Centre d'Etudes Doctorales : Sciences et Techniques de l'Ingénieur

N° d'ordre 01/2020

THESE DE DOCTORAT

Présentée par

Mme: Manal MARZOUQ

Discipline : Génie Électrique

Spécialité : Génie Électrique

Sujet de la thèse : Forecasting and modeling of solar irradiation and its components and the photovoltaic power production using Machine Learning techniques.

Formation Doctorale : Sciences de l'ingénieur Sciences Physiques, Mathématiques et Informatique.

Thèse présentée et soutenue le samedi 02 janvier 2021 à 10h30 devant le jury composé de :

Nom Prénom	Titre	Etablissement	
Driss CHENNOUNI	PES	Ecole Normale Supérieure de Fès	Président
Bekkay HAJJI	PES	Ecole Nationale des Sciences Appliquées d'Oujda	Rapporteur
Rachid EL GOURI	PES	Ecole Nationale des Sciences Appliquées de Kénitra	Rapporteur
Abdelmajid JAMIL	PES	Ecole Supérieure de Technologie de Fès	Rapporteur
Abdellah MECHAQRANE	PES	Faculté des Sciences et Techniques de Fès	Examinateur
Zakia LAKHLIAI Hakim EL FADILI	PES PH	Ecole Supérieure de Technologie de Fès Ecole Nationale des Sciences Appliquées de Fès	Directeurs de thèse

Laboratoire d'accueil : Laboratoire d'Informatique et de Physique Interdisciplinaire.

Etablissement : Ecole Normale Supérieure de Fès.

Aknowledgement

In the Name of Allah, the Most Merciful, the Most Compassionate.

First and foremost, I must acknowledge my limitless thanks to Allah, the Ever-magnificent, the Ever Thankful, for His help and bless by giving me the opportunity, courage and enough energy to carry out and complete this thesis journey.

It is a genuine pleasure to express my deep sense of thanks and gratitude to my supervisor Professor **Zakia LAKHLIAI** for giving me the opportunity to pursue this dream, her kindness, enthusiasm and continuous support. Also, my supervisor Professor **Hakim EL FADILI** my mentor, philosopher and guide. It was a great privilege and honor to work and study under his guidance. His love and dedication for knowledge and science is imperishable. I am extremely grateful for what he offered me, his timely advice and scientific approach and above all, his overwhelming attitude to offer help for his students have been the perpetual fuel all along this thesis journey. I will always remain grateful to this true man of honor.

I thank Professor **Driss CHENNOUNI** for the honor he accorded me by agreeing to chair the jury for this thesis.

My deepest gratitude is also addressed to the reading committee members: Professor **Bekkay HAJJI**, Professor **Rachid EL GOURI**, Professor **Abdelmajid JAMIL** and Professor **Abdellah MECHAQRANE** for their time and interest toward this thesis and their extremely precious comments and suggestions.

I would also like to thank Professor **Driss CHENNOUNI** Director of Computer Sciences and Interdisciplinary Physics Laboratory for providing us with the necessary Software used for our study within this thesis.

I would also like to thank Professor **Abdellah MECHAQRANE** for putting under our disposition the necessary data of the city of Fez from the radiometric station at the Faculty of Sciences and Technologies of Fez.

To my friends, especially **Amina EL YAAGOUBI**, **Meryem EL YAAGOUBI** and **Marouane LAHYAN**, I express my gratitude for their unconditional friendship, support and patience throughout these years.

At the end, my deepest and sincere thanks and gratitude are always to my parents **Bensalem** and **Fatiha** for their constant support and prayers, to my sister **Meryem** my eternal soulmate, to my brothers **Hakim**, **Faiçal** and **Zakaria** for all sacrifices and patience they ever made for me in my journey and to the joy of my life **Kenza** and **Lina**.

Dedication

To my dear Parents for their unconditional love, trust, support and sacrifices.

I owe it all to you.

Abstract

Assessment of solar potential over a location of interest is a crucial step for the successful planning and management of solar energy systems. The main objectives of the present thesis is to develop accurate models for the estimation and forecasting of global, diffuse and direct solar irradiation. This is followed by short term photovoltaic power forecasting using optimized machine learning techniques, especially Evolutionary Artificial neural network compared with implemented k Nearest Neighbors, Decision Trees and Random Forest models.

Our study begins with a systematic review of solar energy estimation and forecasting using Artificial Neural Networks (ANNs) where the most important research papers were analyzed and discussed. The objective is to point out research limitations in this field and to propose recommendations and perspectives for future investigations.

Based on this review, an automatic selection of ANN inputs from available parameters set for the estimation of horizontal daily global solar irradiation is proposed. Experiments were developed using parameters from a radiometric station located in the Faculty of Sciences and Technologies building at Sidi Mohamed Ben Abdellah University, Fez, Morocco. This selection is carried out using Evolutionary Artificial Neural Networks (EANN) with different genetic operators including selection, crossover and mutation taking into account the coding of ANN as an individual in the pool of evolution. Obtained performances show that the proposed model achieves good solar irradiation estimation performance and also allows an automatic selection of the best input parameters for the ANN with lower population size and maximum generation ensuring good computation time.

Besides, we propose a multi-model framework generating multiple models for different time horizons to forecast global, diffuse and direct solar irradiance from 1 to 6 hours ahead for 28 sites in Morocco. Each model evolves simultaneously forecasting history and ANN architecture, searching automatically for the most accurate forecasting. A useful zoning scenario is investigated presenting a particular interest in the case of the lacking of solar irradiance measurements in some areas, and where future values of solar irradiance can be obtained using data from stations belonging to the same zone. A comparative study is then conducted with three other models (smart persistence, regression trees and random forest). The proposed framework shows good accuracy for solar irradiance forecasting and proves high generalization capabilities against different climatic conditions.

Finally, a preliminary work is presented to forecast photovoltaic power production using machine-learning techniques. The proposed EANN approach, is used to forecast short term solar power output of a photovoltaic system placed at the National School of Applied Sciences of Fez. Results of the proposed model are promising compared to two other machine learning models: regression trees and random forest.

Keywords: daily global solar irradiation estimation; Evolutionary Artificial Neural Networks; Automatic selection of inputs; photovoltaic energy production forecasting; Short term forecasting; global, diffuse and direct solar irradiance; Multi-model framework ; Climatic zoning; Machine learning models

Résumé

L'évaluation du potentiel solaire sur un site d'intérêt est une étape cruciale pour la planification et la gestion optimale des systèmes d'énergie solaire. L'objectif principal de la présente thèse est de développer des modèles plus précis pour l'estimation et la prévision de l'irradiation solaire globale, diffuse et directe. En plus d'une prévision à court terme de l'énergie photovoltaïque en utilisant des techniques de Machine Learning optimisées, en particulier les Réseaux de Neurones Artificiels Evolutifs comparés aux modèles implémentés : k plus proches voisins, arbres de décisions et forêts aléatoires.

Notre travail commence par une étude bibliographique systématique de l'estimation et de la prévision de l'énergie solaire à l'aide des Réseaux de Neurones Artificiels (RNA), où les articles de recherche les plus pertinents sont analysés et discutés. L'objectif est de mettre en évidence les limites de la recherche dans ce domaine et de proposer des recommandations et des perspectives pour les futurs travaux de recherche.

Sur la base de cette étude, nous proposons une sélection automatique des entrées RNA à partir d'un ensemble de paramètres disponibles pour l'estimation de l'irradiation solaire globale horizontale journalière. Les expériences ont été développées en utilisant différents paramètres collectés de la station radiométrique située dans le bâtiment de la Faculté des Sciences et Techniques de l'Université Sidi Mohamed Ben Abdellah de Fès au Maroc. Cette sélection est effectuée à l'aide de Réseaux de Neurones Artificiels Evolutifs (RNAE) avec différents opérateurs génétiques, y compris la sélection, le croisement et la mutation en tenant compte du codage de l'ANN en tant qu'individu dans le bassin de l'évolution. Les performances obtenues montrent que le modèle proposé permet d'obtenir de bonnes performances d'estimation de l'irradiation solaire et permet également une sélection automatique des meilleurs paramètres d'entrée pour le RNA avec un temps de calcul minimal.

En outre, nous proposons un Framework multi-modèle pour différents horizons temporels de 1 à 6 heures à l'avance afin de prévoir l'irradiation solaire globale, diffuse et directe pour 28 sites au Maroc. Chaque modèle généré fait évoluer simultanément l'historique des prévisions et l'architecture du RNA, en recherchant automatiquement la prévision la plus précise. Un scénario de zonage utile est proposé, présentant un intérêt particulier dans le cas de l'absence de mesures de l'irradiation solaire dans certaines régions, et où les valeurs futures de l'irradiation solaire peuvent être obtenues en utilisant les données de stations appartenant à la même zone. Une étude comparative est ensuite menée avec trois autres modèles (Smart Persistence, arbres de décision et forêts aléatoires). Le Framework proposé présente une bonne précision pour la prévision de l'irradiation solaire avec une grande capacité de généralisation pour différentes conditions climatiques.

Enfin, nous entamons une étude pour la prévision de l'énergie photovoltaïque à l'aide de techniques Machine Learning en s'inspirant de l'approche RNAE proposée précédemment afin de prévoir la production d'énergie solaire à court terme d'un système photovoltaïque placé à l'École Nationale des Sciences Appliquées de Fès. Les résultats du modèle proposé sont prometteurs comparés à deux autres modèles du Machine Learning : les arbres de décisions et forêts aléatoires.

Mots clés : estimation de l'irradiation solaire globale journalière ; réseaux de neurones artificiels évolutifs ; sélection automatique des entrées ; prévision de l'énergie photovoltaïque ; prévision à court terme ; irradiation solaire globale, diffuse et directe ; Framework multi-modèle ; zonage climatique ; modèles de machine learning

ملخص

يعد تقييم الإمكانات الشمسية على موقع ما خطوة حاسمة للتخطيط والإدارة الناجحين لأنظمة الطاقة الشمسية. تتمثل الأهداف الرئيسية للأطروحة الحالية في تطوير نماذج دقيقة لتقدير وتوقع الإشعاع الشمسي العام والمنتشر والمباشر. نقوم بعد ذلك بتنبؤ قصير المدى للطاقة الكهروضوئية باستخدام تقنيات التعلم الآلي المحسّنة، وخاصة الشبكة العصبية الاصطناعية التطورية مقارنة بنماذج أخرى للتعلم الآلي.

تبدأ الأطروحة بدراسة ببليوغرافية منهجية لتقدير الطاقة الشمسية والتنبؤ بها باستخدام الشبكات العصبية الاصطناعية (ANNs) حيث تم تحليل ومناقشة أهم الأوراق البحثية بهدف تحديد قيود البحث في هذا المجال واقتراح توصيات ووجهات نظر للأبحاث المستقبلية.

بناءً على هذه الدراسة، نقترح اختيار تلقائي لمدخلات ANN من المعلمات المتاحة المحددة لتقدير الإشعاع الشمسي اليومي الأفقي العام حيث تم تطوير التجارب باستخدام معلمات من محطة قياس إشعاعي تقع في مبنى كلية العلوم والتقنيات بجامعة سيدي محمد بن عبد الله، فاس، المغرب. يتم إجراء هذا الاختيار باستخدام الشبكات العصبية الاصطناعية التطورية (EANN)مع مشغلين وراثيين مختلفين بما في ذلك الاختيار والتقاطع والطفرة مع مراعاة تشفير ANN كفرد في مجموعة التطور. تُظهر النتائج التي تم الحصول عليها أن النموذج المقترح يحقق أداءً جيدًا لتقدير الإشعاع الشمسي ويسمح أيضًا بالاختيار التلقائي لأفضل معلمات الإدخال لشبكة ANN مع ضمان وقت جيد للحساب.

إلى جانب ذلك، نقترح إطارًا متعدد النماذج يولد نماذج لآفاق زمنية مختلفة للتنبؤ بالإشعاع الشمسي العام والمنتشر و المباشر من 1 إلى 6 ساعات في 28 موقعًا في المغرب. يتطور كل نموذج في وقت واحد لتاريخ التنبؤ وبنية الANN ، ويبحث تلقائيًا عن التنبؤ الأكثر دقة. نقترح سيناريو تقسيم مفيد في حالة نقص قياسات الإشعاع الشمسي في بعض المناطق، وحيث يمكن الحصول على القيم المستقبلية للإشعاع الشمسي باستخدام بيانات من محطات تنتمي إلى نفس المنطقة. ثم أجريت دراسة مقارنة مع ثلاثة نماذج أخرى للتعلم الآلي. يُظهر الإطار المقترح دقة جيدة للتنبؤ بالإشعاع الشمسي ويثبت قدرات التعميم العالية في ظروف مناخية مختلفة.

أخيرًا، تم تقديم عمل تمهيدي للتنبؤ بالطاقة الكهروضوئية باستخدام تقنيات التعلم الآلي. يتم استخدام نموذج EANN المقترح للتنبؤ بإنتاج الطاقة الشمسية على المدى القصير لنظام الكهروضوئية الموجود في المدرسة الوطنية للعلوم التطبيقية بفاس. نتائج النموذج المقترح واعدة مقارنة بنموذجين آخرين للتعلم الآلي.

الكلمات المفتاحية: التقدير العام اليومي للإشعاع الشمسي؛ الشبكات العصبية الاصطناعية التطورية. الاختيار التلقائي للمدخلات؛ التنبؤ بالطاقة الكهروضوئية؛ التنبؤ على المدى القصير للإشعاع الشمسي العالمي والمنتشر والمباشر؛ إطار متعدد النماذج؛ التقسيم المناخي؛ نماذج التعلم الآلي

List of Figures

1.1 Sun-Earth geometric relationships. 30

1.2 Movement of the Earth's rotation around the Sun. 30

1.3 Celestial coordinate system for a location P on Earth. 32

1.4 Illustraton of the horizontal coordinate system. 35

1.5 The three main components of solar irradiation. 37

2.1 Approximation of the biological neuron by the artificial neuron. 42

2.2 Perceptron, Rosenblatt model. 45

2.3 Separation performed by a multilayer perceptron. 46

2.4 Simplified diagram of a Perceptron Multilayer. 47

2.5 Different types of transfer functions for the artificial neuron. a) threshold function of the neuron of Mc Culloch and W. Pitts (1943) [46], b) piecewise linear of the Adaline model of Widrow and Hoff (1960)[51], c) sigmoid of a Perceptron Multilayer network of Rosenblatt (1962)[57], d) Gaussian of the RFR network of Moody and Darken (1989)[58]. 52

2.6 Distribution of the reviewed publications according to journal. 56

3.1 Various solar irradiation intensities in different cities of Morocco with a specific localization of Fez city [124]. 73

3.2 Forward stepwise regression flowchart. 75

3.3 Genetic algorithm phases starting from selection to mutation through intermediate population. Where string MB is the mutated version of string B. 76

3.4 Our genetic algorithm representation of the ANN estimator, where l is the number of total inputs. 77

3.5 Illustration of the adopted crossover of two parents using (a) a single cutting point in the (i, i+1) position and (b) two cutting points in the positions $(i, i+1)$ and $(j, j+1)$. 79

3.6 The block diagram of the proposed method. 80

3.7 Measured (target) versus estimated (output) daily global solar irradiation obtained with (a) train and (b) test datasets. 86

3.8 Obtained performances of our proposed algorithm through 20 generations with the corresponding input parameters using 4 individuals in the population. 87

4.1 The adopted thermal climatic zoning map in Morocco [175]. 101

4.2 Flowchart of the proposed evolutionary framework generating multi-HAEANN models. 103

4.3 The adopted artificial neural network forecaster. 104

4.4 Illustration of the generated neural network forecaster from the proposed framework up to 3 hours ahead ($i = 3$). 105

4.5 The proposed chromosome representing the forecasting history and ANN architecture with $M + 1$ parameters where M is the maximum number of hidden layers. 105

4.6 Our proposed chromosome representation of the neural network forecaster with corresponding parameters. 106

4.7 Our proposed crossover operator in the case of 2 cutting points. 108

4.8 The adopted sliding window training procedure for all proposed models (inspired from [177]). 109

4.9 Produced offspring for different generations (Gen 1 to Gen 5) using 4 as population size for global solar irradiance forecasted by HAEANN 1 for Marrakech city. 112

4.10 Comparison of hourly solar irradiance forecasting models for various horizons in terms of NRMSE for all_cities scenario: (a) global horizontal solar irradiance, (b) diffuse solar irradiance and(c) direct normal solar irradiance. 122

4.11 Measured vs forecasted global solar irradiance values using HAEANN 6 for Dakhla city. 1 day from each season is represented belonging to: Autumn (a), Winter (b), Spring (c) and Summer (d). 123

4.12 Decomposition of NRMSE error of the four compared models for each of the six hours for the 10 considered cities for global solar irradiance forecasting. 128

4.13 Decomposition of NRMSE error of the four compared models for each of the six hours for the 10 considered cities for diffuse solar irradiance forecasting. 129

4.14 Decomposition of NRMSE error of the four compared models for each of the six hours for the 10 considered cities for direct normal solar irradiance forecasting. 130

4.15 Decomposition of NRMSE error of the four compared models for each of the six hours for the six zones for global solar irradiance forecasting. 131

4.16 Decomposition of NRMSE error of the four compared models for each of the six hours for the six zones for diffuse solar irradiance forecasting. . . . 132

4.17 Decomposition of NRMSE error of the four compared models for each of the six hours for the six zones for direct normal solar irradiance forecasting. 133

4.18 Decomposition of NRMSE error of the four compared models for each of the six hours for all_cities scenario for forecasting (a) global solar irradiance, (b) diffuse solar irradiance, (c) direct solar irradiance. 134

5.1 Classification of PV power forecasting techniques. 139

5.2 Description of the photoelectric effect. 141

5.3 Description of the photovoltaic effect on a solar cell. 142

5.4 A typical I-V curve for a silicon photovoltaic module. 143

5.5 Flowchart of the used Incremental Inductance algorithm. 147

5.6 Block diagram of our PV system with Power Converter and MPPT Control. 148

5.7 The simulation of our PV system with MPPT control based Incremental Conductance technique. 149

5.8 Block diagram of the proposed PV power forecasting model. 150

5.9 The used forecasting procedure using the immediate history values. 151

5.10 Forecasting using past values from previous days for the same forecasted hour. 151

5.11 Real versus Forecasted values of PV power produced on 13, 14 and 15 of July. 152

List of Tables

2.1 Inclusion criteria used in our study. 55

2.2 Performance indicators used in the underlying studied works. 57

2.3 Our state of art of monthly solar irradiation estimation. 58

2.4 Our state of art of daily/hourly solar irradiation estimation. 59

2.5 Our state of art of solar irradiation forecasting using ANNs. 63

3.1 Input parameters available in our radiometric station. 73

3.2 Combination of inputs and their performances using the heuristic search model. 81

3.3 Individual performance of each ANN input parameter. 82

3.4 Performances of the stepwise regression forward selection method. 83

3.5 The performances and best input combinations versus max generations of the proposed model. 84

3.6 Population size versus max generation of the proposed model. 85

3.7 Performances of the proposed model with different cutting points positions and nature. 85

3.8 Effect of mutation percentage on the performance of our model. 86

3.9 Comparison of the implemented models with our proposed model in terms of performance indicators and found inputs. 86

3.10 Typical and adjusted coefficients for the three implemented empirical models. 89

3.11 Performances of the implemented empirical models. 89

3.12 Performance of the implemented kNN model with k=1 for different combinations of input parameters. 91

3.13 The performances of the kNN model with different number of neighbors (k). 91

3.14 Comparison of implemented models with our EANN model in term of estimation. 92

4.1 Geographical coordinates and their corresponding climate for all studied cities. 100

4.2 Signification of the abreviation used in Table 4.1. 100

4.3 Distribution of the studied Moroccan cities according to the thermal climatic zoning. 101

4.4 Global solar irradiance forecasting performances of HAEANN 1 using different combinations of activation functions in terms of NRMSE(%) and NMAE(%). 110

4.5 Performances of HAEANN 1 according to different crossover operators for global solar irradiance forecasting. 111

4.6 Global solar irradiance forecasting performances of HAEANN 1 according to different population sizes. 112

4.7 Division of datasets for each studied city. 113

4.8 Global horizontal solar irradiance forecasting performances of the six proposed models for ten cities. $NRMSE$ and $NMAE$ rows represent test results. 114

4.9 Diffuse horizontal solar irradiance forecasting performances of the six proposed models for ten cities. $NRMSE$ and $NMAE$ rows represent test results. 115

4.10 Direct normal solar irradiance forecasting performances of the six proposed models for ten cities. $NRMSE$ and $NMAE$ rows represent test results. . . 116

4.11 Train and test datasets division of the studied cities for each considered zone. 117

4.12 Global horizontal solar irradiance forecasting performances of the six proposed models for the six considered zones. $NRMSE$ and $NMAE$ rows represent test results. 118

4.13 Diffuse horizontal solar irradiance forecasting performances of the six proposed models for the six considered zones. $NRMSE$ and $NMAE$ rows represent test results. 119

4.14 Direct normal solar irradiance forecasting performances of the six proposed models for the six considered zones. $NRMSE$ and $NMAE$ rows represent test results. 120

4.15 Constructed train and test datasets of the studied cities for all_cities scenario. 121

4.16 Global, diffuse and direct solar irradiance forecasting performances of the six proposed models for all_cities scenario. 121

4.17 Comparison of the proposed HAEANN 6 model with RT, SP and RF models for global solar irradiance forecasting for 10 cities in terms of NRMSE(%) and NMAE(%). 128

4.18 Comparison of the proposed HAEANN 6 model with regression trees, smart persistence and random forest models for diffuse solar irradiance forecasting for 10 cities in terms of NRMSE(%) and NMAE(%). 129

4.19 Comparison of the proposed HAEANN 6 model with regression trees, smart persistence and random forest models for direct normal solar irradiance forecasting for 10 cities in terms of NRMSE(%) and NMAE(%). 130

4.20 Comparison of the proposed HAEANN model with RT, SP and RF models for global horizontal solar irradiance forecasting for six zones in terms of NRMSE(%) and NMAE(%). 131

4.21 Comparison of the proposed HAEANN 6 model with RT, SP and RF models for diffuse solar irradiance forecasting for six zones in terms of NRMSE(%) and NMAE(%). 132

4.22 Comparison of the proposed HAEANN model with RT, SP and RF models for direct solar irradiance forecasting for six zones in terms of NRMSE(%) and NMAE(%). 133

4.23 Comparison of the proposed HAEANN model with RT, SP and RF models for global, diffuse and direct solar irradiance forecasting for all_cities scenario in terms of NRMSE(%) and NMAE(%). 134

5.1 Characteristics of the used PV module at STC. 146

5.2 Comparison of the proposed HAEANN 6 model with regression trees and random forest models for short term PV power forecasting in terms of NRMSE(%) and NMAE(%). 153

List of Abbreviations

AI	Artificial Intelligence
Ang_c	Ångström coefficient
Alt	Altitude
AC	Alternating current
ANN	Artificial Neural Network
CC	Cloud cover
CI	Clearness index
CI_m	Mean clearness index
D_d	Day duration
D_l	Day length
DC	Direct current
Dec_a	Declination angle
DE_{ces}	Diffuse energy component
Dl_h	Daylight hours
EANN	Evolutionary artificial neural network
GHI	Global horizontal solar irradiation
GI	Global solar irradiation
GI_d	Daily global solar irradiation
GNI	Global normal irradiation
GI_{md}	mean daily solar irradiation
h_d	hour of the day
H_{sl}	Height above sea level
ω	Hour angle

$K_{b,cs}$	Direct clear sky index
$K_{d,cs}$	Diffuse clear sky index
$K_{g,cs}$	Global clear sky index
kNN	k Nearest Neighbors
Lat	Latitude
LC	Lost component
Loc	Location
Loc_t	Local time
LW_{ae}	Long wave atmospheric emission
Mon	Month
N	Number of hidden neurons
P_a	Atmospheric pressure
Pa_{max}	Maximum atmospheric pressure
Pa_{min}	Minimum atmospheric pressure
Per	Period
PM	Particulate matter
P_m	Mean pressure
P_p	Partial pressure
R_d	Rainy days
Rh_m	Mean relative humidity
SC	Sky cover
SD_m	Mean Sunshine duration
SD_{theo}	Theoretical Sunshine duration
SD_{max}	Maximum Sunshine duration
SD_{mes}	Measured Sunshine duration
S_p	Sunshine percentage
TL_f	Link turbidity factor
T_m	Mean temperature
VP_m	Mean vapor pressure
WS_e	Evaporation wind speed
Z_a	Zenithal angle

Contents

Aknowledgement 1

Dedication 3

Abstract 4

List of Figures 7

List of Tables 11

List of Abbreviations 15

Introduction 22

1 Solar radiation generalities and its applications 28

1.1 Introduction . . . 29

1.2 Sun-Earth Astronomical Formulas . . . 29

1.2.1 Earth and Celestial Coordinate Systems . . . 32

1.2.2 Determination of Sun's position . . . 33

1.2.2.1 Equatorial coordinate system . . . 33

1.2.2.2 Horizontal coordinate system . . . 34

1.2.3 Sunrise, Sunset and day length . . . 35

1.3 Solar radiation . . . 36

1.3.1 Solar radiation at the top of the atmosphere . . . 36
1.3.2 Terrestrial solar irradiation . . . 37
1.4 Conclusion . . . 37

2 Systematic review of solar irradiation estimation and forecasting using ANNs 39

2.1 Introduction . . . 40
2.2 Overview on Artificial Neural Networks . . . 40
2.2.1 Definition . . . 41
2.2.2 From the physiological to the mathematical model . . . 41
2.2.2.1 Physiological Consideration . . . 42
2.2.2.2 Origin of nervous system power . . . 42
2.2.2.3 Mathematical model . . . 42
2.2.3 Learning Algorithms . . . 43
2.2.3.1 Definition of learning . . . 43
2.2.3.2 Learning algorithms classification . . . 44
2.2.3.3 The single or monolayer Perceptron . . . 44
2.2.3.4 The XOR problem and limitations of single-layer models . . . 46
2.2.3.5 Multilayer Perceptron and Backpropagation Algorithm . . . 46
2.2.3.6 Neural network architecture optimization . . . 52
2.3 Systematic review of solar irradiation estimation and forecasting using ANNs 55
2.3.1 Inclusion criteria . . . 55
2.3.2 Distribution of literature papers . . . 56
2.3.3 Statistical indicators used in literature . . . 56
2.3.4 State of art of solar irradiation estimation using ANNs . . . 56
2.3.5 State of art of solar irradiation forecasting using ANNs . . . 61
2.4 Discussion and recommendations for future research works . . . 64
2.5 Conclusion . . . 65

3 Automatic selection of input parameters using evolutionary ANNs for daily solar irradiation estimation 66

3.1 Introduction . . . 67

3.2 Theoretical backround on genetic algorithms . . . 69

3.2.1 Genetic operators . . . 70

3.2.1.1 Solution representation (encoding scheme) . . . 70

3.2.1.2 Initial population . . . 70

3.2.1.3 Evaluation function (Fitness) . . . 70

3.2.1.4 Selection and Replacement Strategy . . . 71

3.2.1.5 Crossover . . . 71

3.2.1.6 Mutation . . . 71

3.3 Data and location . . . 71

3.4 Proposed EANN inputs selection model . . . 72

3.4.1 Model 1: ANN inputs selection based on heuristic search method . 74

3.4.2 Model 2: ANN input selection based on forward stepwise regression 74

3.4.3 The proposed model: ANN input selection based on evolutionary algorithms . . . 74

3.4.3.1 Coding neural network individuals . . . 75

3.4.3.2 Evaluation function . . . 77

3.4.3.3 Selection scheme . . . 77

3.4.3.4 Crossover process . . . 78

3.4.3.5 Mutation . . . 78

3.4.3.6 Block diagram and algorithm of the proposed method . . 79

3.5 Experimental results and discussion . . . 80

3.5.1 Performance indicators for model 1 . . . 81

3.5.2 Performance indicators for model 2 . . . 81

3.5.2.1 Individual performance . . . 81

3.5.2.2 Forward stepwise regression selection method performances 82

3.5.3 Performance indicators for the proposed model 83
3.5.3.1 Effect of cutting points number and nature 84
3.5.3.2 Effect of mutation percentage 85
3.6 Esimation performance comparison with other methods 87
3.6.1 Implemented empirical models 88
3.6.2 k Nearest Neighbors (kNN) model 89
3.7 Conclusion 92

4 Short term solar irradiance forecasting via a novel evolutionary multi-model framework and performance assessment based on a zoning strategy 94

4.1 Introduction 95
4.1.1 Background 95
4.1.2 Aim of the study 96
4.2 Data preparation 99
4.3 The proposed framework and models structure 102
4.3.1 The proposed framework 102
4.3.2 Models description 104
4.3.3 Genotype coding 105
4.3.4 Fitness function and selection strategy 106
4.3.5 Crossover operator 107
4.3.6 Mutation operator 108
4.4 Training procedure 109
4.5 Experimental results 109
4.5.1 Global, diffuse and direct solar irradiance forecasting using the proposed models 110
4.5.1.1 The choice of the activation functions 110
4.5.1.2 The choice of the crossover operator 111

4.5.1.3 The choice of the population size . . . 111
4.5.1.4 Performances of the proposed models for individual_city scenario . . . 113
4.5.1.5 Performances of the proposed models for zoning scenario 117
4.5.1.6 Performances of the proposed models for all_cities scenario 119
4.5.1.7 Seasonal behavior . . . 122
4.5.2 Comparison of the proposed models with smart persistence and machine learning techniques . . . 123
4.5.2.1 Smart persistence . . . 124
4.5.2.2 Regression trees . . . 125
4.5.2.3 Random forest . . . 126
4.5.2.4 Comparison results . . . 127
4.6 Conclusion . . . 135

5 Evolutionary framework for short term photovoltaic power forecasting 137

5.1 Introduction . . . 138
5.2 From solar irradiation to renewable energy production . . . 139
5.2.1 Solar Photovoltaics . . . 141
5.2.1.1 Photoelectric effect . . . 141
5.2.1.2 Photovoltaic conversion . . . 142
5.2.1.3 Photovoltaic systems . . . 144
5.3 System description . . . 145
5.4 The proposed model for PV power production forecasting . . . 150
5.5 Experimental results . . . 152
5.6 Conclusion . . . 153

Conclusion and perspectives 155

Journal and Conference Publications 178

Introduction

$\mathcal{F}ossil$ fuels have been the foundation of the global energy system, economic growth and modern lifestyles. The exploitation of fossil fuels lifted global energy use fifty-fold in the last two centuries, shaping the geopolitical environment of the modern world. The geographic concentration of fossil fuels has had a significant impact on the wealth and security of nations.

Following the transformation from wood to coal then to oil and gas, the last decades are witnessing the ongoing third major transformation from oil and gas to renewable energy. Historically, the progress of human civilization accelerated the development of the coal industry, and coal accounted for the largest share in primary energy mix in the 1780s, surpassing wood for the first time. This was the first transformation from wood to coal. In 1886, Daimler invented the internal combustion engine, stimulating a great increase in the demand for oil and gas as efficient energy resources. Progress in geological theory and drilling, completion and refining technologies drove oil and gas production to increase substantially. Accordingly, the share of oil and gas in primary energy mix grew rapidly to more than 50% in 1965. These energy resources replaced coal as the largest energy in the world, recording the second transformation from coal to oil and gas. Today, the production of all types of fossil fuels has risen dramatically in the late 20th and early 21st centuries because of the rapid growth of global energy consumption, driven by an expanding population and increasing prosperity at the same time [1]. With this sustained increase in the economic and social demand for energy and the advent of low-carbon society, the third major transformation from traditional fossil fuel to nonfossil new energy became inevitable [2].

Similar to many other African countries [3, 4], Morocco is facing major challenges regarding the future of its energy system [5], particularly in terms of the electricity sector. The Moroccan policy making process is confronted with substantial increases in energy

demand [6]. Due to the strong dependence on imported energy sources, mostly on fossil fuels, Morocco has been placed in a very energetically insecure and precarious financial situation. Thus, it has become more vulnerable to the effects of global climate change. Morocco mainly imports electricity from Algeria and Spain. The current energy system in Morocco depends on hydrocarbons. Its demand in terms of energy is constantly expanding and it is expected to triple by 2030. Unfortunately, Morocco is the only North African country without nearly any fossil reserves [7]. However, the country has an abundance of renewable resources, especially solar and wind power. On a global scale, Morocco is in fact the 23th most attractive country for RE today according to the RE barometer « Renewable Energy Country Attractiveness Index » issued in 2020. The exploitation of this potential would solve two problems of the country: energy security and climate change problems.

In 2009, Morocco adopted a National Energy Strategy as a roadmap for the transition to a low-carbon energy system that reconciles economic development and social and environmental objectives. The Energy Strategy set five priorities for the energy policy of Morocco: 1) they relate to the development of a diversified and optimized energy mix, 2) the mobilization of domestic resources, including the use of renewable energy (wind/solar), 3) making energy efficiency a national priority, 4) stronger regional co-operation with Europe's and Africa's energy markets, 5) and industrial integration, which promotes the development of local industrial capabilities at all levels of the green technology value chain [8]. This National Energy Strategy set out an ambition for 42% of the total installed power capacity to come from renewable energy in 2020. This was expected to require the commissioning of new plants to bring the total capacity to 2000 MW of solar, 2000 MW of wind and 2000 MW of hydro by 2020.

In 2015, during the 21st session of the UNFCCC's Conference of the Parties (COP21), Morocco announced a further planned increase in the renewables capacity to reach 52% of the total by 2030 (20% solar, 20% wind, 12% hydro). To meet the 2030 target, the country aims to add around 10 GW of RE capacities between 2018 and 2030, consisting of 4560 MW of solar, 4200 MW of wind, and 1330 MW of hydropower capacity [8].

To fulfill these ambitious plans, the Moroccan government has launched energy reforms to foster the development of the country's industry in the sectors of renewable energy and its integration into existing and eventually emerging grids [9] to penetrate regional and international markets.

The biggest challenge in the penetration of renewable energy whether in the case of isolated sites or into the grid is the high variability of the output power of photovoltaic panels and solar thermal power systems, which is strongly affected, by the variability of the solar resource [10]. Hence, higher penetration capacity needs the perfect knowledge of the availability and variability of solar irradiation intensity, which could be a decision making tool for producers and grid operators, since it is a critical for any solar device such sizing photovoltaic systems, analysis of the thermal load on buildings, environmental impact analysis and agriculture [11, 12, 13, 14, 15, 16], and last but not least, prediction errors can have economic consequences on electricity markets [17, 18, 19].

Accordingly, the main objectives of the present thesis is to develop more accurate models for the estimation and forecasting of global solar irradiation and its components in Morocco. In fact, several solar irradiance prediction techniques have been proposed in the literature from stochastic, statistic, physical, empirical models and lately, Artificial Intelligence (AI) models [15]. Among AI techniques, artificial neural networks are well known machine learning systems that have been widely used in many fields of application, including control systems, pattern recognition and classification [20]. The choice of the ANN technique was motivated by their particular capability to modelize complex, non-linear and time varying input-output systems [21] due to their simplicity, generality and good learning ability [22]. Particularly, ANNs have been efficiently used as solar irradiation predictors in different sites over the world taking into account the non-linear relationship between inputs (meteorological, geographical, radiometric and atmospheric parameters) and output (direct, diffuse and global solar irradiation).

In the same context, forecasting of electricity production from a solar system, is very important to help the grid operators managing the electric balance between power demand and supply, to improve embedding of distributed renewable energy sources and, in stand-alone hybrid systems, for the optimum size of all its components and to improve the reliability of the isolated systems. Hence, the ability to accurately predict the generated solar power is a task of utmost importance and relevance for both energy managers and electricity traders, in order to minimize uncertainty and ensure reliable electricity supply at acceptable cost.

In the present thesis, our first contribution begins with a systematic review of solar energy estimation and forecasting using artificial neural networks. Therefore, different research papers based on ANN models used to predict solar radiation were analyzed and

discussed. The objective is to highlight some problematics, to extract research limitations in this field, and to propose recommendations and perspectives for future investigations.

Based on this state of art study, our second contribution focuses on one of the important extracted problematics which is the selection of most relevant input variables for ANN models. This problematic is formulated as a search problem, which can be resolved with Evolutionary Artificial Neural Networks (EANN). In fact, there has been a great interest in recent years in considering learning using ANNs and evolution through evolutionary algorithms. This combination is more effective and efficient and can lead to significantly better intelligent systems instead of relying on each method separately [23]. The key of success of such hybrid algorithms is the incorporation of a local search procedure provided by ANNs into the global one achieved by evolution.

The purpose is to use EANNs to produce an automatic methodology of ANN inputs selection (from a set of available data) for the estimation of daily horizontal global solar irradiation for the city of Fez-Morocco, by developing a model with:

- commonly measured variables like temperature, humidity and rainfall,
- good accuracy that is crucial for solar energy devices.

Our third contribution is dealing with the most important issues using ANN models in solar irradiance forecasting, which are the choice of the optimal number of hidden layers, hidden nodes and input nodes providing better accuracy. In fact, as conventionally known, ANN's architecture has a significant impact on the network's information processing capabilities [24]. Architecture design is then crucial in the successful application of ANNs. Up to now, architecture design is still very much a human expert's job [23, 25]. This manual neural network design is something of black art [26] and it can be assumed that most of such network designs are not optimal.

Hence, a new optimization method of ANN architecture through Evolutionary Algorithms (EAs) is proposed in order to determine automatically the near-optimal topology of the ANN forecaster. Besides, the purpose of this work is to forecast the upcoming hourly global, diffuse and direct solar irradiance at different time horizons based only on endogenous historical data, manipulating data from 28 Moroccan cities with different climatic conditions. The resulting model is baptized HAEANN (History and Architecture EANN) allowing the following contributions:

- the generation of multiple HAEANN models from a multi-model framework for

different time horizons to forecast global, diffuse and direct solar irradiance from 1 to 6 hours ahead. Each of which evolves simultaneously forecasting history and ANN architecture, searching automatically for the most accurate model,

- a useful zoning scenario is investigated presenting a particular interest in the case of the lacking of solar irradiance measurements in some areas, and where future values of solar irradiance can be obtained using data from stations belonging to the same zone.

The fourth contribution within the context of this thesis, is inspired from the aforementioned developed framework for the forecast of the generated power of an existing photovoltaic system at the National School of Applied Technology of Fez. This contribution is considered as a first step for our future research works aiming to accurately forecast solar power production from an optimized PV system.

This thesis is organized into five chapters. Their contents are outlined as follows:

Chapter 1: This chapter is designed to be a concise introduction to discuss solar radiation fundamentals, including energy irradiated from the Sun, the geometrical relationship between Sun and Earth, in order to establish and use predictive models of solar irradiation and its components received on the ground.

Chapter 2: In this chapter, theoretical background on Artificial Neural Networks is firstly presented, followed by a focused systematic review of selected papers in the literature on solar irradiation estimation and forecasting using ANNs, including a meticulous and updated list of researches in this field up until 2020. The aim is to point out the research limitations and to propose recommendations and outlooks for our future research projects.

Chapter 3: This chapter describes the proposed model for the automatic selection of ANN inputs from available parameters set for the estimation of horizontal daily global solar irradiation in the city of Fez (Morocco). This selection is carried out using the Evolutionary Artificial Neural Networks (EANN) with different genetic operators including selection, crossover and mutation taking into account the coding of ANN as an individual in the pool of evolution. The study includes a comparison study in order to validate the efficiency of the proposed EANN model on the selection of input parameters, using an implemented heuristic search and forward stepwise regression models. To compare the performances in term of estimation of the proposed model, four temperature-based and k

Nearest Neighbors (kNN) models are implemented and tested.

Chapter 4: This chapter aims to propose a new evolutionary framework to generate multiple forecasting models based on optimized ANNs in order to accurately predict global, diffuse and direct solar irradiance. The proposed framework offers different time horizons up to 6 hours ahead by the evolution of the forecasting history and ANN architecture. A dataset of 28 Moroccan cities is used in our experiments in order to explore the performances of the proposed models in different climatic conditions. The proposed framework is then evaluated through a zoning scenario giving the ability to our models to accurately forecast solar irradiance and its components in sites where no such data is available. Two other scenarios are used to assess and compare the resulting performances. A comparative study is then conducted with a reference model: smart persistence and two other machine learning models: regression trees and random forest.

Chapter 5: In this chapter, a first attempt for solar power forecasting using machine-learning techniques is presented. The proposed EANN approach is used to forecast short term (up 6 hours ahead) solar power output of an optimized photovoltaic system placed at the National School of Applied Sciences of Fez. Results of the proposed model are compared to two machine learning models: regression trees and random forest.

Chapter 1

Solar radiation generalities and its applications

1.1 Introduction

Solar radiation is the most important natural energy resource because it drives all environmental processes acting at the surface of the Earth. The Sun provides the Earth with an enormous amount of energy. The energy stored by the oceans helps maintain the temperature of the Earth at an equilibrium level that allows for stability for a broad diversity of life.

The amount and intensity of solar radiation reaching the Earth's surface depends on the geometric relationship of the Earth with respect to the Sun. In order to understand solar energy, this chapter is designed to be a concise introduction to discuss solar radiation fundamentals, including energy irradiated from the Sun, the geometrical relationship between the Sun and the Earth, as well as the importance of acquiring reliable solar information for engineering design, operation, and management of solar technologies.

1.2 Sun-Earth Astronomical Formulas

The Sun is a typical middle-aged star with a diameter of 13.927×10^5 km, a mass of 1.989×10^{30} kg and a luminosity of 4×10^{26} W. The Sun is primarily composed of 70% hydrogen and 28% helium. The enormous amount of energy radiated from the Sun derives from the extremely high temperatures within its different layers. The Sun radiates throughout the entire electromagnetic spectrum from the shortest x-rays to long-wavelength radio waves. Thereby, the greatest amount of the radiation falls in the visible range, and the shape of the solar spectrum is quite similar to a blackbody spectrum for an effective temperature near 5.8×10^3 K [27]. This emitted energy is called solar radiation.

Apart from geothermal energy, all forms of energy available in the Earth-atmosphere system result from the transformation of solar radiation. At the top of the atmosphere, the distribution of this energy is not homogeneous because the Earth is orbiting the Sun and the inclination of the Earth's equatorial plane relative to the orbital plane (varying according to latitude and seasons).

Before setting out the developed models for modeling solar radiation, it is necessary to formulate the definitions and elements essential to the proper understanding of its foundations by means of methods for calculating the balance of solar radiation at a point on the Earth's surface.

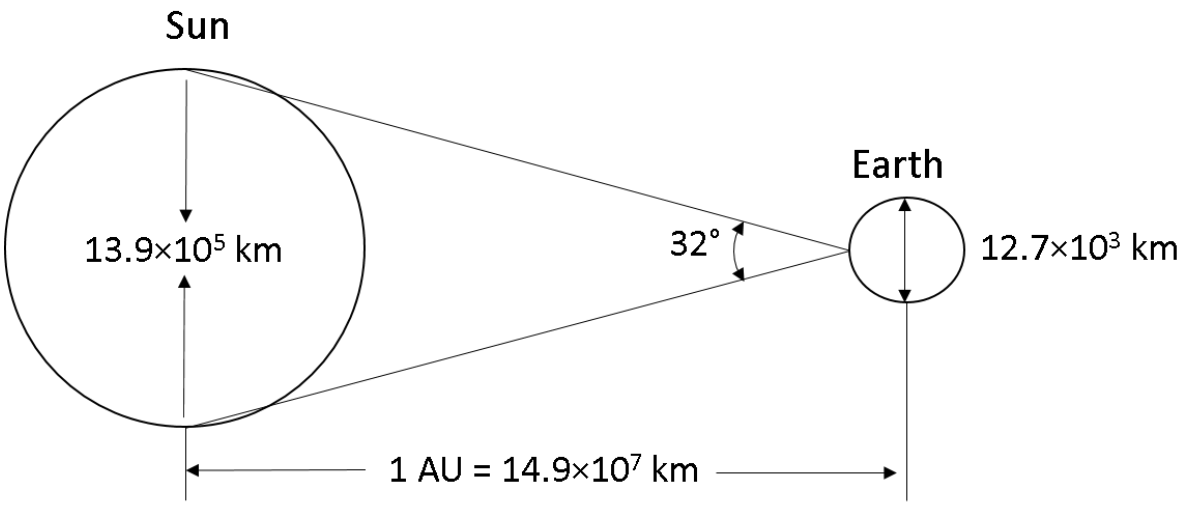


Figure 1.1: Sun-Earth geometric relationships.

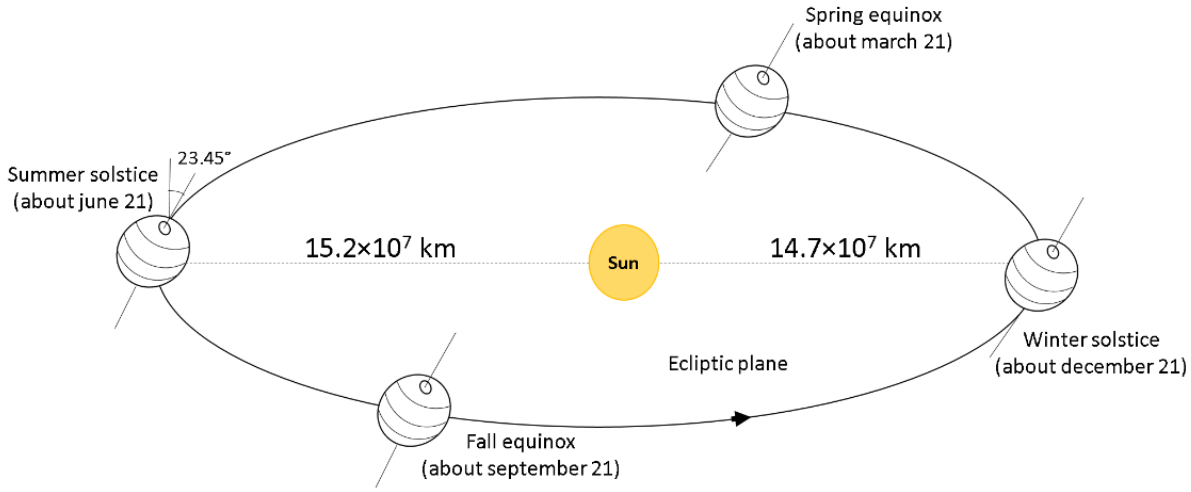


Figure 1.2: Movement of the Earth's rotation around the Sun.

The Sun is about 15×10^7 km away from Earth, so, because thermal radiation travels with the speed of light in vacuum (about 300000 km/s), after leaving the Sun, solar energy reaches our planet in 8 minutes and 20 seconds. As observed from the Earth, the Sun disk forms an angle of 32 min of a degree (Figure 1.1) [28].

The amount and intensity of solar radiation reaching the Earth's surface depends on the geometric relationship of the Earth with respect to the Sun. Figure 1.2 shows this geometric relationship and its effects for different seasons in both hemispheres.

The position of the Sun, at any moment at any place on Earth, can be estimated by two types of calculations: first, by simple equations where the parameters are the day of the year, time, latitude, and longitude, and secondly, by calculations through complex

algorithms providing the exact position of the Sun. Mostly, such algorithms are valid for a limited period varying from 15 to 100 years.

This section includes only calculations from geometry in order to understand the nature of the variant incoming solar radiation.

