World Academy of Science, Engineering and Technology
International Journal of Physical and Mathematical Sciences
Vol:1, No:6, 2007

The Multi-Layered Perceptrons Neural Networks for the Prediction of Daily Solar Radiation

Radouane Iqdour and Abdelouhab Zeroual

***Abstract*—**The Multi-Layered Perceptron (MLP) Neural networks have been very successful in a number of signal processing applications. In this work we have studied the possibilities and the met difficulties in the application of the MLP neural networks for the prediction of daily solar radiation data. We have used the Polack-Ribière algorithm for training the neural networks. A comparison, in term of the statistical indicators, with a linear model most used in literature, is also performed, and the obtained results show that the neural networks are more efficient and gave the best results.

***Keywords*—**Daily solar radiation, Prediction, MLP neural networks, linear model

I. Introduction

THE recent emergence of solar energy as an alternative to the more conventional but non-renewable domestic energy source has resulted in a demand for quantitative information on the size of solar energy sources at specified locations. Such data are requested by engineers, architects and designers of solar systems as they attempt to make effective use of the solar energy that is available. The amount of global solar radiation arriving on a horizontal surface is the minimum information needed.

However, in most cases, the users do not have a long registration of solar data in a given location, or for many locations around the world. Indeed, historical records of sufficient details are not available, even the available data are of questionable quality and have a number of missing values.

Being given that the solar radiation is a stochastic process, it is shown that the sequences of the daily radiation can be described and simulated by models like chain of Markov, or Fourier series, by building statistical models [1]. Other models, which have attracted the attention, are those, which use the recorded data to predict the future observations (prediction) [2]. These models are based on the regression approach and are called Autoregressive (AR), Autoregressive Mobile Averages (ARMA) models. These models are in general developed using Second or Higher Order Statistics mainly applicable for Gaussian and stationary processes and require, in principle, long-term meteorological data. Therefore, it is not always possible to predict the actual solar radiation values for a given location. Thus to use these models we must perform to a transformation on the original data in order to have a Gaussian and stationary time series. Unfortunately, this transformation may influence the prediction precision because the optimal prediction is built on the transformed time series [3]. For this reason we propose to use the neural networks to develop a model able to predict the daily solar radiation, without recourse to this transformation and performing, at the same time, the successful results [4].

The field of ANN has a history of some six decades, but has found solid application only in the past fifteen years, and the field is still developing rapidly. Neural networks have been trained to perform complex functions in various fields of application including pattern recognition, identification, classification, speech and control systems [5]. Among the whole of the existing neural networks, we can say that the Multi-Layered Perceptron is the structure, which makes it possible to carry out the most various applications [6].

The use of the MLP neural networks for the prediction of the solar radiation is undoubtedly related to their conceptual specificities. Indeed, these models are flexible systems, which do not require any particular assumption on the nature of the process binding the input and output variables. Their non-linear character enables them to approximate as a broad class of functions provided that the number of neurons used is rather large. These characteristics to which we can add faculties of training and of auto-adaptation are obviously strongly appreciated for the modelling of complex systems [7].

The works that are interested to the prediction of the solar radiation are increased quickly. Negnevitsky and Le have used the neural networks to generate the hourly solar radiation depending on astronomic and meteoroclimatic conditions. Alawi and Hinai have used them to predict solar radiation in areas not covered by direct measurement instrumentation. The input of the network is selected to be the location, month, mean pressure, mean temperature, mean vapour pressure, mean relative humidity, mean wind speed and mean duration of insulation. Mohandes et al used data from 41 collection stations in Saudi Arabia. The network inputs are latitude, longitude, altitude and sunshine duration. The results obtained in these works, show the efficiency of the neural networks to predict the global solar radiation [8]. Whereas, this way to make confronts to the problem of the absence or rarity of

Radouane Iqdour Department of physics, Faculty of Sciences Cadi Ayyad University Po. Box: 2390, Marrakech 40001, Morocco; phone: 212 44434649; fax: 212- 44- 43- 74- 10; (e-mail :r.iqdour@ucam.ac.ma)

Abdelouhab Zeroual Department of physics, Faculty of Sciences Cadi Ayyad University Po. Box: 2390, Marrakech 40001, Morocco.

measure of the variables, such as humidity, pressure, wind speed, or at least one of them (it is the case of the site of Dakhla) at all locations. Even though, the measures of these variables exist, they are very expensive. To surmount this problem, we propose, in this work, to use a new methodology, which permit us to predict, without having to use other meteorological variables, the future observations of the global solar radiation [9][10].

