

# RPG 2022

## Renewable Power Generation

22-23  
September 2022

Meeting net  
zero carbon

Savoy Place,  
London

Call for papers deadline:  
**10 December 2021**



# Application of an adaptive Bayesian-based model for probabilistic and deterministic PV forecasting

Oveis Abedinia<sup>1</sup>  | Mehdi Bagheri<sup>1</sup>  | Vassilios G. Agelidis<sup>2</sup>

<sup>1</sup> Department of Electrical and Computer Engineering, School of Engineering and Digital Sciences, Nazarbayev University, Kabanbay Batyr Ave. 53, Nur-Sultan 010000 Kazakhstan, Astana, Kazakhstan

<sup>2</sup> Centre for Electrical Power and Energy (CEE), Department of Electrical Engineering, Technical University of Denmark, Kgs. Lyngby, Denmark

## Correspondence

Oveis Abedinia and Mehdi Bagheri, Department of Electrical & Computer Engineering, Nazarbayev University, Astana 010000, Kazakhstan.  
Email: oveis.abedinia@gmail.com;  
mehdi.bagheri@nu.edu.kz

## Funding information

This work was supported in part by Collaborative Research Project (CRP) Grant of Nazarbayev University under grant no. 021220CRP0322

## Abstract

Accurate prediction of solar photovoltaic plant energy generation is essential for optimal planning and operation of modern power systems, and incorporating such plants into the energy sector. In this study, an adaptive Gaussian mixture method (AGM) and a developed variational Bayesian model (VBM) inference through multikernel regression (MkR) are utilized to assist desirable precise prediction. In this model, the MkR processes the multi-resolution solar energy signal, and then the AGM models the complex signals forecasting error. Finally, the proposed model can be optimized, and the concurrent output of the solar energy signal in both probabilistic and deterministic status can be attained through the introduction of the VBM. The solar energy output of an actual plant, including four measurement sites provided the data for the study. The results confirmed that the proposed model delivers higher prediction accuracy for both probabilistic and deterministic forecasts when compared with other well-known models.

## 1 | INTRODUCTION

While renewable energy generation sources continue to contribute an ever-higher percentage in the energy sector, the consumption continues to rise dramatically. The reason is that the electrification of transportation and the heating sectors have been making progress and their demand increased significantly.

However, the unpredictability of renewable energy generation plants and most notably solar energy power plants creates challenges for both power system operation and planning activities. Accurate prediction models for different time frames, namely, a day ahead, an hour ahead, within minutes and also seconds are required for different system functionalities and purposes, such as planning, control, and market pre-dispatch [1–10].

In [3], a correlation of the sun's radiation variabilities in near sites as a function of their distances was investigated. In [4], a sub-hourly variability of photovoltaic (PV) energy based on hourly satellite-derived insolation data was forecasted. In [5], to increase the cloud tracking model, a neural-network-based forecasting engine was presented to predict 15 min ahead.

All the prediction methods should assure suitable level of stability, accuracy and should be able to operate on different time

horizons. Various deterministic methods have been proposed by researchers for solar PV power. However, the enhancement of probabilistic methods can play a crucial role in the prediction process since they can be a suitable model for PV uncertainties and also can reduce some risks in market participation [5]. In [6], the authors proposed quartile regression models to address the abovementioned problem. A Bayesian model was presented in [7], and a hybrid model based on the k-nearest neighbours and the kernel density function was introduced in [8]. Stochastic differential equations were suggested in [9]. Additionally, a probabilistic spatial method was presented in [10]. Applying a series of multiple forecasters rather than a single forecaster to present such models is significant for obtaining improved results. Hypothetically, the single forecasters can be either probabilistic or deterministic [2, 4, 6]. The deterministic models are based on machine learning models such as gradient boosting. It would be better to have a linear collection of various probabilistic forecasters to perform aggregate forecasting, as presented in [11]. This model can be combined with a Bayesian framework that is based on a logarithmic collective [11]. A review on some prediction methods for solar signal is presented in Table 1. Also, Figure 1 shows that among the previous research works between 2010 and 2019, 55% are based on multi-layer model,

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *IET Renewable Power Generation* published by John Wiley & Sons Ltd on behalf of The Institution of Engineering and Technology

