducted by the European Space Agency for the European Organization for the Exploitation of Meteorological Satellites (EUMETSAT). The commissioning test phase that followed focused not only on the new satellite itself but also on a completely new ground segment developed by EUMETSAT. This segment is responsible for commanding and controlling the satellite and processing its data at the organization's headquarters and in centers distributed throughout Europe. During this commissioning period, regular operation of first-generation satellites continued, but users across Europe were able to receive initial data from the new satellite starting from April 2003.



Figure II.12: Second Generation METEOSAT Satellite.

On January 29, 2004, the satellite, renamed METEOSAT-8, was declared fully operational, meaning it was operating at full capacity, and its data could be regularly accessed. This data could thus be used operationally across the satellite's coverage area, including not only Europe but also Africa and the Middle East.

On December 21, 2005, EUMETSAT launched the second MSG-2 satellite, and on July 6, 2006, it moved to 0° longitude, becoming fully operational. It was renamed METEOSAT-9 and replaced the reserve satellite METEOSAT-8.

The main task of these satellites [10] is the continuous observation of the Earth's surface and atmosphere in multiple spectral bands, called channels. They also perform other activities such as:

- Analysis of air masses.
- High-resolution imaging.
- Data collection.
- Transmission and dissemination of meteorological data.
- Operation of a complete data archiving and consultation system.

The main instrument on MSG satellites is a SEVIRI imaging radiometer that scans the Earth's surface to create a new multispectral image in 12 channels every fifteen minutes. There are two visible channels, nine infrared channels, and a twelfth visible HRV (High-Resolution Visible) channel covering only a part of the Earth's disk. The spatial resolution also varies from 2.5 km to 1 km at nadir for the wideband visible channel and from 5 km to 3 km for all other channels. The specifics of the 12 MSG channels and their purposes are summarized in Table II.2.

Table II.2. Characteristics and main objectives of the 12 image acquisition channels of SEVIRI.

Canal	Other Appellation	Central Wavelength (micrometers)	Main Objective
C1	VIS 0.6	0.635	Surface, clouds, wind fields
C2	VIS 0.8	0.810	Surface, clouds, wind fields
C3	NIR 1.6	1.640	Surface, cloud phase
C4	IR 3.9	3.900	Surface, clouds, wind fields
C5	WV 6.2	6.250	Water vapor, high-altitude clouds, atmospheric instability
C6	WV 7.3	7.350	Water vapor, atmospheric instability
C7	IR 8.7	8.700	Surface, clouds, atmospheric instability
C8	IR 9.7	9.660	Ozone
C9	IR 10.8	10.800	Surface, clouds, wind fields, atmospheric instability
C10	IR 12.0	12.000	Surface, clouds, atmospheric instability
C11	IR 13.4	13.400	Cirrus cloud height, atmospheric instability
Large Band C12	HRV	0.4-1.1	Surface, clouds

This increase in MSG's capabilities compared to METEOSAT addresses the evolving needs of immediate forecasting and short-term meteorological risk management, thanks to its ability to characterize active weather systems, their rapid evolution, and their probability of development. Figure II.13 shows three images taken in the VIS 0.6 channel (C1) of the SEVIRI instrument at different times of the day. This channel is sensitive to solar radiation reflected in the visible part of the spectrum, within a narrow spectral window centered on 0.6 µm. Variations in solar illumination during the day are highlighted by the terminator curve between the illuminated and dark parts, visible at 6 a.m. and 6 p.m. UTC. Four channels of SEVIRI, namely VIS 0.6 (C1), VIS 0.8 (C2), NIR 1.6 (C3), and HRV (C12), are primarily sensitive to solar radiation reflected at 0.6 µm, in a similar manner.

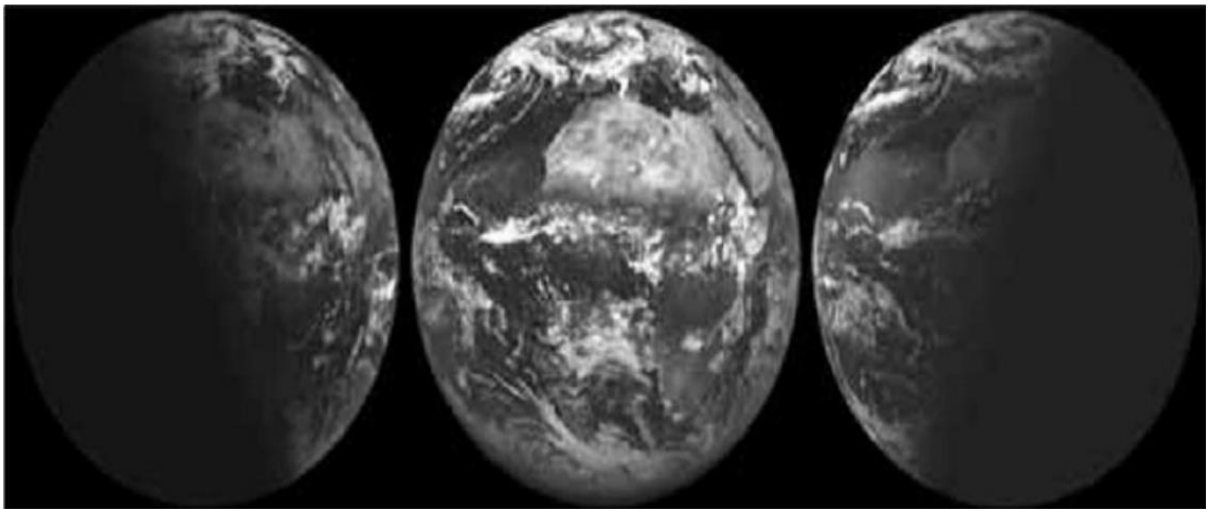


Figure II.13: Sequence of three images captured by SEVIRI in the VIS 0.6 channel (C1) on June 19, 2004, at 06:00, 12:00, and 18:00 UTC.

The electromagnetic radiation perceived by a satellite's sensor passes through the atmosphere during solar illumination (between the sun and the ground) and between the ground and the satellite (radiation reflected by the Earth's surface). This dual passage disrupts solar radiation, causing the amount of solar radiation to vary significantly in space and time, driven by differences in geographical location and cloud cover and composition.

Due to the high cost of investing in and maintaining radiometric measurement stations and the advantages offered by space remote sensing with global coverage, high spatial resolution, and frequent temporal sampling that no other observation system can provide, techniques for estimating solar radiation components based on satellite data are developed to assess solar potential. [11]

C. databases on the web

Freely accessible databases on the web, such as SoDa and NASA-SSE, collect solar radiation data using a combination of satellite observations, atmospheric models, and ground-based measurements. They integrate information from geostationary and polar-orbiting satellites, atmospheric models, and ground-based stations to create comprehensive databases with global coverage. These datasets are valuable for renewable energy applications, climate studies, and scientific research.

- **The SoDa database** is one of the most significant on the web, providing data and services related to ground-received solar radiation. It has been freely accessible since early 2011 and includes 21 years of daily data (1985 to 2005) covering Europe, Africa, and the Atlantic Ocean. These data, with a spatial resolution of 30 km, complement the ground-based measurements from existing meteorological stations, which are often sparse. [12]

- **The NASA-SSE** (Surface Meteorological and Solar Energy) database is a number of databases created by NASA, such as the Goddard Earth Observing Systems (GEOS-1), which used an atmospheric model that was restricted to observations from satellites and radiosondes. The U.S. National Oceanic and Atmospheric Administration's Geostationary Operational Environmental Satellite (GOES) and Polar-orbiting Operational Environmental Satellite (POES), the European Space Agency's Meteorological Satellites (Meteosat), and the Japan Meteorological Agency's geostationary satellites were among the satellites whose observations were analyzed to create these datasets. The NASA-SSE database contains monthly satellite data for a 1°x1° grid (111 km) covering the entire world for a period of 10 years (1983-1993). [13]

II.8. Data Preprocessing

Questioning the data that will be used is the next step. The quantity, organization, and format of the input data must be precisely tailored to each task. Unfortunately, undesirable elements like noise, multiple viewpoints, redundant and inconsistent data, and enormous sizes in terms of features, examples, and dimensions all have a significant impact on real-world databases. As a result, poor data quality will result in poor performance [14].

We shall list the broad categories into which we might subdivide the collection of methods for preprocessing so that readers have a brief summary of the approaches they should be acquainted with after reading this. In order to accomplish this chapter, several subsections will be offered according to the type and variety of techniques that are included under each heading..

II.8.1. Data Preparation

The collection of methods used to correctly initialize data so that it can be used as input for a particular data mining (DM) algorithm is referred to as data preparation. It is vital to emphasize that we choose to utilize the data preparation notation to design specific elements of data preprocessing, even though earlier works have referred to the full set of processes executing data preprocessing tasks using a confusing nomenclature. This is not inaccurate, and we value the terminology, but considering the growing importance of the latter group of methods in the last few years and the clear differences that result from this comprehension, we would prefer to make a clear distinction between data reduction and preparation. Preparing the data is typically a necessary step. It creates new data which fits a DM process from previously useless data.

First of all, improperly prepared data may prevent the DM algorithm from receiving information needed to function or ensure that it reports errors while it is running. In the best of circumstances, the algorithm will produce results, but they won’t make sense or be regarded as reliable information. So what are the fundamental problems in data preparation that need to be fixed? Here, we offer a set of questions covering every kind of procedure that is a member of the data preparation family of techniques, along with the appropriate responses:

• How should the data be cleaned up?— Cleaning the Data.

• How can I deliver precise data?—Transformation of Data.

• How can I add and modify data? —The integration of data.

• How can I scale and unify data? —The normalization of data.

• What should I do about missing data?—Imputation of Missing Data.

• How can I recognize and control noise?— Identification of Noise.

We will next go over each of the aforementioned techniques in brief. An explanation of the different types of data preparation is provided in the figure II.14 below.

II.8.1.1.Data Cleaning

Alternatively known as data cleansing, this process involves reducing unnecessary detail in the data and correcting inaccurate or incomplete data. It is a broad idea that includes or crosses over with other well-known methods of data preparation.

Although missing and noisy data are treated here, they have been divided into two categories so that a more thorough examination of the insightful suggestions made for them can be found later in this work. Additional tasks associated with data cleansing include identifying discrepancies and unclean data—fragments of the original data that don’t make sense.

The latter tasks typically call for human auditing and are more closely tied to the comprehension of the original data.

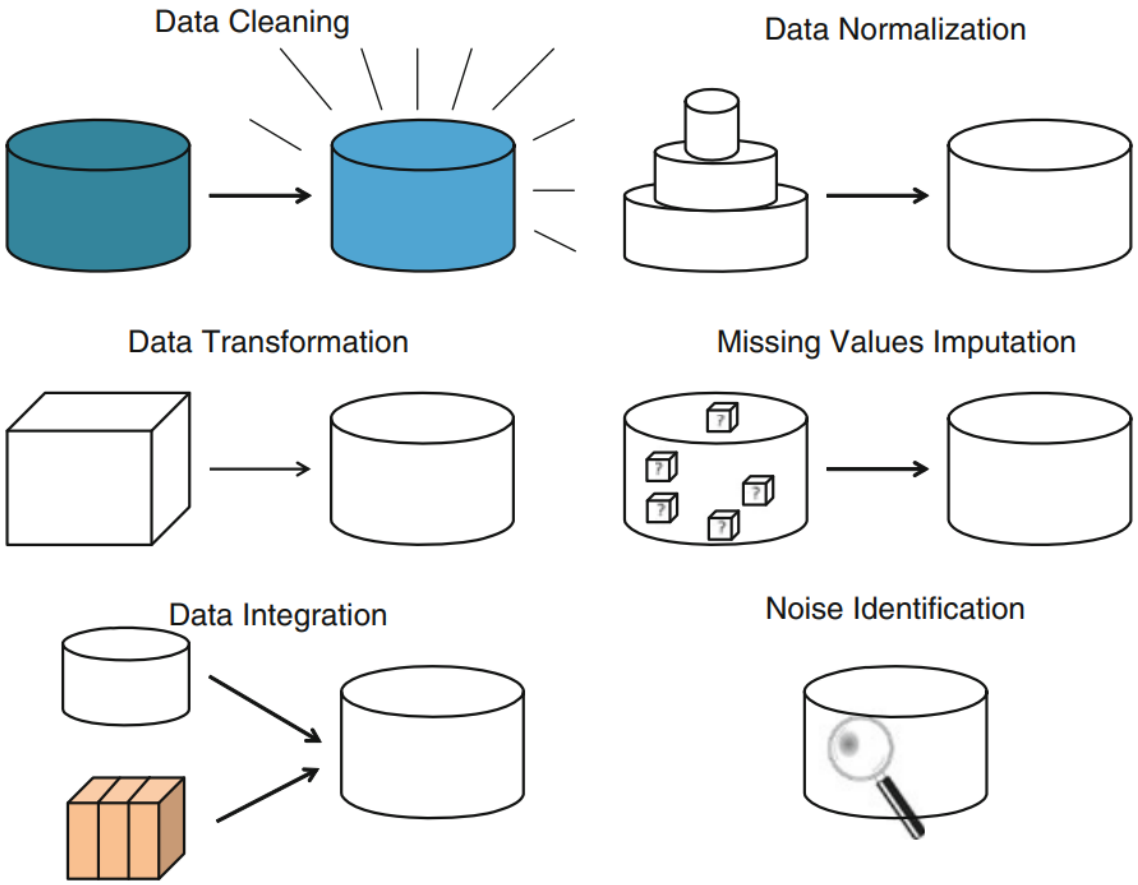


Figure II.14 Data preparation Forms.

II.8.1.2. Transformating of the Data

This preprocessing step's goal is to transform or combine the data in order to maximize the use of the mining process's results. Subtasks within data transformation include smoothing, feature building, normalization, discretization, generalization, and data aggregation or summarization. The bulk of data transformation procedures, including data cleaning, will be broken out into separate tasks since they are referred to as a general data preprocessing family of techniques.

The tasks that require human supervision and are more dependent on the data are the traditional data transformation techniques, such as report generation, new attributes that aggregate existing ones, and generalization of concepts, especially in categorical attributes, like replacing complete dates in the database with year numbers only.

II.8.1.3. Data Integration

It entails merging information from many data sources. This step needs to be done carefully to avoid redundancy and inconsistencies in the final data set. Typical tasks accomplished during data integration include the identification and unification of variables and domains, attribute correlation analysis, tuple duplication, and conflict detection between data values from several sources.

II.8.1.4. Data Normalization

The data analysis may be impacted by the measurement unit chosen. Every attribute ought to have a common scale or range and be expressed in the same measurement units. Normalization is a useful technique in statistical learning methods that aims to assign equal weight to all attributes in the data.

II.8.1.5. Missing Data Imputation

Filling the variables that contain MVs with some intuitive data is the aim of this type of data cleaning. Generally speaking, it is preferable to add a reasonable estimate of a suitable data value rather than leave it blank. [15]

II.8.1.6. Noise Identification

Detecting random errors or variances in a measured variable is the primary goal of this data transformation step, which is also referred to as smoothing. It should be noted that we speak of noise detection rather than noise elimination, which is more closely associated with the IS task in data reduction. After identifying a noisy example, we can implement a correction-based procedure that might entail an underlying operation of some kind. [16]

II.8.2. Data Reduction

The collection of methods that, in one way or another, produce a condensed representation of the original data is known as data reduction. In our opinion, the techniques for data preparation that are required to suitably prepare the data as input for a DM task are what set them apart. As previously stated, this implies that improper data preparation will either prevent the DM algorithms from running or ensure that they produce incorrect results when

they do. When data reduction occurs, the amount of data is reduced but the generated data typically retains the fundamental structure and integrity of the original data.

Can the original data be used as the input of a DM process without undergoing a data reduction process? Yes, but there are other important considerations that need to be made as well; these considerations are just as important as the issues that data preparation addresses. Therefore, it appears to be an optional step at first glance.

But there could be a conflict with this affirmation. Every algorithm has a time complexity that depends on multiple parameters, even though the data integrity is preserved. There exists a direct proportionality between one of these DM parameters and the input database size. The data reduction task becomes equally important as data preparation if the size surpasses the limit, which varies depending on the type of DM algorithms. In such cases, running the algorithm may become prohibitively expensive. As for other factors, like the reduction of complexity and enhancement of model quality, data reduction plays a crucial part once more.

As previously stated, what are the fundamental problems with data reduction that need to be fixed? Again, we offer a set of questions along with the right response for every kind of task that falls under the data reduction techniques:

- How can I make the data less dimensional? —Selection of Features (FS).
- How can I get rid of examples that are contradictory or redundant?—Selection of instances (IS).
- How can I make an attribute’s domain simpler?—Detailization.
- How can I complete data gaps?—Feature extraction and/or generation of instances.

We give a brief description of each of the four previously listed techniques in the section that follows. An example image that demonstrates the various types of data reduction is presented in Figure II.15.

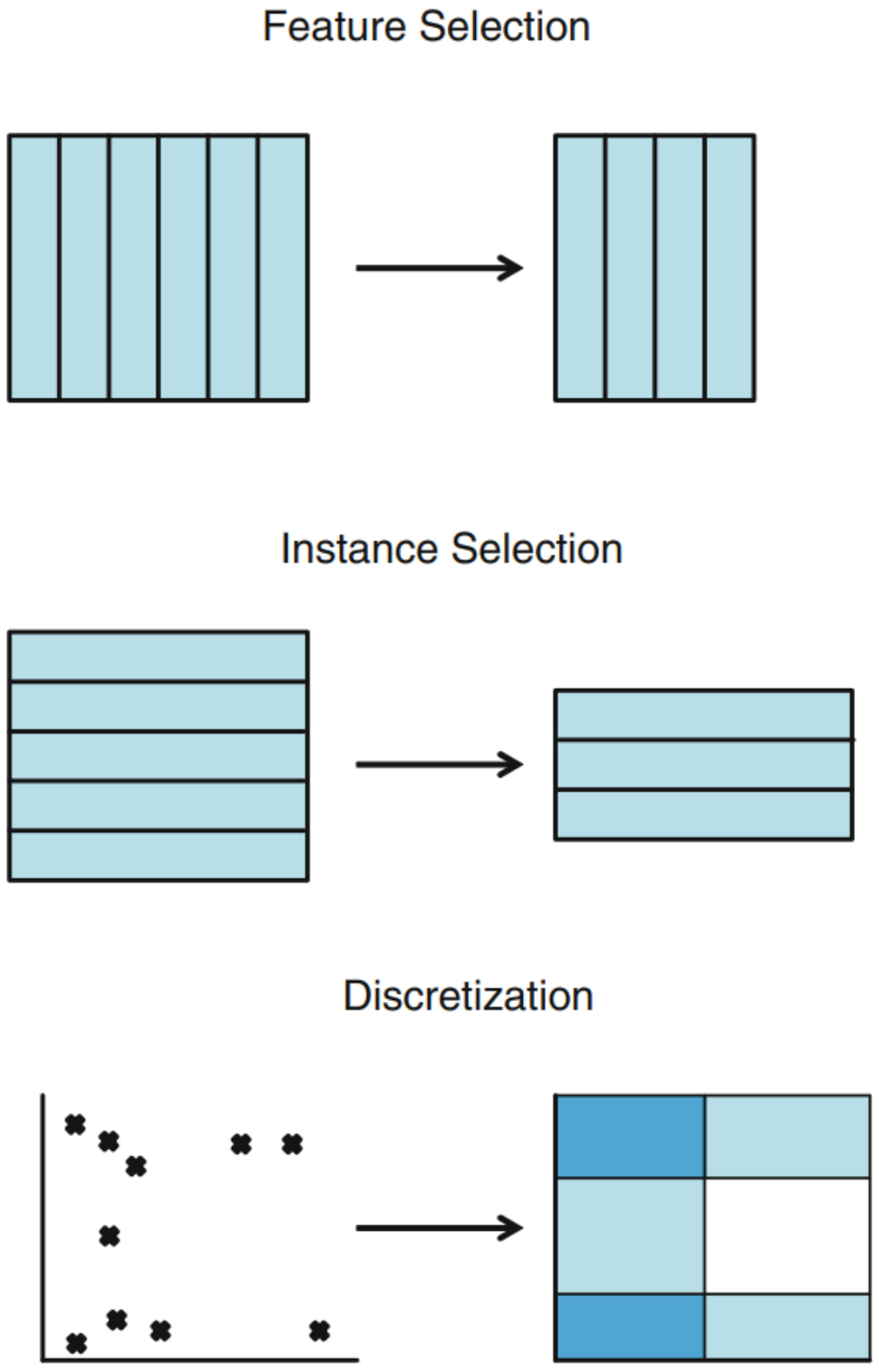


Figure II.15 : Data reduction forms

II.8.2.1. Feature Selection

Reduces the size of the data set by eliminating features (or dimensions) that are irrelevant or unnecessary. Finding the minimum set of attributes that is as close to the original distribution obtained using all attributes is the aim of fuzzy set theory (FS).

An example of this would be the probability distribution of the data output attributes (or classes). It speeds up the learning process and makes it easier to understand the extracted pattern. [17, 18]

II.8.2.2. Instance Selection

Consists of selecting a part of the entire data set to fulfill the initial goal of the DM application as though all of the data had been used. It makes up the family of oriented methods that, by applying heuristics and/or rules, select the optimal subset of examples from the original data in a fairly intelligent manner.

In many DM models, sampling—a term used to describe the random selection of examples—is used to conduct internal validation and prevent over-fitting. [19, 20]

II.8.2.3. Discretization

Through this process, a non-overlapping partition of a continuous domain is obtained by converting numerical attributes into discrete or nominal attributes with a finite number of intervals. In other words, quantitative data is transformed into qualitative data.

The next step is to establish an association between each interval and a discrete numerical value. The data can be used as nominal data in any DM process after discretization has been performed.

Discretization is essentially a hybrid data preprocessing technique that involves tasks related to both data reduction and data preparation, which is particularly noteworthy. A data reduction process is taken into consideration by some sources, while discretization is included in the data transformation category by others. [21]

Because it maps data from a wide range of numeric values to a significantly smaller subset of discrete values, discretization can be thought of as a data reduction technique in practice. While acknowledging the other trend, we have opted to primarily incorporate it into data

reduction. The reasoning behind this is that modern discretization schemes aim to minimize the number of discrete intervals while preserving the further DM process's performance.

Stated differently, basic discretization can be carried out with ease using any kind of data as long as it can be mapped into a simple map between continuous and categorical values using a particular algorithm. The actual challenge, though, is to achieve good reduction without sacrificing the quality of the data, and this is where most of the effort made by researchers is directed.

II.8.2.4. Feature Extraction/Instance Generation

By permitting the alteration of the internal values that represent each example or attribute, it extends the feature as well as IS. In addition to removing attributes, subsets of attributes can be combined or help create artificial substitute attributes during the feature extraction process. In terms of instances, the procedure for creating instances is comparable. It permits the construction or modification of synthetic replacement examples that might more accurately depict the decision boundaries in supervised learning. [22, 20, 23]

II.9. Conclusion

The inconsistent and unpredictable characteristics of solar energy, along with its phase shift in energy needs, highlight how crucial it is to fully understand solar potential before putting in solar capture devices. Ground radiation measurements are the most effective means of assessing solar potential, which emphasizes the necessity of precise radiation data gathering, processing, and analysis. Accurate solar resource estimation is necessary for effective energy planning.

This chapter highlighted the value of radiometric measurements despite Algeria's limited measurement station network and thus insufficient territory coverage.

Therefore, it is essential to optimize the size of proposed installations and model the operation of hypothetical solar energy systems. It's also critical to remember that data preparation is an essential step in this procedure since it guarantees that the collected data is appropriately cleansed, converted, and arranged to yield precise and informative outcomes. This increases the precision and effectiveness of solar energy evaluations and planning projects.

.

II.10. References

1.Bouchouicha K., Oulimar B., (2013), La chaine de mesure radiométrique à l‟Unité de Recherche en Energie Renouvelable en Milieu Saharien d‟Adrar, International Conférence on Energy and Sustainable Development ICESD‟13, 19-20 Février 2013, Adrar, Algeria.

2.OMM. (2008) Guide to meteorological instruments and methods of observation, 7éme edition.

3.McArthur L., (1998), Baseline Surface Radiation Network (BSRN) Operations Manual (Version 1.0), WMO/TD No. 879. Geneva, Switzerland: World Climate Research Program, WMO.

4.Paulescu M, Paulescu E, Gravila P, Badescu V. (2013) Weather modeling and forecasting of PV systems operation. Green Energy and Technology, Edition Springer.

5.Bouchouicha K., (2017), Modélisation multi spectrale des images satellitaire - Application : Quantification du bilan d‟énergie Sol-Atmosphère, thèse Doctorat en science, université U.S.T.O-MB.

6.Faiza M., (2008), Détermination du gisement solaire par traitement d‟images MSG , Thèse Magister, Université Tiziouzou.

7.Notions fondamentales de télédétection, Un cours tutoriel du Centre canadien de télédétection.

8.Gautier C., Diak G., and Masse S., (1980), A simple physical Model to estimate incident solar radiation at the surface from GOES satellite data. J. Appl. Meteor., vol.19, pp. 1005–1012.

9.Site EUMETSAT : www.eumetsat.int, Consulté en Février 2023.

10.Site Internet:http://sinfotech-sig.teledetection.fr/site/index.php,Consulté en Mars 2023.

11.Benatiallah D. (2019), Détermination du gisement solaire par imagerie satellitaire avec intégration dans un système d'information géographique pour le sud d'Algérie. Ph.D. diss., University of Adrar.

12.Younes S, Claywell R, Muneer T. (2005) Quality control of solar radiation data: Present status and proposed new approaches. Energy, 30, p 1533-1549.

13.Journée M, Bertrand Cedric. (2011) Quality control of solar radiation data within the RMIB solar measurements network. Solar Energy, 85, p 72-86.

14.Pyle, D. (1999) Data Preparation for Data Mining. Morgan Kaufmann, San Francisco.

15.Luengo, J., García, S., Herrera, F. (2012) On the choice of the best imputation methods for missing values considering three groups of classification methods. Knowl. Inf. Syst. 32(1), 77–108.

