Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Energy 30 (2005) 1675–1684
ENERGY
www.elsevier.com/locate/energy

Selection of input parameters to model direct solar irradiance by using artificial neural networks

G. López [a,*], F.J. Batlles [b], J. Tovar-Pescador [c]

[a] *Departamento Ingeniería Eléctrica y Térmica, EPS La Rábida, Universidad de Huelva, Palos de la Frontera, 21819 Huelva, Spain*

[b] *Departamento Física Aplicada, Universidad de Almería, 04120 Almería, Spain*

[c] *Departamento Fisica, Universidad de Jaén, Spain*

Abstract

A very important factor in the assessment of solar energy resources is the availability of direct irradiance data of high quality. However, this component of solar radiation is seldom measured and thus must be estimated from data of global solar irradiance, which is registered in most radiometric stations. In recent years, artificial neural networks (ANN) have shown to be a powerful tool for mapping complex and non-linear relationships. In this work, the Bayesian framework for ANN, named as automatic relevance determination method (ARD), was employed to obtain the relative relevance of a large set of atmospheric and radiometric variables used for estimating hourly direct solar irradiance. In addition, we analysed the viability of this novel technique applied to select the optimum input parameters to the neural network. For that, a multi-layer feedforward perceptron is trained on these data. The results reflect the relative importance of the inputs selected. Clearness index and relative air mass were found to be the more relevant input variables to the neural network, as it was expected, proving the reliability of the ARD method. Moreover, we show that this novel methodology can be used in unfavourable conditions, in terms of limited amount of available data, performing successful results.

1. Introduction

Most solar energy applications such as the simulation of solar energy systems require, at the least, a knowledge of hourly values of solar radiation on a tilted and arbitrarily oriented surface. Knowledge of direct irradiance is important in applications where the solar radiation is concentrated, either to raise the temperature of the system, as in solar thermal energy technologies, or to increase the intensity of the electric current in solar cells, as in photovoltaic

* Corresponding author. Fax: +34-959-017-304.
E-mail address: gabriel.lopez@die.uhu.es (G. López).

0360-5442/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.energy.2004.04.035

systems. In the absence of direct irradiance data, this component of solar radiation may be estimated using decomposition models. They calculate direct irradiance from global solar irradiance on a horizontal surface. These models are based on the regressions between two dimensionless indices: the clearness index, k_t (horizontal global irradiance/horizontal extraterrestrial irradiance) and the direct solar transmittance, k_b (direct normal irradiance/extraterrestrial irradiance).

However, the relationship k_b–k_t is far to be one-to-one because of the complex processes affecting solar radiation throughout the atmosphere. In this sense, there is a wide range of values attained by k_b for a given k_t at intermediate clearness index values. The main reason for the high spread observed on the k_b–k_t scatter plots comes from cloud effects. Nevertheless, the above problem has led to include additional atmospheric and meteorological parameters to the decomposition models in order to improve their performance and to account for different climatic conditions. In the literature, there are models that use different input sets for estimating direct transmittance, involving variables such as solar elevation, relative air mass, precipitable water, dew point temperature, temperature, relative humidity, atmospheric turbidity or surface albedo [1–5].

The study of the influence of these variables on k_b–k_t regressions by using traditional statistical methods is complicated and time consuming. In addition, if analytical models are to be developed, it is needed to have a priori information about the structure of the mathematical relationships between variables. These relationships appear to be non-linear and difficult to handle with standard statistical techniques. This is the case for instance to provide a simple continuous function f, in such a way that $k_b = f(k_t)$ for every k_t-value. If new input variables are to be added to the model, the use of the conventional multiple linear regression will become inappropriate, whereas if non-linear regression is to be used, an explicit function should be provided in advance. Moreover, the latter procedures are static in the sense that the nature of the model cannot be changed. Similarly, computer regression programs cannot learn or become smarter.

