



HAL
open science

Periodic Autoregressive Forecasting of Global Solar Irradiation Without Knowledge-based Model Implementation

Cyril Voyant, Jan G de Gooijer, Gilles Notton

► **To cite this version:**

Cyril Voyant, Jan G de Gooijer, Gilles Notton. Periodic Autoregressive Forecasting of Global Solar Irradiation Without Knowledge-based Model Implementation. Solar Energy, 2018. hal-01872340

HAL Id: hal-01872340

<https://hal.science/hal-01872340>

Submitted on 12 Sep 2018

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.

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 Corsica, CNRS UMPR SPE 6134, 20250 Corse, France*

^b *University of Amsterdam, 1018 WB Amsterdam, the Netherlands*

^c *Hospital of Castelluccio, Radiotherapy Unit, BP 85, 20177 Ajaccio, France*

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.

Keywords: Autoregression; Clear sky irradiance; Forecast intervals; Periodic

*Corresponding author. Tel.: +33 4 95 29 36 66; fax: +33 4 95 29 37 97.

E-mail address: cyrilvoyant@gmail.com

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

32 2 Periodic Phenomena

33 Given a stationary time series $\{Y_t; t \in \mathbb{Z}\}$ whose second order moments exist, we define its mean
 34 function by $\mu_t = \mathbb{E}(Y_t)$ and its autocovariance function by $\gamma_{s,t} = \text{Cov}(Y_s, Y_t)$. The process is said
 35 to be periodically correlated (PC) with period H , or periodic stationary (Gladyšhev, 1961), if
 36 the following two conditions

$$\begin{aligned}
 &\mu_t = \mu_{t+h} \text{ for all } t, \text{ and} \\
 &\gamma_{s,t} = \gamma_{s+h,t+h} \text{ for every } s, t \text{ in the index set}
 \end{aligned}
 \tag{1}$$

38 are true for $h = H$ but for no smaller value of h . Note that the case where μ_t and $\gamma_{s,t}$ do not
 39 depend on t reduces to the ordinary covariance stationary time series. Throughout the rest of the
 40 paper, we assume without loss of generality that $\mu_t = 0$. The periodic autocorrelation function
 41 at lag $s = 1, 2, \dots$ and time t is defined by $\rho_{s,t} = \gamma_{s,t}/\gamma_{0,t}$ which is the theoretical counterpart of
 42 the autocorrelation function at lag s of a stationary time series.

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

51 1) The observed time series $\text{GHI}(t)$ ($t = 1, \dots, N$) can be considered as a periodic time series
 52 with two fixed seasonal periods H and D . In this study $H = 24$ hours (h) and $D = 365$ days
 53 (d). For simplicity, we assume that $N/(H \times D) = Y$ is an integer representing the number
 54 of available years. The decomposition of $\text{GHI}(t)$ highlights three new time series: two are
 55 strongly seasonal, and one time series is related to the noise, or irregular component. That
 56 is

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

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

62 The ratio $GHI(t)/CS(t)$ defines the clear sky index $CSI(t) \in [0, 1]$.

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

76 The method of moments based on the well-known Yule-Walker equations and the least squares
77 method in the univariate case are both efficient ways to estimate PAR models. In particular,
78 when the seasonal data and the model for each season are used rather than the annual data,
79 significant gains in parameter efficiency can be achieved. However, the number of estimated
80 parameters is likely to increase with the choice of the season. Thus, in our study it will be
81 easier to consider only the $H = 24$ hours period, giving rise to a parsimonious PAR model with
82 only 24 components rather than estimating a model with $D = 365$ components. Moreover, it is
83 often useful to put linear constraints on the parameters for a given season. Another important
84 advantage is that PAR models can be studied for each season separately. It justifies the use of
85 AIC and BIC information criteria and sample periodic (partial) autocorrelations to identify the
86 optimal model order.

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

95 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 $GHI(t)$ stationary, before calculating the $CS(t)$ 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

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

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

108 Moreover, obtaining accurate $CSI(t)$ series at sunset and sunrise is difficult due to a possible
109 surrounding masking effect such as mountains, buildings, or vegetation. It may also be due
110 to unreliable measurements of solar irradiation at low solar height (instrumental errors due to
111 the cosine response). For these reasons, a pre-processing operation is applied: solar radiation
112 data for which the solar elevation is lower than a threshold of 5° or 10° are removed from the
113 analysis. However, the solar production during these sunset and sunrise periods are often non
114 negligible and their forecasts cannot be avoided. For forecasting tilted solar irradiation, a CS
115 model uses a constant albedo which, in practice, varies seasonally and sometimes during the day
116 (modifications of the land cover, Notton et al., 2006). For our experimental site, the influence
117 of the sea on the reflected and diffused solar radiation differs in the morning and in the evening;
118 Ineichen et al. (1990). Finally, some time lags can occur between the measured and the modeled
119 CS irradiances due to time synchronization or the use of various time scales.

120 4 Data

121 Time series of $GHI(t)$ observations measured at Ajaccio ($41^{\circ}55'N$, $8^{\circ}44'E$, 4 m asl) and Bastia
122 ($42^{\circ}42'N$, $9^{\circ}27'E$, 10 m asl) meteorological stations, Corsica island, France. Both stations are
123 located near the Mediterranean Sea and nearby mountains (more than 1000 m altitude at 40
124 km from the two sites). This specific geographical configuration makes nebulosity or “cloud-like-
125 ness” difficult to forecast. The climate of Corsica is characterized by hot summers with abundant
126 sunshine and mild, dry and clear winters. For Ajaccio, hourly global horizontal solar radiation
127 data are available for the time period 1998–2008 (11 years) and for Bastia the data covers the
128 period 2003–2008 (6 years).

129 A standard cleaning approach is applied to identify and remove bad data. Often mistakes
130 appear in time series of solar data due to problems with the acquisition system; an automatic
131 quality check used in the frame of GEOSS project (Group on Earth Observation System of
132 System) has been applied to the data. The quality of the data (Korany et al., 2016) and the
133 procedure applied to flag suspicious or erroneous measurements is described in detail by David et
134 al. (2016). Both stations are equipped with pyranometers (CM 11 Kipp & Zonen) and standard
135 meteorological sensors (pressure, nebulosity, etc.). A filtering approach is first applied before
136 computing the various time series models. GHI values linked to a solar elevation angle less than
137 10° are excluded from the analysis. In that case, forecasting GHI values will be very difficult
138 due to inaccurate CSI values obtained between periods of sunset and sunrise.

