

# A Novel Probabilistic Generation Model for Grid Connected PV Based Distributed Generation

Muhammad Faisal Nadeem Khan\*, Tahir Nadeem Malik and Intisar Ali Sajjad

*Department of Electrical Engineering, University of Engineering and Technology Taxila, Taxila 47050, Pakistan*

*\*Corresponding Author: faisal.ndeem@uettaxila.edu.pk*

## ABSTRACT

The feasibility of renewable energy resources such as solar energy is marked by uncertainties that make it an unpredictable mean of power generation. To guarantee an uninterrupted power supply, solar irradiance modelling can be taken as a useful step towards meeting the operational challenges of electric power grid. This paper presents a dynamic Probabilistic generation model to estimate and generate the time-coupled solar irradiance patterns. Initially, clustering of yearly solar irradiance measurements is performed to obtain a meaningful grouping of similar days. One-hour time step is considered to construct a time-coupled probabilistic model of solar irradiance data based on a Beta distribution. The parameters of beta distribution are found by considering the variations of irradiance patterns at two successive time steps. The probabilistic model is then used to generate a number of aggregate solar irradiance generation scenarios. The effectiveness of proposed scenario generation approach is evaluated through Average Mean Absolute Percentage Error (AMAPE) and comparison with the probabilistic model already available in the literature.

**Keywords:** Probabilistic; modeling; clustering; solar irradiance; beta parameters.

## INTRODUCTION

Renewable energy resources have gained remarkable attention in the last few years, as they are environment friendly and have no fuel cost (Ela et al., 2013). Among these energy resources, solar Photovoltaic (PV) has the highest growth rate from 2006 to 2011 (Renewables, R., 2012). In a mere decade, solar PV power generation has witnessed global boom, as its installed capacity has reached 227GW. Moreover, it is continuously increasing at an extraordinary rate as year 2015 experienced 25% growth over 2014 (Renewables, R., 2012).

In contrast to conventional energy resources, the renewable energy resources such as wind and solar are intermittent in nature. Consequently, electric power produced from these resources is variable and uncertain. An increasing amount of research pertaining to the integration of solar PV with electric power systems has created a keen interest in understanding the variation of this resource, due to its economic and technical impacts. It is desirable that power system should operate in balanced conditions. However, the uncertainty of solar radiations and limited spinning reserve makes it extremely hard to maintain the balance of power grid. An abrupt change in electric power disturbs the power system stability and power quality (Achleitner et al., 2014). An exact idea of output power produced from solar PV module can only be gained through accurate estimation techniques.

The in-depth study of literature reveals that approximately 75% of the publications on solar irradiance estimation have been done in the last five years (Teke et al., 2015). In 1924, Angström was the first to propose a model for estimating the monthly average daily-global solar radiations using solar radiation and sun shine intensity and duration data in a clear sky (Angstrom et al., 1924). This model can be utilized for modeling both linear and nonlinear characteristics of solar radiations. A number of studies (Khatib et al., 2012, Yorukoglu et al., 2006 & Besharat et al., 2013) have discussed linear and nonlinear models for solar energy modeling. These models present association between solar radiations and other metrological variables. In Aguiar et al. (1992) statistical method is applied to analyze the variation of hourly

solar radiations at various geographical locations and in different climatic conditions. The performance of already existing models for solar radiation prediction was assessed in Gueymard et al. (2000) & Kaplanis et al. (2006). Studies such as Duzen et al. (2012) & Teke et al. (2014) have used empirical models for estimating monthly solar radiations at specific location. As an alternative to conventional modeling methods, Artificial Neural Network (ANN) techniques have become an attractive choice for solar radiation modeling due to their computational efficiency. Different types of ANN techniques such as recurrent, wavelet, Bayesian neural network, and diagonal recurrent wavelet have been exploited to improve the solar radiation estimation (Elminir et al., 2007 & Rahimikhoob et al., 2010). Reference Korachagaon (2012) implemented the Fuzzy logic for solar radiation forecasting and concludes that it produces better results in such cases where the available data for estimation is vague and ambiguous.

In Salameh et al. (1995), a Beta distribution based probabilistic generation model is presented for solar irradiance uncertainty modeling. This model is often employed in power system planning studies (Hung et al., 2014, Hagan et al., 2016, Khatod et al., 2013 & Teng et al., 2013) to determine the electric power output from historical solar radiation data. The standard deviation and mean of historical irradiance data for each hour of the day is utilized to determine the Beta distribution parameters. These parameters are used to model the hourly solar irradiance through Beta probability density function (PDF). The PDF thus obtained is used to determine the output power from the solar panel. In addition to this, Hagan et al. (2016) have proposed a more appropriate segmentation method to model the per hour solar irradiance data through Beta PDF application. Although the probabilistic model employed in Hung et al. (2014), Hagan et al. (2016), Khatod et al. (2013) & Teng et al. (2013) is well-adopted in PV DG planning studies for calculating expected output power, there is a need to develop a dynamic generation model that can deal with the current challenges of power system. The proposed model, in the present study, not only determines expected output power but also provides means to construct solar irradiance patterns. The generated patterns can further be used for scenario analysis in power system studies. Scenario analysis requires generation of several irradiance patterns to combine the characteristic of days under analysis with the irradiance patterns of other days to construct reasonable irradiance patterns, which can be used in power system planning studies such as load flow and reliability analysis.

This paper presents an effective methodology that employs Probabilistic modelling approach reported in Sajjad et al. (2015) and Khan et al. (2017). In comparison with the model presented in Khan et al. (2017), the proposed model is different with the following features: use of k-mean clustering to segregate different irradiance levels of the year; capable of handling irradiance uncertainties for different intensity levels that may represent seasons of the year; well-tuned model parameters to reduce the Average Mean Absolute Percentage Error (AMAPE) between the reference and generated data that represents the AMAPE variation with respect to different irradiance levels; more rigorous validation of the results.

In this paper, solar irradiance patterns (hourly daily irradiance data) generated at successive time steps using measured data. Initially, the collected solar irradiance data is clustered into meaningful groups of similar days. One-hour time step is considered to construct a time-coupled probabilistic model by estimating the moving window-based Beta distribution parameters. The probabilistic model is then used to generate number of solar irradiance patterns. The effectiveness of the proposed method is assessed by investigating the goodness of fitness (GOF) indicator (such as AMAPE) and comparison with the probabilistic model (Hung et al., 2014) readily used in the literature for PV DG planning studies.

## **EXTRACTION OF THE TIME-COUPLED IRRADIANCE PATTERNS**

Historical four-year horizontal solar irradiance data is collected from solar power station installed at Pakistan Engineering Council (PEC) premises. To have uniformity among days of the year, hourly solar irradiance data (6:00 am to 19:00 pm) of each day for complete period is considered and saved in the data matrix. The columns of the data matrix represent hours of the day (14 hours) and the rows represent days of the years, which are termed as irradiance patterns. The irradiance profile plot for a single year is shown in Figure 1.

