IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received April 14, 2019, accepted May 5, 2019, date of publication May 10, 2019, date of current version May 31, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2916168

A New Solar Radiation Model for a Power System Reliability Study

HAMZA ABUNIMA, JIASHEN TEH, (Member, IEEE), AND HUSSEIN JUMMA JABIR
School of Electrical and Electronic Engineering, Engineering Campus, Universiti Sains Malaysia, Nibong Tebal 14300, Malaysia

Corresponding author: Jiashen Teh (jiashenteh@usm.my)

This work was supported by the Universiti Sains Malaysia (USM) Research University Incentive (RUI) under Grant 1001/PELECT/8014099.

ABSTRACT The dwindling number of conventional power resources and its environmental impact has motivated a transition to renewable energy sources, such as solar power. Evaluating the reliability of solar power integration into power networks can help decision makers gauge the feasibility of their solar power projects. However, the stochastic and non-stationary nature of solar radiation is difficult to be modeled and can even hinder an accurate evaluation of reliability. A good solar model for accurately assessing solar-power-integrated systems should be able to retain the original statistical properties of the sampled solar radiation data. Therefore, this paper aims to develop a new robust and easy-to-use methodology for simulating solar radiation. The proposed model was compared with four other models, including the clearness index, auto-regressive moving average, and two probability-distribution-based models. Five statistical tests, namely, F-test, diurnal distributions, partial auto-correlation function (PACF), mean, and standard deviation, were performed for the comparison. The comparison results indicate that the proposed method effectively retains the statistical properties of the original data and outperforms all other models in the tests. Therefore, the proposed model can be used for assessing solar-power-integrated power systems.

INDEX TERMS Solar radiation model, Monte Carlo, power system reliability analysis, renewable energy.

I. INTRODUCTION

In the past few decades, the demand for renewable energy sources has grown steadily partly due to the dwindling fossil fuel supply and the growing energy demand; however, the most important driver of this growth lies in the people's increasing awareness of environmental conservation initiatives [1], [2]. Amongst all possible renewable energy sources, solar power [3] is considered the most advanced and most suitable resource for large-scale electricity generation despite its intermittent nature, which, at the moment, is addressed by using energy storages [4].

Given that solar power generation greatly depends on the availability of sunlight, meteorological studies and modelling of solar radiation are crucial in determining the feasibility of any solar farm project. In other words, before undertaking any solar farm construction projects, the local solar radiation data must be studied to justify the long-term reliability benefits of solar power generation. Unfortunately, the decision to initiate such projects should not be based solely on historical solar radiation data without any statistical processing for two reasons [5]. Firstly, solar radiation records are lacking in certain locations, thereby necessitating the use of synthetically generated values to allow subsequent analyses. Secondly, analysing only the original historical data inherently assumes that the future solar radiation profile is exactly the same as the previous profile, which is unrealistic. A simple solution that uses the average value is also not suitable because the full range of solar radiations is not represented in this value, thereby underestimating and overestimating solar power generation. This error is further aggravated by the inherent uncertainty of solar radiation values as they propagate through time and space. The computation of average solar radiation values can also be easily skewed by the variation in data size [6].

Therefore, a highly sophisticated solar radiation model that retains all important statistical properties whilst allowing a certain degree of uncertainty to represent its natural stochastic behaviour must be developed [5]. As shown in previous studies, these properties include variance, mean, auto-correlation and diurnal distributions [2], [7]. Synthetic data demonstrate the similar behaviours as the original historical data but with an unlimited size, thereby solving the problems related to limited data. Accordingly, synthetic data are very useful in the reliability evaluation of solar power generation, which

The associate editor coordinating the review of this manuscript and approving it for publication was Cristian Zambelli.

necessitates the application of multiple Monte Carlo simulations. In addition, solar radiation is only available during daytime and is always limited to a certain maximum value depending on the location [6]. Solar radiation is also not stationary due to the propagation of irradiation through time [8], [9]. Thus, on top of the previously mentioned properties, a suitable solar radiation model must be time based, non-stationary and able to adhere to local geographical properties.

Meanwhile, time-based models [10]–[13] work well only with stationary data. Therefore, non-stationary data, such as solar radiation, need to be converted into stationary forms through the ''stationarising'' process prior to building a time-series model. Four approached are commonly used for solar radiation modelling; the clearness index [13], Auto-Regressive Moving Average (ARMA) model [14] and two probability distribution approaches [15], [16].

Introduced by Graham, the clearness index is based on the ratio of overall global and extra-terrestrial irradiance on a horizontal plane [13]. When this index is used to measure the fluctuations in solar radiation as a result of the changes in extra-terrestrial solar radiations, the global solar radiation is transformed into stationary data prior to any time-series model fitting [10]. The random terms of the clearness index were sampled from the Gaussian distribution [17], [18]. This index was further popularised after its adoption in various software, including Homer [19], WATGEN [20] and DSSAT [21], all of which are useful for analysing power systems that are integrated with solar and wind power. However, as its major drawback, the clearness index enforces a Gaussian distribution for all solar radiation data without considering the suitability of this distribution [6]. Moreover, the built-in parameters, including solar constant and sunset hour angle, are merely approximations at best, thereby adding unwanted uncertainties to the solar radiation modelling.

The ARMA model comprises two components, namely, Auto-Regressive (AR) and Moving Average (MA), which represent the regression of future and past values and the observation error, respectively [22]. Given that this model only accepts stationary data, stationarising the non-stationary data is required in ARMA model fitting. The ARMA model performs the stationarising process by removing the data trends via a series of detrending and filtering processes [14]. However, this process must be performed with care to avoid enhancing the trends [23]. Therefore, choosing a suitable detrending process is not straightforward and can be computationally expensive for big data. Detrending also inevitably over-reduces the fluctuations in data, thereby leading to information loss [24].

The probability-distribution-based approach, such as the Hourly Mean Solar Radiation (HMSR) method [16], is by far the simplest and easiest-to-use method for modelling solar radiation. This method uses the hourly mean and standard deviation, monthly maximum and minimum irradiance and monthly mean bright sunshine hours to model the historical solar radiation. A more advanced technique based on Pearson distribution, in which the skewness and kurtosis of the probability distribution are considered, has also been proposed [15]. Despite offering a simpler alternative compared with the two aforementioned methods, this technique does not take into account the autocorrelation characteristics of hourly solar radiation values, which are useful for indicating the degree of regression between the future and past values.

