Forecasting method for global radiation time series without training phase: comparison with other well-known prediction methodologies

Cyril Voyant[1,2*], Fabrice Motte[1], Alexis Fouilloy[1], Gilles Notton[1], Christophe Paoli[1,3], Marie-Laure Nivet[1]

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

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

[3] Galatasaray University, Computer Engineering Department Çırağan Cad. Ortaköy, Istanbul, Turkey

***Corresponding author: Cyril Voyant, phone: +33 4 95 29 36 66, fax:+ 33 4 95 29 37 97, email: cyrilvoyant@gmail.com**

Abstract.

Integration of unpredictable renewable energy sources into electrical networks intensifies the complexity of the grid management due to their intermittent and unforeseeable nature. Because of the strong increase of solar power generation the prediction of solar yields becomes more and more important. Electrical operators need an estimation of the future production at various horizons. For nowcasting and short term forecasting, the usual technics based on machine learning need large historical data sets of good quality during the training phase of the predictor. However data are not always available and induce an advanced maintenance of meteorological stations, making the method inapplicable for poor instrumented or isolated sites. In this work, we propose intuitive methodologies and a model based on the Kalman filter use (also known as linear quadratic estimation), able to predict a global radiation time series without the need of historical data. The accuracy of these methods is compared to other classical data driven methods, for different horizons of prediction and time steps. The proposed approach shows interesting capabilities allowing to improve quasi-systematically the prediction. For one to ten hour horizons Kalman model performances are competitive in comparison to more sophisticated models such as ANN which require both consistent historical data sets and computational resources.

Keywords: Time Series forecasting, model free forecasting, Artificial Neural Networks, ARMA, Kalman filter

1. Introduction

An electrical operator shall ensure a precise balance between the electricity production and consumption at any moment. He has often some difficulties to maintain this balance with conventional and controllable energy production system, mainly in small or no interconnected electrical grid (as island ones [1]). The reliability of the electrical system then becomes dependent on the ability of the system to accommodate expected and unexpected changes (in production and consumption) and disturbances while maintaining quality and continuity of service to the customers [2–4]. Then, the energy supplier must manage his system with various temporal horizons (Figure 1) [5].

Figure 1: Prediction scale for energy management in an electrical network.

The integration of non-predictable renewable energy (NPRE) as wind and solar, into an electrical network intensifies the complexity of the grid management and of the continuity of the production/consumption balance due to their intermittent and unpredictable nature [6,7]. The intermittence and the non-controllable characteristics of solar production brings a number of other problems such as: voltages fluctuations, local power quality and stability issues. Forecasting the output power of solar systems is required for a good operating of the power grid or for the optimal management of the energy fluxes occurring into the solar system [8]. It is necessary for estimating the reserves, for scheduling the power system, for congestion management, for optimally managing the storage with the stochastic production and for trading in the electricity market.

Due to the strong increase of solar power generation the prediction of solar yields becomes more and more important. Electrical operators need an estimation of the future production at various time horizons from several minutes to several days, for preparing the production means or planning the start-up of power plants. Various approaches exist to forecast solar irradiance depending on the target forecasted time [9,10]. The literature classifies usually these methods in two main classes: - extrapolation and statistical processes using satellite images or measurements on the ground level dedicated for the “now casting’’ (0–3 h) and the short-term forecasting (3–6 h); -numerical weather prediction (NWP) able to forecasts up to two days ahead and beyond [11].

For nowcasting and short term forecasting, the usual technics are based on machine learning predictions such as Artificial Neural Network (ANN) [10], Support Vector Machines (SVM) [1], AutoRegressive–Moving-Average (ARMA) models [12], etc. A significant inconvenience of these methods is related to the large historic data set required during the training phase of the predictors; thus, in this work, we propose a simple methodology able to predict a global radiation time series without the need of historical data, making the method easily applicable for poor instrumented areas. We suggest to call these intuitive methods in the following “training-less” methods. The accuracy of these methods will be compared against other classical prediction methods, taking into account the time horizon of the prediction and the time granularity of the predicted data.

The rest of the paper is structured as follow: firstly we describe the data used in this work, then the models we use in our irradiance forecasting benchmark and finally the results part shows a comparison of training-less methodologies against data-driven reference models.

2. The Data

The solar data used in the models are global horizontal irradiations (GHI) measured at the meteorological station of Ajaccio (Corsica Island, France, 41°55 N, 8°44 E, 4m asl). This station is equipped with pyranometers (CM 11 Kipp & Zonen) and standard meteorological sensors (pressure, temperature, etc.). It is located near the Mediterranean Sea (100 m) and nearby mountains (1000 m altitude at 40 km from the site). This specific geographical configuration and the island context make the nebulosity difficult to forecast. The Mediterranean climate is characterized by hot summers with abundant sunshine and mild, dry and clear winters. Two data sets have been used, on an hourly basis from 1998 to 2009 (11 complete years) and on a minute basis from 2012 to 2013 (2 complete years). As for all experimental acquisitions, missing values were observed in our data set (less than 2% of the data).

3. The models

This section describes the models used in our irradiance forecasting benchmark. We first describe two data driven methodologies, a linear model and a non-linear model. Then we describe training-less models.

3.1. Time series definition and ideal prediction

In order to introduce the prediction from time series approach we can use the Cartier Perrin theorem [13]. With this theorem a time series is described in a multiplicative mode with:

$$x(t+1) = x_{trend}(t+1).x_{fluc}(t+1) \tag{1}$$

$x_{trend}(t+1)$ is the trend and $x_{fluc}(t+1)$ is a quick fluctuation, in our case x_{trend} is the solar irradiation by clear sky and x_{fluc} the variation of the "real" solar radiation around the clear sky profile called the stochastic part [14]. The very "nature" of those quick fluctuations is left unknown and nothing prevents us from assuming that x_{fluc} is random and/or fractal [15]. Based on this kind of time series definition, an ideal prediction can be obtained from a trend estimation computed with a cubic spline data interpolation based on a tridiagonal linear system. It is solved for the information needed to describe the coefficients of the various cubic polynomials which make up the interpolating spline [16]. Considering the Cartier Perrin theorem the perfect predictor describes the trend while the quick fluctuations are not modelled and the related error is the lowest error than a predictor can generate.

3.2. Data driven methods

Data driven methodologies are technics generally based on machine learning and need large historical data sets of good quality during the training phase of the predictor. We present here the linear model and the nonlinear model which are the most popular and efficient from our point of view [10].

3.2.1. Linear model: AutoRegressive process (AR)

In an AR model, the future value of a variable namely $\widehat{GHI}(t+h)$ is assumed to be a linear combination of several past observations $(t-i)$:

$$\widehat{GHI}_{AR}(t+h)=\phi_0+\sum_{i=0}^{p}\phi_{i+1}GHI(t-i)+\epsilon_t \qquad (2)$$

Where ϵ_t is a white noise with variance $\sigma^2,$ the model's parameters are

, p is called the autoregressive order of the model. One key challenge in the AR model conception is to determine the optimal model order: AutoCorrelation Function (ACF) and Partial AutoCorrelation Function (PACF) are proposed to be used for selecting the best orders [12,17]. In this study, the complexity of the model depends on the autoregressive order p which is optimized using the auto-mutual information factor (see section 3-4 for details).

