PREDICTION OF MAXIMUM SOLAR RADIATION USING ARTIFICIAL NEURAL NETWORKS

S. Kalogirou[1], S. Michaelides[2] and F. Tymvios[2]

[1] Higher Technical Institute, P.O.Box 20423, Nicosia 2152, Cyprus.
[2] Meteorological Service of Cyprus, Nicosia 1418, Cyprus

ABSTRACT

The prediction of solar radiation is very important for many solar applications. Due to the very nature of solar radiation, many parameters can influence both its intensity and its availability and therefore it is difficult to employ analytical methods for such predictions. For this reason, multivariate prediction techniques are more suitable. In the present research, artificial neural networks are utilised due to their ability to be trained with past data in order to provide the required predictions. The input data that are used in the present approach are those which influence mostly the availability and intensity of solar radiation, namely, the month, day of month, Julian day, season, mean ambient temperature and mean relative humidity (RH).

A multilayer recurrent architecture employing the standard back-propagation learning algorithm has been applied. This methodology is considered suitable for time series predictions. Using the hourly records for one complete year, the maximum value of radiation and the mean daily values of temperature and relative humidity (RH) were calculated. The respective data for 11 months were used for the training and testing of the network, whereas the data for the remaining one month were used for the validation of the network. The training of the network was performed with adequate accuracy. Subsequently, the "unknown" validation data set produced very accurate predictions, with a correlation coefficient between the actual and the ANN predicted data of 0.9867. Also, the sensitivity of the predictions to ±20% variation in temperature and RH give correlation coefficients of 0.9858 to 0.9875, which are considered satisfactory. This is considered as an adequate accuracy for such predictions.

INTRODUCTION

Solar radiation is the driving force behind a number of solar energy devices with different operating principles. Therefore, the prediction of solar radiation is very important for the optimum operation of many solar applications. This is particularly important in solar electric generating systems, where accurate predictions of solar radiation allow for a better planning of the operation of an auxiliary system, especially in cases where steam boilers that require many hours to warm-up are used.

Due to the very nature of solar radiation, many parameters can influence both its intensity and its availability and therefore it is difficult to employ analytical methods for such predictions. For this reason, multivariate prediction techniques are more suitable. In this respect, artificial neural networks are utilised in the present research, due to their ability to be trained with past data in order to provide the required predictions.

Artificial neural networks (ANNs) mimic somewhat the learning process of a human brain. Instead of complex rules and mathematical routines, ANNs are able to learn the key information patterns within a multidimensional information domain. In addition, the inherently noisy data does not seem to present a problem. According to Haykin [1], a neural network is a massively parallel distributed processor that has a natural propensity for storing experiential knowledge and making it available for use. It resembles the human brain in two respects: (a) the knowledge is acquired by the network through a learning process, and (b) inter-neuron connection strengths, known as synaptic weights, are used to store the knowledge.

ANN models represent a new method in energy prediction. ANNs operate like a "black box" model, requiring no detailed information about the system. Instead, they learn the relationship between the input parameters and the controlled and uncontrolled variables, by studying previously recorded data, similar to the way a non-linear regression might perform. Another advantage of using ANNs is their ability to handle large and complex systems with many interrelated parameters. They seem to simply ignore excess input data that are of minimal significance and concentrate instead on the more important inputs.

A training set is a group of matched input and output patterns used for training the network, usually by suitable adaptation of the synaptic weights. The outputs are the dependent variables that the network produces for the corresponding input. It is important that all the information needed by the network in order to learn, is supplied to it as a data set. When each pattern is read, the network uses the input data to produce an output, which is then compared to the training pattern, i.e., the correct or desired output. If there is a difference, the connection weights are usually altered in such a direction that reduces the error. After the network has run through all the input patterns, and if the error is still greater than the maximum desired tolerance, the ANN runs again through all the input patterns repeatedly, until all the errors are within the required tolerance. When the training reaches a satisfactory level, the network holds the weights constant. The trained network can then be used to make decisions, identify patterns or define associations in new input data sets not used to train it.