16.Sáez, J.A., Luengo, J., Herrera, F. (2013) Predicting noise filtering efficacy with data complexity measures for nearest neighbor classification. Pattern Recogn. 46(1), 355–364.

17.Liu, H., Motoda, H. (1998) Feature Selection for Knowledge Discovery and Data Mining. Kluwer International Series in Engineering and Computer Science. Kluwer Academic, Boston.

18.Liu, H.,Motoda, H. (2007) ComputationalMethods of Feature Selection. DataMining and Knowledge Discovery Series. Chapman & Hall/CRC, Taylor and Francis, Boca Raton.

19.García, S., Derrac, J., Cano, J.R., Herrera, F. (2012) Prototype selection for nearest neighbor classification: Taxonomy and empirical study. IEEE Trans. Pattern Anal. Mach. Intell. 34(3), 417–435.

20.Liu, H., Motoda, H. (2001) Instance Selection and Construction for Data Mining. Kluwer Academic, Norwell.

21.García, S., Luengo, J., Sáez, J.A., López, V., Herrera, F. (2013) A survey of discretization techniques: taxonomy and empirical analysis in supervised learning. IEEE Trans. Knowl. Data Eng. 25(4), 734–750.

22.Liu, H., Motoda, H. (1998) Feature Extraction, Construction and Selection: A Data Mining Perspective. Kluwer International Series in Engineering and Computer Science. Kluwer Academic, Boston.

23.Triguero, I., Derrac, J., García, S., Herrera, F. (2012) A taxonomy and experimental study on prototype generation for nearest neighbor classification. IEEE Trans. Syst. Man Cybern. Part C 42(1), 86–100.

Chapiter III : Prediction Models

III.1. Introduction

PV systems are contributing more and more to the generation of power. The energy transition will cause this increase to accelerate. However, because solar resources are intermittent, incorporating significant amounts of power into the electrical system presents technical difficulties. PV systems' ability to alter power generation to match expected load patterns sets them apart from conventional systems. If solar energy variation can be forecast with great precision, then this integration can actually provide higher quality services. Grid operators can better account for variable electricity output in their scheduling, dispatching, and power regulation by forecasting PV production.

The efficient functioning and optimization of photovoltaic (PV) energy systems depend on the forecast of global radiation received on a horizontal plane (Gh). Numerous techniques, from sophisticated machine learning algorithms to empirical models, have been devised to forecast solar radiation. These techniques are essential to PV energy forecasting systems because they provide guidance on resource allocation and energy production decisions.

The approach type and the input data employed, which also define the forecast horizon, can be used to classify forecasting approaches broadly. Empirical models, which forecast solar radiation using observed data and empirical relationships, are notable among these techniques. To estimate solar radiation levels, these models frequently take into account variables including location, time of day, and atmospheric conditions.

Deterministic and stochastic models are frequently used in addition to empirical models. While stochastic models use statistical techniques to reflect the stochastic aspect of solar radiation variations, deterministic models seek to directly explain the physical processes regulating solar radiation. Linear and nonlinear models are further classifications for these models. Autoregressive moving average models (ARMA), autoregressive integrated moving average models (ARIMA), and autoregressive models (AR) are examples of linear models that use linear relationships in the data. However, nonlinear models—which encompass artificial intelligence paradigms like fuzzy logic, evolutionary algorithms, and neural networks—offer more flexibility in capturing intricate correlations and patterns in the data.

The ability to learn from data and adjust to changing environmental conditions has made machine learning approaches an effective tool for forecasting solar radiation. Support vector machines, decision trees, random forests, and deep learning architectures are just a few of the many algorithms that are used in these techniques.

Machine learning models can anticipate solar radiation levels with high accuracy and reliability by utilizing satellite imagery, historical weather data, and other pertinent variables. This helps to facilitate the effective operation and administration of solar energy systems.

We go over the primary techniques and strategies for solar radiation prediction in this chapter. Additionally, we include error parameters so that these various models can be compared.

III.2. Prediction Models

III.2.1. Persistence Model

The persistence model enables the prediction of the future behavior of a time series. It assumes that the temporal variable X at time (t+1) is best predicted by its value at time (t):

$$\hat{X}_{t+1} = X_t \quad \text{(III.1)}$$

The persistence model is recognized as the benchmark for solar radiation and PV power forecasting methods, due to its simplicity and universality. This type of predictor is sometimes the only one truly usable because it does not require a historical time series; only the value at time t is needed. Persistence forecasts, also known as naive predictors, are used as comparison models to other more advanced forecasting techniques.

III.2.2. Empirical Models

III.2.2.1. Insolation-Based Models

Empirical models are models based on experimental observations; several types of these models can be distinguished (models based on insolation duration, temperature-based models, etc.).

The first model estimating global radiation is the one proposed by Ångström [1] (1924); he established a simple regression relationship between the ratio of global radiation on a horizontal surface (G_h) and the global radiation under clear sky conditions (G_c), based on the ratio between the average daily insolation duration (S_d) and the maximum insolation duration (S_0). Therefore, for a given location, we have the following relationship:

$$\frac{G_h}{G_c} = a + b\,(\frac{S_d}{S_0}) \qquad \text{(III.2)}$$

Where: a and b are linear constants of the Ångström model. Alternatively, Prescott [2] presents this equation in another more convenient form by replacing the clear-sky radiation with that of extraterrestrial radiation (G_0):

$$\frac{G_h}{G_0} = a + b\,(\frac{S_d}{S_0}) \qquad \text{(III.3)}$$

Often, this latter model is referred to as the Ångström-Prescott model. The regression constants a and b are empirically derived from ground-level measurements; they have a physical significance: constant a represents the total atmospheric transmission on completely overcast days when the ratio $\boldsymbol{\frac{S_d}{S_0}}$ = 0.

A day with a completely clear sky signifies that the ratio $\boldsymbol{\frac{S_d}{S_0}}$ = 1. In the Prescott model, the sum of the constants a and b $(a + b)$ represents the fraction of radiation received under clear sky conditions while accounting for the scattering of solar radiation due to atmospheric effects.

More details on modeling and estimating global radiation based on insolation duration models are provided in [3,4]. Quadratic models [5] and higher-order models [6,7] have also been developed to reduce the sensitivity of the linear model in cases of extremely overcast ($\frac{S_d}{S_0}\approx 0$) or completely clear sky conditions ($\frac{S_d}{S_0}= 1$) [8]:

$$\frac{G_h}{G_0} = a + b\,(\frac{S_d}{S_0}) + c\,(\frac{S_d}{S_0})^2 + d\,(\frac{S_d}{S_0})^3 \qquad \text{(III.4)}$$

There is no consensus on the advantage of a second or third-order regression model over linear Ångström models.

Table (III.1) summarizes some empirical constant values of the Ångström-Prescott model for certain sites worldwide:

Table III.1 : Constants of the Angstrom-Prescott Model for Various Locations

Location	Reference	A	B	C	D
Algeria	[9]	0.309	0.368	-	-
China	[10]	0.2223	0.6529	-	-
Egypt	[11]	0.228	0.527	-	-
India	[12]	0.2281	0.5093	-	-
Jordan	[13]	0.174	0.615	-	-
Libya	[14]	0.1	0.874	-0.255	-
Oman	[15]	0.9428	-1.202	0.9336	-
Pakistan	[16]	0.348	0.32	0.07	-
Spain	[17]	0.184	0.6792	-0.113	-
Turkey	[18]	0.2408	0.3625	0.4597	-0.3708
USA	[19]	0.81	-3.34	7.38	-4.51

III.2.2.2. Temperature-Based Models

Air temperature is a crucial parameter recorded by practically all meteorological stations worldwide; it is considered a common parameter for calculating solar radiation. Since temperature measurements are simple and robust, they are used to estimate the amount of daily solar energy with reasonable accuracy, for energy applications such as PV. Paulescu [20] advocated the use of temperature measurements for solar radiation forecasting; he proposes the following relationship:

$$\frac{I_h}{I_0} = a + b\, t_{max} + \mathrm{c}\ t_{max} + \mathrm{d}\, N + e \qquad \text{(III.5)}$$

Where: N is the cloudiness index, and $a, b, c, d,$ and e are constants whose values are available for most sites worldwide. I_h, I_0 represent the global irradiance on a horizontal plane and the extraterrestrial irradiance, respectively. Therefore, empirical models are the first step to undertake to obtain an estimation of the long-term production of energy conversion systems, and for optimal design and sizing.

III.2.3. Time Series Models

A time series (chronological) is a sequential set of data points, usually measured over successive times. It is mathematically defined as a set of vectors $x(t), t = 0, 1, 2, ..$ where t represents the time measurement. The variables $x(t)$ are treated as random variables. In a time series, measurements taken during an event are organized chronologically in the proper sequence. Univariate time series are those that have just one variable, whereas multivariate time series have multiple variables. A time series may be discrete or continuous.

The evaluation of series observations in continuous time occurs at each time point (temperature measurement, river flow, chemical process concentration, etc.), whereas observations in discrete time are collected at specific points in time (business production, city population density, currency exchange rate, etc.). In a discrete time series, successive observations are often made at equal intervals of time, such as hourly, daily, weekly, monthly, or yearly.

Selecting a suitable model is crucial because it captures the fundamental structure of the data, and the fitted model is then utilized to make predictions. In general, time series models can describe many stochastic processes in a variety of ways. Nonetheless, two popular model types are identified in the literature: the Autoregressive Integrated Moving Average (ARIMA) model and the Autoregressive Moving Average (ARMA) model.

III.2.3.1. ARMA Model

Mathematical models called Autoregressive Moving Averages (ARMAs) are used in fields like banking, engineering, and statistics when a substantial amount of historical data is needed. ARMA models use a two-part polynomial—a Moving Average (MA) portion and an Autoregressive (AR) part—to statistically describe the stationary stochastic process.

The general form of this model is illustrated in the following equation:

$$Y_t = \sum_{i=1}^{p} \emptyset_i \; Y_{t-1} - \sum_{j=1}^{q} \theta_j \; e_{t-1} + \; e_t \qquad \text{III.6}$$

Where: Y_t can be any solar component (G_h, D_h, DNI), $\emptyset_i$ are the parameters of the autoregressive part, θ_j are the parameters of the moving average part, and et is a white noise with a mean equal to zero and a constant variance σe^2.

This model is a great resource for comprehending and projecting future time series values. It is quite adaptable since it may depict various series using various ordering schemes. When the data series exhibits an underlying linear correlation, it has demonstrated its forecasting capacity. However, the stationarity of the time series is the primary prerequisite for using this kind of modeling.

III.2.3.2. ARIMA Model

The application of ARMA-type models requires that time series be stationary, i.e., it is believed that the stochastic process's probability distribution will not change over time. In actuality, a lot of time series, like solar radiation, exhibit periodicity and time-dependent trends (solar irradiation fluctuates seasonally and during the day).

The Autoregressive Integrated Moving Average (ARIMA) model was initially introduced by Box-Jenkins [24] as a way to anticipate non-stationary time data. Statistical models of the ARIMA class can be generally transformed into stationary processes through the application of "differencing." Consequently, ARMA models can be generalized into ARIMA models. They include a moving average component (MA), an autoregressive component (AR), and a differencing term (d); it is commonly denoted by ARIMA (p, d, q) and given by the following mathematical expression:

$$Y_t = (1 - B)^d \; Y_t \tag{III.7}$$

$$Y_t = \sum_{j=1}^{p} \emptyset_j \; Y_{t-j} - \sum_{j=1}^{q} \theta_j \; Z_{t-j} \tag{III.8}$$

Where B is a lag operator, Z_t is an error term distributed as Gaussian white noise, and the parameters p, d, q are determined using various identification tools; the smallest values of p and q are often chosen as a good practice. Once these parameters are determined, the adjustment coefficients $\emptyset_j$ and θ_j are estimated using a minimization procedure [21].

III.2.4. Machine learning

When there is no representation between the input and the output, ML models are employed to find a relationship in data mining, prediction, and pattern recognition problems [22]. Supervised learning techniques in machine learning models include SVM, ANN, linear and nonlinear regression, and ANN [23, 24]. In addition, there are ensemble learning techniques

and unsupervised learning techniques [22]. In supervised learning, the algorithm learns the rule mapping inputs to outputs by using a collection of known inputs and their desired results [25]. The model will then be applied to forecast fresh data. Unsupervised learning algorithms, on the other hand, utilise data inputs to uncover latent structures without any prior knowledge of the intended results. [25]

Figure III.1, which shows various ML techniques, illustrates this. Given the availability of input-output data pairings in this investigation, supervised learning methodology ought to be applied.

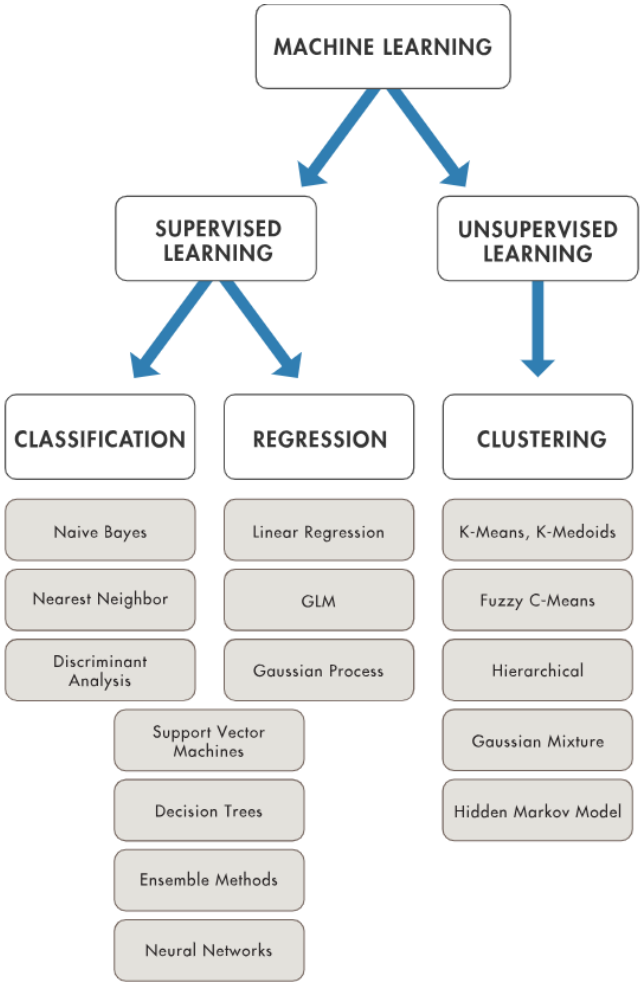


Figure III.1. Machine learning techniques [26]

Let’s start by presenting some supervised ML algorithms such as ANN, SVM, GB, and RF models.

III.2.4.1. Support Victor machine SVM

Vapnik developed the Support Vector Machine (SVM), a supervised learning technique that is applicable to regression. SVMs are categorized as machine learning techniques and are grounded in the structural risk reduction theory. This theory attempts to lower the empirical risk and confidence interval of the learning machine. This aids in the machine's development of powerful adaptability.

In addition to their strong mathematical basis in statistical learning theory, SVMs have demonstrated incredibly competitive performance in a wide range of real-world applications.

SVM techniques, which initially developed to address classification problems, can also be applied successfully to regression problems, also known as SVR (support vector regression) problems. As a result, support vector regression is the name given to the extension of support vector classification to include continuous outcomes. Statistical learning theory gives SVMs a solid mathematical foundation.

With yi as the output result, xi as an input vector, and N as the total amount of input data, SVM may be used to estimate a regression for the data set $\{(x, y)\} = N/i = 1$..

This is accomplished by transforming x into a parameter space using the nonlinear function $\varphi(x)$, and then calculating a regression function in the way that follows:

$$f\ (x) = \ \omega.\varphi\ (x) + b \qquad (\text{III. 9})$$

With its variables, the bias value b and weight vector ω, this function can most closely approximate the actual outcome y with an error tolerance of ε.

The coefficients ω and b are found by minimizing the resulting regularized risk function:

$$R\ (C)\ =\ C\ \sum_i^N L_\varepsilon\ (f\ (x_i),\ y_i)\ =\ \frac{1}{2}\ ||\ \omega\ ||^2 \qquad (\text{III. 10})$$

The term $\frac{1}{2}\ ||\ \omega\ ||^2$ regulates the model's degree of complexity and improves the generalization of the SVM. C (positive trade-off factor) determines the degree of empirical inaccuracy in the user-selected optimization issue.

The main difference between this approach and conventional regression is the use of a novel loss function (ε).

The linear loss function of Vapnik with the ε-insensitivity zone is:

$$L_\varepsilon\,(f(x_i),\,y_i)= 0 \text{ for } |f\,(x_i) - \,y_i| \leq \varepsilon \text{ otherwise } \; |f\,(x_i) - \,y_i| - \;\varepsilon \qquad \text{(III. 11)}$$

Consequently, the loss appears to be equal to zero if the difference between the actual and expected values is less than ε.

If the expected value falls inside the tube, the loss error is equal to zero.

The difference between the expected value and the radius ε of the tube indicates the amount of loss for all other expected sites outside the tube.

To avoid outliers, the gap widths ξ "above" and ξ* "below" the tube ε.

The following lowers the danger:

reduce

$$R\,(\xi\,,\xi^*,\mathrm{b},\omega\,)\;=\;\frac{1}{2}\;||\;\omega\;||^2 + C\sum_{i=1}^{N}\;(\xi_{\,i}\,+\;\xi_i^*) \qquad \text{(III. 12)}$$

subjected to

$$y_i - \;\omega\phi\;(x_i) - \;b_i\;\leq\;\varepsilon + \xi_i$$

$$b_i + \;\omega\phi\;(x_i) - y_i\;\;\leq\;\varepsilon + \xi_i^*$$

$$\xi_{\,i} \qquad\quad \xi_i^* \geq 0$$

The management of empirical risk severity is accomplished by $C\sum_{i=1}^{N}\;(\xi_{\,i}\,+\;\xi_i^*)$. In order to resolve the optimization issue, Lagrange multipliers α and α* have been included within the constraint equations.

The following is an expression of the equation in dual form:

$$R\,(\alpha\,,\alpha^*) = \sum_{i=1}^{N}\mathrm{y}_i(\alpha_i - \;\alpha_i^*) - \varepsilon\sum_{i=1}^{N}\;(\alpha_i - \;\alpha_i^*)\; - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}(\alpha_i - \;\alpha_i^*)\,(\alpha_j - \;\alpha_j^*)\,K\,(\mathrm{x}_i,\mathrm{x}_j) \qquad \text{(III. 13)}$$

With limitations: $\sum_{i=1}^{N} y_i(\alpha_i - \alpha_i^*) = 0$, $0 \leq \alpha_i \leq C$, $0 \leq \alpha_i^* \leq C_i$ =1,2,.....,N. where $K(x_i, x_j) = \phi(x_i).\phi(x_j)$ is a kernel function.

With the help of this function, SVM may construct nonlinear borders by simulating complex separating hyperplanes.

The ideal desirable weights vector of the regression hyperplane is found as follows after the computation of Lagrange multipliers:

$$\omega = \sum_{i=1}^{N} (\alpha_i - \alpha_i^*)\phi(x_i) \qquad \text{(III. 14)}$$

The following is the recast of Eq. (III.9):

$$f(x, \alpha_i, \alpha_i^*) = \sum_{i=1}^{N} (\alpha_i - \alpha_i^*) K(x_i, x_j) + b \qquad \text{(III. 15)}$$

The total number of support vectors is N, and the computed bias, b, is based on training samples. In general, the SVM statistical learning procedure's fundamental mathematical function is:

$$y = f(x) = \sum_{i=1}^{M} \alpha_i.\phi(x_i) = w.\phi(x) \qquad \text{(III. 16)}$$

The nonlinear transformation of $\varphi(x)$ produces the linearly weighted sum of M. The decision-making function of the SVM is

$$y = f(x) = (\sum_{i=1}^{N} \alpha_i.K(x_i, x_j)) - b \qquad \text{(III. 17)}$$

During the training phase, Vectors xi are used, where b and αi are parameters. An independent vector is denoted by xj, while the entire amount of training data is represented by N.

The parameters of αi and b can be obtained by maximizing their respective goal functions. With the appropriate kernel selection, the data could be divided in the feature space even if they were not separable in the original input space (see Table III.2). [27]

.

Table III.2. Some kernel functions

kernel function	Equation
Radial basis function (RB-F)	$K\left(\mathrm{x}_i, \mathrm{x}_j\right) = exp\ \left(-\gamma\, \mathrm{x}_i \,.\, \mathrm{x}_j || \left(\mathrm{x}_i - \mathrm{x}_j||^2 + r\right)\right.$
Polynomial K	$K\left(\mathrm{x}_i, \mathrm{x}_j\right) = \ \left(\gamma \mathrm{x}_i \,.\, \mathrm{x}_j + r\right)^d$
Linear	$K\left(\mathrm{x}_i, \mathrm{x}_j\right) = \ \mathrm{x}_i \,.\, \mathrm{x}_j$
Sigmoid	$K\left(\mathrm{x}_i, \mathrm{x}_j\right) = \ \tanh\ \left(\gamma \mathrm{x}_i \,.\, \mathrm{x}_j + r\right)$

γ ,d, r : are kernel parameters.

The general structure of a SVM is shown in Figure III.2

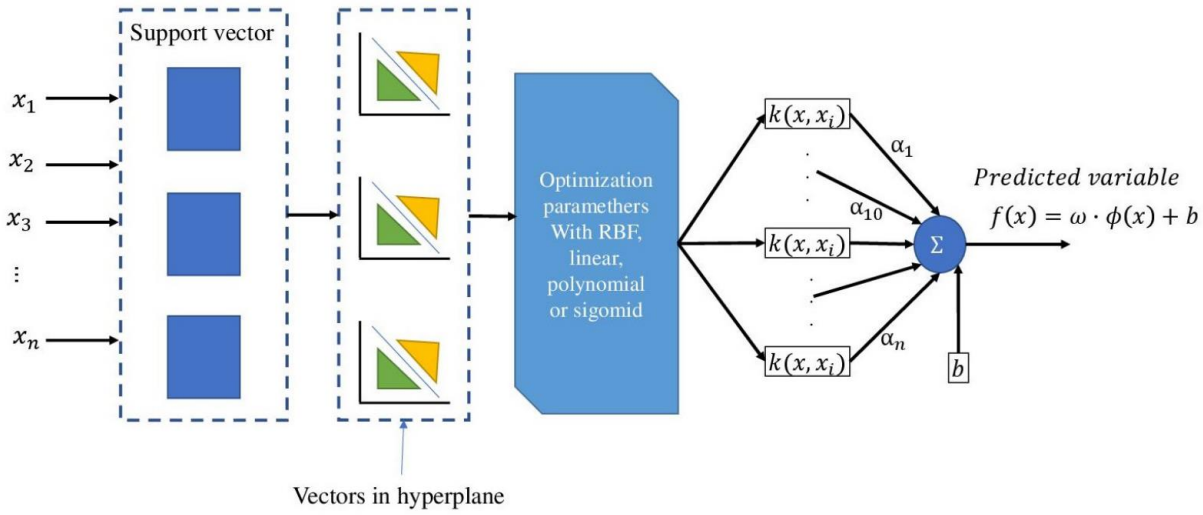


Figure III. 2. Support vector machine [28].

III.2.4.2. Artificial Neural Networks (ANNs)

Inspired by the functioning of the human brain, Artificial Neural Networks (ANNs) are relatively new. They make up an approximate mathematical model for variable relationships. Neural networks are built without imposing any specific structure on the data or assuming anything about the distribution of variables, in contrast to classical techniques.

ANNs are frequently used models in time series forecasting. They are considered useful tools for classification and regression problems, characterized by the ability to correlate nonlinear behaviors, and have been widely employed successfully in forecasting across various domains [29-30].

In data processing, ANNs serve as a technique for approximating complicated systems that is especially helpful when modeling these systems with traditional statistical techniques is challenging.

The initial work on ANNs was conducted by McCulloch and Pitts in 1943 [31]. Their aim was to understand the properties of nervous systems from elementary components. Through models based on simplified neurons (formal neurons), they demonstrated the possibility of constructing systems capable of computing logical functions. In 1949, Donald Hebb [32] initiated the problem of learning where he described how cells learn to modify the intensity of connections (called weights) that link them.

Work on ANNs continued to progress gradually over the years. In 1957, Rosenblatt designed the first neural network dedicated to pattern recognition. In turn, Widrow (1960) developed the Adaline model (Adaptive Linear Element), which originated the widely used gradient descent backpropagation algorithm today with Multilayer Perceptrons.

However, ANNs experienced a setback following criticisms from Minsky and Papert (1969), when they published a book highlighting the inability of single-layer perceptrons to solve nonlinear systems.

After a period of neglect in favor of classical methods, ANNs experienced a revival in the early 1980s, with the work of Hopfield [33, 34] on interconnected networks, and those of McClelland and Rumelhart [35] who introduced the error backpropagation learning algorithm.

III.2.4.2.1. The Formal Neuron

The formal neuron is a mathematical modeling that simulates the principles of biological neuron functioning (Figure III.3), particularly input summation. Each artificial neuron is a basic processor.

It receives a variable number of inputs xi from upstream neurons (Figure III.4). Each of these inputs is associated with a weight wj to produce the product $xj\ wij$, then from this value, a transfer (or activation) function g calculates the neuron state value to form the output yi:

$$y_i = g\ (\sum_{j=1}^{n} x_j\ W_{ij} + b_i) \qquad \text{(III.18)}$$

The weight that is not connected to the inputs is defined by a bias (b_i).