3.2.2. Non-linear model: neural network model (ANN)

ANN is the predominant method in the domain of time series forecasting. Indeed, the availability of meteorological historical data databases and the fact that ANN are data driven approaches capable of performing a non-linear mapping between sets of input and output variables make this modelling tool very attractive. An ANN with *d* inputs, *m* hidden neurons and a single linear output unit defines a non-linear parameterized mapping from an input vector x to an output *y* given by the relationship exposed in equation (3) [18–21]:

$$y=y(\mathbf{x};\mathbf{w})=\sum_{j=1}^{m}w_j\,f\left(\sum_{i=1}^{d}w_{ji}\,x_i\right) \qquad (3)$$

Each of the *m* hidden units are related to the tangent hyperbolic function $f(x)={(e^x-e^{-x})}/{(e^x+e^{-x})}$.

The parameter vector $\mathbf{w}=(\{w_j\},\{w_{ji}\})$ governs the non-linear mapping and is estimated during a phase called training or learning phase. During this phase, the ANN is trained using the data set $\mathcal{D}$ that contains a set of n input and output examples. The second phase, called the generalization phase, consists in evaluating, on the test data set $\mathcal{D}_*$, the ability of the ANN to generalize, that is to say, to give correct outputs when it is confronted with examples that were not seen during the training phase. For our application, the relationship between the output $\widehat{GHI}(t+h)$ and the inputs $\{GHI(t), \mathbf{GHI}(t-1), \cdots, \mathbf{GHI}(t-p)\}$ has the form given in the equation:

$$\widehat{GHI}(t+h) = \sum_{j=1}^{m} w_j f\left(\sum_{i=0}^{p} w_{ji}\, GHI(t-i)\right) \tag{4}$$

As shown by Eq (4), the ANN model is equivalent to a nonlinear autoregressive (AR) model for time series forecasting problems. In a similar manner as for the AR model, the number of past input values p is calculated with the auto mutual information factor (see section 3-4 for details). Careful attention must be put on the building of the model, as a too complex ANN structure will easily overfit the training data. The ANN complexity is in relation with the number of hidden units or conversely the dimension of the vector $\mathbf{w}$. In the present study, the ANN model has been computed with the Matlab© software and its Neural Network toolbox. The Levenberg-Marquardt learning algorithm (approximation to the Newton's method) with a max fail parameter before stopping training equal to 3 was used to estimate the ANN model's parameters. The max fail parameter corresponds to a regularization tool limiting the learning steps after a characteristic number of predictions failures. Consequently it is a means to control the model complexity [14].

3.2.3. Preprocessing steps for AR and ANN models

All the machine learning technics need three steps of preprocessing: the cleaning and filtering; the stationary process and the inputs selection. Concerning the global radiation forecasting, it is a common practice to filter out the data in order to remove night hours in view to compare the predicting methods only on periods between sunrise and sunset. This choice is justified because during these removed periods there is obviously no significant solar radiation to generate electricity. To achieve this preprocessing operation we have chosen to apply a selection criterion based on the solar zenith angle (SZA): solar radiation data for which the solar zenith angle is greater than 80° have been removed [1]. This transformation is equivalent to a filtering related to the solar elevation angle lower than 20°. In addition, this filtering process allows to discard data with less precision as measurement uncertainties associated to pyranometers are typically much higher than ± 3.0% for SZA > 80°. Note that for the sunrise and sunset, the prediction is also very difficult (mainly for the mountainous area due to mask effects) owing to the geographic shield. Make the time series stationary is a data transformation which is often needed before adjust parameters of AR or ANN models. Stationarity means that the statistical characteristics of the time series such as the mean and the autocorrelation structure are constant over time [22]. In this survey, as the initial solar radiation series is not stationary, we used a clear sky model in view to obtain a stationary solar series without daily and yearly periodicities . More precisely, we obtained a de-seasoned series $\{k^*\}$, the so-called clear sky index series, by applying the following data transformation:

$$k^* = \frac{GHI}{GHI_{clsk}} \qquad (5)$$

GHI is the measured solar global irradiation and GHI_{clsk} is the solar global irradiation in clear sky conditions. The global irradiance is then decomposed into a deterministic clear sky component and a stochastic cloud cover component. With this methodology, the models designed in this work predict the stochastic part of the global radiation leaving the deterministic part modelled by the clear sky model. One may notice however that this transformation is not optimal (i.e. the time series of clear sky index may still exhibit some heteroscedasticity, variances are varying with the effects being modelled) and one has to apply more data transformations like differencing to remove the trend and/or stabilize the variance. It must be noted that the same type of pre-processing will be applied to other machine learning methods. Indeed, in practice, it is usually admitted that data normalization (as it is the case here) facilitates the learning process of these methods.

GHI_{clsk} is computed using the simplified "Solis clear sky" model based on radiative transfer calculations and the Lambert-Beer relation [23,24]. The expression of the atmospheric transmittance is valid with polychromatic radiations, however when dealing with global radiation, the Lambert-Beer relation is only an approximation because of the back scattered effects. This model remains a good fitting function of the global horizontal radiation. The use of this model requires fitting parameter (g), corrected extraterrestrial radiation (I_0), solar elevation (h) and total atmospheric optical depth (τ). All these previous parameters depend on the site according to various zones and climates. In this case, the clear sky global horizontal irradiance (GHI_{clsk}) reaching the ground is defined by Eq (6).

$$GHI_{CLSK} = I_0 \exp\left(-\left(\frac{\tau}{sin^g(h)}\right).\sin(h)\right)$$

(6)

One step was common to all models is the number of endogenous inputs to consider, this step is often called “feature selection”. We have chosen to apply the same methodology for all numerical experiments by using the auto-information of the signal. This parameter measures the reduction of uncertainty in $x(t)$ after observing $x(t-i)$ ($i =$ 0,1…N; N number of observations). The Mutual Information (MI) measures non-monotonic and other more complicated relationships between variables [5]. It can be expressed as a combination of marginal and conditional entropies

(respectively) and) as described in the Eq (7).

$$[\![MI(k]\!]^*(t), k^*(t-i)) = H(k^*(t)) - H(k^*(t)|k^*(t-i)) \qquad (7)$$

This quantity should be understood as the amount of randomness of the random variable $k^*(t)$ given that the value is known. The maximum of lagged inputs to consider (i.e. number of inputs of the ANN and AR) lag corresponds to the first minimum of the auto-mutual information [25]. For example, if the first min corresponds to the 10th time lag, the ANN will be constructed with 10 inputs.

3.3. Training-less methods

In the following we use the equation 1 to describe several models called order 0, 1, 2 and 3. These models are construct considering that the best prediction is linked to the trend estimation and that in this case the lowest error occurred is related to the noise or fluctuation part of the time series. After presenting these four models we propose two variants concerning the order 1: one uses a Kalman filter, the other one uses model output statistic (MOS). These variants are not used for the order 0 because it is a very naïve estimation and are not used for the other orders because the models will become so complicated and the goal of this study is to present some easy to implement approaches.