Despite the introduction of these approaches, the previously built solar radiation models are generally lacking in accuracy due to the abovementioned assumptions. Meanwhile, although improving the time resolution of the solar radiation data will increase the precision of solar power generation estimation, doing so will also intensify the fluctuations and non-stationary nature of these data, thereby increasing the difficulty of fitting the original data into the previously mentioned models. This paper aims to address this problem by proposing a new and easy-to-use solar radiation modelling approach that does not have any of the drawbacks mentioned above. The proposed model also fits nicely with the Monte Carlo simulation and, as a result, is suitable for analysing the reliability of solar-power-integrated power systems. To verify its robustness, the proposed model is compared with the three aforementioned approaches, all of which have been introduced only recently yet have already attracted wide usage amongst researchers [13]–[16]. Five statistical tests, namely, F-Test, diurnal distribution, Partial Auto-Correlation Function (PACF), mean and standard deviation, are performed for the verification. The results show that the proposed method can generate solar radiation data that resemble the historical data much better compared with the other aforementioned methods.

The rest of this paper is organised as follows. Section II describes the research data and methodology. Section III investigates the results of the proposed model and verifies the adequacy of the model. Section IV concludes the paper.

II. METHODOLOGY

The algorithm of the proposed model is presented in this section. It was developed in Matlab and simulated using core i5 CPU @ 3 GHz with 4 GB RAM. Table 1 lists the related symbols.

A. SITE LOCATION AND DATA COLLECTION

The global horizontal irradiance data are sourced from the archives of State University of New York (SUNY) project developed by SUNY and the Solar Consulting Services [25]. This project recorded hourly solar irradiation data near Ache, Indonesia from 2000 to 2014 by using Meteosat 5 and 7 satellites. These data are stored in the matrix $\boldsymbol{Y}$ and have a size of 1×131400. In $\boldsymbol{Y}$, the data for the same month of the entire 15 years are further grouped into set $\boldsymbol{M}_\alpha$ such that

$$\boldsymbol{Y} = [\boldsymbol{M}_1, \boldsymbol{M}_2, \ldots, \boldsymbol{M}_\alpha | \, \alpha = 12] \quad \text{and} \tag{1a}$$

$$\boldsymbol{M}_\alpha = \left[\boldsymbol{D}_1, \boldsymbol{D}_2, \ldots, \boldsymbol{D}_y \middle| \, y = 15\right] \tag{1b}$$

where α is the month and $\boldsymbol{M}_\alpha$ is a larger matrix that contains all smaller $\boldsymbol{D}_x$ matrices. The row and column of $\boldsymbol{D}_x$ represent

TABLE 1. Notation system in the proposed algorithm.

Symbol	Description
$\boldsymbol{Y}$	The matrix where the whole original irradiance data is stored
$\boldsymbol{M}_\alpha$	The matrix where the original irradiance data of the month α is stored, where $1 \le \alpha \le 12$
y	The total years number of the available data
$\boldsymbol{D}_x$	The sub-matrix in $\boldsymbol{M}_\alpha$, where the original irradiance data in a certain year x is stored, where $1 \le x \le y$
$m_{h,d}$	The irradiance value at hour h on day d in $\boldsymbol{M}_\alpha$
d'	The corresponding day's number in the corresponding month in the original data
d''	The total number of days in the corresponding month in the original data
y'	The corresponding year's number in the original data
n_α	Number of the significant lags of the data in the month α
k	The total number of clusters
k'_j	The corresponding cluster number where $1 \le j \le k$
c_j	Centre of the cluster k'_j
$\boldsymbol{M}'_\alpha$	The matrix contains the clusters numbers which express the values in $\boldsymbol{M}_\alpha$
$m'_{h,d}$	The cluster number where the value $m_{h,d}$ belongs to
$\boldsymbol{U}^l_h$	A unique possible combination l of the clusters numbers from row n to row h in $\boldsymbol{M}'_\alpha$
P^l_h	The occurrence probability of the combination $\boldsymbol{U}^l_h$
$\boldsymbol{G}^\alpha_{24\times D}$	the new synthetic clusters matrix of the month α for D days
$\boldsymbol{M}''_\alpha$	the new synthetic solar radiation matrix of the month α

the hour and day of the data, respectively, and x donates the year number, where $1 \le x \le y$. y is the total number of years in the available data. For example, given that all months of January from 2000 to 2014, $\boldsymbol{M}_{Jan}$, have a total of 465 days, its size can be determined as (24 × 465). Conveniently, the elements of $\boldsymbol{M}_\alpha$ are denoted by $m_{h,d}$, where h and d represent the corresponding hour and day, respectively. A specific column in $\boldsymbol{M}_\alpha$ can be found as (2)

$$d = d' + d''\left(y' - 1\right)\Big|\, 1 \le y' \le y \tag{2}$$

where $d' \in \left(1, 2, \ldots, d''\right)$ is the day of the month for a particular y^{th} year and d'' is the total number of days in a month. When expanded from (1b), the matrix $\boldsymbol{M}_\alpha$ takes the following form:

$$\boldsymbol{M}_\alpha = \begin{bmatrix} m_{1,1} & m_{1,2} & \cdots & m_{1,d''\times y} \\ m_{2,1} & m_{2,2} & \cdots & m_{2,d''\times y} \\ \vdots & \vdots & \ddots & \vdots \\ m_{24,1} & m_{24,2} & \cdots & m_{24,d''\times y} \end{bmatrix} \tag{3}$$

A simplified example of $\boldsymbol{M}_\alpha$ is given in (4) on which all examples presented in this section will be based.

$$\boldsymbol{M}_1 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 20 & 11 & 23 & 9 & 10 & 19 \\ 54 & 52 & 69 & 53 & 71 & 70 \end{bmatrix} \tag{4}$$

where the zero values represent the hours when no solar radiation is recorded.

