

Research Article

A Model for Hourly Solar Radiation Data Generation from Daily Solar Radiation Data Using a Generalized Regression Artificial Neural Network

Tamer Khatib¹ and Wilfried Elmenreich²

¹Department of Energy Engineering and Environment, An-Najah National University, Nablus, State of Palestine

²Institute of Networked and Embedded Systems, University of Klagenfurt, 9020 Klagenfurt, Austria

Correspondence should be addressed to Tamer Khatib; tamer_khat@hotmail.com

Received 29 June 2015; Accepted 13 September 2015

Academic Editor: Wilfried G. J. H. M. Van Sark

Copyright © 2015 T. Khatib and W. Elmenreich. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper presents a model for predicting hourly solar radiation data using daily solar radiation averages. The proposed model is a generalized regression artificial neural network. This model has three inputs, namely, mean daily solar radiation, hour angle, and sunset hour angle. The output layer has one node which is mean hourly solar radiation. The training and development of the proposed model are done using MATLAB and 43800 records of hourly global solar radiation. The results show that the proposed model has better prediction accuracy compared to some empirical and statistical models. Two error statistics are used in this research to evaluate the proposed model, namely, mean absolute percentage error and root mean square error. These values for the proposed model are 11.8% and -3.1% , respectively. Finally, the proposed model shows better ability in overcoming the sophisticated nature of the solar radiation data.

1. Introduction

Solar energy is the portion of the sun's energy available at the earth's surface for useful applications, such as raising the temperature of water or exciting electrons in a photovoltaic cell, in addition to supplying energy to natural processes. This energy is free, clean, and abundant in most places throughout the year. Its effective harnessing and use are of importance to the world, especially at a time of high fossil fuel costs and degradation of the atmosphere by the use of fossil fuels. Solar radiation data provide information on how much of the sun's energy strikes a location on the earth's surface during a particular time period. These data are needed for effective research into solar energy utilization [1].

In general, solar radiation that reaches the earth surface is called extraterrestrial solar radiation (above the atmosphere). In the meanwhile, the attenuated solar radiation within the atmosphere is called global solar radiation. Global solar radiation incident on a horizontal surface has two components, namely, direct (beam) and diffuse solar radiation. Both components of solar radiation are usually measured by pyranometers, solarimeters, or actinography. Direct (beam)

solar radiation is measured by a pyrheliometer while diffuse solar radiation is measured by placing a shadow band over a pyranometer [1]. In addition, solar radiation can be modeled using different techniques.

Many models of solar radiation were presented in the literature. These methods can be mathematical such as linear and polynomial functions, heuristic methods, fuzzy logic techniques, or other individual methods such as Fourier series and Markov chain. However, recently, artificial intelligence techniques based models such as artificial neural networks (ANNs) were used for solar radiation prediction. According to [1, 2], ANNs were used many times for solar radiation modeling, prediction, and forecasting. Different types of ANNs were utilized for this purpose. Examples for these models are feedback back forward ANN, cascade-forward back propagation ANN, generalized regression ANN, neurofuzzy ANN, and optimized ANN-genetic algorithm. In general, most of the conducted work was done for solar radiation prediction using ground measured meteorological variables such as ambient temperature, sunshine ratio, relative humidity, wind speed, and other solar geometry angles such

as hour angle and angle of declination. The main purpose of the aforementioned models is to generate synthetic solar radiation data at a specific location where there are no measuring devices in order to be utilized in solar energy system design, to restore a solar radiation data set in case of having missing data due to monitoring system outages, or to predict the performance of a solar energy system. In 1990s, ANNs were proposed for predicting monthly or daily solar radiation utilizing monthly or daily meteorological variables due to the availability of such data. However, hourly solar radiation prediction is currently more important in order to optimally design solar energy systems. Hourly solar radiation data can be used to optimally design solar power and thermal systems. By using hourly solar radiation data in the design of solar energy systems, the stochastic nature of the solar radiation is considered. In other words, the reliability of the solar power/thermal systems designed based on hourly solar radiation data is greater than systems designed based on daily or monthly solar radiation profiles [3]. The need for hourly solar radiation data for accurate system's design and control led researchers to utilize hourly meteorological variables for predicting hourly solar radiation. However, there is a big debate regarding the availability of hourly meteorological data such as ambient temperature, relative humidity, and sunshine ratio for this purpose [1]. On the other hand, some of pioneer researchers have proposed empirical equations that can predict hourly solar radiation in terms of daily or monthly solar radiation, hour angle, and sunrise/sunset hour angle. Examples of these models are Liu and Jordan's model [4], Collares-Pereira and Rabel's model [5], Garg and Garg's model [6], Jain's model [7], Baig's model [8], and Kaplanis's model [9, 10]. Proposing these equations has made a big advantage in predicting hourly solar radiation without the need for other meteorological variables. These models are reviewed and discussed in detail in Section 2. Most of these models are either empirical or statistical models that implying complex calculations are required. Therefore, these empirical models can be further enhanced in terms of accuracy and simplicity by utilizing novel learning machine such as generalized artificial neural network (GRNN) where GRNN has been recommended for solar radiation prediction in previous researches according to [1]. There is consequently a need to develop GRNN based models that predict hourly solar radiation using daily or monthly solar radiation without the need for hourly meteorological data. The main objective of this paper is to present a novel model for predicting hourly solar radiation using global solar radiation and other solar angles. This model is developed using a generalized regression artificial neural network and is designed to be more accurate than other models. The proposed model is able to generate hourly solar radiation data from daily solar radiation data at sites where only daily averages of solar radiation are available. These data can be used in optimal sizing of photovoltaic systems. The optimal sizing of such systems requires hourly prediction of system performance for at least one-year time in order to provide optimal sizes of photovoltaic array and storage units, for example. Moreover, such a model can be used to optimally manage photovoltaic based distributed generation (DG) units. The output of DG systems needs to be predicted in

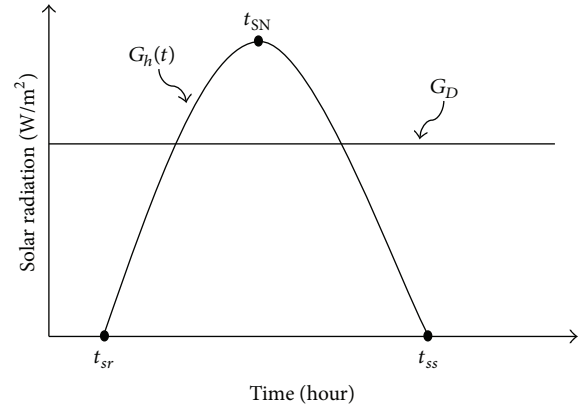


FIGURE 1: Predicting of hourly solar radiation using statistical model.

order to optimally operate the penetrated power system in terms of optimal power flow and system's stability, protection, and power quality. This work is done utilizing solar radiation data for Sohar city, Oman. The city has a deserts climate and it is located on Gulf of Oman with latitude of 24.34 N, longitude of 56.73 E, and elevation of 13 ft. The utilized solar radiation data are measured at Sohar University Weather Station.

2. Hourly Solar Radiation Data Mining

Data mining (knowledge discovery in databases) is the process that attempts to discover patterns in large data sets. Based on this, mean hourly solar radiation data mining is the process that attempts to estimate, predict, or obtain mean hourly solar radiation from a solar radiation data set. This solar radiation data set ideally contains measurements such as mean daily solar radiation and solar angles such as hour angle, sunset angle, and angle of declination. The importance of mean hourly solar radiation data mining is to obtain these data for sites that have only mean daily solar radiation. Mean hourly data represents considerable more information and therefore is more useful for the already mentioned applications. Figure 1 shows a typical profile of mean hourly solar radiation versus time. The mean daily solar radiation is indicated here as a horizontal line.

