Generation of Horizontal Hourly Global Solar Radiation From Exogenous Variables Using an Artificial Neural Network in Fes (Morocco)

H. Loutfi*‡, A. Bernatchou*, R. Tadili*

*Laboratory of Solar Energy and Environment, Faculty of Sciences, University Mohamed V -B.P. 1014 Rabat-Morocco
(Loutfi1964@gmail.com,bernatchou@gmail.com, rtadili@yahoo.fr)

‡Corresponding Author; H. Loutfi, Rue Farhat Hachad, Res. Mabrouka C Apt.13 Ocean- Rabat-Morocco, Tel: +212 (0) 6 10 90 63 11, hloutfi1617@gmail.com

Received: 24.12.2016 Accepted:04.05.2017

Abstract-This paper presents a comparison between multilayer perceptron (MLP) and neural autoregressive with exogenous inputs(NARX) for generating hourly global solar radiation in the city of Fes(Morocco).Results from this analysis are crucial for the design and sizing of any solar energy system.MLP and NARX were created, trained and tested using MATLAB. Four hourly measured variables over five years and one calculated parameter were used as input to the models and horizontal hourly global solar radiation as target. Models with different structures, especially, different combinations of inputs as well as different numbers of hidden neurons were implemented. To evaluate these models, the regression coefficient (R^2)and two error statistics, namely, normalized root mean square error (nRMSE) and normalized mean bias error (nMBE) are used. According to theses statistics, the best model is NARX with 5 inputs. The test of this model over unseen data and its ability to produce authentic forecasts show a good accuracy and prove its generalization capability (nRMSE=15%,nMBE=0.036%and R^2=0.95). For the wide use of the proposed model with available data, different sizes for different periods of data were used for the learning process. Depending on the accuracy required for the generated values, our model gives quite good results with relatively small sets of training data. As such, the proposed model shows a good ability to generate hourly solar radiation values from more available and cheaper data namely temperature and relative humidity.

Keywords- Global solar radiation; artificial neural network;modelling; prediction; NARX

1. Introduction

Morocco is characterised by an intensive solar irradiation, as it lies in a sunny belt, which favours the utilization of solar energy. Indeed, Morocco has launched an ambitious renewable energy programme in 2010,costing around nine billion dollars to be completed by 2020. This project will generate 2,000 megawatts of solar generation capacity. The installation of any solar power system requires high quality solar radiation measurements in order to size and simulate the system’s functioning. Lack of long series of data or poor quality data series can combine errors in plant design, sizing and performance forecasting; thing that impacts negatively on investment. Unfortunately, measures of solar radiation are usually in accurate and rare over the world[1]; especially in Morocco, due to the measuring devices price. There is only a small number of solar radiation stations, that is why there is a lack of solar irradiation measurements over large areas.

However, other meteorological parameters such as ambient temperature, humidity, wind speed are relatively easily measured in a larger number of meteorological stations with relatively low cost. On the other hand, sizing correctly a solar system or simulating its performance require at least, daily or, even better, hourly solar irradiation. Hence, it seems that elaborating relationships between available meteorological data and the solar irradiation ones can be benefit. Previous studies show that Artificial Neural Networks (ANN) are particularly suitable to reach this goal. Recently, (ANN) models have been used in solar radiation modelling for many locations with different climates. Pertaining researches have been done in countriessuch as Greece, Saudi Arabia, Turkey, China, Egypt, Cyprus, Spain, India, Oman, Algeria, the UK and Malaysia [2–35].But no work seems to exist in Morocco. Thus, the purpose of this paper is the generation of horizontal hourly solar irradiations, using ANN’s models, from calculated astronomical variables and measured meteorological data which are cheaper and more available.

The paper is organized as follows. In the first paragraph, we will present a bibliographical review that illustrates the ability of artificial neural networks (ANN) to elaborate non-linear relationships between input and output data. Such relations were developed between meteorological parameters and solar irradiations for different time scales especially monthly , daily or hourly mean values but rarely for short-time step data. In a second paragraph, we will focus on two types of ANN : MLP and NARX models to be compared and statistic parameters for model's evaluation will be presented.

In a third part, we will give a description of our data measurements station where nine meteorological variables are measured. A database will be defined and ten different combinations of five variables will be tested as input to the ANN models. Then, the methodology developed to address different issues will be applied and results will be discussed. At last, conclusions and perspectives for further studies will be presented.

2. Bibliographical Review

Modelling or estimating solar irradiation methods can be widely classified in three categories:

- Models that take into account irradiative transfer modes and solar radiation and earth atmosphere exchanges such as Rayleigh diffusion, absorptions by ozone, aerosols and water vapour [36-39]. In addition to the fact that these models are complex, they only estimate solar irradiation in clear sky conditions.
- The second category concerns methods that uses empirical relations between clearness index, Kt, and sunshine ratio; Kt is defined as a ratio of horizontal global solar irradiation, on extraterrestrial irradiation while the sun ratio is the sunshine duration divided by the theoretical day length. Generally, most of these methods were not very accurate as they used high time steps or averaged data [36],[40-41]. Stochastic models have been also applied at different time scales [42-44].
- The most recent category is based on artificial intelligent methods. Indeed, Artificial Neural Network, ANN, were developed [14], [45-46] for solar radiation studies either for forecasting data series, estimating solar irradiation from exogenous meteorological data or for extrapolating solar irradiation from data measured on other sites.

Our research is part of this last category: making a choice between Multi-Layer Perceptron (MLP) or neural autoregressive with exogenous inputs (NARX) models to estimate hourly global solar irradiation on a horizontal plane, I_g, using exogenous meteorological data which are more available and less costly such as temperature, relative humidity and wind speed.

Indeed, the ANN approach is capable to find both linear and nonlinear relationships between inputs and outputs [47]. ANN models acquire knowledge through the training phases and store this knowledge within inter-neuron connection strengths called synaptic weights.

The ANN models' performance depends on the choice of the best combination of weather variables as input, training algorithm and ANN architecture design. The most important key task in time series prediction is the selection of the input variables, especially that, for non-linear ANN models, there is no systematic approach to adopt [48].

3. Design of the Artificial Neural Network Model

3.1. Neural network

An artificial neural network is an information processing system that is non algorithmic and massively parallel. It is composed of layers of parallel units called neurons. These neurons, being connected by a large number of weighed links, receive inputs over their incoming connections, perform non linear operations generally and output the final results. They have been applied in various aspects of science and engineering [49-57].

There are two major categories of ANN: feed-forward and feedback (recurrent)networks. The main difference between these categories is the existence of one or more loops in recurrent models. While feed-forward networks are organized into layers connected strictly in one direction from the first layer to the last one [58].