TABLE 1 Review of some solar forecasting models

Ref.	Model	Horizon	Input signal	Forecastvariable	Area	Criteria performance	Advantage and disadvantages
[12]	MTM-ANN	24 h	Latitude, longitude and altitude	Radiation	Algeria	RMSE(%) < 8 R2(%) = 92	Obtaining suitable performance compared to AR, ARMA and Markov Chain. But the compared models were used old material.
[13]	HS- MMEF	1 h	Solar signals	Radiation	USA	ImpA(%) = [-1.32-12.6] ImpR(%) = [0.95-17.04]. nRMSE(%) = [10.07-12.43] Copt,h(%) = [6.55-8.02].	Obtaining suitable performance. Many parameters have been used for setting.
[14]	WT-SVM	24 h and 30 days	Rh, T, n and Pr	Radiation	Iran	MAPE(%) = 6.99, MABE (MJ/m2) = 0.84, RMSE (MJ/m2) = 1.4 #RMSE. (MJ/m2) = 7.94, R2(%) = 90.86	High forecasting accuracy compared to ARMA, GP and ANN. The compared models were simple and old.
[15]	Montet WT-MLP-ANFIS	24 h	RH, T, n and Pr	Radiation	Enirate	-	Novel frequency pointer to judge the performance of prediction.
[16]	K-means - DTW-MMF	24 h	Radiation	Radiation	Singapore	RMSE(W/m2) = [81-161]	Suitable results in comparison with ARMA, TDNN.
[17]	K-means clustering-NAR	30 days	Radiation	Radiation	Algeria	RMSE(W h/m2) = 60.24 NRMSE = 0.198	Instability of K-means for choosing clusters and time consuming.
[18]	TB-K-means-ANN	1 h	15 signals	Radiation	11 dissimilar areas	MSE(W/m2) = [0.00193-0.0164] RMSE(W/m2) = [19.5-20.54]	Suitable results in comparison with ARMA. Instability in choosing clusters and time consuming.
[19]	WT-GTSOM- BNN	1 h	Ws, Wd and T	Radiation	Dissimilar USA areas	RMSE (W/m2) = [61.37-92.15] rRMSE = [0.39-0.57]	Outperforms other clustering methods K-means.
[20]	EEMD-LS-SVR-Kmeans-LS-SVR	24 h	Meteorological data	Radiation	China	Std.NRMS(%) = 0.025 Std.MAPE(%) = 0.0032. Std.DS(%) = -0.0098	Suitable results in comparison with other decomposition models. But long time is demanded for prediction model.
[21]	FFA-SVM	30 days	n, Tmax, Tmin	Radiation	Nigeria	RMSE(MJ/m2/day) = 2.49 R2(%) = 0.79 r(%) = 0.89 MAPE = 6.22	Suitable results in comparison with ANN and GP. The mentioned models are simple and easy.

(Continues)

TABLE 1 (Continued)

Ref.	Model	Horizon	Input signal	Forecastvariable	Area	Criteria performance	Advantage and disadvantages
[22]	PSO-ANFIS GA-ANFIS DE-ANFIS	30 days	n, Tmax, Tmin, monthly rainfall, Kt	Radiation	Malaysia	RMSE(Mj/m2) = 0.3065 r(%) = 99.63 R2(%) = 99.21 MABE(Mj/m2) = 0.2482	ANFIS-PSO was better than other coupled models, that is, GA and DE.
[23]	ARMA-GARCH	24 h	Meteorological variables	Radiation	China	MAE(Mj/m2) = [1.26-1.82] RMSE(Mj/m2) = [1.66-2.46] NSE = [0.86-0.87] R2(%) = [86.4-90.7] bR2 = [86e90]	Suitable results. Easy implementation. Just deterministic model and few comparisons to show the validity of proposed model.
[24]	ARMA-NAR	1 h	Meteorological data	Radiation	Algeria, UK, Spain	NRMSE = 0.2 (Algeria) NRMSE = 0.21 (UK) NRMSE = 0.191 (Spain)	Just outperforms single model.
[25]	NWP-ARMA-ANN	1 h	18 input signals	Radiation	Mediterranean area	nRMSE = [16.3-19.9]. R2(%) = [95.04-96.19]	Suitable results in comparison with persistence and stand-alone ANN.

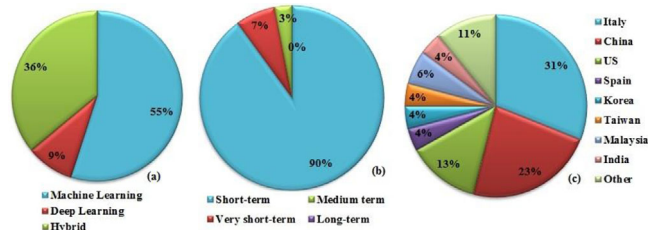


FIGURE 1 Solar energy prediction models (a), Time scale horizons (b), and involved countries for solar energy forecasting

and 36% are based on hybrid methods while 90% are studied in short term domain, in different countries.

The hybrid solar forecasting models can be considered for this problem as they are able to provide speed convergence and accurate forecasting outcomes. Moreover, the single/simple approaches can fail in desired solution finding which proofs the importance of application of hybridization methods. Additionally, it can be mentioned that all the hybrid approaches provide high performance in comparison to the stand-alone methods with various inputs/outputs, in all test cases.

The main goals of a probabilistic forecasting model are to deliver fast prediction along with proper reliability that are at the centre and considered as a core in a forecasting model. The speed can concentrate on the probabilities, and the reliability can focus on the accuracy of the probabilistic predictions. The probabilistic prediction model should be accompanied by sufficient standards to achieve quick and reliable forecasting. Some data and hence, information may be lost during the data transformation. The high-resolution solar energy data recording is the model input to prevent data and information losses.

Benefits of Gaussian mixture method (GMM) can be summarized in easy implementation and requires a small number of parameters. Furthermore, the log-likelihood function is basically simple which is employed to predict the variables. Moreover, the Bayesian model is used in this work. It corresponds to the minimum variational free energy and a lower bound of the Bayesian evidence, which is in turn a vital factor for choosing the model. The Bayesian model can be also effective in approximation method since the variational stochastic complexity displays the distance from the variational posterior distribution to the true Bayesian posterior distribution in terms of Kullback information. Additionally, this model can affect hyperparameters on the learning process while, the VBM minimizes the variational free energy, the derived bounds specify how the hyperparameters impact the learning procedure. Hence, the advantages of proposed model can solve the proposed forecasting problem as complex signals can be incorporated, adaptive spatial or other regularizing priors can be used in this model, and it can be also extended to hierarchical models. In this study, we incorporate the proposed model based on adaptive Gaussian mixture method (AGM) and a developed variational Bayesian model (VBM) inference through multikernel regression (MkR).