139 The comparison between machine learning models and PAR models is done during the last
140 year of the available data set covering $365 \times 24 = 8,760$ observations. For Ajaccio the first 10
141 years (87,600 observations) are used as a training set of the machine learning models, while there
142 are 5 years of daily observations in the training set for Bastia. To reduce computation time,
143 estimation is done only once using the training/estimation period.

144 5 Periodic Autoregression

145 5.1 Model

146 Machine learning models are usually adopted to predict GHI (Voyant et al., 2017). They are
147 often based on the assumption that the data generation process does not change over time, so
148 assuming that the time series under study is stationary. As noted above, the process to make
149 a time series stationary can be complex and, more importantly, can generate uncertainties. In

150 contrast, a time series fitted by a PAR model does not need the stationarity assumption. In fact,
 151 by definition (1) the time series process is said to be periodically stationary. Moreover, PAR
 152 models avoid CS modeling.

153 Let $\{y \in \mathbb{Z} | 1 \leq y \leq Y\}$ denote the set of years with Y the number of within-in-sample
 154 years. Similarly, let $\{d \in \mathbb{Z} | 1 \leq d \leq D\}$ the set of days at year y , and $\{h \in \mathbb{Z} | 1 \leq h \leq H\}$
 155 the set of hours at day d . So, the time index parameter t may be written as $t \equiv t(y, d, h) =$
 156 $H \times D(y - 1) + H(d - 1) + h$. Then, a PAR stochastic process model of period h ($h = 1, \dots, H$)
 157 and order $p(h)$, for the GHI(t) process is defined by

$$158 \quad \text{GHI}(t) = \sum_{k=1}^{p(h)} \phi_k(h) \text{GHI}(t - k) + \varepsilon(t), \quad (3)$$

159 where $\phi_k(h)$ are the autoregressive (AR) coefficients at hour h , and $\varepsilon(t)$ is a periodic white noise
 160 process with $\mathbb{E}(\varepsilon(t)) = 0$ and $\text{Var}(\varepsilon(t)) = \sigma^2(h)$, independent of GHI(t). When $h = d = 1$, model
 161 (3) reduces to a ‘‘classical’’ AR model.

162 Let $\text{GHI}(y, d, h + H) = \text{GHI}(y, d + 1, h)$, $\text{GHI}(y, d + D, h) = \text{GHI}(y + 1, d, h)$, $\phi_k(h + H) =$
 163 $\phi_k(h)$, and $p(h + H) = p(h)$. Then, with $\text{GHI}(t) \equiv \text{GHI}(y, d, h)$ and $\varepsilon(t) \equiv \varepsilon(y, d, h)$, Eq. (3) can
 164 be rewritten as

$$165 \quad \text{GHI}(y, d, h) = \sum_{k=1}^{p(h)} \phi_k(h) \text{GHI}(y, d, h - k) + \varepsilon(y, d, h). \quad (4)$$

166 A convenient way to represent Eq. (4) is by using vector notation,. That is, for each hour
 167 h , the daily and yearly observations are stacked in the $(D \times Y)$ -dimensional column vector
 168 $\mathbf{Y}(h) = (\text{GHI}(1, 1, h), \text{GHI}(1, 2, h), \dots, \text{GHI}(Y, D, h))'$. Model (4) can then be written as

$$171 \quad \mathbf{Y}(h) = \mathbf{X}(h) \mathbf{\Phi}(h) + \boldsymbol{\varepsilon}(h), \quad (5)$$

172 with

$$173 \quad \mathbf{X}(h) = \begin{pmatrix} \text{GHI}(1, 1, h - 1) & \text{GHI}(1, 1, h - 2) & \cdots & \text{GHI}(1, 1, h - p(h)) \\ \text{GHI}(1, 2, h - 1) & \text{GHI}(1, 2, h - 2) & \cdots & \text{GHI}(1, 2, h - p(h)) \\ \vdots & \vdots & & \vdots \\ \text{GHI}(Y, D, h - 1) & \text{GHI}(Y, D, h - 2) & \cdots & \text{GHI}(Y, D, h - p(h)) \end{pmatrix}_{(D \times Y) \times p(h)}$$

$$175 \quad \mathbf{\Phi}(h) = (\phi_1(h), \dots, \phi_{p(h)}(h))',$$

176 where $\boldsymbol{\varepsilon}(h)$ is a $(D \times Y)$ -dimensional column vector containing the stacked daily and yearly
 177 white noise errors corresponding to $\mathbf{Y}(h)$. The parameter vector $\mathbf{\Phi}(h)$ can be estimated by least
 178

179 squares (LS). In particular, an estimator of $\Phi(h)$ is given by $\widehat{\Phi}(h) = (\widehat{\phi}_1(h), \dots, \widehat{\phi}_{p(h)}(h))' =$
 180 $[\{\mathbf{X}'(h)\mathbf{X}(h)\}^{-1}\mathbf{X}'(h)]\mathbf{Y}(h)$ and the corresponding residuals are defined by $\widehat{\varepsilon}(h) = \mathbf{Y}(h) -$
 181 $\mathbf{X}(h)\widehat{\Phi}(h)$.

182 Model (4) can be enlarged by considering yearly varying trends and constants. In the nu-
 183 merical study, we estimate a PAR model with a constant related to a synthetic sequence of
 184 hourly global horizontal irradiation denoted in this study by an approximation of the ‘‘typi-
 185 cal meteorological year’’ (TMY, Grantham et al. (2018)) or climatology. That is $\text{TMY}(d, h) =$
 186 $Y^{-1} \sum_{y=1}^Y \text{GHI}(y, d, h)$. In that case a PAR model can be written as

$$187 \quad \text{GHI}(y, d, h) - \text{TMY}(d, h) = \sum_{k=1}^{p(h)} \phi_k(h) (\text{GHI}(y, d, h - k) - \text{TMY}(d, h - k)) + \varepsilon(y, d, h). \quad (6)$$

189 5.2 Forecasting

190 Assuming that Eq. (6) is known in the sense that all the parameters are known, the optimal
 191 one-step ahead forecast is given by

$$192 \quad \widehat{\text{GHI}}(y, d, h + 1) = \sum_{k=1}^{p(h)} \phi_k(h) (\text{GHI}(y, d, h - k + 1) - \text{TMY}(y, d, h - k + 1)) + \text{TMY}(y, d, h + 1). \quad (7)$$

193
 194 The corresponding forecast error is $\text{GHI}(y, d, h + 1) - \widehat{\text{GHI}}(y, d, h + 1) = \varepsilon(y, d, h + 1)$, and the
 195 forecast error variance is $\sigma^2(h)$. In practice the coefficients $\phi_k(h)$ are replaced by estimates $\widehat{\phi}_k(h)$.

