HAL
open science

Solar Radiation Forecasting Using Ad-Hoc Time Series Preprocessing and Neural Networks.

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

To cite this version:

Christophe Paoli, Cyril Voyant, Marc Muselli, Marie Laure Nivet. Solar Radiation Forecasting Using Ad-Hoc Time Series Preprocessing and Neural Networks.. International Conference on Intelligent Computing (ICIC 2009), Sep 2009, Ulsan, North Korea. pp.898-907, 10.1007/978-3-642-04070-2_95 . hal-00438781

HAL Id: hal-00438781

https://hal.science/hal-00438781v1

Submitted on 4 Dec 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Solar radiation forecasting using ad-hoc time series preprocessing and neural networks [1]

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

[1] University of Corsica, CNRS UMR SPE 6134, (Vignola, Rte des Sanguinaires, 20000 Ajaccio | Campus Grimaldi, 20250 Corte), France
[2] Hospital of Castelluccio, Radiotherapy Unit, BP 85, 20177 Ajaccio, France

{christophe.paoli, cyril.voyant, marc.muselli, marie-laure.nivet}@univ-corse.fr

Abstract. In this paper, we present an application of neural networks in the renewable energy domain. We have developed a methodology for the daily prediction of global solar radiation on a horizontal surface. We use an ad-hoc time series preprocessing and a Multi-Layer Perceptron (MLP) in order to predict solar radiation at daily horizon. First results are promising with nRMSE < 21% and RMSE < 998 Wh/m². Our optimized MLP presents prediction similar to or even better than conventional methods. Moreover we found that our data preprocessing approach can reduce significantly forecasting errors.

Keywords: Time Series, Preprocessing, Seasonality, Multi-Layer Perceptron

1 Introduction

Artificial intelligence techniques are becoming more and more popular in the renewable energy domain [1], [2] and particularly for the prediction of meteorological data such as solar radiation [3], [4] [5] [6]. Thereby many research works have shown the ability of Artificial Neural Networks (ANNs) to predict times series of meteorological data. In this study and according to electricity suppliers, we focus on the prediction of global solar irradiation on a horizontal plane for daily horizon. In this way, we have investigated time series forecasting which is a challenge in many fields. Because it has made tremendous progress in the past twenty years in terms of theory, algorithms and applications, we have chosen to study ANNs. 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 [1], [2]. The originality of our study is to add an ad-hoc time series preprocessing step before using neural networks. Indeed, as seen in [7] a data preprocessing including deseasonalization and detrending can improve ANN forecasting performance. As

[1] A part of this research is founded by the Territorial Collectivity of Corsica.

global solar radiation has a deterministic part, we want to take into account this specificity.

The paper is organized as follow: the section 2 describes the physical phenomena we want to predict and introduces our ad-hoc time series preprocessing. Section 3 presents the neural network architecture we designed. Results are presented and discussed in section 4 where several conventional methods for estimation and modeling of the meteorological data are compared with our methodology. Section 5 concludes and suggests perspectives.

2 Data analysis and preprocessing

There are two approaches that allow quantifying solar radiation: the "physical modeling" based on physical processes occurring in the atmosphere and influencing solar radiation [8], and the "statistic solar climatology" mainly based on time series analysis [8]. As already said we have chosen to combine these two methods in a gray box approach to improve the quality of prediction. In this work, we have used the physical phenomena in an attempt to overcome the seasonality of the resource.

2.1 Meteorological data

Our study proposes to analyze the radiation time series (Wh.m-2) measured at the meteorological station of Ajaccio (METEO FRANCE, Corsica, France, 41°55'N, 8°44'E). The data representing the global solar radiation were measured on a daily basis from January 1971 to December 1989. Thus we have a X_t time series to forecast for time t+1; that is at horizon 1. To achieve this, we choose to use a gray box (or semi-physical) model where time series prediction and modeling are mixed. For time series prediction only past values are used to forecast the future values at a given horizon. In the case of modeling, the different physical processes involved are taken into account in order to represent the variable however the horizon is. In the remainder of this paper, we choose the following naming convention: X_t designate the time series, and $X_{d,y}$ is the modeling of the variables, where d is the day of the year y. In the next section, an explanation of the physical phenomenon is proposed, and then we describe our time series preprocessing.