The Earth has two primary motions, rotation and revolution. The first is linked to the Earth's movement of revolution around the Sun of annual frequency (it takes 365.25 days (one year) to revolve around the Sun), which allows us to understand the phenomenon of the seasons [27]. The second movement is related to the rotation of the Earth on itself around the axis of the poles with a periodicity of 24 hours. The latter is at the origin of the succession of days and nights. It is of course possible to precisely determine the position of the Sun as a function of time and the position of the observer on the Earth.

The Earth revolves around the Sun according to an elliptical trajectory of eccentricity 0.0167. The average Earth–Sun distance of 14.9×10^7 km is defined as the Astronomical Unit (AU), which is used for calculating distances within the solar system. However, Earth orbits the Sun in an elliptic orbit at an average distance of about 150 million km. Due to the elliptic orbit, the speed of Earth is not constant. It reaches its maximum at the summer solstice (aphelion: 1.017 AU or 15.21×10^7 km) and its minimum at the winter solstice (perihelion: 0.983 AU or 14.7×10^7 km). It depends on the number j of the day of the year which varies from 1 to 365 (or 366 for leap years). The Sun subtends an angle of 32° on the Earth at a 1 AU distance.

The Sun-Earth distance r is presented in equation 1.1 [29].

$$\frac{r}{r_0} = 1 + \epsilon \tag{1.1}$$

ϵ designate the relative correction depending on the considered day, which can be approximated using the daily angle ω_j defined in equation 1.2 as [30]:

$$\omega_j = \frac{j \times 2 \times \pi}{365.2422} \tag{1.2}$$

Using equation 1.2, we can redefine the ratio $\frac{r}{r_0}$ with [30] :

$$\frac{r}{r_0} = 1 + 0.3344 \times cos(\omega_j - 0.049) \tag{1.3}$$

1.2.1 Earth and Celestial Coordinate Systems

Angular relations between Sun-Earth requires knowledge to locate a point on Earth's surface. The location of any feature on the surface of the Earth is described by two angles, latitude (φ) and longitude (λ). Latitude φ is defined as the angle formed by the intersection of a line perpendicular to the Earth's surface at a point and the plane of the Equator. Points north of the Equator have positive latitude values, while points south have negative values.

- Latitude values range from -90 to +90°. A meridian, or line of longitude, is formed by a plane that passes through the point and the North and South poles.
- The longitude value is defined by the angle (ranging from -180 to +180°) between the meridian of the specified location and the prime meridian (Greenwich meridian).

Figure 1.3 represents a location P located at the northern hemisphere of the Earth, which is identified by the two parameters: latitude (φ) and longitude (λ)

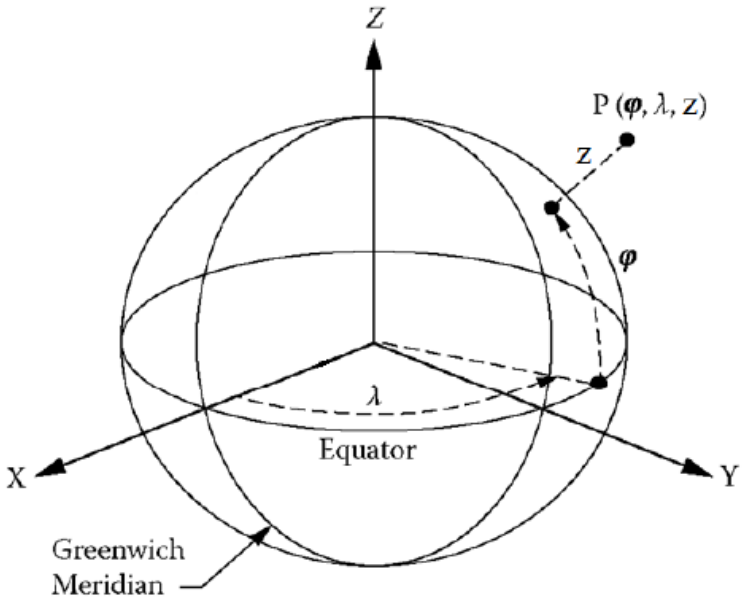


Figure 1.3: Celestial coordinate system for a location P on Earth.

1.2.2 Determination of Sun's position

1.2.2.1 Equatorial coordinate system

In equatorial coordinates, Sun position is determined using solar declination (δ) and the hour angle (ω).

- **The solar declination (δ)**: if the Sun is theoretically observed from the center of the Earth during the annual circulation of the Earth around the Sun, the angle between the Sun's directions with the equatorial plane of the Earth constantly changes. This angle is called the solar declination (degrees), and it varies from -23°27' in the winter solstice to +23°27' in the summer solstice, while at the equinoxes it is null in the northern hemisphere.

  Expressions giving the approximate value of the solar declination with varying degrees of accuracy have been developed by a simple approximation by Cooper (1969) [31] or more precisely by Fourier series according to Spencer (1971) [32].

$$\delta = 23.45 \sin(\frac{284}{365}(j + 284)) \tag{1.4}$$

$$\begin{aligned}\delta = \frac{180^\circ}{\pi} \times (0.006918 - 0.399912 \cos\Gamma + 0.070257 \sin\Gamma - \\ 0.006758 \cos 2\Gamma + 0.000907 sin 2\Gamma - 0.00269 \cos 3\Gamma + 0.00148 \sin 3\Gamma)\end{aligned} \tag{1.5}$$

  with :

$$\Gamma = 2\pi(j - 1)/365 \tag{1.6}$$

- **Angle Hour (ω)** : is defined for a given point on the Earth's surface as the angle between the meridian plane passing through the Sun's center and the vertical plane of the given place (meridian). It is null at local solar noon, afternoon hours being designated as positive and negative before noon.

  The hour angle can be calculated by the below equation that defines the true solar time T_{SV} (hours) in 12:00 h at the highest position of the Sun across the sky during the day.

$$\omega = 15(T_{SV} - 12) \tag{1.7}$$

  The relation between the solar time T_{SV} and the standard time T_l (Local Time) is giving by the relationship:

$$T_{SV} = T_l - DT_l + D_{hg} + E_t \tag{1.8}$$

T_l (hours) : the local time,

DT_l (hours) : the difference between the local and standard time (time zones),

D_{hg} (hours): the time difference (advance of 4 minutes per degree),

E_t (hours): the equation of time.

The values of the equation of time as a function of day's number (j) of the year (N) can be obtained from a simple used formula given in equation 1.9:

$$E_t = \frac{1}{60}(9.87 \sin 2N - 7.35 \cos N - 1.5 \sin N) \tag{1.9}$$

with :

$$N = \frac{360}{365.25}(j - 81) \tag{1.10}$$

The time difference or longitude correction D_{hg} is given as:

$$D_{hg} = \frac{4\lambda}{60} \tag{1.11}$$

λ: the longitude of the specified location.

It should be noted that the longitude is positive if the local meridian is east of the standard meridian and it is negative west of the standard meridian [33, 34].

1.2.2.2 Horizontal coordinate system

The horizontal coordinate system, also known as the Alt/Az system, is a method for describing the exact position of the Sun in the sky. As shown in Figure 1.4, this system is based on two coordinates : solar altitude (SA) and solar azimuth (a).

- **Solar altitude (SA)**: or solar elevation is the angle (measured in degrees) between the horizontal plane and the Sun's direction. The value SA = 0° is at the Sunrise and Sunset. It varies between 90° (zenith) and −90° (nadir). The mathematical expression for solar altitude is given by equation 1.12 [35]:

$$\sin SA = \sin \varphi \times \sin \delta + \cos \varphi \times \cos \delta \times \cos \omega \tag{1.12}$$

It is related to the solar zenith angle θ_z (in degrees), which is the angle between the Sun's direction and the normal (vertical). Where:

$$\cos \theta_z = \sin SA \tag{1.13}$$

and

$$\theta_z + SA = \frac{\pi}{2} = 90° \tag{1.14}$$

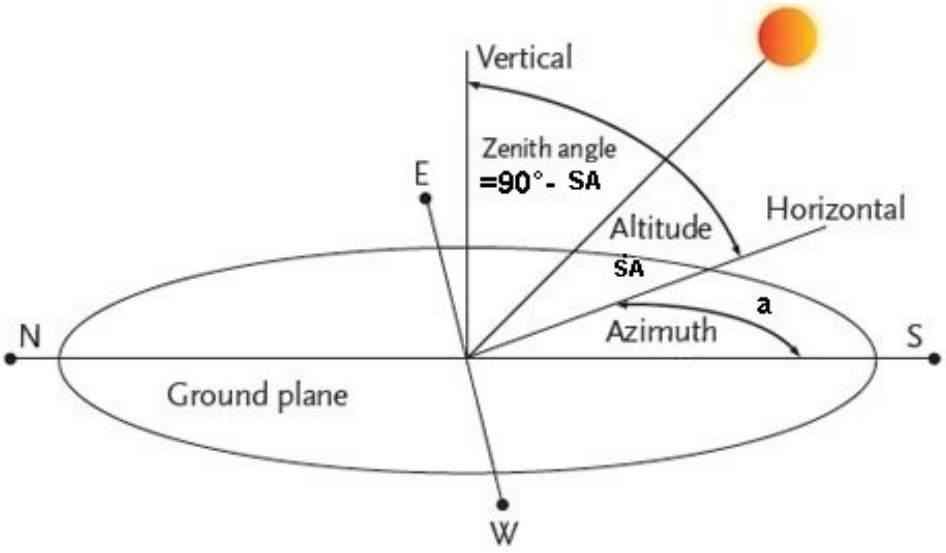


Figure 1.4: Illustraton of the horizontal coordinate system.

- **Solar azimuth (a)** : is the angle of the Sun's rays measured in the horizontal plane between the projection of the beam radiation and the north–south direction line. The mathematical expression for the solar azimuth angle is given by equation 1.15 [35]:

$$\sin a = \frac{\cos\delta \times \sin\omega}{\cos SA} \tag{1.15}$$

1.2.3 Sunrise, Sunset and day length

Day length, is the time each day from the moment the upper limb of the Sun's disk appears above the horizon during Sunrise to the moment when the upper limb disappears below the horizon during Sunset. Day length varies throughout the year as the Earth–Sun geometric relationships change. Such geometrical changes are clearly perceived by the apparent movement of the Sun in the sky during the year [36].

The Sun is said to rise and set when the solar altitude angle is 0. So, the hour angle at Sunset ω_{ss}, can be found by solving Equation 1.12 for ω, when SA=0°.

$$\sin(SA) = \sin(0) = 0 = \sin\varphi \times \sin\delta + \cos\varphi \times \cos\delta \times \cos\omega_{ss} \tag{1.16}$$

$$\cos\omega_{ss} = -\frac{\sin\varphi \times \sin\delta}{\cos\varphi \times \cos\delta} \tag{1.17}$$

which reduces to :

$$\cos\omega_{ss} = -\tan\varphi \times \tan\delta \tag{1.18}$$

Since the hour angle at local solar noon is 0°, with each 15° of longitude equivalent to 1 hour, the Sunrise and Sunset time in hours from local solar noon is then the hour angle at Sunrise is:

$$\cos\omega_{sr} = -\omega_{ss} = \frac{1}{15}\cos^{-1}(-\tan\varphi \times \tan\delta) \tag{1.19}$$

The day length (D_l) is twice the Sunset hour, since the solar noon is at the middle of the Sunrise and Sunset hours. Therefore, the length of the day in hours is formulated as [32]:

$$D_l = \frac{2}{15}\cos^{-1}(-\tan\varphi \times \tan\delta) \tag{1.20}$$

1.3 Solar radiation

1.3.1 Solar radiation at the top of the atmosphere

The amount of solar radiation that is intercepted by the Earth varies because of variations in Sun-Earth distance and, to a much lesser extent, day-to-day variations in the spectrum due to solar activity. The closer the Earth to the Sun, the greater the solar irradiance impinging on a plane normal to the Sun rays and located at the top of the atmosphere.

Throughout the year, the extraterrestrial radiation measured on the plane horizontal to the radiation $G_{toa,h}$, varies in the range of 3.3%. An accurate equation provided by Spencer (1971) [32] is given in Equation 1.21

$$G_{toa,h} = G_{sc} \times E_0(\sin\varphi \times \sin\delta + \cos\varphi \times \cos\delta \times \cos\omega) \tag{1.21}$$

G_{sc} is the solar constant($G_{sc} = 1367W/m^2$), representing the amount of solar energy per unit time, at the mean distance of the Earth from the Sun, received on a unit area of a surface normal to the Sun (perpendicular to the direction of propagation of the radiation) outside the atmosphere.

The eccentricity factor (E_0) is calculated using equation 1.22:

$$E_0 = 1.000110 + 0.034221\cos\Gamma + 0.001280\sin\Gamma + 0.000719\cos 2\Gamma + 0.000077\sin 2\Gamma \tag{1.22}$$

The total radiation, incident on an extraterrestrial horizontal surface during a day can be obtained by the integration of Equation 1.21 over a period from Sunrise to Sunset. The resulting equation is:

$$G_{toa,j} = \frac{G_{sc} \times E_0}{\pi}(\cos\varphi \times \cos\delta \times \cos\omega_{ss} + \frac{\pi \times \omega_{ss}}{18}\sin\varphi \times \sin\delta) \tag{1.23}$$

1.3.2 Terrestrial solar irradiation

The actual amount of solar radiation that reaches a particular place on the Earth is extremely variable. In addition to the regular daily and annual variation due to the apparent motion of the Sun, irregular variations have to be taken into account that are caused by local atmospheric conditions, such as clouds. These conditions particularly influence the direct and diffuse components of solar radiation. The direct component of solar radiation is that part of the Sunlight that directly reaches the surface. Scattering of the Sunlight (by atmospheric constituents, such as clouds and dust) in the atmosphere generates the diffuse component. Reflected radiation is reflected from surface features (see Figure 1.5). We use the term global radiation to refer to the total solar radiation, which is made up of these three components.

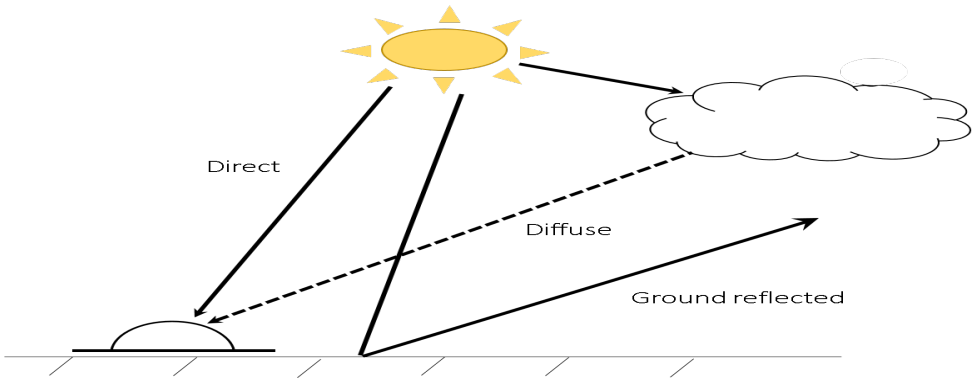


Figure 1.5: The three main components of solar irradiation.

1.4 Conclusion

This chapter concluded the prior knowledge on the main astronomical equations that had to be acquired in order to establish and use predictive models of solar irradiation and its components received on the ground, showing the influence of the terrestrial atmosphere on the received solar irradiation on any location on Earth.

From the above, we can certainly assume that accurate knowledge of the available solar irradiation is a key issue in any solar energy system. However, long-standing records of solar irradiation data are not available in many places because of the cost and maintenance of the measuring instruments. Thus, different type of estimating models including Artificial Intelligence (AI) techniques are becoming useful as alternate approaches for estimating and forecasting solar irradiation and its components specifically, Artificial Neural

Networks (ANNs) which will be developed in our work.

In the next chapter, we will present theoretical background on ANN techniques with a meticulous and updated state of the art on the use of ANNs in the field of solar irradiation estimation and forecasting.

Chapter 2

Systematic review of solar irradiation estimation and forecasting using ANNs

2.1 Introduction

Solar irradiation data is of a great importance in solar energy research and engineering. In fact, the perfect knowledge of the availability and variability of solar irradiation intensity is very critical, especially for solar devices such as flat plate collectors [37], sizing photovoltaic systems, analysis of the thermal load on buildings, environmental impact analysis and agriculture [11, 12, 13, 14, 15, 16]. However, there is a limitation of number of meteorological stations that measure solar irradiation with reliable and calibrated pyranometers, due to the cost and maintain difficulty of solar irradiation measuring instruments [38, 39]. Moreover, most stations only have limited measured parameters such as relative humidity, temperature, wind speed and Sunshine duration. Hence, since these parameters are more readily available around the globe, they are used to estimate solar irradiation, as an alternative way, to generate these needed data using appropriate models. Since the first models [40, 41], Artificial Intelligence (AI) techniques are becoming useful as alternate approaches. Specifically, Artificial Neural Networks (ANNs) perform a computational simulation with a high potential to modelize complex, nonlinear, and time-varying input-output systems [21]. ANNs have been widely employed to estimate and to forecast solar irradiation using various meteorological and geographical parameters.

In this chapter, theoretical background on Artificial Neural Networks is firstly presented, followed by a focused review on selected papers in the literature on solar irradiation estimation and forecasting using ANNs. Our purpose is to include recent researches in this field up until 2020 with a meticulous list of important works. The aim is to help future researchers to easily identify published papers by specifying the prediction horizon, the used inputs, ANN architecture and the corresponding obtained performance indicators. Another contribution is to point out the research limitations and propose recommendations and outlooks for our future research projects.

2.2 Overview on Artificial Neural Networks

One of the most important approaches to automatic information processing is based on the processing of information by the brain. The main hypothesis behind the development of artificial neural networks is that the intelligent behavior is underpinned by a set of mental mechanisms. These mechanisms being based on neurophysiological processes, the

structure of the nervous central system is therefore the basis for the development of all intelligent behavior.

In the field of artificial neural networks two major paradigms can be distinguished: connectionism and neuromimetics. It should be noted that, as Touzet pointed out in [42], neural networks designate only mathematical models, it is neither a research field nor a scientific discipline. Connexionnism and neuromimetics are both fields of research which manipulate neural network models but with different objectives. The objective pursued by neural network engineers and researchers is to improve the capabilities of computer science by using models with highly connected components. On the other hand, neuromimeticians manipulate artificial neural network models in order to verify their biological theories of how the central nervous system works.

Since the present work will be centered around neural networks models for the estimation and forcasting of solar radiation, it is therefore important to address some concepts of neuroscience that can be very useful in the remaining of this thesis.

2.2.1 Definition

The artificial neural network (ANN) is an assembly of elementary units called neurons, each of which performs simple processing of an input set to produce an output. The response of the network is the output calculated on the output layer, also called the decision layer. According to Haykin [43], neural networks are parallel distributed processors that have the property of storing knowledge and making it available for use. Artificial neural networks resemble the human brain in two aspects:

- knowledge is acquired through a learning process,
- inter-neuronal connection weights, or synaptic weights, are used to store knowledge.

2.2.2 From the physiological to the mathematical model

In this subsection, we will review some aspects of the biological behaviour of the neuron. We then present the mathematical modeling of this behavior to arrive at the so-called artificial neuron, which is the subject of any design using the neural networks.

2.2.2.1 Physiological Consideration

The brain consists of about 10^{12} neurons (one thousand billion), with 1000 to 10000 synapses (connections) per neuron [44]. The neuron is a cell composed of a cell body and a nucleus. The cell body branches off to form dendrites. It is through the dendrites that information is carried from the outside to the body of the neuron. The information processed by the neuron then travels along the axon to be transmitted to other neurons. The transmission between two neurons is not direct. In fact, there is an inter-cellular space of a few dozen Angstroms (10^{-9}m) between the axon of the afferent neuron and the dendrites of the efferent neuron. The junction between two neurons is called the synapse.

2.2.2.2 Origin of nervous system power

The performance of the human brain can be explained by :

- the density of connections: each neuron communicates, on average, with 10^4 other neurons, and
- the parallel processing of information by arround 10^{11} neurons, and not by the individual performances theirselves.

2.2.2.3 Mathematical model

After several approximations, scientists obtain a mathematical modelling of the biological neuron as shown in Figure 2.1 [44]:

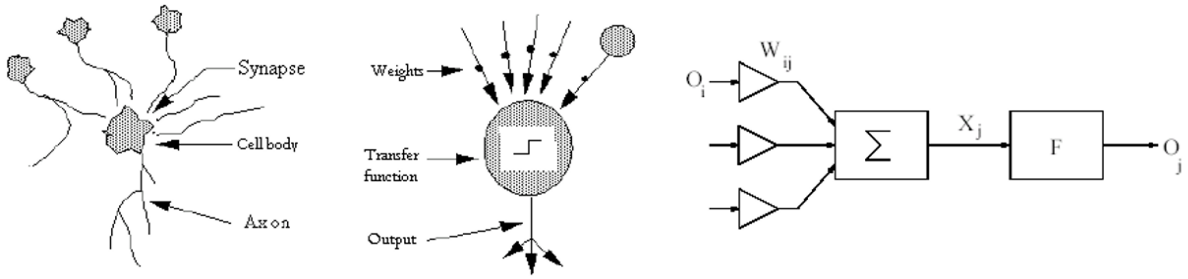


Figure 2.1: Approximation of the biological neuron by the artificial neuron.

The behaviour of the artificial neuron consists of two phases:

- the first phase, called activation, represents the calculation of the weighted sum of the inputs, also called post-synaptic potential, according to the following expression:

$$a_i = \sum_{j=1}(w_{ij}x_j) \tag{2.1}$$

- the second phase is the calculation of the transfer function F, which can take several different forms.

2.2.3 Learning Algorithms

In this subsection, we will present one of the most important properties of neural networks which is learning using different algorithms.

2.2.3.1 Definition of learning

From literature, there is two types of definitions of learning : from the physiological point of view and from the algorithmic one.

- Learning in the physiological sense is the mechanism by which the nervous system adapts to the outside world. This adaptation is mainly achieved by modifying synaptic connections. Unfortunately, the laws governing these modifications are very poorly understood.
- In the algorithmic sense, learning is a phase in the development of a neural network during which the behavior of the network is modified until the desired behavior is achieved. In order to carry out learning, algorithms use samples corresponding to observations on the system called "behavioral examples".

In the majority of current algorithms, the variables modified during learning are the connection weights. Learning is the modification of network weights in order to tune the network response to behavioural examples and experience. It is often impossible to decide a priori the values of the network connection weights for a given application.

At the end of the learning process, the weights are found. Then comes the use phase. Some network models are improperly called continuous learning models. In this case, it is true that learning never stops. Generally, we can always distinguish between a learning phase (updating the behaviour) and a use phase.

2.2.3.2 Learning algorithms classification

Two large classes have been defined [44] according to whether the learning is supervised or unsupervised. This distinction is based on the form of the learning examples.

- in the case of supervised learning, the behaviour examples are pairs (input, associated output).
- whereas for unsupervised learning only the values of the Inputs are available. On the other hand, we distinguish between two main types of learning algorithms according to the architecture of the used neural network :
- the single perceptron or single-layer perceptron algorithm [44] where the network architecture contains a single processing layer.
- the backpropagation algorithm [45] where the network used contains additional layers called "hidden layers".

In the following sections we will explain the details of these two main types of algorithms, as well as their limitations.

2.2.3.3 The single or monolayer Perceptron

Definition

The first neural network model was introduced by McCulloch and Pitts in 1943 [46]. The neurons in this model are binary. They receive excitatory and inhibitory signals from upstream neurons to perform summation. If this sum is greater than a threshold, the neuron is in an active state and sends a signal to the other neurons. If not, it is in an inactive state and does not send a signal.

The perceptron is called a monolayer because it contains only one processing layer.

Hebb Rule

After the description of this first neuron model, a way to perform learning was needed. Based on Pavlovian conditioning for animals, Donald O. Hebb introduced the first learning rule in 1949 [47], which allows synaptic weights to be modified. The principle of this rule, known as the Hebb rule, is that if two connected neurons are activated at the same time, then the connection between them is strengthened. Otherwise, it is not modified.

Single perceptron threshold principle

This learning rule was the starting point for the work of Rosenblatt [48] who developed the first version of a neural model with a learning algorithm that is well known today, namely the Perceptron. It is a network with two layers (an input layer and an output layer (see Figure 2.2)) of the feedforward type, i.e. the signals are propagated from layer l to layer $l+1$ without any propagation of the signals backwards. The role of the input layer neurons is to provide the network with external data. Each neuron of the output layer performs a weighted sum of its inputs :

$$o_i = \psi(a_i) = \psi\left(\sum_{k=1}^{N}(w_{ik}\xi_k)\right) \tag{2.2}$$

where:

ξ_k is the kth component of the sample vector,

w_{ik} is the weight between node k and i,

a_i is the activation of the node i,

ψ is the activation function (threshold function):

$$\psi = \begin{Bmatrix} +1 & if & x \geq 0 \\ -1 & if & x < 0 \end{Bmatrix} \tag{2.3}$$

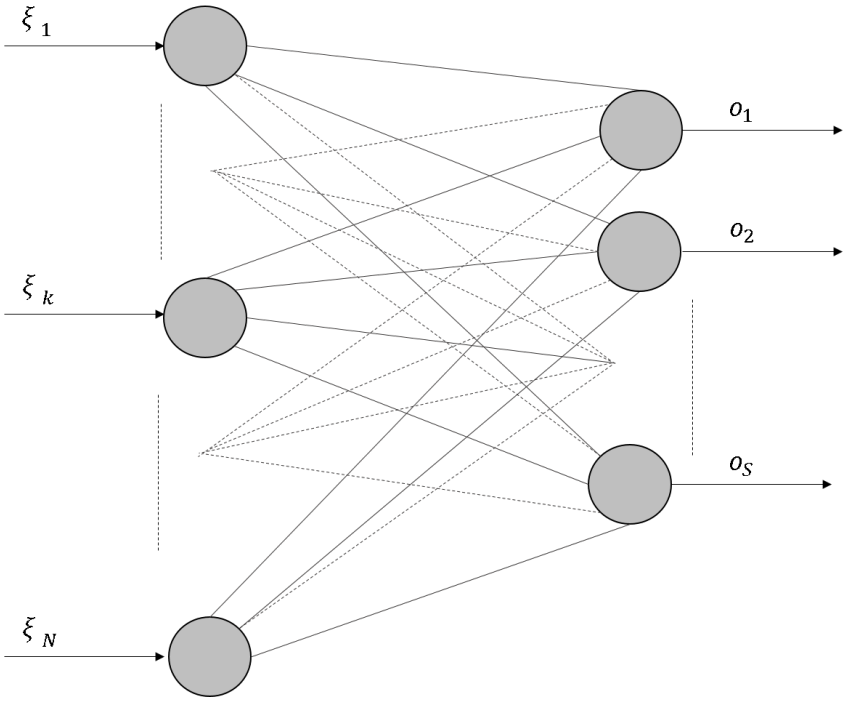


Figure 2.2: Perceptron, Rosenblatt model.

2.2.3.4 The XOR problem and limitations of single-layer models

The XOR problem ("or" exclusive) is certainly one of the simple examples of limitation of single-layer networks. Indeed, this problem, however trivial, is not linearly separable, and therefore cannot be solved by a linear perceptron. Indeed it is impossible to discriminate using a straight line, the two classes representing the two outputs 0 and 1 (Figure 2.3).

A solution [44] consists in drawing two separating lines in the plane using two perceptrons (Figure 2.3). Each isolates points (0,0) and (1,1) respectively. A third perceptron then combines these two lines to achieve the desired separation.

The resulting array is a multilayer perceptron. In addition to its input and output neurons, it has an intermediate layer called the hidden layer.

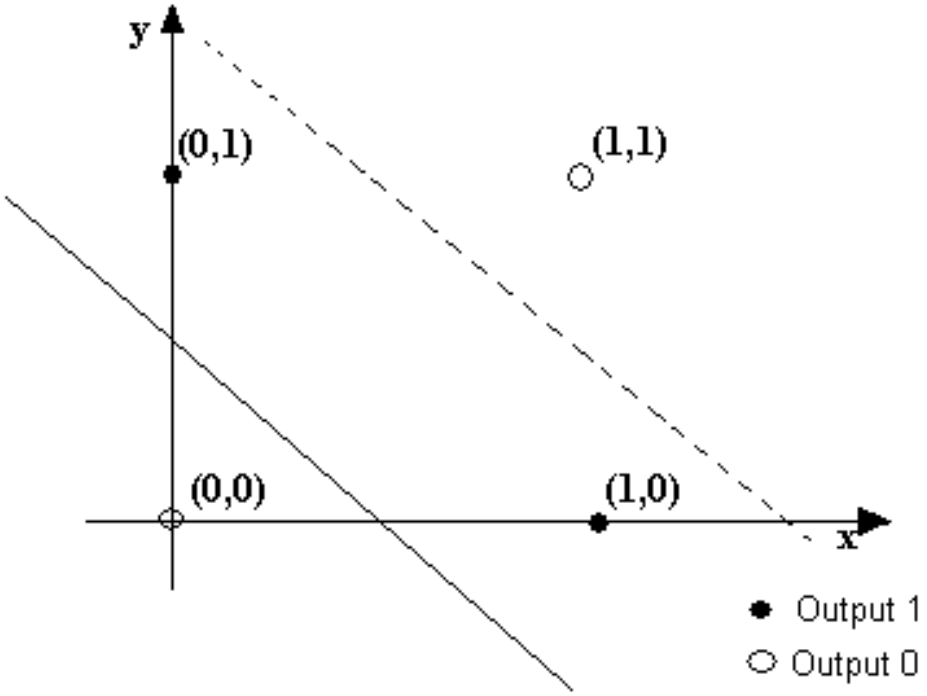


Figure 2.3: Separation performed by a multilayer perceptron.

2.2.3.5 Multilayer Perceptron and Backpropagation Algorithm

Definition

A multilayer perceptron (MLP) (see Figure 2.4) is a neural network with, in addition to the input and output layers, one or more hidden layers. The structure of the network is defined by the set $N_1, ...N_l, ...N_L$ where L is the total number of layers and N_l is the

number of neurons of layer l such that the connections are made between two adjacent layers N_l and N_{l+1} by the synaptic weights w_{ij} linking cell j to cell i.

A multilayer perceptron can use continuous, discrete, boolean, or a mixture of these different types of inputs and outputs. Layer 1 is called the input or retina layer, L is called the output or decision layer. The number of cells on the input and output layer depends on the modeling of the problem. The number of hidden layers as well as the number of nodes on each layer are set by the user, often by "trial/error". We will discuss below the existing methods of searching these variables for a given problem.

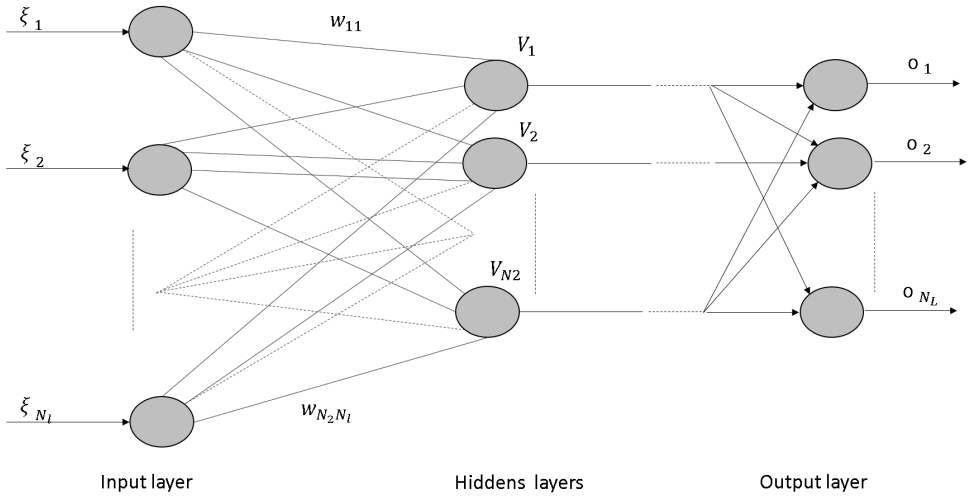


Figure 2.4: Simplified diagram of a Perceptron Multilayer.

In Figure 2.4:

ξ_i is the ith component of the input,

v_i is the output of the hidden layers,

o_i is the output of the output layer.

Principle of Backpropagation

After many years of stagnating research, several teams of researchers were able to separately develop a learning algorithm adapted to multilayer perceptrons called Error Gradient Backward Propagation, respectively: LeCun [49], Werbros [47], and Rumelhart et al [50].

Backpropagation is an extension of the Wirdrow-Hoff rule [51] to multilayer networks. The learning rule performs a descent of the gradient in the weight space, so we must

"learn" the values of the W_{ij} weights that minimize the quadratic error of the network:

$$E(W) = \sum_r E^r = \frac{1}{2} \sum_r \sum_i^{N_L} (t_i^r - o_i^r)^2 \tag{2.4}$$

where:

$$E^r = \frac{1}{2} \sum_i^{N_L} (t_i^r - o_i^r)^2 \tag{2.5}$$

is the quadratic error when presenting the pattern ξ^r.

In multilayer networks, only the output nodes have information that can be directly used to change their weights, but no external signal directly specifies whether the states of the hidden cells are correct or not. This problem, known as the "credit assignment problem" make difficult the update of the synaptic weights.

In order to resolve this problem, the idea of backpropagating the error is as follow [44]:

Each hidden unit is responsible for a fraction of the error committed by each of the output units.

The error for the hidden unit=$\sum_{i \in output} W_{ij} \delta i$ where δi is the error for the output unit i.

As the error is propagated backwards through the network, each of the weights undergoes a modification as follows:

$$w_{ij} = w_{ij} + \Delta w_{ij} \tag{2.6}$$

There are various variants of backpropagation, the OnLine algorithm [44] (stochastic descent) suggests updating the weights after the presentation of each of the patterns.

$$\Delta w_{ij} = -\eta \frac{\partial E^r}{\partial W_{ij}} \tag{2.7}$$

The OffLine algorithm [44] (Batch), on the other hand, recommends modifying the weights only after all the learning patterns have been presented to the network.

$$\Delta w_{ij} = -\eta \sum_r \frac{\partial E^r}{\partial W_{ij}} \tag{2.8}$$

the OnLine algorithm is faster than the OffLine one when the train dataset is of a great size.

For the input data r, the output value of the hidden unit j is given by :

$$v_j^r = \psi(a_j^r) = \psi\left(\sum_k w_{jk}\xi_k^r\right) \tag{2.9}$$

The units of output have the value:

$$o_j^r = \psi(a_i^r) = \psi\left(\sum_k w_{ij}v_j^r\right) \tag{2.10}$$

The resulting algorithm is presented in Algorithm 1.

Algorithm 1 : Backpropagation algorithm

1. Initialization

   The weights and biases are initialized randomly.

2. Presentation of a sample

   we present a sample with its corresponding target :

   $$\xi_i^r, t_j^r \text{ with } \begin{cases} i \in \{1...N_l\} \\ j \in \{1...N_L\} \end{cases}$$

3. "Forward" propagation of signals (relaxation).

   The potential of the neurons is calculated according to:

   $$o_j^r = \psi\left(\sum_k w_{jk}C_k^r\right) \tag{2.11}$$

   $$c_k^r = \begin{cases} \xi_k^r & l = 2 \\ v_k^r & 2 < l \leq L \end{cases}$$

4. "Reverse" propagation of the error

   coefficient δ_i^r calculation for the output layer :

   $$\delta_i^r = \psi_i^{'}(a_i^r)(t_i^r - o_i^r) \tag{2.12}$$

   coefficients for the hidden layers are then calculated using :

   $$\delta_j^r = \psi_j^{'}(a_j^r)\sum_{i=1}^{N_l+1}\delta_i^r w_{ij} \tag{2.13}$$

5. Update the weights using

   $$\Delta w_{ij} = \eta\delta_i^r c_i^r \tag{2.14}$$

   Where η is the learning coefficient such as $0 < \eta < 1$

6. Presentation of the next sample until the the end of train dataset.

Stop criterion of the algorithm

Generally there are two stopping criteria [44] to decide to end of learning phase:

- when the error calculated on the set of training data is below a fixed threshold, or
- when the error is above a threshold after a given number of training cycles, which avoids learning the network indefinitely if the network is wrongly selected.

Convergence theorem of the MLP

Multi layer neural networks have the general property of being parsimonious universal approximators. They are in fact two distinct properties detailed below.

(a) The universal approximation property.

The property of universal approximation has been demonstrated by [45], [52] and [53].

Theorem 1 [52]:

Any sufficiently regular bounded function can be approximated uniformly, with arbitrary precision, in a finite domain of its variable space, by a neural network with a layer of finite number of hidden neurons, all having the same activation function, and a linear output neuron.

Theorem 2 [52]:

When discriminating by a given function, the number of misclassified points can be made arbitrarily small.

This property, which is only an existence theorem and does not give a method to find the parameters of the network, is not specific to neural networks. For neural networks, the number of hidden neurons must be chosen appropriately to obtain the desired accuracy.

The following property is specific to neural networks.

(b) property of parsimony

An approximation is said to be parsimonious if it allows the most satisfactory results to be obtained with a minimum number of adjustable parameters.

In this regard, several central results relating to MLP have been published: [54, 55] highlighting the property that has aroused great interest among researchers compared with other approximation techniques:

A. Barron in [54] shows that an approximation is more parsimonious if this approximation depends in a non-linear way on its adjustable parameters. This result applies to neural networks with sigmoidal activation function since the output of these neurons is not linear with respect to synaptic weights. This property shows the interest of neural networks compared to other approximators such as polynomials whose output is a linear function of the adjustable parameters: for the same number of inputs, the number of adjustable parameters to be determined is lower for a neural network than for a polynomial.

Since the number of parameters depends directly on the number of examples available to the designer, the latter must be large, networks that are parsimonious in terms of parameters are also parsimonious in terms of the number of examples; however, these those samples are :

- expensive: especially if they are the results of experiments on an industrial process,
- their collection can be time-consuming (for example, if one wants to model an economic process based on indices published monthly or annually).

The mathematical properties described are fundamental since they provide a solid foundation for supervised learning neural networks in all their applications.

Activation function

The choice of the transition function is a priori left to the designer of the neural application. The only imperative constraint is that f is derivable. However, it is desirable to respect the following constraints [56]:

1. the function is non-linear

A fully linear network can always be reduced to an equivalent network without an intermediate layer. We thus lose all the advantage of a multilayer structure to have parsimony.

2. the function is strictly increasing

This condition is expressed as $f' > 0 \forall x \in \Re$, because it is desirable that the derivative does not become null in order to avoid a learning block (null gradient). Moreover, changes in the sign of the derivative cause sudden inversions of the gradient which destabilize learning: the function must be monotonous.

3. the function is saturating

This condition is expressed as $|f(\pm\infty)| < \infty$. It corresponds to biological observations, and moreover, the presence of saturations gives the network a high robustness against noise.

4. the function has a quasi-linear range

This can be useful when the application requires a certain degree of accuracy. Several functions are used for this purpose, Figure 2.5 shows the most common ones.

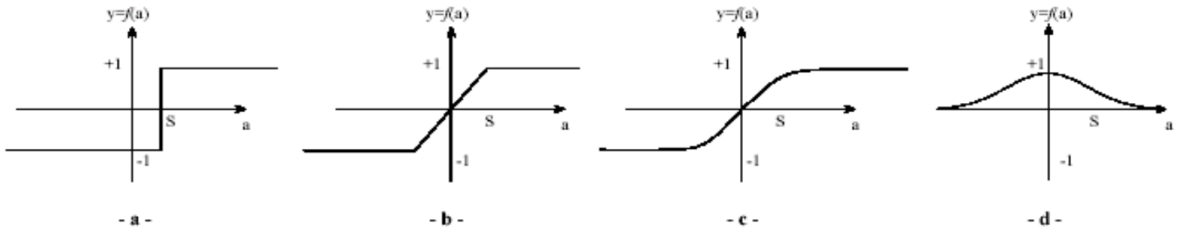


Figure 2.5: Different types of transfer functions for the artificial neuron. a) threshold function of the neuron of Mc Culloch and W. Pitts (1943) [46], b) piecewise linear of the Adaline model of Widrow and Hoff (1960)[51], c) sigmoid of a Perceptron Multilayer network of Rosenblatt (1962)[57], d) Gaussian of the RFR network of Moody and Darken (1989)[58].

2.2.3.6 Neural network architecture optimization

The choice of a neural architecture for a given application is a complex problem. We present the dilemma that manages this choice called the bias/variance dilemma and then we present the classes of existing methods for determining a neural architecture.

• Bias/variance dilemma

The most natural idea would be to choose the largest possible number of hidden neurons, since this is what ensures the greatest precision in the uniform approximation of a function. However, we must not forget that the real problem we are facing is not a problem of uniform approximation, but a problem of fitting a function to a finite number of points, so we must guarantee :

- that the function performed by the neural network passes through the points of the learning set,
- but also that it is able to generalize, i.e. to recognize unknown samples.

As a summary, an insufficiently complex network will tend to perform insufficient modeling of class boundaries while a complex network will tend to learn data by an over-fitting or over-learning, which will result in limited generalization capabilities.

• Methods for determining neural architecture

Two methods are often used to determine the architecture of a neural network:

- the empirical method, by successive approximations,
- the use of architectural optimization algorithms.

a. Empirical Method

The empirical approach consists in comparing the performances obtained for different architectures, and then retaining only the one that corresponds to the best compromise between performance and architectural cost. If the choice of the architecture definition parameters is given to the designer, he can use algorithms to compare the generalization capacities of several architectures from the learning dataset [59, 60].

Several authors have tried to give a priori rules for the size of the hidden layer equal to :

- that of the input layer.
- 75% of this input layer.
- the square root of the product of the number of neurons in the input and output layer.

The determination of architecture by hand is more of an art than a science. Indeed, in [26] the author describes this manual attribution as "Black Art", concluding that the architectures produced are not optimal.

b. Automatic optimization methods

In the framework of this study, we limit ourselves to presenting a few algorithms that can be used for multilayer architectures.

• Growth algorithms of neural networks

The goal of growth algorithms (or constructive algorithms) [44] is to perform learning with, at the beginning, a very small architecture (in terms of number of neurons), and to increase it according to the performances obtained during learning.

For a long time, growth methods have been criticized for being very time consuming and giving rather poor results in generalization [61]. In order to avoid this problem, growth algorithms exploit the relationship between the size of a network and the learning base like in the "Selective Learning with Flexible neural architecture" [62].

• Pruning algorithms

In degeneracy methods (or destructive methods), the starting point is an oversized network that is simplified during learning (by reducing the number of neurons and connections). The most commonly used principles consist either in removing weights or neurons while performing a control by a sensitivity measure, or in using for learning a cost function that includes a factor to sanction the complexity of the network.

The sensitivity measure used by the Gradual Reduction Method (GRM) algorithm for neuron reduction [63] integrates the notions of variance and cross-validation between neurons. The general principle is to eliminate neurons for which the output state does not fluctuate sufficiently according to the data in the learning base. The elimination phases are then alternated with learning phases.

This succession of pruning and learning phases often entails a high computation cost because it amounts to carrying out several parallel training processes, each one with a new architecture.

• Genetic algorithms

The evolutionary method is a metaphor for the principle of natural selection and Darwin's Law of Survival of the Strongest [64]. Since in this thesis the choice of neural network optimization has focused on evolutionary methods, we devote a dedicated section 3.2 in the next chapter to explain the general principle of these methods and their main applications to neural networks.

As already mentioned, the present work will focus on the use of optimized ANNs in the estimation and forecasting of solar irradiation and its components. In the next section, a systematic and updated review in the field of solar irradiation modeling using ANNs is presented to point out the latest findings and the research limitations in this specific axe of research.

2.3 Systematic review of solar irradiation estimation and forecasting using ANNs

In order to establish our systematic review, we adopt the following requirements as described in the literature [65]:

- the review identifies the type of information searched, critiqued and reported within known timeframes.
- it must defines the search terms, search strategies (including database names, platforms, dates of search) and limits all need to be included in the review.

2.3.1 Inclusion criteria

In order to ensure a high quality of the publications included in our review, an automatic search was performed on the databases of the most prestigious publishers with additional criteria illustrated in Table 2.1. Multiple keywords were used as "search terms" such as: solar irradiation estimation, ANN solar irradiation prediction, solar radiation modeling using artificial neural networks.

Table 2.1: Inclusion criteria used in our study.

Electronic databases	IEEE Xplorer SpringerLink Elsevier ScienceDirect ResearchGate
Language	English
Publication period	1998 - 2020

This was followed by including works cited in previous selected publications in order to have a list that is as exhaustive as possible. During our search process, conference articles, working papers, commentaries, and book review articles were excluded [66].

2.3.2 Distribution of literature papers

In the following section, we present the distribution of the reviewed publications according to journal publisher (Figure 2.6).

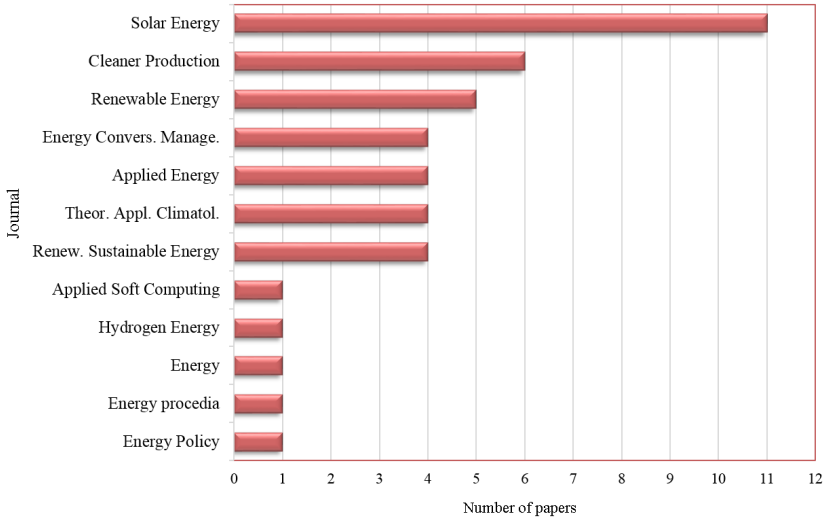


Figure 2.6: Distribution of the reviewed publications according to journal.

2.3.3 Statistical indicators used in literature

To evaluate the ANN estimation and forecasting performances, many well-known accuracy indices are adopted in the literature. Our detailed inspection of the publications under study allowed the extraction of the list dressed in Table 2.2.

2.3.4 State of art of solar irradiation estimation using ANNs

In existing research works, we can find different estimation horizons: monthly, daily and hourly. In fact, monthly prediction allows a realization of a pre-sizing of solar devices, while the daily and hourly solar irradiation values are necessary for a reliable and a precise sizing.

Indeed, our first criteria of study is to classify papers by the prediction horizon. The obtained results of monthly and daily/hourly solar irradiation estimation classification are illustrated in Tables 2.3 and 2.4 respectively. In these Tables we have classified the publications chronologically by specifying the estimated component, the ANN architecture, the location and the corresponding performance evaluation indicators.

Table 2.2: Performance indicators used in the underlying studied works.

