Solar Resource Assessment in Morocco with three Machines Learning Models and Ordinary Kriging

Mohamed Chaibi*‡, EL Mahjoub Benghoulam*, Mohamed Berrada, Abdellah El Hmaidi*****

*University of Moulay Ismail, Faculty of Science, Department of Physics, Team of Renewable energy and energy efficiency, BP 11201, Zitoune, Meknes, Morocco

**University of Moulay Ismail, ENSAM, Laboratory of Mathematical and Computational Modeling, Marjane II, BP 15290, Al Mansour, 50000, Meknes, Morocco.

***University of Moulay Ismail, Faculty of Science, Department of Geology, Laboratory of Water Sciences and environmental engineering, BP 11201, Zitoune, Meknes, Morocco

‡ Corresponding Author; Mohamed Chaibi, Faculty of Science, Department of Physics, BP 11201, Zitoune, Meknes, Morocco,

IJASR 2021
VOLUME 4
ISSUE 3 MAY – JUNE **ISSN: 2581-7876**

Abstract: Accurate assessment of the solar resource is an essential step for successful planning, design, and operation of solar energy systems. In this study, the potential of three machine learning methods, i.e. Gaussian Process(GP), Support Vector Machines Regression (SVR), and Artificial Neural Networks (ANN), was evaluated for estimating monthly solar irradiation (H) in Morocco with4 inputs variables (latitude, average temperature(T), average relative humidity(RH), and month of the year). Then, the Ordinary Kriging (OK) method was used with 86 data to generate monthly and annual solar maps of the country at a regional scale. The results revealed that the three methods showed close results in terms of prediction accuracy, the values of the coefficient of determination (R^2) for GP, SVRs, and ANNs are (0.960,0.966, and 0.963)for the testing phase. Also, the GP model was the most stable with an increase of 7.1% in the testing root mean squared error (RMSE) compared to 36.7% and 26.6% for the SVR and ANN algorithms. Nevertheless, its computational cost is approximately 14 and 24 much higher than the computational cost of the SVR and ANN models, respectively. Based on the solar maps, and the subdivision of the solar potential in seven classes ranging between poor and superb, the results confirmed that Morocco has a large solar resource dominated by excellent (5-5.57kWh/m^2. day) and outstanding (5.57-6.08 kWh/m^2. day) classes. The outstanding solar resource was found in the southern and southeastern regions of the country.

Keywords: Solar resource, Gaussian Process, Artificial Neural Networks, Support Vector Machines Regression, Ordinary Kriging, Outstanding.

1. Introduction

Global solar radiation reaching the Earth's surface is critical for a variety of applications, including meteorology, hydrology, crop production estimation, atmospheric physics, and, most significantly, the design and use of renewable solar energy [1].Solar energy is widely harnessed in different regions around the world due to its exceptional nature, which is abundant, environmentally sustainable, and inexhaustible. This helps to increase sustainability and mitigate some of the negative environmental problems caused by the excessive use of fossil fuels [2]. With abundant solar resources (an average of 5.3kWh/m2) and under annual sunshine durations ranging from 2700h in the north to approximately 3500h in the south, Morocco has one of the most ambitious energy targets in the world [3].This goal is to produce 52% of the country's electricity in 2030 from renewable sources: 20% of them from solar energy [4].

A precise assessment of the solar resource is a critical component of any solar project, as it is needed for design, power production estimation, and decision-making.[5]. This assessment requires knowledge of the solar radiation intensity and its spatial distribution.The most accurate method of acquiring this information is through remote measurements taken at a specific location. However, due to the high cost of calibrating and maintaining these instruments, solar radiation data are scarce in a large number of meteorological stations worldwide, especially in

developing countries such as Morocco[6].For this aim,a vast number of models have been suggested to estimate solar radiation such as empirical models, models developed from satellite images, clear sky models, and machine learning models [7].Moreover,a precise solar resource assessment can help support the developmentof solar power systems in phases of exploration, prospection, site selection, and pre-feasibility evaluation[8]. This assessment is done by drawing solar irradiation maps allowing the exploration and the analysis of the variability of the solar potentialin a country [5].

Machine learning algorithms have been used widely throughout the past decade to estimate solar radiation, with the ANN model being the most popular.Şenkal [9] applied ANN technique to estimate global solar radiation using measured dataof five stations in Turkey. The ANN technique had betterresults compared to other meteorological methods. Rumbayan et al.[10] Successfully estimatedthe monthly solar irradiation in Indonesia by using ANN method. The results indicated that the ANN method canbe useful in remote locations for island sites in Indonesia which lackground measurement.In another study[11],ANN was applied using 13 inputs to predict the monthly solar radiation in 45 locations over Italy.The best results are obtained with 7 inputs(top of atmosphere radiation, day length, number of rainy days, rainfall, latitude,time period, and altitude).Wang et al. [12] also investigated three types of ANN models, Multilayer Perceptron (MLP), Generalized Regression Neural Network (GRNN), and Radial Basis Neural Network (RBNN) for predicting daily solar irradiation at 12 stations in different climatic zones of China. The results showed that the ANN models could produce satisfactory solar irradiation estimates at most stations, and the MLP and RBF models provided better accuracy than the GRNN model.Kumar et al. [13] compared the performance of ANN models to the corresponding regression models for estimating monthly global solar irradiation and concluded that the ANN models are better than the regression models.

Recently, SVR techniques have gained the attention of researchers in the field of solar energy.Chen et al.[14]evaluated seven SVRs modelsusing different inputsat three stations in the Liaoning province in China. The authors found that all the SVRs models outperformedsignificantlythe studied empirical models.Belaid and Mellit [15]applied successfully SVRs to estimate daily and monthly global solar horizontal radiation inGhardaïa(Algeria).Theproposed SVRs models have the same performances compared with ANNs models and other models published in the literature.Inref [16] three machine learning methods were compared (adaptive neuro-fuzzy inference system (ANFIS), ANN, and SVR) to predict daily horizontal global solar radiation in the YucatánPeninsula, México. The SVRapproachperformed better than the ANN and ANFIS technique.Olatomiwa et al.[17]investigated the potential of the SVR techniquefor global solar radiation modeling in a semi-arid environment in Nigeria.It was found that the SVR model outperformed both the ANFIS and the empirical models. Rohani et al.[18]have evaluated the performance of the Gaussian Process (GP) model for daily and monthly H prediction at Mashhad of Iran. Results of comparison between the GP model and other machines learning models in literature showed that this new model can be used to predict daily and monthly solar radiation with high accuracy.

The interpolation techniques are used to estimate solar radiation at an unsampled site from measurements made at nearby stations[19]. In general, interpolation techniques are classifiedas deterministic and geostatistical.Deterministic methods create surfaces using mathematical functions, based on the degree of similarity or the degree of smoothness. Examples of these techniques are the spline-functions or weighted averages [20].On the other hand, geostatistical techniques generate the prediction surfaces using statistical models. These methods quantify the spatial autocorrelation among sampling data and evaluate the uncertainty of the obtained results[21].Among these methods, one of the most widely used are the kriging methods.Ertekin et al.[22] used universal Kriging to map the annual solar potential in Turkey.Cross-validation statistics indicated that this technique was sufficiently reliable to predict the spatial variability of the global solar radiation across the country.A comparative study between OK and Residual Kriging was carried out for the mapping of solar radiation in southern Spain. It was found that the Residual Kriging model outperformed the OK model especially, in the autumn and winter months [20].Inorder to determine the areas and the periods of the year with the greatest solar resourceMexico's golf course, Escobedo et al. [23] employed OK to construct monthly and annual solar maps of this region. In ref [24] the monthly solar radiationis estimated firstby a sunshine duration model. Then a topographical factor is applied before using the modeled data for mapping the solar resource in Koreausing OK.Oummi et al[25] characterized solar irradiation resource in Morocco with Neural Kriging using three inputs (latitude, longitude, and elevation), the authors indicated that the most interesting sites for solar power plant installation in the country are located in the southern, southeastern and some internal territories.From this brief review of the three models used in this study for modeling solar radiation, it can be noticed that the GP model is rarely employed.Besides, to the best of the authors' knowledge,there is no comprehensive assessment of Moroccan solar resources at regional level.

The present study aims to achieve the following objectives: Firstly, to predict the monthly solar irradiation in Morocco using three machine learning techniques GP, SVR, and ANN,based on four input variables (latitude, month of the year,average temperature T, and relative humidity RH).Secondly,to compare thethree methods in terms of accuracy, stability, and computational time. Thirdly,to construct monthly and annual solar maps using OK at a regional scale, and to assign each region to its corresponding solar class.

2. Methodology

2.1. Study area and data collection

Morocco is located in the subtropical zone of northwest Africa between latitudes 21N and 36N and longitudes 1W and 17W. It is bordered by the Atlantic Ocean to the west, the Mediterranean Sea to the north, Algeria to the east, and Mauritania to the south, with an area of750810 Km²[26].Since 2015, Morocco is divided into 12 regions. These regions are presented in Fig.1. The climate of the country varies according to the seasons and has many nuances: mediterranean to the north, oceanic to the west, continental ininternal regions,and Saharian to the south[27].

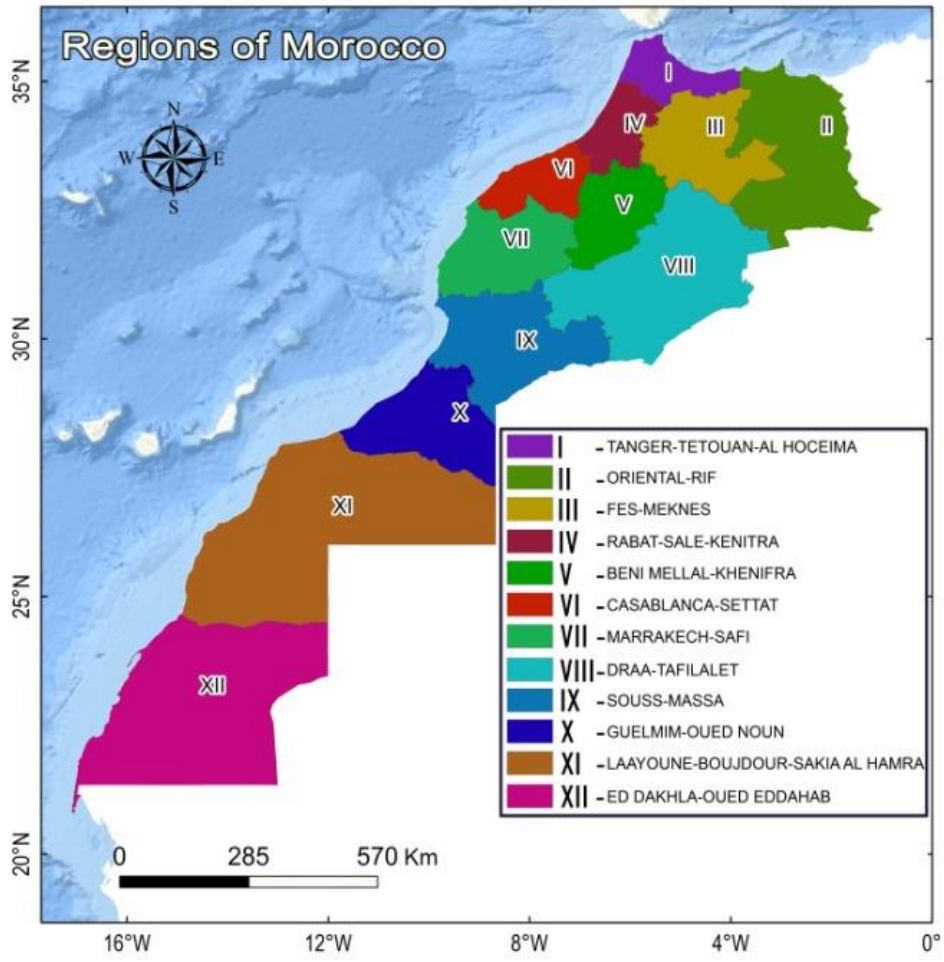


Fig 1. The 12 administrative Regions of Morocco.