2.2 Physical phenomenon

We can observe in Figure 1 that the global radiation consists of three types of radiation: direct radiation, diffuse radiation and ground-reflected radiation [10]. The ground-reflected radiation does not concern us because we try to predict the radiation on a horizontal surface where the ground reflected radiation does not make sense. For clear sky, ie without cloud cover, global radiation is relatively easy to model because it is primarily due to the distance from the sun sensor [9], [10], [11], [12], [13]. It is not the same, when there are clouds near the detectors. Indeed, these are mostly stochastic phenomena, which depend on the weather site.

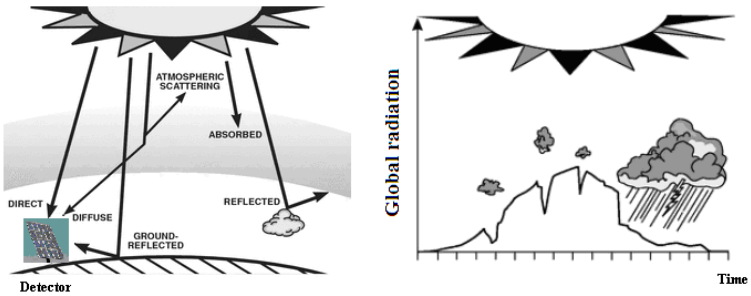

Fig. 1. Origin of the three types of radiation: direct radiation, diffuse radiation and ground-reflected radiation on a detector at ground level (left of the figure). Modification of the global irradiation profile accordingly to clouds cover (right of the figure).

The spectral analysis of the global radiation series highlights the high periodicity of the phenomenon (almost 365 days period). As proposed in [7], it appeared wise to make the time series stationary as much as possible. Deseasonalize and detrend a series allows to eliminate seasonal and trend components without changing the other information. In the present study, we have not considered the solar irradiation like a trend process but only like a seasonal process. The choice of the methodology used depends on the nature of the seasonality. As in our case the seasonality is very pronounced and repetitive, so very deterministic and not stochastic. Moreover, it is possible to physically quantify the components of our irradiation series. Like we have seen in figure 1, the global solar irradiance on horizontal plane depends on direct and diffuse radiation. This specificity allows us to apprehend the periodicity of the phenomenon, and to reduce the non-stationarity of the series.

In the next section, we propose to stationarize radiation data to overcome the deterministic component which is easily quantifiable. Thus we devote to the prediction of cloud cover included in the global radiation of the site.

2.3 Ad-hoc time series preprocessing

In fact, by dividing the series by daily extraterrestrial radiation [14], we can quantify the annual periodicity. It is the first step of our stationarization process. In first approximation, it is possible to quantify the deterministic component of global radiation by the extraterrestrial solar radiation alone (H_0). Thus we apply on the original series $X_{d,y}$ (where "d" is the day and "y" is the year) the ratio to trend method. This leads to a new series ($S_{d,y}$), known as series of index clarity:

$$S_t = S_{d,y} = x_{d,y} \quad / \quad H_0^d \ . \tag{1}$$

After this step a new rigid seasonality is updated, we can lift it with the use of periodic seasonal factors [14]. This treatment aims to create a new distribution without periodicity. Although this pre treatment tends to stationarize the time series, a test of Fisher shows that seasonality was not optimal. According Bourbonnais [14],

after using a ratio to trend method (H_o in this case) to correct rigid seasonalities, we can use a ratio to moving average. This second ratio can be applied when there is no analytical expression of the trend. In our case, we find that H_o led a new seasonality which is difficult to model. That's why we made a moving average ratio to overcome the seasonality. In the case of flexible seasonality, ie random in amplitude or period, the filtering techniques by successive moving averages are recommended.

