

RESEARCH ARTICLE

WILEY

A cyclostationary model for temporal forecasting and simulation of solar global horizontal irradiance

Soumya Das¹  | Marc G. Genton¹  | Yasser M. Alshehri² | Georgiy L. Stenchikov²

¹Statistics Program, King Abdullah University of Science and Technology, Thuwal, 23955-6900, Saudi Arabia

²Earth Science and Engineering Program, King Abdullah University of Science and Technology, Thuwal, 23955-6900, Saudi Arabia

Correspondence

Soumya Das, Statistics Program, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia.
Email: soumya.das@kaust.edu.sa

Abstract

As part of *Saudi Vision 2030*, a major strategic framework developed by the Council of Economic and Development Affairs of Saudi Arabia, the country aims to reduce its dependency on oil and promote renewable energy for domestic power generation. Among the sustainable energy resources, solar energy is one of the leading resources because of the endowment of Saudi Arabia with plentiful sunlight exposure and year-round clear skies. This essentializes to forecast and simulate solar irradiance, in particular global horizontal irradiance (GHI), as accurately as possible, mainly to be utilized by the power system operators among many others. Motivated by a dataset of hourly solar GHIs, this article proposes a model for short-term point forecast and simulation of GHIs. Two key points, that make our model competent, are: (1) the consideration of the strong dependency of GHIs on aerosol optical depths and (2) the identification of the periodic correlation structure or cyclostationarity of GHIs. The proposed model is shown to produce better forecasts and more realistic simulations than a classical model, which fails to recognize the GHI data as cyclostationary. Further, simulated samples from both the models as well as the original GHI data are used to calculate the corresponding photovoltaic power outputs to provide a comprehensive comparison among them.

KEYWORDS

cyclostationary, energy production, forecast, periodically correlated, renewable energy, solar global horizontal irradiance

1 | INTRODUCTION

The rapidly growing population and economy of Saudi Arabia demand energy resources for desalination, electricity generation, and process heating. Although conventional energy production using petroleum has been widely implemented, it is, notably, the primary cause of environmental pollution and thus of declining public health. Therefore, the Kingdom finds it essential to examine local sustainable energy resources, for domestic use and export, and greater purposes. Indeed, on April 25, 2016, a strategic framework called *Saudi Vision 2030* was announced for reducing Saudi Arabia's dependence on fossil fuels, with the promotion of non oil-based energy production as the key goal. This announcement eventually led to the launch of the *Renewable Energy Project* on March 27, 2018, with a funding of \$200 billion.

Saudi Arabia is a land of deserts with plentiful sunlight and year-round clear skies. Because of such a geographically strategic location, solar energy has become a promising resource for adequate and sustainable power generation in Saudi

Arabia. Consequently, many researchers have analyzed various solar resource measurements; see, for example, Myers et al. (1999), AlYahya and Irfan (2014), Pazheri (2014), Almasoud and Gandayh (2015), Zell et al. (2015), AlYahya and Irfan (2016), Hejase and Assi (2016) and references therein, and Salam and Khan (2018). Besides, it is also important to provide advanced statistical methodologies to produce reliable forecasts and realistic simulations of solar irradiances. Accurate predictions of solar irradiances can help estimate future solar energy production needs, thus assisting photovoltaic (PV) power system operators to adopt early measures, such as the introduction of new market products or increment in procuring operating reserves, to mitigate impacts caused by variations in solar irradiance (Navid and Rosenwald 2013, Li and Zhang 2020); refer to Yang et al. (2018) and Yang (2019) for an extensive review on solar forecasting. On the other hand, when solar irradiance data are unavailable, realistic simulations can effectively help save large amounts of funds through, for example, efficient commitment in petroleum dispatching.

Among the three different solar irradiances—direct normal irradiance (DNI), global horizontal irradiance (GHI), and diffuse horizontal irradiance (DHI)—GHI data are more easily available because DNI (and DHI) measurements require Sun tracking, therefore, appear more costly and require frequent maintenance; see Lee et al. (2017). Moreover, the GHI is more relevant than DHI or DNI to energy production, as it characterizes the entire incoming solar energy. We recall that the GHI represents the total amount of radiation received by a solar measurement mechanism; it can be expressed in terms of DHI and DNI as

$$\text{GHI} = \text{DHI} + \text{DNI} \cdot \cos(\theta),$$

where θ is the solar zenith angle.

Considering the significant impact of solar irradiances on energy production in Saudi Arabia, in this article, we fit an appropriate statistical model that can provide reliable short-term forecasts and generate realistic simulations. For this, we consider a dataset of hourly solar GHIs comprising observations of 33 consecutive days measured at the King Abdullah University of Science and Technology (KAUST) solar monitoring station, Thuwal, Saudi Arabia; we formally introduce the dataset in Section 2.

Solar radiation attenuates as it passes through the Earth's atmosphere because of scattering and absorption by atmospheric components such as aerosols and clouds, significantly increasing DHI while decreasing DNI and GHI; see Foyo-Moreno et al. (2014). Furthermore, the Sun's position above the horizon directly affects the solar radiation. In our context, Saudi Arabia mostly experiences clear-skies; hence GHI mainly depends on aerosols, generally quantified in terms of aerosol optical depth (AOD), and solar zenith angle, which is the angle between the zenith and the center of the Sun's disk. Mathematically, if θ_t represents the zenith angle at time point t , then

$$\cos(\theta_t) = \sin(\Phi) \sin(\delta_t) + \cos(\Phi) \cos(\delta_t) \cos(h_t),$$

where Φ denotes the local latitude, δ_t represents the declination of the Sun, and h_t is the solar hour angle at time point t . Therefore, we incorporate AOD and zenith angle in our GHI model. However, this incorporation also introduces our first challenge, that is, AOD is not consistently monitored at all times. While GHI and zenith angle were measured every hour during the 33-day measurement at the KAUST solar station, AOD was mostly measured during the daytime each day, with a few missing observations during this period. We handle this issue by partitioning the 24-hours interval into the daily time-interval during which AOD was almost always measured and the daily time-interval during which it was almost never measured. For the former interval, we interpolate the few missing AOD data using splines. For the latter interval, we do not consider AOD when modeling GHI. Indeed, during the latter interval, most AODs have significantly less magnitudes and, therefore, become difficult to be properly measured by the monitoring mechanism. That is why, we anticipate that disregarding AODs in the latter interval has minimal, if any, impact on forecasting.

The second challenge is related to the non-stationarity of the covariance structure. As detailed in Subsection 2.2, the GHI data show a clear departure from stationarity through both its mean and covariance functions. The mean function departs from stationarity by being periodic over time, which is expected because of the periodic nature of the motions of the Sun. However, interestingly enough, the correlation structure of the GHIs also exhibits a periodic pattern. This observation motivates us to fit its correlation structure using a periodically correlated or cyclostationary stochastic process; see Hurd and Miamiee (2007) and Napolitano (2019) for an introduction of cyclostationary processes. We recall that a random sequence is called cyclostationary if it shows a periodic pattern in its mean and covariance structures. That is, to put mathematically, we call a second order stochastic process $\{X_t, t \in \mathbb{Z}\}$ cyclostationary (or periodically correlated) with period T if, for every $s, t \in \mathbb{Z}$,

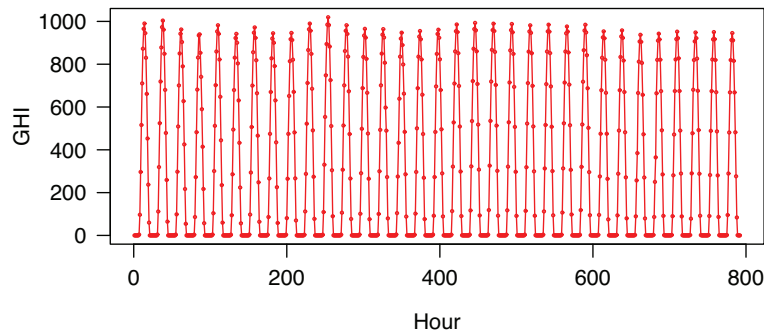


FIGURE 1 Plot of the hourly solar GHIs, $\{y_t, t = 1, \dots, n\}$, as observed from 00:00 on June 2, 2013 to 23:00 of July 4, 2013

$$m(t) = E(X_t) = m(t + T) \quad \text{and} \quad R(s, t) = E[\{X_s - m(s)\} \{X_t - m(t)\}] = R(s + T, t + T)$$

holds, where T is the smallest positive integer that satisfies the above two identities. Upon identifying the observed series as cyclostationary, we fit a periodic autoregressive moving average (PARMA) model to the GHI data and provide forecasts several hours ahead. We compare the forecasts with a classical stationary model that does not take account of the cyclostationary structure of the series. Then, we perform simulations based on each model and use the simulated samples to estimate power outputs by a standard PV power system to provide a further comprehensive comparison. As a result, in all the scenarios, the cyclostationary model is shown to perform better than the stationary one. That is, the proposed model can produce more dependable forecasts, realistic simulations, and, thence, more reliable power estimates than the other one.

In brief, we model the deterministic part of the GHI using the covariates AOD and zenith angle, while the stochastic part is modeled using a cyclostationary PARMA process. The remaining article is organized as follows. In Section 2, we introduce the GHI, AOD, and zenith angle datasets, present preliminary analyses, identify the periodic correlation structure of the GHI data, determine the relation between GHI and AOD after accounting for the solar zenith angle of the Sun, and finally propose the model. In Section 3, we fit the proposed model, forecast 48 hours ahead, generate simulated series, and calculate the PV power outputs using the simulated series. The forecasts, simulations, and PV power estimates achieved using the proposed model are compared with those obtained using a classical stationary model. In Section 4, the findings of this article are summarized and their significance and possible future extensions are discussed.