3.2. Multilayer Perceptron (MLP)

MLPs are the most commonly used type of feed-forward networks. A schematic diagram of the basic architecture is shown in Fig. 1a. Each neuron k in the hidden layer (fig. 1b) sums up its inputs x_i after weighting them with the strengths of the respective connections w_{ki} from the input layer and calculates its output y_k as follows:

$$y_k = \mathrm{f}\left(\sum_{n=1}^{N} \mathrm{w_{ki}}\, \mathrm{x_i}\right) \qquad (1)$$

f is a transfer function that can be a sigmoid, hyperbolic tangent or radial basis function. The final output in the last layer is computed similarly.

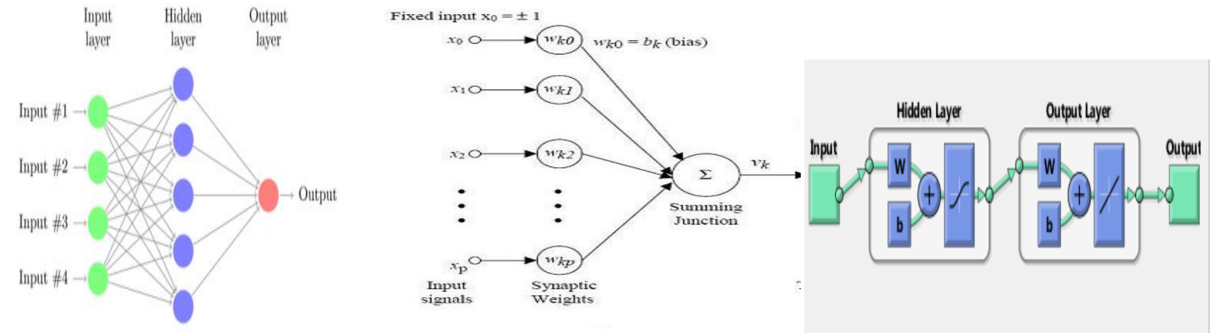
The neural network adopted in our study is a multi-layer feed-forward back-propagation network. It was designed and trained using MATLAB's code and MATLAB's neural network toolbox. A simplified schematic diagram of this network is shown in Fig. 1c; the main characteristics of this model to be mentioned:

-There is one hidden layer (the user can change the numbers of hidden neurons)

-The transfer function adopted is a sigmoidal while the output node has a linear activation function

-The training algorithm is back-propagation based on a Levenberg-Marquardt(LM) minimization method which is the most commonly used [59].

-The learning procedure is controlled by a cross-validation technique based on a random division of the initial set of data in 3 subsets (training, validation process control and testing).



(a) (b) (c)

Fig. 1. a) Multi-layer perceptron (MLP) network **b**) Architecture of an artificial neuron **c**)Diagram of the used MLP in MATLAB

3.3. Recurrent models:NARX

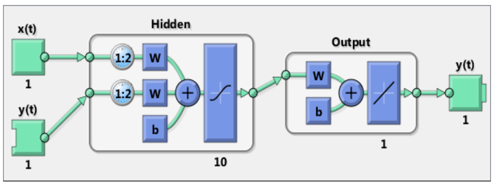
From recurrent dynamic networks we adopted the nonlinear autoregressive network with external inputs (NARX). The equation characterizing NARX model is:

$$y(t) = f\begin{pmatrix} y(t-1), y(t-2), \dots, y(t-d), x(t-1), \\ x(t-2), \dots, x(t-d) \end{pmatrix} \quad (2)$$

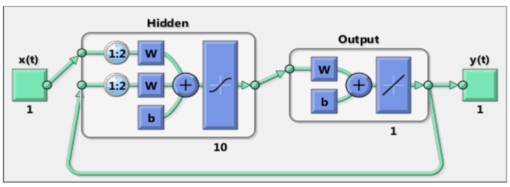
Where *y*(*t*) is the output signal, x(t) the input signal, t denotes time and d the delay. *y*(*t*) is regressed on its previous values and previous values of an independent input signal x(t). NARX model can be carried out by using a feed-forward neural network to approximate the function *f*. Fig. 2a illustrates a diagram of the resulting network.

The output of the NARX network can be considered as an estimate of the output of a nonlinear dynamic system to be modeled. This output is fed back to the input of the network as part of the standard NARX design (Fig. 2b). The true output being available during the training, it will be benefit to create a series-parallel architecture, in which the true output is fed back instead of the estimated one (Fig. 2a). This ensures that the resulting network has a purely feed-forward design and static back-propagation can be adopted for training.

Here also, the activation function for the hidden layer is a sigmoid while for the output unit a linear function is used.



(a)



(b)

Fig. 2. a) Diagram of NARX series- parallel architecture **b**) Diagram of NARX closed loop architecture

The training process is done in open loop with the LM algorithm.

In order to accomplish multistep-ahead prediction, once the learning achieved, the architecture is converted to closed loop. For this network '*d*' initial inputs and outputs are needed as initial conditions [59]. In this study, the delay number was fixed at the value of two to reduce the number of solar radiation values needed as input to the model.

3.4. Model evaluation :

To evaluate the quality of estimation and ANN performances several parameters can be used such as [60]:

Root means square error (RMSE) and normalized RMSE (nRMSE) expressed as:

$$RMSE = \sqrt{\sum_{i=1}^{N} (y_i - x_i)^2 / N} \quad ;$$

$$nRMSE = \left[\sqrt{\sum_{i=1}^{N} (y_i - x_i)^2 / N}\right] \Big/ \bar{x} \quad (3)$$

Determination coefficient R^2defined by the equation :

$$R^2 = \left[\sum_{i=1}^{N} (y_i - \bar{y})(x_i - \bar{x})\right]^2 \Big/ \left\{\left[\sum_{i=1}^{N} (y_i - \bar{y})^2\right]\left[\sum_{i=1}^{N} (x_i - \bar{x})^2\right]\right\} \quad (4)$$

Mean bias error (MBE) and normalized mean bias error (nMBE) defined by :

$$MBE = \sum_{i=1}^{N} (y_i - x_i) \Big/ N \;;\; nMBE = \sum_{i=1}^{N} (y_i - x_i) \Big/ N\bar{x} \quad (5)$$

y_i denotes estimated values, x_i the measured ones; their respective averages values are defined as $\bar{x} = \sum_{i=1}^{N} x_i / N$; $\bar{y} = \sum_{i=1}^{N} y_i / N$ with N, the data number.

- Root Mean Square Error (RMSE) shows the difference between the measured values and the predicted ones; it indicates the scattering of data around linear lines. The approximation is better if RMSE is minimal (tends to 0).

- The determination coefficient (R^2) expresses the correlation between the real values and the estimated ones; the best approximation corresponds to the highest R^2 (closer to 1).

- MBE and nMBE describe the direction of the error bias, they show if the model tends to under-estimates (MBE<0) or over-estimates (MBE>0) the real values.

4. Database Development

4.1. Meteorological data description:

In this paper, we used data measured in a radiometric station supervised by our laboratory and installed in Faculty of science and technology in Fes(Morocco) (latitude:33°56'-N; longitude: 4°99'-W; altitude=579 m). Fes is a city with cold Winters and dry and hot Summers.

