

Prediction of Daily Global Solar Radiation using Resilient-propagation Artificial Neural Network and Historical Data: A Case Study of Hail, Saudi Arabia

Sahbi Boubaker

Department of Electronics Engineering
University of Hail
Hail City, Saudi Arabia
s.boubaker@uoh.edu.sa

Souad Kamel

Department of Computer & Network
Engineering, University of Jeddah
Jeddah, Saudi Arabia
skamel@uj.edu.sa

Mourad Kchaou

Department of Electrical Engineering
University of Hail
Hail, Saudi Arabia
m.kchaou@uoh.edu.sa

Abstract—In this paper, several different Feed Forward Artificial Neural Networks (FFANNs) were used for forecasting the one-day-ahead Global Horizontal Irradiation (GHI) in Hail region, Saudi Arabia. The main motivation behind predicting GHI is that it is a critical parameter in sizing and planning photovoltaic water pumping systems. The novelty of the proposed approach is that it employs only the historical values of the GHI itself as explanatory variables and a fast training algorithm (resilient-propagation). In terms of performance metrics, the rp-trained FFANNs provided better results than Quasi-Newton (bfg) algorithm trained FFANNs for almost all the studied combinations of the FFANN structure. It has been also shown that increasing the number of neurons per layer didn't improve the performance. Medium structures with fast training algorithms are recommended.

Keywords—global horizontal irradiation (GHI); forecasting; feed-forward artificial neural network (FFANN); resilient-propagation (rp) training algorithm

I. INTRODUCTION

Hail region (Saudi Arabia) has a semi-arid climate where solar energy is abundant, while water needs to be pumped from relatively deep wells. Due to the new electricity pricing policy adopted in Saudi Arabia at the beginning of 2018, where the price of one KWh increased from 0.05SAR to 0.18SAR (1USD=3.75SAR), solar energy can constitute an alternative solution replacing classical electricity. Predicting as accurately as possible the future amounts of solar energy is of high importance in designing stand-alone or grid-connected solar plants. Given the fact that solar energy resources measurement stations remain relatively expensive, one solution is the design of computerized numerical forecasters. Moreover, global horizontal irradiation (GHI) is characterized by a high level of variability due to the climatic and geographical factors affecting it. Accordingly, researchers have resorted to mathematical models to describe such complex and uncertain relationships. An increasing interest in more reliable and accurate forecasting approaches has been observed among the solar energy research community during the last few years. Various techniques have been used in the literature to forecast

GHI including time-series (TS) and machine learning (ML). Artificial Neural Networks (ANNs) as flexible tools having high capabilities to map inputs (patterns) to outputs (targets) are popular in the domain of forecasting. However, accuracy is found to be varying from one study to another and to be highly related to the nature and size of the available datasets. The present paper focuses on forecasting the GHI in Hail region using different structures of a Feed-Forward ANN (FFANN) based only on the historical records of the GHI itself. Thus, the GHI is expressed as a symbolic form $GHI(t)=f(GHI(t-1), GHI(t-2), \dots, GHI(t-d))$. This form is a dynamic description of the next day GHI. Once this form is identified, it will be used in a recursive manner to predict the future value of the GHI. Based on trial-and-error procedure, different FFANN structures, input combinations, and training algorithms have been explored in order to improve the accuracy and reduce the computational burden. It should be noted that the results presented in this paper are a part of a complete study aiming at sizing, implementing and controlling a pilot photovoltaic water pumping system (PVWPS) in Hail, Saudi Arabia.

II. LITERATURE REVIEW

In this section, a critical discussion is presented in order to situate the contribution of our paper.