2 | PRELIMINARY ANALYSIS AND MODEL DETERMINATION

2.1 | GHI data from KAUST

The analyzed GHI dataset contains hourly solar GHIs in watts per square meter (Wh/m^2) from 00:00 on June 2, 2013 to 23:00 on July 4, 2013, that is, over a period of 33 consecutive days. Therefore, the dataset contains a total of $33 \times 24 = 792$ regularly spaced data-points. The data were collected at the KAUST solar monitoring station, Thuwal, Saudi Arabia, and prepared by King Abdullah City for Atomic and Renewable Energy (K.A.CARE) as part of the Renewable Resource Monitoring and Mapping (RRMM) program. They can be accessed via the website <http://rratlas.kacare.gov.sa/>. For convenience of analysis, we represent the GHI data using a trajectory $\{y_t, t = 1, \dots, n\}$, where the subscript t indicates the t th time point and $n = 792$ denotes the sample size.

2.2 | Preliminary analysis of GHI dataset

We begin our analysis by plotting the trajectory of solar GHIs; see Figure 1. Eyeballing the plot suggests no significant (increasing or decreasing) trend over time but a clear periodic structure in the data. The first six hours (0–5) and final three hours (21–23) of each day represent early morning and late night hours, justifying the zero-valued GHI values for those hours.

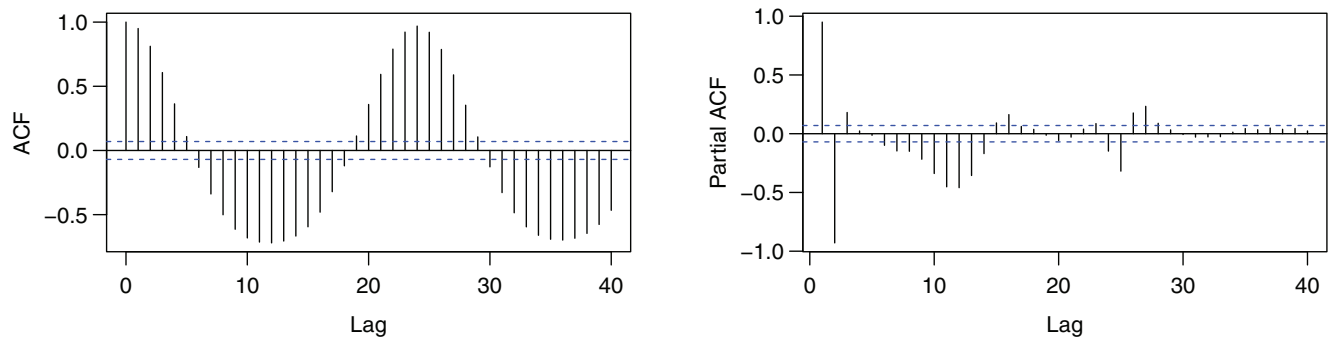


FIGURE 2 ACF (left) and PACF (right) plots of the hourly solar GHI data $\{y_t, t = 1, \dots, n\}$

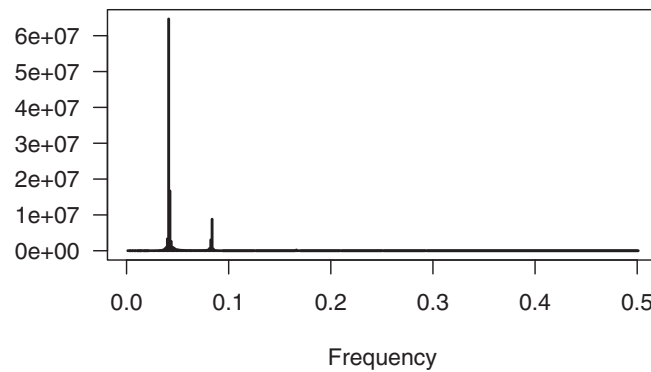


FIGURE 3 Periodogram plot of the hourly solar GHI data $\{y_t, t = 1, \dots, n\}$

To summarize the strength of the relationship between an observation of the time series $\{y_t, t = 1, \dots, n\}$ and observations at previous time points, the autocorrelation function (ACF) and partial ACF (PACF) are plotted in Figure 2. From the figure, it may seem that a standard stationary autoregressive moving average (ARMA) model suits the data; however, as we will see later, this can also be indicative to the presence of a “hidden” periodic structure in the covariance function.

Both Figure 1 and the ACF plot in Figure 2 indicate periodicity with a period of 24 in the series. This observation is expected as the data-points were collected hourly. However, we further methodically verify this conjecture using the classical tool of plotting the periodogram; see Figure 3. The most significant peak of the periodogram plot corresponds to the frequency ≈ 0.042 , affirming the dominant period of $0.042^{-1} \approx 24$ for the underlying series.

Considering the period of 24 for the time series, we provide a graphical summary of the entire dataset in the top panel of Figure 4 using periodic boxplots of the time series. The bottom left and right panels of Figure 4 show the estimates of the sample periodic mean and standard deviation, respectively, which are defined in the sequel, along with their 95% confidence intervals. By noting that the time series contains $n = NT = 792$ observations, where the number of cycles, $N = 33$, and the length of period, $T = 24$, we calculate the sample periodic mean of the time series as

$$\hat{m}_y(t) = \frac{1}{N} \sum_{j=0}^{N-1} y_{t+jT}, \quad t = 1, \dots, T,$$

and the sample periodic variance (or the square of the sample periodic standard deviation) as

$$\hat{R}_y(t, t) = \frac{1}{N-1} \sum_{j=0}^{N-1} \{y_{t+jT} - \hat{m}(t+jT)\}^2, \quad t = 1, \dots, T.$$

In addition to the sample periodic mean and standard deviation, we calculate the sample periodic autocorrelation functions (PeACFs) of the time series $\{y_t, t = 1, \dots, n\}$ for lag $h = 1, \dots, 12$, and the corresponding 95% confidence intervals. Figure 5 plots the PeACFs for $h = 1, \dots, 6$. The PeACF plots for $h = 7, \dots, 12$ are provided in Supplementary Material. The PeACF is defined as

$$\rho_s(h) = \frac{\gamma_s(h)}{\sqrt{\gamma_s(0)\gamma_{s-h}(0)}}, \quad s = 1, \dots, T, \quad h \geq 0, \quad (1)$$

where $\gamma_s(h) = R(rT + s, rT + s - h)$ at season s and lag $h \geq 0$, and the sample counterpart of that is defined as

$$\hat{\rho}_s(h) = \frac{\hat{\gamma}_s(h)}{\sqrt{\hat{\gamma}_s(0)\hat{\gamma}_{s-h}(0)}},$$

for $s = 1, \dots, T$ and $h = 0, \dots, (N-1)T + s - 1$, where $\hat{\gamma}_s(h) = \sum_{r=r^*}^{N-1} (y_{rT+s} - \bar{y}_s)(y_{rT+s-h} - \bar{y}_{s-h})$ and r^* is the smallest integer satisfying $r^*T + s > h$; see Shao and Lund (2004).

Figures 4 and 5 clearly indicate that the estimated periodic mean, standard deviation, and sample PeACFs at various lags significantly vary with the time of the day. Thus, the data are certainly not stationary; rather a cyclostationary model seems appropriate.