This station consists of pyranometers SP-Lite (Kipp & Zonen) to measure global and diffuse solar radiation, an anemometer with a van(wind monitor 05103) to measure wind speed and direction and a thermo-hygrometer (HMP45C) having a sensor to measure temperature and another sensor for relative humidity. In addition, the station has a pluviometer to measure precipitations and some radiometers to measure spectral components of solar radiation, in particular the UV and the active photosynthetic radiation APR. All these instruments are related to a data acquisition board(CR10X) with a storage module. Data are measured and recorded every 5seconds and then converted to hourly averages.

For this study, five years of hourly data recorded from the 1st January 2010 to 31th December 2014 are available with43824data records. The measured variables are: horizontal global solar radiation (I_g); diffuse radiation (I_d); temperature (T); pressure (P); relative humidity (H_r); precipitations(p); cumulative precipitations (Cp); wind speed (W_s) and wind direction (W_d).

In order to extract outlier values, each parameter was examined. Then data before sunrise and bellow sunset periods were also deleted to avoid the mask effect or a no-reliable response of pyranometers at high zenith angle on solar data[52]. Therefore, we have definitely 20544 records for each variable mentioned above. All these variables can be used as input parameters to the ANN models.

4.2. Database development

As mentioned before, the choice of the best combination of inputs is a prerequisite stage as there is no systematic rules. However, we must take into account some criteria such as:

-Parsimony which consists in developing the simplest ANN architecture with a minimum of inputs, hidden layers and hidden neurons while keeping high performances.

-Avoid redundant inputs (they contain the same information), choose the best correlated variables to solar irradiation. Indeed, too many inputs can reduce the model efficiency[62].

In this study, we limited the number of inputs to five variables: relative humidity H_r(%), temperature T(°C), wind speed Ws (m/s), number of sunshine, calculated based on measured data, N(hours) and horizontal extraterrestrial solar radiation I_{he}(kWh/m^2).Having hourly global solar radiation as the target variable, these inputs will be used to train, validate and test the models

4.2.1. Calculation of I_{he}

To calculate horizontal extraterrestrial solar radiation I_{he} for each hour two geographical parameters must be calculated: solar declination '*dec*' and zenith angle A_z.

-Solar declination 'dec':

'dec' is defined as the angle between the Sun's rays and the Earth's equatorial plane, it depends on the day number d_n via the day angle G defined by[38]:

$$G = 2\mathrm{pi}(d_n - 1)/365 \qquad (6)$$

with pi=180°

$$dec(rad) = 0.006918 - 0.399912cosG + 0.070257sinG - 0.006758cos2G + 0.000907sin2G + 0.002697cos3G + 0.00148sin3G \qquad (7)$$

-Zenith angle A_z:

A_z represents the angle between the sun and a horizontal surface [39]; when it is low, the incident radiation is less absorbed and the solar radiation is maximal because its cosines is maximal and the optical path is minimal. Thus, it influences the quality and the quantity of solar radiation; A_z is calculated at the middle of the considered hour by:

$$cosA_Z = \sin(dec)\sin(La) + \cos(dec)\cos(La)\cos(W) \quad (8)$$

with 'La' the latitude and W the hour angle calculated from the true solar time t_s by:

$$W(degrees) = 15(12 - t_s) \qquad (9)$$

To calculate the horizontal extraterrestrial solar radiation I_{he} in Wh. m^{-2}, we calculate first of all extraterrestrial solar radiation I_e in W.m^{-2}and then we compute I_{he} by integration of I_e on the time period. I_e is given by[39]:

$$I_e = I_0 E_0\, cosA_Z \qquad (10)$$

where I_0 is the solar constant (I_0=1367 Wm^{-2}), E_0, called eccentricity, is a dimensionless factor that permits to take into account the variation of the earth-sun distance [39]:

$$E_0 = 1.00110 + 0.034221\cos G + 0.001280sinG + 0.000719cos2G + 0.000077sin2G \qquad (11)$$

4.2.2. Calculation of The sunshine hours N

The number of sunshine hours is a measured variable. It is strongly correlated with solar radiation so, as it is not available in our records, it will be computed from measured data. For a specific hour, the normal direct beam solar radiation I_b is calculated from the horizontal global (I_g) and diffuse(I_d) solar radiation by Eq.(12):

$$I_b = (I_g - I_d)/cos\,Az \qquad (12)$$

According to the World Meteorological Organization, If I_b is greater than 120W.m^{-2} then this hour is considered as a sunshine hour, summation of sunshine hours over all the day gives the value of N for this day.

Thus, there are 31 (2^5-1=31) possible associations of five meteorological variables as input, to test the network with all combinations was arduous. Hence, we only consider the ten most significant combinations of variables to simplify the process. In Table 1. each inputs combination is associated to the corresponding model number.

5. Results and Discussion

5.1. Choice of the best model

In this section, ten structures of MLP and NARX models were created. The hidden neurons' number was varied from 5 to 35. For NARX models, The delay number was fixed at the value of two to reduce the number of solar radiation values needed as input to the model[62].

16090 records (about 80%) are used in training the proposed ANN models. Meanwhile, the remaining data 4365 records (about 20%) are used in the test process. The model development and training were done using MATLAB line code and MATLAB GUI's ANN toolbox. It is also worth mentioning that every ANN models have been tested repeatedly (up to 10 times) in order to provide best performance of the model.

Tables 2. and 3. present the best achieved results, in terms of nRMSE, nMBE and R^2, for both MLP and NARX ANN models. For the network structure's identification used is in the second column of Tables 2 and 3, the numbers indicate respectively the number of neurons in the input layer (number of inputs), the number of neurons in the hidden layer and the number of neurons in the output layer.

From Table 2., the minimum nRMSE=15.1%, is obtained for model 1 (5 inputs) while the worst configuration (model 8) has a nRMSE= 20.03%. Results for model 6 (nRMSE=16%) are satisfying with such small number of inputs(only one measured input (temperature)).

We notice that the increase of the number of inputs does not, necessarily, improve NARX's performances: some NARX structures with a lower number of inputs are better than architectures with a higher number of inputs. Model 6, for example, with 2 inputs is better than model 2 with 4 inputs. So, adding new inputs may decrease the model performances.

In Table 2., we remark that when configurations contain I_{he} results are better; for example comparison between models 1 and 2 and between models 3 and 5 (I_{he} is strongly correlated with solar radiation).