For predicting daily global solar radiation in arid Northwest of China, a back-propagation ANN tuned by particle swarm optimization (BP-PSO) has been investigated in [1], using data from eight meteorological stations. The BP-PSO has shown good performance with a coefficient of determination R^2 varying between 0.5630 and 0.9678. Sunshine duration has been found to be the most affecting global solar radiation. However, the study presented a kind of now-casting (assessment) exercise rather than a forecasting (prediction) one. Based on multiple neural networks and using a photovoltaic panel (PV) model associated with an irradiance forecast, a PV yield prediction system has been presented in [2]. The designed GHI forecaster has shown a mean absolute percentage error (MAPE) of 3.4% on a sunny day and 23% on a cloudy for Stuttgart, Germany. The developed approach is found to

Corresponding author: Sahbi Boubaker

require a large meteorological data set which is usually unavailable. A quaternion-valued neural network (QVNN) based method has been developed for forecasting GHI in [3]. Presenting the meteorological variables to the quaternion domain in a unique variable including the latitude, the longitude and the time indexes has resulted on three times reduction of the input-output layers which decreased computation time. This method has been successfully applied on real datasets from Tamarnasset, Algeria. Several empirical equations and ANN-based algorithms were used for estimating the solar radiation for stations of Aristotle University, Greece [4]. Using daily meteorological data such as temperature, radiation and humidity as explanatory variables, ANN and multi-linear regression (MLR) have been found able to improve the accuracy.

Prediction and forecasting of monthly mean daily GHI has been implemented using ANNs. Nonlinear autoregressive (NAR) with exogenous input (NARX) and hybrid time-series have been evaluated in selected regions from Nigeria. The hybrid model involving the current month number has been found to be the most accurate and reliable. In fact, the coefficient of determination (R^2) has reached 0.96 in Abuja when using the hybrid model [5]. Since solar radiation is scarcely measured in Turkey, an approach using ANN, adaptive neuro-fuzzy inference system (ANFIS), and MLR was implemented in order to forecast GHI in [6]. Using a variance factor analysis procedure, calendar month number, average temperature, average relative humidity and extraterrestrial radiation have been found the most affecting GHI variables. ANN was more accurate outperforming ANFIS, MLR and a set of empirical equations. A hybrid approach including wavelet multi-resolution procedure combined to ANN has been used in [7] for modeling solar radiation. Wavelets have been applied to the original time-series and decomposed them into simple parts. After that, various ANN types (MLP, ANFIS, NARX and generalized ANN (GRNN)) have been used to reconstruct the original signals. The coefficient of determination R^2 has been improved (6.84%) for the MLP and the RMSE (2.78%) for the GRNN without wavelet transformation. The case study was Abudhabi, United Arab Emirates. Further testing was declared to be interesting in other geographical locations.

For sizing a stand-alone PV system at Almadinah, Saudi Arabia [8], a radial-basis ANN has been utilized for predicting global solar radiation. The RBF using the sunshine duration and the temperature as explanatory variables has provided a coefficient of determination of 98.80%. A set of artificial intelligence methods including ANNs have been compared when used for assessing global solar radiation in 12 locations over Iran in [9]. The group method handling data (GMHD) has been found to outperform the other methods in terms of R^2 , RMSE and MSE. ANN-based methods combined to several feature selection techniques have been devised for predicting one-day-ahead GHI in different locations in Saudi Arabia [10]. Further improvements will be obtained when taking into consideration the uncertainties of the used data. A back-propagation ANN (BPNN) optimized by genetic algorithm (GA) and PSO for predicting daily diffuse solar radiation in Beijing, China has been implemented in [11]. Based on several external variables such as temperature and humidity, the PSO-

BPNN has been found to be more accurate than GA-BPNN and BPNN. Combining BPNN to a global optimization algorithm has helped in overcoming the problem of local search algorithms based on gradient descent.

One can conclude that the accuracy of the obtained results is strongly linked to factors such as the size and quality of datasets, the geographic location, the climatic conditions and the design of the method itself. It could be also noted that combined methods look to be more suitable since they allow superposing the advantages of each one of the individual methods.

III. MATERIALS AND METHODS

The data used to perform this study were collected from King Abdullah City for Atomic and Renewable Energy (KACARE) in Saudi Arabia. About two years of daily GHI covering the period from January 1, 2015 to November 30, 2016 from a measurement station located in Hail City (Figure 1) were utilized. The data were divided into subsets: 85% were used for training and 15% for testing of the ANN models.

