Conference Proceedings

Solar World Congress 2015
Daegu, Korea, 08 – 12 November 2015

Intermittence forecasting of the solar resource in Corsica

Pierrick Haurant, Auline Rodler, Ghjuvan-Antone Faggianelli , Marc Muselli, Philippe Poggi

University of Corsica, UMR SPE CNRS 6134, F-20000 Ajaccio, France

Abstract

Island electrical small-scale grids are sensitive to variations in power production. In case of important integration of solar power in the energy mix, the solar resource intermittency becomes a high risk to the grid's stability. That is why a good knowledge of the variations is the first step to the massive optimized PV integration in the energy mix. This paper focuses on the forecasting of the solar resource variability. First, the solar resource variability characterization will be presented. It is based on a typological classification method that relies on variability scales which enable to distinguish the different dynamics. Afterwards, the predictability of variations is studied through forecasting models as a simple persistence, the k-Nearest Neighbors and Artificial Neural Networks (ANN). In this regard, time series of intervals classified according to their dynamics of variations have been generated and the forecasting performance, for different horizons and with different models were compared.

Keywords: Solar resource, variations, classification, forecasting, k-Nearest Neighbors, Artificial Neural Networks

1. Introduction

Island electrical grids are typical: they are small-scale and sensitive to variations in power production. In case of important integration of solar power in the energy production mix, the solar resource intermittency becomes a high risk to the grid's stability. This is the reason why, in France, it is considered that no more than 30% of the production power must be supplied by power plants using intermittent renewable energy sources. In Corsica Island, this threshold has already been reached in 2012, as a large amount of PV plants have been connected to the grid. This threshold is a constraint to the development of renewable energies on the island. Therefore, the aim would be to overpass this threshold without destabilizing the electrical grid.

A good understanding of the solar resource in Corsica, its variability, its territorial mitigation and forecasting are important challenges for the electrical grid's manager to overpass this threshold. The optimization of the territorial mitigation may present an opportunity to the development of renewable energies and their integration to a small electrical grid which is not connected to the mainland power grid. We can anticipate that the smoothing of the electrical photovoltaic power production by the compensation of the territorial mitigation will be insufficient to stabilize the grid. Finally, it seems essential to compensate the solar variability using other means of production or storage solutions.

The level of solar variation has to be forecasted so that the electrical grid manager can anticipate and can choose a control strategy concerning all the means of power production. In this context, the solar resource intermittence characterization in Corsica, the use of the territorial mitigation and the forecasting of variations are challenges to the grid's manager. For this, a new method of typological classification of variations has been developed and will be introduced.

2. Typological classification of the radiation

Indicators such as the clearness index k_t, the air-mass-corrected clearness index k'_t or the clear-sky index k_{Cls} give information on the sky's clearness or cloudiness. A segmentation of these indexes domains into two

doi:10.18086/swc.2015.07.03 Available at http://proceedings.ises.org

or three intervals are the result of a first classification. In the literature, we can find numerous segmentations of daily k_t values that have succeeded to different classifications (De Miguel et al., 2001; Li and Lam, 2001; Mefti et al., 2008; Notton et al., 2004; Rigollier et al., 2004). This first classification method has the advantage to be easy but has the disadvantage to be static: fluctuations of irradiations and the cloudy dynamic are smoothen by a daily mean and cannot be considered. Consequently, a changeable weather day can have the same daily clearness index as an invariably cloudy one. Other typological methods can consider the clouds dynamics by completing the discrimination indexes of the different types of radiation. Muselli et al. (2000) have used a set of discriminating parameters derived from k_t hourly and daily values in order to characterize three types of days (clear sky, overcast sky, cloudy sky), using classification method of Ward. Daily dynamic of the sunshine has been considered by a parameter entitled "Integral of the squared second derivative of hourly clearness index profile". Other typological methods consider cloudy dynamic : the mean clearness index segmentation is completed by the segmentation of mathematical quantities that quantify the signals roughness such as the fractal dimensions (Maafi and Harrouni, 2003) or using Dirichtlet decompositions (Soubdhan et al., 2009) or even the wavelet decompositions (Woyte et al., 2007).

All these new methods present the limit of discriminating by days, while in a single day we can have different weather dynamics. That is why these classifications unable an energy manager to anticipate the variability of the production: the knowledge of the category of a day to come does not allow knowing precisely the irradiation profile. It seems valuable to consider days as successions of irradiation profiles that can be subject to a classification.

3. Methodology: Variation scales

Irradiance dynamic regimes distinguish themselves by the variations form observed that can be characterized by two criteria: their duration and amplitude. The aim of this work is to set up a classification method of these regimes depending on these characteristics.

