Daily Global Solar Radiation estimation for Gran Canaria Island using Artificial Neural Networks

L. Mazorra Aguiar [1], P. Lauret [2], F. Díaz[1], A. Ortegón[3] and R. Pérez-Suárez

[1] Department of Electrical Engineering. University of Las Palmas de Gran Canaria
Campus Tafira, 35017 Las Palmas de Gran Canaria (Spain)
Phone/Fax number: +0034 928 451989, e-mail: **luis.mazorra@ulpgc.es**, **felipe.diaz@ulpgc.es**

[2] Laboratoire PIMENT from the University of La Réunion, France
email: philippe.lauret@univ-reunion.fr

[3] Instituto Tecnológico de Canarias (ITC), Canary Islands, Spain
email: aortegon@itccanarias.org

Abstract. Forecasting of global solar radiation is an important tool for power systems planning and operation, especially in island grids. The aim of this paper is the analysis of an artificial neural network as a reliable method to obtain a daily forecast for solar radiation. Some different tests are proposed to obtain the optimal ANN that will capture the underlying physical process that generates the data. In the present study, the available data come from seven measuring stations throughout the Gran Canaria Island along six years. ANN was trained and tested only with past ground measurement solar radiation and other meteorological data available at measurement stations as inputs.

Key words

Artificial Neural Networks, Forecasting, Perceptron, Solar radiation.

1. Introduction

Global solar irradiation is one of the main variables for determining solar energy production in a region. Forecasting of global solar radiation is, therefore, an important tool for power system planning and operating [1] [2]. Insular electrical grids, such as the one from Gran Canaria, are generally not connected with any other big territories and all electrical energy must be generated and consumed inside the island. In addition, solar radiation presents a high spatio-temporal variability in Canary Islands, so an efficient forecasting is becoming an important issue [3] [4]. In the present paper, the available data come from seven measuring stations throughout the island along six years.

In the following pages an artificial neuronal network (ANN) model is proposed for forecasting daily solar radiation. In the NN approach, a long-term data is needed to get the best approach for forecasting solar radiation. ANN is a data-driven approach capable of performing a nonlinear mapping between a set of input and output variables, making this method very attractive [5]. The connection between units has an associated weight, which is modified in a training process. Solar radiation can be treated as time series and statistical models can be built only with historical data. Many papers describe efficient solar radiation forecasting results of ANNs [6] [7] [8]. Besides solar radiation, other data as daylight hours, humidity and temperature are also taken into account in the proposed ANN. Several simulations for many areas around the world are described in scientific bibliography. Rehman and Mohandes [9] use temperature and relative humidity for forecasting daily global solar irradiation, Kemmoku et al. [10] use a multi-stage ANN working with atmosphere pressure and other meteorological data as inputs and Sfetsos and Coonick [11] introduce temperature, wind speed and pressure in order to forecast hourly radiation data. Similar experiments using different combinations of meteorological data as length of the day, daily average temperature, humidity, latitude, longitude and solar radiation have been taken for both hourly and daily solar radiation [12] [13] [8].

This work is divided into the following parts: the first section makes an introduction of the Gran Canaria measurement station and the datasets available for the study. The second section provides a brief description of a neural network that focuses on a multilayer perceptron structure, and presents the model implementation applied in this study. The next section shows different tests for global solar radiation forecasting. Finally, summary and conclusions are shown in last section.

2. Solar radiation data

Ground hourly datasets used in this paper were obtained from seven stations in Gran Canaria Island (Spain) owned by the Canary Islands Technology Institute (Instituto Tecnológico de Canarias, I.T.C.). These measurement stations are equipped for measuring global solar horizontal radiation, relative humidity and temperature. Solar irradiance is measured with a secondary standard pyranometer CMP-11 of Kipp & Zonen, with a 3% accuracy for daily sum of GHI. In Table I, the location of the different stations can be seen. An essential step in any solar radiation survey is a quality assessment of the data. The involved measurements stations use widely recognized equipment and follow periodic calibration and maintenance. In addition, negative data and data over top of atmosphere (ToA) were filtered out. The hourly and daily GHI average were computed for hours and days with more than 50% of data available.