2.1. Empirical Models for Calculating Mean Hourly Solar Radiation. According to the literature, there are some empirical models developed for calculating hourly solar radiation from daily solar radiation. In [4], Liu and Jordan proposed the following to calculate hourly solar radiation:

$$\frac{G_h}{G_D} = \frac{(\pi/24) (\cos \omega - \cos \omega_s)}{\sin \omega_s - (2\pi\omega_s/360) \cos \omega_s}, \quad (1)$$

where G_h is the mean hourly solar radiation, G_D is the mean daily solar radiation, ω is the hour angle, and ω_s is the sunset hour angle.

The hour angle (ω) is the angular displacement of the sun from the local point and it is given by the following:

$$\omega = 15 (\text{AST} - 12 \text{ hour}), \quad (2)$$

where AST is the apparent or true solar time and it is given by the daily apparent motion of the true, or observed, sun. AST is based on the apparent solar day, which is the interval between two successive returns of the sun to the local meridian. Apparent solar time can be calculated as follows:

$$\text{AST} = \text{LST} + \text{EoT} + \left(4 \frac{\text{min}}{\text{degree}}\right) [(\text{LSMT} - \text{LOD})], \quad (3)$$

where LST is the local standard time, LOD is the longitude, LSMT is the local standard meridian time, and EoT is the equation of time.

The local standard meridian (LSMT) is a reference meridian used for a particular time zone and is similar to the prime meridian, used for Greenwich Mean Time. LSMT is given by the following:

$$\text{LSMT} = 15^\circ * \text{Time zone in GMT}. \quad (4)$$

In the meanwhile, the equation of time (EoT) is the difference between apparent and mean solar times, both taken at a given longitude at the same real instant of time. EoT is given by the following:

$$\text{EoT} = 9.87 \sin(2B) - 7.53 \cos(B) - 1.5 \sin(B), \quad (5)$$

where B is a factor and it can be calculated by

$$B = \frac{360^\circ}{365} (N - 81), \quad (6)$$

where N is the day number and it is defined as the number of days elapsed in a given year up to a particular date (e.g., 2nd February corresponds to 33).

The sunset hour angle can be calculated using the following:

$$\omega_{ss, sr} = \cos^{-1}(-\tan \phi \tan \delta), \quad (7)$$

where ϕ is the latitude and δ is the angle of declination. The angle of declination is the angle between the Earth-sun vector and the equatorial plane and it is calculated as follows:

$$\delta = 23.45 \sin \left[\frac{360(284 + N)}{365} \right]. \quad (8)$$

On the other hand, Collares-Pereira and Rabel verified the previous model in [5] and propose the following for calculating mean hourly solar radiation:

$$\frac{G_h}{G_D} = (a + b \cos \omega) \frac{(\pi/24) (\cos \omega - \cos \omega_s)}{\sin \omega_s - (2\pi\omega_s/360) \cos \omega_s}, \quad (9)$$

where the coefficients a and b are defined as follows:

$$\begin{aligned} a &= 0.409 + 0.5016 \sin(\omega_s - 60), \\ b &= 0.6609 - 0.4767 \sin(\omega_s - 60). \end{aligned} \quad (10)$$

In addition to that, H. P. Garg and S. N. Garg checked the adequacy of the Liu-Jordan correlation in [6] to estimate

the hourly horizontal global radiation for various Indian stations as follows:

$$\begin{aligned} \frac{G_h}{G_D} &= \frac{(\pi/24) (\cos \omega - \cos \omega_s)}{\sin \omega_s - (2\pi\omega_s/360) \cos \omega_s} \\ &\quad - 0.008 \sin 3(\omega - 0.65). \end{aligned} \quad (11)$$

In addition, Jain in [7] suggested calculating hourly solar radiation as follows:

$$G_h = r_t \cdot G_D, \quad (12)$$

where r_t is a Gaussian function to fit the recorded data. The authors of [7] established the following relation for global irradiation:

$$r_{\text{AST}} = \frac{1}{\sigma \sqrt{2\pi}} \exp \left[-\frac{(\text{AST} - 12)^2}{2\sigma^2} \right], \quad (13)$$

where r_{AST} is the ratio of hourly to daily global solar radiation and σ is a factor that is defined by

$$\sigma = \frac{1}{r_{\text{AST}(\text{AST}=12)} \sqrt{2\pi}}. \quad (14)$$

Later, Baig et al. in [8] have proposed a model that is based on Jain's model in [7]. Baig et al. modified Jain's model to better fit the recorded data during the start and the end periods of a day. In this model, r_{AST} is estimated by the following:

$$\begin{aligned} r_{\text{AST}} &= \frac{1}{\sigma \sqrt{2\pi}} \left(\exp \left[-\frac{(\text{AST} - 12)^2}{2\sigma^2} \right] \right. \\ &\quad \left. + \cos \left[180 \frac{(\text{AST} - 12)}{(S_o - 1)} \right] \right). \end{aligned} \quad (15)$$

Another model was proposed by Kaplanis in [9]. The authors assumed that daily solar radiation profile can be described as follows:

$$I = a + b \cdot \cos \left(\frac{2\pi t}{24} \right), \quad (16)$$

where a and b are parameters to be determined for any site and for any day. The followed methodology was represented by integrating (16) over t , from sunrise (t_{sr}) to sunset (t_{ss}) as below:

$$\int_{t_{sr}}^{t_{ss}} G_h dt = 2a(t_{sr} - 12) + \frac{24b}{\pi} \sin \left(\frac{2\pi t_{ss}}{24} \right). \quad (17)$$

Later, Kaplanis has proposed two improvements for his model in [10]. In [10], a statistical model for calculating hourly solar radiation in terms of the variation of the daily solar attention as well as air mass is done. The proposed statistical model is based on three constants which need to be calculated based on actual data. Similarly in [10], a stochastic prediction model for hourly solar radiation is presented. The model calculates hourly solar radiation in terms of the average global solar radiation and the standard deviation of the hourly solar

radiation from the average daily solar radiation. The model is developed based on historical data. Here, also the models presented in [10] are only valid for the case study. Here, also the models presented in [10] are location dependent.

In addition to that, in [11], the authors presented a correlation between G_h and G_D according to Figure 1 as follows:

$$\frac{G_h}{G_D} = a + bt + ct^2, \quad (18)$$

where a , b , and c are coefficients that can be determined by any curve fitting tool. However, the drawback of this model is that it is a location dependent model whereas such a type of models is devoted to a specific region. This is because the coefficients a , b , and c are calculated based on a specific solar radiation profile.

In addition, in [12], statistical model for calculating hourly global solar radiation on horizontal surface was developed. This model represents the hourly solar radiation as a function of extraterrestrial solar radiation as well as a sky transmission function. The proposed sky transmission function is presented as two transmission functions indicating the daily and the hourly variation. At this point, the hourly and the daily transmission variation functions are estimated as statistical relations in terms of day number, hour of the day, and location latitude and longitude based on ground measurements of environmental parameters for a specific location. After all, the inputs of the model were the hour of day, day number, optimized sky transmission function, solar constant, and location coordination. The main drawback of this work is that the relations proposed for the daily and hourly transmission functions are location dependent. Moreover, the utilized extraterrestrial solar radiation data are based on satellite measurements which might not be accurate. More statistical methods were provided. In [13], a model for generating hourly solar radiation as a function of the clearness index is proposed. The development of the model is based on an assumption that the relation between clearness index values and solar radiation can be described by a Gaussian function. In the meanwhile, in [14], the authors proposed a trigonometrical function for predicting daily solar radiation values from monthly solar radiation values.

2.2. Proposed Generalized Regression Artificial Neural Network Model. Artificial neural networks, ANNs, are nonalgorithmic and intensely parallel information processing systems. They learn the relationship between input and output variables by mastering previously recorded data. An ANN usually consists of parallel elemental units called neurons. Neurons are connected by a large number of weighted links which pass signals or information. A neuron receives and combines inputs and then generates the final results in a nonlinear operation. The term ANN usually refers to a Multilayer Perceptron (MLP) Network; however, there are many other types of neural networks, including Probabilistic Neural Networks (PNNs), General Regression Neural Networks (GRNNs), Radial Basis Function (RBF) Networks, Cascade Correlation, Functional Link Networks, Kohonen networks, Gram-Charlier networks, Learning Vector Quantization, Hebb

networks, Adaline networks, Heteroassociative networks, Recurrent Networks, and Hybrid Networks [15].