Performance indicator	**Formula**
R (correlation coefficient)	$\frac{\sum_{i=1}^{n}(\theta_i-\frac{1}{n}\sum_{i=1}^{n}\theta_i)(t_i-\frac{1}{n}\sum_{i=1}^{n}t_i)}{\sqrt{\sum_{i=1}^{n}(\theta_i-\frac{1}{n}\sum_{i=1}^{n}\theta_i)(t_i-\frac{1}{n}\sum_{i=1}^{n}t_i)}}$
R^2 (coefficient of determination)	$1-(\frac{\sum_{i=1}^{n}(t_i-\theta_i)^2}{\sum_{i=1}^{n}(t_i-\frac{1}{n}\sum_{i=1}^{n}\theta_i)^2})$
MSE (Mean Square Error)	$\frac{1}{n}\sum_{i=1}^{n}(\theta_i-t_i)^2$
RMSE (Root Mean Square Error)	$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\theta_i-t_i)^2}$
NRMSE/RRMSE (Normalized / Relative Root Mean Square Error)	$\frac{100}{\frac{1}{n}\sum_{i=1}^{n}\theta_i}\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\theta_i-t_i)^2}$
MAPE (Mean Absolute Percentage Error)	$\frac{1}{n}\sum_{i=1}^{n}\lvert\frac{\theta_i-t_i}{t_i}\rvert\times 100$
MBE (Mean Bias Error)	$\frac{1}{n}\sum_{i=1}^{n}(\theta_i-t_i)$
RMBE (Root Mean Bias Error)	$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\theta_i-t_i)}$
MAE (Mean Absolute Error)	$\frac{1}{n}\sum_{i=1}^{n}\lvert\theta_i-t_i\rvert$

Table 2.3: Our state of art of monthly solar irradiation estimation.

Component	Ref	Year	Inputs	The ANN architecture	Performance indicators	Location
Global	[67]	2016	G_{toa}, GHI, Lat, T_m, Rh_m, TL_f, Ang_c, Z_a	One hidden layer (8-35-1)	RMSE(Wh/m^2) R^2(%) 5.75 99.9	Algeria
Global	[68]	2016	G_{toa}, D_l, R_d, Rf, Lat, Alt	One hidden layer (7-4-1)	MAPE(%) NRMSE(%) about 1.67 1.01	Italy
Global	[69]	2016	Lat, Long, Alt, Mon, SD_m, T_m	One hidden layer (6-3-1)	MAPE(%) R^2(%) around 5.0 >99	Turkey
Global	[70]	2014	T_{max}, T_{min}, T_m,Rh_m,Rf, SD_m	One hidden layer (6-9-1)	RMSE: from 0.0031 to 0.0035 MBE: from -0.0003 to 0.0011	Thailand
Global	[71]	2014	T_{max}, T_{min}, T_m, Alt, SD_m	One hidden layers(5-10-1)	MAPE(%)=6.89	India
Diffuse	[72]	2008	K_t, S_p	One hidden layer (2-5-1)	R^2 RMSE (MJ/m^2) 0.90 0.75	China
Global	[39]	2008	SD_m, T_m, CC, Lat, Long, Alt	One hidden layer (6-15-1)	R MBE(MJ/m^2) RMSE(MJ/m^2) 0.97 0.059 0.385	Uganda
Beam	[73]	2006	Lat, Long, Alt, Mon, SD_m, Rh_m, Rf	One hidden layer (7-3-1)	RMSE(%): from 1.65 to 2.79	India
Global	[74]	2004	Lat, Long, Alt, Mon, SD_m, T_m	Two hidden layers (6-2-5-1)	MAPE(%) R^2 RMSE(%) 6.735 0.99 4.465	Turkey
Global	[75]	2004	Lat, Long, Alt, Mon, SD_m, T_m	(6-N/A-1)	MAPE (%) R^2(%) $\leq$ 6.73 99.89	Turkey
Global	[76]	2002	Lat, Long, Alt, Mon, SD_m	(5-N/A-1)	RMSE= 0.83	Oman
Global	[77]	2000	Lat, Long, Alt, Mon, SD_m	One hidden layer (4-10-1)	MAPE=10.1	Saudi Arabia
Global	[78]	1998	Loc, Mon, P_m, T_m, VP_m, Rh_m, WS, SD_m	One hidden layer (8-15-1)	MAPE(%) R^2(%) 5.43 95	Sultanate of Oman
Global	[79]	1998	Lat, Long, Alt, SD_m	One hidden layer (4-10-1)	MAPE(%):from 6.5 to 19.10	Saudi Arabia

Table 2.4: Our state of art of daily/hourly solar irradiation estimation.

Component	Ref	Year	Input parameters	ANN architecture	Performance indicators	Location
Global/Hourly	[80]	2020	T_m, h_d, Rh_m, Rf, WS	multineural network approach	R(%):from 96.24 to 98.41 RMSE(%):from 15.20 to 22.50	Morocco
Global/Daily	[81]	2019	G_toa, Rh_m, Rf, $\triangle T$	One hidden layer (4-9-1)	R^2(%) NRMSE(%) MAPE(%) 97.50 17.85 18.46	Morocco
Global/Hourly	[82]	2018	Loc_T, Rh, P, SD_{max}, WS	One hidden layer (5-150-1)	R(%)=98.87 $RMSE(Wh/m^2)$=41.09	Iran
Diffuse/Daily	[83]	2017	Mon, SD_m, T_m, Rf, WS, Rh_m, GHI	One hidden layer (7-5-1)	R(%) RMSE MAE(%) 95.30 82.40 70.30	China
Global/Daily	[84]	2017	Lat, Long, Alt, Loc_T,WS, Rh, G_{toa}, SD, D_y	One hidden layer (9-15-1)	R^2=0.96 RMSE(Wh/m^2)=38.89	Turkey
Global/Daily	[85]	2016	T_{max}, T_{min}, Rh, WS, $PM_{2.5}$, PM_{10}	Two hidden layers (6-12-24-1)	R^2=0.97 MAPE(%)=3.13 RMSE(Wh/m^2)=0.214	Iran
Global/Daily Beam/Hourly	[86]	2016	Lat, Long, T_m, SD, Rf, Dec_a, Dl_h ω, CI, Dec_a, GNI	One hidden layer (7-10-1) (4-5-1)	R^2(%) MAPE(%) RMSE(Wh/m^2) 99.18 4.57 160.30 99.40 5.57 17.70	Italy
Global/Daily	[87]	2016	Lat, Long, Alt, Rh, P, CI, T_m	One hidden layer (7-10-1)	R^2(%)=98.89 MAPE(%)=2.56 RMSE($Wh/m^2/day$)=223	Zimbabwe
Global/Hourly	[88]	2016	Dec_a, G_{toa}, SA, GHI	Two hidden layers (4-18-13-1) tilted planes 45 (4-18-13-1) tilted planes 60	R^2(%) MAE RMSE 92.42 0.028 0.055 93.02 0.026 0.055	Iran
Diffuse/ Daily Hourly	[89]	2015	CI, D_l CI, h_d, SA	One hidden layer (2-10-1) (3-10-1)	R(%) $NRMSE$(%) 95.96 18.73 94.32 19.73	Morocco
Global/Daily	[90]	2010	D_y, T_m, Rh, SD, WS_e	Two hidden layers (5-3-3-1)	R^2(%) $MAPE$(%) 98.57 5.21	Iran
Global/Daily	[91]	2008	D_y, Lat, Long, Alt, T_m, SD	One hidden layer (6-N/A-1)	$MBE(Wh/m^2)$:from -916.6 to 608.3 $RMSE(Wh/m^2)$:from 388.8 to 1113.8	China
Global/Daily	[92]	2008	K_t, D_y, Alt	One hidden layer (3-10-1)	MBE(%)=0.20 $RMSE$(%)=6.00	Spain
diffuse/ Hourly Daily	[93]	2007	GI, LW_{ae}, T, Rh, P_a GI, G_{toa}, SD	One hidden layer Hourly scale (5-40-1) Daily scale (3-40-1)	$MBE(Wh/m^2)$: from -44.44 to 61.11	Egypt
Global	[94]	2005	SD_{meas}, SD_{theo}, T_{max}	Two hidden layers (3-46-23-1)	MBE(%) RMSE(%) 0.12 5.67	Athens
Diffuse/Hourly	[95]	2004	GI, DE_{ces}, CI, DE_{ces}, Rh, LW_{ae}, SA, Z_a, P_p	One hidden layer (9-N/A-1)	MBE(Wh/m^2)=3.06 RMSE(Wh/m^2)=43.05	Brazil

*N/A : not available

The presented papers in Tables 2.3 and 2.4 worked on the estimation of solar irradiation and most of them focused on a horizontal plane excepted M. Shaddel et al. in [88] where the authors estimated solar irradiation on tilted planes for two different angles (45° and 60°) in Iran. It also show that most of the studied works used Sunshine duration as an input, given the proved correlation between Sunshine duration and solar irradiation [71, 96]. In fact, Ångström (1924) in [96] was the first to suggest a linear relationship between global solar irradiation and Sunshine duration. As well as A. K. Yadav et al. in [71] where the authors found that Sunshine duration is one of the most relevant input parameters for predicting solar radiation using artificial neural networks. On the other hand, recent studies made an effort to use parameters that are routinely recorded at a large number of meteorological stations such as E. F. Alsina et al. in [68] where the used parameters were: top of atmosphere irradiation, day length, number of rainy days, average rainfall, latitude and altitude. In another important work [97], the authors proposed a prediction model using daily air temperature, relative humidity, atmospheric pressure, wind speed and Earth temperature providing good accuracy in the prediction of solar irradiation. Furthermore, we note that all works used one or two hidden layers, which is sufficient for modeling and estimating solar irradiation [78, 98, 99]. Other studies extended the architecture to two hidden layers such as for instance the work of C. Renno in [86] where the authors used twelve and twenty four neurons in the first and second hidden layer respectively. In M. Shaddel et al. study [88], the authors used an ANN with two hidden layers, the first one with eighteen neurons and the second contain thirteen neurons.

Tables 2.3 and 2.4 show that most of the studied works used Sunshine duration as an input, given the proved correlation between Sunshine duration and solar irradiation [71, 96]. On the other hand, recent studies made an effort to use parameters that are routinely recorded at a large number of meteorological stations such as minimum temperature, maximum temperature, wind speed and relative humidity, which also provides a good accuracy in the estimation of solar irradiation [68]. We can also note that the most used inputs are altitude, longitude, latitude (for the studies using different stations), temperature (minimum, maximum and average) and relative humidity. Furthermore, we note that all studied works used one or two hidden layers, which is sufficient for modeling and estimating solar irradiation [78, 98, 99].

Concerning performance evaluation indicators, the most used in examined papers are

RMSE, MAPE, R^2 and MBE. The found MAPE values falls in the range [2.56 - 19.10] which is good estimation accuracy according to [100]. All the models show good performances with a coefficient of determination (R^2) between 0.90 and 0.99.

2.3.5 State of art of solar irradiation forecasting using ANNs

In this subsection, a particular attention will be given to ANNs and their applications in forecasting solar irradiation (aiming to predict future solar irradiance values using past or present data). Table 2.5 contains several papers focusing on forecasting solar irradiation and its components.

In this Table, as for the previous section, the selected papers are chronologically classified, by specifying the forecasted component, the used input parameters, the forecasted horizon, performance indicators and the corresponding performances.

In fact, since the late 1990s, ANNs have seen increased application in this field. Kemmoku et al. proposed in [101] a single and multistage neural networks to forecast the insolation of the next day where exogenous data were used to train these models. In [102], the authors describe a neural network approach for modeling short-term prediction to forecast daily solar radiation. They found that for the summer season, an ANN model with 13 inputs can be sufficient to forecast with good accuracy from 10 to 20 days ahead, while in winter time, 9 inputs are more suitable to predict solar radiation up to 8 days ahead. In [103], Sfetsos et al. compared different artificial intelligence based approaches. ANNs among other models were used to forecast global solar irradiance on a horizontal level for 1 h ahead. The best prediction results were found using the multivariate Levenberg Marquardt artificial neural network. Mellit et al. in [104] suggested a feed-forward network using wavelets as activation functions to find a suitable model to predict daily total solar radiation. Martín et al. presented a comparison of various models (autoregressive, ANNs and fuzzy logic) to predict half daily values of solar irradiance for the next three days. In this study, they transformed global solar irradiance to two new variables: clearness index (k_t) and Lost Component (LC). They found that the best approach to forecast half daily values of solar irradiance is neural network models with LC as input parameter [105]. In [12], Mellit et al. proposed a multilayer perceptron model to forecast the solar irradiance for 24 h using values of the mean daily solar irradiance, air temperature and day of the month. Marquez et al. developed a medium-term solar irradiance forecasting model by adopting predicted meteorological variables from the US National

Weather Service's forecasting database as inputs to an Artificial Neural Network (ANN) model. An input selection scheme was applied using gamma test and genetic algorithms. They concluded that, as expected, the proposed model loses accuracy with increasing forecasting horizon [106]. Chen et al. used both fuzzy logic and ANNs techniques to forecast hourly solar irradiance under different sky and temperature conditions. In their approach, the authors classified sky conditions and temperature information as different fuzzy sets based on fuzzy rules. Then, this obtained data with the past values of solar radiation were used to train the basic neural network [107]. Benmouiza et al. proposed a methodology combining k-means algorithm and Non-linear Auto Regressive (NAR) neural network to provide better results in forecasting hourly horizontal global solar radiation. The NAR method was tested with different architectures to create the global predictor. However, as highlighted by the authors, this method needs a considerable calculation time in the preparation phase of the NAR network and shows some limitations in presence of heavy clouds in the sky [108]. In [109], Pedro et al. present a global solar irradiance forecasting model based on the optimization of the k nearest neighbors and artificial neural networks techniques for 15 min time horizon up to 2 hours. In this study, the impact of different microclimate on the forecast error and forecasting ability using endogenous data is investigated. An interesting paper was proposed by Ghofrani et al. in [110], where the authors present a hybrid forecasting method that combines a new clustering selection algorithm and Multi Layer Perceptron Neural Networks (MLPNNs) to forecast global horizontal irradiance for three different horizons 1, 24 and 48 hours. The proposed method provides more accurate forecasts as compared to direct clustering of hourly data. Lately, Benali et al. compared the Random Forest (RF) and artificial neural networks models with the Smart Persistence (SP) model to forecast hourly global, beam and diffuse solar radiation up to six hours ahead. In their study RF model outperformed the other two models [111].

Table 2.5: Our state of art of solar irradiation forecasting using ANNs.

Component	Ref	Year	Inputs	Horizon	Performance indicators	Location
Global Diffuse Direct	[111]	2019	preceding values of: $K_{g,cs}$ $K_{d,cs}$ $K_{b,cs}$	Up to 6 hours ahead	NRMSE(%):from 22.57 to 34.85	France
GHI	[110]	2016	Hourly GHI	1h, 24h and 48h ahead	RMSE(W/m^2):from 57.56 to 139.28	USA
DNI			Hourly DNI		MAE(W/m^2):from 23.12 to 77.32	
Global	[109]	2015	SA, TL_f	from 15 min to 2 h	RMSE(W/m^2):from 41.3 to 90.2	USA
Global	[108]	2013	GHI	multi-step ahead	RMSE(Wh/m^2)=64.34 R^2(%)=93.33	Algeria
Global Global	[107]	2013	sky conditions, T_m	Hourly One day ahead	MAPE(%):from 6.03 to 9.65	Singapore
GHI DNI	[106]	2011	SC, RF, T_{max}, T_{min}	One day ahead	NRMSE(%): from 15 to 22 NRMSE(%): from 28 to 35	USA
Global	[12]	2010	GI_{md}, T_m, D	24 h ahead	Sunny days: R^2(%):from 98 to 99 cloudy days R^2(%):from 94 to 96	Italy
Global	[105]	2010	LC	half daily values	RMSE(Wh/m^2):from 29 to 39	Spain
Global	[104]	2006	5 preceding values of daily GHI	One day ahead	R(%)=97	Algeria
Global	[103]	2000	preceding values of GHI	1 hour ahead	RMSE(Wh/m^2):from 29 to 39	France
Global	[102]	2000	8 preceding values of GHI	1 hour ahead	MSE(MJ/m^2):from 0.19 to 0.22	Athens
Global	[101]	1999	P_{amin}, P_{max}, CI_m, T at 18:00 h, P_a at 18:00 h	One day ahead	Mean error(%):about 20	Japan

2.4 Discussion and recommendations for future research works

In the following section, we will point out some observations and problems that we have noted during our study of the papers with our corresponding recommendations.

• To test and validate solar energy estimation and forecasting models, long term weather data are required. However, such data are not easily available because of the high cost of measuring devices and the difficulty in accessibility of the measuring sites which puts severe limitation in conceiving reliable and accurate models. Through our inspection of the studied literature we have noticed the lack of a standard database having a large number of input types with the recording intervals of data. Also, the training and test subsets in such database need to be statistically representative in order to have accurate models (one rule of thumb is that the training set size should be 10 times the network weights to accurately classify data with 90% accuracy [112]). In all the ANN models, the number of hidden layers and corresponding neurons is determined experimentally (which may require large computational analysis) and there is no mentioned systematic method to optimize this number. This task remains an open and challenging problematic and must be addressed in future works. Genetic algorithms, Particle Swarm optimization, simulated annealing techniques can be considered as optimization techniques for this aim.

• From published literature, the correct choice of geographical and meteorological input parameters is essential to predict solar radiation with reliability and better accuracy. Unless few studies working on this problematic [71, 113], there is yet no automatic methodology carrying out the selection of most relevant input variables for ANN models.

• In the work of [71], the effect of Sunshine duration on the prediction accuracy has been highlighted. This study should be generalized to see the effect of each variable on the overall ANN model performances.

• Regarding Tables 2.3 and 2.4 showing the fewness of papers on diffuse and beam solar irradiation prediction using ANN and due to the importance of those components for the energy applications, more studies are required in future works.

• Unfortunately, limited attention has been given to the comparison between ANN and other prediction models [94, 114, 103].

• As indicated in [100], different ANN models need to be developed using latitude, longitude, altitude, extraterrestrial radiation as input parameters and checked for accu-

racy. This will be useful for those locations where no meteorological stations have been established even if it is found that latitude and longitude have minimum effect on solar radiation prediction as demonstrated in [71].

2.5 Conclusion

In this chapter, a first part was dedicated to present the needed theoretical background and basic knowledge on artificial neural networks, essential for the development of our ANN-based models. Due to their high potential to simulate the complex, nonlinear and time-varying input-output systems compared to other existing models, ANN models are found to predict solar irradiation accurately. Therefore, different research papers based on ANN models are used to predict solar irradiation. In the second part of this chapter a state of the art is provided in this particular matter. Our study provides an updated review to help future investigations in this field. It uses a meticulously selected bibliography and its corresponding prediction horizon, used inputs, ANN architecture and performance indicators. Furthermore, the objective behind analyzing these publications was to highlight difficulties such as the lack of a standard database (having a large number of input types with data recording intervals) and the absence of a systematic method optimizing the ANN architecture. Also we have noticed the fewness of papers working on the estimation of the solar irradiation components (diffuse and direct) using artificial neural networks. Moreover, unless few studies working on this problematic, there is no methodology to carry out the selection of most relevant input variables for ANN models.

In this optic, the next chapter will present our proposed ANN model optimized by genetic algorithms, which is developed to solve the problematic related to the manual selection of the input parameters feeded to the ANN estimator.

Chapter 3

Automatic selection of input parameters using evolutionary ANNs for daily solar irradiation estimation

3.1 Introduction

Morocco is located in northern of Africa and has an abundant solar resources approximating 5.3 kWh/m^2/day [115] which brings increasing number of investments aiming at developing Thermal and Photovoltaic solar energy systems. Since the design of any cost-effective solar energy system is related to reliable data, developing techniques for accurate estimation of solar irradiation is crucial. In fact, as discussed in the previous chapter, various estimation models exist in the literature including empirical, stochastic, statistic and artificial intelligence (AI) models [15]. Among AI techniques, artificial neural networks are well known machine learning systems that have been widely used in many fields of application, including control systems, pattern recognition and classification [20]. ANNs have the particular capability to modelize complex, non-linear and time varying input-output systems [21] due to their simplicity, generality and good learning ability [22]. Particularly, ANNs have been efficiently used as solar irradiation predictors in different sites over the world taking into account the non-linear relationship between inputs (meteorological, geographical, radiometric and atmospheric parameters) and output (direct, diffuse and global solar irradiation). As cited in [116], one of the most important challenges in using ANNs for solar irradiation estimation is the selection of the most relevant input variables for a set of available data giving accurate prediction [71, 113]. In fact, as pointed out by A. Qazi et al. in [66] choosing appropriate input parameters to estimate solar irradiation will improve reliability and accuracy of the resulting ANN predictor. Additionally, an ANN model with less input parameters will guarantee reduced computational time for prediction and consequently leads to technical and economic benefits.

Except few studies dealing with this issue [71, 113, 117], there is not yet an automatic methodology carrying out this selection. In fact, Lopez et al. in [113] proposed the Automatic Relevance Determination method (ARD) based on Bayesian framework applied for ANN to obtain the relative relevance of the inputs. The authors then selected optimum input parameters and deduced that clearness index and relative air mass were the most relevant input variables to the neural network. An important work pointing out this problem was developed by Yadav et al. in [71], where the authors proposed the WEKA software (Waikato Environment for Knowledge Analysis) to evaluate the most influencing input parameters in prediction of solar irradiation using ANNs. In this model, it has been found that temperature, maximum temperature, minimum temperature, altitude above

sea level and sunshine hours are the most relevant input variables. Recently Mohamadi et al. in [117] adopted an Adaptive Neuro Fuzzy Inference System (ANFIS) technique as a selection procedure. They determined the most influential parameters combination among nine inputs for three cities with different climatological and geographical conditions. An alternative method was proposed by Marquez et al. in [106] where a pre-processing phase is carried out to select the relevant inputs that will be fed to the ANN.

In this chapter, the problem of finding the most relevant inputs to an ANN as a global horizontal solar irradiation (GHI) estimator is formulated as a search problem which can be resolved with evolutionary artificial neural networks (EANNs). In fact, there has been a great interest in recent years in considering learning using ANNs and evolution through evolutionary algorithms which are two forms of adaptation. This combination is more effective and efficient and can lead to significantly better intelligent systems instead of relying on each method separately [23]. The key of success of such hybrid algorithms is the incorporation of a local search procedure provided by ANNs into the global one achieved by evolution.

The purpose of this work is to use EANNs to produce an automatic methodology of ANN inputs selection (from a set of available data) for the estimation of daily horizontal global solar irradiation by developing a model with:

- commonly measured variables like temperature, humidity and rainfall,
- good accuracy that is crucial for solar energy conversion applications.

In this automatic selection, we use evolutionary algorithms (based on genetic algorithms) as a selection methodology instead of conventional and other pre-processing methods. In our model a new coding strategy of the ANN individuals is proposed with specific genetic operators that guarantee the production of valid offspring in the decoding phase. Our genetic population is a set of ANN individuals represented in the form of chromosomes where each of which is a subset of input parameters. The resulting algorithm begins with a selection phase followed by a coding strategy and the application of crossover and mutation operators. In order to assess the validity of the proposed model, we will compare it with other methods of selection (heuristic search, forward stepwise regression) and methods of estimation (empirical models, kNN model). To achieve this study aim, we used measurements of meteorological parameters and global horizontal solar irradiation (GHI) for the city of Fez (Morocco) carried out by a local radiometric station of the Faculty of Sciences and Technologies.

3.2 Theoretical backround on genetic algorithms

Genetic algorithms (GA), more commonly referred to as evolutionary algorithms (EA), are adaptive methods commonly used in optimization problems. They are based on the process of genetic evolution of biological organisms through generations according to Darwin's theory of evolution [64].

The principle of this theory is that biological organisms live together in an environment where they reproduce and share the same resources such as food and shelter from predators. Thus, stronger individuals are more likely to survive against predators or find a partner to reproduce. Those that have succeeded to survive and find partners will generate relatively more offspring. The weaker ones will have few or no offspring. As a result, the genes of the most adapted individuals will be passed on to many individuals in subsequent generations. The combination of the best genes from different ancestors can sometimes produce "super-individuals" who adapt even better than their parents. In this way, species evolve and become more and more adapted to their environment.

By simulating this principle, evolutionary algorithms applied to an optimization problem evolve a set of candidate solutions, called a population of individuals. An individual represents a possible solution to the given problem. Each individual is assigned a "fitness" that measures the quality of the solution it represents, often it is the value of the function to be optimized. Then, a new population of possible solutions is produced by selecting the best parents of the current "generation" to crossover and mutate. The new population contains a higher proportion of characteristics of the best individuals of the previous generation.

So evolution is performed through two basic steps:

- selection of individuals based on their fitness,
- reproduction of the selected individuals.

In this way, from generation to generation, the best genes are propagated in the population, combining or exchanging the best characteristics. By favoring the best individuals, the most promising areas of the research space are explored. If the algorithm is well designed (coding, individual evaluation function and other parameters), the population will converge to a minimal steady state.

3.2.1 Genetic operators

To use the concept of genetic algorithms in optimization problems, the following steps are needed :

1. a system of encoding the possible solutions or chromosome structure,
2. an initial population of solutions,
3. a function to evaluate a solution's fitness,
4. a method of selection to be used to produce new solutions,
5. crossover and mutation operators to create new solutions from existing ones.

3.2.1.1 Solution representation (encoding scheme)

A chromosome representation is necessary to describe each individual in the GA population. The representation scheme determines how the problem is structured in the GA. Each chromosome is usually expressed in a string of genes in binary, real number or other forms, and its range is usually determined by specifying the problem [118].

3.2.1.2 Initial population

Before strarting to evolve, the first step of any genetic algorithm is to generate a set of possible solutions as an initial generation or population to the problem. Initial solutions are sometimes created randomly or by using a specific algorithm.

3.2.1.3 Evaluation function (Fitness)

The fitness function is used to selectively reproduce the "strongest" individuals to produce new individuals for the next generation. To give more chances to small neuronal structures (containing a low number of nodes), a bias is introduced for these structures by allocating them more learning time than large structures. This approach has already been introduced in [119].

A subset called "validation subset" extracted from the learning set is used to measure the performance of the genetic algorithm. The fitness function is, therefore, the performance of the neural networks measured on this set for the total number of learning epochs. The value of this function corresponds to the assigned performance indicator.

3.2.1.4 Selection and Replacement Strategy

Selection tends to give high probabilities for the most suitable individuals in a population in order to produce even more suitable offspring. It is precisely from this where the principle of "survival of the fittest" derives.

There are several strategies that can be adopted in the production of a new population. In the generation cycle, an individual in the current population is replaced by a new offspring. It is very likely that the best individual from the current generation will not be reproduced in the new generation. Hence, it is interesting to introduce the strategy of elitism [120] which consists in copying the best or a subset of the best chromosomes from the current generation to the next one. This strategy of elitism allows to increase the dominance of the best individuals in a population, thus improving the chance to obtain best solutions.

3.2.1.5 Crossover

The essence of any crossover operator is the exchange of components from both parents to form new offspring. After selection, crossover causes a structured, yet randomized exchange of genetic material between solutions, with the possibility that "good" solutions can generate "better" ones. Crossover occurs with a probability PC (the crossover probability or crossover rate). When the soltutions are not subjected to crossover, they remain unmodified. Notable crossover techniques include the single-point, the two-point, and the uniform types [121].

3.2.1.6 Mutation

Mutation involves the modification of the value of each "gene" of a solution with some probability p (the mutation probability). The role of mutation in GAs has been restoring lost or unexplored genetic material into the population to prevent the premature convergence of the GA to suboptimal solutions.

3.3 Data and location

The present chapter is devoted to the estimation of solar irradiation in the city of Fez (see the map in Figure 3.1) which is classified as Csa (hot-summer mediterannean cli-

mate) according to the Köppen climate classification [122]. The measurement of GHI at ground surface and other meteorological parameters was performed by a radiometric station placed on the roof of the Faculty of Sciences and Technologies building at Sidi Mohamed Ben Abdellah University, Fez, Morocco (latitude 33° 56' N, longitude 4° 59' W, altitude 579 m). The device collecting the solar data used in this work is a kipp & Zonen model CM-11 pyranometer (with a measurement uncertainty of ±5%). For measuring the rainfall, the rain gauge (±3%) were used while air temperature ($\pm 0.8°C for T > -20°C$) and relative humidity (3%) were measured by means of a thermohygrometer. The wind speed (1%) and direction (3%) were measured with an anemometer. All the sensors are connected to a data logger measurement system (Campbell Scientific, CR1000 model). The collected data cover the period from 2009 to 2015 and were divided into 75% for models training and 25% for testing. Before applying our model, a quality control procedure [123] was performed on dataset which consists of a set of four main tests:

- the first test eliminates all the data recorded at a solar altitude lower than 5° to treat the intrinsic error of the pyranometer caused during Sunrise and Sunset,
- the second test represents physical limits of the clearness index and the diffuse fraction which must be bounded between 0 and 1,
- in the third test, the global horizontal irradiation and the diffuse horizontal irradiation are compared and limited by their values in clear and cloudy sky,
- the fourth test is a statistical analysis, where we retain only the data values that respect some conditions (see [123]).

A recapitulation of our available input data are drawn in Table 3.1 with an abbreviation of each parameter and its corresponding index useful in the experimental result section.

3.4 Proposed EANN inputs selection model

Before introducing our proposed EANN model, we present some models which can be used to select relevant inputs in the estimation of solar irradiation. In the first model baptized heuristic search, we implement a method based on selecting combinations referring to the state of the art chapter where we test subsets of inputs having good correlations with solar irradiation. In the second implemented model in our case study, we adopt a forward stepwise regression search to favorize selecting small subset of inputs fed to ANNs. After that, we introduce the proposed EANN model which automatically extract relevant inputs

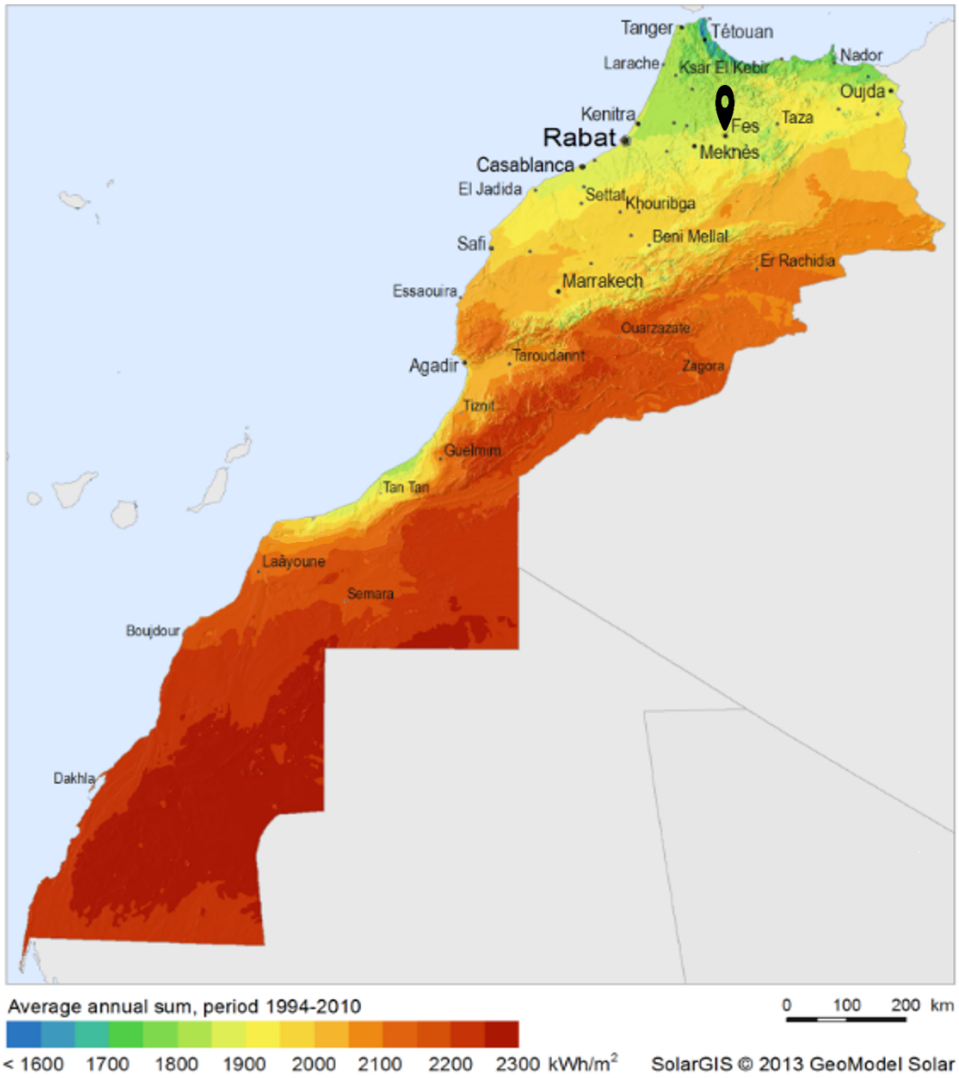


Figure 3.1: Various solar irradiation intensities in different cities of Morocco with a specific localization of Fez city [124].

Table 3.1: Input parameters available in our radiometric station.

Num	Abbreviation	Input parameter	Num	Abbreviation	Input parameter
1	Y	Year	9	$S\Delta T$	Root square temperature (°C)
2	Dy	Day of the year	10	Rh_{max}	Max relative humidity (%)
3	Mon	Month	11	Rh_{min}	Min relative humidity (%)
4	Mon_d	Day of month	12	T_{max}	Max Temperature (°C)
5	Rh_m	Mean relative humidity (%)	13	T_{min}	Min Temperature (°C)
6	Rf	Rainfall(mm)	14	WS	Wind speed (m/s)
7	SA	Solar altitude (°)	15	WD	Wind direction (°)
8	ΔT	Daily temperature gradient (°C) (Difference between daily maximum and minimum temperatures)	16	G_{toa}	Solar irradiation at the top of the atmosphere

using the evolution paradigm combined with the estimation power of ANNs.

3.4.1 Model 1: ANN inputs selection based on heuristic search method

Searching the best subset of input variables is a difficult task. The obvious method to select this subset is to systematically test the relevance of all possible subsets. This exhaustive search must explore $(2^d - 1)$ combinations and hence, involves a high computational burden, even if it guarantees the optimal determination for a given ANN model [125].

Rather than evaluating all possible combinations, most works in the literature [68, 88, 92] settle for using various combinations of inputs and evaluate the corresponding performances. Our first proposed selection method is a heuristic search based on the known correlation of some inputs with solar irradiation deduced principally from Tables 2.3 and 2.4 from the state of art chapter presented earlier. Section 3.5.1 explores the obtained results using combination of inputs like: maximum temperature, minimum temperature, relative humidity and solar irradiation at the top of the atmosphere.

3.4.2 Model 2: ANN input selection based on forward stepwise regression

In the literature, backward selection starts with all input variables and removes the least relevant variables one at the time. The algorithm stops if the performance drops below a given threshold by removal of any of the remaining input variables [92, 126]. In this work, we adopt a forward stepwise regression which starts by training d single-input ANN models and selects the input variable that maximizes the model performance-based optimality criterion. Then, the selection is terminated when the addition of another input variable fails to improve the performance of the ANN model (see Figure 3.2).

3.4.3 The proposed model: ANN input selection based on evolutionary algorithms

As already defined in section 3.2, genetic algorithms are optimization models based on Darwin's theory of natural evolution. Based on the principal of survival of the fittest, a genetic algorithm starts with a population of potential solutions individuals (phenotype) previously coded as a set of chromosomes (genotype). New individuals are generated by the reproduction (crossover) of the genetic information of the parents. During this process

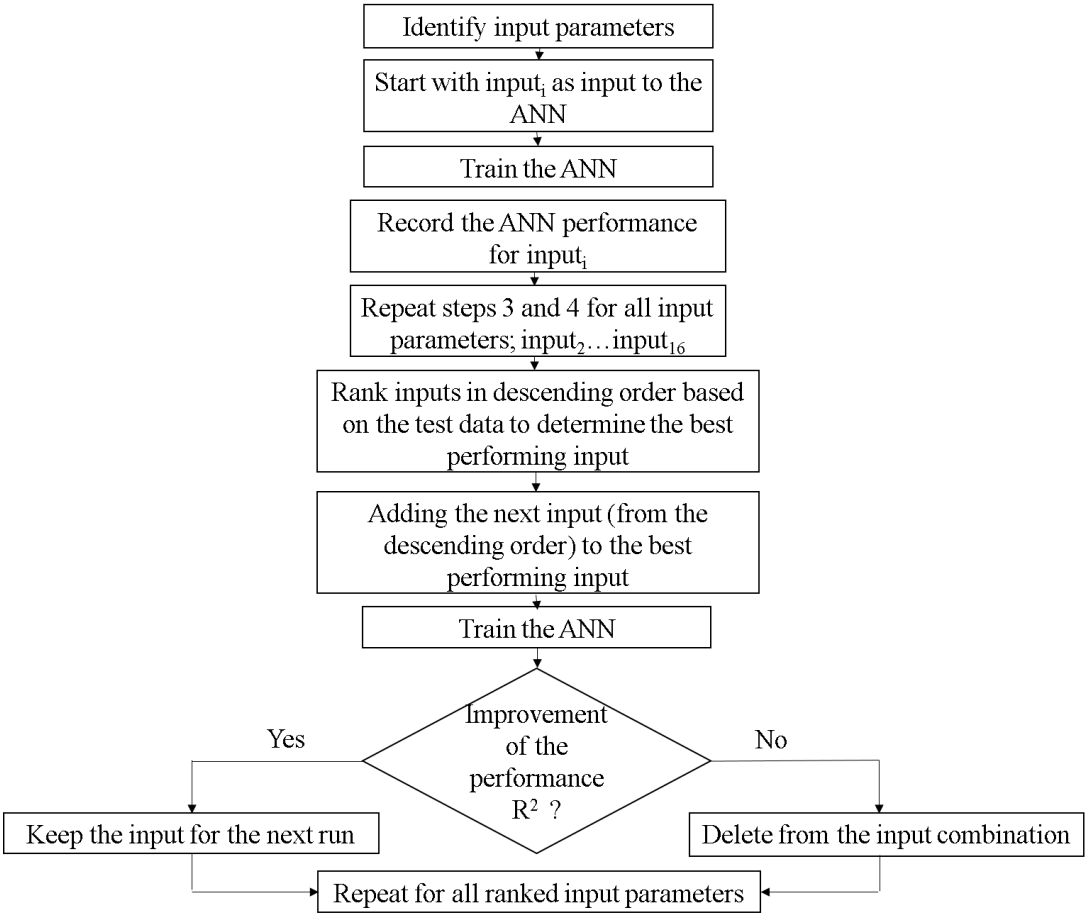


Figure 3.2: Forward stepwise regression flowchart.

the quality of any solution is evaluated using the fitness function. Typically, a GA uses selection, crossover and mutation operators to generate new solutions in the search space. This process is repeated for a predetermined number of generations or until a satisfactory solution has been found according to the defined fitness function.

The process starts from an initial population. Selection is then carried out to create an intermediate population for which crossover is applied. Mutation is then performed to produce the next generation as shown in Figure 3.3.

3.4.3.1 Coding neural network individuals

Representing the genetic structure of a neural network is not as straightforward. As reported in [127], many considerations have to be taking into account.

In our genetic algorithm representation, the chromosome is a finite sequence of 0's and 1's. Each chromosome represents a subset of inputs, the k-th bit denoting the presence or absence of the k-th variable. Hence, the chromosome length, noted l, is always equal to the

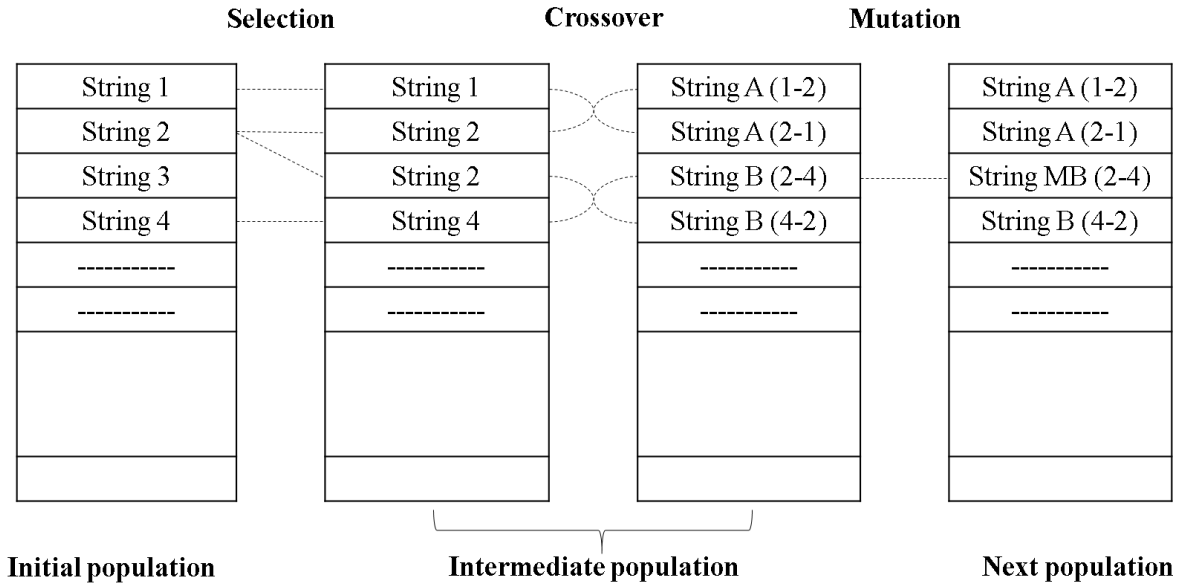


Figure 3.3: Genetic algorithm phases starting from selection to mutation through intermediate population. Where string MB is the mutated version of string B.

number of input parameters regardless of the subset generated by the genetic algorithm. Our representation scheme is illustrated in Figure 3.4.

This representation of the phenotype is straightforward and the decoding procedure involves neural network individuals having one node in the output layer (GHI estimator) and one hidden layer. Our architecture takes into consideration the fact that one hidden layer with sufficient number of hidden neurons is capable of approximating any continuous multivariate function [127].

We have adopted this representation in order to evolve neural network structure easily and to create new individuals by applying the proposed crossover operator. Moreover, this coding strategy is easily scalable to large networks. In fact, adding additional inputs to the network does not necessitate any major modifications in the structure of the model: the length of the chromosome is equal to the number of input parameters.

The basic GA operators are selection, crossover and mutation. Before addressing the chosen operators for our developed model, the following section points out the adopted evaluation function.

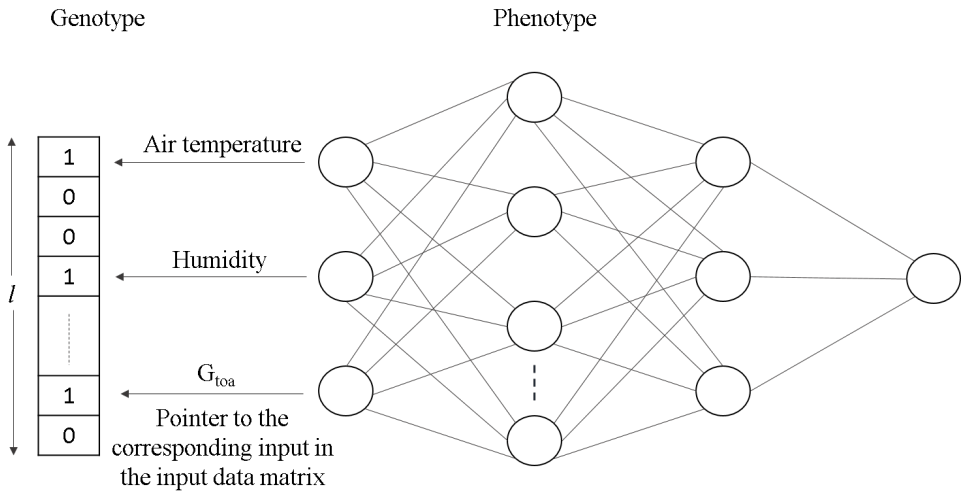


Figure 3.4: Our genetic algorithm representation of the ANN estimator, where l is the number of total inputs.

3.4.3.2 Evaluation function

The fitness function is used to evaluate each candidate solution in the GA. During each generation, individuals are selected for reproduction based on their fitness.

In this chapter, the fitness function of an individual ind_i is the maximum value of its corresponding performance over the total number of epochs:

$$f(ind_i) = \max_{j \leq 1 \leq N_{epochs}} R^2(ind_i) \tag{3.1}$$

N_{epochs}: total number of epochs
R^2: coefficient of determination calculated for the test dataset

3.4.3.3 Selection scheme

Selection in GA aims at giving higher probabilities to fitter individuals in a population so that they can produce hopefully fitter offspring. This is where the principle of survival of the fittest comes in [128].

In this work we used the roulette wheel selection where each individual has a selection probability proportional to its fitness [129].

The generation gap defines the replacement strategy i.e. the percentage of the population to be replaced during each generation. A value of G=0.5 means that half of the individuals in each population survive into the next generation. In our case the entire

population is replaced during each generation in order to avoid the problem of super individuals [130] where few individuals dominate the whole population due to their extremely high selection probabilities. Such elite individuals reduce the diversity of the next population.

We define, for a giving population j, the best individual as:

$$f_{best}^{j} = \max_{1 \leq i \leq N_p} f(ind_i) \tag{3.2}$$

N_p: population size

Finally, we define the selected individual over N_g generations as:

$$f_s = \max_{1 \leq j \leq N_g} f_{best}^{j} \tag{3.3}$$

3.4.3.4 Crossover process

In most cases, crossover consists of selecting a number of crossover points along the parent's chromosomes. Then, the operator exchanges segments of the two chromosomes delimited by cutting points [131] to form new offspring. In our work, one and two cutting points crossover are used, it's realized by cutting the strings at the same (fixed) or random positions. The offspring are hence generated by taking one or more segment parts from each parent as illustrated in Figure 3.5. The procedure for two point's crossover involves an exchange of the segments in the positions $(i, i+1)$ and $(j, j+1)$.

Our coding strategy and crossover operator guarantee the production of valid offspring in the decoding phase. The crossover percentage used in our algorithm is 1 in order to ensure the adopted replacement strategy.