An alternative way to avoid the above problems is to employ artificial neural networks (ANN). Among the multiple utilities of the ANN (such as pattern recognition and classification, function approximation, prediction, etc.), we emphasise their growing use for data analysis, offering an effective alternative to more traditional statistical techniques in many scientific fields. Particularly, in the meteorological and solar energy resources fields, ANN based models have been successfully developed to model different solar radiation variables improving the existing statistical approaches [6–11]. In this sense, it is very important to have simple and accurate models relating solar radiation variables to each one or/and to atmospheric and meteorological variables. This is because most studies for the utilisation of solar energy require data that are seldom available at the location of interest and need to be estimated from whatever limited data are at hand there. This is the common problem exhibited in using data of the direct normal component of the solar irradiance [5]. Since neural networks are highly non-linear and require no prior assumptions concerning the data relationships, they have become an useful tool to tackle the direct solar irradiance modelling.

Nevertheless, despite the advantages in using ANN, they present some drawbacks. One is concerning with ANN, as ANN are able to include any superfluous input variable to the model, increasing complexity and given erroneous information about the true variables that affect the dependent one. The other one is the need to divide the data set into three subsets (training, testing and validation sets) which may become a problem if only a few data are available. The Bayesian method of ARD [12,13] for multilayer feedforward perceptron networks (MLP) solves both problems simultaneously: it provides the relative importance of different inputs to the ANN and avoids the need to use separate testing and validation data, thanks to the inclusion of regularization coefficients inside the ANN structure.

In this preliminary work, we test the sensibility of the ARD method for selecting the more relevant input parameters affecting the direct solar transmittance. For this purpose, we have tried to collect the commonest radiometric and meteorological variables used by empirical models, such as it was explained above, as well as other additional parameters, such as wind speed and pressure, which are not expected to be significant a priori. In this sense, since wind speed could account for seasonal variations of atmospheric turbidity [14], and thus to present some significance, we have taken into account a site with low annual levels of aerosol particles, as we had proved in a previous study.

2. Methodology

The neural network selected here is a MLP [15] with one single hidden layer and one single output. Fig. 1 shows the topology of this ANN. Every layer is formed from elemental units named neurones or nodes. Neurons in the input layer only receive the input signals $\vec{x}$ and distribute them forward to the network. In the following layers, each neuron receives a signal, which is a weighted sum of the outputs of the nodes in the layer below. Inside each neuron, an activation function controls its output. The activation function used for the hidden units is the hyperbolic tangent (tanh) whereas the identity function has been used for the output ones. Such a network determines a non-linear mapping from an input vector (radiometric and atmospheric variables) to the output (direct solar transmittance) parameterised by a set of network weights,

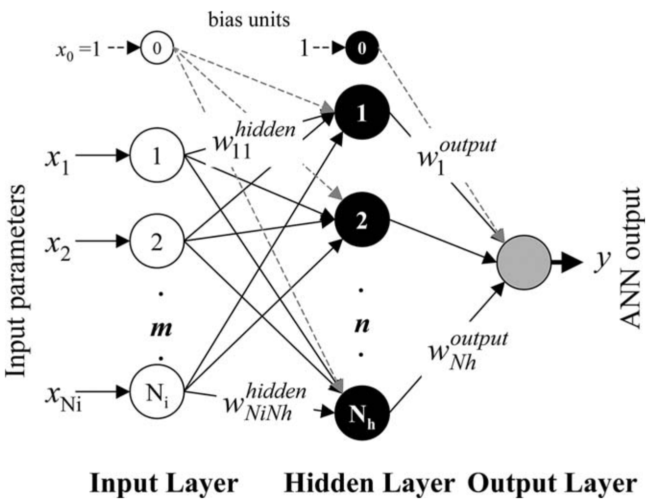

Fig. 1. Topology of the multilayer perceptron.

which we refer to as the vector of weights $\vec{w}$. Following notation from Fig. 1, the output y given by this ANN is the following notation from Fig. 1

$$y = \sum_{n=1}^{N_h} w_n^{\text{output}} \tanh\left(\sum_{m=0}^{N_i} w_{mn}^{\text{hidden}} x_m\right) + w_0^{\text{output}} \tag{1}$$