In measurement stations in Gran Canaria there are available data from year 1998 to 2008 (except in C0-Pozo Izquierdo where there are data from 2001). In this paper meteorological data use as inputs are date (day of the year), daily average of relative humidity (percentage of absolute humidity), daily average of temperature (ºC), global horizontal irradiation GHI (daily solar radiation accumulate -Wh/m^2-) and sunshine duration (describe as the number of daily hours when solar radiation is over 120 W/m2).

Table I. – Measurement Stations location

Location	Lat. (º)	Long. (º)	Altitude (m)
C0 – Pozo Izquierdo	27.8175	-15.4244	47
C1 – Las Palmas	28.1108	-154269	17
C2 – La Aldea	27.9901	-15.7907	197
C4 – Maspalomas	27.7716	-15.5841	265
C5 – Sta. Brígida	28.0337	-15.4991	525
C6 – Mogán	27.8839	-15.7216	300
C7 – Gáldar	28.1681	-15.6865	40

It is important that all meteorological data used as inputs must be available in every days used in training and testing datasets. If any data are not available, this day is filtered out. Once datasets were filtered the final number of daily data are 1896 for C0, 2355 for C1, 1021 for C2, 1389 for C4, 1543 for C5 and 1893 for C6. Once data were filtered, station C7 - Gáldar has been discarded because there was not enough good data.

3. Artificial Neural Network description

Artificial neural network (ANN) [5] simulates biological nervous systems, composed of a set of units, called neurons, which are connected one to each other. It is a method that learns the association between a set of input data and the corresponding outputs. All units receive an input from other units or an external data source. The connection between units has an associated weight, which is modified in a training process. The optimal weights are obtained by optimizing a cost function. A widely method

consist of minimizing the mean square error between the output value and the expected value (target). Each neuron receives input variables affecting by these weights and sum them to produce an output. The sum is passed through a nonlinear activation function or transfer function to limit the amplitude of the output. Usually, the activation function applied is the hyperbolic tangent function:

$$f(x) = \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \qquad (1)$$

A. MLP: A system able to establish a link between two data sets

The multilayer perceptron (MLP) is the best-known architecture and with a greater number of applications in ANN. This structure consists of an input layer, at least one hidden layer, and an output layer with no feedback or lateral connections between them. The hidden layer is made of several non-linear neurons, while output layer uses a linear activation function. The input data set, 'x', is transmitted forward through the weights from the input layer to the hidden layer. The hidden layer neurons transform the received signal by a non-linear activation function in an output value. This is transmitted through other weights to the output layer and provides network final output, 'y'. See (2). The weights of the network are set by a random initialization and then are optimized during the training phase by minimizing an error function (cost function). The most common error function is the average of square difference between the network output 'y' and the desired output, target set 't'. See (3).

$$y = y(x; w) = \sum_{j=0}^{h} \left[w_j f\left(\sum_{i=0}^{d} w_{ji} \cdot x_i \right) \right] \qquad (2)$$

$$E(w) = \frac{1}{2} \sum_{i=1}^{N} (y_i - t_i)^2 \qquad (3)$$

The final parameters are set during the training process by minimizing the error between neural network output and real data by backpropagation algorithm. In this process the training data set consists in pairs of input-output vectors with 'N' number of patterns. The backpropagation algorithm is an iterative training method that uses scale conjugate gradient optimization method. The scale conjugate gradient calculates network output and a final error for each iteration. Then it updates weights Δw and recalculates the final error with w + Δw as network weights. If the error has decreased the new weights are used and process begins again, otherwise, new weights are calculated.

Once the network is trained, a test set (part of the data not used during the learning phase) is used to evaluate the ability of the NN to give correct outputs. This generalization error is usually given by %rRMSE. So, once the network has been trained it can provide good results with data that has never seen in its training phase.

B. Model implementation

The structure of the final neural networks is so flexible that it is necessary to find the most adequate to the pairs of input-output data. The accuracy of ANNs to approach continuous functions depends on the network structure. If

the network approximates the noise of the function, ANNs used to lead to get good training set results but poor error results when new data are used (called overfitting). The structure will determine the possible generalization of the neural networks results. In order to avoid the overfitting problem it is generally used regularization techniques and control the complex of the ANN [15].