Moreover, a mean-max discrepancy method is considered that addresses the kernel function to obtain the appropriate features in Hilbert space. In the next step, the proposed improved

mean-max discrepancy (IMD) function maps the features to the Hilbert space. The kernel function could also map the processed data into the Hilbert space within the required time-frame. The construction of the MkR must use this technique, which employs multiresolution information. The MkR processes the multiresolution solar energy signal and then the AGM will model the complex signal's forecasting error. Finally, the proposed model can be optimized, and the concurrent output of the solar energy signal in probabilistic and deterministic status can be attained by utilizing variational Bayesian methods that cause each forecast to possess a continuous probability density function (PDF).

The key considerations and contributions of this paper can be summarized as follows:

- Applying an improved version of a mean-max discrepancy model based on allocating a weight for the samples' functions.
- Utilizing the AGM method to model the complex signals of the forecasting error.
- Integrating the VBM model to optimize the proposed model.
- Deriving concurrently, the deterministic and probabilistic forecasts from the forecast distributions.
- Applying the AGMs as posterior distributions of the solar energy forecasts.

The remainder of the paper is organized as follows. Section 2 introduces the proposed forecasting model. Section 3 presents the estimation criteria for the prediction process. The case study is presented in Section 4. Section 5 presents the numerical results and analysis, and finally Section 6 summarizes the conclusions of the paper.

## 2 | PROPOSED FORECASTING MODEL

### 2.1 | Proposed IMD model

In this section, the proposed method is introduced. The mechanism of mean-max discrepancy starting with mapping of  $x_i$  and  $x_j$  (as samples) according to the  $\phi(\mathbf{x})$  function based on a random variable  $\mathbf{x} \in \mathbf{R}^d$ . Then, the kernel function  $K(\mathbf{x}_i, \mathbf{x}_j) = \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle$  assesses two samples' similarity, where the inside creation of the candidates is determined by  $\langle \cdot, \cdot \rangle$ . The Gaussian kernel that is the most common form of the kernel function is computed as follows:

$$K(x_i, x_j) = \exp\left(\frac{\|x_i - x_j\|^2}{-2\sigma^2}\right). \quad (1)$$

Based on (1), the Euclidean distance between  $\mathbf{x}_i$  and  $\mathbf{x}_j$  is the criterion for computing the similarity between them. However, according to the accidental parameters of  $\mathbf{X}$  and  $\mathbf{Y}$  by the  $p(\mathbf{X})$  and  $p(\mathbf{Y})$  PDF, respectively, it is not appropriate to use the Euclidean distance. The distance between distributions  $p(\mathbf{X})$  and  $p(\mathbf{Y})$  is measured by mean-max dependency utilization.

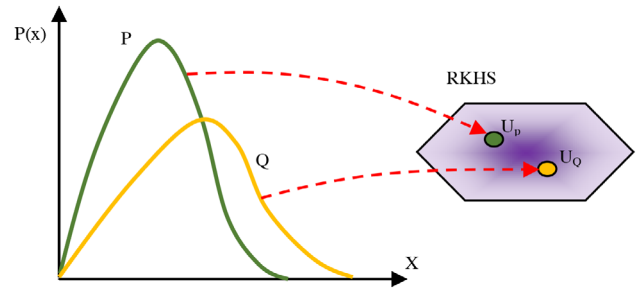


FIGURE 2 Concept of the mean-max discrepancy strategy

The description is provided in Figure 2 and the calculation is provided in [26].

$$MMD(p(\mathbf{X}), p(\mathbf{Y})) = \|\mu_X - \mu_Y\|_H^2. \quad (2)$$

In (2), the  $\mu_X$  introduces the kernel embedding in *RKHS* for the  $p(\mathbf{X})$  distribution and can be obtained as;

$$\mu_X := E_X[\phi(\mathbf{X})] = E_X[K(\mathbf{X}, \cdot)] = \int \phi(\mathbf{x}) dp(\mathbf{x}). \quad (3)$$

The practical application of  $p(\mathbf{X})$  is indirect. In other words, a set of samples of  $\{X_1, X_2, \dots, X_N\}$ , which are prolated by  $p(\mathbf{X})$ , are used instead of  $p(\mathbf{X})$ . Then, the computation of the empirical kernel embedding can be obtained through:

$$\mu_X = \frac{1}{N} \sum_{i=1}^N K(X_i, \cdot). \quad (4)$$

Accordingly, the unbiased evaluator of the mean-max dependency can be calculated as shown in [26].

$$\begin{aligned} \bar{MMD}(p(\mathbf{X}), p(\mathbf{Y})) &= \frac{1}{N_X(N_X - 1)} \sum_{i=1}^{N_X} \sum_{j \neq i}^{N_X} K(X_i, X_j) \\ &+ \frac{1}{N_Y(N_Y - 1)} \sum_{i=1}^{N_Y} \sum_{j \neq i}^{N_Y} K(Y_i, Y_j) \\ &- \frac{2}{N_X N_Y} \sum_{i=1}^{N_X} \sum_{j=1}^{N_Y} K(X_i, Y_j). \end{aligned} \quad (5)$$