ANNs have recently been used to predict the amount of solar radiation based on meteorological variables such as sunshine ratio, temperature, and humidity [1]. However, up to now, no one has used ANNs to find the correlation between mean hourly solar radiation and mean daily solar radiation. Therefore, in this paper, a generalized regression artificial neural network (GRNN) is proposed for this purpose. The GRNN is the most recommended type of ANN for solar radiation prediction according to Khatib et al. in [16]. The generalized regression neural network (GRNN) is a probabilistic based network. This network makes classification where the target variable is definite, and GRNNs make regression where the target variable is continuous. GRNN falls into the category of PNNs. This neural network, like other PNNs, needs only a fraction of the training samples an MLP would need. The additional knowledge needed to obtain the fit in a satisfying way is relatively small and can be done without additional input by a user. This makes GRNNs a useful tool for performing prediction and comparison of system performance in practice. The probability density function used in GRNNs is the normal distribution. Each training sample, X_j , is used as the mean of a normal distribution function given by the following:

$$y(x) = \frac{\sum_{i=1}^n Y_i \exp(-D_i^2/2\sigma^2)}{\sum_{i=1}^n \exp(-D_i^2/2\sigma^2)} \quad (19)$$

$$D_i^2 = (X - X_i)^T \cdot (X - X_i).$$

D_i is the distance between the training sample and the point of prediction; it is used as a measure of how well each training sample represents the position of prediction, X . If the distance, D_i , between the training sample and the point of prediction is small, $\exp(-D_i^2/2\sigma^2)$ becomes larger. For $D_i = 0$, $\exp(-D_i^2/2\sigma^2)$ becomes 1.0 and the point of evaluation is represented best by this training sample. A larger distance to all the other training samples causes the term $\exp(-D_i^2/2\sigma^2)$ to become smaller and therefore the contribution of the other training samples to the prediction is relatively small. The term $Y_i * \exp(-D_i^2/2\sigma^2)$ for the i th training sample is the largest and contributes strongly to the prediction. The standard deviation or smoothness parameter, s , is subjected to a search. For a large smoothness parameter, the possible representation of the point of evaluation by the training sample is possible for a wider range of X . For a small smoothness parameter, the representation is limited to a narrow range of X .

GRNNs consist of input, hidden, and output layers. The input layer has one neuron for each predictor variable. The input neurons standardize the range of values by subtracting the median and dividing by the interquartile range. The input neurons then feed the values to each of the neurons in the hidden layer. In the hidden layer, there is one neuron for each case in the training data set. The neuron stores the values of the predictor variables for each case, along with the target value. When presented with a vector of input values from

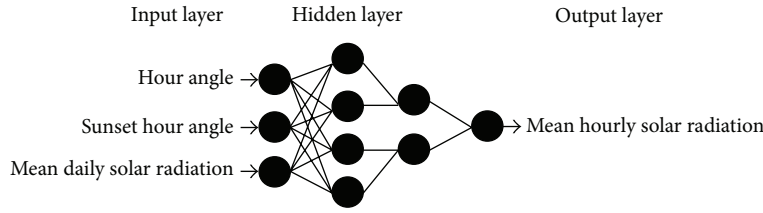


FIGURE 2: Proposed model for hourly solar radiation prediction.

the input layer, a hidden neuron computes the Euclidean distance of the test case from the neuron's center point and then applies the RBF kernel function using the sigma value. The resulting value is passed to the neurons in the pattern layer. However, the pattern (summation) layer has two neurons: one is the denominator summation unit and the other is the numerator summation unit. The denominator summation unit adds the weights of the values coming from each of the hidden neurons. The numerator summation unit adds the weights of the values multiplied by the target value for each hidden neuron. The decision layer divides the value accumulated in the numerator summation unit by the value in the denominator summation unit and uses the result as the predicted target value [15].

Based on the previous models ((1), (9), and (11)), it is clear that the hourly solar radiation value is a function of parameters such as mean daily solar radiation, hour angle, and sunset/sunrise hour angle. Based on this, the GRNN illustrated in Figure 2 is proposed for estimating mean hourly solar radiation. The input layer of the network has three inputs: mean daily solar radiation, hour angle, sunset hour angle. Meanwhile, the output layer has one node which is mean hourly solar radiation.

With regard to the number of neurons in the hidden layer, there is no way to determine the optimal number of hidden neurons without training several networks and estimating the generalization error of each. A low number of hidden neurons cause high training and generalization error due to underfitting and high statistical bias, but a large number of hidden neurons cause high generalization error due to overfitting and high variance [17]. There are some rules of thumb for choosing the number of the hidden nodes. Blum claims in [18] that the number of neurons in the hidden layer ought to be somewhere between the input layer size and the output layer size. Swinger in [19] and Berry and Linoff in [20] claim that the hidden layer will never require more than twice the number of the inputs. In addition, Boger and Guterman suggest in [21] that the number of hidden nodes should be 70–90% of the number of input nodes. Additionally, Caudill and Butler recommend in [15] that the number of hidden nodes equals the number of inputs plus the number of outputs multiplied by (2/3). Based on these recommendations, the number of neurons in the hidden layer of our model should be between 2 and 4. In this research, we used 4 hidden nodes.

2.3. Model Evaluation Criteria. To evaluate the proposed GRNN model and the other models, two statistics errors are used: mean absolute percentage error (MAPE) and root mean

square error (RMSE). MAPE is an indicator of accuracy. MAPE usually expresses accuracy as a percentage and is defined by the following formula:

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \frac{I - I_p}{I}, \quad (20)$$

where I is the measured value and I_p is the predicted value. The resultant of this calculation is summed for every fitted or forecasted point in time and divided again by the number of fitted points, n . This formula gives a percentage error, so one can compare the error of fitted time series that differ in level.

In addition, prediction models were evaluated using RMSE. RMSE provides information about the short-term performance of the models and is a measure of the variation of the predicted values around the measured data. RMSE indicates the scattering of data around linear lines. Moreover, RMSE shows the efficiency of the developed network in predicting future individual values. A large positive RMSE implies a large deviation in the predicted value from the measured value. RMSE can be calculated as follows:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (I_{p_i} - I_i)^2}, \quad (21)$$

where I_{p_i} is the predicted value, I_i is the measured value, and n is the number of observations [1].

2.3.1. Sensitivity Analysis. For any prediction model that deals with stochastic data such as solar radiation, the sensitivity analysis is important for many reasons such as testing the robustness of the results in the presence of uncertainty. Moreover, such an analysis can provide better understanding of the relation between the model's output(s) and input(s). This understating may lead to a model enhancement by identifying the model's input(s) which case significant uncertainty in the output. According to [22–25], the sensitivity analysis can be defined as the study of how the uncertainty in the output of a mathematical model or system can be apportioned to different sources of uncertainty in its inputs. One of the popular methods is automated differentiation method, where the sensitivity parameters are found by simply taking the derivatives of the output with respect to the input.

In this research, the hourly solar radiation can be described as a function of daily solar radiation and sunrise or sunset hour angle and hour as follows:

$$G_h = f(G_D, \omega_{ss, sr}, \omega). \quad (22)$$

However, following the models presented in (1) and (9), the presented model can be described as below:

$$r_h = \frac{G_h}{G_D} = f(\omega_{ss, sr}, \omega). \quad (23)$$

Assume that we are calculating the sensitivity of species r_h with respect to every parameter in the model in (23). Thus, it is required to calculate the time-dependent derivatives as follows:

$$\frac{\partial r_h}{\partial \omega_{ss, sr}}, \frac{\partial r_h}{\partial \omega}. \quad (24)$$

There are different methods to estimate the value of each deferential part. One of these methods is Taylor method. This method states that Taylor approximation of y around a given point (x_o, y_o) —which stands for the first derivative—can be given by the following:

$$r_{h1} = r_{ho} + \frac{\partial r_h}{\partial \omega_{ss, sr}} (\omega_{ss, sr1} - \omega_{ss, sr_o}). \quad (25)$$

Anyway, nowadays such a problem can be also solved by many kinds of software such as MATLAB using functions such as *ParameterInputFactors* and *SensitivityAnalysis*. However, in this research, we used simpler and popular method as well which is the scatter plots method. This method is represented by scatter plots of the output against input(s) individually. This method gives a direct visual indication of sensitivity. Moreover, quantitative measures can also be provided by measuring the correlation between the output and each input.