Table 1. Best combinations of input parameters

Model	Input parameters	Model	Input parameters
1	N-T-H_r-W_s- I_{he}	6	T- I_{he}
2	N-T-H_r-Ws	7	T-H_r
3	N-T-H_r- I_{he}	8	N-W_s
4	T-H_r-W_s- I_{he}	9	N-T- I_{he}
5	N-T-H_r	10	H_r-T- I_{he}

Table 2. NARX models' performances

Model	Network structure	R^2	nRMSE(%)	nMBE(%)
1	5-10-1	0.95	15.10	0.036
2	4-10-1	0.91	19.47	-0.164
3	4-12-1	0.95	15.23	0.066
4	4-10-1	0.94	16.06	-0.126
5	3-7-1	0.92	19.50	0.089
6	2-12-1	0.94	16.17	-0.044
7	2-10-1	0.91	19.76	-0.123
8	2-20-1	0.92	20.03	0.389
9	3-10-1	0.94	15.83	0.055
10	3-15-1	0.95	16.15	-0.009

Table 3. MLP models' performances

Model	Network structure	R^2	nRMSE(%)	nMBE(%)
1	5-30-1	0.88	23.31	-0.189
2	4-10-1	0.38	53.61	0.589
3	4-15-1	0.88	23.80	0.174
4	4-22-1	0.87	24.36	-0.014
5	3-21-1	0.34	55.34	-0.567
6	2-15-1	0.81	29.90	0.064
7	2-15-1	0.32	56.11	-0.028
8	2-22-1	0.23	59.81	0.246
9	3-20-1	0.83	27.20	0.079
10	3-15-1	0.87	24.73	-0.002

Table 3. shows also that model 1 surpasses all the developed MLP models with an RMSE = 23.31%; for the worst configuration nRMSE=59.81%.

Remarks mentioned before for NARX models are also valid for MLP models. Concerning the comparison between the two types of ANN model; results from Tables 2. and 3 show that NARX's performances are better than the MLP ones over the ten models. This result can be explained by the fact that: dynamic networks(NARX) are generally more powerful than static networks(MLP). Indeed, the embedded memory is particularly significant in recurrent NARX neural networks. So, they can be trained to learn sequential or time-varying patterns(by creating a spatial representation of temporal pattern, putting time delays into the neurons or their connections, employing recurrent connections). Thus, in NARX network, the output depends not only on the current input to the network, but also on the previous inputs, outputs, or states of the network. That is why it can depict the complex dynamic and sequential patterns in solar time series.

It is worth mentioning that MLP models fail completely with configurations without I_{he} (R^2<0.4) while NARX models give quite acceptable results(R^2 > 0.90).

So, if we can provide the two values of hourly solar radiation as input, NARX models give the best results.

5.2. Hourly solar radiation prediction

To make further analysis, for the rest of our study, we will use the best NARX's model (model 1(5-10-1)) to generate a hole year of horizontal hourly global solar radiation series. These predicted solar radiation data will be compared to the remaining 4365 real data (1 year of solar day). As mentioned before, data used for the test was not used in training in order to check the ability of our model to predict future data and ensure its proper evaluation over unseen data. In Fig. 3, a regression plot illustrates correlation between the measured and the predicted hourly global solar radiation using this model. This figure indicates a good fit with a considerably high coefficient of correlation (R= 0.97) [63]. This high correlation value implies that the proposed model makes accurate predictions.

In addition, Fig. 4a,4b, 4c and 4d show the predicted and the measured hourly global solar irradiation versus time, it represent also the error defined as measured value minus the predicted one versus time. Each figure represents a season of the year. The four figures present generated values for a whole year.

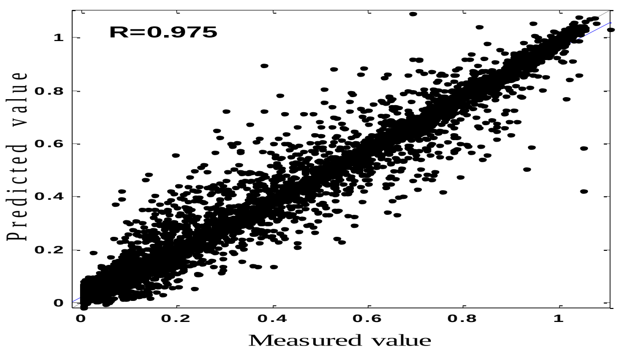
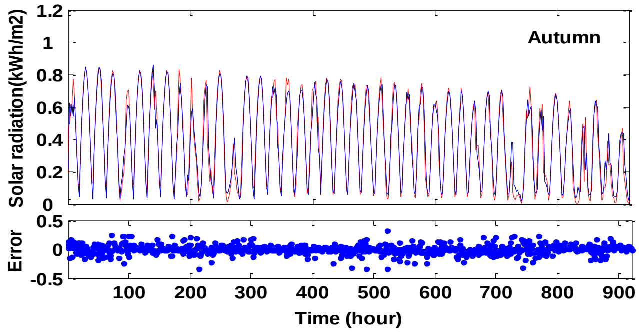
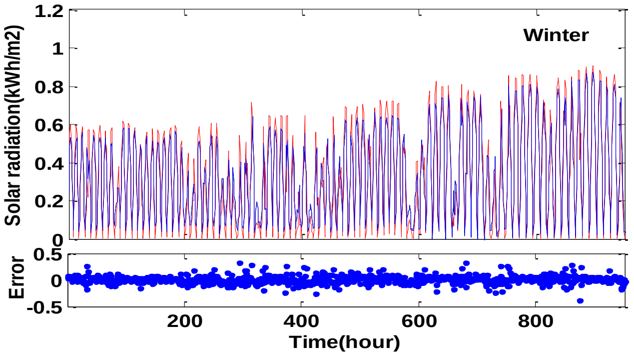


Fig. 3. Regression plot of measured and predicted solar radiation using model 1(5-10-1)

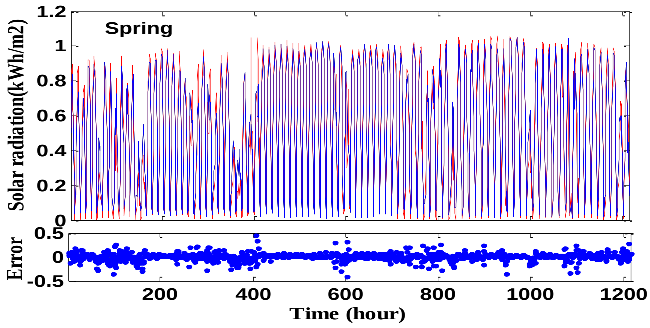
From these figures, it can be noted that the proposed NARX model predicts the hourly solar radiation successfully. However, it seems that the prediction's accuracy depends on the period of the year. Indeed, the quality of fit differs from a season to another and as it can be seen the error scatter is getting closer and closer to zero from winter to summer.