196 The optimal ℓ -step ahead forecast ($\ell > 1$) is given by

$$197 \quad \widehat{\text{GHI}}(y, d, h + \ell) = \sum_{k=1}^{p(h+\ell-1)} \phi_k(h + \ell - 1) (\widehat{\text{GHI}}(y, d, h - k + \ell - 1) - \text{TMY}(d, h - k + \ell - 1)) \quad (8)$$

$$198 \quad + \text{TMY}(y, d, h + \ell).$$

200 The corresponding ℓ -step ahead forecast error is given by

$$201 \quad \text{GHI}(y, d, h + \ell) - \widehat{\text{GHI}}(y, d, h + \ell) = \varepsilon(y, d, h + \ell) + \sum_{i=1}^{\ell-1} \left(\varepsilon(y, d, h + i) \prod_{j=i}^{\ell-1} \phi_{h+j}^* \right),$$

203 where $\phi_h^* = \sum_{k=1}^{p(h)} \phi_k(h)$. The forecast error variance is given as

$$204 \quad \sigma^2(h, \ell) = \sigma^2(h) \left\{ 1 + \sum_{i=1}^{\ell-1} \left(\prod_{j=i}^{\ell-1} (\phi_{h+j}^*)^2 \right) \right\}. \quad (9)$$

206 5.3 Forecast intervals (FIs)

207 On the assumption that the forecast errors are normally distributed, a standard Box-Jenkins FI
 208 for an ℓ -step ahead forecast of GHI is of the form $\widehat{\text{GHI}}(y, d, h + \ell) \pm z_{\alpha/2}\sigma(h, \ell)$ where $z_{\alpha/2}$ denotes
 209 the $(1 - \alpha/2)$ th percentile of the standard normal distribution. However, a preliminary analysis of
 210 both time series indicated non-normality (p -values < 0.001) of the distribution of $\widehat{\text{GHI}}(y, d, h + \ell)$
 211 with significant values of the sample skewness and kurtosis. In this case, bootstrapping FIs
 212 (BFIs) is a possible alternative.

213 Now, assuming a PAR model is correctly specified, the residuals from a fitted PAR model
 214 are asymptotically uncorrelated (McLeod, 1994). This result implies that bootstrapping can be
 215 carried for each hour h separately. BFIs for ARs have received quite some attention; see, e.g.,
 216 Pan and Politis (2016) for a recent review. In this paper, we use their Algorithm 3.5 (backward
 217 bootstrap with fitted residuals). Monte Carlo simulations performed by these authors show that
 218 this algorithm performs well in terms of coverage. For a fixed h , the resulting ℓ step ahead
 coverage probability $(1 - \alpha)$ will be denoted by

$$219 \quad \text{BFI}_\alpha(\ell) = [\widehat{\text{GHI}}_B^{(\alpha/2)}(t + \ell), \widehat{\text{GHI}}_B^{(1-\alpha/2)}(t + \ell)],$$

220 where $\widehat{\text{GHI}}_B^{(\alpha/2)}(t + \ell)$ and $\widehat{\text{GHI}}_B^{(1-\alpha/2)}(t + \ell)$ are, respectively, the $(\alpha/2)$ th and $(1 - \alpha/2)$ th
 221 percentiles of the empirical bootstrapped distribution function of $\text{GHI}(t + \ell)$ based on B bootstrap
 222 replicates. We set $B = 1,000$ in our computation of FIs (Section 7.3).
 223

224 5.4 Forecast Evaluation

225 We use three error measures to compare the out-of-sample forecasting performance of the models.
 226 The first criterion is the normalized root mean square error (nRMSE, unitless); it is a commonly
 227 used error metric for evaluating point forecasts of GHI; see, e.g., Lauret et al. (2015). The
 228 measure is defined as $\text{nRMSE}(\ell) = [\sum_{t=1}^{N(\ell)} (\widehat{\text{GHI}}(t) - \text{GHI}(t))^2]^{1/2} / \sum_{t=1}^N \text{GHI}(t)$, where $N(\ell)$ is
 229 the number of out-of-sample forecasts depending on forecast horizon $\ell = 1, \dots, 6$. The second
 230 measure is the normalized mean interval length (nMIL; unitless). For a coverage probability
 231 $(1 - \alpha)$ and given set of ℓ -step ahead forecasts $\{\widehat{\text{GHI}}(t + \ell)\}$, this measure is defined as the
 232 difference between the upper $\widehat{\text{GHI}}_U^{(1-\alpha)}(t + \ell)$ and lower $\widehat{\text{GHI}}_L^{(1-\alpha)}(t + \ell)$ limits. The resulting
 233 statistic is given by $\text{nMIL}^{(1-\alpha)}(\ell) = \sum_{t=1}^{N(\ell)} (\widehat{\text{GHI}}_U^{(1-\alpha)}(t) - \widehat{\text{GHI}}_L^{(1-\alpha)}(t)) / \sum_{t=1}^{N(\ell)} \text{GHI}(t)$. As a
 234 third evaluation measure, we consider the ℓ -step ahead prediction interval coverage $\text{PICP}^{(1-\alpha)}(\ell)$.
 235 This measure is defined by the probability that the target value of an input pattern lies between
 236 the forecast limits. That is, $\text{PICP}^{(1-\alpha)}(\ell) = 100 \times (1/N(\ell))\#(j)$ with $\#(j)$ the number of j 's

Table 1: Summary of forecasting models. Short-hand notation: SP = smart persistence, AR = autoregressive, MLP = multilayer perceptron, PAR = periodic AR, TMY = typical meteorological year, CS = clear sky (Solis) model.

Knowledge-based model	Predictor type	Form of the ℓ -step ahead forecast $\widehat{\text{GHI}}(y, d, h + \ell)$
With	SP1	$\text{GHI}(y, d, h) \frac{\text{CS}(d, h + \ell)}{\text{CS}(d, h)}$
	AR(p)	$\sum_{i=1}^p \frac{\phi_i \text{GHI}(y, d, h + \ell - i)}{\text{CS}(d, h + \ell - i)} \text{CS}(d, h + \ell)$
	MLP*	$\sum_{j=1}^m w_j f \left(\sum_{i=0}^p \frac{w_{ji} \text{GHI}(y, d, h - i)}{\text{CS}(d, h - i)} \right) \text{CS}(d, h + \ell)$
Without	SP2	$\text{GHI}(y, d, h) - \text{TMY}(d, h) + \text{TMY}(d, h + \ell)$
	PAR($p(h)$)	Equation (7)

* An AR-NN model of order p with m hidden neurons and a single linear output.