3.4.3.5 Mutation

Generally, mutation consists of flipping one randomly chosen bit from zero to one or vice versa. Other mutation types exist in the literature such as boundary, uniform, non-uniform and Gaussian which may differ for the adopted encoding strategy (binary or real-valued encoding) [132]. Often, the mutation is controlled by the mutation rate, which is very low for GAs in comparison with the crossover rate [130]. For a given chromosome, the mutation percentage is defined as the number of genes that will be randomly flipped

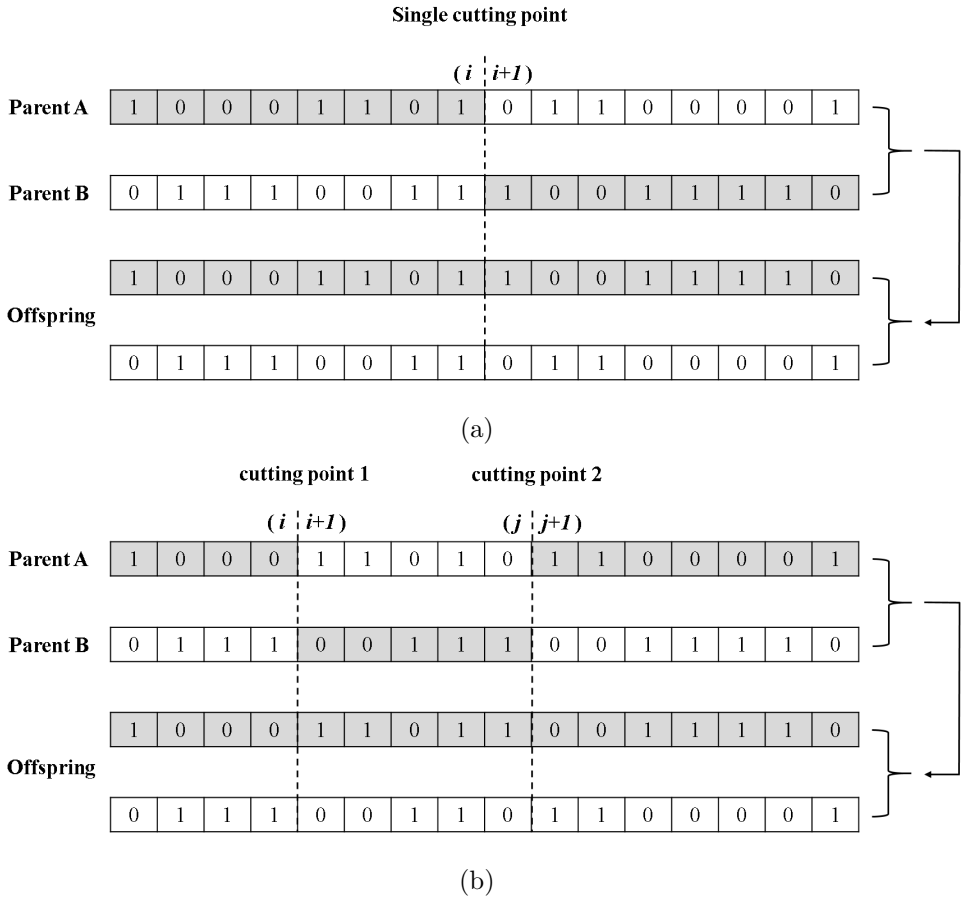


Figure 3.5: Illustration of the adopted crossover of two parents using (a) a single cutting point in the (i, i+1) position and (b) two cutting points in the positions $(i, i+1)$ and $(j, j+1)$.

over the total number of genes. As used and assessed in some theoretical studies, we adopt a mutation percentage of $1/l$ where l is the length of the chromosome [133, 134] and flipping as mutation strategy.

3.4.3.6 Block diagram and algorithm of the proposed method

In the following, we present an illustration of our method (Figure 3.6) and the steps of the proposed evolutionary algorithm for neural network estimator (Algorithm 2), used to select the most relevant variables for the estimation of global solar irradiation.

Note that the satisfying performance can be fixed previously depending on the required precision of the solar energy system.

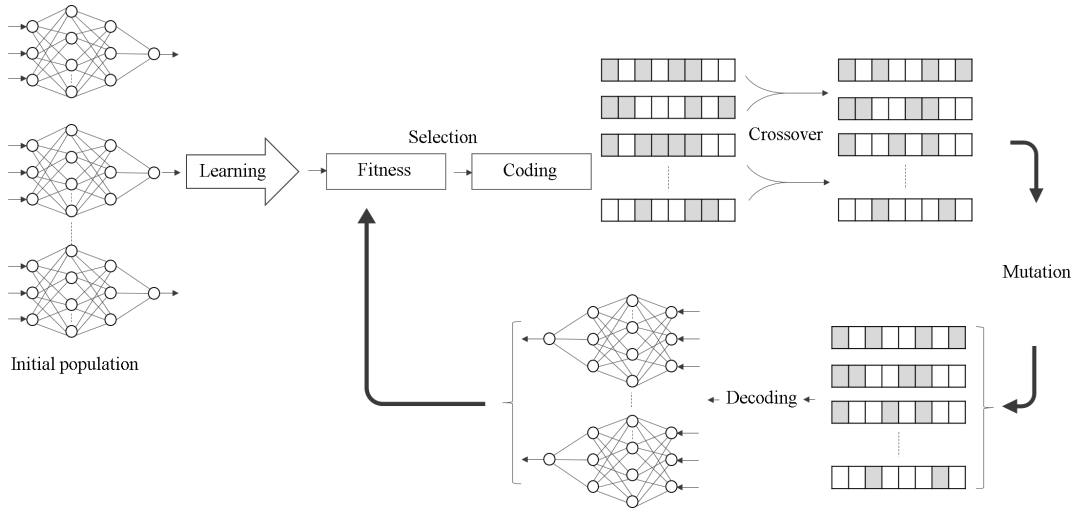


Figure 3.6: The block diagram of the proposed method.

Algorithm 2 : The proposed evolutionary algorithm for artificial neural network inputs selection.

1. Generate initial population of size N_p.
2. Train the individuals using BP over N_{epochs} epochs.
3. Evaluate each member of the population using equation 3.1.
4. Apply the selection operator by the roulette wheel method.
5. Apply the crossover to the selected parents based on the fixed or random position schemes.
6. Mutate with respect to the previously cited rate.
7. Replace the parents by the produced offspring to form the next generation.
8. Decode the descendants and go to step 2.

Repeat step 2 to 8 until N_g generations or satisfying performance R^2 is reached.

Obtain the selected individual as described in equation 3.3 and note the corresponding inputs.

3.5 Experimental results and discussion

In this section, we will present the performances of the two selection models (heuristic and forward stepwise regression selection) explained previously. We will then present an exploration of the proposed EANN model. This exploration contains the effect of maximum generations, population size, cutting points positions and nature and finally the mutation percentage. Note that all the experimental results are obtained for the test dataset.

Table 3.2: Combination of inputs and their performances using the heuristic search model.

Combination															Performance indicators		
Y	Mon	D_m	Rf	ΔT	Rh_m	SA	G_{toa}	T_{max}	T_{min}	D_y	Rh_{max}	Rh_{min}	$S\Delta T$	WS	WD	R^2	MAPE	NRMSE
X							X	X								95.69	31.83	22.55
				X	X		X			X						96.79	22.31	20.01
	X			X	X	X										96.98	21.22	19.35
		X			X		X						X			96.81	21.80	19.90
					X	X							X			96.79	22.53	19.93
								X	X	X	X	X				96.68	22.04	20.36
								X	X		X	X				94.57	27.60	25.53
						X	X	X	X							96.37	24.93	21.30
				X	X	X									X	96.95	22.98	19.41
				X		X	X								X	96.74	22.25	20.05
				X	X	X					X			X	X	96.87	23.69	19.78
				X	X	X										97.16	21.18	19.31
X	X	X				X	X									87.02	134.25	33.52
	X						X	X	X	X						96.24	24.53	21.75
X	X		X									X	X	X		93.60	33.28	27.62

3.5.1 Performance indicators for model 1

The purpose of heuristic search is to build and explore different combination possibilities based principally on the synthesis established in the state of art section. Table 3.2 shows, through 10 runs, the experimental results for multiple input subsets to choose the best combination based on their performance indicators.

The best combination is obtained for ΔT, SA and Rh_m as inputs with a performance of 97.16% for R^2 and 19.31% for NRMSE.

3.5.2 Performance indicators for model 2

In this model, after an individual performance measurement, the single most relevant candidate variable is selected first, and then forward stepwise regression selection proceeds by iteratively identifying the next most relevant candidate and evaluating whether the variable should be selected or discarded.

3.5.2.1 Individual performance

In the first step of this model, we start with evaluating individually the average performance of each input variable (obtained for 10 runs). Table 3.3 summarizes the obtained performance indicators for the test dataset sorted decreasingly.

Table 3.3: Individual performance of each ANN input parameter.

Variable	R^2 (%)	MAPE (%)	NRMSE (%)
Dy	93.65	44.17	26.97
Mon	93.63	44.92	26.92
T_{max}	92.92	39.18	28.59
G_{toa}	92.91	45.99	28.48
SA	92.90	45.85	28.55
Rh_m	91.50	43.67	30.98
Rh_{min}	91.37	37.49	31.69
$S\Delta T$	89.27	40.08	35.19
ΔT	88.90	39.83	35.90
T_{min}	87.92	63.21	35.96
Rh_{max}	87.42	65.46	36.53
Rf	84.99	56.17	40.42
WS	81.50	72.08	44.15
WD	77.62	92.41	47.82
D_m	76.53	92.39	48.51
Y	74.95	66.11	50.38

From Table 3.3 we can see that day of the year and month, indicating the season, give the best performances followed by the meteorological parameter T_{max} (maximum temperature).

3.5.2.2 Forward stepwise regression selection method performances

The second phase of our forward stepwise regression model is to trials increasingly larger subsets by adding/discarding inputs until the best combination is reached. The obtained results of this methodology are dressed in Table 3.4.

Best combination obtained with this model is D_y, T_{max}, SA and Rh_m with R^2=96.54% and 20.45% for NRMSE.

Table 3.4: Performances of the stepwise regression forward selection method.

Combination	Performance indicators		
	R^2 (%)	MAPE (%)	NRMSE (%)
Dy	93.65	44.17	26.97
Dy, Mon	93.42	44.64	27.51
Dy, T_{max}	95.54	30.05	23.21
Dy, T_{max}, G_{toa}	95.52	29.95	23.32
Dy, T_{max}, SA	95.56	29.65	23.19
Dy, T_{max}, SA, Rh_m	96.54	25.99	20.45
Dy, T_{max}, SA, Rh_m, Rh_{min}	95.59	25.77	23.42
Dy, T_{max}, SA, Rh_m, SΔT	96.42	23.68	21.18
Dy, T_{max}, SA, Rh_m, ΔT	96.52	23.23	20.88
Dy, T_{max}, SA, Rh_m, T_{min}	96.48	23.93	20.96
Dy, T_{max}, SA, Rh_m, Rh_{max}	95.90	27.72	22.40
Dy, T_{max}, SA, Rh_m, Rf	93.99	24.53	26.24
Dy, T_{max}, SA, Rh_m, WS	86.62	28.38	39.15
Dy, T_{max}, SA, Rh_m, WD	95.67	27.93	22.94
Dy, T_{max}, SA, Rh_m, D_m	96.31	27.26	21.16
Dy, T_{max}, SA, Rh_m, Y	88.01	46.19	37.33

3.5.3 Performance indicators for the proposed model

To verify the effectiveness of the proposed selection method, a first set of experimental results is carried out using population size N_p=50, crossover rate=1, mutation rate=0.1, mutation percentage=$1/l$ and cutting points=1 as initial parameters. More exploration of those parameters will be presented in the remainder of this section. Note that as for the previous sections, all the results obtained were performed 10 times to avoid random effects because of the initial weight initialization and the random nature of genetic operators. This procedure is generally used to ensure the reliability of the estimation model [116, 92]. Table 3.5 summarizes the obtained performance indicators for different maximum generations. After several tests with different hidden neuron's numbers we found that 9 nodes in the hidden layer offer a good trade-off between performance and computation time. Hence, all the individuals in the population have 9 nodes in the hidden layer.

This first preliminary set of experiments show higher performances (even for low num-

Table 3.5: The performances and best input combinations versus max generations of the proposed model.

Max generations	Performance indicators			Best combination
	R^2 (%)	MAPE (%)	NRMSE (%)	
10	97.25	21.60	18.58	2-4-5-9-15
20	97.30	19.69	18.48	6-7-9-15-16
30	97.33	22.37	18.24	7-9-11-14-15-16
40	97.25	22.26	18.59	2-3-8-15
50	97.29	22.05	18.41	7-9-15
60	97.37	19.34	18.27	6-7-9-15
70	97.24	19.89	18.82	2-4-6-9-15-16
80	97.31	19.39	17.99	1-3-4-5-8-14-15
90	97.23	21.56	18.52	1-2-5-6–10-13-14-15-16
100	97.24	22.42	18.61	2-3-4-9-10-12-13-14-15-6
150	97.11	22.14	19.14	1-4-5-8-9-11-12
200	97.07	20.97	19.08	2-4-5-6-9-12-13-14
250	97.43	20.53	19.93	2-7-9-10-14-15

ber of generations) compared with those obtained with model 1 and 2. The obtained performances are greater than 97% for R^2.

In order to explore the effect of the population size, a second set of experiments was conducted. The results are listed in Table 3.6 where experiments are performed up to 100 as maximum generations.

Table 3.6 shows that we obtain good performances with population size $N_p = 4$ over 20 generations which in addition gives low computational burden in comparison with higher population sizes.

3.5.3.1 Effect of cutting points number and nature

Based on the results obtained in the previous section, this subsection investigates the impact of one and two cutting points crossover and its nature (fixed or randomly chosen). Table 3.7 presents the resulting performances for a population of 4 individuals and 20 maximum generations with different explored positions (this Table shows only 3 positions

Table 3.6: Population size versus max generation of the proposed model.

Max gen	Population size											
	4			16			32			100		
	R^2	NRMSE	MAPE	R^2	NRMSE	MAPE	R^2	NRMSE	MAPE	R^2	NRMSE	MAPE
10	96.81	19.87	22.31	97.07	19.07	23.80	96.95	19.42	24.13	97.13	18.96	20.89
20	97.50	17.85	18.46	97.04	19.15	24.18	97.23	18.65	21.97	97.20	18.74	21.66
30	97.09	19.18	19.59	96.94	19.64	22.07	97.15	18.79	21.72	97.27	18.56	19.26
40	97.07	19.21	19.29	97.07	19.15	21.71	97.31	18.17	23.15	97.28	18.50	18.54
50	97.00	19.27	22.57	97.01	19.26	21.33	97.13	19.13	21.32	97.37	18.35	19.07
60	96.78	20.08	24.85	97.24	19.86	18.67	97.15	18.79	21.72	97.36	18.12	21.33
70	97.07	19.06	19.29	97.30	18.54	19.54	97.14	18.74	22.54	97.41	17.89	21.27
80	97.11	19.36	20.97	97.38	18.14	20.87	97.23	18.78	19.45	97.28	18.59	19.95
90	97.07	19.11	23.10	97.30	18.54	19.53	97.49	17.68	22.06	97.32	18.02	24.4
100	96.96	19.41	22.43	97.19	18.58	20.74	97.25	18.66	20.59	97.26	18.59	19.95

giving the best performances). Best combination of inputs is obtained for one cutting point fixed at position 8. Performing a cutting in the middle of the genotype favorises more exchange of the building blocks enhancing the final performance of offspring.

Table 3.7: Performances of the proposed model with different cutting points positions and nature.

Performance indicator	One cutting point crossover				Two cutting points crossover	
	Fixed			Random	Fixed	Random
	Pos 4	Pos 8	Pos 12		(4-5) & (12-13)	
R^2 (%)	97.00	97.50	97.42	96.89	97.33	96.83
NRMSE (%)	19.53	17.85	17.88	19.80	19.98	19.92
MAPE (%)	20.36	18.46	21.99	19.91	21.83	25.83
Best combination	6-8-15-16			5-6-7-8-16	5-6-7-8-11-15	4-9-10-15-16

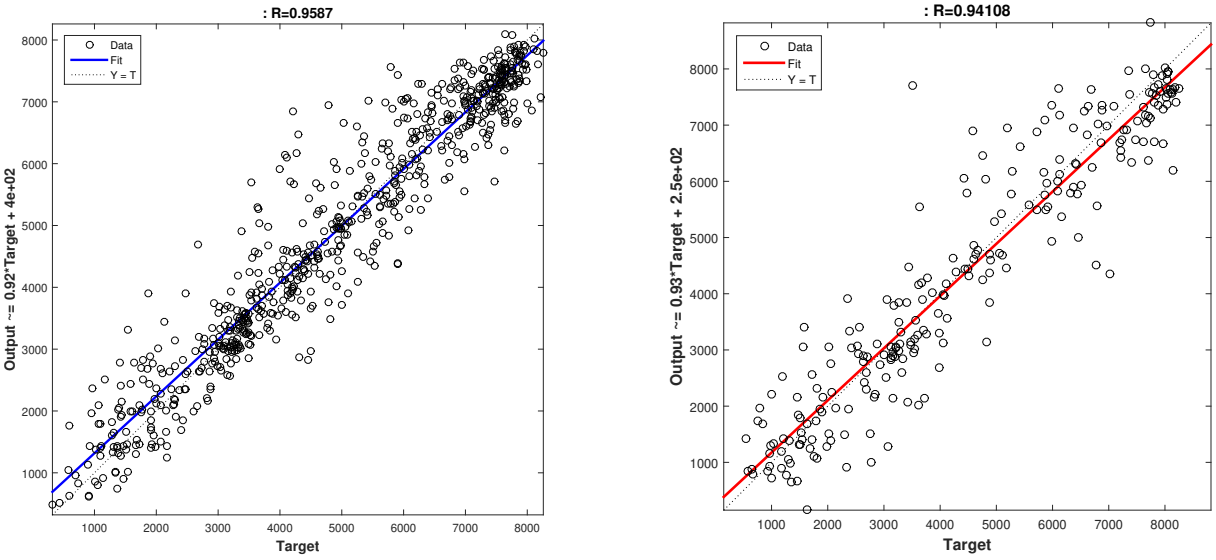
*Pos : Cutting point position

3.5.3.2 Effect of mutation percentage

In this subsection, we carry out an investigation on the effect of mutation percentage on the proposed model's performances. Table 3.8 shows the performance of the model using a population of 4 individuals and 20 maximum generations with different mutation percentage. The best result is obtained with $1/l$ which preserves the balance between the diversity brought by the mutation operator in one hand and the behavioral link between parents and their descendants on the other hand.

Table 3.8: Effect of mutation percentage on the performance of our model.

Mutation percentage	Performance indicators			Best combination
	R^2 (%)	NRMSE (%)	MAPE (%)	
$1/l$	97.50	17.85	18.46	6-8-15-16
$2/l$	96.97	19.09	19.63	4-8-7-12-14
$3/l$	96.84	19.83	21.48	8-10-11-16
$5/l$	96.52	20.77	24.93	1-2-9
$6/l$	96.69	20.41	22.63	2-3-5-8-10-11



(a) (b)

Figure 3.7: Measured (target) versus estimated (output) daily global solar irradiation obtained with (a) train and (b) test datasets.

Table 3.9: Comparison of the implemented models with our proposed model in terms of performance indicators and found inputs.

Model	Input parameters	R^2 (%)	NRMSE (%)	MAPE (%)
Model 1	ΔT, SA, Rh_m	97.16	19.31	21.18
Model 2	D_y, T_{max}, SA, Rh_m	96.54	20.45	25.99
Our proposed model	Rf , WD, ΔT, G_{toa}	97.50	17.85	18.46

Figure 3.7 shows the measured versus the estimated daily global solar irradiation with a population size of 4 and 20 as maximum generation.

Launching our proposed algorithm as described in Algorithm 2 gives the following

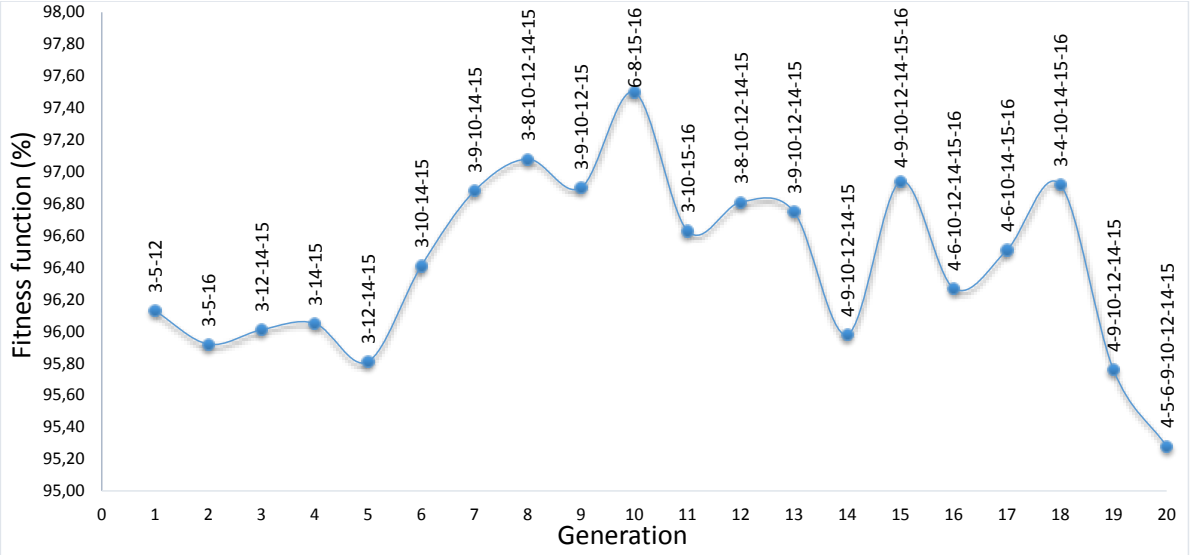


Figure 3.8: Obtained performances of our proposed algorithm through 20 generations with the corresponding input parameters using 4 individuals in the population.

performances (Figure 3.8).

Table 3.9 summarizes the performance of our model compared with the two previously described selection models. Our model apart from giving good performances as an estimator, allows an automatic selection of the best inputs for the ANN with lower population size and maximum generation ensuring good computation time.

3.6 Esimation performance comparison with other methods

In this section, a second set of experiments was performed in order to compare the performances in term of estimation of the proposed EANN model with some well known empirical models. To accomplish this comparative study, we choose to implement three empirical models. In fact, various other empirical models exist in the literature, however they assume the availability of other meteorological parameters not available in our case study. The comparison contains also the implementation of a rarely used model which is the k Nearest Neighbors model. A brief state of the art of the aforementioned estimation models is established in the next section followed by the corresponding experimental results.

3.6.1 Implemented empirical models

One of the most popular family of models in solar irradiation estimation are empirical models [135]. These models exploit empirical relationships between solar irradiation and existing climatological parameters. Ångström (1924) [96] pioneered this approach and then Ångström-Prescott models were developed based on Sunshine duration and have been widely used to predict solar irradiation [71, 136, 137]. However, most stations have only limited measured parameters (air temperature, humidity and rainfall) which make Sunshine duration based models inapplicable in those cases.

Additionally, due to the common availability of daily maximum and minimum air temperatures, several temperature-based models have been proposed to estimate solar irradiation. These models assume that a clear sky will increase the daily maximum temperature due to higher shortwave radiation, while minimum air temperature will decrease with reduced long wave emission from the atmosphere, which make the difference between daily maximum and minimum air temperature an indicator of cloudiness [138].

The first temperature-based model for estimating global solar irradiation was established by Hargreaves and Samani [139], who estimated GI as a function of temperature range and extra-terrestrial irradiation. Samani in [140] proposed a polynomial model for the estimation of GI based on maximum and minimum temperature, on data from 65 weather stations in the United States. The equation for this model is:

$$GI = G_{toa}[(a(T_{max} - T_{min})^2 + b(T_{max} - T_{min}) + c)\sqrt{T_{max} - T_{min}}] \tag{3.4}$$

Li et al. in [141] proposed another polynomial model for estimation of GI, where the empirical coefficients are: $a = 0.043$, $b = -0.040$ and $c = -0.072$. This model estimates the global solar irradiation as:

$$GHI = G_{toa}(aT_{max} + bT_{min} + c) \tag{3.5}$$

Mellit et al. in [142] proposed an other model which is formulated as:

$$GHI = aG_{toa}\sqrt[4]{T_{max} - T_{min}} + b \tag{3.6}$$

Wu et al. proposed in [143] a relationship for the estimation of GI as a function of air temperature and rainfall where Rf_t is the transformed rainfall data, which is calculated using the following conditions: if $Rf > 0$ then $Rf_t = 1$ else if $Rf_t = 0$, where Rf is the

Table 3.10: Typical and adjusted coefficients for the three implemented empirical models.

Model	Typical coefficients				Adjusted coefficients			
	a	b	c	d	a	b	c	d
Empirical model 1	0.0018	0.0433	0.0402	–	-0.0001	0.0069	0.1062	–
Empirical model 2	0.0430	-0.0400	-0.0720	–	0.0390	-0.0340	-0.0610	–
Empirical model 3	0.5800	0.2830	–	–	0.3620	-749.60	–	–
Empirical model 4	-0.1580	0.1930	0.0050	0.1550	-0.1100	0.1930	0.0100	0.1550

Table 3.11: Performances of the implemented empirical models.

Empirical model	Performance indicators					
	Typical coefficients			Adjusted coefficients		
	R^2 (%)	NRMSE (%)	MAPE (%)	R^2 (%)	NRMSE (%)	MAPE (%)
Empirical model 1	29.97	110.05	322.58	95.50	24.29	23.73
Empirical model 2	88.22	41.57	31.91	95.25	25.16	24.09
Empirical model 3	77.73	50.52	128.24	95.97	22.02	30.10
Empirical model 4	93.97	26.99	33.27	95.78	23.79	22.92

measured daily total rainfall. This model stipulates that:

$$GHI = G_{toa}(a + b(T_{max} - T_{min}) + c\frac{(T_{max} - T_{min})}{2} + dRf_t) \tag{3.7}$$

More details about those temperature based models can be found in [144].

In this chapter, due to our available data, the four pre-cited models were implemented in order to evaluate their performances and compare them with the proposed EANN model. These models were particularly selected on the basis of their wide applicability to different climates and regions. When applying those models in our estimation problem, we used the typical coefficient values and then, we adjusted those coefficients to the site under study in order to determine the best results. These coefficients are dressed in Table 3.10. The performances of those estimation models are shown in Table 3.11.

3.6.2 k Nearest Neighbors (kNN) model

kNN algorithm is a non-parametric method used essentially for classification and regression [145]. Only few papers were devoted to the use of kNN in the field of estimation, especially for solar irradiation [146, 147]. The aim of this section is to use the kNN method as a predictor of solar irradiation. The main idea of the resulting model is to determine

which of the k instances in the training dataset are most similar to a new sample according to a distance metric. In this work we used Euclidean distance to measure the distance between a sample x to estimate and the training samples:

$$d(x, T^j) = \sqrt{\sum_{i=1}^{N}(x_i - T_i^j)^2} \tag{3.8}$$

N is the number of inputs for a given sample,
x_i is the ith input of the sample to estimate.

In the case of one neighbor, the estimation of the sample x noted y is obtained by:

$$y(x) = t_m \tag{3.9}$$

t_m is m^{th} measured daily global solar irradiation

where the index m is obtained by:

$$T^m = \min[d(x, y^j)], \quad j = 1, \cdots, M \tag{3.10}$$

M is size of the training dataset.

In the case of the k Nearest Neighbors model, the estimation of new sample is taken as the average of the values of its k nearest neighbors according to the same distance metric adopted in this work. Table 3.12 shows the obtained performances with different combination sets using one neighbor.

After that, the effect of k (the number of neighbors) is evaluated and dressed in Table 3.13. This evaluation uses the best combination obtained for one neighbor: Rh_{max}, Rh_{min}, ΔT and SA with R^2=96.11% and NRMSE=21.98%.

The best result is obtained for k=6. Even with an increased number of k the performances are approximately the same with R^2=96.63% and NRMSE=20.27%.

As a summary of obtained results, Table 3.14 shows the evaluation performances of each implemented estimation model compared with our method. Results show that in term of estimation, our proposed EANN model outperforms the other studied models.

Table 3.12: Performance of the implemented kNN model with k=1 for different combinations of input parameters.

Combination																Performance indicators		
Y	Mon	D_m	Rf	ΔT	Rh_m	SA	G_{toa}	T_{max}	T_{min}	D_y	Rh_{max}	Rh_{min}	$S\Delta T$	WS	WD	R^2	MAPE	NRMSE
			X	X		X	X									92.14	40.16	31.02
				X	X		X			X						91.53	44.44	32.03
					X		X	X	X							91.65	41.58	32.04
						X		X	X					X		93.51	28.89	28.51
								X	X	X	X	X		X		94.65	32.54	27.28
	X										X	X	X	X		84.94	41.55	43.43
X				X		X	X									88.56	48.28	31.75
				X		X					X	X			X	94.00	30.84	26.16
				X		X					X	X				96.11	26.43	21.98
				X			X							X		92.12	40.20	31.11
											X	X	X		X	89.99	44.49	35.09
						X	X	X	X	X				X	X	89.43	48.46	36.29
					X	X								X	X	93.61	35.51	28.19
				X		X								X		94.65	27.82	26.05
X											X	X	X		X	90.59	44.14	34.04
	X	X			X	X	X									91.41	42.39	32.49

Table 3.13: The performances of the kNN model with different number of neighbors (k).

Performance indicators				Performance indicators			
k	R^2 (%)	MAPE (%)	NRMSE (%)	k	R^2 (%)	MAPE (%)	NRMSE (%)
2	96.24	23.91	21.58	13	96.39	25.39	20.87
3	96.44	24.23	20.93	14	96.38	25.61	20.90
4	96.59	23.97	20.44	15	96.39	25.56	20.83
5	96.62	24.13	20.33	16	96.40	25.49	20.83
6	96.63	24.16	20.27	17	96.43	25.36	20.73
7	96.50	24.67	20.64	18	96.45	25.33	20.68
8	96.49	24.68	20.64	19	96.44	25.55	20.67
9	96.50	24.73	20.58	20	96.46	25.56	20.62
10	96.45	25.20	20.72	21	96.45	25.58	20.65
11	96.45	25.07	20.73	22	96.41	25.78	20.76
12	96.42	25.14	20.79	23	96.42	25.88	20.72

Table 3.14: Comparison of implemented models with our EANN model in term of estimation.

Estimation model	R^2 (%)	NRMSE (%)	MAPE (%)
Empirical model 1	95.50	24.29	23.73
Empirical model 2	95.25	25.16	24.09
Empirical model 3	95.97	22.02	30.10
Empirical model 4	95.78	23.79	22.92
kNN model	96.63	20.27	24.16
Our proposed model	97.50	17.85	18.46

3.7 Conclusion

This chapter aims to consolidate the vision of the Moroccan project for the transition to a cleaner energy production which plans to increase the contribution of renewable energies to 42% of electricity production by 2020. This will prevent the emission of about 3.7 million tonnes of CO_2 per year [115]. For fulfilling this objective of developing thermal and photovoltaic solar energy systems, this chapter proposes an EANN model in the field of solar irradiation estimation. Our model allows an automatic selection of the best inputs for the ANN and recorded excellent performances in term of estimation of global solar irradiation in the city of Fez, using commonly measured meteorological parameters. Those performances are obtained with reduced population size and maximum generation yielding low computational burden.

Our selection method was compared with two models of selection (heuristic search and forward stepwise regression) and then, validated against three empirical models and kNN model, showing good performance in term of estimation. Results in term of crossover operator show that a cutting in the middle of the genotype favors more exchange of the building blocks of the parent chromosomes enhancing the final performance of the offspring. Concerning the mutation percentage, the best result is obtained with flipping a single bit in the chromosome which preserves the compromise between the population diversity and the behavioral relationship between parents and offspring. In our case study the proposed model gives a performance of 97.50% for R^2 and 17.85% for NRMSE with rainfall, wind direction, daily temperature gradient and global solar irradiation at the top of the atmosphere as the best combination of input parameters.

Giving the promising obtained results of our proposed EANN model in the estimation of the daily global solar irradiation for the city of Fez, in the next chapter, a new extended version of the proposed EANN model will be presented. The proposed model will be developed aiming to:

- predict future values (forecast from 1 to 6 hours ahead) of solar irradiation,
- forecast hourly instead of daily values of solar irradiation,
- forecast global, diffuse and direct solar irradiation values,
- assess the model's generalization capability in different climatic zones of Moroccan cities.

Chapter 4

Short term solar irradiance forecasting via a novel evolutionary multi-model framework and performance assessment based on a zoning strategy

4.1 Introduction

4.1.1 Background

Morocco is the largest net energy importer in the North Africa region with over 90% of its energy from foreign sources [148, 115]. To overcome this dependency, the country is making great efforts in renewable energy development and its integration into existing and eventually emerging grids [9].

The biggest challenge in the penetration of renewable energy into the grid is the high variability of the output power of photovoltaic panels and solar thermal power systems which is strongly affected by the variability of the solar resource [10]. Hence, higher penetration capacity needs accurate forecasting of solar irradiance since forecasting errors can have economic consequences on electricity markets [17, 18, 19].

In fact, several solar irradiance forecasting techniques have been proposed in the literature and numerous review papers presented different classifications of these techniques [149, 150, 151]. The most recent and well established classification is proposed by Yang et al. in [152]. This study categorized these methods into five major classes: time series, regression, Numerical Weather Prediction (NWP), sky image-based forecasting and machine learning techniques.

A time series forecasting method uses a conceptual model to predict future values based on observed ones, usually modeled through a stochastic process which introduces time as one of its explicative variables [153, 154]. Regression is a statistical model which is widely used to investigate the linear or non-linear relationship between a target and independent variables. This makes regression very useful in solar forecasting in terms of modeling exogenous variables. However, forecast horizons using this method are often limited [152]. NWP models use measurements of weather and the dynamic modeling of the atmosphere to infer local cloud formation probability to finally simulate the irradiance fluxes at multiple levels in the atmosphere [155, 152]. The sky image-based models focus on the determination of clouds cover and cloud optical depth at a designated time and space. These clouds information have the strongest influence on solar irradiance. Hence, extrapolation of the clouds motion leads to a forecast of cloud positions which allows the forecast of solar irradiance [156].

Recently, Machine Learning techniques (ML) are the most used models in solar fore-

casting [152]. They can be broadly defined as learning-based computational methods using information from empirical data to improve performance or to make accurate predictions [157]. Depending on the characteristics of the output, they can be used for different applications such as classification, clustering and regression [157, 158, 159].

ANNs, as one of the most popular ML approaches, have proven to be a useful tool to perform computational simulation with a high potential for modeling non-linear functions and complex systems with unknown behavioral rules [160, 21, 161]. Especially in the field of solar irradiance estimation and forecasting, ANNs are the most frequently used models [158, 116].

4.1.2 Aim of the study

As recently highlighted in [25], among the most important issues using ANN model in solar irradiance forecasting are the choice of the optimal number of hidden layers, hidden nodes and input nodes providing better accuracy. In fact, as conventionally known, ANNs architecture has a significant impact on the network's information processing capabilities [24], especially on the speed and performance of backpropagation trained networks [23]. Architecture design is then crucial in the successful application of ANNs. Up to now, architecture design is still very much a human expert's job [25, 23]. This manual neural network design is something of black art [26] and it can be assumed that most of such network designs are not optimal.

To overcome the unstructured process of determining good network architecture [162], other families of connectionist methods have been proposed. They fall into two general categories: some methods for network growing called constructive algorithms which initially adopt a simple network and add nodes and connections, while destructive methods start with a large network and remove the unneeded elements [163, 164]. As well indicated in [165], such hill climbing methods are vulnerable against local optimum solutions and explore only restricted topological subsets instead of investigating the complete set of network architectures.

In the field of solar irradiance modeling, several authors highlighted these issues [66, 116, 152, 25] and most of the proposed models used a tedious trial and error process to determine the best topology [103, 105, 12, 108, 161].

In this chapter, a new optimization method of ANN architecture through evolution-

ary algorithms (EAs) is proposed in order to determine automatically the near-optimal topology of the ANN forecaster. In fact, an important class of connectionist methods and broadly used in the literature is called evolutionary optimization. This has led to the development of Evolutionary Artificial Neural Networks (EANNs) in which evolution is another fundamental form of adaptation in addition to learning [166, 167]. EANNs can hence be regarded as a general framework for adaptative systems [166] i.e systems that can change their architectures and learning rules adaptatively without human intervention [166]. EANNs have been shown to possess several advantages over conventional methods of training [168]. Consequently, there has been a great interest in the last decades in the use of evolutionary approaches in the training of artificial neural networks [169, 170, 171]. As precedently and pertinently suggested by Miller et al. [26], the main reason for evolutionary computation to be good candidate for this neural adaptation is that the fitness associated with the space of topologies is complex, noisy, non-differentiable, multi-modal and deceptive.

In our proposed model, the Backpropagation (BP) learning algorithm is integrated as a new search operator into EA in order to improve individuals locally before they are evaluated and give rise to another generation [171]. The basic idea of this proposed hybrid evolutionary method is a global sampling by evolutionary algorithm over the space of alternative solutions with respect to network topologies, while the BP algorithm proceeds by locally searching the immediate neighborhood of the current solution. The interaction between global and local search is performed via a proposed crossover operator where we explore new topologies using EA and the BP. The proposed crossover operator explores large possible solutions and yields valid individuals through some constraints on the cutting point positions.

Furthermore, in order to ensure the coding scalability [172], the search of the neural architecture is obtained via simple and compact representation of the phenotype (potential solutions) composed of the number of layers and neurons in each layer. Moreover, in order to guarantee the population diversity during evolution, the selection scheme uses the roulette wheel method and a non-overlapping as a replacement strategy. This non elitist strategy in which the offspring population replace entirely the parent population tends to make the search more explorative rather than exploitative [173].

Besides, the purpose of this chapter is to forecast the upcoming hourly solar irradiance at different time horizons based only on endogenous historical data (baptized forecasting

history). Using an optimal number of historical data for the prediction allow more efficient forecasting models. Only few authors address the importance of this parameter and its impact on the forecasting. In fact, Mihalakakou et al. in [102] determined empirically the number of past measured values used as inputs to predict the future values of total solar irradiance. The obtained results show that 13 inputs are needed to forecast from 10 to 20 days ahead in summer, while in winter 9 past values of solar irradiance are used to forecast insolation up to 8 days. In [174], Wu et al. used cross validation experiment in order to find the optimal length of window size with an optimal value of 6. Benali et al. proposed the auto-mutual information method to determine the size of the forecasting history (called input matrix). The authors showed that the number of input data for predicting global, bean and diffuse solar irradiance are 6, 7 and 10 [111].

In our work, forecasting history is considered as a supplementary parameter in our evolutionary algorithm. In order to evolve this parameter, it was included and coded in the first part of our final adopted chromosome. The objective is to automatically explore the best forecasting history combined with the best ANN architecture. The resulting model is baptized HAEANN (History and Architecture EANN).

In this work, a new short term multi-model framework is proposed allowing the generation of multiple HAEANN models for different time horizons to forecast global, diffuse and direct solar irradiance from 1 to 6 hours ahead. Each of which evolves simultaneously forecasting history and ANN architecture, searching automatically for the most accurate model.

In the present work, a useful zoning scenario is investigated presenting a particular interest in the case of the lacking of solar irradiance measurements in some areas, and where future values of solar irradiance can be obtained using data from stations belonging to the same zone. Two other partitioning scenarios are used to assess and compare the resulting performances. Those three scenarios manipulating data from 28 Moroccan cities with different climatic conditions are described as follows:

- first scenario called “individual_city scenario” in which, for each considered city, train and test phases are performed with the same city dataset,
- second scenario where sites belonging to a considered zone are used as training data to forecast solar irradiance of cities from the same zone. This data partitioning is named “zoning scenario”,
- third scenario where we build a single train dataset with different cities (belonging

to different zones) and a single test dataset with the remaining cities. Given that, the union of train and test datasets form the whole available database containing 28 Moroccan cities, then the resulting partitioning will be titled "all_cities scenario".

Finally, a comparative study is conducted with a reference model namely Smart Persistence model (SP) and two well-known machine learning techniques namely Regression Trees (RT) and Random Forest (RF) models. This comparison, developed for these scenarios, is presented as averages of the obtained errors for 6 hours ahead forecasting, followed by a detailed visualization of each hour error for global, diffuse and direct solar irradiance forecasting.

As a summary, the main contributions of our work include the following aspects:

1. a multi-model framework aiming an automatic determination of history forecasting and ANN architecture is proposed to forecast global, diffuse and direct solar irradiance,
2. three scenarios are inspected and compared to evaluate the validity of the zoning scenario in the case of sites with no solar irradiance data,
3. the generated HAEANN models (from the proposed framework) are used to forecast solar irradiance for different horizons from 1 to 6 hours using only endogenous data,
4. a reference model with two machine learning based models are implemented and compared with our proposed model, considering the three aforementioned scenarios to forecast global solar irradiance and its components up to 6 hours ahead.

4.2 Data preparation

In this work, hourly horizontal global, horizontal diffuse and normal direct solar irradiance data have been acquired from Meteonorm software, from January 1st to December 31th 2005, available for 28 Moroccan cities. Table 4.1 dresses the geographical coordinates and corresponding climate and signification (see Table 4.2), according to the Köppen climate classification [122], for all studied cities.

A pre-processing step was applied in order to remove night hours and to conserve only diurnal hours. For this, solar irradiance data with solar altitude lower than 10° were removed [111].

Table 4.1: Geographical coordinates and their corresponding climate for all studied cities.

Site	Latitude(°)	Longitude(°)	Altitude(m)	Climate	Site	Latitude(°)	Longitude(°)	Altitude(m)	Climate
Agadir	30.38	-9.57	23	BSh	Marrakech	31.62	-8.03	466	BSh
Al Hoceima	35.18	-3.85	14	Csa	Meknes	33.88	-5.53	549	Csa
Beni Mellal	32.36	-6.4	468	Csa	Midelt	32.68	-4.73	1515	BSk
Casablanca	33.57	-7.67	55	Csa	Nador	35.15	-2.91	16	Csa
Dakhla	25.48	29	106	BSh	Ouarzazate	30.93	-6.9	1140	BWh
Elayoune	27.16	-13.21	64	Csa	Oujda	34.68	-1.83	545	Csa
El Jadida	33.21	-8.5	0	Csa	Rabat	34.05	-6.77	75	Csa
Errachidia	31.93	-4.4	1045	BWh	Safi	32.28	-9.23	45	BSh
Essaouira	31.52	-9.78	8	BSh	Salé	32.28	-9.23	45	Csa
Fez	33.93	-4.98	579	Csa	Sidi Ifni	29.36	-10.18	66	BWh
Ifrane	33.5	-5.17	1665	Csa	Tanger	35.74	-5.83	85	Csa
Kenitra	34.3	-6.6	14	Csa	Taza	34.22	-4	510	Csa
Khouribga	32.9	-6.95	774	Csa	Tetouan	35.58	-5.33	10	Csa
Larache	35.18	-6.13	49	Csa	Tan Tan	28.45	-11.15	200	BWh

Table 4.2: Signification of the abreviation used in Table 4.1.

Climate	Signification
BSh	Hot semi-arid climate
Csa	Hot summer mediterranean climate
BSk	Cold semi-arid climate
BWh	Hot desert climate

In order to evaluate the proposed models, in the case of sites with no solar irradiation data, a zoning partitioning of the considered Moroccan cities was adopted (see Table 4.3). In fact, climatic zones in Morocco have been delimited through different studies realized by the National Meteorology Directorate (DMN) and the Moroccan Energy Efficiency Agency (AMEE). Subsequently, AMEE has proposed a unique climatic zoning including six climatic zones, which is in line with current administrative limits. Other zoning configurations are possible, however, this partitioning was proposed for an effective application of the new thermal regulation in Morocco (see Figure 4.1) [175]. This unique zoning is based on hourly annual climate files and the results of simulations carried out by the building annual thermal requirements of heating and air conditioning in Morocco [176].

Table 4.3: Distribution of the studied Moroccan cities according to the thermal climatic zoning.

Climatic zones	Cities	Number of cities
Zone 1	Agadir - Elayoune - Casablanca - Dakhla - El Jadida - Essaouira - Kenitra - Rabat - Safi - Sidi Ifni- Salé	11
Zone 2	Al Hoceima - Larache - Nador - Tanger - Tetouan	5
Zone 3	Beni Mellal - Fez - Khouribga - Meknes - Oujda - Taza	6
Zone 4	Ifrane - Midelt	2
Zone 5	Marrakech	1
Zone 6	Errachidia - Ouarzazate - Tan Tan	3
		28

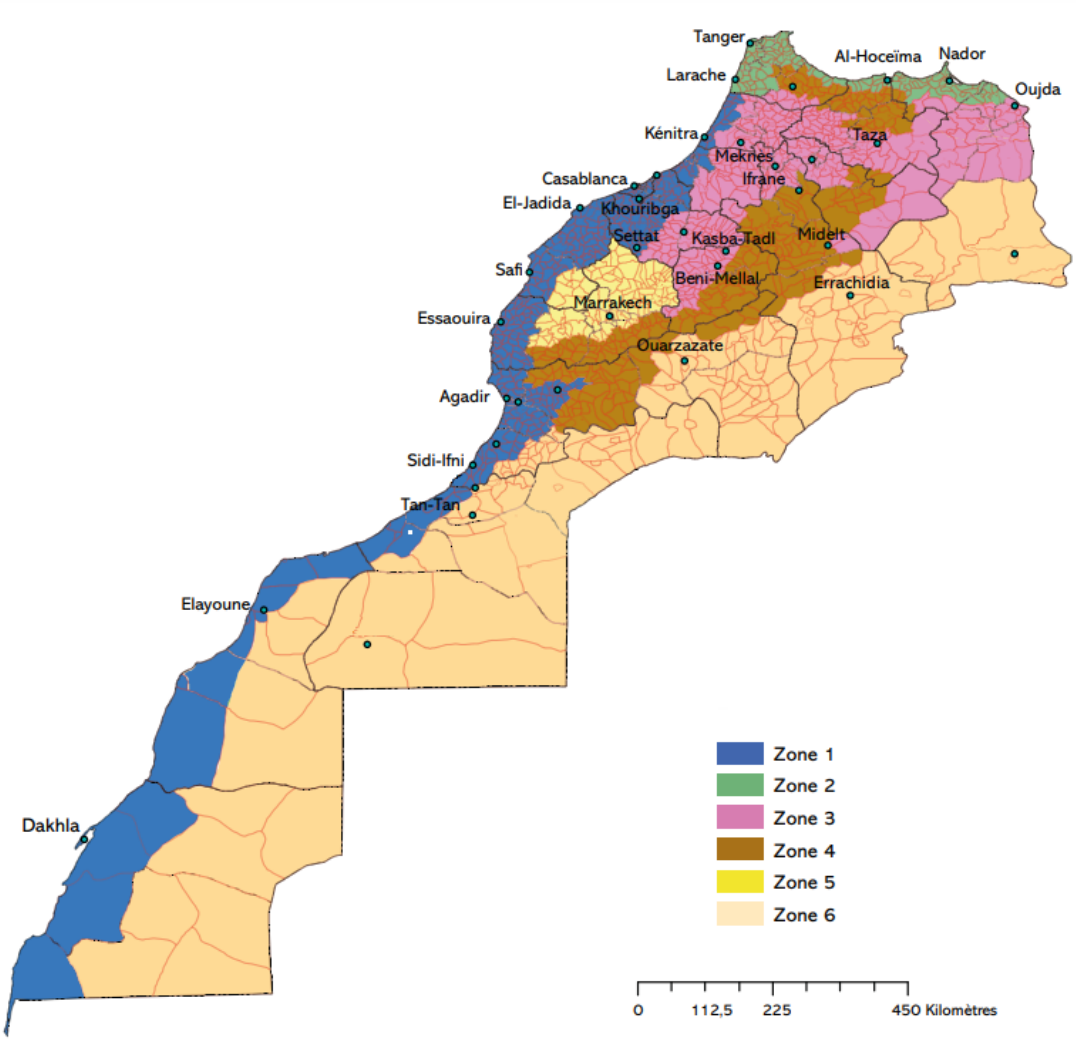


Figure 4.1: The adopted thermal climatic zoning map in Morocco [175].

4.3 The proposed framework and models structure

4.3.1 The proposed framework

Before describing the details of the generated models, it is worth noting that our framework aims to generate H HAEANN models with ascendant number of outputs up to H (H being the forecasted horizon i.e the future global, diffuse and direct solar irradiance values to predict). Hence, the HAEANN model i corresponds to a forecasting of i hours ahead. For instance, for H=6 (forecasting up to 6 hours ahead), the generated models by the framework are:

- HAEANN 1: forecasts 1 hour ahead,
- HAEANN 2: forecasts simultaneously 1 and 2 hours ahead,
- ...,
- HAEANN 6: forecasts simultaneously 1, 2, 3, 4, 5 and 6 hours ahead.