The variation scales dt are defined as the time interval for which relative variations of solar irradiance $\mathrm{G(t)}$ are calculated. In this study, we are interested in the relative absolute variation $\mathrm{dG(t)}$ between two successive moments separated by a time dt, neglecting existing variations in the interval $[\mathrm{t} - \mathrm{dt}\,; \mathrm{t}]$ and following this relation:

$$\mathrm{dG(t)} = \left|\frac{\mathrm{G(t)} - \mathrm{G(t-dt)}}{\mathrm{G(t-dt)}}\right| \qquad \text{(eq. 1)}$$

3.1. Variations categories

Variation scales allow discriminating different variation dynamics. Indeed, variations which are too slow to be perceptible for a value of dt can be revealed for a superior value.

A typology can be built on scale variation following this procedure: considering a given variations scale, the issue will be to calculate and compare variations at a variability threshold noted $\mathrm{S_{var}}$. Variations higher than this threshold will integrate the equivalent class that could be described as the class of the «perceptible variations at this scale ». Then, for example, two different variations scales $\mathrm{dt_1}$ and $\mathrm{dt_2}$ lead to three classifications, from the close to zero variations set to high variations one. It is convenient to choose the scales separated by larger orders of magnitudes, in order to discriminate completely different variation profiles.

3.2. Intervals delimitation

In a majority of cases, even in a very dynamic cloudy regime, variations are only punctually above the fixed threshold. Then, the classification conditions generate a large number of intervals of different classes following themselves. Two thresholds have been introduced in this optic. The first threshold called intra-interval threshold and noted $\mathrm{S_{intra}}$, represents a maximal duration between two variations so that they can be grouped in the same interval. The other threshold is called inter-interval threshold and noted $\mathrm{S_{inter}}$. It is comparable to a minimal duration of intervals.

The use of these two thresholds gives birth to a sequential procedure of interval determination following three non-commutative steps:

1. The instability instants set t_{inst} are defined as a set of instants for which variations exceed the variability threshold S_{var};
2. All the instants between two variations separated by the inferior laps time of the intra-interval threshold S_{intra} are considered fluctuant : it is the grouping inter-interval step or homogenization step;
3. The intervals considered too short integrate adjacent intervals: it is the inter-interval grouping. If adjacent intervals to the studied one are of different classes, the inter-intervals grouping is done under a condition: if the interval is of maximal class, it will be concatenated to the adjacent one of nearest inferior class, else it will integrate automatically to the nearest interval of the superior class. These conditions are coherent with the electric network managers' point of view: it is convenient to outrank marginal events in order to consider in the worst case to limit the risks and to upgrade marginal events. However, in the first case where the over-classification is impossible, we consider that the oddness or the short variation duration takes over their intensity, the interval is downgraded.
4. The classification method and the variability interval definition and sunshine regimes need the determination of four specific parameters:
   - A variability threshold S_{var} above which we consider that there is instability; it has been fixed to 10 % in our study.
   - Scale variation values dt for the discrimination of the different sunshine regimes; it seems better to use different order of magnitudes in order to distinguish different dynamics. The magnitudes of a few seconds can find out strong and narrowed variations while those of a few minutes find strong variations and also slower and stronger ones. Thus, considering these observations, it seems appropriate to have two scales: $dt_1 = 1$ s and $dt_2 = 300$ s leading to a partition in three classes:
     - Class 0 (noted C_0): the variations of this class are perceptible neither for $dt = 1$ s, nor for $dt = 300$ s. There are no or very few variations ;
     - Class 1 (C_1): variations are perceptible only for $dt = 300$ s. These variations are slow and not very intense ;
     - Class 2 (C_2): variations are perceptible for $dt = 1$ s and $dt = 300$ s. They are narrowed and deep.
   - Intra- and inter- grouping interval thresholds (S_{intra} and S_{inter}), which we can evaluate the limits of the intervals. No objective criterion has been found for the thresholds determination. It is the reason why these thresholds are determined depending on the application we want to do with the intervals and the temporal resolution necessary to this application. The smaller the thresholds' values are, the smaller and more numerous will be the resulting intervals.

Figure 1 is an example of days separated into intervals ranked for $S_{var} = 0.1$; $dt_1 = 1$ s and $dt_2 = 300$ s and $S_{inter} = S_{intra} = 900$ s.

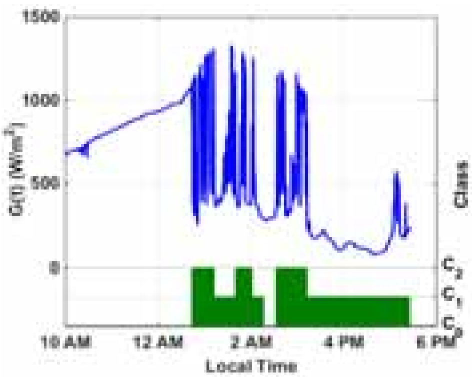

Figure 1 : An example of a daily irradiance curve and the variability classes associated