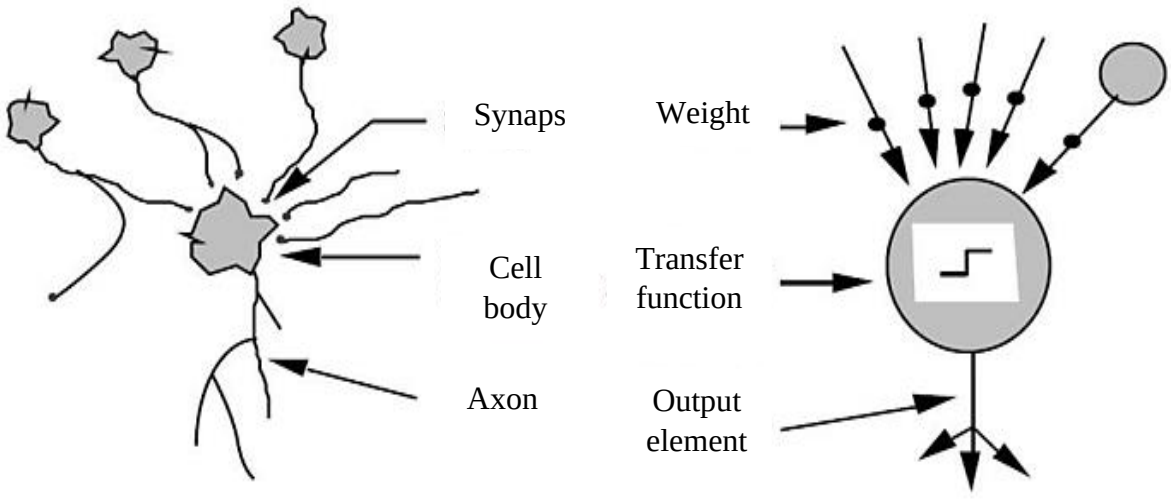


Figure III-3: Correspondence between the biological neuron and the artificial neuron [30]

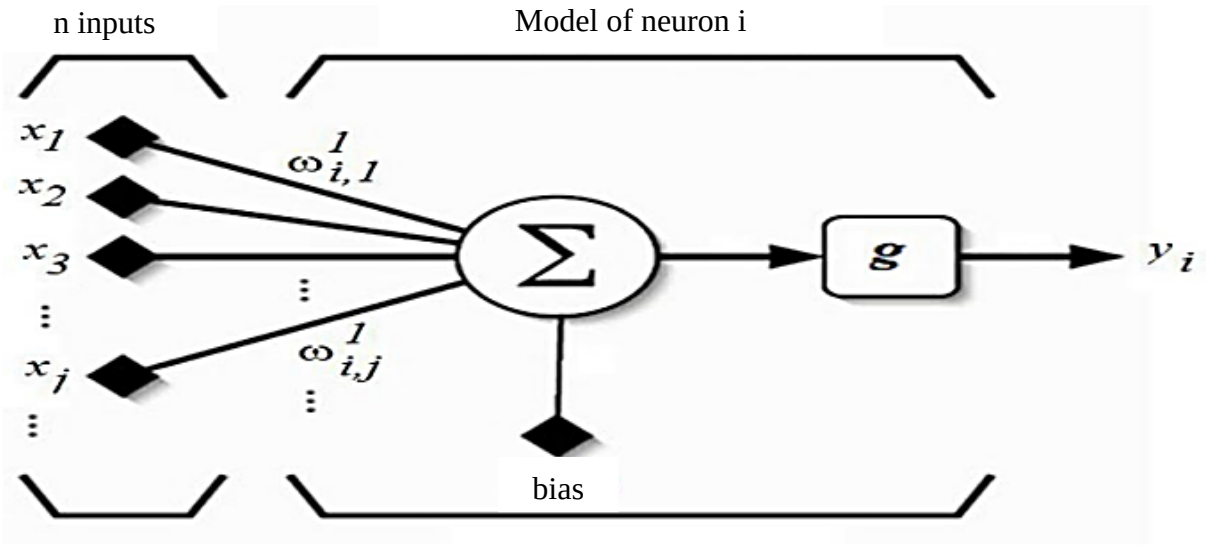


Figure III.4: Principle of the formal neuron [30-33]

However, there are several types of activation functions used to limit the amplitude of the neuron's output signal and reproduce the threshold effect observed in biological neurons. The most commonly used functions are:

- The linear function, which directly assigns its input to its output: g(x) = x,

- The hyperbolic tangent function, expressed as: $g(x) = tanh(x) = \frac{(e^x - e^{-x})}{(e^x + e^{-x})}$,

- The standard sigmoid function: $g(x) = \frac{1}{(1 + e^{-x})}$

It can be noted that the sigmoid transfer function has a quasi-linear regime around 0 and is nonlinear elsewhere. This characteristic allows the network to adapt to both linear and nonlinear problems.

III.2.4.2.2. Types of Neural Networks

Generally, the use of a single isolated neuron has no practical effect or interest. However, when it is linked and connected to other neurons, it is possible to form a network capable of solving complex problems such as pattern recognition, time series prediction, etc.

The most commonly used network in the literature is the Multi-layer Perceptron (MLP). In this network, neurons are organized to form multiple layers.

The number of layers varies from 3 to 4, including an input layer, an output layer, and one or two hidden layers, with information flowing only from the input layer to the output layer.

The input layer contains as many neurons as the dimension of the input space, and the output layer contains a single neuron for the prediction value. Finally, one or more hidden layers contain a variable number of neurons.

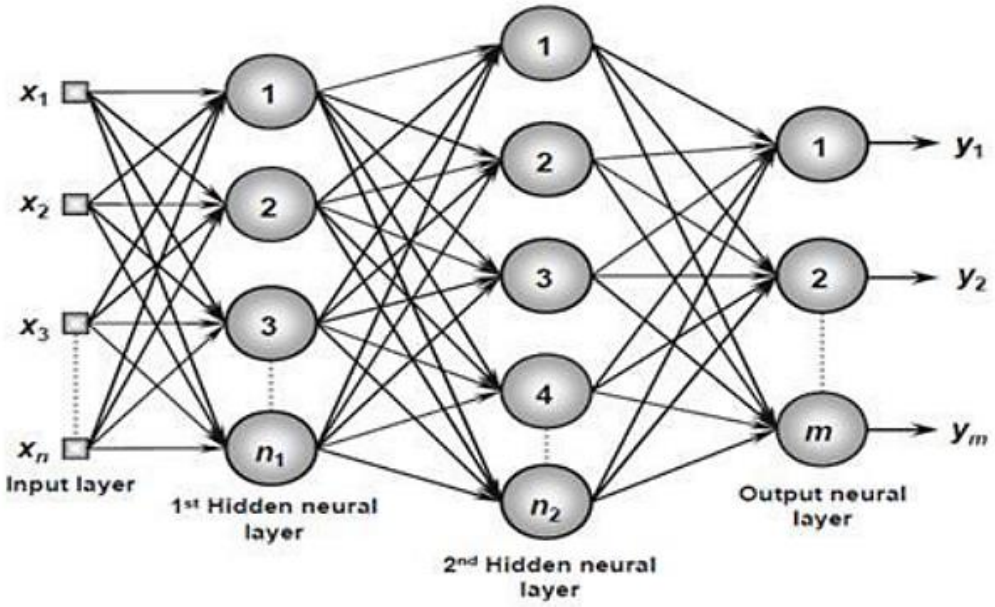


Figure III.5: Example of an MLP network with multiple layers [36]

Another category of ANNs is that of recurrent networks, which have recurrent connections that feedback one or more outputs to the inputs, unlike feedforward networks (such as MLP) where information flows from input to output without any backward feedback. The best example of this type of recurrent network is the one developed by Hopfield [33], referred to as the Hopfield network in general.

The properties of this neuronal architecture are as follows:

- Ability to memorize relationships,
- Dynamic behavior,
- Easy implementation on analog media,
- Capability to store information.

A recurrent network with one of its output signals sent back to the intermediate layer is shown in the accompanying figure.

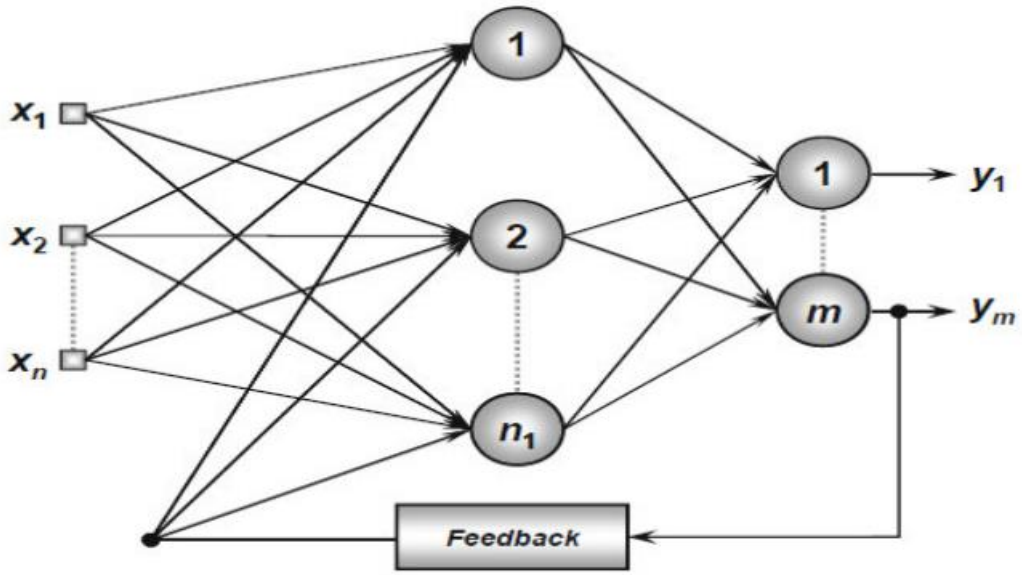


Figure III.6: Example of a recurrent network [36]

Radial Basis Function (RBF) networks are another family of ANNs;

They can also be used in almost all types of problems solved by MLP networks. In contrast to the latter, which can consist of multiple intermediate layers, the typical structure of RBF consists of a single layer with a Gaussian activation function. The training process used to modify the weights of the two layers of neurons in RBF networks is one of its key features.

The activation function, which is usually a radial basis function like the Gaussian function, that the neurons in the intermediate layer use is another feature of RBF networks. With supervised training, RBF networks are also a part of the MLP design.

It can be confirmed from Figure III.7 that information flows through the structure from the input layer to the intermediate layer (neurons with Gaussian activation function) and finally to the output layer of neurons (linearly activated neurons).

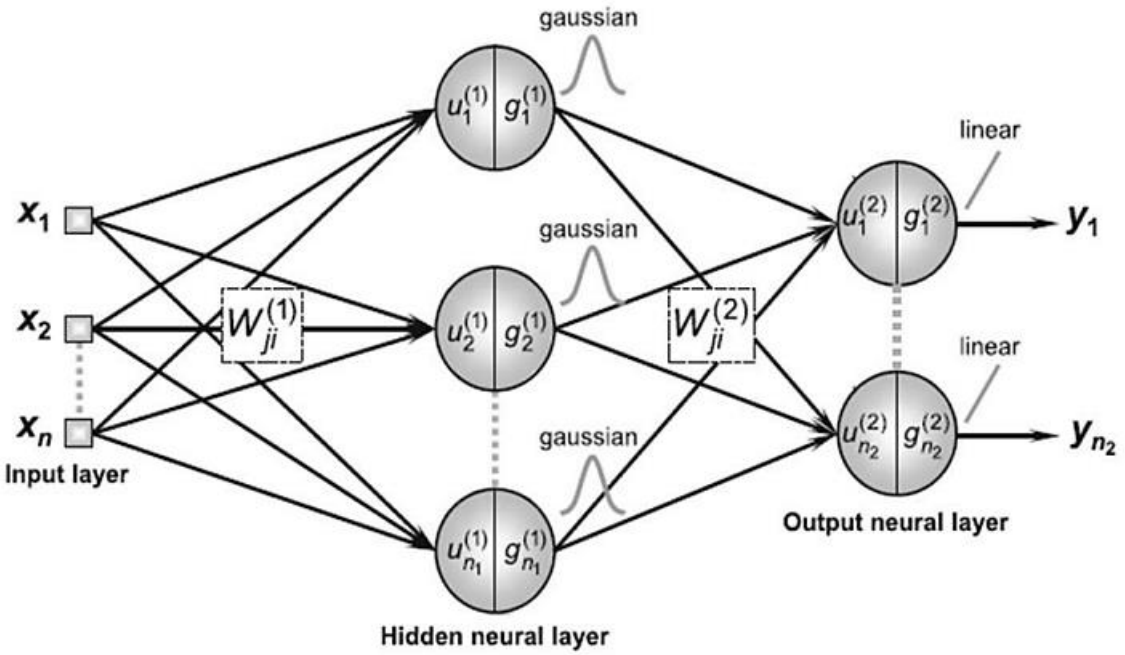


Figure III.7: Typical Configuration of an RBF Network [36]

One of the most relevant characteristics of ANNs is their learning or "training," which involves adjusting the network weights based on input data until the outputs closely match the desired input dataset.

There are three modes of learning:

- Supervised learning, where the network is forced to converge to a specific final state while being given a pattern. The network adapts by comparing the calculated result with the expected output.

- Unsupervised learning, which does not require prior knowledge of the desired output; the network is free to converge to any final state when given an input.

- Reinforcement learning, similar to supervised learning but instead of providing desired results, the network is given a grade (or score) measuring its performance after iterations.

III.2.4.3. Gradient Boosting

An ensemble learning method called the gradient boosting (GB) algorithm joins a series of weak learners to create a strong model. By gradually introducing new learners, each of which raises the accuracy of the one before it, the GB algorithm seeks to increase prediction accuracy. Additionally, GB is considered a gradient descent algorithm for regression problems, evaluating weak learner shortcomings using gradients in the loss function.

While regression trees are commonly used as weak learners, their combination in an ensemble significantly improves accuracy. The primary objective is to find the approximation function $f\hat{}\,(X)$ that minimizes the loss function $L(f(X))$.

$$\hat{f}(X) = \arg min_f \;\; L(f(X)) \tag{III.19}$$

where X is the input vector.

We consider that $f_0\,(X)$ and $f_{t-1}\,(X)$ are the initial and current approximations, respectively, of $f\hat{}(X)$; the algorithm updates the approximation function $f_t\,(X)$ using the gradient $g_t\,(X)$ and the step size ρ_t ($\rho_t > 0$) as follows [37]:

$$f_t\,(X) = f_{t-1}\,(X) - \rho_t \,.\, g_t\,(X) \tag{III.20}$$

the process of updating continues until the number of iterations is reached. The negative gradient $-g_t\,(X)$ is approximated using the parametric function $h(X;\,\theta_t)$, which is described for the regression terminal node J as

$$h(X;\,\theta_t) = \sum_{j=1}^{J} \bar{y}_j \; I_{\,[X \in R_j]} \tag{III.21}$$

where θ_t is a parameter vector, $\bar{y}_j$ and R_j are, respectively, the mean and separated input space of the J-th terminal node of the regression tree [38].

the accuracy of the GB model depends on hyperparameter values such as the number of trees m, the learning rate η, and the maximum number of splits (MaxNumSplits). These hyper-parameters are tuned by exploring predefined values to find the optimal architecture of the GB model [39].

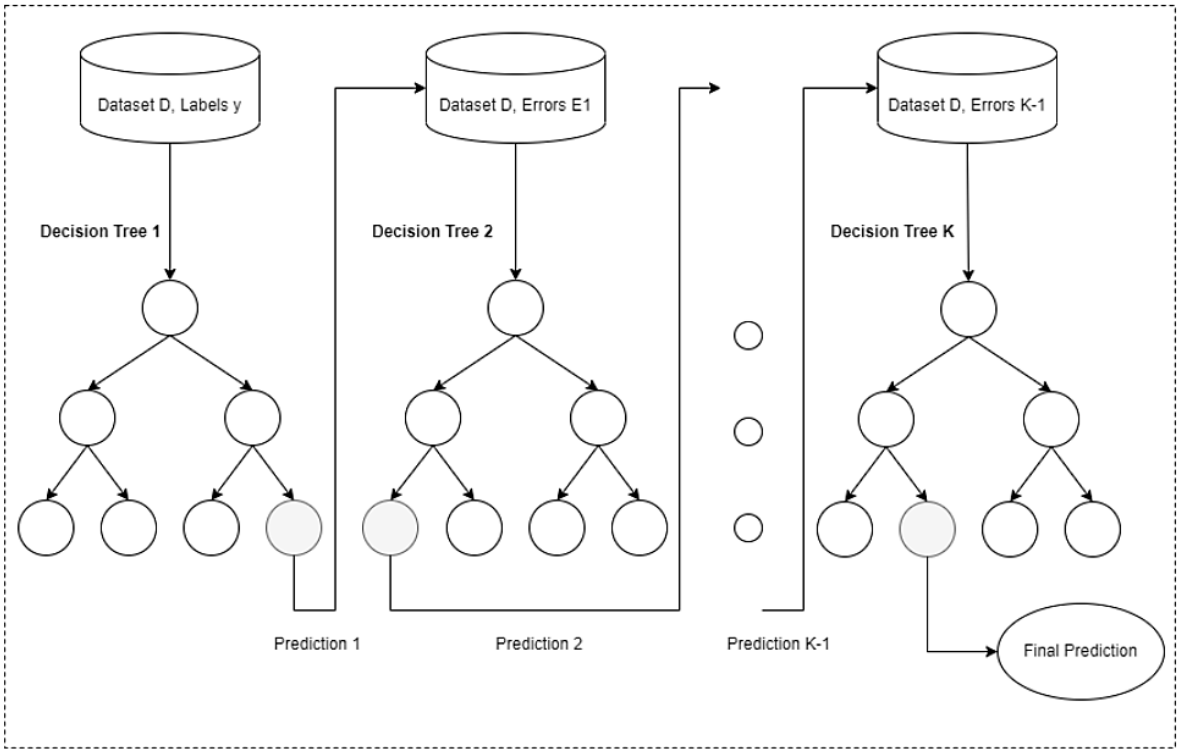


Figure III.8: The general structure of the gradient boosting algorithm [40].

III.2.4.4. Random Forest.

Tree-based models, particularly the random forest (RF) algorithm, are increasingly used in solar radiation forecasting due to their high performance [41, 42, 43]. RF, an ensemble learning technique, combines multiple decision trees to enhance accuracy and address individual tree shortcomings [43, 44]. It predicts outputs by averaging the values from all trees, resulting in more precise predictions [45].

RF employs two key concepts: bagging and random feature selection. Bagging randomly divides the dataset with replacement, training each tree with different subsets [42]. Random feature selection occurs at each node split, where a random sampling of features is chosen [43]. Trees are trained with bootstrap samples, with one-third left out as out-of-bag (OOB) data for validation [42].

At each node, binary splits with randomly selected variables grow the trees until the optimal structure is achieved. The final prediction is the average output of all trees [46]. RF performance depends on hyperparameters like the number of trees (m), variables selected at each split (NumVarSplits), minimum samples per leaf (MinLeafSize), and maximum splits (NumMaxSplits).

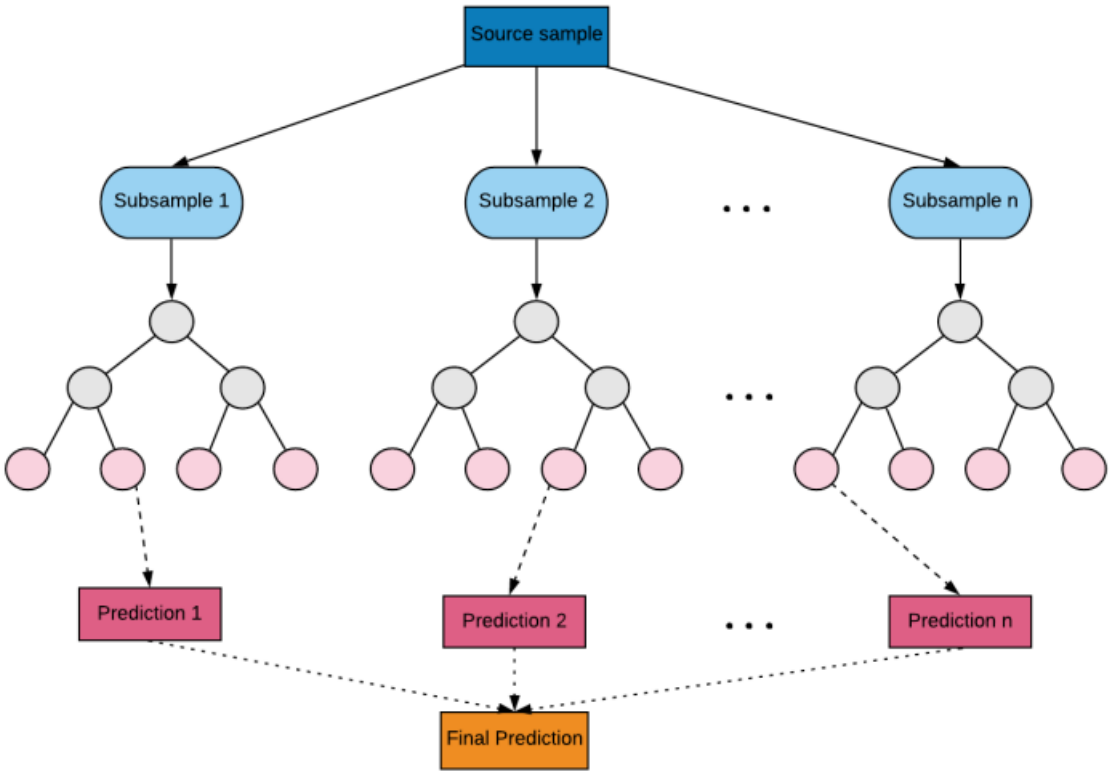


Figure III.9: The general structure of random forest model. [47].

These hyperparameters should be optimized to avoid overfitting or underfitting. Therefore, grid search is used to find the optimal combination of hyperparameters for the RF model [39]. This involves manually defining and testing multiple combinations, selecting the model with the most accurate predictions as the best [48].

III.2.4.5. Linear Regression model

Regression analysis is a statistical technique for investigating and modelling the relationship between variables [49]. In fact, the regression analysis is the most widely used statistical technique.

- **The simple linear regression**

The simple linear regression model used is a model with a single independent variable x that has a relationship with a response variable y that is a straight line. This simple linear regression model is given by

$$y = \beta_0 + \beta_1 x + \varepsilon \quad \text{(III.22)}$$

where the intercept β_0 and the slope β_1 are unknown constant and ε is a random error. The errors are assumed to have mean zero and unknown variance σ^2. The parameters β_0 and β_1 are unknown and must be estimated using sample data.

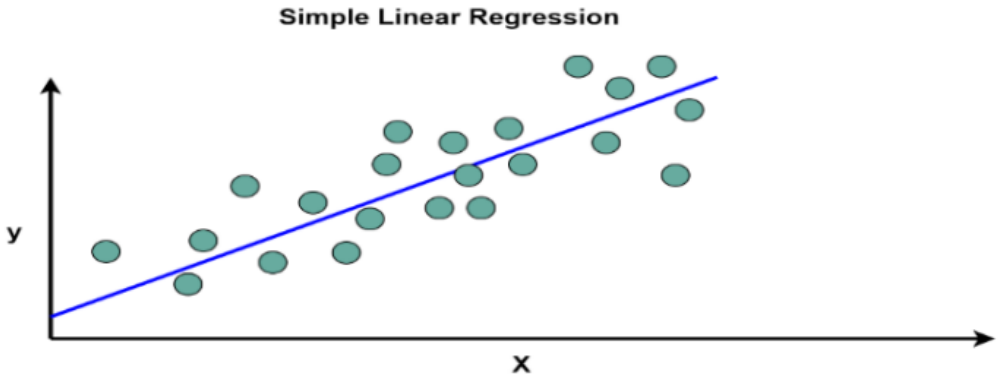


Figure III.10: simple linear regression structure [50]

The simple linear regression equation is also called the least squares regression equation. It tells the criterion used to select the best fitting line, namely the sum of the squares of the residuals should be least. That is, the least squares regression equation is the line for which the sum of squared residuals $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ is a minimum. [51]

- **Multiple linear regression**

A In the multiple linear regression model, the relationship between multiple independent variables (x_1, x_2, x_3, x_4) and a dependent variable $(\hat{y})$ is examined.

The regression function used in this example is the first order, and it is assumed that each independent variable is a linear relationship with the dependent variable, as in (1)

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + b_4 x_4 \quad \text{(III.23)}$$

where b_0 is the y-axis cut point of the modified regression curve, b_1 is the coefficient x_1 of the first guess variable, b_2 is the coefficient x_2 of the first guess variable, and so on. [52]

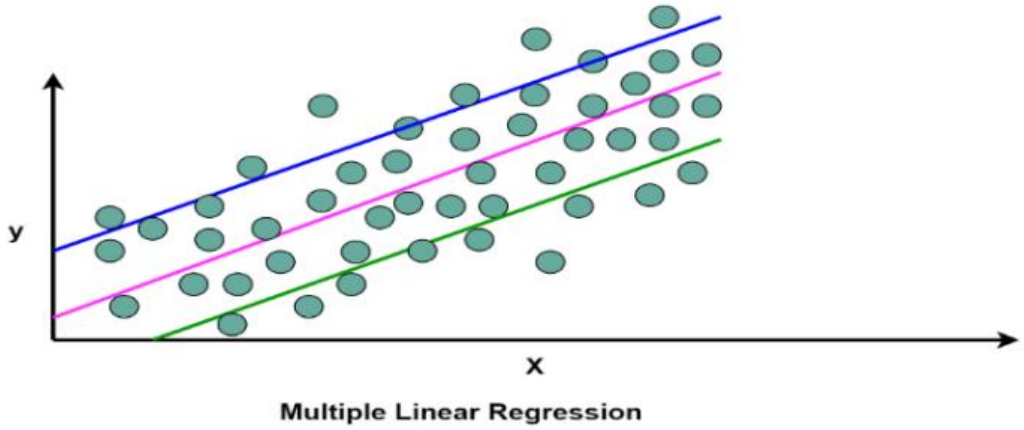


Figure III.11: multiple linear regression structure. [50]

III.3.Hybrid models

Building hybrid models involves combining multiple machine learning algorithms or modeling techniques to leverage the strengths of each approach. When it comes to predicting solar radiation, hybrid models can be particularly effective in capturing the complex relationships and patterns present in solar radiation data. Here's how optimization algorithms can be useful in enhancing the performance of machine learning models for predicting solar radiation:

1. **Parameter Optimization:** Optimization algorithms can be used to fine-tune the parameters of individual models within the hybrid framework. For example, techniques like grid search or Bayesian optimization can help identify the optimal hyperparameters for each component model, such as neural network architectures or support vector machine kernels.

