Solar Energy 174 (2018) 121–129

Contents lists available at ScienceDirect

Solar Energy

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

Periodic autoregressive forecasting of global solar irradiation without knowledge-based model implementation

Cyril Voyant[a,c,*], Jan G. De Gooijer[b], Gilles Notton[c]

[a] University of Reunion Island – PIMENT Laboratory, 15, Avenue René Cassin, BP, 97715 Saint-Denis Cedex, France
[b] University of Amsterdam, 1018 WB Amsterdam, The Netherlands
[c] Hospital of Castelluccio, Radiotherapy Unit, BP 85, 20177 Ajaccio, France

ARTICLE INFO

Keywords:
Autoregression
Clear sky irradiance
Forecast intervals
Periodic

ABSTRACT

Reliable forecasting methods increase the integration level of stochastic production and reduce cost of intermittence of photovoltaic production. This paper proposes a solar forecasting model for short time horizons, i.e. one to six hours ahead. In this time-range, machine learning methods have proven their efficiency. But their application requires that the solar irradiation time series is stationary which can be realized by calculating the clear sky global horizontal solar irradiance index (CSI), depending on certain meteorological parameters. This step is delicate and often generates additional uncertainty if conditions underlying the calculation of the CSI are not well-defined and/or unknown. As a novel alternative, we introduce a so-called periodic autoregressive (PAR) model. We discuss the computation of post-sample point forecasts and forecast intervals. We show the forecasting accuracy of the model via a real data set, i.e., the global horizontal solar irradiation (GHI) measured at two meteorological stations located at Corsica Island, France. In particular, and as opposed to methods based on CSI, a PAR model helps to improve forecast accuracy, especially for short forecast horizons. In all the cases, PAR is more appropriate than persistence, and smart persistence. Moreover, smart persistence based on the typical meteorological year gives more reliable results than when based on CSI.

1. Introduction

Solar energy, mainly photovoltaic, is an energy resource which plays an increasingly important role in the electrical energy production due to its abundance, cleanness and cost effectiveness characteristics with limited environmental consequences. On the other hand, solar power has a fluctuating generation profile because of its inherent cyclic and time varying nature, leading to limitations on stability and trustworthiness of solar power grid systems (Shamshirband et al., 2015). To reduce the inconvenience of this stochastic and intermittent nature, and to improve the inclusion of solar power plants, an efficient forecasting method of solar radiation is paramount. Moreover, this intermittent character gives rise to additional production costs compared with conventional production, from 1 to 8€/MWh with an average value around 6€/MWh (Notton et al., 2018). Thus, a reliable production forecasting method decreases the average annual operating costs. In addition, it reduces the reserve shortfalls and it reduces curtailments. Several methods are available for forecasting depending on the time horizon and time resolution (Notton and Voyant, 2018).

This paper concerns forecasting at short time horizons, i.e., one to six hours ahead with a one hour time step. In this time-range, machine learning methods have proven their accuracy. But their application requires that a solar irradiation time series is stationary which can be realized by calculating conditions for clear sky (CS) solar irradiation, depending on certain meteorological parameters (Diagne et al., 2013; Lauret et al., 2015). The use of a CS solar radiation model, however, induces an important source of error because this type of model depends on meteorological parameters which vary month by month or during a day. To avoid this difficulty, the purpose of the paper is to present a forecasting model which does not require a CS model, and which can be easily implemented in practice.

The remainder of the paper is organized as follows. In Section 2, we introduce the concept of periodically correlated processes and we provide arguments why global horizontal solar irradiation measurements are periodic seasonal time series. In Section 3, we discuss problems induced by a CS model. Section 4 provides details about the data under study. The periodic autoregressive (PAR) model is introduced in Section 5, and expressions for point forecasts, forecast intervals, and forecast evaluation measures are given. Section 6 provides some information about alternative forecasting models. Section 7 shows PAR

* Corresponding author at: University of Reunion Island – PIMENT Laboratory, 15, Avenue René Cassin, BP, 97715 Saint-Denis Cedex, France.
E-mail address: cyrilvoyant@gmail.com (C. Voyant).

https://doi.org/10.1016/j.solener.2018.08.076
Received 8 May 2018; Received in revised form 7 August 2018; Accepted 27 August 2018

identification and PAR forecasting results. It includes results of a comparative forecasting experiment. Lastly, Section 8 offers some concluding remarks.