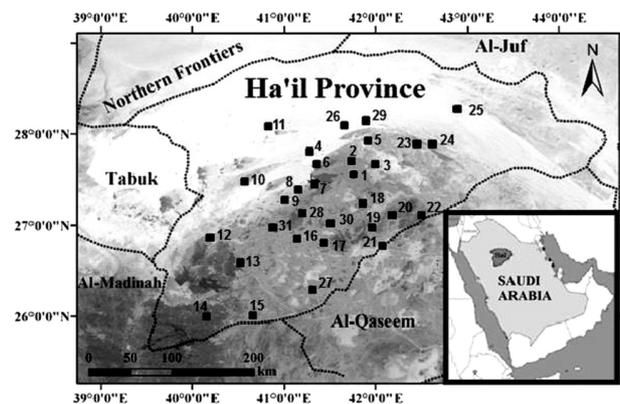


Fig. 1. Hail region location

A. ANN Principles

Basically, ANNs are computing agents that have the ability to imitate the human brain functions. Structurally, an ANN is composed of nodes, weights and activation functions. In its basic form, an ANN includes three layers (input, hidden and output). The main strength of an ANN is that given a set of inputs (patterns) and outputs (targets), it performs a training process to map the inputs to the outputs. This process is called training which can be supervised or unsupervised.

B. ANNs Used in our Study

In this study, we considered the case where the one-day-ahead GHI is affected only by its own previous values. Thus, the ANN inputs are respectively $GHI(t-1)$, $GHI(t-2)$, ..., $GHI(t-d)$. The output is the $GHI(t)$ and d is the model order (Figure 2). The number of neurons in each layer is determined by trial-and-error procedure. The basic multi-layer perceptron (MLP) [14] was used first with a small number of neurons. If the ANN fails to provide good performance, the number of hidden layers and the number of their neurons are increased in a reasonable manner. Weighted-inputs are summed with biases

and fed to a transfer function which converts them to an output of the input layer. This output plays after the role of input to the next layer. This process is repeated until reaching the last layer (commonly called output layer). In general, an ANN operates following seven steps: data collection, dataset pre-processing, creation of the network structure, configuration of the network, definition of the ANN training parameters, training and testing of the ANN.

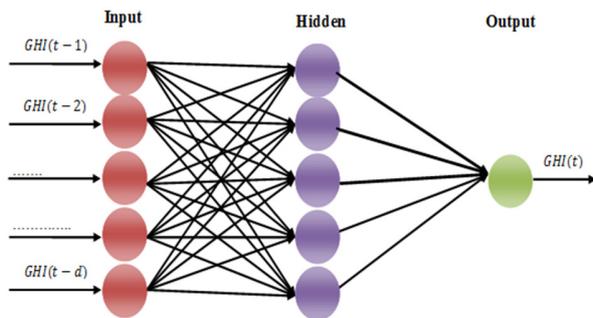


Fig. 2. Proposed FFANN for predicting GHI

The multi-layer FFANN is described by the following set of equations:

$$\begin{aligned}
 a^1 &= f^1(W^{1,1}X+b^1) \\
 a^2 &= f^2(W^{2,1}a^1+b^2) \\
 &\vdots \\
 a^j &= f^j(W^{j,j-1}a^{j-1}+b^j) \\
 &\vdots \\
 a^N &= f^N(W^{N,N-1}a^{N-1}+b^N)
 \end{aligned} \tag{1}$$

where: $a^j, j=1:N$ are the outputs of the respective layers. a^1 is the output of the input layer and a^N is the output of the output layer, $W^{j,j-1}$ are the weights of the j^{th} layer, b^j are the biases of the j^{th} layer, f^j is the transfer function of the j^{th} layer. Tangent sigmoid (tansig) and logsig are commonly used for input and hidden layers whereas purelin is usually used for the output layer. It should be noted that $N=3$ layers (1 input, 1 hidden and 1 output) are used in most cases. If the number of hidden layers is more than one, we will be in the case of deep learning (a special class of machine learning).

During the training phase, the training algorithm updates iteratively the ANN weights as in (2):

$$W(k) = W(k-1) + 2\alpha\Delta W(k-1)X^T \tag{2}$$

where k is the iteration index, α the learning rate and $\Delta W(k-1)$ is the error function related to the weights.