The most popular learning algorithms are the back-propagation and its variants. The Back-Propagation (BP) algorithm is one of the most powerful learning algorithms in neural networks. The training set must be a representative collection of input-output examples. Back-propagation training is a gradient descent algorithm. It tries to improve the performance of the neural network through reducing the total error by changing the weights along its gradient. A review of applications of ANNs in renewable energy engineering, together with a detailed description of BP is given in [2].

A number of other researchers have used neural networks for the prediction of solar radiation. Alawi and Hinai [3] have used ANNs to predict solar radiation in areas not covered by direct measurement instrumentation. In their research, the input data to the network were the location, month, mean pressure, mean temperature, mean vapour pressure, mean relative humidity, mean wind speed and mean duration of sunshine. The ANN model predicts solar radiation with an accuracy of 93% and mean absolute percentage error of 7.3.

Mohandes *et al.* [4] used data from 41 recording stations in Saudi Arabia. The respective data set for 31 stations was used to train a neural network and the data for the other 10 for testing the network. The input values to the network are latitude, longitude, altitude and sunshine duration. The results for the testing stations obtained are within 16.4% and indicate the viability of this approach for spatial modelling of solar radiation.

Kemmoku *et al.* [5] used a multistage ANN to predict the insolation of the next day. The input data to the network are the average atmospheric pressure, predicted by another ANN, and various weather data of the previous day. The results obtained have shown a prediction accuracy of 20%.

The objective of this paper is to produce a neural network model for the prediction of maximum solar radiation based on simple data, which express the time of the year, season and some measurable weather parameters such as the mean ambient temperature and relative humidity.

DATA COLLECTION

Despite the abundance of records for ambient temperature and rainfall, which cover almost a century in most meteorological stations operating in Cyprus, data for solar radiation are scarce and refer mostly to the area of

Nicosia. Data for one complete year of measurements are used in the present study to train a neural network in order to predict the maximum radiation.

Using the hourly collected records, the maximum value of radiation and the mean daily values of temperature and relative humidity (RH) were calculated. The respective data for 11 months were used for the training and testing of the network, whereas the data for the remaining one month, consisting of values for the first two weeks of July and the last two weeks of December, were used for the validation of the network.

The input data that are used in the present approach are those which influence mostly the availability and intensity of solar radiation, namely, the month, day of month, Julian day, season, mean ambient temperature and mean relative humidity (RH). For example, increased values of relative humidity may mean either increased moisture in the atmosphere or cloudiness, both leading to a reduction of the available radiation. Additionally, the input data, i.e. mean temperature and RH can be predicted with other statistical or analytical methods and used in the present modelling approach.

ARTIFICIAL NEURAL NETWORKS FOR MODELLING THE SYSTEM

Various network architectures, such as with three, four and five-layers, a number of recurrent type, and a number of feedforward ones have been investigated aiming at finding the one that yields the best overall performance. Also a number of different network sizes and learning parameters have been tried. The architecture that was ultimately selected is shown in Figure 1. It is of the recurrent type and is composed of four slabs, one of which is hidden and one is used for dampened feedback. The extra slab is connected to the hidden layer. This architecture is commonly called "Jordan Elman recurrent network". It holds the contents of one of the layers as it existed when the previous pattern was trained. In this way, the network sees the previous knowledge it had about previous inputs. This extra slab is sometimes called the network's "long term" memory. The long-term memory remembers the hidden layer, which contains features detected in the raw data of previous patterns. Recurrent neural networks are particularly suitable for prediction of sequences so they are excellent for time series data, as in the present case. A back-propagation network with standard connections responds to a given input pattern with exactly the same output pattern, every time the input pattern is presented. A recurrent network may respond to the same input pattern differently at different times, depending upon the patterns that have been presented as inputs just previously. Thus, the sequence of the patterns is as important as the input pattern itself. Recurrent networks are trained in the same manner as standard back-propagation networks, except that patterns must always be presented in the same order, i.e., random selection is not allowed.

