Optimization of an artificial neural network (ANN) dedicated to the daily global radiation and PV plant production forecasting using exogenous data

C.Voyant[1,2], M.Muselli[1,*], C.Paoli[1], ML.Nivet[1]

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

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

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 and a 1.175 kWp PV plant production. Different forecasting methods are compared: a naïve forecaster like persistence, an ANN with preprocessing using only endogenous inputs and an ANN with preprocessing using endogenous and exogenous inputs (like temperature, pressure, nebulosity, insulation, wind speed and direction etc.). The endogenous case is easily computed: the use of Partial Auto Correlation Factor (PACF) allows to optimize the number of lag time to consider. For the exogenous variables, we have applied a Pearson correlation coefficient method to optimize the number of considered input neurons. Although intuitively the use of meteorological data in the input layer of the MLP can only increase the quality of prediction, the obtained results are relatively mixed. The use of exogenous data generates a decrease of nRMSE between 0.5% and 1% for the both studied locations. The absolute error (RMSE) is decreased by 52 Wh/m²/day in the simple endogenous case and 335 Wh/m²/day for the persistence forecast. The results are similar to the concrete case of a tilted PV wall (1.175 kWp), endogenous and exogenous data ANN inputs allow decreasing the nRMSE by 1% on a 6 months-cloudy period for the DC power production (January-June). Moreover, the use of exogenous data shows an interest only in cloudy period (winter season). In summer, endogenous data as inputs on a preprocessed ANN are sufficient. By comparison to a naïve forecaster as persistence, an ANN with endo and exogenous data improves the DC electrical power energy prediction by 9% (nRMSE). Next step of this work will drive to shorter horizons; power predictor of meteorological data should have a greater impact.

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

*: corresponding author: M.MUSELLI, Tel: 33 4 95 52 41 41, email: marc.muselli@univ-corse.fr

1 Introduction

We present the results of the prediction of global radiation using Artificial Neural Networks (ANN) which are a popular artificial intelligence technique in the forecasting domain [1-4]. In previous studies [5,6], we have demonstrated that an optimized ANN with endogenous inputs can forecast the solar radiation with acceptable errors. In this present paper, our aim was to answer to the following question: does the use of exogenous variables (like temperature, humidity, wind speed and direction, pressure gradient etc.) increase the quality of prediction? We tried to answer to this question with sites located on the Island of Corsica (France). The island is characterized by a Mediterranean climate and a hilly terrain. The paper will be organized as follow: in section 2, the ad-hoc time series preprocessing description and in section 3, the results of the endogenous and exogenous optimization. The result of the predictions was shown in the Section 4 for horizontal radiation and on an 80° tilted PV plant in the section 5. The conclusions of this work will be presented in the last part.

2 Methodology

An Artificial Neural Network (ANN) is made up by simple processing units, the neurons, which are connected in a network by synaptic strengths, where the acquired knowledge is stored. In a Feed forward Neural Network (FNN) also known as a Multi Layer Perceptron (MLP), neurons are grouped in layers and only forward connections exist. A typical feed forward neural network consists of an input, a hidden and an output layers. Each component includes a neuron, weights and a transfer function. 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 by $y_i = f(\sum_{j=1}^{n} x_j w_{ij})$ where i is a neuron index in the hidden layer and j is an index of an input to the neural network. Training is the process of modifying the connection weights in some orderly fashion using a suitable learning method. According to previous experimentations [5], we utilized a stationarization method to increase the prediction quality witch consists to use the clear sky model for clear sky index. There are a lot of methods to determine this model. In our case, we have preferred to use the simplified "Solis clear sky" model [7] based on radiative transfer calculations and the Lambert-Beer relation. In this case, the clear sky global horizontal irradiance (H_{gh}) reaching the ground is defined by:

$$H_{gh,clearsky} = H_0.e^{-(\tau/\sin^b(h))}.\sin(h) \tag{1}$$

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 to determine the daily solar radiation modelling $H^d_{gh,clearsky}$. A series of tests (not presented in this paper) has allowed to validate the Solis model on horizontal global radiation. We obtain a relation of stationarization "equivalent" to Eq.(2) where X is the measure and S the new time series, (d is the day of the year y) :

$$S_{d,y} = \frac{X_{d,y}}{H^d_{gh,clearsky}} \tag{2}$$

This treatment aims to create a new distribution without periodicity. Further the new series generated is equivalent to a nebulosity series. Ideally, the values are fixed to 1 and decrease with the occurrence of a cloud. The problem of ANN optimization is composed by 3 independent subparts:

- choice of the hidden layer number and activation function,
- choice of the endogenous lag number,
- choice of the exogenous lag numbers for each parameter (daily pressure variation; wind direction; humidity; insulation; nebulosity; precipitation; mean pressure; min-max-mean temperatures, night temperature; wind speed).