3. Results and Discussion

In this research, the utilized solar radiation data were measured using rugged solar radiation transmitter (model: WE300, sensor size: 7.6 cm diameter. × 3.8 cm long). The detector of this sensor is high-stability silicon photovoltaic (blue enhanced). Meanwhile, the output range of this sensor is 4 to 20 mA and the measuring range is 0 to 1500 W/m² and the spectral response is in the range of 400 to 1100 nm. The accuracy of this sensor is ±1% full scale with warming up time up to 3 seconds. The operating voltage of this sensor is in the range of 10 to 36 V_{DC}. Using this sensor, solar radiation data are measured and recorded every 5 minutes. Then, these data have been converted to hourly averages with 43800 data records. In this research, an hourly solar radiation data set consisting of 43800 records (5 years) is used. 35040 records (4 years) are used in training the proposed GRNN model. Meanwhile, the 5th year data are used to test the developed model. The model development and training were done using MATLAB line code which is provided in the Appendix of this paper.

To test the proposed model, a control data set containing 8760 records of hourly solar radiation and hour angle is used. The average daily solar radiation and the sunset hour angle are calculated based on these records. The calculated daily solar radiation, sunset hour angles, and hour angles are then used

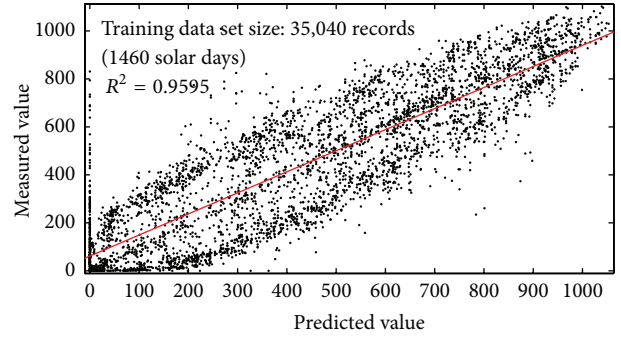


FIGURE 3: Correlation between generated and measured values using the proposed model.

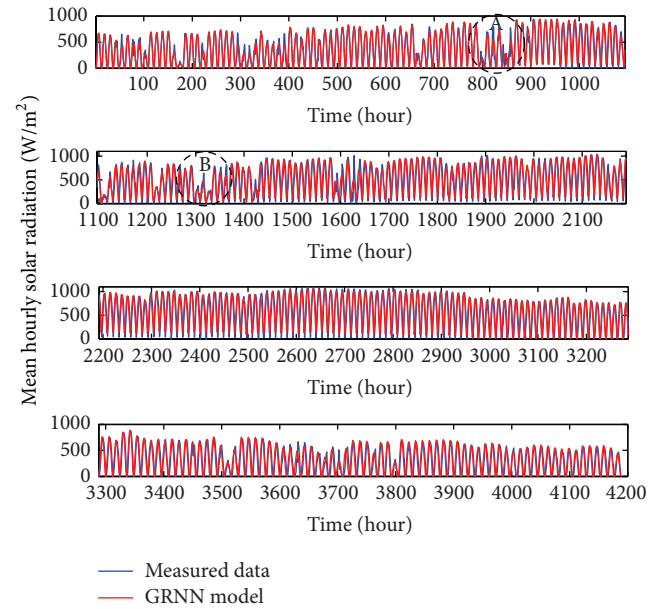


FIGURE 4: Prediction results of the proposed GRNN model (Part A).

as inputs for the proposed GRNN model. Meanwhile, the output of the GRNN model is compared to the actual hourly solar radiation data. To ensure proper evaluation, the control data set is not used in training the proposed GRNN model in order to check the ability of the proposed model for predicting future and foreign data. It is also worth mentioning that the GRNN model has been tested repeatedly (up to 10 times) in order to provide average performance of the proposed model.

Figure 3 shows the correlation between the predicted and the measured data using the proposed GRNN model. From the figure, it is clear that the correlation value is about 96%, which is considerably high. This high correlation value implies that the proposed model makes accurate predictions.

In addition, Figure 4 shows the prediction results for a whole year. From Figure 4, it can be noted that the proposed GRNN model predicts the hourly solar radiation successfully. However, to provide deeper analysis of the proposed GRNN model, two zones, namely, A and B, are chosen from Figure 4 and illustrated more clearly in Figure 5. The selection of these zones is done based on the amount of cloud cover. These

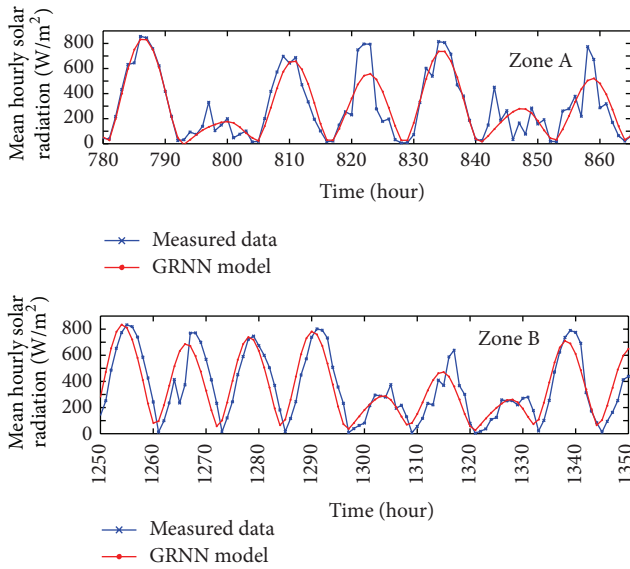


FIGURE 5: Prediction results of the proposed GRNN model (Part B).

zones represent cloudy days where the proposed model prediction accuracy is expected to be low due to unstable solar radiation levels. From Figure 5, the accuracy of the proposed model for predicting the hourly solar radiation is acceptable whereas the generated values of the hourly solar radiation are close to the actual values even on totally overcast days. However, there is a time shift in some cases. This time shift is due to the difference between the calculated hour angle and the read hour angle that the solar radiation is measured at. This problem is usually solved by adding shifting constants to the models. It is assumed that each cycle in Figure 5 represents a whole solar day, where the first cycle is day 1 and the last is day 15. From Figure 5, it can be noticed that, on clear days such as 1, 3, 5, 8, 10, 11, and 15, the prediction is accurate and acceptable. On the other hand, for cloudy days such as 2, 6, 7, 12, 13, and 14, the prediction accuracy is lower than the previous case but still acceptable as most of the solar radiation prediction models' accuracies degraded on cloudy days [1–3].

In order to validate the proposed model, we did two types of comparison; first, a comparison is between the proposed model and some location dependent models. From the conducted literature review, we found two location dependent models which are presented in (16) and (18). The model presented in (16) assumes that the hourly solar radiation can be described by trigonometric functions. In the meanwhile, the model presented in (18) assumes that the averages hourly ratios of hourly solar radiation to daily solar radiation in average can be described by a polynomial function of the second degree. In this research, we used the same data used to train the proposed model in developing these models. Figure 6 shows the development of these models.