3.3.1. Order 0: the persistence

The persistence model [26] can be consider as an order 0 model (figure 2). This naïve method is the most cost-effective forecasting model, it provides a reference against which more sophisticated models can be compared. Using the naïve approach produced forecasts are equal to the last observed value.

$$\widehat{GHI}_{order_0}(t+h) = GHI(t) \qquad (8)$$

It simply states that future GHI values will be equal to observed GHI at time t (i.e. the atmospheric conditions and solar irradiation remain unchanged between current time t and future time *t*+*h*).

3.3.2. Order 1 : scaled persistence

In equation 1, considering that $x_{fluc}(t) = x_{fluc}(t+1)$ we define a scaled persistence or an order_1 training-free model defined by:

$$\hat{x}(t+1) = x(t).\frac{x_{trend}(t+1)}{x_{trend}(t)} \qquad (9)$$

Using the clear sky model and applying the previous equation to solar irradiation estimation this order_1 equation training-free can be replaced by:

$$\widehat{GHI}_{order_1}(t+1) = GHI(t).\frac{GHI_{clsk}(t+1)}{GHI_{clsk}(t)} \qquad (10)$$

Note that another model using an additive mode exists:

$$\widehat{GHI}(t+1) = GHI(t) - (GHI_{clsk}(t) - GHI_{clsk}(t+1)) \qquad (11)$$

The model described by equation (11) will be not used in this paper because preliminary tests showed that the multiplicative model (equation (10)) is more efficient.

3.3.3.Order 2 and order 3

Taylor series is a series expansion of a function about a point. With this tools, it is possible to construct an order_2 model without training phase considering that the predicted value of solar irradiation is the intersection of the tangent at $x_{fluc}(t)$ concerning the point t_0 and the curve $t = t_0 + 1$. To better understand the different orders of prediction Fig. 2. illustrates the methodologies.

Figure 2. Illustration of the prediction method without training phase using a model of order 0, 1, 2 and 3.

For t_0 we obtain:

$$x_{fluc}(t)\big|_{t_0} = x_{fluc}(t_0) + \frac{dx_{fluc}(t_0)}{dt(t-t_0)} \qquad (12)$$

The intersection with the line $t = t_0 + 1$ induces that:

$$\tilde{x}_{fluc}(t_0+1) = x_{fluc}(t_0) + \frac{dx_{fluc}(t_0)}{dt}(t_0 + 1 - t_0) = x_{fluc}(t_0) + \frac{dx_{fluc}(t_0)}{dt} \qquad (13)$$

Another form is:

$$\tilde{x}_{fluc}(t+1) = x_{fluc}(t) + \frac{x_{fluc}(t) - x_{fluc}(t-1)}{t-t+1} = 2x_{fluc}(t) - x_{fluc}(t-1) \qquad (14)$$

Replacing x_{fluc} by GHI/GHI_{clsk} in equation (14) we obtain:

$$\widehat{GHI}_{order_2}(t+1) = 2GHI(t)\frac{GHI_{clsk}(t+1)}{GHI_{clsk}(t)} - GHI(t-1)\frac{GHI_{clsk}(t+1)}{GHI_{clsk}(t-1)}$$

(15)

It is possible to extend this methodology using an order 2 Taylor development (approximation of an unknown and noisy function), then:

$$x_{fluc}(t)\Big|_{t_0} = x_{fluc}(t_0) + \frac{dx_{fluc}(t_0)}{dt}(t-t_0) + \frac{1}{2}\frac{d^2x_{fluc}(t_0)}{dt^2}(t-t_0)^2 \qquad (16)$$

The intersection with the curve $t = t_0 + 1$, gives:

$$\hat{x}_{fluc}(t_0+1) = x_{fluc}(t_0) + \frac{dx_{fluc}(t_0)}{dt} + \frac{1}{2}\frac{d^2x_{fluc}(t_0)}{dt^2} \qquad (17)$$

For all t0, we have:

$$\hat{x}_{fluc}(t+1) = x_{fluc}(t) + \frac{x_{fluc}(t) - x_{fluc}(t-1)}{1} + \frac{1}{2}\frac{d\,[\![(x]\!]_{fluc}(t) - x_{fluc}(t-1)\big)}{dt} = 2x_{fluc}(t) - x_{fluc}(t-1) + \cdots\frac{1}{2}\frac{dx_{fl}}{d}$$

(18)

Replacing x$_{fluc}$ by GHI/GHI$_{clsk}$ in equation (18) we obtain:

$$\widehat{GHI}_{order_3}(t+1) = \frac{5}{2}GHI(t)\frac{GHI_{clsk}(t+1)}{GHI_{clsk}(t)} - 2GHI(t-1)\frac{GHI_{clsk}(t+1)}{GHI_{clsk}(t-1)} + \frac{1}{2}GHI(t-2)\frac{GHI_{clsk}(t+1)}{GHI_{clsk}(t-2)}$$

(19)

3.3.4. Add-on methods: Kalman filter and model output statistic (MOS)

In this paper, in order to improve the reliability of all the predictions we propose to use the recursive Kalman method and a classical model output statistic.

3.3.4.1. Kalman filter

. We will apply this method only on the order 1 but this approach is also usable with order 2 and 3. Indeed for the order 0, there is not great interest to use it because of the simplicity of prediction and committed error. For the other orders, formalism becomes too complicated and so far from the goal of the paper, ie provide a simple and easy to use method. Kalman method is an add-on able to complete the prediction from an ad-hoc filtering. The Kalman filter is a recursive estimator. This means that only the estimated state from the previous time step and the current measurement are needed to compute the estimate for the current state. The Kalman filter can be written as a single equation, however it is most often constructed with two distinct phases: prediction and update. The prediction phase uses the state estimated from the previous timestep (t-1) to produce an estimation of the state at the current timestep (t). This predicted state estimated is also known as the *a priori* state estimated because, although it is an estimation of the state at the current timestep, it does not include observation information from the current timestep. In the update phase, the current *a priori* prediction is combined with current observation information to refine the state estimated. This improved estimation is termed the *a posteriori* state estimated. With this methodology, the forecasting algorithm becomes [27]:

$$GHI(t+1) = A(t).GHI(t) + \omega(t) \tag{20}$$

with ω a multivariate normal distribution with covariance Q (=ℵ (0,Q)) and $A(t) = \frac{GHI_{clsk}(t+1)}{GHI_{clsk}(t)}$. Note that is the case of orders 2 or 3, it is possible to determinate A(t) from equation 15 and 19.

