



HAL
open science

Uncertainties in global radiation time series forecasting using machine learning: The multilayer perceptron case

Cyril Voyant, Gilles Notton, Christophe Paoli, Alexis Fouilloy, Fabrice Motte, Christophe Darras

► To cite this version:

Cyril Voyant, Gilles Notton, Christophe Paoli, Alexis Fouilloy, Fabrice Motte, et al.. Uncertainties in global radiation time series forecasting using machine learning: The multilayer perceptron case. Energy, 2017, 125, pp.248-257. 10.1016/j.energy.2017.02.098 . hal-01470994

HAL Id: hal-01470994

<https://hal.science/hal-01470994>

Submitted on 18 Feb 2017

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

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

Uncertainties in global radiation time series forecasting using machine learning: The multilayer perceptron case

Cyril Voyant^{1,2*}, Gilles Notton¹, Christophe Paoli^{1,3}, Alexis Fouilloy¹, Fabrice Motte¹, Christophe Darras¹

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

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

³Galatasaray University, Çırağan Cad. No:36, 34349 Ortaköy/İstanbul, Turkey

Abstract.

As global solar radiation forecasting is a very important challenge, several methods are devoted to this goal with different levels of accuracy and confidence. In this study we propose to better understand how the uncertainty is propagated in the context of global radiation time series forecasting using machine learning. Indeed we propose to decompose the error considering four kinds of uncertainties: the error due to the measurement, the variability of time series, the machine learning uncertainty and the error related to the horizon. All these components of the error allow to determinate a global uncertainty generating prediction bands related to the prediction efficiency. We also have defined a reliability index which could be very interesting for the grid manager in order to estimate the validity of predictions. We have experimented this method on a multilayer perceptron which is a popular machine learning technique. We have shown that the global error and its components are essential to quantify in order to estimate the reliability of the model outputs. The described method has been successfully applied to four meteorological stations in Mediterranean area.

Keywords: Time Series Forecasting, Processing, Artificial Neural Networks, interval, Energy Prediction, Stationarity

*Correspondingauthor: Cyril Voyant, tél: +33 4 95 52 41 30, fax (33 4 95 52 53 28) email: voyant@univ-corse.fr

1 Introduction

Solar radiation is one of the principal energy sources for physical, biological and chemical processes, occupying the most important role in many engineering applications[1]. The process of converting sunlight to electricity without combustion allows to create power without pollution. The major problem of such energy source is its intermittence and its stochastic character which make difficult their management into an electrical network[2]. Thereby, the development of forecasting models is necessary to ideally use this technology[3]. By considering their effectiveness, it will be possible for example to identify the most optimal locations for developing a solar power project or to maintain the grid stability and security of a power management system[4]. Thus the solar energy forecasting is a process used to predict the amount of solar energy available for various time horizons[5]. Several methods have been developed by experts around the world and the mathematical formalism of Times Series (TS[6]) has been often used for the short term forecasting (among 6 hours ahead) [5,7]. TS is a set of ordered numbers that measures some activities over time[8]. It is the historical record of global horizontal irradiance with measurements taken at equally spaced intervals with a consistency in the activity and the method of measurement. Some of the best predictors found in literature are Autoregressive and moving average[9–11], Bayesian inferences [12,13], Markov chains [14,15], k-Nearest-Neighbors predictors[16], support vector machine[3,17], regression tree[18,19], or artificial neural network (ANN) [20,21]. All these approaches are related to the machine learning application[22]. The most often used is the last presented method: the artificial neural network and particularly the multilayer perceptron (MLP[23]). In the present study, we focus on this prediction method, the goal being to detail the uncertainties related to the global radiation prediction[24].

The paper is organized as follow: Section 2 describes the data and material needed to conduct our experiments. In the section 3, we propose to define the different component of the error generated through the MLP used. These uncertainties can be decomposed into several components that will be explained and developed. In section 4, the results of the error decomposition will be exposed for five meteorological sites in order to quantify the reliability of the predictions. The last section will allow to draw conclusions about the present study.

2 Data and material

In this work, measured hourly horizontal global radiation data from meteorological ground stations are used to forecast global horizontal solar irradiation (GHI) for a specific horizon[11]. All the measurements used are obtained from the French Meteorological Organization (Météo-France) data base and from measurement realized in the frame of the Tilos H2020 project (<http://www.tiloshorizon.eu>). Four sites are studied: Ajaccio, Bastia, Montpellier and Marseille in France. As for all experimental acquisitions, missing values are observed, here, this represents less than 2% of the data. A classical cleaning approach is then operated in order to identify and remove this data[16].

2.1 Data

In Corsica Island, the data used to build the models are GHI measured in the meteorological stations of Ajaccio (41°55'N, 8°44'E, 4m asl) and Bastia (42°42'N, 9°27'E, 10m asl). They are located near the Mediterranean Sea and nearby mountains (1000 m altitude at 40km from the sites). The data representing the global horizontal solar radiation were measured on an hourly basis from 1998 to 1999 (exactly two years). The two last studied stations are Montpellier (43.6°N and 3.9°E, 2 masl) and Marseille (43.4°N and 5.2°E, 5 masl) concerning the years 2008 and 2009. All these stations are equipped with pyranometers (CM 11 from Kipp&Zonen). The choice of these particular places is explained by their closed geographical and orographical configurations. These stations are located near the Mediterranean Sea and mountains. This specific geographical configuration of the four French meteorological stations makes cloudness difficult to forecast. Mediterranean climate is characterized by hot summers with abundant sunshine and mild, dry and clear winters. Irradiance nighttime values are not being used, the first morning data forecast are operated with the day before evening data.

2.2 Prediction methodology

We chose to develop error propagation in the GHI prediction for the most common used predictor: the MLP. The base of this model is the time series approach (TS). A TS $x(t)$ can be defined by a linear or non-linear model called f_n (see Equation 1 where $t = n, n-1, \dots, p+1, p$ with n , the number of observations and p the number of parameters of the model ; $n \gg p$; h is the horizon of prediction and ϵ_{t+h} the committed error) [25].

$$x(t+h) = f_n(x(t), x(t-1), \dots, x(t-p+1)) + \epsilon_{t+h} \quad \text{Eq1}$$

To estimate the f_n model, a stationarity hypothesis is often necessary. This condition usually implies a stable process [26,27]. This notion is directly linked to the fact that whether certain feature such as mean or variance change over time or remain constant. Previous studies [3,28,29] show that the use of clear sky index (CSI) allows to make stationary the time series and so to correctly use the MLP forecasting.

