IMPROVED GENERATION OF HOURLY SOLAR IRRADIATION ARTIFICIAL SERIES USING NEURAL NETWORKS*

L. Hontoria†, J. Riesco‡, P. Zufiria‡, J. Aguilera†

†Grupo Jaén de Técnica Aplicada. Dpto. De Electrónica. Universidad de Jaén
Alfonso X El Sabio, 23700 Linares.(Jaén) (Spain)

‡Grupo de Redes Neuronales
Dpto. De Matemática Aplicada a las Tecnologías de la Información E.T.S.I.T.-U.P.M.
Ciudad Universitaria s/n. 28040 Madrid (Spain)

Abstract

This paper presents a new neural network approach for the generation of synthetic hourly irradiation series, a relevant problem in the photo-voltaic field. The neural model employed is the well known Multi-Layer Perceptron (MLP) paradigm, in a feedback architecture, using a record of historical values for the supervised network training. The method is based on the MLP ability to extract, from a sufficiently general training set, the existing relationships between variables whose interdependence is unknown a priori. Simulation results are compared to other methods, and show that the generated values follow the average tendency of the real values. Though the method has been developed using data values from Madrid, it can be generalised to any location. Even more, the proposed methodology is of general applicability to the estimation of highly complex temporal series.

I. INTRODUCTION

The synthetic generation of hourly or daily solar irradiation values is often the only practical way to obtain radiation data for any given location. Measured historic sequences are not available for most of the countries and regions, and if they exist, they usually have gaps in the records, or are of insufficient length.

Several mathematical radiation models and methods have been developed [2, 3, 4] to generate sequences of values, which try to preserve the same statistical properties (average, variance, type of noise probability density function, etc.) and sequential characteristics (autocorrelation function) as those of the historical records (i.e., those observed in nature). The output of these models may be used for example for the construction of typical meteorological years or to provide computer-generated data sequences for the analysis and design of photo-voltaic (PV) converters, which is usually performed using numerical simulation tools [4, 11]. For the study of PV systems with a high storage capacity, daily radiation data will usually suffice as the storage attenuates the effects of hourly variations; but for PV systems with one or two-hour response time, such as peak plants or PV plants which return energy to the network at maximum charge instants, hourly series are required.

The radiation models can be classified into two groups: spectral and *universal* models. The first ones have the inconvenient that are local and can be only implemented in very few locations, where some data are available. However the universal ones can be used (theoretically) in any site. The models proposed by Aguiar and Collares [2] and by Graham and Hollands [3, 4] can be included in this second group.

The models proposed by Aguiar and Collares [2] and by Graham and Hollands [3, 4], referred from now on as AC and GH respectively for short, may be considered as paradigms in the field of hourly [2, 4] radiation modelling. They are auto-regressive first-order models [14], based on a stochastic disaggregation methodology, that generate hourly irradiation series making use of daily values. These values can be obtained from historic records (which are more common than hourly records) or using daily generation methods [3] in turn (which are more validated than hourly methods). Complex empirical

* This work has been partially supported by Proyecto Multidisciplinar de I+D 14908 of the UPM, and Project PB-97-0566-C02-01 of Programa Sectorial de PGC of Dirección General de Enseñanza Superior e Investigación Científica of the MEC.

expressions are proposed to relate the hourly and daily values, obtaining the parameters in the formulae via regression analysis [20] over historical data.

The main criticisms to these methods are the high computing requirements to obtain each value of the series, and the geographical location dependency of the method with the place where data has been retrieved for the construction of the model [17] (though they intend to be universal).

In this paper, a new neural network approach using a *Multi-Layer Perceptron* (MLP) [10, 16, 20] in a feedforward-feedback architecture [12] is proposed for the generation of hourly solar radiation series, and its results are compared to the AC and GH methods. The approach is based in the MLP capability for approximating any continuous function defined on a compact set within a prescribed error margin. It may be proved that it suffices to employ a MLP with a hidden layer, a required number of neurones and an appropriate training procedure [6]. In practice, selection of appropriate topology as well as training algorithms may become a big challenge [15, 17]. This work improves previous results were an a-priori knowledge of the initial values of the series to be generated was assumed [15, 18, 21].

Another important aspect addressed in this paper is the possibility of employing the presented architecture with a reduced knowledge of the problem to be considered. In that sense the paper defines a simple design methodology with quite general applicability.