Eighty-six locations representing different climatic regions in the country are used for collecting monthly weather datafrom RETScreen[28]and NASA[29]databases. These databasesincorporate observations from satellites (1983-2005) and weather monitoring stations.The geographical locations of sites used in the present study arepresented in Fig .2.where solar irradiation data for 76 sites are provided by NASA and 10 from weather ground stations.

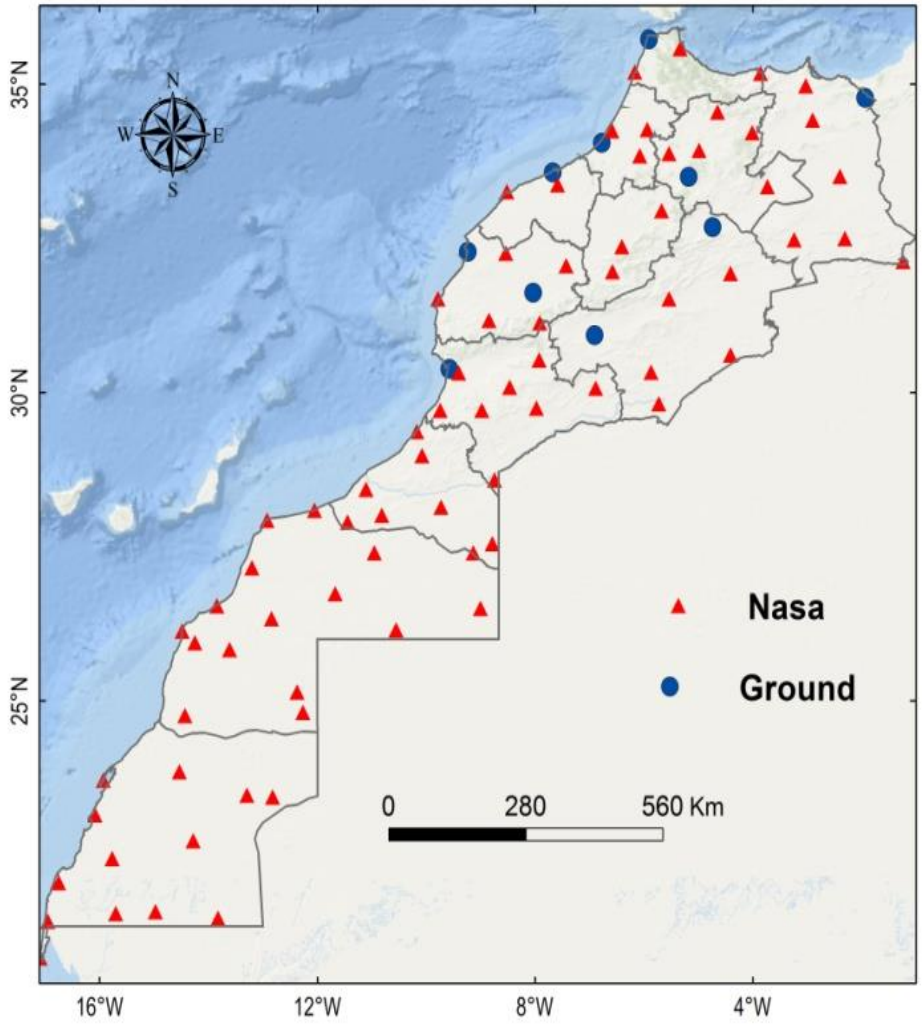
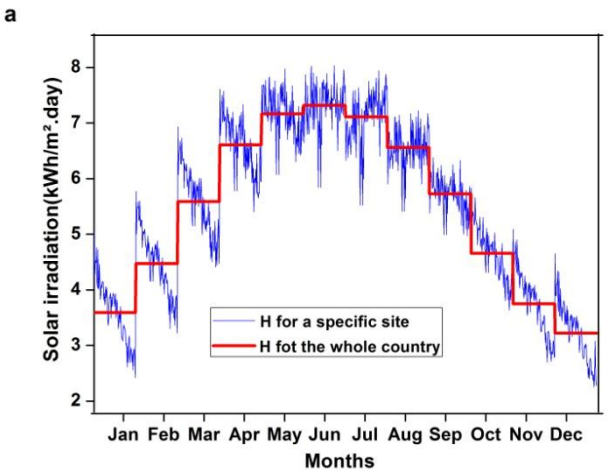


Fig 2. Geographical map of Morocco indicating the sites selected for the study.

The monthly variations of RH, T, and H with increasing latitude varying between 21N and 36N, and for the whole country are shown in Fig.3. A clear dependence of Hon latitude can be observed. For the same latitudes, there are variationsin the climatic conditions represented by the parameters T and RH. Table 1 presents the monthly coefficient of variation (CV) forthe three parameters. The solar radiation H showed the lowestvariability compared to T and RH,the CV varies between 5.4% in September and 16% in December. Themaximum of CV relative to temperature was found to be 32% in January while the minimum was found to be11.8% in August and September. Concerning RH, its CV variesbetween25% in January and 46% in July.The solar radiation depends stronglyon months,the highest values for the whole territory are recorded in the months of May, June, and July with values equals to 7.16kWh/m².day, 7.3kWh/m².dayand 7.1kWh/m².dayrespectively, while the lowest values are observed during November, December, and January with respective values of 3.74kWh/m².day, 3.22kWh/m².dayand 3.59kWh/m².day. The temperature variation is almost similar to that of Hbut the highest values occur in July and August with an average value of 27.38°C and 27.48°C respectively for the whole country and the lowest in January with an average value of 12.48°C. This is due to the inertia of theearth. The variation in RH isapproximately inverse to that of T and H.The maximum RH occurs in the winter months especially in December and January withrespective values of 55.12% and 55.53%,while the minimum in the summer months especially in July with a value of 41.74%.

Table 1: Monthly values of CV for H, T, and RH.

Months	H (%)	T (%)	RH (%)
January	15	32	25
February	13	26	29
March	10	20	31
April	7.8	17	32
May	5.7	14	34
June	5,7	13	38
July	6.7	12	46
August	5.6	11.8	42
September	5.4	11.8	33.7
October	9.9	15	29
November	14.9	22	26
December	16	30	26



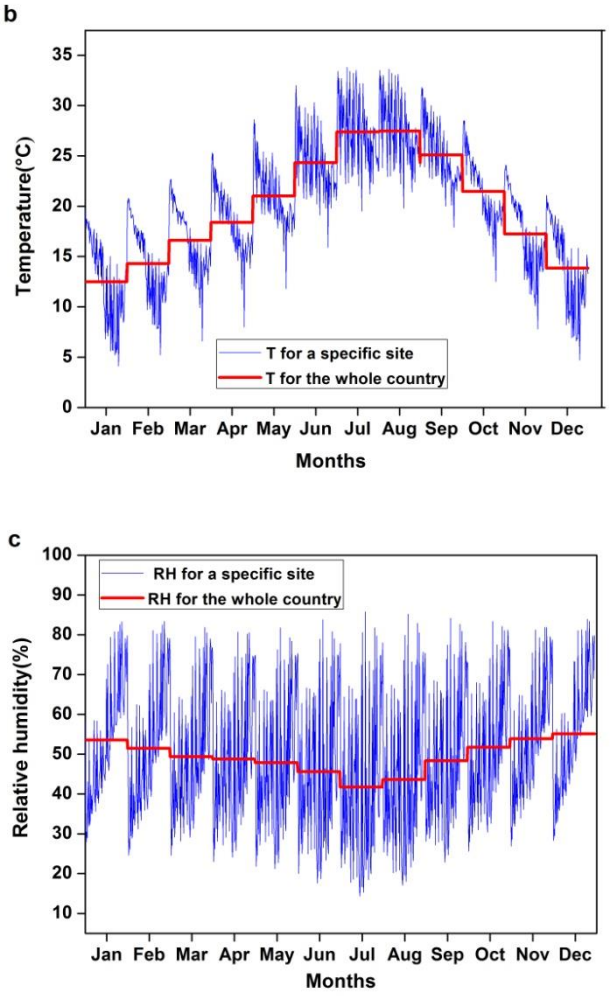


Fig 3. Monthly variation of (a) solar irradiation, (b) temperature and (c) relative humidity.

2.2. Artificial Neural Networks

Artificial Neural Networks are among the methods of artificial intelligence whose design has been inspired by the functioning of biological neurons. An ANN is considered as a black box that includes a series of equations that aims to map a set of input vectors $\{X_i\}$,i=1: N to a set of output vectors $\{Z_i\}$,i=1:Nand improves the outcome by repeated training[30].In the literature, several ANNsmodels are developed. Among these models, Multi-Layer Perceptron (MLP) networks are the most widely used. The structure of the (MLP) network consists of an input layer, an output layer, and one or several hidden layers.

An elementary node k called neuron in an MLP network is shown in Fig.4.It includes a weighted sum of the input vectors$X_i \in R^m$, followed by a nonlinear operation with an activation functionf_k.The output of the neuron k is defined as

$a_k = f_k(\sum_{j=1}^{m} w_{jk}x_{ji} + b_k)$ (1)

where$x_{ji}$is the jth component of the input vector X_i, w_{jk} is the jth component of the weight vector $W$and b_k is the bias.