2.2.1 Stationary process

In previous studies [30,31], it was demonstrated that the clear sky index calculated with the simplified Solis model [32] is the most reliable for our locations. The Solis model generates a clear sky hourly irradiation (CS) expressed by Eq. (2), the use of this model requires fitting parameter (g), extraterrestrial radiation (I_0), solar elevation (h) and total measured atmospheric optical depth (τ):

$$CS(t) = I_0(t) \cdot \exp\left(\frac{-\tau}{\sin^g(h(t))}\right) \cdot \sin^2(h(t)) \quad \text{Eq2}$$

The simplified “Solis clear sky” model is based on radiative transfer calculations and the Lambert-Beer relation [32]. 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 scattering effects. According to [33] this model remains a good fitting function of the global horizontal radiation. The new computed time series (CSI) can be directly used with the MLP forecasting and is described by the equation 3:

$$CSI(t) = GHI(t)/CS(t) \quad \text{Eq 3}$$

2.2.2 MLPprediction

Although a large range of different architectures of ANN is available [34], MultiLayer Perceptron (MLP) remains the most popular [35]. In particular, feed-forward MLP networks with two layers (one hidden layer and one output layer) are often used for modeling and forecasting time series. Several studies [20,36,37] validated this approach based on ANN for the non-linear modeling of time series. To forecast the time series, a fixed number p of past values are set as inputs of the MLP, the output is the prediction of a future value [38]. Considering the initial time series equation (Equation 1), this equation can be adapted to the non-linear case of one hidden layer MLP with b related to the biases, f and g to the activation function of the output and hidden layer, and ω to the weights. The number of hidden nodes (H) and the number of the input node (In) allow to detail this transformation. The number of layer 1 and 2 is given in superscript. (Equation 4):

$$\widehat{CSI}(t + 1) = f(\sum_{i=1}^H y_i \omega_i^2 + b^2) \text{ with } y_i = g(\sum_{j=1}^{In} CSI(t - j + 1) \omega_{ij}^1 + b_i^1) \quad \text{Eq 4}$$

In the presented study, the MLP has been computed with the Matlab© software and its Neural Network toolbox. The characteristics chosen and related to previous work are the following: one hidden layer, the activation functions are the continuously and differentiable hyperbolic tangent (hidden) and linear (output), the Levenberg-Marquardt learning algorithm with a max fail parameter before stopping training equal to 5 (early stopping tool allowing the stop the learning when the error increases consecutively 5 times). This algorithm is an approximation to the Newton's method. The prediction of the GHI is obtained using the equation:

$$\widehat{GHI}(t + 1) = \widehat{CSI}(t + 1).CS(t + 1) \quad \text{Eq 5}$$

To customize the input layer of the MLP we choose the use of the mutual information to determine In as described in [3,39,40]. According the results obtained in these papers, we use H equal to In for all the experiments conducted in this study. Furthermore in order to improve the learning of the MLP, it is a common practice to filter out the data removing night hours. Indeed we consider only periods between sunrise and sunset [41,42]. 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 [3]. This transformation is equivalent to a

filtering related to the solar elevation angle lower than 10°. All the simulations are related to the Matlab software and NNtoolbox use.

3 Error decomposition

In these section, we propose to decompose the error considering four kinds of uncertainties: the error due to the measurement, the error due to the variability of the time series, the error related to the machine learning uncertainty and the error related to the horizon.

3.1 Error due to the measurement (σ_{meas})

In experimental sciences, there is no perfect measure. Experiments can only be marred with significant errors more or less depending on the selected protocol or the quality measuring instruments [24]. Assess the uncertainty measurement is a complex task that is the subject of a complete branch called metrology. The uncertainty associated with a measurement result allows to provide a quantitative indication of the quality of this result [43].

In this section, we will show that it is possible to quantify the impact of a measurement error (or precision) on the MLP output. A MLP with 2 inputs and 2 hidden neurons (H=2 and In=2) is considered here in order to understand the methodology[2]. The output of this MLP can be defined with the following formula:

$$\widehat{CSI}(t+1) = \sum_{i=1}^H (g(\sum_{j=1}^n CSI(t-j+1) \omega_{ij}^1 + b_i^1)) \omega_i^2 + b^2 = \omega_1^2 \tanh(CSI(t) \omega_{11}^1 + CSI(t-1) \omega_{12}^1 + b_{11}^1) + \omega_2^2 \tanh(CSI(t) \omega_{21}^1 + CSI(t-1) \omega_{22}^1 + b_{21}^1 + b_{22}^1) \quad \text{Eq 6}$$

In order to calculate the uncertainty propagation of a MLP related to measurement error, we propose to use two methods: (i) the classical variables differentiation and (ii) the differentiation of $\log(\widehat{CSI}(t+1))$.

Using the classical variables differentiation, we obtained the following formula for the measurement error:

$$\sigma_{meas}^2(\widehat{CSI}(t+1)) = \left(\frac{\partial \widehat{CSI}(t+1)}{\partial CSI(t)}\right)^2 \sigma^2(CSI(t)) + \left(\frac{\partial \widehat{CSI}(t+1)}{\partial CSI(t-1)}\right)^2 \sigma^2(CSI(t-1)) = u \cdot \sigma^2(CSI(t)) + v \cdot \sigma^2(CSI(t-1)) \quad \text{Eq 7}$$

$$\text{With } u = (\omega_1^2 \omega_{11}^1 (\tanh(CSI(t) \omega_{11}^1 + CSI(t-1) \omega_{12}^1 + b_{11}^1)^2 - 1) + \omega_2^2 \omega_{21}^1 (\tanh(CSI(t) \omega_{21}^1 + CSI(t-1) \omega_{22}^1 + b_{21}^1 + b_{22}^1)^2 - 1)) \quad \text{Eq 8}$$

Eq 8

and $v = (\omega_1^2 \omega_{12}^1 (\tanh(CSI(t)\omega_{11}^1 + CSI(t-1)\omega_{12}^1 + b_1^1)^2 - 1) + \omega_2^2 \omega_{22}^1 (\tanh(CSI(t)\omega_{21}^1 + CSI(t-1)\omega_{22}^1 + b_2^1)^2 - 1)) / 2$

On the two parameters u and v , and considering that $\sigma^2(CSI(t-1))$ et $\sigma^2(CSI(t))$ are equivalent and equal to the pyranometer uncertainty ($\sigma^2(CSI)$), the global error is:

$$\sigma_{meas\ 1}^2(\widehat{CSI}(t+1)) = \sigma^2(CSI)((\omega_1^2 \omega_{11}^1 \mathfrak{S}_1 + \omega_2^2 \omega_{21}^1 \mathfrak{S}_2)^2 + (\omega_1^2 \omega_{12}^1 \mathfrak{S}_1 + \omega_2^2 \omega_{22}^1 \mathfrak{S}_2)^2) \quad \text{Eq 10}$$

$$\text{Considering that } \mathfrak{S}_n = \tanh(CSI(t)\omega_{n1}^1 + CSI(t-1)\omega_{n2}^1 + b_n^1)^2 - 1 \text{ for } n \in [1, H] \quad \text{Eq 11}$$

It follows that the uncertainty of a MLP related to measurement error is:

$$\sigma_{meas\ 1}(\widehat{CSI}(t+1)) = \sigma(CSI)((\omega_1^2 \omega_{11}^1 \mathfrak{S}_1 + \omega_2^2 \omega_{21}^1 \mathfrak{S}_2)^2 + (\omega_1^2 \omega_{12}^1 \mathfrak{S}_1 + \omega_2^2 \omega_{22}^1 \mathfrak{S}_2)^2)^{1/2} \quad \text{Eq 12}$$

The maximum value will be reached when the output of the hidden nodes will be equal to 1 ($\mathfrak{S}_i = 1 \forall i$). In this case, we obtain:

$$\sigma_{meas\ 1, max}(\widehat{CSI}_{t+1}) = \sigma(CSI)((\omega_1^2 \omega_{11}^1 + \omega_2^2 \omega_{21}^1)^2 + (\omega_1^2 \omega_{12}^1 + \omega_2^2 \omega_{22}^1)^2)^{1/2} \quad \text{Eq 13}$$

Note that this formalism is applied only because the orthogonality hypothesis of the inputs have been done (inputs independent) without this approximation the computing is impossible. The generalization for H hidden neurons and I input nodes gives:

$$0 \leq \sigma_{meas\ 1}(\widehat{CSI}(t+1)) = \sigma(CSI)(\sum_{j=1}^I (\sum_{i=1}^H \omega_i^2 \omega_{ij}^1 \mathfrak{S}_i)^2)^{1/2} \leq \sigma(CSI)(\sum_{j=1}^I (\sum_{i=1}^H \omega_i^2 \omega_{ij}^1)^2)^{1/2} \quad \text{Eq 14}$$