The framework parameters can be mathematically described by the quadruplet: ($Max_history$, M, Max_hid_nodes, H) defined as follow:

- $Max_history$: the maximum value of hours which can be used as history (input data of our models) to forecast future solar irradiance values,
- M : maximum number of hidden layers in the generated models of the framework,
- Max_hid_nodes : maximum number of hidden nodes in each layer in the generated models of the framework,
- H : forecasting horizon.

The flowchart of the proposed framework is illustrated in Figure 4.2. This process is an EANN which principally resolves the problematic of the choice of ANN architecture when using ANNs as irradiance forecasting models. The aim of our framework is to generate H models obtained from H adaptation cycles with an evolution of forecasting history and ANN architecture. Beginning from an initial evolution pool containing randomly generated ANN forecaster individuals, our framework evolves simultaneously the forecasting history and ANN architecture through proposed genetic operators ending with a final evolution pool. This evolutionary framework is constructed based on Algorithm 3.

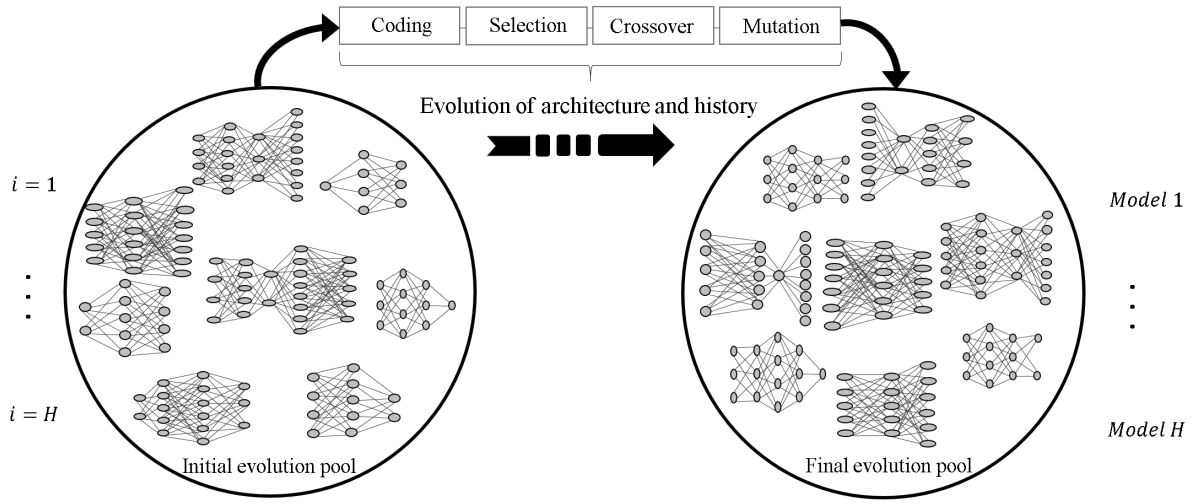


Figure 4.2: Flowchart of the proposed evolutionary framework generating multi-HAEANN models.

Algorithm 3 :The proposed algorithm of our evolutionary multi-HAEANN models framework.

1. choose the framework parameters (*Max_history*, *M*, *Max_hid_nodes*, *H*),
2. for each H model:

2.1. launch the evolutionary process with an initial evolution pool of size p by:

a. coding the forecasting history and ANN architecture in two parts of the chromosome according to Figure 4.5,

b. selecting best individuals referring to the fitness function defined in equation 4.3,

c. mating the individuals using the crossover operator with at least two cutting points considering the constraints explained in equations 4.6 and 4.7,

d. applying the mutation operator on two parts of the chromosome as explained in subsection 4.3.6,

2.2. produce the next generation using the non-overlapping replacement strategy (see subsection 4.3.4),

2.3. for each next generation, verify the stopping criteria :

(the best fitness $<$ threshold*) or ($N_G{}^+$ generations reached),

2.4. note for the best individual:

a. the forecasting history,

b. the ANN architecture,

c. the forecasted global, diffuse or direct solar irradiance values.

*Threshold is selected based on the desired forecasting accuracy.

^+N_G is the predetermined number of generations for the genetic algorithm.

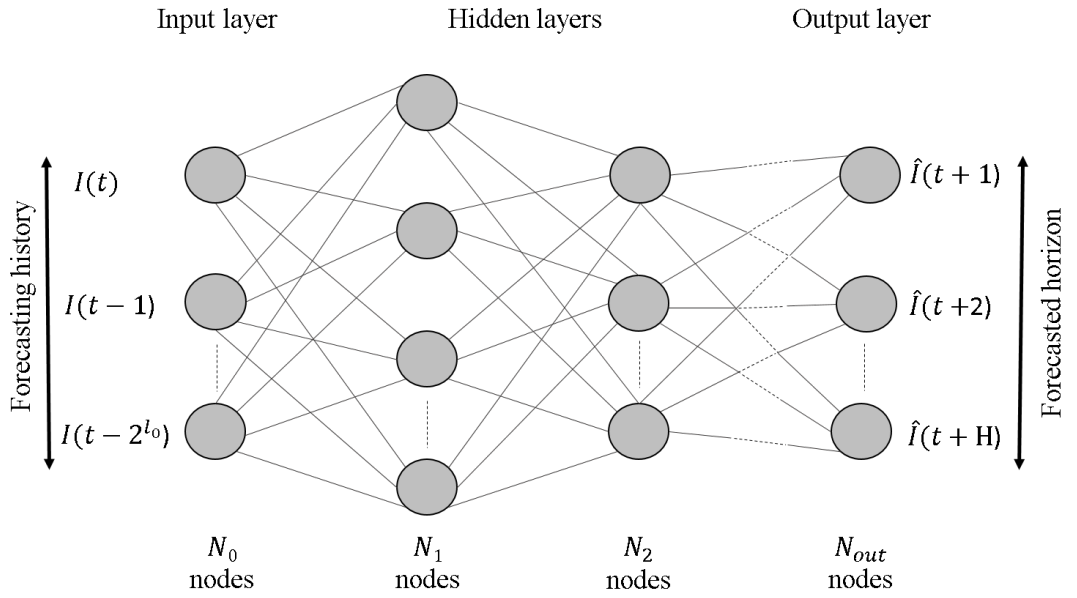


Figure 4.3: The adopted artificial neural network forecaster.

In the present work, the H models will be evaluated and compared. However, we can directly choose i hours to forecast and the evolutionary process will automatically generate the best ANN architecture and forecasting history.

4.3.2 Models description

In order to forecast one or multiple future global, diffuse or direct solar irradiance values noted $\hat{I}(t+i)/i = 1, \cdots, H$, we use a historical data (forecasting history) $I(t-j)/j = 0, \cdots, 2^{l_0}$. Where l_0 and $Max_history$ are related using equation 4.1.

Figure 4.3 shows the structure of each phenotype representing the evolved ANN forecasting model. The ANNs adopted in this work are the multilayer perceptrons (MLP), which is a particular family among the feed-forward artificial neural networks. In the output layer we use multiple nodes, each one corresponds to a forecasted hour from $t+1$ to $t+H$. N_i is the number of neurons in each layer where $N_i \leq 2^{l_i}$, $i = 0, \cdots, M$ and $N_{out} \leq H$.

Figure 4.4 illustrates the generated ANN model forecaster using the proposed framework in the case of three hours forecasting.

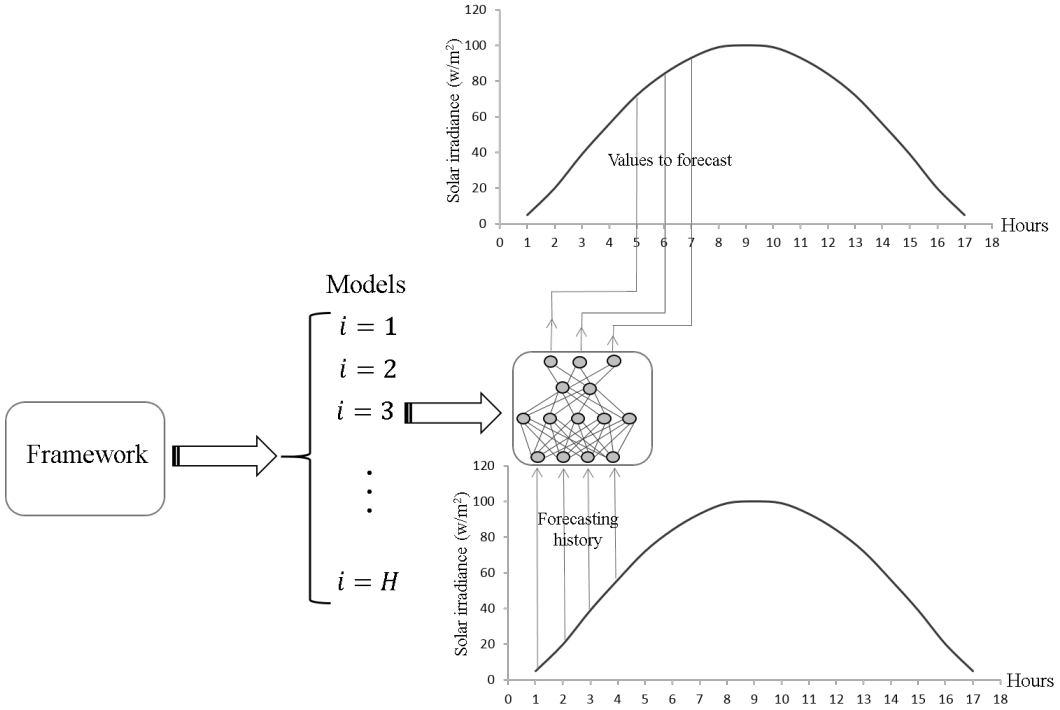


Figure 4.4: Illustration of the generated neural network forecaster from the proposed framework up to 3 hours ahead ($i = 3$).

4.3.3 Genotype coding

In this subsection, we describe the genotype coding where a simple and compact representation of the chromosome is adopted. Figure 4.5 shows the coding chromosome of the proposed ANN forecasters with different lengths representing the forecasting history and ANN architecture.

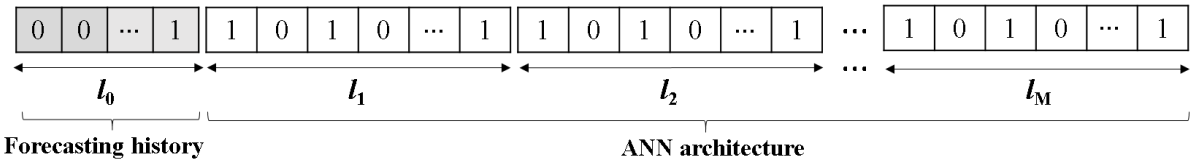


Figure 4.5: The proposed chromosome representing the forecasting history and ANN architecture with $M + 1$ parameters where M is the maximum number of hidden layers.

Related to the framework parameters, we must respect the following conditions:

$$2^{l_0} \leq Max_history \tag{4.1}$$

$$2^{l_i} \leq Max_hid_nodes, \quad i = 1, \cdots, M \tag{4.2}$$

The first part of the chromosome encodes the forecasting history with a length l_0. The remaining parts represent the ANN architecture. Each subpart is the number of neurons in the corresponding hidden layer with a length l_i where $i = 1, \cdots, M$ and M is the maximum number of hidden layers. Figure 4.6 shows an illustration of our coding scheme, this correspondence will be adopted in the remaining of this work to represent each individual in our population. For instance, an individual noted $7-9-3$ uses 7 hours as forecasting history, 2 hidden layers with 9 and 3 hidden neurons in each one respectively.

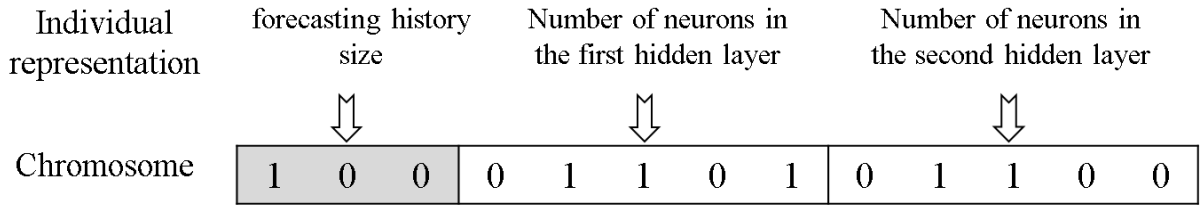


Figure 4.6: Our proposed chromosome representation of the neural network forecaster with corresponding parameters.

Note that, in the case of a forecasting history N_0, the coded value in the chromosome is $f_{db}(N_0 - 1)$ where f_{db} is the conversion function from decimal to binary basis system. This encoding scheme is used to avoid invalid individuals (with only zero binary values) which may occur after crossover in the produced generations. Hence, one hour as a history is represented by zeros in the first part of the chromosome.

4.3.4 Fitness function and selection strategy

The fitness function used in our model corresponds to the minimum Normalize Root Mean Square Error (NRMSE) of a potential solution x_k. This function is expressed as:

$$f(x_k) = \min_{1 \leq j \leq N} NRMSE(x_k), \quad k = 1, \cdots, p \tag{4.3}$$

Where N is the total number of epochs used for ANN training and p is the population

size and where :

$$NRMSE = \frac{100}{\frac{1}{n}\sum_{i=1}^{n} t_i} \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\theta_i - t_i)^2} \tag{4.4}$$

Normalized Mean Absolute Error (NMAE) performance indicator is also considered in this study, where the NMAE corresponding to the minimum Normalize Root Mean Square Error (NRMSE) of a potential solution x_k is expressed in equation 4.5.

$$NNMAE = \frac{100}{\sum_{i=1}^{n} t_i} \sum_{i=1}^{n} |\theta_i - t_i| \tag{4.5}$$

with :

n is the number of samples,

t_i is the ith measured value,

θ_i is the ith forecasted value.

To breed new individuals in the mating pool, the fitness function is used to select best individuals. The roulette wheel strategy is adopted as a selection method. This strategy gives better chances to the fittest individuals of the population to produce the next generation, and also allows the remaining individuals to survive in the next mating pool.

In order to guarantee the population diversity, we use the non-overlapping strategy where the search is more explorative and where parents and offspring do not coexist in the same population [173].

4.3.5 Crossover operator

Figure 4.7 illustrates the proposed crossover operator where we need to use at least two cutting points in order to evolve both forecasting history and ANN architecture. More cutting points can be chosen in the general case. The first and second cutting points (P_1 and P_2 respectively) are constrained to the following conditions:

$$0 < P_1 < l_0 \tag{4.6}$$

$$l_0 < P_2 < L, \qquad L = \sum_{j=0}^{M} l_j \tag{4.7}$$

Where L is the total length of the chromosome.

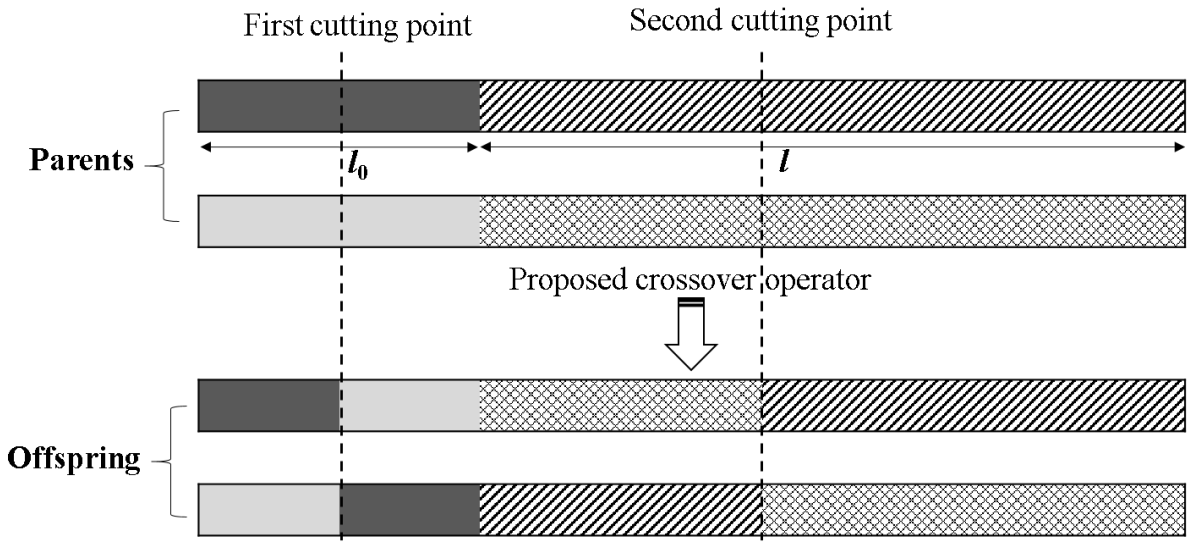


Figure 4.7: Our proposed crossover operator in the case of 2 cutting points.

More exploration of other crossover strategies is illustrated in the experimental results (section 4.5.1.2) to see the effectiveness of our proposed operator.

4.3.6 Mutation operator

The mutation operator aims to introduce changes in the structure of potential solutions in order to guarantee diversity. As mutation may produce major changes in the structure of individuals, its rate must be tuned appropriately in order to preserve the inheritance link between parents and offspring. This tradeoff is expressed by the mutation rate and mutation percentage [81]. The mutation is applied with a mutation rate of $1/l_0$ in the first part of the chromosome having length with the condition:

$$0 \leq length \leq l_0 \tag{4.8}$$

For the second part of the chromosome ($l_0 \leq length \leq L$), the mutation rate of $1/l$ is used where:

$$l = \sum_{j=1}^{M} l_j \tag{4.9}$$

4.4 Training procedure

In each individual of our populations, we use sliding window approach across the training samples as explained in Figure 4.8. In fact, this approach has a changing starting point in the training sample which is rolled forward by removing the first data of the sample and adding the latest one to the end. Another horizon step ahead forecast is then made. This process is repeated until all the training dataset is attained. Contrarily to rolling window approach, our technique assumes that old data is no longer relevant to the prediction of the next sample [177].

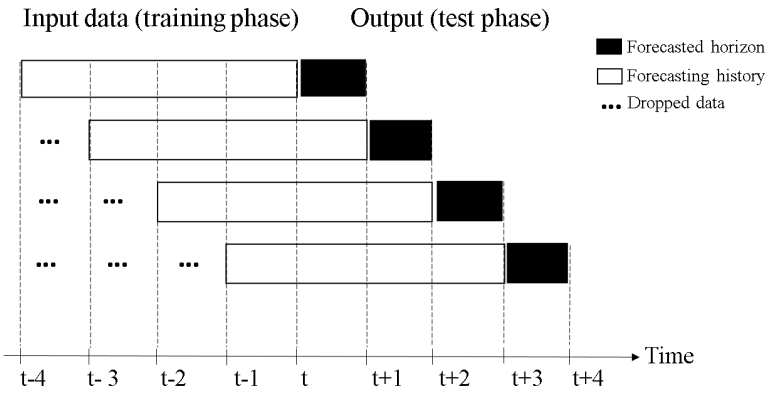


Figure 4.8: The adopted sliding window training procedure for all proposed models (inspired from [177]).

4.5 Experimental results

In the present section, we will adopt the following methodology to present the obtained results of our models. First, we will explore different: combinations of activation functions, crossover operators and population sizes, that will be used in the remaining of the exposed results, based on the NRMSE and NMAE performances. Then, the models are evaluated in terms of datasets constructed from three data partitioning scenarios: individual_city scenario, zoning scenario and all_cities scenario for global, diffuse and direct solar irradiation. In the first scenario, for each city, train and test phases are realized with the same city data. For the second scenario, cities of a particular zone are used as training dataset to predict solar irradiance of other cities belonging to the same zone. In the third scenario, a single train dataset is constructed from different cities (from different zones) to forecast solar irradiance and its components of the remaining cities. This will

yield three subsections developing our experimental exploration. In order to compare the performances of the generated models, we choose to use the same cities in the test phase for the three scenarios. Seasonal behavior of the proposed models is then investigated. Finally, a comparison study is established with a reference model: smart persistence and two machine learning techniques: regression trees and random forest models.

4.5.1 Global, diffuse and direct solar irradiance forecasting using the proposed models

4.5.1.1 The choice of the activation functions

In artificial neural networks modeling, the input-output behavior is greatly determined by the used activation functions [23]. In this subsection, an investigation of the best choice of activation functions adopted in our HAEANN proposed model is first conducted. Three of the most used transfer functions are investigated, including the positive linear (poslin), hyperbolic tangent sigmoid (tansig) and log-sigmoid (logsig) functions for hidden layers [156]. The linear (purelin) function is adopted for the output layer [156]. In our ANN forecasters, the Levenberg-Marquardt backpropagation algorithm is used.

Table 4.4 summarizes the obtained results for a chosen city (Marrakech), zone (Zone 6) and for all_cities scenarios using HAEANN 1 model for forecasting global horizontal solar irradiance for one hour ahead.

Table 4.4: Global solar irradiance forecasting performances of HAEANN 1 using different combinations of activation functions in terms of NRMSE(%) and NMAE(%).

Activation functions combination	Marrakech			Zone 6			all cities scenarios		
	Best individual	NRMSE	NMAE	Best individual	NRMSE	NMAE	Best individual	NRMSE	NMAE
One hidden layer									
logsig	7 - 4	6.90	4.17	7 - 8	6.25	3.36	7 - 17	8.54	4.55
poslin	6 - 12	8.60	5.78	6 - 13	7.16	3.77	6 - 14	9.11	4.91
tansig	8 - 5	6.63	3.70	8 -16	5.92	2.83	8 - 9	8.47	4.50
Two hidden layers									
logsig - poslin	7 - 10 - 16	7.02	4.65	6 - 17 - 31	6.88	3.92	7 - 11 - 5	8.79	4.67
tansig - tansig	7 - 3 - 13	6.81	3.82	7 - 7 - 8	6.02	3.26	8 - 29 - 6	8.44	4.48
tansig - logsig	8 - 5 - 5	7.18	3.82	8 - 25 - 10	6.05	3.01	8 - 25 -14	8.52	4.51
logsig - logsig	4 - 13 - 12	8.86	5.71	8 - 9 - 17	6.87	3.22	7 - 20 - 7	8.59	4.58

As shown in Table 4.4, results for hyperbolic tangent sigmoid transfer function in the hidden layers show better performances for the three tested scenarios. Hence, this activation function (tansig) is chosen for the hidden layers and purelin function for the

output layer in all remaining results.

4.5.1.2 The choice of the crossover operator

Various types of crossover operators exist in the literature, the choice of an operator rather than others depends on the problem nature and also the used encoding scheme [178]. In this subsection, three different crossover operators, appropriate for our encoding scheme, are tested: fixed cutting points, random cutting points and uniform crossover operators.

Table 4.5 presents the resulting performances for a chosen city (Marrakech) and zone (Zone 6) and for all_cities scenarios with different crossover operators using model 1 (HAEANN 1) for global solar irradiance forecasting.

Table 4.5: Performances of HAEANN 1 according to different crossover operators for global solar irradiance forecasting.

Crossover operator	Marrakech		Zone_6		All_cities	
	NRMSE(%)	NMAE(%)	NRMSE(%)	NMAE(%)	NRMSE(%)	NMAE(%)
Fixed 2-points crossover	6.63	3.70	5.92	2.83	8.44	4.48
Fixed 3-points crossover	6.79	4.27	6.23	3.37	9.20	4.97
Random 3-points crossover	6.84	4.44	5.93	3.14	9.03	4.86
Unifom crossover	6.56	3.80	5.97	3.01	8.59	4.49

Table 4.5 shows that the fixed 2-points crossover operator presents the best performances of offspring in terms of NRMSE and NMAE in the three tested scenarios. Therefore, this crossover operator will be used in the remaining results.

4.5.1.3 The choice of the population size

In order to explore the population size, NRMSE and NMAE metrics of the best individuals versus population size (varying from 4 to 24) for a chosen city (Marrakech) and zone (Zone 6) and for all_cities scenario are dressed in Table 4.6 using model 1 (HAEANN 1).

From Table 4.6, we can see that even if we increase the population size, the results remain approximately the same. Hence, using 4 as a population size will consume less computation time and guarantees good performances.

Table 4.6: Global solar irradiance forecasting performances of HAEANN 1 according to different population sizes.

Population size	Marrakech		Zone 6		All_cities	
	NRMSE(%)	NMAE(%)	NRMSE(%)	NMAE(%)	NRMSE(%)	NMAE(%)
4	6.63	3.70	5.92	2.83	8.21	4.16
8	6.84	4.16	6.04	3.01	8.28	4.24
16	6.66	3.79	5.92	2.79	8.24	4.20
24	6.73	3.84	5.92	3.12	8.27	4.24

In order to visualize the impact of the chosen population size on the diversity of the produced offspring, the following experiments were conducted where we represent the forecasting history and ANN architecture of each individual in generated populations using 4 as population size (see Figure 4.9).

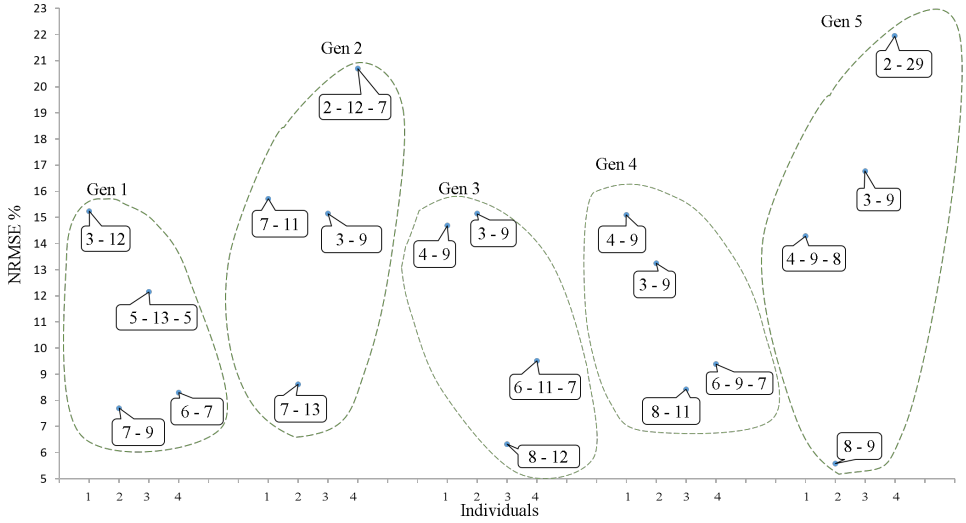


Figure 4.9: Produced offspring for different generations (Gen 1 to Gen 5) using 4 as population size for global solar irradiance forecasted by HAEANN 1 for Marrakech city.

From Figure 4.9, we can see that the use of 4 as population size guarantees great genetic diversity and increasing performance of the best individuals through generations.

These results and same investigations on other cities show that this population size is a good tradeoff between diversity and computation time. Hence, all the remaining results will be presented and explored using the population size 4.

In each of the proposed models, after several tests, a set of experimental parameters were approved: crossover rate= 1, mutation rate= 0.1 and a mutation percentage of

$2/L$ conducted with N_G= 50. For the ANN forecaster, the Levenberg-Marquardt back-propagation (LM) algorithm was used, with hyperbollic tangent sigmoid as an activation function in hidden layers and linear activation function in the output layer. The total number of epochs used for ANN training is equal to 100. The initial weights of the ANN forecaster are randomly generated and the presented results are given through 10 runs to ensure the reliability of the forecasting models [116].

As already proven, one or two hidden layers with an adequate number of neurons is sufficient to model any solution surface [78], besides, most works in the field of solar irradiance forecasting using ANNs used up to two hidden layers to achieve a good prediction accuracy [179, 180, 181]. Hence, we choose to set two as maximum number of hidden layers in our framework parameters ($M = 2$).

4.5.1.4 Performances of the proposed models for individual_city scenario

In each of the proposed models, for the studied cities, the repartition of the train and test datasets respect the division of 70% and 30% respectively which gives the corresponding sizes dressed in Table 4.7 where we dispose of 4380 pre-processed measures for each city.

Table 4.7: Division of datasets for each studied city.

Train dataset size	Test dataset size
3060	1320

Table 4.8, 4.9 and 4.10 show results for the six proposed models generated by our framework for the forecasting of horizontal global, diffuse and direct normal solar irradiance respectively. The presented results for each model i are the average performance obtained for the i hours (called hours average). Note that, to avoid exhaustive presentation of obtained results related to all cities, only performances of 10 cities are dressed. This choice is motivated by the fact that those same cities will be used in the test dataset of the other scenarios. For each model of each city, the best individual is represented by its forecasting history and ANN architecture with its corresponding performances.

Table 4.8: Global horizontal solar irradiance forecasting performances of the six proposed models for ten cities. $NRMSE$ and $NMAE$ rows represent test results.

		El jadida	Errachidia	Al Hoceima	Essaouira	Fez	Midelt
	Best individual	8 - 5	8 - 5	8 - 5	8 - 13	7 - 13	8 - 5
HAEANN 1	$NRMSE(\%)$	7.94	9.34	7.83	7.26	5.63	10.01
	$NMAE(\%)$	4.55	5.16	4.49	4.41	3.28	5.59
	Best individual	6 - 7	8 - 5	7 - 5 - 13	8 - 12	6 - 13 - 7	7 - 13
HAEANN 2	$NRMSE(\%)$	10.74	11.32	10.21	9.64	7.10	11.59
	$NMAE(\%)$	6.88	7.00	5.60	5.84	4.17	6.54
	Best individual	6 - 12	7 - 9	7 - 9	8 - 15 - 4	8 - 15 - 7	8 - 5
HAEANN 3	$NRMSE(\%)$	11.65	13.07	10.46	11.20	7.99	13.83
	$NMAE(\%)$	7.22	7.90	6.50	6.99	4.89	8.63
	Best individual	7 - 9	8 - 9	7 - 12	7 - 14	8 - 12	8 - 7 - 7
HAEANN 4	$NRMSE(\%)$	12.67	13.18	10.77	12.06	8.89	14.46
	$NMAE(\%)$	7.75	8.07	6.91	7.83	5.92	8.73
	Best individual	8 - 9	7 - 9	8 - 7	8 - 12	8 - 12	8 - 12
HAEANN 5	$NRMSE(\%)$	12.89	14.39	11.75	12.38	9.54	15.58
	$NMAE(\%)$	7.99	8.99	7.81	8.01	6.57	9.40
	Best individual	6 - 7	8 - 13	6 - 7	8 - 18	8 - 12 - 12	7 - 8
HAEANN 6	$NRMSE(\%)$	13.89	14.75	12.59	12.49	10.9	16.89
	$NMAE(\%)$	9.18	9.32	8.27	7.78	7.05	11.11

		Marrakech	Dakhla	Oujda	Tanger	Cities_average
	Best individual	8 - 5	8 - 5	8 - 7	8 - 2	-
HAEANN 1	$NRMSE(\%)$	6.63	5.41	8.24	7.63	7.59
	$NMAE(\%)$	3.70	3.70	4.49	4.71	4.41
	Best individual	8 - 13 - 5	8 - 4	7 - 15 - 4	8 - 7	-
HAEANN 2	$NRMSE(\%)$	8.25	5.67	9.33	7.99	9.18
	$NMAE(\%)$	5.16	4.03	5.83	4.73	5.58
	Best individual	8 - 5 - 8	7 - 7	8 - 15	7 - 15 - 4	-
HAEANN 3	$NRMSE(\%)$	9.40	5.97	9.77	9.62	10.29
	$NMAE(\%)$	5.61	4.52	5.76	5.28	6.33
	Best individual	8 - 5	8 - 5	7 - 12	8 - 9	-
HAEANN 4	$NRMSE(\%)$	10.52	6.16	11.27	9.79	10.98
	$NMAE(\%)$	6.85	4.59	7.03	6.09	6.98
	Best individual	7 - 7	8 - 5	8 - 7	8 - 12	-
HAEANN 5	$NRMSE(\%)$	11.11	6.46	12.06	9.96	11.61
	$NMAE(\%)$	7.12	4.85	8.02	5.88	7.46
	Best individual	8 - 5	7 - 5 - 5	7 - 12 - 4	8 - 16	-
HAEANN 6	$NRMSE(\%)$	11.36	7.25	13.07	11.02	12.49
	$NMAE(\%)$	7.37	5.66	8.83	6.67	8.12

Table 4.9: Diffuse horizontal solar irradiance forecasting performances of the six proposed models for ten cities. $NRMSE$ and $NMAE$ rows represent test results.

		El jadida	Errachidia	Al Hoceima	Essaouira	Fez	Midelt
	Best individual	6 - 5 - 7	7 - 12	8 - 7 - 2	8 - 7 - 2	7 - 12	8 - 13
HAEANN 1	NRMSE	16.18	14.01	13.81	13.82	14.21	17.79
	NMAE	11.58	9.29	10.61	10.60	10.33	12.84
	Best individual	8 - 9 - 2	7 - 13	8 - 11	8 - 11	8 - 12 - 16	8 - 5
HAEANN 2	NRMSE	18.92	19.38	18.20	17.36	18.79	21.89
	NMAE	13.87	12.84	13.10	13.30	13.55	15.51
	Best individual	8 - 9	7 - 9	8 - 13 - 7	7 - 13 - 7	6 - 15	8 - 5
HAEANN 3	NRMSE	22.72	21.49	20.72	20.21	20.53	25.44
	NMAE	16.40	14.54	14.96	15.61	15.37	18.34
	Best individual	7 - 9	8 - 9	7 - 9	8 - 9	8 - 9	6 - 7
HAEANN 4	NRMSE	24.27	23.02	22.05	20.71	20.73	28.95
	NMAE	18.02	15.54	16.28	16.16	15.24	20.96
	Best individual	8 - 9	7 - 9	8 - 11	8 - 16	7 - 9	6 - 14
HAEANN 5	NRMSE	25.36	25.52	22.93	21.39	21.99	31.26
	NMAE	18.80	17.21	16.71	16.49	16.49	22.71
	Best individual	6 - 4 - 14	7 - 9	8 - 5 - 7	8 - 9 - 16	7 - 12	6 - 7
HAEANN 6	NRMSE	26.75	27.16	24.31	21.96	23.53	33.65
	NMAE	19.95	18.66	17.98	16.40	17.59	24.64

		Marrakech	Dakhla	Oujda	Tanger	Cities_average
	Best individual	6 - 7	8 - 4	7 - 12	8 - 5	-
HAEANN 1	NRMSE	16.93	18.59	19.38	18.86	15.36
	NMAE	11.85	13.52	12.52	11.703	11.48
	Best individual	6 - 7	8 - 9	7 - 5	8 - 13	-
HAEANN 2	NRMSE	20.79	21.22	22.56	22.13	20.12
	NMAE	14.93	15.76	15.97	14.65	14.35
	Best individual	7 - 14 - 7	8 - 9	7 - 7	6 - 14	-
HAEANN 3	NRMSE	23.33	23.33	25.02	26.19	22.89
	NMAE	16.79	17.68	17.98	17.91	16.56
	Best individual	7 - 9	8 - 9	8 - 9	6 - 12	-
HAEANN 4	NRMSE	25.03	25.10	27.01	28.08	24.50
	NMAE	18.50	19.26	19.31	19.69	17.89
	Best individual	6 - 7	8 - 9	7 - 10	8 - 9	-
HAEANN 5	NRMSE	27.52	25.91	28.83	29.04	25.98
	NMAE	20.61	19.91	20.59	20.31	18.98
	Best individual	8 - 13	8 - 5	6 - 7	8 - 12	-
HAEANN 6	NRMSE	28.71	26.62	30.79	31.37	27.49
	NMAE	21.57	20.69	21.88	22.20	20.16

Table 4.10: Direct normal solar irradiance forecasting performances of the six proposed models for ten cities. $NRMSE$ and $NMAE$ rows represent test results.

		El jadida	Errachidia	Al Hoceima	Essaouira	Fez	Midelt
	Best individual	6 - 7	6 - 5	8 - 5	6 - 12	6 - 5 - 7	6 - 7
HAEANN 1	NRMSE	19.29	14.01	18.19	18.61	14.10	19.96
	NMAE	12.69	7.69	11.81	13.59	10.27	12.06
	Best individual	5 - 5	7 - 5 - 5	8 - 5	8 - 9	8 - 5 - 7	8 - 5
HAEANN 2	NRMSE	23.36	16.34	21.01	22.81	17.13	23.37
	NMAE	15.95	9.60	14.27	15.25	12.56	14.61
	Best individual	7 - 7	8 - 9	8 - 7	6 - 14	6 - 13	7 - 12
HAEANN 3	NRMSE	25.52	17.75	22.96	25.13	20.28	26.56
	NMAE	17.51	11.03	15.50	17.27	15.35	17.35
	Best individual	5 - 12	6 - 12	7 - 9	6 - 12	7 - 9	7 - 9
HAEANN 4	NRMSE	27.08	19.19	24.86	26.75	21.38	28.39
	NMAE	19.02	12.14	17.56	18.55	16.05	18.97
	Best individual	8 - 9	7 - 9	8 - 9	5 - 5	8 - 9	8 - 9
HAEANN 5	NRMSE	28.22	20.32	25.72	28.06	22.73	29.98
	NMAE	19.81	13.21	17.81	19.57	16.92	20.08
	Best individual	5 - 2 - 5	6 - 5 - 4	8 - 5	5 - 13 - 7	6 - 7	6 - 5
HAEANN 6	NRMSE	29.39	21.32	26.13	29.08	23.87	31.18
	NMAE	21.14	14.25	18.16	20.54	17.9	21.67

		Marrakech	Dakhla	Oujda	Tanger	Cities_average
	Best individual	5 - 5 - 12	8 - 5	8 - 5 - 7	7 - 12	-
HAEANN 1	NRMSE	15.06	12.29	16.72	13.61	16.18
	NMAE	9.39	8.14	9.98	8.10	10.37
	Best individual	2 - 12	8 - 6	8 - 13 - 1	8 - 5 - 4	-
HAEANN 2	NRMSE	18.06	13.72	18.78	15.93	19.05
	NMAE	11.83	9.82	12.35	10.16	12.64
	Best individual	8 - 14	8 - 6	8 - 15 - 4	6 - 14	-
HAEANN 3	NRMSE	19.84	15.30	21.46	18.19	21.29
	NMAE	13.17	11.28	14.07	12.38	14.49
	Best individual	4 - 2 - 25	7 - 4	8 - 9	7 - 9	-
HAEANN 4	NRMSE	21.70	15.99	22.14	19.37	22.69
	NMAE	15.02	12.14	15.06	13.09	15.76
	Best individual	8 - 3	7 - 9	6 - 7	6 - 14	-
HAEANN 5	NRMSE	22.91	16.98	24.07	20.57	23.96
	NMAE	15.93	12.90	17.15	14.50	16.79
	Best individual	2 - 12	8 - 12	8 - 13	7 - 12	-
HAEANN 6	NRMSE	24.15	17.92	24.59	21.29	24.89
	NMAE	17.15	13.52	17.16	14.75	17.62

From Table 4.8, the six models give good performances for the ten cities. In almost all cases, the models tend to use 6, 7 or 8 hours of forecasting history. Also, as shown in this table, our choice of two hidden layers as a maximum provides good results. In addition, even with one hidden layer, good performances can be achieved (e.g. a NRMSE of 7.26% and a NMAE of 4.41% is obtained with one hidden layer with 13 hidden neurons for Essaouira city using HAEANN 1). Besides, the hours average performances of the proposed models decrease with the increase of the forecasted horizon (e.g. a NRMSE of

9.34% and a NMAE of 5.16% are obtained using HAEANN 1 and of 14.75% and 9.32% using HAEANN 6 for Errachidia city).

Considering individual_city scenario performances, for the global, diffuse and direct solar irradiance, the best results for the six models were obtained for Dakhla city. On the other hand, the city of Midelt shows lower performances for all components.

From Table 4.9 and 4.10, we can notice that the models performances are less satisfying than with global solar irradiance forecasting. Thus, forecasting performances for diffuse and direct solar irradiance are less accurate than global solar irradiance. Particularly, because these components are more sensitive to meteorological conditions and because the variation of the direct radiation intensity is more rapid and of a greater magnitude.

An additional cities_average column is added where we calculate the average of models performance for the ten cities. These values will endly be used to compare models performances according to the three scenarios.

4.5.1.5 Performances of the proposed models for zoning scenario

Instead of using the neighboring stations data to forecast target cities based on distance criteria [161], the aim of this subsection is to present experimental results related to forecasting global solar irradiance of studied cities using data of other cities belonging to the same zone. Hence, the constructed train and test datasets of this scenario do not contain the same data cities, except for zone 5 (which contain only one city: Marrakech). Results of this portioning is given in Table 4.11.

Table 4.11: Train and test datasets division of the studied cities for each considered zone.

Zone	Train cities	Train dataset size	Test cities	Test dataset size	Total size
1	Agadir - Elayoune - Casablanca - Kenitra - Safi - Sidi Ifni - Salé - Rabat	35040	Dakhla - Eljadida - Essaouira	13140	48180
2	Larache - Tetouan - Nador	13140	Al Hoceima - Tanger	8760	21900
3	Beni Mellal - Khouribga - Meknes - Taza	17520	Fez - Oujda	8760	26280
4	Ifrane	4380	Midelt	4380	8760
5	Marrakech	3060	Marrakech	1320	4380
6	Ouarzazate - Tan Tan	8760	Errachidia	4380	13140

From Tables 4.12, 4.13 and 4.14, we can see that all generated models, trained with the zoning partitioning, succeeded to forecast global, diffuse and direct solar irradiance with NRMSE less than 17%, 26% and 22% respectively. These results show almost similar

Table 4.12: Global horizontal solar irradiance forecasting performances of the six proposed models for the six considered zones. $NRMSE$ and $NMAE$ rows represent test results.

		Zone 1	Zone 2	Zone 3	Zone 4	Zone 5	Zone 6	Zones_average
	Best individual	6 - 9	8 - 19	8 - 9	7 - 13	8 - 5	8 - 16	-
HAEANN 1	$NRMSE(\%)$	7.88	9.58	10.76	9.52	6.63	5.92	8.38
	$NMAE(\%)$	4.55	5.29	5.43	5.15	3.70	2.83	4.49
	Best individual	8 - 31	8 - 13 - 7	8 - 14	7 - 9	8 - 13 - 5	8 - 16	-
HAEANN 2	$NRMSE(\%)$	9.19	10.86	12.94	11.55	8.25	7.07	9.98
	$NMAE(\%)$	5.18	6.22	6.77	7.19	5.16	3.90	5.74
	Best individual	6 - 12 - 8	8 - 13 - 5	8 - 13	8 - 15	8 - 5 - 8	8 - 9	-
HAEANN 3	$NRMSE(\%)$	11.04	11.98	14.39	12.82	9.40	7.93	11.26
	$NMAE(\%)$	6.64	6.93	8.35	7.84	5.61	4.69	6.68
	Best individual	8 - 12 - 8	8 - 9	8 - 25	7 - 15 - 5	8 - 5	7 - 13	-
HAEANN 4	$NRMSE(\%)$	11.07	13.09	15.98	14.48	10.52	8.63	12.30
	$NMAE(\%)$	6.68	7.9	8.62	9.28	6.85	5.11	7.41
	Best individual	7 - 13 - 4	8 - 29	8 - 9	8 - 9	7 - 7	8 - 8	-
HAEANN 5	$NRMSE(\%)$	11.56	13.51	16.31	15.29	11.11	9.25	12.84
	$NMAE(\%)$	7.06	8.11	9.75	10.05	7.12	5.67	7.96
	Best individual	8 - 25	8 - 10 - 7	8 - 23	8 - 13 - 5	7 - 5	8 - 23	-
HAEANN 6	$NRMSE(\%)$	12.10	14.19	16.88	15.54	11.35	9.57	13.27
	$NMAE(\%)$	7.45	8.59	9.73	10.43	7.36	5.65	8.20

performances compared to the obtained results for individual_city scenario (see Tables 4.8, 4.9 and 4.10). These performances can be explained by the fact that the proposed models use greater number of data having the same climatological conditions, even with training data from other cities belonging to the same zone. We can note, for instance, that the city of Midelt presents better results for all models as part of zone 4, than evaluated individually in the previous scenario. However, the test dataset of zone 3 composed of Fez and Oujda cities gives for global solar irradiance forecasting using HAEANN 1 a NRMSE and NMAE of 10.76% and 5.43% respectively, while it gives 6.94% and 3.89% as a mean values of the performances of the two cities taken separately (5.63% and 3.28% for Fez and 8.24% and 4.49% for Oujda).

For this zoning scenario, zones 2 and 3 present the highest errors for the 6 models and zone 6 shows the best performances for both global and direct solar irradiance forecasting. For the diffuse component, zone 4 presents the highest error.

Note that as for the first scenario, an additional zones_average column is added where

Table 4.13: Diffuse horizontal solar irradiance forecasting performances of the six proposed models for the six considered zones. $NRMSE$ and $NMAE$ rows represent test results.

		Zone_1	Zone_2	Zone_3	Zone_4	Zone_5	Zone_6	zones_average
	Best individual	8 - 9	8 - 15	8 - 15	6 - 7	6 - 7	8 - 9	-
HAEANN 1	NRMSE	16.39	18.83	20.92	19.57	16.93	17.80	18.41
	NMAE	11.38	12.79	13.22	13.50	11.85	11.21	12.33
	Best individual	8 - 12	8 - 25	8 - 23	7 - 4	6 - 7	8 - 9	-
HAEANN 2	NRMSE	19.36	22.32	24.63	24.79	20.79	22.15	22.34
	NMAE	13.74	15.55	16.02	17.78	14.93	14.43	15.41
	Best individual	7 - 11 - 5	8 - 13 - 16	8 - 13 - 5	7 - 9	7 - 14 - 7	8 - 25	-
HAEANN 3	NRMSE	22.32	25.12	27.29	27.94	23.33	25.33	25.22
	NMAE	16.13	17.84	18.64	20.00	16.79	16.66	17.68
	Best individual	8 - 9	8 - 13	8 - 12 - 7	7 - 9	7 - 9	8 - 13	-
HAEANN 4	NRMSE	24.00	27.32	30.03	30.99	25.03	28.69	27.68
	NMAE	17.58	19.96	20.65	22.56	18.50	19.32	19.76
	Best individual	8 - 25	8 - 9	8 - 13	8 - 9	6 - 7	8 - 13	-
HAEANN 5	NRMSE	24.96	29.26	31.99	32.72	27.52	30.92	29.56
	NMAE	18.42	21.58	22.79	24.33	20.61	20.87	21.43
	Best individual	7 - 13 - 5	7 - 24	8 - 13 - 4	8 - 12	8 - 13	8 - 13	-
HAEANN 6	NRMSE	26.77	30.83	33.69	34.49	28.71	32.52	31.17
	NMAE	19.90	22.95	23.81	25.34	21.57	22.33	22.65

we calculate the average of models performance for the six zones. This metric will be used to compare the three scenarios performances.

4.5.1.6 Performances of the proposed models for all_cities scenario

Instead of rolling the data of all available cities on train and test dataset with multiple rounds of different partitions (each one with different distribution of cities), we choose to build the train and test datasets with cities belonging to all considered zones (see Table 4.15). This choice with different zones (diverse climates) is used to evaluate the generalization capacity of our proposed models. The obtained results are dressed in Table 4.16.

In most cases, we can observe that our models tend to use 8 hours as forecasting history with two hidden layers. Moreover, even if the train and test datasets were constructed from different climatic zones, the proposed models show good generalization ability for the six models with a NRMSE less than 14%, 31% and 29% for the global, diffuse and

Table 4.14: Direct normal solar irradiance forecasting performances of the six proposed models for the six considered zones. $NRMSE$ and $NMAE$ rows represent test results.