B. PROPOSED SOLAR MODELLING ALGORITHM

The procedures of the proposed solar modelling algorithm are explained in detail as follows.

Step 1: The PACF of the sampled solar irradiation data is determined for all $\boldsymbol{M}_\alpha$ to identify the number of statistically significant lags n_α. The lags of the example matrix in (4) is assumed to be $n_1 = 1$.

Step 2: The k-means clustering algorithm [26] is applied on each row of $(\boldsymbol{M}_\alpha | 1 \le \alpha \le 12)$. Recall from (3) that an arbitrary row of $\boldsymbol{M}_\alpha$ is $\left[m_{h,1}, m_{h,2}, \ldots, m_{h,d''\times y}\right]$. Then, k-means clustering is achieved by minimising the following objective function:

$$Q\left(k'_j\Big|_{j=1}^{k}\right) = \sum_{j=1}^{k} \sum_{m\in k'_j} \left\| m - c_j \right\|^2 \tag{5}$$

such that all radiation data that belong to the same hour (i.e. row) are grouped into k clusters, each cluster is denoted by $\left(k'_j \Big| 1 \le j \le k\right)$ and the centre of each cluster is denoted by c_j. A special case is observed when the entire row has an equal value. In this situation, all values are grouped into a single cluster. This situation is encountered most often at hours with zero solar irradiation.

After the clustering, the new matrix $\boldsymbol{M}'_\alpha$ presented in (6) is derived from $\boldsymbol{M}_\alpha$. In this matrix, the solar irradiation values are replaced with cluster numbers. The mean and variance of each cluster in each row are also determined and kept for later use in **Step 5**.

$$\boldsymbol{M}'_\alpha = \begin{bmatrix} m'_{1,1} & m'_{1,2} & \cdots & m'_{1,d''\times y} \\ m'_{2,1} & m'_{2,2} & \cdots & m'_{2,d''\times y} \\ \vdots & \vdots & \ddots & \vdots \\ m'_{24,1} & m'_{24,2} & \cdots & m'_{24,d''\times y} \end{bmatrix} \tag{6}$$

If all elements in both non-zero rows of the example presented in (4) can be classified into two clusters, then the resulting $\boldsymbol{M}'$ is

$$\boldsymbol{M}'_1 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 2 & 1 & 2 & 1 & 1 & 2 \\ 1 & 1 & 2 & 1 & 2 & 2 \end{bmatrix} \tag{7}$$

Step 3: At each row (denoted h) of matrix $\boldsymbol{M}'_\alpha$, all unique possible combinations of n previous values up until the values in the current row are determined and stored in the matrix $\boldsymbol{U}^l_h$ as

$$\boldsymbol{U}^l_h = \begin{bmatrix} m'_{h-n} & m'_{h-n+1} & \cdots & m'_h \end{bmatrix}^{\boldsymbol{T}} \tag{8}$$

where l denotes the unique combination number, which highest value can be computed as

$$l = k^{n+1} \tag{9}$$

The purpose of this step is to ensure that the number of times when $\boldsymbol{U}_h^l$ is encountered can be calculated. Afterwards, its probability of occurrence, P_h^l, is determined by dividing the number of encounters by the number of columns in $\boldsymbol{M}_\alpha^{'}$. Given the influence of lagging terms, one must note that $h \geq n+1$.

Following the example in (7) and given that $n_1 = 1$, the range $2 \leq h \leq 4$ applies. Therefore, at $h = 2$,

$$\boldsymbol{U}_2^1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \text{and} \tag{10a}$$

$$P_2^1 = \frac{6}{6} = 1 \tag{10b}$$

at $h = 3$,

$$\boldsymbol{U}_{\boldsymbol{3}}^1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \boldsymbol{U}_{\boldsymbol{3}}^2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \quad \text{and} \tag{11a}$$

$$P_{\boldsymbol{3}}^1 = \frac{3}{6} = \frac{1}{2}, \quad P_{\boldsymbol{3}}^2 = \frac{3}{6} = \frac{1}{2} \tag{11b}$$

at $h = 4$,

$$\boldsymbol{U}_4^1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \boldsymbol{U}_4^2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \boldsymbol{U}_4^3 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \boldsymbol{U}_4^4 = \begin{bmatrix} 2 \\ 2 \end{bmatrix} \quad \text{and} \tag{12a}$$

$$P_4^1 = \frac{2}{6}, P_4^2 = \frac{1}{6}, P_4^3 = \frac{1}{6}, P_4^4 = \frac{2}{6}. \tag{12b}$$

Step 4: All the obtained $\boldsymbol{U}_h^l$ and P_h^l from the previous step are used to guide the random generation of the new synthetic cluster number matrix $\boldsymbol{G}_{24\times D}^\alpha$:

$$\boldsymbol{G}_{24\times D}^\alpha = \begin{bmatrix} g_{1,1} & g_{1,2} & \cdots & g_{1,D} \\ g_{2,1} & g_{2,2} & \cdots & g_{2,D} \\ \vdots & \vdots & \ddots & \vdots \\ g_{24,1} & g_{24,2} & \cdots & g_{24,D} \end{bmatrix} \tag{13}$$

where D is the number of generated solar radiation days. The initialisation process of $\boldsymbol{G}^\alpha$ requires all values from $h-n$ to $h-1$ to be assigned the value of 1. On the one hand, the few initial solar radiation values are always equal to 0 due to night time and can only be grouped into a single cluster. Secondly, these values are needed in constricting $\boldsymbol{G}^\alpha$ given that $h \geq n+1$. Under these settings, the rows that require initialisation are guided by the sampled historical solar radiation data without needing inputs from the user.