Regularization techniques have been used in order to obtain better predictions for new inputs. This consists in adding to the cost function a penalty term, also known as weight decay term, to drives to nearly zero the irrelevant weight.

$$S(w) = E(w) + \mu \cdot \frac{1}{2} \sum_{i=1}^{m} w_i^2 \qquad (4)$$

Where '**m**' is the total number of parameters and '**μ**' is the regularization coefficient, which can be neither too large nor too small because it produces an overfitting or underfitting respectively.

The number of input and output units is determined by the problem, while the number of hidden units depends of the application. For this reason, cross validation (CV) techniques are used to find the optimal value of 'μ' and the best architecture. First of all, the data set has to be divided in two sets: training set and test set. For the CV only the first set is used to evaluate the variation of the generalization error; the parameters that correspond to the minimum validation error are the optimal. Since data is often scarce, there might not be enough to set aside for a validation sample. The solution could be split the training data into K subsets. For each fold, a network will be trained on all the subsets (training set) except one (validation set) which is used to test the model. Then, this process is repeated multiple times using all possible splits of the data into training and validation set. For each fold, the network it is trained using different regularization parameters and it is calculated the error related to the training set and validation set. The average of the CV metric across the folds determines the CV error. The regularization coefficient and number of hidden units that determine a minimum error with validation set are the optimal values to obtain the best architecture. This technique is known as K-fold cross-validation.

4. Solar radiation forecasting with ANN

This section shows the different tests that have been made to obtain the best forecasting model. In all of them the output vector is the forecast daily global solar radiation GHI for the next day and the input vector contains the daily mean values of some variables as: daylight hours, humidity, temperature, day of the year and global radiation for day before [8][9].

In order to evaluate the performance of each method, we used the standard error metrics: the Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). Both metrics are widely used in solar forecasting community [16], as well as their relative metrics (rRMSE and rMAE), obtained dividing the absolute error by the average of the daytime GHI data. In this paper, to study the quality of the forecasting methods, we provide the relative errors.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (GHI_{forecast,i} - GHI_{measured,i})^2} \quad (5)$$

$$MAE = \frac{1}{N} \left| \sum_{i=1}^{N} (GHI_{forecast,i} - GHI_{measured,i}) \right| \qquad (6)$$

First, a multilayer perceptron network MLP was trained to forecast daily GHI for the next day (G_{+1}) based on the days of the year, humidity, temperature, daily irradiation, daylight hours (d, W, T, G, n) from the day before. Another MLP was trained for forecasting daily GHI for the next day based on the days of the year, humidity, temperature, daily irradiation, daylight hours from the previous day and two days before (d, W, W_{-1}, T, T_{-1}, G, G_{-1}, n, n_{-1}). Several ANNs were trained using combination of all meteorological data presented in this paper for the previous day and two days before, in order to find which ones of them are more relevant for the optimal forecasting. In all cases, a multilayer perceptron with one hidden layer, tangent hyperbolic activation function and a linear (purelin) output was used.

In this paper, a 10 fold cross validation has been used. Each group consists of a training set (90% of the training data) and a validation set (10% of the training data). Before CV, it has been used 10% of the whole data as the test set. These data are used with the final ANN to study the generalization of the forecasting when new data are presented. Several ANNs has been training using different number of hidden units (from 2 to 30) and different weight decay parameter (from ln $\mu=e^{-2}$ to ln $\mu=e^6$). The CV technique leads us to optimal ANN configuration depending on the validation error. Fig. 1 shows the results of CV technique for station C0-Pozo Izquierdo. Fig. 1 compares forecasting error for validation dataset with the logarithm of weight decay parameter for different hidden units. For station C0 we can see that best result is obtained for 2 hidden units and a weight decay parameter μ=e0=1. Weight decay parameter is an important variable for validation error because it is easily observed that for bigger values the results are worse. All stations show best results for weight decay parameter between 1 and 2. On the other hand, number of hidden units is not relevant for error results. Simulations with different 'nh' show similar results, so the best ANN architecture is always with a few 'nh' (between 2 and 5 depend on the station).
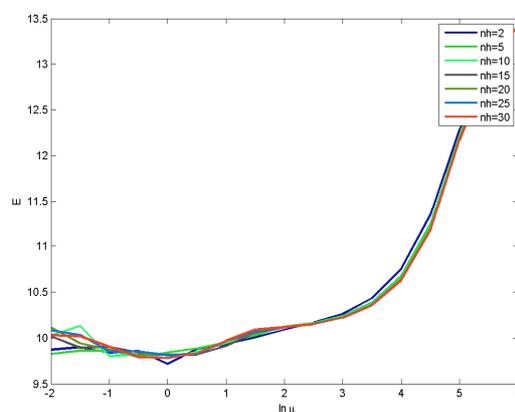