Finally, the resemblance among two accidental parameters can be presented as:

$$K(\mu_X, \mu_Y) = \exp\left(-\frac{\bar{MMD}}{2\sigma_\mu^2}\right). \quad (6)$$

#### 2.1.1 | Enhanced MMD

Due to non-existence bias between distributions, Equation (5) will be evaluated. It is the estimator calculation that can cause

unbiased samples from the solar signal candidates' distribution. In other words, if the samples are biased in our target distribution, that is, the solar signal, this evaluation can guess the variance among the simulations and the biased empirical distribution. Thus, our model will internally recalculate the biases. To address this problem, an enhanced version of this model is presented while if the normalized constant of Z is assumed, the unbiased estimator can be calculated among the  $P(y)$  and  $P(x)$  by allocating a weight for the functions of the samples based on  $\lambda(x)P(x)$  and the likelihood ratio of  $P(x)/\lambda(x)P(x) = 1/\lambda(x)$ , as follows:

$$\begin{aligned} \bar{MMD}_w &= \frac{1}{N_X(N_X - 1)} \sum_{i=1}^{N_X} \sum_{j \neq i}^{N_X} \frac{1}{\lambda(X_i)\lambda(X_j)} K(X_i, X_j) \\ &+ \frac{1}{N_Y(N_Y - 1)} \sum_{i=1}^{N_Y} \sum_{j \neq i}^{N_Y} K(Y_i, Y_j) - \frac{2}{N_X N_Y} \\ &\times \sum_{i=1}^{N_X} \sum_{j=1}^{N_Y} \frac{1}{\lambda(X_j)} K(X_i, Y_j). \end{aligned} \quad (7)$$

However, as the exact value of  $\lambda(x)$  is not known, (7) must change to (8) based on the self-normalized weights:

$$\begin{aligned} IM_w &= \sum_{i=1}^{N_X} \sum_{j \neq i}^{N_X} \omega(X_i)\omega(X_j)K(X_i, X_j) \\ &+ \frac{1}{N_Y(N_Y - 1)} \sum_{i=1}^{N_Y} \sum_{j \neq i}^{N_Y} K(Y_i, Y_j) \\ &- \frac{2}{N_Y} \sum_{i=1}^{N_X} \sum_{j=1}^{N_Y} \omega(X_i)K(X_i, Y_j), \end{aligned} \quad (8)$$

where,  $\omega(X_i) = \frac{1/\lambda(X_i)}{\sum_{j=1}^{N_X} 1/\lambda(X_j)}$ .

## 2.2 | The classic model of the Gaussian mixture

The classic version of the Gaussian mixture (GM) is composed of K Gaussian members [27]. Therefore, to obtain the probability density (PD), those Gaussian constituents should be added as obtained by:

$$p(x) = \sum_{n=1}^K p(n)p(x|n) = \sum_{n=1}^K \pi_n N\left(x \middle| \mu_n, \sum_n \right) \quad (9)$$

Taking the logarithmic function was necessary due to the small result of the multiplication of multiple probabilities. The

maximum likelihood function is expressed as follows:

$$\max \sum_{i=1}^N \log \left( \sum_{n=1}^K \pi_n N(x_i | \mu_n, \sigma_n) \right). \quad (10)$$

*Phase one:* The primary GM variable or the variables from the prior iterative evaluation based on the PD for each K-GC generated signal are used [27].

$$\gamma(i, n) = \frac{\pi_n N(x_i | \mu_n, \Sigma_n)}{\sum_{j=1}^K \pi_j N(x_i | \mu_j, \Sigma_j)}. \quad (11)$$

*Phase two:* The results from the first step by estimating the parameters of the Gaussian model are used.

$$\mu_n = \frac{1}{N} \sum_{i=1}^N \gamma(i, n)x_i, \quad (12)$$

$$\Sigma_n = \frac{1}{N_n} \sum_{i=1}^N \gamma(i, n)(x_i - \mu_n)(x_i - \mu_n)^T, \quad (13)$$

$$N_n = \sum_{i=1}^N \gamma(i, n). \quad (14)$$

*Phase three:* The first two phases are repeated until the likelihood function value is less than the assumed threshold. In other words, the first two phases are repeated to obtain a stable value for the likelihood function.

### 2.2.1 | Adaptive GM strategy

The classic GM method possesses a fixed mixing degree. The main idea of the developed algorithm comprises three aspects. Firstly, for a Gaussian component to be useful, it should not have a light weight, and the related space, as well as the other Gaussian component (GC) should not be less than a specific value. Secondly, both GCs must avoid weight values being too large and thirdly, both GCs must avoid distance values being too small due to the merged GC. Additionally, if a GC's weight is large and its variance value is very large, the GC that shows multiple fragments of the signal specification distribution will be scattered.