2. Periodic phenomena

Consider a time series process $\{Y_t, t \in \mathbb{Z}\}$ whose second moments exist. The process is said to be periodically correlated (PC) with period H, or periodic covariance stationary (Gladhyšhev, 1961; Pagano, 1978), if the following two conditions

$$\mu_t \equiv \mathbb{E}(Y_t) = \mu_{t+h} \text{ for all } t \in \mathbb{Z}, \text{ and}$$
$$\gamma_{s,t} \equiv \mathbb{E}[(Y_s - \mu_s)(Y_t - \mu_t)] = \gamma_{s+h,t+h} \text{ for every } s, t \in \mathbb{Z} \quad (1)$$

are true for $h = H$ but for no smaller value of h. In other words, the mean and autocovariance functions are periodic functions of time with period h: that is, the first and second order moments of the process do depend on the period and the lag, but not on the absolute time. The periodic autocorrelation function at lag $s = 1, 2, \ldots$ and time t is defined by $\rho_{s,t} = \gamma_{s,t}/\gamma_{0,t}$.

Global horizontal solar irradiation (GHI, in W/m^2) measurements can be viewed as periodic seasonal time series. In general, a seasonal pattern appears when a time series is influenced by seasonal factors, e.g., the month of the year, the day of the week, or the hour of the day (Hokoi et al., 1990). As can be seen from Eq. (1), the seasonality is always of a fixed and known period, and hence, the time series is called periodic. In general, the average length of cycles is longer than the length of a seasonal pattern, and the magnitude of cycles tends to be more variable than the magnitude of seasonal patterns (Franses and Paap, 1994). From these observations, we deduce the following two properties.

(1) The observed time series $\mathrm{GHI}(t)$ $(t = 1, \ldots, N)$ can be considered as a periodic time series with two fixed seasonal periods H and D. In this study $H = 24$ h (h) and $D = 365$ days (d). For $H = 24$, this condition is formally tested in Section 7.1 using p-values of the sample periodic autocorrelation function.
For simplicity, we assume that $N/(H \times D) = Y$ is an integer representing the number of available years. $\mathrm{GHI}(t)$ can be decomposed into three new time series: two are strongly seasonal, and one time series is related to the noise, or irregular component. That is

$$\{\mathrm{GHI}(t), t \in \mathbb{Z}\} = \{f(S_{24h}(t), S_{365d}(t), \varepsilon(t)), t \in \mathbb{Z}\}. \quad (2)$$

2) The function $f(\cdot)$ defines the type of decomposition: additive, multiplicative or hybrid. Usually the multiplicative mode is preferred, and the term $S_{24h} \times S_{365d}$ at time t is a proxy of the so-called CS global irradiation, i.e., $\mathrm{CS}(t) = \{S_{24h}(t) \times S_{365d}(t), t \in \mathbb{Z}\}$.

The ratio $\mathrm{GHI}(t)/\mathrm{CS}(t)$ defines the clear sky index $\mathrm{CSI}(t) \in [0, 1]$. Note, however, that for a cloudy period CSI can be greater than 1 due to cloud enhancement.