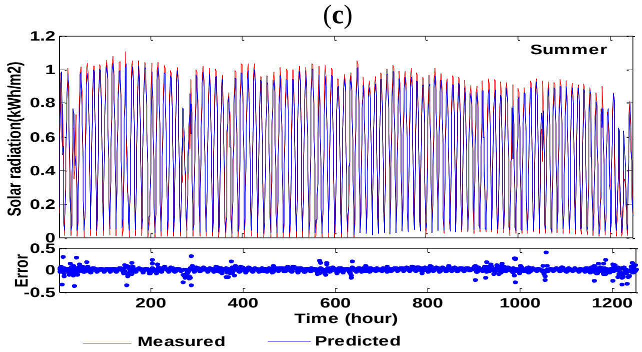
(**a**)



(**b**)



(**c**)



(**d**)

Fig. 4. Measured and predicted solar radiation (**a**) for Autumn (**b**) for Winter (**c**) for Spring (**d**) for Summer

For a deeper analysis, we zoom in zones of good and bad predictions. We notice that good fitting (Fig. 5a) occurs over days with important amount of solar radiation. Indeed, during summer nRMSE is about 12% for example. Zones with bad fit (Fig. 5b) correspond to cloudy days, the prediction's accuracy decrease due to unstable solar radiation levels. For days with weak insolation, the prediction's accuracy is lower than on sunny days but still acceptable as most of the solar radiation prediction model's accuracies degraded on cloudy days [64–66].

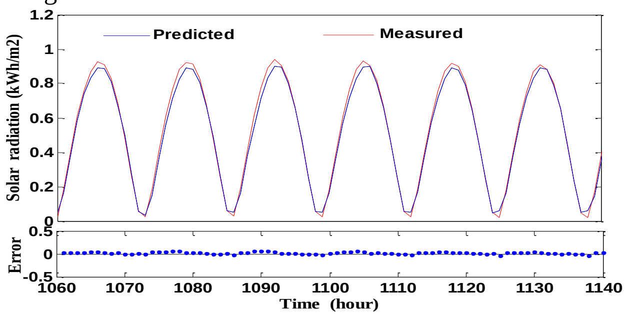
5.3. *Impact of the size of training set*

For the best use of our approach, either for long time prediction (generating of synthetic data) or for short term prediction (completing missing data in solar radiation series for example), we will check if such a large data set (about 5 years, 43824 records) is a must to develop such a model.

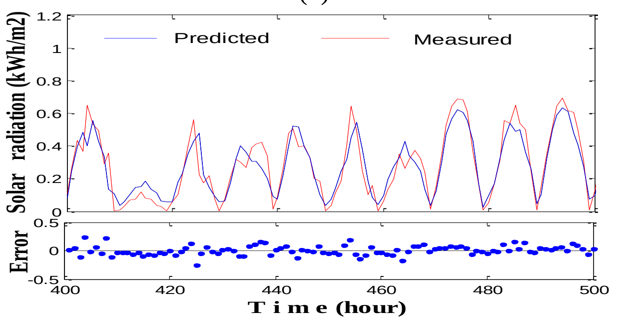
It is all about utilizing available data to develop accurate model with excellent ability to predict future data. As a matter of fact, there are two important issues to be considered when deciding the size of a training data set for a solar radiation prediction model. These issues are the uncertainty nature of solar radiation and the day number nature of the year. Hence, is it possible to forecast hourly solar radiation in July using a model which is trained over data for January for example ?.

To investigate this issue, we have trained the proposed model using data sets with different periods of relatively small sizes.

Fig. 6a, 6b, 7a, 7b and 8a, 8b show results of this practice. The right part of these figures shows the prediction's performance of the model by comparing its outputs to the actual values while the left part illustrates the model's errors histogram.



(**a**)



(**b**)

Fig. 5(**a**) Zoom in of a zone of summer(**b**) Zoom in of a zone of winter

For Fig. 6a and 6b, our model was trained over January's data (310 records) and tested over data from July; it is clear from Fig. 6b that prediction is not accurate (nRMSE=27%). Fig. 6a. shows that instances where the error is greater than 30% exceed 60% probably. It means that training set was not sufficient so that the network retains climate's patterns.

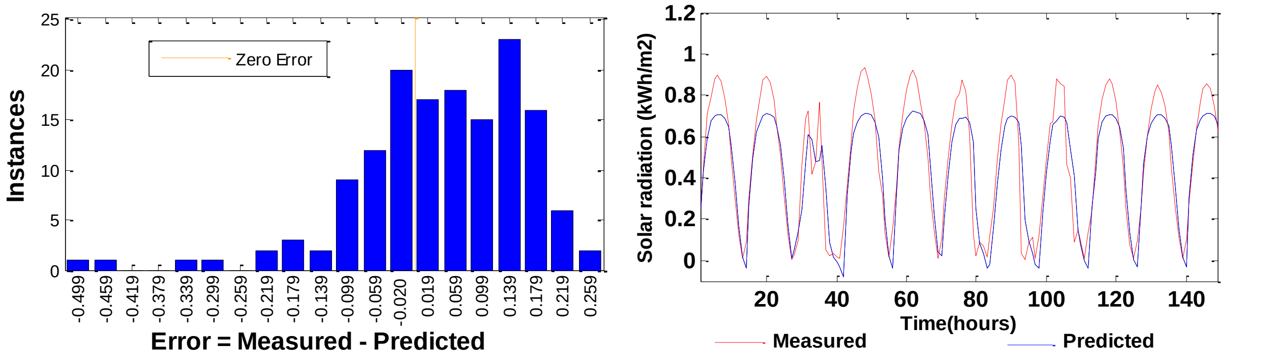
-In Fig. 7a and 7b, our model was trained over July's data(433 records) and tested over data from January. We notice an enhancement in Fig. 7b comparing to Fig. 6b (nRMSE=18.5%) even if we have a slight increase in the size of the learning set. In Fig. 7a., more than 90% of instances have an error less than 23%. This amelioration can be interpreted by the fact that over summer period, accuracy of models is very high (nRMSE=12% over summer).

-Fig. 8a and 8b show results when the learning process is done with data from Jun to August and test over data for September. Here also there is a significant improvement in the quality of fit in Fig. 8b. It is seen clearly in Fig. 8a that, as far as the training set increases the errors histogram is shifted towards the zero. This enhancement over the above examples is achieved as both the size and period of records for training were concerned.

Finally, for more analysis, we have tested the capability to generate long series values using our model with different learning sets: one year, two years and three years. Table 4. summarizes results through R^2 and nRMSE for all the studied cases. It shows that the model's performance was further improved when 4356 records (1 year solar days) were used in the training; R^2=0.93 prediction's accuracy is then acceptable. It is clear that, as far as the model is trained using more data, the accuracy will be better. However, it seems that one year of records for training can be sufficient as the fact of adding two years of solar days reduces the nRMSE by only 1.22% and increases R^2 by 1%. In fact, such a model can be developed using a relatively small data set depending on what accuracy needed for the generated data.