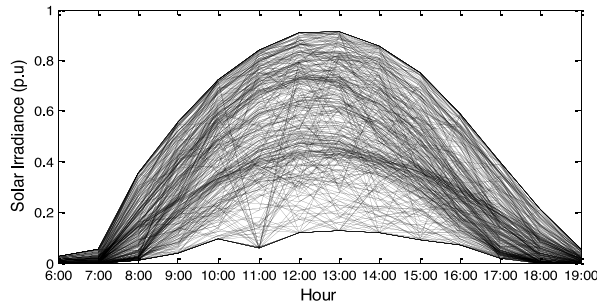


Figure 1. Solar Irradiance Profile for complete year.

### Clustering

Clustering provides an opportunity to obtain useful information from available data. It is a process that involves the partitioning and arrangement of larger set of data points into smaller groups known as clusters. The number of clusters plays a significant role in understanding the behavior of certain process. An inappropriate choice of the number of clusters will invalidate the whole process (Jimenez et al., 2016). In this paper, k-means clustering algorithm is used to group the available data in four clusters as explained in later sections. K-means is one of the simplest unsupervised learning algorithms that solve the well-known clustering problem (Nassar et al., 2015). K-means clustering is carried out by minimizing the Euclidean squared error function given in equation (1).

$$\min \left( \sum_{j=1}^k \sum_{i=1}^n \|x_i^j - c_j\|^2 \right) \tag{1}$$

where  $x_i^j$  and  $c_j$  are representations of cluster data point and centroid, respectively. The four clusters of the above data are shown in Figure 2. These clusters are named as high, medium, low, and random intensity irradiance data clusters. Irradiance patterns that belong to each cluster are arranged separately in four distinct files. The specific cluster of  $N$  irradiance patterns is presented by the following matrix representation.

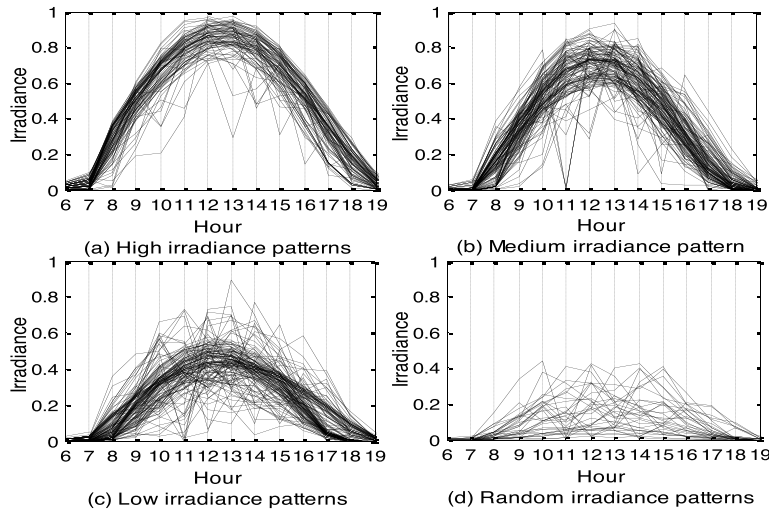


Figure 2. Irradiance patterns of various clusters

$$I = \begin{bmatrix} I_{11} & I_{12} & \dots & I_{1K} \\ I_{21} & I_{22} & \dots & I_{2K} \\ \vdots & \vdots & \dots & \vdots \\ I_{N1} & I_{N2} & \dots & I_{NK} \end{bmatrix} \tag{2}$$

Here matrix  $I$  contains the irradiance patterns of specific cluster and  $k$  represents the overall time step number. Matrix  $I$  can also be written as a series of column vectors,  $I_r = [I_{1r}, I_{2r}, \dots, I_{Nr}]^T$  via equation (3); here  $T$  indicates vector transposition. This form of  $I$  is useful to carry out further analysis in this paper.

$$I = [I_1 \ I_2 \ I_3 \ \dots \ I_K] \quad (3)$$

### Probabilistic representation of the irradiance at each time step

The key aspects that should be considered while generation of time coupled solar irradiance patterns at consecutive time steps from reference data are as follows:

- The probability distribution function of the generated irradiance patterns and reference data should be consistent for the same time step.
- The variation in time between successive time steps should represent the variations that appear in individual irradiance patterns of reference data. Hence, an appropriate representation of time coupling among next irradiance values must be exploited.

At first, upper bound and lower bound for each time step  $k = 1, 2, \dots, K$  (i.e., each column of all patterns) are calculated using 95% confidence interval and are saved in dedicated vectors.

$$I^{UB_{95\%}} = [I_1^{UB_{95\%}} \ I_2^{UB_{95\%}} \ \dots \ I_S^{UB_{95\%}}] \quad (4)$$

$$I^{LB_{95\%}} = [I_1^{LB_{95\%}} \ I_2^{LB_{95\%}} \ \dots \ I_S^{LB_{95\%}}] \quad (5)$$

where  $I^{UB_{95\%}}$  and  $I^{LB_{95\%}}$  are upper and lower bounds of solar irradiance. For analysis, the width of the amplitude range  $c^{95\%}$  is calculated as

$$c^{95\%} = I^{UB_{95\%}} - I^{LB_{95\%}} \quad (6)$$

To gain information about the values lying in the amplitude ranges, a sliding window is introduced for each time step. The width of sliding window  $\Delta c$  can be reduced via user defined parameter  $l_c$  known as window factor. Several values of  $l_c$  are examined to achieve better outcome. Subsequently, a window factor of 10 is suggested in this paper for analysis.

$$\Delta c = \frac{c^{95\%}}{l_c} = \{\Delta c_k, \ k = 1, 2, \dots, K\} \quad (7)$$

where  $\Delta c_k$  is the sliding window width for corresponding  $I_k$ . The probability distribution of the solar irradiance that occurs at consecutive time steps can be accomplished through sliding window, which attempts to achieve this initially by considering any point lying inside the window (all these points are considered as the same starting point). To construct the probability distribution at the next time step, the size of sliding window should be large enough to hold the significant amount of points. However, the size of sliding window should not be too large; otherwise, data points with relatively different irradiance values will become part of the same sliding window, and this must be avoided.

### Moving window based probabilistic representation of the irradiance for the next time step

The flow chart of the methodology for data clustering and determination of solar irradiance level at the next time step is presented in Figure 3. This involves extraction of irradiance patterns, normalization, determination of Beta parameters, and application of Piecewise Cubic Hermite interpolation polynomial (PCHIP) for smoothening of variations associated with Beta parameters.