In general, both models show good fitting of the average daily hourly r_d ratios with R -square values of 0.9413 for the model presented in (16) and 0.9721 for the model presented in (18). Despite these high R -square values, these models are not expected to predict hourly solar radiation accurately for two

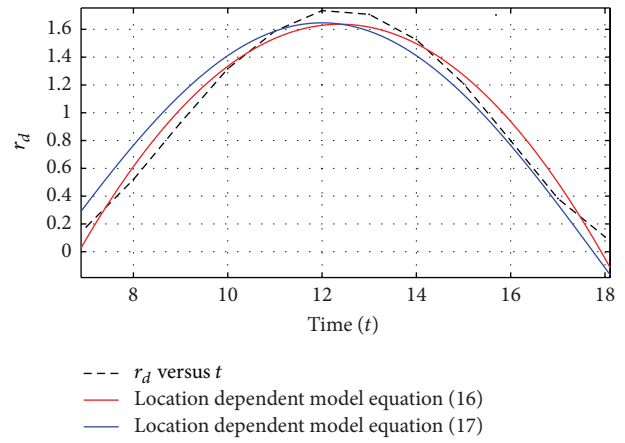


FIGURE 6: Development of two location dependent models.

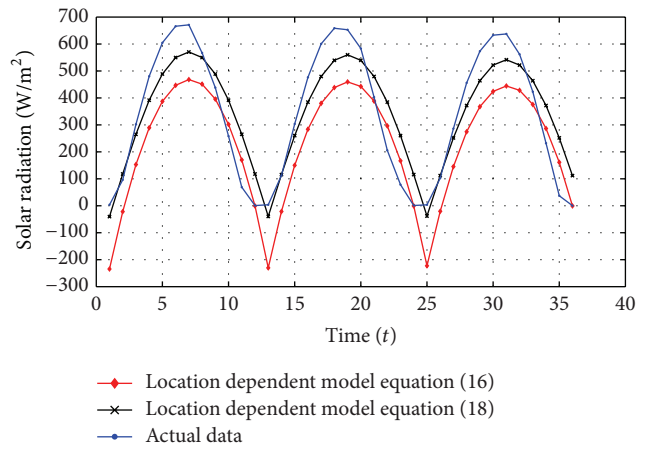


FIGURE 7: Prediction results of location dependent models.

reasons; the assumption of describing the hourly solar radiation by these mathematical functions is not accurate. Secondly, the developed models fit perfectly the average day but they may be unable to fit individual days. Figure 7 shows three days prediction results of these location dependent models.

From Figure 7, it is very clear that the prediction of these random days was inaccurate. Anyway, after ignoring the extreme underestimations of these models in Figure 7, we found that the average MAPE for the model presented in (16) is about 60% while it is about 40% for the model presented in (18).

However, for more fair comparison, the proposed model is compared with more accurate empirical models. A comparison between the proposed model and Liu-Jordan and Collares-Pereira models is conducted in this research. Figure 8 shows a sample of the comparison conducted for 8 solar days. From Figure 8, it is clear that the three models can predict hourly solar radiation data accurately in clear sky days. However, Liu-Jordan model generated underestimated values sometimes. In addition, the generated curves are slightly shifted in some days which caused overestimated values in the morning and underestimated value in the afternoon. On the other hand, generated profiles using Collares-Pereira model

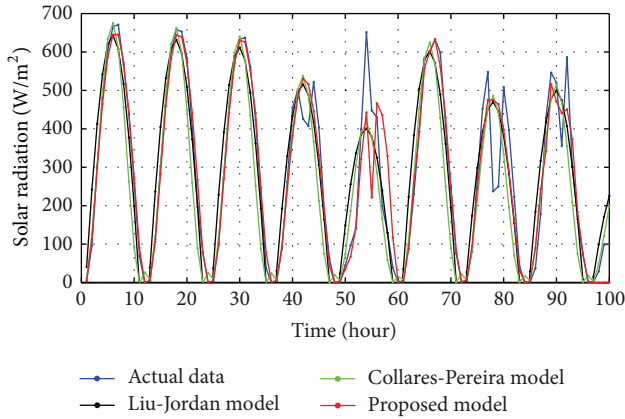


FIGURE 8: Comparison between the proposed model and other models.

are sometimes narrower than the actual one which caused underestimations in the afternoon. As for the proposed model, it can be seen that it is more accurate than the other two models. However, the three models are not able to predict the solar radiation accurately in the cloudy days as compared to their ability in the clear sky days.

Based on the results of the whole testing year, the generated synthetic solar radiation data profiles by the three models seem to be shifted from the actual solar radiation profile in some days. This is due to the difference between the calculated hour angle and the hour angle at which the solar radiation is measured. Therefore, the prediction accuracy of these models can be significantly improved by adding a shifting coefficient to the original models depending on the climate zone. This conclusion has been previously found by the authors of [4] whereas such a practice (adding a shifting coefficient) has been proposed to the original Liu-Jordan model [4]. The authors of [4] discussed the validity of Liu-Jordan model in predicting hourly solar radiation utilizing actual data for different sites with close latitude values. The authors concluded that, for hourly diffuse solar radiation prediction, the model presented by Erdinc and Uzunoglu in [3] performed accurately as the ratio of hourly diffuse to daily diffuse radiation is insensitive to the shade ring correction. On the other hand, this model showed more inaccuracy when predicting hourly global solar radiation because of the increase of beam radiation with incidence angle caused by the atmospheric attenuation. Therefore, the authors of [4] decided to combine the new data they have used with corresponding points from [3] in which each data base is weighted according to the number of years of observation. This fitting process resulted in a correlation term added to the model presented by [3]. The proposed model by [4] was tested using data for countries with different latitudes and found relatively accurate. However, the authors suggested that latitude independence is a good correlation practice for improving the prediction accuracy of the model presented by Erdinc and Uzunoglu in [3].

Table 1 shows an evaluation of the three models using MAPE and RMSE. It is clear that the proposed model has the best accuracy prediction whereas it exceeds the other models

TABLE 1: Evaluation of the proposed GRNN model.

Model	MAPE	RMSE (W/m^2)	RMSE (%)	R^2
Liu-Jordan	27.4%	64.7	22.1	80.3
Collares-Pereira	22.5%	59.1	20.7	85.7
GRNN (training)	11.4%	45.0	14.8	98.3
GRNN (testing)	11.8%	46.1	15.1	96.0

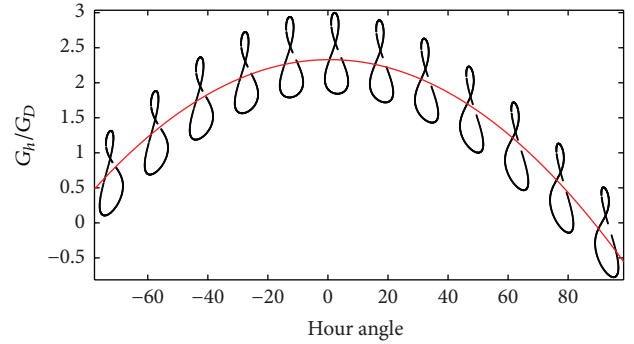


FIGURE 9: ω versus r_h for Liu-Jordan model.

by the MAPE and RMSE. This implies that the proposed model is more powerful in predicting hourly solar radiation according to the MAPE value. Moreover, it has the ability to predict future data based on the RMSE value.

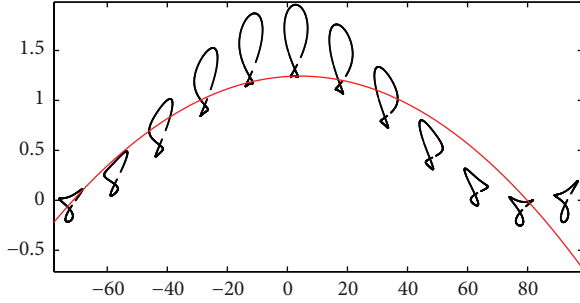
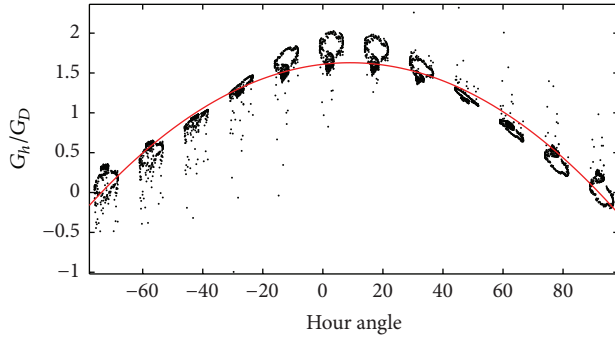
As for the sensitivity analysis, in this research, the proposed model as well as Liu-Jordan and Collares-Pereira models is tested in terms of sensitivity by plotting scatter plots for each model with respect to hour angle factor (ω versus r_h). Then, the correlation value for each data set is provided. Figures 9–11 show the scatter plots of the three models as described in (23).