In this work, we applied Multi-Layer Networks (MLP) for the prediction of the daily solar radiation measured in a horizontal surface on the site of Dakhla in Morocco. The training algorithm used to estimate the weights of the neurons is the Polak-Ribiere conjugate gradient algorithm [11].

II. Multilayer Perceptron and the Training Algorithms

The most common neural network model is the MLP. This type of neural network is known as a supervised network because it requires a desired output in order to learn. The notation R-S1-S2-S refers to a MLP with two hidden layers. The first layer have R neurons is called the input layer, the last is the output layer equipped with S neurons and the intermediate layers are the hidden layers with S1, S2 neurons. Each neuron of a layer is connected to all the neurons of the following layer (feed-forward neural network) [12].

We associate a weighting coefficient (synaptic weight) to each connection. These weights are stored in the matrices of weight noted W1, W2 and W3 (example: R-S1-S2-S network). The element (i,j) of a weight matrix represents the connection weight connecting neuron i of the downstream layer to neuron j of the upstream layer. Each layer (except that of input) is connected to a special cell with a constant output of value 1. The corresponding weights are stored in a vector called bias and noted b1, b2 and b3 in the case of two hidden layers. Each neuron i of the first hidden layer computes his input net1[i] and his output (its activation) a1[i] as follow:

$$\text{net1}[I] = \langle W1[i], p \rangle + b1[i] \quad (1)$$

$$a1[i] = F1(\text{net1}[i]) \quad (2)$$

Where W1[i] is the i^{th} line of W1, $\langle\ \rangle$ is the notation for the usual scalar product, p is the input vector and F1 is the activation function associates to the first hidden layer, the activation functions are non linear and of sigmoid type, i.e:

$$F(x) = 1/(1+e^{-x}) \quad (3)$$

The activations stored in the vector a1 are propagated to the cells of the following layer. In a similar way, we calculate the second activation vector:

$$a2[i] = F2(\langle W2[i], a1 \rangle + b2[i]) \quad (4)$$

This mechanism continues to the last layer and makes it possible to obtain the output vector t corresponding to the input p. The network inputs consist of vectors of size R stored in a matrix P with N columns. Each column p of P is associated a desired vector q of the output Q with size S stored in a matrix Q. The outputs computed by the network are stored in a matrix T.

The required goal, is the learning of associations (p,q): the network must restore the desired output q (or an output rather close to Q) when the form p is presented as a input. The training of the MLP networks consists in computing the weights of connections between the neurons in order to minimise a square criterion E:

$$E = \frac{1}{N} \sum_{p=1}^{N} E_p \quad (5)$$

$$E_p = \frac{1}{2} \sum_{j=1}^{S} (Q[j][p] - T[j][p])^2 \quad (6)$$

For the real applications, we do not know the pace of the error function E in the space of the weights, what has as consequence when the non-linear training algorithm converges; we are never assured that the obtained minimum is global.

Backpropagation algorithm was created by generalizing the Widrow-Hoff learning rule to multiple-layer networks and non-linear differentiable transfer functions [13].The standard backpropagation is a gradient descent algorithm in which the network weights are moved along the negative of the gradient of the performance function E.

An iteration of this algorithm can be written:

$$w(k+1) = w(k) - \varepsilon g_w(k) \quad (7)$$

Where:

$g_w(k) = (\partial E/\partial w_1(k), ..., \partial E/\partial w_n(k))^T$: The gradient error evaluated in w(k).

n: The number of the connections of the network.

k : index of the iteration.

$w(k) = (w_1(k), ..., w_n(k))^T$: The weight vector in the iteration k.

ε : Learning rate ($\varepsilon > 0$).

Parameter ε regulates the size of the gradient step. The performance of the algorithm is very sensitive to the proper setting of the learning rate. If the learning rate is set too high, the algorithm may oscillate and becomes unstable. If the learning rate is too small, the algorithm will take too long to converge. The greatest disadvantage of this algorithm is that it does not even ensure convergence towards a local minimum.