In practice, it is very difficult to choose the suitable training algorithm in terms of accuracy and computation time. The choice is conditioned depending on many factors including the quality and size of the datasets, the number of weights and biases to be adjusted, the level of accuracy and the complexity of the problem to be solved. Commonly, Levenberg-Marquardt (LM), scaled-conjugate gradient (SCG) and resilient propagation (rp) are used.

IV. RESULTS AND DISCUSSION

All computations are carried out using Matlab2013a run on an Intel Core i3 processor with a CPU of 2.40Ghz with 4GB RAM. First, the collected data were pre-processed by dividing them by the maximum value of the daily GHI. The aim of this operation is to transform them into values within the interval $[0, 1]$ which is most suitable for the ANN activation functions. At the end of the training and the validation phases, the results are multiplied by the same maximum value to return-back to their regular values. To evaluate the performance of the developed ANN algorithms, three performance metrics, namely the mean absolute percentage error (MAPE), the coefficient of determination (R^2) and the RMSE were used (see [13, 14] for definitions and details). After several runs, a three-layer (1 input, 1 hidden and 1 output) ANN was adopted. The number of neurons in each layer varied. The resilient-propagation (rp) and the Quasi-Newton (bfg) are used as training algorithms. The tansig and the linear transfer functions (purelin) have been used respectively for the input, hidden and output layers. Several combinations of inputs have been roughly tested. It has been found that better results were obtained when using the six previous daily GHIs as inputs for forecasting the one-day-ahead GHI. The obtained results for some combinations of ANN structures are summarized in Table I for the training and testing phase, in Table II for the training phase and in Table III for the testing phase.

TABLE I. PERFORMANCE OF THE TRAINING AND TESTING PHASES

Case #	ANN structure	Training algorithm	MAPE	R^2	RMSE
1	(10, 5, 1)	bfg	6.8554	0.8594	6.7894
2	(10, 5, 1)	rp	7.7818	0.8112	7.8657
3	(20, 15, 1)	bfg	5.1914	0.9166	5.2281
4	(20, 15, 1)	rp	6.9134	0.8576	6.8316
5	(50, 30, 1)	bfg	7.9008	0.8069	7.9563
6	(50, 30, 1)	rp	5.6850	0.9058	5.5566

Jan. 1, 2015 to Nov. 30, 2016

TABLE II. PERFORMANCE OF THE TRAINING PHASE

Case #	ANN structure	Training algorithm	MAPE	R^2	RMSE
1	(10, 5, 1)	bfg	6.9162	0.8606	6.8871
2	(10, 5, 1)	rp	7.9192	0.8129	7.9784
3	(20, 15, 1)	bfg	4.7008	0.9363	4.6561
4	(20, 15, 1)	rp	6.9653	0.8597	6.9087
5	(50, 30, 1)	bfg	8.1239	0.8050	8.1450
6	(50, 30, 1)	rp	5.3614	0.9157	5.3549

Jan. 1, 2015 to Aug. 26, 2016

TABLE III. PERFORMANCE OF THE TESTING PHASE

Case #	ANN structure	Training algorithm	MAPE	R^2	RMSE
1	(10, 5, 1)	bfg	6.3815	0.7840	6.0898
2	(10, 5, 1)	rp	6.7113	0.7098	7.0590
3	(20, 15, 1)	bfg	9.0127	0.6188	8.0899
4	(20, 15, 1)	rp	6.5091	0.7698	6.2873
5	(50, 30, 1)	bfg	6.1631	0.7506	6.5435
6	(50, 30, 1)	rp	7.5339	0.7337	6.7616