Observe from property (1) that a solar irradiation time series contains only seasonal patterns. These components can be deleted by seasonal adjustment using a ratio to trend (detrending), divided by an estimate of a $\mathrm{CS}(t)$ series (Grantham et al., 2016, 2018) or, if estimation is difficult, divided by a moving average of the series (Voyant et al., 2011). Alternatively, one can adopt a classical seasonal autoregressive integrated moving average (SARIMA) model. Implicit in such models is the assumption of homogeneity or time invariance, i.e. the seasonally differenced series is sure to become stationary. However, many seasonal time series cannot be filtered, standardized or differenced to achieve second-order stationarity because the series exhibits a strong seasonal behavior such that the entire autocorrelation structure of the series depends on the season, hence the homogeneity assumption fails (Tiao and Grupe, 1980). In fact, the majority of GHI time series satisfy the property of periodic stationarity (Ula and Smadi, 1997), stating that their sample mean and sample autocorrelation function are periodic with respect to time. A more realistic family of models characterizing those kind of seasonal time series is the PAR model.

The method of moments based on the well-known Yule-Walker equations and the least squares method in the univariate case are both efficient ways to estimate PAR models. However, the number of estimated parameters is likely to increase with the choice of the season. Thus, in our study it will be easier to consider only the $H = 24$ h period, giving rise to a parsimonious PAR model with only 24 components rather than estimating a model with $D = 365$ components. Moreover, it is often useful to put linear constraints on the parameters for a given season.

In the next section, we will focus on two approaches to take seasonality into account. The first approach uses a white box model (WM) based on the knowledge model which we couple with the stochastic modeling of $\mathrm{CSI}(t)$. This approach is often called grey box modeling, or in short-hand notation GM. The second approach uses the previously measured data and any knowledge-based model, and we call it a black box model or BM. Observe that a GM (= WM + BM) is often more interesting to analyze than a BM since it encompasses a semi-physical model. But adopting the GM can add an additional uncertainty if the parameters of the model are not well-defined, and thus decreasing the reliability of the GM.

3. Clear sky (CS) model

For a temporal forecast horizon up to and including six hours ahead, a CS solar irradiation model is often used to make the time series $\mathrm{GHI}(t)$ stationary, before calculating the $\mathrm{CS}(t)$ index (Lauret et al., 2015; Voyant et al., 2015). The chosen CS model in this study is the Solis model (Mueller et al., 2004; Ineichen, 2008). The CS global horizontal irradiance reaching the ground is defined by

$$\mathrm{CS}(t) = I_0(t)\exp\left(\frac{-\tau}{\sin^g(\eta(t))}\right)\sin(\eta(t)). \quad (3)$$

Here I_0 is the extraterrestrial radiation (depends on the day of the year), η is the solar elevation angle (depends on the hour of the day), τ is the global total atmospheric optical depth (depends on the day of the year and the hour of the day), and g is a fitting parameter. In order to be well-defined, the CS model requires meteorological parameters (Gelaro et al., 2017) to characterize the state of the sky such as, for instance, the aerosol optical depth (AOD) and the water vapor column defining the total AOD. These parameters are difficult to obtain. Moreover, they fluctuate in a large range from one year to another and during a day from one hour to the next. Thus, the average value of these parameters do not accurately reflect the CS condition at a given time t. Indeed Voyant et al. (2015, Fig. 5), showed the impact of AOD values on the forecast accuracy, as measured by the normalized mean absolute forecasting error (nMAE) for Ajaccio. Specifically, these authors obtain an nMAE value of 11% in the optimized parameter case, and 18% with very ill-optimized parameters, so an increase of 7 percentage points.

Moreover, obtaining accurate $\mathrm{CSI}(t)$ series at sunset and sunrise is difficult due to a possible surrounding masking effect such as mountains, buildings, or vegetation. It may also be due to unreliable measurements of solar irradiation at low solar height (instrumental errors due to the cosine response). For these reasons, a pre-processing operation is applied: solar radiation data for which the solar elevation is lower than 10° are excluded from the analysis. However, the solar production during these sunset and sunrise periods are often non negligible and their forecasts cannot be avoided. For forecasting tilted solar irradiation, a CS model uses a constant albedo which, in practice, varies seasonally and sometimes during the day (modifications of the land cover, Notton et al., 2006). For our experimental site, the influence of the sea on the reflected and diffused solar radiation differs in the