The changes of research direction are done the ones per rapport to the others perpendicularly, which generates behaviour oscillatory. We try to attenuate these oscillations by the addition of a term of moment:

$$\Delta w(k) = w(k+1) - w(k) = -\varepsilon g_w(k) + \mu \Delta w(k-1) \quad (8)$$

The additional term privileges the directions of descent where the variations of weight are done in the same direction $(\Delta w(k)\Delta w(k) > 0)$, which causes to filter the oscillations of the algorithm. Another improvement consists in using a learning rate varying during the training process (adaptive) [14].

III. Training Algorithms

The previous section presents two backpropagation training algorithms: gradient descent, and gradient descent with momentum. These two methods are often too slow for practical problems. In this section, we describe the Polack-Ribière algorithm that can converge from ten to one hundred

times faster than the algorithms discussed previously. This faster algorithm belongs to the category of algorithms that use standard numerical optimisation techniques.

A. The Polack-Ribière Algorithm

The basic backpropagation algorithm adjusts the weights in the steepest descent direction (negative of the gradient). In this direction the performance function is decreasing most rapidly. It turns out that, although the function decreases most rapidly along the negative of the gradient, this does not necessarily produce the fastest convergence. In the conjugate gradient algorithms a search is performed along conjugate directions, which produces generally faster convergence than steepest descent directions.

The Backpropagation algorithm is used to calculate derivatives of performance E with respect to the weight and bias variables w. Each variable is adjusted according to the following equation:

$$w(k+1) = w(k) + \alpha(k)\, s(k) \quad (9)$$

Where s(k) is the search direction. The parameter $\alpha(k)$ is selected to minimize the performance along the search direction.

The first search direction is the negative of the gradient of performance:

$$s(0) = -g_w(0) \quad (10)$$

In succeeding iterations the search direction is computed from the new gradient and the previous search direction according to the formula:

$$s(k) = -g_w(k) + \gamma(k)\, s(k-1) \ (k \geq 1) \quad (11)$$

The parameter $\gamma(k)$ can be computed in several different ways. For the Polak-Ribière variation of conjugate gradient it is computed according to:

$$\gamma(k) = \langle g_w(k) - g_w(k-1),\, g_w(k) \rangle / \langle g_w(k-1), g_w(k-1) \rangle \quad (12)$$

B. Procedure of Training and Criteria of Error

In this section, we present the plan of the used training procedure (fig.1). We have employed for the learning phase the Polack Ribière backpropagation algorithm who belongs to the conjugate gradient algorithms type. The phases of initialisation and perturbation of the weights are done according to the following procedure: We start by perturbing randomly several times the origin of the space of the weights, what generates a cloud of points. We mark the point that minimises the error function E. This point will be the centre of a new perturbance whose amplitude is smaller than the previous, and so on. This procedure stops after a number of perturbances fixed of advance and keeps the last centre like initial point for the training procedure. The training algorithm (Polak-Ribière) stops after stabilisation on a local minimum (in practice less than 1000 iterations are sufficient). The procedure retains the best configuration of weight obtained starting from the kmax initialisations (kmax = 50).

The criteria selected to measure the performances of the neural networks are as follows:

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(Y(i) - \hat{Y}(i))^2 \quad (13)$$

$$\text{RMS} = \sqrt{MSE} \quad (14)$$

$$\text{NER} = \frac{RMS}{\sigma(Y)} \quad (15)$$

$$\text{MAPE} = \frac{100}{N}\sum_{i=1}^{N}\left|Y(i) - \hat{Y}(i)\right| \quad (16)$$

Where Y is the original time series, $\hat{Y}$ is the predicted time series and N is the size of the time series Y.