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

In our case as 2m+1 = 365 days, we obtain m = 182. To complete the process, then we use the 365 seasonal factors (y_d). These are in fact coefficients which allow to overcome rigid seasonality by a moving average ratio described above. In order to not 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 equation 5). The transition coefficients (N = 18, the number of years of history) and the average coefficients of the regular 365 days are given by the equations 3 and 4. A new series seasonally adjusted that represents only the stochastic component of global radiation is given by the equation 6.

$$y_d = \frac{1}{N} \left(\sum_{y=1}^{N} y_{d,y} \right) . \tag{3}$$

$$\bar{y}_d = \frac{1}{365} \left(\sum_{y=1}^{365} y_d \right) . \tag{4}$$

$$y_d^* = y_d \quad / \quad \overline{y_d} \, . \tag{5}$$

$$S_{d,y}^{corr} = S_{d,y} \quad / \quad y_d^* \, . \tag{6}$$

In the next section, we present the RNA architecture used.

3 Neural network architecture and design

The search for the ideal network structure is a complex and crucial task. We have adopted a feed forward Multi-Layer Perceptron (MLP), which is the most commonly used in the renewable energy domain [1], [2]. In order to determine the best network configuration, we have tried to study all the parameters available in this network architecture. To perform this optimization, we have considered the practice hypothesis that parameters are orthogonal. We have optimized parameters by considering each other constants.

We used the Matlab® software and its neural network toolbox to implement our network. The Matlab® training and testing data sets were set respectively to 80% and

20%. As a result of this iterative process, the selected network (see figure 2) has three neuron layers: input, hidden and output layer. There was no significant difference in the use of 1, 2 and 3 hidden layer architectures. One hidden layer was used in order to minimize the complexity of the proposed ANN model. We tried several input layer configurations: 1, 2, 4, 8 and 15 variables. Best results were obtained with 8 inputs which received as input the endogenous entries S_{t-1},.. , S_{t-8} normalized on $\{0,1\}$. The same process was used to fix the number of hidden layer neurons. We found that 3 neurons were sufficient. Finally, we have one neuron on the output layer $\hat{S}_t$. Concerning the transfer functions the best results were obtained with the Gaussian (hidden layer) and linear (output layer) function. Regarding the training algorithm, many experiences have enabled us to choose the Levenberg–Marquardt optimization (second-order algorithm) with 5000 epochs and a decrease factor (mu) of 0.5. Default values have been use for all other parameters. The early stopping technique was set to the maximum validation failure (the parameter max_fail = 5).

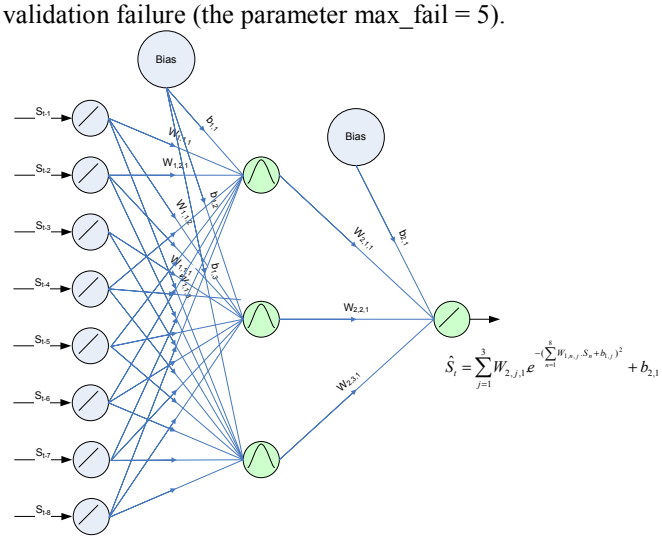

Fig. 2. Architecture of the optimized MLP.

The learning has concerned the years 1971 to 1987 and the performance function was mean square error – MSE.

4 Results and discussions

Figure 3 summarizes the protocol that has allowed us to conduct our experiences and validate our approach. A first treatment (step 1 of the figure 3) allows to clean the series of atypical points. We have replaced them by the average over the 19 years for the hours and the days corresponding to the problems.