The result of this division of the GC should be two components. The process can be presented as follows:

1) the GC's weight is defined as  $\pi_i$ , which should be smaller than the threshold  $\theta_i$ . Furthermore, the space between the GCs should be greater than  $D_1$ . After omitting the defined component, that is,  $\pi_i$ , its Gaussian component weight was allocated to the other components. Accordingly, the related degree of GM can be  $-1$ . In this manner, if the multiple components are removed, the lowest GC weights should be removed. Specifically, the space of the  $i^{\text{th}}$  and  $j^{\text{th}}$  GCs are computed via the

following formula:

$$D_{ij} = \left\| \bar{\mu}_i - \bar{\mu}_j \right\| = \sqrt{\sum_{l=1}^N (\mu'_i - \mu'_j)^2}. \quad (15)$$

In (15), the dimension of mean GC is introduced by  $l = 1, 2, \dots, N$ .

2) A smaller space among the GCs when compared to  $D_2$  as the threshold and lower weights of the two GCs when compared to  $\theta_2$  showed that the piece of information conveyed by the two GCs was identical. Thus, the variables can be changed as follows:

$$\begin{cases} \pi_T = \pi_i + \pi_j, a_i = \frac{\pi_i}{\pi_T}, a_j = \frac{\pi_j}{\pi_T} \\ \mu_T = a_i \mu_i + a_j \mu_j, \Sigma_T = a_i \Sigma_i + a_j \Sigma_j \end{cases}. \quad (16)$$

The  $i^{\text{th}}$  and the  $j^{\text{th}}$  Gaussian components were substituted with the new Gaussian component. Therefore, the related degree of the GM can be  $-1$ . If the merging conditions were met for multiple GCs, then those Gaussian components with the nearest distance would be the target of merging.

3) The Gaussian component conveyed too much information when the GC's weight was larger than the defined threshold  $\theta_3$ , while the  $\delta$  was larger as the variance. It needed to be separated and its parameters changed as follows:

$$\begin{aligned} \pi_T &= \frac{1}{2} \pi_i + \frac{1}{2^n}, \pi_{T+1} = \frac{1}{2} \pi_i - \frac{1}{2^n} \\ \mu_T &= \mu_i + aE, \mu_{T+1} = \mu_i - aE \\ \Sigma_T &= \Sigma_{T+1} = (1 + \beta)^{-1} \Sigma_i. \end{aligned} \quad (17)$$

As is evident in the formula, the diagonal of  $\Sigma_i$  was  $a = \frac{\Sigma_i^*}{2^n}$ ,  $\Sigma_i^*$  with the maximum value. The dimension and all elements of the matrix  $E = [1, \dots, N]$  were  $N$  and  $1$ , respectively. The original Gaussian component was replaced by the most recent two Gaussian components  $\{\pi_T, \mu_T, \Sigma_T\}$  and  $\{\pi_{T+1}, \mu_{T+1}, \Sigma_{T+1}\}$ . Hence, the related degree can be set to  $+1$ . In this manner, if an appropriate condition is prepared for multiple GCs, then the largest GC weight can be selected.

The first parameter should go through the updating process; then, the expectation maximization (EM) algorithm should be used for training. The number of training iterations should reach the preset value; thus, the aforementioned two steps should be repeated continuously until the value is reached.

### 2.3 | Developed variational Bayesian model

Independent observations ( $n$ ) should be denoted by  $\mathbf{x} = (x_1, \dots, x_n)$ . In our context, the Dirichlet distribution is possible by allowing  $p(\theta)$  be the past dispersal of a Bayesian model [28]. In the following one can see the formula  $\theta \sim D(\alpha_1, \dots, \alpha_K)$ . In this case, the hyperparameters are defined by  $\alpha_k > 0, k = 1, \dots, K$ .

Moreover, the marginal likelihood is introduced as:

$$m(\mathbf{x}) := \int_{\Theta_K} p(\mathbf{x} | \theta) p(\theta) d\theta = \int_{\Theta_K} \sum_{Z \in \mathfrak{S}} p(\mathbf{x} | \mathbf{z}) p(\mathbf{z} | \theta) p(\theta) d\theta. \quad (18)$$

It is very important to estimate this quantity because it is used for model selection. In general settings, although Markov Chain Monte Carlo (MCMC) estimation of  $m(\mathbf{x})$  is probable, it is complicated [28–30].

#### 2.3.1 | Classic variational estimator model

The lower-bound ( $L$ ) of  $\log m(\mathbf{x})$  is used to compose the variational Bayesian (VB) methods. Therefore, an approximate distribution  $q$  should be diverged from the true posterior distribution  $p$ . The divergence should be in a free-form minimization of the Kullback–Leibler (KL) divergence, summarized as  $KL(q || p)$ .

$$\log m(\mathbf{x}) = L + KL(q || p). \quad (19)$$

Since KL is positive, it is obvious that the KL's divergence minimization is matched by the lower constraint's maximization. The optimization should be done within a restricted range of distributions to manage the problem as follows:

$$\mathfrak{S} = \left\{ g(\theta, \mathbf{z}) = g(\theta)g(\mathbf{z}) : g(\mathbf{z}) = \prod_{i=1}^n \prod_{k=1}^K \phi_{ik}^{z_{ik}} \right\}. \quad (20)$$