From Figures 9–11, it can be seen that the proposed model is more efficient in considering the uncertainty of the solar radiation for this case. In Figure 9, the values of r_h factor are more concentrated around specific points with some deviated extreme points. These extreme points are the unexpected values of solar radiation due to some reasons such as clouds and dust particles. Moreover, there is no symmetric nature of these values which means that the uncertainty problem of such data is relatively overcome by the proposed model. On the other hand, both Liu-Jordan and Collares-Pereira models resulted in symmetric behavior of the data regardless of any external conditions which caused prediction inaccuracy in some cases. Table 2 shows the R -square values of each model.

From Table 2, it is also clear that the proposed model exceeds the other empirical models. In general, the advantage of the proposed model as compared to the empirical models is that it is a machine with the ability of learning and handling huge data sets with stochastic nature. As a fact, heuristic techniques such as GRNN are more efficient in handling stochastic data subject to a prior concrete training. However, the empirical models exceed the proposed model in case of having short historical data that is not enough to train the proposed model.

TABLE 2: Sensitivity analysis of the proposed GRNN model.

Scatter plot	R-square values			Fitting function
	Liu-Jordan	Collares-Pereira	GRNN model	
ω versus G_h/G_D	0.78	0.75	0.90	Polynomial/quadratic

FIGURE 10: ω versus r_h for Collares-Pereira model.FIGURE 11: ω versus r_h for the proposed GRNN model.

Finally, utilizing such a large data set (5 years, 43800 records) is not a must to develop such a model. It is all about utilizing available data to develop accurate model with excellent ability to predict future data. As a matter of fact, there are two important issues to be considered when deciding the size of a training data set for a solar radiation prediction model. These issues are the uncertainty nature of solar radiation and the day number nature of the year. In other words, it is possible to predict hourly solar radiation in January (winter) using a model that is trained based on data for June (summer). However, it will not be accurate as compared to a model that is trained based on the whole year time with small step records data. In order to address this issue, we have trained the proposed model using data sets with three relatively small sizes 84, 348, and 684 records. Figures 12(a), 12(b), 13(a), 13(b), 14(a), and 14(b) show the result of this practice. The first part of Figures 10, 11, and 12 shows the prediction performance of the model by comparing its output to the actual values. It is worth mentioning that in these figures the days where there is no red line during them mean that the model generated values of zero only (the model failed to predict the solar radiation profile). On the other hand, the second part of Figures 12, 13, and 14 shows model accuracy using the correlation factor R^2 .

TABLE 3: Impact of training data set size on model's accuracy.

Training data set size (records)	Av. overall MAPE	R^2
84 (7 solar days)	64.3%	0.1354
348 (29 solar days)	36.8%	0.2263
684 (57 solar days)	24.1%	0.6583
35,040 (1460 solar days)	11.8%	0.9595

From Figures 12(a), 12(b), 13(a), 13(b), 14(a), and 14(b), it is clear that the proposed model did not perform well when it is trained using only 84 records (7 solar days). However, the model was able to predict a number of days accurately according to Figure 12(a). These results are seconded by Figure 12(b) where the correlation value between the generated and the measured values is very low due to the high rate of model shortages (the days where the model generates values of zero only). The mean absolute percentage error of the model that developed based on 84 records is very high where it is 100% in 70% of the generated days. Meanwhile, the prediction accuracy of the other days where the model was able to generate synthetic solar radiation data was fine with MAPE in the range of 10–13%. In general, the average whole MAPE for this model is 64.3% with R^2 value of 0.1354. Here, the empirical models show superior performance as these models do not need any prior training.

On the other hand, according to Figures 13(a) and 13(b), the performance of the proposed model was significantly enhanced when it is trained using 348 records (29 solar days). Meanwhile, Figures 14(a) and 14(b) show that the performance of the model was further improved when 684 records (57 solar days) were used in the training. The model which has been trained based on 348 records was able to work probably during 72% of the testing days, while the 684 records based model was able to work probably during 86% of the testing days. The average whole MAPE values of both models are 36.8% and 24%, respectively. These results can be also concluded from Figures 13(b) and 14(b) where R^2 values are significantly increased. R^2 values for 348 records based model and 684 records based model are 0.2263 and 0.6583, respectively. Finally, by comparing all of these figures to Figure 3, it can be realized that the correlation value is much better than all the previous models and consequently the model is supposed to be more accurate in predicting solar radiation values. Table 3 summaries the aforementioned results.

As a conclusion, as far as the model is trained using more data, the accuracy will be better. However, such a model can be developed using a relatively small data set (about 70 solar days) which is usually not difficult to obtain. As mentioned before, the proposed model has the advantage of avoiding the complex calculation of the empirical and statistical model's

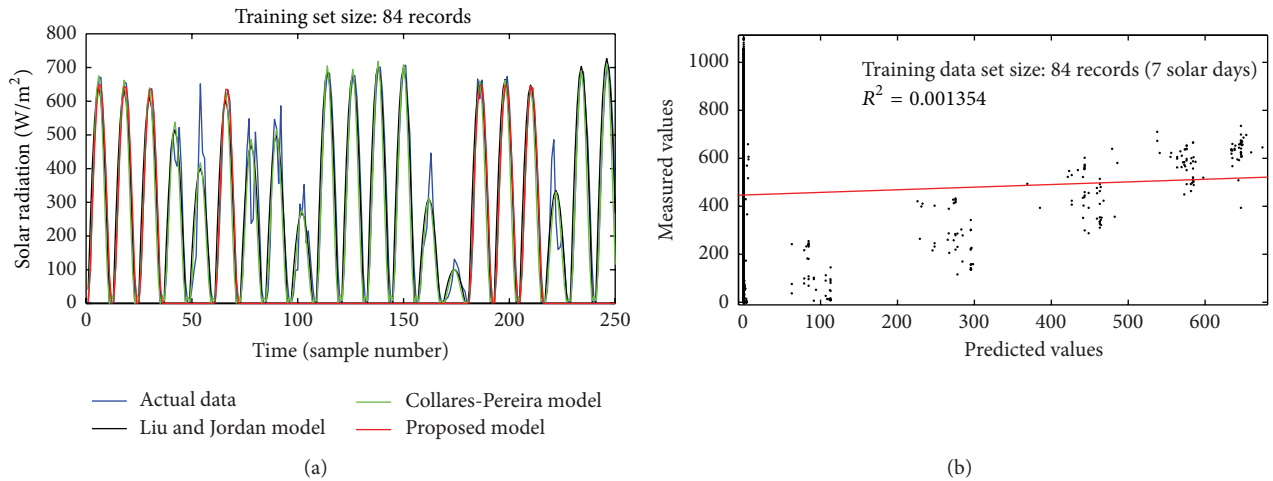


FIGURE 12: (a) Proposed model performance considering small sizes of training data set. (b) Correlation between generated and measured values using training data set size of 7 solar days.