The initial weights are denoted by w_j and w_{ji} .

237 lying in the interval $[\widehat{\text{GHI}}_L^{(1-\alpha)}(t), \widehat{\text{GHI}}_U^{(1-\alpha)}(t)]$. Several test statistics can be based on PICP;
 238 see, e.g., De Gooijer (2017, Chapter 10).

239 6 Some Alternative Models

240 It is interesting to compare the proposed PAR forecasting model with existing naïve and reference
 241 models. Naïve models will be denoted by SP which stands for smart persistence. They are simple
 242 models related to the knowledge-based model CS (SP1) and to TMY (SP2) as described in Table
 243 1. Detailed information about CS and SP1 can be found in Lauret et al. (2015), Voyant et al.
 244 (2015), Mueller et al. (2004), and Ineichen (2008) and the references therein. The CS model
 245 under study is the so-called Solis model which is known to give good results; see, e.g., Lauret
 246 et al. (2015), Ruiz-Arias et al. (2017), and Voyant et al. (2015). Two reference models are an
 247 AR model and a multilayer perceptron (MLP) belonging to the class of artificial neural network
 248 models. The main equations are given in Table 1; see, e.g., Voyant et al. (2011), Voyant et al.
 249 (2014), and Voyant et al. (2017).

250 7 Empirical Results

251 7.1 Exploratory Analysis

252 The presence of periodic correlation in the series $\text{GHI}(t)$ can be detected by the sample periodic
 253 autocorrelation function at time lag s and period h , i.e. $r_{s,h} = c_{s,h}/(c_{0,h}c_{0,h-s})^{1/2}$ where $c_{s,h} =$
 254 $Y^{-1} \sum_y (\text{GHI}(y, d, h) - \text{TMY}(d, h)) (\text{GHI}(y-s, d, h) - \text{TMY}(d, h))$ ($y = 1, \dots, Y; h = 1, \dots, H$).
 255 This statistic can also be used to test the null hypothesis: $\rho_{s,h} \equiv \rho_s$, ($s = 1, 2, \dots$). Using
 256 the function `peracf` in the R-`perARMA` package, we reject the null hypothesis for all values $s =$
 257 $1, \dots, 30$ (p -values < 0.001). This indicates that both $\text{GHI}(t)$ series are properly PC and is not
 258 just an amplitude modulated stationary sequence. In addition, we reject the null hypothesis
 259 $\rho_{s,h} = 0$ (p -values < 0.001). So the series are not PC white noise.

260 A suitable PAR model can be selected by examining plots of the sample periodic partial
 261 autocorrelation (perPACF), say $\hat{\rho}_{\cdot,s,h}$, or by using an information criterion such as AIC or BIC.
 262 Sakai (1982) showed that if the correct order is $p(h)$ for period h , the estimate of the asymptotic
 263 standard deviation of $\hat{\rho}_{\cdot,s,h}$ equals $1/\sqrt{n}$, $s > p(h)$. The order $p(h)$ can be identified by finding
 264 the lowest lag for which the sample perPACF cuts off. BIC may be factored to obtain a separate
 265 criterion for each period. Thus

$$266 \quad \text{BIC}(p(1), \dots, p(H)) = \sum_{h=1}^H \text{BIC}(p(h)) \quad (10)$$

267 with

$$269 \quad \text{BIC}(p(h)) = \log(\hat{\sigma}_h^2(p(h))) + \frac{\log(n)}{n} p(h), \quad (11)$$

271 where $\hat{\sigma}_h^2(p(h))$ corresponds to the LS estimator of the residual variance, and $n = 3,650$ (1,825)
 272 in the case of Ajaccio (Bastia). Thus, the total minimization of Eq. (10) can be established
 273 by minimizing each $\text{BIC}(p(h))$ about $p(h)$. Replacing $\log(n)$ in Eq. (11) by 2 gives AIC. It
 274 should be pointed out that for identifying subsets of PAR models with AIC or BIC a local search
 275 algorithm (e.g., a genetic algorithm, Ursu and Turkman, 2012) is recommended to avoid lengthy
 276 computations.

277 Table 2 summarizes the significance of the sample perPACF values for Ajaccio at a 5%
 278 nominal significance level, in terms of two indicator symbols and for $h = 4, \dots, 20$. Since global
 279 solar radiation is zero during the night, depending on the season, the sunrise and the fluctuation
 280 of sunshine, sample perPACF results for $h = 1, 2, 3$ and $h = 21, \dots, 24$ are not included. Observe
 281 that for almost all values of h , significant values of $\hat{\rho}_{\cdot,s,h}$ occur at lags $s = 1, 2, 3$. For $h =$

Table 2: Indicator pattern of statistically significant values of the sample perPACF for Ajaccio; “+” indicates a value $> 1.96n^{-1/2}$ and “-” a value $< -1.96n^{-1/2}$ with $n = 3,650$.

	Lags (s)																													
h	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
4									-													+	+	+	+	+				
5	+										+	+	+											+	+	-				
6	+	-											+	+																
7	+	-													+	+														
8	+	-																												
9	+	-																												
10	+																													
11	+		+																											
12	+		+																											
13	+																													
14	+		+																											
15	+		+			+	+	+																						
16	+					+		+	+																					
17	+	-												-										+						
18	+	-													-									+	+					
19	+	-																												
20	+																													

282 4, . . . , 20, AIC selects a PAR model of order (24, 25, 25, 26, 18, 2, 3, 1, 3, 7, 4, 7, 8, 12, 26, 25, 26) at
 283 forecast horizon $\ell = 1$ while BIC suggests a PAR(24, 25, 25, 14, 2, 2, 2, 1, 3, 4, 1, 3, 8, 9, 26, 25, 2)
 284 model. These results are broadly in agreement with the orders selected by perPACF for the
 285 GHI series of Ajaccio. Note, BIC prefers low order PAR models. Similar results are observed
 286 for the GHI series of Bastia at $\ell = 1$. Forecast results will be based on the PAR model orders
 287 determined by AIC and BIC.