This can be written in the matrix form as

$a_k = f_k(W^T X_i + b_k)$ (2)

where$X_i = [x_{1i}x_{2i} \dots \dots \dots x_{mi}]^T$ and $W = [w_{1k}w_{2k} \dots \dots \dots w_{mk}]^T$

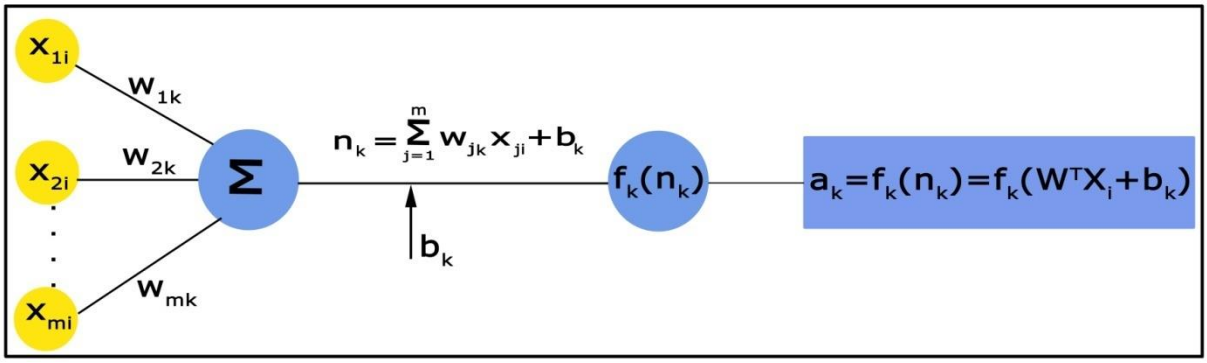


Fig 4. Structure of a single neuron.

The nonlinear activation function used in this study is the hyperbolic tangent, given by

$$f(n_k) = tanh(n_k) = \frac{e^{n_k} - e^{-n_k}}{e^{n_k} + e^{-n_k}} \quad (3)$$

For an MLP network with one hidden layer with hyperbolic tangent neurons followed by an output layer of linear neurons.The calculated output $h(X)$ is given by

$$h(X) = LW^*(tanh(IW^*X + B_1)) + B_2 \quad (4)$$

whereIW andB_1 are the inputs weights and bias matrix for connection from input to hidden layer respectively, andLW and B_2 are layer weights and bias matrix between the hidden layer and output layer.An example of an MLP network with one hidden layer is depicted in Fig.5.[31].

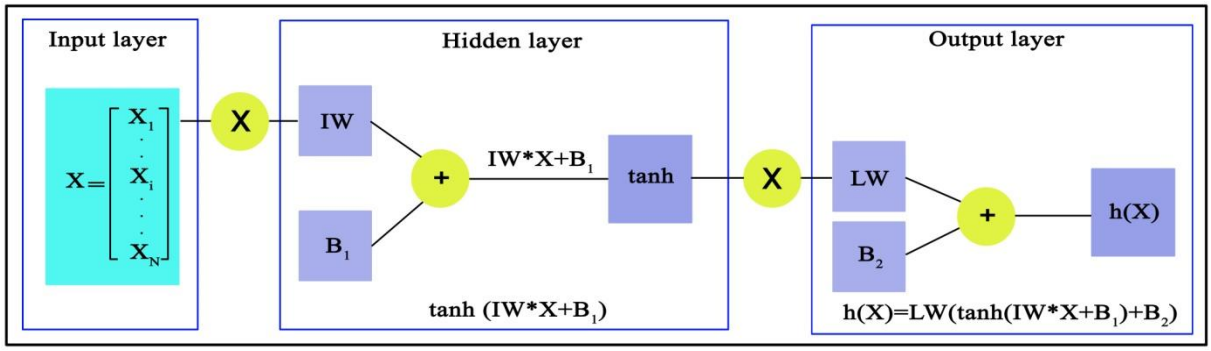


Fig 5. Structure of an MLP network with one hidden layer.

The goal of ANN learning consists of adjusting the weights and bias with a learning algorithmto minimize a cost function defined by:

$$E = \frac{1}{2}\sum_{i=1}^{N} e_i^2 = \frac{1}{2}\sum_{i=1}^{N}(h(X_i) - Z_i)^2 \quad (5)$$

where Nthenumber of patterns, $h(X_i)$ is the calculated output,Z_i is the desired output and $e_i = h(X_i) - Z_i$ is the training error[32].

There are different learning algorithms. A popular algorithm is the steepest descent algorithmbut this algorithm is often too slow for practical problems. Faster algorithms such as conjugate gradient, quasi-Newton, and Levenberg–Marquardt (LM)are more efficient[33].

In this study, the (LM) algorithm has been used. It presents the advantages of Newton's methods and thesteepestgradient descent;in this case, the weights are adjusted as follows:

$$w_{l+1} = w_l - [J^T J + \mu I]^{-1} J^T e \quad (6)$$

whereJis theJacobian matrix that contains the first derivatives of the network errors for the weights and biases,e is a vector of network errors, I is the identity matrix, and $\mu$is the step of learning [30].

2.3. Support Vectors Machines Regression

Support vector machines are powerful supervised learning techniques for both classification and regression. These techniques are based on statistical learning theory and the principle of structural risk minimization[14].In the context of regression, thebasic idea is to transform the input space into a high-dimensional space and learn a linear regression in the new space via the kernel trick[16].Considering the problem of approximating the set of data D= $\{X_i, Z_i\}$,i=1: N, the goal is to find the best function g defined as follows:

$$g(X) = W^T \emptyset(X) + b \quad (7)$$

Where$\emptyset$ is a nonlinear function that maps the inputs data into the high dimensional space, W and b represent the weights and the biasthat are determined by minimizing the regularized risk function

$$R = \frac{1}{2} W^T W + C \sum_i^N L_\varepsilon(g(X_i), Z_i) \quad (8)$$

The minimization of the term$\frac{1}{2} W^T W$will make the function as flat as possible, and regulate the degree of model complexity. The second term is the empirical error and C is a positive trade-off parameter between the degree of the empirical error and the model flatness. The ε −insensitive loss function$L_\varepsilon$is defined as

$$L_\varepsilon(g(X_i), Z_i) = \begin{cases} 0 \; for \; |g(X_i) - Z_i| \leq \varepsilon \\ |g(X_i) - Z_i| - \varepsilon otherwise \end{cases} \quad (9)$$

Where ε is the radius of a region calledthe ε-tube.The errors positioned in the ε-tube are ignored and only the predicted points outside this region cause a loss equals to the magnitude of the difference between thepredicted value and the radius ε of the tube. Positiveslack variables ξ_i and ξ_i^* are introduced to measure the deviation of the training sample outside the ε-tube.

The primal formulation of the regression problem can be formulated as follows

$$Min \frac{1}{2} ||w||^2 + C \sum_{i=1}^N (\xi_i + \xi_i^*) \quad (10)$$

$$\text{Subjected to} \begin{cases} Z_i - W^T \emptyset(X_i) - b_i \leq \varepsilon + \xi_i \\ Z_i - W^T \emptyset(X_i) - b_i \geq -\varepsilon - \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases}$$

By introducing the Lagrange multipliers $\alpha$and α^*this problem can be reformulated into dual problem formalism and written as follows

$$R(\alpha, \alpha^*) = \sum_{i=1}^N y_i(\alpha_i - \alpha_i^*) - \varepsilon \sum_{i=1}^N (\alpha_i - \alpha_i^*) - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) K(X_i, X_j) \quad (11)$$

With constraints

$$\sum_{i=1}^N (\alpha_i - \alpha_i^*) = 0 \qquad 0 \leq \alpha_i \leq C \qquad 0 \leq \alpha_i^* \leq C \qquad i = 1{:}N$$

where$K(X_i, X_j) = \emptyset(X_i)^T \emptyset(X_j)$,K denotes the kernel function whose value equals the inner product of the nonlinear mapping functions of the inputs X_i and X_j[15]. The kernel trick allows SVRs to compute this inner product without having to explicitly compute the map function $\emptyset$.After calculating Lagrange multipliers, the optimal desired weights vector of the regression is found as follows

$$W = \sum_{i=1}^N (\alpha_i - \alpha_i^*) \emptyset(X_i) \quad (12)$$

And the regression function is given by

$$g(X) = \sum_{i=1}^N (\alpha_i - \alpha_i^*) K(X_i, X) + b \quad (13)$$

Any function satisfying Mercers conditions can be used as a kernel function[14]. In this work a Radial Basis Function (RBF) has been used, it is defined as follows

$$K(X_i, X_j) = exp(-\frac{||X_i - X_j||^2}{2\sigma^2}) \quad (14)$$

where$||X_i - X_j||^2$is the squared Euclidean distance between the two input vectors $X_i$and X_j, andσ is the bandwidth parameter of the RBF function[15].

2.4. Gaussian Process

Gaussian processes regression (GP) is a well-established powerful nonparametric framework for nonlinear regression[34]. GP is an extension of multivariate normal random variables parameterized by a mean function m(X) and a covariance matrix k(X,X')

$GP\sim \quad \mathbf{N} \quad (m(X), k(X, X^{'})$ (15)

where **N** denotes the standard normal distribution[35].

Let us now assume that we have observed a set of data D=$\{X_i,\ Z_i\}\ i = 1{:}\ N$. A GPmodel for this data set can be specified as follows $f(X_i) = Z_i + \varepsilon_i$ with $\varepsilon_i = N(0, \sigma^2)$is Gaussian noise. In this study, we assume a zero mean function in the Gaussian Process prior and we choose squared exponential as a covariance function defined as

$$k(X_i, X_j) = \sigma_f^2 exp[-\frac{1}{2}\frac{(X_i - X_j)^T(X_i - X_j)}{\sigma_l^2}] \quad (16)$$

whereσ_f is the signal standard deviation, andσ_l is the characteristic length scale. Thehyperparameter $\sigma_f$specifies the maximum allowable covariance, while $\sigma_l$represents the rate of decay in correlation as points becomefarther away from each other[18].The function values $Z_{1:N}$jointly follow a multivariate Gaussian distribution as$Z_{1:N}\sim\mathbf{N}(0, K)$where covariance matrix K is given as[36]:

$$K = \begin{bmatrix} k(X_1, X_2) & \cdots & k(X_1, X_N) \\ \vdots & \ddots & \vdots \\ k(X_N, X_1) & \cdots & k(X_N, X_n) \end{bmatrix} \quad (17)$$

For a new data point X^*then the joint distribution and the properties of the updated GP are given by

$$\begin{bmatrix} Z_{1:N} \\ Z^* \end{bmatrix} \sim \mathbf{N}(0, \begin{bmatrix} K + \sigma^2 I & k \\ k^T & k(X^*, X^*) + \sigma^2 I \end{bmatrix}) \quad (18)$$

where$k = [k(X^*, X_1) \ldots\ldots k(X^*, X_N)]^T$

$p(Z^*/X^*, D)\sim\mathbf{N}(m(X^*), \sigma^2(X^*))$ (19)

where the predictive mean and the variance aregiven as[36]:

$m(X^*) = k^T[K + \sigma^2 I]^{-1} Z_{1:N}$ (20)

$\sigma^2(X^*) = k(X^*, X^*) - k^T[K + \sigma^2 I]^{-1} k$ (21)

2.5. Ordinary Kriging

Kriging is a geostatistical techniquebased on a variogram model that describes the spatial continuity of data.Kriging aimsto estimate the value of a regionalized variable $Z(s_0)$ at an unknown location from a set of scattered points with observed values $Z(s_i)$[37].Among several techniques of Kriging, OK is the most general and widely used[24].