The activation function used for each slab is also shown in Figure 1. The activation function used in the input slab is linear (i.e., of the form y=x) whereas in the hidden and output slabs are of the sigmoid form given by the logistic formula:

$$y = \frac{1}{1+e^{-x}} \tag{1}$$

Six input elements have been used corresponding to the values of the input parameters listed above. The learning procedure was implemented by using the back-propagation algorithm. The learning rate and the momentum factor were set to a constant value of 0.1. The weights were initialised to a value of 0.3. In back-propagation networks, the number of hidden neurons determines how well a problem can be learned. If too many are used, the network will tend to try to memorise the problem and thus not generalise well later. If too few are used, the network will generalise well but may not have enough "power" to learn the patterns well. The procedure for getting the right number of hidden neurons is not scientifically justified but it is a matter of trial and error. In this case the number of hidden neurons was estimated by applying the following formula:

$$\text{Number of hidden neurons} = \tfrac{1}{2}\,(\text{inputs} + \text{outputs}) + \sqrt{\text{number of training patterns}} \tag{2}$$

This is an empirical formula which has been used in a number of renewable energy engineering problems for modelling and prediction with very good results [2].

From a total of 334 patterns, 66 were randomly selected to be used as test patterns and the remaining 268 were used for training the network. For the training data set a correlation coefficient equal to 0.9811 was obtained. Regarding the correlation coefficient the closer to unity is the better the mapping.

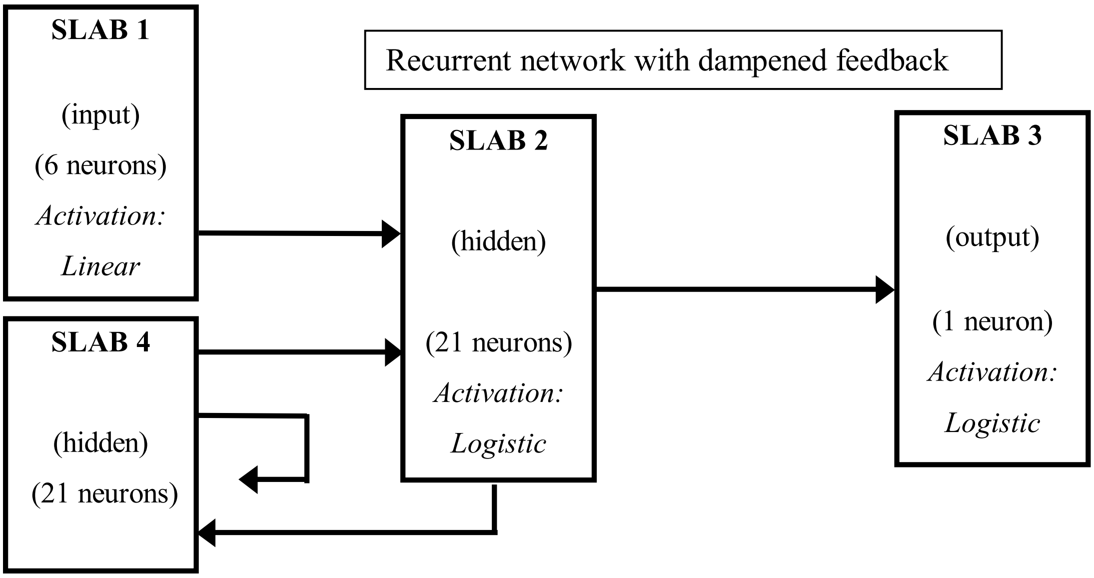

Figure 1: The selected neural network architecture.

RESULTS/VALIDATION

Once a satisfactory degree of input-output mapping has been reached, the network training is frozen and a set of completely unknown test data was applied for validation. The trained network was used to predict the maximum solar radiation for the days contained in the validation data set. It should be noted that the network has not seen these data before, i.e., these cases were completely unknown to the network. The correlation coefficient obtained for the validation data set is 0.9867. In this respect, the closer to unity this value is the better the prediction accuracy. A graph of the actual and ANN predicted values of maximum solar radiation for the validation data set is shown in Figure 2. As expected, the prediction is much more accurate during the summer period when most of the days are almost always cloud free.