288 7.2 Forecasting

289 The nRMSE forecasting results are summarized in Table 3, where PAR-AIC (PAR-BIC) denote
 290 the PAR models selected by AIC (BIC). Observe that a simple model of persistence (P) has been
 291 added to the set of models. It is defined by $\widehat{\text{GHI}}(y, d, h + \ell) = \text{GHI}(y, d, h)$. It assumes that
 292 all ℓ -step ahead forecast values of GHI are equal to GHI at hour h , independent of y and d . It
 293 is also worth mentioning that the three models “simple SP”, “smart SP1”, and “smart SP2” are
 294 well-specified at both sites, i.e. their specification has benefited from long work experience.

295 Initially, the AOD and w parameters are updated each month from the aeronet database
 296 (<https://aeronet.gsfc.nasa.gov/>) where AOD values fluctuate between 0.1 and 0.5 and w be-
 297 tween 0.2cm and 0.9cm. If we select constant and ill-chosen values; AOD = 0.9 and $w = 0.9$, the

Table 3: nRMSE(ℓ) values ($\ell = 1, \dots, 6$). MLP results are based on the best of 10 training sets with random initialization. Bold values show the best results in each row.

Horizon (ℓ)	Climatology TMY	Persistence			AR	MLP	PAR	
		Simple P	Smart SP1	Smart SP2			AIC	BIC
Ajaccio								
1	0.3554	0.3409	0.1978	0.1950	0.1965	0.1961	0.1908	0.1838
2	0.3354	0.5938	0.3120	0.2683	0.2642	0.2661	0.2640	0.2639
3	0.3554	0.7974	0.4299	0.3062	0.3037	0.3027	0.2982	0.3008
4	0.3554	0.9563	0.5666	0.3313	0.3307	0.3267	0.3160	0.3149
5	0.3554	1.0699	0.7054	0.3486	0.3497	0.3419	0.3259	0.3282
6	0.3554	1.1449	0.8346	0.3585	0.3627	0.3540	0.3353	0.3383
Bastia								
1	0.3606	0.3609	0.2435	0.2103	0.2231	0.2235	0.1968	0.1972
2	0.3606	0.6184	0.4013	0.2942	0.2931	0.2909	0.2660	0.2678
3	0.3606	0.8261	0.5545	0.3384	0.3298	0.3269	0.2995	0.3048
4	0.3606	0.9853	0.7095	0.3629	0.3515	0.3456	0.3205	0.3271
5	0.3606	1.0963	0.8540	0.3745	0.3642	0.3593	0.3331	0.3425
6	0.3606	1.1676	0.9821	0.3781	0.3725	0.3677	0.3409	0.3495

298 one-step ahead ($\ell = 1$) nRMSE increases to 0.2394 for SP1, to 0.2300 for MLP, and to 0.2481 for
299 AR. In this case, the accuracy of the CS is poor and widely influences the forecasts. For $\ell = 6$
300 the results are worse, i.e. 0.9927 (SP1), 0.4404 (MLP), and 0.4571 (AR). This confirms that the
301 accuracy of the CS model plays an important role in establishing its reliability, and that a PAR
302 model is all the more an interesting alternative.

303 As compared to Ajaccio, the nRMSE results are slightly worse for Bastia. The reason is that
304 as the nebulosity becomes more important for this site, the CS model is very difficult to optimize.
305 Moreover, the training set covers only 5 years of data. In summary, even when a CS model is
306 well-specified, PAR models give the best forecasting results for all horizons. GM models (AR
307 and MLP) give good nRMSE results, albeit with some variability. Overall, SP2 based on TMY
308 gives lower nRMSE results than SP1 based on CS. The use of TMY with PAR methodology
309 improves the forecasting results. In particular, for Ajaccio nRMSE(1) = 0.1838 (PAR-BIC) with
310 TMY and 0.2125 without. This effect is equivalent for the other forecast horizons ℓ , i.e. the gain
311 of including TMY in a PAR model specification comes close to 4 percentage points.

312 7.3 Forecasts intervals

313 The reliability of a FI can be assessed by nMIL and PICP. An optimal methodology is associated
 314 with a high PICP (close to 100%) and a low nMIL. The choice of a lower (L) and upper (U) FI
 315 limit is essential here. To explore the null hypothesis that the $GHI(y, d, h)$ data comes from a
 316 symmetrical distribution (skewness) with an unknown median, we used the `triplestest` available
 317 from MATLAB Central. At the 5% nominal significance level, the test statistic indicates that
 318 the null hypothesis cannot be rejected for $h = 6, 7, 8, 17$ and 20 . For the remaining 12 series,
 319 we obtained small p -values casting doubt about the validity of the null hypothesis. Nevertheless,
 320 we set $L_{t+\ell}(h) \equiv -U_{t+\ell}(h)$ throughout the rest of the analysis.

321 Figure 1 shows a plot of $nMIL(\ell)$ versus $PICP(\ell)$ ($\ell = 1, \dots, 6$) for Ajaccio aggregated over
 322 all values of $h \in [4, 20]$. Clearly, the best compromise between these accuracy measures is for
 323 $1 \leq U_{t+\ell}(h) \leq 2$ with PICP in the range $[0.7, 0.8]$ and $nMIL \in [0.3, 0.6]$ for $U_{t+\ell}(h) = 1$, while
 324 $PICP \in [0.8, 0.9]$ and $nMIL \in [0.7, 1.2]$ for $U_{t+\ell}(h) = 2$. Interestingly, mixed results emerge from
 325 using BFIs. For instance, at $\alpha = 0.05$ and taken across all values of h , $nMIL(1) = 0.02$ and
 326 $PICP(1) = 0.46$ with the mean length of the $BFI_{0.05}(1)$ interval ranging between 1.97 ($h = 4$)
 327 and 302.9 ($h = 13$).

328 The bootstrapped nMIL and PICP values may be tentatively compared with those obtained
 329 when $U_{t+1}(h) = 1.96$ which corresponds to a 95% FI if, for the moment, the forecast errors are
 330 assumed to be normally distributed. Then $nMIL(1) = 1.1$, $PICP(1) = 0.92$ and the mean length
 331 of the FI, averaged over all h , equals 84.9 . So, with bootstrapping the value of nMIL in this case
 332 is considerably smaller than in the non-bootstrapping case, but with bootstrapping there is also
 333 a marked decrease in the value of PICP. This observation applies to all values of ℓ .