Once the MLP topology is fixed, the ANN learns the relationship between input and output parameters from examples. For that, the network must be trained. The network training is the procedure where the weights are adjusted based on training data. The training data is a set of patterns consisting of input and corresponding output values, so-called target values. The training method used in this work is named as supervised training. The patterns in the training set are presented to the network one at a time, and following a random sequence for optimal learning. For each sample, outputs obtained by the network are compared with the desired outputs. After the entire training pattern has been processed, the weights are updated. This updating is done in such a way that a measure of the error in the network's results is minimised. As a consequence, the ANN is able to include all input parameters presented to: relevant and unrelevant parameters. This is an undesirable outcome in order to obtain a simple model. To avoid this situation, different techniques have been developed. One of the most efficient methods is based on Bayesian framework for ANN.

Bayesian framework for ANN is based on a probabilistic interpretation of network training, providing a useful method for determining the relevance of input parameters named as ARD method [16–19]. The Bayesian approach considers a priori probability distribution over the network weights $p = (\vec{w}|\vec{\alpha})$ given the vector of hyperparameters $\vec{\alpha}$ (which represents the weight decay regularizers). For that, all the weights associated with the same input node are grouped and a hyperparameter, α_g, is introduced for each of them (see Fig. 2). This hyperparameter controls the prior distribution, which is achieved by using a Gaussian distribution with zero mean and variance $\sigma_g^2 = 1/\alpha_g$, from which the random weights in the gth weight group are sampled.

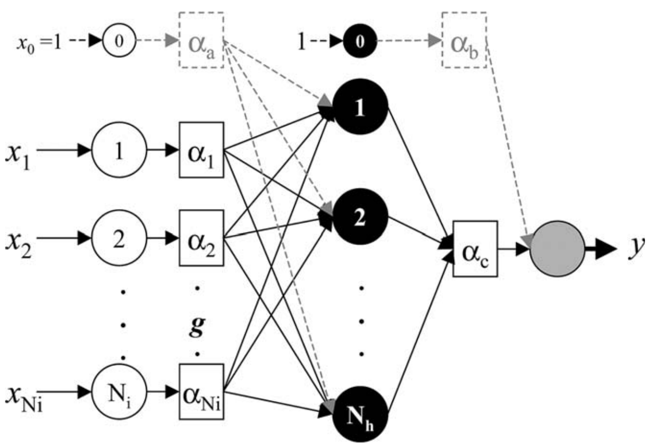

Fig. 2. Graphical representation of the multilayer perceptron with automatic relevance determination. The hyperparameters $\{\alpha_1, \ldots, \alpha_{N_i}\}$ control the weights connecting each input to the hidden layer.

A prior distribution $p(\vec{\alpha})$ determines the initial values of α_g. Three additional hyperparameters are also defined to account for the remainder groups of the network parameters (input bias, hidden bias and hidden weights). These control the complexity of the model. After, once the training data set, D, has been presented to the network, the posterior weight and $\vec{\alpha}$ distributions are found by using the Bayes theorem

$$p(\vec{w}|\vec{\alpha}, D) = \frac{p(D|\vec{w})p(\vec{w}|\vec{\alpha})}{p(D|\vec{\alpha})} \tag{2}$$

$$p(\vec{\alpha}|D) = \frac{p(D|\vec{\alpha})p(\vec{\alpha})}{p(D)} \tag{3}$$

where $p(D|\vec{w})$ is the likelihood and $p(D|\vec{\alpha})$ is termed as the *evidence* for $\vec{\alpha}$. These probability distributions are determined using MacKay's evidence procedure [12].

In this paper, the evidence procedure is carried out '*s*' iterative sessions. Each session involves two steps. The first step calculates the weights maximising the likelihood by means of a standard training of the ANN. The training algorithm selected was the scaled conjugate gradient [20], which performs '*c*' network training cycles. In the second step, the hyperparameters are re-estimated '*n*' times to provide the maximum evidence. The weights of an input with a large α_g are closed to zero, and thus the corresponding input is not as relevant as in order to explain the variability of the dependent variable. Thus, ARD allows the network to determine automatically the importance of each input, effectively turning off those that are not relevant. Furthermore, the early 'stopping' procedure for network training is avoided. The ARD method was performed using MATLAB [21] code in conjunction with several routines developed by Bishop and Nabney [22,23].