For many years, we have studied the ANN optimization in daily horizon [6], this experience allows us to used previous results like the model of ANN for this kind of problem and the learning algorithm. The ANN used is so a FFN, we have simulated it with the Matlab software and the NN toolbox. These characteristics are: 1 hidden layer, the activation function are hyperbolic tangent (hidden) and linear (output), the learning algorithm is the Levenberg-Marquardt model (with max fail parameter equal to 5, μ decreases and increases respectively to 0.1 and 0.001, and goals equal to zero), the normalization is done between 0 and 1; the ratio of train, validation and test periods represent respectively 80%, 10% and 10%. We have learned the ANN during the 8 first years and we have computed the global solar radiation during the 2 last years.

3 Endogenous and exogenous parameters optimization

When we study the autocorrelation of a time series, the question is to know the term which differ significant of zero. In the case of FFN the parameter witch determines the best time lag of the endogenous net-input to use, is the PACF. In practice, the last correlation coefficient (R_k) different to zero (T-test methodology) induces k endogenous clear sky index like network inputs. The methodology chosen for the exogenous variable selection is similar to the previous one. We have used a correlation criterion (Pearson correlation). In our experimental sample size, the significant threshold for the T-test indicates a very low limit. Indeed, the limit is below 0.1 for the sample above 1000 (for the 0.05 critical alpha level). This methodology is not realistic in our case, the threshold for the coefficient R should be more important to select a limited number of exogenous inputs. Thereafter we have chosen a threshold R = 20%, only the higher correlations have been chosen. If we use the previous notation for the clear sky index (S_t) and for an exogenous variable (y_t representing nebulosity, temperature…), the correlation between the variables is:

$$R_k^y = \frac{\sum_{t=k+1}^{N}(S_t - \bar{S})(y_{t-k} - \bar{y})}{\sqrt{\sum_{t=k+1}^{N}(S_t - \bar{S})^2 \sum_{t=k+1}^{N}(y_{t=k+1} - \bar{y})^2}} \quad (3)$$

4 Results for horizontal radiation

4.1 Bastia

To determinate the input number of our FFN, it's necessary to used in a first step the PACF described in the last section. PACF allows quantifying the endogenous number of inputs. The purpose of this study is not to demonstrate the interest of use an ANN technique, but to search a mechanism for his optimisation. The first step was to stationnarize the global radiation (transition to clear sky index), then to find the inputs number (4 endogenous, 1 for humidity, 1 for insulation and 1 for nebulosity of previous days). We have verified that the use of each "exo-input" improves the quality of forecast. In details, the results show that the use of the insulation lag1, or humidity, or nebulosity, increases not only the mean of the error prediction but also the robustness of the methodology. The variance of the results is lower with the use of exo-inputs. The other interesting element is the equivalence of the insulation, humidity and nebulosity when they are used separately. For Bastia, the gap of the exo-input utilization is minimal, but is real. The nRMSE is 25.43% against 25.85%, the RMSE is reduce by 20 Wh/m² (1233 Wh/m²/day vs 1253 Wh/m²/day), and the mean absolute error by 51 Wh/m² (957 Wh/m²/day vs 1008 Wh/m²/day). A simplest forecaster as the persistence leads to an error nRMSE = 31.17% (MAE = 1081 Wh/m²/day and RMSE = 1569 Wh/m²/day). This site is well known to be very difficult to predict, this result was already checked in previous studies [8]. The second site mentioned in this article (Ajaccio) is reputed to be easier to predict.

4.2 Ajaccio

The optimization methodology is the same that used in the previous case. Applying the autocorrelation function to determine the "interesting" time lags, we obtain only two parameters (k=2) correlated with the clear sky index (S_{t+1} correlated with S_t and S_{t-1}). The experiment shows that we have to consider only the exo-inputs such as insulation and nebulosity at a time lag 1. The other variables are equivalent and don't contribute to add information for increase the quality of the prediction. Like Bastia case, only 3 neurons on hidden layer are necessary. Add more neurons complicate the system without obvious improvements. For Ajaccio, the exogenous methodology generates good results, the nRMSE is 21.54% against 22.50%, the RMSE is reduce by 52 Wh/m² (1087 Wh/m²/day vs 1139 Wh/m²/day), and the mean absolute error by 73 Wh/m² (839 Wh/m²/day vs 912 Wh/m²/day). The simplest forecaster is the persistence and the error generated is a nRMSE = 27.07% (MAE=971 Wh/m²/day and RMSE=1422 Wh/m²/day).

