Forecasting of preprocessed daily solar radiation time series using neural networks

Christophe Paoli, Cyril Voyant, Marc Muselli, Marie Laure Nivet

Forecasting of preprocessed daily solar radiation time series using neural networks

Christophe Paoli[1], Cyril Voyant[1,2], Marc Muselli[1*], Marie-Laure Nivet[1]

[1] University of Corsica, CNRS UMR SPE 6134, 20250 Corte, France

[2] Hospital of Castelluccio, Radiotherapy Unit, BP 85, 20177 Ajaccio, France

Abstract. In this paper, we present an application of Artificial Neural Networks (ANNs) in the renewable energy domain. We particularly look at the Multi-Layer Perceptron (MLP) network which has been the most used of ANNs architectures both in the renewable energy domain and in the time series forecasting. We have used a MLP and an ad-hoc time series preprocessing to develop a methodology for the daily prediction of global solar radiation on a horizontal surface. First results are promising with nRMSE ~ 21% and RMSE ~ 3.59 MJ/m². The optimized MLP presents predictions similar to or even better than conventional and reference methods such as ARIMA techniques, Bayesian inference, Markov chains and k-Nearest-Neighbors. Moreover we found that the data preprocessing approach proposed can reduce significantly forecasting errors of about 6% compared to conventional prediction methods such as Markov chains or Bayes inferences. The simulator proposed has been obtained using 19 years of available data from the meteorological station of Ajaccio (Corsica Island, France, 41°55'N, 8°44'E, 4 m above mean sea level). The predicted whole methodology has been validated on a 1.175 kWc mono-Si PV power grid. Six prediction methods (ANN, clear sky model, combination …) allow to predict the best daily DC PV power production at horizon d+1. The cumulated DC PV energy on a 6-months period shows a great agreement between simulated and measured data ($R^2 > 0.99$ and nRMSE < 2%).

Keywords: Time Series Forecasting, Preprocessing, Artificial Neural Networks, PV Plant Energy Prediction

[*] Corresponding author : Marc MUSELLI, Université de Corse, UMR SPE 6134, Route des Sanguinaires, 20000 AJACCIO
Tél : 33 4 95 52 41 30, Fax : 33 4 95 45 33 28, email : marc.muselli@univ-corse.fr

Nomenclature:

$\hat{x}_t$, $\hat{x}_{d,y}$	Time series model at time *t* or at day *d* and year *y*	H_0^d	Extraterrestrial solar radiation coefficient for day *d* [MJ/m²]
x_t, $x_{d,y}$	Time series data at time t or at day *d* and year *y*	$H_{gh,clearsk}^d$	clear sky global horizontal irradiance [MJ/m²] integrated on the day *d*
k_p, m_q	Parameters of ARMA process	$y_{d,y}$	Moving average of the stationary global radiation for the day *d* and the year *y*
K_p ,M_Q	Parameters specific to seasonal ARMA process	y_d	Intermediate seasonal factor for day *d*
d, D	Non seasonal and seasonal difference	y_d*	Final seasonal factor for day *d*
$S_{d,y}^{corr}$, S_t	Stationary time series (clearness index) for the day *d* and the year *y* or for the time *t*	E_{pv}	PV plant power (MJ)
$S_{d,y}^{corr^*}$, S_t^*	Stationary time series (clear sky index) for the day *d* and the year *y* or for the time I	η_{PV}	PV Plant efficiency (%)
b	Fitting parameter of the Solis clear sky model	I_β	Daily global radiation (tilt of β) [MJ/m²]
τ	Global total atmospheric optical depth	S	Surface of PV wall [m²]
h	Solar elevation angle	Δ_p, Δ_m	Tendency of prediction and tendency of measure [MJ]
$H_{gh,clearsky}$	Clear sky global horizontal irradiance [MJ/m²]		

1 Introduction and Review

An optimal use of the renewable energy needs its characterization and prediction in order to size detectors or to estimate the potential of power plants. In terms of prediction, electricity suppliers are interested in various horizons to estimate the fossil fuel saving, to manage and dispatch the power plants installed and to increase the integration limit of renewable energy systems on non-interconnected electrical grids. In this study, we focus on the prediction of global solar irradiation on a horizontal plane for daily horizon, which might interest electricity suppliers. In this way, we have investigated both in time series forecasting, which is a challenge in many fields, and in artificial intelligence techniques, which are becoming more and more popular in the renewable energy domain [Mellit A., Kalogirou, SA.. 2008. Artificial intelligence techniques for photovoltaic applications: A review. Progress in Energy and Combustion Science 1-1, 52-76] and particularly for the prediction of meteorological data such as solar radiation [Mubiru, J. 2008. Predicting total solar irradiation values using artificial neural networks. Renewable Energy 33-10 2329-2332a ; Mubiru, J., Banda, E., 2008. Estimation of monthly average daily global solar irradiation using artificial neural networks. Solar Energy, 82-2, 181-187Kalogirou, SA., 2001. Artificial neural networks in renewable energy systems applications: a review. Renewable and sustainable energy reviews 5, 373-401Hocaoglu, FO., Gerek, ON., Kurban, M., 2008. Hourly solar forecasting using optimal coefficient 2-D linear filter and feed-forward neural networks. Solar energy 82-8, 714-726]. Thereby many research works have shown the ability of Artificial Neural Networks (ANNs) to predict time series of meteorological data. Moreover, if we compare to conventional algorithms based on linear models, ANNs offer an attractive alternative by providing nonlinear parametric models. Through the proposed study, we will particularly look at the Multi-Layer Perceptron (MLP) network which has been the most used of ANN architectures in the renewable energy domain [Mellit A., Kalogirou, SA.. 2008. Artificial intelligence techniques for photovoltaic applications: A review. Progress in Energy and Combustion Science 1-1, 52-76] and time series forecasting [Crone, SF., 2005. Stepwise Selection of Artificial Neural Networks Models for Time Series Prediction Journal of Intelligent Systems, Department of Management Science Lancaster University Management School Lancaster, United Kingdom Faraway, J., Chatfield, C., 1995. Times series forecasting with neural networks: a case study, Research report 95-06 of the statistics group, University of BathHu, Y., Hwang, J., 2002. Handbook of neural network signal processing. ISBN 0-8493-2359-2 Jain, K., Jianchang, M., Mohiuddin, KM., 1996. Artificial neural networks: A tutorial, IEEE Computer, 29-3, 31–44]. We compare this

tool with other classical predictors like ARIMA, Bayesian inference, Markov chains and k-Nearest-Neighbors predictors. During this study, we use an ad-hoc time series preprocessing step before using neural networks. Indeed, a data preprocessing including seasonal adjustment can improve ANN forecasting performances [Zhang, GP., Qi, M., 2005. Neural network forecasting for seasonal and trend time series, European Journal of Operational Research 160, 501-514]. Finally we validate the simulator on a 1.175 kWp mono-Si PV power grid connected on a wall of our laboratory. This validation step presents both additional difficulties: the PV modules are 80° tilted to horizontal and are localized at 10 km from the original station.

The paper is organized as follow: section 2 describes the context in which this research was done and the data we have used and introduces time series forecasting, presents several conventional prediction and modeling methods of daily irradiation data including ANNs. All these predictors are then tested and compared with the same 19 years data set in the section 3 (data used for training and validation of the ANN comes from the meteorological station of Ajaccio airport). At last, the global approach is validated and discussed on real tilted PV modules. Section 4 concludes and suggests perspectives.

2 Methodology

2.1 Context

In this work, measured global daily radiation data taken from a meteorological ground station are used to forecast global solar irradiation for the next day. The global radiation consists of three types of radiation: direct, diffuse and ground-reflected [Liu, B.Y.H., Jordan, R.C., 1962. Daily insolation on surfaces tilted towards the equator. Trans SHRAE; 67, 526–541]. The ground-reflected radiation does not concern us because we try to predict the radiation on a horizontal surface. For clear sky, global radiation is relatively easy to model because it is primarily due to the distance from the sun to the sensor [Hay, J.E., Davies, J.A., 1980. Calculation of the solar radiation incident on an inclined surface. In: Proc. First Canadian Solar radiation workshop, 59–72 Ineichen, P., Guisan, O., Perez, R., 1990. Ground-reflected radiation and albedo. Solar Energy;44-4, 207–214Liu, B.Y.H., Jordan, R.C., 1962. Daily insolation on surfaces tilted towards the equator. Trans SHRAE; 67, 526–541 ; Perez, R., Ineichen, P., Seals, R., 1990. Modelling daylight availability and irradiance components from direct and global irradiance. Solar Energy 44-5, 271–289 ; Reindl, D.T., Beckman, W.A., Duffie, J.A., 1990. Evaluation of

hourly tilted surface radiation models. Solar Energy; 45-1, 9-17]. With cloudy sky, these are mostly stochastic phenomenon, which depend on the local weather.

The study proposes to analyze the radiation time series ($MJ.m^{-2}$) measured at the meteorological station of Ajaccio (Corsica Island, France, 41°55'N, 8°44'E, 4 m above mean sea level) maintained by the French meteorological organization (Météo-France). Ajaccio has a 'Mediterranean' climate, hot summers with abundant sunshine and mild, dry, clear winters. It is located near the sea but there is also relief nearby (~ 30 km). This geographical configuration can make nebulosity difficult to forecast.

The data used represent the global horizontal solar radiation and are available on an hourly basis for a period from January 1971 to December 1989. A compact XARIA-C station (DEGREANE HORIZON), certified by Météo-France, is used to record horizontal radiation (Kipp & Zonen CMP6) with a 3 seconds time sampling. The sensor works in a range 0 – 90000 J/m² and requires a quality control every year. Only average measures over one hour are stored in the national database. These last values are integrated to extract a daily time series of the global horizontal radiation.

2.2 Time series forecasting

A time series [Faraway, J., Chatfield, C., 1995. Times series forecasting with neural networks: a case study, Research report 95-06 of the statistics group, University of Bath] is a collection of time ordered observations x_t, each one being recorded at a specific time t (period). Time series can appear in a wide set of domains such as Finance, Production or Control, just to name a few. In first approximation, a time series model ($\hat{x}_t$) assumes that past patterns will occur in the future. In fact a time series model could be used only to provide synthetic time series statistically similar to the original one. The modeling of the series begins with the selection of a suitable mathematical model (or class of models) for the data. Then, it is possible to predict future values of measurements [Brockwell, P. J., Davis, R. A., 2006. Time series: theory and methods. ISBN 0387974296 (USA). 577p. Springer series in statistics, second edition].