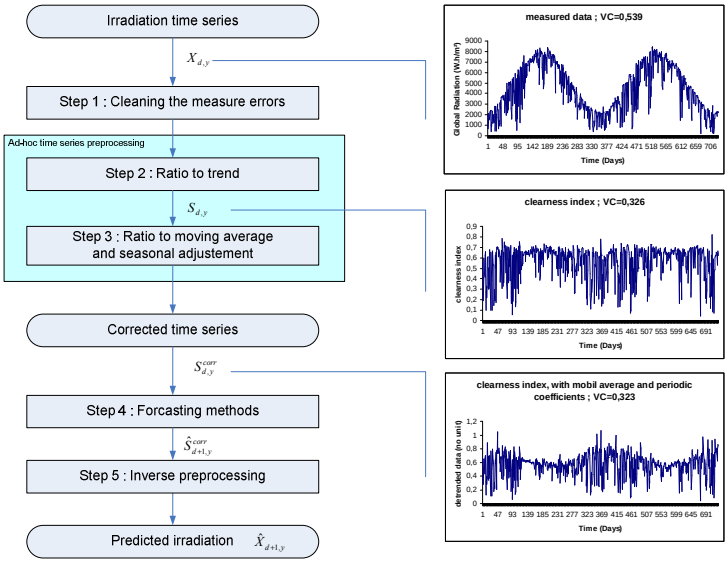

Fig. 3. Summarize of the protocol followed to obtain the predicted irradiation.

Steps 2 and 3 have been described in the previous section and lead to a series corrected. In the step 4 we compare classical forecasting methods outlined in the next section with our optimized MLP. Finally step 5 allows to reverse the preprocessing treatment and to obtain the prediction of global irradiation.

4.1 The classical forecasting methods

In order to measure the effectiveness of our approach, we have decided to compare it with the following classical forecasting methods.

The ARIMA techniques [15], [16] 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). After several experiments, we have obtained and decided to use an ARMA (2,2). Bayesian inference [17], [18] is another classical technique. In this method evidences or observations are used to update or to newly infer the probability that a hypothesis may be true. We have identified that the prediction was better if we had 50 classes and an order of 3. Some authors have tried to use so-called Markov process [19], specifically the Markov chains, which is a stochastic process. The mean idea of this technique is that 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. In our case, we obtained 50 for the dimension of the transition matrix (number of class) and an order of 3 for the chain (determination of the prediction lag). The k-nearest neighbors algorithm (k-NN) [20] 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 ones 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. In our study we choose a k equal to 10. The following section presents the results obtained.

4.2 Results

To determine whether our network was really interesting in terms of daily prediction of solar radiation, we compared its performances with the forecasting results obtained with a naive predictor (average over 18 years of the day considered), order 3 Markov chains, order 3 Bayesian inferences, an order 10 k-NN, an order 8 AR without preprocessing, an ARMA(2,2) with preprocessing. The predicted results for each combination were compared statistically using three parameters: the Root Mean Square Error (RMSE), the normalized RMSE (nRMSE), and the Mean Bias Error (MBE):

$$RMSE = \sqrt{\frac{1}{N}\sum_{i+1}^{N}(C_i - M_i)^2}\,. \tag{6}$$

$$nRMSE = \sqrt{\frac{1}{N}\sum_{i+1}^{N}(C_i - M_i)^2} \quad / \quad \sqrt{\frac{1}{N}\sum_{i=1}^{N}(M_i)^2}\,. \tag{7}$$

$$MBE = \frac{1}{N}\sum_{i+1}^{N}(C_i - M_i)\,. \tag{8}$$

Table 1 presents results we have obtained in the case of an annual error for daily prediction of global solar radiation.

Table 1. Annual error for all prediction methods, years 1988 and 1989, 8 simulations

Prediction methods	nRMSE	Confidence interval
Naïve predictor	26 %	0%
Markov Chain (order 3)	25,11 %	0%
Bayes (order 3)	25,16 %	0%
k-NN (order 10)	25,20 %	0%
AR(8) without preprocessing	21,18 %	0,2%
ARMA(2,2) with preprocessing	20,31 %	<0,1%
ANN[8,3,1] without preprocessing	20,97 %	0,15%
ANN[8,3,1] with preprocessing	20,17 %	0,1%

We highlight that the predictors other than ARMA and ANN give the same results, slightly higher than those obtained with a naive predictor. Even without preprocessing, ARMA and ANN are the best predictors. The preprocessing improves the quality of prediction and allows access to the 20% error. C_i and M_i are respectively the i_{th} calculated and measured values and N is the total number of

observation. Table 2 details in the MLP case, the annual prediction errors obtained for the years 1988 and 1989.