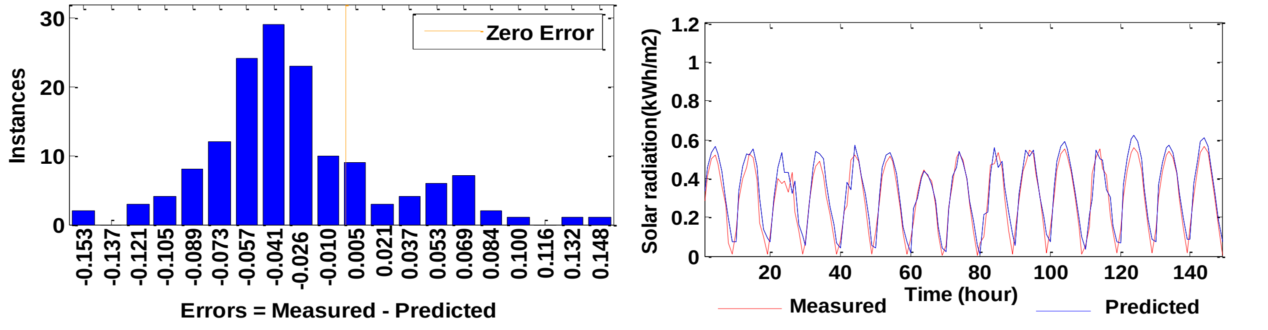
Table 4. Impact of training data set's size on model's accuracy

Period of training	Period of prediction	R^2	nRMSE (%)
310 records(January)	149 values (July)	0.82	27
433 records(July)	149 values (January)	0.90	18.50
1262 records (Jun-August)	366 values (September)	0.91	17.30
4356 records (1 year)	4356 values (1 year)	0.93	16.32
8715 records (2 years)	4356 values (1 year)	0.94	16.23
13085 records (3 years)	4356 values (1 year)	0.94	15.71



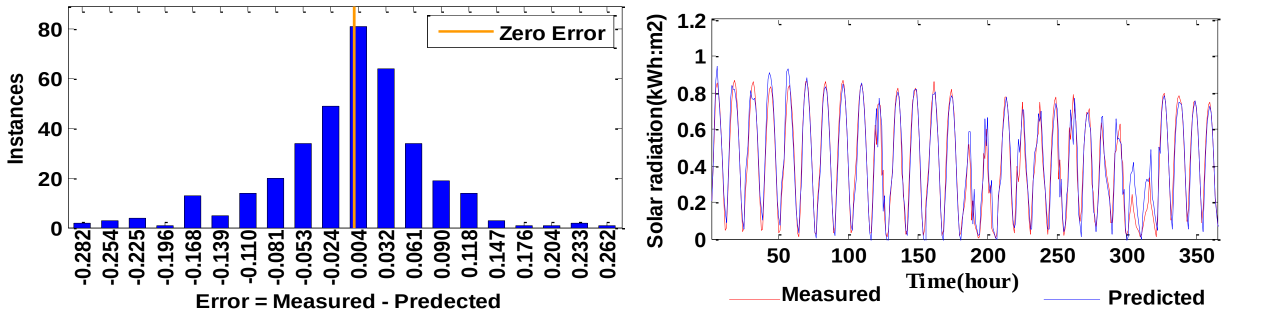
(a) (b)

Fig. 6(**a**).Errors histogram (**b**)Accuracy of Prediction of 15 solar days of July with model trained over January



(a) (b)

Fig. 7(**a**)Errors histogram (**b**)Accuracy of Prediction of 15 solar days of January with model trained over July



(a) (b)

Fig. 8(**a**)Errors histogram (**b**)Accuracy of prediction of September with model trained over Jun to August

5.4. Comparison with literature

Comparing our results with other studies was not an easy task, because most of the papers concern the estimation of monthly or daily average values of solar irradiations. Indeed, the accuracy obtained for monthly average values are not comparable with daily, hourly or shorter time scale data. In fact, due to the compensating effect of the sky anisotropy phenomena, longer is the time step, better is the efficiency of estimation.

Despite this, we will compare our results with at least two main studies. Indeed, Yang et al. [58] extended the most well-known correlation of Angstrom Prescott [68], that calculates monthly average values from only sunshine duration, to daily and hourly values of solar irradiation for sites in USA and Saudi Arabia. Concerning ANN models, even if daily solar irradiations seems to be the most estimated in literature; a recent study[62] considered hourly data but with a different approach. The RMSE obtained by our model(60Wh.m^{-2}) is in the range of the one found by Yang et al. [67] for hourly data (between 49 and 79 Wh. m^{-2}) this is on the one hand. On the other hand, our nRMSE=15.1% is comparable to those found in [62] for different ANN models(between 13.33% and 20.23%). Our model can be then considered as very correct.

6. Conclusion

To overcome the lack of accurate long series of solar data, needed for solar systems optimal sizing and design, an approach to predict hourly global solar radiation from cheaper meteorological data was presented in this paper research. Ten different associations of five meteorological variables were used to develop two types of ANN models. The model that gives best performances is a neural autoregressive with external inputs NARX with five inputs. It was used for predicting hourly solar radiation from more available and cheaper meteorological data(T, H_r, N,W_s). The average prediction accuracy of the proposed model is about 15% with *R*-square value of 0.95. Furthermore, different training sets were used, results showed that the proposed model needed a concrete prior training in order to show accurate prediction. It is also concluded that such a model can be developed, with relatively accepted prediction accuracy (nRMSE=17%), using only about three months' data with an hourly step. Indeed, such accuracy could be sufficient for some applications. From a learning set of one year of hourly data , the proposed model can be satisfactory to estimate long sets of hourly global solar radiation.

Finally, our model can provide synthetic solar radiation series to be used in optimal sizing and planning of solar energy systems.

We look forward to apply this approach in further studies using data from other locations to develop a model that represents all Moroccan's locations.

References

[1] M.A. Atwater, J.T. Ball, "A numerical solar radiation model based on standard meteorological observations", Sol. Energy 21, pp. 163–170, 1978.

[2] D. Elizondo, G. Hoogenboom, and R. W. McClendon, "Development of a neural network model to predict daily solar radiation" Agricultural and Forest Meteorology, vol. 71, no.1-2, pp. 115–132, 1994.

[3] D. B. Williams and F. S. Zazueta, "Solar radiation estimation via neural network" in Proceedings of the 6th International Conference on Computers in Agriculture, pp. 140–146, ASAE, Cancun, Mexico, 1994.

[4] D. Williams and F. Zazueta, "Solar radiation estimation via neural network" in Proceedings of the 6th International Conference on Radiation. Computers in Agriculture, pp. 1143–1149, Cancun, Mexico, 1996.

[5] M. Mohandes, S. Rehman, and T. O. Halawani, "Estimation of global solar radiation using artificial neural networks" Renewable Energy, vol. 14, no. 1–4, pp. 179–184, 1998.

[6] S. M. Al-Alawi and H. A. Al-Hinai, "An ANN-based approach for predicting global radiation in locations with no direct measurement instrumentation" Renewable Energy, vol. 14, no.1–4, pp. 199–204, 1998.

[7] A. Guessoum, S. Boubkeur, and A. Maafi, "A global irradiation model using radial basis function neural network" in Proceedings of the World Renew Energy Congress (WREC '98), pp. 332–336, Oxford, UK, 1998.