The paper is organised as follows. Section II presents some specific aspects related with the generation of irradiation series, the use of MLP based architecture for time series prediction; finally the proposed neural network model is presented. In sections III and IV the main results are shown and compared to the results of other classical methods (AC, GH). The main conclusions and some suggestions for further research are outlined in section V.

II. GENERATION OF SOLAR IRRADIATION SERIES USING NEURAL NETWORKS

II.A. CLEARNESS INDEX

The best currently available radiation models for irradiation series generation use the *atmospheric transmittance* or *clearness index*, (denoted k_t for hourly events and K_T for daily events) as the variable to be modelled instead of the irradiation variable itself. The hourly clearness index is defined as:

$$k_{th} = \frac{G_h}{B_h} \in [0,1] \tag{1}$$

where B_h and G_h are the solar irradiations on a horizontal plane during hour h, outside the earth's atmosphere and on the earth surface respectively. B_h behaves in a deterministic way (depending only on solar time and geographical latitude) and thus it can be accurately calculated for any hour at a given location [11]. It is the atmospheric transparency, k_{th}, that induces randomness to the solar irradiation measured on earth, and its probability distributions behave in a quasi-universal manner [2, 4] (while G_h is obviously specific to each location).

II.B. MULTI-LAYER PERCEPTRON BASED TIME SERIES PREDICTION

Since several years ago, neural networks are increasingly used in different scientific and technical fields [1, 5, 7, 8, 10]. For instance, as a computation and learning paradigm, they can be used for many types of applications. One of the most appealing properties of some neural network paradigms is their potential use for functional approximation purposes [6]. The *Multi-Layer Perceptron* (MLP) is the most widely used type of neural network for approximation tasks; it is classified as a feedforward type neural network, whose topology defines several layers of neurones. The MLP, in static contexts, is usually trained via a supervised procedure, being the existence of a very efficient training method, the backpropagation algorithm [16, 20], one of its great advantages.

Also, in dynamic contexts, the identification and control of non-linear plants, as well as the Time Series Prediction (TSP) have been successfully addressed via feedback architectures of supervised neural models [12, 13, 19].

The problem of TSP via MLP makes use of the time series $\{s_n\}$, for obtaining the function G (assuming that such function exists) which relates each series value with the previous p values:

$$\widehat{s}_{n+1} = G(s_{n-p+1},\ldots,s_n) \cong MLP(s_{n-p+1},\ldots,s_n) \tag{2}$$

By training a MLP with p inputs and 1 output, with a training set representative enough, the MLP will be able to find the desired relationship (in case that it exists) just approximating the function G. Once the approximation is performed, future values can be computed via feedback of the predictions whenever they are available. Such method is called prediction by network evolution.

II.C. PROPOSED GENERATION METHOD

The methodology employed for Times Series Prediction and system identification via MLP in [9, 12, 13, 15, 17, 18, 21], defines the framework of the method developed for the problem addressed here: the generation of hourly clearness index series $\{k_t\}$.

For our computational experiments, a set of hourly irradiation values k_t measured in Madrid between 1978 and 1986 (16 values per day), and its corresponding daily values K_T have been used. As a first approach, in order to evaluate the quality of a generated series, the first 8 years were considered as a *training set*, and the 9th year was employed for testing the validity of the generated series.

The proposed method has been developed via a step by step inclusion of the available associated information. The great advantage of this MLP based methodology is that explicit knowledge of the relationship among all the information sources is not needed. Such information sources can be progressively incorporated in different steps upon the proposed method. The details of this step-by-step procedure can be found in [15, 17, 18, 21]. The final procedure, employing a MLP in a mixed feedback-feedforward configuration, is shown in figure 1.

A day by day prediction method is employed with a dependency of each hourly value on the three previous hours of the same day. Therefore, in order to generate the 16 hourly values $\{kt\}$ of a given day, a method of prediction by network evolution with window p=3 is used; within each day, the window values are initialised to 0 (which are indeed the physical real values).

In order to keep the monthly stationality, an input (d_n) was added to the MLP, containing the distance (days) between the value to be generated and the day with maximum value in the $\{k_t\}$ annual distribution. Another input (h_n) indicates the hour order number of the k_t value (ensuring the hourly stationality). Both inputs (d_n and h_n) are normalised to the range [0,1]. The K_T values were also used as an input to the MLP, since the hourly values of a given day are physically related to the daily clarity index K_T corresponding to that day. Finally, a last input (s) taking the values {0,1}, indicates whether an hour is between sunrise and sunset[1] (k_t should be different from 0) or not (k_t has to be 0).