4. Forecasting of solar irradiance variation

In literature we can find a large amount of forecasting models to predict solar irradiance at different horizons (Kemmoku et al., 1999 ; Sözen et al., 2005 ; Mellit et al., 2005 ; Cao and Cao, 2006 ; Hocaoglu et al., 2008 ; Zervas et al., 2008 ; Chaabene et Ben Ammar, 2008 ; Paoli et al., 2010 ; Voyant et al., 2011 ; Marquez and Coimbra, 2011). These forecast concerns punctual irradiance values or the solar irradiance between two moments. For these two configurations no information is available on the short solar variations. Ideally, the aim would be to forecast for different small horizons in the range of the minute, on different lags at the same time. The shortest developed horizons until know are of 5 minutes on a single lag (Voyant, 2011). On the one hand, such a horizon does not seem sufficiently close to evaluate fast variations. On the other hand, a single lag does not allow to know a variation tendency (except if we refer us to past measures). Finally, this prediction does not leave the time to the grid manager to act on the system when a fluctuating regime is recognized. The ranks forecast seem interesting as it means that the variations of a future interval would be forecasted. The forecast methods which we have chosen are statistical based on:

- The k-Nearest Neighbor (k-NN): it is a technique following the philosophy that a succession of events of the same nature will lead always to a same consequence. Here k is the number of sequences. The principle of the k-NN consists to look for in a set of data identical events to the last observed. Then it is supposed that the future event will be the most often observed one after equivalent sequences observed in the historic.
- The Artificial Neural Network (ANN) is a learning algorithm inspired by biological neural networks. It represents a collection of neurons which are mathematical and computer representations of biological neurons. The mathematical representation is like an algebraic function which evaluates a weighted sum of the inputs matched to a bias, which is independent with the inputs.

Either the model has predicted the correct class or it has forecasted the wrong one, so that the performance of these models will be based on the percentage of times the class forecasted was the correct one. The models results will be compared to a primitive model, named the persistence. This model tends to reproduce the present event to the studied horizons.

4.*1 Set of data*

The classified irradiance data used in this study to test the different forecasting models were measured in 11 sites dispatched in Corsica Island (Fig. 2) during 2 years. The measurements are sampled at 1 s and synchronized. As a first step, the data collected in all sites have been concatenated in a same set in order to make a large sample as it is necessary for forecasting.

Figure 2 : Lucciana (42° 39’ 49” ; 9° 25’28” ; 60) - 2 Oletta (42° 39’ 36” ; 9° 19’ 45” ; 52) - 3 Calvi (42° 33’ 38” ; 8° 44’ 48”; 31) - 4 Corte (42° 18’ 04” ; 9° 09’ 57” ; 381) - 5 Piana (41° 16’ 03” ; 8° 41’ 37” 12) - 6 Cargèse (42° 08’ 40” ; 8° 35’ 58” ; 30) ; 7 Ghisonaccia (42° 03’ 54” ; 9° 22’ 14” ; 65) - 8 Ajaccio (41° 55’ 49” ; 8° 45’ 23” ; 2) - 9 Sainte Lucie (41° 41’ 59” ; 9° 20’ 12” ; 66) - 10 Propriano (41° 39’ 43” ; 8° 55’ 02” 17) - 11 Bonifacio (41° 22’ 17” ; 9° 12’ 10” ; 46)

The variations classes have been determined according to the method previously introduced, using a variation threshold of 10 % and variations scales of 60 s and 300 s. The predictions realized in this study are based on time series defined as a sequence of observations for regular time steps. These observations are done based on identical acquisition methods and data processing. In our case, the observed days need to be divided into constant time intervals for which we attribute fluctuation rank. Each measure has to be ranked using the method defined above. Then each time interval composed by ranked measures will get the mostly present fluctuation rank. The time series time step will be chosen based on the forecast demand in terms of temporal resolution and horizons.

80 % of the set of ranked intervals were used for the learning phase and 20 % were used for the assessments models performance.

4.2 Forecast with a persistence model

The persistence model is a simple model which considers that the event x_t at time t is repeated at time $t + h$, where h represents the horizon of the forecast:

$$\hat{x}_{t+h} = x_t \qquad \text{(eq. 2)}$$

This model has been implemented for ranked intervals of 15 min, 30 min and one hour. The model has been used to forecast horizons from 15min to maximum 6 hours. Figure 3 presents performance of the persistence model applied to these three types of intervals for the entire horizons predicted. First of all, we observe that whatever the size of the used ranked intervals, the predictions performance decrease when more lags are used. Above a number of lag the performance will not change much. For an interval of 15 min, we have a performance which changes from 83.3 % for lag 1, to 50.1 % for lag 6 to 33.3 % for lag 16. The performance does not change much and reach 32.9 % at lag 24. In the same way, the performance of the predictions for 30 min intervals changes from 72 % at lag 1 to 39.1 % for lag 6, decreasing then slightly. Finally, the correct forecast for a one hour interval will change between 66.9 % at lag 1 to 42.9 % at lag 6.