The fundamental technique applied here is that the historical data of the current time step (used to define the width of the sliding windows) and information of preceding time step are used for computation of the irradiance level

at any time step. To do so, sliding window  $\Delta c_k$  related to step  $k$ , is moved from the initial to the final position. This determines each time the probability distribution of irradiance reached at the next time step  $(k+1)$ , by starting from any point in the amplitude ranges of sliding window. A user defined parameter  $M_c = 1000$  (normally  $M_c > l_c$ ) is used to initialize the number of central positions for the sliding window. The computation of displacement step  $s$  is extremely important for the movement of sliding window and it is determined as

$$s = \frac{c^{95\%}}{M_c - 1} \tag{8}$$

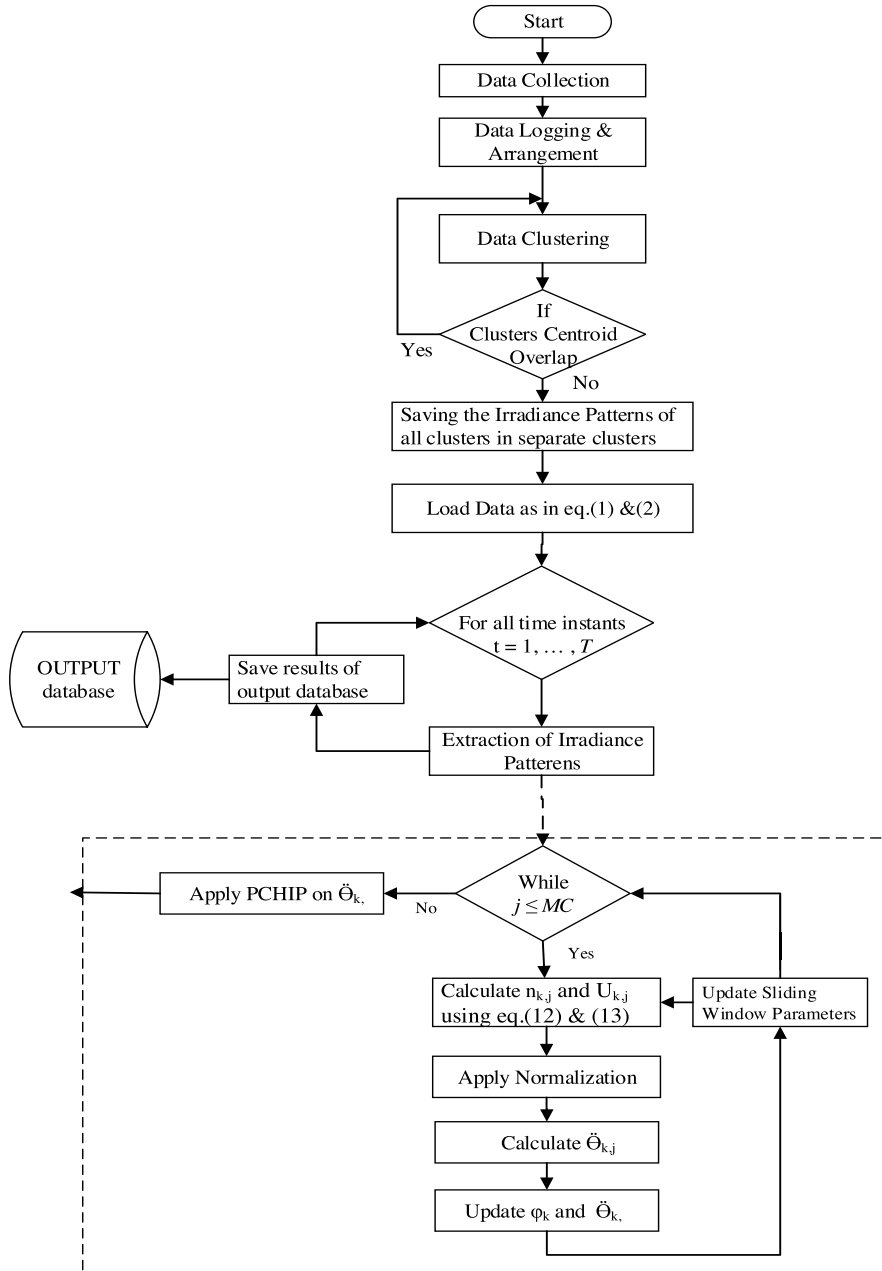
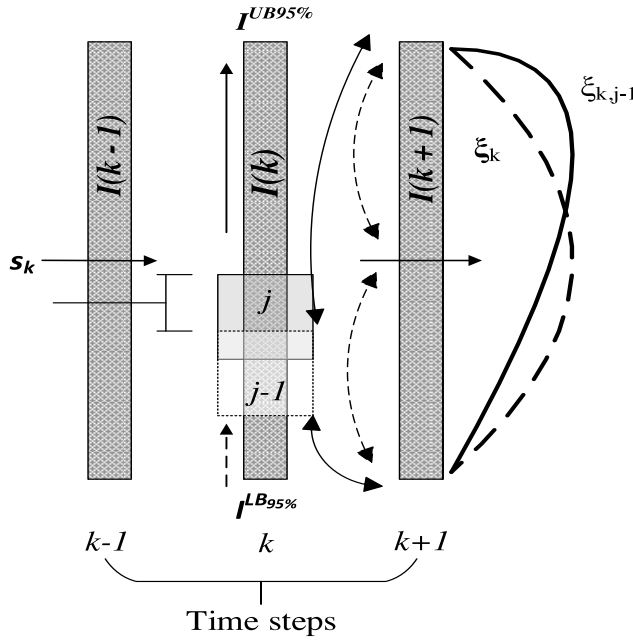


Figure 3. Flowchart to calculate probability distributions of irradiance patterns at next time step.

Initially, for time step  $k$ , the sliding window is placed at such a position that its central value is congruent with lower bound  $I_k^{LB95\%}$ . Thus, the irradiance values in the range  $I_k^{LB95\%} - (\Delta c_k/2)$  to  $I_k^{LB95\%} + (\Delta c_k/2)$  are considered by taking the relevant points from reference irradiance patterns and generating a set of values achieved at next time step  $k+1$ . Hence, the parameters corresponding with the probability distribution of the values reached at time step  $k+1$  are calculated and stored. The process to find out the beta distribution parameters for the next step is the same; however, in this case, the central value of sliding window will become  $I_k^{LB95\%} + s_k$  where  $s_k$  is the displacement step of sliding window. Figure 4 demonstrates the basic methodology of irradiance pattern generation.



**Figure 4.** Basic methodology of irradiance pattern generation model.

Irradiance values lesser than  $I_k^{LB95\%} - (\Delta c_k/2)$  and greater than  $I_k^{UB95\%} + (\Delta c_k/2)$  are not considered, as these irradiance values are excessively small or large to coincide with the generation of the time-varying irradiance patterns. In generic position  $j = 1 \dots M_c$  of sliding window at time step  $k$ , the points lying inside the sliding window are those belonging to the range with minimum value.

$$I_{k,j}^{min} = I_k^{LB95\%} + j \cdot s_k - \frac{\Delta c_k}{2} \tag{9}$$

The irradiance maximum value in the range is

$$I_{k,j}^{max} = I_k^{LB95\%} + j \cdot s_k + \frac{\Delta c_k}{2} \tag{10}$$

Thus, equations (9) & (10) are used to calculate the moving average

$$I_{k,j}^{avg} = I_k^{LB95\%} + j \cdot s_k \tag{11}$$

Now consider a set  $\phi_k$  that consists of data points reached by any irradiance pattern that belongs to the reference irradiance patterns at the preceding time step. Subsequently, those points that belong to the set  $\phi_k$  and associated with points that fall in the  $j^{th}$  sliding window in one of the reference irradiance patterns form the subset  $U_{k,j}$ . These irradiance values are obtained from the data matrix  $I$  as follows:

$$n_{k,j} = \{n | I_{k,j}^{min} \leq I_{nk} \leq I_{k,j}^{max} \forall m\} \tag{12}$$

$$U_{k,j} = \{I_{n(k+1)} | \forall n \in n_{k,j}\} \tag{13}$$

where  $n_{k,j}$  is the set of observation numbers of solar irradiance points within sliding window,  $n$  is observation number,  $I_{nk}$  is the solar irradiance at time step  $k$  for observation  $n$ , and  $U_{k,j}$  is solar irradiance points for observation number  $n_{k,j}$ .