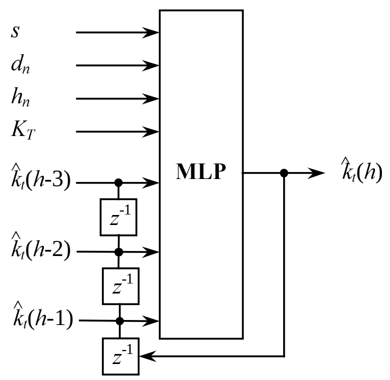

Figure 1 MLP Architecture for Clearness Indexes Prediction

[1] For each day of a year, at a given location, sunrise and sunset hours may be deterministically computed.

The used MLP topology has therefore 7 inputs, a single hidden layer with 15 neurons, and just 1 neuron in the output layer to generate the predicted k_t value. The network layers are fully connected, but with input s directly connected to the output, instead of the hidden layer (feed-through connection). All the neurons have sigmoidal activation functions.

Different optimisation schemes were employed for the supervised training of the network studied. Although some schemes slightly improve the performance and results, the proper selection of the neural model inputs showed to be a more relevant design issue than the training method. The best results were obtained using the backpropagation algorithm [16, 20] (with momentum term and random data presentations), eventually combined with second order optimisation methods [15] during the last few epochs.

One of the great advantages of employing the proposed methodology for the generation of radiation series is that once the method is developed from historical data of a prescribed place, it may be applied to new places just by repeating the training procedure with their corresponding data. Also, assuming that the values of the series to be predicted behave, as stated previously, in a quasi-universal manner [2, 4], the trained network could be used to generate values for any location (provided enough data, from different locations with differing climate characteristics, were used in the training procedure).

III. RESULTS

In order to test the quality of the method, hourly values series were generated for the 8 training years as well as for the test (9th) year (the later to be used after the MLP training). As a quality measure, the Mean Relative Variance (MRV) was used. This parameter, commonly employed in the digital signal processing community, quantifies the relative error, and it is defined as the quotient between the prediction error signal power and the AC power of the signal to be predicted:

$$MRV = \frac{\sum_h (k_{th} - \hat{k}_{th})^2}{\sum_h (k_{th} - \bar{k}_t)^2} \tag{3}$$

The MRV obtained, after 40 training epochs, was 0.1022 for the test year and 0.1026 (mean) for the training years (see Table 1), proving that the method emulates quite well the deterministic component of the series (see Figure 2). Nevertheless, the shape of the resulting series does not have the characteristic rippling of the real series. This is due to the fact that the employed training set has input/output pairs such that quite different desired output values correspond to the same (or very similar) input values. Therefore, after training, the MLP performs an averaging among such output values.

IV. COMPARISONS WITH OTHER METHODS

Comparisons have been carried out with the AC [2] and GH [4] models for hourly radiation. They suggest that the variation in k_t events consists of two components, a trend (or mean) component and a random component: $k_t = k_{tm} + \alpha$, identifying and characterising the sets $\{k_{tm}\}$ and $\{\alpha\}$ for all possible values of K_T via regression analysis.

The neural series generator can be successfully compared with the trend values (k_{tm}) computed using AC and GH methods, as shown in Table 1.

	MRV	
Method	**Training Years**	**Test Year**
MLP	0.1026	0.1022
GH (Original parameters[4])	0.1523	0.1607
GH (Madrid regressions)	0.1281	0.1253
AC (Original parameters[2])	0.1524	0.1597
AC (Madrid regressions)	0.1484	0.1512