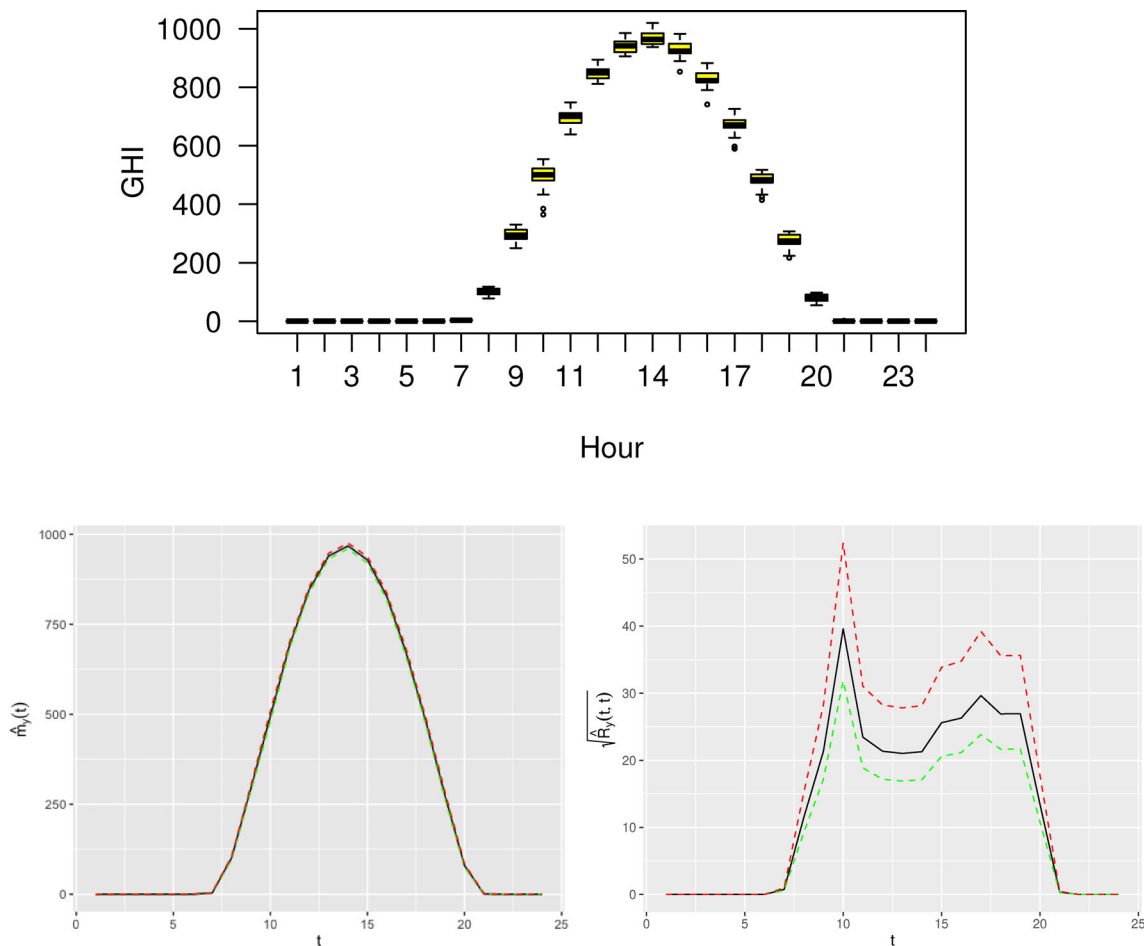


FIGURE 4 Periodic boxplots (top), estimated periodic mean (in black at bottom left), and estimated periodic standard deviation (in black at bottom right) of the solar GHI data. In the bottom panels, the dashed green and red curves in the plots of estimated periodic mean and standard deviation represent the lower and upper 95% confidence intervals, respectively

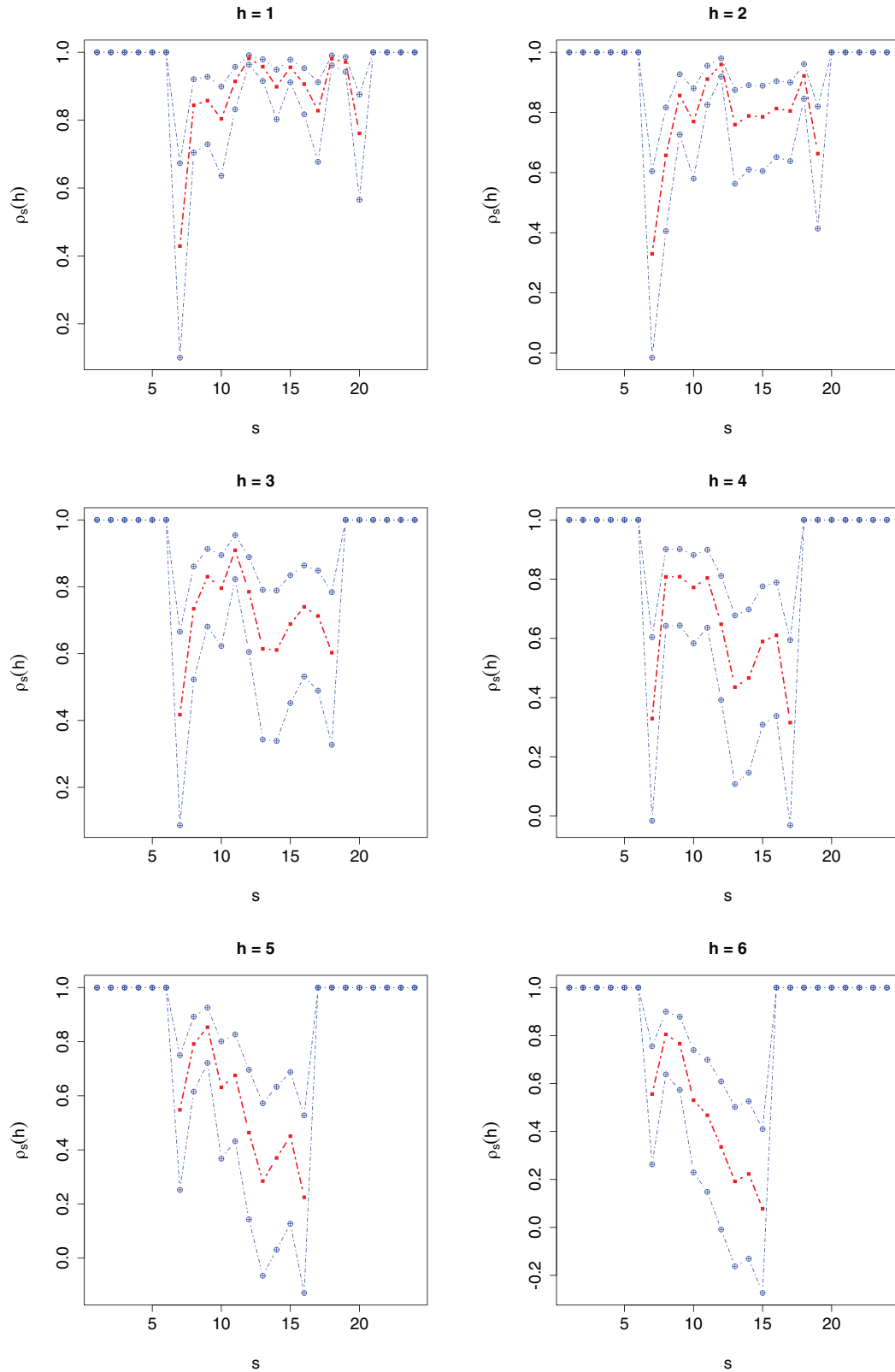


FIGURE 5 Plots of the sample PeACFs (red) and corresponding 95% confidence intervals (blue) of the time series $\{y_t, t = 1, \dots, n\}$ for lag $h = 1, \dots, 6$, respectively

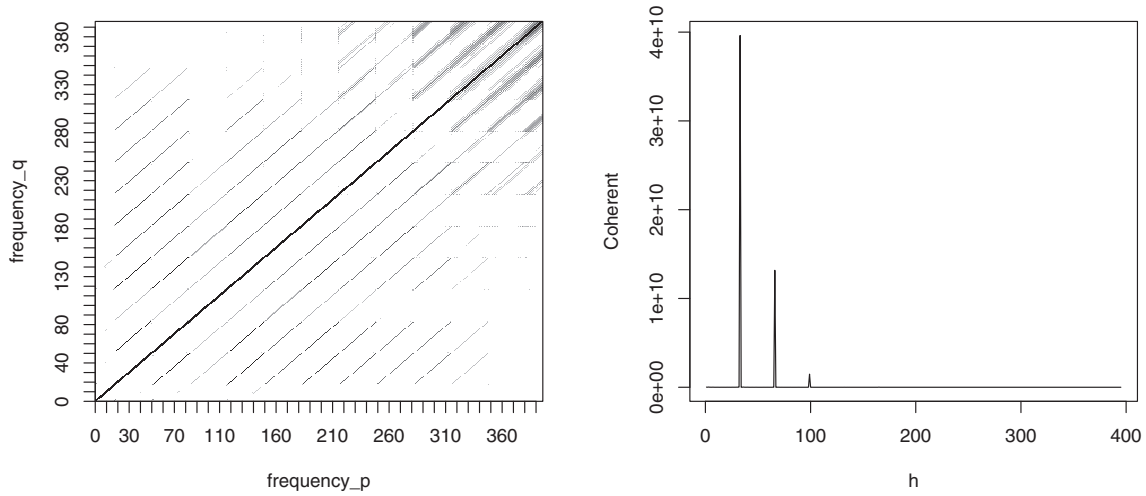


FIGURE 6 Sample squared coherence (left) and coherent statistic (right) plot of the solar GHI data $\{y_t, t = 1, \dots, n\}$

2.3 | Identifying periodic correlation in GHI series

In this subsection, we carry out two popular coherence-based methods to determine the cyclostationarity of the series $\{y_t, t = 1, \dots, n\}$. We begin by introducing the sample squared coherence (Hurd and Gerr 1991),

$$|C(p, q, M)|^2 = \frac{\left| \sum_{m=0}^{M-1} d_n(\omega_{p+m}) \overline{d_n(\omega_{q+m})} \right|^2}{\sum_{m=0}^{M-1} |d_n(\omega_{p+m})|^2 \sum_{m=0}^{M-1} |d_n(\omega_{q+m})|^2},$$

where $d_n(\omega_k) = \frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} y_t \exp(-i\omega_k t)$ is the discrete Fourier transform of $\{y_t, t = 1, \dots, n\}$, M is the smoothing parameter, and $\omega_k = 2\pi k/n$ for $k = 0, \dots, n-1$. In a sample squared coherence plot, lines parallel to the main diagonal at distance $h = kn/T, k = 0, \dots, T-1$, indicate that the series is cyclostationary with period T . Other methods for detecting cyclostationarity can be found in Hurd and Gerr (1991), Martin and Kedem (1993), Dandawate and Giannakis (1994), Bloomfield et al. (1994), Broszkiewicz-Suwaj (2003), Broszkiewicz-Suwaj et al. (2004), Wang et al. (2005), Mahmoudi and Maleki (2017), and Pries et al. (2018).

The sample squared coherence of the GHI is plotted in the left panel of Figure 6 with the choice of smoothing parameter, $M = 32$; refer to Hurd and Miamiee (2007) regarding the choice of the smoothing parameter. The figure shows clear lines parallel to the main diagonal, indicating that the series is cyclostationary. Notably, lines as clear as those observed in the plot corresponding to GHI are rarely observed in cyclostationarity analysis of real-world data. This finding implies that the GHI can be modeled very well using a cyclostationary process. Moreover, it can be considered an ideal real-life example of a cyclostationary process.