334 Observe that the above results are taken over all values of h jointly. Since the selected PAR
 335 orders are different for each h , it is interesting to present nMIL and PICP values for each h
 336 separately. Table 4 shows these results for $U_{t+1}(h) = 1, 1.5$, and 2 . We see that for $h = 8, \dots, 16$
 337 the one-step ahead PICP values for Ajaccio and Bastia are very close to 100%. The nMIL values
 338 on the other hand increase for $h = 8, \dots, 16$ with values for Bastia about two times larger than
 339 those for Ajaccio. At this point it is worth noting that the choice of the upper and lower FI bands
 340 depends on the operational system in place, and on the prerogative of the network manager. But
 341 the results in Figure 1 and Table 4 show that PAR models can result in quite accurate one-step
 342 ahead forecasts with high confidence.

343 Table 5 shows the impact of TMY on the PAR forecasts. Introducing TMY in a PAR model

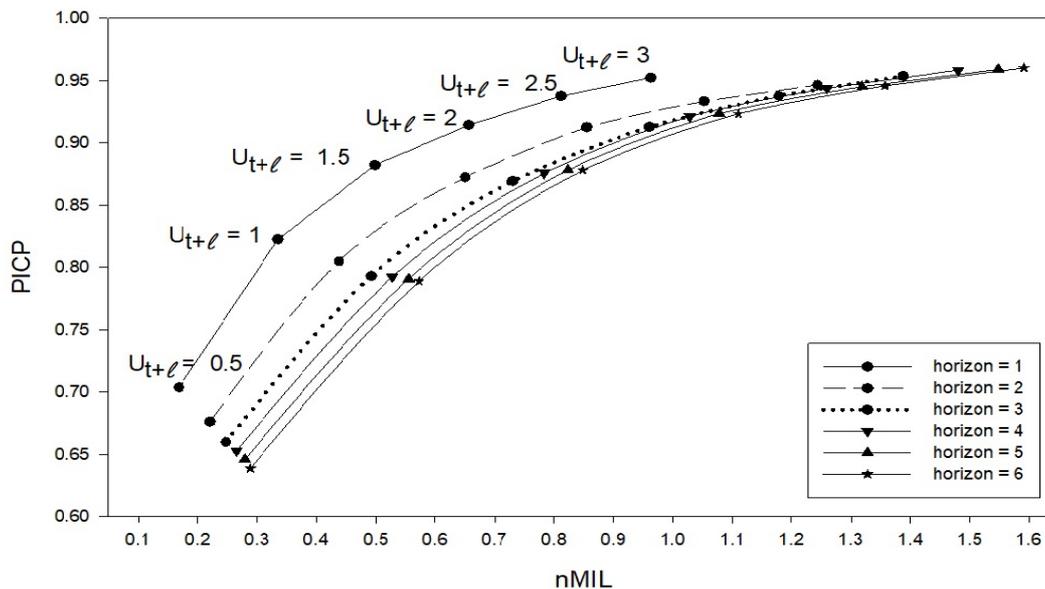


Figure 1: $nMIL(\ell)$ versus $PICP(\ell)$ for various values of $U_{t+\ell}(h)$ ($\ell = 1, \dots, 6$) for Ajaccio. PAR orders are selected by BIC.

Table 4: $nMIL$ and $PICP$ values (in percentages) for different values of $U_{t+1}(h)$ ($h = 4, \dots, 20$).

h	Ajaccio						Bastia					
	$U_{t+1}(h) = 1$		$U_{t+1}(h) = 1.5$		$U_{t+1}(h) = 2.0$		$U_{t+1}(h) = 1$		$U_{t+1}(h) = 1.5$		$U_{t+1}(h) = 2.0$	
	nMIL	PICP	nMIL	PICP	nMIL	PICP	nMIL	PICP	nMIL	PICP	nMIL	PICP
4	0.0	1.1	0.0	1.1	0.0	1.1	0.4	0.0	0.6	0.0	0.9	0.0
5	0.7	37.5	1.1	37.5	1.5	37.5	1.4	24.1	2.1	24.1	2.7	24.1
6	6.6	62.5	9.7	62.5	12.8	62.5	12.8	55.9	18.8	55.9	24.6	55.9
7	16.4	78.6	24.2	78.6	31.9	78.6	32.5	78.4	48.0	78.4	63.2	78.4
8	27.7	98.4	40.7	98.4	53.3	98.4	58.9	97.5	86.6	97.5	113.2	97.5
9	40.8	98.9	60.5	98.9	79.9	98.9	84.0	99.5	124.4	99.5	163.9	99.5
10	49.3	99.2	73.5	99.2	97.4	99.2	105.1	100.0	156.5	100.0	207.0	100.0
11	58.3	99.7	86.9	99.7	115.1	99.7	132.7	99.7	197.6	99.7	261.4	99.7
12	63.3	100.0	94.2	100.0	124.6	100.0	132.2	99.7	196.7	99.7	260.1	99.7
13	64.2	100.0	95.3	100.0	125.5	100.0	136.9	100.0	203.4	100.0	268.6	100.0
14	60.1	100.0	89.0	100.0	117.1	100.0	128.0	99.7	188.9	99.7	247.7	99.7
15	53.1	100.0	78.6	100.0	103.4	100.0	111.5	99.2	164.1	99.2	214.6	99.2
16	38.8	97.3	57.0	97.3	74.7	97.3	75.6	95.6	11.2	95.6	145.5	95.6
17	24.6	83.6	35.8	83.6	46.7	83.6	47.4	80.5	69.6	80.5	91.1	80.5
18	12.6	58.4	18.5	58.4	24.3	58.4	23.3	59.2	34.5	59.2	45.5	59.2
19	4.2	34.0	6.2	34.0	8.2	34.0	8.1	44.4	12.0	44.4	15.9	44.4
20	0.5	0.0	0.8	0.0	1.1	0.0	1.2	21.4	1.8	21.4	2.4	21.4

Table 5: Forecast interval evaluation measures for the fitted PAR models; $U_{t+\ell}(h) = 2$.