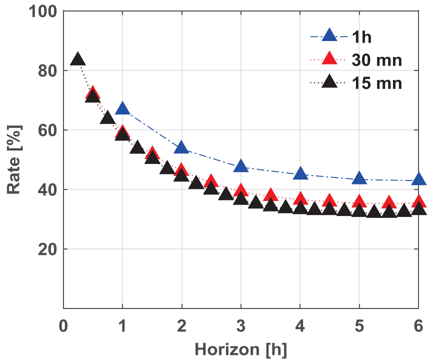

Figure 3 : Correct prediction rates obtained with a persistence as a function of horizons, using ranked intervals of 15 mn, 30 mn and 1 h.

For a same horizon we notice that longer (in terms of time) ranked intervals allow always better predictions for all the cases: for a horizon of 30 min the predictions with 30 min intervals (at lag 1) are better than with 15 min intervals (at lag 2). We have performance of 72 % for 30 min intervals and of 70.7 % for 15 min intervals. In the same way, one hour interval allow better predictions than with 30 and 15 min intervals, giving respectively 66.9 %, 58.8 % and 58 % of performance.

4.2. K-NN Forecast method

The k-NN method developed here is based on a historic of the successions of variation classes. The method consists in finding similar sequences of the k last past events in this historic. Then, the prediction at horizon h corresponds to the most frequently observed class that follows the sequences at lag h.

The optimization of the k-NN passes by the determination of the sequences size $k \in [1 ; 8]$ and of the time series time step used as input. As the tests realized with the persistence model the intervals of 15 min, 30 min and one hour have been tested. The first results have shown that it is convenient to choose a k-NN model with k = 4:

- For a time step of 15 minutes (Fig. 4), the existence of forecast impossible for $k \in [5 ; 8]$ puts into question their usability. Some sequences preceding the class to be predicted have not been found in the historic so no class forecast has been possible. Taking k = 4 allowed all the predictions and give better results for h ∈ [1 ; 4]. It gave close results to the best results of the other horizons.

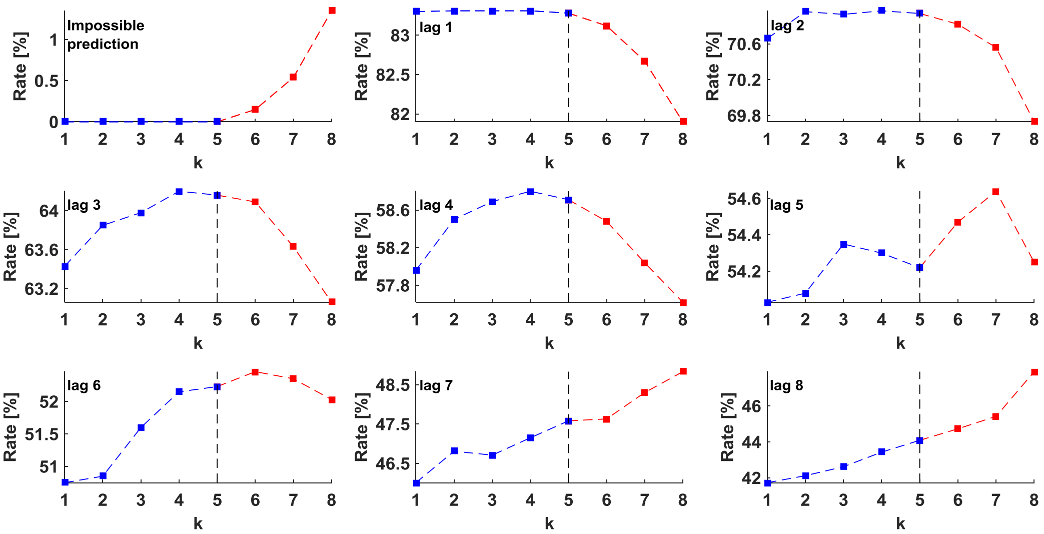

Figure 4 : Correct prediction rate as a function of the sequence size of the k-NN for different horizons, using 15 mn intervals

- For 30 minutes intervals (Fig. 5), the existence of impossible forecast for k ∈ [4 ; 8] show that these models are not reliable. The most performant model whatever the lag is 4-NN.