At time t, an observation (or measurement) z(t) of the true state GHI(t) is made according to:

$$z(t) = H(t).GHI(t) + v(t) \tag{21}$$

where *v* is the observation noise which is assumed to be zero mean Gaussian white noise with covariance *R* (=ℵ (0,*R*)) [28]. The initial state, and the noise vectors at each step are all assumed to be mutually independent. In our problem, the prediction defining the state vector is defined by:

$$\widehat{GHI}(t|t-1) = \widehat{GHI}(t-1|t-1).A(t-1) \tag{22}$$

$\widehat{GHI}(t|t)$ is the predicted value of GHI given the measured value of GHI at time (t). It is in fact, an *a posteriori* state estimated at time *t* given observations up to and including time *t.*

Then the *a posteriori* error covariance matrix P (a measure of the estimated accuracy of the state estimated) is calculated [29], in this study a windows width (*WW*) is used to compute the matrix P, in order to take into account the seasonality of the signal.

$$P(t|t-1) = A(t-1).P(t-1|t-1).A(t-1)^T + Q \tag{23}$$

From this last equation, we define the filter gain *K* which is then computed:

$$K(t) = P(t|t-1).H(t).(H(t).P(t|t-1) + R)^{-1} \tag{24}$$

A correction factor is then introduced and defined by [30]:

$$\widehat{GHI}(t|t) = \widehat{GHI}(t|t-1) + K(t).\left(z(t) - H(t).\widehat{GHI}(t|t)\right) \tag{25}$$

$$P(t|t) = P(t|t-1) - K(t).H(t).P(t|t-1) \tag{26}$$

The prediction for the horizon 1 becomes:

$$\widehat{GHI}_{kalman}(t+1|t) = A(t).\widehat{GHI}(t|t) = \frac{GHI_{clsk}(t+1)}{GHI_{clsk}(t)}.\widehat{GHI}(t|t) \tag{27}$$

Note that the approach is easily generalizable for other horizons *h* ($\widehat{GHI}_{kalman}(t+h|t)$).

3.3.4.2. Model output statistic (MOS)

The last tested methods is related to a basic MOS (model output statistic) methodology [3]. This methodology consists in adding a correction at t+1 taking into account the gap measured between the predicted and measured value at t. In this case the prediction becomes:

$$\widehat{GHI}_{MOS}(t+1) = \widehat{GHI}_{order_1}(t+1).\frac{GHI(t)}{\widehat{GHI}_{order_1}(t)} \tag{28}$$

4. Numerical experiments and validation set-up

The goal of this paper is to compare the accuracy of some machine learning against training-less techniques. We predict future values of solar irradiation from only past values of solar irradiation i.e. no exogenous variables are used. In other words, all forecasting methods described in this work seek to find a generic model F of the form shown in Eq (29).

$$\widehat{k^*}(t+h) = F(k^*(t), k^*(t-1), \cdots, k^*(t-p)) \text{ for } h = 1, 2, \ldots 6 \quad (29)$$

Where the sign ^ is used to identify the forecasted variable and the sequence $\{k^*(t), k^*(t-1), \cdots, k^*(t-p)\}$ represents the time series of p past values of the clear sky index. The forecast horizon denoted by the letter h usually ranges from 1h to 6h (intraday solar forecasting). In our case, as mentioned above, the variable of interest is the clear sky index k^*. From this forecast value of the clear sky index k*, GHI forecasts can be easily obtained by using Eq (5) and (6). The ANN and ARMA statistical methods previously described are supervised learning methods or data-driven approaches. As a consequence, the techniques rely on the information content embedded in the training data in order to produce forecasts on unseen data [31]. More precisely, the models' parameters (or in other words an approximation to the function F) are determined with the help of pairs of input and output examples contained in the training data. Once the model is fitted, the model can be evaluated on a test data set. In our context, $\mathcal{D} = \{\mathbf{x_i}, y_i\}_{\mathbf{i=1}}^{\mathbf{n}}$ represents the training data set. The vector $\mathbf{x_i}$ contains the p past values of the clear sky index for training sample i and y_i refers to the corresponding value of the clear sky index for the horizon h of interest. The column vector inputs for all n training cases can be aggregated in the so-called $n \times p$ design matrix $\mathbf{X}$ and the corresponding model's outputs (or targets) are collected in the vector $\mathbf{y}$ so we can write $\mathcal{D} = \{\mathbf{X}, \mathbf{y}\}$. Similarly, we have $\mathcal{D}_* = \{\mathbf{X}_*, \mathbf{y}_*\}$ for the test data set. During this study a k-fold methodology has been used. In k-fold cross-validation, the original sample is randomly partitioned into k equal sized subsamples [32]. k-fold cross validation should be employed to estimate the accuracy of the model induced from a classification algorithm, because the accuracy resulting from the training data of the model is generally too optimistic [33]. Among these k subsamples, a single subsample is retained as the validation data for testing the model, and the remaining $k-1$ subsamples are used as training data. The cross-validation process is then repeated *k-1* times (the *folds,* k=10 in our case), with - as a result - each of the *k* subsamples used exactly once as the validation data. The *k* results from the folds can then be averaged (or otherwise combined) to produce a single estimation. The advantage of this method over repeated random sub-sampling (see below) is that all observations are used for both training and validation, and each observation is used for validation exactly once [34]. 10-fold cross-validation is commonly used, but generally *k* remains an unfixed parameter.

In this work, we chose to report the accuracy of the different forecasting methods by solely using the nRMSE ($n\mathrm{RMSE} = \sqrt{}(\mathbf{E}[(\mathbf{x}\hat{}-\mathbf{x})^{\uparrow 2}])/\langle\mathbf{x}\rangle$) and the forecast skill ($s = 100.(1 - 〖RMSE〗_{\downarrow method}/〖RMSE〗_{\downarrow persistence})$) . It is the two most used error criterion in solar radiation forecasting [1].With this last metric the scaled persistence model has a forecast skill s= 0% [1]. A value of s= 100% denotes a perfect forecast. Negative values of s indicate that the forecasting model fails to outperform the scaled persistence model while positive values of s means that the forecasting method improves on scaled persistence. Further, the higher the skill score is, the better the improvement is.

5. Results

This section details the main results we obtained on the previous presented models for two time granularity (1 hour and 1 minute) and for 10 horizons (t+1 to t+10).

5.1.Hourly data

In this first part of the results, we show the importance of parameters optimization for the Kalman filter method. Then results of our benchmark for 1 hour horizon are presented. Finally we study the accuracy of the models varying the horizon from 1 hour to 10 hours.

5.1.1. Kalman parameters optimization

In the hourly case, the use of the Kalman filtering in general and in the Ajaccio data case in particularly, shows that determination of the *R* coefficient (equation 24) is very important and allows to deeply modify the output of the modelling. In the original Kalman filter theory, the covariance is computed over all past predictions and measurements. In practice, we find that modifying the data set size of this computing can improve results. Two parameters to be customized are very essential: the R coefficient and the windows width (WW) of the variance computing. WW allows to consider a seasonal aspect of the filtering, indeed, it is easy to imagine that in summer or winter, variance changes and should be adjusted. We have tested 2000 values of R (from 1 to 2000) and 50 WW (from 0 to 5000, by step of 100, in hour). The figure 3 shows three typical curves of the R-WW optimization, for clarity others curves are not shown.

Figure 3. Optimization of the R parameters and windows width (WW). Lines represent the lowest value of each curve

The best configuration is obtained with a WW equal to 300 hours and a R parameter close to 400. In the following of the manuscript we will use these values for $\widehat{GHI}_{kalman}$ computing.

5.1.2. Results for the 1 hour horizon

In the Table 1 are shown results of prediction for all the 8 predictors described previously. As an ideal reference we also proposed to consider the minimal error given by a perfect evaluation of the trend of the signal to be predicted (see section 3.1).