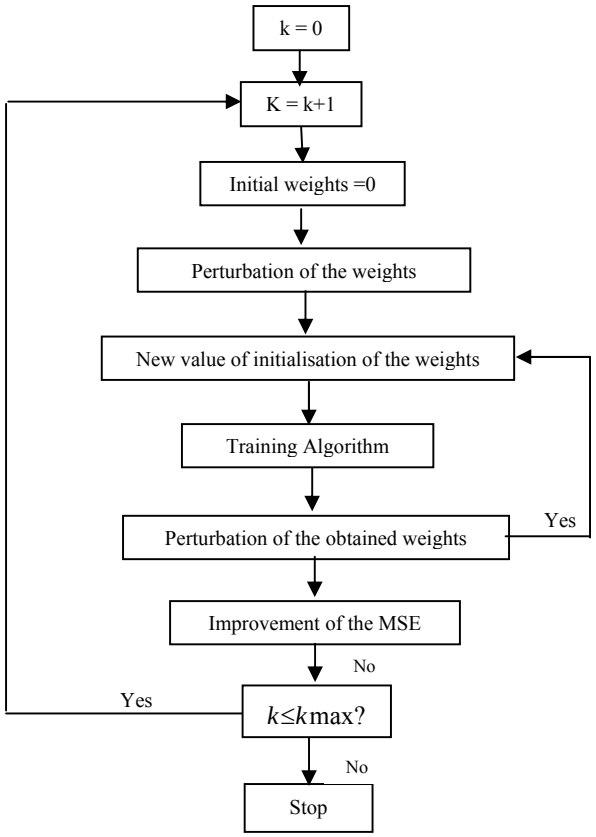

Fig.1 Procedure of training

The choice of an error criterion as measure of the performance is a delicate point because it depends directly on the problem to treat. Many studies compare the performances of the networks on the basis of the Mean Square Error (MSE) and the Root Mean Square (RMS) criterion. But these criteria inform us only at the "distance" separate the predicted series from the observed series. The major disadvantage of these measurements is the dependence on the used scale, what makes them not very reliable in practice. The Normalised ERror (NER) Criterion normalise the RMS per rapport to the measuring unit and thus removes the disadvantage. Contrary to the RMS error, the Mean Absolute Percentage Error (MAPE) criterion penalises less heavily the large errors committed by the network [15].

C. The Structure of the Network

Being given that there are no (or very little) theoretical results usable practically for the determination of neural network architecture, it became usual to proceed empirically on the matter. A commonly allowed empirical rule is that the number of weights in the network cannot exceed the tenth of the number of the training example. Recent developments based on the study of the dimension confirm the recommendation of Widrow [16]. Moreover, generally we use one or two hidden layers, although theoretically, only one hidden layer allows already having the property of universal approximator. We construct various networks with one and two layers hidden by holding account of the Widrow remark. The architecture that will be retained is that which gives the best results on the validation set (Cross-Validation Technique).

The construction of the various networks used follows the following diagram:

The networks with only one hidden layer were constructing by successively adding two additional neurons on this one. This technique, which has of course the advantage of decreasing the number of architecture to be validated, is founded on the intuitive idea that the addition of two neurons instead of only one will not generate too large differences in performances between two architectures consecutively constructed.

For the networks with two hidden layers, we considered only the triangular structures, those for which the number of neurons on a layer is superior to the one of the following layer. This choice is justified by their frequent use in the literature.

IV. Application to Solar Process

It is needed, in practice, for a solar system designer to use daily or hourly radiation sequences recorded during many years. But in most cases we do not have a long registration of solar data in a given location. For many locations around the world, historic records of sufficient detail are either not available or the available data are not good quality and have an appreciable number of missing values. To solve this problem, it is necessary to have an empirical model to generate or predict data having similar features than measured data. For this reason we consider in this work a set of data representing the daily solar radiation recorded during three years on a horizontal surface at Dakhla in Morocco. The set of training and the one of validation are constructed from the measurements of two years. The measurements of the third year are left for the test phase.

In fig.2, we represent the times series of daily solar radiation data recorded over one year (Y(t) t=1,…,365). We can remark easily a non-linear "progression", which indicates that this time series come from a "non stationary" process. Thus, to use the second or the high order statistics to model this process, we need to make the time series stationary. Unfortunately, this transformation is responsible, in many cases, of some errors in the prediction or the generation of the data. Furthermore, this transformation may cause the loss of a part of useful information.