### Normalization

It is necessary to perform normalization of the set  $U_{k,j}$  in the period (0, 1) to utilize the simple form of Beta distribution, as given in equation (14).

$$\bar{U}_{k,j} = \left[ \frac{(1 - 2\mathfrak{h})(U_{k,j} - I_k^{LB95\%})}{I_k^{UB95\%} - I_k^{LB95\%}} \right] + \mathfrak{h} \quad \mathfrak{h} = 10^{-4} \tag{14}$$

Here  $\bar{U}_{k,j}$  represents a set that contains the normalized solar irradiance values for  $n_{k,j}$  and  $\mathfrak{h}$  is the normalization significance level incorporated in this equation for the representation of Beta probability distribution such that its values are within 0 and 1.

### Determination of Beta Distribution Parameters through maximum likelihood

To obtain the beta distribution parameters, the subsequent step is to perform beta fitting on normalized subset  $\bar{U}_{k,j}$ . The Beta distribution function used in this paper is presented as follows:

$$Beta(\alpha, \beta): prob(y|\alpha, \beta) = \frac{y^{\alpha-1} (1-y)^{\beta-1}}{B(\alpha, \beta)}, \tag{15}$$

$$y \in (0,1) \wedge \alpha, \beta > 0$$

where  $\alpha$  and  $\beta$  are shape parameters and  $B(\alpha, \beta)$  is normalized function defined by equation

$$B(\alpha, \beta) = \frac{\Gamma(\alpha) \Gamma(\beta)}{\Gamma(\alpha + \beta)} \tag{16}$$

We will first consider a set  $Y_{k,j}$  of beta distribution parameters  $\alpha_{k,j}$  and  $\beta_{k,j}$ . Then a Maximum Likelihood Estimation technique (Dougherty et al., 1989) is utilized for estimation of Beta parameters. This will result in a  $Y_k$  from  $Y_{k,j}$  that corresponds to  $\phi_k$ . Equation (17) exhibits the implementation of discussed technique while vector  $n_{k,j}$  has the length  $m_{k,j}$ .

$$Y_{k,j} = \max[\log Q(\alpha_{k,j}, \beta_{k,j}, \bar{U}_{k,j})] \tag{17}$$

where  $[\log Q(\alpha_{k,j}, \beta_{k,j}, \bar{U}_{k,j})]$  is obtained from equation (18),  $\alpha_{k,j}$  and  $\beta_{k,j}$  are the parameters of beta distribution, and  $\log Q(\cdot)$  is the log likelihood function.

$$\log Q = m_{k,j} \left[ \log \left( \frac{1}{B(\alpha_{k,j}, \beta_{k,j})} \right) \right] - \left[ (1 - \alpha_{k,j}) \sum_{e=1}^{m_{k,j}} \log(\bar{U}_{k,j,e}) \right] - \left[ (1 - \beta_{k,j}) \sum_{e=1}^{m_{k,j}} \log(1 - \bar{U}_{k,j,e}) \right] \tag{18}$$

### Piecewise cubic Hermite interpolation polynomial

The data points included in the sliding window are used to estimate Beta distribution parameters at consecutive time steps. At times, the number of data points is somewhat less; therefore, it becomes difficult to determine the Beta distribution parameters that could lead to high values. The resulting Beta parameters then undergo interpolation for smoothing the variations associated with these parameters and ensure continuity. Piecewise cubic Hermite interpolation polynomial (PCHIP) is an excellent option to preserve the shape of Beta distribution parameters. Hence, the resulting  $Y_k(a_k)$  is continuous and maintains the characteristics of parameters (Nassar et al., 2015 & Dougherty et al., 1989).

We will now assume an interval  $[I_k^{LB95\%}, I_k^{UB95\%}]$ , having discrete data points  $I_{k\Delta s}^{LB95\%} = I_{k,1}^{avg}, I_{k,2}^{avg}, \dots, I_{k,M_c}^{avg} = I_k^{UB95\%}$ . Here  $a_k$  is the solar irradiance level such that  $I_k^{LB95\%} \leq a_k \leq I_k^{UB95\%}$ . Then,  $\xi_k(a_k)$  is the resultant PCHIP that includes all the data points  $(I_{k,j}^{avg}, Y_{k,j})$  so that  $\xi_k(a_k)$  is continuous.

$$\omega_{k,j}(a_k) = w_j + x_j(a_k - I_{k,j}^{avg}) + y_j(a_k - I_{k,j}^{avg})^2 + z_j(a_k - I_{k,j}^{avg})^2(a_k - I_{k,j+1}^{avg}) \quad (19)$$

$$\xi_k(a_k) = \begin{cases} \omega_{k,1}(a_k) & a_k \in [P_{k,1}^{avg}, P_{k,2}^{avg}] \\ \omega_{k,2}(a_k) & a_k \in [P_{k,2}^{avg}, P_{k,3}^{avg}] \\ \vdots & \vdots \\ \omega_{k,M_c-1}(a_k) & a_k \in [P_{k,M_c-1}^{avg}, P_{k,M_c}^{avg}] \end{cases} \quad (20)$$

The local and global cubic polynomials in equations (19) & (20) are solved for all unknown coefficients using known values of  $\omega_{k,j}(a_k)$  to determine  $\xi_k(a_k)$ . Here,  $\xi_{k,j}$  is the PCHIP function at time step  $k$ ,  $a_k$  is the bounded solar irradiance variable, and  $w, x, y, z$  are coefficients of  $\omega_{k,j}$ . A resulting  $\xi_k(a_k)$  for a complete set of values within lower and upper limits for  $a_k$  and at certain time instant is shown in Figure 5.

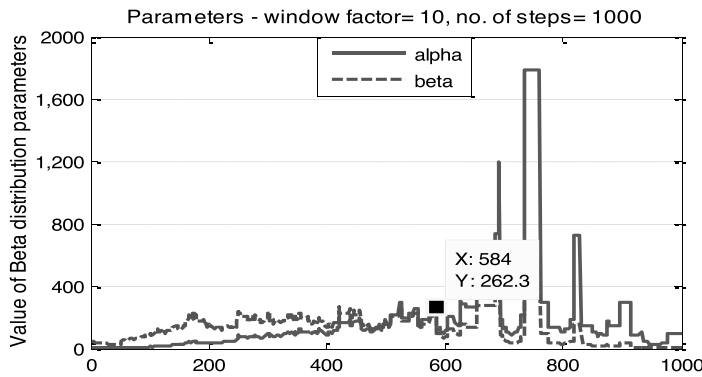


Figure 5. Beta distribution parameters after application of PCHIP fitting.

### GENERATION OF THE SOLAR IRRADIANCE PATTERNS

The basic methodology for irradiance patterns generation is shown in Figure 6.