Table 2. Annual prediction error for the years 1988 and 1989 with our MLP

	Arithmetic Mean	**95% confidence interval**
nRMSE	20,2 %	0,1 %
RMSE	997,971	6,333
MBE	-104,239	28,794
R^2	0,801	0,002
Monthly average error	3,9 %	0,4 %

The confidence interval is calculated after 10 simulations. Given the small size of the confidence intervals, we can say that there are very few local minimums. The monthly average error represents the error for the value of the irradiation. We obtain that the combination of the prediction of global radiation received after 1 month is different from an average of 4% of the aggregate measured. The negative MBE means that we underestimate the solar potential on averaged over the year. Since we have an atypical day of low irradiation then there is a tendency to overestimate. The determination coefficient R^2 is greater than 0,8. Figure 4 shows the errors of prediction and distinguishes the seasons for the years 1988 and 1989. As we can see best results; i.e. less important error, 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.

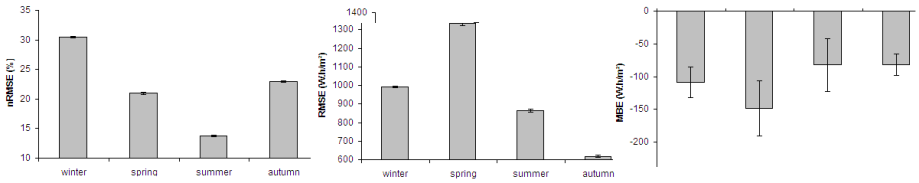

Fig. 4. Seasonal errors for the daily prediction of the years 1988 and 1989 (mean with 95% confidence interval).

There should be a compromise between RMSE and nRMSE. The nRMSE are useful for comparison and optimization. But for the interpretation of energy, we must look at the RMSE. The spring season is the most difficult to predict. The absolute error is consistent. However, we find that in summer the error does not exceed 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 will always have the same prediction error, whatever the season. Finally, Figure 5 compares the real data of solar radiation with the results obtained with our MLP with preprocessing. The error of prediction is also drawn. The increase in errors at the beginning of the cycle corresponds to the spring when the cloud disturbances are very important. We can see it is very difficult to predict the radiation cause of very noisy variable. Figure 6 shows the correlation between experimental and simulated global irradiation.

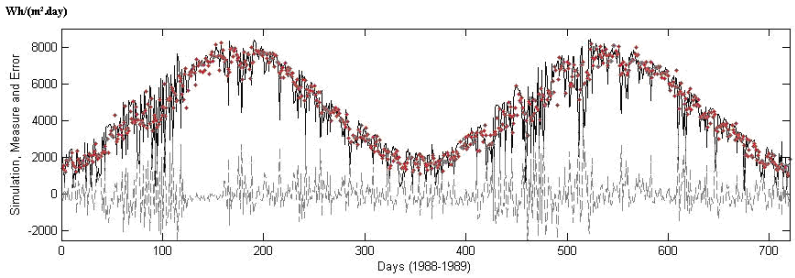

Fig. 5. Real data of solar radiation and results obtained with our MLP with preprocessing and error of prediction. Dashed line is the error; solid line is the real data of solar radiation and red points are the prediction.

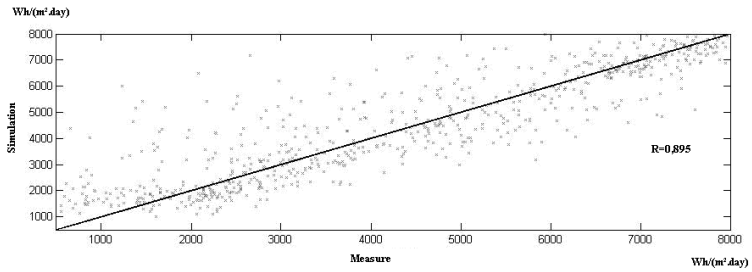

Fig. 6. Correlation between experimental and simulated global irradiation.