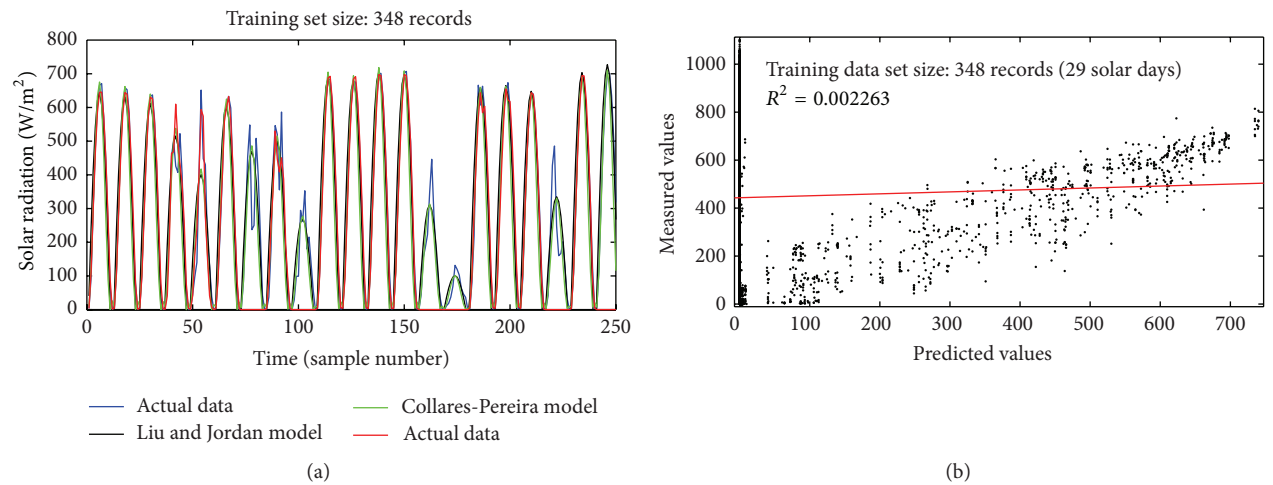


FIGURE 13: (a) Proposed model performance considering small sizes of training data set. (b) Correlation between generated and measured values using training data set size of 29 solar days.

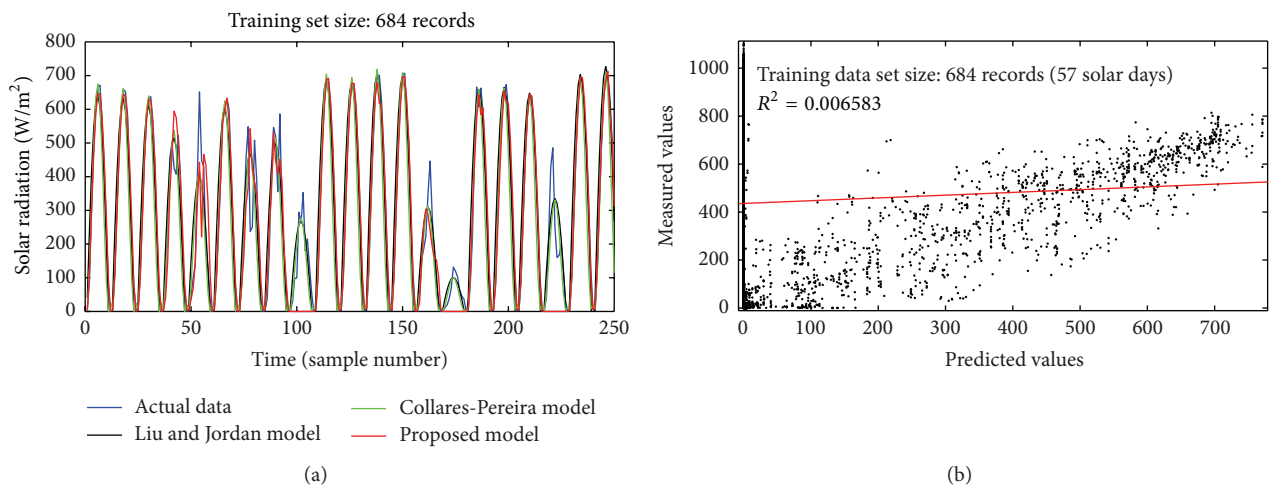


FIGURE 14: (a) Proposed model performance considering small sizes of training data set. (b) Correlation between generated and measured values using training data set size of 57 solar days.

```

% Start
clc
clear
close all
% %-----Input Data-----
fileName='Site.xlsx';
sheetName='hourly Solar radiation data';
L=; % add a latitude value
LOD=; % add a longitude value
% T means testing data
SRi=xlsread(fileName, sheetName, 'C3:C4358'); % hourly solar radiation
Wi=xlsread(fileName, sheetName, 'D3:D4358'); % hour angle
Wsi=xlsread(fileName, sheetName, 'E3:E4358'); % sun set angle
GD=xlsread(fileName, sheetName, 'G3:G4358'); % daily solar radiation
DN=xlsread(fileName, sheetName, 'F3:F4358'); % day number (day & month)
h=xlsread(fileName, sheetName, 'I3:I4358'); % hour
SRi_T=xlsread(fileName, sheetName, 'C4359:C8546');
Wi_T=xlsread(fileName, sheetName, 'D4359:D8546');
Wsi_T=xlsread(fileName, sheetName, 'E4359:E8546');
GD_T=xlsread(fileName, sheetName, 'G4359:G8546');
DN_T=xlsread(fileName, sheetName, 'F4359:F8546');
h_T=xlsread(fileName, sheetName, 'I4359:I8546');
SR=SRi*1000;
SR_T=SRi_T*1000;
W=(Wi_T*(pi/180));
Ws=(abs(Wsi_T*(pi/180)));
%=====Proposed GRNN Model=====
inputs=[L LOD Wi, Wsi, GD DN h];
I=inputs';
targets=SR;
T=targets';
net=newgrnn(I,T);
test=[L LOD Wi_T, Wsi_T, GD_T DN_T h_T];
Test=test';
x_predicted=sim(net,Test);
X_P=x_predicted';
%=====Liu and Jordan Model=====
R_LiuJordan((((pi)/24)*(cos(W)-cos(Ws)))/(sin(Ws)-((pi*Ws)/180).*cos(Ws)));
SR_LiuJordan=R_LiuJordan.*GD_T
%=====Collares-Pereira and Rabel Model=====
a=0.409+(0.5016*sin(Ws-60));
b=0.6609+(0.4767*sin(Ws-60));
R_Collares((a + (b.*cos(W))).((((pi)/24)*(cos(W)-cos(Ws)))/(sin(Ws)-((pi*Ws)/180).*cos(Ws)));
SR_Collares=R_Collares.*GD_T;
%=====Plot data=====
plot (SR_T)
xlim([0 100])
hold on
plot (SR_LiuJordan, 'k')
xlim([0 100])
hold on
plot (SR_Collares, 'g')
xlim([0 35])
hold on
plot (X_P, 'red')
xlim([0 100])
%-----Models evaluation-----

```

```

n=length(X_P);
E3_Hour=SR_T-X_P;
GRNN_MAPE=abs(E3_Hour./SR_T);
GRNN_meanMAPE1=sum(GRNN_MAPE)/n;
GRNN_meanMAPE=GRNN_meanMAPE1*100;
GRNN_RMSE=sqrt(sum((X_P-SR_T).^2/n))
GRNN_MBE=sum(X_P-SR_T)/n
SUM=(sum(X_P)./n);
liner_RMSE_Percentage=(GRNN_RMSE/SUM)*100
liner_MBE_Percentage=(GRNN_MBE/SUM)*100
% End

```

PSEUDOCODE 1

parameters. Moreover, in case of training the proposed model well, the model will be able to handle the uncertainty issue in solar radiation much better than the empirical and statistical models.