		Zone_1	Zone_2	Zone_3	Zone_4	Zone_5	Zone_6	zones_average
	Best individual	8 - 9	5 - 13 - 5	8 - 9	6 - 7	5 - 5 - 12	8 - 9	-
HAEANN 1	NRMSE	16.82	20.34	19.68	18.34	15.06	12.66	17.15
	NMAE	10.84	14.49	12.92	10.74	9.39	6.89	10.88
	Best individual	8 - 13 - 2	3 - 12 - 5	7 - 13	7 - 9	2 - 12	8 - 9	-
HAEANN 2	NRMSE	19.76	24.39	23.38	22.31	18.06	14.94	20.47
	NMAE	13.21	17.32	15.91	14.81	11.83	8.59	13.61
	Best individual	8 - 8 - 24	3 - 12	8 - 9	7 - 12	8 - 14	8 - 15	-
HAEANN 3	NRMSE	21.63	26.17	25.51	24.56	19.84	16.56	22.38
	NMAE	14.66	18.56	17.58	16.66	13.17	9.95	15.10
	Best individual	7 - 27	3 - 12	8 - 15 - 3	7 - 13	4 - 2 - 25	8 - 13	-
HAEANN 4	NRMSE	23.29	27.99	27.55	26.48	21.70	17.79	24.13
	NMAE	16.34	19.84	19.34	18.48	15.02	10.87	16.65
	Best individual	5 - 28	5 - 10 - 5	8 - 29	8 - 9	8 - 3	7 - 13 - 18	-
HAEANN 5	NRMSE	25.11	28.59	28.69	28.24	22.91	19.08	25.44
	NMAE	18.10	20.36	20.28	20.16	15.93	12.00	17.81
	Best individual	8 - 28	5 - 13 - 7	8 - 9 - 5	8 - 9	2 - 12	8 - 9	-
HAEANN 6	NRMSE	25.48	29.78	29.99	28.98	24.15	19.67	26.34
	NMAE	18.12	21.10	21.63	20.99	17.15	12.62	18.60

direct solar irradiance respectively.

It is worth mentioning that as for all previous results, increasing the forecasted horizon involves a decrease of the hours average models performances. In order to see the impact of forecasted horizon on the hourly performance of the models, we dressed the errors obtained for each hour of each model for global, diffuse and direct solar irradiance forecasting (see Figure 4.10).

From Figure 4.10, we can see that for the six models, even with the inclusion of ascendant number of forecasted horizon hours, the obtained errors remain approximately equal for the same forecasted hour. This property, gives the choice among different models depending on users needs preserving the same precision.

We can clearly see that for the three scenarios, global solar irradiance is forecasted with a better accuracy compared with diffuse and direct components. It is probably due to the fact that in global solar irradiance, the two components, diffuse and direct, have compensating effects (when diffuse increases, direct decreases) and the speed of variation

Table 4.15: Constructed train and test datasets of the studied cities for all_cities scenario.

Train Cities	Train Climatic zones	Number of cities	Train dataset size
Agadir - Elayoune - Casablanca - Kenitra - Safi - Sidi Ifni - Salé - Rabat	Zone 1	8	35040
Larache - Nador - Tetouan	Zone 2	3	13140
Beni Mellal - Khouribga - Meknes - Taza	Zone 3	4	17520
Ifrane	Zone 4	1	4380
Ouarzazate - Tan Tan	Zone 6	2	8760
		18	78840
Test Cities	**Test climatic zones**	**Number of cities**	**Test dataset size**
Dakhla - El Jadida - Essaouira	Zone 1	3	13140
Al Hoceima - Tanger	Zone 2	2	8760
Fez - Oujda	Zone 3	2	8760
Midelt	Zone 4	1	4380
Marrakech	Zone 5	1	4380
Errachidia	Zone 6	1	4380
		10	43800

Table 4.16: Global, diffuse and direct solar irradiance forecasting performances of the six proposed models for all_cities scenario.

	Global horizontal solar irradiance			Diffuse horizontal solar irradiance			Direct normal solar irradiance		
	Best individual	NRMSE	NMAE	Best individual	NRMSE	NMAE	Best individual	NRMSE	NMAE
HAEANN 1	8 - 29 - 6	8.44	4.48	8 - 12	18.1	12.05	8 - 12	17.94	11.29
HAEANN 2	8 - 9 - 8	9.94	5.51	8 - 13 -20	21.41	14.52	8 - 11 - 8	20.92	13.75
HAEANN 3	8 - 13	11.18	6.37	8 - 9	25.13	17.72	8 - 25	23.39	16.04
HAEANN 4	8 - 13 - 7	12.09	7.12	8 - 13 - 16	26.58	18.84	8 - 9	26.00	18.52
HAEANN 5	7 - 13	13.41	8.21	8 - 9 - 4	29.31	21.30	6 - 26	27.43	19.70
HAEANN 6	8 - 15 - 12	13.76	8.37	8 - 9	31.02	22.70	8 - 13 - 7	28.30	20.49

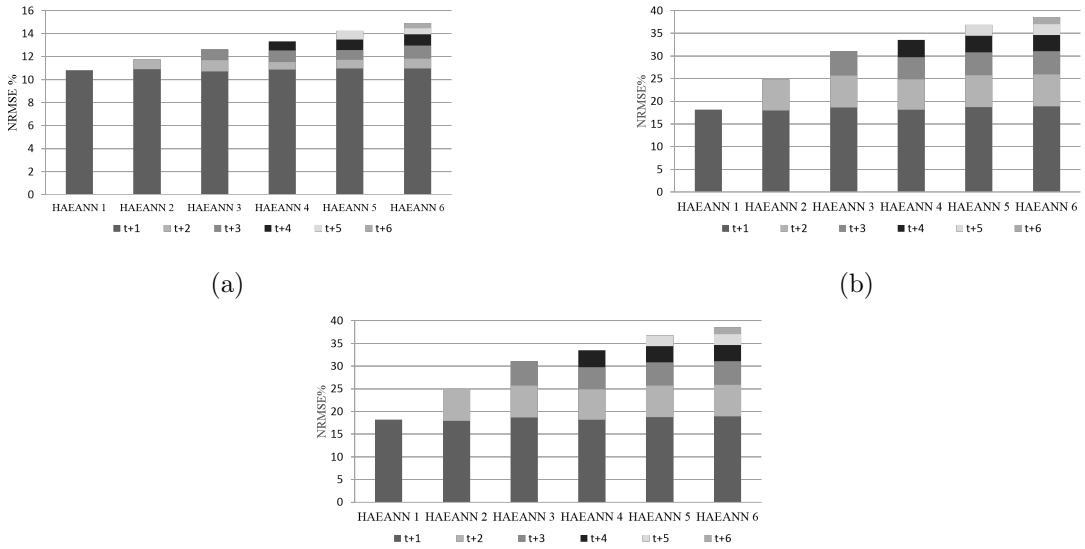


Figure 4.10: Comparison of hourly solar irradiance forecasting models for various horizons in terms of NRMSE for all_cities scenario: (a) global horizontal solar irradiance, (b) diffuse solar irradiance and(c) direct normal solar irradiance.

of global horizontal solar irradiance is less important than for direct and diffuse solar irradiance.

Furthermore, from all above obtained results for global, diffuse and direct solar irradiance forecasting, we note that the zoning and all_cities scenarios present good and almost similar average performances compared to the individual_city scenario. Those results prove the generalization ability of the proposed models. Therefore, the proposed zoning scenario can be considered as a good alternative in sites where no solar irradiance data is available and where future values can be forecasted from stations belonging to the same zone.

4.5.1.7 Seasonal behavior

Before addressing the comparison study with other machine learning techniques, a seasonal behavior of the proposed models can be illustrated in Figure 4.11 where measured and predicted global solar irradiance values are drawn for the four seasons using HAEANN 6 for Dakhla city. The aim is to observe the impact of the meteorological variability on the accuracy of the proposed models.

Figure 4.11 demonstrates the efficiency of the proposed model to forecast global solar

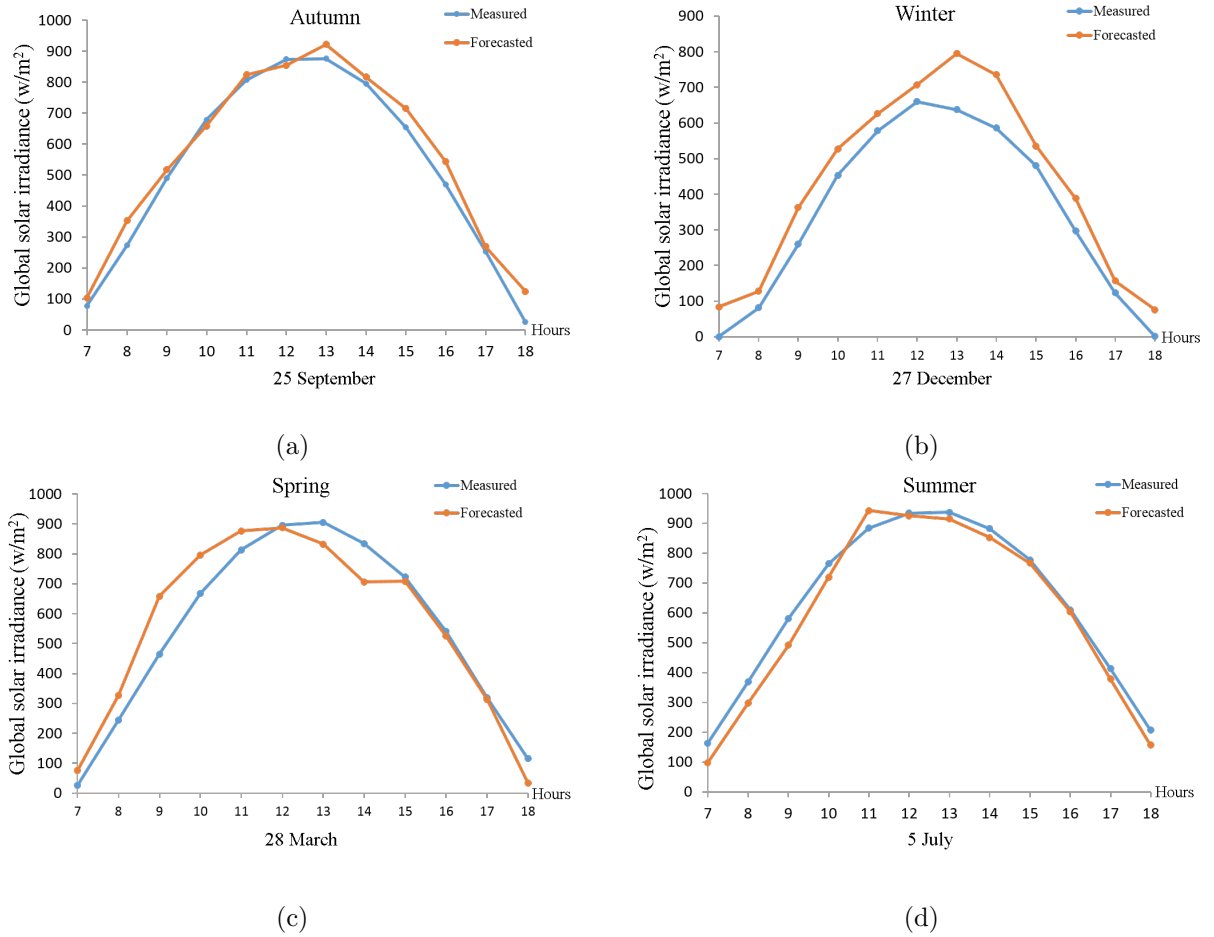


Figure 4.11: Measured vs forecasted global solar irradiance values using HAEANN 6 for Dakhla city. 1 day from each season is represented belonging to: Autumn (a), Winter (b), Spring (c) and Summer (d).

irradiance in different weather conditions. However, a higher accuracy is noticeable in summer season since the forecasting of most solar radiation models become more accurate during this season [110]. For the spring and autumn seasons, forecast error is partially due to formation and dissolution of clouds, while for winter season, low Sun elevation is the major reason why our HAEANN model forecasts solar irradiance with relatively low accuracy [182].

4.5.2 Comparison of the proposed models with smart persistence and machine learning techniques

In this subsection, smart persistence, regression trees and random forest techniques are implemented and compared with our proposed models. In this comparison, the proposed HAEANN 6 is used for forecasting hourly global, diffuse and direct solar irradiance up to

6 hours ahead. The same three aforementioned scenarios are used in this study. Before addressing experimental comparison results, a brief overview of those three techniques is first given.

4.5.2.1 Smart persistence

Smart persistence model is a naive model based on the clear sky index. This model, which is an improved version of persistence model, assumes that the sky conditions will remain constant (instead of irradiance itself) [183]. In this model, the forecasting algorithm predicts solar irradiance as the product of the clear sky index (k_t) and global solar irradiance in clear sky conditions in the forecasted point, as follow :

$$global\ solar\ irradiance(t+h) = k_t(t+h) * clear\ sky\ global\ solar\ irradiance(t+h) \quad (4.10)$$

With the assumption that:

$$k_t(t+h) = k_t(t) \quad (4.11)$$

and

$$k_t(t) = \frac{global\ solar\ irradiance(t)}{clear\ sky\ global\ solar\ irradiance(t)} \quad (4.12)$$

The smart persistence model is widely used as the baseline for validating more complex models and demonstrated to be fairly accurate in short term horizons [184].

Numerous models of clear sky solar irradiance are available in the literature and the selection of a given model depends heavily on the availability of the inputs used by each one. In this work, one of the most known clear-sky models is used [185, 186], namely the ESRA (European Solar Radiation Atlas) model [29]. In this model, the scattering effects and absorption by the atmospheric gases are gathered by the Linke turbidity factor. This model contemplates the attenuation of the beam component and the increase in the diffuse solar component in actual sky conditions [187].

In this model, the global solar irradiance for clear sky is split into two parts: the direct normal component B_{cs} and the diffuse component D_{cs} where each component is determined separately.

In accordance with ESRA model, the clear sky direct irradiance on a horizontal surface B_{cs} is calculated as follows:

$$B_{cs} = G_0 * exp(-0.8662 * T_L(AM2) * \delta_R) \quad (4.13)$$

where G_0 is the extraterrestrial irradiance, $T_L(AM2)$ is the Linke turbidity factor at an air mass equal to 2, m is the relative optical air mass and δ_R is the integral Rayleigh optical thickness.

The diffuse horizontal irradiance for clear sky D_{cs} also depends on the Linke turbidity factor T_L at any solar elevation. D_{cs} is determinated by:

$$D_{cs} = G_0 * T_{rd} * F_d \tag{4.14}$$

where T_{rd} represents the diffusion of air molecules as a function of Linke turbidity factor. F_d is the diffuse angular function which accommodates the increased diffusion that occurs at increasing zenith angles (θ_z) [185].

Then, the direct and diffuse irradiances under clear sky conditions can be summed to yield the global clear sky horizontal irradiance.

$$clear\ sky\ global\ solar\ irradiance = B_{cs} * cos(\theta_z) + D_{cs} \tag{4.15}$$

A full detailed description of this model and its coefficients may be found in [29] .

4.5.2.2 Regression trees

Decision Trees (DT) are a popular supervised learning method that can be used for both regression and classification. They are non-parametric hierarchical models where the local region is identified for each data point by a sequence of recursive splits that result in a decision which predicts the target value. A decision tree is illustrated as a growing binary tree composed of a root and nodes. The tree is grown by repeated binary splitting of the data. Each split is defined by a simple rule, usually based on a single explanatory variable and the homogeneity of the resulting two nodes. The terminal nodes represent the groups of data formed by the tree and are also called the leaves of the tree [188]. Regression Trees (RT) are generally used to predict the values of a continuous variable from one or more continuous and/or categorical predictor variables [189]. The final result of the algorithm is usually obtained by averaging the decision nodes (leaves) outputs of all observations. The dataset is generally divided into two distinct subsets, a training set and a test set. The former is used for deriving the regressor, while the latter is used to measure its accuracy.

Only few studies considered regression trees in solar radiation modeling. In [190], Hassan et al. used different machine learning algorithms for estimating daily horizontal global solar radiation at Earth's surface. The same authors adopted multiple tree-based ensemble methods to estimate global, diffuse and beam irradiance components in daily and hourly time-scales [191].

4.5.2.3 Random forest

Random Forests (RF), initially developed by Breiman in 2001 [192], are a bagging based method that operate as an ensemble machine learning technique of a large number of individual decision trees. They are defined by a sequence of rules that split the data in a way that most optimally reduces variation. Each tree receives a random subset of training samples, and then the algorithm randomly selects a subset of variables at each split in the trees. According to Breiman, random forests are an effective tool in prediction which avoid over fitting thanks to the law of large numbers. Furthermore, this model, assembling the individual predictors strength and their correlations ability, gives more accurate prediction results [192].

Same as for decision trees model, random forest technique was not often explored in solar radiation modeling. Huang et al. introduced a daily variability index to quantify the daily variability of solar irradiance using observation data collected by the Bureau of Meteorology at four sites across Australia. Among the implemented models, RF technique demonstrated the ability to provide the best performances compared to statistical ones [193]. Urraca et al. investigated multiple models including the Support Vector Regression (SVR) and RF models to forecast global solar irradiance for 1 hour horizon in a site of southeast Spain. The authors found that machine learning methods outperformed the conventional persistence model [194]. In [191], Hassan et al. explored the potential of four tree-based methods: random forest, decision trees, bagging and gradient boosting for estimating global, diffuse and beam solar irradiance. Then, the authors compared their performances with most used machine learning techniques (Multi-Layer Perceptron and Support Vector Regression) for daily and hourly time-scales. Results indicated that the proposed tree-based models provided accurate estimation. In the work of Ibrahim et al. [195], the firefly algorithm was used to optimize the random forests technique by searching for the best number of trees and leaves per tree in the forest. Obtained results of the proposed model are more accurate when compared with conventional random forests

model, conventional and optimized artificial neural networks.

Most recently, Benali et al. implemented RF, ANN and smart persistence models to forecast hourly global solar irradiance and its components up to six hours ahead. The authors concluded that the RF method gives the best results for global diffuse and beam solar irradiance forecasting [111].

4.5.2.4 Comparison results

In this subsection, global, diffuse and direct solar irradiance forecasting results are presented first as averages of the obtained errors for the 6 hours ahead, followed by a detailed visualization for each error of these hours. Those results exhibits the performances of the proposed HAEANN 6 compared with Smart Persistence (SP), Regression Trees (RT) and Random Forest (RF) models.

For smart persistence model, the necessary parameters (site altitude, zenith angle (θ_z) and linke turbidity factor (T_L)) for the calculation of the clear sky solar irradiance for the 28 cities were provided from the Meteonorm software.

For regression trees and random forest models, several experimental trials were tested to determine the optimal parameters, the best performance are obtained in all the presented results using the parameters as follow. For RT, the main model parameters are:

- min_sample_splits=2 (the minimum number of samples required to split an internal node),
- min_samples_leaf=1 (the minimum number of samples required to be at a leaf node),
- max_features=8 (the number of inputs to consider when looking for the best split).

For RF, the main model parameters are:

- n_estimators =50 (the number of trees in the forest),
- min_sample_splits=2 (the minimum number of samples required to split an internal node).

Table 4.17 presents comparative results between our proposed HAEANN 6 model, SP, RT and RF models for global horizontal solar irradiance forecasting in terms of hours average NRMSE(%) and NMAE(%) for the individual_city scenario. The corresponding

NRMSE error of the four compared models is decomposed for each of the six hours for the 10 considered cities for global solar irradiance forecasting (see Figure 4.12).

Table 4.17: Comparison of the proposed HAEANN 6 model with RT, SP and RF models for global solar irradiance forecasting for 10 cities in terms of NRMSE(%) and NMAE(%).

	Eljadida		Errachidia		Al Hoceima		Essaouira	
	NRMSE	NMAE	NRMSE	NMAE	NRMSE	NMAE	NRMSE	NMAE
Regression trees model	19.72	13.58	27.07	20.49	20.03	13.20	15.49	9.29
Smart persistence	17.15	9.45	18.36	10.66	20.23	10.99	16.60	8.75
Random forest model	15.59	9.61	17.12	12.53	14.30	9.18	14.08	8.07
The proposed HAEANN 6	13.89	9.18	14.75	9.32	12.59	8.27	12.49	7.78
	Fez		Midelt		Marrakech		Dakhla	
	NRMSE	NMAE	NRMSE	NMAE	NRMSE	NMAE	NRMSE	NMAE
Regression trees model	21.89	12.57	26.99	18.74	20.56	13.57	16.98	13.26
Smart persistence	24.05	12.52	20.75	11.57	15.97	8.48	16.24	12.56
Random forest model	11.74	8.18	18.49	12.03	12.74	8.70	10.38	7.90
The proposed HAEANN 6	10.90	7.05	16.89	11.11	11.36	7.37	7.25	5.66
	Oujda		Tanger					
	NRMSE	NMAE	NRMSE	NMAE				
Regression trees model	21.30	14.04	17.50	11.38				
Smart persistence	24.62	13.28	19.73	10.69				
Random forest model	16.40	10.13	12.39	8.01				
The proposed HAEANN 6	13.07	8.83	11.02	6.67				

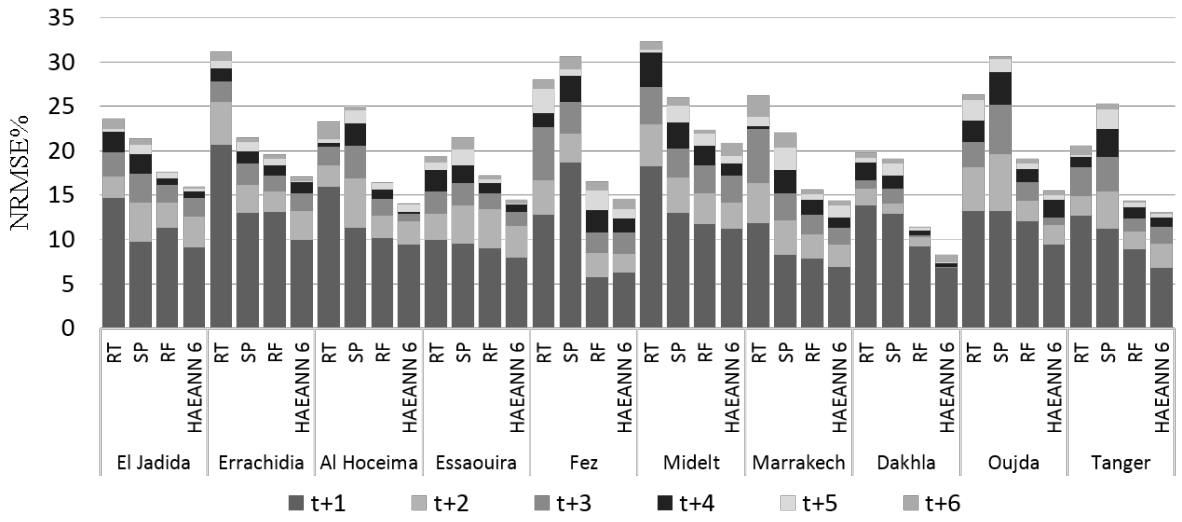


Figure 4.12: Decomposition of NRMSE error of the four compared models for each of the six hours for the 10 considered cities for global solar irradiance forecasting.

Table 4.18: Comparison of the proposed HAEANN 6 model with regression trees, smart persistence and random forest models for diffuse solar irradiance forecasting for 10 cities in terms of NRMSE(%) and NMAE(%).

	Eljadida		Errachidia		Al Hoceima		Essaouira	
	NRMSE	NMAE	NRMSE	NMAE	NRMSE	NMAE	NRMSE	NMAE
Regression trees model	38.27	29.00	39.58	30.24	35.82	27.10	33.05	23.32
Smart persistence model	31.45	22.77	50.09	30.32	41.29	28.85	46.64	31.98
Random forest model	27.82	20.56	28.66	21.23	26.12	18.80	23.57	17.64
The proposed HAEANN 6	26.75	19.95	27.16	18.66	24.31	17.98	21.96	16.40
	Fez		Midelt		Marrakech		Dakhla	
	NRMSE	NMAE	NRMSE	NMAE	NRMSE	NMAE	NRMSE	NMAE
Regression trees model	44.05	34.16	46.54	34.10	42.45	31.44	43.32	31.71
Smart persistence model	46.41	31.87	50.79	32.24	48.45	31.84	49.34	29.68
Random forest model	24.84	18.76	33.67	24.76	29.14	21.84	29.86	23.01
The proposed HAEANN 6	23.53	17.59	33.65	24.64	28.71	21.57	26.62	20.69
	Oujda		Tanger					
	NRMSE	NMAE	NRMSE	NMAE				
Regression trees model	49.36	35.27	47.66	33.73				
Smart persistence model	49.72	32.43	48.45	32.70				
Random forest model	31.89	23.02	31.34	22.10				
The proposed HAEANN 6	30.79	21.88	31.37	22.20				

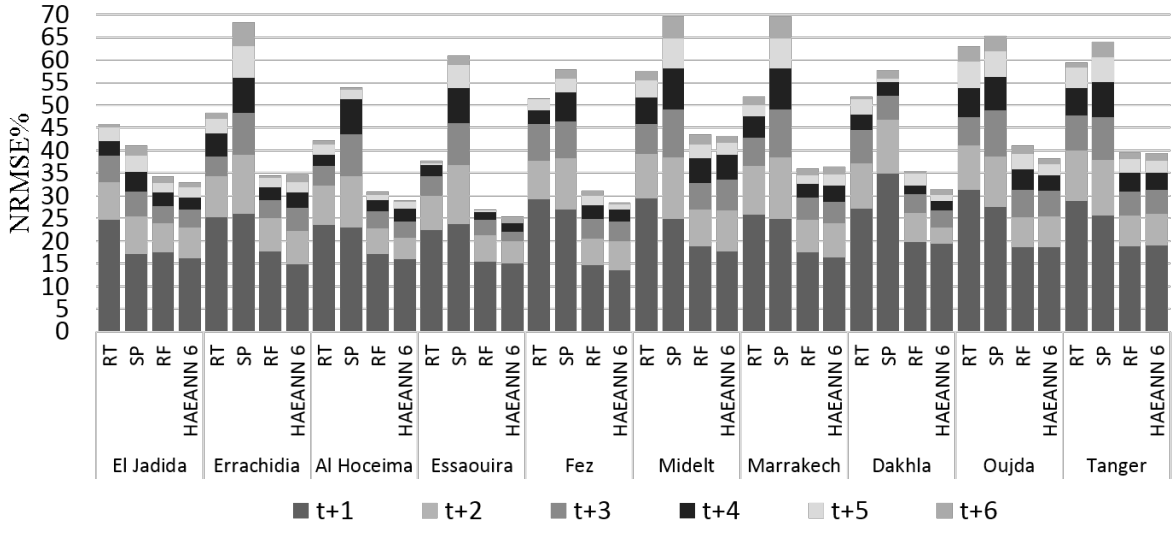


Figure 4.13: Decomposition of NRMSE error of the four compared models for each of the six hours for the 10 considered cities for diffuse solar irradiance forecasting.

Table 4.19: Comparison of the proposed HAEANN 6 model with regression trees, smart persistence and random forest models for direct normal solar irradiance forecasting for 10 cities in terms of NRMSE(%) and NMAE(%).

	Eljadida		Errachidia		Al Hoceima		Essaouira	
	NRMSE	NMAE	NRMSE	NMAE	NRMSE	NMAE	NRMSE	NMAE
Regression trees model	40.74	29.53	30.84	21.19	38.88	27.67	39.16	28.53
Smart persistence model	42.27	31.94	29.57	24.78	38.26	26.39	31.63	22.91
Random forest model	30.03	21.35	24.36	17.00	28.02	19.81	30.85	21.72
The proposed HAEANN 6	29.39	21.14	21.32	14.25	26.13	18.16	29.08	20.54
	Fez		Midelt		Marrakech		Dakhla	
	NRMSE	NMAE	NRMSE	NMAE	NRMSE	NMAE	NRMSE	NMAE
Regression trees model	44.05	34.16	42.71	30.27	36.13	25.86	22.37	16.51
Smart persistence model	43.58	32.80	36.07	30.33	35.40	27.55	18.03	14.04
Random forest model	24.84	18.76	32.74	21.96	25.06	17.61	17.40	13.03
The proposed HAEANN 6	23.87	17.90	31.18	21.67	24.15	17.15	17.9	13.52
	Oujda		Tanger					
	NRMSE	NMAE	NRMSE	NMAE				
Regression trees model	36.93	25.37	31.55	21.78				
Smart persistence model	40.73	31.86	37.02	28.86				
Random forest model	25.66	17.62	23.44	16.12				
The proposed HAEANN 6	24.59	17.16	21.29	14.75				

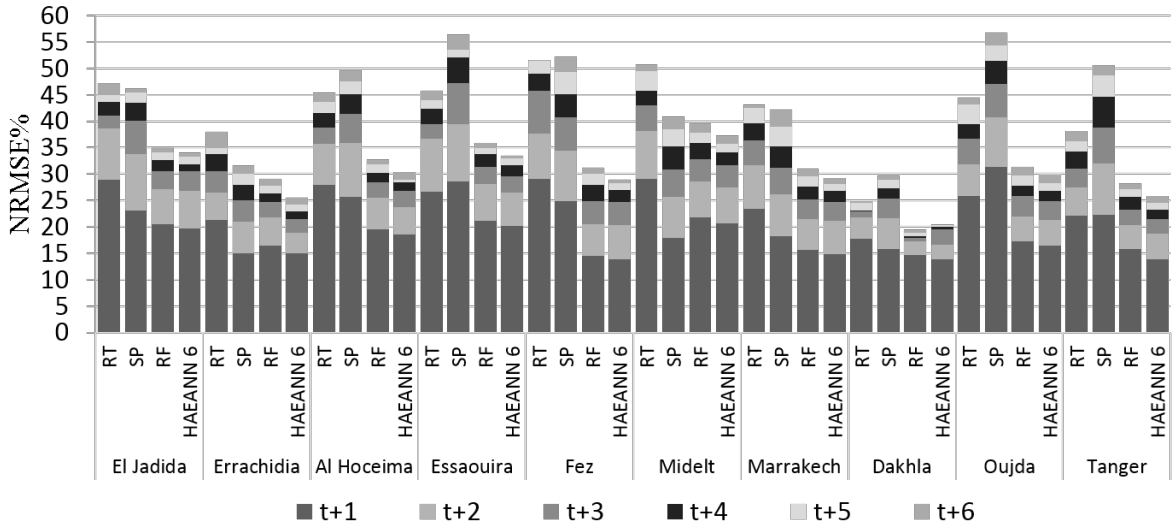


Figure 4.14: Decomposition of NRMSE error of the four compared models for each of the six hours for the 10 considered cities for direct normal solar irradiance forecasting.

From Tables 4.17, 4.18 and 4.19 and Figures 4.12, 4.13, and 4.14 for global, diffuse and direct irradiance forecasting respectively, it is clear that the RF model exhibit better

performances than the RT and SP models, while the proposed model outperforms these models for the ten cities.

Tables 4.20, 4.21 and 4.22 summarize the comparison forecasting results for global, diffuse and direct solar irradiance of the considered zones respectively with the corresponding representation of each hour NRMSE performance (see Figures 4.15, 4.16, and 4.17,).

Table 4.20: Comparison of the proposed HAEANN model with RT, SP and RF models for global horizontal solar irradiance forecasting for six zones in terms of NRMSE(%) and NMAE(%).

	Zone 1		Zone 2		Zone 3	
	NRMSE	NMAE	NRMSE	NMAE	NRMSE	NMAE
Regression trees model	18.91	11.24	20.09	15.02	23.70	14.35
Smart persistence model	16.16	11.92	19.48	14.38	17.86	13.15
Random forest model	13.96	7.93	15.62	9.10	18.33	10.10
The proposed HAEANN 6	12.10	7.45	14.19	8.59	16.88	9.73
	Zone 4		Zone 5		Zone 6	
	NRMSE	NMAE	NRMSE	NMAE	NRMSE	NMAE
Regression trees model	25.50	21.54	20.56	13.57	25.97	16.96
Smart persistence model	33.25	23.93	15.97	8.48	18.54	12.53
Random forest model	17.25	10.75	12.74	7.70	18.28	11.43
The proposed HAEANN 6	15.54	10.43	11.35	7.36	9.57	5.65

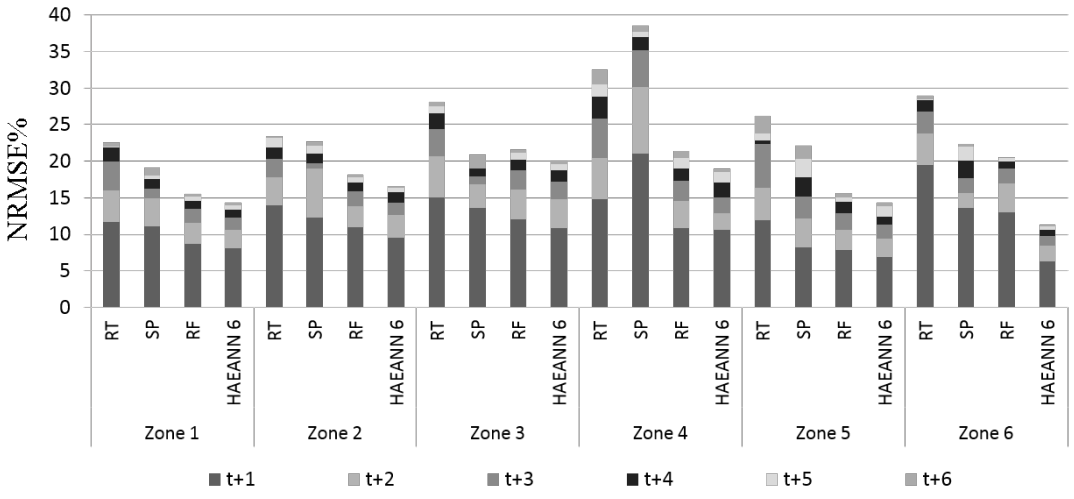


Figure 4.15: Decomposition of NRMSE error of the four compared models for each of the six hours for the six zones for global solar irradiance forecasting.

Table 4.21: Comparison of the proposed HAEANN 6 model with RT, SP and RF models for diffuse solar irradiance forecasting for six zones in terms of NRMSE(%) and NMAE(%).

	Zone 1		Zone 2		Zone 3	
	NRMSE	NMAE	NRMSE	NMAE	NRMSE	NMAE
Regression trees model	38.68	28.08	42.96	30.98	47.13	32.65
Smart persistence model	38.65	30.26	47.65	36.60	48.76	35.16
Random forest model	27.28	20.12	33.39	23.93	35.61	24.71
The proposed HAEANN 6	26.77	19.90	30.83	22.95	33.69	23.81
	Zone 4		Zone 5		Zone 6	
	NRMSE	NMAE	NRMSE	NMAE	NRMSE	NMAE
Regression trees model	55.43	39.45	42.45	31.44	36.45	26.16
Smart persistence model	58.87	39.79	29.14	21.84	44.36	34.33
Random forest model	36.47	26.96	29.14	21.84	25.07	18.01
The proposed HAEANN 6	34.49	25.34	28.71	21.57	32.52	22.33

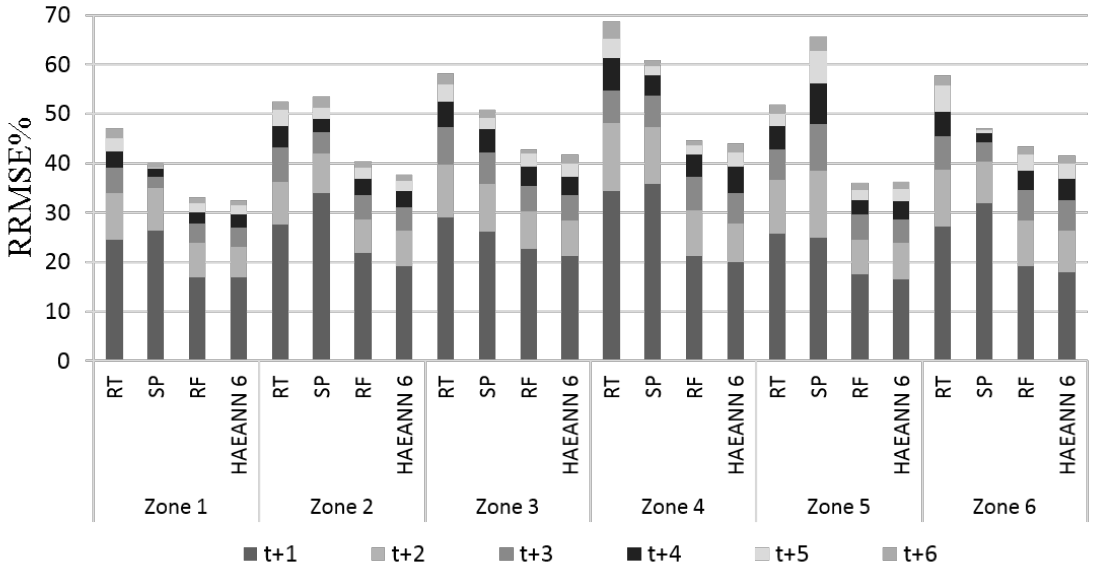


Figure 4.16: Decomposition of NRMSE error of the four compared models for each of the six hours for the six zones for diffuse solar irradiance forecasting.

Table 4.22: Comparison of the proposed HAEANN model with RT, SP and RF models for direct solar irradiance forecasting for six zones in terms of NRMSE(%) and NMAE(%).

	Zone 1		Zone 2		Zone 3	
	NRMSE	NMAE	NRMSE	NMAE	NRMSE	NMAE
Regression trees model	38.68	28.08	42.96	30.98	47.13	32.65
Smart persistence model	38.65	30.26	47.65	36.60	48.76	35.16
Random forest model	27.28	20.12	33.39	23.93	35.61	24.71
The proposed HAEANN 6	26.77	19.90	30.83	22.95	33.69	23.81
	Zone 4		Zone 5		Zone 6	
	NRMSE	NMAE	NRMSE	NMAE	NRMSE	NMAE
Regression trees model	55.43	39.45	42.45	31.44	36.45	26.16
Smart persistence model	58.87	39.79	29.14	21.84	44.36	34.33
Random forest model	36.47	26.96	29.14	21.84	25.07	18.01
The proposed HAEANN 6	34.49	25.34	28.71	21.57	32.52	22.33

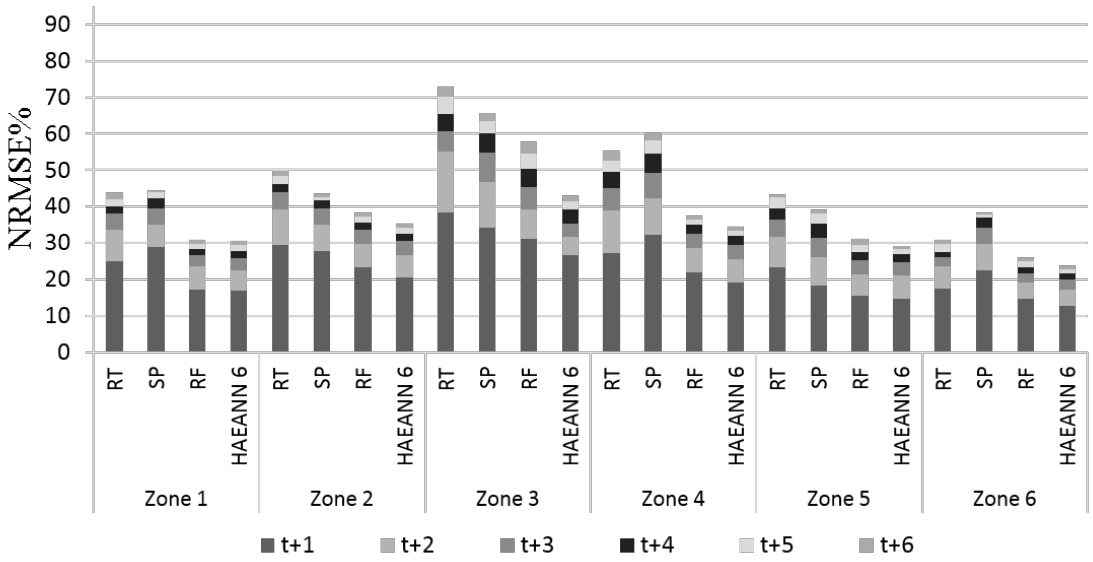
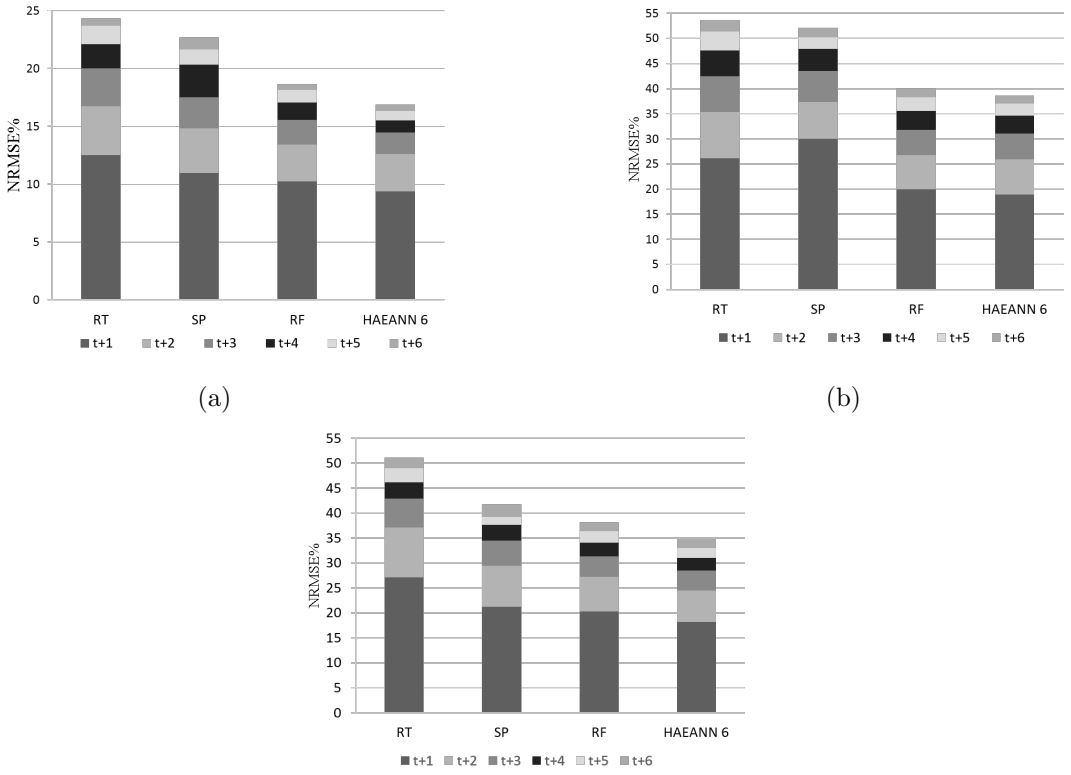


Figure 4.17: Decomposition of NRMSE error of the four compared models for each of the six hours for the six zones for direct normal solar irradiance forecasting.

Table 4.23 and Figure 4.18 summarize the comparison forecasting results for global horizontal solar irradiance and its components for the all_cities scenario with the corresponding representation of each hour NRMSE errors.

Table 4.23: Comparison of the proposed HAEANN model with RT, SP and RF models for global, diffuse and direct solar irradiance forecasting for all_cities scenario in terms of NRMSE(%) and NMAE(%).

	GHI		DHI		DNI	
	NRMSE	NMAE	NRMSE	NMAE	NRMSE	NMAE
Regression trees model	19.88	11.99	42.75	30.07	42.24	29.01
Smart persistence model	17.70	12.80	43.52	30.71	33.97	24.03
Random forest model	15.51	8.94	32.09	22.65	31.25	22.09
The proposed HAEANN 6	14.19	8.56	31.02	22.70	28.30	20.49



(a) (b)

(c)

Figure 4.18: Decomposition of NRMSE error of the four compared models for each of the six hours for all_cities scenario for forecasting (a) global solar irradiance, (b) diffuse solar irradiance, (c) direct solar irradiance.

Considering the results for the zoning scenario presented in Tables 4.20, 4.21 and 4.22 and Figures 4.15, 4.16 and 4.17 for global, diffuse and direct solar irradiance respectively with the forecasting results for the all_cities scenario represented in Table 4.23 and Figure 4.18, it is obvious that the highest performances for forecasting global solar irradiance

and its components were scored by the proposed model, and it is followed by the random forest, smart persistence and regression trees models respectively.

In the three considered scenarios, the proposed HAEANN model outperforms regression trees, smart persistence and random forest models for global, diffuse and direct solar irradiance forecasting.

4.6 Conclusion

Insuring the operationalization of any renewable solar energy system and its integration into the grid require power scheduling and unit commitment during distribution and transmission [196]. Therefore, accurate forecasting of solar resources is a key issue for planning and management of solar energy generation technologies. In this context, a multi-model framework based on evolutionary artificial neural networks (producing a set of models baptized HAEANN models) has been proposed in this chapter to accurately forecast global horizontal solar irradiance up to 6 hours ahead. Through an evolutionary process, an automatic search of the near-optimal forecasting history and ANN architecture parameters is achieved, overcoming the unstructured manual process of searching optimal solutions. Those parameters were represented by a simple and compact chromosome, evolved by an adapted set of genetic algorithm operators including roulette wheel selection strategy, two points crossover and bit-flip mutation operators.

Global, diffuse and direct solar irradiance data from 28 cities covering the geography of Morocco were used for training and testing the proposed models. Considering the same cities in the test dataset, two scenarios were explored: individual_city scenario and all_cities scenario, aiming to assess the validity of the considered zoning scenario using the proposed models.

In the individual_city scenario, the proposed models were tested using data from each city to forecast the upcoming solar irradiance in this same site, obtaining average levels of NRMSE accuracies for forecasting global, diffuse and direct solar irradiance respectively from 7.59%, 15.36% and 16.18% for HAEANN 1 (t+1) to 12.49%, 27.49% and 24.89% for HAEANN 6 (as an average value from t+1 to t+6).

For the zoning scenario, sites belonging to a considered zone were used as training data to forecast solar irradiance of other cities from the same zone. This scenario gives almost similar performances as for the first scenario, with averages of accuracies for NRMSE from

8.38%, 18.41% and 17.15% for HAEANN 1 to 13.27%, 31.17% and 26.34% for HAEANN 6 for forecasting global, diffuse and direct solar irradiance respectively. Especially, this zoning scenario presents an attractive alternative in some geographical areas with no historical solar irradiance data, where future values of solar irradiance can be obtained using data from stations belonging to the same zone.

Then, the all_cities scenario was inspected by building a single train dataset with sites belonging to different zones and a single test dataset with the remaining cities. Resulting average NRMSE performances for forecasting global, diffuse and direct solar irradiance are respectively from 8.44%, 18.10% and 17.94% for HAEANN 1 to 13.76%, 31.02 and 28.30% for HAEANN 6. This all_cities scenario manipulating data with heterogeneous climatic properties in the training phase, presented approximately the same average performances as the zoning and individual_city scenario, showing good accuracy for solar irradiance forecasting and proving the high generalization capabilities of the proposed HAEANN models. However, it needs the availability of a considerable database which is rarely accessible in most countries around the world.

For all scenarios, the six HAEANN models give good hourly performances even with the inclusion of ascendant number of forecasted horizon hours. This property, gives the choice among different models depending on users needs with the same accuracy.