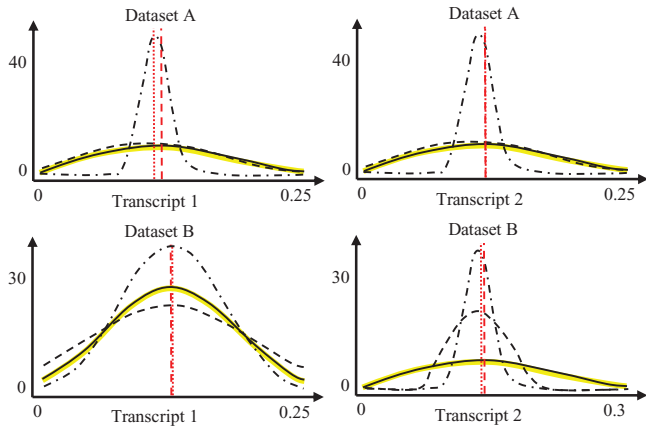
The varied parameters are  $\phi = (\phi_{ik} : i = 1, \dots, n, k = 1, \dots, K)$ . In (18),  $\mathbf{z}$  and  $\theta$  are independent because the mixture models are based on the VB methodology. Additionally, parameterizing is performed only for the distribution of  $\mathbf{z}$ , while the explicit analytical evaluation of the boundary is used for the approximate distribution of  $\theta$  as a byproduct.

By considering the distributions in (18), [29] imitates a constraint to the logarithm of (18) in the following manner:

$$\begin{aligned} \log p(\mathbf{x} | \theta) &= \log \sum_{Z \in \mathfrak{S}} p(\mathbf{x}, \mathbf{z} | \theta) = \log \sum_{Z \in \mathfrak{S}} \frac{p(\mathbf{x}, \mathbf{z} | \theta) q(\mathbf{z})}{q(\mathbf{z})} \\ &= \log i_{q(\mathbf{z})} \frac{p(\mathbf{x}, \mathbf{z} | \theta)}{q(\mathbf{z})} \geq i_{q(\mathbf{z})} \left\{ \frac{p(\mathbf{x}, \mathbf{z} | \theta)}{q(\mathbf{z})} \right\} \\ &= \sum_{i=1}^n \sum_{k=1}^K \phi_{ik} \{ \log f_k(x_i) + \log \theta_k - \log \phi_{ik} \}. \end{aligned} \quad (21)$$

Considering (18) for all  $q(\mathbf{z}) \in \mathfrak{S}, \theta \in \Theta_K$ , we can write the following:

$$\begin{aligned} m(\mathbf{x}) &\geq \int_{\Theta_K} \exp \left( \sum_{i=1}^n \sum_{k=1}^K \phi_{ik} \{ \log f_k(x_i) + \log \theta_k - \log \phi_{ik} \} \right) \\ &\quad \times p(\theta) d\theta. \end{aligned} \quad (22)$$



**FIGURE 3** The kernel density of  $\theta_k | \mathbf{x}$ ; MCMC (Yellow line), normal VB (black dashed dotted line), Dirichlet (black dashed line), generalized Dirichlet (black solid line), predicted MCMC (red dashed line) and VB means (red dotted line)

In the above equation, the analytical assessment can be presented as follows:

$$\log m(\mathbf{x}) \geq \mathfrak{F}_1(\phi) := c + \sum_{k=1}^K \log \Gamma \left( a_k + \sum_{i=1}^n \phi_{ik} \right) + \sum_{i=1}^n \phi_{ik} \{ \log f_k(x_i) - \log \phi_{ik} \}. \quad (23)$$

$C$  and  $\Gamma(\cdot)$  denote a known constant (not dependent on  $\phi$ ) and the gamma function, respectively. Moreover, according to the significance of the analytical integration of (22),  $\theta$  is approximately distributed, as follows:

$$q(\theta) = D(\gamma_k; k = 1, \dots, K). \quad (24)$$

In the above formula,  $\gamma_k = \alpha_k + \sum_{i=1}^n \phi_{ik}$ . Then, the element in  $\mathfrak{F}_1(\phi)$  could be maximized by using an EM-revised  $\phi$ . The maximization of Equation (23) occurs using the proposed algorithm in the slope. Therefore,  $-\underset{p}{(\theta_k | \mathbf{x})}$  indicates an estimate of the density of future bordering in  $k = 1, \dots, K$  rising from extended MCMC executions that are assumed to be the truth [30].

An evaluation of  $-\underset{p}{(\theta_k | \mathbf{x})}$  against the VB estimate in Equation (24) is illustrated by the red lines in Figure 3. The primary downside of this technique is that it leads to variance inaccuracy, although this approach displays favourable performance in terms of the means of posterior (the black and red dashed perpendicular lines). In Equation (24), the KL convergence between the combined posterior  $p(\theta, \mathbf{z} | \mathbf{x})$  and the circulations measured in Equation (21) is increased. Therefore, the circulation would be improved. Nonetheless, it does not serve as the ‘‘superlative’’ Dirichlet estimate of the posterior  $p(\theta | \mathbf{x})$ , considering the grey curves. The next section presents methods for achieving an improved estimate.

### 2.3.2 | Estimating a non-increased posterior

Based on varied inferential points of view, the hidden variables’ ( $\mathbf{z}$ ) independency and the mixture model weights ( $\theta$ ) are assumed for convenience; however, this could be perceived as an oversimplified approach.