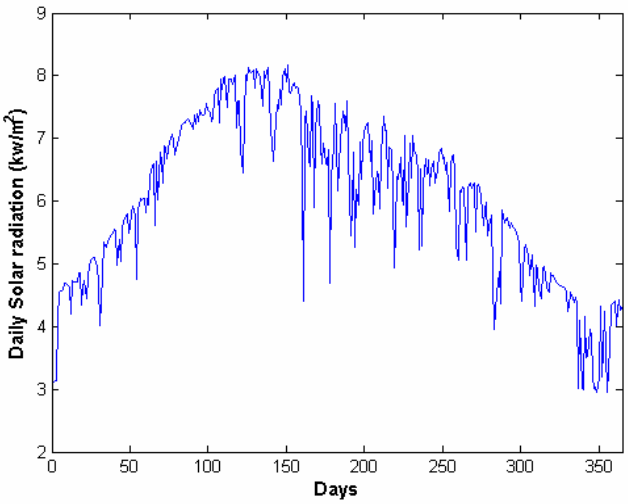

Fig. 2 Annual variation of daily global solar radiation

A. Inputs and Outputs of the Network

Each pattern of training is constituted of ν delayed values $Y(t-k\tau)$ $(0\leq k\leq \nu-1)$ and the corresponding desired output $Y(t+\tau)$. The parameter τ is the sampling period of the time series and represents the number of days separating two consecutive observations of the time series. The minimal value of τ is imposed by the structure of the database and is equal to one day in our case.

We use a neural network as predictor of order ν: from ν values $Y(t)$, $Y(t-\tau)$,…, $Y(t-(\nu-1)\tau)$, we try to estimate $Y(t+\tau)$. For the moment, we take the value of the delay time τ to one day and the width of window ν to 5 days. Each displacement of the window on {Y(t)} generates an additional training pattern (Sliding Window Technique, Time Delay Neural Network).

B. Results and Discussion

The table I, presents the obtained results on the validation set for eight selected networks. We remark that the network most efficient is network 5-3-2-1. We notice also that the differences in performances between the selected networks go from 0.7% to 7% for RMS, from 5% to 11% for NER and for 4% to 13% for MAPE. With regard to the convergence speed we note that the Polak-Ribière algorithm converges quickly (see table. I). Indeed, it needs less than 10 s to converge and the time of convergence increases proportionally with the number of neurons in the hidden layers.

TABLE I
THE PERFORMANCES OF THE MLP NEURAL NETWORKS ON THE VALIDATION PHASE

Networks	Time(s)	RMS	NER	MAPE%
5-0-1	4.42	0.4551	0.1629	31.96
5-2-1	5.94	0.4452	0.1593	31.55
5-4-1	6.22	0.4437	0.1588	30.33
5-6-1	6.84	0.4366	0.1562	30.05
5-8-1	9.55	0.4546	0.1627	31.69
5-3-2-1	8.48	0.4256	0.1503	29.83
5-4-2-1	9.93	0.4424	0.1583	30.49
5-4-3-1	9.31	0.4433	0.1586	30.18

Once the appropriate network is selected and its parameters are estimated, the identified MLP neural network model is used for the test data. In fig.3, we present the evolution of measured and predicted times series of the daily solar radiation. We notice that the two time series have the same behaviour.

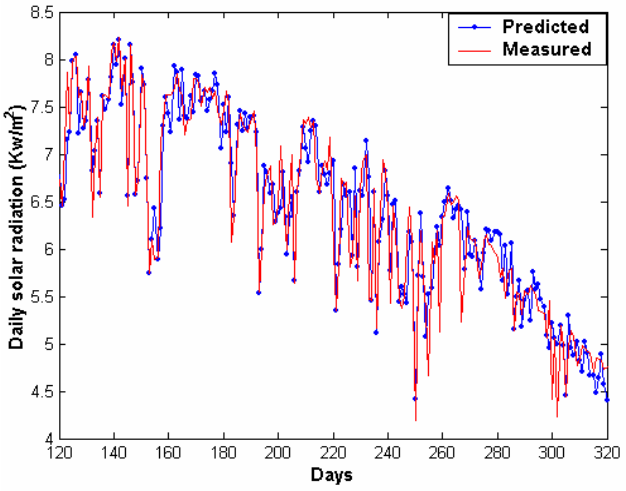

Fig. 3 measured and predicted daily solar radiation

To show the impact of the initialisation method on the training, we represent on the fig.4 the evolution of the MSE calculated, for the network 5-3-2-1 on the Training set for 50 different initialisations of the weights. The maximum deviation obtained is 5 % which is considerable.