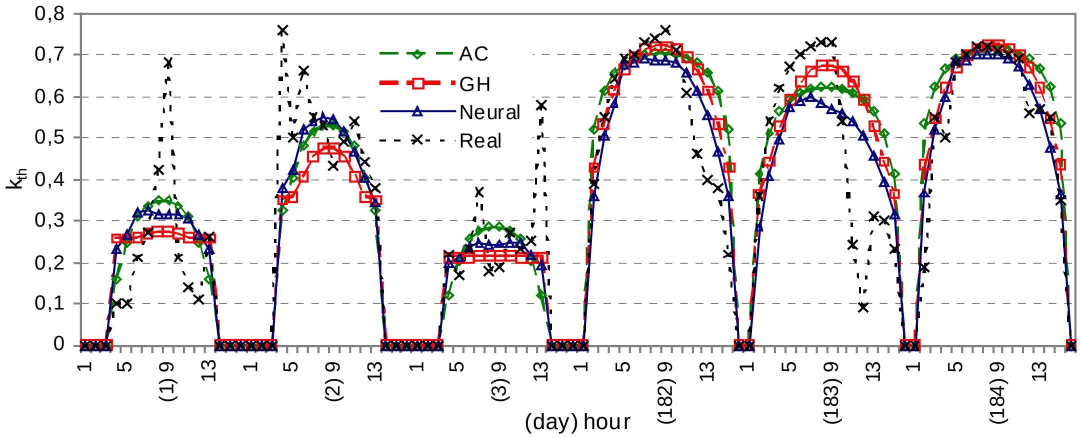

Figure 2. Real Series versus Predicted ones for Days 1-3 (winter) and 182-184 (summer)

Table 1. MRV Values for Different Methods (only trend component)

To make fair comparisons, since the parameters for k_{tm} calculations given in [2, 4] serve for any location (they are universal), a non linear regression over the same data used to train the network was done, hence obtaining specific parameters for Madrid (which are in fact quite different to those in [2, 4]). The results, also given in Table 1, show that in order to get good prediction values with the AC and GH methods, the regressions which link the k_{tm} and K_T values corresponding to each locality must be computed. Furthermore, these computations are far more costly than the neural network training process: 1030 CPU seconds compared to 790 seconds for 40 training epochs (which correspond to the results shown in the table); in addition, MLP with just 2 epochs (45 seconds) provides a MRV under 0.12 (i.e. better than those of the AC and GH methods).

From an academic point of view it is very interesting to note the MLP capability for finding relationships among variables of different nature and that its does not assume any a priori model, being advantageous versus a non-linear regression approach. In our example, making use of an appropriate training set, the MLP was able to relate information from hour, day, daily clarity index value, and 3 previous values of the hourly clarity index in order to generate a new k_t index value.

For the sake of emulation completeness, the stochastic rippling was emulated, as a first approach, using both GH and AC methods for the random component (α). In figure 3 real values for some days and the corresponding simulated values are depicted.

V. CONCLUSIONS

A methodology based on neural networks has been presented for generating time series following the average tendency of the hourly radiation series k_t in a given place. Such methodology is based on the possibility of implicitly employing information associated with the problem, without knowing the existing relationships between different variables and sources of information; i.e. the proposed method does not assume any a priori model, as opposed to the standard approximation techniques where polynomial regression techniques are employed. Furthermore, test results have shown that the neural network approach outperforms those methods.

The proposed methodology could be easily extended to the generation of minute scale radiation series, a field that still lacks a reference theoretical frame and for which data are becoming available nowadays [2], and to the generation of day scale radiation, a field where some good results are being obtained nowadays. The same neural network based framework is also being used for the variance prediction of the k_t random index. Preliminary experiments have shown promising results.

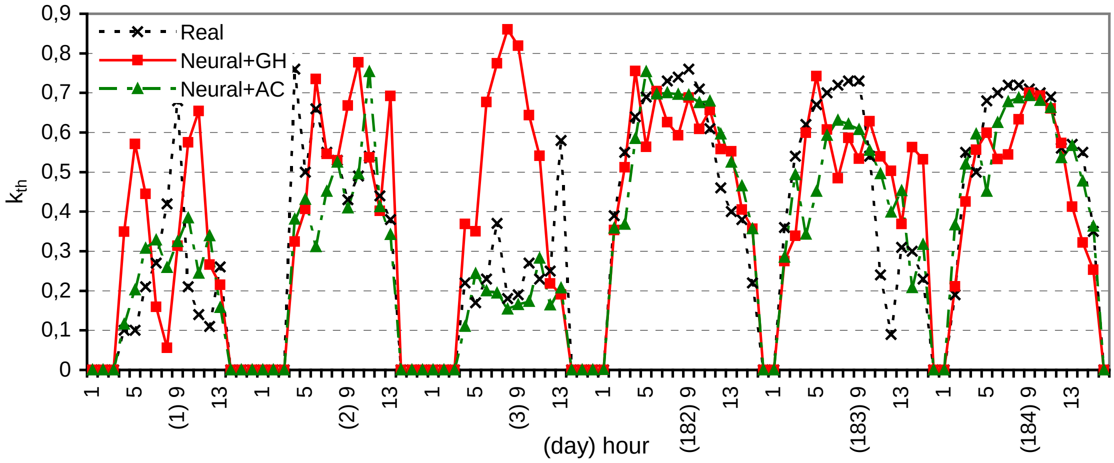

Figure 3. Real Series versus Predicted ones with Rippling. Days 1-3 (winter) and 182-184 (summer)

REFERENCES

1 M. Agarwal. *A* Systematic Classification of Neural-Network-Based Control. IEEE Control Systems Magazine, vol. 17, No. 2, pp. 75-93, April 1997.