Following the example in (7), given that $h \geq 2$, only the first row of $\boldsymbol{G}^1$ needs to be initialised as follows:

$$\boldsymbol{G}^1 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \tag{14a}$$

Afterwards, the second row is filled based on (10a) and (10b), thereby indicating that this row has 100% probability of obtaining a value of 1:

$$\boldsymbol{G}^1 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \tag{14b}$$

The third row is sampled based on (11a) and (11b), thereby suggesting that there are equally 50% that the previous row value of all ones will be followed by either 1 or 2:

$$\boldsymbol{G}^1 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 & 2 & 1 & 2 & 1 & 2 & 2 & 2 & 1 \end{bmatrix} \tag{14c}$$

The final row is sampled based on (12a) and (12b), thereby suggesting that the value 1 has about 33% and 17% probability to be followed by 1 and 2, respectively. These percentage probabilities are reversed if the third row takes the value of 2 instead of 1. Therefore, the remaining matrix $\boldsymbol{G}^1$ can be built as

$$\boldsymbol{G}^1 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 & 2 & 1 & 2 & 1 & 2 & 2 & 2 & 1 \\ 1 & 2 & 2 & 1 & 2 & 1 & 1 & 2 & 2 & 1 & 2 & 1 \end{bmatrix} \tag{14d}$$

Step 5: When the matrix $\boldsymbol{G}^\alpha$ is completed, the cluster numbers are randomly converted into solar radiation values based on the normal distributions of the respective clusters in each row. The mean and variance of these distributions were obtained earlier in **Step 2**. By using these values, the new synthetic solar radiation matrix $\boldsymbol{M}_\alpha^{''}$ can be obtained.

The example developed so far in (14d) becomes

$$M_1^{''} = \begin{bmatrix} 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 \\ 10.2 & 18.5 & 10.0 & 10.7 & \ldots & 9.1 \\ 53.0 & 68.8 & 70.7 & 52.6 & \cdots & 53.5 \end{bmatrix} \tag{15}$$

where the values in the first two rows are equal to zero given that the mean and variance of their historical values are also equal to zero.

The proposed algorithm is illustrated in Fig 1.

C. MODEL VALIDATION

To verify its ability to generate solar radiation data that statistically resemble the original data, the proposed model was compared with the four models discussed in Section I, namely, the clearness index [13], ARMA modelling [14], the HMSR model [16] and the Pearson distribution model [15]. Several statistical tests, including F-test, diurnal distribution, PACF [2], mean and standard deviation [27], were performed for the evaluation. All solar radiation values recorded at night were excluded from the mean and standard deviation tests because these values are always equal to zero, thereby masking the test results.

The F-test (F_0) was performed to test the null hypothesis H_0, that is, the original and simulated datasets do not show any differences in their normal distribution and variance [2]. F_0 is calculated as:

$$F_0 = \frac{\sigma_1^2}{\sigma_2^2} \tag{16}$$

where σ_1 and σ_2 are the standard deviations of the original and simulated data, respectively. The *P-value* of F_0 is calculated according to [28] to determine whether the null hypothesis is accepted or rejected. H_0 is accepted if *P-value*> 0.05 and rejected if *P-value* ≤ 0.05.

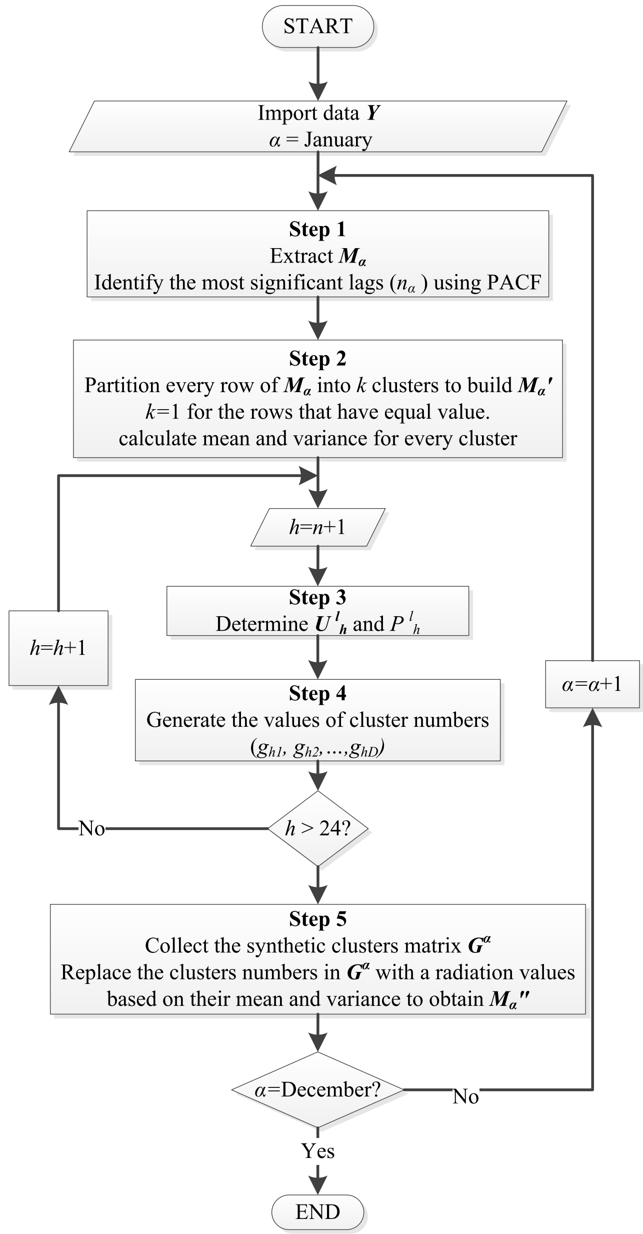


FIGURE 1. Flowchart of the proposed model.

The diurnal distribution test was performed to test the ability of the models to accurately simulate the diurnal patterns of the original data [2].

The PACF test was performed to compare the lag terms between the simulated and original data. Firstly, the PACF of the data was calculated as follows by using the Yule–Walker equation [29]:

$$\begin{bmatrix} 1 & r_1 & \cdots & r_{p-2} & r_{p-1} \\ r_1 & 1 & \cdots & r_{p-3} & r_{p-2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ r_{p-1} & r_{p-2} & \cdots & r_1 & 1 \end{bmatrix} \begin{bmatrix} \emptyset_{p1} \\ \emptyset_{p2} \\ \vdots \\ \emptyset_{pp} \end{bmatrix} = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_p \end{bmatrix} \tag{17}$$

where $\emptyset_{pp}$ is the partial autocorrelation coefficient, r_p is the auto-covariance function and p is the lag term. Afterwards, the differences in the lag terms of the original and simulated data are determined by calculating Root Mean Square Error (RMSE) of their PACF as

$$RMSE = \sqrt[2]{\frac{\sum_{i=1}^{p} (\emptyset_i^O - \emptyset_i^g)^2}{p}} \tag{18}$$

where $\emptyset_i^O$ and $\emptyset_i^g$ are the partial autocorrelation coefficients of the original and simulated data at lag i, respectively.