2.3 Most popular forecasting methods

We have identified in the time series forecasting literature the following methods: ARIMA (classical method), Bayesian inference, Markov chains and k-Nearest-Neighbors predictors. We present in the next sub-sections each method with the model obtained after a specific optimization.

2.3.1 ARIMA techniques

The ARIMA techniques [Hamilton, J.D., 1994. Times series analysis. ISBN 0-691-04289-6] are reference estimators in the prediction of global radiation field. It is a stochastic process coupling autoregressive component (AR) to a moving average component (MA). This kind of model is commonly called SARIMA (p, d, q)(P, D, Q) and is defined as follows (p,d,q, P,D and Q are integers and B the lag operator):

$$k_p(B)K_p\left(B^S\right)(1-B)^d\left(1-B^S\right)^D X_t = m_q(B)M_Q\left(B^S\right)u_t \quad (1)$$

Where, X_t is a time series, k_p are the parameters of the autoregressive part, m_q are the parameters of the moving average part and u_t is an error term distributed as a Gaussian white noise, K_p and M_Q are parameters of seasonal autoregressive and seasonal moving average part. The d and D are non seasonal and seasonal difference and S is seasonal period. The optimization of these parameters must be made depending on the type of the series studied. To help us in this study, we chose to use grocer, a very complete toolbox compatible with the Scilab® software [Dubois, E., Michaux, E., 2008: "Grocer: an econometric toolbox for Scilab", available at http://dubois.ensae.net/grocer.html]. The criterion adopted to consider when an ARMA model 'fits' to the time series is the normalized root mean square error obtained by: nRMSE = $\sqrt{<(x-y)^2>}/\sqrt{<(x)^2>}$ where x represents the measurement and y the prediction, using averaged values. This parameter is generated by the prediction of two years of radiation not used during the ARMA parameters calculation step. After several experiments the model best suited to the study of global radiation is an ARMA (2,2), its representation is:

$$X_t = u_t + \sum_{i=1}^{2} k_i.X_{t-i} + \sum_{j=1}^{2} m_j.u_{t-j} \quad (2)$$

Where k1=-1.47, k2=0.47, m1=-1.19 and m2=0.22. A Student's T-test (introduced by William Sealy Gosset in 1908) has been used to verify that these coefficients were significantly different from zero and residual auto-correlogram tests have been computed to verify white noise error terms.

Moreover, we also wanted to study a simplest model among the different ARIMA techniques. We have chosen to study the AR type and we have found that the most interesting in this case was the AR(8).

2.3.2 Bayesian inference

The second classical technique we have chosen is called Bayesian inference [Diday, E., Lemaire, L., Pouget, J., Testu, F., 1982. Éléments d'analyse de données, Dunod, Paris ; Celeux, G., Nakache, J.P., 1994. Analyse

discriminante sur variables qualitatives ISBN 2840540274, Polytechnica, 270 p., Paris Pole, A., West, M., Harrison, J., 1994. Applied Bayesian forecasting and time series analysis. ISBN 0412044013. Chapman and Hall/CRC.]. In this method, evidences or observations are used to update or to newly infer the probability that a hypothesis may be true. To estimate the probability that the series is in the state y_k at time t, it is useful to use the Bayes theorem. This ideology is expressed mathematically by the following formula (where $\aleph$ represents the initial and known measured, k is the class and y_k the value of this class, P is the conditional probability):

$$X_t = y_{k^*} = \arg\max_k \left(P(X_t = y_k \mid \aleph)\right) = \arg\max_k \left(P(X_t = y_k).P(\aleph \mid X_t = y_k)\right) \tag{3}$$

A practical way to solve this equation is to make the assumption of conditional independence as follows:

$$P(\aleph \mid X_t = y_k) = \prod_{j=1}^{J} P(X_{t-j} \mid X_t = y_k) \tag{4}$$

which leads to:

$$X_t = y_{k^*} = \arg\max_k \left(P(X_t = y_k).\prod_{j=1}^{J} P(X_{t-j} \mid X_t = y_k) \right) \tag{5}$$

To use this kind of predictor, we must establish the conditional probability table of the series and so quantify the last term of the equation. For the optimization we have used Matlab® software and we have been able to identify that the best prediction was obtained with 50 classes ($0 < k < 51$) and an order of $J = 3$.

2.3.3 Markov chains

In forecasting domain, some authors have tried to use so-called Markov processes [Logofet, D.O., Lesnaya, E.V., 2000. The mathematics of Markov models: what Markov chains can really predict in forest successions. Ecological Modelling 126, 285-298Muselli, M., Poggi, P., Notton, G., Louche, A., 2001. First Order Markov Chain Model for Generating Synthetic 'Typical Days' Series of Global Irradiation in Order to Design PV Stand Alone Systems. Energy Conversion and Management 42-6, 675-687], specifically the Markov chains. A Markov process is a stochastic process with the Markov property. Markov property means that, given the present state, future states are independent of the past states. In other words, the description of the present state fully captures all the information that could influence the future evolution of the process. Future states will be reached through a probabilistic process instead of a deterministic one. The proper use of these processes needs to calculate firstly the matrix of transition states. The transition probability of state i to the state j is defined by $p_{i,j}$. The family of these numbers is called the transition matrix of the Markov chain R, we have:

$$p_{i,j} = \mathrm{P}(X_t = j \mid X_{t-1} = i) \tag{6}$$

the formulation of the prediction can be resumed by:

$$X_{t+n} = X_t.R^n \tag{7}$$

The choice of the dimension of the transition matrix (number of class), and order of the chain (determination of the lag prediction) was done on Matlab®. We obtained an optimal prediction for 50 classes and an order of 3:

$$X_t = y_{k^*} = \arg\max_k \left((X_{t-1}.R + X_{t-2}.R^2 + X_{t-3}.R^3).e_k \right) \tag{8}$$

where e_k represents the set of basic vectors of the transition matrix.

2.3.4 K-Nearest Neighbors

The k-nearest neighbors' algorithm (k-NN) [Sharif, M., Burn, D.H., 2006. Simulating climate change scenarios using an improved K-nearest neighbor model. Journal of Hydrology, 325 1-4,179-196Yakowitz, S.. 1987. Nearest neighbors method for time series analysis. Journal of Time Series Analysis 8 : 235-247.] is a method for classifying objects based on closest training examples in the feature space. k-NN is a type of instance-based learning, or lazy learning where the function is only approximated locally and all computation is deferred until classification. It can also be used for regression. Unlike previous models, this tool does not use a learning base. The method consists in looking into the history of the series for the case the most resembling to the present case. By considering a series of observations X_t, to determine the next term $X_{t+1,}$ we must find among anterior information, which minimize the quantity defined on Eq.(9) (d is the quadratic error).

$$r_o = \arg\min_r \left(d(X_t, X_{t-r}) + d(X_{t-1}, X_{t-r-1}) + ... + d(X_{t-k}, X_{t-r-k}) \right) \tag{9}$$

In this study we have chosen a *k* equal to 10. After this argument of the minimum search, the prediction can be written:

$$X_{t+1} = X_{t-ro+1} \tag{10}$$

2.4 Artificial neural networks for prediction

Artificial neural networks (ANN) are intelligent systems that have the capacity to learn, memorize and create relationships among data. We present in this section an overview of this methodology in the context of time series prediction. An ANN is made up by simple processing units, the neurons, which are connected in a network

by a large number of weighted links where the acquired knowledge is stored and over which signals or information can pass. An input x_j is transmitted through a connection, which multiplies its strength by a weight w_{ij} to give a product $x_j w_{ij}$. This product is an argument to a transfer function f, which yields an output y_i represented as: $y_i = f(\sum x_j w_{ij})$ where i is an index of neurons in the hidden layer and j is an index of an input to the neural network [Mubiru, J. 2008. Predicting total solar irradiation values using artificial neural networks. Renewable Energy 33-10 2329-2332Cortez, P., Sollari Allegro, F., Rocha, M., Neves, J., 2002. Real-Time Forecasting by Bio-Inspired Models. Proceeding (362) Artificial Intelligence and Applications.]. In a Multi Layer Perceptron (MLP), neurons are grouped in layers and only forward connections exist. This provides a powerful architecture, capable of learning any kind of continuous nonlinear mapping, with successful applications ranging from Computer Vision, Data Analysis or Expert Systems, etc. [Cortez, P., Rocha, M., Neves, J., 2001. Evolving Time Series Forecasting Neural Network Models. Proceeding of int. symposium on adaptive systems: evolutionary computation and probabilistic graphical models.Faraway, J., Chatfield, C., 1995. Times series forecasting with neural networks: a case study, Research report 95-06 of the statistics group, University of Bath Jain, K., Jianchang, M., Mohiuddin, KM., 1996. Artificial neural networks: A tutorial, IEEE Computer, 29-3, 31–44Hu, Y., Hwang, J., 2002. Handbook of neural network signal processing. ISBN 0-8493-2359-2 ; Crone, SF., 2005. Stepwise Selection of Artificial Neural Networks Models for Time Series Prediction Journal of Intelligent Systems, Department of Management Science Lancaster University Management School Lancaster, United Kingdom]. One of the most interesting characteristic of ANNs is their ability to learn and to model a phenomenon. Figure 1 gives the basic architecture for a MLP application to time series forecasting during the learning task. To establish the prediction, a fixed number *p* of past values are set as inputs of the MLP, the output is the prediction of the future value of the time series. This method called “sliding-window technique” uses a moving time window to select N times, p inputs data for training.