In geostatistics, the variables are considered as spatial random variables, their values are just one of many infinitely that ispossible. For each, however, we have only a single realization. To make inferences, many realizations are required, so we consider that the process is stationary[38], strict stationarity assumes that the mean and the variance of the variable Z must depend only on the lag-distance h, this condition is very difficult to be verified[39]. Therefore, another weak assumption is considered and known as intrinsic stationarity. Itis expressed in two parts.

The expectation of the deviations is zero

$$E\big(Z(s + h) - E(s)\big) = 0 \quad (22)$$

The variance of the deviations depends only on h

$$Var\big(Z(s + h) - Z(s)\big) = E\left[\big(Z(s + h) - Z(s)\big)^2\right] = 2\gamma(h) \quad (23)$$

The estimated value $\widehat{Z}(s_0)$ by the ordinary Krigingat a non-sampled site $s_0$is a linear combination of the adjacent observations in sites s_i.

$\hat{Z}(s_0) = \sum_i \lambda_i Z(s_i)$(24)

To determine the optimal Kriging weights λ_i,the Kriging estimator must be unbiased $E(\hat{Z}(s_0) - Z(s_0)) = 0$ which is guaranteed when $\sum_i \lambda_i$ =1and optimal when $Var(\hat{Z}(s_0) - Z(s_0))$ is minimal. To obtain this objective, first, an experimental variogram is created from the scattered set of points. It can be written as follows

$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(s_i + h) - Z(s_i)]^2$ (25)

where N (h) represents the points separated by the lag-distance h[20].

After that a well-defined modeled variogram is fitted using the trend in the experimental model, it is an increasing function of h. In this study, an exponential model has been used to fit the experimental variogram.

$\gamma(h) = C_0 + C_1 \left[1 - exp(\frac{-h}{a})\right]$ (26)

whereC_0,$C_1$and$a$ are called, respectively, nugget, partial sill, and range[24].The quantity $C_0 + C_1$ represents the value at which the model first flattens out and it is called sill.The nugget C_0 is caused by sampling errors and the variation at a very short scale.The range a is the distance at which no spatial autocorrelation exists between samples[38].
Finally, theoptimal weightsλ_i are determined by solving the system of equations

$$\begin{cases} \sum_{i=1}^{n} \lambda_i \gamma(s_i, s_j) - \mu = \gamma(s_i, s_0) \\ \sum_{i=1}^{n} \lambda_i = 1 \end{cases} (27)$$

where μ is theLagrangemultiplier [39].

2.6. Implementation of models:

In this study three script files were written in Matlab 2016b version for the development of the GP, SVRs,and ANNstechniques, while the ArcGIS 10.4.1 was used for the implementation of the OK technique.Before applying the ANNs and SVRs models, the data were standardized and scaled to the range [-1, 1] for the ANN model and between [0, 1] for the SVR model. For all models, the databaseis divided into two parts: 60% for training and 40% fortesting. To obtain the optimum architecture of the ANN model, the number of neurons in the hidden layer was changed from 1 to 30.In the case of OK, the best variogram model is selected through cross-validation, which consists of temporarily eliminating a point in the data set and then estimating its value by Kriging using the remaining data and the variogram model that has been adjusted. This operation is repeated for all points.

2.1. Statistical indicators for evaluation

The performance of the predictive models was evaluated using the coefficient of determination R^2,the root mean squared error RMSE, and themean absolute error MAE

$R^2 = 1 - \frac{\sum_{i=1}^{N} (\hat{Z}_i - Z_i)^2}{\sum_{i=1}^{N} (Z_i - \bar{Z}_i)^2}$(28)

$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (\hat{Z}_i - Z_i)^2}$ (29)

$MAE = \frac{1}{N} \sum_{i=1}^{N} (|\hat{Z}_i - Z_i|)$(30)

whereN represents the total number of evaluating data,Z_i and$\hat{Z}_i$represent the measured and predicted solar radiation, respectively,and$\bar{Z}_i$is the mean of measuredvalues.The ideal R^2 value for a model is 1while the ideal RMSE and MAE values are 0[40].

In the case of the OK, cross-validation is evaluated through the normalizedMean error NME, root mean squared error RMSE,and the normalized root of the mean squared error NRMSE

$\mathrm{NME} = \frac{1}{\mathrm{N}} \sum_{\mathrm{i}=1}^{\mathrm{N}} (\frac{\hat{\mathrm{Z}}(\mathrm{s_i}) - \mathrm{Z}(\mathrm{s_i})}{\hat{\sigma}(\mathrm{s_i})})$ (31)

$\mathrm{RMSE} = \sqrt{\frac{1}{\mathrm{N}} \sum_{\mathrm{i}=1}^{\mathrm{N}} (\hat{\mathrm{Z}}(\mathrm{s_i}) - \mathrm{Z}(\mathrm{s_i}))^2}$ (32)

$$NRMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\frac{(\hat{Z}(s_i)-Z(s_i))^2}{\hat{\sigma}(s_i)^2}} \quad (33)$$

where$Z(s_i)$ is the observed solar irradiation in a locations_i,$\hat{Z}(s_i)$ is the estimated solar irradiation at the same location,$\hat{\sigma}(s_i)^2$ is the Krigingvariance, and N is the total number of observations.The OK is better whenthe NMEand RMSE are close to zero and NRMSE is close to 1[41].

3. Results and discussion

3.1. Predictivemodels

Performances of GP, SVR, and ANN models are summarized in table 2.All models gave generally low values of the RMSE and MAE and high values of R^2. Moreover, for all models, it is seen thatthe RMSE and MAE values increasedwhile the R^2 decreased relatively in the testingphase.

It is seen fromTable 2 that the SVR model outperforms both GP and ANN models with respect to MAE statistics in the training and testing stages.The values of the training MAE for GP, SVR and ANN are (0.206, 0.167, 0.174 kWh/m².day) while the values of the testing MAE are (0.234, 0.214, 0.232 kWh/m².day).

Predicted values of the solar irradiation by the 3 machine learning models for training and testing phases are plotted against the measured values in Fig. 6. As can be seen in Fig (6-a), all points fall in the diagonal line, the SVR model has the highest coefficient of determination (R^2=981), it is followed by the ANN model (R^2=0.975) and finally the GP model (R^2=0.965). In the testing stage Fig (6-b), the 3 models showed close results, the values of the coefficient of determination for SVR, GP, and ANN models are 0.966, 0.960, and 0.962, respectively.

Table2: Statistical performances of SVR, GP, and ANN models.

Models	Structure	R^2		RMSE(kWh/m².day)		MAE(kWh/m².day)	
		Training	Testing	Training	Testing	Training	Testing
GP	Squared exponential	0.965	0.960	0.280	0.3	0.206	0.234
SVR	RBF Kernel	**0.981**	**0.966**	**0.207**	**0.283**	**0,167**	**0.214**
ANN	20 neurons	0.975	0.963	0.235	0.297	0,174	0.232

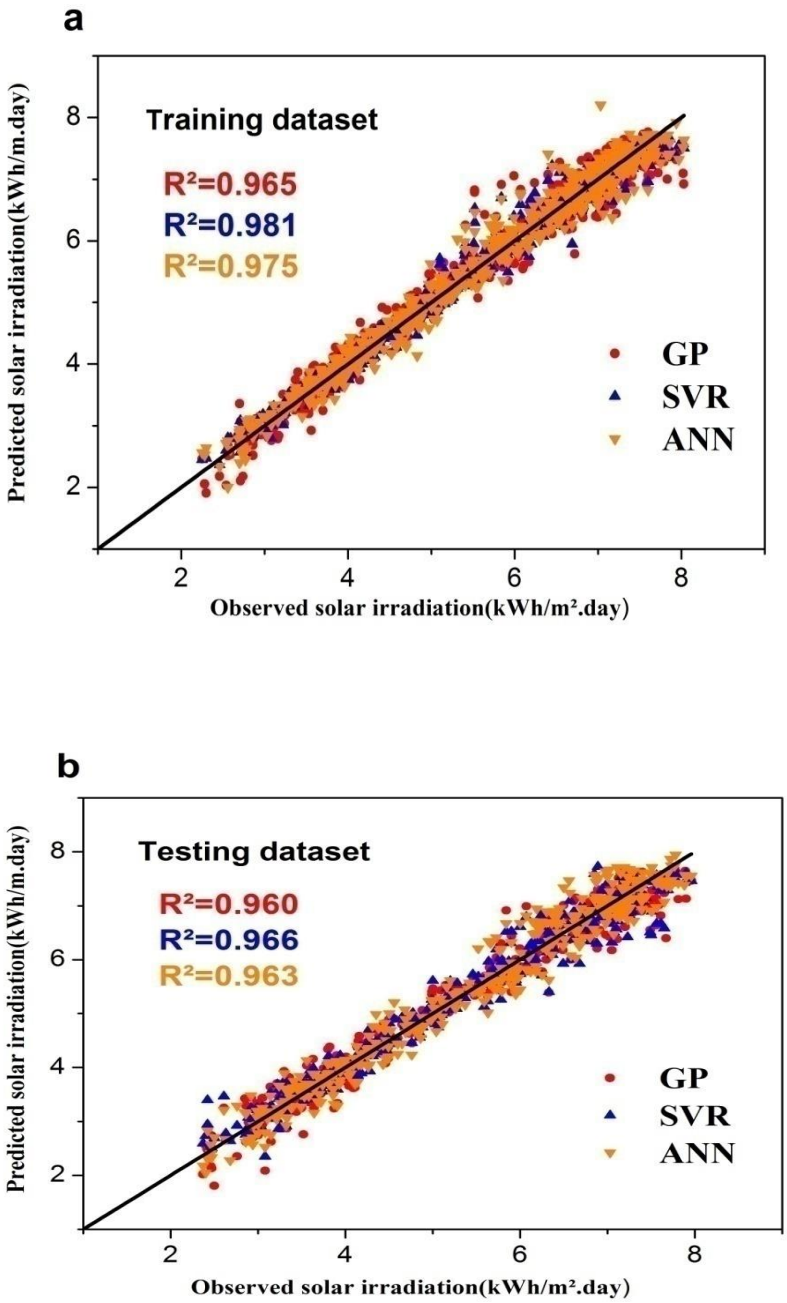


Fig 6. Correlation plots for the 3 models between predicted and observed solar irradiation for (a) training datasets, (b) testing datasets.