Table 1. Results (nRMSE) in the hourly case of the 8 predictors among the 3 types of forecasters (models, trainless models and recursive trainless model), in bold the best results for each column.

We can see that the best predictors (in the Annual case) are AR and NN but these are also the more elaborated methods. However the gain related to a very simple method as $\widehat{GHI}_{order_1}$ is very small, less than 1% for the annual error estimation. Note that AR and NN are computed with 80% of the data set used during the training phase. All the other forecasters don't need any training data. We see also that among the trainless model, the order_1 is the best and that it is not necessary to improve the theory and to consider higher orders. The Kalman filter improves quasi-systematically the order_1 prediction. We see also that the MOS prediction is not efficient, this result seems logic because the noise is important in the global radiation time series, with less noisy series it may be a good predictor. Note that concerning the quarter 3 the normal and recursive order 1 give very good results comparing to the perfect predictor (less than 3 percentage points). It is essentially during the second and last quarters that errors are very large, during these seasons it is ANN and AR which give the best results. In the figure 4 are shown the profiles of predictions for the best models (NN) and all the trainless models.

Figure 4. Predictions (crosses) Vs measures (lines) for 200 hours in Ajaccio

In this figure we see that visually it very difficult to compare all the predictors, the difference of performance is small, all the five predictors seem very interesting to predict the global solar radiation.

5.1.3. Importance of the size of the training set

In the previous results, NN and ARMA are related to a training step done with 80% of data. We propose now how to rank predictors considering the data available. Indeed, sometimes in global radiation forecasting, only a few data are available and the prediction method must should efficient nevertheless. In figure 4 we see the trend of the skill score considering the size of available data. The four best predictors are studied: NN, AR, Order_1 and Kalman estimations.

Figure 4. Influence of the training set on the prediction error

When there are few data (less than 150 days) we see that the trainless models are better than NN modelling. We can consider that before 600 days AR is the most interesting and after it is preferable to use NN. In parallel, the simple methods like order_1 and order_1 Kalman are very good alternative when only one year or less of data is available.

5.1.4. Results concerning the horizon considered

Because the prediction of the global forecasting is not only interesting 1 hour in advance, in this part we propose to study the impact of the considered horizon on the four best predictors. In addition, we compare ANN, AR, order_1 and Kalman filter with the irradiance forecasting based on the NWP estimation called “Climatology” in the following. The model used is AROME, a small scale numerical prediction model, operational at Meteo-France (French national meteorological service since December 2008). It was designed to improve short range forecasts of severe events such as intense

Mediterranean precipitations, severe storms, fog and urban heat during heat waves. Figure 5 shows the nRMSE related to the horizon.

Figure 5. nRMSE evolution in term of time horizons

Previously (1 hour horizon case) we've shown that trainless model order_1 and Kalman filter were similar to NN and AR models. When horizon increases, his effect becomes false, from the 2 hours horizon, the two last one are very better than the first ones. Moreover the NWP is better than NN and AR from the 6 hours horizon and better than the order_1 and Kalman filter from the 3 hours horizon.

5.2. Minute data

In order to test the sensitivity of the proposed Kalman filter to variation of meteorological conditions, we used a one minute data set while keeping the horizon of solar irradiation prediction equal to one hour. Each minute a prediction is done. Predictions are then integrated over the horizon. The *WW* has also been optimized for this time step and three other custom parameters were add and optimized, allowing to improve the forecasting performances: Kini (0.1-1 by step of 0.1) Kgap (0.1-1 by step of 0.1) and gap_error (1 - 50 by step of 5). Kini consist in reinitializing the Kalman filter gain K each morning (seasonal aspects being partly conserved with other parameters such as the covariance matrix P). The Kgap and gap_error consist in forcing the Kalman gain to the defined Kgap value if the difference between the predicted and the measure values of the previous time step is bigger than the defined gap_offset. Parameters optimization has been conducted this time step for R, *WW*, Kini, Kgap, gap_error based on the minimization of the nRMSE. Kini, Kgap and gap_error have been set to respectively 0.9, 0.4 and 20. The value of R appears to be independent of the time step of the data set and we found the same value of 400 as for the one hour database. One the other hand, the *WW* appears to be strongly dependent of the database time step. The optimum *WW* value is much lower (10 hours instead of 300 hours) when using the one minute database as the model is much more sensitive to the change of meteorological thought the day. In that case the daily aspect is more important than the seasonal one. Some additional works will be conducted in order to combine these both aspects. The figure 6 presents the results of the Kalman filter for 3 typical days: sunny, partially cloudy and cloudy.

Figure 6. Predictions (crosses) Vs measures (lines) for 3 typical days in Ajaccio, one minute time step database.

The table 2 presents the one hour horizon forecasting comparison of four tested models for the one minute time step (among the 3 types of forecasters). As expected, trainless models results are slightly better for smaller data set time step while the results of machine learning predictors failed to predict efficiently at this time step for short term horizons.

Table 2. Results (nRMSE) in the hourly case for one minute database time step.

When increasing the horizon, the recursive trainless model continues to give some good results comparing to the others tested models, as presented on figure 7.

Figure 7. nRMSE evolution in term of time horizons for one minute time step

We can observe that Kalman filter method gives better results than order 1 model for all the tested horizons. Note that AR and NN give the same results and that the curves are similar.

6. Conclusion

In the scope of satisfying electrical operator needs, several kinds of short term prediction methodologies: a naïve model, a linear model, non-linear models and models without training phase have been described. The difference of performance being very small, it is very difficult to compare all the presented predictors. All of them seem interesting to predict the global solar radiation depending on use situation: timestep, horizon, size of the training set, etc. Applying the parsimony concept, a model using the recursive Kalman method of order 1 is used performing the prediction from an ad-hoc filtering. The Kalman filter is a recursive estimator, this means that only the estimated state from the previous time step and the current measurement are needed to compute the estimate for the current state. In contrast to batch estimation techniques, no history of observations and/or estimates is required. Kalman filter is applied here for short-term forecasting for both one hour and one minute data sets timestep. The presented approach presents interesting results as it allows to improve quasi-systematically the order_1 prediction and Kalman model performances are, for one hour's horizons, competitive with much more complicated models such as ANN which require both consistent historical data sets (at least 200 days) and computational resources (time consuming, Matlab toolboxes, etc.). The optimized Kalman filter just required, to be fully operational, an historical database corresponding to the optimums *WW* identified: 300 h for a one hour data (seasonal influence) set and 10 h for a one minute data set (daily influence). Thus this method is easily applicable for poor instrumented or isolated sites. This work has been done over well-known data set. Then, the model will be tested on several distinct geographical spots such as Tilos Island in the framework of the Horizon 2020 TILOS project (http://www.tiloshorizon.eu/). Some additional work will be conduct in order to have simultaneously the seasonal and the daily influence.

Acknowledgment

This work is supported by European Union's Horizon 2020 research and innovation programme under grant agreement No 646529 through the TILOS project (Technology innovation for the Local Scale, Optimum integration of Battery Energy Storage). The data sets were provides by the CNRS UMR SPE 6134 Laboratory.