The second method is based on the differentiation of $\log(\widehat{CSI}(t+1))$. This method is simpler but less efficient and does not take into account the error compensation. In the case we use the following formulas to determine the uncertainty ($\sigma_{meas\ 2}$):

$$d(\log(\widehat{CSI}(t+1))) = \frac{\sigma_{meas\ 2}(\widehat{CSI}(t+1))}{\widehat{CSI}(t+1)} = \frac{\sigma(CSI)(\omega_1^2 \mathfrak{S}_1 (\omega_{11}^1 + \omega_{12}^1) + \omega_2^2 \mathfrak{S}_2 (\omega_{21}^1 + \omega_{22}^1))}{\widehat{CSI}(t+1)} \quad \text{Eq 15}$$

thus:

$$\begin{aligned} \sigma_{meas\ 2}(\widehat{CSI}(t+1)) &= \sigma(CSI)(\omega_1^2 \mathfrak{S}_1 (\omega_{11}^1 + \omega_{12}^1) + \omega_2^2 \mathfrak{S}_2 (\omega_{21}^1 + \omega_{22}^1)) \\ &= \sigma(CSI)(\omega_1^2 \omega_{11}^1 \mathfrak{S}_1 + \omega_2^2 \omega_{21}^1 \mathfrak{S}_2 + \omega_1^2 \omega_{12}^1 \mathfrak{S}_1 + \omega_2^2 \omega_{22}^1 \mathfrak{S}_2) \end{aligned} \quad \text{Eq 16}$$

For H hidden neurons and I input nodes, we obtain the following generalization:

$$\sigma_{meas\ 2}(\widehat{CSI}(t+1)) = \sigma(CSI)(\sum_{j=1}^I \sum_{i=1}^H \omega_i^2 \omega_{ij}^1 \mathfrak{S}_i) \quad \text{Eq 17}$$

Combining Eq (16) and (14), we see that $\sigma_{meas\ 1}(\widehat{CSI}(t+1)) \leq \sigma_{meas\ 2}(\widehat{CSI}(t+1))$. In order to take into account compensation only present in meas1 case computing, we think it is preferable to use the first form and in the following we use the uncertainty of the measurement with the equation 18 (with H the number of hidden neurons and I the number of input nodes):

$$0 \leq \sigma_{meas}(\widehat{GHI}(t+1)) = \sigma(GHI)(\sum_{i=1}^H \omega_i^2 \sum_{j=1}^{In} \omega_{ij}^1 \mathfrak{S}_i) \leq \sigma(GHI)(\sum_{i=1}^H \omega_i^2 \sum_{j=1}^{In} \omega_{ij}^1) \quad \text{Eq 18}$$

3.2 Error due to the quick fluctuations of the time series

We choose to define the error due to the variability of the time series. Indeed, the quick fluctuations of the series are very difficult to predict and generate error in the prediction. One of the possibility to define this kind of uncertainty will be called the inherent error and the second one the variability error.

3.2.1 Inherent error (σ_{inh})

The Cartier and Perrin theorem [6], which is stated in the language of nonstandard analysis, allows to understand the existence of trends for time series [47]. The time series $GHI(t)$ may then be decomposed as a sum where $GHI_{trend}(t)$ is the trend and $GHI_{fluc}(t)$ is a “quickly fluctuating” function around 0 also called the inherent noise of the time series.

$$GHI(t) = GHI_{trend}(t) + GHI_{fluc}(t) \quad \text{Eq 19}$$

The nature of those quick fluctuations is left unknown and nothing prevents us from assuming that $GHI_{fluc}(t)$ is random and/or fractal. The forecast of the trend is possible on a “short” time interval under the assumption of a lack of abrupt changes, whereas the forecast of the fluctuation term at a given time instant is meaningless and should be abandoned. Based on this kind of time series definition, an ideal prediction can be obtained from a trend estimation $GHI_{trend}(t)$. In our case, we choose to compute it with a classical non-linear fit based on cubic spline data interpolation based on a tridiagonal linear system [48]. It is solved for the information needed to describe the coefficients of the various cubic polynomials which make up the interpolating spline. 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. The inherent error is so computed with the equation 20:

$$\sigma_{inh} = \widehat{GHI}(t+1) \cdot nRMSE(GHI_{trend}(t) - GHI(t)) \text{ with } nRMSE = \frac{\sqrt{E[(\widehat{GHI} - GHI)^2]}}{E(GHI)} \quad \text{Eq 20}$$

3.2.2 Variability error (σ_{var})

The knowledge of the volatility of the series at time t provides an a priori information about the variability, and so, about the expected error obtained with machine learning predictions [51]. Indeed, it is possible to define the volatility[49] via $Vol1(t) = |CSI(t) - \widehat{CSI}(t)|$. It is also possible to define another form of volatility ($Vol2(t)$) integrating the logarithm in order to define the log return which has the nice property of log-normality[50]. These parameters are constructed in order to take into account the intermittency in the CSI series represented by GHI_{fluc} in the equation 19

$$Vol2(t) = |\log(CSI(t)) - \log(CSI(t-1))| \quad \text{Eq 21}$$

In the following, $Vol2$ will be used to compute the volatility. Note that in order to take into account the daily seasonality of the series which would modify the results, the volatility is computed with the CSI and not with the GHI. In previous studies[6,50], it has been shown that volatility is linked to the error of prediction ($nRMSE = f(Vol2(t))$). With this argument, the error variability (σ_{var}) is defined by equation 22 where g is a non-linear function depending on the considered site.

$$\sigma_{var} = \widehat{GHI}(t+1) \cdot g(Vol2(t)) \quad \text{Eq 22}$$

3.3 Error related to the machine learning uncertainty

Another type of error that can be generated during the prediction is related to machine learning approach itself[44]. Indeed, in supervised learning applications and statistical learning theory, the out-of-sample error is a measure of how accurately an algorithm is able to predict outcome values for new data. Because learning algorithms are evaluated on finite samples [45], the evaluation of a learning algorithm may be sensitive to sampling error [46]. As a result, measurements of prediction error on the current data may not provide much information about predictive ability on new data. Generalization error can be minimized by avoiding overfitting in the learning algorithm. In the next subsection, the various kind of uncertainties generated under a classical MLP prediction of GHI will be described.

3.3.1 Sampling error (σ_{samp})

The MLP parameters are determined in using the pairs of input and output examples contained in the training data. In fact, this property is related to all the machine learning method. Once the model is fitted, the model can be evaluated on a test data set. In our context, $\mathcal{D} = \{\mathbf{x}_i, y_i\}_{i=1}^n$ represents the training data set. The vector x_i contains the p past values of the clear sky index (taken as inputs of the model) for training sample i and y_i refers to the corresponding value of the clear sky index for the considered horizon h (in the case of horizon 1 hour, $y_i = CSI(t_i + 1)$). The column vector inputs for all n training cases can be aggregated in the so-called $n \times p$ design matrix X and the corresponding model's outputs (or targets) are collected in the vector y so we can write $D = \{X, y\}$. To overcome the problem of sampling phase, often, a k -fold methodology is used [50]. In k -fold cross-validation, the original sample is randomly partitioned into k equal sized subsamples [52]. 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 [53]. The cross-validation process is then repeated $k-1$ times (the folds), 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 and the standard deviation (σ_{samp}) of the results can be computed.