The mean solar radiation value over a particular interval γ was calculated as follows after excluding all night values for the same reason mentioned previously:

$$Mean^{\gamma} = \frac{\sum_{j=1}^{R} irradiance\ at\ hour\ j}{R} \tag{19}$$

where R is the total number of hours in interval γ, excluding those hours when the irradiance values are equal to zero. Each interval is four months long. This interval size was selected in order to report the results in this paper due to space limitations. We have found that the variations in interval size will not affect the outcomes of the comparison between the mean values of the original and simulated data.

The standard deviation of the monthly solar radiation values was calculated to measure the amount of dispersion in the dataset. A good solar model should be able to generate solar values whose standard deviations match those of the original data. The similarity between these standard deviations was measured by comparing their RMSE values.

III. RESULTS AND DISCUSSIONS

The proposed model shown in Fig. 1 was applied to the original data. Statistical tests were also performed to examine how the simulated data can retain the statistical characteristics of the original data. The data simulated by the compared models in [13]–[16] reveal that 5000 sets of simulated data are large enough to perform a proper comparison and this is also repeated in this paper.

A. F-TEST

The *P-values* of the proposed and compared models were calculated for each month and the results are shown in Table 2, where the letters 'R' and 'A' indicate that the test rejects or accepts H_0, respectively. Only the data simulated by the proposed model are fully accepted over the 12 months. All *P-values* are close to 1, thereby indicating that the simulated data match the original data. In the other models, acceptance rate of the test results is either much lower or non-existence. In the few cases where the test is accepted as shown by the clearness index method, the *P-values* are all significantly lower than the values given by our proposed model. This further strengthens that although our proposed model and the clearness index at sometimes both accept the test, our

TABLE 2. F-test results for the proposed and compared models.

	Proposed model		Clearness index		ARMA		HMSR		Pearson Dis.	
	H_0	p	H_0	p	H_0	p	H_0	p	H_0	p
Jan	A	0.99	R	≈0	R	≈0	R	≈0	R	≈0
Feb	A	0.99	R	≈0	R	≈0	R	≈0	R	≈0
Mar	A	0.77	R	≈0	R	≈0	R	≈0	R	≈0
Apr	A	0.98	A	0.61	R	≈0	R	≈0	R	≈0
May	A	0.97	R	≈0	R	≈0	R	≈0	R	≈0
Jun	A	0.99	R	≈0	R	≈0	R	≈0	R	≈0
Jul	A	0.96	R	0.004	R	≈0	R	≈0	R	≈0
Aug	A	0.94	R	≈0	R	≈0	R	≈0	R	≈0
Sep	A	0.94	R	≈0	R	≈0	R	≈0	R	≈0
Oct	A	0.98	A	0.684	R	≈0	R	≈0	R	≈0
Nov	A	0.98	R	≈0	R	≈0	R	≈0	R	≈0
Dec	A	0.94	R	≈0	R	≈0	R	≈0	R	≈0

TABLE 3. RMSE of the datasets simulated over the year by the models in terms of diurnal distribution, PACF, and standard deviation.

	Proposed model	Clearness index	ARMA	HMSR	Pearson Dis.
RMSE of the diurnal distributions (w/m^2)	0.1	74	7.0	10.3	4.5
RMSE of PACF	0.02	0.06	1.047	1.03	1.04
RMSE of standard deviation (w/m^2)	12.0	22.1	47.7	24.9	27.4

proposed model is able to simulate data that matches more with the original data than the clearness index is able to.

B. DIURNAL DISTRIBUTION TEST

The hourly average of the data simulated by the proposed and compared models was calculated for each month to examine the diurnal distributions. The hourly average irradiance of the data simulated for January was compared with that of the original data as shown in Fig. 2. All models except for the clearness index model showed similar diurnal distributions with the original data. In order to highlight the differences between these models, RMSE of the diurnal distribution values between the simulated and original data over the 12 months was calculated and shown in Fig. 3 and Table 3. The proposed model attains the lowest error value of 0.1 w/m^2 followed by the Pearson distribution model, which attains an error value of 4.5 w/m^2. However, this value is still 45 times higher than the error value of the proposed model. As expected, the clearness index model registers the highest error value of 74 w/m^2. The results from Fig. 3 indicate that the proposed model can generate data that are consistently similar to the original data in terms of diurnal distribution. The same conclusion is also drawn from the remaining months, which are not reported in this paper due to space limitations.

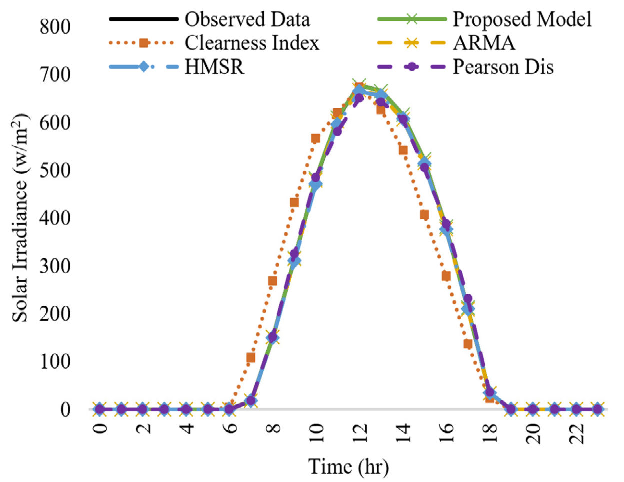


FIGURE 2. Original and simulated diurnal distributions of hourly average irradiance in January.