The earlier computations in this paper have provided the probability distributions of the irradiance levels at each time step and the probabilistic link between these time steps. The outcome of these calculations is utilized to produce irradiance patterns that may be employed in scenario analysis. This requires the representation of appropriately identified random patterns for the irradiance level of a specified number of days. Let us consider a starting point  $h_k$  at  $k=1$  to generate the irradiance patterns. Then, the PCHIP results reported in equation (20) are used to determine the parameters of beta distribution at time step  $k+1$ . To differentiate among the parameters of reference and generated irradiance patterns, there is a need to introduce new variables for generated irradiance patterns. Let the parameters of



Beta distribution be  $\alpha_{h_k}$  and  $\beta_{h_k}$  for irradiance level  $h_k$ ; and for the possible irradiance at time step  $k+1$  the randomly generated cumulative probability is  $p_{k+1}$ . The normalized irradiance level  $\bar{h}_{k+1}$  is calculated by using the inverse Beta cumulative distribution function (CDF)  $O^{-1}(\cdot)$  and its mathematical expression is given below:

$$\bar{h}_{k+1} = O^{-1}(p_{k+1} | \alpha_{h_k}, \beta_{h_k}) \{ \bar{h}_{k+1} : O(\bar{h}_{k+1} | \alpha_{k_t}, \beta_{k_t}) = p_{k+1} \} \quad (21)$$

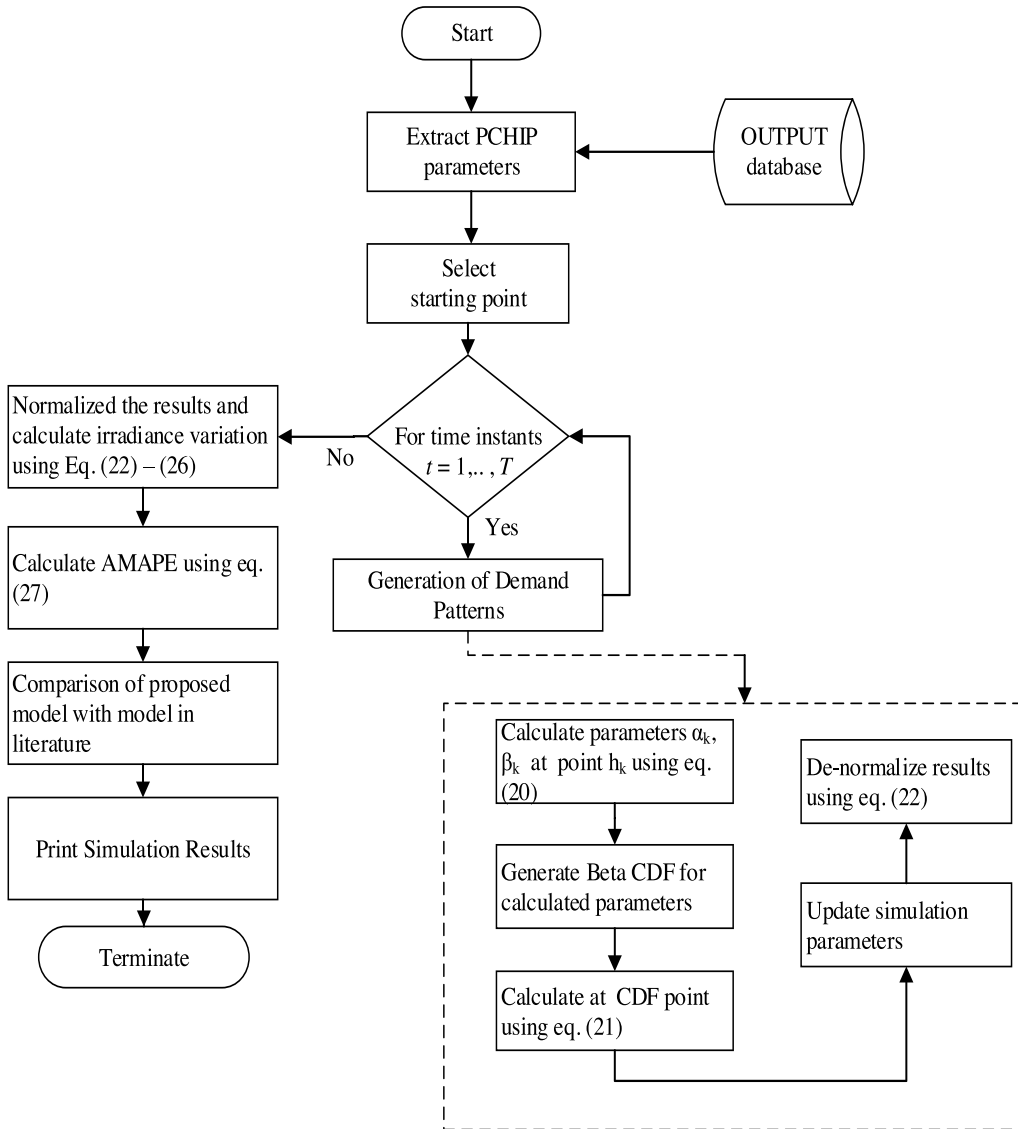


Figure 6. Irradiance Pattern generation procedure.

The irradiance  $\bar{h}_{k+1}$  is regarded as the current irradiance level. To obtain the complete irradiance pattern for all the time steps, the same procedure is repeated. The real values of solar irradiance can be retrieved by denormalizing  $\bar{h}_{k+1}$  which results in  $h_{k+1}$ .

$$h_{k+1} = \left[ \frac{(\bar{h}_{k+1} - \bar{h})(I_k^{UB95\%} - I_k^{LB95\%})}{(1 - 2\bar{h})} \right] + I_k^{LB95\%} \quad (22)$$

Since there are  $N$  irradiance patterns in a specific cluster of reference data set, a set of fixed number of irradiance patterns are generated. The irradiance patterns of generated data set can be represented by the following equation.

$$H = \begin{bmatrix} h_{11} & h_{12} & \dots & h_{1t} \\ h_{21} & h_{22} & \dots & h_{2t} \\ \vdots & \vdots & \dots & \vdots \\ h_{N1} & h_{N2} & \dots & h_{Nt} \end{bmatrix} \quad (23)$$

$$H = [h_1 \ h_2 \ \dots \ h_t] \quad (24)$$

where  $h_j = [h_{1j} \ h_{2j} \ \dots \ h_{Nj}]^T$  for  $j = 1, \dots, K$ . The change in irradiance from one-time step to the next time step can be determined through equations (25) & (26).

$$\Delta H = [\Delta h_1 \ \Delta h_2 \ \dots \ \Delta h_K] \quad (25)$$

$$\Delta h_K = h_K - h_{K-1} \quad (26)$$

## CASE STUDIES AND GOF ASSESSMENT

### Case study application

The performance of the proposed irradiance pattern generation model is investigated by utilizing the hourly solar irradiance data measured at the solar power station installed at Pakistan engineering council premises. This data set is composed of four years irradiance patterns sampled on hourly basis. The collected data is then arranged in the matrix form where each row represents the daily irradiance profile and the columns represent the hours. These irradiance patterns undergo clustering using K-means clustering algorithm. To find the suitable number of clusters of the reference data, the centroids for various cluster selections are investigated. The results for centroids of four, six, and eight clusters are shown in Figure 7.