4. Conclusion

In this research, a generalized regression artificial neural network based model was presented for predicting hourly solar radiation using daily solar radiation. This model has three inputs, namely, mean daily solar radiation, hour angle, and sunset hour angle. The output layer has one node, which is the generated mean hourly solar radiation. Five years of data for hourly solar radiation were used to train and develop the model running under MATLAB. The results showed that the proposed model has better prediction accuracy as compared to existing empirical and statistical models especially in dealing with special location dependent cases. The average prediction accuracy of the proposed model is about 11% with R -square value of 0.96. Furthermore, the proposed model showed superiority in terms of prediction sensitivity as compared to some empirical models. However, the results showed that the proposed model needed a concrete prior training in order to show prediction superiority. It is also concluded that such a model can be developed with relatively accepted prediction accuracy (24%) using about two months' data with an hourly step. Finally, the proposed model can be used in predicting the performance of solar energy systems such as photovoltaic system and solar water heater. Moreover, it can be used to generate long term data in order to be used in optimal sizing and planning of solar energy systems.

Appendix

See Pseudocode 1.

Nomenclature

ANNs: Artificial neural networks
 AST: Apparent or true solar time
 δ : Angle of declination
 N : Day number
 GD: Distributed generation

EoT: Equation of time
 G_D : Mean daily solar radiation
 G_h : Mean hourly solar radiation
 GRNN: Generalized artificial neural network
 ϕ : The latitude
 LLP: Loss of load probability
 LOD: Longitude
 LSMT: Local standard meridian time
 LST: Local standard time
 MAPE: Mean absolute percentage error
 MLP: Multilayer Perceptron Network
 PNNs: Probabilistic Neural Networks
 PV: Photovoltaic
 RBF: Radial Basis Function Networks
 RMSE: Root mean square error
 r_t : The ratio of hourly to daily global solar radiation
 t_{sr} : Sunrise time
 t_{ss} : Sunset time
 ω : Hour angle
 $\omega_{ss, sr}$: Sunset/sunrise hour angle.

Conflict of Interests

The authors hereby confirm that there is no conflict of interests in the paper with any third part.


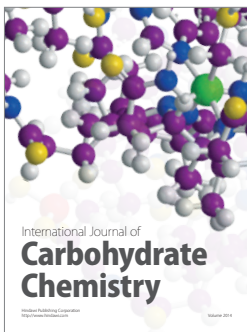
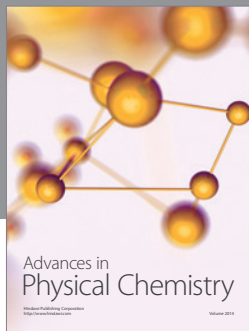
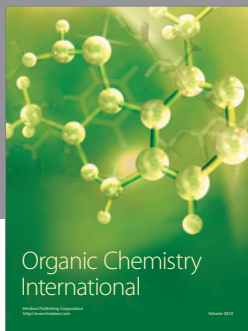
Acknowledgments

This work is supported by Lakeside Labs, Klagenfurt, Austria, and funded by the European Regional Development Fund (ERDF) and the Carinthian Economic Promotion Fund (KWF) under Grant 20214|22935|34445 (Project Smart Microgrid).

References

- [1] T. Khatib, A. Mohamed, and K. Sopian, "A review of solar energy modeling techniques," *Renewable and Sustainable Energy Reviews*, vol. 16, no. 5, pp. 2864–2869, 2012.
- [2] A. Mellit and S. A. Kalogirou, "Artificial intelligence techniques for photovoltaic applications: a review," *Progress in Energy and Combustion Science*, vol. 34, no. 5, pp. 574–632, 2008.

- [3] O. Erdinc and M. Uzunoglu, "Optimum design of hybrid renewable energy systems: overview of different approaches," *Renewable and Sustainable Energy Reviews*, vol. 16, no. 3, pp. 1412–1425, 2012.
- [4] B. Y. H. Liu and R. C. Jordan, "The interrelationship and characteristic distribution of direct, diffuse and total solar radiation," *Solar Energy*, vol. 4, no. 3, pp. 1–19, 1960.
- [5] M. Collares-Pereira and A. Rabl, "The average distribution of solar radiation-correlations between diffuse and hemispherical and between daily and hourly insolation values," *Solar Energy*, vol. 22, no. 2, pp. 155–164, 1979.
- [6] H. P. Garg and S. N. Garg, "Improved correlation of daily and hourly diffuse radiation with global radiation for Indian stations," *Solar & Wind Technology*, vol. 4, no. 2, pp. 113–126, 1987.
- [7] P. C. Jain, "Estimation of monthly average hourly global and diffuse irradiation," *Solar and Wind Technology*, vol. 5, no. 1, pp. 7–14, 1988.
- [8] A. Baig, P. Akhter, and A. Mufti, "A novel approach to estimate the clear day global radiation," *Renewable Energy*, vol. 1, no. 1, pp. 119–123, 1991.
- [9] S. N. Kaplanis, "New methodologies to estimate the hourly global solar radiation: comparisons with existing models," *Renewable Energy*, vol. 31, no. 6, pp. 781–790, 2006.
- [10] S. Kaplanis and E. Kaplani, "A model to predict expected mean and stochastic hourly global solar radiation $I(h;n_j)$ values," *Renewable Energy*, vol. 32, no. 8, pp. 1414–1425, 2007.
- [11] O. P. Singh, S. K. Srivastava, and G. N. Pandey, "Estimation of hourly global solar radiation in the plane areas of Uttar Pradesh, India," *Energy Conversion and Management*, vol. 38, no. 8, pp. 779–785, 1997.
- [12] P. K. Pandey and M. L. Soupir, "A new method to estimate average hourly global solar radiation on the horizontal surface," *Atmospheric Research*, vol. 114–115, pp. 83–90, 2012.
- [13] R. Aguiar and M. Collares-Pereira, "TAG: a time-dependent, autoregressive, Gaussian model for generating synthetic hourly radiation," *Solar Energy*, vol. 49, no. 3, pp. 167–174, 1992.
- [14] R. Festa, C. F. Ratto, and D. DeGol, "A procedure to obtain average daily values of meteorological parameters from monthly averages," *Solar Energy*, vol. 40, no. 4, pp. 309–313, 1988.
- [15] M. Caudill and C. Butler, *Understanding Neural Networks: Computer Explorations*, vol. 1 of *Basic Networks*, The MIT Press, Cambridge, Mass, USA, 1993.
- [16] T. Khatib, A. Mohamed, K. Sopian, and M. Mahmoud, "Assessment of artificial neural networks for hourly solar radiation prediction," *International Journal of Photoenergy*, vol. 2012, Article ID 946890, 7 pages, 2012.
- [17] S. Geman, E. Bienenstock, and R. Doursat, "Neural networks and the bias/variance dilemma," *Neural Computation*, vol. 4, no. 1, pp. 1–58, 1992.
- [18] A. Blum, *Neural Networks in C++*, Wiley, New York, NY, USA, 1992.
- [19] K. Swingler, *Applying Neural Networks: A Practical Guide*, Academic Press, London, UK, 1996.
- [20] M. Berry and G. Linoff, *Data Mining Techniques*, John Wiley & Sons, New York, NY, USA, 1997.
- [21] Z. Boger and H. Guterman, "Knowledge extraction from artificial neural network models," in *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics. Computational Cybernetics and Simulation*, vol. 4, pp. 3030–3035, IEEE, Orlando, Fla, USA, October 1997.
- [22] D. Thevenard and S. Pelland, "Estimating the uncertainty in long-term photovoltaic yield predictions," *Solar Energy*, vol. 91, pp. 432–445, 2013.
- [23] D. P. Finn, D. Connolly, and P. Kenny, "Sensitivity analysis of a maritime located night ventilated library building," *Solar Energy*, vol. 81, no. 6, pp. 697–710, 2007.
- [24] M. U. Siddiqui, A. F. M. Arif, A. M. Bilton, S. Dubowsky, and M. Elshafei, "An improved electric circuit model for photovoltaic modules based on sensitivity analysis," *Solar Energy*, vol. 90, pp. 29–42, 2013.
- [25] X.-G. Zhu, Z.-H. Fu, X.-M. Long, and Xin-Li, "Sensitivity analysis and more accurate solution of photovoltaic solar cell parameters," *Solar Energy*, vol. 85, no. 2, pp. 393–403, 2011.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