$$\sigma_{samp}(t+1) = \left(\frac{1}{k} \cdot \sum_{i=1}^k (\widehat{GHI}_k(t+1) - \langle \widehat{GHI}(t+1) \rangle)^2 \right)^{1/2} \quad \text{Eq 23}$$

3.3.2 Initialization learning error (σ_{ini})

With MLP, there are a lot of methods to initialize weights before to use a back propagation method. The most often used is certainly the random initialization. In order to consider only the initialization which minimize the error of prediction under a test sample. The problem is that a global minimum is maybe not reached. Gradient-based minimization of the cost function during the learning phase is relatively fast, but for complex problems, the training may find local minima of the error function that are far from the global minimum [55]. In order to take into account the phenomenon, we propose to define the standard deviation of the outputs related to 50 trainings which 50 random initializations (arbitrary choice). With this approach, we propose

that more is important the number of local minimum, less it is evident to find the global minimum. With this assumption, the initialization learning error is described by the equation 24(for l=50 random initializations).

$$\sigma_{ini}(t+1) = \left(\frac{1}{l} \cdot \sum_{i=1}^l (\widehat{GHI}_i(t+1) - \langle \widehat{GHI}(t+1) \rangle)^2 \right)^{1/2} \quad \text{Eq 24}$$

3.4 Horizon error (σ_{hor})

It is obvious to understand that more the prediction horizon is important less the prediction is efficient. The prevailing view is that the evidence for long-horizon GHI predictability is significantly stronger than that for short horizons. All researchers plotting an GHI time series autocorrelation observed that the link between GHI (t) and GHI(t+i) (with i>1) decreases when GHI is made stationary without seasonal effect. This lack of correlation is more generally explained by the Hurst exponent or the Lyapunov horizon [55]. To overcome this problem, we propose to compute $\sigma_{hor}(h)$ considering the observed error for each horizon in a test sample, as described in equation 25.

$$\sigma_{hor}(h) = \widehat{GHI}(t+h) \cdot \frac{nRMSE(GHI(t+h))}{nRMSE(GHI(t+1))} = \widehat{GHI}(t+h) \cdot \alpha(h) \quad \text{Eq 25}$$

3.5 Global error of prediction (σ_{tot})

In our assumption, all the previous σ terms are independent random variables that are normally distributed (and therefore also jointly so), then their sum is also normally distributed and the global form of the standard deviation $\sigma_{tot}(t+1)$ becomes (in this equation the quick fluctuations are taken into account with σ_{var} , but it is also possible to consider σ_{inh}):

$$\sigma_{tot}(t+1) = \sqrt{(\sigma_{meas})^2 + (\sigma_{samp}(t+1))^2 + (\sigma_{ini}(t+1))^2 + (\sigma_{var}(t+1))^2} \quad \text{Eq 26}$$

Considering that there is a persistence of the variability for a short horizon, $\sigma_{var}(t) = \sigma_{var}(t+1)$

thus:

$$\sigma_{tot}(t+1) = \sqrt{(\sigma_{meas})^2 + (\sigma_{samp}(t+1))^2 + (\sigma_{ini}(t+1))^2 + (\sigma_{var}(t))^2} \quad \text{Eq 27}$$

with $\sigma_{meas} = \sigma(CSI)(\sum_{j=1}^{Ne}(\sum_{i=1}^{Nc} \omega_i^2 \omega_{ij}^1)^2)^{1/2}$, σ_{samp} and σ_{ini} are computed respectively with k-fold and 50 random initializations and $\sigma_{var} = \widehat{GHI}(t+1) \cdot g(Vol2(t))$. For an easier computing, it is also possible to use σ_{inh} replacing $\sigma_{var}(t)$ with a less robust result (Equation 20) but not dependent on the instant of the prediction.

$$\sigma_{tot}(t+1) = \sqrt{(\sigma_{meas})^2 + (\sigma_{samp}(t+1))^2 + (\sigma_{ini}(t+1))^2 + (\sigma_{inh})^2} \quad \text{Eq 28}$$

It is possible to define a prediction band taking into account all the uncertainties (Eq 29).

$$\widehat{GHI}(t+1) = \widehat{GHI}_{MLP}(t+1) \pm \sigma_{tot}(t+1) \quad \text{Eq 29}$$

Such prediction intervals were often proposed in the literature [6,56,57]; they refer to machine learning method $(\sigma_{ML}(t+1))^2 = (\sigma_{samp}(t+1))^2 + (\sigma_{ini}(t+1))^2$ [56,57] or to volatility and $\sigma_{var}(t)$ [6] but rarely both to the two kinds of uncertainty and never concerning σ_{meas} . Note that in the case of other machine learning methods used the term σ_{ini} can be equal to zero (e.g. support vector regression, regression tree etc.). The ideal case would be to systematically propose a confidence interval of prediction related to the three sorts of uncertainty (with $\sigma_{TS} = \sigma_{var}(t)$ or σ_{inh} considering the desired reliability).

$$\sigma_{tot}(t+1) = \sqrt{(\sigma_{meas})^2 + (\sigma_{ML}(t+1))^2 + (\sigma_{TS}(t))^2} \quad \text{Eq 30}$$

Now, considering the horizon of prediction, we define the new global uncertainty with the equation 30 with $\sigma_{hor} = \widehat{GHI}(t+h) \cdot \alpha(h)$.

$$\sigma_{tot}(t+h) = \sqrt{(\sigma_{meas})^2 + (\sigma_{ML}(t+1))^2 + (\sigma_{TS})^2 + (\sigma_{hor}(h))^2} \quad \text{Eq 31}$$

4 Results

All the previous uncertainty estimations will be computed for the 4 sites: Ajaccio, Bastia, Montpellier, Marseille.

In the end of this section we will be able to propose a global prediction interval for all predictions and locations.

4.1 Computing of σ_{meas}

Performance ratio or performance index calculations are more relevant when they are based on accurate independent data from a pyranometer than when they are based on a reference cell with lower accuracy and the same inherent flaws as the panel itself. The pyranometer used for the five meteorological stations measures global horizontal solar irradiance with 1% accuracy. In Eq 12, $\sigma(GHI)$ can be taken as $1\% \cdot \widehat{GHI}(t+1)$. The term $\sum_{i=1}^H \omega_i^2 \sum_{j=1}^n \omega_{ij}^1$ for the 5 sites and for 50 simulations (i.e. for 50 MLP configurations for each site), is (considering an average value with 95% confident interval) $\sum_{i=1}^H \omega_i^2 \sum_{j=1}^n \omega_{ij}^1 = 1.02 \mp 0.3$. For each site the results are very close and contribute a little bit to the global uncertainty. Note that the number of inputs or of hidden nodes don't modify (or weakly) the result concerning $\sigma_{meas}(\widehat{GHI}(t+1))$. The uncertainty related to the error measurement is calculated using Eq 18 and we obtained the following result:

$$\sigma_{meas}(\widehat{GHI}(t+1)) = 1.02\% \cdot \widehat{GHI}(t+1) \approx 1\% \cdot \widehat{GHI}(t+1) \quad \text{Eq 3}$$

As for all the sites, the value of σ_{meas} is consistent, in the next we consider only the equation 31 for each location.

4.2 Computing of σ_{var} and σ_{inh}

In a previous study realized for three insular sites (Corsica, Guadeloupe and Reunion) [50], it has been shown that the absolute log-return was correlated with the forecasting error obtained with a MLP prediction in Ajaccio as described in the figure 1.