The training and the testing RMSE, as well as its percentage increase, are shown in Fig.7, for the 3 studied models. The values of the training RMSE for GP, SVR and ANN are (0.28, 0.207, 0.235 kWh/m².day) while the values of the testing RMSE are (0.3, 0.283, 0.297 kWh/m².day).Based on these results, the six algorithms can be ranked based on their prediction accuracy from best to worst as follows: SVR, ANN, and GP.

The stability of machine learning models is also an important factor toconsider when these techniques are adopted [42]. As seen from Fig.7, the GP algorithm is the most stable among the 3 models with the less percentage increase in the testing RMSE (7.1%) while the SVR model exhibited the largest increase in the testing RMSE (36.7%). Considering their stability, the algorithms can be ordered as follows: GP, ANN, and SVR.

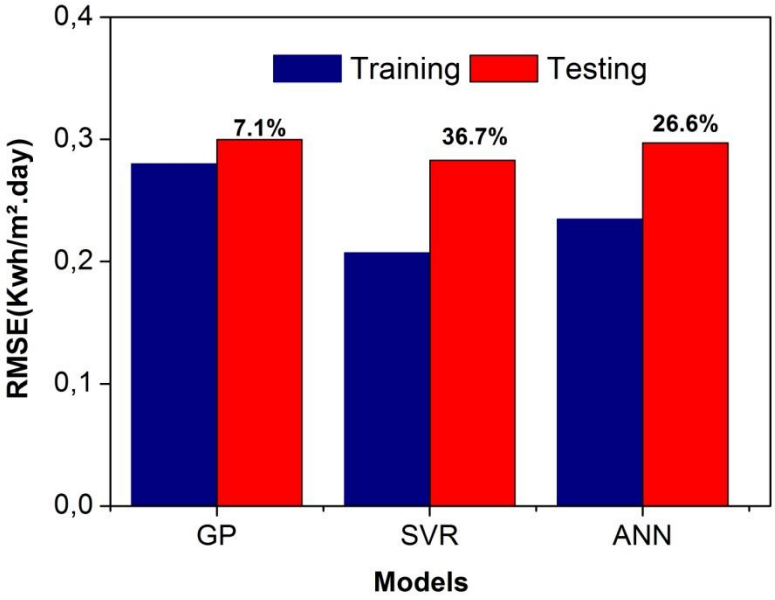


Fig 7. Percentage increase in testing RMSE over training RMSE for the 3 machine learning models.

Even if the prediction accuracy and the stability are primarily considered when using the machine learning models, the computational time of these models is also an important factor, especially when a huge amount of data is available [1].Fig.8 presents the averaged computational time of the 3 models using a single sample containing all dataset.The results indicated that the average time consumed by the GPmodel was much higher than those of the other algorithms, The GP time cost is approximately 14 and 24 much higher than the computational cost of the SVR and ANN models, respectively.

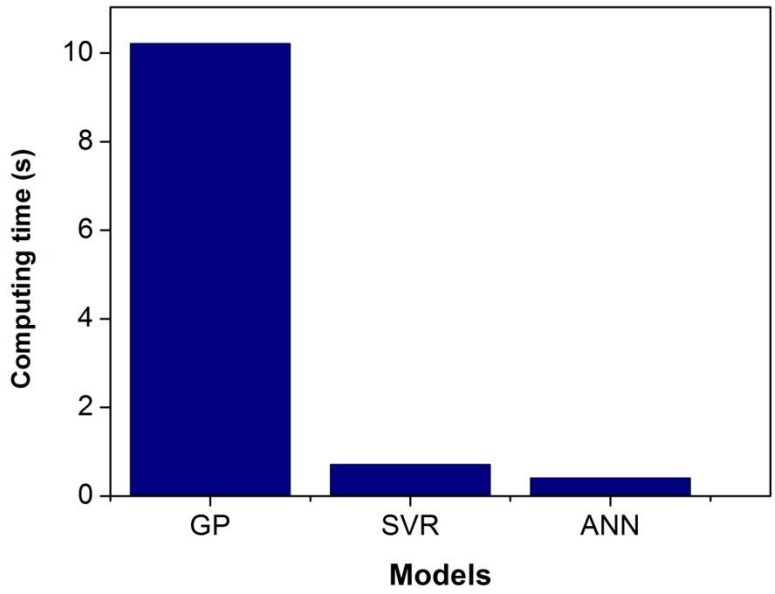


Fig 8. Comparison of computational time used of the 3 machine learning models

Based on the above results, the 3developed modelsGP, SVR, and ANN can be used with high reliability to estimate the monthly solar irradiation in Morocco from a few, simple, and quickly available parameters (T, RH,latitude,and the month of the year). Considering both the prediction accuracy and stability the GP model is highly recommended for predicting monthly H under different climatic conditions of Morocco.

3.2. Ordinary Kriging results

The statistical indicators for the OK model are shown in table 3.The values of the NME are close to zero for all months, which means that the Kriging estimator is unbiased. The values of the RMSE are small and vary between 0.205kWh/m².dayinJanuary and 0.366kWh/m².dayin June, and the NRMSE error values are close to 1 for all months indicating the good performance ofvariogram modeling. The resulting cross-validationerrors in the case of the annual solar irradiation are -0.00128kWh/m².dayfor NME, 0.248kWh/m².dayfor RMSE, and 1.018 for NRMSE. These results show that the OK gives fair estimates for the monthly and the annual solar irradiation in Morocco.

Table3: Statistical indicators for the OK model.

Months	NME	RMSE(kWh/m².day)	NRMSE
January	-0.078	0.205	0.98
February	-0.001	0.259	1.06
March	-0.0168	0.303	1.02
April	-0.0093	0.329	0.99
May	-0.017	0.352	0.938
June	0.0029	0.366	0.9
July	0.0042	0.362	0.93
August	0.0093	0.35	0.936
September	-0.006	0.28	1.02
October	-0.01	0.254	1.086
November	-0.016	0.236	1.1
December	-0.029	0.232	1.09
Annual	-0.0012	0.248	1.018

3.3. Solar resource

The OK technique is used to construct monthly and annual maps ofthe solar resource in Morocco andvisualizeits spatial distribution. For each month, the variability of the solar irradiationis due to the latitude coordinate and the climatic conditions.

The solar resource in autumn is shown in Fig.9.As can be seen from this figure, solar irradiation decreases gradually from September to November.It varies between 4.97kWh/m².dayand 6.42kWh/m².dayin September, between 3.63kWh/m².day and 5.83kWh/m².dayin October, and between 2.7kWh/m².day and 5.09kWh/m².dayin November. The highest values are observed during September in autumn equinox across the country. Moreover, the southern regions XII, XI, and X receive more solar radiation than the north and the northeast regions I, II, and III, especially in November and October. This season is characterized by fairly high temperatures and low relative humidity in the south and south-east and by lower temperatures and relatively high humidity in the north and internal regions.

Fig.10displays the distribution of the solar resource in winter, the solar irradiationincreasesmoderately from December to February. The lowest values during the whole year are recorded in December at the winter solstice;it varies from a minimum value of 2.25kWh/m².day to a maximum value of 4.25kWh/m².day. At this month the sun is at its lowest position in the sky and the days are the shortest.On the other hand, the highestvalues in this season are observed in February with solar radiation ranging from 3.22kWh/m².day to 5.77kWh/m².day.Regions II,Iare the

less irradiated areas of the country; this is due to higher cloud activityresulting fromhigh relative humidity and very low temperatures. Consequently, this is the time of the year when abundant rains are observed.

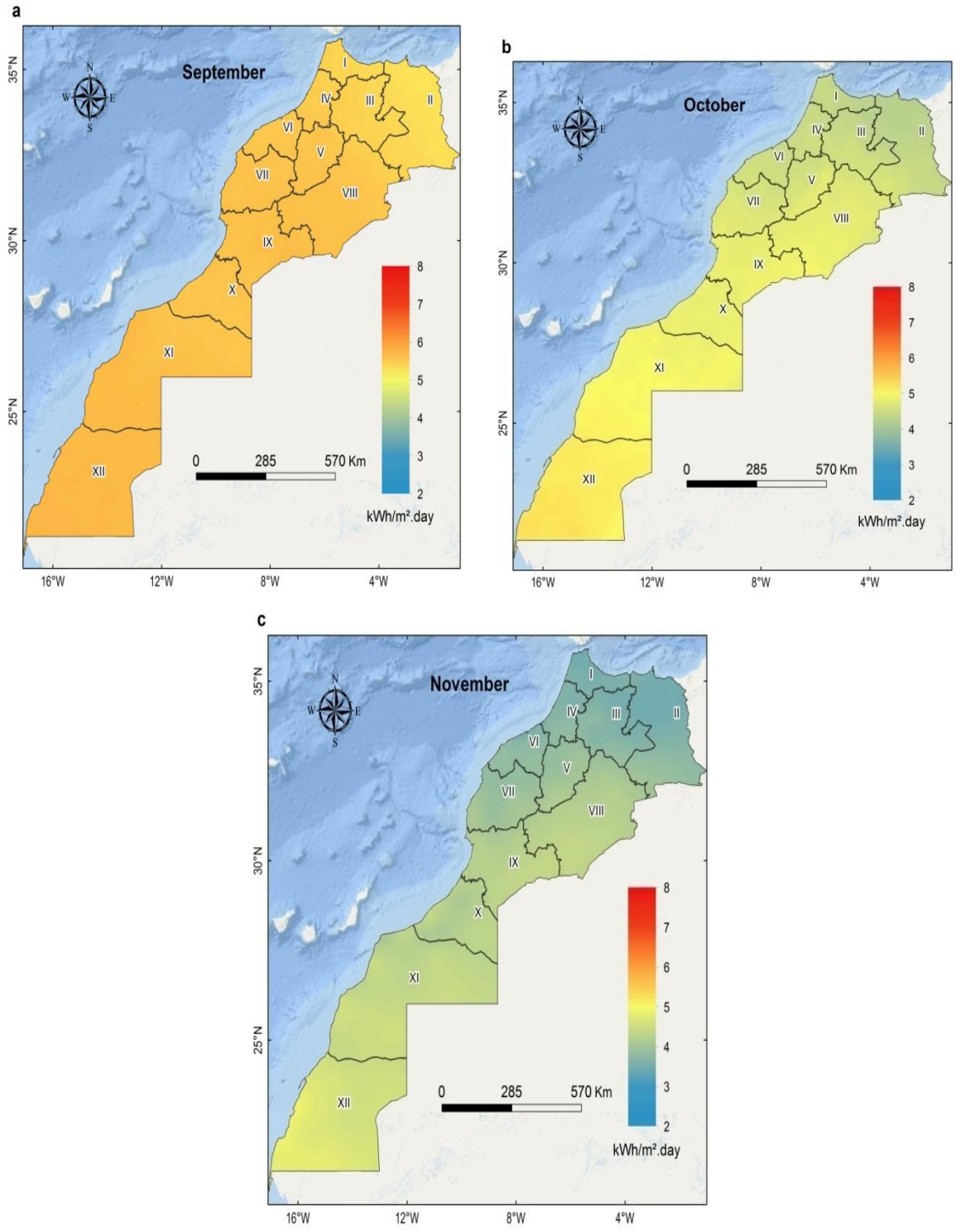


Fig 9. Solar irradiation maps for autumn months (a) September, (b) October, and (c) November.