[8] L. Hontoria, J. Riesco, P. Zufiria, and J. Aguilera, "Improved generation of hourly solar radiation artificial series using neural networks" in Proceedings of the Engineering Applications of Neural Networks (EANN '99), Warsaw, Poland, 1999.

[9] P. Zufiria, A. Vazquez, ´ J. Riesco, J. Aguilera, and L. Hontoria, "A neural network approach for generating solar radiation artificial series" in Proceedings of the International Work- Conference on Artificial and Natural Neural Networks (IWANN'99), Alicante, Spain, June 1999.

[10] M. Mohandes, A. Balghonaim, M. Kassas, S. Rehman, and T. O. Halawani, "Use of radial basis functions for estimating monthly mean daily solar radiation" Solar Energy, vol. 68, no.2, pp. 161–168, 2000.

[11] L. Hontoria, J. Riesco, P. Zufiria, and J. Aguilera, "Application of neural networks in the solar radiation field. Obtainment of Solar Radiation Maps" in Proceedings of the 16th European Photovoltaic for Chemical Engineers, vol. 3, pp. 385–408, Elsevier, Amsterdam, The Netherlands, 2000.

[12] A. Sfetsos and A. H. Coonick, "Univariate and multivariate forecasting of hourly solar radiation with artificial intelligencetechniques" Solar Energy, vol. 68, no. 2, pp. 169–178, 2000.

[13] G. Mihalakakou, M. Santamouris, and D. N. Asimakopoulos, "The total solar radiation time series simulation in Athens,using neural networks" Theoretical and Applied Climatology,vol. 66, no. 3-4, pp. 185–197, 2000.

[14] S. A. Kalogirou, "Artificial neural networks in renewable energy systems applications: a review" Renewable and Sustainable Energy Reviews, vol. 5, no. 4, pp. 373–401, 2000.

[15] L. Hontoria, J. Aguilera, J. Riesco, and P. Zufiria, "Recurrent neural supervised models for generating solar radiation synthetic series" Journal of Intelligent and Robotic Systems, vol.31, no. 1–3, pp. 201–221, 2001.

[16] L. Hontoria, J. Aguilera, and P. Zufiria, "Generation of hourly irradiation synthetic series using the neural network multilayer perceptron" Solar Energy, vol. 72, no. 5, pp. 441–446, 2002.

[17] A. S. S. Dorvio, J. A. Jervase, and A. Al-Lawati, "Solar radiation estimation using artificial neural networks" Applied Energy, vol. 74, pp. 307–319, 2002.

[18] F. Tymvios, C. P. Jacovides, and S. C. Michaelides, "The total solar energy on a horizontal level with the use of artificial neural networks" in Proceedings of the 6th Hellenic Conference of Meteorology, Climatology and Atmospheric Physics, pp. 26–28, Ioannina, Greece, September 2002.

[19] S. Kalogirou, S. C. Michaelides, and F. S. Tymvios, "Prediction of maximum solar radiation using artificial neural networks" in Proceedings of the 7th World Renewable Energy Congress (WREC '02), 2002.

[20] K. S. Reddy and M. Ranjan, "Solar resource estimation using artificial neural networks and comparison with other correlation models" Energy Conversion and Management, vol. 44, no. 15, pp. 2519–2530, 2003.

[21] A. Sozen, E. Arcakl.ogl . u, M. Ozalp, and EG Kan.t, “Use of artificial neural networks for mapping the solar potential in Turkey” Applied Energy, vol. 77, pp. 273–286, 2004.

[22] A. Sozen, E. Arcakl.oglu, . and M. Ozalp, “Estimation of solar potential in Turkey by artificial neural networks using meteorological and geographical data” Energy Conversion and Management, vol. 45, no. 18-19, pp. 3033–3052, 2004.

[23] A. Mellit, M. Benghanem, A. Hadj-Arab, and A. Guessoum, “Modeling of global solar radiation data from sunshine duration and temperature using the Radial Basis Function networks” in Proceedings of the IASTED International Conference on Modelling, Identification and Control (MIC ’04), Grindelwald, Switzerland, February 2004.

[24] L. Hontoria, J. Aguilera, and P. Zufiria, “An application of the multilayer perceptron: solar radiation maps in Spain” Solar Energy, vol. 79, no. 5, pp. 523–530, 2005.

[25] A. Sozen, E. Arcaklyo ´ gblub, M. Ozalpa, and N. C. Agclarc, “Forecasting based on neural network approach of solar potential in Turkey” Renewable Energy, vol. 30, pp. 1075–1090, 2005.

[26] F. S. Tymvios, C. P. Jacovides, S. C. Michaelides, and C.Scouteli, “Comparative study of Angstroms and artificial neural networks methodologies in estimating global solar radiation” Solar Energy, vol. 78, no. 6, pp. 752–762, 2005.

[27] A. Mellit, M. Benghanem, and M. Bendekhis “Artificial neural network model for prediction solar radiation data: application for sizing stand-alone photovoltaic power system”, in Proceedings of the IEEE Power Engineering Society General Meeting, vol. 1, pp. 40–44, June 2005.

[28] G. Lopez, F. J. Batlles, and J. Tovar-Pescador, “Selection of input parameters to model direct solar irradiance by using artificial neural networks” Energy, vol. 30, no. 9, pp. 1675–1684, 2005.

[29] L. F. Zarzalejo, L. Ramirez, and J. Polo, “Artificial intelligence techniques applied to hourly global irradiance estimationfrom satellite-derived cloud index” Energy, vol. 30, no. 9, pp.1685-1697, 2005.

[30] S. Alam, S. C. Kaushik, and S. N. Garg, “Computation of beam solar radiation at normal incidence using artificial neural network” Renewable Energy, vol. 31, no. 10, pp. 1483–1491, 2006.

[31] H. K. Elminir, Y. A. Azzam, and F. I. Younes, “Prediction of hourly and daily diffuse fraction using neural network, as compared to linear regression models” Energy, vol. 32, no. 8, pp. 1513–1523, 2007.

[32] G. Lopez ´ and C. A. Gueymard, “Clear-sky solar luminous efficacy determination using artificial neural networks” Solar Energy, vol. 81, no. 7, pp. 929–939, 2007.

[33] T. Khatib,A.Mohamed,K. Sopian,and M. Mahmoud“Solar Energy Prediction for Malaysia UsingArtificial Neural Networks ” International Journal of Photoenergy, Article ID 419504, 2012.

[34] M.Omar; A.Dolara; G.Magistrati; M.Mussetta;E.Ogliari;F.Viola “Day-ahead forecasting for photovoltaic power using artificial neural networks ensembles”, International Conference on Renewable Energy Research and Applications (ICRERA),pp.1152 - 1157, 2016. IEEE.

[35] N. Kumar , S. P. Sharma, U. K. Sinha, Y. Nayak “Prediction of Solar Energy Based on Intelligent ANN Modeling”; International Journal of Renewable Energy Research ; Vol.6 ; No.1; pp. 183-188, 2016.