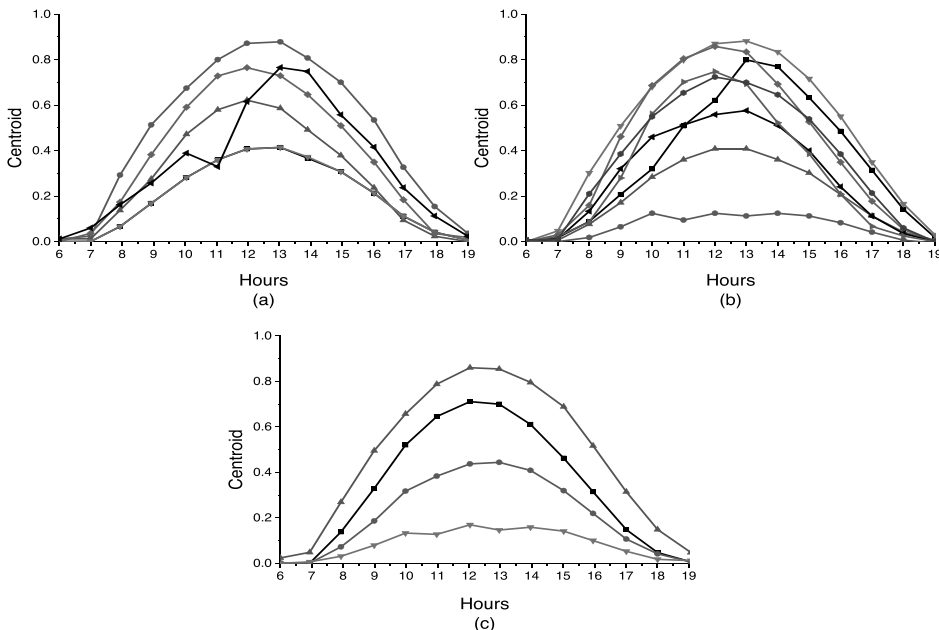


Figure 7. Centroid for (a) six, (b) eight, and (c) four clusters.

It is evident from the figure that the centroids for eight and six numbers of clusters are overlapping.

Hence, these cluster selections do not provide an information about the exact trend of irradiance data for the specified period. However, in case of four clusters, no overlapping of centroids is observed. This demonstrates that each cluster is independent of the other and holds unique information. Therefore, four clusters are considered as an appropriate choice to perform current study.

Irradiance patterns corresponding to each cluster are utilized to determine the parameters of beta distribution as shown in Figure 5 by employing the method in Figure 3. Different parameters for the generation model are given in Table 1. At times, there is a possibility that, at initial steps of sliding window, a small number of irradiance data points are available to compute the parameters of beta distribution. In such situations, the probability of occurrence of irradiance level at time step  $k+1$  is assumed as uniformly distributed among the upper and lower bounds. Therefore, the value of Beta parameters is equal to one for a limited number of initial solar irradiance amplitude steps, and this presents uniform distribution as an exclusive case of Beta distribution. It can be seen from Figure 5 that the parameters of Beta distribution behave in a different way at different time instants. This shows that the data set  $U_{k,j}$  can behave in two different ways: it can be either a leptokurtic distribution or a negatively skewed distribution. The former is a result of very high values of Beta distribution and the latter occurs in situations when  $\alpha_{nk} > \beta_{nk}$  such that more data points are located near the upper bounds.

There are three significant positions of sliding window at which variations among the beta distribution parameters can be observed. Firstly, a stark difference in the values of Beta distribution parameters becomes visible in the final steps (irradiance amplitude step 1000) of sliding window where the alpha values are considerably large than the beta values. Secondly, the converse behavior of beta distribution becomes visible during the initial steps where the beta parameter values are larger than alpha. Finally, a small variation can be observed in beta parameters at the central part of irradiance amplitude steps, which demonstrates that the majority of the points are equally distributed around the mean.

The method explained in Figure 6 is used to generate the irradiance patterns from the extracted beta distribution parameters for each cluster. The reference and generated irradiance patterns are shown in Fig. 8 (a) & (b), respectively. The bold line in Fig. 8(c) shows the upper and lower bounds of the reference irradiance patterns. It can be seen from Fig. 8(d) that the variation in amplitude of generated irradiance patterns has almost the same trend as that of reference irradiance patterns.

**Table 1.** Irradiance patterns generation model.

Parameter	Value
Confidence limit percentile	95%
Amplitude window central points $M_c$	1000
Window factor $l_c$	10
Time step duration	1 hour
Starting point	random, between 95% percentile limits

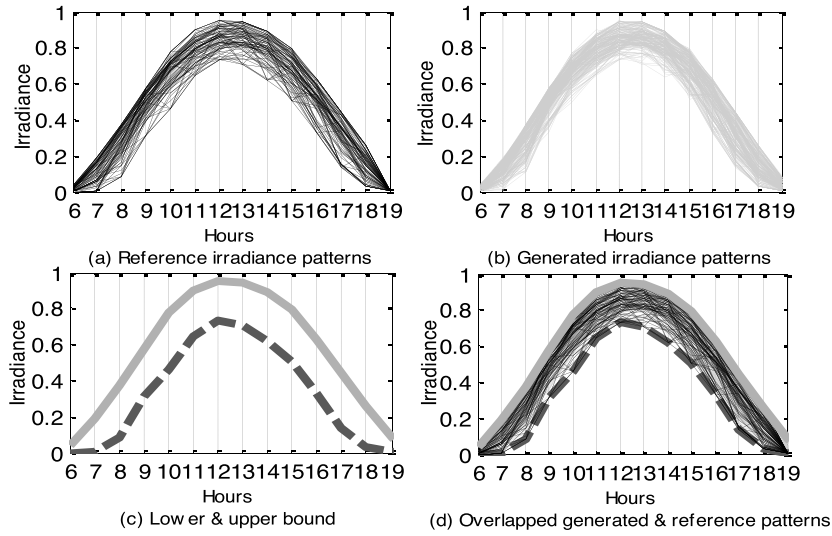


Figure 8. Generated and Reference irradiance patterns.

### GOF assessment for case studies

The accuracy of a model determines its value for use. The validity of the proposed method for solar irradiance pattern generation from the reference dataset is assessed through the application of a GOF test. For this purpose, the Average Mean Absolute Percentage Error (AMAPE) mean parameter is selected to determine the statistical error between generated and reference data (Wu et al., 2010 & Conejo et al., 2005). AMAPE expresses the difference between measured and estimated values and is an indicator of accuracy. The smaller this indicator is, the better the performance of the model is.

$$AMAPE_{mean} = 100 \times \frac{1}{S} \left( \sum_{t=i}^{i+S} \left[ \frac{|G_t - R_t|}{(\sum_{s=i}^{i+S} R_t)} \right] \right) \tag{27}$$

where S is the total number of hours considered in a day, and in this paper, it is termed as time step and S=14

$G_t$  generated irradiance patterns

$R_t$  reference irradiance patterns

S time step number

The calculations are based on 1-hour time interval step and are used to calculate  $AMAPE_{mean}$  for all the clusters. In general view, AMAPE values for solar radiation modelling are between 2.4966% and 6.934%, respectively (Angstrom et al., 1924 & Premalatha et al., 2016). The results for  $AMAPE_{mean}$  of each cluster are given in Table 2.