Horizon (ℓ)	PICP(%)		Gain(%)	nMIL(%)		Gain(%)
	Without TMY	With TMY		Without TMY	With TMY	
Ajaccio						
1	91.0	91.4	+0.4	70.3	65.7	-4.6
2	89.3	91.3	+2.0	92.5	85.6	-6.9
3	88.2	91.2	+3.0	104.5	96.0	-8.5
4	87.2	92.1	+4.9	112.6	102.8	-9.8
5	86.8	92.4	+5.6	118.4	107.8	-10.6
6	86.6	92.3	+5.7	122.5	111.0	-11.5
Bastia						
1	95.0	94.7	-0.3	78.3	70.4	-7.9
2	94.3	94.3	0	102.3	91.6	-10.7
3	94.1	94.1	0	114.6	102.4	-12.2
4	94.0	94.0	0	122.3	109.3	-13.0
5	94.0	94.1	+0.1	127.6	113.8	-13.8
6	94.2	94.2	0	131.0	116.8	-14.2

344 specification, reduces the values of $nMIL(\ell)$ significantly; between 4.6% and 11.5% for Ajaccio,
345 and between 7.9% and 14.2% for Bastia. Interestingly, for Bastia the $PICP(\ell)$ values remain
346 constant and close to 95% across all values of ℓ . This phenomenon is quite different for Ajaccio
347 with decreasing values of $nMIL(\ell)$ and increasing values of $PICP(\ell)$.

348 8 Concluding Remarks

349 We have introduced the PAR family of models and applied these models to forecast GHI time
350 series. PAR models are intended for periodic time series. They are not restricted to stationary
351 time series. In contrast, classical models like AR, and MLP need a preprocessing transformation
352 to make the GHI series stationary. Thereby, in the PAR case, the use of a knowledge-based
353 model is not necessary. Usually a CS radiation model is fitted to the data to remove certain
354 components, and hence the initial time series is transformed in a new stationary series of CSIs.
355 This transformation is efficient if all parameters are perfectly known. But if this is not the case,
356 the introduction of a CS model can become ineffective and can even give rise to an additional
357 source of error. The results obtained in this study show that even if a CS model is well-specified,
358 AR and MLP coupled with CS are less efficient than a PAR model. Indeed, for all the horizons,

359 PAR gives the best results even if the improvement is small compared with other models. If a
360 CS model is not perfectly known for a site, a PAR model becomes an interesting alternative with
361 minimal forecast errors. The model allows FIs and through the use of upper and lower forecast
362 bands the length of the FIs can be adjusted to gain forecast efficiency.

363 A related study is by Pedro et al. (2018) based on 5 minutes GHI and direct normal irradi-
364 ance (DNI) data obtained from Folsom, CA, USA. With the purpose of building an intra-hour
365 forecasting model, they use two machine learning algorithms, i.e. k-nearest-neighbors (kNN) and
366 gradient boosting. At a nominal coverage of 80% and a 30 minute forecast horizon, the best result
367 is obtained for a kNN method with cloud cover information derived from sky images (Pedro et
368 al., 2018, Table 6). The reported values are PICP = 81.5% and nMIL = 9.9%, with nMIL called
369 PINAW (Prediction Interval Normalized) by these authors. In contrast, a PAR model gives
370 84.7% (PICP) and 73.5% (nMIL) at a one hour forecast horizon while with bootstrapping we
371 have 49.5% (PICP) and 3.9% (nMIL). It is evident that these results should be interpreted with
372 care. Especially, since data location, data frequency, forecast horizon, and forecasting methods
373 are different. Nevertheless, it is clear that PAR models perform pretty well in this special case.
374 Moreover, PAR models need no additional forecast information in the form of sky images.

375 In summary, while PAR models are known since many years, they are rarely used for GHI
376 forecasting. A PAR model is a good and recommended model when a forecasting tool must be
377 developed for a new meteorological site. In particular, if there is not sufficient hindsight and
378 historical data for which a CS model is not well-defined by, for instance, lack of meteorological
379 data. This frequently happens when a new project of photovoltaic plant implementation occurs
380 at a new site. Some further research is needed for greater validation of the proposed methodology.
381 For instance, comparing PAR forecasts with other models/methods, using GHI data obtained
382 from a larger set of meteorological sites, and studying different time granularities

383 References

- 384 David, M., Ramahatana, F., Trombe, P.-J., Lauret, P., 2016. Probabilistic forecasting of the
385 solar irradiance with recursive ARMA and GARCH models. *Sol. Energy*, 133, 55–72. DOI:
386 10.1016/j.solener.2016.03.064.
- 387 Diagne, M., David, M., Lauret, P., Boland, J., Schmutz, N., 2013. Review of solar irradiance
388 forecasting methods and a proposition for small-scale insular grids. *Renew. Sustain. Energy*
389 *Rev.*, 27, 65–76. DOI: 10.1016/j.rser.2013.06.042.
- 390 De Gooijer, J.G., 2017. *Elements of Nonlinear Time Series Analysis and Forecasting*. Springer
391 *Series in Statistics*, Springer-Verlag, New York. DOI: 10.1007/978-3-319-43252-6.
- 392 Franses, P.H., Paap, R., 1994. Model selection in periodic autoregressions. *Oxf. Bull. Econ. Stat.*
393 56(4), 421–439. DOI: 10.1111/j.1468-0084.1994.tb00018.x.
- 394 Gelaro, R., McCarty, W., Suárez, M.J., Todling, R., Molod, A., Takacs, L., Randles, C.A.,
395 Darmenov, A., Bosilovich, M.G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper,
396 C., Akella, S., Buchard, V., Conaty, A., da Silva, A.M., Gu, W., Kim, G.-K., Koster, R.,
397 Lucchesi, R., Merkova, D., Nielsen, J.E., Partyka, G., Pawson, S., Putman, W., Rienecker,
398 M., Schubert, S.D., Sienkiewicz, M., Zhao, B., 2017. The modern-ear retrospective analysis for
399 research and applications, Version 2 (MERRA-2). *J. Climate*, 30, 5419–5454. DOI: 10.1175/
400 JCLI-D-16-0758.1.
- 401 Gladhyšhev, E.G., 1961. Periodically correlated random sequences. *Sov. Math. Doklady*, 2, 385–
402 388.
- 403 Grantham, A.P., Gel, Y.R., Boland, J.W., 2016. Nonparametric short-term probabilistic fore-
404 casting for solar radiation, *Sol. Energy*, 133, 465–475. DOI: 10.1016/j.solener.2016.04.011.
- 405 Grantham, A.P., Pudney, P.J., Boland, J.W., 2018. Generating synthetic sequences of global
406 horizontal irradiation. *Sol. Energy*, 162, 500–509. DOI: 10.1016/j.solener.2018.01.044.
- 407 Hokoi, S., Matsumoto, M., Ihara, T., 1990. Statistical time series models of solar radiation
408 and outdoor temperature – Identification of seasonal models by Kalman filter. *Energy Build.*,
409 15(3-4), 373–383. DOI: 10.1016/0378-7788(90)90011-7.
- 410 Ineichen, P., 2006. Comparison of eight clear sky broadband models against 16 independent data
411 banks. *Sol. Energy*, 80(4), 468–478. DOI: 10.1016/j.solener.2005.04.018.