3. Data and measurements

The data set was collected at one radiometric station located at Desert Rock (36.63° N, 116.02° W, 1007 m) (USA), between 1998 and 1999. This station is part of NOAA's *Surf*ace *Rad*iation budget network (SURFRAD) [24]. The data set contains records of global and direct solar irradiance, temperature, relative humidity, surface pressure and wind speed. Global irradiance was measured by means of an Eppley ventilated pyranometer model PSP, whereas an Eppley normal-incident pyrheliometer (NIP) was employed to measure the direct irradiance. Measurements of temperature T, relative humidity RH, and pressure p, were registered by means of standard sensors housed in a radiation shield. A propeller anemometer/vane combination was used to account for the wind. Two additional parameters were derived from the measured data: the dew point temperature T_{d}, and the precipitable water, w. Solar position was taken into account by using the cosine of solar zenith angle, θ_{z}. The relative optical air mass, m_r, was also obtained.

Desert Rock was selected among several stations due to its clear atmosphere, in the sense of low amount of aerosols, high altitude, and no seasonal changes in vegetation. The first two conditions minimise the attenuation of the direct irradiance by aerosols (one of the main attenuation sources), and the third one ensures a constant albedo. These facilities allow to analyse the relationships between other variables and direct solar irradiance in a more easy way. Hourly averaged values were obtained for all variables. Due to cosine response problems of radiometric sensors, we have only used cases corresponding to solar zenith angles less than 85°. Consequently, 4880 h were employed. The input and output values were linearly scaled to lie in the range (0,1) using the maximum and minimum recorded value for each variable.

4. Results

To develop an ANN based model, several free parameters must be fixed before the training stage. For our multilayer perceptron based models, these free parameters are the input and hidden unit numbers, N_i and N_h, respectively. Because of computer limitations, N_h may be bounded and depend on the training pattern number, N_p (or vice-versa). At the same time, the training pattern number depends on the selected location and the amount of available measurements there. Once these network parameters are defined, the network training is determined by the above noted training parameters: *c*, *n* and *s*. We will analyse the ARD results by varying these parameters, but *n*, which was set to 5.

Fig. 3 shows the log-values of the hyperparameters versus training session, *s*, for two multilayer perceptrons with 10 and 2 hidden units, MLP10 and MLP2, respectively. For each one, 30 and 200 training cycles, *c*, were accomplished using the whole database (4880 h). A noted important feature is the constant values of the hyperparameters after several training sessions. This behaviour is more marked if the number of the hidden units is low. This is because a more complex ANN needs more training cycles to capture the underlying relationships between the