In order to have an idea on the influence of the following parameters: the size of the window (ν) and the period of sampling (τ). We have preceded a modification of the values of these parameters, the tables II and III show the obtained results.

TABLE II
INFLUENCE OF THE DELAY (τ) ON THE RESULTS

τ	1	2	3	4	5
MSE	0.1812	0.2815	0.2919	0.3181	0.3011

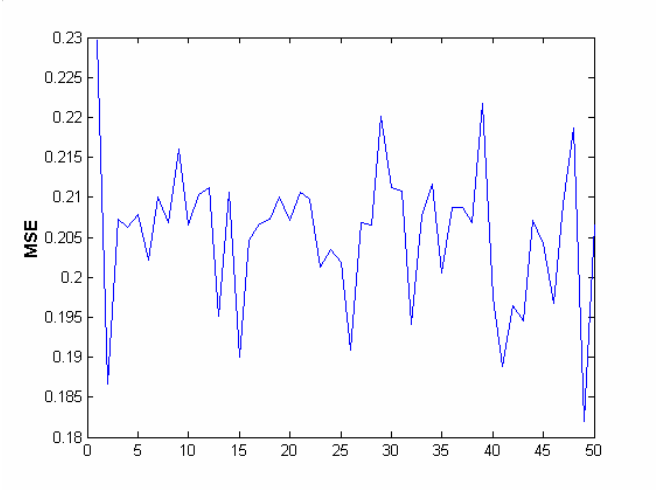

Fig. 4 MSE after training of the neural network starting from 50 different initialisations of the weights

According to the table II and table III, we notice that the parameter τ do not have a great influence on the results except for τ =1. This can be due to the lack of information at the time sampling higher than 1. Whereas for the second parameter (ν) we note that more the size of the window is large more the results are better. This can be explained by the fact that there is more information in the input of the networks what enables it to better predict the daily solar radiation.

TABLE III
INFLUENCE OF THE WINDOW SIZE (ν) ON THE RESULTS

ν	2	3	4	5	6
Network	2-4-2-1	3-3-2-1	4-4-2-1	5-5-3-1	6-5-4-1
MSE	0.2268	0.1944	0.1908	0.1834	0.1671

C. MLP Neural Networks vs. Regression Model

The prediction is very useful in solar energy applications because it permits to generate solar data for locations where measurements are not available. This section compares the quality of prediction using the identified MLP neural network techniques and linear ARMA model.

To identify the linear ARMA model able to generate solar data with similar character than those recorded; we use the Box-Jenkins procedure. This methodology is composed in three stages: the model identification, the parameters estimate and the model validation.

The first stage consists on determining the ARMA model order on the basis of the graphs of the Autocorrelation Functions (ACF) and the Partial Autocorrelation Function (PACF). In the second stage we estimate the model parameters, several methods are employed in the literature, the method of least squares is the simplest and mostly used. The third stage aims at to know if the model chosen in the identification stage can be considered as valid. In the affirmative case, the residual time series is a realisation of a

white noise [17].

Before using the Box-Jenkins method to identified a ARMA linear model it is necessary to transform the time series of the daily solar radiation (Y(t)) to a stationary process using differentiation operator ∇^p (differentiation operator of order p).

After filtering the low frequency component witch is responsible of the non stationarity of the time series Y(t) (differentiation of the time series Y(t)), the time series obtained Z(t) seems to be stationary ($Z(t) = \nabla^p Y(t)$ with p =1). The methodology of the Box-Jenkins is then applied on the Z (t).

TABLE IV
THE AIC CRITERIA FOR FIVE SELECTED MODELS

	MA(1)	ARMA(2,1)	ARMA(1.2)	MA(2)	ARMA(1,1)
AIC	503.71	502.68	506.75	499.25	503.59

The pace of the ACF and PACF functions does not permit to identify only one model. So we brought to consider several models. In order to choose the most adequate model we compute the Akaike Information criterion (AIC).

The Table IV presents five selected models and their corresponding AIC criterion. According to this table, we can conclude that the most adequate model is the Moving Average (MA) of order 2.

The identified linear model (MA (2)) is then used to generate the daily solar radiation data measured on the site of Dakhla. The comparison between the regression model and the identified MLP neural network model proves the superiority of the MLP neural network in the prediction of the daily global solar radiation. Indeed, the value of the RMSE = 0.56 kw/m^2 gives by the MLP neural network model is less of 10% of this produces by the regression model.