Fig. 1 C0 Validation dataset error compare with logarithm of weight decay parameter for different hidden units number

Fig. 2 compares measure GHI and forecast GHI for both training and testing data set. These images show overfitting if training datasets points are situated on the main diagonal while testing dataset points are spread. All station in Gran Canaria are not affected by overfitting for the optimal ANN architecture obtained during CV training process. Once the number of hidden unit and weight decay parameter was fix, we studied the best number of input variable for forecasting day ahead solar irradiation. The results obtained using five inputs from the day before (d, W, T, G, n) and nine inputs from the previous day and two days before (d, W, W_{-1}, T, T_{-1}, G, G_{-1}, n, n_{-1}) are very similar for all stations, Table II. For training dataset the results are better if we use more inputs, while for testing dataset ANNs with nine inputs don't obtain better forecasting errors. For all station similar results as shown in Table II were obtained, so it easily to understand that the best option is using only 5 inputs from day before data.
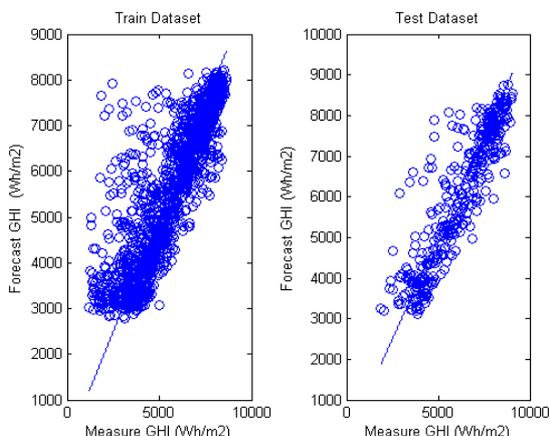

Fig. 2 C0 station data. Forecast GHI vs Measure GHI data for training and testing datasets

Table II. – Error results for C0 station

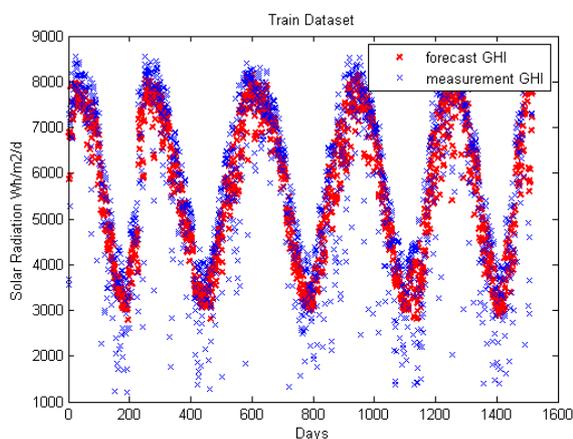| Architecture	Train %rRMSE	Train %rMAE	Test %rRMSE	Test %rMAE
5 inputs – 5 nh	17.01	12.14	12.65	9.34
9 inputs – 5 nh	16.84	11.96	12.76	9.33

Fig. 3 C0 station data. Training dataset measure and forecast GHI values

Fig. 3 and 4 show the comparison of the measured data and the neural network approach for station C0. The ANN obtain very good results for forecasting daily solar radiation data using meteorological variables as inputs. Only the low and estrange daily data are difficult to forecast.
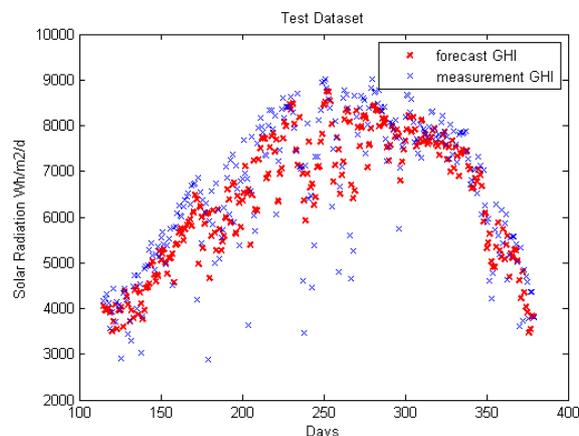