References

[1] Lauret P, Voyant C, Soubdhan T, David M, Poggi P. A benchmarking of machine learning techniques for solar radiation forecasting in an insular context. Sol Energy 2015;112:446–57. doi:10.1016/j.solener.2014.12.014.

[2] CRE. Cahier des charges de l'appel d'offres portant sur des installations au sol de production d'électricité à partir de l'énergie solaire. Ministère de l'Ecologie, de l'Energie, du Développement durable et de l'Aménagement du territoire; 2009.

[3] Diagne M, David M, Lauret P, Boland J, Schmutz N. Review of solar irradiance forecasting methods and a proposition for small-scale insular grids. Renew Sustain Energy Rev 2013;27:65–76. doi:10.1016/j.rser.2013.06.042.

[4] Lai TM, To WM, Lo WC, Choy YS. Modeling of electricity consumption in the Asian gaming and tourism center--Macao SAR, People's Republic of China. Energy 2008;33:679–88. doi:doi: DOI: 10.1016/j.energy.2007.12.007.

[5] Voyant C, Notton G, Paoli C, Nivet ML, Muselli M, Dahmani K. Numerical weather prediction or stochastic modeling: an objective criterion of choice for the global radiation forecasting. Int J Energy Technol Policy 2014.

[6] Espinar B, Aznarte J-L, Girard R, Moussa AM, Kariniotakis G. Photovoltaic Forecasting: A state of the art, OTTI - Ostbayerisches Technologie-Transfer-Institut; 2010, p. Pages 250-255-ISBN 978-3-941785-15-1.

[7] Lara-Fanego V, Ruiz-Arias JA, Pozo-Vázquez D, Santos-Alamillos FJ, Tovar-Pescador J. Evaluation of the WRF model solar irradiance forecasts in Andalusia (southern Spain). Sol Energy 2012;86:2200–17. doi:10.1016/j.solener.2011.02.014.

[8] Badescu V. Modeling solar radiation at the earth's surface: recent advances. Springer; 2008.

[9] Mellit A, Kalogirou S. Artificial intelligence techniques for photovoltaic applications: A review. Prog Energy Combust Sci 2008;34:574–632. doi:10.1016/j.pecs.2008.01.001.

[10] Voyant C, Paoli C, Muselli M, Nivet M-L. Multi-horizon solar radiation forecasting for Mediterranean locations using time series models. Renew Sustain Energy Rev 2013;28:44–52. doi:10.1016/j.rser.2013.07.058.

[11] Inness P, Dorling S. NWP Models – the Basic Principles. Oper. Weather Forecast., John Wiley & Sons, Ltd; 2012, p. 53–107.

[12] Hamilton J. Time series analysis. Princeton N.J.: Princeton University Press; 1994.

[13] Join C, Fliess M, Voyant C, Chaxel F. Solar energy production: Short-term forecasting and risk management. ArXiv160206295 Cs Q-Fin 2016.

[14] Paoli C, Voyant C, Muselli M, Nivet M-L. Forecasting of preprocessed daily solar radiation time series using neural networks. Sol Energy 2010;84:2146–60. doi:10.1016/j.solener.2010.08.011.

[15] Join C, Voyant C, Fliess M, Muselli M, Nivet M-L, Paoli C, et al. Short-term solar irradiance and irradiation forecasts via different time series techniques: A preliminary study, 2014.

[16] Andersson L-E, Elfving T. Interpolation and approximation by monotone cubic splines. J Approx Theory 1991;66:302–33. doi:10.1016/0021-9045(91)90033-7.

[17] De Gooijer JG, Hyndman RJ. 25 years of time series forecasting. Int J Forecast 2006;22:443–73. doi:10.1016/j.ijforecast.2006.01.001.

[18] Al-Alawi SM, Al-Hinai HA. An ANN-based approach for predicting global radiation in locations with no direct measurement instrumentation. Renew Energy 1998;14:199–204. doi:10.1016/S0960-1481(98)00068-8.

[19] Benghanem M, Mellit A, Alamri SN. ANN-based modelling and estimation of daily global solar radiation data: A case study. Energy Convers Manag 2009;50:1644–55. doi:10.1016/j.enconman.2009.03.035.

[20] Chaouachi A, Kamel RM, Ichikawa R, Hayashi H, Nagasaka K. Neural Network Ensemble-Based Solar Power Generation Short-Term Forecasting. World Acad Sci Eng Technol 2009;54.

[21] Mellit A, Pavan AM. A 24-h forecast of solar irradiance using artificial neural network: Application for performance prediction of a grid-connected PV plant at Trieste, Italy. Sol Energy 2010;84:807–21. doi:10.1016/j.solener.2010.02.006.

[22] Faraday J, Chatfield C. Times Series Forecasting with Neural Networks: A Case Study 1998.

[23] Ineichen P. A broadband simplified version of the Solis clear sky model. Sol Energy 2008;82:758–62. doi:10.1016/j.solener.2008.02.009.

[24] Mueller RW, Dagestad KF, Ineichen P, Schroedter-Homscheidt M, Cros S, Dumortier D, et al. Rethinking satellite-based solar irradiance modelling: The SOLIS clear-sky module. Remote Sens Environ 2004;91:160–74. doi:10.1016/j.rse.2004.02.009.

[25] these cyril voyant pdf free ebook download. n.d.

[26] Voyant C, Muselli M, Paoli C, Nivet M-L. Numerical weather prediction (NWP) and hybrid ARMA/ANN model to predict global radiation. Energy 2012;39:341–55. doi:10.1016/j.energy.2012.01.006.

[27] Grewal MS. Kalman filtering. Springer; 2011.

[28] Julier SJ, Uhlmann JK. New extension of the Kalman filter to nonlinear systems. AeroSense'97, International Society for Optics and Photonics; 1997, p. 182–193.

[29] Shumway RH, Stoffer DS. An approach to time series smoothing and forecasting using the EM algorithm. J Time Ser Anal 1982;3:253–264.

[30] Harvey AC. Forecasting, structural time series models and the Kalman filter. Cambridge university press; 1990.

[31] Bengio Y, Grandvalet Y. No unbiased estimator of the variance of k-fold cross-validation. J Mach Learn Res 2004;5:1089–1105.

[32] Rodriguez JD, Perez A, Lozano JA. Sensitivity analysis of k-fold cross validation in prediction error estimation. IEEE Trans Pattern Anal Mach Intell 2010;32:569–575.

[33] Wong T-T. Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. Pattern Recognit 2015;48:2839–46. doi:10.1016/j.patcog.2015.03.009.

[34] Wiens TS, Dale BC, Boyce MS, Kershaw GP. Three way k-fold cross-validation of resource selection functions. Ecol Model 2008;212:244–255.