2. **Feature Selection:** Optimization algorithms can aid in selecting the most relevant features from the input data for each component model. This can involve techniques like genetic algorithms or simulated annealing to search for the subset of features that maximize predictive performance while minimizing complexity.

3. **Model Combination:** Optimization algorithms can optimize the combination of different models in the hybrid framework. Ensemble learning methods, such as boosting or stacking, can be used to combine the predictions of multiple base models in an optimal way, leveraging their complementary strengths to improve overall performance.

4. **Hyperparameter Tuning:** Optimization algorithms can also be applied to optimize the hyperparameters of the ensemble learning techniques themselves. This can involve

tuning parameters such as the learning rate or regularization strength in boosting algorithms, or the choice of meta-learner in stacking models.

5. **Adaptive Learning:** Some optimization algorithms, such as evolutionary algorithms or reinforcement learning, can adaptively adjust the parameters of the hybrid model over time in response to changes in the underlying data distribution or environmental conditions. This can help ensure that the model remains robust and effective in dynamic environments.

By leveraging optimization algorithms in these ways, hybrid models for predicting solar radiation can achieve higher accuracy and robustness compared to individual models or traditional forecasting techniques. Additionally, they can be more flexible and adaptable to changing conditions, making them valuable tools for renewable energy management and grid integration.

III.4. Performance and Evaluation of Predictors

The accuracy of solar radiation models is crucial not only for the final stages of solar projects but especially in the initial stages before any model construction. There are numerous statistical tools used to verify the effectiveness of a predictor.

The accuracy of predicted models is judged by statistical indicators such as the coefficient of determination (R^2), mean bias error (MBE), root mean square error (RMSE), mean absolute error (MAE), etc. These tests are generally applied in comparing solar radiation models. These tools are detailed as follows:

III.4. 1. Coefficient of Determination (R^2)

The coefficient of determination is the proportion of variability in a dataset that is accounted for by a statistical model, where variability is quantitatively measured as the sum of squared deviations. Most often, it is defined as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - Y_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \quad \text{(III.24)}$$

The coefficient of determination is a statistic that provides information about the goodness of fit of a model. In regression, the R-squared coefficient is a statistical measure indicating how close the regression line is to the actual data points.

An R-squared value of 1 indicates that the regression line fits the data perfectly, whereas an R-squared value of 0 signifies that there is no correlation between the model and the actual data points.

III.4.2. Root Mean Square Error (RMSE)

The root mean square error provides information about the performance of predictors by allowing the comparison of actual deviations between estimated values and measured values. The lower the RMSE value, the better the predictor's performance. This statistical indicator is given by the following relationship:

$$\text{nRMSE} = \frac{1}{N}\left[\sum_{i=1}^{N}(y_i - Y_i)^2\right]^{1/2} \quad \text{(III.25)}$$

III.4. 3. Mean Bias Error (MBE)

The MBE is defined as the arithmetic mean of errors. If its value is equal to zero, it does not imply that the model provides error-free estimates. Thus, the MBE provides a measure of the general tendency of a given estimator, i.e., over-estimation (positive values) or underestimation (negative values). However, smaller MBE values indicate better model performance. The MBE is given by the following equation:

$$\text{nMBE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - Y_i) \quad \text{(III.26)}$$

III.4.4. Mean Absolute Percentage Error (MAPE)

The MAPE expresses the accuracy in percentage of the error. Since this number is a percentage, it may be easier to understand than other statistics. This error is defined by:

$$MAPE = \frac{100}{N}\sum_{i=1}^{N}\left|\frac{y_i - Y_i}{\bar{Y}_i}\right| \quad \text{(III.27)}$$

In general terms, these error formulations provide quantitative measures that have the same physical units as the dependent variable "Y_i". In some cases, normalized formulations of these errors (nMBE), (nRMSE) are required; they are defined as follows:

$$\text{nRMSE} = \frac{1}{N}\left[\sum_{i=1}^{N}\left(\frac{y_i - Y_i}{\bar{Y}_i}\right)^2\right]^{1/2} \quad \text{III.28}$$

$$\text{nMBE} = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{y_i - Y_i}{\bar{Y}_i}\right) \quad \text{III.29}$$

It is noted that "y_i" represents the measured values, and "Y_i" represents the estimated values [53]

III.5. Conclusion

Precisely forecasting solar radiation is essential for incorporating photovoltaic (PV) systems' growing contribution to the generation of electricity, especially in light of the current energy transition. Grid operators can improve service quality by managing variable electricity output more effectively through the use of forecasted solar energy variation. PV energy forecasting involves a variety of techniques, such as advanced algorithms for machine learning and empirical models. These approaches use statistical techniques and observed data to represent variability in solar radiation, providing modeling flexibility for complex relationships and patterns.

The basic methods for forecasting solar radiation have been described in this chapter, along with error parameters that can be used to compare various forecasting models. With the use of satellite imagery and historical meteorological data, machine learning algorithms have become highly effective methods for forecasting solar radiation. Such insights help make well-informed decisions while optimizing solar energy systems.

III.6. References

1. Ångström A. (1924) Solar and terrestrial radiation. Q J Roy Meteor Soc, 50, 121-125.
2. Prescott JA. (1990) Evaporation from a water surface in relation to solar radiation, Philosophical Transactions of the Royal Society, 64, 114-120.
3. Besharat F. Dehghan A. Faghih A. (2013) Empirical models for estimating global solar radiation: A review and case study. Renew Sustain Energy Rev, 21, 798-821.
4. Pandey CK. Katiyar AK. (2013) Solar Radiation: Models and Measurement Techniques. Journal of Energy. Volume 2013, Article ID 305207, 8 pages.
5. Ogelman H. Ecevit A. Tasdemiroglu E. (1984) A new method for estimating solar radiation from bright sunshine data. Solar Energy, 33, 619–625.
6. Samuel T. (1991) Estimation of global radiation for Sri Lanka. Solar Energy, 47(5), 333-337.
7. Bahel V., Bakhsh H., Srinivasan R., (1987) A correlation for estimation of global solar radiation. Energy, 12(2), 131-135.
8. Benson RB., Paris MV., Sherry JE., Justus CG., (1984) Estimation of daily and monthly direct, diffuse and global solar radiation from sunshine duration measurements. Solar Energy, 32(4), 523-535.
9. Chegaar M., Chibani A., (2001) Global solar radiation estimation in Algeria. Ener Conv & Mangt, 42, 967-973.
10. Li H, Ma W, Lian Y, Wang X, Zhao L. (2013) Global solar radiation estimation with sunshine duration in Tibet, China. Renewable Energy, 36(11), 3141-3145.
11. El-Metwally M. (2005) Sunshine and global solar radiation estimation at different sites in Egypt. Journal of Atmospheric and Solar-Terrestrial Physics, 67(14), 1331-1342.
12. Katiyar AK, Pandey CK. (2010) Simple correlation for estimating the global solar radiation on horizontal surfaces in India. Energy, 35(12), 5043-5048.
13. Al-saad MA. (1990) Characteristic distribution of global solar radiation for Amman, Jordan. Solar & Wind Technology, 7(2-3), 261-266.
14. Said R, Mansor M, Abuain T. (1998) Estimation of global and diffuse radiation at Tripoli. Renewable Energy, 14(1), 221-227.
15. Ampratwum DB, Dorvlo AS. (1999) Estimation of solar radiation from the number of sunshine hours. Applied Energy, 63(3), 161-167.

16. Firoz A, Intikhab U. (2004) Empirical Models for the Correlation of Monthly Average Daily Global Solar Radiation with Hours of Sunshine on a Horizontal Surface at Karachi, Pakistan. Turk J Phys, 28(5), 301-307.
17. Almorox J, Hontoria C. (2004) Global solar radiation estimation using sunshine duration in Spain. Ener Conv & Manag, 45(9), 1529-1535.
18. Ulgen K, Hepbasli A. (2002) Comparison of solar radiation correlations for İzmir, Turkey. Inter J Enr Resear, 26(5), 413-430.
19. Gladius L. (1992) An empirical relation for estimating global irradiation for Tennessee, U.S.A. Ener Conv & Mangt, 33(12), 1097-1099.
20. Baduscu V. (2008) Modeling Solar Radiation at the Earth's Surface. Edition Springer.
21. Box GEP. Jenkins GM. Reinsel GC. (2008) Time series analysis. forecasting and control. 4th ed. Edition Wiley.
22. C. Voyant, G. Notton, S. Kalogirou, M.L. Nivet, C. Paoli, F. Motte, and A. Fouilloy. (2017) Machine learning methods for solar radiation forecasting: A review. Renewable Energy, 105, 569-582.
23. S. Karasu, A. Altan, Z. Saraç, and R. Hacıoğlu. (2017) Prediction of wind speed with non-linear autoregressive (NAR) neural networks. IEEE 25th Signal Processing and Communications Applications Conference, 1-4.
24. S. Karasu, A. Altan, Z. Saraç, and R. Hacıoğlu. (2017) Estimation of fast varied wind speed based on NARX neural network by using curve fitting. International Conference on Advanced Technology & Sciences (ICAT'17), 551-557.
25. R. H. Inman, H. T. C. Pedro, and C. F. M. Coimbra. (2013) Solar forecasting methods for renewable energy integration. Progress in Energy and Combustion Science, vol. 39, no. 6, 535–576.
26. What Is Machine Learning? How It Works, Techniques & Applications. (2024) Available online: https://ch.mathworks.com/discovery/machine-learning.html (accessed on 2 April 2024).
27. Vapnik, V. (2013) The Nature of Statistical Learning Theory. Springer Science & Business Media, Berlin, Heidelberg, Germany.
28. Deo, R.C.; Wen, X.; Qi, F. (2016) A wavelet-coupled support vector machine model for forecasting global incident solar radiation using limited meteorological dataset. Appl. Energy, 168, 568–593.

29. Bishop CM. Neural networks for pattern recognition. (Year not mentioned) Clarendon Press. Oxford edition.
30. Touzet C. (1992) Les réseaux de neurones artificiels : introduction au connexionnisme, cours et exercices.
31. Mc Culloch WS. Pitts WH. (1943) A logical calculus of the immanent in nervous activity. Bulletin of Mathematical biophysics, 5, 115-133.
32. Hebb DO. (1949) The organization of behavior, Wiley Edition.
33. Hopfield JJ. (1982) Neural Networks and Physical Systems with Emergent Collective Computational Abilities. Proceedings of the National Academy of Sciences, 79, 2554-2558.
34. Hopfield JJ. (1984) Neurons with Graded Response Have Collective Computational Properties like Those of Two-State neurons. Proceedings of the National Academy of Sciences, 81, 3088-3092.
35. McCleland JL. Rumelhart DE. (1986) A distributed model of human learning and memory. Journal of experimental psychology, 114, 159-188.
36. Da Silva IN. Spatti DH. Flauzino RA. Bartocci Liboni LH. Reis Alves SF. (2016) Artificial neural networks: a practical course. Edition Springer.
37. M. A. Hassan, A. Khalil, S. Kaseb, and M. A. Kassem. (2017) Exploring the potential of tree-based ensemble methods in solar radiation modeling. Applied Energy, 203, 897–916.
38. J. Lee, W. Wang, F. Harrou, and Y. Sun. (2020) Reliable solar irradiance prediction using ensemble learning-based models: a comparative study. Energy Conversion and Management, 208, Article ID 112582.
39. A. Torres-Barrán, Á. Alonso, and J. R. Dorronsoro. (2019) Regression tree ensembles for wind energy and solar radiation prediction. Neurocomputing, 326-327, 151–160.
40. Sarmas, E.; Spiliotis, E.; Dimitropoulos, N.; Marinakis, V.; Doukas, H. (2023) Estimating the Energy Savings of Energy Efficiency Actions with Ensemble Machine Learning Models. Appl. Sci., 13, 2749.
41. G. M. Yagli, D. Yang, and D. Srinivasan. (2019) Automatic hourly solar forecasting using machine learning models. Renewable and Sustainable Energy Reviews, 105, 487–498.

42. I. Anwar and T. Khatib. (2017) A novel hybrid model for hourly global solar radiation prediction using random forests technique and firefly algorithm. Energy Conversation Management, 138, 413–425.
43. L. Benali, G. Notton, A. Fouilloy, C. Voyant, and R. Dizene. (2019) Solar radiation forecasting using artificial neural network and random forest methods: application to normal beam, horizontal diffuse and global components. Renewable Energy, 132, 871–884.
44. L. Breiman. (2001) Random forests. Machine Learning, 45(1), 5–32.
45. L. Breiman. (1996) Bagging predictors. Machine Learning, 24(2), 123–140.
46. M. W. Ahmad, M. Mourshed, and Y. Rezgui. (2017) Trees vs Neurons: comparison between random forest and ANN for high-resolution prediction of building energy consumption. Energy and Buildings, 147, 77–89.
47. Dmitrievsky, M. (2018) Random Decision Forest in Reinforcement Learning; MetaQuotes Language 5 (MQL5).
48. Available online: https://www.mql5.com/en/articles/widget/3856 (accessed on 3 April 2024).
49. Montgomery, D. C. and Peck, E. A (1992) “Introduction to linear regression analysis” 2nd ed. Wiley. New York.
50. Rasel Uddin Md., (2023) ; COVID-19 Pandemic Data Analysis and Prediction Using Machine-Learning Algorithms ; September.
51. Ibrahim, S., Daut, I., Irwan, Y. M., Irwanto, M., Gomesh, N., & Farhana, Z. (2012) Linear Regression Model in Estimating Solar Radiation in Perlis. Energy Procedia, 18, 1402–1412.
52. Karasu S., Altan A., Sarac Z., hacioglu R., (2017). prediction of solar radiation based on machine learning methods the journal o f cognitive systems volume 02 number 01 JUNE 2017 p 16-20.
53. Gairaa K. (2018) Prédiction du rayonnement solaire en Algérie à partir des modèles hybrides ; These De Doctorat LMD.

Chapter IV : Solar Radiation Forecasting Based On A New FFA-ANN Model

IV.1. Introduction

The precise forecasting of solar radiation plays a pivotal role in optimizing the layout and operation of thermal and photovoltaic systems, thereby facilitating the generation of clean and renewable energy. In this context, artificial neural networks (ANNs) have emerged as a versatile and effective tool for solar radiation prediction due to their ability to yield accurate results with a limited number of input parameters.

Research conducted across various geographical regions has consistently demonstrated the efficacy of ANNs in solar radiation forecasting. For instance, studies in Algeria, Greece, Italy, India, Spain, and Iran have showcased ANNs' capability to accurately predict solar radiation patterns, often surpassing traditional forecasting methods.

Furthermore, support vector machines (SVMs) have gained prominence for their effectiveness in addressing nonlinear regression problems, making them valuable tools for solar radiation prediction. Hybrid approaches, which combine SVMs with optimization algorithms such as the firefly algorithm, have shown promising results, demonstrating improved performance compared to standalone models.

In light of these advancements, this chapter seeks to contribute to the ongoing research in solar radiation forecasting by experimentally evaluating a hybrid firefly algorithm-based (FFA-ANN) model in the context of Algeria. By comparing this hybrid approach to standalone ANN and SVM models using meteorological data from selected study regions, we aim to provide insights into the effectiveness and applicability of hybrid machine learning techniques in solar radiation prediction. Through this analysis, we endeavor to enhance our understanding of renewable energy forecasting methodologies, thereby contributing to the advancement of sustainable energy practices.

IV.2. Literature review of the models used

For thermal and solar photovoltaic systems to be set out and run as efficiently as possible, an accurate and reliable sun radiation forecast is essential. This is essential for the production of clean and renewable energy. Artificial neural networks (ANNs) are widely used in solar radiation forecasting because of their diverse structures and ability to produce outstanding

effectiveness with a small number of accessible factors as inputs [1]. A Literature review of the models used is shown in Table IV.1

Table IV.1: Literature review of the models used

Study	Location	Model(s)	Key Findings
[2]	Algeria	ARMA-ANN	Combined model (ARMA-ANN) outperformed ARMA by ~18% in MAE, and ANN by < 3%
[3]	Algeria	Multilayer ANN	Logistic sigmoid function with 15-neuron hidden layer suitable for hourly sun irradiation estimation
[4]	Greece	ANN, MLR, Empirical	Combining MLR and ANN with temperature differential and extraterrestrial radiation increases accuracy; ANN is more sophisticated
[5]	Algeria	Multitask Neural Network	Sigmoidal units yielded better results; multitask model simpler and more efficient
[6]	Isparta, Turkey	RF, ANN, k-NN, DL	ANN and DL perform best; dummy variables increase performance for ANN and DL, decrease for RF and k-NN
[7]	Evora, Portugal	ANN	Two models: instantaneous and period predictions; improvements in R2, MAE, and RMSE
[8]	France	ANN	Exogenous data useful in winter; endogenous data adequate in summer for MLP
[9]	Algeria	ANN	High precision in estimating global sun radiation based on air temperature and humidity
[10]	Kocaeli, Turkey	ANN	Accurate solar radiation forecast; important for energy enterprises and climate research
[11]	Algeria	SVM-WT, SVM-FFA	SVM-WT more efficient than SVM-FFA; hybrid approaches perform better than standard SVM
[12]	Algeria	SVM-R	Effective in estimating daily global solar irradiation using sunshine ratio
[13]	Spain	Neural Computation Algorithms	Structure sigmoid unit-product unit with evolutionary training most effective
[14]	Algeria	SVM	Good agreement between expected and observed solar radiation values with minimal parameters
[15]	Kerman, Iran	SVM-WT, Empirical, ANN, SVM-RBF	SVM-WT more accurate and efficient compared to other models
[16]	Saudi Arabia	SVM	Effective GSR prediction using temperature and humidity measurements

[17]	Turkey	SVM, FRF	FRF-SVM with multiple kernels effective for long-term forecasting with complex variables
[18]	Mexico	ANN, SVM, ANFIS	SVM outperformed ANN and ANFIS in forecasting daily global sun radiation
[19]	India	SVM-R	Superior results with temperature, air pressure, day number, wind speed, and humidity combinations
[20]	Air Pollution Zones	SVM-PSO, SVM BAT, SVM-WOA	Hybrid models improved forecasting accuracy and convergence rate
[21]	China	SVM-FFA, Empirical, CNQR	SVM-FFA outperformed other models; CNQR recommended for balance of cost and accuracy
[22]	Iran	SVM-FFA	Exceptional performance using various datasets and parameters; advantage over empirical models
[23]	Nigeria	SVM-FFA, ANN	SVM-FFA outperformed ANN in horizontal GSR predictions using temperature and sunshine duration

This study aims to investigate our proposed hybrid firefly algorithm-based (FFA-ANN) model as a hybrid machine learning approach for modeling and predicting solar radiation through simulation. We compare the FFA-ANN model to two standalone models using meteorological data from six study regions in Algeria.

IV.3. Models and materials

IV.3.1. Areas of research and datasets

The southwest region of Algeria has been selected as the research area. The climatic characteristics of the provinces of Adrar, Eloued, Ouargla, Tamanrasset, Timimoun, and Bechar were taken into consideration when choosing them as study locations. There is a lot of potential solar radiation in this area. Over 3,900 hours of insolation are experienced in this region at its longest [24]. In most of Algeria (Figure IV.1), the average amount of solar energy collected on horizontal surfaces is more than five kWh/m^2, or around 2,263 kWh/m^2/year in the southern region of the country [25].

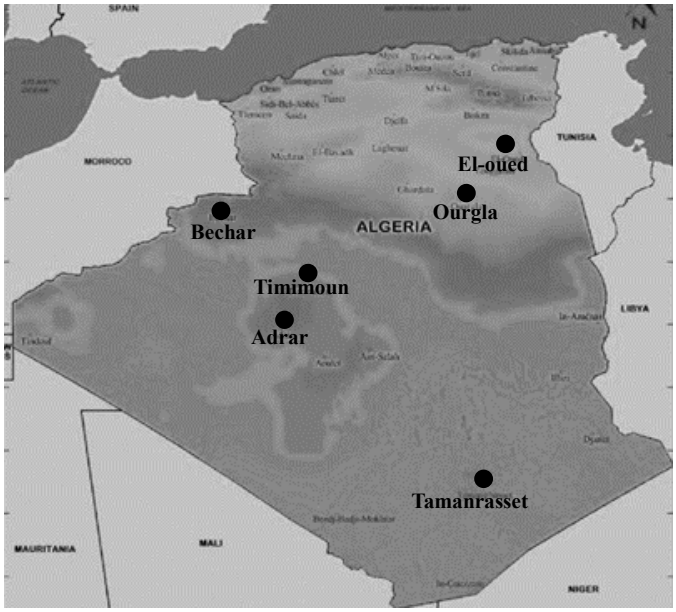


Figure IV.**1.** shows the locations of every site.

The selected provinces and their locations are listed in Table IV.2.

Table IV.**2.** Coordinates of the location.

Provinces	Latitude (°)	Longitude (°)	Altitude (m)
Adrar	26.489	-1.358	286
El-oued	33.512	6.783	63
Ouargla	30.998	6.766	178
Tamanrasset	24.335	4.455	810
Timimoun	30.024	0.849	431
Bechar	31.386	-2.012	785

We collect an enormous amount of data—enough to create a representative dataset. The datasets used span an eleven-year period (from January 1, 2010, to December 31, 2021). They are composed of nine factors, which are relevant to the aforementioned study topics and are

shown in Table IV.2. The information came from the SODA (Simple Ocean Data Assimilation) database [26].

Table IV.**3.** The Parameters used.

Parameters	Abbreviation	Unit	Category	Type
The Year	Y	/	Meteorological	Numerical
Day of the Year	D	/		
Wind speed	WS	m/s		
Atmospheric pressure	BP	hPa		
Average Temperature	T_{avg}	°C		
Relative humidity	RH	%		
Declination	DE	Degree (°)	Astronomical	
Hour Angle	HA	Degree (°)		
Extraterrestrial solar irradiation	H_0	Wh/m^2		

To forecast DGSR, we employed meteorological and astronomical factors. The datasets are divided into two subgroups: 20% are put aside for testing and validation, while the remaining 80% are used for sample training the model.

IV.3.2 *Methods and models*

IV.3.2.1. ANN

An artificial neural network (ANN) is an abstract computing method that mimics brain activity [27]. The Wij sum of the input signal yi is used by each neuron in the i th hidden layers to determine its weight. This weight is then used to a nonlinear activation function to produce the output signal uj for each hidden layer. The structure of this function is:

$$uj = \sum_{i=0}^{n} Wij\, yi \qquad \text{(IV.1)}$$

For the purpose of forecasting solar radiation, the multilayer feed-forward neural network (MLF) with the back-propagation algorithm (BP) has been used most frequently [28]. Problems that are nonlinearly separable are represented using this method. MLF is made up of an input layer, several hidden layers, and an output layer at the end. Each neuron contributes a threshold term or a bias to the total before nonlinearity shifts the sum to form an output. Weights Wjk with Wij link each layer.

This nonlinear transition is referred to as the node's activation function. Logistic sigmoid (Eq. 2) and linear functions (Eq. 3) are commonly employed by the hidden and output layers in MLFs, respectively [29]:

$$f(w) = \frac{1}{1+e^{-w}} \quad \text{(IV.2)}$$

$$f(x) = x \quad \text{(IV.3)}$$

The weighted total of the input is shown by w, and the input to the output layer is represented by x.

BP is the term used to describe an error that is computed at the output layer and propagated backward to the input layer [30].

IV.3.2.2. ANN architecture

The MLP design and structure of the ANN multilayer neural network are generally influenced by the database (input and output). Five hidden layers are employed in this model to estimate the daily solar radiation. The produced ANN's features are listed in Table IV.4 and Figure IV.2.

Table IV.4**.** The characteristics of ANN.

Activation function		Traning algorithm
Hidden layer: Tansig	Output layer: Pureline	Trainlm

For every model, we used a script file produced by the MATLAB software in each of the six cities we chose.

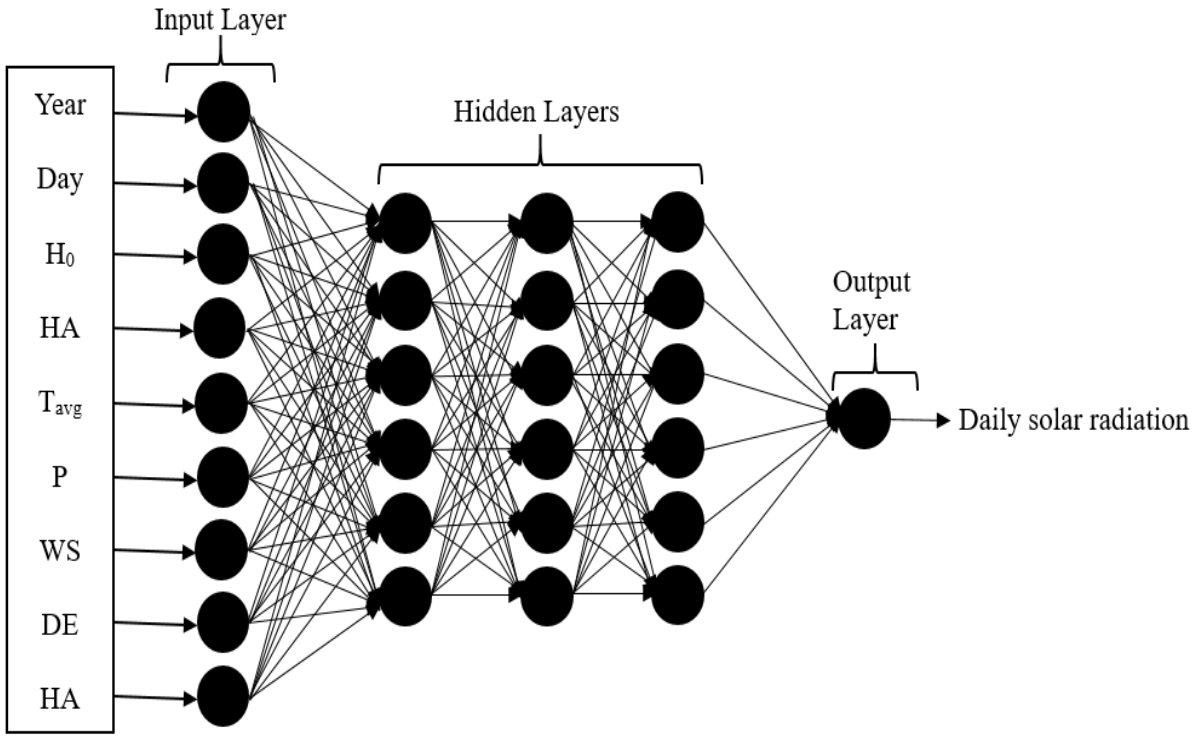


Figure IV.2**.** The model based on ANNs

IV.3.2.3. The SVM model

MATLAB 2021a was used to optimize and train the SVMs utilizing the leaner app regression. After a number of trial-and-error experiments, the cubic and linear kernels were determined to be the best in this study. SVM hypothesis states that the slack parameter (ε) determines SVM performance.