Figure 1

In order to determine the best network configuration, we have tried to study all the parameters available in this network architecture. The principal parameters witch influence the number of local minima and the complexity of the network and its learning are: the number of input (these are the neurons which represent the input layer of the neural net); the number of hidden layer and their number of neurons; the activation (or transfer) function; the learning algorithm and the comparison function used during the learning phase. Additionally we have to think about the normalization of data, the learning sampling size and data distribution between the

learning, test and validation phases. For convenience we have chosen to optimize the parameters separately with an intuitive order of preference for all the ANN used in this work. Thus the chosen parameter optimization follows the pattern below (witch is adapted from the method described by Kalogirou, SA., 2001. Artificial neural networks in renewable energy systems applications: a review. Renewable and sustainable energy reviews 5, 373-401):

1. Number of input: 1 to 15 inputs
2. Number of hidden layers and neurons : 1 to 3 hidden layers, and 1 to 50 neurons by layer
3. Transfer function of hidden layer : linear (with or without saturating), exponential and sigmoid
4. Normalization of data : size of the interval for standardization between 0.6 to 1
5. Learning sample size : 500 to 6000 samples
6. Learning algorithm and parameters : all gradient descent, conjugate gradient, and quasi-Newton algorithms
7. Comparison function : MSNE (mean squared normalized error), MSE (mean square error), SSE (sum of square error), MAE (mean absolute error)
8. Data distribution (Learning - Test - Validation) : 60-20-20 and 80-10-10
9. Customization => new study on the number of hidden neurons, and the hidden layer to complete the process

As a result of this iterative process, the selected network has 3 layers: input, hidden and output layers. There was no significant difference in the use of 1, 2 and 3 hidden layers architectures. Considering this fact one hidden layer was used in order to minimize the complexity of the proposed ANN model. We tried several input layer configurations, best results were obtained with 8 inputs which received the endogenous entries S_{t-1},.. , S_{t-8} normalized on {0,1}. We found that 3 neurons on the hidden layer were sufficient. Finally, as we wanted to predict the solar radiation at horizon 1, we have one neuron on the output layer $\hat{S}_t$. Concerning the activation functions the best results were obtained with the Gaussian (hidden layer) and linear (output layer) functions. Concerning the training algorithm, many experiences enabled us to choose the Levenberg–Marquardt optimization (second-order algorithm) with 5000 epochs and μ decrease factor (learning rate) of 0.5, all other parameters have default values. The ANN configuration leads to an output signal corresponding to a nonlinear auto-regression to the data entry:

$$\hat{S}_t = \sum_{j=1}^{3} W_{2,j,1} e^{-\left(\sum_{n=1}^{8} W_{1,n,j} . S_{t-n} + b_{1,j}\right)^2} + b_{2,1} \qquad (11)$$

where parameters $W_{i,j,k}$ represent interconnection weights between layers and $b_{i,j}$ bias correction coefficients. The reader can find in the Appendix A the full list of these values for the most efficient ANN predictor.

We used the Matlab® software and its neural network toolbox to implement the network. The Matlab® learning, testing and validation were set respectively to 80%, 10% and 10%. These 3 steps constitute the training of the network, after which the weights and bias values are optimized. The training has concerned the years 1971 to 1987 and the performance function was the Mean Square Error (MSE). In the learning step, the data are presented to the ANN following a sequential order. The predicted values of global radiation are compared to the years 1988-89. Those years were not used during training.

3 Results and discussion

3.1 Results with non-stationary time series

In this section we examine if the proposed network was really interesting in terms of daily irradiation prediction. Indeed, we have compared its performance with the forecasting results obtained with a naive predictor (365 values representing the average over 17 years of the considered day), order 3 Markov chains, order 3 Bayesian inferences, an order 10 k-NN, an order 8 AutoRegressive AR(8). Table 1 and Figure 2 present results obtained in the case of prediction for a period of 2 years (1988-1989). The main comparison function used is the nRMSE for daily global radiation prediction (d+1). Moreover, classical statistical parameters enable us to evaluate the prediction quality: the absolute error is measured by the Root Mean Square Error (RMSE), the Mean Bias Error (MBE) witch gives an idea about the error in term of under or over estimation, determination coefficient (R^2) and Mean Absolute Error (MAE).

Table 1, Figure 2

As can be seen the predictors other than AR(8) and ANN give the same results, slightly better than those obtained with a naive predictor. Although AR and ANN could be considered as the best predictors, nRMSE are important considering from the point of view of energetic applications such as grid management with decentralized grid connected systems. In order to improve these results, the possibility to determine a signal pretreatment as input of AR and ANN is discussed. As proposed in [Diday, E., Lemaire, L., Pouget, J., Testu, F., 1982. Éléments d'analyse de données, Dunod, Paris], it appeared interesting to make stationary the time

series as much as possible. So we try to find a treatment in order to eliminate seasonal components without changing the other information. This treatment is called a seasonal adjustment process. In the present study, we have considered the solar irradiation like a seasonal phenomenon. The choice of the methodology used depends on the nature of the seasonality. In our case the seasonality is very pronounced and repetitive, so very deterministic and not stochastic. In the next sections, we present how we have used the physical modeling of radiation to determine a stationarization methodology for the daily signal.

3.2 Result with stationary time series

3.2.1 Preprocessing proposed

In this section we present all the steps of an ad-hoc time series preprocessing. Figure 3 summarizes the different steps we want to experiment. The 'classical methods' like Knn, Bayesian inferences and also naïve predictor don't require the use of preprocessing. In the literature, very few references related to way to make the data stationary. Concerning the Markov chains, we assumed that the simplest model without preprocessing was sufficient in comparison with others 'classical methods'. For the preprocessing, a first treatment (step 1) allows us to clean the series of non-typical points related to sensor maintenances or absence of measurement representing, 4.1% of measurements were missing and replacing by the hourly average over the 19 years for the given day.

Figure 3

Steps 2 and 3 will be described in the next section and lead to a series corrected utilizing clear sky or extraterrestrial normalizations. In case of stationary transformation by clear sky model the term $S_{d,y}^{corr}$ will be replaced by $S_{d,y}^{corr^*}$. The step 4 consists to use one of the forecasting methods which have been outlined in the previous section with the optimized MLP. Finally, step 5 allows to reverse the preprocessing treatment and to obtain the prediction of global irradiation. The VC parameter represents the signal dispersion. In the remainder of this paper, we choose the following naming convention: X_t designates the time series, and $X_{d,y}$ the modeling of the variables, where d is the day of the year y. To remove the seasonal dependencies, we chose to study both the "clearness index" and "clear sky index". According to Bird and Hulstrom [Bird, E., Hulstrom, R. L., 1981. A

Simplified Clear Sky Model for Direct and Diffuse Insolation on Horizontal Surfaces, SERI/TR-642-761, Solar Energy Research Institute, Golden, USA, Colorado], the global radiation on an horizontal plane under clear sky (standard Earth's atmosphere without cloud) is connected to the extraterrestrial radiation and solar elevation angle. In daily case, the periodic phenomenon of irradiation is only due to the extraterrestrial solar radiation coefficient for day d (H_0^d) [Badescu, V., 2008. Modelling Solar radiation at the earth surface, recent advances. ISBN: 978-3-540-77454-9, Viorel Ed.]. Thus we apply on the original series $X_{d,y}$ the ratio to trend method [Bourbonnais, R., Terraza, M., 2008. Analyse des séries temporelles. ISBN 9782100517077, 318p., Dunod Ed., Paris]. This leads to a new series $S_{d,y}$, known as *clearness index* series (often noted as K in scientific press):

$$S_{d,y} = \frac{X_{d,y}}{H_0^d} \tag{12}$$

In parallel, we wanted to test another method to remove seasonal effects, which is done by using the clear sky model for *clear sky index.* There are a lot of methods to determine this model. In our case, we preferred the simplified "Solis clear sky" model [Ineichen, P., 2008. A broadband simplified version of the Solis clear sky model. Solar Energy, 82-8, 758-762] based on radiative transfer calculations and the Lambert-Beer relation [Mueller, R.W., Dagestad, K.F., Ineichen, P., Schroedter-Homscheidt, M., Cros, S., Dumortier, D., Kuhlemann, R., Olseth, J.A., Piernavieja, G., Reise, C., Wald, L., Heinemann, D., 2004. Rethinking satellite-based solar irradiance modelling: The SOLIS clear-sky module. Remote Sensing of Environment 91, 160-174]. In this case the clear sky global horizontal irradiance ($H_{gh,clearsky}$) reaching the ground is defined by:

$$H_{gh,clearsky} = H_0 . e^{-(\tau / \sin^b(h))} . \sin(h) \tag{13}$$

where τ is the global total atmospheric optical depth (-0.37 in our case), h is the solar elevation angle and b is a fitting parameter (0.35 for us). The daily integration of the $H_{gh,clearsky}$ parameter allows us to determine the daily solar ratio for modeling $H^d_{gh,clearsky}$. A series of test (not presented in this paper) has allowed to validate the Solis model on horizontal and 80° tilted daily global radiation We obtain a relation "equivalent" to Eq.(12):

$$S^*_{d,y} = \frac{X_{d,y}}{H^d_{gh,clearsky}} \tag{14}$$

These treatments aim to create a new distribution without periodicity. Although the previous pre-treatment tends to make the time series stationary, a test of Fisher shows that seasonality was not optimal. According to Bourbonnais and Terraza [Bourbonnais, R., Terraza, M., 2008. Analyse des séries temporelles. ISBN 9782100517077, 318p., Dunod Ed., Paris], after using a ratio to trend method (Eqs 12 and 14) to correct rigid

seasonalities, a ratio to moving average can be used. This second ratio can be applied when there is no analytical expression of the trend. In this case, we find that H_o led a new seasonality which is difficult to model. That is the reason why we considered a moving average ratio to overcome this flexible seasonality:

$$y_{d,y} = \frac{S_{d,y}}{\frac{1}{2.m+1}.\sum_{i=-m}^{m} S_{d+i,y}} \tag{15}$$

In the present case, 2m+1 = 365 days (period of original time series), we obtain m = 182. The term $S_{d,y}$ is equivalent to $S^*_{d,y}$, they refer only to the chosen stationarization methodology. To complete the process, we use the 365 seasonal factors (y_d). These are coefficients which allow overcoming seasonality by a moving average ratio described above [Bourbonnais, R., Terraza, M., 2008. Analyse des séries temporelles. ISBN 9782100517077, 318p., Dunod Ed., Paris]. In order not to distort the series, we have considered that the total sum of the components of the series is the same before and after the report (final seasonal factors y_d* of the Eq.18). The transition coefficients (N = 18, the number of years of history) and the average coefficients of the regular 365 days are given by Eq.(16). A new series seasonally adjusted that represents only the stochastic component of global radiation is given by Eq.(19).

$$y_d = \frac{1}{N}\left(\sum_{y=1}^{N} y_{d,y}\right) \tag{16}$$

$$\bar{y}_d = \frac{1}{365}\left(\sum_{d=1}^{365} y_d\right) \tag{17}$$

$$y_d^* = \frac{y_d}{\bar{y}_d} \tag{18}$$

$$S_{d,y}^{corr} = \frac{S_{d,y}}{y_d^*} \quad and \quad S_{d,y}^{corr^*} = \frac{S_{d,y}^*}{y_d^*} \tag{19}$$

Thereafter, the juxtaposition of different values of $S_{d,y}^{corr}$ or $S_{d,y}^{corr^*}$ will be resumed on stationary time series noted S_t (extraterrestrial stationarization) and S_t^* (clear sky stationarization). After this step we needed to verify the effectiveness of pre-treatment on the time series. To achieve this we used a Fast Fourier Transform (FFT). The spectrum of the original series shows an important peak for the value 365 days that disappears after the pre-treatment.