Aug. 27, 2016 to Nov. 30, 2016

To investigate the performance of the FFANN for forecasting daily GHI in Hail region, six cases have been

considered. Three levels of investigations have been drawn. At first, the effect of the ANN structure was studied (more particularly the number of neurons per layer). Then, the used training algorithm in terms of convergence characteristics, result accuracy, and training time and finally, the performance of the ANN for different data subsets (training and testing subset, training subset, and testing subset). By analyzing the results of case#1 and case#2 where a relatively small number of neurons have been utilized, we can conclude that the bfg algorithm performs better than the rp algorithm. By increasing the number of neurons per layer (case#3 and case#4), an improvement of performance for both training algorithms for the three data subsets has been obtained. However, the rp training algorithm has provided better results during the testing phase (MAPE=6.5091, $R^2=0.7698$ and RMSE=6.2873). This result is very interesting since it concerns the testing phase because naturally the training dataset has to provide good agreement when used in testing phase. The relatively high number of neurons (case#5 and case#6) didn't improve the results for the bfg algorithm in the three subsets. In addition, the bfg algorithm took a lot of time in training the ANN (in case#6, the training time was more than one hour). The rp algorithm was way faster than the bfg. Based on the above described results, it can be concluded that for a small number of neurons per layer it is better to use the Quasi-Newton (bfg) algorithm. For a medium number of neurons, rp can be chosen for its fastness and bfg can be chosen for its accuracy. In the case of high numbers of neurons, the rp training algorithm performs better in terms of computation time. Figures 3 and 4 show the graphs of the forecasted GHI, the observed (real) GHI, and the scattered plot of the obtained results during the testing phase.

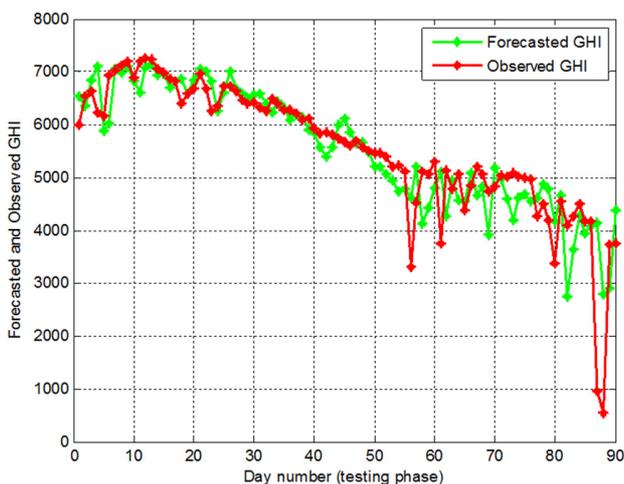


Fig. 3. Forecasted and observed GHI for the testing phase (FFANN trained by resilient-propagation algorithm)

V. CONCLUSIONS

The suitability of FFANNs to predict daily GHI was evaluated using data from the King Abdullah City for Atomic and Renewable Energy (KACARE) station located in Hail College of Technology.

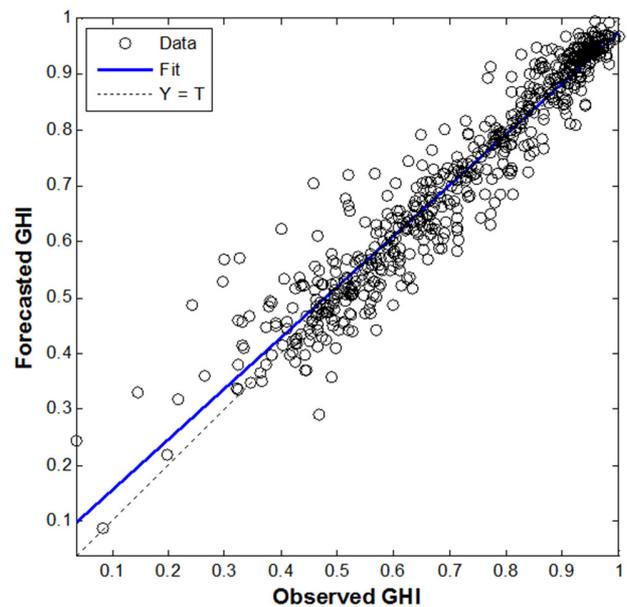


Fig. 4. Scattered plot (case#6)

The obtained results were compared with the use of statistical criteria such as MAPE, R^2 , and RMSE. The best results were recorded when using a medium structure (20, 15, 1) of neurons for the input, hidden and output layers and the Quasi-Newton (bfg) training algorithm. However, the resilient-propagation (rp) algorithm is recommended for structures with high number of neurons in different layers. Further improvements will be achieved in future works by using deep learning (DL) algorithms with more data. Also, other explanatory variables such as climatic conditions and storm effects will be investigated. The results of this study can be extended to photovoltaic power system modeling as in [15].