5 Experience on a tilted photovoltaic wall

A frontage PV system has been installed recently in our laboratory (Ajaccio-Vignola). It has a nominal power of 6.525 kW composed by respectively 1.8 kW and 4.725 kW amorphous and mono-crystal PV modules built in 6 independent power subsystems. PV power predictions from ANN methodology described in this paper have been computed from one of this whole PV plant on a frontage side exposed to the south (azimuth null) and tilted at 80°. The PV system is composed by 9 SUNTECH 175S-24Ac for a 1.175 kWp nominal power connected to a 1.85 kW SUNNY BOY SMA inverter for PV production on the grid. For the PV power calculation, we use in first approximation, a linear production based on a constant PV plant efficiency η_{PV} ~ 15%, with E_{PV} (Wh) = η_{PV} I_β S.; I_β is the daily global irradiation on the PV system (β = 80°), S is the usable surface of the PV system under consideration (S = 10.125 m²). To predict this energy, we have used 10 years of global horizontal irradiation available on the site of Ajaccio, like learning set. Classical models are used to compute tilted irradiation for an 80° angle. We used the Climed-2 [8] method to determine the diffuse fraction then the classical transformation to tilt the beam component, and the Klutcher equation to tilt the diffuse part [8]. The couple of exogenous data and the ANN estimation designed previously has been used to predict the energy produced by the PV plant of our laboratory. The distance is about 10 km from the place of obtaining the series of training. The exogenous data are, like described on the previous part, the lag1 of nebulosity and insulation. We have chosen to compare the ANN methodology with only endogenous data, endo- and exogenous data and persistence. The test of prediction is done with extreme conditions, because the period chosen for prediction is 6 months between January and June 2009 where the weather was very cloudy. The standard deviation of the DC PV electrical power measured is 299 Wh, while the means is equal to 666 Wh. The variation coefficient is close to 50%. Tables 1-3 show the results of the prediction tests for the three methodologies used (persistence, ANN with endo-inputs and ANN with endo & exo-inputs).

	Jan-June	Jan-feb	March-April	May-June
nRMSE	*0.331*	*0.375*	*0.361*	*0.164*
RMSE (Wh)	*216.35*	*272.66*	*247.70*	*90.11*
MBE (Wh)	*9.62*	*36.40*	*13.01*	*-13.43*
MAE (Wh)	*161.09*	*233.84*	*195.39*	*161.09*

Table 1: DC PV power prediction (80° tilted PV plant). ANN with endogenous and exogenous input.

	Jan-June	Jan-feb	March-April	May-June
nRMSE	*0.342*	*0.379*	*0.379*	*0.164*
RMSE (Wh)	*223.64*	*276.06*	*260.08*	*89.90*
MBE (Wh)	*-9.24*	*5.73*	*-4.69*	*-25.73*
MAE (Wh)	*172.07*	*242.16*	*213.54*	*172.07*

Table 2: DC PV power prediction (80° tilted PV plant). ANN with only endogenous input.

	Jan-June	Jan-feb	March-April	May-June
nRMSE	*0.423*	*0.470*	*0.471*	*0.188*
RMSE (Wh)	*276.59*	*342.36*	*323.20*	*102.81*
MBE (Wh)	*1.76*	*-1.77*	*2.51*	*3.16*
MAE (Wh)	*184.46*	*256.75*	*235.59*	*184.46*

Table 3: DC PV power prediction (80° tilted PV plant). Persistence case.

6 Conclusion

The results shown in this paper are interesting. On the site of Bastia, the use of the exogenous data on ANN inputs increases a little the prediction quality (only 0.5%). At Ajaccio, the nRMSE is improved by 1%. The last configuration based on the coupling of endogenous and exogenous data begin to be interesting for a power manager. The next studies will be based on the hourly time step, in order to verify the idea that the interest of exogenous data increase when the time step of time series decreases.

7 References

1. J. Faraway, C.Chatfield. Times series forecasting with neural networks: a case study, Research report 95-06 of the statistics group, University of Bath (1995)
2. P. Ineichen, “A broadband simplified version of the Solis clear sky model,” *Solar Energy*, vol. 82, 2008, pp. 758-
3. A. Mellit, S.A. Kalogirou, L. Hontoria, S. Shaari. Artificial intelligence techniques for sizing photovoltaic systems: A review. Renewable and Sustainable Energy Reviews 13-2 (2009), 406-419
4. J. Mubiru, E. Banda. Estimation of monthly average daily global solar irradiation using artificial neural networks. Solar Energy, 82-2 (2008) 181-187
5. C. Paoli, C. Voyant, M. Muselli, et M. Nivet, “Solar radiation forecasting using ad-hoc time series preprocessing and neural networks,” *0906.0311*, 2009.
6. C. Voyant, M. Muselli, C. Paoli, M.L. Nivet, et P. Poggi, “Predictability of PV power grid performance on insular sites without weather stations: use of artificial neural networks,” *0905.3569*, Mai. 2009.
7. R.W. Mueller, K.F. Dagestad, P. Ineichen, M. Schroedter-Homscheidt, S. Cros, D. Dumortier, R. Kuhlemann, J.A. Olseth, G. Piernavieja, C. Reise, L. Wald, et D. Heinemann, “Rethinking satellite-based solar irradiance modelling: The SOLIS clear-sky module,” *Remote Sensing of Environment*, vol. 91, Mai. 2004, pp. 160-174.
8. G. Notton, P. Poggi, et C. Cristofari, “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, 2006, pp. 1816-1829.