412 Ineichen, P., 2008. A broadband simplified version of the solis clear sky model. *Sol. Energy*, 82(8),
413 758–762. DOI: 10.1016/j.solener.2008.02.009.

414 Ineichen, P., Guisan, O., Perez, R., 1990. Ground-reflected radiation and albedo. *Sol. Energy*,
415 44(4), 207–214. DOI: 10.1016/0038-092x(90)90149-7.

416 Korany, M., Boraiy, M., Eissa, Y., Aoun, Y., Abdel Wahab, M.M., Alfaro, S.C., Blanc, P., El-
417 Metwally, M., Ghedira, H., Hungershofer, K., and others, 2016. A database of multi-year
418 (2004–2010) quality-assured surface solar hourly irradiation measurements for the Egyptian
419 territory. *Earth Syst. Sci. Data*, 8(1), 105–113. DOI: 10.5194/essdd-8-737-2015.

420 Lauret, P., Voyant, C., Soubdhan, T., David, M., Poggi, P., 2015. A benchmarking of machine
421 learning techniques for solar radiation forecasting in an insular context. *Sol. Energy*, 112,
422 446–457. DOI: 10.1016/j.solener.2014.12.014.

423 McLeod, A.I., 1994. Diagnostic checking of periodic autoregression models with applications. *J.*
424 *Time Ser. Anal.*, 15(2), 221–233. DOI: 10.1111/j.1467-9892.1994.tb00186.x.

425 Mueller, R.W., Dagestad, K.F., Ineichen, P., Schroedter-Homscheidt, M., Cros, S., Dumortier,
426 D., Kuhlemann, R., Olseth, J.A., Piernavieja, G., Reise, C., Wald, L., Heinemann, D., 2004.
427 Rethinking satellite-based solar irradiance modelling: The SOLIS clear-sky module. *Remote*
428 *Sens. Environ.*, 91(2), 160–174. DOI: 10.1016/s0034-4257(04)00069-0.

429 Notton, G., Poggi, P., Cristofari, C., 2006. Predicting hourly solar irradiances on inclined
430 surfaces based on the horizontal measurements: Performances of the association of well-
431 known mathematical models. *Energy Convers. Manag.*, 47(13-14), 1816–1829. DOI: 10.1016/
432 j.enconman.2005.10.009.

433 Notton, G., Nivet, M.-L., Voyant, C., Paoli, C., Darras, C., Motte, F., Fouilloy, A., 2018.
434 Intermittent and stochastic character of renewable energy sources: Consequences, cost of
435 intermittence and benefit of forecasting. *Renew. Sustain. Energy Rev.*, 87, 96–105. DOI:
436 10.1016/j.rser.2018.02.007.

437 Notton, G., Voyant, C., 2018. Forecasting of intermittent solar energy resource (Chapter 3). In
438 I. Yahyaoui (Ed.), *Advances in Renewable Energies and Power Technologies*, Volume 1: Solar
439 and Wind Energies, Elsevier, pp. 77–114. DOI: 10.1016/b978-0-12-812959-3.00003-4.

440 Pan, L., Politis, N., 2016. Bootstrap prediction intervals for linear, nonlinear and nonparametric
441 autoregressions. *J. Statist. Plann. Inference*, 177, 1–27. DOI: 10.1016/j.jspi.2014.10.003.

442 Pedro, H.T.C., Coimbra, C.F.M., David, M., Lauret, P., 2018. Assessment of machine learning
443 techniques for deterministic and probabilistic intra-hour solar forecasts. *Renew. Energy*, 123,
444 191–203. DOI: 10.1016/j.renene.2018.02.006.

445 Ruiz-Arias, J.A., Gueymard, C.A., Cebecauer, T., 2017. Worldwide multi-model intercomparison
446 of clear-sky solar irradiance predictions. *AIP Conference Proceedings* 1850, 140018. DOI: 10.
447 1063/1.4984526.

448 Sakai, H., 1982. Circular lattice filtering using Pagano’s method. *IEEE Trans. Accoustics, Speech,*
449 *and Signal Processing*, 30(2), 279–287. DOI: 10.1109/tassp.1982.1163874.

450 Shamshirband, S., Mohammadi, K., Yee, P.L., Petković, D., Mostafaeipour, A., 2015. A
451 comparative evaluation for identifying the suitability of extreme learning machine to pre-
452 dict horizontal global solar radiation. *Renew. Sustain. Energy Rev.*, 52, 1031–1042. DOI:
453 10.1016/j.rser.2015.07.173.

454 Tiao, G.C., Grupe, M.R., 1980. Hidden periodic autoregressive moving average models in time
455 series data. *Biometrika*, 67(2), 365–373. DOI: 10.1093/biomet/67.2.365.

456 Ula, T.A., Smadi, A.A., 1997. Periodic stationarity conditions for periodic autoregressive moving
457 average processes as eigenvalue problems. *Water Resour. Res.*, 33(8), 1929–1934. DOI: 10.
458 1029/97wr01002.

459 Ursu, E., Turkman, K.F., 2012. Periodic autoregressive model identification using genetic algo-
460 rithms. *J. Time Ser. Anal.*, 33, 398–405. DOI: 10.1111/j.1467-9892.2011.00772.x.

461 Voyant, C., Muselli, M., Paoli, C., Nivet, M.-L., 2011. Optimization of an artificial neural network
462 dedicated to the multivariate forecasting of daily global radiation. *Energy*, 36(1), 348–359. DOI:
463 10.1016/j.energy.2010.10.032.

464 Voyant, C., Haurant, P., Muselli, M., Paoli, C., Nivet, M.-L., 2014. Time series modeling and
465 large scale global solar radiation forecasting from geostationary satellites data. *Sol. Energy*,
466 102, 131–142. DOI: 10.1016/j.solener.2010.08.011.

467 Voyant, C., Soubdhan, T., Lauret, P., David, M., Muselli, M., 2015. Statistical parameters as a
468 means to a priori assess the accuracy of solar forecasting models. *Energy*, 90, 671–679. DOI:
469 10.1016/j.energy.2015.07.089.

470 Voyant, C., Notton, G., Kalogirou, S., Nivet, M.-L., Paoli, C., Motte, F., Fouilloy, A., 2017.
471 Machine learning methods for solar radiation forecasting: A review. *Renew. Energy*, 105, 569–
472 582. DOI: 10.1016/j.renene.2016.12.095.