Another related method for confirming cyclostationarity requires the plotting of the coherent statistic; see the right panel of Figure 6. The coherent statistic is defined as $|C(h, 0, n)|^2$; see Hurd and Gerr (1991) for a detailed discussion. In the coherent statistic plot, a significant peak at $h = h_0$ indicates that the underlying process is cyclostationary with a period of n/h_0 . In our scenario, the plot exhibits a prominent peak at $h = 33$, implying that the series is cyclostationary, with period $T = 792/33 = 24$.

Although the above discussion confirms the cyclostationarity of the data, the possibility that the data comes from merely a cyclostationary white noise or an amplitude-modulated stationary process persists. Therefore, we conduct two additional tests on the PeACF of the series, as defined in (1), to substantiate that the underlying process is a proper cyclostationary one. First, we consider the hypothesis

$$H_0^{(h)} : \rho_s(h) = 0 \quad \text{for } s = 1, \dots, T,$$

for different $h \neq 0$. If $H_0^{(h)}$ is rejected for some $h \neq 0$, we conclude that the sequence is not cyclostationary white noise; see Dudek et al. (2015). This is the case for the considered series $\{y_t, t = 1, \dots, n\}$ as the hypothesis $H_0^{(h)}$ is rejected for each $h \neq 0$.

Next, to check that the process is not simply an amplitude-modulated stationary sequence, we test the null hypothesis

$$\tilde{H}_0^{(h)} : \rho_s(h) = \rho(h) \quad \text{for } s = 1, \dots, T,$$

for different $h \neq 0$. Here, if $\tilde{H}_0^{(h)}$ is rejected for some $h \neq 0$, then we cannot decline that the sequence comes from a proper cyclostationary process; see Dudek et al. (2015). For the series $\{y_t, t = 1, \dots, n\}$, the hypothesis $\tilde{H}_0^{(h)}$ also gets rejected for all $h \neq 0$ indicating that the observed data are not generated by an amplitude-modulated stationary process. Therefore, we can assume that the data are generated by a proper cyclostationary process.

2.4 | AOD and zenith angle data from KAUST

The variations in AOD are considered at 500 nm wavelength from 06:00 to 15:00 each day from June 2, 2013 to July 4, 2013 (i.e., on the same days as the GHI dataset), with some missing data points. Therefore, the dataset contains 33 consecutive days of observations. The dataset was measured at the AERONET site on the KAUST campus.

Prior to further analysis, we interpolate the missing data points. For this, we fit an interpolating (natural) cubic spline for each of the 33 days and estimate the missing data points by evaluating the value of the fitted curve at each of those time points. We recall that natural cubic splines are piece-wise cubic polynomials that are twice continuously differentiable; see McKinley and Levine (1998). The completed AOD dataset, with 330 data points (10 hours for each of the 33 days), is plotted in the top panel of Figure 7. We note that the data points are not regularly spaced because the AOD was observed only from 06:00 to 15:00 each day. A few significant fluctuations are observed in the AOD plot. Those are either caused by severe dust storms or heavy rains.

The dataset containing the zenith angles (in degrees) at which the GHIs are observed is also considered from 06:00 to 15:00 each day from June 2, 2013 to July 4, 2013. The middle panel of Figure 7 plots the zenith angle dataset. The data were collected along with GHI at the KAUST solar monitoring station.

Hereafter, we use x_t and z_t to indicate AOD and zenith angle, respectively, at time point t .

2.5 | Effect of AOD as a regressor on GHI

To understand how AOD affects GHI, we plot AOD versus GHI (observed from 06:00 to 15:00 each day from June 2, 2013 to July 4, 2013), after accounting for the zenith angle of the Sun. In other words, we plot $x_t / \cos(z_t)$ versus $y_t / \cos(z_t)$; see the bottom panel of Figure 7.

From the plot, an exponential decay in GHI is evident as AOD increases; see also Pai (1966). This finding affirms the dependence of GHI on AOD and provides insights for GHI model construction with AOD incorporated in its mean function.

2.6 | Model determination

So far, the analyses confirm the following points:

1. Although the GHI has been observed at every hour, AOD was mostly observed during the daytime each day in the time range from June 2, 2013 to July 4, 2013. This suggests that our model for GHI should be partitioned into two scenarios, that is, when AOD is available and when it is unavailable.
2. Figure 1 suggests that the GHI takes zero values between 20:00 each day and 6:00 the following day.
3. GHI exhibits exponential decay with increasing AOD.

Considering the above three observations, we propose the following model.

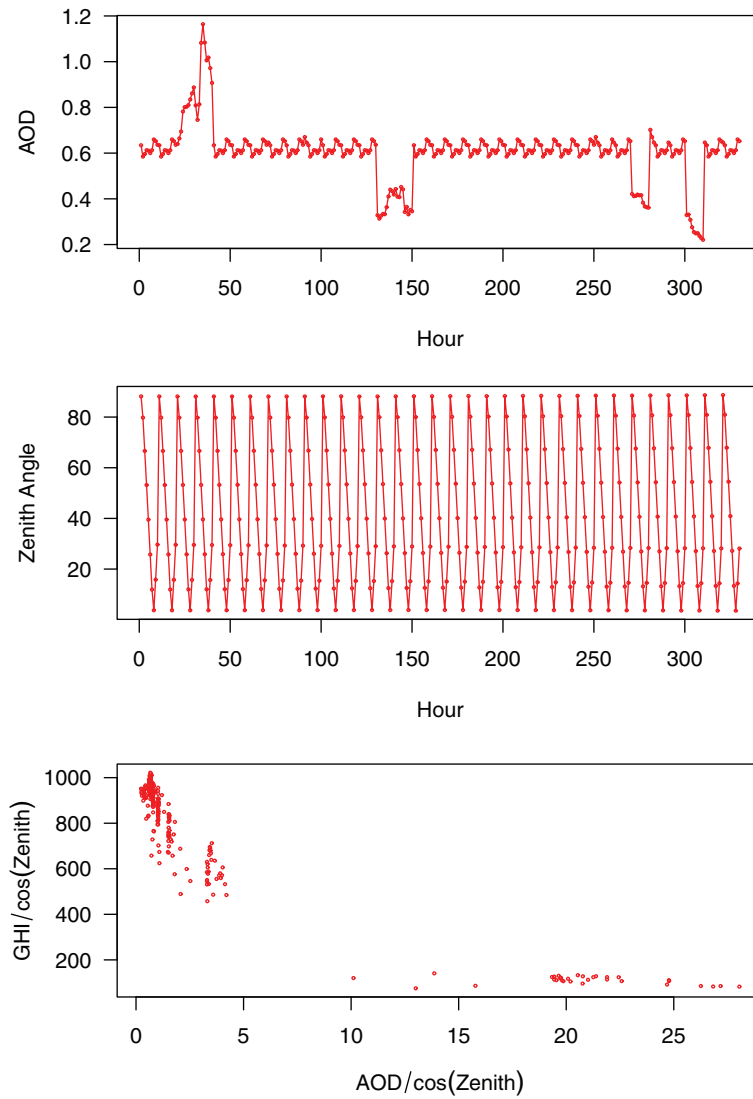


FIGURE 7 Top panel: plot of the AOD dataset; middle panel: plot of the zenith angle dataset; bottom panel: scatter plot between AOD and GHI dataset, after accounting for the solar zenith angle. Each of these datasets are considered from 06:00 to 15:00 each day from June 2, 2013 to July 4, 2013

Proposed Model: Let the time domain \mathbb{D} on which the GHIs are observed be partitioned into four disjoint and exhaustive subsets:

- \mathbb{D}_1 : 00:00–05:00 each day (early-morning hours),
- \mathbb{D}_2 : 06:00–15:00 each day (daytime hours),
- \mathbb{D}_3 : 16:00–20:00 each day (evening hours), and
- \mathbb{D}_4 : 21:00–23:00 each day (late-night hours).

We then consider the model

$$y_t = \begin{cases} 0, & \text{if } t \in \mathbb{D}_1, \\ \alpha + \beta \exp \{ -x_t / \cos(z_t) \} \cos(z_t) + \epsilon_t, & \text{if } t \in \mathbb{D}_2, \\ \xi_t, & \text{if } t \in \mathbb{D}_3, \\ 0, & \text{if } t \in \mathbb{D}_4, \end{cases} \quad (2)$$

where ϵ_t and ξ_t are cyclostationary residuals. The parameters α and β are unknown constants.

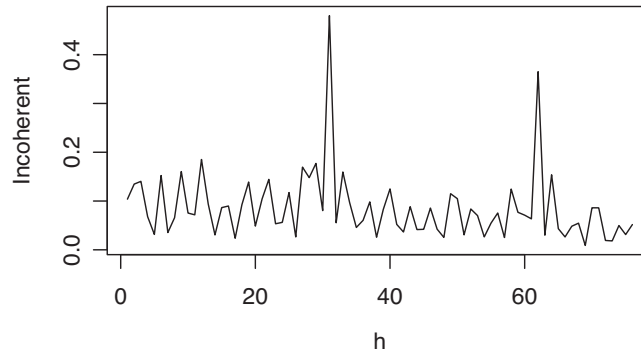


FIGURE 8 Incoherent statistic plot of $\{\xi_t, t \in \mathbb{D}_3\}$

3 | MODEL FITTING

This section describes the fitting of model (2), in which we estimate the unknown parameters α and β and fit appropriate cyclostationary models for ϵ_t and ξ_t . For this purpose, we use the first 31 days (i.e., 744 hours) of the considered time-interval for fitting and reserve the remaining 2 days (i.e., 48 hours) for validating the model.