V. CONCLUSION

In this paper we have studied the possibilities of the MLP neural networks for the prediction of daily solar radiation. Several architectures were tested by explaining the difficulties related to the manipulation of these models. The Polack-Ribière algorithm is employed for the identification of the parameters of the networks. The performances obtained were compared on the basis of criterion RMS, NER and MAPE.

A new approach is proposed for the prediction of the future observations of the daily solar radiation, without using other meteorological variables. This new methodology can be used in unfavourable conditions, in terms of limited amount of available data, performing successful results. So we can conclude that the MLP neural networks are able to fit the daily solar radiation data, and can be used to fill missing data in daily solar radiation databases. Additionally, the proposed model can be generalized and used in different locations.

In order to have an idea about the impact of the sampling period and the window size parameters on the precision of the prediction, we carried out a modification of these parameters; the results show that the two parameters influence the prediction capacity of the MLP neural networks.

A comparison between the MLP neural networks and the conventional model MA is then performed. The obtained results show that the MLP neural networks, and thus the new methodology improve the accuracy of the prediction

REFERENCES

[1] H. Bouhadou, MM. Hassani, A. Zeroual, Wilkinson AJ, "Stochastic Simulation of weather data using higher order statistics", *Renewable Energy* , 12(1), 1997, pp. 21-37.

[2] S. Safi, A. Zeroual, "MA system identification using higher order cumulants: application to modeling solar radiation", *Journal of Statistical Computation and Simulation*, Vol.72 (7) , 2002, pp. 533-548,.

[3] S. Safi, A. Zeroual, MM. Hassani, "Prediction of global daily solar radiation using higher order statistics", *Renewable Energy*, 27, 2002, pp. 647-666.

[4] S. Safi, A. Zeroual, "Modelling solar data using high order statistics" *A.M.S.E. Advances in Modelling & Analysis*, Vol. 6, N°1, 2, Advances D-2001, pp. 1-16..

[5] M. T. Hagan, H. B. Demuth, and M. H. Beale, Neural Network Design, Boston, MA: PWS Publishing, 1996.

[6] J.J. Hopfield "Neural Networks and Physical Systems with Emergent Collective Computational Abilities", Proceeding of the Natl. Acd. Sci., n°79, 1982, pp. 25554-2558.

[7] Hornik K., Stinchcombe M., White H. "Multilayer Feedforward Networks are universal Appoximators" *Neural Networkd*, n°2, pp .359-366.

[8] Soteris A. Kalogirou "Artificial neural networks in renewable energy systems applications: a review" *Renewable and Sustainable Energy Reviews* 5, 2001, 373–401.

[9] R. Iqdour, A. Zeroual, "Prédiction de l'irradiation journalière à l'aide des réseaux de neurones MLP", *International Conference on Approximation Methods and numerical Modeling in environment and Natural Resources MAMERN 2005,* May 9-11, 2005 Oujda - Morocco.

[10] R. Iqdour, A. Zeroual, "An application of the MLP neural networks to the prediction of daily solar radiation", *IVème Conférence Internationale en Recherche Opérationnelle Théorie et Applications CIRO 05*, 22-26 Mai 2005 Marrakech – Morocco.

[11] E. Polak, *Computational Methods in Optimisation: a Unified Approach*, Editions Academic Press.

[12] J. Dayhoff *Neural Networks Architectures*, Editions Van Norstrand Reynold.

[13] B. Widrow, "Adaline and madaline", *Proceedings of the 1st International Conference on Neural Networks*, pp.143-158.

[14] T. Tollenaere, "Fast Adaptive Backpropagation with good Scaling Properties" *Neural Networks*, n°3, pp.561-573.

[15] J.S. Armstrong, F. Collopy, "Error Measures for Generalzing About Forecasting Methods: Empirical Comparison ", *International Journal of Forecast in*, n°8, pp.69-80.

[16] D.R.Hush, B.G. Horne, "Progress in Supervised Neural Networks", *IEEE Signal Processing Magazine*, n°10, pp.8-39.

[17] G.E.P. Box, G.M. Jenkins, *Time Series Analysis*, *Forecasting and Contro*l, Editions Holden Day, 1970.