3.2.2 Comparison of ARIMA and ANN

In Table 2, the statistical errors nRMSE obtained in the case of pre-treatment with the two best predictors (AR/ARMA and ANN) are shown. The results are systematically (but slightly) better with preprocessing. The reduction of the error prediction is low, but the difference between the two methods is significant, given the average and the confidence interval. A nRMSE reduction from 21.0% to 20.2% represents a gain of ~1%.

Table 2

Table 3 details in both ANN pre-processing normalizations (clear sky and clearness indexes) the annual prediction errors obtained for the years 1988 and 1989. In this study (global horizontal radiation) no significant differences between the both stationarization methodologies ("clearness index" and "clear sky index") are noted.

Table 3

The confidence interval is calculated after 8 training-simulations providing information on the prediction robustness. The weights are initiated before each simulation. With small confidence intervals, it can be said that there are very few local minimums. The ANN learning step error is obtained for RMSE < 10^{-4} (fixed before the treatment), leading to the stop of the learning. The monthly average error is the error (nRMSE) generated by monthly average values of the global radiation prediction. It's in fact the error of the monthly sum of global radiation. As can be seen the prediction is different from an average of 4% of the measured data. The negative MBE (-0.37MJ/m² and -0.32 MJ/m²) means that the solar potential is underestimated over the year. With non-ordinary low irradiation days (thick cover cloud during the whole days causing a very low sunshine duration), a tendency to overestimate is noted. The determination coefficient R^2 is greater than 0.8 that is good in relation to the noise present in measurements. There should be a compromise between RMSE and nRMSE. The nRMSE, computed from the mean global radiation obtained on the season, are useful for comparison and optimization (20.17% for clearness index and 20.25% for clear sky index). But for the absolute interpretation of received energy we must look at the RMSE. In order to better understand the predictor performances, Figure 4 shows the errors of prediction and distinguishes the seasons for the years 1988 and 1989 for a clear sky pre-processing. Best results in term of forecast are obtained in summer. These results can be used for example by energy managers who need to avoid using hydraulic power plants in dry season.

Figure 4

The spring season is the most difficult to predict with accuracy given the climate instabilities of this period. The absolute error is consistent. However, we find that in summer the error does not exceed 3.24 MJ / m² (900 Wh/m²), while the irradiation is important. MBE are found negative, which indicates an underestimation. The MBE is not significantly different from one season to another. Thus we always have the same prediction error, whatever the season. We systematically overestimate the days when the irradiation is minimal (winter). Moreover it is very difficult to predict the days when the irradiation had to be theoretically important. We would undoubtedly have improved the results optimizing an ANN by season, but it would complicate the process, and tend to decrease the procedure robustness.

After presenting the pre-treatment to be performed on the series and after verifying its effectiveness, we propose in the next section to validate this process on a facade PV system.

3.3 Frontage PV

In order to validate the approach proposed, we decided to use the previous simulator on a real frontage PV system installed recently in our laboratory (Vignola near Ajaccio-Campo del'Oro). Two additional difficulties have to be overcome: the PV modules are tilted 80° to horizontal and are located at 10 km from the meteorological station of Ajaccio-Campo del'Oro.

3.3.1 System description

The system has a nominal power of 6.525 kWp composed by 1.8 kWp and 4.725 kWp amorphous and mono-crystal PV modules respectively built in 6 independent power subsystems (Figure 5). PV power predictions from ANN methodology described in this paper have been computed from the central part of the PV plant on the front side exposed to the south (azimuth zero) and tilted at 80° (Index "PV" on Figure 5). The PV system consists of 9 SUNTECH 175S-24Ac with 1.175 kWp nominal power connected to a 1.85 kW SUNNY BOY SMA inverter for PV production on the grid. The irradiance sensor used (Figure 5) is an INGENIEURBÜRO SI-12TC calibrated by the PTB Braunschweig (German national metrology Institute): scale range between 0 and 1200 W/m² requiring an annual quality control for calibration.

Figure 5

For the daily PV power calculation, we use in first approximation, a classical linear production based on a constant PV plant efficiency $\eta_{PV} \sim 15.3\%$ ($R^2 = 0.9967$), with

$$E_{PV}\,(MJ) = \eta_{PV}\, I_\beta\, S \qquad (20)$$

Where I_β is the daily global irradiation on the PV system ($MJ.m^{-2}$, $\beta = 80°$), S is the usable surface of the PV system under consideration ($S = 10.125\ m^2$).

We propose to predict the DC electrical energy produced by PV modules, at 1 day horizon between the 15 January 2009 and the 15 June 2009. Unfortunately, no historical irradiation data (on synoptic Campo del'Oro station and on our laboratory) are available for $\beta = 80°$. So as it is shown in Figure 6 we planed to use an ANN trained on the site of Campo del'Oro (10 km from the PV modules) with horizontal global radiation data. The training phase described in section 3 is performed; only the prediction phase is different because we need to transpose the tilted data to horizontal data for use the ANN (and the reverse process as output).

Figure 6

3.3.2 Tilted global radiation

Before continuing, it is necessary to develop a methodology to convert the horizontal data for the tilt angle of the PV. This approach used the PVSYST software version 4 (www.pvsyst.com) to determine the global horizontal and inclined clear sky radiation. This software is very recognized in the field of renewable energies [Bouhouras, A.S., Marinopoulos, A.G., Labridis, D.P., Dokopoulos, P.S., 2009. Installation of PV systems in Greece–Reliability improvement in the transmission and distribution system, Electric Power Systems Research ; Qoaider, L., Steinbrecht, D., 2009 "Photovoltaic systems: A cost competitive option to supply energy to off-grid agricultural communities in arid regions," Applied EnergySpanos, I., Duckers, L., 2004. Expected cost benefits of building-integrated PVs in UK, through a quantitative economic analysis of PVs in connection with buildings, focused on UK and Greece, Renewable energy, vol. 29, p. 1289–1303.]. Eq. 21 has been built to convert the horizontal global radiation where β is the angle of inclination and I the global radiation. On the basis of a 6-months period, this equation leads to a nRMSE = 14% from measured and simulated daily tilted irradiation between 80° to 0° (RMSE = 2.4 MJ / m²). To validate this methodology, we have compared the tilt results with a

methodology more sophisticated based on the Climed-2 and Klucher models. The first model is used to determine the diffuse and beam fraction, and the second to tilt the diffuse component. In hourly case, the Climed2-Klutcher methodology is very convincing, [Noorian, A.M., Moradi, I., Kamali, G.A., 2008. Evaluation of 12 models to estimate hourly diffuse irradiation on inclined surfaces, Renewable Energy, vol. 33, p. 1406–1412 ; Notton, G., Poggi, P., Cristofari, C., 2006. Predicting hourly solar irradiations on inclined surfaces based on the horizontal measurements: Performances of the association of well-known mathematical models, Energy Conversion and Management, vol. 47, p. 1816–1829.Notton, G., Cristofari, C., Poggi, P., 2006. Performance evaluation of various hourly slope irradiation models using Mediterranean experimental data of Ajaccio, Energy Conversion and Management, vol. 47, p. 147–173.] but in daily case (integration of hourly data), the results are similar with the methodology proposed in this article. For both methodologies the nRMSE values are almost identical. Because the Climed-Klutcher methodology is only applicable on some locations near the Mediterranean Sea, we have chosen to develop in this article only the PVSYST method which is applicable everywhere. In the Eq 21, the diffuse and ground reflected radiations are implied in the bracketed term:

$$I_\beta = I_{\beta=0} \cdot \left[\frac{I_\beta}{I_{\beta=0}} \right]^{clearsky} \tag{21}$$

3.3.3 Predictor models and Results

In order to conduct an objective study, several forecasting experiments have to be completed in parallel. Their comparisons allow to generate interpretable results of PV power prediction (Eq.20). We have chosen to use models that have shown the best results in the horizontal case presented in previous sections, and also a naïve predictor for quantify the prediction quality. The selected models are:

A. Prediction omitting cloud cover: clear sky modeling

The Solis model (section 3.2.1) is used and coupled with the PVSYST process (Eq.21) to determine the tilted clear sky global radiation on Vignola site.

B. Prediction by average values

Daily solar radiation averages ($\beta = 0°$) are computed from the 19 years of historical data available on the site of Campo del'Oro, then inclined at 80° (Eq.21)

C. ANN *without* preprocessing

ANN is used with time series not made stionary, directly on the raw data.

D. ANN *with the clearness index*

This method follows the schematic of Figure 6. The time series is made stationary with the processing based on extraterrestrial radiation.

E. ANN with clear sky index

Equivalent to D but the processing used is the ratio to clear sky model (Solis)

F. ARMA(2,2) with clear sky index

Equivalent to E but the ANN are replaced by the linear methodology (ARMA)

Identical irradiation data from Campo del'Oro meteorological station are used as inputs for the processes B-F except model A where a physical model SOLIS computes clear sky conditions. Table 4 presents errors obtained on the period of prediction (January-June) and details of the bi-monthly errors. As can be seen on the 6-months period (January 15 – June 15, 2009), the ANN with clear sky index (E) methodology slightly outperforms the other. Indeed, all results computed are consistent (nRMSE=34.1%, RMSE=0.81 MJ, MBE=-0.03 MJ and MAE=0.62 MJ). But the differences are small among C, D, E and F experiments i.e. between the use of ANN or ARMA predictor, with or without preprocessing. The max value of RMSE is 0.85 MJ (C, ANN without preprocessing), the min is 0.81 MJ (E, ANN clear sky) and the difference is only 0.04 MJ/day. Only clear sky (A) and average (B) models are the worst forecasters (respectively nRMSE=46.6% and 47.6%). Bi-monthly errors show different behavior for prediction processes. The methods with clear sky index are the most relevant for the months for which the signal is very noisy (January-February prediction with E, ANN, and March-April with F, ARMA). In the case of May-June prediction the simple method, ANN without preprocessing (C) is the most relevant with nRMSE=15.5% while the sophisticate methods with preprocessing (D, E and F) generate nRMSE > 16.2%. The MBE factor is very weak for C, D, E and F (max = -0.11 MJ in F case is very acceptable). In the case of the wall-PV power system with 80° tilt, the maximum radiation value is obtained during the winter months concluding that the "E" process is the most relevant. ANN and ARMA perform almost similar, denoting the stochastic nature of the time series and thus the impossibility to predict the cloud effect on solar radiation.