We begin by fitting a cyclostationary model to $\{\xi_t, t \in \mathbb{D}_3\}$ while extensively discussing the required steps to do so, in Subsection 3.1. Then, in Subsection 3.2, we fit the model for the domain \mathbb{D}_2 using an iterative algorithm.

3.1 | Model fitting for evening hours

The above analysis indicates that the underlying process of the considered sequence of solar GHI is cyclostationary with period $T = 24$. Generally, this finding should also imply that the partial series, when $t \in \mathbb{D}_3$, is also cyclostationary with period $T_1 = 5$ (because the partial series considers 5 hours of each day). To ensure the validity of this statement, we plot the incoherent statistic for $\{\xi_t, t \in \mathbb{D}_3\}$ in Figure 8. The incoherent statistic is defined as

$$\delta(h, M) = \frac{1}{L+1} \sum_{j=0}^L |C(jM, jM+h, M)|^2,$$

where M is the smoothing parameter, $L = \lfloor (n_1 - T_1 - 1)/M \rfloor$, and n_1 is the cardinality of the set $\mathbb{D}_3 = 5 \times 31 = 155$; see Hurd and Gerr (1991). In the incoherent statistic plot, a spike at $h = h_0$ implies that the underlying process is cyclostationary with a period of n_1/h_0 .

For the current scenario, we considered the value of the smoothing parameter, $M = 32$. The plot in Figure 8 shows a significant peak at $h = 31$, implying the cyclostationarity of the underlying process, with period $T_1 = 155/31 = 5$.

Considering that the series $\{\xi_t, t \in \mathbb{D}_3\}$ is cyclostationary, a PARMA model should fit the process, $\{\xi_t, t \in \mathbb{D}_3\}$, reasonably well. Prior to fitting, we remove the mean of the series,

$$\hat{m}_\xi(t) = \frac{1}{N_1} \sum_{j=0}^{N_1-1} \xi_{t+jT_1}, \quad t = 1, \dots, T_1,$$

where N_1 = number of cycles present in the series = 31, to have $\{\tilde{\xi}_t = \xi_t - \hat{m}_\xi(t), t \in \mathbb{D}_3\}$; the demeaned series is plotted in Figure 9.

We recall that a zero mean time series $\{\tilde{\xi}_t, t \in \mathbb{D}_3\}$ is PARMA(p, q) with period T_1 if it satisfies

$$\tilde{\xi}_t = \sum_{j=1}^p \phi_j(t) \tilde{\xi}_{t-j} + \sum_{k=1}^q \theta_k(t) \delta_{t-k} + \sigma(t) \delta_t, \quad (3)$$

where $\delta_t \stackrel{iid}{\sim} (0, 1)$, that is, $\{\delta_t\}$ are drawn independently from a random variable with mean zero and variance one. Here $\phi_j(t) = \phi_j(t + T_1)$ for $j = 1, \dots, p$, $\theta_k(t) = \theta_k(t + T_1)$ for $k = 1, \dots, q$, and $\sigma(t) = \sigma(t + T_1)$ are the model parameters.

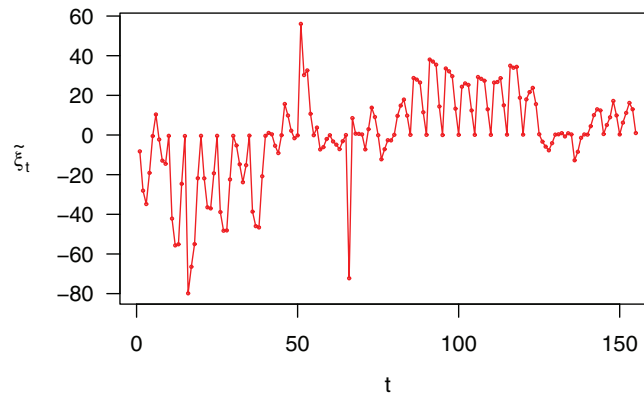


FIGURE 9 Plot of the demeaned series $\{\tilde{\xi}_t, t \in \mathbb{D}_3\}$. The x-axis plots the time (in hours) within the domain \mathbb{D}_3

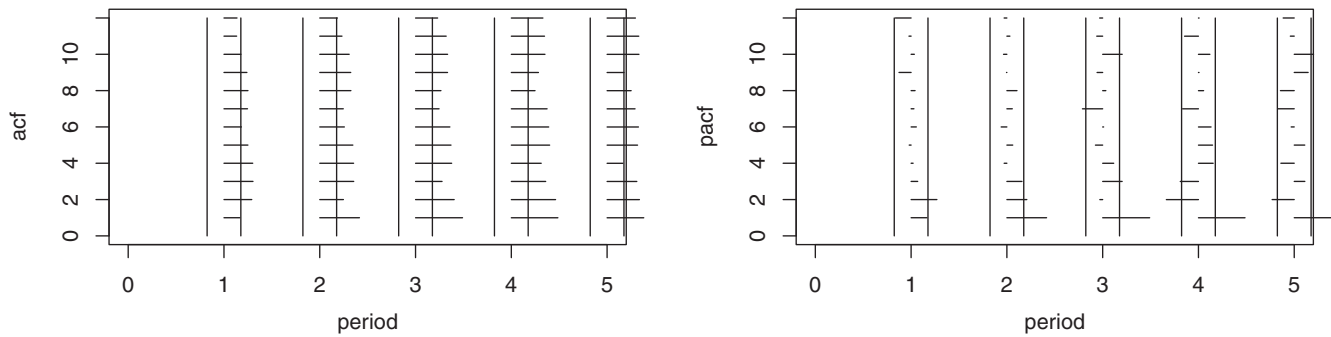


FIGURE 10 PeACF (left) and PePACF (right) plots of $\{\tilde{\xi}_t, t \in \mathbb{D}_3\}$

TABLE 1 Parameter estimates for the model (3)

	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$
$\hat{\phi}_1(t)$	-26.866	0.634	0.998	0.876	0.049
$\hat{\phi}_2(t)$	1.781	30.768	-0.019	-0.402	-0.015
$\hat{\sigma}(t)$	4.929	3.569	2.291	1.529	0.434

To fit a PARMA model to $\{\tilde{\xi}_t, t \in \mathbb{D}_3\}$, we must determine the periodic AR and MA orders, p and q , respectively. For this, we use the PeACF and periodic PACF (PePACF) plots illustrated in the left and right panels of Figure 10, respectively. A smooth periodic movement is observed in the PeACF plot while a sudden drop is observed in the PePACF plot. Therefore, we deduce a $\text{PAR}(p)$ (i.e., $\text{PARMA}(p, 0)$) model is a reasonable fit to the demeaned process.

To determine the order p of the PAR model, we note the significant spikes (at 95% confidence levels) in the PePACF plot up to a lag of 2; this suggests that a $\text{PAR}(2)$ model should fit $\{\tilde{\xi}_t, t \in \mathbb{D}_3\}$ significantly well. Now, to estimate the parameters $\phi_1(t)$, $\phi_2(t)$, and $\sigma(t)$ of the model (3) for $t = 1, \dots, T_1$, we use the Yule-Walker method; see Gladyshev (1961), Pagano (1978), Brockwell et al. (1991), and Box et al. (2011). The parameter estimates, $\hat{\phi}_1(t)$, $\hat{\phi}_2(t)$, and $\hat{\sigma}(t)$ for $t = 1, \dots, T_1$, are listed in Table 1.

3.2 | Model fitting for daytime hours

To estimate the parameters α and β , we use the least squares technique because it does not require any distributional assumption of the residuals, $\{\epsilon_t, t \in \mathbb{D}_2\}$. However, estimating the parameters using a least squares method is not straightforward because the residuals are periodically correlated and, thus, we need to use a generalized least

TABLE 2 Parameter estimates of the demeaned PAR(1) model obtained in the 17th iteration

	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$	$t = 7$	$t = 8$	$t = 9$	$t = 10$
$\varphi^{(17)}(t)$	0.002	6.702	2.320	1.787	0.977	1.084	1.042	1.073	0.931	0.926
$\omega^{(17)}(t)$	0.808	3.600	4.251	5.078	5.024	3.116	3.046	3.907	3.860	3.459

squares (GLS) method after estimating the covariance matrix of the residuals. At the same time, α and β must be known to estimate the covariance matrix. Therefore, we use an iterative method in which we start with ordinary least squares (OLS) estimates (which does not consider the correlated structure of the residuals) of the parameters, calculate the covariance matrix of the residuals, update the parameter estimates using GLS, and so on.