FIGURES

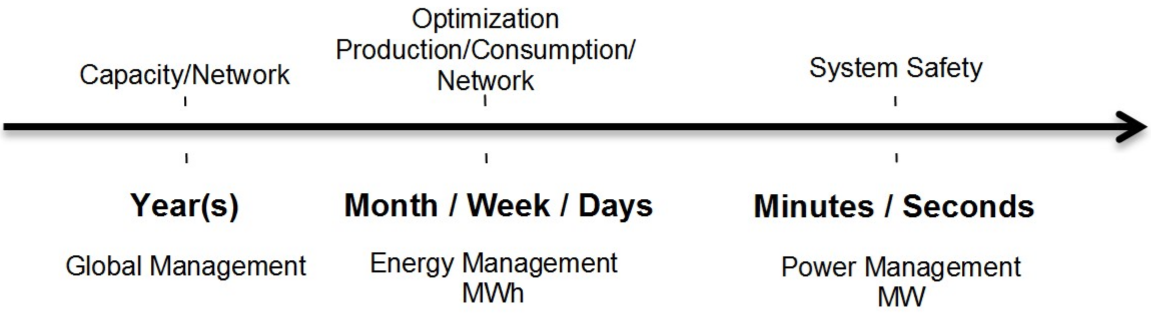

Figure 2: Prediction scale for energy management in an electrical network.

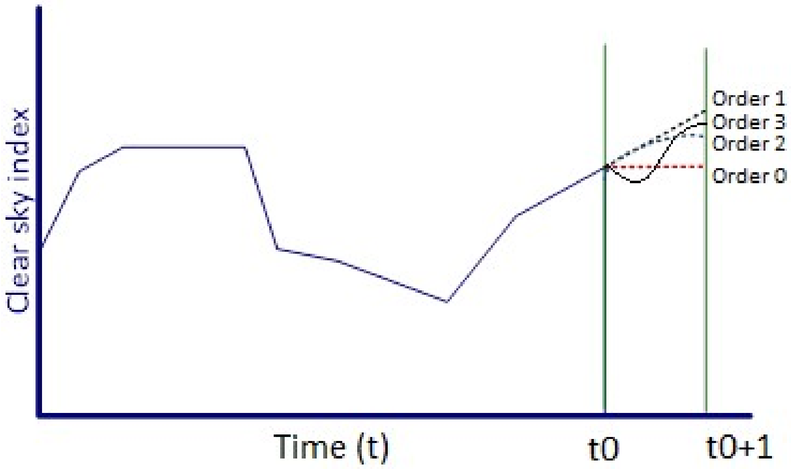

Figure 2. Illustration of the prediction method without training phase using a model of order 0, 1, 2 and 3.

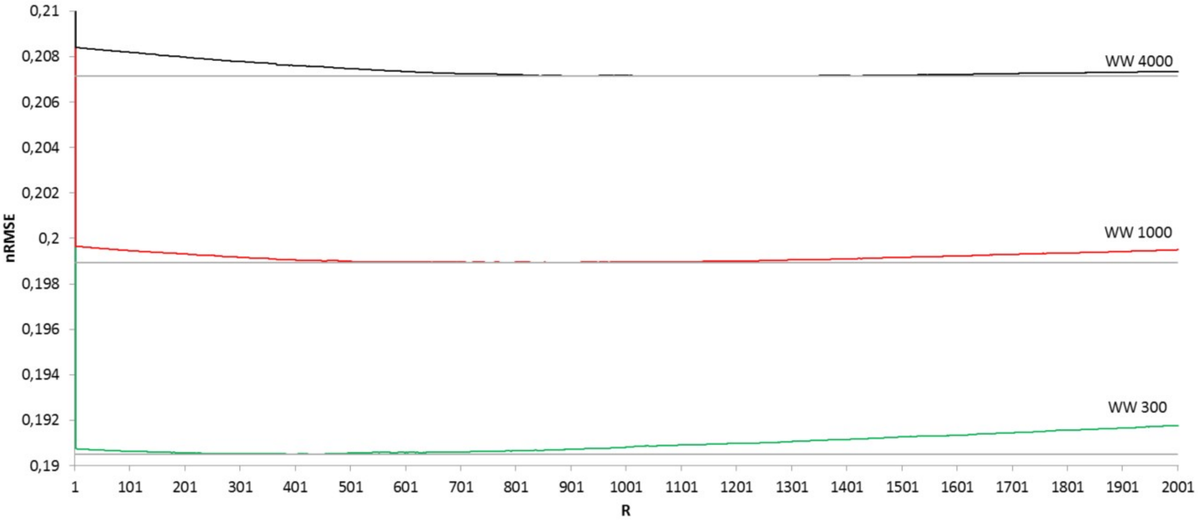

Figure 3. Optimization of the *R* parameters and windows width (*WW*). Lines represent the lowest value of each curve