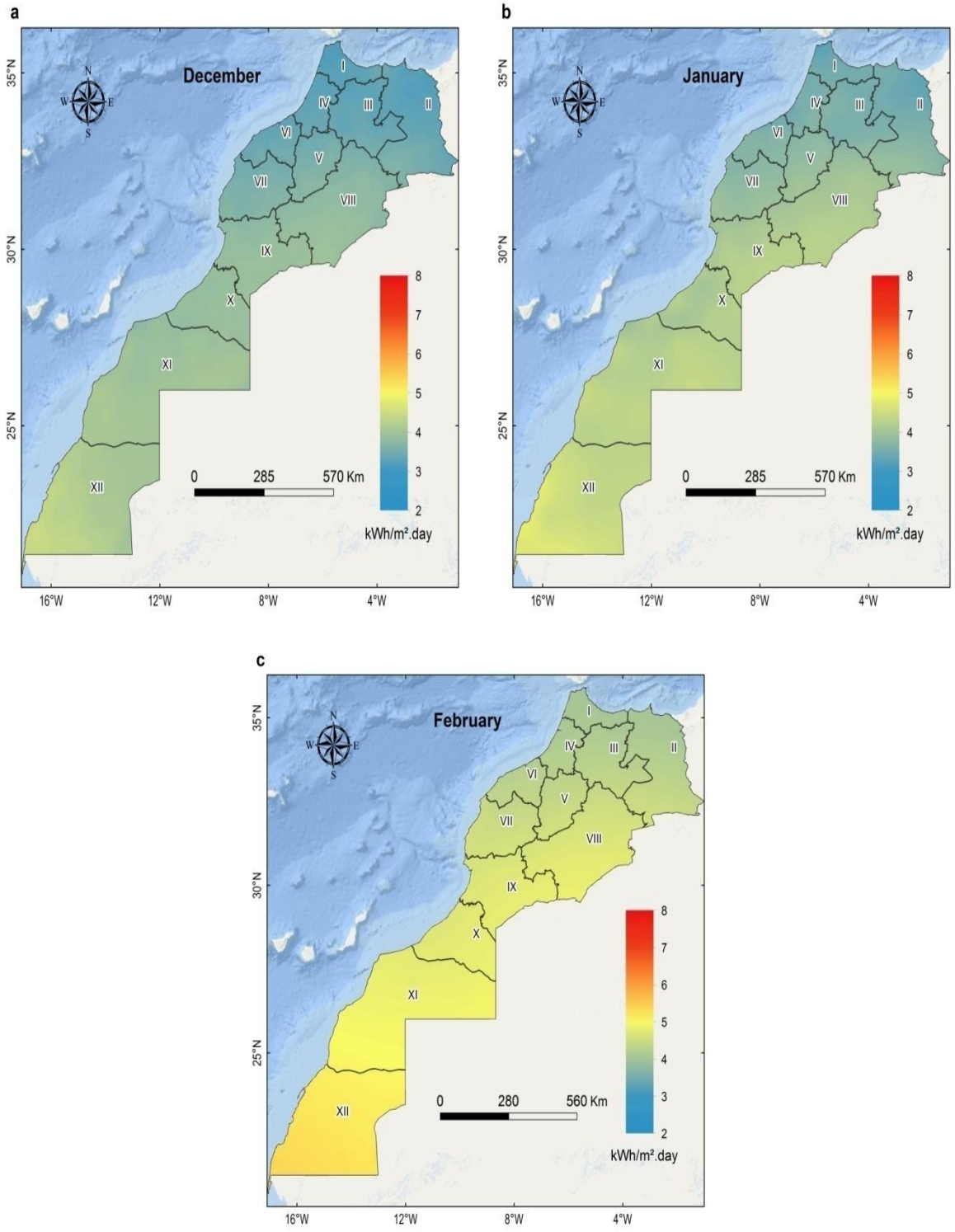


Fig 10. Solar irradiation maps for winter months (a) December, (b) January and (c) February.

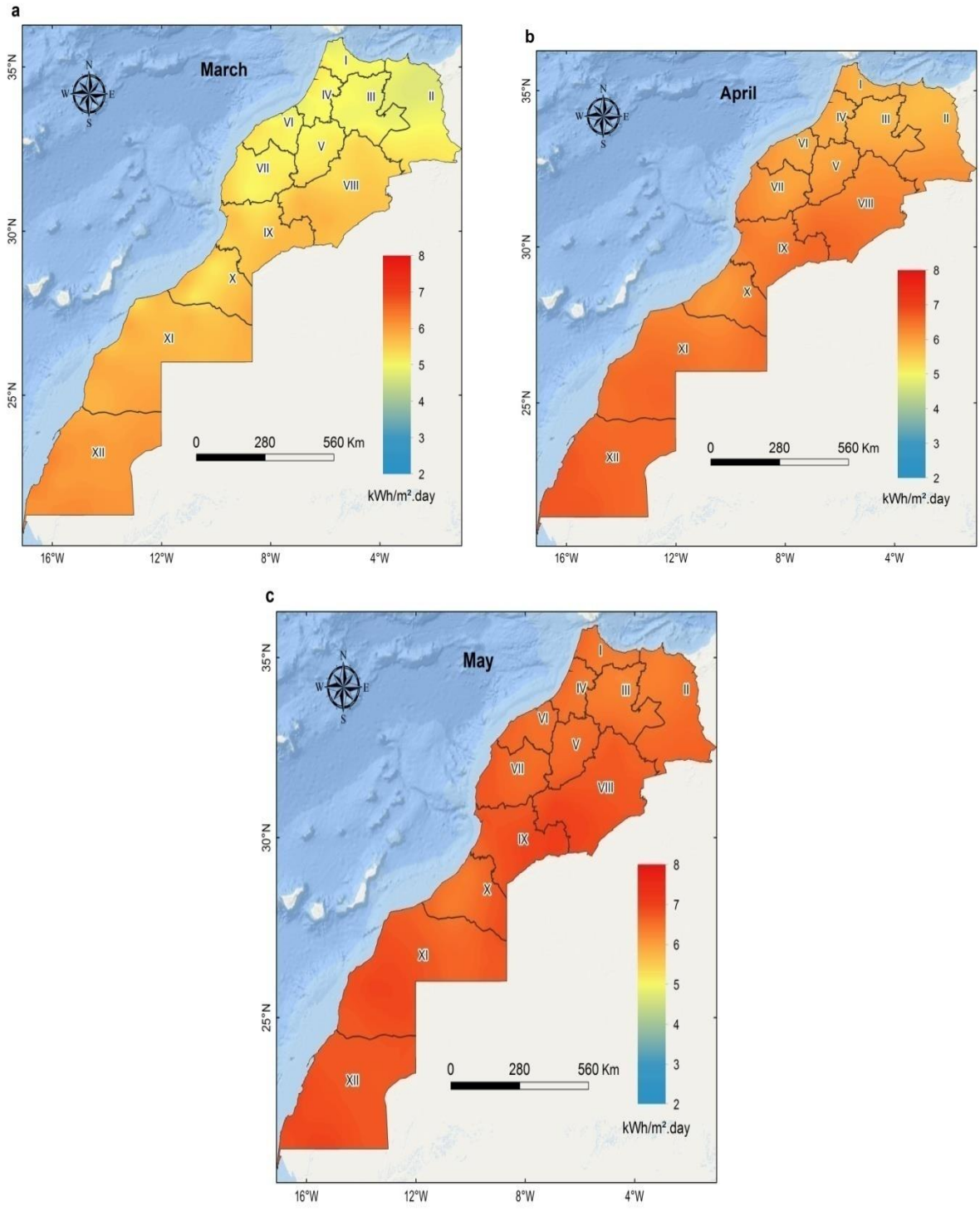


Fig 11. Solar irradiation maps for spring months (a) March, (b) April, and (c) May.