Before providing the iterative algorithm, we note that, similar to Subsection 3.1, the $\{\epsilon_t, t \in \mathbb{D}_2\}$ is cyclostationary with period, $T_2 = 10$, because we are now considering 10 hours of each day. Hereafter, the cardinality of the set \mathbb{D}_2 is denoted by $n_2 (= 31 \times 10 = 310)$. Also, we have decided to fit a PAR(1) model to the residuals, $\{\epsilon_t, t \in \mathbb{D}_2\}$, after conducting an analysis similar to that in Subsection 3.1. The superscript “(k)” in the algorithm denotes the k th iteration, while “(0)” denotes the initial estimates. So, the algorithm to estimate α and β then goes as follows:

- Step 1: Find the OLS estimates, $\hat{\alpha}^{(0)}$ and $\hat{\beta}^{(0)}$, of α and β , respectively.
- Step 2: Extract the residuals $\hat{\epsilon}_t^{(0)} = y_t - \hat{\alpha}^{(0)} - \hat{\beta}^{(0)} \exp\{-x_t / \cos(z_t)\} \cos(z_t)$, $t \in \mathbb{D}_2$.
- Step 3: Demean the estimated residuals to obtain $\{\epsilon_t^{(0)} = \hat{\epsilon}_t^{(0)} - \hat{m}_{\hat{\epsilon}}^{(0)}(t), t \in \mathbb{D}_2\}$, where $\hat{m}_{\hat{\epsilon}}^{(0)}(t) = \frac{1}{N_2} \sum_{j=0}^{N_2-1} \hat{\epsilon}_{t+jT_2}^{(0)}$, $t = 1, \dots, T_2$, and $N_2 =$ number of cycles present in the series ($= 31$).
- Step 4: Fit a PAR(1) model to $\{\epsilon_t^{(0)}, t \in \mathbb{D}_2\}$: $\epsilon_t^{(0)} = \varphi^{(0)}(t)\epsilon_{t-1}^{(0)} + \omega^{(0)}(t)\zeta_t^{(0)}$, $t \in \mathbb{D}_2$, where, $\zeta_t^{(0)} \stackrel{iid}{\sim} (0, 1)$. Here, $\varphi^{(0)}(t) = \varphi^{(0)}(t + T_2)$ and $\omega^{(0)}(t) = \omega^{(0)}(t + T_2)$ are model parameters.
- Step 5: Estimate the parameters and denote them by $\hat{\varphi}^{(0)}(t)$ and $\hat{\omega}^{(0)}(t)$ for $t = 1, \dots, T_2$.
- Step 6: Calculate the covariance matrix of $\{\epsilon_t^{(0)}, t \in \mathbb{D}_2\}$ and denote it as $\Omega^{(0)}$. The derivation of the covariance function, $\Omega_{s,t}^{(0)}$, is presented in the Appendix.
- Step 7: Estimate the parameters α and β using GLS and denote them as $\hat{\alpha}^{(1)}$ and $\hat{\beta}^{(1)}$, respectively, where

$$\begin{pmatrix} \hat{\alpha}^{(1)} \\ \hat{\beta}^{(1)} \end{pmatrix} = \left\{ \Lambda^T (\Omega^{(0)})^{-1} \Lambda \right\}^{-1} (\Omega^{(0)})^{-1} \mathbf{y},$$

$$\text{where } \Lambda = \begin{pmatrix} 1 & \dots & 1 \\ e^{-x_1 / \cos(z_1)} \cos(z_1) & \dots & e^{-x_{n_2} / \cos(z_{n_2})} \cos(z_{n_2}) \end{pmatrix}^T \text{ and } \mathbf{y} = (y_t)_{t \in \mathbb{D}_2}.$$

- Step 8: Repeat Steps 2–7 until convergence. That is, stop at k th step if the Euclidean distance,

$$\sqrt{(\hat{\alpha}^{(k)} - \hat{\alpha}^{(k-1)})^2 + (\hat{\beta}^{(k)} - \hat{\beta}^{(k-1)})^2},$$

is less than some predefined small quantity.

For the considered dataset, the initial OLS estimates are $\hat{\alpha}^{(0)} = 124.47$ and $\hat{\beta}^{(0)} = 1538.27$, both being statistically significant (i.e., different from zero). Their final estimates, obtained at the 17th iteration, are $\hat{\alpha}^{(17)} = -5.29$ and $\hat{\beta}^{(17)} = 1091.08$, matching the estimates of the 16th iteration up to 6 decimal places. The parameters of the PAR(1) model fitted to the residuals are considered to be the estimates achieved in the final iteration. Table 2 presents the estimates of the demeaned PAR(1) model.

3.3 | Short-term forecast and importance of considering periodic correlation structure

As we complete the fitting of our proposed model (2), this subsection carries out a diagnostic checking by comparing the last 48 hours of true solar GHI values (which were not used while modeling) with the corresponding values predicted

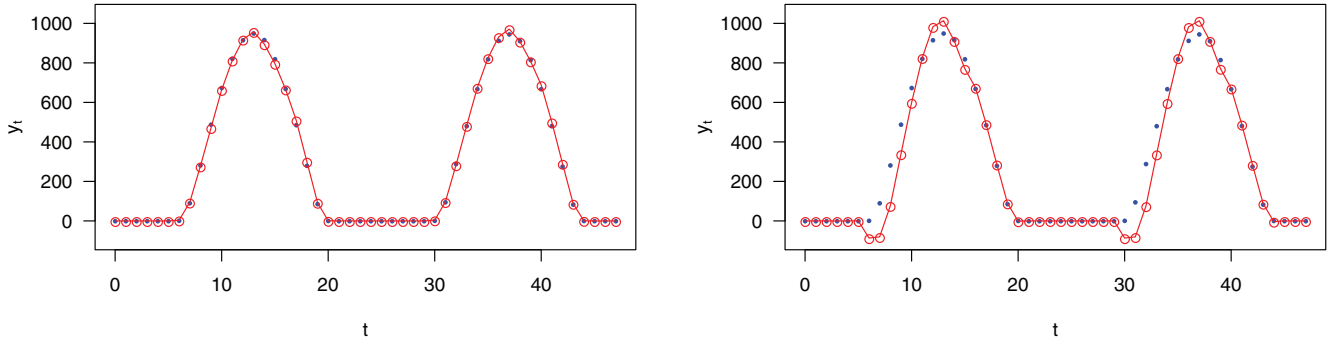


FIGURE 11 Left: Plot of the last 48 hours of observed solar GHIs (blue) and the corresponding predicted values (red) using model (2). Right: Plot of the last 48 hours of observed solar GHIs (blue) and the corresponding predicted values (red) without considering the periodic structure of the correlation

using our model. The GHIs are predicted as

$$y_t^{\text{pred}} = \begin{cases} 0, & \text{if } t \in \mathbb{D}_1, \\ \hat{\alpha}^{(17)} + \hat{\beta}^{(17)} \exp \{-x_t / \cos(z_t)\} \cos(z_t) + \hat{m}_{\varepsilon}^{(16)}(t) + \hat{\varphi}^{(16)}(t) \varepsilon_{t-1}^{(16)}, & \text{if } t \in \mathbb{D}_2, \\ \hat{m}_{\xi}(t) + \hat{\phi}_1(t) \tilde{\xi}_{t-1} + \hat{\phi}_2(t) \tilde{\xi}_{t-2}, & \text{if } t \in \mathbb{D}_3, \\ 0, & \text{if } t \in \mathbb{D}_4. \end{cases}$$

The left panel of Figure 11 plots the forecasted GHIs (in red) superimposed on the observed values (in blue). The plot suggests that the proposed model performs quite well in evaluating short-term forecasts. The success of the model is expected because the correlation structure of the GHI was well explained through cyclostationarity, as also noted in Figure 6 of Section 2.3.

In contrast, a negligent analysis of the ACF and PACF plots in Figure 2 might disregard the periodic structure of the correlation and suggest a classical stationary model, such as ARMA, to fit the stochastic part of the data. An analysis similar to the above can be performed to determine the coefficients of the ARMA model. In other words, for $t \in \mathbb{D}_2$, an ARMA model with appropriate AR and MA orders (AR(10), in our case) can be fitted to the stochastic part, ε_t , and its coefficients along with the coefficients of the deterministic part, α and β , can be estimated using an iterative algorithm similar to that described in Subsection 3.2. For $t \in \mathbb{D}_3$, an AR(8) model is fitted to the data after examining the corresponding ACF and PACF plots. Hereafter, we call this model ‘model 2’, our proposed model being ‘model 1’. To evaluate the ‘model 2’, we employ it to forecast the same 48 hours of GHIs as before; see right panel of Figure 11. From the plot, we observe that the ‘model 2’ significantly underperforms compared with the first model, as further verified in the next subsection.

3.4 | Simulation and energy production

As mentioned in the introduction, GHI is instrumental in power estimation; therefore, when GHI data are unavailable, realistic simulations can aid the respective authorities. Considering that the ‘model 1’ produces reliable forecasts, we can also use it to generate GHI series:

$$y_t = \begin{cases} 0, & \text{if } t \in \mathbb{D}_1, \\ \hat{\alpha}^{(17)} + \hat{\beta}^{(17)} \exp \{-x_t / \cos(z_t)\} \cos(z_t) + \hat{m}_{\varepsilon}^{(16)}(t) + \varepsilon_t, & \text{if } t \in \mathbb{D}_2, \\ \hat{m}_{\xi}(t) + \hat{\phi}_1(t) \tilde{\xi}_{t-1} + \hat{\phi}_2(t) \tilde{\xi}_{t-2} + \hat{\sigma}(t) \delta_t, & \text{if } t \in \mathbb{D}_3, \\ 0, & \text{if } t \in \mathbb{D}_4, \end{cases}$$

where $\varepsilon_t = \hat{\varphi}^{(16)}(t) \varepsilon_{t-1} + \hat{\omega}^{(16)}(t) \zeta_t$, $t \in \mathbb{D}_2$. The noise terms, ζ_t and δ_t , are needed to be drawn from independent distributions with mean zero and variance one.