A seasonal behavior was also inspected in this study, aiming to illustrate the effect of seasonality on our models accuracy. In fact, obtained results show the efficiency of the HAEANN models to forecast solar irradiance in different weather conditions with better accuracy during summer season, compared to other seasons (where low Sun elevation with formation and dissolution of clouds are more frequent).

Finally, to validate the efficiency of the proposed HAEANN models, their forecasting performances were compared with a reference model: Smart Persistence (SP) and two tree-based machine learning models: Regression Trees (RT) and Random Forest (RF). In this comparison study, the same three above-mentioned scenarios were used to forecast global solar irradiance and its components up to 6 hours ahead. Obtained results reveal that in all tested scenarios, the proposed HAEANN models outperformed the implemented SP, RT and RF models.

In the next chapter, a multi-step evolutionary framework is presented to forecast short term (from 1 to 6 hours ahead) PV power produced from an optimized PV system.

Chapter 5

Evolutionary framework for short term photovoltaic power forecasting

5.1 Introduction

The electricity produced by renewable energy sources is constantly increasing worldwide thanks to government policies and technological advancements. Challenges of controlling and maintaining energy from inherently intermittent sources and of the grid-connected renewable energy sources penetrating the distribution systems involves many aspects: efficiency, reliability, safety, stability of the grid and ability to forecast energy production. Forecasting PV power production is very important to help the grid operators managing the electric balance between power demand and supply, to improve embedding of distributed renewable energy sources and, in stand-alone systems, to optimally size all its components and to improve the reliability of the isolated systems.

Hence, the ability to accurately predict the generated PV power is a task of utmost importance and relevance for both energy managers and electricity traders, in order to minimize uncertainty and ensure reliable electricity supply at acceptable cost.

There are mainly two approaches for PV power forecasting (see Figure 5.1). The first uses analytical equations to model the PV system where most efforts are devoted to obtain accurate irradiance forecasts, which is the main factor related to the power production. This approach is called "physical" ("PV performance", "parametric" or "white box") method. On the other hand, the second approach directly predicts the PV power output using statistical and machine learning methods. Furthermore, a combination of the two approaches leads to a "hybrid" or "grey box" model.

In the physical method, we can directly obtain PV power forecast with no need of historical data. Even if these models can be very simple, more complex ones require additional parameters [197]. For machine learning methods, historical data about meteorological conditions and PV power production (exogenous inputs) or only PV power production (endogenous inputs) are necessary to train the models. However, they show better forecasting performances than physical models [198].

Since irradiance follows similar patterns to those of power [197], similar forecasting models used earlier in solar irradiance forecasting are adapted in this chapter. For this, a first attempt for solar power forecasting using machine learning techniques is presented. The proposed EANN approach is used to forecast short term (6 hours ahead) PV power production of an optimized photovoltaic system.

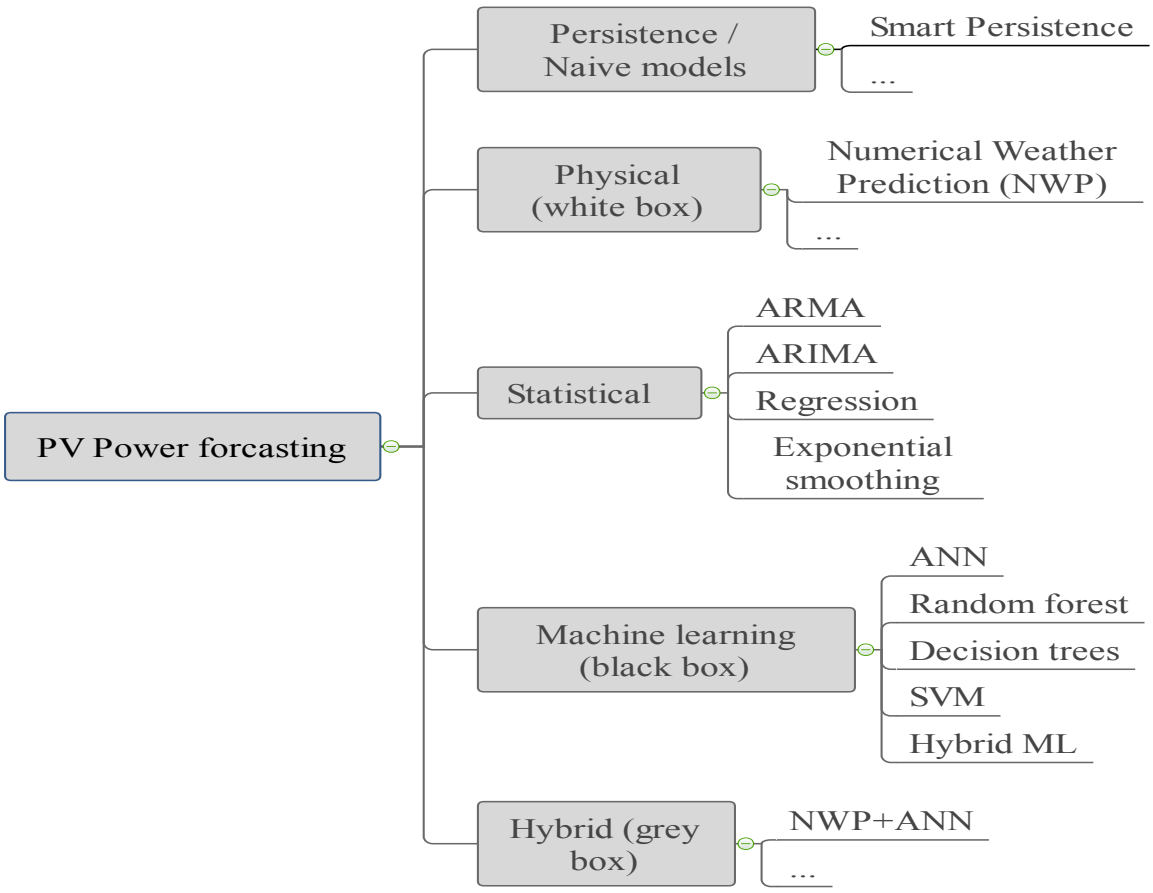


Figure 5.1: Classification of PV power forecasting techniques.

The main contributions of this chapter can be summarized as follows:

1. the basic knowledge towards solar energy production and its different types and systems with a focus on solar Photovoltaics are first presented.
2. the studied PV system is then optimized using the incremental inductance MPPT algorithm to enhance the photovoltaic system efficiency,
3. a multi-step model forecasting, producing multi-output at a time, from 1 to 6 hours ahead for PV power forecast using only endogenous inputs is presented,
4. a comparative study is finally performed with two other machine learning techniques.

5.2 From solar irradiation to renewable energy production

The term solar energy is generally used to refer to methods of collecting light and turning it directly into a useful form of energy. Depending on the used technology, it can do

so in a variety of ways in order to provide heat, light, hot water, electricity, cooling to houses, buildings or even industrial complexes. The most common types of solar energy are Thermal solar energy and solar Photovoltaics.

Thermal solar energy applications are usually concerned with harvesting thermal energy through the use of solar collectors that employ active mechanical components (such as pumps) to collect and transport heat. Solar energy is converted into thermal energy using a special kind of heat exchanger known as a solar collector. These solar collectors can be classified into two general categories: (i) non-concentrating and (ii) concentrating. In the non-concentrating type, the collector area (the area that intercepts the solar irradiation) is the same as the absorber area (the area that absorbs the radiation). Flat plate collectors (FPC) and evacuated tube collectors (ETC) are non-concentrating type collectors. These collectors are mainly designed for solar hot water and industrial process heat applications, which require energy delivery at temperatures in the range of 60-250°C. These collectors use both diffuse and direct solar radiation and do not require tracking of the Sun. They are mechanically simpler than concentrating collectors and require less maintenance [199].

In the concentrating type solar collector, various types of mirrors, reflectors or concentrators are used to concentrate the solar energy and they provide higher temperatures (i.e., 400-1000°C) than non-concentrating type collectors. Compound parabolic concentrator (CPC), central receiver or solar tower, parabolic trough collector and parabolic dish collectors are concentrating type collectors and are known as concentrated solar power (CSP) systems. The concentrating type collectors are classified into three types: (i) non-imaging (ii) line focusing and (iii) point focusing collectors. Non-imaging collectors have a low concentration ratio; line-focusing collectors have intermediate concentration ratio and point focusing collectors have high concentration ratio. The CPC collector is a non-imaging concentrating collector. Whereas, parabolic trough and central receiver are line focusing and point focusing collectors, respectively. These CSP systems are mainly used for solar thermal power generation.

In the previous paragraphs, a brief initiation to thermal solar energy was presented and more details could be found in [28]. In the next subsection, our focus will be specifically on solar photovoltaic basics, which will be useful in this Chapter.

5.2.1 Solar Photovoltaics

In this subsection, the basics of photovoltaic power generation are presented, starting from the fundamentals of the photoelectric effect, the photovoltaic conversion and behavior of the solar cell to different types of photovoltaic systems.

5.2.1.1 Photoelectric effect

The photoelectric effect was first discovered by German physicist Heinrich Rudolf Hertz (1887). He experimentally realized that a metallic material exposed to light could emit electrons (see Figure 5.2). This discovery was called the "photoelectric effect". Unfortunately, Hertz was not able to theoretically explain the phenomenon, because some characteristics of this effect could not be formalized with the classical theory of electromagnetism of the time.

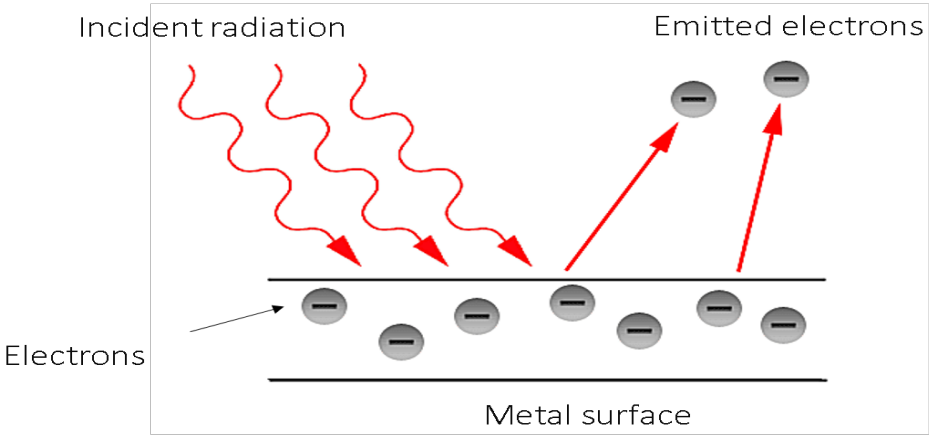


Figure 5.2: Description of the photoelectric effect.

Meanwhile, work on the physics of photoelectric phenomena has expanded. In the 1920s and 1930s, quantum mechanics laid the theoretical foundation for our present understanding of photovoltaic (PV). By applying the hypothesis of the existence of the photon, Albert Einstein was able to explain the quantum nature of the photoelectric effect.

A major step forward in solar-cell technology came in the 1940s and early 1950s when a method (called the Czochralski method) was developed for producing highly pure crystalline silicon. In 1954, work at Bell Telephone Laboratories resulted in a silicon photovoltaic cell with a 4% efficiency. Bell Labs soon bettered this to a 6% and then 11%

efficiency, heralding an entirely new era of power-producing cells [32].

5.2.1.2 Photovoltaic conversion

The photovoltaic conversion is based on the photoelectric effect, that is, on the conversion of the light energy coming from the Sun into electrical energy. To carry out this conversion, devices called solar cells are used, constituted by semiconductor materials in which a constant electric field has been created artificially (by means of a p-n junction). A p-n junction provides an electrical field that will cause the material and cause the vacancies (holes) left by the excited electrons to move in the opposite direction. If the electrons and holes reach the respective edges of the semiconductor material, the device is capable of delivering electrical power to an external circuit (see Figure 5.3). The motion of electrons or holes receives competition from recombination processes (electrons being recaptured into vacancies), giving importance to such factors as overall dimensions and electron mobility in the used material.

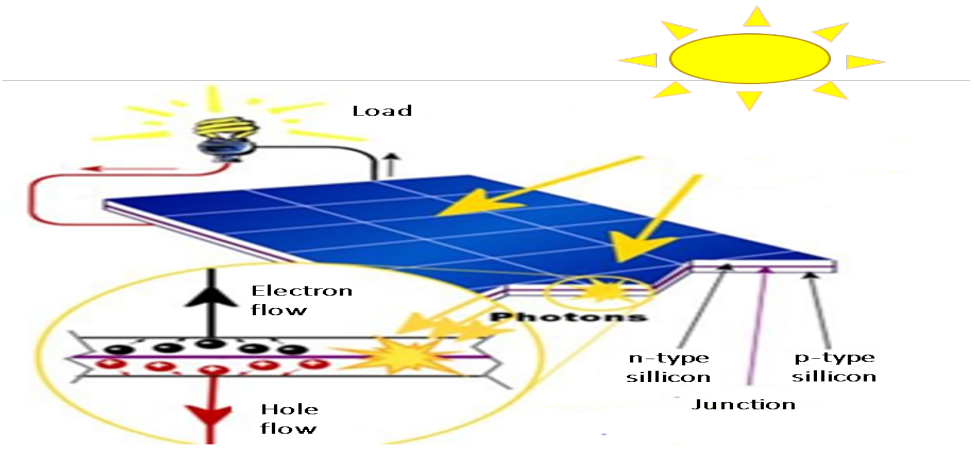


Figure 5.3: Description of the photovoltaic effect on a solar cell.

As already mentioned, most photovoltaic cells use semiconductors to harvest the electron-hole pairs created by the collision of photons in the material. Each material has its own energy gap (forbidden energy band between the valence band and the conduction band [32]). If the material gap is large, few photons will have enough energy to create current, but across the cell, the open-circuit voltage will be large and make it even easier to harness electrical energy. Conversely, a material with a small gap absorbs more photons but has a lower voltage at its terminals. Until today, silicon is the most widely used material for building cells in PV systems and the theoretical maximum efficiency for

a single silicon junction is around 29%.

The familiarity of current and voltage relationship of photovoltaic modules under real operating conditions is essential for the determination of their power output. Normally, the cells are mounted in modules, and multiple modules are used in arrays to get desired power output. Individual modules may have cells connected in series and parallel combinations to obtain the required current and voltage. Similarly, the array of modules may be arranged in series and parallel connections. When the cells or modules are connected in series, the voltage is additive, and when they are attached in parallel, the currents are additive [34, 200].

When the output voltage V=0, the current is the short circuit current (I_{sc}) and when the current I=0, the output voltage is the open circuit voltage (V_{oc}). These values can change depending on the used material, the temperature and irradiation. Figure 5.4 shows the typical characteristics of a PV module under standard test conditions (an irradiation of 1000 W/m^2 and a cell temperature of 25°C).

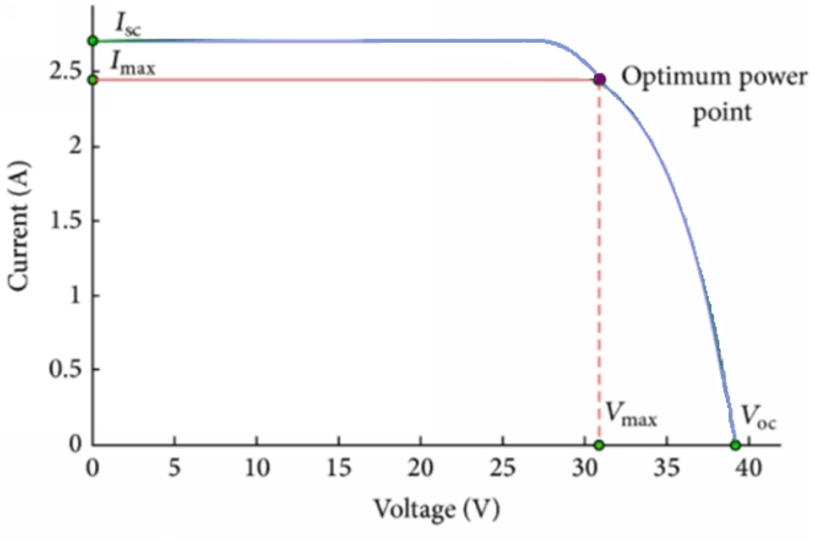


Figure 5.4: A typical I-V curve for a silicon photovoltaic module.

The power output of PV modules could be predicted from the behavior of current-voltage, I-V characteristic curve. Many models are available in the literature to modelize the I-V curves of PV modules. We can cite in particular the single diode model [201] and the double diode model [202]. To determine the average electrical power generated by a module (P) during a time Δt, we will tend to favor the linear relationship between the optimal theoretical peak power (PC) of the module (for an irradiation of 1000W/m^2

under standard test conditions (STC)) and the electrical power produced. In practice, the concept of PV efficiency (electric power/radiative power) is added to establish the equation for the generated electrical power which is formulated as:

$$P = \frac{GI}{\Delta t} \times S \times \eta_{PV} \tag{5.1}$$

where :
GI is the global solar irradiation,
η_{PV} is PV efficiency,
S is the exposed surface of the modules.

The PV efficiency may also vary depending on the type of the photovoltaic solar cells (amorphous, polycrystalline or monocrystalline sillicon).

It is important to note that the produced PV power increases with the increase of solar irradiation and that it decreases when the temperature increases.

5.2.1.3 Photovoltaic systems

Photovoltaic refers to the direct generation of electricity by solar irradiation. No matter what kind of photovoltaic systems are examined, the most fundamental element is the individual photovoltaic cell. Individual cells are assimilated into modules, modules are assembled into arrays, which are integrated into systems with a wide range of components. Functional uses in an increasing hierarchy of complexity of photovoltaic systems are as follows: 1) direct coupled, 2) systems with battery storage, 3) hybrid power systems, 4) systems connected to the electrical grid and 5) systems for utility power production.

The first three systems in this list are generally stand-alone systems.

A direct coupled photovoltaic system supplies DC power directly to the load or AC power, via an inverter, directly to the load. Such systems are frequently used for specific, usually low-kilowatt requirements–applications for which grid electricity is not available. With no grid connection, backup generation, or storage capability, stand-alone systems are completely dependent on solar irradiation for power generation and operation. Nonetheless, these are important “niche” applications for photovoltaics. One of the disadvantages of direct-coupled photovoltaic systems is their inability to operate except when exposed to solar irradiation. A photovoltaic system with battery storage eliminates this disadvantage and permits operation during the night hours or when the solar irradiation is insufficient to meet the required electrical demand. The charge controller, a key component, directs

the output of the photovoltaic array to the load and to battery recharging when electricity is being generated, when the array is not generating, it extracts electrical energy from the battery to the load.

A photovoltaic hybrid system occurs if a stand-alone system with battery storage is provided with a non grid source of electricity. The system schematic for a hybrid system is similar to a stand-alone system with battery storage with the addition of a non-grid source of electricity and a rectifier. The rectifier is used to convert AC from an AC generator to DC for battery charging. When photovoltaic output and/or the battery are insufficient to meet the loads, the "other" source of electricity is then used to make up the difference. Such a system provides great flexibility, but for grid-connected systems the utility rate structure can be economically constraining [203].

Photovoltaic systems can be used to supply electricity to the grid. In any photovoltaic system connected to the grid, the power from the solar arrays must be compatible with the voltage, frequency, and phase of the grid. IEEE Standard 1547 (Interconnecting Distributed Resources with Electric Power Systems) delineates the requirements for grid connectivity of photovoltaic systems.

One issue with grid-connected systems is the price per kilowatt hour paid by the utility to the system owner. Photovoltaic systems with high installed kilowatts may be owned by utilities, but low-kilowatt residential and commercial systems are likely to be privately owned. Since photovoltaic generating is considered as green power, many utilities are willing to participate in a net metering arrangement. A formal agreement with the utility is needed by any owner contemplating selling power to the utility.

5.3 System description

In this work, an existing PV installation at the National School of Applied Sciences of Fez is used. The photovoltaic panel of the studied installation is composed of two independent modules, each module includes 36 (4x9) monocrystalline silicon cells and a Single Phase Bridge Inverter. Table 5.1 shows the characteristics of the used PV panel at Standard Test Conditions (STC).

While waiting for the implementation and recording of measured data, the PV system is simulated using MATLAB-SIMULINK in order to generate the produced PV energy output. To enhance the photovoltaic system efficiency, it is important to improve the

Table 5.1: Characteristics of the used PV module at STC.

Characteristic	Value
Maximum power, P_{max}	135.05 W
Voltage at P_{max},V_{mp}	17.7 V
Current at P_{max}, I_{mp}	7.63 A
Short-circuit current, I_{sc}	8.18 A
Open-circuit voltage, V_{co}	22.4 V
Temperature coefficient of open-circuit voltage V_{oc}, K_v	-38 $mV/°C$
Temperature coefficient of short-circuit current I_{sc}, K_i	6.5 $mA/°C$
The number of cells per module	36

overall performance at the system level. One area of immense interest is the maximum power point tracker (MPPT). MPPT aims to ensure that at any environmental condition (particularly solar irradiation and temperature), the maximum power is extracted from the PV panels by matching its I-V operating point with the load. The maximization of the PV power always remains a significant challenge. Researchers have proposed different MPPT algorithms to maximize PV power, namely Fractional Short-Circuit Current, Fractional Open-Circuit Voltage, Fuzzy Logic, Artificial Neural Network, Perturb and Observe, and incremental inductance [204, 205]. Fractional Short-Circuit Current and Fractional Open-Circuit Voltage are the simplest MPPT algorithms, which are based on the linearity of short-circuit current or open-circuit voltage to the maximum power point current or voltage. However, these techniques isolate the PV panel to measure short-circuit current or open-circuit voltage. Therefore, the loss of energy is increased due to the periodic isolation of the panel [206]. One of the most used techniques in the literature is the Incremental Inductance algorithm.

The Incremental Inductance method is based on the fact that the slope of the power curve of the panel is zero at the MPP, positive to the left and negative to the right [207, 208, 209]. Since:

$$\frac{dP}{dV} = \frac{d(I \times V)}{dV} = V + \frac{dI}{dV} = V + \frac{\Delta I}{\Delta V} \tag{5.2}$$

Where $P = I \times V$ and:

$$\begin{cases} \frac{\Delta I}{\Delta V} = -\frac{I}{V} & at\ the\ MPP \\ \frac{\Delta I}{\Delta V} > -\frac{I}{V} & left\ of\ the\ MPP \\ \frac{\Delta I}{\Delta V} < -\frac{I}{V} & right\ of\ the\ MPP \end{cases} \tag{5.3}$$

The MPP can be tracked by comparing the instantaneous conductance to the incremental conductance, as shown in Figure 5.5.

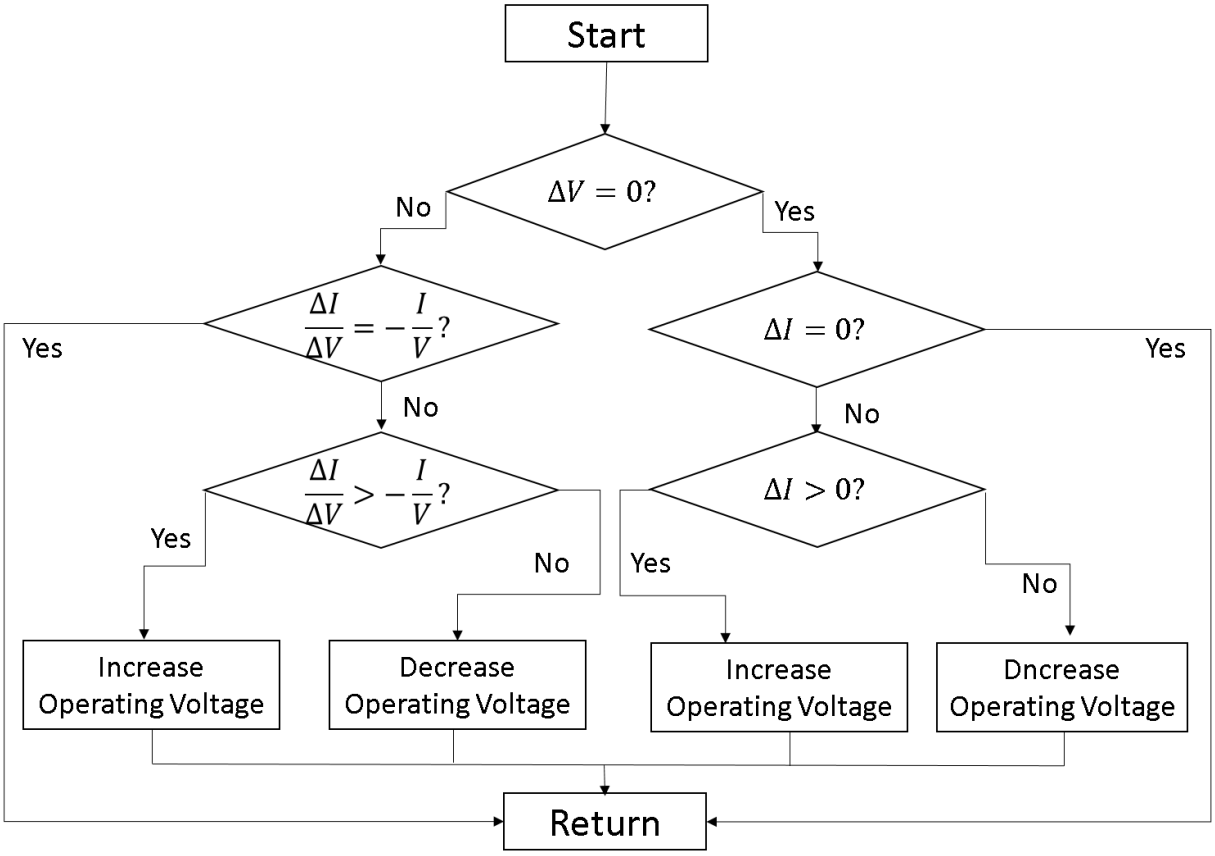


Figure 5.5: Flowchart of the used Incremental Inductance algorithm.

Figure 5.6 presents the block diagram of our PV system with Power Converter and MPPT Control. Figure 5.7 shows the detailed MATLAB-SIMULINK scheme of our photovoltaic system with MPPT control based on the incremental conductance method. A Boost converter controlled by the MPPT algorithm is inserted between the PV panel and the load. The inputs of MPPT controller are voltage and current of the PV module, while the output is PWM (pulse width modulation) for controlling the duty cycle (α) of the

BOOST converter.

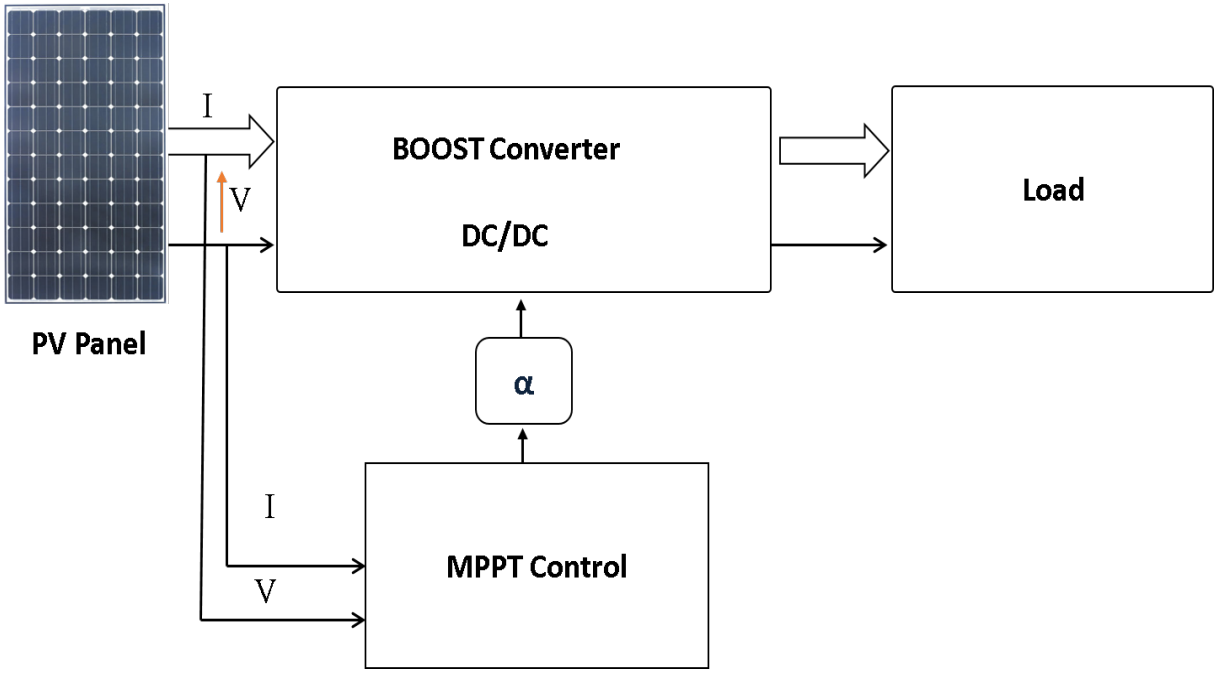


Figure 5.6: Block diagram of our PV system with Power Converter and MPPT Control.

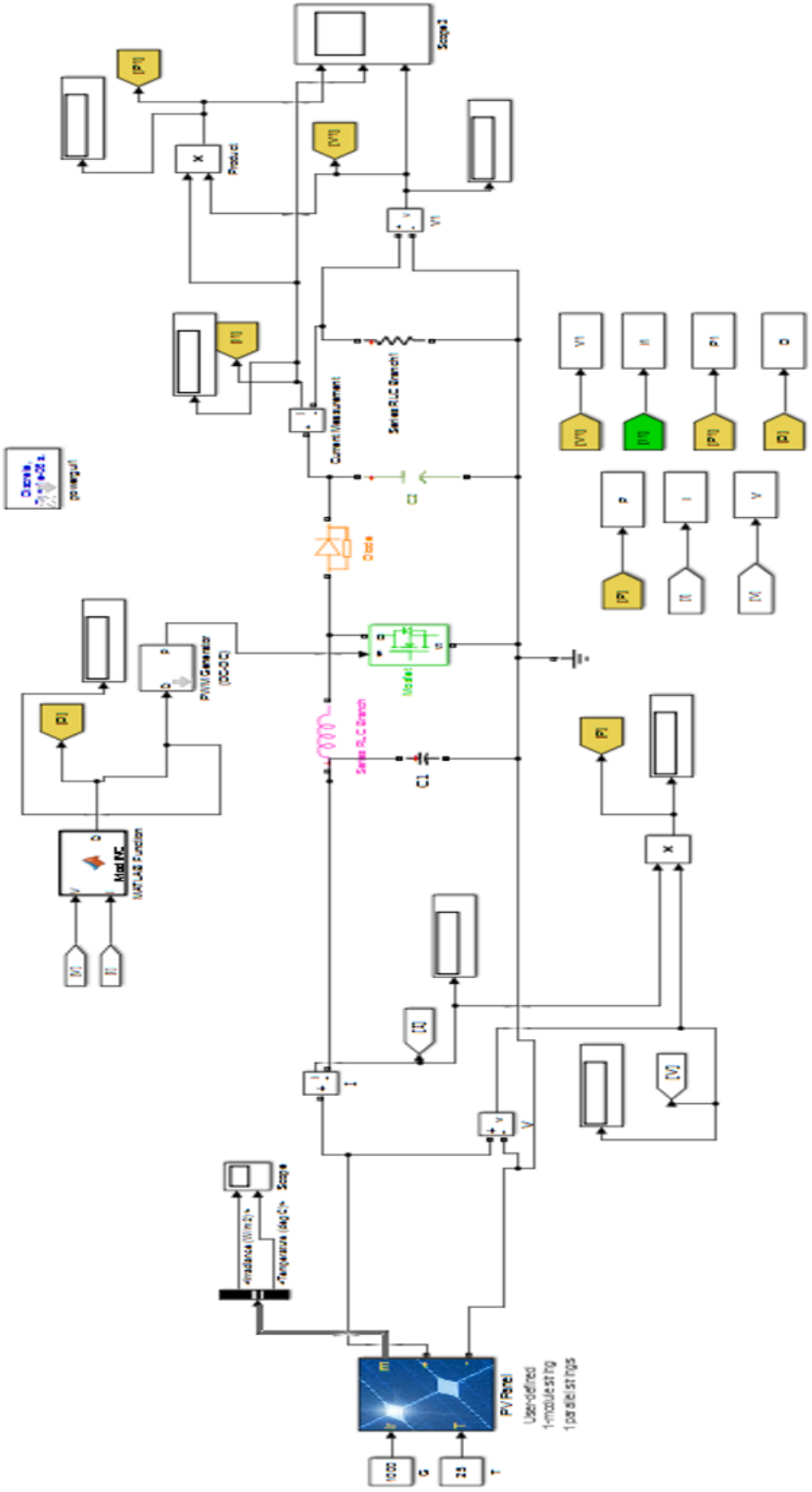


Figure 5.7: The simulation of our PV system with MPPT control based Incremental Conductance technique.

5.4 The proposed model for PV power production forecasting

In this section, detailed description of the proposed forecasting model is presented. Our model is a multi-step model forecasting, producing multi-output at a time, from 1 to 6 hours ahead, contrarily to other one step forecasting which sums up forecasted values to generate H power values ahead [210]. Figure 5.8 shows the block diagram of the proposed PV power forecasting model where the PV power is obtained as an output of the optimized PV system previously described in section 5.3 using hourly global solar irradiance and temperature values as system inputs. Then, historical PV power values (from t to $t-2^{(l_0)}$) are recorded and fed to our proposed forecasting model where l_0 is the length of the forecasting history encoded in the first part of the chromosome (see Figure 4.5). The proposed forecasting model is generated from the aforementioned evolutionary framework, which is an EANN that principally resolves the problematic of the choice of ANN parameters when using ANNs as PV power forecasting models. The framework generates H models obtained from H adaptation cycles with an evolution of forecasting history and ANN architecture. We start with an initial population containing randomly generated ANN power forecaster individuals, our framework evolves simultaneously the forecasting history and ANN architecture through proposed genetic operators ending with a final population. This evolutionary framework is constructed based on Algorithm 3 (Section 4.3.1). In order to forecast one or multiple future produced PV power values noted $\hat{P}(t+i)/i = 1, \cdots, H$, we use a historical data (forecasting history) $P(t-j)/j = 0, \cdots, 2^{l_0}$. Through evolution, our model test different configurations of historical data to determine which one produces the best forecasting performances.

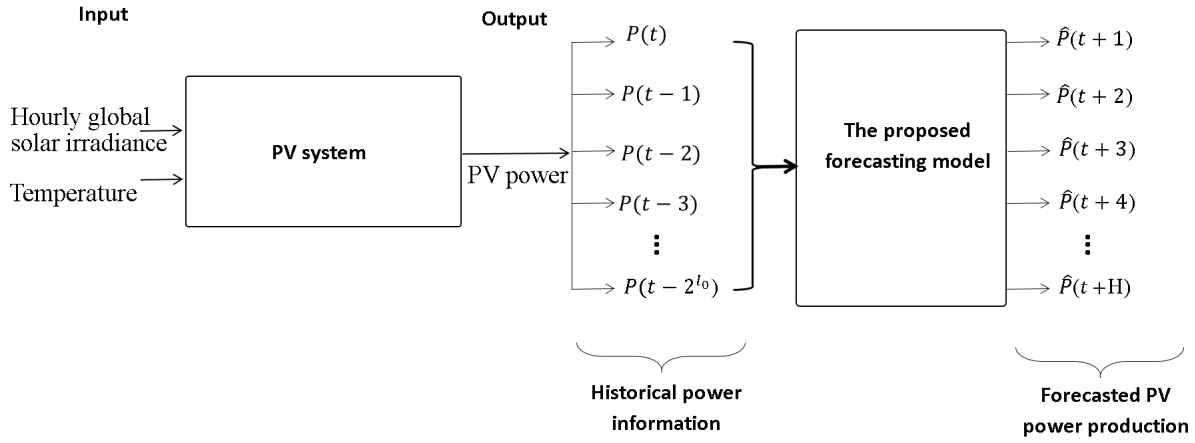


Figure 5.8: Block diagram of the proposed PV power forecasting model.

The procedure used in our model adopt the immediate history values as explained in Figure 5.9. Conventional methodology forecast future values using past ones from previous days for the same forecasted hour (see Figure 5.10). Unfortunately, this methodology requires multiple previous days of data to forecast one hour, which needs important dataset sizes.

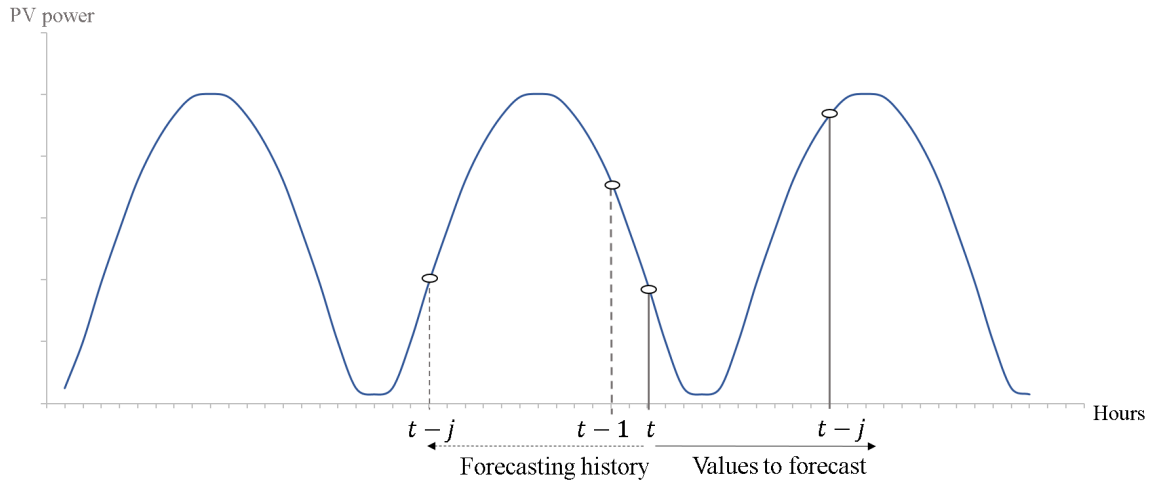


Figure 5.9: The used forecasting procedure using the immediate history values.

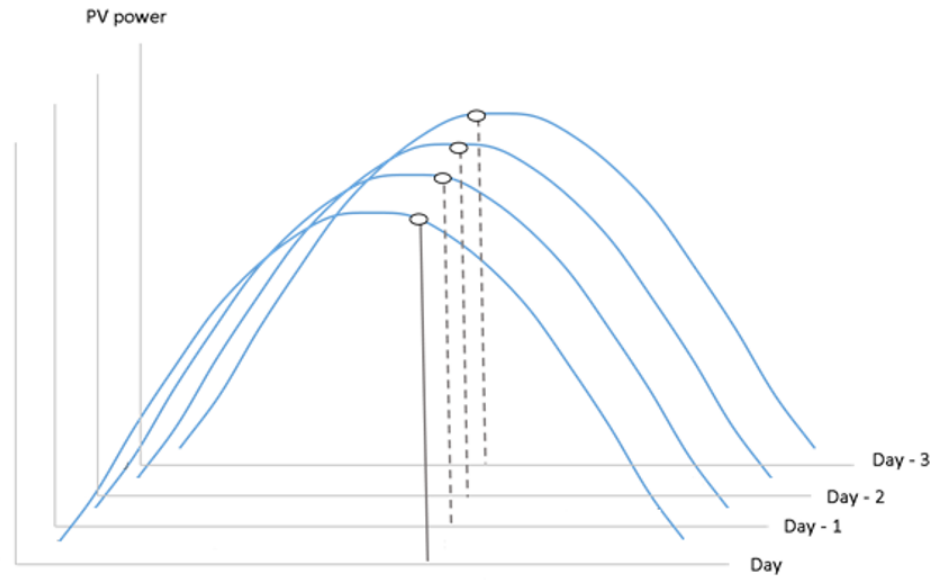


Figure 5.10: Forecasting using past values from previous days for the same forecasted hour.

5.5 Experimental results

One year of hourly solar irradiance and temperature data have been acquired from Meteonorm software, from January 1st to December 31th related to the city of Fez. These data were used as input parameters to our PV system. The output of this simulated PV system is then recorded and divided into train and test datasets which will be used by our proposed forecasting model. Note that, only endogenous inputs are used in this forecasting process. 70% of the collected data are used to train the Evolutionary Artificial Neural Networks and 30% to test them.

Figure 5.11 shows real versus forecasted values of PV power production for three days (13, 14 and 15 of July) using our proposed model.

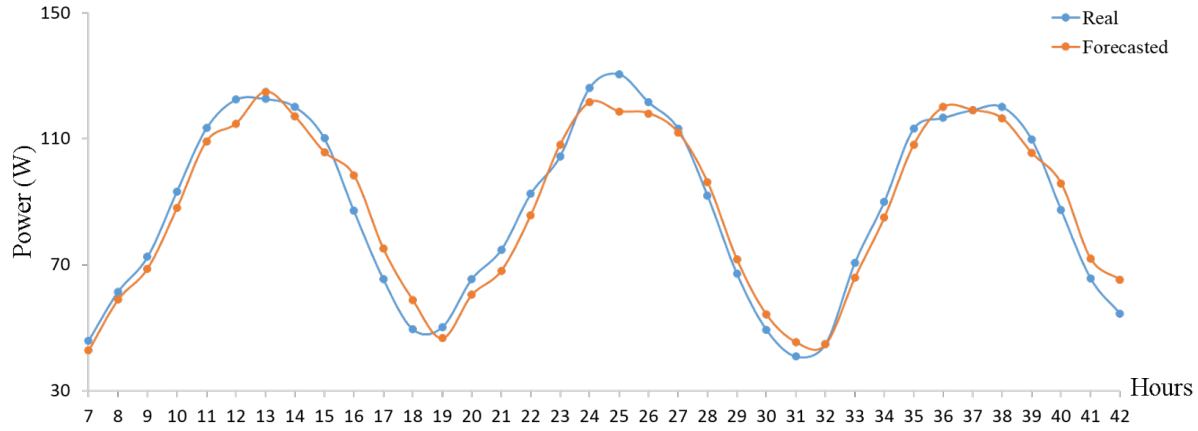


Figure 5.11: Real versus Forecasted values of PV power produced on 13, 14 and 15 of July.

We can see from Figure 5.11 the good accuracy of our proposed model.

Table 5.2 presents comparative results between our proposed model (producing 6 hours ahead: HEANN 6), RT and RF models for PV power production forecasting in terms of NRMSE(%) and NMAE(%). The corresponding NRMSE error of the three compared models is dressed for each of the six hours.

For regression trees and random forest models, several experimental trials were tested to determine the optimal parameters, the best performance are obtained in the presented results using the parameters as follow. For RT, the main model parameters are:

- min_sample_splits=3 (the minimum number of samples required to split an internal node),
- min_samples_leaf=1 (the minimum number of samples required to be at a leaf

node),

- max_features=8 (the number of inputs to consider when looking for the best split).

For RF, the main model parameters are:

- n_estimators =100 (the number of trees in the forest),
- min_sample_splits=3 (the minimum number of samples required to split an internal node).

Table 5.2: Comparison of the proposed HAEANN 6 model with regression trees and random forest models for short term PV power forecasting in terms of NRMSE(%) and NMAE(%).

Forecasting model	Performance indicators	1h	2h	3h	4h	5h	6h
Regression trees model	NRMSE(%)	23.21	27.59	47.95	32.11	32.92	64.44
	NMAE(%)	15.59	18.84	22.88	22.04	22.65	26.99
Random forest model	NRMSE(%)	16.59	19.61	22.74	21.36	22.62	25.95
	NMAE(%)	10.89	13.25	15.12	14.29	15.15	16.54
Our proposed model	NRMSE(%)	15.31	19.65	17.51	18.94	20.19	20.49
	NMAE(%)	9.99	11.78	12.92	13.61	13.83	14.05

From Table 5.2, we can clearly see that the proposed PV power production forecasting model for 6 hours ahead presents better performances than the regression trees and random forest models with a NRMSE from 15.31% for 1 hour ahead to 20.49% for 6 hours ahead using 8 previous values as forecasting history and one hidden layer with 12 hidden nodes.

5.6 Conclusion

In the first part of this chapter, we presented the basic knowledge towards solar energy production and its different types and systems with a focus on solar photovoltaics. In the second part a multi-step short term approach for forecasting the power energy generated by an optimized solar PV system for 6 hours ahead is proposed. In order to use enhanced photovoltaic system efficiency, our PV system is first optimized using the incremental inductance MPPT algorithm. The outputs of this simulated PV system is then collected using one year of hourly data of solar irradiation and temperature for the city of Fez,

Morocco. Next, the recorded PV power values are fed to our evolutionary neural network based model: HAEANN. This proposed model is trained to forecast 6 hours ahead with only endogenous inputs. Our model presents good accuracy with a NRMSE from 9.99% for 1 hour ahead to 14.05% for 6 hours ahead. Finally, through a comparison study with regression trees and random forest models, our model outperforms the two other implemented machine-learning techniques.

Conclusion and perspectives

Given the huge potential for renewable energy (solar and wind) available in Morocco, the building of a diversified energy mix benefiting renewable energy has been planned in order to satisfy growing electricity demand, preserve the environment and reduce dependence on energy from other countries. To achieve this transition, the Moroccan project planned for 2030 is marching steadily towards fulfilling its renewable energy goal of generating 52% of its electricity from renewable sources.

The present thesis aims to consolidate the vision of the Moroccan project for the transition to a cleaner energy production. As presented in the introduction section, the biggest challenge in the penetration of renewable energy, whether in the case of isolated sites or into the grid, is the high variability of the output power of photovoltaic panels and solar thermal power systems, which is strongly affected, by the variability of the solar resource.

This study is therefore part of the search for accurate predictors of solar and PV resource in Morocco, in order to quantify the availability and variability of solar energy to facilitate its management and its integration. Our goal is to develop accurate predictors based on machine learning techniques, especially optimized Artificial Neural Networks (ANN).

Our first contribution begins with a systematic review of solar energy estimation and forecasting using artificial neural networks. In fact, ANN models are found to predict solar radiation accurately, due to their high potential to simulate the complex, nonlinear and time-varying input-output systems compared to other existing models. Therefore, different research papers based on ANN models used to predict solar radiation were discussed. Our study provides an updated review to help future investigations in this field. Furthermore, the objective behind analyzing these publications was to highlight important solar prediction using ANNs problematics. The main ones are the lack of a standard database

and the absence of a systematic method to optimize the ANN architecture. Moreover, we have pointed out the problematic of the selection of most relevant input variables for ANN models.

Our second contribution proposes an Evolutionary Artificial Neural Network (EANN) model, based on genetic algorithms in the field of solar irradiation estimation. Our model allows an automatic selection of the best inputs for the ANN and recorded excellent performances in term of estimation using commonly measured meteorological parameters, validated for daily solar irradiation in the city of Fez, Morocco. Those performances are obtained with reduced population size and maximum generation yielding low computational burden. Our selection method was compared with two models of selection: heuristic search and forward stepwise regression, then, validated against three empirical temperature based models and kNN model, showing good performance in term of estimation. Our proposed model gives a performance of 97.50% for R^2 and 17.85% for NRMSE with rainfall, wind direction, daily temperature gradient and global solar irradiation at the top of the atmosphere as the best combination of input parameters.