Table 4

These two processes present no significant differences; only these two predictors are considered in the following. Figure 7 shows the cumulative daily predicted PV power versus measured ones from January 15 to June 15, 2009. This type of chart is very interesting to check the gap between measurement and simulation over

the long term (seasonal variations and under or overestimations of the PV energy). All figures show a good correlation between measured and predicted data ($R^2 > 0.99$). We understand easily that C, D, E and F methods are similar and although nRMSE are important, the accumulated predictions are almost identical to the measures (2.1% for ANN without preprocessing, 1.5% for ANN with clearly sky and clarity index and also for ARMA with clear sky index).

Figure 7

It is important to understand why the daily errors are significant. The first hypothesis is based on the high frequency noise series. In fact, the sampling frequency is the same as the frequency of the noise. It seems very unlikely that the ANN can predict “extra-ordinary days" at least if the previous day‘s cloud cover is ordinary. The second hypothesis is that ANNs (and ARMA) don’t take risks and propose irradiation value centered, on a mean value with a small standard deviation. The output of the network is then an “improved average” that fits the precursor trend of previous lags. Table 5 presents the decomposition for the ANN prediction errors (“E” process).

Table 5

The decomposition of the error is quite significant. We understand easily that the total error committed is important: the combination of preprocessing and modeling data (tilting, normalization for ANN) induces an error of about 20%. Furthermore the year 2009 (Jan-June) is not representative, there was an escalation of 13% error due to specific meteorological conditions (rainy period). Moreover, the 80° module tilt is very destabilizing, because the winter is a period of high climatic instability, but also a period where irradiation (80°) is the most important.

To try to compare all prediction methods (A-F) in a last step, we tried to present a new factor that reflects the tendency of prediction (the first derivative linked to the signal predicted slope). In the case of the measured signal, we use the Δ_m coefficient defined as follows (I is the measured values of PV electricity energy): $\Delta_m = I(t+1) - I(t)$, and in the case of prediction using the coefficient Δ_p ($\hat{I}$ represents the values predicted for the DC PV energy): $\Delta_p = \hat{I}(t+1) - I(t)$.

It is interesting to note that a Student's T-test between the coefficient Δ_m and Δ_p for ANN and ARMA shows that predictions with the ANN process better represent the reality (t = 0.364 and p = 0.71 for ANN vs t = 0.584 and p = 0.56 for ARMA), recall that *p*-value is the probability that the averages between the forecasting and measured values are identical. So according to this indicator, on average, ANN is more representative than ARMA for the trend estimation. Table 6 presents the coefficient comparing the quality of prediction based on the ratio $\Delta = \Delta_m / \Delta_p$ for all prediction methods (A-F). This type of parameter "Δ" allows determining 4 interpretation intervals:

- <0 ; number of very bad interpretation, for example : $\Delta_m > 0$, $\Delta_p < 0$, and vice versa,
- $[0;1-\varepsilon[$ and $]1+\varepsilon;+\infty[$: area where one commits an error of prediction; take $\varepsilon = 0.2$. In the first interval on a $\Delta_p > \Delta_m$ and the other interval $\Delta_p < \Delta_m$,
- $[1-\varepsilon;1+\varepsilon]$: good prediction zone, the sign and the absolute value is correct.

Table 6

Table 6 presents the "Δ" values and its results for the 4 classes considering for each interval, the number of events obtained on the studied period.

With this methodology we must recognize that ANN and ARMA are a little different. A-B processes are not efficient for daily DC PV energy predictions (Figure 8a). It must be said that it is very difficult to distinguish the quality of prediction made with ANN or ARMA (Figure 8b). Only the Student test used with the coefficient Δ_m and Δ_p allows separating the prediction quality of the two models with ANN slightly higher than ARMA.

Figure 8

4 Conclusion and perspectives

In this paper, an ANN prediction approach developed to determine global irradiation at daily horizon (d+1) which can help electrical managers with grid-PV power systems connected is presented. This prediction model has been compared to other prediction methods (AR, ARMA, k-NN, Markov Chains, etc.). We have used an ad-hoc time series preprocessing (based on clearness or clear sky indexes) and a time series prediction designed

MLP. Although the location was very specific, with the proximity to the sea and the mountain that can greatly affect nebulosity, we have obtained relevant results.

Without pre-processing, AR(8) and ANN models presents better daily RMSE of about 3.6 MJ/m² (998 Wh/m^{-2}) and nRMSE ~ 21% compared to Markov chain, Bayes, k-NN methods where nRMSE ~ 25 - 26%. However, annual pre-processing ANN methods based on clearness index and clear sky index reduce forecasting errors of about 5-6% (nRMSE ~ 20%) compared to classical predictors as Markov chains. The choice of a pre-processing built on clearness index (with H_0^d) of clear sky index ($H_{gh,clearsky}^d$) leads to comparable results.

These simulation tools have been successfully validated on the DC energy prediction of a 1.175 kWp mono-Si PV power system connected to the grid. On a seasonal point of view, ANN with clear sky pre-processing, represents for winter months (January, February) an adequate solution (nRMSE ~ 37%). For summer months, ANN without pre-processing gives the best results (nRMSE ~ 15 %). Cumulative simulated and measured PV productions are in very good agreement ($R^2 > 0.99$) validating the whole prediction process (nRMSE < 2% for a 6-months period).

Finally, a new differential variable Δ has been introduced to study predicted tendency errors based on the first derivative predicted signal. According to a Student test, ANN and ARMA simulated daily irradiation profiles confirm the good accuracy for the predicted tendency at d + 1. These two methodologies are similar.

In the future, this tool (ANN) could eventually help the system manager, for implanting new PV systems especially on isolated electrical grid where 30% for renewable energy power system represents an integration limit. With operational prediction tools, this threshold can be increased to 50% allowing to limit the utilization of fossil plants to supply electricity.

Moreover, it seems important to study shorter time horizons (hour). As a matter of fact, electrical managers are also interested to horizons that can range from ½ hour to several hours: from 3 hours to 24 hours. In this new configuration, other ANN architecture types have to be studied: time delay, recurrent ANNs, etc.

In the long-term, it would be also very interesting to study a network trained on an urban data, and used on other site with equivalent geographical feature, and maybe combine both ANN and Geographic Information Systems (GIS) approaches.

References

Badescu, V., 2008. Modelling Solar radiation at the earth surface, recent advances. ISBN: 978-3-540-77454-9, Viorel Ed.

Bird, E., Hulstrom, R. L., 1981. A Simplified Clear Sky Model for Direct and Diffuse Insolation on Horizontal Surfaces, SERI/TR-642-761, Solar Energy Research Institute, Golden, USA, Colorado

Bouhouras, A.S., Marinopoulos, A.G., Labridis, D.P., Dokopoulos, P.S., 2009. Installation of PV systems in Greece–Reliability improvement in the transmission and distribution system, Electric Power Systems Research

Bourbonnais, R., Terraza, M., 2008. Analyse des séries temporelles. ISBN 9782100517077, 318p., Dunod Ed., Paris

Brockwell, P. J., Davis, R. A., 2006. Time series: theory and methods. ISBN 0387974296 (USA). 577p. Springer series in statistics, second edition

Celeux, G., Nakache, J.P., 1994. Analyse discriminante sur variables qualitatives ISBN 2840540274, Polytechnica, 270 p., Paris

Cortez, P., Rocha, M., Neves, J., 2001. Evolving Time Series Forecasting Neural Network Models. Proceeding of int. symposium on adaptive systems: evolutionary computation and probabilistic graphical models.

Cortez, P., Sollari Allegro, F., Rocha, M., Neves, J., 2002. Real-Time Forecasting by Bio-Inspired Models. Proceeding (362) Artificial Intelligence and Applications.

Crone, SF., 2005. Stepwise Selection of Artificial Neural Networks Models for Time Series Prediction Journal of Intelligent Systems, Department of Management Science Lancaster University Management School Lancaster, United Kingdom

Diday, E., Lemaire, L., Pouget, J., Testu, F., 1982. Éléments d'analyse de données, Dunod, Paris

Dubois, E., Michaux, E., 2008: "Grocer: an econometric toolbox for Scilab", available at http://dubois.ensae.net/grocer.html

Faraway, J., Chatfield, C., 1995. Times series forecasting with neural networks: a case study, Research report 95-06 of the statistics group, University of Bath

Hamilton, J.D., 1994. Times series analysis. ISBN 0-691-04289-6

Hay, J.E., Davies, J.A., 1980. Calculation of the solar radiation incident on an inclined surface. In: Proc. First Canadian Solar radiation workshop, 59–72

Hocaoglu, FO., Gerek, ON., Kurban, M., 2008. Hourly solar forecasting using optimal coefficient 2-D linear filter and feed-forward neural networks. Solar energy 82-8, 714-726

Hu, Y., Hwang, J., 2002. Handbook of neural network signal processing. ISBN 0-8493-2359-2

Ineichen, P., 2008. A broadband simplified version of the Solis clear sky model. Solar Energy, 82-8, 758-762

Ineichen, P., Guisan, O., Perez, R., 1990. Ground-reflected radiation and albedo. Solar Energy;44-4, 207–214

Jain, K., Jianchang, M., Mohiuddin, KM., 1996. Artificial neural networks: A tutorial, IEEE Computer, 29-3, 31–44

Kalogirou, SA., 2001. Artificial neural networks in renewable energy systems applications: a review. Renewable and sustainable energy reviews 5, 373-401

Liu, B.Y.H., Jordan, R.C., 1962. Daily insolation on surfaces tilted towards the equator. Trans SHRAE; 67, 526–541

Logofet, D.O., Lesnaya, E.V., 2000. The mathematics of Markov models: what Markov chains can really predict in forest successions. Ecological Modelling 126, 285-298

Mellit A., Kalogirou, SA.. 2008. Artificial intelligence techniques for photovoltaic applications: A review. Progress in Energy and Combustion Science 1-1, 52-76

Mellit, A., Kalogirou, S.A., Hontoria, L., Shaari, S. 2009. Artificial intelligence techniques for sizing photovoltaic systems: A review. Renewable and Sustainable Energy Reviews 13-2, 406-419

Mubiru, J. 2008. Predicting total solar irradiation values using artificial neural networks. Renewable Energy 33-10 2329-2332

Mubiru, J., Banda, E., 2008. Estimation of monthly average daily global solar irradiation using artificial neural networks. Solar Energy, 82-2, 181-187

Mueller, R.W., Dagestad, K.F., Ineichen, P., Schroedter-Homscheidt, M., Cros, S., Dumortier, D., Kuhlemann, R., Olseth, J.A., Piernavieja, G., Reise, C., Wald, L., Heinemann, D., 2004. Rethinking satellite-based solar irradiance modelling: The SOLIS clear-sky module. Remote Sensing of Environment 91, 160-174

Muselli, M., Poggi, P., Notton, G., Louche, A., 2001. First Order Markov Chain Model for Generating Synthetic 'Typical Days' Series of Global Irradiation in Order to Design PV Stand Alone Systems. Energy Conversion and Management 42-6, 675-687

Noorian, A.M., Moradi, I., Kamali, G.A., 2008. Evaluation of 12 models to estimate hourly diffuse irradiation on inclined surfaces, Renewable Energy, vol. 33, p. 1406–1412

Notton, G., Poggi, P., Cristofari, C., 2006. Predicting hourly solar irradiations on inclined surfaces based on the horizontal measurements: Performances of the association of well-known mathematical models, Energy Conversion and Management, vol. 47, p. 1816–1829.