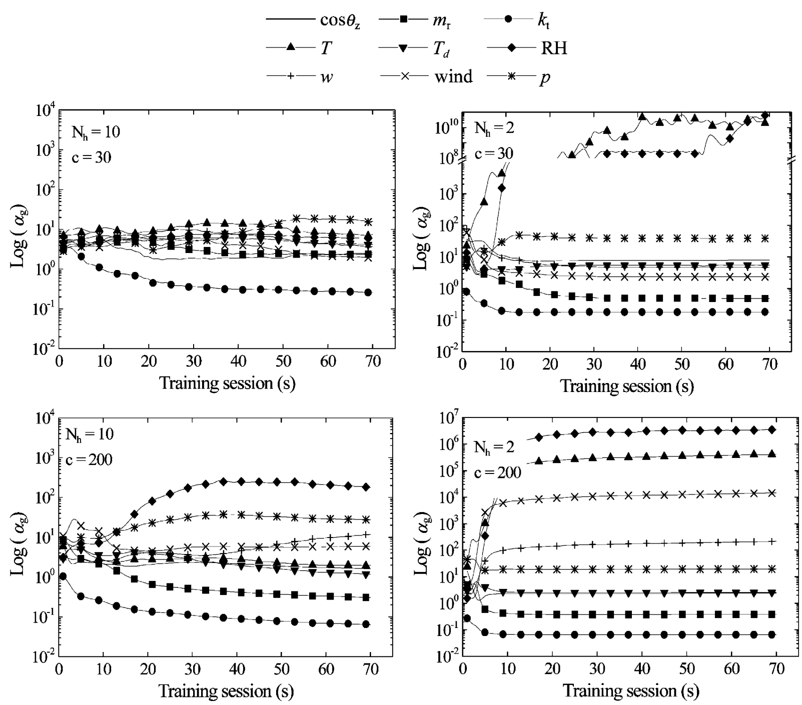

Fig. 3. Log-values of the hyperparameters versus training session, *s*. Two MLPs with 10 and 2 hidden units (N_h) were trained using 30 and 200 training cycles, *c*. The whole database (4880 h) was employed.

inputs and the outputs. In this sense, it is observed how the values of the hyperparameters obtained with MLP10 and trained with only 30 cycles are close to each other, providing no useful information. If training cycles c is increased, the ARD method seems to differentiate the relevance of the inputs in a better way. Therefore, a prescription to employ the ARD method for selecting the more relevant inputs is to use an ANN with few hidden units and try to perform a large number of training cycles. It is important to note that, as opposite to the standard ANN training, increasing the training cycles does not lead to overfit to the data due to the regularization terms introduced via the hyperparameters.

For all cases, the clearness index, k_t, is the more relevant input and the next one is the relative air mass, m_r. It is interesting to note that by using ANN with the ARD method, it is not needed to search the significance of higher degree terms in k_t or m_r (as needed by employing multiple linear regression techniques), since the non-linear relationships between these variables and k_b is implicitly explained by the network. On the other hand, relative humidity appears to be the minor relevant input. Similarly, the temperature hyperparameter is high when MLP2 is utilized, pointing out to discard this variable as input to the k_b–k_t regressions, at least only when data from one climatic condition is involved. Wind speed, pressure, precipitable water and dew point temperature present hyperparameters with two and three orders of magnitude higher than that corresponding to k_t. Therefore, the relative relevance of these variables is almost null against the clearness index and they could be excluded from the model.

However, it is interesting to analyse the relative relevance of these parameters (based on the results given by the MLP2 trained with 200 cycles), to check the reliability of the ARD method. Dew point temperature is the more significant one among them. Since both precipitable water and dew point temperature account for the attenuation of direct irradiance by atmospheric water vapor, this finding points out that: (a) this process affects slightly the k_b–k_t regressions (as it could be expected), and (b) dew point temperature appears to be more appropriate than precipitable water to be included in an ANN model. On the other hand, the higher relative relevance of pressure against relative humidity or temperature was not expected a priori. The reason for that could be due to the ability of pressure to distinguish cloudy (associated with cyclonic conditions with low pressure values) from cloudless conditions (associated with anti-cyclonic conditions with high pressure values) and, in this way, to explain a few of the variance of direct transmittance. In fact, Kemmoku et al. [25] employed pressure as an input parameter to a multi-stage neural network to forecast daily insolation levels. Nevertheless, the significance level of this parameter (as proved by its corresponding hyperparameter value) is too low to perform any improvement in the models.

In order to test the sensibility of the ARD method against the amount of input information, we perform another trial similar to the above but using 1080 h selected by random sampling. The results shown in Fig. 4 agree with the first ones. The clearness index and the relative air mass present the smallest values of the hyperparameters. However, although the hyperparameter of the relative humidity for MLP2 is large, pressure becomes the minor relevant input, as a possible consequence of the decreasing information utilized and the low significance level exhibited in the above trial. Moreover, the hyperparameters associated with T_{d}, w, wind speed and $\cos\theta_z$ increase their values around two orders of magnitude against those presented using the whole data set, corroborating their low relevance. These outcomes point to ARD like a robust method to determine the most relevant inputs when only few data are available.