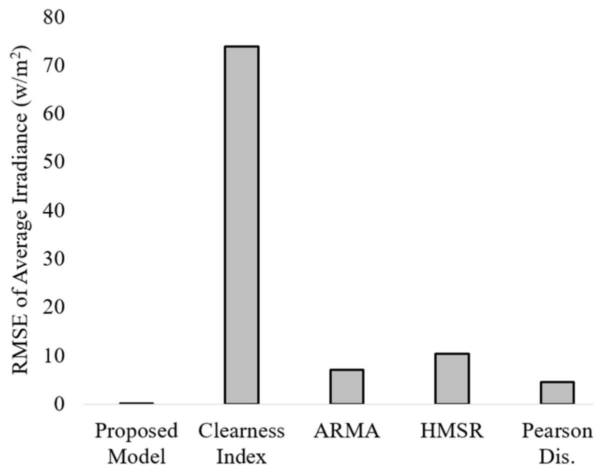


FIGURE 3. RMSE of the diurnal distribution of the dataset simulated by all models over 12 months.

C. PACF TEST

The PACF of the original data was compared with the data simulated by the five models, including the proposed model Table 3 and Fig. 4 show the RMSE values of the PACF coefficients over 12 months. The proposed model attained the lowest RMSE value of 0.02, followed by the clearness index model. However, the RMSE of the clearness index model is still 200% larger than that of the proposed model. The RMSE of ARMA, HMSR and Pearson distribution models are similar and are all larger than the proposed model by about 5000%. Given its low RMSE value, the proposed model can retain the partial autocorrelation characteristic of the original data the most compared with the other tested models.

D. MEAN TEST

The mean μ irradiance value over January–April, May–August and September–November were calculated by using (19). The mean values of the datasets simulated by all models were compared with those of the original data

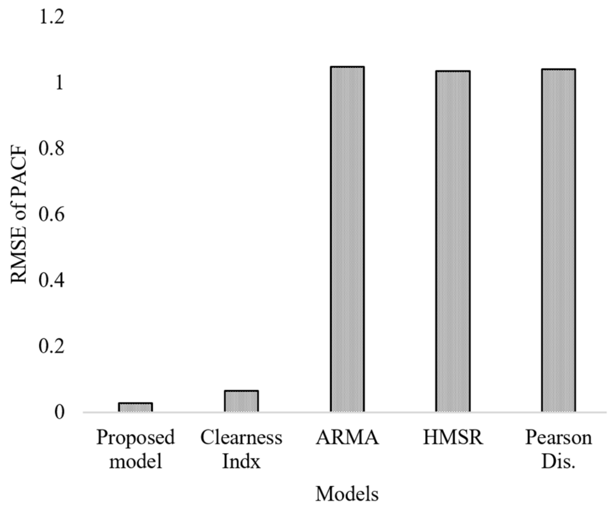


FIGURE 4. RMSE of PACF of the dataset simulated by all models over 12 months.

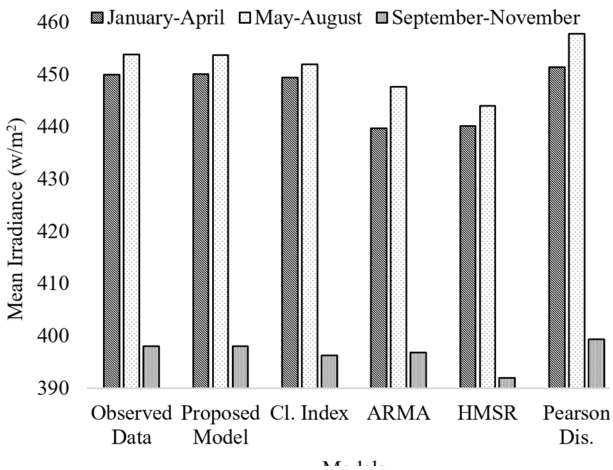


FIGURE 5. Daily mean values of the datasets over three intervals.

and the results are shown in Fig. 5. The mean irradiances of the original data are 449.9, 453.8 and 398 w/m^2 for the three aforementioned intervals, respectively, whilst those of the simulated data are 450, 453.6 and 397.9 w/m^2. The average percentage difference between the original data and the data simulated by the proposed model is only about 0.02%, thereby indicating a high similarity between the two datasets. The data that showed the second highest degree of similarity was simulated by the clearness index model with an average percentage difference of 0.3%. However, such degree of similarity is still 15 times greater than that of the proposed model. The other models showed a worse performance and attained mean irradiances that are far from those of the original data.

E. STANDARD DEVIATION

The monthly standard deviation σ was calculated for the datasets simulated by the five models and they are shown

TABLE 4. Monthly standard deviation of generated solar values.

	Historical data (w/m^2)	Proposed model (w/m^2)	Clearness index (w/m^2)	ARMA (w/m^2)	HMSR (w/m^2)	Pearson Dis. (w/m^2)
Jan	280	280.4	320.1	244.5	267.0	251.4
Feb	315	305.2	321.4	268.9	290.9	286.1
Mar	330	315.3	326.3	279.4	298.9	299.3
Apr	323	311.5	298.3	268.19	293.0	299.0
May	313	296.6	299.0	261.0	281.7	276.4
Jun	306	286.6	289.4	252.5	272.0	272.4
Jul	302	289.0	285.9	257.8	273.7	272.9
Aug	311	297.4	308.8	256.8	282.6	281.1
Sep	310	303.7	328.2	260.7	286.9	284.9
Oct	297	298.4	289.9	253.0	282.6	281.0
Nov	277	283.5	313.1	242.2	269.7	259.1
Dec	253	264.5	287.3	211.0	252.9	241.8

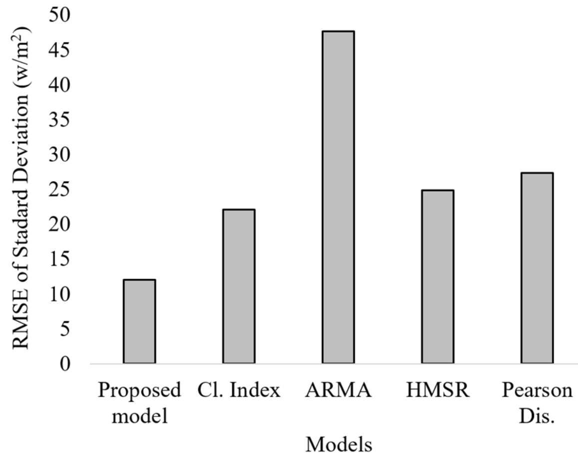


FIGURE 6. RMSE of the monthly standard deviation of the five models over 12 months.