We systematically overestimate the days when the irradiation is minimal (winter). The points that lie at the very top of the line y = x shows that 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 procedure, and tend to decrease the robustness of the procedure. The next section concludes this paper and suggests prospects.

6 Conclusion and perspectives

This paper has developed an ANN predictor approach to determine global irradiation at daily horizon in order to help electrical managers. We have used an ad-hoc time series preprocessing 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. Seasonal RMSE are less than 998 Wh/m-2 (nRMSE < 21%). ANN processes presents a great interest compared to classical stochastic predictor like ARIMA. Moreover we found that our data preprocessing approach reduce significantly forecasting errors.

The next step of our work will be to validate our predictor on real photovoltaic system. It was recently installed in our laboratory and we are awaiting data. In the future, it seems important to study shorter time horizons. As 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. Thus others ANN architecture types have to be studied: recurrent ANNs, adaptative ANNs, etc. An ongoing study will be based on the implementation of exogenous variables on the input neurons like METARs data (pressure gradient, temperature, etc.). Determining the relevant variables could be done by the random probe method [21].

References

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

2 Mellit, A., Kalogirou, S.A., Hontoria, L., Shaari, S.: Artificial intelligence techniques for photovoltaic applications: A review. Progress in Energy and Combustion Science (2008)

3 Mubiru, J.: Predicting total solar irradiation values using artificial neural networks. Renewable Energy, Volume 33, Issue 10, Pages 2329-2332 (2008)

4 Mubiru, J., Banda, E.J.K.B.: Estimation of monthly average daily global solar irradiation using artificial neural networks. Solar Energy, Volume 82, Issue 2, Pages 181-187 (2008)

5 Kaligirou, S.A.: Artificial neural networks in renewable energy systems applications: a review. Renewable and sustainable energy review, 5, 373-401 (2001)

6 Hocaoglu, F.0., Gerek, O.N., Kurban M.: Hourly solar forecasting using optimal coefficient 2-D linear filter and feed-forward neural networks. Solar energy. (2008)

7 Zhang, G.P., Qi, M.: Neural network forecasting for seasonal and trend time series, European journal of operational research, 160, Pages 501-514 (2005)

8 Badescu, V.: Modelling Solar radiation at the earth surface. (2008)

9 Reindl DT, Beckman WA, Duffie JA. Evaluation of hourly tilted surface radiation models. Solar Energy; 45:9–17 (1990)

10 Liu BYH, Jordan RC. Daily insolation on surfaces tilted towards the equator. Trans SHRAE; 67:526–41 (1962).

11 Hay JE, Davies JA. Calculation of the solar radiation incident on an inclined surface. In: Proc. first Canadian solar radiation workshop, pages 59–72 (1980)

12 Perez R, Ineichen P, Seals R. Modelling daylight availability and irradiance components rom direct and global irradiance. Solar Energy ; 44:271–9 (1990)

13 Ineichen, P., Guisan, O., Perez, R.: Ground-reflected radiation and albedo. Solar Energy;44:207–14 (1990)

14 Bourbonnais, R. and Terraza, M.: Analyse des séries temporelles. (2004)

15 Hamilton, J. D.: Times series analysis. ISBN 0-691-04289-6 (1994)

16 Poggi, P., Muselli, M., Notton, G., Cristofari, C., Louche, A.: Forecasting and simulating wind speed in Corsica by using an autoregressive model. Energy Conversion and Management, Vol 44,Iss 20, pp. 3177-3196 (2003)

17 Diday, E., Lemaire, L., Pouget, J. Testu, F.: Éléments d'analyse de données, Dunod (1982)

18 Celeux, G., Nakache, J.P.: Analyse discriminante sur variables qualitatives (1994)

19 Muselli, M., Poggi, P., Notton, G. Louche. A.: 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 (2001)

20 Mohammed, S., Donald, H.B.: Simulating climate change scenarios using an improved K-nearest neighbor model (2006)

21 Dreyfus, G.: Random probes for variable selection, Multiple Simultaneous Hypothesis Testing, Paris (2007)