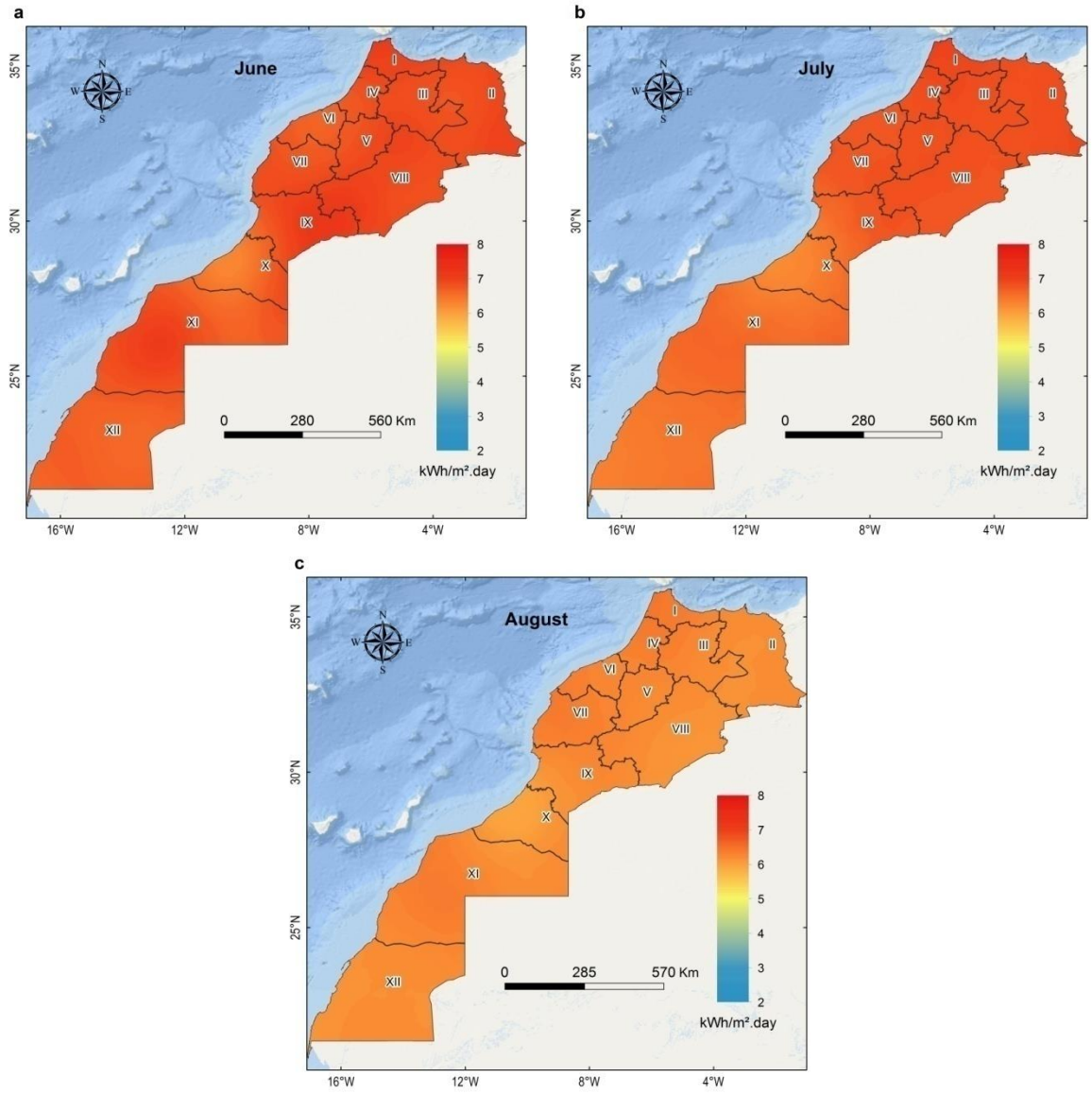


Fig 12. Solar irradiation maps for summer months (a) June, (b) July and (c) August.

In spring, solar irradiation continues to increase;the spatial distribution of thesolar radiation in this season is shown in Fig.11. The lowest values are observed in March at equinox with a minimum value of4.41kWh/m².day, and a maximum value of 6.93kWh/m².dayand always with the same spatial distribution.

As the sun gets higher in the sky, the high latitude regions in the center and the north of the country receivesmore solar radiation. The highest amounts of solar radiation are registered during May with valuesgreater than 6.07kWh/m².day. In this season, the temperatures are high across the country and the relative humidity showsan increase from south to north.

As is indicated in Fig.12, in summer, solar radiation decreases from June to August, and the highest solar potential is shifted to the northern part of the country.Regions I, IV, and IIbecome the most irradiated, while region XII is the less irradiated.The month of June at summer solstice is characterized by the highest values in the whole country ranging between 5.84kWh/m².dayand8kWh/m².day; this is due to the high position of the sun in the sky and the long hours of daylights. However, the month of August is characterized by the lowest solar resource, with a recording of 5.41kWh/m².dayto7.21kWh/m².day. During this season, solar radiation is generally very important acrossMorocco. Summer in the Saharian Desert is characterized by very high temperatures and a clear sky. In the northern and coastalregions, temperatures reach their peaks and the season is considered as the driest of the whole year.

The solar resource is expressed recently in terms of annual solar irradiation classes ranging from class1 (the lowest) to class7 (the highest)[8].This classification is presented in table 4.

Table 4: Classes of solar irradiation potential.

Solar irradiation class	Solar resource potential	Annual solar irradiation(kWh/m².day)
1	Poor	<3.26
2	Marginal	3.26-3.88
3	Fair	3.88-4.49
4	Good	4.49-5
5	Excellent	5-5.57
6	Outstanding Suberb	5.57-6.08
7	Superb	> 6.08

The geographic distribution of annualsolar resource in Morocco is shown in Fig.13,the solar potential varies between 4.6kWh/m².dayand6.33kWh/m².day.Based on this map and the solarresource classification,it can be noticed that the country has a high potential for solar energy,with more than 96 % of its total area is covered by excellent and outstanding classes.

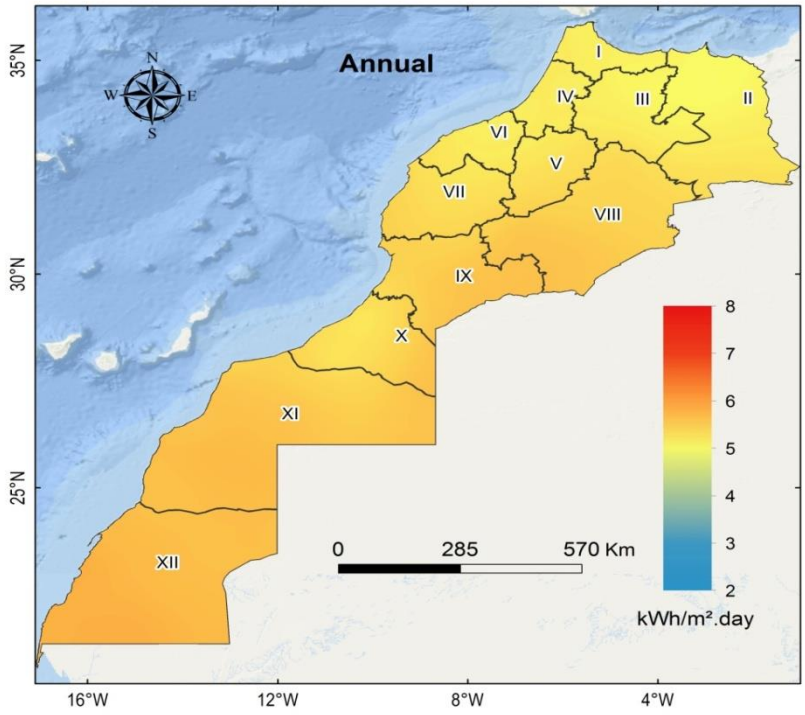


Fig 13. Annual solar irradiation map.

The ranking of the twelve regions according to their mean solar irradiation is displayed in Fig.14. It is clear from this figure that all regions have a solar resource that exceeds 5 kWh/m².day.Additionally, region XII has the most intense solar resource with all its area corresponds to the outstanding class, this class is also dominant in regions XI (90%) and X (73%) and represents (54%) and (23%) of the total areas of the regions VIII and X, respectively. Besides, all other regions correspond to the excellent potential, except for some areaswith good potential in region II (14% of the region).

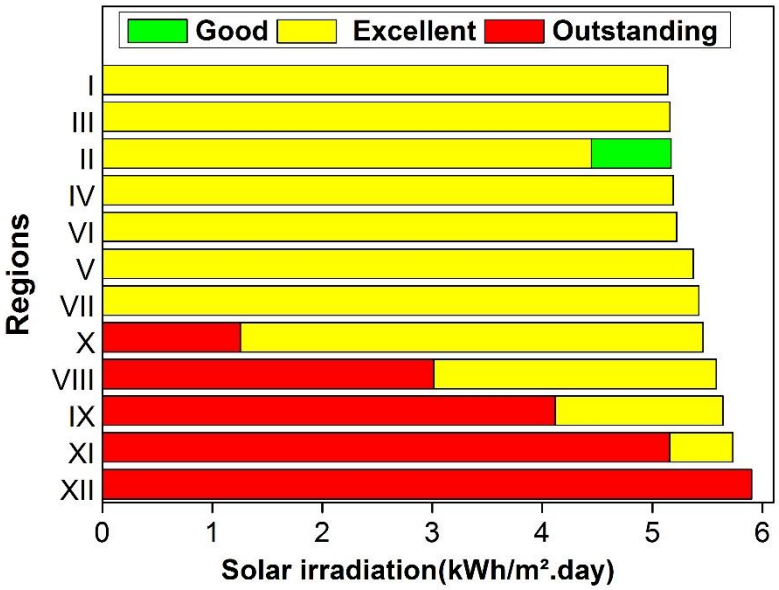


Fig 14. Ranking of regions according to theirsolar irradiation potential.

The monthly and annual solaratlasesobtained by the OK method are in agreement with othermaps constructed with other databases and spanning different periods,like high-resolution maps obtained from SolarGIS company[43]covering the period 1994-2014,and those published by Ouammi et al[25]with data coming from (CM-SAF)-PVGISdatabase representing the period from 1998 to 2010.Thisconfirms the high stability and non-volatility of Moroccan solar resourcefora long periodwhich is an essential factor for solar applications.In comparison to neighboring countries, the Moroccan solar potential is higher than those of Spain and Portugal which are dominated by the good and the fair classes[8],as well as the solar potential of Tunisia covered by the excellent, good, and fair classes[7]. However, it is lower than the solar potential of Algeria which 90 % of its total land is covered by the excellent, outstanding, and superb classes[44].

Even though the great potentialities of the Moroccan solar resource, the total photovoltaic (PV) capacity installed by the end of 2018 was only (205MW)which is lower than the installed capacity in Algeria(410MW), Portugal (670MW), and Spain (4744 MW)which is one of the world leaders in terms of installed PV technologies[45].At the same time, Morocco is the African leader in CSP(Concentrated solar power) systems with 500 MW installed capacity thanks to sites such as Noor Ouarzazate located in region VIII [45].

This simple and preliminary assessmentof the solar resource in Morocco reveals its immense potentialitiesespecially in regions XII, XI, and IX of the country. These regions are promising locations to host PV solar systems plants. However, some additional factors and constraints must be considered like environmental conditions that affect strongly the performances of PV modulesincluding dust accumulation, air temperatures, wind speed and direction, and relative humidity[46], as well asland use and economic conditions.Studies with Multi-Criteria Decision-Makingare needed in these regions to choose suitable solar sites like those conducted in regions II[4]and VIII[47].

Conclusion

In this study,three machine learning techniquesGP,SVR, and ANN wereused and compared to predict monthly solar irradiation in 86 zones of Morocco using few and simple parameters.The three models had comparable prediction accuracy and can be applied successfully to estimate solar irradiation in any area in Morocco. The GP model provided the best combination of prediction accuracy and stability. However, it exhibited much computational cost. The spatial distribution of the solar potential was assessed and mapped using the OK;the results showed that Morocco has a large solar potential dominated by the excellent and the outstanding classes, particularly in the southern regions XII, XI, and IX.Further evaluation is required using more data collected in ground weather stations and using other geostatistical techniques that consider the effect of topography, as well as using the Multi-Criteria Decision-Making approach to select the suitable sites to host solar projects in the sunniest regions in the country.