in Table 4. It can be seen that the standard deviation values resulted from the proposed model and the clearness index model are the closest to the original data. The highest contrast between the proposed model and the original data occurred in Jun, by about 19.61 w/m^2, while for the clearness index model, it occurred in November by about 29.5 w/m^2. The smallest contrast given by the proposed model data occurred in January is about 0.344 w/m^2, and 2.41 w/m^2 in May for the data simulated by the clearness index model. In order to clearly highlight the performance of each model, RMSE of the standard deviation resulting from the five models was calculated over 12 months as shown in Fig. 6 and Table 6. The proposed model simulated standard deviation with the lowest RMSE of about 12 w/m^2, thereby suggesting that the proposed model manages to retain the dispersion pattern of the original data. The second lowest RMSE was attained by the clearness index model, but this value was double that attained by the proposed model.

IV. CONCLUSION

This paper presents a new methodology for modelling solar radiation that successfully retains the statistical properties of the original data. This method generates data at each hour based on the probability of occurrence of solar radiation at n previous hours. The value of n is computed according to the PACF of the original data. The proposed method was then applied on the solar radiation data imported from the SUNY project conducted in Ache, Indonesia from 2000 to 2014.

To verify its goodness, the proposed method was compared with four other methods, including the clearness index, ARMA, HMSR method, and Pearson distribution. Five tests, including the F-test, diurnal distribution, PACF, mean and standard deviation, were then performed for the evaluation.

The F-test results show that the proposed method can generate the same normal distribution and variance as the original data. Although all the tested methods obtained the same diurnal distribution curves, the proposed method showed the lowest RMSE value when compared with the original data, thereby suggesting that this method can accurately simulate the diurnal distribution of the original solar radiation data.

The PACF of the original data was accurately represented by the proposed methodology. The PACF test revealed that the proposed method outperforms the other four methods in this aspect. Therefore, the proposed method can generate synthetic solar radiation that is suitable for chronological simulation, such as the sequential Monte Carlo simulation. Standard deviation and mean are other statistical characteristics that are highly retained by the proposed method.

The superiority of the proposed method in all tests irrefutably show its efficiency in modelling solar radiation. Future works can verify the ability of this method to simulate other metrological parameters required in reliability studies, such as wind and temperature.

REFERENCES

[1] H. Abunima, J. Teh, C.-M. Lai, and H. J. Jabir, ''A systematic review of reliability studies on composite power systems: A coherent taxonomy motivations, open challenges, recommendations, and new research directions,'' *Energies*, vol. 11, no. 9, p. 2417, 2018.

[2] R. Billinton, H. Chen, and R. Ghajar, ''Time-series models for reliability evaluation of power systems including wind energy,'' *Microelectron. Rel.*, vol. 36, no. 9, pp. 1253–1261, 1996.

[3] B. Zhou *et al.*, ''Multiobjective generation portfolio of hybrid energy generating station for mobile emergency power supplies,'' *IEEE Trans. Smart Grid*, vol. 9, no. 6, pp. 5786–5797, Nov. 2018.

[4] M. A. Hassan, A. Khalil, S. Kaseb, and M. A. Kassem, ''Exploring the potential of tree-based ensemble methods in solar radiation modeling,'' *Appl. Energy*, vol. 203, pp. 897–916, Oct. 2017.

[5] M. J. Brook and B. A. Finney, ''Generation of bivariate solar radiation and temperature time series,'' *Solar Energy*, vol. 39, no. 6, pp. 533–540, 1987.

[6] L. Linguet, Y. Pousset, and C. Olivier, ''Identifying statistical properties of solar radiation models by using information criteria,'' *Solar Energy*, vol. 132, pp. 236–246, Jul. 2016.

[7] R. Posadillo and R. L. Luque, ''Approaches for developing a sizing method for stand-alone PV systems with variable demand,'' *Renew. Energy*, vol. 33, no. 5, pp. 1037–1048, 2008.

[8] J. Huang, M. Korolkiewicz, M. Agrawal, and J. Boland, ''Forecasting solar radiation on an hourly time scale using a coupled autoregressive and dynamical system (CARDS) model,'' *Solar Energy*, vol. 87, pp. 136–149, Jan. 2013.

[9] J. Tovar-Pescador, ''Modelling the statistical properties of solar radiation and proposal of a technique based on boltzmann statistics,'' in *Modeling Solar Radiation at the Earth's Surface*. Berlin, Germany: Springer, 2008, ch. 3, p. 5591.

[10] K. Gairaa, A. Khellaf, Y. Messlem, and F. Chellali, ''Estimation of the daily global solar radiation based on Box–Jenkins and ANN models: A combined approach,'' *Renew. Sustain. Energy Rev.*, vol. 57, pp. 238–249, May 2016.

[11] A. Azadeh, S. F. Ghaderi, S. Tarverdian, and M. Saberi, ''Integration of artificial neural networks and genetic algorithm to predict electrical energy consumption,'' *Appl. Math. Comput.*, vol. 186, no. 2, pp. 1731–1741, 2007.

[12] A. S. Dobakhshari and M. Fotuhi-Firuzabad, ''A reliability model of large wind farms for power system adequacy studies,'' *IEEE Trans. Energy Convers.*, vol. 24, no. 3, pp. 792–801, Sep. 2009.

[13] V. A. Graham, K. G. T. Hollands, and T. E. Unny, ''A time series model for K_t with application to global synthetic weather generation,'' *Solar Energy*, vol. 40, no. 2, pp. 83–92, 1988.

