Renewable and Sustainable Energy Reviews 92 (2018) 343–352

Contents lists available at ScienceDirect

Renewable and Sustainable Energy Reviews

journal homepage: www.elsevier.com/locate/rser

Solar irradiation nowcasting by stochastic persistence: A new parsimonious, simple and efficient forecasting tool

Cyril Voyant[a,b,*], Gilles Notton[a]

[a] University of Corsica/CNRS UMR SPE 6134, Campus Grimaldi, 20250 Corte, France
[b] CHD Castelluccio, Radiophysics Unit, B.P85, 20177 Ajaccio, France

ARTICLE INFO

Keywords:
Prediction
Machine learning
Forecasting
Persistence
Bias-variance

ABSTRACT

Simple, naïve, smart or clearness persistences are tools largely used as naïve predictors for the global solar irradiation forecasting. It is essential to compare the performances of sophisticated prediction approaches with that of a reference approach generally a naïve methods. In this paper, a new kind of naïve "nowcaster" is developed, a persistence model based on the stochastic aspect of measured solar energy signal denoted stochastic persistence and constructed without needing a large collection of historical data. Two versions are proposed: one based on an additive and one on a multiplicative scheme; a theoretical description and an experimental validation based on measurements realized in Ajaccio (France) and Tilos (Greece) are exposed. The results show that this approach is efficient, easy to implement and does not need historical data as the machine learning methods usually employed. This new solar irradiation predictor could become an interesting tool and become a new member of the solar forecasting family.

1. Introduction

1.1. Interest of solar irradiation forecasting

Over the last ten years, energy market was boosted with the advent of renewable energies and in particular thanks to solar energy. The main interest of this kind of primary energy is to be easily and cleanly transformed into electricity particularly via photovoltaic conversion [1], which is the most flexible form of energy [2]. The main problem concerning the use of solar energy is its continuous variability relating both to time and space [3,4]. The variability can be divided into two components, the first one denoted deterministic part and the second one stochastic or random part. If the deterministic component is generated by the movements of rotation and revolution of the Earth [5], the stochastic component is generated by weather and cloud occurrences [6]. Solar energy intermittency has a great influence on the output power of photovoltaic (*PV*) plants, which can fluctuate significantly in short intervals (related to the random part) and in long intervals (related to daily and yearly seasonal effects) [7]. This no-controllable intermittence has negative consequences on the management of the electrical distribution and stability (forcing to limit the penetration rate of such intermittent energy systems) and on the kWh production costs [8]. One way to solve or to reduce this problem is to forecast this PV output power [9]. A good forecast helps the grid manager to plan the other energy capabilities to compensate for the *PV* plants power variations [10]. The forecasting quality of the output PV plant is strongly linked to the global horizontal irradiation (*GHI*) forecasting accuracy [11]. Some authors go even further and consider the problem of *PV* output power forecasting and the solar irradiance forecasting problem as equal [12]. In this paper, a new forecasting tool is developed and tested in view to assist the electrical grid manager by predicting easily *GHI*.

1.2. Prediction and parsimony

Time series forecasting [13] consists to estimate possible events or their evolutions by using as tools the past and the present. Before exposing the deferent tools available in order to nowcast *GHI*, it is important to define the "time series" term and the word "prediction" related to this kind of mathematical tools [14,15].

Definition 1.1.. Time series: A univariate time series is a sequence of measurements of the same variable collected over time. Most often, the measurements are made at regular time intervals. The common notation concerning a time series of *GHI* measurement is $GHI = \{GHI(t): t \in T\}$ where T is the index set.

Remark 1.1.. *GHI* (nondeterministic) time series may be analyzed by assuming they are partly the manifestations of stochastic (random)

* Corresponding author at: University of Corsica/CNRS UMR SPE 6134, Campus Grimaldi, 20250 Corte, France.
E-mail address: voyant@univ-corse.fr (C. Voyant).

https://doi.org/10.1016/j.rser.2018.04.116
Received 5 October 2017; Received in revised form 22 April 2018; Accepted 28 April 2018

processes [16–18] which is a statistical phenomenon consisting of a collection of random variables ordered in time and evolving according to a priori unknown probabilistic laws.

To succeed a time series prediction, only four conditions must be fulfilled:

- a certain regularity in the functioning of the studied process,
- this regularity must provide information on the future,
- the method chosen to establish the prediction captures a part of this regularity,
- the prediction will be efficient if and only if the "**noise**" or past irregularities are excluded as far as possible.