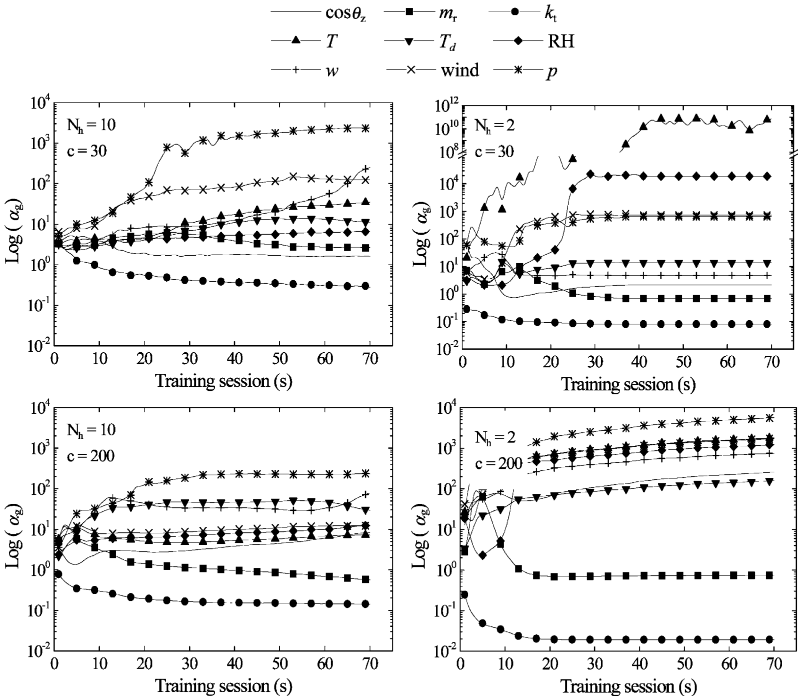

Fig. 4. Log-values of the hyperparameters versus training session, s. Two MLPs with 10 and 2 hidden units (N_h) were trained using 30 and 200 training cycles, c. The whole database (1080 h) was employed.

Finally, it is important to note that although $\cos\theta_z$ and w are found as irrelevant variables, this is only because of m_r and T_d were more suitable to be computed by the ANN. If these latter parameters have been removed, then $\cos\theta_z$ and w would become more relevant.

5. Conclusions

The present work has shown the powerful nature of the novel Bayesian method of the automatic relevance determination to evaluate the more relevant input parameters in modelling direct solar irradiance by using ANNs. In this sense, the relevant variables for estimating the direct irradiance are the clearness index and the relative air mass, which are in agreement with the existing studies, demonstrating the reliability of the ARD method. We also show that the best option to obtain the more relevant input variables from the ARD method is to use ANN with a low number of hidden units.

On the other hand, we have found that this methodology can be applied to locations with limited amount of measurements and avoid the delicate problem of dividing the data set into three subsets (training, validation and test sets). Under these conditions, ARD provides reliable information about the more relevant inputs. In addition, the use of Bayesian framework for ANN is portable for modelling any radiometric variables from other parameters, providing the best option against traditional statistical techniques and other ANN based models.

Acknowledgements

This work was accomplished thanks to the project (REN/2001)-3890-C02-01/CLI of the Ministerio de Ciencia y Tecnología of Spain. The authors are grateful to SRRB's staff for the facilities offered for providing the SURFRAD databases.