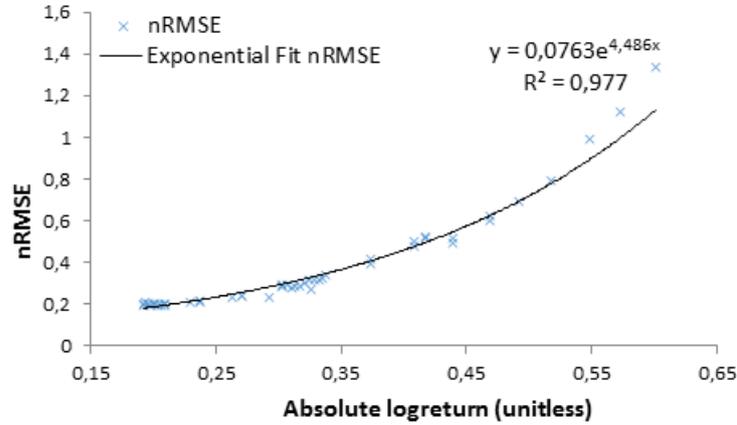


Figure 1. Link between volatility (absolute log return) and prediction error (nRMSE) for Ajaccio

This relation allows to estimate for Ajaccio, the σ_{var} parameter considering Eq 22. We obtain that for this site an exponential form of $\sigma_{var}(t)$:

$$\sigma_{var}(t) = \widehat{GHI}(t+1) * 0.0763 \exp(4.486 \cdot vol2(t)) \quad \text{Eq 32}$$

For the other locations similar equations are obtained, but in the next, concerning the intrinsic variability of the series, we choose to consider σ_{inh} because of the previous exponential fit, very large values of $\sigma_{var}(t)$ are observed when very quickly cloud occurrences are observed. The prediction band becomes in this case so large and inconsistent and we prefer in this example consider only σ_{inh} .

σ_{inh} estimates the uncertainty related to the studied time series and based on the definition of the $GHI_{trend}(t)$ (Eq 20), the values $nRMSE(GHI_{trend}(t) - GHI(t))$ for each station are shown in the table 1 .

Site	σ_{inh} in %
Ajaccio	13.98
Bastia	14.82
Marseille	10.13
Montpellier	12.28
Average	12.80

Table 1. Values of the σ_{inh} uncertainty coefficient

The value of σ_{inh} could be approximated for all locations by $\sigma_{inh} = 12.2\% * \widehat{GHI}(t+1)$, but in order to customize the prediction band it is necessary choose value related to each site.

4.3 Computing of σ_{ML} (σ_{samp} and σ_{ini})

The error components σ_{samp} and σ_{ini} are computed together. For each site, we compute 50 times the weights learning phase and also we generate 50 different training sets (see k-fold methodology in section 3.4).

SITE	$\sqrt{\sigma_{samp}^2 + \sigma_{ini}^2}$
AJACCIO	55.8
BASTIA	75.8
MARSEILLE	72.6
MONTPELLIER	57.4

Table 2. Value of $\sqrt{\sigma_{samp}^2 + \sigma_{ini}^2}$ for all the studied sites (in Wh/m²)

The values of the uncertainty are different for each sites and must be considered separately, for all cities the parameter $\sqrt{\sigma_{samp}^2 + \sigma_{ini}^2}$. In the figure 2 the profile of GHI and $\sqrt{\sigma_{samp}^2 + \sigma_{ini}^2}$ is shown.

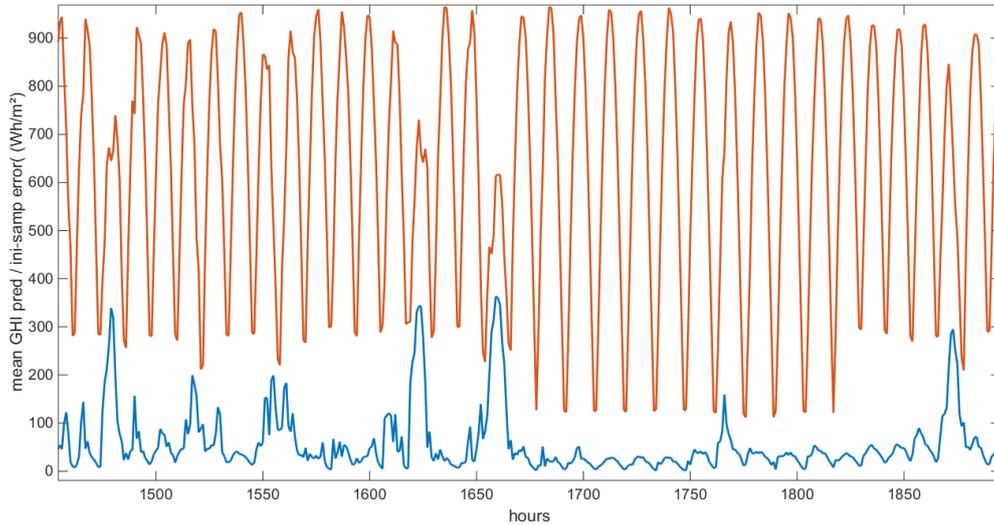


Figure 2. GHI measured and uncertainties related to initialization and sampling for Ajaccio

We can see that when there is no cloud occurrence this parameter is very low but can reach 350Wh/m² for other conditions. In Fig. 3, the maximum and minimum output values predicted are reported for 50 trained MLP in summer for Ajaccio and compared to the measure.

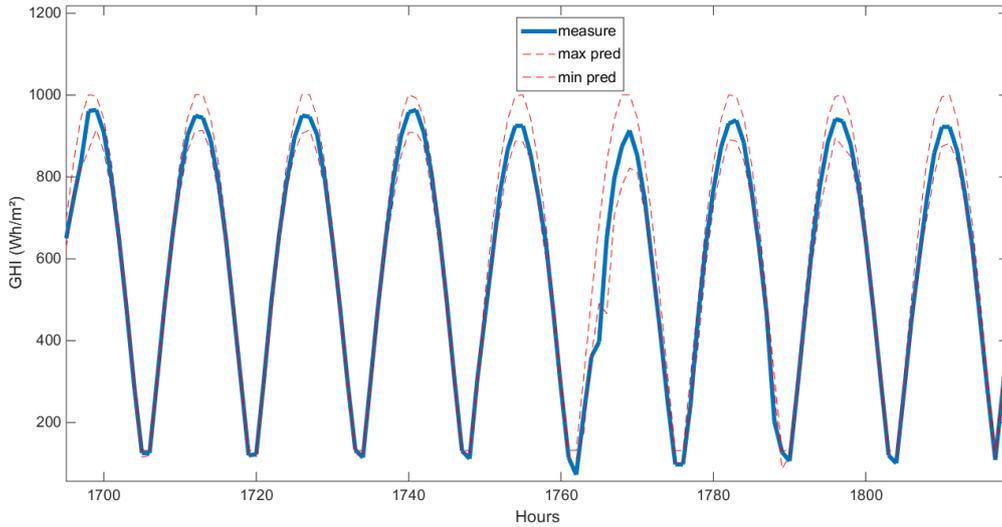


Figure 3. Representation of the sampling and initialization uncertainties on the GHI prediction for Ajaccio

We have ran 50 simulations to take into account the initialization and the sampling effects and thus consider the upper and lower output for each step. The prediction becomes the average of all the outputs, the max and min value of prediction (related to the 50 trained MLP) are represented by the dashed lines.

4.4 Computing of σ_{hor}

The estimation of this error component depends on the chosen time horizon as described in the section 3.6. The impact of the horizon on the $\alpha(t)$ (equation 23) is shown in figure 4.