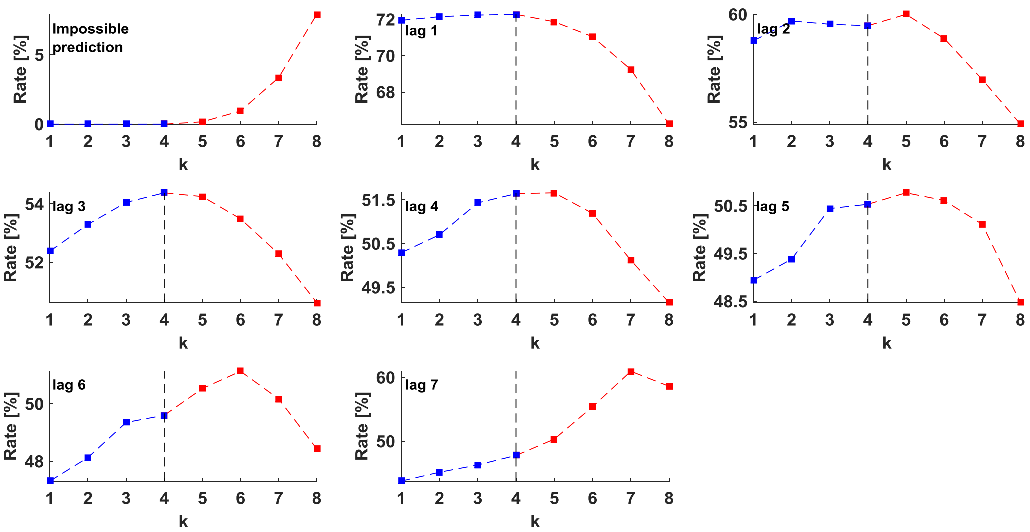

Figure 5 : Correct prediction rate as a function of the sequence size of the k-NN for different horizons, using 30 mn intervals

- Concerning the one hour interval (Fig. 6), impossible forecasts for k ∈ [5 ; 8] exist, so that we do not consider them. Then, the best forecasts are always reached with k = 3 or k = 4. We can notice that the performance of the 4-NN model are always higher than of the 3-NN for lags 1, 2 and 4 with small differences between 0.1-0.3 points. The 4-NN model has much better performance than model 3-NN for lags 3, 5 and 6 for differences between 0.5 and 7.2 points. In general, the 4-NN model presents the best results.

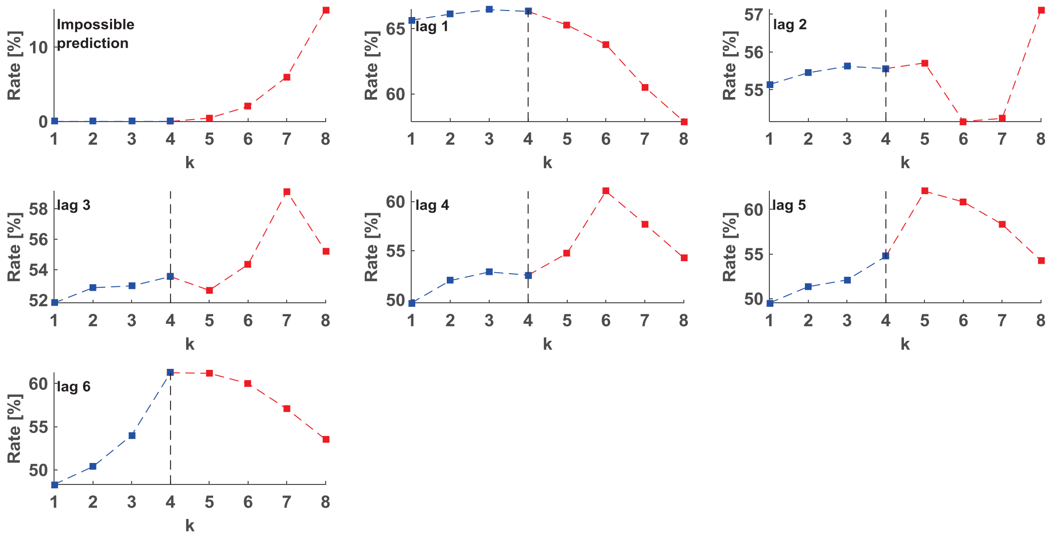

Figure 6: Correct prediction rate as a function of the sequence size of the k-NN for different horizons, using 1 h intervals

Finally, the performance of the 4-NN for the three ranked intervals studied are summed up on figure 7.

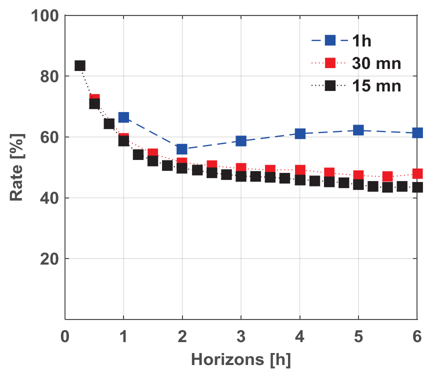

Figure 7 : Correct prediction rate obtained with a k-NN (k = 4) as a function of horizons, using ranked intervals of 15 mn, 30 mn and 1 h.

As the persistence, the longest ranked intervals give better results for distant horizons:

- The 15 minutes intervals allow 71.0 % of correct forecasts for 30 min horizons (at lag 2) while we have 72.3 % of good predictions, for the same horizon, with 30min intervals (à lag 1).
- The 15 minutes intervals give 58.8 % of correct predictions at one hour horizon (at lag 4), against 59.5 % with 30 minutes intervals (at lag 2) and 66.3 % with one hour intervals (at lag 1).