Table 2. Goodness of fitness assessment.

GOF	Parameter	Cluster Intensity	% Index Value
AMAPE	Mean	Medium	1.58
		High	2.23
		Random	9.58
		Low	3.23

These results clearly demonstrate that the  $AMAPE_{mean}$  values accomplished in this research work are predominantly small and are in the range mentioned in Angstrom et al. (1924) & Premalatha et al. (2016). The  $AMAPE_{mean}$  of medium intensity irradiance cluster is 1.58 %, which is least among the four, while its value for low and high intensity irradiance clusters is 2.23% and 3.23%, respectively. In case of random irradiance cluster irradiance patterns are irregular and do not follow a certain trend. Consequently,  $AMAPE_{mean}$  in this case is highest among other clusters, i.e., 9.58%. The AMAPE values of high, medium, and low intensity clusters illustrate that the proposed model is suitable for solar irradiance modelling. Furthermore, the results presented in Table 2 are in accordance with the AMAPE values available in the literature for different case studies (Wu et al., 2010 & Conejo et al., 2005).

### COMPARISON OF PROPOSED MODELLING TECHNIQUE

The probabilistic model proposed in this paper for solar irradiance modelling is validated by comparing with the model that has been employed in various PV DG planning studies (Hung et al., 2014, Hagan et al., 2016, Khatod et al., 2013 & Teng et al., 2013). To perform comparison, hourly output power is calculated for each cluster using both models. At a particular site, solar irradiance, ambient temperature, and PV module specifications are important factors in determining the output power. Matrix  $H$  in equation (23) comprises generated solar irradiance patterns where the columns of this matrix represent hours of the day (14 hours) and the rows represent days of the years. The mean values of the generated solar irradiance patterns for each hour and PV module specifications (Teng et al., 2013) given in Table 3 are used to determine the expected output power by the following equation:

**Table 3.** Specifications of PV module.

PV Module Specification	Values
Watt peak (W)	220
Open circuit voltage (V)	36.6
Short circuit current (A)	8.38
Voltage at maximum power (V)	28.36
Current at maximum power (A)	7.76
Voltage temperature coefficient (mV/°C)	0.1278
Current temperature coefficient(mA/°C)	0.00545
Nominal cell operating temperature (°C)	43

$$P_{Sy}(s_{ay}) = N * FF * V_y * I_y \tag{28}$$

and

$$T_{Cy} = T_A + s_{ay} \left( \frac{N_{OT} - 20}{0.8} \right)$$

$$I_y = s_{ay} [I_{SC} + K_i (T_C - 25)]$$

$$V_y = V_{oc} - K_v * T_{Cy}$$

$$FF = \frac{V_{MPP} * I_{MPP}}{V_{OC} * I_{SC}}$$

where the total number of PV modules is represented by  $N$ ,  $T_A$  and  $T_{Cy}$  are ambient and cell temperature (°C), respectively,  $k_v$  and  $K_i$  are voltage (V/°C) and current coefficient (A/°C), respectively,  $N_{OT}$  is the nominal operating

coefficient of cell  $I$  ( $^{\circ}\text{C}$ ), fill factor is denoted by  $FF$ ,  $I_{SC}$  and  $V_{OC}$  are short circuit current (A) and open circuit voltage, respectively, (V),  $I_{MPP}$  and  $V_{MPP}$  are current (A) and voltage (V) at maximum power point, and  $s_{ay}$  is the hourly irradiance mean.

It can be seen in Figure 9 that output power determined through the proposed model and probabilistic model (Hung et al., 2014) is identical for the same irradiance data. The comparison of results for high and medium intensity irradiance clusters is given in Figure 9. Moreover, random and low irradiance clusters also exhibited the same trend. Similar output power in case of both models demonstrate that the results of proposed model are in accordance with the model well adopted in the literature. Hence, it is suggested that the proposed model is well suited for power system planning studies. Furthermore, the proposed model provides an opportunity to generate solar irradiance patterns, which are not achievable through the probabilistic model in (Hung et al., 2014).

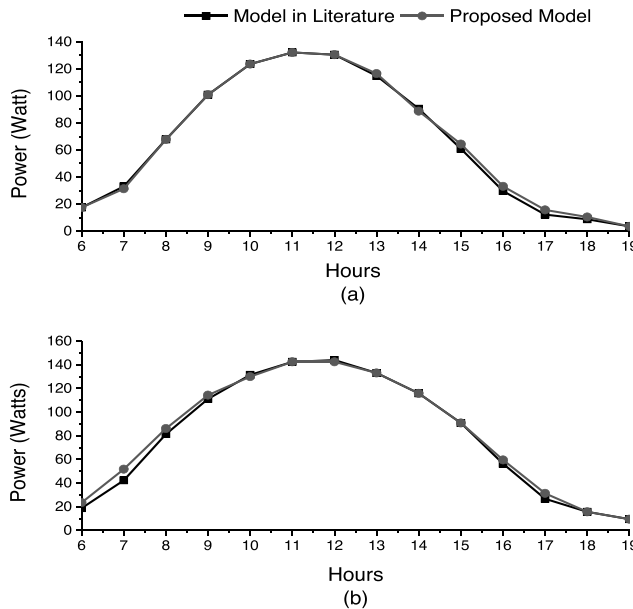


Figure 9. Comparison of proposed model with the model in literature.

## CONCLUSION

The present research employs Probabilistic model based on Beta Distribution to generate hourly solar irradiance patterns from measured solar irradiance data. The feasibility of Beta Distribution is evident from its flexibility in terms of parameters as it can change itself according to the solar irradiance profile. Prior to applying this model, yearly solar irradiance data was arranged into meaningful clusters through K means Clustering Algorithm to achieve better outcome. The Beta distribution based time-coupled probabilistic modelling is performed considering hourly data. The variations of irradiance patterns from one-time step to another are used to determine the beta distribution parameters. Hence, solar irradiance patterns are generated through probabilistic modelling.  $AMAPE_{mean}$  is used to calculate the error between generated and reference irradiance patterns, and the results demonstrate that the proposed model is suitable for solar irradiance modeling. Moreover, the hourly output power produced from the irradiance data generated through the proposed model and that from the Beta probabilistic model available in the literature are almost similar. Therefore, the proposed model can be used in power system planning studies. The Probabilistic modelling performed in this work is not merely a prediction of solar radiation, but also generates solar irradiance patterns for scenario analysis in power system.