ACKNOWLEDGMENT

The authors would like to thank Hail University for providing the grant number 160811 to support this study.

REFERENCES

- [1] Y. Zhang, N. Cui, Y. Feng, D. Gong, X. Hu, "Comparison of BP, PSO-BP and statistical models for predicting daily global solar radiation in arid Northwest China", *Computers and Electronics in Agriculture*, Vol. 164, Article ID 104905, 2019
- [2] S. P. Durrani, S. Balluff, L. Wurzer, S. Krautter, "Photovoltaic yield prediction using an irradiance forecast model based on multiple neural networks", *Journal of Modern Power Systems and Clean Energy*, Vol. 6, No. 2, pp. 255–267, 2018
- [3] L. S. Saoud, F. Rahmoune, V. Tourtchine, K. Baddari, "A novel method to forecast 24 h of global solar Irradiation", *Energy Systems*, Vol. 9, pp. 171–193, 2018
- [4] V. Z. Antonopoulos, D. M. Papamichail, V. G. Aschonitis, A. V. Antonopoulos, "Solar radiation estimation methods using ANN and empirical models", *Computers and Electronics in Agriculture*, Vol. 160, pp. 160–167, 2019
- [5] C. G. Ozoegwu, "Artificial neural network forecast of monthly mean daily global solar radiation of selected locations based on time series and month number", *Journal of Cleaner Production*, Vol. 216, pp. 1–13, 2019

-
- [6] H. Citakoglu, "Comparison of artificial intelligence techniques via empirical equations for prediction of solar radiation", *Computers and Electronics in Agriculture*, Vol. 118, pp. 28-37, 2015
- [7] S. Hussain, A. AlAlili, "A hybrid solar radiation modeling approach using wavelet multiresolution analysis and artificial neural networks", *Applied Energy*, Vol. 208, pp. 540-550, 2017
- [8] M. Benghanem, A. Mellit, "Radial basis function network-based prediction of global solar radiation data: application for sizing of a stand-alone photovoltaic system at al-Madinah, Saudi Arabia", *Energy*, Vol. 35, pp. 3751-3762, 2010
- [9] A. Khosravi, R. O. Nunes, M. A. H. Assad, L. Machado, "Comparison of artificial intelligence methods in estimation of daily global solar radiation", *Journal of Cleaner Production*, Vol. 194, pp. 342-358, 2018
- [10] M. Almarachi, "Investigating the impact of feature selection on the prediction of solar radiation in different locations in Saudi Arabia", *Applied Soft Computing*, Vol. 66, pp. 250-263, 2018
- [11] X. Xue, "Prediction of daily diffuse solar radiation using artificial neural networks", *International Journal of Hydrogen Energy*, Vol. 42, pp. 28214-28221, 2017
- [12] S. Boubaker, "Identification of nonlinear Hammerstein system using mixed integer-real coded particle swarm optimization: application to the electric daily peak-load forecasting", *Nonlinear Dynamics*, Vol. 90, No. 2, pp. 797-814, 2017
- [13] S. Boubaker, "Identification of monthly municipal water demand system based on Autoregressive Integrated Moving Average model tuned by Particle Swarm Optimization", *Journal of Hydroinformatics*, Vol. 19, No. 2, pp. 261-281, 2017
- [14] S. Shah, H. N. Nagraja, J. Chakravorty, "ANN and ANFIS for short term load forecasting", *Engineering, Technology & Applied Science Research*, Vol. 8, No. 2, pp. 2818-2820, 2018
- [15] F. Mavromatakis, Y. Franghiadakis, F. Vignola, "Modeling Photovoltaic Power", *Engineering, Technology & Applied Science Research*, Vol. 6, No. 5, pp. 1115-1118, 2016