5. Artificial Neural Networks Forecast method

ANN have been applied to the time series representing the classes of variations. These classes are forecasted for different sizes of intervals and different horizons. In all the cases, only 6 lags have been forecasted in order to limit the complexity of the studied models.

The optimization of the ANN parameters is a major stake. Therefore, we have focused on the study of the number and the type of inputs as well as the number of hidden neurons N_c, conserving the rest of the ANN's structure (architecture, activation function, number of hidden layer, etc.). The optimization has been done by fixing one of these two parameters (i.e. the number and natures of the inputs or the number of hidden neurons) and varying the other one, following the method developed by Voyant et al. (2011). To start with, the inputs are defined and then the numbers of neurons of the unique layer hidden N_c are defined. The inputs have been optimized following two axes fixing $N_c = 1$:

1. The nature of the inputs, associating to the endogenous inputs exogenous inputs. The exogenous inputs used in this study are spatiotemporal indices: the measured instants of the events are detailed and are geographically located, since the historic is composed of measures done on 11 different sites on a period above one year and presenting discontinuities. Some successions of events can be linked to the site of measure and a temporality of the phenomenon can exist: some events successions are seasonal or daily. Three types of indexes have been studied independently each one from another :
   - A time index I_{tps} resuming the hour, the day, the month and the year of the event ;
   - A spatial index I_{geo} represented by the altitude of the measures site. This index is more representative than the geographic coordinates as the sites are very close to each other so that the coordinates would not be representative ;
   - A spatiotemporal index I_{Gcls} : the global radiation for clear sky daily conditions : Gcls, simulated by the ESRA model (Rigollier et al., 2000). The radiation depends on the position and the instant.
2. The number of inputs: we try to limit the number of inputs in order to reduce the complexity of the model. An iterative procedure has been set up to eliminate the unnecessary inputs. If we consider eight endogenous inputs and as many exogenous inputs a neural network is generated. The performance of the neural network is calculated and the weights between the inputs and the hidden units are examined. The input associated to the weakest weight is eliminated as it is considered unnecessary. This procedure is repeated until we have only one input. The model kept will be the one which maximizes the number of correct predictions.

We obtain from these optimizations different input choices according to the sizes of intervals. The model choice is difficult since their ranks are different from a lag to another: a model can be the best for prediction at a lag whereas it is the worst for predictions at another lag. A mean of the forecast on the six lags has been chosen to discriminate the different models. Only I_{tps} and I_{Gcls} indexes improve the ANN performance for 15 min classed intervals and lags higher than 1 (Fig. 8). The models using I_{tps} give the best results for lag 3 and 4 whereas I_{Gcls} increase the success rate for lag 2, lag 5 and lag 6, giving close results to the best ones for the other lags. These models need respectively 14 and 9 inputs. In this way, according to the Okham razor principle, it seems more interesting to keep the model with I_{Gcls} index for most cases. Only forecasts at lag 1 need endogenous data.

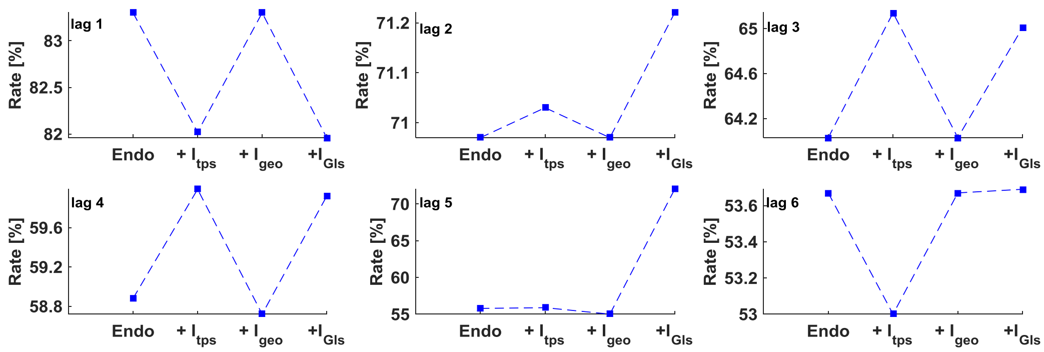

Figure 8: Correct prediction rate as a function of the type of ANN inputs for different horizons, using 15 mn intervals

For 30 minutes intervals, the indices tend to damage the models performance (Fig. 9). In this way the model using two endogenous inputs present the best performance for all the lags except for lag 3. This model will be put forward especially as it needs few inputs.