This approach could be moderated by amalgamating the hidden variables from the estimation. First, if the standard differential bound ( $L_1$ ) is compared with a novel bound ( $L_2$ ) from the relegation of the same estimated circulation over the hidden space, it is found to be inferior. This is confirmed in the subsequent scheme [31].

**Scheme 2.3.2.1**  $q(\theta)$  is designated and

$$\mathfrak{F}_2 := \int_{\Theta_K} \{ \log p(\mathbf{x} | \theta) + \log p(\theta) - \log q(\theta) \} q(\theta) d\theta, \quad (25)$$

$$\log m(x) \geq \mathfrak{F}_2 \geq \mathfrak{F}_1. \quad (26)$$

The equality holds only under the condition that  $q(\theta, \mathbf{z}) = p(\theta, \mathbf{z} | \mathbf{x})$ ,  $\forall \theta, \mathbf{z}$ .

*Proof.* Showing that Equation (25) is a lower bound (LB) of  $\log m(\mathbf{x})$  is simple. Certainly, the LB between  $q$  and  $p$  is signified as follows:

$$\begin{aligned} L &= \log m(\mathbf{x}) KL(q(\theta) \| p(\theta | \mathbf{x})) \\ &= \log m(\mathbf{x}) - \int \{ \log q(\theta) - \log p(\mathbf{x} | \theta) \\ &\quad - \log p(\theta) + \log m(\mathbf{x}) \} q(\theta) d\theta \\ &= \int_{\Theta_K} \{ \log p(\mathbf{x} | \theta) + \log p(\theta) - \log q(\theta) \} q(\theta) d\theta =: \mathfrak{F}_2. \end{aligned} \quad (27)$$

Next, the second inequality of Equation (26) is proved by the log-sum variation  $\forall \theta \in \Theta$ , as follows:

$$\begin{aligned} q(\theta) \log \frac{q(\theta)}{p(\theta | \mathbf{x})} &= \left( \sum_{\mathbf{z}} q(\theta, \mathbf{z}) \right) \log \frac{\sum_{\mathbf{z}} q(\theta, \mathbf{z})}{\sum_{\mathbf{z}} p(\theta, \mathbf{z} | \mathbf{x})} \\ &\leq \sum_{\mathbf{z}} q(\theta, \mathbf{z}) \log \frac{q(\theta, \mathbf{z})}{p(\theta, \mathbf{z} | \mathbf{x})}. \end{aligned}$$

If both sides of the previous discrimination are integrated, we conclude the following:

$$\begin{aligned} \int q(\theta) \log \frac{q(\theta)}{p(\theta | \mathbf{x})} d\theta &\leq \int \sum_{\mathbf{z}} q(\theta, \mathbf{z}) \log \frac{q(\theta, \mathbf{z})}{p(\theta, \mathbf{z} | \mathbf{x})} d\theta \\ &\Leftrightarrow KL(q(\theta) \| p(\theta | \mathbf{x})) \leq KL(q(\theta, \mathbf{z}) \| \log p(\theta, \mathbf{z} | \mathbf{x})). \end{aligned} \quad (28)$$



Provided that Equation (28) is replaced by Equation (27), it could be concluded that  $\wp_2 \geq \wp_1$ . Note that if and only if  $q(\theta, \mathbf{z}) = p(\theta, \mathbf{z} | \mathbf{x}) = \forall \theta, \mathbf{z}$  the log-sum inequality perceived as equality; this could accomplish the proof [31].

Considering the same estimation of  $\theta | \mathbf{x}$ ,  $\wp_2$  as an LB could be a more accurate approximate calculation. Marginalizing over  $\mathbf{z}$  and neglecting the factorization presumption make it different than the initial bound. A more accurate estimate is not a remedy for this problem; however, an estimated circulation inside a distinct category of circulations should be determined to improve this lower bound. A discussion of it is provided in the following.  $f$  signifies every subcategory of circulations with  $q(\theta) \in f$ . Equation (26) suggests the following:

$$\min_{f \in \mathcal{F}} KL(f(\theta) \| p(\theta | \mathbf{x})) \leq KL(q(\theta, \mathbf{z}) \| p(\theta, \mathbf{z} | \mathbf{x})). \quad (29)$$

Therefore, first, a fixed  $f$  that comprises  $q(\theta)$  must be specified. Then, optimization to find the best approximating circulations to construct an improved border of the bordering probability is performed. The parameter space of “ $f$ ” is represented through  $\Delta_f$  and adopts  $\delta \in \Delta_f$ . The boundary in the calculation is converted to the following formula by knowing that the estimation “ $f$ ” is changing in the set  $f$ :

$$\wp_2(\delta) = \int_{\Theta_k} \{\log p(\mathbf{x} | \theta) + \log p(\theta) - \log f(\theta; \delta)\} f(\theta; \delta) d\theta. \quad (30)$$

Logging  $m(x)$  for all  $\delta \in \Delta_f$  is performed by Equation (30). There is an emphasis on the inability to compute Equation (30) with a static  $\delta$ . Nevertheless, Equation (30) can be estimated through a suppositional estimate because it represents the random variable mean  $g(\theta) := \log p(\mathbf{x} | \theta) + \log p(\theta) - \log f(\theta; \delta)$ ,  $\theta \sim f(\cdot; \delta)$ . Consequently, the unbiased drive could be written as follows:

$$\max_{\delta \in \Delta_f} \wp_2(\delta) = \max_{\delta \in \Delta_f} E_{\delta} g(\theta), \theta \sim f(\cdot; \delta) \in f. \quad (31)$$