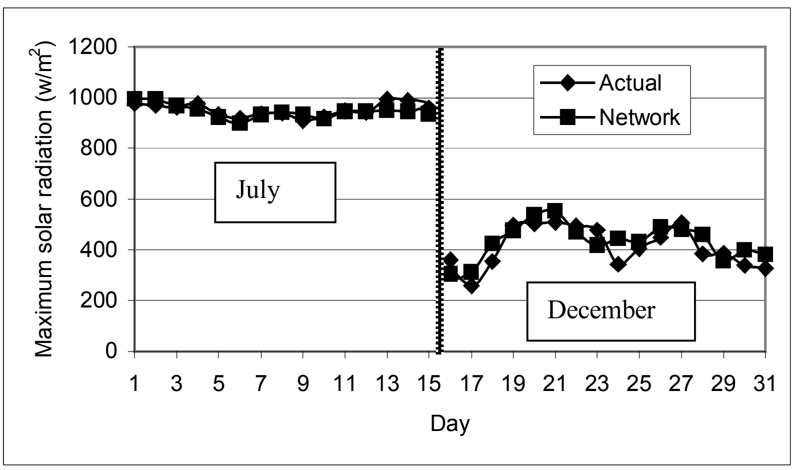

Figure 2: Comparison of the actual and ANN predicted values of maximum solar radiation for the validation data set (first two weeks of July and last two weeks of December).

In order to check the sensitivity of the present model, the values of the mean ambient temperature and relative humidity were deliberately varied to ±20% of the recorded values. The subsequent predictions of the present neural network model are shown in Table 1. As can be seen, variations in correlation coefficients of 0.9858 to 0.9875 were obtained, which is considered satisfactory. This is considered as an adequate accuracy for such predictions, which strengthens the robustness of the model. It should be noted that in actual use, predicted values of mean ambient temperature and relative humidity would be used.

The training of the network required about 2 minutes to run on a Pentium 400 MHz machine, for a total of 1990 learning epochs. The subsequent predictions for the unknown cases (validation data set) require less than a second on the same machine.

TABLE 1
SENSITIVITY ANALYSIS

Parameter	Correlation coefficient
+20% variation in ambient temperature	0.9858
-20% variation in ambient temperature	0.9875
+20% variation in relative humidity	0.9867
-20% variation in relative humidity	0.9861

CONCLUSIONS

The ANN modelling presented here was able to predict the maximum solar radiation with acceptable accuracy. A multilayer recurrent architecture, employing the standard back-propagation learning algorithm has been applied. This methodology is considered suitable for time series predictions.

The training of the network was performed with adequate accuracy. A correlation coefficient of 0.9867 has been obtained when completely unknown data were presented to the network. This is considered adequate and thus the neural network can be used effectively for this type of prediction. Also, the sensitivity of the predictions to ±20% variation in temperature and RH, give correlation coefficients of 0.9858 to 0.9875, which are considered satisfactory.

At this stage the work was confined primarily to the investigation of the suitability of artificial neural networks for predicting the maximum solar radiation. In order for the network to be of significant use to scientists and engineers working in the field of meteorology and solar energy, it needs to be enriched with more training cases and more diverse weather conditions. As the Meteorological Service collects more data in the coming years, these will be used to retrain the network so as to produce a useful operational tool for radiation prediction. It is anticipated that this prediction methodology will be improved further, as the training database expands, covering more unusual cases.

It should be noted that radiation values are site specific, i.e. they depend on the cloud cover and clearness index of the particular site and therefore a different neural model need to be trained for each site. Apart from the prediction of the maximum solar radiation value, the present method can also be used to fill missing values of a radiation database when the mean temperature and RH are known. Subsequently, it would be much easier for meteorologists to fill missing hourly values based on expert knowledge.

Additionally the authors believe that other parameters related to solar radiation, such as the total daily global radiation, can be predicted in the same way, by using artificial neural networks.

REFERENCES

1. Haykin S. (1994) *Neural Networks: A Comprehensive Foundation*, Macmillan, New York.
2. Kalogirou S. (2001) *Renewable & Sustainable Energy Reviews*, **5**, 373.
3. Alawi S.M. and Hinai H.A. (1998) *Renewable Energy*, **14**, 199.
4. Mohandes M., Rehman S. and Halawani T.O. (1998) *Renewable Energy*, **14**, 179.
5. Kemmoku Y., Orita S., Nakagawa S. and Sakakibara T. (1999) *Solar Energy*, **66**, 193.