In this work, the parameter (ε) was optimized with 30 iterations of the Bayesian algorithm, changing the values of ε from 1.7529 to 259485.0259. An overall solution can be reached by repeating the following two steps: (i) simulating a surrogate function; (ii) determining where to query next by maximizing an acquisition function. This involves adding up a few query points together with their function value.

Convergence guarantees only hold when the acquisition function's overall optimizer is identified and selected as the next query point in each round. Nevertheless, local optimizers for the acquisition function are also used in practice because finding an overall optimizer can occasionally be challenging and time-consuming. Table IV.5 displays the optimum values of the parameter ε for the SVM models.

Table IV.5**.** Parameters optimized for SVM

Station	Kernel function	Optimum values ε
Adrar	Cubic	2.0255
El-oued	Cubic	24.6855
Ouargla	Linear	2.682
Tamanrasset	Cubic	16.3562
Timimoun	Cubic	7.6419
Bechar	Linear	5.4471

IV.3.2.4. FFA

Xin-She Yang first suggested FFA in late 2007 and early 2008 [32, 33]. It is known as an algorithm for metaheuristics based on evolutionary populations. FFA draws inspiration from swarm behavior in nature, including that of insects, birds, and schools of fish. FFA is a beneficial algorithm that outperforms other metaheuristic algorithms, such genetic algorithms, according to a plethora of recent study [32–35]. FFA, which gets its inspiration from actual fireflies, has three flashing characteristics and idealized rules that can be summed up as follows:

1. Fireflies engage with different fireflies regardless of gender or sex since the insects are all unisex.

2. The attraction of a firefly decreases with increasing separation from another firefly and is inversely connected with brightness. A less brilliant firefly moves closer to the brighter one. The firefly moves at random when there isn't a brighter firefly nearby.

3. The objective function's value determines how brilliant the firefly is.

A firefly's attraction decreases with distance (r) from its source (another firefly) while employing the FFA because light is absorbed by the surroundings. The brightness $I(r)$ variations are: is:

$$I(r) = \frac{I_0}{r^2} \quad \text{(IV. 4)}$$

The brightness of the source is I_0. Equation (IV.8) can be used to approximate the cumulative influence of absorption to a Gaussian form with the fixed light absorption coefficient γ, preventing singularity at r = 0.

$$I(r) = I_0 e^{-\gamma r^2} \quad \text{(IV. 5)}$$

Given that the attractiveness of a firefly is directly correlated with light intensity, its attractiveness function can be described as

$$B(r) = B_0 e^{-\gamma r^2} \quad \text{(IV. 6)}$$

At r = 0, the initial attraction is represented by B_0

As demonstrated in [34], [32], and [36], the distance r_{ij} between two firefly positions, x_i and x_j, can be expressed as either a Cartesian or Euclidian distance; that is

$$r_{ij} = \|x_i - x_j\| = \sqrt{\sum_{d=1}^{D} (x_{i,d} - x_{j,d})^2} \quad \text{(IV. 7)}$$

D is the total number of dimensions; k is the d th component of the spatial coordinates x_i with respect to the i^{th} firefly.

The motion of firefly i toward firefly j can be explained as follows when firefly j is more luminous than firefly i:

$$x_i = x_i + \beta_0 * exp(-\gamma r_{ij}^2)(x_i - x_j) + \alpha * (rand - 0.5) \quad \text{(IV. 8)}$$

The first phrase describes the firefly's current location; the second describes its brightness-based attractiveness to nearby fireflies; and the third describes the firefly's irregular movement when there isn't a brighter firefly around. Both a randomization parameter (α) and a random number (rand) have values equal to $\in[0,1]$.

In the FFA, absorption coefficient γ results in two distinct scenarios:

When $\gamma = \infty$, the fireflies exhibit minimal attraction to brightness and are nearly undetectable to one another. As such, the FFA operates in a manner akin to a random walk technique.

On the other hand, when γ = 0, the brightness does not decrease with increasing distance (r) and the attraction coefficient stays constant (β = β0). In this case, the FFA and the PSO approach are consistent.

IV.3.2.5. The Hybrid FFA-ANN model

The fundamental steps of the FFA are as follows:

- Step 1: First, the FFA sets the most important parameters (β0, γ, α) at their initial values. It also sets the population size P and the maximum generation number (MGN), which is the FFA's termination criterion (see Table IV.6).
- Step 2: We establish the primary population, $xi, i = \{1, \ldots, P\}$, at random. We then compute the associated objective function for each solution $f(xi)$ in the population to determine its fitness.
- Step 3: To fulfill the termination condition, the following steps are repeated until the necessary number of MGN iterations is reached.

Selecting the ideal subset of input variables is difficult. The obvious way to select this subset is to systematically test the fit of each prospective subgroup. While this exhaustive search guarantees the optimal result for a specific ANN model, it requires the investigation of (2d -1) combinations and thus requires a large processing burden. The choice of the ANN's inputs determines how well it can predict sun radiation.

The ANN model in this work was created and trained with several combinations of inputs (see to Figure IV.3).

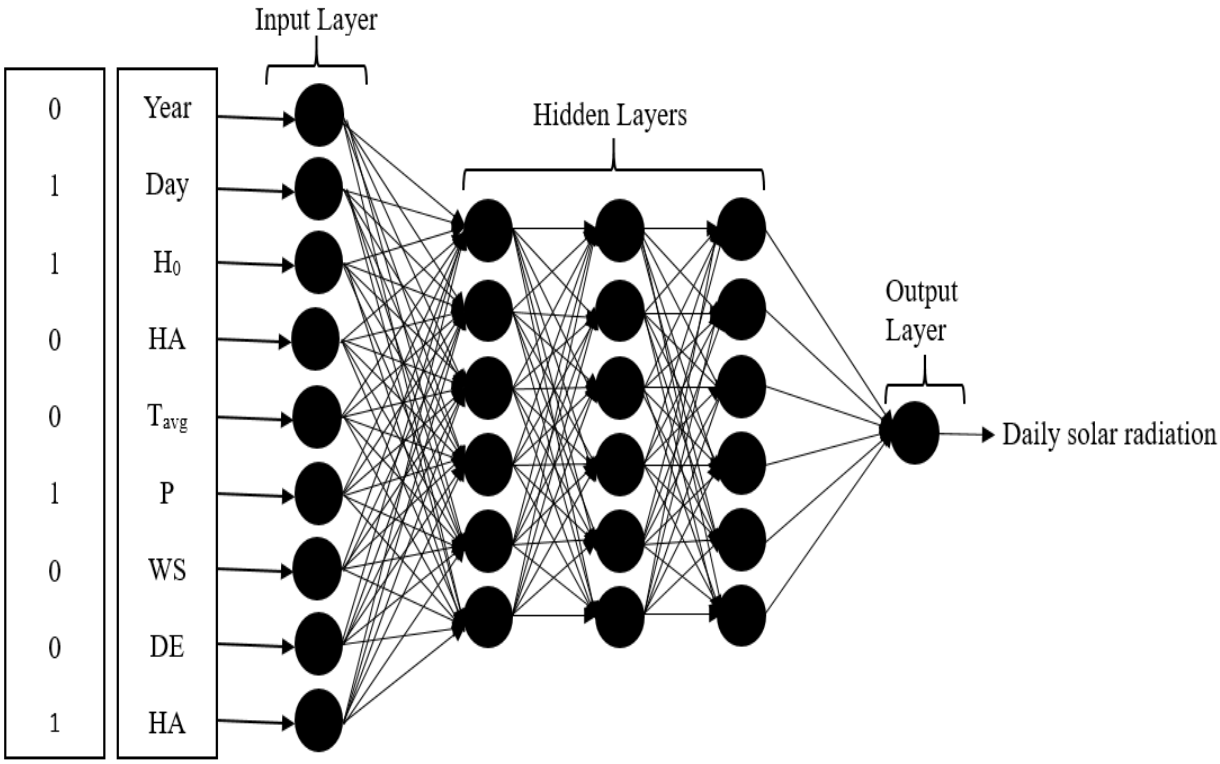


Figure IV.**3. The** FFA-ANN model.

In order to train and test the ANN model, we use a hybrid machine learning approach in our proposed model that is based on the FFA-ANN. This algorithm favors choosing the small and most relevant set of inputs with strong correlations to sun irradiation.

A paradigm of evolution based on the FFA is automatically employed to select the most pertinent inputs.

Table IV.6**.** The parameters of the Firefly algorithm.

Description	Population size	Maximum generation	A	γ	Delta	β_0
value	10	50	1	0.6	0.8	1

IV.3.3. Evaluation metrics

The effectiveness of the models was investigated using metrics that are frequently employed in evaluation results [37]. Table IV.7 below provides the definitions for these indices, where N is the total number of observations, G_{Sim} is the anticipated value, and G_{Act} is the actual value.

Table IV.7. Evaluation metrics.

Indice	Optimal value	Equation	
rRMSE %	0	$= \frac{1}{\bar{G}_{Act}} \sqrt{\frac{\sum_{i=1}^{N}(G_{Sim,i} - G_{Act,i})^2}{N}}$	(III.28)
MAPE %	0	$= \frac{100}{N} \sum_{i=1}^{N} \left\| \frac{G_{Sim,i} - G_{Act,i}}{X_{Act,i}} \right\|$	(III.27)
R	1	$= \frac{\sum_{i=1}^{N}(G_{Sim,i} - \bar{G}_{Sim})(G_{Act,i} - \bar{G}_{Act})}{\sqrt{\sum_{i=1}^{N}(G_{Sim,i} - \bar{G}_{Sim})^2} \times \sqrt{\sum_{i=1}^{N}(G_{Act,i} - \bar{G}_{Act})^2}}$	(IV.9)

IV.4. Findings and conversation

Six different locations in Algeria were given daily sun radiation forecasts using the ANN, FFA-ANN, and SVM models. The statistical performance indicators for each model are shown in Table IV.8.

The results clearly show that the proposed hybrid FFA-ANN approach gave very high accuracy and performed very well for the following cities: Ardar (MAPE = 6.62; rRMSE = 10.10; R = 0.911); Ouargla (MAPE = 7.9; rRMSE = 11.47; R = 0.923); Tammanrasset (MAPE = 6.29; rRMSE = 9.35; R = 0.905); Bechar (MAE = 8.23; rRMSE = 11.76; R = 0.920); EL-Oued (MAPE = 8.26; rRMSE = 11.96; R = 0.9321); Timimoun (MAPE = 6.88; rRMSE = 10.48; R = 0.928).

Table IV.**8.** Statistical indices

		MAPE%	rRMSE%	R
Adrar	ANN	7.37	10.81	0.895
	FFA-ANN	6.62	10.01	0.911
	SVM	6.09	10.54	0.908
El oued	ANN	9.49	13.17	0.917
	FFA-ANN	8.26	11.96	0.9321
	SVM	7.41	12.43	0.9320
Ouargla	ANN	8.02	11.84	0.918
	FFA-ANN	7.9	11.47	0.923
	SVM	8.96	13.37	0.898
Tamanrasset	ANN	6.91	10.01	0.890
	FFA-ANN	6.29	9.35	0.905
	SVM	5.79	9.77	0.903
Timimoun	ANN	7.87	11.59	0.911
	FFA-ANN	6.88	10.48	0.928
	SVM	6.21	11.05	0.926
Bechar	ANN	8.38	12.06	0.915
	FFA-ANN	8.23	11.76	0.920
	SVM	9.73	13.99	0.888

Table IV.9 and Figure 4 make it evident that the FFA-ANN hybrid technique performed better than the stand-alone ANN. The ideal configuration for every site when utilizing the ANN approach is displayed in Table IV.9.

Table IV.**9.** Optimal combination of FFA-ANN.

Location	The best combination [Y, D, H_0, DE, T_{avg}, RH, BP, HA, WS]
Adrar	[0.0.1.1.1.1.1.0.1]
El-oued	[0.0.1.1.1.1.1.0.1]
Ouargla	[0.0.1.1.1.1.1.1.0]
Tamanrasset	[0.0.1.1.1.1.1.1.1]
Timimoun	[0.0.1.1.1.1.1.0.1]
Bechar	[0.0.1.1.1.1.0.0.0]

Based on Table IV.8 's results, we can infer that every city has a unique optimal combination of input data. Nonetheless, when we observe certain inputs' involvement in every combination for every city, we become aware of their significance for the ANN model. These are the inputs: RH, Tavg, DE, and H0.

The accuracy of the model was not always improved when the same inputs were used for the SVM model, hence the best combinations for the ANN model might not always be the best combinations for the SVM model. Variations in the optimum feature set can be caused by aspects of a model's unique qualities as well as the intricacy of the modeling techniques.

A hybrid FFA-ANN technique was created for this study in order to quantify global sun radiation. The two primary statistical output metrics, MAPE and rRMSE, were employed to compare and assess the performances of our suggested model. Table IV. 8 and Figure IV.4 present the results of the overall average statistical analysis of rRMSE and MAPE for both models that were considered.

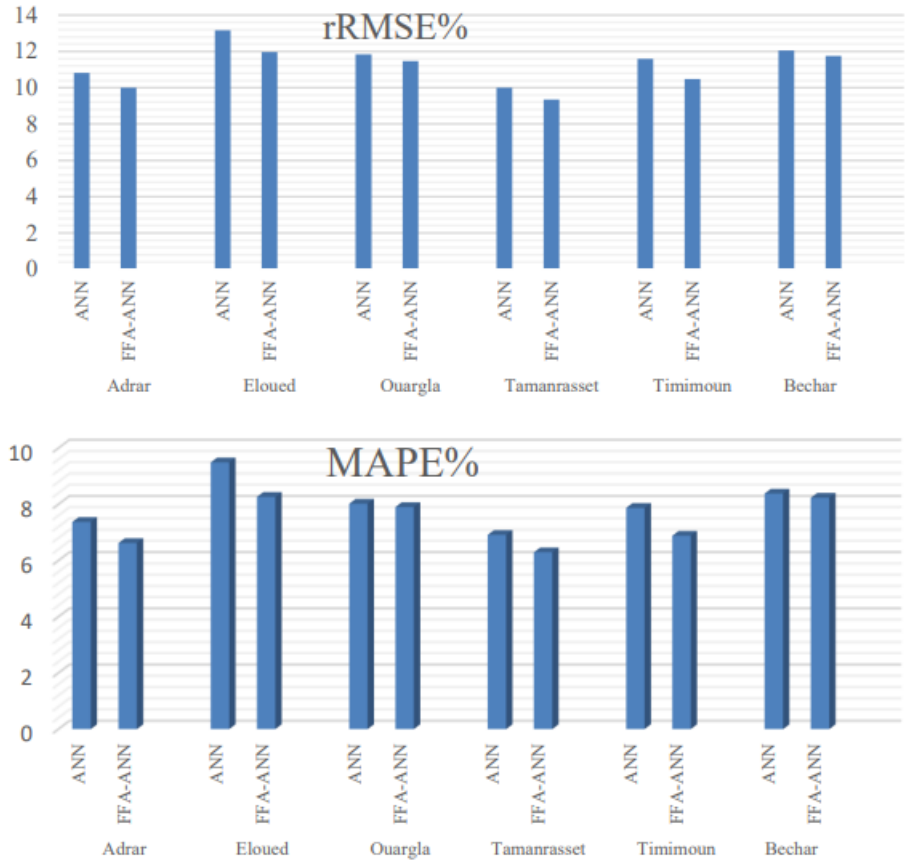


Figure IV.**4.** The FFA-ANN and ANN's MAPE (%) and rRMSE (%).

The results presented in both figures indicate that the FFA-ANN hybrid model beat the stand-alone ANN model when the suggested firefly algorithm was used to select the optimal combination ANN model.

Figure IV.5 displays the expected and actual values of the most accurate model, FFA-ANN, over the course of the study. Because of its great accuracy, this model is suggested for ANN model optimization and solar radiation regression. Its benefit is that it provides the optimal combination for every place, making it easier and less time-consuming to collect the necessary data in the future.

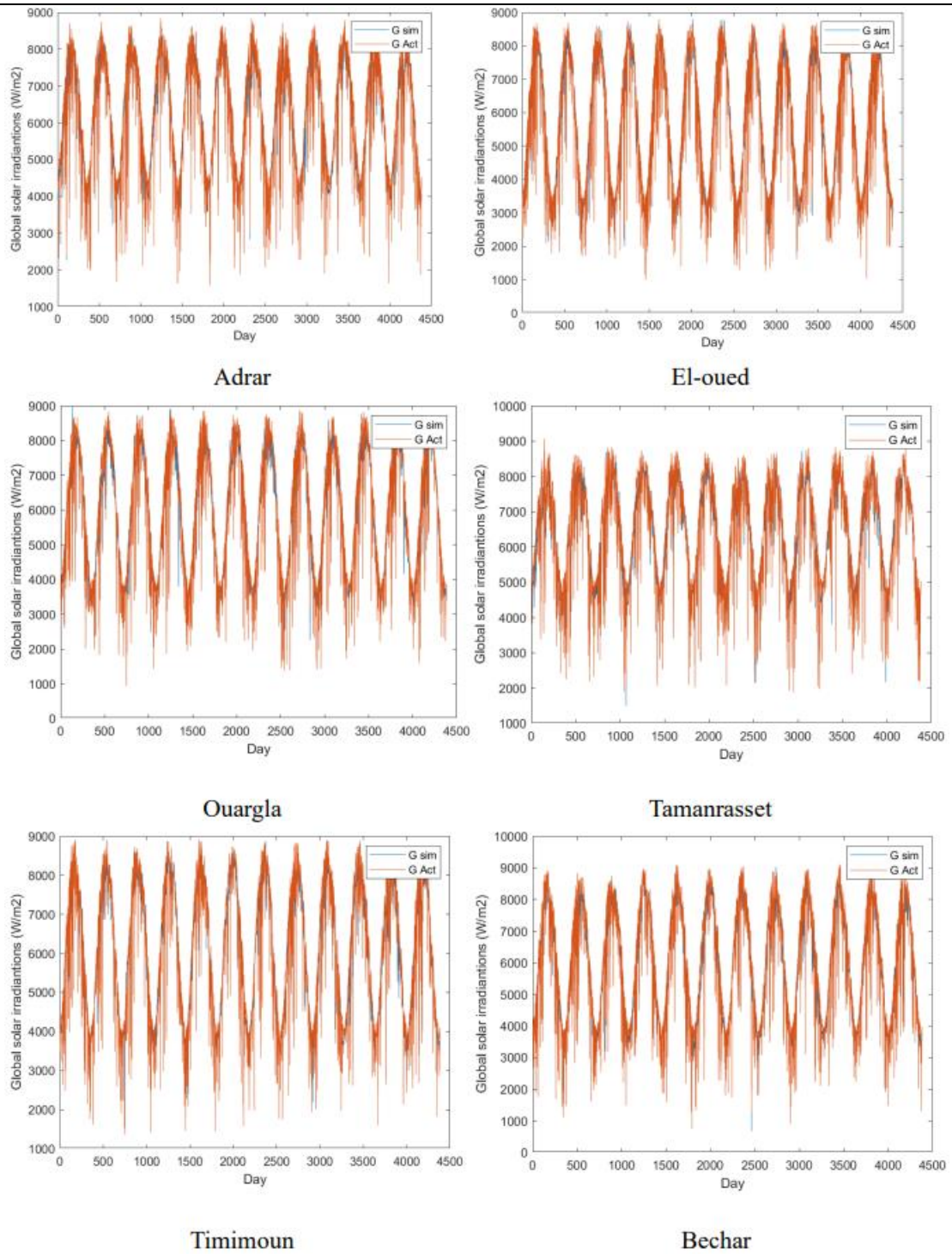


Figure IV.**5.** Forecasted and gathered values for the FFA-ANN approach.

The relationship between the projected and actual values across the six chosen cities during the study was also displayed by scatter plots (see Figure IV.6) of the FFA-ANN models, which

predicted DGSR and displayed the forecasted data distributed as a series of points close to the linear (red) ideal fit.

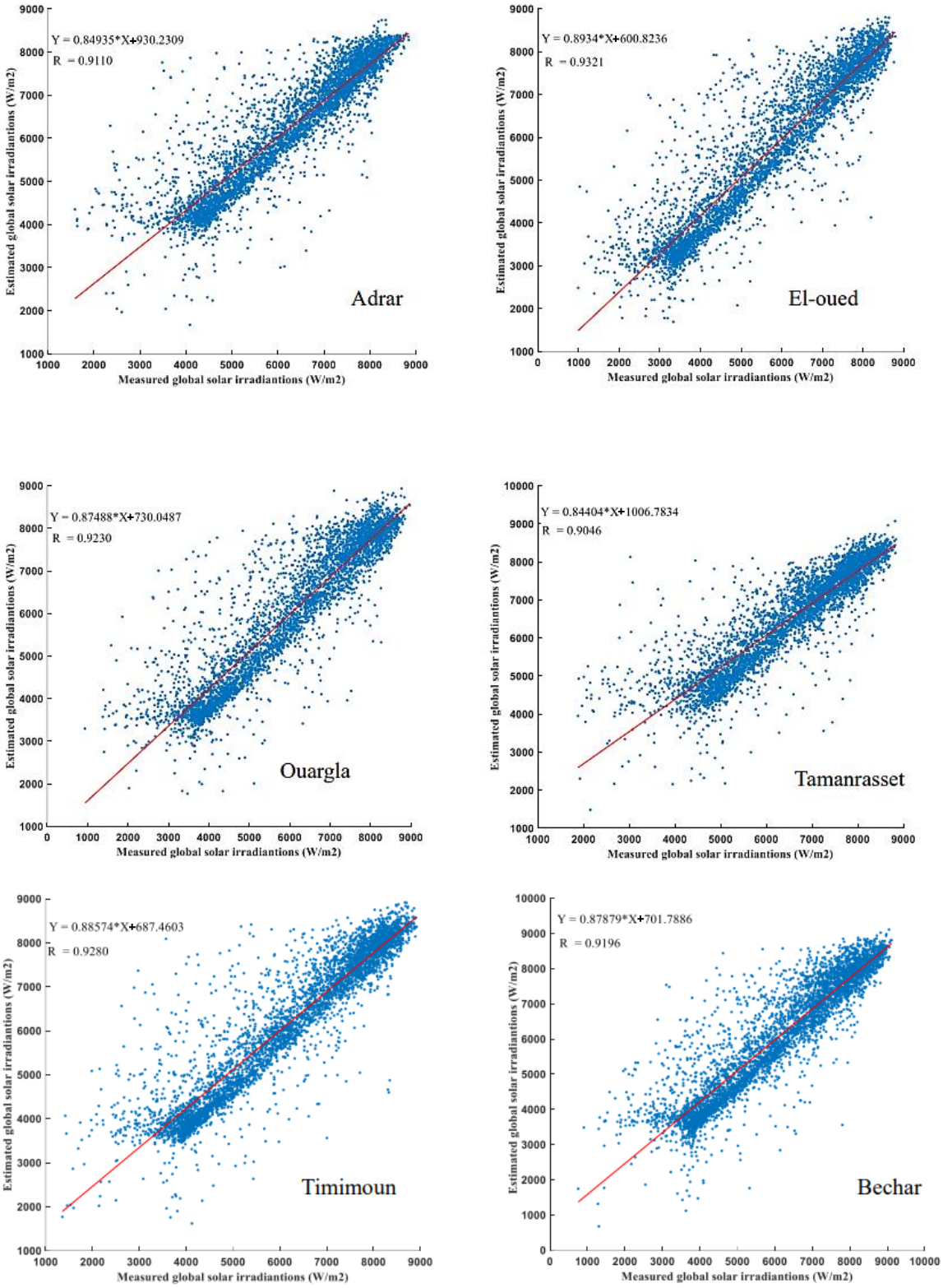


Figure IV.**6.** FFA-ANN models' scatter plot.

Because FFA-ANN requires several runs of the ANN model, which results in a considerably longer running time, it has a greater computational cost than typical machine learning (ML) models with hyperparameter optimization, such as stand-alone ANN and SVM. But in terms of simulation efficacy, typical ML models perform well, but FFA-ANN is better at capturing intricate interactions, which results in the best accuracy.

Table IV.10 displays several models used to predict sun irradiation. The proposed hybrid FFA-ANN model is evaluated with neural networks and empirical techniques. The dependability of the predictions made by the previously described and current approaches is assessed using three statistical index errors: R, rRMSE, and MAPE. The results of the suggested inquiry show high and compatible effectiveness in terms of predictability and improved forecasting capabilities when compared to the other strategies. The outcomes are similar amongst them.

Table IV.**10.** Statistics assess errors across the different models.

Reference	Model	Time step	Location	R	rRMSE%	MAPE%
[38]	ANN	Daily	Nevs¸ehir city, Turkey	0.965	14.10	15.92
[38]	k-NN	Daily	Karaman city, Turkey	0.941	25.19	30.24
[38]	SVM	Daily	Tokat city, Turkey	0.938	18.43	23.37
[39]	Empirical	Daily	Tamanrasset city, Algeria	0.891	-	-
[40]	ANN-FFA	Hourly	Klang Valley, Malaysia	-	23.47	13.00
The presented Model	FFA-ANN	Daily	ELoued city, Algeria	0.9321	11.96	8.26

IV.5. Conclusions

In this chapter, we conducted a comprehensive study focusing on six cities in southern Algeria: Adrar, El-Oued, Ouargla, Timimoun, Tamanrasset, and Bechar. Our objective was to examine the accuracy of three distinct models— (ANN), (FFA-ANN), and (SVM)—in estimating (DGSR) over an eleven-year period spanning from 2010 to 2021. We utilized 9 input parameters as input data for our analysis.