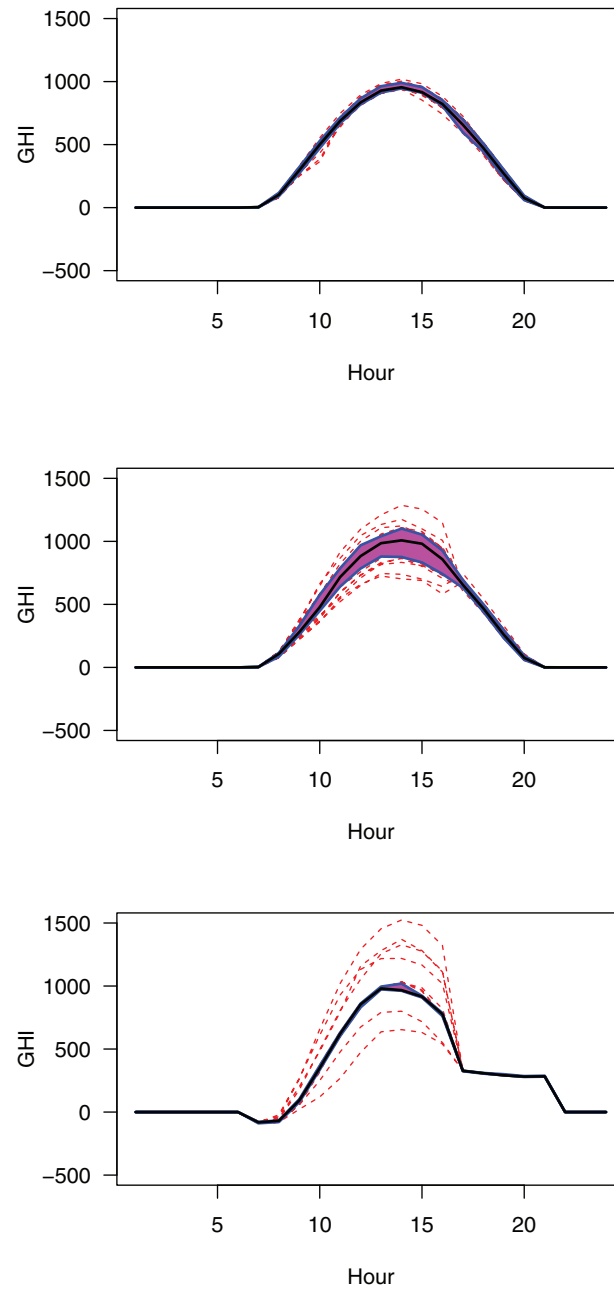


FIGURE 12 Functional boxplots of 33 days of the observed GHI data considering each day as a function (top panel), simulated series using our proposed model (middle panel), and simulated series using a model that does not consider the periodic correlation structure of the GHI data (bottom panel)

Similarly, we generate GHI series using the ‘model 2’. To provide a comprehensive visualization and, thence, compare the simulated samples from the two models with the true data, we present Figure 12. In the top panel of Figure 12, we display a functional boxplot (see Sun and Genton (2011)) of the 33 days of GHI data considering each day a function. Then, we simulate observations of 33 days using each model and put their functional boxplots in the middle and bottom panels, respectively.

From both Figures 11 and 12, we observe high volatility (or, high uncertainties) and even unrealistic negative values while considering the ‘model 2’. On the other hand, the ‘model 1’, that takes the cyclostationarity of the series into account, achieves more realistic forecasts and simulations than the other one.

Considering that the GHI data are mostly used by power system operators as discussed in the introduction, we next estimate the direct current (DC) power outputs by a standard PV system for each of the 14 hours during sunlight (i.e., 06:00–19:00) for the 33-day period for each of the three scenarios presented in Figure 12.

TABLE 3 Specifications to calculate the DC power output

Maximum power (P_{\max})	$106 \pm 5\%$ Wp
Temperature at NOCT (T_{NOCT})	47°C
Reference cell temperature (T_{ref})	25°C
Number of cells in series (s)	4
Number of cells in parallels (p)	2
Tilt angle (θ)	22.3°
Fixed temperature coefficient (γ)	$-0.5\%/^\circ\text{C}$

To compute the DC power output, we use the *PVWatts module model* provided by the National Renewable Energy Laboratory (NREL), a national laboratory of the U.S. Department of Energy, Office of Energy Efficiency and Renewable Energy (see Dobos (2014)):

$$P_{dc} = \frac{G_T}{1000} P_{dc0} \{1 + \gamma(T_{\text{cell}} - T_{\text{ref}})\},$$

where G_T is the global solar irradiance on a tilted surface, P_{dc0} represents a pre-specified nameplate of the DC rating, γ indicates the fixed temperature coefficient, T_{cell} is the computed cell temperature, and T_{ref} is the reference cell temperature. The considered dataset collected at the KAUST solar monitoring station also includes data for air temperature. To obtain the cell temperature, T_{cell} , from the available air temperature data, we use the following model (see Rawat et al. (2016)):

$$T_{\text{cell}} = \text{Temp} + \frac{G_T}{800} (T_{\text{NOCT}} - 20),$$

where Temp indicates the air temperature measured at KAUST and T_{NOCT} is the nominal operating cell temperature (NOCT). To compute the tilted solar irradiance G_T , we used the model provided in Dobos (2014). A standard PV module specifications, used for computing various terms in the above formulae, are summarized in Table 3.

After computing the power outputs for the three sets of GHI series, we calculate the percentage changes in the power outputs with respect to the original GHI series for each model. That is, to put mathematically, consider the estimated power using the true value of GHI at hour h of the d th day is denoted as $\{w_{hd}, h = 6, \dots, 19 \text{ and } d = 1, \dots, 33\}$. Similarly, we denote the estimated power outputs for the ‘model 1’ and ‘model 2’ as $\{u_{hd}, h = 6, \dots, 19, d = 1, \dots, 33\}$ and $\{v_{hd}, h = 6, \dots, 19, d = 1, \dots, 33\}$, respectively. Now, to summarize the overall behaviors of these three sets of power outputs, we calculate the hourly averages of each of these three scenarios, that is,

$$\left\{ \bar{w}_h = \frac{1}{33} \sum_{d=1}^{33} w_{hd}, h = 6, \dots, 19 \right\}, \quad \left\{ \bar{u}_h = \frac{1}{33} \sum_{d=1}^{33} u_{hd}, h = 6, \dots, 19 \right\}, \quad \text{and} \\ \left\{ \bar{v}_h = \frac{1}{33} \sum_{d=1}^{33} v_{hd}, h = 6, \dots, 19 \right\}.$$

Finally, we calculate the percentage relative change between each model’s values and the true values as

$$\left\{ p_h = \frac{|\bar{w}_h - \bar{u}_h|}{\bar{w}_h} \times 100, h = 6, \dots, 19 \right\} \quad \text{and} \quad \left\{ q_h = \frac{|\bar{w}_h - \bar{v}_h|}{\bar{w}_h} \times 100, h = 6, \dots, 19 \right\}.$$

The percentage changes, p_h and q_h , are presented in Table 4.

We observe from Table 4 that, for almost all hours (except the 18th hour), the percentage change with respect to the true power output is smaller for the ‘model 1’, sometimes quite significantly, than for the ‘model 2’. This implies a lesser deviation from the truth while simulating from our model than from the other. For power system operators, this indicates a low associated risk when making a plan or commitment in petroleum dispatching. Therefore, the proposed model can significantly contribute to the economical and reliable operation of power systems.

TABLE 4 Percentage changes of the calculated power outputs using the two considered models with respect to the power output while using the original data

h	6	7	8	9	10	11	12	13	14	15	16	17	18	19
p_h	0.082	0.016	0.009	0.021	0.007	0.009	0.005	0.003	0.002	0.007	0.012	0.014	0.025	0.039
q_h	55.964	1.571	0.666	0.315	0.113	0.035	0.101	0.081	0.036	0.038	0.717	0.421	0.023	2.240

Note: h indicates the hour of the day, and p_h and q_h are the percentage changes of the proposed model and the model that does not consider the periodic correlation structure, respectively.

4 | DISCUSSIONS

In this article, we have provided a model for the short-term (48 hours ahead) forecasting and simulation of solar GHI. For this, we have first recognized the strong dependency between GHI and AOD, after accounting for solar zenith angle. This has helped us in modeling the mean function or the deterministic part of the GHI series. On the other hand, we have identified the periodic correlation structure of the series, which has helped us modeling the random perturbations of the series using a cyclostationary process.

We have shown that our model, that takes cyclostationarity into account, achieves short-term forecasts and simulations that are more accurate and realistic than those of a classical stationary model. The simulated samples are then used to calculate solar power forecasts by a standard PV system. Power output estimates based on dependable methodologies for the forecasting and simulation of solar GHI are expected to significantly benefit power system operators and related authorities of the Kingdom, particularly because Saudi Arabia is still in the initial stage of renewable energy development.

Although we have provided a model only for the solar GHI at a single station (i.e., the KAUST solar monitoring station), our model can also be extended to a cyclostationary spatio-temporal scenario, encompassing other stations in Saudi Arabia. This extension will be investigated in a future project.

ACKNOWLEDGMENT

The authors thank the King Abdullah City for Atomic and Renewable Energy (K.A.CARE) for providing the solar irradiance observational dataset.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from King Abdullah City for Atomic and Renewable Energy. Restrictions apply to the availability of these data, which were used under license for this study. Data are available from <https://rratlas.kacare.gov.sa/> with the permission of King Abdullah City for Atomic and Renewable Energy.