Notton, G., Cristofari, C., Poggi, P., 2006. Performance evaluation of various hourly slope irradiation models using Mediterranean experimental data of Ajaccio, Energy Conversion and Management, vol. 47, p. 147–173.

Perez, R., Ineichen, P., Seals, R., 1990. Modelling daylight availability and irradiance components from direct and global irradiance. Solar Energy 44-5, 271–289

Pole, A., West, M., Harrison, J., 1994. Applied Bayesian forecasting and time series analysis. ISBN 0412044013. Chapman and Hall/CRC.

Qoaider, L., Steinbrecht, D., 2009 “Photovoltaic systems: A cost competitive option to supply energy to off-grid agricultural communities in arid regions,” Applied Energy

Reindl, D.T., Beckman, W.A., Duffie, J.A., 1990. Evaluation of hourly tilted surface radiation models. Solar Energy; 45-1, 9-17

Sharif, M., Burn, D.H., 2006. Simulating climate change scenarios using an improved K-nearest neighbor model. Journal of Hydrology, 325 1-4,179-196

Spanos, I., Duckers, L., 2004. Expected cost benefits of building-integrated PVs in UK, through a quantitative economic analysis of PVs in connection with buildings, focused on UK and Greece, Renewable energy, vol. 29, p. 1289–1303.

Yakowitz, S.. 1987. Nearest neighbors method for time series analysis. Journal of Time Series Analysis 8 : 235-247.

Zhang, GP., Qi, M., 2005. Neural network forecasting for seasonal and trend time series, European Journal of Operational Research 160, 501-514

Acknowledgements

This work was partly supported by the Territorial Collectivity of Corsica. We thank the French National Meteorological Organization (Météo-France) which has supervised the data collection from their data bank for Campo del'Oro synoptic station. We thank Jean PANIGHI (University of Corsica) for the supply of irradiance sensors technical characteristics.

List of captions:

Figure 1. MLP application to time series forecasting (bias nodes are not displayed).

Figure 2. Comparison between measures and forecasts for the daily global horizontal radiation. (a): average value; (b): Markov chain; (c): Bayesian inference; (d): knn; (e): AR ; (f): ANN. The thick line represents the graph y = x and the normal line, the linear regression.

Figure 3. Summarize of the protocol followed to obtain the predicted irradiation. VC is the variation coefficient (ratio to standard deviation to the means).

Figure 4. Seasonal errors for the daily prediction of the years 1988 and 1989 (mean with 95% confidence interval).

Figure 5. Frontage PV system (6.525 kWp). The PV plant under study is mentioned by the "PV" mark. Tilted (80°) daily solar radiation I_β is measured on the top of the system facing south.

Figure 6. PV energy prediction methodology based on a daily horizontal irradiation ANN simulator

Figure 7. Cumulative predictions and measures, (a): ANN without preprocessing, (b): ANN with clearness index, (c): ANN with clear sky index and (d): ARMA with clear sky index.

Figure 8. Distribution of Δ-values (X-axis) representing the quality of prediction. (a) : A-B processes not relevant, (b) : C-F processes corresponding to best ANN and ARMA simulators

Table 1. Evaluation of the prediction quality for all prediction methods, forecasting years 1988 and 1989, 1 day horizon

Table 2. Annual errors for ARMA and ANN prediction methods (nRMSE ± 95% CI), forecasting years 1988 and 1989, 1 day horizon

Table 3. Annual prediction error for the years 1988 and 1989 with our MLP

Table 4. Bi-monthly prediction errors and 6-months prediction errors on the electrical energy from PV plant (1.175 kWp, $\beta = 80°$) for 1 day horizon from 15 January 2009 to 15 June 2009. The mean and variance are the characteristics of the global radiation time series on the period.

Table 5. Decomposition of the nRMSE for electrical energy prediction ("E" process).

Table 6. Tendency of prediction with Δ-values distributions for all processes

APPENDIX A

MLP weight values after training with clarity index for the global radiation (most efficient configuration). W1 corresponds to the matrix where each element $W1_{n,j}$, is the weight related to the number n of input neurons (from 1 to 8) and the number j of hidden neurons (from 1 to 3). W2 corresponds to a vector with 3 elements representing the weights between the 3 hidden nodes and the output. B1 is the bias vector related to each hidden neuron (3 values) and B2 the output bias value.

$$W1 = \begin{pmatrix} 1.26 & 0.02 & 0.26 & -0.001 & -0.08 & 0.34 & 0.12 & 0.13 \\ 0.02 & -0.19 & 0.29 & -0.89 & 0.41 & 0.64 & 0.78 & -0.85 \\ 0.21 & 0.50 & -0.14 & -0.09 & -0.61 & -0.48 & 0.49 & -0.02 \end{pmatrix}$$

$$W2 = \begin{pmatrix} 0.47 & -0.07 & -0.07 \end{pmatrix}$$

$$B1 = \begin{pmatrix} -0.90 \\ -0.53 \\ 1.02 \end{pmatrix}$$

$$B2 = (-0.13)$$

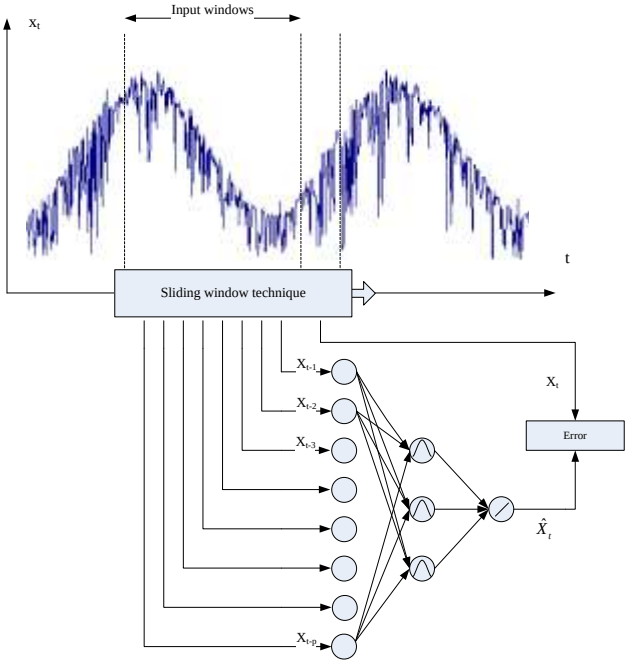

Input windows
x_t
t
Sliding window technique
X_{t-1}
X_{t-2}
X_{t-3}
X_{t-p}
x_t
Error
$\hat{X}_t$

Figure 1.

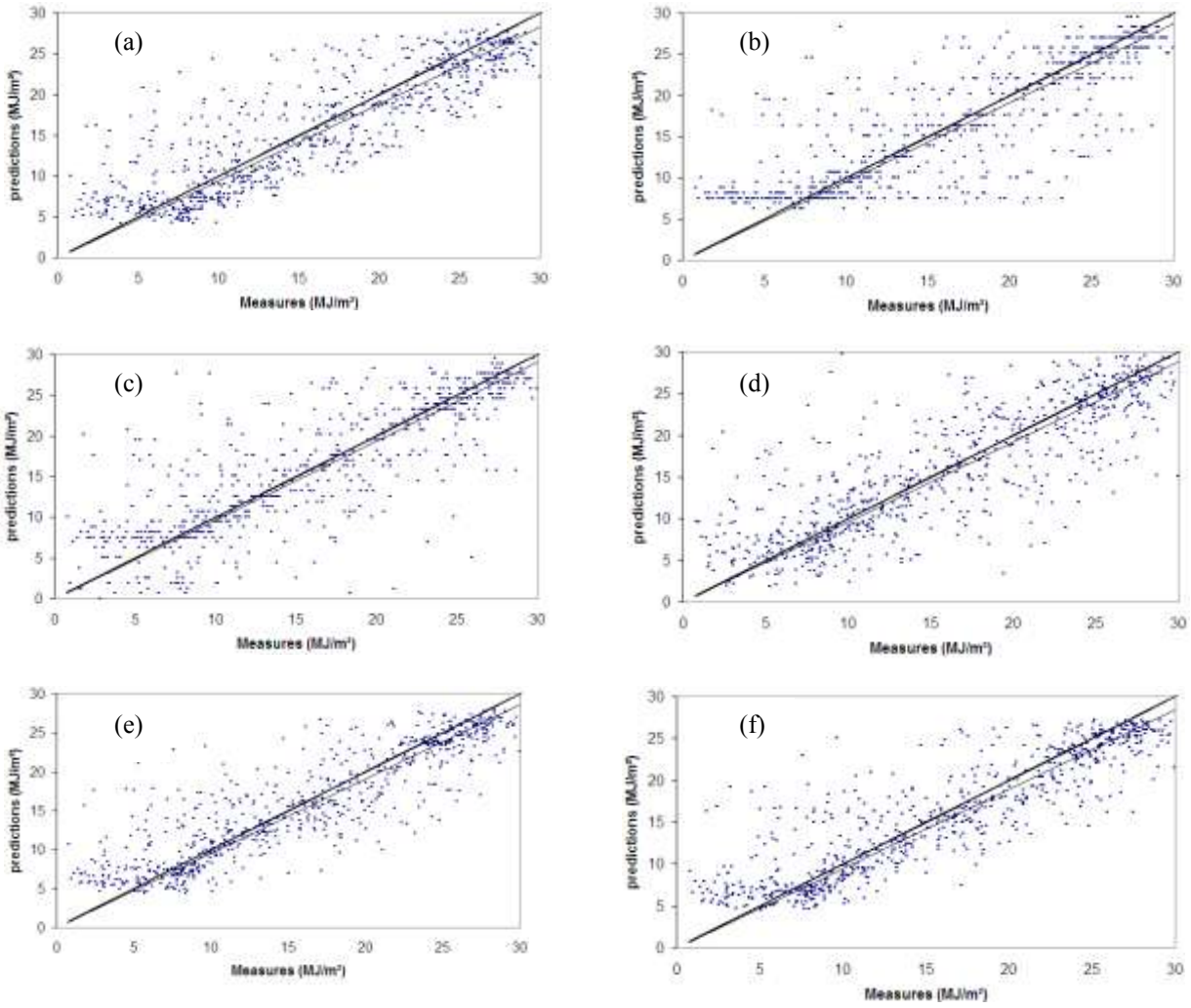


Figure 2.