[14] L. F. N. Lourenço, M. B. de Camargo Salles, M. M. F. Gemignani, M. R. Gouvea, and N. Kagan, ''Time Series modelling for solar irradiance estimation in northeast Brazil,'' in *Proc. IEEE 6th Int. Conf. Renew. Energy Res. Appl. (ICRERA)*, Nov. 2017, pp. 401–405.

[15] O. O. Amusat, P. R. Shearing, and E. S. Fraga, ''On the design of complex energy systems: Accounting for renewables variability in systems sizing,'' *Comput. Chem. Eng.*, vol. 103, pp. 103–115, Aug. 2017.

[16] R. M. Moharil and P. S. Kulkarni, ''Reliability analysis of solar photovoltaic system using hourly mean solar radiation data,'' *Solar Energy*, vol. 84, no. 4, pp. 691–702, 2010.

[17] R. Aguiar and M. Collares-Pereira, ''TAG: A time-dependent, autoregressive, Gaussian model for generating synthetic hourly radiation,'' *Solar Energy*, vol. 49, no. 3, pp. 167–174, 1992.

[18] J. Polo, L. F. Zarzalejo, R. Marchante, and A. A. Navarro, ''A simple approach to the synthetic generation of solar irradiance time series with high temporal resolution,'' *Solar Energy*, vol. 85, no. 5, pp. 1164–1170, 2011.

[19] (2019). *Homer Grid 1.3 User Manual*. [Online]. Available: https://www.homerenergy.com/products/grid/docs/1.3/index.html

[20] R. Billinton and Bagen, ''Generating capacity adequacy evaluation of small stand-alone power systems containing solar energy,'' *Rel. Eng. Syst. Saf.*, vol. 91, no. 4, pp. 438–443, 2006. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0951832005000797

[21] K. Střelcová, *Bioclimatology and Natural Hazards*. Amsterdam, The Netherlands: Springer, 2008.

[22] J. Teh and I. Cotton, ''Reliability impact of dynamic thermal rating system in wind power integrated network,'' *IEEE Trans. Rel.*, vol. 65, no. 2, pp. 1081–1089, Jun. 2016.

[23] P. M. T. Broersen, *Automatic Autocorrelation and Spectral Analysis*. London, U.K.: Springer, 2006.

[24] Z. Cao, Y. Wang, and D. Li, *Probabilistic Approaches for Geotechnical Site Characterization and Slope Stability Analysis*. Berlin, Germany: Springer, 2017.

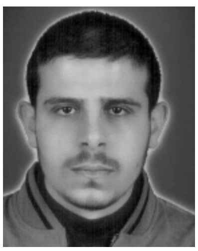
[25] *SUNY India Direct Normal Irradiance*, NSRDB, Colorado, CO, USA, 2015.

[26] S. Lloyd, ''Least squares quantization in PCM,'' *IEEE Trans. Inf. Theory*, vol. 28, no. 2, pp. 129–137, Mar. 1982.

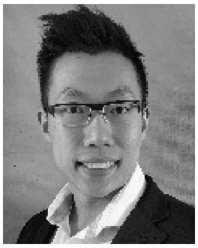
[27] R. I. Putri, A. Priyadi, and M. H. Purnomo, ''Stochastic Petri Nets for very short-term wind speed modeling,'' in *Proc. IEEE Int. Conf. Comput. Intell. Virtual Environ. Meas. Syst. Appl.*, Jun. 2015, pp. 1–4.

[28] D. Montgomery and G. C. Runger, *Applied Statistics and Probability for Engineers*. Hoboken, NJ, USA: Wiley, 2010.

[29] B. D. McCullough, ''Algorithm choice for (partial) autocorrelation functions,'' *J. Econ. Social Meas.*, vol. 24, nos. 3–4, pp. 265–278, 1998.



HAMZA ABUNIMA received the B.S. degree in electrical engineering from the Islamic University of Gaza, Gaza, Palestine, in 2011, and the M.Sc. degree in electrical engineering (industrial power) from the Universiti Teknikal Malaysia, Melaka, Malaysia, in 2015. He is currently pursuing the Ph.D. degree in power systems and energy conversion with the Universiti Sains Malaysia (USM), Penang, Malaysia.



JIASHEN TEH received the B.Eng. degree (Hons.) in electrical and electronic engineering from Universiti Tenaga Nasional (UNITEN), Selangor, Malaysia, in 2010, and the Ph.D. degree in electrical and electronic engineering from the University of Manchester, Manchester, U.K., in 2016.

Since 2016, he has been a Senior Lecturer/Assistant Professor with the Universiti Sains Malaysia (USM), Penang, Malaysia. In 2018, he was appointed and served as an Adjunct Professor with the Green Energy Electronic Center, National Taipei University of Technology (Taipei Tech), Taipei, Taiwan. He is currently an Adjunct Professor with the Intelligent Electric Vehicle & Green Energy Center, National Chung Hsing University (NCHU), Taichung, Taiwan. His research interests include probabilistic modelling of power systems, grid-integration of renewable energy sources, and reliability modelling of smart grid networks.

Dr. Teh is a Chartered Engineer (CEng) conferred by the Engineering Council, U.K., and The Institution of Engineering and Technology (IET), a member of the IEEE Power and Energy Society, The Institution of Engineers Malaysia (IEM) and a Registered Engineer in the Board of Engineers Malaysia (BEM). He received the outstanding publication awards from USM, in 2017 and 2018. He is also a regular Invited Reviewers for the *International Journal of Electrical Power and Energy Systems*, IEEE Access, the IEEE Transactions on Industry Applications, the IEEE Transactions on Vehicular Technology, the IEEE Transactions on Reliability, the IEEE Transactions on Industrial Electronics, and The *IET Generation, Transmission and Distribution*.

HUSSEIN JUMMA JABIR was born in Tal Muhammad, Baghdad, Iraq, in 1977. He received the B.Sc. degree in electrical engineering from the University of Technology, Baghdad, Iraq, in 2003, and the M.Tech. degree in electrical power engineering from Jawaharlal Nehru Technological University, Hyderabad, India, in 2014. He is currently pursuing the Ph.D. degree in electrical power systems engineering with the Universiti Sains Malaysia, Nibong Tebal, Malaysia. His research interests include electrical power systems reliability and load shaping.