ORCID

Soumya Das  <https://orcid.org/0000-0002-6690-3128>

Marc G. Genton  <https://orcid.org/0000-0001-6467-2998>

REFERENCES

- Almasoud, A., & Gandayh, H. M. (2015). Future of solar energy in Saudi Arabia. *Journal of King Saud University-Engineering Sciences*, 27(2), 153–157.
- AlYahya, S., & Irfan, M. A. (2014). *New solar radiation atlas for Saudi Arabia*. Proceedings of the 2014 International Conference on Renewable Energy Research and Application (ICRERA). (pp. 245–249). Milwaukee; IEEE.
- AlYahya, S., & Irfan, M. A. (2016). Analysis from the new solar radiation atlas for Saudi Arabia. *Solar Energy*, 130, 116–127.
- Bloomfield, P., Hurd, H. L., & Lund, R. B. (1994). Periodic correlation in stratospheric ozone data. *Journal of Time Series Analysis*, 15(2), 127–150.
- Box, G. E., Jenkins, G. M., & Reinsel, G. C. (2011). *Time series analysis: Forecasting and control* (Vol. 734). John Wiley & Sons.
- Brockwell, P. J., Davis, R. A., & Fienberg, S. E. (1991). *Time series: Theory and methods*. Springer Science & Business Media.
- Broszkiewicz-Suwaj, E. (2003). *Methods for determining the presence of periodic correlation based on the bootstrap methodology* (Technical Report Research Report HSC/03/2). Wrocław University of Technology.
- Broszkiewicz-Suwaj, E., Makagon, A., Weron, R., & Wylomańska, A. (2004). On detecting and modeling periodic correlation in financial data. *Physica A: Statistical Mechanics and Its Applications*, 336(1-2), 196–205.
- Dandawate, A. V., & Giannakis, G. B. (1994). Statistical tests for presence of cyclostationarity. *IEEE Transactions on Signal Processing*, 42(9), 2355–2369.

- Dobos, A. P. (2014). *Pvwatts version 5 manual* (Technical Report, NREL/TP-6A20-62641). National Renewable Energy Lab. (NREL), Golden, CO.
- Dudek, A. E., Hurd, H., & Wójtowicz, W. (2015). *PARMA models with applications in R*. In *Cyclostationarity: Theory and methods-II* (pp. 131–153). Springer.
- Foyo-Moreno, I., Alados, I., Antón, M., Fernández-Gálvez, J., Cazorla, A., & Alados-Arboledas, L. (2014). Estimating aerosol characteristics from solar irradiance measurements at an urban location in southeastern Spain. *Journal of Geophysical Research: Atmospheres*, 119(4), 1845–1859.
- Gladyshev, E. (1961). *Periodically correlated random sequences*. In *Doklady Akademii Nauk* (Vol. 137, pp. 1026–1029). Russian Academy of Sciences.
- Hejase, H. A., & Assi, A. H. (2016). *Estimation of global and diffuse horizontal irradiance in Abu Dhabi, United Arab Emirates*. In *Renewable energy in the service of mankind* (Vol. II, pp. 3–14). Springer.
- Hurd, H. L., & Gerr, N. L. (1991). Graphical methods for determining the presence of periodic correlation. *Journal of Time Series Analysis*, 12(4), 337–350.
- Hurd, H. L., & Miamee, A. (2007). *Periodically correlated random sequences: Spectral theory and practice* (Vol. 355). John Wiley & Sons.
- Lee, H.-J., Kim, S.-Y., & Yun, C.-Y. (2017). Comparison of solar radiation models to estimate direct normal irradiance for Korea. *Energies*, 10(5), 594.
- Li, B., & Zhang, J. (2020). A review on the integration of probabilistic solar forecasting in power systems. *Solar Energy*, 207, 777–795.
- Mahmoudi, M. R., & Maleki, M. (2017). A new method to detect periodically correlated structure. *Computational Statistics*, 32(4), 1569–1581.
- Martin, D. E., & Kedem, B. (1993). Estimation of the period of periodically correlated sequences. *Journal of Time Series Analysis*, 14(2), 193–205.
- McKinley, S., & Levine, M. (1998). Cubic spline interpolation. *College of the Redwoods*, 45(1), 1049–1060.
- Myers, D. R., Wilcox, S., Anderberg, M., Al-Awaji, S. H., Al Abbadi, N. M., & bin Mahfoodh, M. Y. (1999). *Saudi Arabian solar radiation network and data for validating satellite remote sensing systems*. In *Earth observing systems IV* (Vol. 3750, pp. 503–513). International Society for Optics and Photonics.
- Napolitano, A. (2019). *Cyclostationary processes and time series: Theory, applications, and generalizations*. Academic Press.
- Navid, N., & Rosenwald, G. (2013). *Ramp capability product design for MISO markets*. In *Market development and analysis*.
- Pagano, M. (1978). On periodic and multiple autoregressions. *The Annals of Statistics*, 6(6), 1310–1317.
- Pai, S.-I. (1966). *Radiation gas dynamics*. Springer.
- Pazheri, F. (2014). Solar power potential in Saudi Arabia. *International Journal of Engineering Research and Applications*, 4(9), 171–174.
- Pries, A., Ramírez, D., & Schreier, P. J. (2018). LMPIT-inspired tests for detecting a cyclostationary signal in noise with spatio-temporal structure. *IEEE Transactions on Wireless Communications*, 17(9), 6321–6334.
- Rawat, R., Kaushik, S., & Lamba, R. (2016). A review on modeling, design methodology and size optimization of photovoltaic based water pumping, standalone and grid connected system. *Renewable and Sustainable Energy Reviews*, 57, 1506–1519.
- Salam, M. A., & Khan, S. A. (2018). Transition towards sustainable energy production—A review of the progress for solar energy in Saudi Arabia. *Energy Exploration & Exploitation*, 36(1), 3–27.
- Shao, Q., & Lund, R. (2004). Computation and characterization of autocorrelations and partial autocorrelations in periodic ARMA models. *Journal of Time Series Analysis*, 25(3), 359–372.
- Sun, Y., & Genton, M. G. (2011). Functional boxplots. *Journal of Computational and Graphical Statistics*, 20(2), 316–334.
- Wang, J., Chen, T., & Huang, B. (2005). Cyclo-period estimation for discrete-time cyclo-stationary signals. *IEEE Transactions on Signal Processing*, 54(1), 83–94.
- Yang, D. (2019). A guideline to solar forecasting research practice: Reproducible, operational, probabilistic or physically-based, ensemble, and skill (ROPES). *Journal of Renewable and Sustainable Energy*, 11(2), 022701.
- Yang, D., Kleissl, J., Gueymard, C. A., Pedro, H. T., & Coimbra, C. F. (2018). History and trends in solar irradiance and PV power forecasting: A preliminary assessment and review using text mining. *Solar Energy*, 168, 60–101.
- Zell, E., Gasim, S., Wilcox, S., Katamoura, S., Stoffel, T., Shibli, H., Engel-Cox, J., & Al Subie, M. (2015). Assessment of solar radiation resources in Saudi Arabia. *Solar Energy*, 119, 422–438.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Das, S., Genton, M. G., Alshehri, Y. M., & Stenchikov, G. L. (2021). A cyclostationary model for temporal forecasting and simulation of solar global horizontal irradiance. *Environmetrics*, 32(8), e2700. <https://doi.org/10.1002/env.2700>

APPENDIX A

Derivation of the covariance function mentioned in Subsection 3.2. Consider a representative PAR(1) model with period T_2 ,

$$\varepsilon_t = \phi(t)\varepsilon_{t-1} + \omega(t)\zeta_t, \quad (\text{A1})$$

where $\zeta_t \overset{iid}{\sim} (0, 1)$, and $\phi(t) = \phi(t + T_2)$ and $\omega(t) = \omega(t + T_2)$ are the model parameters. The theoretical condition for cyclostationarity of (A1) is $\left| \prod_{t=0}^{T_2-1} \phi(t) \right| < 1$. In other words, the derivation below is valid if and only if this condition holds; see Hurd and Miamee (2007). We also define

$$B_0(t) = 1 \quad \text{and} \quad B_k(t) = \prod_{j=0}^{k-1} \phi(t-j), \quad k = 1, \dots, T_2.$$

Then, for $s > t$, the covariance function of $\{\varepsilon_t\}$ can be calculated as

$$\begin{aligned} E(\varepsilon_s \varepsilon_t) &= E[\{\phi(s)\varepsilon_{s-1} + \omega(s)\zeta_s\} \varepsilon_t] \\ &= \phi(s)E(\varepsilon_{s-1} \varepsilon_t) = \phi(s)\phi(s-1)E(\varepsilon_{s-2} \varepsilon_t) = \dots = \sigma_\varepsilon^2(t) \prod_{j=t+1}^s \phi(j), \end{aligned}$$

where $\sigma_\varepsilon^2(t) = E(\varepsilon_t^2)$. To obtain $\sigma_\varepsilon^2(t)$, we start from (A1),

$$\begin{aligned} \sigma_\varepsilon^2(t) &= B_1^2(t)\sigma_\varepsilon^2(t-1) + \omega^2(t) \\ &= B_2^2(t)\sigma_\varepsilon^2(t-2) + B_1^2(t)\omega^2(t-1) + \omega^2(t) \\ &\vdots \\ &= B_{T_2}^2(t)\sigma_\varepsilon^2(t-T_2) + B_{T_2-1}^2(t)\omega^2(t-T_2+1) + \dots + B_1^2(t)\omega^2(t-1) + \omega^2(t). \end{aligned}$$

Therefore,

$$\sigma_\varepsilon^2(t) = \frac{1}{1 - B_{T_2}^2(t)} \sum_{j=0}^{T_2-1} B_j^2(t)\omega^2(t-j),$$

because $\sigma_\varepsilon^2(t - T_2) = \sigma_\varepsilon^2(t)$. ■