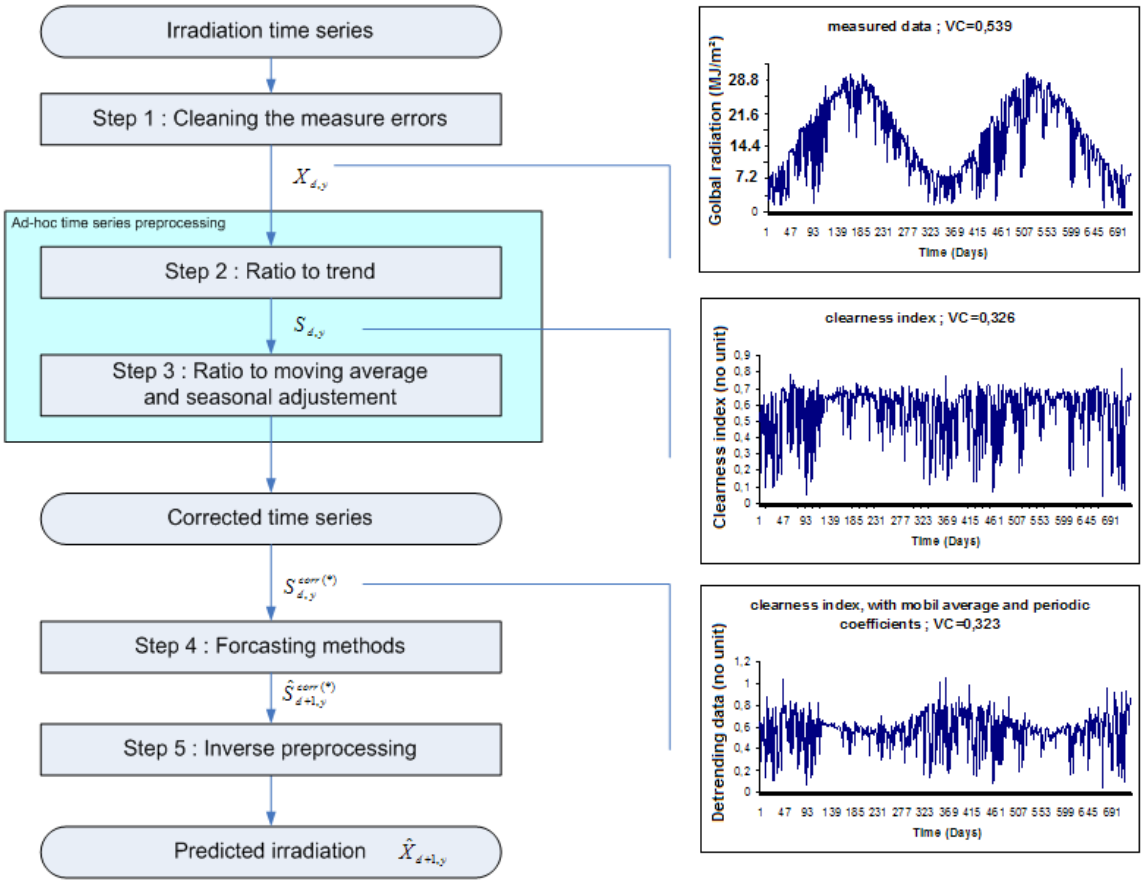

Irradiation time series
Step 1 : Cleaning the measure errors
$X_{d,y}$
Ad-hoc time series preprocessing
Step 2 : Ratio to trend
$S_{d,y}$
Step 3 : Ratio to moving average and seasonal adjustement
Corrected time series
$S^{corr(*)}_{d,y}$
Step 4 : Forcasting methods
$\hat{S}^{corr(*)}_{d+1,y}$
Step 5 : Inverse preprocessing
Predicted irradiation $\hat{X}_{d+1,y}$
measured data ; VC=0,539
Golbal radiation (MJ/m²)
28.8
21.6
14.4
7.2
0
1 47 93 139 185 231 277 323 369 415 461 507 553 599 645 691
Time (Days)
clearness index ; VC=0,326
Clearness index (no unit)
0,9
0,8
0,7
0,6
0,5
0,4
0,3
0,2
0,1
0
1 47 93 139 185 231 277 323 369 415 461 507 553 599 645 691
Time (Days)
clearness index, with mobil average and periodic coefficients ; VC=0,323
Detrending data (no unit)
1,2
1
0,8
0,6
0,4
0,2
0
1 47 93 139 185 231 277 323 369 415 461 507 553 599 645 691
Time (Days)

Figure 3.

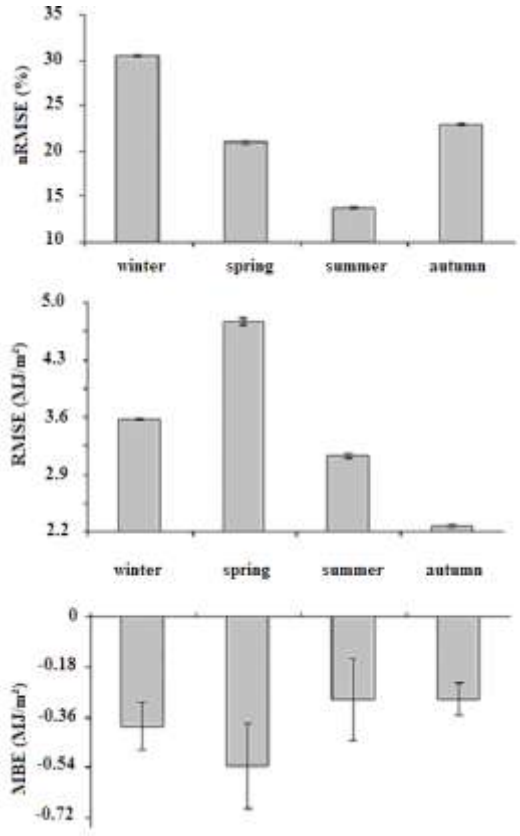

35
30
25
20
15
10
nRMSE (%)
winter
spring
summer
autumn
5.0
4.3
3.6
2.9
2.2
RMSE (MJ/m²)
winter
spring
summer
autumn
0
-0.18
-0.36
-0.54
-0.72
MBE (MJ/m²)

Figure 4.

Figure 5.

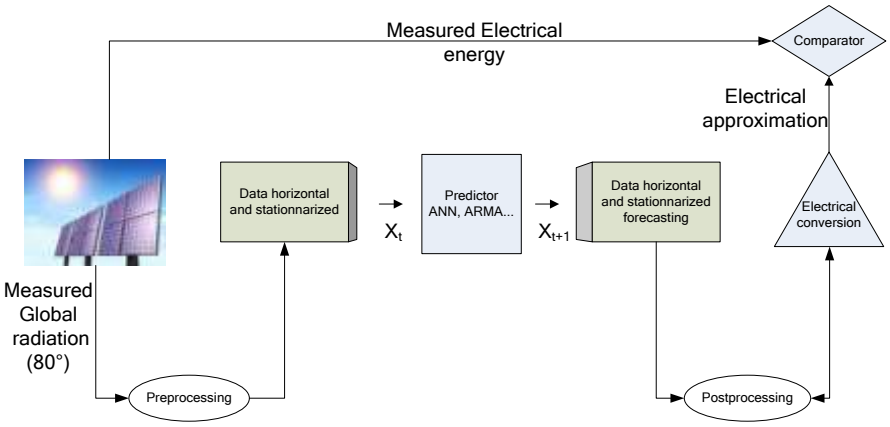

Measured Electrical
energy
Comparator
Electrical
approximation
Data horizontal
and stationnarized
Predictor
ANN, ARMA...
Data horizontal
and stationnarized
forecasting
Electrical
conversion
X_t
X_{t+1}
Measured
Global
radiation
(80°)
Preprocessing
Postprocessing

Figure 6.