Our proposed hybrid FFA-ANN model was developed with the primary aim of optimizing the ANN model created by our team. Despite obtaining promising results from both the ANN and SVM models, our findings indicated that the FFA-ANN hybrid method consistently outperformed the standalone ANN-based model across various statistical indicators, (R), (rRMSE), and (MAPE).

Furthermore, our investigation revealed significant insights into the input parameters influencing the ANN model's forecasting accuracy for DGSR. Specifically, the model exhibited a reliance on variables such as (H_0), (DE), (Tavg), and (RH) when predicting DGSR.

As a consequence, our study suggests that the FFA-ANN model developed holds promise as a reliable forecasting tool for DGSR in regions characterized by dry climates, such as southern Algeria. Additionally, the model's performance highlights its potential utility in guiding decisions related to the installation and optimization of solar energy systems.

IV.6. References

1. Kabouche N, Chellali F, Recioui A (2021) A review on solar radiation assessment and forecasting in Algeria: (Part 2; Solar radiation forecasting). Algerian J Sign Syst 6: 130–146. https://doi.org/10.51485/ajss.v6i3.141
2. Gairaa K, Khellaf A, Messlem Y, et al. (2016) Estimation of the daily global solar radiation based on Box–Jenkins and ANN models: A combined approach. Renew Sustain Energy Rev 57: 238–249. https://doi.org/10.1016/j.rser.2015.12.111
3. Benatiallah D, Benatiallah A, Bouchouicha K, et al. (2020) Prediction of hourly solar radiation using artificial neural networks. Algerian J Env Sci Technol 6: 1236–1245. Available from: https://www.aljest.net/index.php/aljest/article/view/272.
4. Antonopoulos VZ, Papamichail DM, Aschonitis VG, et al. (2019) Solar radiation estimation methods using ANN and empirical models. Comp Electron Agric 160: 160–167. https://doi.org/10.1016/j.compag.2019.03.022
5. Amiri B, Gómez-Orellana AM, Gutiérrez PA, et al. (2020) A novel approach for global solar irradiation forecasting on tilted plane using hybrid evolutionary neural networks. J Clean Produc 287: 125577. https://doi.org/10.1016/j.jclepro.2020.125577
6. B. Güzel, O. Sevli, and E. Okatan, "Forecasting Solar Radiation Based on Meteorological Data Using Machine Learning Techniques: A Case Study of Isparta Province," Uluslararası Mühendislik Araştırma ve Geliştirme Dergisi / Inter J. Eng Res and Dev, vol. 15, no. 2, pp. 704-713, Jun. 2023. doi: 10.29137/umagd.1268055
7. S. Pereira, P. Canhoto, and R. Salgado, "Development and assessment of artificial neural network models for direct normal solar irradiance forecasting using operational numerical weather prediction data," Energy and AI, vol. 15, p. 100314, Nov. 2023. doi: 10.1016/j.egyai.2023.100314
8. Voyant C, Muselli M, Paoli C, et al. (2011) Optimization of an artificial neural network dedicated to the multivariate forecasting of daily global radiation. Energy 36: 348–359 https://doi.org/10.1016/j.energy.2010.10.032
9. Hasni A, Sehli A, Draoui B, et al. (2012) Estimating global solar radiation using artificial neural network and climate data in the south-western region of Algeria. Energy Proc 18: 531–537. https://doi.org/10.1016/j.egypro.2012.05.064
10. M. Sari and M. Arici, "Global Solar Radiation Forecasting with Artificial Neural Networks," in Green Energy, Environment and Sustainable Development,vol 1, pp 275- 285, 2023. doi:10.3233/ATDE230300
11. Mohammadi K, Shamshirband S, Danesh AS, et al. (2020) Retracted article: Horizontal global solar radiation estimation using hybrid SVM-firefly and SVM-wavelet algorithms: A case study. Natural Hazards 102: 1613–1614. https://doi.org/10.1007/s11069-015-2047-5

12. Guermoui M, Rabehi A, Gairaa K, et al. (2018) Support vector regression methodology for estimating global solar radiation in Algeria. European Phys J Plus 133: 22. https://doi.org/10.1140/epjp/i2018-11845-y
13. Guijo-Rubio D, Durán-Rosal AM, Gutiérrez PA, et al. (2020) Evolutionary artificial neural networks for accurate solar radiation prediction. Energy 210: 118374. https://doi.org/10.1016/j.energy.2020.118374
14. Belaid S, Mellit A (2016) Prediction of daily and mean monthly global solar radiation using support vector machine in an arid climate. Energy Conver Manag 118: 105–118. https://doi.org/10.1016/j.enconman.2016.03.082
15. Shamshirband S, Mohammadi K, Khorasanizadeh H, et al. (2016) Estimating the diffuse solar radiation using a coupled support vector machine–wavelet transform model. Renew Sustain Energy Rev 56: 428–435. https://doi.org/10.1016/j.rser.2015.11.055
16. Bakhashwain JM (2016) Prediction of global solar radiation using support vector machines. Int J Green Energy 13: 1467–1472. https://doi.org/10.1080/15435075.2014.896256
17. Baser F, Demirhan H (2017) A fuzzy regression with support vector machine approach to the estimation of horizontal global solar radiation. Energy 123: 229–240 https://doi.org/10.1016/j.energy.2017.02.008
18. Queja VH, Almoroxa J, Arnaldob JA, et al. (2017) ANFIS, SVM and ANN soft-computing techniques to estimate daily global solar radiation in a warm sub-humid environment. J Atmos Sol-Terrestrial Physic 155: 62–70. https://doi.org/10.1016/j.jastp.2017.02.002
19. Bhola P, Bhardwaj S (2019) Estimation of solar radiation using support vector regression. J Inf Optim Sci 40: 339–350. https://doi.org/10.1080/02522667.2019.1578093
20. Fan J, Wu L, Ma X, et al. (2019) Hybrid support vector machines with heuristic algorithms for prediction of daily diffuse solar radiation in airpolluted regions. Renew Energy 145: 2034–2045. https://doi.org/10.1016/j.renene.2019.07.104
21. Liu Y, Zhou Y, Chen Y, et al. (2020) Comparison of support vector machine and copula-based nonlinear quantile regression for estimating the daily diffuse solar radiation: A case study in China. Renew Energy 146: 1101–1112. https://doi.org/10.1016/j.renene.2019.07.053
22. Shamshirband S, Mohammadi K, Tong CW, et al. (2016) Retracted article: A hybrid SVM-FFA method for prediction of monthly mean global solar radiation. Theor Appl Climatol 125: 53–65. https://doi.org/10.1007/s00704-015-1482-2
23. Olatomiwa L, Mekhilef S, Shamshirband S, et al. (2015) A support vector machine–firefly algorithm-based model for global solar radiation prediction. Sol Energy 115: 632–644. https://doi.org/10.1016/j.solener.2015.03.015
24. Benatiallah D, Bouchouicha K, Benatiallah A, et al. (2019) Forecasting of solar radiation using an empirical model. Algerian J Renewable Energy Sustainable Dev 1: 212–219.

https://doi.org/10.46657/ajresd.2019.1.2.11

25. Benatiallah D, Benatiallah A, Harouz A, et al. (2016) Development and modeling of a geographic information system solar flux in Adrar, Algeria. Int J Syst Mod Simul 1: 15–19. Available from: https://www.aljest.net/index.php/aljest/article/download/524/500.
26. SODA data. Available from: www.soda-pro.com/web-services#meteodata.
27. Haykin S, Lippmann R (1994) Neural networks, a comprehensive foundation. Int J Neural Syst 5: 363–364. https://doi.org/10.1142/S0129065794000372
28. Yadav AK, Chandel SS (2014) Solar radiation prediction using artificial neural network techniques: A review. Renew Sustain Energy Rev 33: 772–781. http://doi.org/10.1016/j.rser.2013.08.055
29. Ata R (2015) Artificial neural networks applications in wind energy systems: A review. Renew Sustain Energy Rev 49: 534–562. https://doi.org/10.1016/j.rser.2015.04.166
30. Esmaeelzadeh SR, Adib A, Alahdin S (2015) Long-term streamflow forecasts by adaptive neuro-fuzzy inference system using satellite images and K-fold crossvalidation (case study: Dez, Iran). KSCE J Civ Eng 19: 2298–2306. https://doi.org/10.1007/s12205-014-0105-2
31. Vapnik V (2013) The nature of statistical learning theory. Springer Sc & Bus Media, Berlin, Heidelberg, Germany 267–287. Available from: https://statisticalsupportandresearch.files.wordpress.com/2017/05/vladimir-vapnik-the-nature-of-statistical-learning-springer-2010.pdf
32. Yang XS (2008) Nature-inspired metaheuristic algorithms. Luniver Press, UK 81–89. Available from: https://www.academia.edu/457296/Nature_inspired_metaheuristic_algorithms
33. Chantasut N, Charoenjit C, Tanprasert C (2004) Predictive mining of rainfall predictions using artificial neural networks for chao phraya river. 4th Inter Conf of The Asian Feder of Info Tech in Agriculture and The 2nd World Cong on Comp in Agriculture and Natural Res, Bangkok, Thailand 117–122. Available from: https://www.thaiscience.info/Journals/Article/NETC/10438493.pdf
34. Lukasik S, Zak S (2009) Firefly algorithm for continuous constrained optimization tasks. In Proceedings of the Inter Conf on Comp and Comput Intelligence (ICCCI 09). Springer, Wroclaw, Poland 5796: 97–106, Available from: https://link.springer.com/chapter/10.1007/978-3-642-04441-0_8
35. Yang XS (2010) Firefly algorithm, stochastic test functions and design optimization, Int J Bio-Inspired Comp 2: 78–84. https://doi.org/10.1504/IJBIC.2010.032124
36. Yang XS (2009) Firefly algorithms for multimodal optimisation, Proc. 5th Sympo on Stoch Algo, Found and Appl, (Eds. O. Watanabe and T. Zeugmann), Lect Notes in Comp Sc 5792: 169–178. https://doi.org/10.1007/978-3-642-04944-6_14
37. Stone RJ (1993) Improved statistical procedure for the evaluation of solar radiation

estimation models. Sol Energy 89: 51–91. https://doi.org/10.1016/0038-092X(93)90124-7

38. Agbulut Ü, Etem Gürel A, Biçen Y (2021) Prediction of daily global solar radiation using different machine learning algorithms: Evaluation and comparison. Renew Sustain Energy Rev 135: 110114. https://doi.org/10.1016/j.rser.2020.110114
39. Ihaddadene N, El Hacen Ould Ahmedou M, Jed B, et al. (2019) Daily global solar radiation estimation based on air temperature: Case of study south of Algeria. E3S Web of Conf 80: 01002. https://doi.org/10.1051/e3sconf/20198001002
40. Anwar Ibrahim I, Khatib T (2017) A novel hybrid model for hourly global solar radiation prediction using random forests technique and firefly algorithm. Energy Conv Manag 138: 413–425. https://doi.org/10.1016/j.enconman.2017.02.006

Chapter V : Solar Radiation Estimation Based On A New GA-ANN model

Chapter V: Solar Radiation Estimation Based On A New GA-ANN model

V. 1. Introduction

Emerging as a viable renewable energy source, solar power provides a sustainable substitute for conventional fossil fuels. For solar energy systems to be designed and operated as efficiently as possible, global solar radiation (GSR) must be estimated accurately. However, the investigation of machine learning models as a potential remedy was prompted by the fact that direct solar radiation measurement can be expensive and complicated.

Because Artificial Neural Networks (ANNs) can handle complex interactions with few input parameters, they have become popular for solar radiation prediction. Furthermore, hybrid models that combine fuzzy logic-based techniques like Adaptive Neuro-Fuzzy Inference Systems (ANFIS) and genetic algorithms (GAs) with ANNs have demonstrated potential in improving prediction accuracy.

The comparison of three different methods for predicting daily GSR in southern Algeria—ANN, GA-ANN, and ANFIS—is the main objective of this chapter. In particular, we assess how well each model captures the complex dynamics of solar radiation patterns. The hybrid GA-ANN model that has been suggested seeks to improve the predictive power of the ANN by utilizing GA-based optimization.

V.2. Literature review of the models used

Through a comprehensive analysis of these models, our study aims to shed light on their strengths and weaknesses in capturing solar radiation patterns. By evaluating their performance using meteorological data from selected research regions in Algeria's desert, we seek to provide valuable insights for optimizing solar energy system design and operation. Solar power is growing more appealing in this century, given the effects on the environment of using fossil fuels, which make things worse [1.] Research shows that green energy sources like solar and wind can effectively replace conventional energy sources to meet global demand without environmental harm. [2] Accurate global solar radiation (GSR)

data is crucial for optimizing solar power systems. However, due to high costs and measurement difficulties, this data is often unavailable in many countries. [3]

Machine learning (ML) models handle complex nonlinear issues and are widely used by experts globally. [4] Due to the stochastic nature of weather, solar energy from sunlight is not easily schedulable. [5] Evaluating solar radiation components is crucial for solar power implementation. [6] Predicting solar radiation is essential for large-scale and grid-connected solar energy systems to balance energy demand and supply.

Modern research indicates that many solar systems operate in an off-on mode during clear or cloudy weather, making it critical to predict solar radiation at various intervals. [5] Artificial neural networks (ANNs) are often used for solar radiation predictions due to their effectiveness with limited parameters. [7]

The adaptive neuro-fuzzy inference system (ANFIS) combines ANN and fuzzy inference. It reduces the search space by splitting a representation related to prior knowledge into subgroups, then adjusts fuzzy parameters using backpropagation. The result is an adaptive ANN functioning like a first-order Takagi-Sugeno inference system. [8] Genetic algorithms (GAs) are widely used for optimization in both machine learning and research. [9]

A Literature review of the models used is shown in Table V.1.

Table V.1: A Literature review of the models used.

Study	Location	Model(s)	Key Findings
[10]	China	MLP, GRNN, RBNN (ANN methods)	RBNN and MLP more precise in predicting solar radiation across different climate zones
[11]	Algeria	ARMA-ANN	Combined model (ARMA-ANN) outperforms other models in precision using GSR data
[12]	Ghardaia, Algeria	ANN	ANN model best for GSI estimation with four weather input factors; sunshine duration crucial for predictions
[13]	UAE	MLP, RBF (ANN methods)	MLP-ANN approach shows great accuracy with four weather-related variables
[14]	Adrar, Algeria	ANN	15 neurons in hidden layer using logistic sigmoid function useful for assessing GSR intensities
[15]	Greece	Empirical Equations, MLR, ANN	ANN models produce consistent results with MLR methods using similar input variables
[16]	Algeria	Multi-task Hybrid Evolutionary ANN	Sigmoidal units yield superior results; topology determined using evolutionary algorithm
[17]	North Egypt	ANFIS, Empirical Techniques	ANFIS model provides superior GSR estimations compared to empirical techniques
[18]	Irán	ANFIS	Sunshine duration significant input; ANFIS accurately determines daily Hd
[19]	China	ANFIS, ANN, Iqbal Models	ANFIS accurately estimates daily Hd using R2 and RMSE for comparison
[20]	Saudi Arabia	ANFIS	Subtractive clustering technique improves ANFIS model performance for GHI prediction
[21]	Algeria	ANFIS	Fifteen ANFIS-based models tested; Gaussian membership function used for GSR models
[22]	Singapore	Genetic Algorithm (GA)	GA-based multi-model framework outperforms alternative algorithms for solar radiation prediction
[23]	Turkey	Fuzzy Genetic (FG) Method	FG approach outperforms ANFIS and ANN in one-month forward sun radiation prediction
[24]	Iran	ANFIS optimized with PSO, Ant Colony, GA	All proposed methods accurately estimate global horizontal irradiance
[25]	China	GA, Deep Neural Networks	GA optimizes deep learning predictions by determining optimal window sizes and units
[26]	China	GA Optimized Wavelet Neural Networks	GAO-WNN accurately estimates daily sun radiation
[27]	China	GA, PSO, Back Propagation Neural Network (BPNN)	PSO enhances BPNN's capacity to accurately estimate daily Hd

The main goal of our study is to simulate daily global solar radiation (GSR) using a hybrid GA-ANN model. We aim to forecast GSR and compare its performance with two other methods: a stand-alone ANN and an ANN-based ANFIS. The models will use meteorological data from Adrar, Ourgla, and Bechar in Algeria's desert.

V.3. Material and Models

V.3.1 Study Areas and Data Collection

The study region is located in Algeria's desert (Adrar, Ouargla, and Bechar); the locations that are under investigation are listed in Table IV.2. The area features significant solar energy potential [28].

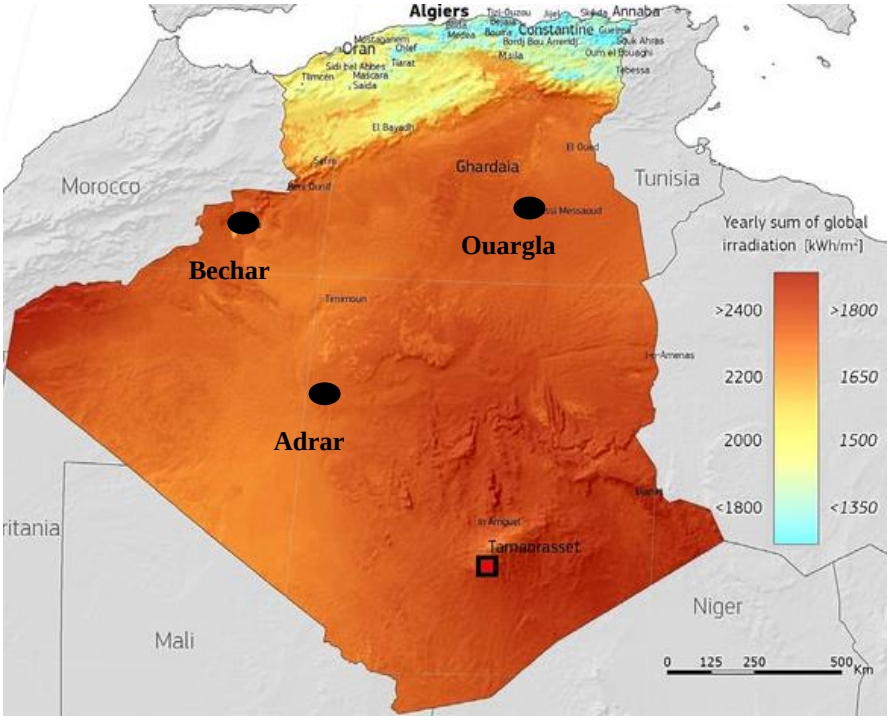


Figure V.**1:** Solar potential and sites locations **[29]**

We compiled a large database for testing and training our model using daily data from 2019 to 2021 from the SODA database [30]. With N = 1096 observations for each

variable—Day of the Year, Average Temperature, Relative Humidity, Declination, Hour Angle, and Extraterrestrial Solar Irradiation—20% of the data was reserved for testing, and the rest for training. These six inputs collectively provide crucial information on atmospheric characteristics, solar geometry, and environmental conditions, enabling accurate solar radiation forcasting at specific locations and times.

V.3.2 Data Preprocessing

By understanding past patterns, this technique reduces inadequate training issues and computation costs. Preprocessing inputs enhances the model's efficiency. Various strategies can be used; in our paper, we chose normalization [31]. This method selects a subset of inputs and limits data to a range of zero to one, reducing regression error, increasing precision, and maintaining data correlation.

$$I_{\text{norm}} = \frac{I_{\text{now}} - I_{\text{min}}}{I_{\text{max}} - I_{\text{min}}} \quad \text{(V.1)}$$

I_{norm} and ; I_{now} are normalized and original data respectively. I_{norm} lies between 0 and 1 based on the minimum I_{min} and maximum I_{max} values of the input data.

V.3.3. ANN Model

In our approach, we design an ANN to anticipate daily GSR. The characteristics are illustrated in Fig. V.2 and Table V.2.

Table V.2: Training Details and Features of the Model

Aspect	Description
Network Architecture	With a "tansig" activation function for the hidden layer, this feedforward neural network has five hidden layers, each with five neurons.
Data Preprocessing	The mapminmax function is used to apply min-max normalization to the input (P) and target (T) data.
Training Parameters	'Levenberg-Marquardt Backpropagation' is the training algorithm, with a maximum of 1000 epochs and a goal error of 0.000001.
Training Process	80% of the provided data was used for training, and the remaining 20% was used for testing.
Monitoring Training	A training record (tr), which tracks both training and testing errors, is used to track the progress of the training process.

In our case, we used one neuron in the output layer to represent the anticipated daily GSR for each of the three selected areas (see Fig. V.2). For all models, we used a script generated by MATLAB.

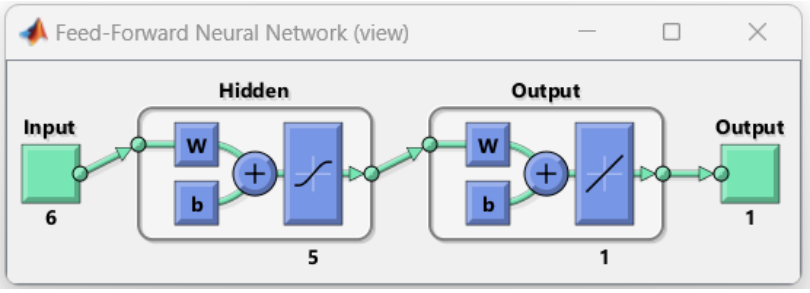


Figure V.**2:** Characteristics of our ANN

V.3.4 *ANFIS Model*

ANFIS combines ANN and fuzzy logic algorithms, with fuzzy variables processed by nodes across different feed-forward network layers, similar to distributed variable fuzzy inference systems (FIS) [32]. The ANFIS consists of 5 layers (see Table V.3 and Fig. V.3):

Table V.3: The five layers of the ANFIS model.

Layer	Description
1st Fuzzification	- Uses fuzzy C-means clustering for adaptive nodes - Generates membership grades for linguistic labels based on input values - Defines premise parameters using specific formulas. - Ensures membership functions confine outputs to the range of 0 to 1.
2nd Rule Activation	- Nodes calculate rule firing strength by multiplying membership grades. - Determines rule products for each pair of linguistic labels.
3rd Normalization	- Nodes normalize firing strengths from the previous layer.
4th Consequent	- Each node squares the normalized firing strength and combines it with resulting parameters.
5th Output	- Aggregates weighted products from the previous layer to obtain the final output [33].

We used Gaussian and Linear Membership Functions Types respectively

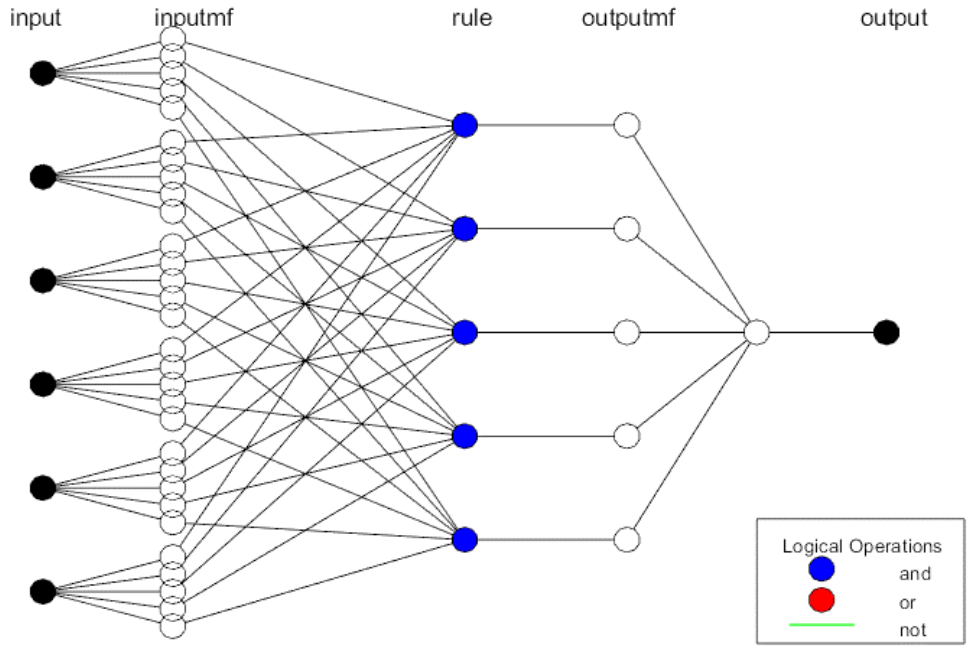


Figure V.3**:** Our ANFIS model.

V.3.5. ANN - *Genetic Algorithm*

These heuristic combinatorial search methods, known as genetic algorithms (GAs), aim to mimic biological evolution processes by simulating chromosome structures [34]. The GA process involves selecting chromosomal sequences to represent solutions and applying a fitness function to evaluate their suitability for the next generation.

Genetic algorithms differ from traditional optimization methods in several key ways [35]:

- GAs encode parameter sets rather than individual parameters.
- They use a population to explore the solution space rather than a single point.
- GAs rely on objective function evaluations rather than derivatives.
- They employ probabilistic evolution rules rather than deterministic rules.

The hybrid model's objective is to enhance ANN performance by initially training it with data and subsequently optimizing it using a GA, as depicted in Fig. V.4. This approach involves:

1. Generating training and testing data for the ANN.
2. Using a hybrid ANN-GA training function to refine the ANN's weights and biases through a combination of backpropagation and GA.
3. Applying the optimized ANN to make predictions on new input data.