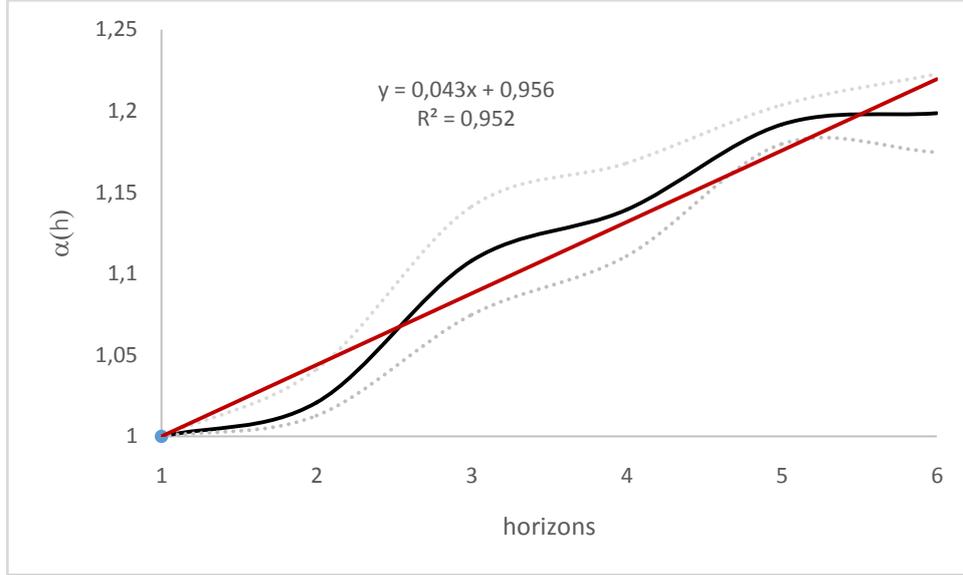


Figure 4. Correlation between horizon (in hour) and $\alpha(t)$

Dash lines represent max and min values for all stations while black curve represent the average and red line represent the linear fit. From this equation 23 we can model the horizon uncertainty adding the term $\sigma_{hor}(h)$ to the global error of prediction (linear fit).

$$\sigma_{hor}(h) = \widehat{GHI}(t+h)(0.0439 * h + 0.9561) \quad \text{Eq 33}$$

In the next, for each site, we will use to $\alpha(t)$ computed with the data related to the studied site.

4.5 Computing of σ_{tot}

The previous components allows to calculate the global uncertainty and to propose two prediction bands: UB for upper band and LB lower band[46]. Thus, the quality of the prediction can be defined by the triplet $\{\widehat{GHI}(t+h); LB; UB\}$ [44]. We can also estimate the reliability of the prediction considering that the prediction is efficient when $UB-LB$ is very lower than $\widehat{GHI}(t+h)$ and inefficient when $UB-LB$ is equal or upper to $\widehat{GHI}(t+h)$ value. From this hypothesis, we can define the reliability η as $\eta(t+1) = 1 - (UB(t+1) - LB(t+1)) / \widehat{GHI}(t+h)$. Lower is this parameter, more efficient is the prediction. We construct a reliability index between 0 and 1 considering that if $(UB(t+1) - LB(t+1)) / \widehat{GHI}(t+h) > 1$, then $\eta(t+1) = 0$, i.e. the prediction is not sure. The final prediction becomes:

$\langle \widehat{GHI}(t+h) \rangle = \langle \widehat{GHI}(t+h) \rangle$, average of 50 simulations (50 training and initialization weights, 50 different training sets)

$$-LB = \sqrt{(\sigma_{meas})^2 + (\widehat{GHI}_{min}(t+h) - \langle \widehat{GHI}(t+h) \rangle)^2 + (\sigma_{inh})^2 + (\sigma_{hor}(h))^2} \quad \text{Eq 34}$$

$$-UB = \sqrt{(\sigma_{meas})^2 + (\widehat{GHI}_{max}(t+h) - \langle \widehat{GHI}(t+h) \rangle)^2 + (\sigma_{inh})^2 + (\sigma_{hor}(h))^2} \quad \text{Eq 35}$$

With $\sigma_{meas}(\widehat{GHI}(t+1)) = 1\% \cdot \widehat{GHI}(t+1)$, $\sigma_{inh} = \widehat{GHI}(t+1) \cdot nRMSE(GHI_{trend}(t) - GHI(t))$, $\sigma_{hor}(h) = \widehat{GHI}(t+h) \cdot \alpha(h)$ and $\widehat{GHI}_{min/max}(t+h)$ are the min and max values of the 50 predictions generated with 50 simulations. The figure 5 shows for Ajaccio an example of the prediction bands, considering all the kind of uncertainty with horizon $h=1$ hour. Line represents measurement and dashed lines the upper and lower bands concerning each kind of uncertainties.

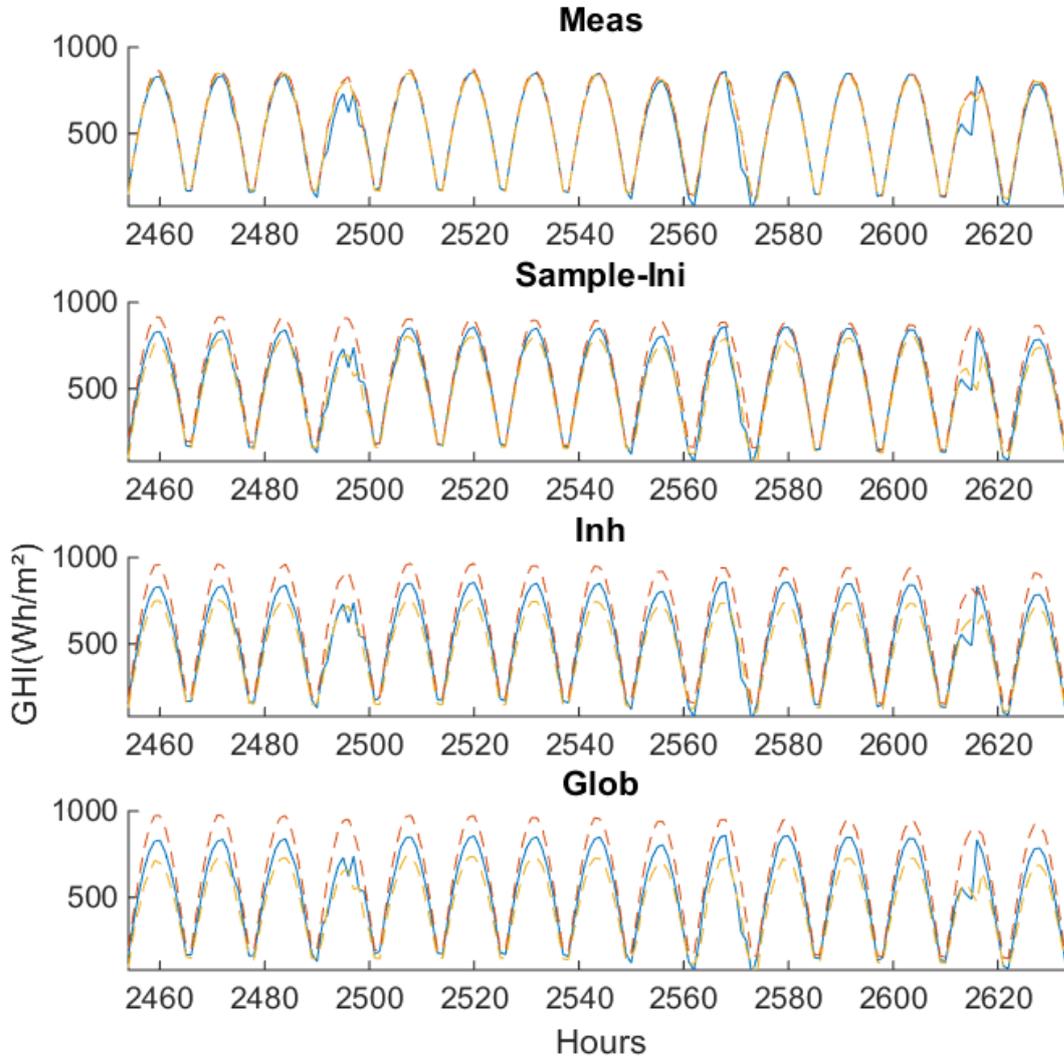


Figure 5. Uncertainty in the GHI predictions for the horizon $h=1$ for Ajaccio case