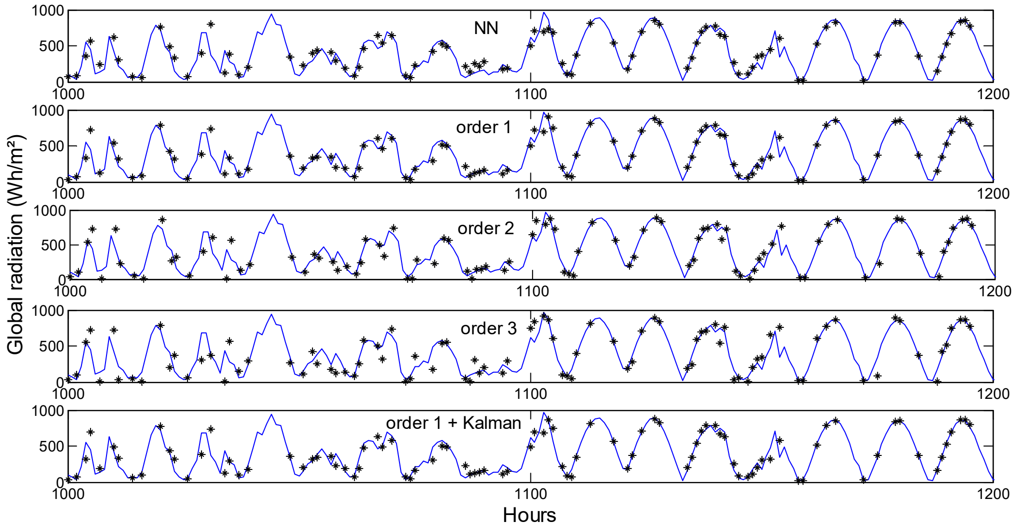

Figure 4. Predictions (crosses) Vs measures (lines) for 200 hours in Ajaccio

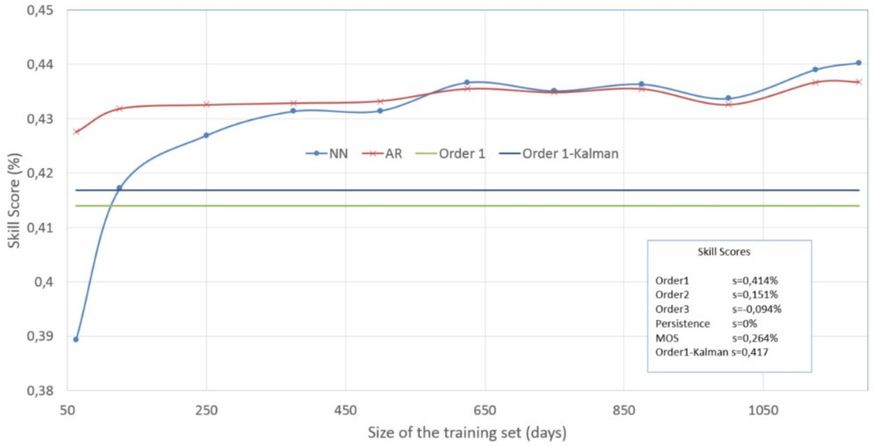

Figure 5. Influence of the training set on the prediction error

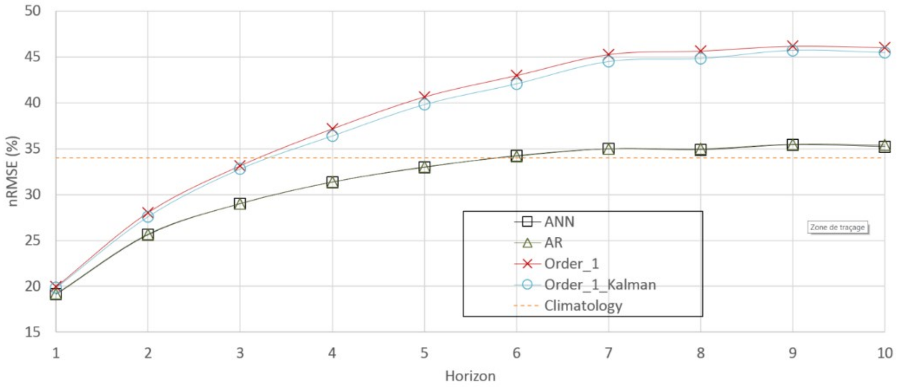

Figure 6. nRMSE evolution in term of time horizons

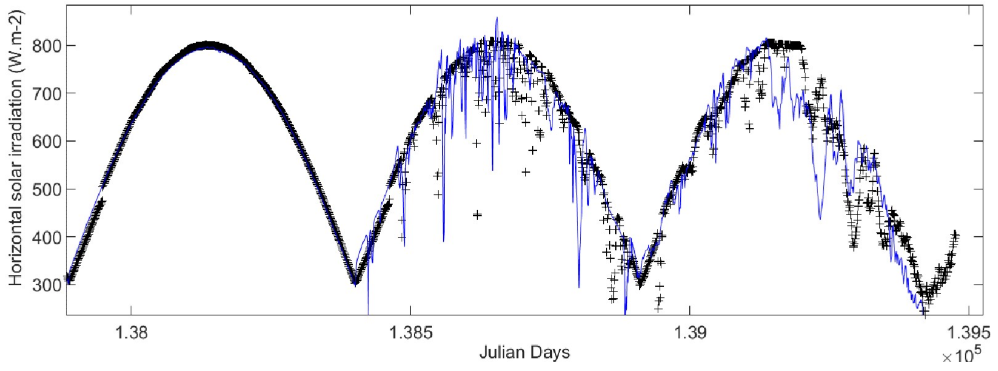

Figure 7. Predictions (crosses) Vs measures (lines) for 3 typical days in Ajaccio, one minute time step database.

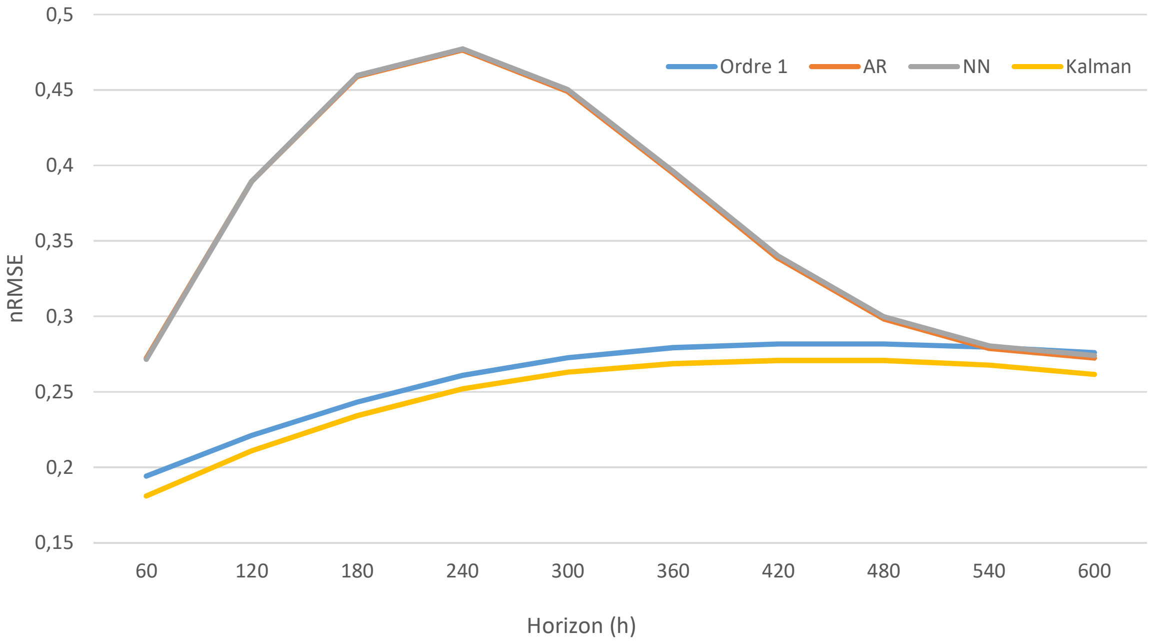

Figure 8. nRMSE evolution in term of time horizons for one minute time step

TABLES

k-fold=10		nRMSE				
		Annual	Quater 1	Quater 2	Quarter 3	Quater 4
models	Lowest error	0,0993	0,0916	0,1043	0,0918	0,1059
	AR	0,1953	0,1722	0,2349	0,1259	0,2203
	NN	**0,1947**	0,1707	**0,2320**	0,1287	**0,2184**
trainless models	persistence (order_0)	0,3427	0,3065	0,3933	0,3092	0,3687
	order_1	0,2022	0,1721	0,2391	**0,1186**	0,2344
	order_2	0,2909	0,2514	0,3111	0,1932	0,3621
	order_3	0,3745	0,3137	0,3762	0,2748	0,4546
	MOS	0,2530	0,2050	0,2888	0,1651	0,3085
resursive trainless models	order_1_Kalman	0,2010	**0,1706**	0,2391	0,1197	0,2330

Table 1. Results (nRMSE) in the hourly case of the 8 predictors among the 3 types of forecasters (models, trainless models and recursive trainless model), in bold the best results for each column.

k-fold=10		nRMSE				
		Annual	Quarter 1	Quarter 2	Quarter 3	Quarter 4
models	Ideal predictor	0.0993	0.0916	0.1043	0.0918	0.1059
	AR	0.2724	0.2990	0.2602	0.1985	0.3012
	NN	0.2715	0.2968	0.2598	0.1985	0.3006
trainless models	Order_1	0.1941	0.2298	**0.1790**	**0.0876**	0.2356
recursive trainless models	Order_1_Kalman_	**0.1812**	**0.2266**	0.1795	0.0933	**0.2328**

Table 2. Results (nRMSE) in the hourly case for one minute database time step.