Acknowledgments

The authorsare grateful to the NASA and RETScreen websites for providing solar irradiation data. They would also like to thankMohamed Tarik (Paul Scherrer Institute, Switzerland) for his valuable guidance.

References

1. J. Fan et al., 'Empirical and machine learning models for predicting daily global solar radiation from sunshine duration: A review and case study in China', Renew. Sustain. Energy Rev., vol. 100, pp. 186–212, 2019.
2. A. Teke, H. B. Yıldırım, and Ö. Çelik, 'Evaluation and performance comparison of different models for the estimation of solar radiation', Renew. Sustain. Energy Rev., vol. 50, pp. 1097–1107, 2015.
3. T. Kousksou et al., 'Renewable energy potential and national policy directions for sustainable development in Morocco', Renew. Sustain. Energy Rev., vol. 47, pp. 46–57, 2015.
4. A. A. Merrouni, F. E. Elalaoui, A. Mezrhab, A. Mezrhab, and A. Ghennioui, 'Large scale PV sites selection by combining GIS and Analytical Hierarchy Process. Case study: Eastern Morocco', Renew. Energy, vol. 119, pp. 863–873, 2018.
5. J. Polo, 'Solar global horizontal and direct normal irradiation maps in Spain derived from geostationary satellites', J. Atmospheric Sol.-Terr. Phys., vol. 130–131, pp. 81–88, Aug. 2015.
6. L. Olatomiwa, S. Mekhilef, S. Shamshirband, and D. Petković, 'Adaptive neuro-fuzzy approach for solar radiation prediction in Nigeria', Renew. Sustain. Energy Rev., vol. 51, pp. 1784–1791, 2015.
7. M. Chelbi, Y. Gagnon, and J. Waewsak, 'Solar radiation mapping using sunshine duration-based models and interpolation techniques: Application to Tunisia', Energy Convers. Manag., vol. 101, pp. 203–215, 2015.
8. R. Prăvălie, C. Patriche, and G. Bandoc, 'Spatial assessment of solar energy potential at global scale. A geographical approach', J. Clean. Prod., vol. 209, pp. 692–721, 2019.
9. O. Şenkal, 'Solar radiation and precipitable water modeling for Turkey using artificial neural networks', Meteorol. Atmospheric Phys., vol. 127, no. 4, pp. 481–488, 2015.
10. M. Rumbayan, A. Abudureyimu, and K. Nagasaka, 'Mapping of solar energy potential in Indonesia using artificial neural network and geographical information system', Renew. Sustain. Energy Rev., vol. 16, no. 3, pp. 1437–1449, 2012.
11. E. F. Alsina, M. Bortolini, M. Gamberi, and A. Regattieri, 'Artificial neural network optimisation for monthly average daily global solar radiation prediction', Energy Convers. Manag., vol. 120, pp. 320–329, 2016.
12. L. Wang, O. Kisi, M. Zounemat-Kermani, G. A. Salazar, Z. Zhu, and W. Gong, 'Solar radiation prediction using different techniques: model evaluation and comparison', Renew. Sustain. Energy Rev., vol. 61, pp. 384–397, 2016.
13. R. Kumar, R. K. Aggarwal, and J. D. Sharma, 'Comparison of regression and artificial neural network models for estimation of global solar radiations', Renew. Sustain. Energy Rev., vol. 52, pp. 1294–1299, 2015.
14. J.-L. Chen, G.-S. Li, and S.-J. Wu, 'Assessing the potential of support vector machine for estimating daily solar radiation using sunshine duration', Energy Convers. Manag., vol. 75, pp. 311–318, 2013.
15. S. Belaid and A. Mellit, 'Prediction of daily and mean monthly global solar radiation using support vector machine in an arid climate', Energy Convers. Manag., vol. 118, pp. 105–118, 2016.
16. V. H. Quej, J. Almorox, J. A. Arnaldo, and L. Saito, 'ANFIS, SVM and ANN soft-computing techniques to

estimate daily global solar radiation in a warm sub-humid environment', J. Atmospheric Sol.-Terr. Phys., vol. 155, pp. 62–70, Mar. 2017.
17. L. Olatomiwa, S. Mekhilef, S. Shamshirband, and D. Petkovic, 'Potential of support vector regression for solar radiation prediction in Nigeria', Nat. Hazards, vol. 77, no. 2, pp. 1055–1068, 2015.
18. A. Rohani, M. Taki, and M. Abdollahpour, 'A novel soft computing model (Gaussian Process with K-fold cross validation) for daily and monthly solar radiation forecasting (Part: I)', Renew. Energy, vol. 115, pp. 411–422, 2018.
19. L. Zou, L. Wang, A. Lin, H. Zhu, Y. Peng, and Z. Zhao, 'Estimation of global solar radiation using an artificial neural network based on an interpolation technique in southeast China', J. Atmospheric Sol.-Terr. Phys., vol. 146, pp. 110–122, 2016.
20. H. Alsamamra, J. A. Ruiz-Arias, D. Pozo-Vázquez, and J. Tovar-Pescador, 'A comparative study of ordinary and residual kriging techniques for mapping global solar radiation over southern Spain', Agric. For. Meteorol., vol. 149, no. 8, pp. 1343–1357, 2009.
21. A. M. Martín and J. Dominguez, 'Solar Radiation Interpolation', in Solar Resources Mapping, Springer, 2019, pp. 221–242.
22. C. Ertekin and F. Evrendilek, 'Spatio-temporal modeling of global solar radiation dynamics as a function of sunshine duration for Turkey', Agric. For. Meteorol., vol. 145, no. 1–2, pp. 36–47, Jul. 2007.
23. Q. Hernández-Escobedo, E. Rodríguez-García, R. Saldaña-Flores, A. Fernández-García, and F. Manzano-Agugliaro, 'Solar energy resource assessment in Mexican states along the Gulf of Mexico', Renew. Sustain. Energy Rev., vol. 43, pp. 216–238, 2015.
24. J.-K. Park, A. Das, and J.-H. Park, 'A new approach to estimate the spatial distribution of solar radiation using topographic factor and sunshine duration in South Korea', Energy Convers. Manag., vol. 101, pp. 30–39, Sep. 2015.
25. A. Ouammi, D. Zejli, H. Dagdougui, and R. Benchrifa, 'Artificial neural network analysis of Moroccan solar potential', Renew. Sustain. Energy Rev., vol. 16, no. 7, pp. 4876–4889, Sep. 2012.
26. Y. El Mghouchi, T. Ajzoul, and A. El Bouardi, 'Prediction of daily solar radiation intensity by day of the year in twenty-four cities of Morocco', Renew. Sustain. Energy Rev., vol. 53, pp. 823–831, Jan. 2016.
27. 'Accueil | Maroc Météo'. [Online]. Available: http://www.marocmeteo.ma/. [Accessed: 21-Jan-2021].
28. N. R. Canada, 'RETScreen', 10-Mar-2010. [Online]. Available: https://www.nrcan.gc.ca/energy/software-tools/7465. [Accessed: 18-Mar-2019].
29. 'POWER Data Access Viewer'. [Online]. Available: https://power.larc.nasa.gov/data-access-viewer/. [Accessed: 18-Mar-2021].
30. K. L. Priddy and P. E. Keller, Artificial Neural Networks: An Introduction. SPIE Press, 2005.
31. I. Estiati, F. B. Freire, J. T. Freire, R. Aguado, and M. Olazar, 'Fitting performance of artificial neural networks and empirical correlations to estimate higher heating values of biomass', Fuel, vol. 180, pp. 377–383, 2016.
32. M. Paulescu, E. Paulescu, P. Gravila, and V. Badescu, Weather modeling and forecasting of PV systems operation. Springer Science & Business Media, 2012.
33. A. Sözen, E. Arcaklioğlu, M. Özalp, and E. G. Kanit, 'Use of artificial neural networks for mapping of solar potential in Turkey', Appl. Energy, vol. 77, no. 3, pp. 273–286, 2004.
34. C. E. Rasmussen, 'Gaussian processes in machine learning', in Summer School on Machine Learning, 2003, pp. 63–71.
35. J. Q. Shi and T. Choi, Gaussian Process analysis for functional data. Chapman and Hall/CRC, 2011.
36. T. T. Joy, S. Rana, S. Gupta, and S. Venkatesh, 'A flexible transfer learning framework for Bayesian optimization with convergence guarantee', Expert Syst. Appl., vol. 115, pp. 656–672, 2019.
37. M. Abzalov, Applied Mining Geology, vol. 12. Springer, 2016.
38. M. A. Oliver and R. Webster, 'A tutorial guide to geostatistics: Computing and modelling variograms and kriging', CATENA, vol. 113, pp. 56–69, Feb. 2014.
39. S. Rehman and S. G. Ghori, 'Spatial estimation of global solar radiation using geostatistics', Renew. Energy, vol. 21, no. 3, pp. 583–605, 2000.
40. H. K. Elminir, F. F. Areed, and T. S. Elsayed, 'Estimation of solar radiation components incident on Helwan site using neural networks', Sol. Energy, vol. 79, no. 3, pp. 270–279, 2005.
41. R. Benavides, F. Montes, A. Rubio, and K. Osoro, 'Geostatistical modelling of air temperature in a mountainous region of Northern Spain', Agric. For. Meteorol., vol. 146, no. 3–4, pp. 173–188, Oct. 2007.
42. M. A. Hassan, A. Khalil, S. Kaseb, and M. A. Kassem, 'Exploring the potential of tree-based ensemble methods in solar radiation modeling', Appl. Energy, vol. 203, pp. 897–916, 2017.

43. 'Atlas de la Ressource Solaire au Maroc'. [Online]. Available: https://solaratlas.masen.ma/map?c=28.886494:-9.05077:5&s=31.625815:-7.989137. [Accessed: 05-Mar-2021].
44. M. R. Yaiche, A. Bouhanik, S. M. A. Bekkouche, A. Malek, and T. Benouaz, 'Revised solar maps of Algeria based on sunshine duration', Energy Convers. Manag., vol. 82, pp. 114–123, 2014.
45. 'International Renewable Energy Agency (IRENA), Abu Dhabi', International Renewable Energy Agency (IRENA), Abu Dhabi.
46. S. A. Said, G. Hassan, H. M. Walwil, and N. Al-Aqeeli, 'The effect of environmental factors and dust accumulation on photovoltaic modules and dust-accumulation mitigation strategies', Renew. Sustain. Energy Rev., vol. 82, pp. 743–760, 2018.
47. M. Tahri, M. Hakdaoui, and M. Maanan, 'The evaluation of solar farm locations applying Geographic Information System and Multi-Criteria Decision-Making methods: Case study in southern Morocco', Renew. Sustain. Energy Rev., vol. 51, pp. 1354–1362, 2015.