[36] F.S. Tymvios, C.P. Jacovides, S.C. Michaelides, C. Scouteli, “Comparative study of angstroms and artificial neural networks methodologies in estimating global solar radiation”, Solar Energy 78,pp. 752–762, 2005.

[36] J.A. Davies, D.C. McKay, “Estimating solar irradiance and components”, Solar Energy 29, pp.55–64,1980.

[38] C. Gueymard, “Critical analysis and performance assessment of clear sky solar irradiance models using theoretical and measured data”, Sol. Energy 51 (1993) 121–138.

[39] M. Iqbal, “An Introduction to Solar Radiation”, Academic Press, Canada, 1983. ISBN 0-12-373752-4.

[40] M. Tiris, C. Tiris, E. Ture, “Correlations of monthly-average daily global, diffuse and beam radiations with hours of bright sunshine in Gebze, Turkey”, Energy Conversion Management. 37, pp.1417-1421, 1996.

[41] G. Lopez, M.A. Rubio, F.J. Batlles, “Estimation of hourly direct normal from measured global solar irradiance in Spain”, Renew. Energy 21, pp. 175–186, 2000.

[42] H. Loutfi; A. Khtira “Stochastic analysis and generation of synthetic sequences of daily global solar irradiation: Rabat site (Morocco)”, Renewable energy; Vol:2; pp.129–138, 1992

[43] P.Poggi, G. Notton, M. Muselli and A. Louche “Stochastic study of hourly total solar radiation in Corsica using a Markov model”, International Journal of Climatology; Vol.20; pp. 1843–1860, 2000.

[44] R. Aguiar and M. Collares-Pereira, “TAG: a time-dependent, autoregressive, Gaussian model for generating synthetic hourly radiation” Solar Energy, vol. 49, no. 3, pp. 167–174, 1992.

[45] A. Mellit, “Artificial intelligence technique for modelling and forecasting of solar radiation data: a review”, Int. J. Artif. Intell. Soft Comput. 1, pp.52–76, 2008.

[46] S.A. Kalogirou, A. Sencan, “Artificial Intelligence Techniques in Solar Energy Applications. Theory and Applications”, ISBN 978-953-307-142-8, 2010.

[47] S. Haykin, “Neural Networks: a Comprehensive Foundation”, second ed., Prentice-Hall, New Jersey, 1999.

[48] A.K. Yadav, S.S. Chandel, “ Solar irradiation prediction using artificial neuralnetwork techniques: A review” , Renew. Sustain. Energy Rev. 33, pp.772–781, 2014.

[49] S.Haykin, "Neural Networks and Learning Machines", 3rd edition, Pearson Education,Inc.,New Jersey, 2009.

[50] A.Benyamina; S.Moulahoum; I.Colak; R.Bayindir “Hybrid fuzzy logic-artificial neural network controller for shunt active power filter” International Conference on Renewable Energy Research and Applications (ICRERA), pp. 837 - 844, 2016. IEEE.

[51] J. M.Malof; L.M.Collins; K.Bradbury; R.G.Newell“A deep convolutional neural network and a random forest classifier for solar photovoltaic array detection in a erial imagery” International Conference on Renewable Energy Research and Applications(ICRERA),pp.650 - 654, 2016. IEEE.

[52] M. Yesilbudak; M. Çolak; R. Bayindir “A review of data mining and solar powerprediction”, International Conference on Rene wable Energy Research and Applications (ICRERA), pp. 1117 - 1121, 2016. IEEE.

[53] M. Alabdul Salam, A. Aziz, A. H. Alwaeli, H.A. Kazem “ Optimal Sizing of Photovoltaic Systems using HOMER for Sohar, Oman ”, International Journal of Renewable Energy Research, Vol.3, No.3,pp. 470-475, 2013.

[54] P. S. Kumar and D. Lopez “Forecasting of Wind Speed Using Feature Selection and Neural Networks”, International Journal of Renewable Energy Research , Vol.6, No.3 , pp.833-837 , 2016.

[55] S. Balluff , J. Bendfeld et S. Krauter “ Short term wind and energy prediction for offshore wind farms using neural networks”, International Conference on Renewable Energy Research and Applications(ICRERA), pp. 379 – 382, 2015.IEEE.

[56] A. R Finamore, V. Calderaro, V. Galdi, A. Piccolo, G. Conio, and S. Grasso “A day-ahead wind speed forecasting using data-mining model-a feed-forward NN algorithm”. In Renewable Energy Research and Applications (ICRERA), 2015 International Conference on (pp. 1230-1235). IEEE.

[57] , J.Yang, H. Rivard, &R. Zmeureanu, "Building Energy Predication with AdaptiveArtificial Neural Networks", Ninth International IBPSA Conference Montréal, August15–18, 2005.

[58] A. Jain, J. Mao, and K.Mohiuddin, “Artificial Neural Networks: A Tutorial”, IEEEComputer, Vol. 29, No.3, pp.31–44,March 1996.

[59]TheMathWorks(2013).MATLAB.URL:http://www.mathworks.com/products/neuralnet

[60] T. Krishnaiah, S. Srinivasa Rao, K.S. Reddy, “Neural network approach for modelling global solar radiation”, Journal of Applied Sciences Research, 3(10), pp.1105– 1111,2007.

[61] A. De Miguel, J. Bilbao, R. Aguiar, H. Kambezidis, E. Negro, "Diffuse solar irradiationmodel evaluation in the North Mediterranean belt area", Solar Energy 70.pp.143–153, 2001

[62] K. Dahmani, G. Notton, C. Voyant, R. Dizene, M.L. Nivet, C. Paoli, W. Tamas, “Multilayer perceptron approach for estimating 5-min and hourly horizontal global irradiation from exogenous meteorological data in locations without solar measurements”, Renewable Energy, Vol. 90, pp. 267–282, 2016

[63] “Analyse neural network performance after training” https://fr.mathworks.com/help/nnet/ug/analyze-neural-network-performance-after-training.html

[64] Aziz Ahmad , Tim Anderson "Global Solar Radiation Prediction using Artificial Neural NetworkModels for New Zealand" Proceedings of the 52nd Annual Conference, Australian Solar Energy Society - Melbourne, pp 141-150, 2014.

[65] T. Khatib, A. Mohamed,and K. Sopian,“A review of solar energymodeling techniques” Renewable and Sustainable EnergyReviews, vol. 16, no. 5, pp. 2864–2869, 2012.

[66] A. Mellit and S. A. Kalogirou, “Artificial intelligence techniquesfor photovoltaic applications: a review” Progress in Energy andCombustion Science, vol. 34, no. 5, pp. 574–632, 2008.

[67] K. Yang, T. Koike, B. Ye, “Improving estimation of hourly, daily and monthlysolarradiation by importing global data sets”, Agric. For. Meteorol. 137, pp.43–55, 2006.

[68] A. Angstrom, “ On the computation of global radiation from records of sunshine”,Ark. Geofys. 2, pp.471–47, 1956.