Forecasting the solar irradiation from 1 h to 6 h (defining the nowcasting [12]) is currently done using statistical or machine learning methods coupled to time series analysis. Many papers show that these methods yield similar results [19–21], none appears to outperform other and sometimes simple methods propose very similar results. According to a review analysis [22], it seems that it is not interesting to predict with very complex methods because a gain of tenths of a percent on the forecasting performances has only a small (but not negligible) impact on the grid management. Moreover, in [8], authors model a fictive solar plant with a nominal capacity of 1000 kW and show that a large nRMSE reduction from 32% to 28% (−4% points) allows a financial saving close to 9%, so 70€ per day for the considered installation. In fact, the electricity grid operator needs a reliable tool which is adaptable for all horizons (between 5 min and 6 h). The ideal case is to elaborate a tool which does not require a large learning history [15] in order to be quickly deployed on any site. In this paper we propose a new very simple and parsimonious tool based on the persistence of stochastic signal. Note that if in the operational case, the prediction with persistence does not need a large historical data (only a few hours), the present study is a retrospective comparison and is operated with historical data. The idea behind parsimonious models stems from the 14th century and the formulation of the Occam's razor [23] stating that “we should use no more parameters than necessary to explain the model well. There is generally a tradeoff between goodness of fit and parsimony. Models with many parameters (as machine learning tools [24,25]) tend to have a better fit than high parsimony models (as persistence), however this is not usually a good thing. Indeed, adding more parameters usually results in a good model fit for the data at hand, but that same model will likely be useless for predicting other data sets. In [26] (pp. 103–104), sentences summarize the interpretation related to simple models results: «Sometimes a simple model will outperform a more complex model … Nevertheless, I believe that deliberately limiting the complexity of the model is not fruitful when the problem is evidently complex. Instead, if a simple model is found that outperforms some particular complex model, the appropriate response is to define a different complex model that captures whatever aspect of the problem led to the simple model performing well. It is essential to correctly study the simple models before to elaborate more sophistical approaches. Reference models should be well chosen to truly and objectively decide on the quality of the forecast.

2. Machine learning or simple models of persistence

Machine learning [27] is a branch of artificial intelligence [28]. It concerns the construction and the study of systems that can learn from data sets, giving to computers the ability to learn without being explicitly programmed.

2.1. Models definitions

With the machine learning tools based predictions, the system is built from a random output (denoted variable y) and a set of random input (denoted variables $x = \{x_1 \ldots x_n\}$). Using a learning sample $\{y_i, x_i\}_1^N$ of known values of pairs (y,x), the aim is to obtain and estimate a model function $f^*(x)$, among all the functions $f(x)$ available and which allows to map (as well as possible!) x to y. The objective is reached after an optimization of the expected value ($\mathbb{E}$) of some specified loss functions $L(y, f(x))$ over the joint distribution of all (y,x) pairs:

$$f^*(x) = \underset{f}{argmin}(\mathbb{E}(L(y, f(x))) \tag{1}$$

In a regression problem, the loss function $L(y, f(x))$ includes usually 2-norm or 1-norm distances respectively computed from the squared-error $(y - f(x))^2$ (Euclidean norm giving more importance to large deviations or outliers) and the absolute error $|y - f(x)|$ (absolute-value norm giving importance to the trend gap). Typically in the supervised cases, the machine learning methods are confronted to bias-variance tradeoff and are very user dependent and difficult to make a good use [29]. Is machine learning is overhyped? This question was recently asked in [30], it may be time to consider other methods of modeling. The simplest method of forecasting the weather, persistence, relies upon today's conditions to forecast the conditions tomorrow. This can be a valid way of forecasting the weather when it is in a steady state, such as during the summer season when clouds are rare. This method of forecasting strongly depends upon the presence of a stagnant weather pattern. Therefore, with a fluctuating weather pattern, this method of forecasting becomes inaccurate. It can be useful in both short range forecasts and long range forecasts. The time series of global horizontal irradiation (*GHI*) is composed by a stochastic part (Cf previous section); often when a machine learning method is used, a strong condition is necessary: the stationarity of the input data [31]. That means that the joint distribution of $GHI(t)$ and $GHI(t + h)$ does not depend on t but only on h ($t,h \in \mathbb{N}^*$). To our knowledge, it is not proved that the tools used to make the *GHI* time series stationary allows to correctly respect this condition [32]. It is legitimate to ask: can we really use these methods even if the results are consistent? we have of course not the answer and we would be very embarrassed to answer "no" to this question given that we ourselves abundantly study the forecast of *GHI* via the data driven, machine learning, artificial intelligence and others statistical methods. What is sure is that with the persistence there are both advantages: directly usable (without learning and without need of historical data) and any hypotheses or conditions concerning the model building. The “classical” persistence is not really adapted to the forecast [29] while the smart persistence (integrating a knowledge-based model using a clear sky model taking into account the sun position and the average conditions of sky state) allows to greatly improve the prediction [19].

Definition 2.1.. Simple persistence: the term persistence (or simple persistence) in time series context is related to the notion of memory properties of time series, the model is built for the horizon (look-ahead time) h as , where t is a time index and ε denotes the residual. The forecast $\hat{GHI}$ obtained with this model is $\hat{GHI}(t + h) = GHI(t)$, which states that the expected value at horizon h is equal to the most recent measured value.

Definition 2.2.. Smart persistence: This model is based on the same assumption than persistence model but is corrected for the deterministic diurnal variation in solar irradiance, using a knowledge-based model $K(t, h)$: $\hat{GHI}(t + h) = GHI(t).K(t, h)$

2.2. A short literature review on persistence

Numerous studies show the efficiency of these naïve predictions: the persistence. In [33] the persistence is extremely detailed and authors wrote "It has been found that for short time horizons, beating persistence models is a difficult task " and demonstrated that, often, the persistence is the best method to use for the short-casting (< 1 h) and the now-casting (1–6 h). In several studies, the simple persistence