2 R. Aguiar and M. Collares-Pereira. TAG : A Time-Dependent, Autorregressive, Gaussian Model for Generating Synthetic Hourly Radiation. Solar Energy, Vol. 49, No. 3, pp. 167-174, 1992.

3 V. A. Graham, K. G. T. Hollands and T. E. Unny. A Time Series Model for Kt with Application to Global Synthetic Weather Generation. Solar Energy, Vol. 40, No. 3, pp. 269-279, 1988.

4 V. A. Graham and K. G. T. Hollands. A Method to Generate Synthetic Hourly Solar Radiation Globally. Solar Energy, Vol. 44, No. 6, pp. 333-341, 1990.

5 S. Haykin. Neural Networks. A Comprehensive Foundation. Macmillan Publishing Company, 1994.

6 K. Hornik, M. Stinchcombe and H. White. Multilayer Feedforward Networks Are Universal Approximators. Neural Networks, Vol. 2, no. 5, pp. 359-366, 1989.

7 D. R. Hush and B. G. Horne. Progress in Supervised Neural Networks. What's New Since Lippmann?. IEEE S.P. Magazine, pp. 8-39, January 1993.

8 T. Kohonen. Self-Organizing Maps. Springer Verlag, Berlin Heidelberg, 1995.

9 A. S. Lapedes and R. M. Farber. Non Linear Signal Processing Using Neural Networks: Prediction and System Modeling. Technical Report, Los Alamos National Laboratory, 1987.

10 R. P. Lippmann. An Introduction to Computing with Neural Nets. IEEE ASSP Magazine, pp. 4-22, April 1987.

11 E. Lorenzo. Electricidad Solar Fotovoltaica. ETSI Telecomunicación (U.P.M. Madrid), 1991.

12 K. S. Narendra and K. Parthasarathy. Identification and Control of Dynamical Systems Using Neural Networks. IEEE Transactions on Neural Networks, vol. 1, n. 1, pp. 4-27, March 1990.

13 K. S. Narendra and K. Parthasarathy. Gradient Methods for the Optimization of Dynamical Systems Containing Neural Networks. IEEE Transactions on Neural Networks, vol. 2, n. 2, pp. 252-262, March 1991.

14 M. B. Priestley. Non-Linear And Non-Stationary Time Series Analysis. Academic Press, 1988.

15 J. Riesco. Optimización Global : Aplicación en el Entrenamiento de Redes Neuronales. U.P.M., E.T.S.I. Telecomunicación, Madrid 1993.

16 D. Rumelhart and J. L. MacClelland. Learning Internal Representations by Error Backpropagation. Chapter 8 from Parallel distributed Processing. Vol.1: Foundations. The MIT Press, 1986.

17 A. Vázquez-López. Identificación de Sistemas mediante Redes Neuronales para Control de Robots. U.P.M., E.T.S.I. Telecomunicación, Madrid 1992.

18 A. Vázquez-López and P. J. Zufiria. Generación Artificial de Series de Radiación Solar Mediante Perceptrón Multicapa. Actas V Conferencia de la Asociación Española para la Inteligencia Artificial (CAEPIA 93), pp. 196-205, 16-18 Noviembre 1993.

19 A. S. Weigend, D. E. Rumelhart and B.A.Huberman. Back-Propagation, Weight-Elimination and Time Series Prediction. Chapter in Proceedings of the 1990 Connectionist models Summer School. Morgan Kaufman, 1990.

20 P. Werbos. Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences. Ph.D. dissertation, Committee on Appl. Math., Harvard Univ., Cambridge, MA, Nov. 1974.

21 P. J. Zufiria, A. Vázquez, J. Riesco, J. Aguilera and L. Hontoria. A Neural Network Approach for Generating Solar Irradiation Artificial Series. Proc. of the (IWANN'99). 4-5 June, 1999. Alicante (Spain).