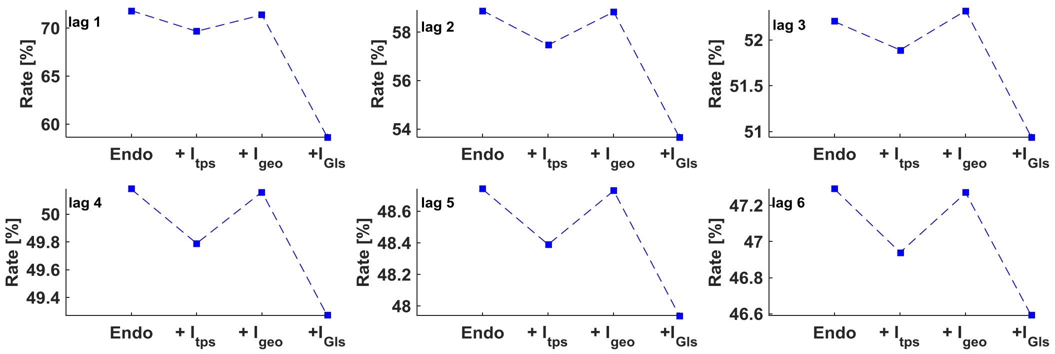

Figure 9: Correct prediction rate as a function of the type of ANN inputs for different horizons, using 30 mn intervals

Finally, concerning one hour intervals, the model using three endogenous inputs is the most promising. It allows better statistics for all the lags forecasted and needs few inputs.

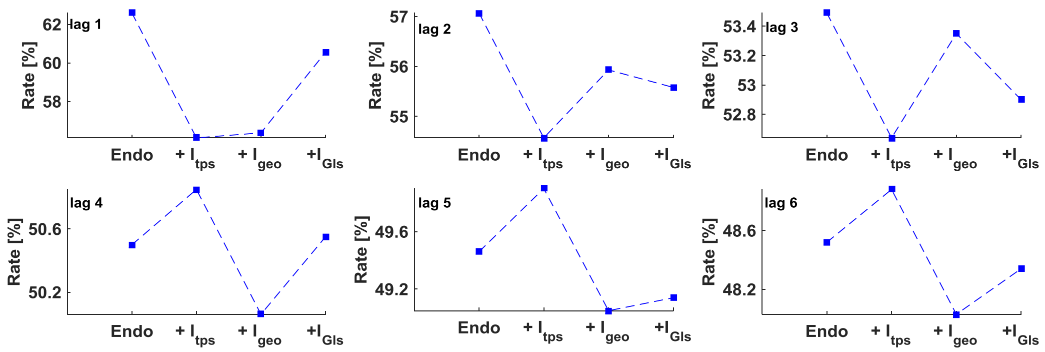

Figure 10: Correct prediction rate as a function of the type of ANN inputs for different horizons, using 1 h intervals

Concerning the number of hidden units in the unique layer, we must note that the number of parameters to evaluate for a ANN strongly increases with the number of hidden units: for a hidden layer with N_c neurons, N_e inputs and N_s outputs, the ANN will count $N_w = N_c(N_e + N_s)$ weights and N_c bias, a total $N_{tot} = N_c(N_e + N_s + 1)$ of parameters. It is important to limit at most N_c.

$N_c \in [1\,;8]$ have been tested for 15 min ranked intervals. We had slightly best performance for the three first lag with $N_c = 6$, $N_c = 7$ and $N_c = 5$ and the three following lags are a bit better forecasted with $N_c = 3$ and $N_c = 2$. But the small enhancements of the performance have to be relativized as the models are more complex so that $N_c = 1$ can be privileged since it is less complex and its performance are near the best more complex models.

For intervals of 30 minutes, 3 or 4 hidden neurons are needed to maximize the performance of the ANN at lag 1, lag 2 and lag 3 and that $N_c = 1$ shows the best results for predictions at lag 4, lag 5 and lag 6. $N_c = 1$ can be privileged: this model, which is the less complex, shows the best compromise between performance and complexity.

In the same way and for the same reasons, the ANN with $N_c = 1$ is preferred among $N_c \in [1\,;5]$ for one hour intervals.

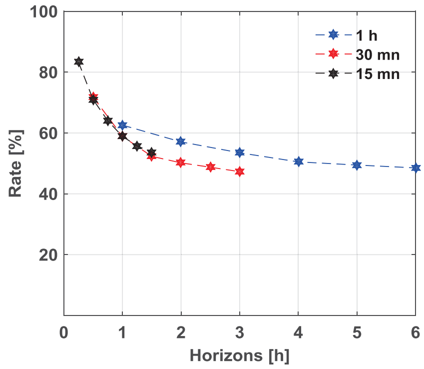

Figure 11 : Correct prediction rates obtained with an ANN as a function of horizons, using ranked intervals of 15 mn, 30 mn and 1 h.

6. Conclusions

The correct prediction rates for the studied horizons obtained with the optimized models are summed up in figure 12.