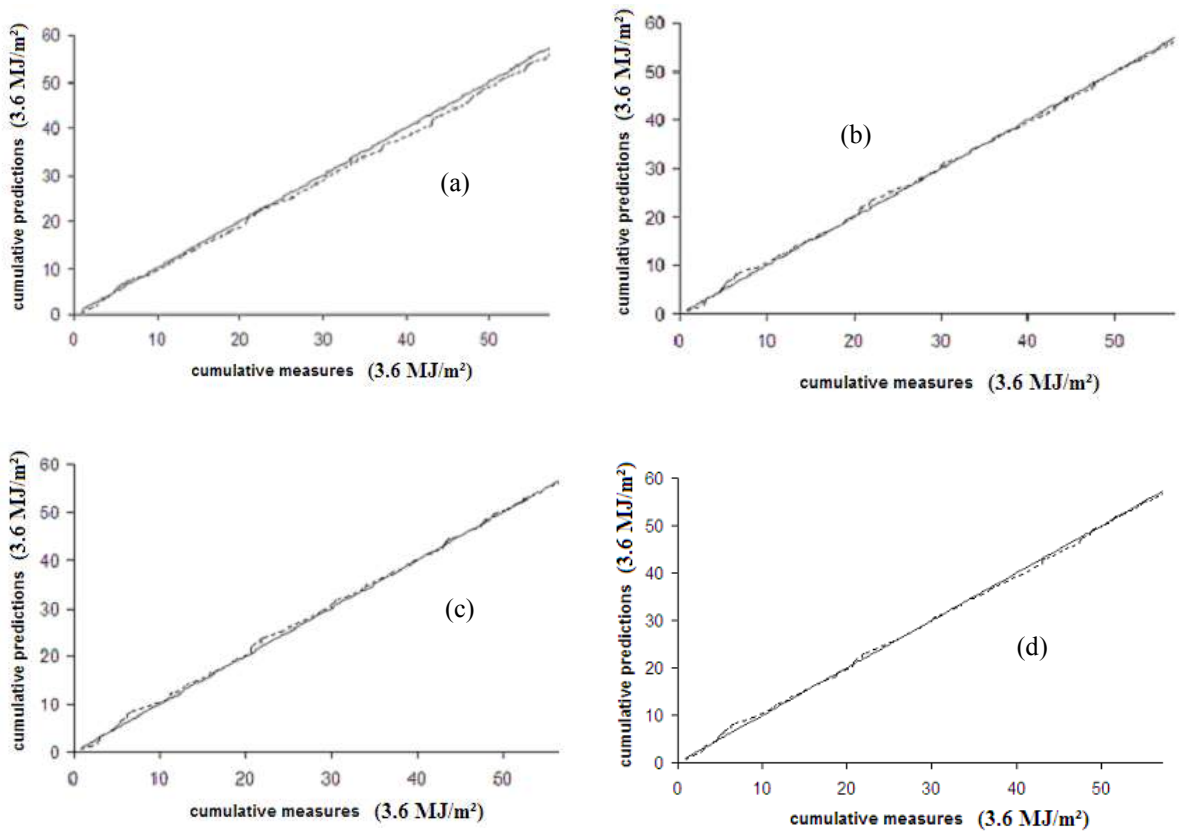
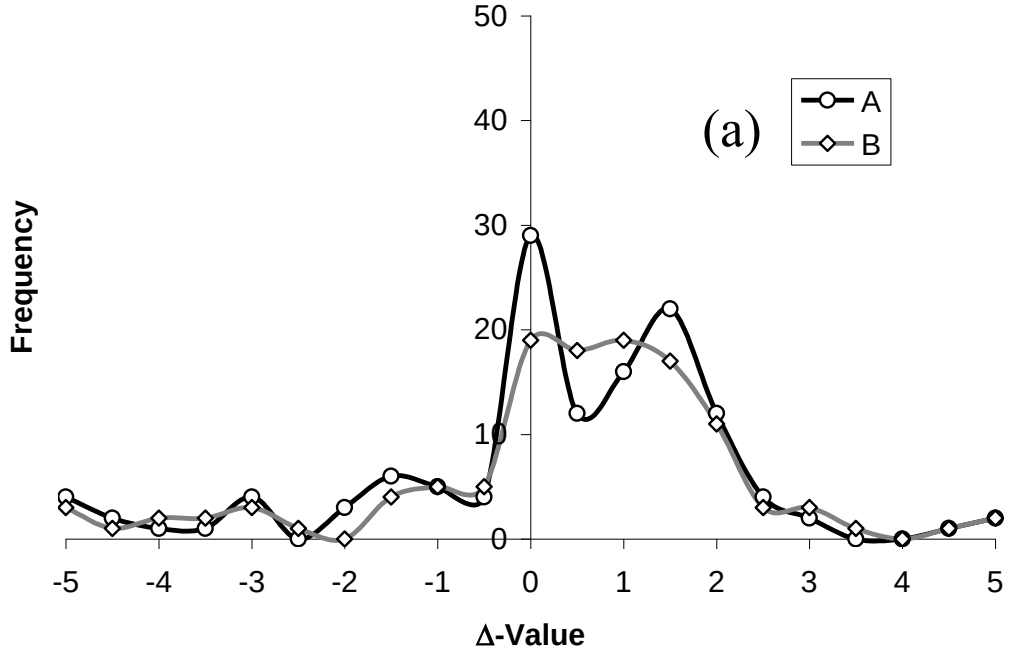

cumulative predictions (3.6 MJ/m²)
cumulative measures (3.6 MJ/m²)
(a)
(b)
(c)
(d)
0
10
20
30
40
50
60

Figure 7.


(a)
A
B
Frequency
Δ-Value
50
40
30
20
10
0
-5
-4
-3
-2
-1
0
1
2
3
4
5

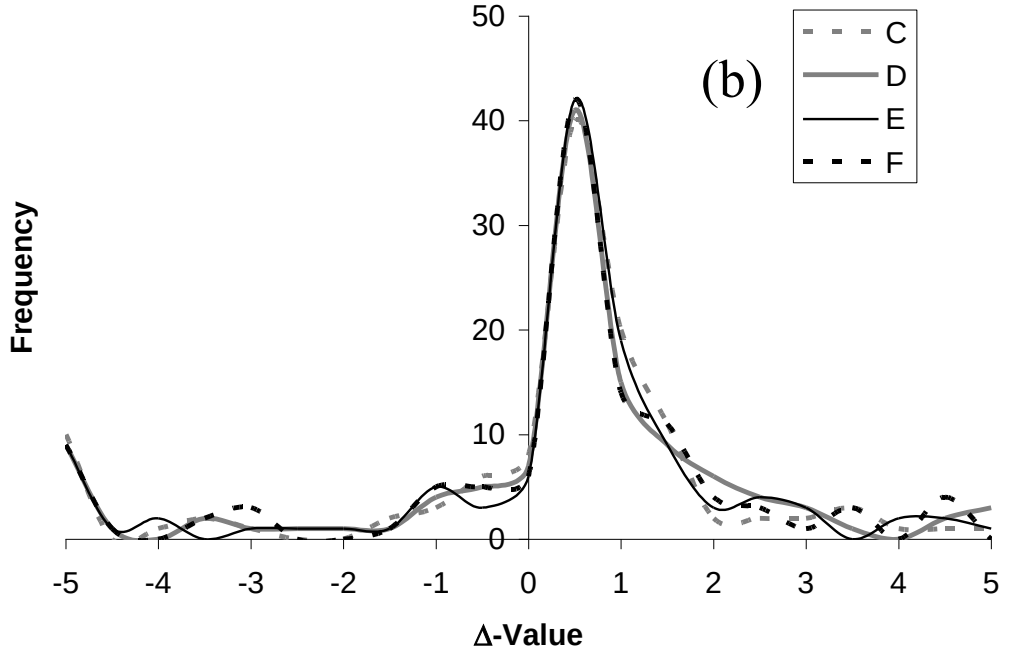

(b)
C
D
E
F
Frequency
Δ-Value
50
40
30
20
10
0
-5
-4
-3
-2
-1
0
1
2
3
4
5

Figure 8.

	nRMSE (%) ± 95% IC	**RMSE (MJ/m²)**	**MAE (MJ/m²)**	**R²**	**MBE (MJ/m²)**
Naïve predictor (persistence)	26.13±0.00	4.65	3.03	0.69	-0.001
Naïve predictor (daily average)	22.52±0.00	4.01	3.11	0.75	-0.39
Markov Chain (order 3)	25.85±0.00	4.59	3.03	0.69	-0.23
Bayes (order 3)	25.57±0.00	4.55	3.01	0.69	-0.13
k-NN (order 10)	25.20±0.00	4.48	3.19	0.70	-0.03
AR(8)	21.18±0.00	3.77	2.85	0.78	-0.61
ANN[8,3,1]	20.97±0.15	3.73	2.82	0.79	-0.58

Table 1.

	Clearness index methodology	**Clear sky index methodology**
AR(8) without preprocessing	21.18 ± 0%	21.18 ± 0%
ARMA(2,2) with preprocessing	20.31 ± 0%	20.32 ± 0%
ANN[8,3,1] without preprocessing	20.97 ±0.15%	20.97 ±0.15%
ANN[8,3,1] with preprocessing	20.17 ± 0.1%	20.25 ± 0.1%

Table 2.

	Clearness index methodology	**Clear sky index methodology**
nRMSE (%)	20.17 ± 0.1%	20.25 ± 0.1%
RMSE (MJ/m²)	3.59	3.60
MAE (MJ /m²)	2.65	2.67
MBE (MJ /m²)	-0.37	-0.32
R^2	0.801	0.790
Monthly average error (%)	3.9	4.1

Table 3.

	A clear sky	B average	C ANN	D ANN Clearness	E ANN clear sky	F ARMA clear sky
Jan-Feb	Mean : 2.39 MJ/m²			Std Dev : 0.062		
nRMSE (%)	56.4	57.1	41.7	38.7	**37.9**	38.4
RMSE (MJ)	1.47	1.49	1.09	1.01	**0.99**	1.01
MBE (MJ)	1.04	1.02	-0.13	**-0.01**	0.02	-0.04
MAE (MJ)	1.06	1.12	0.98	0.90	**0.87**	0.89
March-April	Mean : 2.27 MJ/m²			Std Dev : 0.059		
nRMSE (%)	49.3	45.1	38.3	38.3	37.9	**37.7**
RMSE (MJ)	1.22	1.11	0.95	0.95	0.94	**0.93**
MBE (MJ)	0.79	0.58	-0.01	-0.03	-0.02	**0.01**
MAE (MJ)	0.79	0.76	0.76	0.77	0.77	**0.75**
May-June	Mean : 1.95 MJ/m²			Std Dev : 0.031		
nRMSE (%)	17.9	18.9	**15.5**	16.2	16.4	16.4
RMSE (MJ)	0.35	0.37	**0.31**	0.32	0.32	0.33
MBE (MJ)	0.20	0.08	**-0.03**	-0.07	-0.09	-0.12
MAE (MJ)	0.66	0.79	0.63	0.63	**0.62**	0.62
January-June	Mean : 2.19 MJ/m²			Std Dev : 0.055		
nRMSE (%)	46.6	47.6	35.6	34.7	**34.1**	34.2
RMSE (MJ)	1.11	1.13	0.85	0.82	**0.81**	**0.81**
MBE (MJ)	0.65	0.64	-0.05	**-0.03**	**-0.03**	-0.05
MAE (MJ)	0.66	0.79	0.63	0.63	**0.62**	**0.62**

Table 4.

	nRMSE
Preprocessing and modeling	20%
Specificity of year 2009	13%
Electrical conversion	~ 1%
Total observed error	**34%**

Table 5.

Δ-values	A	B	C	D	E	F
< 0	59	45	33	32	30	32
0 - 0.8	17	24	54	53	56	53
0.8 - 1.2	22	23	11	5	7	8
> 1.2	38	44	38	46	43	43

Table 6.