We can see that σ_{meas} is the parameter the less interesting for the bands construction and that it is necessary to consider the coupling of σ_{inh} and σ_{ML} (related to σ_{samp} and σ_{ini}) for a good prediction interval definition. For other sites the obtained curves are similar and no more information is observed. In the figure 6, the top curve (same prediction configuration that previously) compares the average prediction $\widehat{GHI}(t+h)$ (marks) versus the GHI measurement (line). The bottom curve shows the associated reliability index ($\eta(t+1)$).

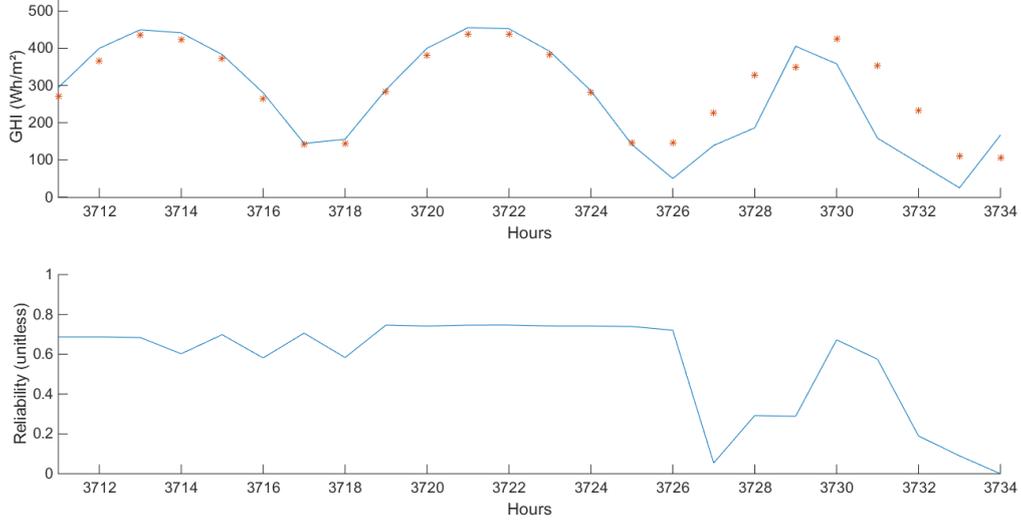


Figure 6. Comparison for the horizon $h=1$ for Ajaccio of GHI predictions (mark) and GHI measurement (line) on the top and associated reliability index in the bottom

We see that when the variability is low (two first day from 3711 to 3726) the reliability is important (close to 70%) but when cloud occurs the value is much lower and can reach 0%.

5 Conclusions

In this paper we have shown that it is possible to compute a prediction band in the context of global radiation time series forecasting using machine learning. We have defined for a popular machine learning technique, the multilayer perceptron, four kinds of uncertainties: the error due to the measurement, the variability of time series, the machine learning uncertainty (initialization and sampling) and the error related to the horizon. In literature, rarely both to the two kinds of uncertainty σ_{ini} and σ_{ML} are studied, and never σ_{meas} . We have also defined a reliability index which could be very interesting for the grid manager in order to estimate the validity of predictions. The described method has been successfully applied to four meteorological stations in Mediterranean area. In practice it is certainly not necessary to take into account all the proposed components. The ranking of the different uncertainty terms is: $\sigma_{meas} (\sim 1\%) < \sigma_{ML} (\sim 5\%) < \sigma_{inh} (\sim 10\%) < \sigma_{hor} (5 - 20\%)$. For

an one time step horizon, consider only σ_{ML} and σ_{inh} seems to be sufficient, for a larger horizon ($>3h$) it is essential to add the horizon component σ_{hor} . Note that for the predictors as SVM, ARMA, Gaussian process, regression tree (and derived) σ_{ML} is easily obtained computing only σ_{sample} . We are sure that it is possible to generalize the approach to other sites and other machine learning tools. Thereby in future, we will try to apply the methodology to other time granularities and predictor as SVM, regression tree or random forest. In order to increase the interest of the LB and UB methodology, it will be also possible to physically bounded the upper and lower bands using very performant clear sky modelling in order to obtain the global solar irradiation (maximum value of the band) and the diffuse part of the solar irradiation (minimum value of the band).

6 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.

7 References

- [1] Hoff TE, Perez R. Modeling PV fleet output variability. *Sol Energy* 2012;86:2177–89. doi:10.1016/j.solener.2011.11.005.
- [2] 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.
- [3] 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.
- [4] Mellit A, Kalogirou SA, Hontoria L, Shaari S. Artificial intelligence techniques for sizing photovoltaic systems: A review. *Renew Sustain Energy Rev* 2009;13:406–19. doi:10.1016/j.rser.2008.01.006.
- [5] Lorenz E, Remund J, Müller SC, Traunmüller W, Steinmaurer G, Pozo D, et al. Benchmarking of different approaches to forecast solar irradiance. 24th Eur. Photovolt. Sol. Energy Conf. Hambg. Ger., vol. 21, 2009, p. 25.
- [6] Join C, Fliess M, Voyant C, Chaxel F. Solar energy production: Short-term forecasting and risk management. ArXiv160206295 Cs Q-Fin 2016.
- [7] Bjön Wolff, Elke Lorenz, Oliver Kramer. Statistical Learning for Short-Term Photovoltaic Power Predictions (chapter) 2015:in print.