References

[1] Garrison JD. A study of the division of global irradiance into direct and diffuse irradiance at thirty-three U.S. sites. Solar Energy 1985;35(4):341–51.
[2] Maxwell AL. A quasi-physical model for converting hourly global horizontal to direct normal insolation. Report SERI/TR-215-3087, Solar Energy Research Institute, Golden (CO); 1987.
[3] Reindl DT, Beckman WA, Duffie JA. Diffuse fraction correlations. Solar Energy 1990;45:1–7.
[4] Perez R, Iniechen P, Maxwell E, Seals R, Zelenka A. Dynamic models for hourly global to direct irradiance conversion. ISES World Congress, Denver, CO. 1991.
[5] López G, Rubio MA, Batlles FJ. Estimation of hourly direct normal from measured global solar irradiance in Spain. Renewable Energy 2000;21:175–86.
[6] Elizondo D, Hoogenboom G, McClendon RW. Development of a neural network model to predict daily solar radiation. Agricultural and Forest Meteorology 1994;71:115–32.
[7] Williams DB, Zazueta FS. Solar radiation estimation via neural network. In: ASAE, editor. Sixth International Conference on Computers in Agriculture, Cancun, Mexico. 1994, p. 140–6.
[8] Mohandes M, Rehman S, Halawani TO. Estimation of global solar radiation using artificial neural networks. Renewable Energy 1998;14:179–84.
[9] López G, Martínez M, Rubio MA, Batlles FJ. Estimation of hourly global photosynthetically active radiation using artificial neural networks. Agricultural and Forest Meteorology 2001;107:279–91.
[10] López G, Martínez M, Rubio MA, Tovar J, Barbero J, Batlles FJ. Estimation of the hourly diffuse fraction using a neural network based model. In: Instituto Geofisico do Infante, Luis D, editors. Second Asamblea Hispano-Portuguesa de Geodesia y Geofísica, Lagos, Portugal. 2000, p. 425–6 (in Spanish).
[11] López G, Martínez M, Rubio MA, Batlles FA. Estimación de la radiación global solar diaria mediante un modelo de red neural. In: López A, López R, López E, Casares F, Universidad de Córdoba, editors. Ninth Congreso Ibérico de Energía Solar, III Jornadas Técnicas sobre Biomasa, Córdoba, Spain. 2002, p. 115–8 (in Spanish).
[12] MacKay DJC. Bayesian non-linear modelling for the energy prediction competition. ASHRAE Transaction 1994;100:1053–62.
[13] Neal RM. Bayesian learning for neural networks. Lecture notes in statistic. New York: Springer-Verlag; 1996.
[14] Rapti AS. Atmospheric transparency, atmospheric turbidity and climatic parameters. Solar Energy 2000;69(2):99–111.
[15] Rumelhart DE, Hinton GE, Williams RJ. Learning internal representations by error propagation. In: Rumelhart DE, McClelland, PDP Research Group, editors. Parallel distributed processing: explorations in the microstructure of cognition. Foundations, vol. I. Cambridge (MA): MIT Press; 1986, p. 318–62.
[16] Lampinen J, Vehtari A. Bayesian approach for neural networks—review and case studies. Neural Networks 2001;14:257–74.
[17] Penny WD, Roberts SJ. Bayesian neural networks for classification: how useful is the evidence framework? Neural Networks 1999;12:877–92.
[18] Vivarelli F, Williams KI. Using Bayesian neural networks to classify segmented images. Fifth IEEE International Conference on Artificial Neural Networks. 1997, p. 263–8.
[19] Thodberg HH. A review of Bayesian neural networks with application to near infrared spectroscopy. IEEE Transactions Neural Networks 1996;7:56–72.
[20] Moller M. A scaled conjugate gradient algorithm for fast supervised learning. Neural Networks 1993;6(4): 525–533.
[21] MATLAB, The MathWorks, Inc. Natick, MA; 1999.

[22] Bishop C. Neural networks for pattern recognition. London: Oxford University Press; 1995.
[23] Neural Computing Research Group, NETLAB Neural Network Software. Aston University: Birmingham, UK. See also: http://www.ncrg.aston.ac.uk/netlab/.
[24] Augustine JA, DeLuisi JJ, Long CN. SURFRAD—a national surface radiation budget network for atmospheric research. The Bulletin of the American Meteorological Society 2000;81:2341–58.
[25] Kemmoku Y, Orita S, Nakagawa S, Sakakibara T. Daily insolation forecasting using a multi-stage neural network. Solar Energy 1999;66(3):193–9.