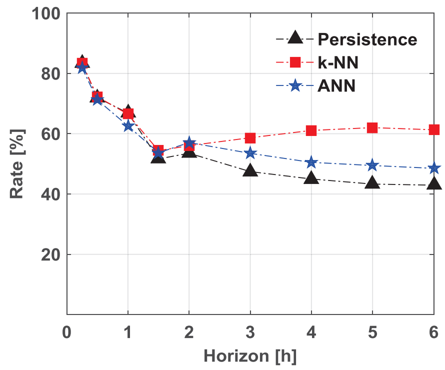

Figure 12: Correct prediction rate of the best model of each category as a function of horizons

We can see that whatever the model (persistence, k-NN or ANN), it is necessary to implement classified intervals of different durations as input to forecast variability classes at more or less far horizons:

1. for horizons above 1 hour, we can use classified intervals of 15 or 30 minutes. These inputs give close results whatever the model. However, it is interesting to input 15 minutes intervals for a best time resolution;

2. for horizons over 1 hour, it is necessary to input 1 hour intervals.

Besides, we can observe that for predictions at close horizon (between 15 minutes and 1 hour), the three studied models give quite similar results. For these horizons the persistence must be preferred to respect the Okham razor principle. For horizons between 1 hour and 2 hours, ANN gives best results and must be favored. Finally for horizons over 2 hours, the k-NN offers best performance and must be used. Thus an hybridation of these models should achieve better performance for prediction at all horizons.

In future works, Markow chains could be considered since it is possible to evaluate probability of transition from a class to another.

7. References

Cao, J., and Cao, S., 2006. Study of forecasting solar irradiance using neural networks with preprocessing sample data by wavelet analysis. Energy, 31(15), 3435–3445.

Chaabene, M., and Ben Ammar, M., 2008. Neuro-fuzzy dynamic model with Kalman filter to forecast irradiance and temperature for solar energy systems. Renewable Energy, 33(7), 1435–1443.

De Miguel, A., Bilbao, J., Aguiar, R., Kambezidis, H., and Negro, E., 2001. Di use solar irradiation model evaluation in the North Mediterranean Belt area. Solar Energy, 70(2), 143–153.

Hocaoglu, F. O., Gerek, m. N., and Kurban, M., 2008. Hourly solar radiation forecasting using optimal coefficient 2-D linear filters and feed-forward neural networks. Solar Energy, 82(8), 714–726.

Kemmoku, Y., Orita, S., Nakagawa, S., and Sakakibara, T., 1999. Daily insolation forecasting using a multi-stage neural network. Solar Energy, 66(3), 193–199.

Li, D., and Lam, J, 2001. An analysis of climatic parameters and sky condition classification. Building and Environment, 36(4), 435–445.

Maafi, A., and Harrouni, S., 2003. Preliminary results of the fractal classification of daily solar irradiances. Solar Energy, 75(1), 53–61.

Marquez, R., and Coimbra, C. F., 2011. Forecasting of global and direct solar irradiance using stochastic learning methods, ground experiments and the NWS database. Solar Energy, 85(5), 746–756.

Mefti, A., Adane, A., and Bouroubi, M., 2008. Satellite approach based on cloud cover classification: estimation of hourly global solar radiation from meteosat images. Energy Conversion and Management, 49(4), 652–659.

Mellit, A., Benghanem, M., Arab, A. H., and Guessoum, A., 2005. A simplified model for generating sequences of global solar radiation data for isolated sites : Using artificial neural network and a library of Markov transition matrices approach. Solar energy, 79(5), 469–482

Muselli, M., Poggi, P., Notton, G., and Louche, A., 2000. Classification of typical meteorological days from global irradiation records and comparison between two Mediterranean coastal sites in Corsica Island. Energy Conversion and Management, 41(10), 1043–1063.

Notton, G., Cristofari, C., Muselli, M., and Poggi, P., 2004. Calculation on an hourly basis of solar diffuse irradiations from global data for horizontal surfaces in Ajaccio. Energy Conversion and Management, 45(18-19), 2849–2866.

Paoli, C., Voyant, C., Muselli, M., et Nivet, M., 2010. Forecasting of preprocessed daily solar radiation time series using neural networks. Solar Energy, 84(12), 2146–2160.

Rigollier, C., Lefèvre, M., and Wald, L., 2004. The method Heliosat-2 for deriving shortwave solar radiation from satellite images. Solar Energy, 77(2), 159–169.

Soubdhan, T., Emilion, R., and Calif, R., 2009. Classification of daily solar radiation distributions using a mixture of Dirichlet distributions. Solar Energy, 83(7), 1056–1063.

Sözen, A., ArcaklIoglu, E., Özalp, M., and Caglar, N., 2005. Forecasting based on neural network approach of solar potential in Turkey. Renewable Energy, 30(7), 1075–1090.

Voyant, C., Muselli, M., Paoli, C., and Nivet, M, 2011. Optimization of an artificial neural network dedicated to the multivariate forecasting of daily global radiation. Energy, 36(1), 348–359.

Woyte, A., Belmans, R., and Nijs, J., 2007. Fluctuations in instantaneous clearness index: Analysis and statistics. Solar Energy, 81(2), 195–206.

Zervas, P., Sarimveis, H., Palyvos, J., and Markatos, N., 2008. Prediction of daily global solar irradiance on horizontal surfaces based on neural-network techniques. Renewable Energy, 33(8), 1796–1803.