In the next contribution, a multi-model framework based on evolutionary artificial neural networks, producing a set of models baptized HAEANN models, has been proposed to accurately forecast global, diffuse and direct solar irradiance up to 6 hours ahead. Through an evolutionary process, an automatic search of the near-optimal forecasting history and ANN architecture parameters is achieved, overcoming the unstructured manual process of searching optimal solutions. Those parameters were represented by a simple and compact chromosome, evolved by an adapted set of genetic algorithm operators.

For training and testing the proposed models, global, diffuse and direct solar irradiance data from 28 cities covering the geography of Morocco were used. Two scenarios were explored: "individual_city scenario" and "all_cities scenario", aiming to assess the validity of a considered "zoning scenario" using the proposed models.

In the zoning scenario, where sites belonging to a considered zone were used as training data to forecast solar irradiance of other cities from the same zone, we obtained averages of accuracies for NRMSE from 8.38%, 18.41% and 17.15% for HAEANN 1 to 13.27%, 31.17% and 26.34% for HAEANN 6 for forecasting global, diffuse and direct solar irradiance respectively. This zoning scenario presents an attractive alternative in some geographical areas with no historical solar irradiance data, where future values of solar irradiance can be obtained using data from stations belonging to the same zone.

Then, the all_cities scenario was inspected by building a single train dataset with sites belonging to different zones and a single test dataset with the remaining cities. Resulting average NRMSE performances for forecasting global, diffuse and direct solar irradiance are respectively from 8.44%, 18.10% and 17.94% for HAEANN 1 to 13.76%, 31.02 and 28.30% for HAEANN 6. This all_cities scenario manipulating data with heterogeneous climatic properties in the training phase, presented approximately the same average performances as the zoning and individual_city scenario, showing good accuracy for solar irradiance forecasting and proving the high generalization capabilities of the proposed HAEANN models. However, it needs the availability of a considerable database which is rarely accessible in most countries around the world.

A seasonal behavior was also inspected in this study, aiming to illustrate the effect of seasonality on our models accuracy. In fact, obtained results show the efficiency of the HAEANN models to forecast solar irradiance in different weather conditions with better accuracy during summer season, compared to other seasons (where low Sun elevation with formation and dissolution of clouds are more frequent).

Finally, to validate the efficiency of the proposed HAEANN models, their forecasting performances were compared with a reference model: Smart Persistence (SP) and two tree-based machine learning models: Regression Trees (RT) and Random Forest (RF). In this comparison study, the same above-mentioned scenarios were used to forecast global solar irradiance and its components up to 6 hours ahead. Obtained results reveal that in all tested scenarios, the proposed HAEANN models outperformed the implemented SP, RT and RF models.

Our last contribution concerns forecasting of PV power production using evolutionary artificial neural networks based model. The adopted methodology was described as follows:

1. the PV system is first optimized using the incremental inductance MPPT algorithm to enhance the photovoltaic system efficiency,
2. a multi-step model forecasting, producing multi-output at a time, from 1 to 6 hours ahead for PV power forecast using only endogenous inputs is presented,
3. a comparative study is finally performed with two other machine learning techniques.

Results show encouraging performances and open new outlooks in PV power forecasting.

Our main perspectives for future works are to attain more forecasting accuracy of the proposed HAEANN models using exogenous inputs selected by our first automatic selection model for solar irradiation and PV power production. In fact, one should expect even more expressive gains when using more input parameters. Furthermore, the presented forecast of PV power production for the city of Fez can be extended to different sites with different climatic zones. Another important perspective for our research works is to extend the forecasting horizons to 24 hours ahead.

Bibliography

[1] C. Zou, Q. Zhao, G. Zhang, B. Xiong, Energy revolution: From a fossil energy era to a new energy era, Natural Gas Industry B 3 (1) (2016) 1–11.

[2] R. G. Newell, D. Raimi, Global energy outlook comparison methods: 2020 update, https://www.rff.org/publications/reports/global-energy-outlook-comparison-methods-2020/, [Online; accessed 01-September-2020] (2020).

[3] S. Adesola, F. Brennan, Energy in Africa, Springer, 2019.

[4] M. E. Eshun, J. Amoako-Tuffour, A review of the trends in ghana's power sector, Energy, Sustainability and Society 6 (1) (2016) 9.

[5] K. Choukri, A. Naddami, S. Hayani, Renewable energy in emergent countries: lessons from energy transition in morocco, Energy, Sustainability and Society 7 (1) (2017) 25.

[6] B. Schinke, J. Klawitter, D. Zejli, T. Barradi, I. Garcia, A. Leidreiter, Background paper: country fact sheet morocco, Energy and Development at a Glance 2016 (2016) 58.

[7] S. Hanger, N. Komendantova, B. Schinke, D. Zejli, A. Ihlal, A. Patt, Community acceptance of large-scale solar energy installations in developing countries: Evidence from morocco, Energy Research & Social Science 14 (2016) 80–89.

[8] P. IEA, Energy policies beyond IEA countries: Morocco 2019, https://www.iea.org/reports/energy-policies-beyond-iea-countries-morocco-2019, [Online; accessed 01-Septemebr-2020] (2020).

[9] L. E. Jones, Renewable energy integration: practical management of variability, uncertainty, and flexibility in power grids, Academic Press, 2017.

[10] R. Marquez, C. F. Coimbra, Proposed metric for evaluation of solar forecasting models, Journal of solar energy engineering 135 (1) (2013) 011–016.

[11] Y. Sukamongkol, S. Chungpaibulpatana, W. Ongsakul, A simulation model for predicting the performance of a solar photovoltaic system with alternating current loads, Renewable energy 27 (2) (2002) 237–258.

[12] A. Mellit, A. M. Pavan, A 24-h forecast of solar irradiance using artificial neural network: Application for performance prediction of a grid-connected pv plant at trieste, italy, Solar Energy 84 (5) (2010) 807–821.

[13] L. Wong, W. Chow, Solar radiation model, Applied energy 69 (3) (2001) 191–224.

[14] T. Muneer, S. Younes, S. Munawwar, Discourses on solar radiation modeling, Renewable and Sustainable Energy Reviews 11 (4) (2007) 551–602.

[15] A. Mellit, S. A. Kalogirou, Artificial intelligence techniques for photovoltaic applications: A review, Progress in energy and combustion science 34 (5) (2008) 574–632.

[16] Z. Sen, Solar energy fundamentals and modeling techniques: atmosphere, environment, climate change and renewable energy, Springer Science & Business Media, 2008.

[17] R. Blaga, A. Sabadus, N. Stefu, C. Dughir, M. Paulescu, V. Badescu, A current perspective on the accuracy of incoming solar energy forecasting, Progress in Energy and Combustion Science 70 (2019) 119–144.

[18] G. Energy, Western wind and solar integration study, Tech. rep., National Renewable Energy Lab.(NREL), Golden, CO (United States) (2010).

[19] S. Venkataraman, G. Jordan, R. Piwko, L. Freeman, U. Helman, C. Loutan, G. Rosenblum, M. Rothleder, J. Xie, H. Zhou, M. Kuo, Integration of renewable resources: Operational requirements and generation fleet capability at 20 percent RPS, Tech. rep. (2010).

[20] S. R. Kulkarni, B. Rajendran, Spiking neural networks for handwritten digit recognition-supervised learning and network optimization, Neural Networks 103 (2018) 118–127.

[21] A. Gani, K. Mohammadi, S. Shamshirband, H. Khorasanizadeh, A. S. Danesh, J. Piri, Z. Ismail, M. Zamani, Day of the year-based prediction of horizontal global solar radiation by a neural network auto-regressive model, Theoretical and applied climatology 125 (3-4) (2016) 679–689.

[22] W. Zhou, Verification of the nonparametric characteristics of backpropagation neural networks for image classification, IEEE Transactions on geoscience and remote sensing 37 (2) (1999) 771–779.

[23] X. Yao, Evolving artificial neural networks, Proceedings of the IEEE 87 (9) (1999) 1423–1447.

[24] D. W. Opitz, J. W. Shavlik, Generating accurate and diverse members of a neural-network ensemble, in: Advances in neural information processing systems, 1996, pp. 535–541.

[25] Y. Kwon, A. Kwasinski, A. Kwasinski, Solar irradiance forecast using naïve Bayes classifier based on publicly available weather forecasting variables, Energies 12 (8) (2019) 1529.

[26] G. F. Miller, P. M. Todd, S. U. Hegde, Designing neural networks using genetic algorithms., in: ICGA, Vol. 89, 1989, pp. 379–384.

[27] V. Badescu, Modeling solar radiation at the earth's surface, Vol. 1, Springer, 2014.

[28] S. A. Kalogirou, Solar energy engineering: processes and systems, Academic Press, 2013.

[29] C. Rigollier, O. Bauer, L. Wald, On the clear sky model of the esra—european solar radiation atlas—with respect to the heliosat method, Solar energy 68 (1) (2000) 33–48.

[30] A. Oumbe, Exploitation des nouvelles capacités d'observation de la terre pour évaluer le rayonnement solaire incident au sol, Ph.D. thesis, École Nationale Supérieure des Mines de Paris (2009).

[31] P. Cooper, The absorption of radiation in solar stills, Solar energy 12 (3) (1969) 333–346.

[32] J. Spencer, Fourier series reprensentation of the position of the sun, Search 2 (5) (1971) 172.

[33] M. Iqbal, An introduction to solar radiation, Elsevier, 2012.

[34] J. R. Brownson, Solar energy conversion systems, Academic Press, 2013.

[35] T. Muneer, Solar radiation and daylight models, Routledge, 2007.

[36] R. Foster, M. Ghassemi, A. Cota, Solar energy: renewable energy and the environment, CRC press, 2009.

[37] M. Iqbal, Study of canadian diffuse and total solar radiation data. i. monthly average daily horizontal radiation, Sol. Energy;(United States) 22 (1).

[38] T. Khatib, A. Mohamed, K. Sopian, A review of solar energy modeling techniques, Renewable and Sustainable Energy Reviews 16 (5) (2012) 2864–2869.

[39] J. Mubiru, E. Banda, Estimation of monthly average daily global solar irradiation using artificial neural networks, Solar energy 82 (2) (2008) 181–187.

[40] A. Whillier, The determination of hourly values of total solar radiation from daily summations, Archiv für Meteorologie, Geophysik und Bioklimatologie, Serie B 7 (2) (1956) 197–204.

[41] B. Y. Liu, R. C. Jordan, The interrelationship and characteristic distribution of direct, diffuse and total solar radiation, Solar energy 4 (3) (1960) 1–19.

[42] C. Touzet, les réseaux de neurones artificiels, introduction au connexionnisme, 1992.

[43] S. Haykin, Neural Networks, A comprehensive foundation, Prentice-Hall, New Jersey, 1999.

[44] J.-L. Beuchat, Reconnaissance de caractères manuscrits à l'aide de réseaux neuromimétiques, Tech. rep., IDIAP (1997).

[45] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proceedings of the IEEE 86 (11) (1998) 2278–2324.

[46] W. S. McCulloch, W. Pitts, A logical calculus of the ideas immanent in nervous activity, The bulletin of mathematical biophysics 5 (4) (1943) 115–133.

[47] D. O. Hebb, The organization of behavior: a neuropsychological theory, J. Wiley; Chapman & Hall, 1949.

[48] F. Rosenblatt, The perceptron: a probabilistic model for information storage and organization in the brain., Psychological review 65 (6) (1958) 386.

[49] Y. LeCun, Une procedure d'apprentissage ponr reseau a seuil asymetrique, Proceedings of Cognitiva 85 (1985) 599–604.

[50] D. E. Rumelhart, G. E. Hinton, R. J. Williams, Learning internal representations by error propagation, Tech. rep., California Univ San Diego La Jolla Inst for Cognitive Science (1985).

[51] B. Widrow, M. E. Hoff, Adaptive switching circuits, Tech. rep., Stanford Univ Ca Stanford Electronics Labs (1960).

[52] G. Cybenko, Approximation by superpositions of a sigmoidal function, Mathematics of control, signals and systems 2 (4) (1989) 303–314.

[53] K.-I. Funahashi, On the approximate realization of continuous mappings by neural networks, Neural networks 2 (3) (1989) 183–192.

[54] A. R. Barron, Universal approximation bounds for superpositions of a sigmoidal function, IEEE Transactions on Information theory 39 (3) (1993) 930–945.

[55] K. Hornik, M. Stinchcombe, H. White, P. Auer, Degree of approximation results for feedforward networks approximating unknown mappings and their derivatives, Neural Computation 6 (6) (1994) 1262–1275.

[56] G. Burel, Reseaux de neurones en traitement d'images-des modèles théoriques aux applications industrielles, Ph.D. thesis (1991).

[57] F. Rosenblatt, Principles of neurodynamics: Perceptions and the theory of brain mechanisms.

[58] J. Moody, C. J. Darken, Fast learning in networks of locally-tuned processing units, Neural computation 1 (2) (1989) 281–294.

[59] N. Murata, S. Yoshizawa, S.-i. Amari, Network information criterion-determining the number of hidden units for an artificial neural network model, IEEE transactions on neural networks 5 (6) (1994) 865–872.

[60] Y. Wada, M. Kawato, Estimation of generalization capability by combination of new information criterion and cross validation, in: IJCNN-91-Seattle International Joint Conference on Neural Networks, Vol. 2, IEEE, 1991, pp. 1–6.

[61] J. Hérault, C. Jutten, Réseaux neuronaux et traitement du signal, Hermes, 1994.

[62] B.-T. Zhang, An incremental learning algorithm that optimizes network size and sample size in one trial, in: Proceedings of 1994 IEEE International Conference on Neural Networks (ICNN'94), Vol. 1, IEEE, 1994, pp. 215–220.

[63] S. Yamamoto, T. Oshino, T. Mori, A. Hashizume, J. Motoike, Gradual reduction of hidden units in the back propagation algorithm, and its application to blood cell classification, in: Proceedings of 1993 International Conference on Neural Networks (IJCNN-93-Nagoya, Japan), Vol. 3, IEEE, 1993, pp. 2085–2088.

[64] C. Darwin, On the origin of species london, UK: John Murray 62.

[65] D. Denyer, D. Tranfield, Producing a systematic review.

[66] A. Qazi, H. Fayaz, A. Wadi, R. G. Raj, N. Rahim, W. A. Khan, The artificial neural network for solar radiation prediction and designing solar systems: a systematic literature review, Journal of cleaner production 104 (2015) 1–12.

[67] M. Laidi, S. Hanini, A. Rezrazi, M. R. Yaiche, A. A. El Hadj, F. Chellali, Supervised artificial neural network-based method for conversion of solar radiation data (case study: Algeria), Theoretical and applied climatology 128 (1-2) (2017) 439–451.

[68] E. F. Alsina, M. Bortolini, M. Gamberi, A. Regattieri, Artificial neural network optimisation for monthly average daily global solar radiation prediction, Energy conversion and management 120 (2016) 320–329.

[69] Ö. Çelik, A. Teke, H. B. Yıldırım, The optimized artificial neural network model with levenberg–marquardt algorithm for global solar radiation estimation in eastern mediterranean region of turkey, Journal of cleaner production 116 (2016) 1–12.

[70] J. Waewsak, C. Chancham, M. Mani, Y. Gagnon, Estimation of monthly mean daily global solar radiation over bangkok, thailand using artificial neural networks, Energy Procedia 57 (2014) 1160–1168.

[71] A. K. Yadav, H. Malik, S. Chandel, Selection of most relevant input parameters using weka for artificial neural network based solar radiation prediction models, Renewable and Sustainable Energy Reviews 31 (2014) 509–519.

[72] Y. Jiang, Prediction of monthly mean daily diffuse solar radiation using artificial neural networks and comparison with other empirical models, Energy policy 36 (10) (2008) 3833–3837.

[73] S. Alam, S. Kaushik, S. Garg, Computation of beam solar radiation at normal incidence using artificial neural network, Renewable Energy 31 (10) (2006) 1483–1491.

[74] A. Sözen, E. Arcaklioğlu, M. Özalp, Estimation of solar potential in turkey by artificial neural networks using meteorological and geographical data, Energy Conversion and Management 45 (18-19) (2004) 3033–3052.

[75] A. Sözen, E. Arcaklioğlu, M. Özalp, E. G. Kanit, Use of artificial neural networks for mapping of solar potential in turkey, Applied Energy 77 (3) (2004) 273–286.

[76] A. S. Dorvlo, J. A. Jervase, A. Al-Lawati, Solar radiation estimation using artificial neural networks, Applied Energy 71 (4) (2002) 307–319.

[77] M. Mohandes, A. Balghonaim, M. Kassas, S. Rehman, T. Halawani, Use of radial basis functions for estimating monthly mean daily solar radiation, Solar Energy 68 (2) (2000) 161–168.

[78] S. Al-Alawi, H. Al-Hinai, An ann-based approach for predicting global radiation in locations with no direct measurement instrumentation, Renewable Energy 14 (1-4) (1998) 199–204.

[79] M. Mohandes, S. Rehman, T. Halawani, Estimation of global solar radiation using artificial neural networks, Renewable energy 14 (1-4) (1998) 179–184.

[80] M. A. Jallal, S. Chabaa, A. Zeroual, A new artificial multi-neural approach to estimate the hourly global solar radiation in a semi-arid climate site, Theoretical and Applied Climatology 139 (3-4) (2020) 1261–1276.

[81] M. Marzouq, Z. Bounoua, H. El Fadili, A. Mechaqrane, K. Zenkouar, Z. Lakhliai, New daily global solar irradiation estimation model based on automatic selection of

input parameters using evolutionary artificial neural networks, Journal of Cleaner Production 209 (2019) 1105–1118.

[82] A. Khosravi, R. Koury, L. Machado, J. Pabon, Prediction of hourly solar radiation in abu musa island using machine learning algorithms, Journal of Cleaner Production 176 (2018) 63–75.

[83] X. Xue, Prediction of daily diffuse solar radiation using artificial neural networks, international journal of hydrogen energy 42 (47) (2017) 28214–28221.

[84] H. B. Yildirim, Ö. Çelik, A. Teke, B. Barutçu, Estimating daily global solar radiation with graphical user interface in Eastern Mediterranean region of Turkey, Renewable and Sustainable Energy Reviews.

[85] M. Vakili, S. R. Sabbagh-Yazdi, S. Khosrojerdi, K. Kalhor, Evaluating the effect of particulate matter pollution on estimation of daily global solar radiation using artificial neural network modeling based on meteorological data, Journal of Cleaner Production 141 (2017) 1275–1285.

[86] C. Renno, F. Petito, A. Gatto, ANN model for predicting the direct normal irradiance and the global radiation for a solar application to a residential building, Journal of Cleaner Production 135 (2016) 1298–1316.

[87] K. Chiteka, C. Enweremadu, Prediction of global horizontal solar irradiance in zimbabwe using artificial neural networks, Journal of Cleaner Production 135 (2016) 701–711.

[88] M. Shaddel, D. S. Javan, P. Baghernia, Estimation of hourly global solar irradiation on tilted absorbers from horizontal one using artificial neural network for case study of mashhad, Renewable and Sustainable Energy Reviews 53 (2016) 59–67.

[89] B. Ihya, A. Mechaqrane, R. Tadili, M. Bargach, Prediction of hourly and daily diffuse solar fraction in the city of fez (morocco), Theoretical and applied climatology 120 (3-4) (2015) 737–749.

[90] M. Behrang, E. Assareh, A. Ghanbarzadeh, A. Noghrehabadi, The potential of different artificial neural network (ann) techniques in daily global solar radiation modeling based on meteorological data, Solar Energy 84 (8) (2010) 1468–1480.

[91] J. C. Lam, K. K. Wan, L. Yang, Solar radiation modelling using anns for different climates in china, Energy Conversion and Management 49 (5) (2008) 1080–1090.

[92] J. Bosch, G. Lopez, F. Batlles, Daily solar irradiation estimation over a mountainous area using artificial neural networks, Renewable Energy 33 (7) (2008) 1622–1628.

[93] H. K. Elminir, Y. A. Azzam, F. I. Younes, Prediction of hourly and daily diffuse fraction using neural network, as compared to linear regression models, Energy 32 (8) (2007) 1513–1523.

[94] F. Tymvios, C. Jacovides, S. Michaelides, C. Scouteli, Comparative study of Ångström's and artificial neural networks' methodologies in estimating global solar radiation, Solar energy 78 (6) (2005) 752–762.

[95] J. Soares, A. P. Oliveira, M. Z. Božnar, P. Mlakar, J. F. Escobedo, A. J. Machado, Modeling hourly diffuse solar radiation in the city of são paulo using a neural-network technique, Applied energy 79 (2) (2004) 201–214.

[96] A. Ångström, Solar and terrestrial radiation. Report to the international commission for solar research on actinometric investigations of solar and atmospheric radiation, Quarterly Journal of the Royal Meteorological Society 50 (210) (1924) 121–126.

[97] S. M. Mousavi, E. S. Mostafavi, A. Jaafari, A. Jaafari, F. Hosseinpour, Using measured daily meteorological parameters to predict daily solar radiation, Measurement 76 (2015) 148–155.

[98] M. Benghanem, A. Mellit, S. Alamri, Ann-based modelling and estimation of daily global solar radiation data: A case study, Energy conversion and management 50 (7) (2009) 1644–1655.

[99] N. Premalatha, A. Valan Arasu, Prediction of solar radiation for solar systems by using ann models with different back propagation algorithms, Journal of applied research and technology 14 (3) (2016) 206–214.

[100] A. K. Yadav, S. Chandel, Solar radiation prediction using artificial neural network techniques: A review, Renewable and sustainable energy reviews 33 (2014) 772–781.

[101] Y. Kemmoku, S. Orita, S. Nakagawa, T. Sakakibara, Daily insolation forecasting using a multi-stage neural network, Solar Energy 66 (3) (1999) 193–199.

[102] G. Mihalakakou, M. Santamouris, D. Asimakopoulos, The total solar radiation time series simulation in athens, using neural networks, Theoretical and Applied Climatology 66 (3-4) (2000) 185–197.

[103] A. Sfetsos, A. Coonick, Univariate and multivariate forecasting of hourly solar radiation with artificial intelligence techniques, Solar Energy 68 (2) (2000) 169–178.

[104] A. Mellit, M. Benghanem, S. A. Kalogirou, An adaptive wavelet-network model for forecasting daily total solar-radiation, Applied Energy 83 (7) (2006) 705–722.

[105] L. Martín, L. F. Zarzalejo, J. Polo, A. Navarro, R. Marchante, M. Cony, Prediction of global solar irradiance based on time series analysis: Application to solar thermal power plants energy production planning, Solar Energy 84 (10) (2010) 1772–1781.

[106] R. Marquez, C. F. Coimbra, Forecasting of global and direct solar irradiance using stochastic learning methods, ground experiments and the nws database, Solar Energy 85 (5) (2011) 746–756.

[107] S. Chen, H. Gooi, M. Wang, Solar radiation forecast based on fuzzy logic and neural networks, Renewable Energy 60 (2013) 195–201.

[108] K. Benmouiza, A. Cheknane, Forecasting hourly global solar radiation using hybrid k-means and nonlinear autoregressive neural network models, Energy Conversion and Management 75 (2013) 561–569.

[109] H. T. Pedro, C. F. Coimbra, Short-term irradiance forecastability for various solar micro-climates, Solar Energy 122 (2015) 587–602.

[110] M. Ghofrani, M. Ghayekhloo, R. Azimi, A novel soft computing framework for solar radiation forecasting, Applied Soft Computing 48 (2016) 207–216.

[111] L. Benali, G. Notton, A. Fouilloy, C. Voyant, R. Dizene, Solar radiation forecasting using artificial neural network and random forest methods: Application to normal beam, horizontal diffuse and global components, Renewable energy 132 (2019) 871–884.

[112] J. Principe, N. Euliano, C. Lefebvre, Hebbian learning and principal component analysis, in: Neural and Adaptive Systems: Fundamentals Threough Simulation, John Wiley and Sons, 1999.

[113] G. Lopez, F. Batlles, J. Tovar-Pescador, Selection of input parameters to model direct solar irradiance by using artificial neural networks, Energy 30 (9) (2005) 1675–1684.

[114] T. Khatib, A. Mohamed, M. Mahmoud, K. Sopian, Modeling of daily solar energy on a horizontal surface for five main sites in malaysia, International Journal of Green Energy 8 (8) (2011) 795–819.

[115] H. Othieno, J. Awange, Energy Resources in Africa, Springer, 2016.

[116] M. Marzouq, H. El Fadili, Z. Lakhliai, K. Zenkouar, A review of solar radiation prediction using artificial neural networks, in: Wireless Technologies, Embedded and Intelligent Systems (WITS), 2017 International Conference on, IEEE, 2017, pp. 1–6.

[117] K. Mohammadi, S. Shamshirband, C. W. Tong, K. A. Alam, D. Petković, Potential of adaptive neuro-fuzzy system for prediction of daily global solar radiation by day of the year, Energy Conversion and Management 93 (2015) 406–413.

[118] A. H. Wright, Genetic algorithms for real parameter optimization, in: Foundations of genetic algorithms, Vol. 1, Elsevier, 1991, pp. 205–218.

[119] D. Whitley, T. Starkweather, C. Bogart, Genetic algorithms and neural networks: Optimizing connections and connectivity, Parallel computing 14 (3) (1990) 347–361.

[120] K. F. Man, K. S. Tang, S. Kwong, Genetic algorithms for control and signal processing, Springer Science & Business Media, 2012.

[121] W. M. Spears, K. A. De Jong, An analysis of multi-point crossover, in: Foundations of genetic algorithms, Vol. 1, Elsevier, 1991, pp. 301–315.

[122] A. J. Arnfield, Köppen climate classification, `http://www.britannica.com/science/Koppen-climate-classification`, [Online; accessed 12-April-2018] (2009).

[123] Z. Bounoua, M. Marzouq, A. Mechaqrane, Assessment of a quality control procedure of hourly solar irradiations at Fez city, Morocco, in: IOP Conference Series: Earth and Science, Vol. 161, IOP Publishing, 2018, p. 012010.

[124] Solargis geomodel, solar global horizontal irradiation map of morocco, http://fr.wikipedia.org/wiki/Fichier:SolarGIS-Solar-map-Morocco-en.png, [Online; accessed 20-August-2018] (2013).

[125] B. V. Bonnlander, A. S. Weigend, Selecting input variables using mutual information and nonparametric density estimation, in: Proceedings of the 1994 International Symposium on Artificial Neural Networks (ISANN'94), 1994, pp. 42–50.

[126] P. Van De Laar, S. Gielen, T. Heskes, Input selection with partial retraining, in: International Conference on Artificial Neural Networks, Springer, 1997, pp. 469–474.

[127] J. Branke, Evolutionary algorithms for neural network design and training, in: In Proceedings of the First Nordic Workshop on Genetic Algorithms and their Applications, 1995, pp. 145–163.

[128] J. H. Holland, Adaptation in Natural and Artificial Systems, University of Michigan, Michigan/U.S.A., 1975.

[129] D. Golmohammadi, R. C. Creese, H. Valian, J. Kolassa, Supplier Selection Based on a Neural Network Model using Genetic Algorithm, IEEE transactions on neural networks 20 (9) (2009) 1504–1519.

[130] X. Yao, Global optimisation by evolutionary algorithms, in: Parallel Algorithms/Architecture Synthesis, 1997. Proceedings., Second Aizu International Symposium, IEEE, 1997, pp. 282–291.

[131] E. Cantú-Paz, C. Kamath, Evolving neural networks to identify bent-double galaxies in the first survey, Neural networks 16 (3-4) (2003) 507–517.

[132] X. Yu, M. Gen, Introduction to evolutionary algorithms, London Limited: Springer-Verlag, 2010.

[133] T. Blickle, L. Thiele, A comparison of selection schemes used in genetic algorithms (1995).

[134] T. Back, Evolutionary Algorithms in Theory and Practice: Evolution Strategies, Evolutionary Programming, Genetic Algorithms, Oxford university press, 1996.

[135] H. C. Bayrakçı, C. Demircan, A. Keçebaş, The development of empirical models for estimating global solar radiation on horizontal surface: A case study, Renewable and Sustainable Energy Reviews.

[136] K. Bakirci, Models of solar radiation with hours of bright sunshine: a review, Renewable and Sustainable Energy Reviews 13 (9) (2009) 2580–2588.

[137] M. Yorukoglu, A. N. Celik, A critical review on the estimation of daily global solar radiation from sunshine duration, Energy Conversion and Management 47 (15-16) (2006) 2441–2450.

[138] J. Almorox, M. Bocco, E. Willington, Estimation of daily global solar radiation from measured temperatures at Cañada de Luque, Córdoba, Argentina, Renewable Energy 60 (2013) 382–387.

[139] G. H. Hargreaves, Z. A. Samani, Estimating potential evapotranspiration, Journal of the Irrigation and Drainage Division 108 (3) (1982) 225–230.

[140] Z. Samani, Estimating solar radiation and evapotranspiration using minimum climatological data, Journal of Irrigation and Drainage Engineering 126 (4) (2000) 265–267.

[141] M.-F. Li, H.-B. Liu, P.-T. Guo, W. Wu, Estimation of daily solar radiation from routinely observed meteorological data in Chongqing, China, Energy Conversion and Management 51 (12) (2010) 2575–2579.

[142] M. Benghanem, A. Mellit, A simplified calibrated model for estimating daily global solar radiation in Madinah, Saudi Arabia, Theoretical and Applied Climatology 115 (1-2) (2014) 197–205.

[143] G. Wu, Y. Liu, T. Wang, Methods and strategy for modeling daily global solar radiation with measured meteorological data–a case study in Nanchang station, China, Energy Conversion and Management 48 (9) (2007) 2447–2452.

[144] J. Fan, B. Chen, L. Wu, F. Zhang, X. Lu, Y. Xiang, Evaluation and development of temperature-based empirical models for estimating daily global solar radiation in humid regions, Energy 144 (2018) 903–914.

[145] N. S. Altman, An introduction to Kernel and Nearest-Neighbor nonparametric regression, The American Statistician 46 (3) (1992) 175–185.

[146] Y. Zhang, J. Wang, Gefcom2014 probabilistic solar power forecasting based on K-Nearest Neighbor and Kernel density estimator, in: Power & Energy Society General Meeting, 2015 IEEE, IEEE, 2015, pp. 1–5.

[147] H. T. Pedro, C. F. Coimbra, Nearest-Neighbor methodology for prediction of intra-hour global horizontal and direct normal irradiances, Renewable Energy 80 (2015) 770–782.

[148] worldbank, Morocco to make history with first of its kind solar plant, http://www.worldbank.org/en/news/feature/2015/11/20/morocco-to-make-history-with-first-of-its-kind-solar-plant, [accessed: 20-December-2019] (2015).

[149] R. H. Inman, H. T. Pedro, C. F. Coimbra, Solar forecasting methods for renewable energy integration, Progress in energy and combustion science 39 (6) (2013) 535–576.

[150] A. Tuohy, J. Zack, S. E. Haupt, J. Sharp, M. Ahlstrom, S. Dise, E. Grimit, C. Mohrlen, M. Lange, M. G. Casado, et al., Solar forecasting: methods, challenges, and performance, IEEE Power and Energy Magazine 13 (6) (2015) 50–59.

[151] C. Coimbra, J. Kleissl, R. Marquez, Overview of solar-forecasting methods and a metric for accuracy evaluation in: Solar energy forecasting and resource assessment (2013).

[152] D. Yang, J. Kleissl, C. A. Gueymard, H. T. Pedro, C. F. Coimbra, History and trends in solar irradiance and PV power forecasting: A preliminary assessment and review using text mining, Solar Energy 168 (2018) 60–101.

[153] J. G. De Gooijer, R. J. Hyndman, 25 years of time series forecasting, International journal of forecasting 22 (3) (2006) 443–473.

[154] N. Ding, Y. Besanger, F. Wurtz, G. Antoine, P. Deschamps, Time series method for short-term load forecasting using smart metering in distribution systems, in: 2011 IEEE Trondheim PowerTech, IEEE, 2011, pp. 1–6.

[155] C. Voyant, M. Muselli, C. Paoli, M.-L. Nivet, Numerical weather prediction (NWP) and hybrid ARMA/ANN model to predict global radiation, Energy 39 (1) (2012) 341–355.

[156] J. O. Kamadinata, T. L. Ken, T. Suwa, Sky image-based solar irradiance prediction methodologies using artificial neural networks, Renewable Energy 134 (2019) 837–845.

[157] M. Mohri, A. Rostamizadeh, A. Talwalkar, Foundations of machine learning, MIT press, 2018.

[158] C. Voyant, G. Notton, S. Kalogirou, M.-L. Nivet, C. Paoli, F. Motte, A. Fouilloy, Machine learning methods for solar radiation forecasting: A review, Renewable Energy 105 (2017) 569–582.

[159] J. Polo, L. Martín-Pomares, A. Sanfilippo, Solar Resources Mapping: Fundamentals and Applications, Springer, 2019.

[160] R. Reed, I. Marks, Rj (1999). neural smithing. supervised learning in feedforward artificial neural networks, Cambridge, MA: MIT Press.

[161] F.-V. Gutierrez-Corea, M.-A. Manso-Callejo, M.-P. Moreno-Regidor, M.-T. Manrique-Sancho, Forecasting short-term solar irradiance based on artificial neural networks and data from neighboring meteorological stations, Solar Energy 134 (2016) 119–131.

[162] J. Kamruzzaman, Artificial neural networks in finance and manufacturing, IGI Global, 2006.

[163] M. R. Azimi-Sadjadi, S. Sheedvash, F. O. Trujillo, Recursive dynamic node creation in multilayer neural networks, IEEE Transactions on Neural Networks 4 (2) (1993) 242–256.

[164] K. Chen, L. Wang, Trends in neural computation, Vol. 35, Springer, 2006.

[165] P. J. Angeline, G. M. Saunders, J. B. Pollack, An evolutionary algorithm that constructs recurrent neural networks, IEEE transactions on Neural Networks 5 (1) (1994) 54–65.

[166] X. Yao, M. M. Islam, Evolving artificial neural network ensembles, IEEE Computational Intelligence Magazine 3 (1) (2008) 31–42.

[167] C.-K. Goh, E.-J. Teoh, K. C. Tan, Hybrid multiobjective evolutionary design for artificial neural networks, IEEE Transactions on Neural Networks 19 (9) (2008) 1531–1548.

[168] Y. Liu, X. Yao, T. Higuchi, Evolutionary ensembles with negative correlation learning, IEEE Transactions on Evolutionary Computation 4 (4) (2000) 380–387.

[169] B. A. Hoverstad, On the effect of network modularity on evolutionary search, in: 2009 IEEE Congress on Evolutionary Computation, IEEE, 2009, pp. 3191–3198.

[170] J.-B. Mouret, S. Doncieux, Evolving modular neural-networks through exaptation, in: 2009 IEEE congress on evolutionary computation, IEEE, 2009, pp. 1570–1577.

[171] K. Bandurski, W. Kwedlo, A Lamarckian hybrid of differential evolution and conjugate gradients for neural network training, Neural processing letters 32 (1) (2010) 31–44.

[172] J. urgen Branke, Evolutionary algorithms in neural network design and training - a review, Proc. of the First Nordic Workshop on Genetic Algorithms and their Applications (1NWGA), (1995) 145–164.

[173] Y. Liang, K.-S. Leung, Genetic algorithm with adaptive elitist-population strategies for multimodal function optimization, Applied Soft Computing 11 (2) (2011) 2017–2034.

[174] J. Wu, C. K. Chan, Prediction of hourly solar radiation with multi-model framework, Energy conversion and management 76 (2013) 347–355.

[175] ADEREE, Règlement thermique de construction au Maroc (RTCM), http:///www.amee.ma/sites/default/files/inline-files/Reglement_thermique_de_construction_au_Maroc.pdf, [accessed: 25-March-2020] (2014).

[176] I. Merini, A. Molina-García, M. S. García-Cascales, M. Ahachad, Energy efficiency regulation and requirements: Comparison between morocco and spain, in: International Conference on Advanced Intelligent Systems for Sustainable Development, Springer, 2018, pp. 197–209.

[177] B. H. Morantz, T. Whalen, G. P. Zhang, A weighted window approach to neural network time series forecasting, in: Neural networks in business forecasting, IGI Global, 2004, pp. 251–265.

[178] A. Umbarkar, P. Sheth, Crossover operators in genetic algorithms: a review., ICTACT journal on soft computing 6.

[179] T. Kmet, M. Kmetova, A 24h forecast of solar irradiance using echo state neural networks, in: Proceedings of the 16th International Conference on Engineering Applications of Neural Networks (INNS), ACM, 2015, pp. 1–5.

[180] J. Mubiru, Predicting total solar irradiation values using artificial neural networks, Renewable Energy 33 (10) (2008) 2329–2332.

[181] F. Wang, Z. Mi, S. Su, H. Zhao, Short-term solar irradiance forecasting model based on artificial neural network using statistical feature parameters, Energies 5 (5) (2012) 1355–1370.

[182] J. Kleissl, Solar energy forecasting and resource assessment, Academic Press, 2013.

[183] A. Kumler, Y. Xie, Y. Zhang, A physics-based smart persistence model for intra-hour forecasting of solar radiation (PSPI) using GHI measurements and a cloud retrieval technique, Solar Energy 177 (2019) 494–500.

[184] J. Huertas Tato, M. Centeno Brito, Using smart persistence and random forests to predict photovoltaic energy production, Energies 12 (1) (2019) 100.

[185] N. Engerer, F. Mills, Validating nine clear sky radiation models in australia, Solar Energy 120 (2015) 9–24.

[186] X. Sun, J. M. Bright, C. A. Gueymard, B. Acord, P. Wang, N. A. Engerer, Worldwide performance assessment of 75 global clear-sky irradiance models using principal component analysis, Renewable and Sustainable Energy Reviews 111 (2019) 550–570.

[187] B. Beckers, Solar energy at urban scale, John Wiley & Sons, 2013.

[188] G. De'Ath, Multivariate regression trees: a new technique for modeling species–environment relationships, Ecology 83 (4) (2002) 1105–1117.

[189] Y. Z. Engin, K. Turhan, A. Örem, Prediction of some complete blood count parameters' values using boosted regression tree, in: 2012 International Symposium on Innovations in Intelligent Systems and Applications, IEEE, 2012, pp. 1–4.

[190] M. A. Hassan, A. Khalil, S. Kaseb, M. Kassem, Potential of four different machine-learning algorithms in modeling daily global solar radiation, Renewable Energy 111 (2017) 52–62.

[191] M. A. Hassan, A. Khalil, S. Kaseb, M. Kassem, Exploring the potential of tree-based ensemble methods in solar radiation modeling, Applied Energy 203 (2017) 897–916.

[192] L. Breiman, Random forests, Machine learning 45 (1) (2001) 5–32.

[193] J. Huang, A. Troccoli, P. Coppin, An analytical comparison of four approaches to modelling the daily variability of solar irradiance using meteorological records, Renewable Energy 72 (2014) 195–202.

[194] R. Urraca, J. Antonanzas, M. Alia-Martinez, F. Martinez-de Pison, F. Antonanzas-Torres, Smart baseline models for solar irradiation forecasting, Energy conversion and management 108 (2016) 539–548.

[195] I. A. Ibrahim, T. Khatib, A novel hybrid model for hourly global solar radiation prediction using random forests technique and firefly algorithm, Energy Conversion and Management 138 (2017) 413–425.

[196] S. Eftekharnejad, G. T. Heydt, V. Vittal, Optimal generation dispatch with high penetration of photovoltaic generation, IEEE Transactions on Sustainable Energy 6 (3) (2014) 1013–1020.

[197] J. Antonanzas, N. Osorio, R. Escobar, R. Urraca, F. J. Martinez-de Pison, F. Antonanzas-Torres, Review of photovoltaic power forecasting, Solar Energy 136 (2016) 78–111.

[198] G. Graditi, S. Ferlito, G. Adinolfi, Comparison of photovoltaic plant power production prediction methods using a large measured dataset, Renewable Energy 90 (2016) 513–519.

[199] P. Horta, I. FhG, Process heat collectors: State of the art and available medium temperature collectors, International Energy Agency Task 49.

[200] A. Q. Jakhrani, A. R. H. Rigit, R. Baini, S. R. Samo, L. P. Ling, et al., Investigation of solar photovoltaic module power output by various models, NED University Journal of Research (2012) 25–35.

[201] J. Rosell, M. Ibanez, Modelling power output in photovoltaic modules for outdoor operating conditions, Energy conversion and management 47 (15-16) (2006) 2424–2430.

[202] P. Singh, S. Singh, M. Lal, M. Husain, Temperature dependence of i–v characteristics and performance parameters of silicon solar cell, Solar Energy Materials and Solar Cells 92 (12) (2008) 1611–1616.

[203] B. K. Hodge, Alternative energy systems and applications, John Wiley & Sons, 2017.

[204] A. Amir, J. Selvaraj, N. A. Rahim, Study of the mpp tracking algorithms: Focusing the numerical method techniques, Renewable and Sustainable Energy Reviews 62 (2016) 350–371.

[205] A. Gupta, Y. K. Chauhan, R. K. Pachauri, A comparative investigation of maximum power point tracking methods for solar pv system, Solar energy 136 (2016) 236–253.

[206] D. Verma, S. Nema, A. Shandilya, S. K. Dash, Maximum power point tracking (mppt) techniques: Recapitulation in solar photovoltaic systems, Renewable and Sustainable Energy Reviews 54 (2016) 1018–1034.

[207] T.-C. Yu, T.-S. Chien, Analysis and simulation of characteristics and maximum power point tracking for photovoltaic systems, in: 2009 International Conference on Power Electronics and Drive Systems (PEDS), IEEE, 2009, pp. 1339–1344.

[208] A. Chermitti, O. Boukli-Hacene, S. Mouhadjer, Design of a library of components for autonomous photovoltaic system under matlab/simulink, International Journal of Computer Applications 53 (14).

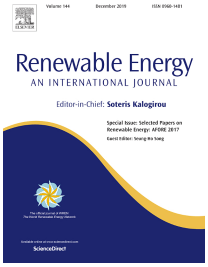
[209] M. A. Masoum, H. Dehbonei, E. F. Fuchs, Theoretical and experimental analyses of photovoltaic systems with voltageand current-based maximum power-point tracking, IEEE Transactions on energy conversion 17 (4) (2002) 514–522.

[210] Z. Li, S. Rahman, R. Vega, B. Dong, A hierarchical approach using machine learning methods in solar photovoltaic energy production forecasting, Energies 9 (1) (2016) 55.

Journal and Conference Publications

Journal publications:

1. **Marzouq, M.**, El Fadili, H., A., Zenkouar, K., Lakhliai, Z. and Amouzg, M. (2020). "Short term solar irradiance forecasting via a novel evolutionary multi-model framework and performance assessment for sites with no solar irradiance data." Renewable Energy, 157, 214 – 231. DOI: 10.1016/j.renene.2020.04.133



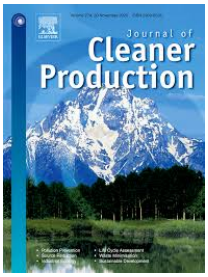
CiteScore: 11.2

Impact Factor: 6.274 (ClarivateAnalytics, 2019)

5-Year Impact Factor: 5.964 (ClarivateAnalytics, 2019)

SCImagoJournal Rank (SJR): 2.052

2. **Marzouq, M.**, Bounoua, Z., El Fadili, H., Mechaqrane, A., Zenkouar, K., and Lakhliai, Z., (2019). "New daily global solar irradiation estimation model based on automatic selection of input parameters using evolutionary artificial neural networks". Journal of cleaner production, 209, 1105-1118. DOI: 10.1016/j.jclepro.2018.10.254



CiteScore: 10.9

Impact Factor: 7.246 (ClarivateAnalytics, 2019)

5-Year Impact Factor: 7.494 (ClarivateAnalytics, 2019)

SCImagoJournal Rank (SJR): 1.886

3. **Marzouq, M**., Bounoua, Z., Mechaqrane, A., El Fadili, H., Lakhliai, Z., and Zenkouar, K., (2018). "ANN-based modelling and prediction of daily global solar irradiation using commonly measured meteorological parameters." IOP Conference Series: Earth and Environmental Science. Vol. 161. No. 1. IOP Publishing, 2018. (SJR=0.82). DOI:10.1088/1755-1315/161/1/012017
4. Bounoua, Z., **Marzouq, M.**, and Mechaqrane, A., (2018). "Assessment of a quality

control procedure of hourly solar irradiations at Fez city, Morocco" IOP Conference Series: Earth and Environmental Science. Vol. 161. No. 1. IOP Publishing, 2018. (SJR=0.82). DOI:10.1088/1755-1315/161/1/012010

Conference papers:

5. **Marzouq, M.**, El Fadili, H., Lakhliai, Z., Abdellah Mechaqrane and Zenkouar, K. "New distance weighted k Nearest Neighbor model for hourly global solar irradiation estimation". The 5th edition of the IEEE International Conference on WIreless Technologies, embedded and intelligent Systems-WITS-National School of Applied Sciences, Fez, Morocco, 04, 05 and 06 April 2019.
6. **Marzouq, M.**, Bounoua, Z., Mechaqrane, A., El Fadili, H., Lakhliai, Z. and Zenkouar, K. "Prediction of daily global solar irradiation using commonly measured meteorological parameters in the city of Fez". International Conference on Renewable energies and Energy Efficiency REEE'2017, Faculty of Sciences and Technologies of Fez, Morocco, 8 and 9 November 2017.
7. Bounoua, Z., **Marzouq, M.**., and Mechaqrane, A. "Assessment of a quality control procedure of hourly solar irradiations at Fez city, Morocco". International Conference on Renewable energies and Energy Efficiency REEE'2017, Faculty of Sciences and Technologies, Fez, Morocco, 8 and 9 November 2017.
8. **Marzouq, M.**, El Fadili, H., Lakhliai, Z. and Zenkouar, K. "A review of solar radiation prediction using artificial neural networks". The 5th edition of the **IEEE** International Conference on Wireless Technologies, Embedded and Intelligent Systems (WITS), 19 and 20 April, 2017. DOI:10.1109/WITS.2017.7934658
9. Marzouq, M., El Fadili, H. and Lakhliai, Z. "Prediction of solar radiation and its components: State of art ". 2ème Colloque franco-marocain sur les Energies Renouvelables et leur intégration aux réseaux de transport et de distribution , Université Sidi Mohammed Ben Abdellah , Fès, Maroc,11 et 13 Avril 2017.
10. Habou Laouali I., Qassemi H., **Marzouq, M.**, Ruano A., Dosse Bennani S. and El Fadili H. "A Survey on Computational Intelligence Techniques for Non Intrusive Load Monitoring". The 2nd **IEEE** International Conference on Electronics, Control, Optimization and Computer Science (ICECOCS), FSK, ENSA-Kenitra, Morocco, 2nd and 3rd December 2020. Accepted communication.
11. Ruano A., Qassemi H., Habou Laouali I., **Marzouq, M.**, Dosse Bennani S. and El Fadili H. "A Model-based Predictive Control approach for Home Energy Management

Systems. First results". The 6th edition of the International Conference on WIreless Technologies, embedded and intelligent Systems-WITS-National School of Applied Sciences, Fez, Morocco, 14, 15 and 16 Octber 2020. Accepted communication.

12. Habou Laouali I., Qassemi H., **Marzouq, M.**, Ruano A., Dosse Bennani S. and El Fadili H. "A Non linear autoregressive neural network model for forecasting appliance power consumption". The 6th edition of the International Conference on WIreless Technologies, embedded and intelligent Systems-WITS-National School of Applied Sciences, Fez, Morocco, 14, 15 and 16 Octber 2020. Accepted communication.