Fig. 4 C0 station data. Testing dataset measure and forecast GHI values

5. Summary and conclusions

It is observed that, for each station, there is a neural network that best suited to their conditions. As it is shown for Pozo Izquierdo's measurement station, ANNs forecasts next day global solar radiation and fits the real data. The relative root mean square error comparing the results obtained from the artificial neural network once it was trained and actual data from measurement stations are around 12-20% in southern stations (C0,C2,C4 and C6) and 17-24% in northern stations (C1 and C5). As it is observed, the results are different between northern, affected by cloudiness in summer, and southern station. It will be a future way train separates ANNs between northern and southern stations.

References

[1] M. Wittmann, H. Breitkreuz, M. Schroedter-Homscheidt y M. Eck, «Case studies on the use of solar irradiance forecast for optimized operation strategies of solar thermal power plants,» *Selected Topics in Applied Earth Observations and Remote Sensing IEEE Journal,* vol. 1, nº 1, pp. 18-27, 2004.

[2] D. Heinemann, E. Lorenz y M. Girodo, «Forecasting of solar radiation. Solar Energy resource managmente for electricity generation from local level to global scale.,» *New York: Nova Science Publishers,* 2006.

[3] F. Díaz Reyes, Modelo numérico para la generación de mapas de radiación solar y u aplicación al aprovechamiento de energía solar fotovoltaica y termoeléctrica, Las Palmas de Gran Canaria: Tesis Doctoral Universidad de Las Palmas de Gran Canaria, 2013.

[4] M. Diagne, M. David, J. Boland, N. Schmutz and P. Lauret, "Post-processing of solar irradiance forecasts from WRF model at Reunion Island," *Solar Energy,* vol. 105, pp. 99-108, 2014.

[5] C. Bishop, Neural Networks for Pattern Recognition, Oxford: Oxford University Press, 1995.

[6] L. Hontoria, J. Aguilera and P. Zufiria, "Generation of hourly irradiation synthetic series using the neural network multilayer perceptron," *Solar Energy,* vol. 72, no. 5, pp. 446-2002, 2002.

[7] P. Lauret, M. David, E. Fock, A. Bastide and C. Riviere,

"Bayesian and Sensitivity Analysis Approaches to Modeling the Direct Solar Irradiance," *Solar Energy Engineering,* vol. 128, pp. 394-405, 2006.

[8] A. Mellit and A. Pavan, "A 24-h forecast of solar irradiance using artificial neural network: Application for performance prediction of a grid-connected PV plant at Trieste, Italy," *Solar Energy,* vol. 84, pp. 807-821, 2010.

[9] Shafiqur Rehman, Mohamed Mohandes, "Artificial neural network estimation of global solar radiation using air," *Energy Policy,* vol. 36, pp. 571-576, 2008.

[10] Y. Kemmoku, S. Orita, S. Nakagawa y T. Sakakibara, «Daily insolation forecasting using a multi-stage neural network,» *Solar Energy,* vol. 66, nº 3, pp. 193-199, 1999.

[11] A. Sfetsos y A. Coonick, «Univariate and multivariate forecasting of hourly solar radiatin with artificial intelligence techniques,» *Solar Energy,* vol. 68, nº 2, pp. 169-178, 2000.

[12] M. Mohandes, S. Rehman y T. O. Halawani, «Estimation of global solar radiation using artificiaol neural networks,» *Renewable Energy,* vol. 14, nº 1-4, pp. 179-184, 1998.

[13] A. Ghanbarzadeh, A. Noghrehabadi, E. Assareh y M. Behrang, «Solar radiation forecasting based on meteorological data using artificial neural networks,» de *7th IEEE International Conference on Industrial Informatics*, Indin, 2009.

[14] C. M. Bishop, Neural Networks for Pattern Recognition, Oxford: Clarendon Press, 1995.

[15] P. Lauret, E. Fock, R. Randrianarivony and J. Manicom-Ramsamy, "Bayesian neural network approach to short time load forecasting," *Energy Conversion & Managment,* vol. 49, no. 5, pp. 1156-1166, 2008.

[16] C. Willmott and K. Matsuura, "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance," *Climate Research,* vol. 30, pp. 79-82, 2005.