## REFERENCES

- Achleitner, S., Kamthe, A., Liu, T. & Cerpa, A.E. 2014, April.** SIPS: Solar irradiance prediction system. In Proceedings of the 13th international symposium on Information processing in sensor networks (pp. 225-236). IEEE Press.
- Aguiar, R. & Collares-Pereira, M. 1992.** Statistical properties of hourly global radiation. *Solar Energy*, **48**(3): 157-167.
- Angstrom, A. 1924.** Solar and terrestrial radiation. Report to the international commission for solar research on actinometric investigations of solar and atmospheric radiation. *Quarterly Journal of the Royal Meteorological Society*, **50**(210): 121-126.
- Besharat, F., Dehghan, A.A. & Faghih, A.R. 2013.** Empirical models for estimating global solar radiation: A review and case study. *Renewable and Sustainable Energy Reviews*, **21**: 798-821.
- Conejo, A.J., Plazas, M.A., Espinola, R. & Molina, A.B. 2005.** Day-ahead electricity price forecasting using the wavelet transform and ARIMA models. *IEEE transactions on power systems*, **20**(2): 1035-1042.
- Dougherty, R.L., Edelman, A.S. & Hyman, J.M. 1989.** Nonnegativity-, monotonicity-, or convexity-preserving cubic and quintic Hermite interpolation. *Mathematics of Computation*, **52**(186): 471-494.
- Duzen, H. & Aydin, H. 2012.** Sunshine-based estimation of global solar radiation on horizontal surface at Lake Van region (Turkey). *Energy Conversion and Management*, **58**: 35-46.
- Elminir, H.K., Azzam, Y.A. & Younes, F.I. 2007.** Prediction of hourly and daily diffuse fraction using neural network, as compared to linear regression models. *Energy*, **32**(8): 1513-1523.
- Ela, E., Diakov, V., Ibanez, E. & Heaney, M. 2013.** Impacts of variability and uncertainty in solar photovoltaic generation at multiple timescales (No. NREL/TP-5500-58274). National Renewable Energy Laboratory (NREL), Golden, CO.
- Gueymard, C. 2000.** Prediction and performance assessment of mean hourly global radiation. *Solar Energy*, **68**(3): 285-303.
- Hagan, K.E., Oyebanjo, O.O., Masaud, T.M. & Chaloo, R. 2016, February.** A probabilistic forecasting model for accurate estimation of PV solar and wind power generation. In *Power and Energy Conference at Illinois (PECI), 2016 IEEE* (pp. 1-5). IEEE.
- Hung, D.Q., Mithulananthan, N. & Lee, K.Y. 2014.** Determining PV penetration for distribution systems with time-varying load models. *IEEE Transactions on Power Systems*, **29**(6): 3048-3057.
- Jiménez-Pérez, P.F. & Mora-López, L. 2016.** Modeling and forecasting hourly global solar radiation using clustering and classification techniques. *Solar Energy*, **135**: 682-691.
- Kaplanis, S.N. 2006.** New methodologies to estimate the hourly global solar radiation; Comparisons with existing models. *Renewable Energy*, **31**(6): 781-790.
- Khatib, T., Mohamed, A. & Sopian, K. 2012.** A review of solar energy modeling techniques. *Renewable and Sustainable Energy Reviews*, **16**(5): 2864-2869.
- Khatod, D.K., Pant, V. & Sharma, J. 2013.** Evolutionary programming based optimal placement of renewable distributed generators. *IEEE Transactions on Power systems*, **28**(2): 683-695.
- Khan, M.F.N. & Malik, T.N. 2017.** Probabilistic generation model for optimal allocation of PV DG in distribution system with time-varying load models. *Journal of Renewable and Sustainable Energy*, **9**(6): 065503.
- Korachagaon, I. 2012.** Estimating Global Solar Radiation Potential for Brazil by Iranna-Bapat's Regression Models. *International journal of Emerging Technology and Advanced Engineering*, **2**(2).
- Nassar, M.E. & Salama, M.A. 2015, May.** A novel probabilistic load model and probabilistic power flow. In *Electrical and Computer Engineering (CCECE), 2015 IEEE 28th Canadian Conference on* (pp. 881-886). IEEE.
- Premalatha, N. & Valan Arasu, A. 2016.** Prediction of solar radiation for solar systems by using ANN models with different back propagation algorithms. *Journal of applied research and technology*, **14**(3): 206-214.
- Rahimikhoob, A. 2010.** Estimating global solar radiation using artificial neural network and air temperature data in a semi-arid environment. *Renewable Energy*, **35**(9): 2131-2135.
- Renewables, R. (2012).** Global status report. Renewable Energy Policy Network for the 21st Century.

- Salameh, Z.M., Borowy, B.S. & Amin, A.R. 1995.** Photovoltaic module-site matching based on the capacity factors. IEEE transactions on Energy conversion, **10**(2): 326-332.
- Sajjad, I.A., Chicco, G. & Napoli, R. 2015.** Probabilistic generation of time-coupled aggregate residential demand patterns. IET Generation, Transmission & Distribution, **9**(9): 789-797.
- Teng, J.H., Luan, S.W., Lee, D.J. & Huang, Y.Q. 2013.** Optimal charging/discharging scheduling of battery storage systems for distribution systems interconnected with sizeable PV generation systems. IEEE Transactions on Power Systems, **28**(2): 1425-1433.
- Teke, A. & Yıldırım, H.B. 2014.** Estimating the monthly global solar radiation for Eastern Mediterranean Region. Energy conversion and management, **87**: 628-635.
- Teke, A., Yıldırım, H.B. & Çelik, Ö. 2015.** Evaluation and performance comparison of different models for the estimation of solar radiation. Renewable and Sustainable Energy Reviews, **50**: 1097-1107.
- Wu, L. & Shahidehpour, M. 2010.** A hybrid model for day-ahead price forecasting. IEEE Transactions on Power Systems, **25**(3): 1519-1530.
- Yorukoglu, M. & Celik, A.N. 2006.** A critical review on the estimation of daily global solar radiation from sunshine duration. Energy Conversion and Management, **47**(15): 2441-2450.

**Submitted:** 27/10/2017

**Revised:** 14/09/2018

**Accepted:** 12/11/2018



## نموذج احتمالي مستحدث لتوليد الطاقة المركزة على الشبكة الكهروضوئية

محمد فيصل نديم خان، طاهر نديم مالك وإنتصار علي سجاد  
قسم الهندسة الكهربائية، جامعة الهندسة والتكنولوجيا تاكسيلا، باكستان

### الخلاصة

تتميز الجدوى الاقتصادية من موارد الطاقة المتجددة مثل الطاقة الشمسية بعدم اليقين الذي يجعلها وسيلة غير متوقعة لتوليد الطاقة. ولضمان توفير الطاقة دون انقطاع، يمكن استغلال الأشعة الشمسية لمواجهة التحديات التشغيلية لشبكة الطاقة الكهربائية. يقترح هذا البحث نموذج توليد احتمالي ديناميكي لتقدير وتوليد أنماط الأشعة الشمسية المقترنة بالوقت. في البداية، تم تجميع قياسات الأشعة الشمسية السنوية للحصول على مجموعة ذات مغزى لأيام مماثلة. وتعتبر الخطوة الزمنية التي تستغرق ساعة واحدة بمثابة إنشاء نموذج احتمالي مقترن بالوقت لبيانات الأشعة الشمسية ارتكازاً على توزيع بيتا. تم العثور على معلمات توزيع بيتا من خلال النظر في الاختلافات في أنماط الإشعاع في اثنين من الخطوات الزمنية المتعاقبة. ثم تم استخدام النموذج الاحتمالي لتوليد عدد من أنماط الإشعاع الشمسي الكلي. تم تقييم فعالية الطريقة المقترحة من خلال متوسط خطأ النسبة المئوية المطلقة (AMAPE) ومقارنتها مع النموذج الاحتمالي المنشور بالفعل في المطبوعات العلمية.