- [8] 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.
- [9] Abrahart RJ, See L. Neural Network vs. ARMA Modelling: constructing benchmark case studies of river flow prediction, 1998.
- [10] Mora-López L, Sidrach-de-Cardona M. Multiplicative ARMA models to generate hourly series of global irradiation. *Sol Energy* 1998;63:283–91. doi:10.1016/S0038-092X(98)00078-4.
- [11] 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.
- [12] Lauret P, Fock E, Randrianarivony RN, Manicom-Ramsamy JF. Bayesian neural network approach to short time load forecasting. *Energy Convers Manag* 2008;49:1156–1166.
- [13] Pole A, West M, Harrison J. Applied Bayesian forecasting and time series analysis. New York: Chapman and Hall; 1994.
- [14] Logofet D. The mathematics of Markov models: what Markov chains can really predict in forest successions. *Ecol Model* 2000;126:285–98. doi:10.1016/S0304-3800(00)00269-6.
- [15] Kumar U, Jain VK. Time series models (Grey-Markov, Grey Model with rolling mechanism and singular spectrum analysis) to forecast energy consumption in India. *Energy* 2010;35:1709–16. doi:10.1016/j.energy.2009.12.021.
- [16] 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.
- [17] Krömer P, Musilek P, Pelikán E, Krč P, Juruš P, Eben K. Support Vector Regression of multiple predictive models of downward short-wave radiation. 2014 Int. Jt. Conf. Neural Netw. IJCNN, 2014, p. 651–7. doi:10.1109/IJCNN.2014.6889812.
- [18] Tso GKF, Yau KKW. Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks. *Energy* 2007;32:1761–8. doi:10.1016/j.energy.2006.11.010.
- [19] Lahouar A, Ben Hadj Slama J. Day-ahead load forecast using random forest and expert input selection. *Energy Convers Manag* 2015;103:1040–51. doi:10.1016/j.enconman.2015.07.041.
- [20] Mellit A, Kalogirou SA, Hontoria L, Shaari S. Artificial intelligence techniques for sizing photovoltaic systems: A review. *Renew Sustain Energy Rev* 2009;13:406–19. doi:10.1016/j.rser.2008.01.006.
- [21] 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.
- [22] Aler R, Martín R, Valls JM, Galván IM. A Study of Machine Learning Techniques for Daily Solar Energy Forecasting Using Numerical Weather Models. In: Camacho D, Braubach L, Venticinque S, Badica C, editors. *Intell. Distrib. Comput. VIII*, Springer International Publishing; 2015, p. 269–78.
- [23] Costa MA, Braga A de P, Menezes BR de. Improving generalization of MLPs with sliding mode control and the Levenberg-Marquardt algorithm. *Neurocomputing* 2007;70:1342–7. doi:10.1016/j.neucom.2006.09.003.
- [24] Ahlburg D. Error measures and the choice of a forecast method. *Int J Forecast* 1992;8:99–100. doi:10.1016/0169-2070(92)90010-7.
- [25] Voyant C, Muselli M, Paoli C, Nivet M-L. Optimization of an artificial neural network dedicated to the multivariate forecasting of daily global radiation. *Energy* 2011;36:348–59. doi:10.1016/j.energy.2010.10.032.
- [26] Hornik K, Stinchcombe M, White H. Multilayer feedforward networks are universal approximators. *Neural Netw* 1989;2:359–66. doi:10.1016/0893-6080(89)90020-8.
- [27] Cybenko G. Approximation by superpositions of a sigmoidal function. *Math Control Signals Syst* 1989;2:303–14. doi:10.1007/BF02551274.
- [28] Voyant C, Muselli M, Paoli C, Nivet M-L, Poggi P, Haurant P. Predictability of PV power grid performance on insular sites without weather stations : use of artificial neural networks, Hambourg: 2009.
- [29] Voyant C, Muselli M, Paoli C, Nivet M-L. Hybrid methodology for hourly global radiation forecasting in Mediterranean area. *Renew Energy* 2013;53:1–11. doi:10.1016/j.renene.2012.10.049.
- [30] Paoli C, Voyant C, Muselli M, Nivet M-L. Solar Radiation Forecasting Using Ad-Hoc Time Series Preprocessing and Neural Networks. *Emerg. Intell. Comput. Technol. Appl.*, vol. 5754, Springer Berlin / Heidelberg; 2009, p. 898–907.

- [31] Paoli C, Voyant C, Muselli M, Nivet M-L. Use of Exogenous Data to Improve Artificial Networks Dedicated to Daily Global Radiation Forecasting, Valencia, Spain: 2010.
- [32] 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.
- [33] 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.
- [34] 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.
- [35] Kalogirou S. Artificial neural networks in renewable energy systems applications: a review. *Renew Sustain Energy Rev* 2001;5:373–401. doi:10.1016/S1364-0321(01)00006-5.
- [36] Diazrobles L, Ortega J, Fu J, Reed G, Chow J, Watson J, et al. A hybrid ARIMA and artificial neural networks model to forecast particulate matter in urban areas: The case of Temuco, Chile. *Atmos Environ* 2008;42:8331–40. doi:10.1016/j.atmosenv.2008.07.020.
- [37] Zhang G. Forecasting with artificial neural networks: The state of the art. *Int J Forecast* 1998;14:35–62. doi:10.1016/S0169-2070(97)00044-7.
- [38] Crone SF. Stepwise Selection of Artificial Neural Networks Models for Time Series Prediction 2005;15.
- [39] Huang D, Chow TWS. Effective feature selection scheme using mutual information. *Neurocomputing* 2005;63:325–43. doi:10.1016/j.neucom.2004.01.194.
- [40] Jiang A-H, Huang X-C, Zhang Z-H, Li J, Zhang Z-Y, Hua H-X. Mutual information algorithms. *Mech Syst Signal Process* 2010;24:2947–60. doi:10.1016/j.ymsp.2010.05.015.
- [41] Badescu V. Modeling solar radiation at the earth's surface: recent advances. Springer; 2008.
- [42] Paulescu M, Paulescu E, Gravila P, Badescu V. Weather Modeling and Forecasting of PV Systems Operation. London: Springer London; 2013.
- [43] Hopson TM. Assessing the Ensemble Spread–Error Relationship. *Mon Weather Rev* 2014;142:1125–42. doi:10.1175/MWR-D-12-00111.1.
- [44] 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.
- [45] 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.
- [46] McAleer M, Medeiros MC. Realized Volatility: A Review. *Econom Rev* 2008;27:10–45. doi:10.1080/07474930701853509.
- [47] Chen B, Gel YR, Balakrishna N, Abraham B. Computationally efficient bootstrap prediction intervals for returns and volatilities in ARCH and GARCH processes. *J Forecast* 2011;30:51–71. doi:10.1002/for.1197.
- [48] Voyant C, Soubdhan T, Lauret P, David M, Muselli M. Statistical parameters as a means to a priori assess the accuracy of solar forecasting models. *Energy* 2015;90:671–679.
- [49] Wilks DS, Wilks DS. Statistical Methods in the Atmospheric Sciences An Introduction. Burlington: Elsevier Science; 2014.
- [50] Divya U, Pasupathi C. Survey on Machine Learning Approaches for Solar Irradiation Prediction. *Int J Eng Sci Res Technol* 2014;3:478–82.
- [51] Gneiting T, Katzfuss M. Probabilistic Forecasting. *Annu Rev Stat Its Appl* 2014;1:125–51. doi:10.1146/annurev-statistics-062713-085831.
- [52] Wiens TS, Dale BC, Boyce MS, Kershaw GP. Three way k-fold cross-validation of resource selection functions. *Ecol Model* 2008;212:244–55. doi:10.1016/j.ecolmodel.2007.10.005.
- [53] 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.
- [54] Hu Y, Hwang J-N. Handbook of neural network signal processing. Boca Raton: CRC Press; 2002.
- [55] Ding R, Li J. Nonlinear finite-time Lyapunov exponent and predictability. *Phys Lett A* 2007;364:396–400. doi:10.1016/j.physleta.2006.11.094.
- [56] David M. Probabilistic forecasting of the solar irradiance with recursive ARMA and GARCH models n.d.
- [57] Trapero JR. Calculation of solar irradiation prediction intervals combining volatility and kernel density estimates. *Energy* 2016;114:266–274.

8 Tables

Table 1. Values of the σ_{inh} uncertainty coefficient

Table 2. Value of $\sqrt{\sigma_{samp}^2 + \sigma_{ini}^2}$ for all the studied sites (in Wh/m²)

9 Figures

Figure 1. Link between volatility (absolute log return) and prediction error (nRMSE) for Ajaccio

Figure 2. GHI measured and uncertainties related to initialization and sampling for Ajaccio

Figure 3. Representation of the sampling and initialization uncertainties on the GHI prediction

Figure 4. Correlation between horizon (in hour) and $\alpha(t)$

Figure 5. Uncertainty in the GHI predictions for the horizon h=1 for Ajaccio

Figure 6. Comparison for the horizon h=1 for Ajaccio of GHI predictions (mark) and GHI measurement (line) on the top and associated reliability index in the bottom