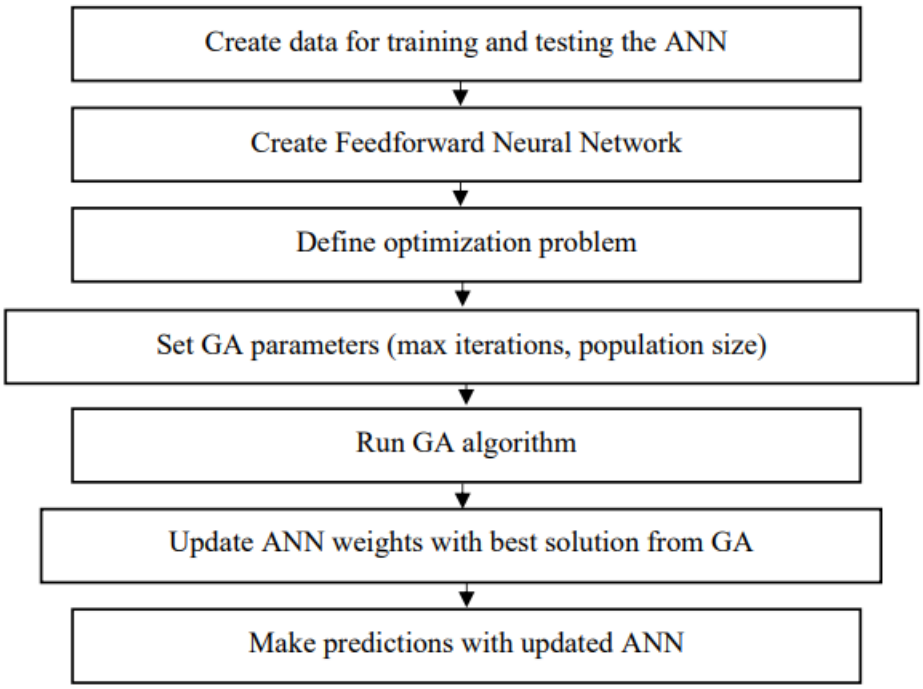


Figure V.4**:** The steps of GA-ANN approach

- The hybrid ANN-GA training function optimizes an ANN model using a genetic algorithm (GA). It defines an optimization problem with a cost function that minimizes the RMSE by adjusting weights and biases. The GA parameters, such as population size and iterations, are set to run the algorithm. The best solution updates the ANN weights for accurate predictions.

Table V.4: Parameters of ANN-GA method

Number of hidden layers	Max Iterations	Population size	Selection method	Crossover percentage	Mutation percentage
5	100	1000	Roulette wheel	0.6	0.4

V.4. Findings and discussions

We'll apply metrics (Equations III.28, III.27, and IV.13) from the previous chapter to assess how well our ANN, ANFIS, and GA-ANN models predicted daily GSR across three Algerian locations. Table V.5 summarizes key performance indicators. Results indicate the combined GA-ANN approach performed effectively with satisfactory precision.

Table V.5: Model's overall performance

City	Adrar			Ouargla			Bechar		
Model	ANN	ANFIS	GA-ANN	ANN	ANFIS	GA-ANN	ANN	ANFIS	GA-ANN
R	0.8909	0.9206	0.9008	0.8865	0.9304	0.9076	0.8839	0.9017	0.8877
rRMSE %	10.33	9.05	10.05	13.11	10.76	12.32	13.88	13.07	13.85
MAPE %	7.11	5.87	6.80	9.05	7.27	8.43	10.09	9.24	10.01

The ANN, ANFIS, and GA-ANN models were developed to predict daily GSR. Results from Table V.5 and Fig. V.5 demonstrate that combined approaches like GA-ANN

and ANFIS outperformed the standalone ANN model. Scatter plots (Fig. 7) reveal a tight clustering of predicted data points around the optimal fit line (in red), indicating strong correlation between predicted and observed values across the three cities studied.Statistical metrics show ANFIS as the most accurate model with an R value of 0.9206 in Adrar, followed closely by GA-ANN at 0.9008. In Bechar, ANFIS achieved 0.9017 compared to GA-ANN's 0.8839. In Ouargla, ANFIS led with 0.9304 while GA-ANN followed at 0.9076.

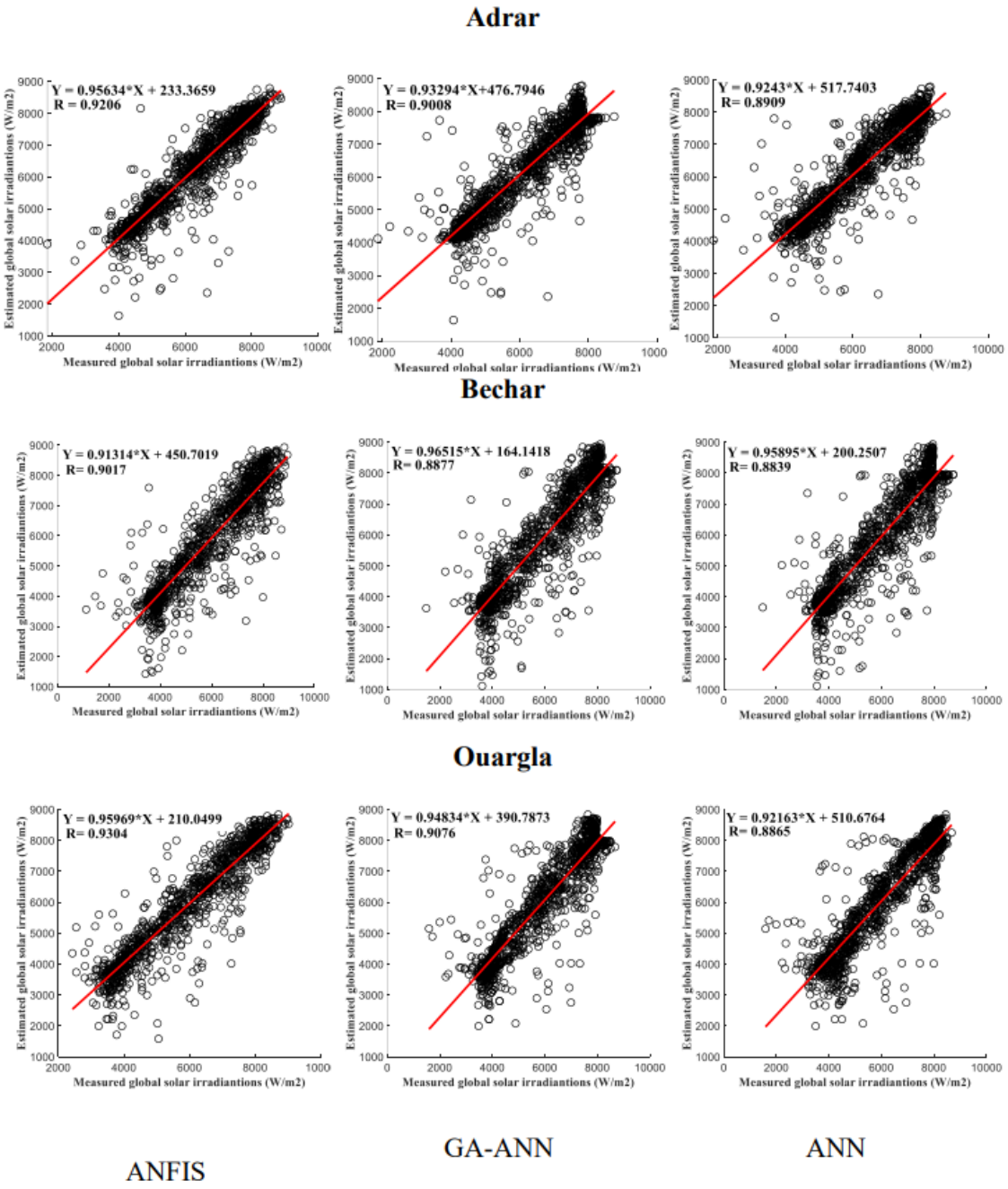


Figure V.5: Outcomes scatter plot.

In addition, for this research, two hybrid approaches, GA-ANN and ANFIS, they were designed with the purpose of predicting GSR. The two essential statistical outcome metrics, rRMSE and MAPE, have been used to compare and assess the effectiveness of our suggested model.

Table V.5 along with Fig.V.6 present the results of the overall statistic evaluation for rRMSE as well as MAPE across both of the considered models. Fig.V.6 demonstrate that the ANFIS hybrid model outperformed the hybrid GA-ANN model, with:

- In Adrar: (MAPE= 5.87; rRMSE= 9.05);

-In Bechar: (MAPE= 7.27; rRMSE= 10.76);

-In Ouargla: (MAPE= 9.24; rRMSE= 13.07).

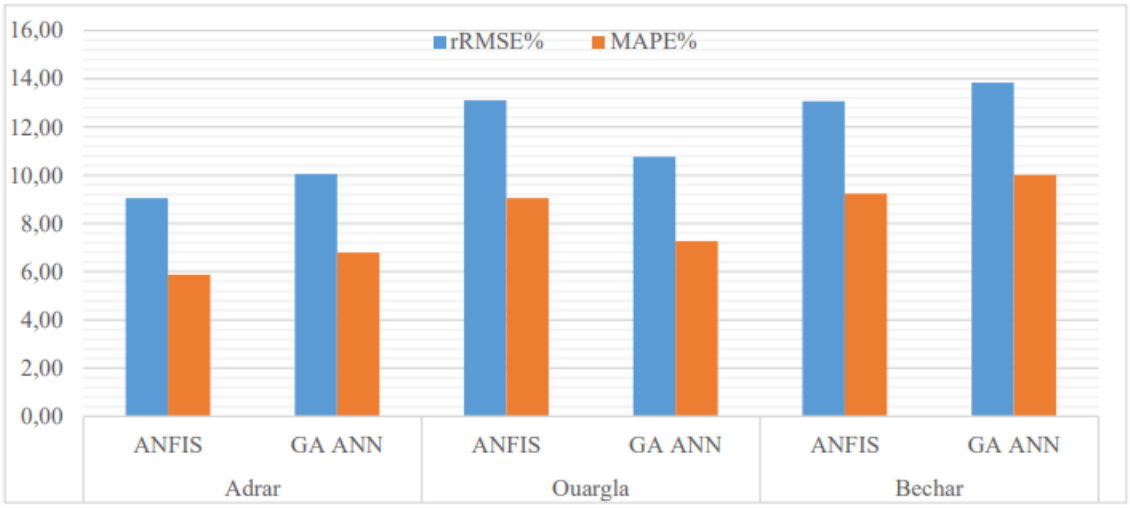


Figure V.6: rRMSE (%) and MAPE (%) of the models

We found that both the hybrid GA-ANN and ANFIS approaches surpassed the standalone ANN model. To conduct a comprehensive evaluation of their effectiveness, the error histograms of several hybrid models, along with specific indicators like the mean (μ) along with standard deviation (σ), are presented in Fig.V.7.

Based on Fig.V.7, it is obvious that ANFIS provided the best forecasts among the hybrid approaches by providing the lowest standard deviation indicator (σ) values, with (σ = 562.98) in Adrar, (σ = 770.96) in Bechar, and (σ = 639.8201) in Ouargla. It suggests that the predictions of that model are more consistent and less scattered around the mean. This

is often considered a sign of stability in the model's performance.

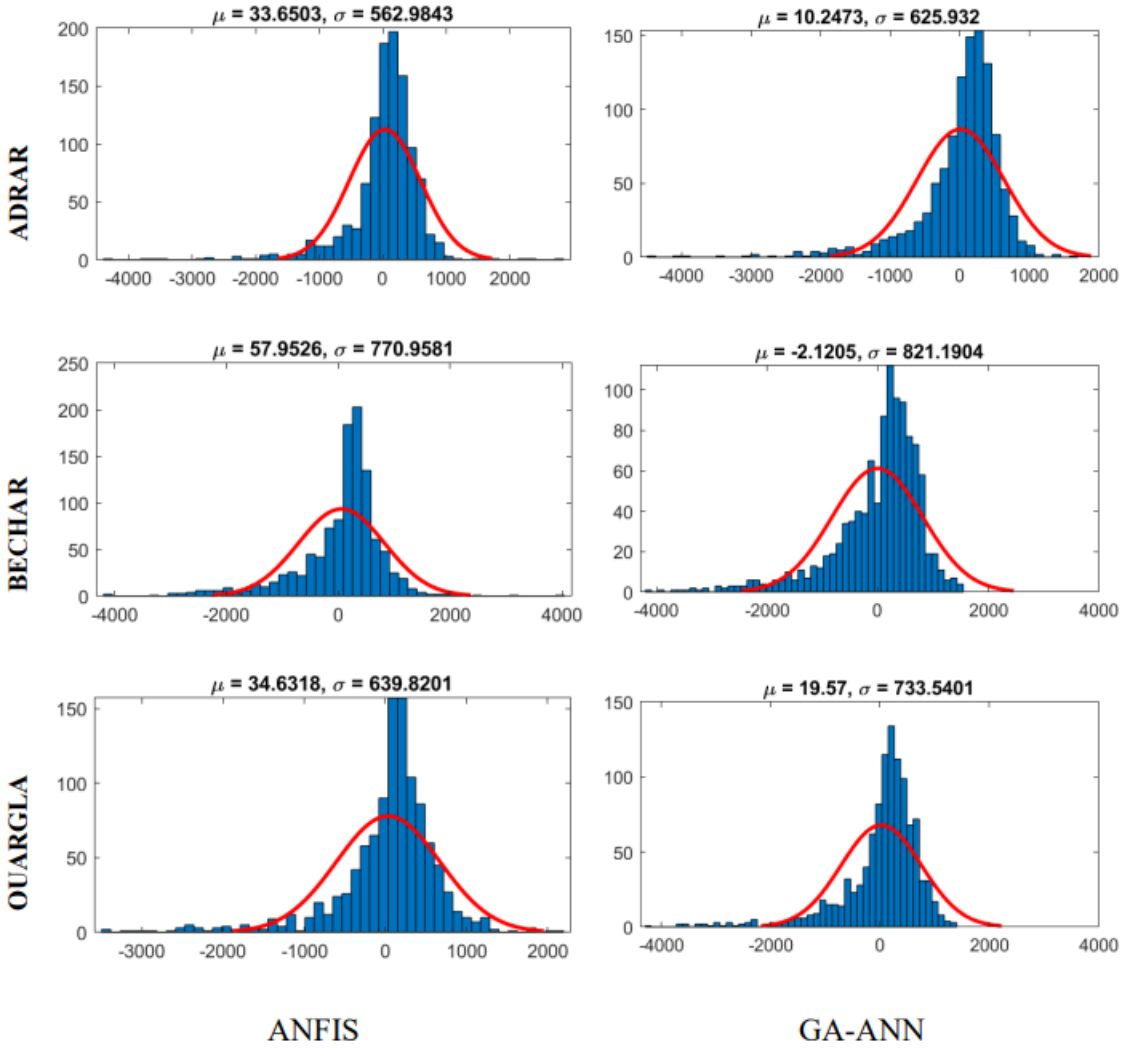


Figure V.7: The overall result's values of error histogram

In Fig.V. 8, representing residuals vs. fitted values, the ANFIS model exhibits a better distribution of points around the horizontal line passing through zero compared to the GA-ANN model. This suggests better performance for the predictive model corresponding to the ANFIS model. The enhanced accuracy is reflected in a tighter clustering of points around zero, indicating precise predictions and homogeneity in variance. This may result from model refinement in representing the relationship between independent and dependent variables, potentially incorporating influential variables. The ANFIS model demonstrates stability and robustness in the model's predictions.

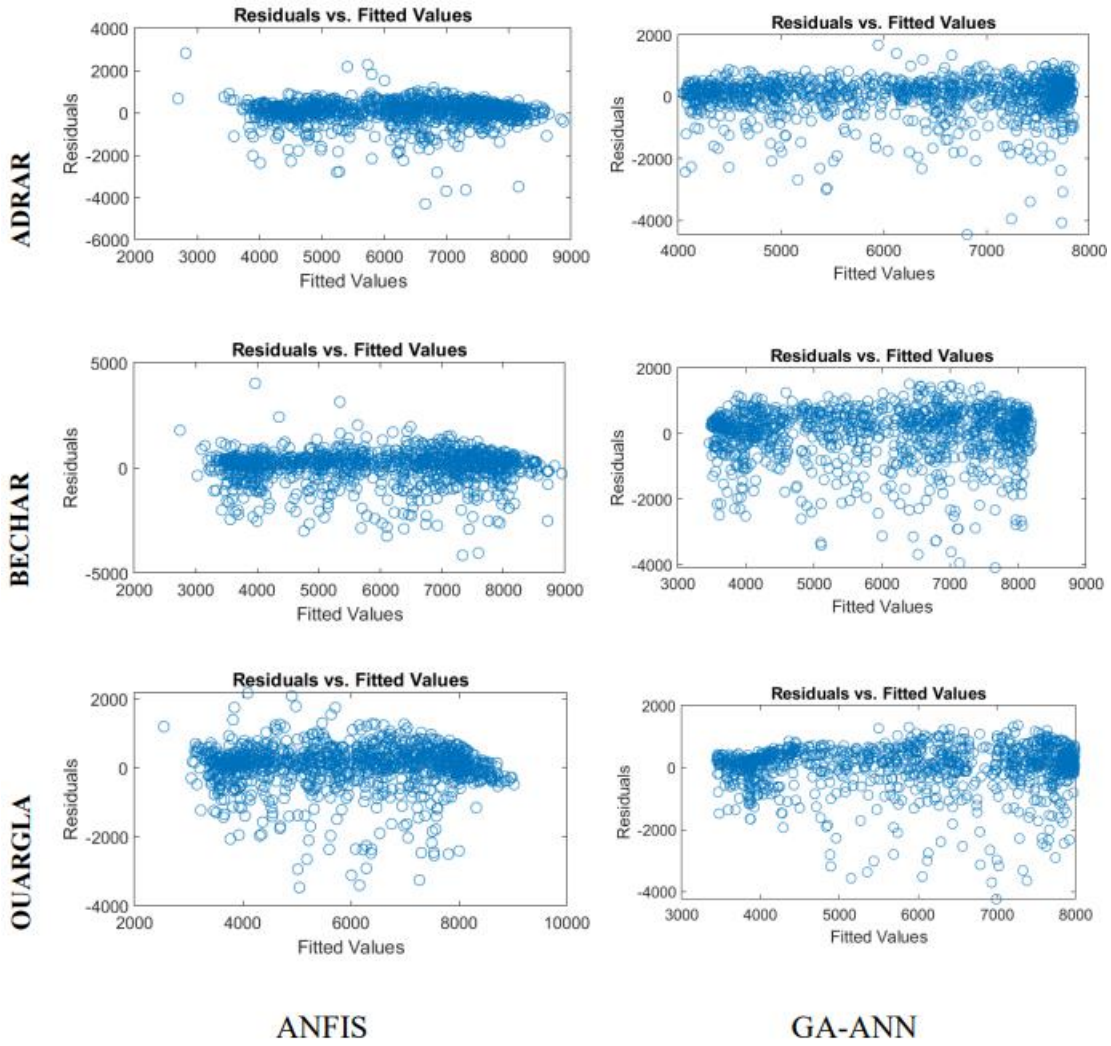


Figure V.8: The overall result's residuals vs. fitted Values.

Simiraley in Fig. V.9 shows the error plot. The ANFIS model in all three cities shows a more consistent error pattern with lower deviations from zero. This is consistent with the lower values of MAE, MSE, and RMSE, as well as the higher values of Nash-Sutcliffe Efficiency (NSE), which outperform the other combined GA-ANN approach in terms of prediction performance. With the best outcomes of the accuracy performance in each location for the ANFIS model of:

- In Adrar: (MAE= 365.79 J/m²/day; RMSE= 563.73 J/m²/day and NSE= 0.85);

- In Bechar: (MAE= 546.64 J/m²/day; RMSE= 772.78 J/m²/day and NSE= 0.81);

-In Ouargla: (MAE= 432.61 J/m²/day; RMSE= 640.471 J/m²/day and NSE= 0.87).

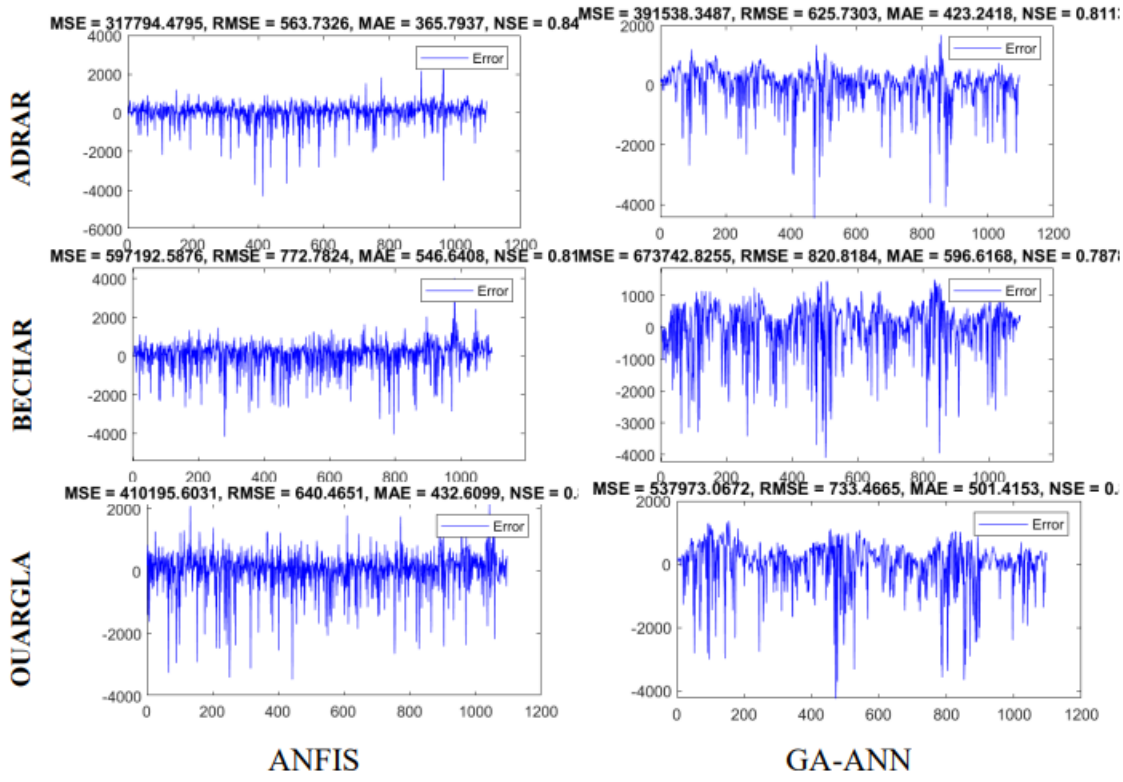


Figure V.9 Error plot of models performance

Fig. V.10 shows the QQ (Quantile-Quantile) plot, which is a useful tool for determining the normality of residuals while assessing the predictive performance of multiple models. A model is deemed more dependable when its residuals closely resemble a straight line in the QQ plot, showing that it complies with the normality assumption. The plots for both the ANFIS and hybrid GA-ANN models show that ANFIS performed better than the GA-ANN in all locations. The result shows that ANFIS residuals resemble better a straight line in the QQ plot, showing that ANFIS is more accurate at estimating Global Solar Radiation (GSR). ANFIS' higher performance, combined with its adaptability to complicated patterns and reduced reliance on manual adjustments, make it the preferable choice for accurate and reliable GSR forecasts in this scenario.

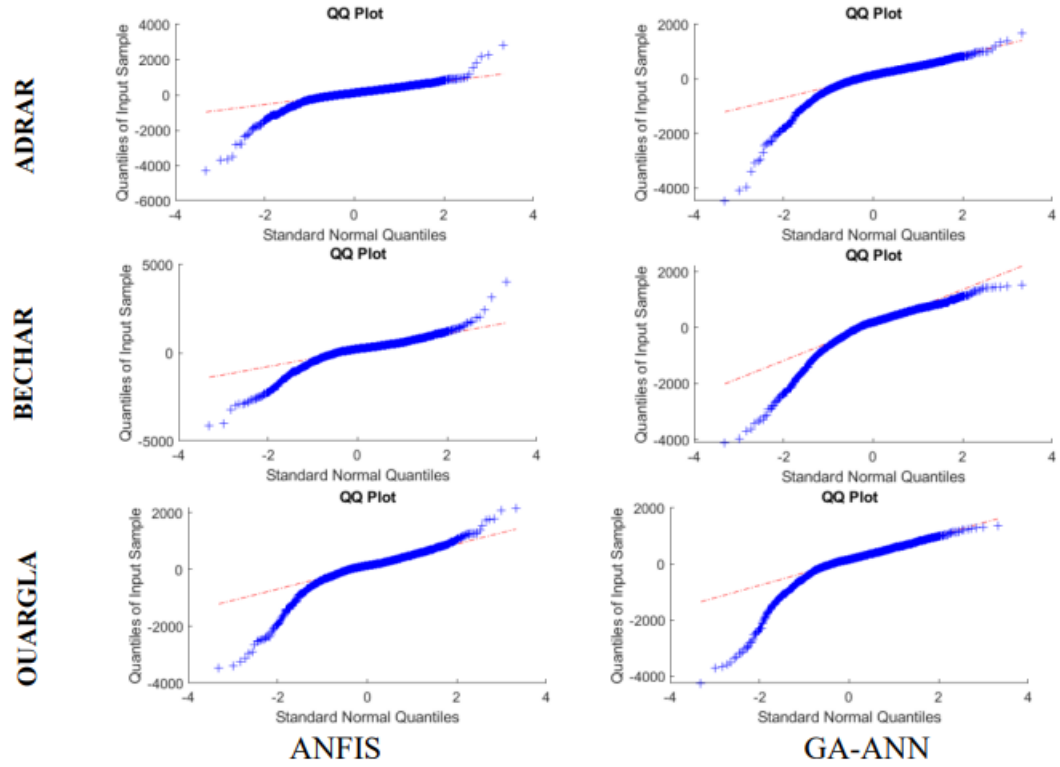


Figure V.10: Residuals Quantile-Quantile for ANFIS and GA-ANN models Normality Assessment

We can establish that in our study, the hybrid ANFIS model demonstrated the highest accuracy in predicting daily GSR based on overall data outcomes. We also conducted an extensive analysis to evaluate the performance of both hybrid ANN-based models (ANFIS and GA-ANN) during testing and training phases. Table V.5 displays the statistical indicators of performance for every model.

The ANFIS model outperforms the GA-ANN model in forecasting daily solar radiation. ANFIS demonstrates enhanced linear correlation and precision, particularly evident in Fig. V.11, where its regression line shows a steeper slope and lower intercept for lower radiation values. In contrast, GA-ANN exhibits a higher intercept and slightly shallower slope, suggesting less accuracy, especially for lower radiation levels. ANFIS is preferred for daily solar radiation prediction due to its superior accuracy and stronger predictive capabilities, particularly in lower radiation conditions. We can certainly establish that in this research, the hybrid ANFIS model resulted the Best.

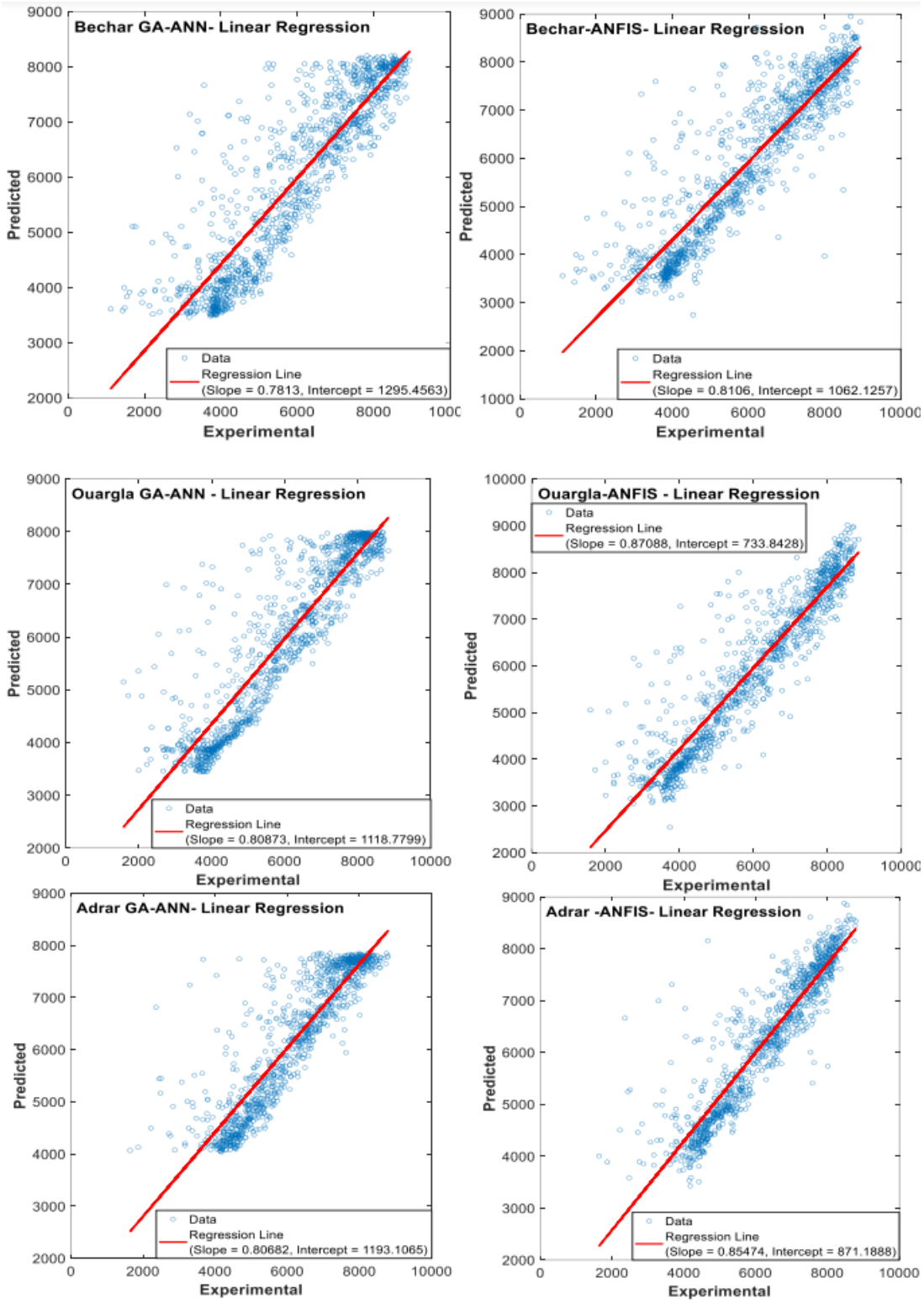


Figure V.11: Intercept tests and linear slope of the models.

Certainly, based on our study, it is evident that the hybrid ANFIS model demonstrated the highest accuracy in predicting daily GSR across all evaluations. Subsequently, we conducted an in-depth analysis to assess the performance of the two hybrid ANN-based models (ANFIS and GA-ANN) during both phases.

Table V.6 shows performance metrics for both models across all sites. ANFIS performed better in both phases.

Table V.6: Performance statistical indicators

		Training			Testing		
City	**Model**	**R**	**rRMSE%**	**MAPE%**	**R**	**rRMSE%**	**MAPE%**
ADRAR	ANFIS	**0.9224**	**8.99**	**5.75**	**0.9196**	**9.28**	**6.34**
	GA-ANN	0.8971	10.21	6.86	0.9017	9.35	6.51
OUARGLA	ANFIS	**0.9298**	**10.81**	**7.14**	**0.9374**	**10.54**	**7.78**
	GA-ANN	0.9095	12.26	8.43	0.9005	12.56	8.40
BECHAR	ANFIS	**0.9056**	**12.67**	**8.88**	**0.8987**	**14.63**	**10.74**
	GA-ANN	0.8894	13.61	9.92	0.8811	14.98	10.79

While fuzzy logic serves as a mathematical framework for managing imprecision and uncertainty, its typical applications are in control systems and decision-making processes, differing from optimization algorithms like GAs that excel in fields such as machine learning by finding optimal solutions.

When evaluating the models developed in this study, it becomes apparent that the combination of fuzzy logic with ANNs outperformed the approach of integrating ANNs with GA. In the context of forecasting solar radiation using the same set of inputs, ANFIS appears to be an appropriate approach. Its ability to handle data-driven patterns both and rule-based, along with enhanced interpretability through fuzzy logic, positions it as a strong contender.

ANFIS seamlessly combines linguistic rules with data-driven learning, adeptly addressing challenges posed by nonlinear data relationships. Importantly, it minimizes the requirement for extensive manual tuning, rendering it well-suited for scenarios encompassing intricate and rule-based data patterns.

Despite employing identical inputs, the hybrid GA-ANN model enhances the weights of the artificial neural network (ANN) through genetic algorithms (GA). This model offers flexibility and customization in terms of ANN architecture and hyperparameters, with its performance largely contingent on the evolutionary algorithm's capacity to optimize weights.

This model has the ability to excel in activities that need a high level of customization and optimization, particularly in complicated domains. In our case, this model succeeds in outperforming the standalone ANN model by optimizing its weights.

We may conclude that, not often, the application of optimization algorithms such as GA might not be the best option when developing a hybrid ANN model. It is undeniable that optimization algorithms do enhance the stand-alone model's performance.

Table V.7 shows various types of models used to forecast GSR. ANNs and other machine learning approaches can be compared to the proposed GA-ANN method. The calculated assessment indice (R) is used to assess the predictability of the stated and current methodologies.

Comparing with other methods, the results of our study demonstrate similar effectiveness in terms of predictability and forecasting proficiency.

Table V.7: The performance of the models compared to other studies.

Reference	Model	Time step	Location	R
[36]	ANN	daily	France	0.7800
[37]	ANN	daily	Iran	0.8940
[38]	ANFIS	daily	Nigeria	0.8540
[27]	PSO-BPNN	daily	China	0.9530
[11]	ARMA-ANN	daily	Algeria	0.9140
[39]	FFA-ANN	daily	Algeria	0.9321
The presented study	ANFIS	daily	Algeria	0.9304
The presented study	GA-ANN	Daily	Algeria	0.9076

V.5. Conclusions

Three methods for forecasting daily global solar radiation (GSR) over three cities in southern Algeria (Adrar, Ouargla, and Bechar) have been thoroughly evaluated in this chapter: ANN, GA-ANN, and ANFIS. Using six input variables, we assessed the models' the suitability for solar radiation prediction in arid locations during a three-year period (2019–2021).

Our results show that the hybrid GA-ANN model outperforms the standalone ANN model across key statistical parameters, suggesting that the hybrid model, which aims to refine the ANN technique through optimization of ANN weights, has promise. However, the greater computational cost of the GA-ANN model should be taken into account in real-world applications.

However, when it comes to daily GSR predicting over all examined locations, ANFIS outperforms the GA-ANN model, proving the usefulness of fuzzy logic in combination with ANNs. This emphasizes how crucial it is to take into account different approaches when building hybrid models to forecast solar radiation.

In conclusion, this study's model offers a useful resource for precisely predicting daily GSR in areas with arid conditions. Its use involves assisting decision-making concerning the setup and maintenance of solar power plants, which advances sustainable energy methods in arid areas.

V.6. References

1. Zhang, J., Zhao, L., Deng, S., Xu, W., & Zhang, Y. (2017). A critical review of the models used to estimate solar radiation. Renew. Sustain. Energy Rev, 70, 314–329. doi:10.1016/j.rser.2016.11.124
2. Zendehboudi, A., Baseer, M. A., & Saidur, R. (2018). Application of support vector machine models for forecasting solar and wind energy resources: A review. J. Clean Prod, 199, 272–285. doi:10.1016/j.jclepro.2018.07.164
3. Despotovic, M., Nedic, V., Despotovic, D., & Cvetanovic, S. (2015). Review and statistical analysis of different global solar radiation sunshine models. Renew. Sustain. Energy Rev, 52, 1869–1880. doi:10.1016/j.rser.2015.08.035
4. Zhou, Y., Liu, Y., Wang, D., Liu, X., & Wang, Y. (2021). A review on global solar radiation prediction with machine learning models in a comprehensive perspective. Energy Conv Manag, 235, 113960. doi:10.1016/j.enconman.2021.11396
5. Qazi, A,. Fayaz ,H,. Wadi, A,. Raj, R.G,. Rahim, N.A,. (2015). The artificial neural network for solar radiation prediction and designing solar systems: a systematic literature review. J Clean Prod. doi: 10.1016/j.jclepro.2015.04.041.
6. Guermoui, M., Melgani, F., Gairaa, K., & Mekhalfi, M. L. (2020). A comprehensive review of hybrid models for solar radiation forecasting. J Clean Prod, 120357. doi:10.1016/j.jclepro.2020.120357
7. Kabouche, N., Chellali, F., & Recioui, A. (2021). A review on solar radiation assessment and forecasting in Algeria: (Part 2; Solar radiation forecasting). Algerian J. Sign Syst, 6(3), 130-146. DOI:https://doi.org/10.51485/ajss.v6i3.141.
8. Queja, V.H,. Almoroxa, J,. Arnaldob, J. A,. Saitoc, L,. (2017). ANFIS. SVM and ANN soft-computing techniques to estimate daily global solar radiation in a warm sub-humid environment. J. Atmos. Sol–Terres Physics, 155, 62–70
9. Liu, Y., Wang, R., Yi, G. (2009). Research a Novel Optimization Mechanism of Parameters Based on Hybrid NN and GA. In: Yu, W., He, H., Zhang, N. (eds) Advances in Neural Networks – ISNN 2009. ISNN 2009. Lecture Notes in Computer

Science, vol 5552. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-01510-6_52

10. Wang, L,. Kisi, O,. Zounemat-Kermani, M,. Salazar, G.A,. Zhu, Z,. et al. (2016). Solar radiation prediction using different techniques: model evaluation and comparison. Renew. Sustain. Energy Rev, 61384–397.
11. Gairaa, K,. Khellaf, A,. Messlem, Y,. Chellali, F,. (2016). Estimation of the daily global solar radiation based on Box–Jenkins and ANN models: A combined approach. Renew. Sustain. Energy Rev, 57, 238–249. doi:10.1016/j.rser.2015.12.111
12. Benkaciali, S,. Haddadi, M,. Khellaf, A,. Gairaa K,. Guermoui, M,. (2016). Evaluation of the global solar irradiation from the artificial neural network technique. Rev Energies Renouv, 19, 617-631
13. Hejase, H. A. N., Al-Shamisi, M. H., & Assi, A. H. (2014). Modeling of global horizontal irradiance in the United Arab Emirates with artificial neural networks. Energy, 77, 542–552. doi:10.1016/j.energy.2014.09.064
14. Benatiallah D., Benatiallah A., Bouchouicha K., Nasri B. (2020). Prediction du rayonnement solaire horaire en utilisant les reseaux de neurone artificiel, Algerian J. Env. Sc. Technology, 6(1), 1236-1245
15. Antonopoulos V. Z,. Papamichail, D. M,. Aschonitis V. G,. Antonopoulos A.V,. (2019). Solar radiation estimation methods using ANN and empirical models. Comp. Electro. Agric, 160, 160-167 https://doi.org/10.1016/j.compag.2019.03.022
16. Amiri, B,. Gómez-Orellana A. M,. Gutiérrez P.A,. Dizène, R,. Hervás-Martínez, C,. Dahmani, K,. (2020). A Novel Approach for Global Solar Irradiation Forecasting on Tilted Plane using Hybrid Evolutionary Neural Networks. J.Clean. Prod, 35623-7 DOI: https://doi.org/10.1016/j.jclepro.2020.125577
17. Kassem A. S,. Aboukarima, A.M,. El Ashmawy, N. M,. Zayed M. F,. (2016). Comparison of Empirical Models and an Adaptive Neuro Fuzzy Inference System for Estimating Hourly Total Solar Radiation on Horizontal Surface at Alexandria City. Egypt; Adv. Res, 7(5), 1-17.
18. Mohammadi, K,. Shamshirband, S,. Petković, D,. Khorasanizadeh, H,. (2016). Determining the most important variables for diffuse solar radiation prediction using

adaptive neuro-fuzzy methodology ;case study: City of Kerman, Iran. Renew. Sustain. Energy Rev, 53, 1570–1579

19. Xue, X. (2019). Daily diffuse solar radiation estimation using adaptive neuro-fuzzy inference system technique. Numerical Heat Transfer, Part B: Fundamentals, 1–14. doi:10.1080/10407790.2019.1690879
20. Hussein, H.s,. (2020). Global solar radiation prediction using a combination of subtractive clustering algorithm and adaptive neuro-fuzzy inference system: a case study. J. eng science, 15, 1652 – 1669.
21. Benatiallah, D,. Bouchouicha, K,. Benatiallah, A,. Harouz, A,. Nasri B,. (2020). Estimation of global solar radiation using adaptive neuro-fuzzy inference system. Algerian J. Env. Sc. Tech, 7(4), 2097-2106
22. Wu, J., Chan, C. K., Zhang, Y., Xiong, B. Y., & Zhang, Q. H. (2014). Prediction of solar radiation with genetic approach combing multi-model framework. Renew Energy, 66, 132–139. doi:10.1016/j.renene.2013.11.064
23. Kisi, O. (2014). Modeling solar radiation of Mediterranean region in Turkey by using fuzzy genetic approach. Energy, 64, 429–436. doi:10.1016/j.energy.2013.10.009
24. Khosravi, A., Nunes, R. O., Assad, M. E. H., & Machado, L. (2018). Comparison of artificial intelligence methods in estimation of daily global solar radiation. J. Clean Prod, 194, 342–358. doi:10.1016/j.jclepro.2018.05.147
25. Bendali, W., Saber, I., Bourachdi, B., Boussetta, M., & Mourad, Y. (2020). Deep Learning Using Genetic Algorithm Optimization for Short Term Solar Irradiance Forecasting. 2020 Fourth International Conference On Intelligent Computing in Data Sciences (ICDS). doi:10.1109/icds50568.2020.926868
26. Wang, J., Xie, Y., Zhu, C., & Xu, X. (2011). Daily solar radiation prediction based on Genetic Algorithm Optimization of wavelet neural network. 2011 International Conference on Electrical and Control Engineering. doi:10.1109/iceceng.2011.6057583
27. Xue, X. (2017). Prediction of daily diffuse solar radiation using artificial neural

networks. Inter. J. Hydrogen Energy, 42(47), 28214–28221. doi:10.1016/j.ijhydene.2017.09.15

28. Benatiallah, D., Benatiallah, A., Bouchouicha, K., Hamouda, M., & Nasri, B. (2018). An empirical model for estimating solar radiation in the Algerian Sahara. AIP Conference Proceedings 1968, 030031 (2018); doi: 10.1063/1.5039218
29. Achour, L., Bouharkat, M., & Behar, O. (2018). Performance assessment of an integrated solar combined cycle in the southern of Algeria. Energy Reports, 4, 207–217. doi:10.1016/j.egyr.2017.09.003
30. SODA data. Available at: www.soda-pro.com/webservices#meteodata
31. Shi, J., Lee, W.-J., Liu, Y., Yang, Y., & Wang, P. (2012). Forecasting Power Output of Photovoltaic Systems Based on Weather Classification and Support Vector Machines. IEEE Transactions on Industry Applications, 48(3), 1064–1069. doi:10.1109/tia.2012.2190816
32. Jang, J.-S.R., (1993). ANFIS: adaptive-network-based fuzzy inference system. Syst. Man Cybern. IEEE Trans, 23, 665–685. http://dx.doi.org/10.1109/21.25654
33. Abdulshahed, A.M., Longstaff, A.P., Fletcher, S., (2015). The application of ANFIS prediction models for thermal error compensation on CNC machine tools. Appl. Soft Comput, 27, 158–168.
34. Preis, A., & Ostfeld, A. (2008). A coupled model tree–genetic algorithm scheme for flow and water quality predictions in watersheds. J. Hydrology, 349(3-4), 364–375. doi:10.1016/j.jhydrol.2007.11.013
35. Goldberg, D. (1989). Genetic Algorithms in Search, Optimization, and Machine Learning. Addison-Wesley, Reading, MA.
36. Voyant, C., Muselli, M., Paoli, C., Nivet, M.-L. (2011). Optimization of an artificial neural network dedicated to the multivariate forecasting of daily global radiation. Energy, 36(1), 348–359. doi:10.1016/j.energy.2010.10.032
37. Ramedani, Z., Omid, M., Keyhani, A., Shamshirband, S., & Khoshnevisan, B. (2014). Potential of radial basis function based support vector regression for global solar radiation prediction. Renew Sustain Energy Rev, 39, 1005–1011.

doi:10.1016/j.rser.2014.07.108

38. Olatomiwa, L., Mekhilef, S., Shamshirband, S., & Petković, D. (2015). Adaptive neuro-fuzzy approach for solar radiation prediction in Nigeria. Renew. Sustain. Energy Rev, 51, 1784–1791. doi:10.1016/j.rser.2015.05.068

39. Djeldjli, H,. Benatiallah, D,. Tanougast, C,. Benatiallah, A,. (2023). Solar radiation forecasting based on ANN. SVM and a novel hybrid FFA-ANN model: A case study of six cities south of Algeria. AIMS Energy, 12(1),62-83. DOI: 10.3934/energy.2024004

General conclusion and perspective

With its limitless potential and renewable nature, solar energy is becoming a major force in the world energy scene, providing a long-term substitute for conventional fossil fuels. We have conducted a thorough investigation of solar radiation and its consequences for the production of renewable energy throughout this thesis. Our trip has shed light on the difficulties and possibilities that come with using solar radiation, highlighting the need for precise forecast models to optimize its use.

The unpredictable and fluctuating nature of solar radiation presents a major obstacle to the steady integration of solar power into national electrical systems. Sophisticated forecasting models that can capture the dynamic nature of solar energy are necessary to address this difficulty.

Our research has concentrated on creating hybrid forecasting models that combine artificial intelligence methods, metaheuristic algorithms, and other strategies to improve prediction accuracy.

Our research intends to construct hybrid forecasting models that can provide accurate estimates of solar radiation by utilizing different input characteristics from different sites in Algeria. By means of thorough examination and testing, we have attempted to enhance these models and realize their complete capability in forecasting variations in solar irradiation. Our study's conclusions highlight how crucial it is to gather, process, and analyze data with integrity in order to evaluate solar potential and optimize the installation and functioning of solar energy systems.

It is clear from considering the conclusions in each chapter that precise solar radiation forecasting is necessary for the effective management and operation of photovoltaic systems. Our investigation of several approaches—from sophisticated machine learning algorithms to empirical models—has yielded insightful information on the intricacies of solar radiation variability. Through a comprehensive analysis of the performance of multiple forecasting models, we have pinpointed potential opportunities to improve prediction precision and dependability.

The hybrid approaches—like the FFA-ANN and GA-ANN models—show a great deal of promise for enhancing forecasting precision and directing solar energy deployment

decisions. In addition, our analysis emphasizes the necessity of taking into account other approaches, such ANFIS, in order to tackle the difficulties presented by arid climate zones.

The knowledge gained from our research helps to enhance sustainable energy practices by facilitating better decision-making in energy planning and management.

Accurate solar radiation forecasting models enhance the efficiency of solar energy production and help identify suitable locations for building solar power plants without the need for measurement devices. This reduces operational costs, boosts renewable energy stability, and thus supports the economy while achieving sustainable development.

As a conclusion, this thesis offers proof of the importance of accurate solar radiation forecasting for maximizing the efficiency and stability of solar energy systems. By developing and evaluating hybrid forecasting models, we hope to provide all pertinent stakeholders with the knowledge and tools they require to effectively handle the complexities associated with the production of solar energy. While we work to build a future that is more ecologically friendly and sustainable. The study emphasizes how important accurate solar radiation forecasting is to maximizing the dependability and efficiency of solar energy systems.

In the future, we will be investigating new hybrid methods to improve solar radiation forecast models while making sure that they are accurate and flexible. By pursuing this objective, we hope to contribute to a more sustainable and eco-friendly future by giving all parties concerned essential information and resources for managing solar energy.

الملـــخـــص

يلعب التنبؤ بالإشعاع الشمسي دورًا محوريًا في تصميم أنظمة الطاقة الشمسية وتحسينها، خاصة في المناطق التي تتميز بالمناخ الجاف مثل جنوب الجزائر. تبحث هذه الأطروحة في أداء نماذج التنبؤ المختلفة في تقدير الإشعاع الشمسي العالمي اليومي (G) عبر ست مدن في المنطقة: أدرار، ورقلة، بشار، الوادي، تيميمون، وتمنراست. تم تقييم أساليب متميزة - الشبكات العصبية الاصطناعية (ANN)، وشبكة (GA-ANN) القائمة على دمج الخوارزمية التحسين الجينية، وشبكة (FFA-ANN) القائمة على دمج خوارزمية التحسين اليراع - للتأكد من دقتها وموثوقيتها. تكشف الدراسة أن نماذج التنبؤ الهجينة، التي قمنا بتطويرها FFA-ANN وGA-ANN، تتفوق على نماذج ANN المستقلة من حيث دقة التنبؤ عبر جميع المواقع. في حين يقوم النموذج GA-ANN على تحسين اوزان الشبكة في التنبؤ بـ GSR ، يقوم النموذجFFA-ANN باختيار احسن مجموعة مدخلات مناسبة لكل موقع لضمان احسن أداء للنموذج، حيث حقق النموذجين قيمًا جيدة جدا في المؤشرات الإحصائية مثل معامل التحديد (R=0.9076 لـGA-ANN ،و R=0.9321 لـ FFA-ANN)، متوسط النسبة المئوية للخطأ المطلق (MAPE) ، والجذر النسبي يعني خطأ مربع (rRMSE). والجدير بالذكر أن FFA-ANN يُظهر أداءً قويًا في تقدير GSR اليومي، مما يُظهر قدرته على التكيف مع الظروف المناخية المختلفة.علاوة على ذلك، يسلط البحث الضوء على أهمية معلمات المدخلات مثل الإشعاع الشمسي خارج الأرض (H_0)، والانحراف، ومتوسط درجة الحرارة (T_{avg})، والرطوبة النسبية (RH) في تعزيز القدرات التنبؤية لنموذج ANN في المناطق الجافة. تؤكد النتائج على إمكانات نماذج التنبؤ الهجينة في تحسين التنبؤ بالإشعاع الشمسي وتسهيل اتخاذ القرارات في تركيبات نظام الطاقة الشمسية وحجمها. وبشكل عام، تساهم هذه الأطروحة برؤى قيمة في تطوير وتطبيق نماذج التنبؤ الهجينة للتنبؤ بالإشعاع الشمسي في المناخات القاحلة. ومن خلال الاستفادة من تقنيات التحسين المتقدمة والتحليل الشامل للبيانات، تضع الدراسة الأساس لتعزيز كفاءة وفعالية أنظمة الطاقة الشمسية في جنوب الجزائر والمناطق المماثلة في جميع أنحاء العالم.

الكلمات المفتاحية: نموذج هجين، شبكة عصبية اصطناعية، خوارزمية اليراعة، إشعاع شمسي، ، خوارزمية جينيةGA .

Abstract

Solar irradiation prediction is essential to the planning and development of solar energy systems. especially in regions characterized by a dry climate such as southern Algeria. This thesis investigates the performance of different forecast models in predicting daily global solar radiation (GSR) across six cities in the region: Adrar, Ouargla, Béchar, El Oued, Timimoun, and Tamanrasset. Artificial neural networks (ANN), GA-ANN, and FFA-ANN—are evaluated for their accuracy and reliability. The study reveals that the hybrid prediction models we developed, FFA-ANN and GA-ANN, outperform standalone ANN models in terms of prediction accuracy across all sites. While the GA-ANN model optimizes the network weights when predicting GSR, the FFA-ANN model chooses the best combination set of inputs for each site to ensure the best performance of the model. The two models achieved very good statistical indicators results , the coefficient of determination (R=0.9076 for GA-ANN, and R=0.9321 for FFA-ANN), and the average absolute percentage error (MAPE), and relative root mean square error (rRMSE). Furthermore, the research highlights the significance of input parameters such as extraterrestrial solar irradiation (H_0), declination, average temperature (T_{avg}), and relative humidity (RH) in enhancing the predictive capabilities of the ANN model. The findings underscore the potential of hybrid forecasting models in optimizing solar radiation forecasting and facilitating informed decision-making in solar energy system installations and sizing. Overall, this thesis contributes valuable insights into the development and application of hybrid forecasting models for solar radiation forecasting in arid climates. By leveraging advanced optimization techniques and comprehensive data analysis, the study lays the foundation for enhancing the efficiency and effectiveness of solar energy systems in southern Algeria and similar regions worldwide.

Keywords: Hybrid model, Artificial neural network, Firefly algorithm, Solar radiation, Genetic algorithm.