### 2.3.3 | Structure of “ $f$ ”

Remaining with the Dirichlet function to have the best differential estimation, especially in the case of the joint posterior, should be considered. The subclass of the Dirichlet circulations family should be searched to find an obvious choice for  $f$ . Furthermore, it would be useful to consider a family with an even wider series, that is, a Dirichlet group of circulations that is representative of the whole [29]. A generalized Dirichlet circulation could be expressed via the VB solution as follows

$$q(\theta) = D(\gamma_1, \dots, \gamma_K) \equiv \mathfrak{D}D(\gamma_1, \dots, \gamma_{K-1}; \gamma_1^+, \dots, \gamma_{K-1}^+), \quad (32)$$

in which  $\gamma_l^+ := \sum_{j=l+1}^K \gamma_j$ ,  $l = 1, \dots, K-1$ . Subsequently, two different parameters of  $f$  are elaborated in Equation (31),

signified via  $f_D$  and  $f_{\mathfrak{D}D}$  with  $f_D \subset f_{\mathfrak{D}D}$ . Consider that parameters  $(\gamma, \gamma^+) \rightarrow (\tilde{\gamma}, \tilde{\gamma}^+)$  must be changed to perform this calculation. The same mean value is retained as the original VB circulation to simplify the optimization. This is sensible by assuming replication of the study; this circulation is fairly precise in valuing the posterior means.

The first alteration is based on only one variable  $\delta \in \mathbb{R}$ , as follows:

$$(\tilde{\gamma}_k, \dots, \tilde{\gamma}_k^+) = (e^{\delta} \gamma_k, \dots, e^{\delta} \gamma_k^+), k = 1, \dots, K-1. \quad (33)$$

Considering that this alteration suggests the following:

$$\begin{aligned} \tilde{\gamma}_{k+1} + \tilde{\gamma}_k^+ &= e^{\delta} \gamma_{k+1} + e^{\delta} \gamma_{k+1}^+ = e^{\delta} (\gamma_{k+1} + \gamma_{k+1}^+) = e^{\delta} \gamma_k^+ \\ &= \tilde{\gamma}_k^+, \forall k = 1, \dots, K-1. \end{aligned}$$

As a result, with respect to every  $\delta$ , the consequential circulation affiliates to the subsection of the Dirichlet group, as follows:

$$f_D := \{D(e^{\delta} \gamma_1, \dots, e^{\delta} \gamma_k) : \delta \in \mathbb{R}\}. \quad (34)$$

The additional alteration reduces the limit of the residual inside the Dirichlet family; currently,  $\delta = (\delta_1 \dots \delta_{K-1})$ .

$$(\tilde{\gamma}_k, \tilde{\gamma}_k^+) = (e^{\delta_k} \gamma_k, e^{\delta_k} \gamma_k^+), k = 1, \dots, K-1. \quad (35)$$

This leads to the subsequent subsection of the Dirichlet family that is representative of the whole, as follows:

$$\begin{aligned} f_{\mathfrak{D}D} &:= \{\mathfrak{D}D(e^{\delta_1} \gamma_1, \dots, e^{\delta_{K-1}} \gamma_{K-1}; e^{\delta_1} \gamma_1^+, \dots, e^{\delta_{K-1}} \gamma_{K-1}^+) : \\ &\delta_k \in i, k = 1, \dots, K\}. \end{aligned} \quad (36)$$

Considering that the number of parameters in Equation (34) is one and the number of parameters in Equation (36) is then  $K-1$ . Furthermore, every  $f \in f_{\mathfrak{D}D}$  maintains the following:

$$\begin{aligned} E_{\theta_k} &= \frac{\tilde{\gamma}_k}{\tilde{\gamma}_k + \tilde{\gamma}_k^+} \prod_{j=1}^{k-1} \frac{\tilde{\gamma}_j}{\tilde{\gamma}_j + \tilde{\gamma}_j^+} = \frac{e^{\delta_k} \gamma_k}{e^{\delta_k} \gamma_k + e^{\delta_k} \gamma_k^+} \\ &\times \prod_{j=1}^{k-1} \frac{e^{\delta_j} \gamma_j}{e^{\delta_j} \gamma_j + e^{\delta_j} \gamma_j^+} \\ &= \frac{\gamma_k}{\gamma_1 + \gamma_1^+} = \frac{\gamma_k}{\sum_{j=1}^K \gamma_j}, \end{aligned} \quad (37)$$

where  $\forall k = 1, \dots, K$ ; because  $f_D \subset f_{\mathfrak{D}D}$ , it also conforms to  $f \in f_D$ . Therefore, the circulation of  $q(\theta)$  is associated with both families' circulations in Equations (34) and (36). Finally, noticing that both parameterizations are the same when  $K=2$ , and a Dirichlet circulation that has the potential to be representative of the whole converts to a Dirichlet that is not

generalizable. To make the best use of Equation (31), the subsequent suppositional estimation algorithm was applied.

### 2.3.4 | Suppositional estimation model

Considering that the unbiased function in Equation (32) can be assessed for any assumed assessment of  $\delta$ . Assuming that  $\theta_m, m = 1, \dots, M$  is an unselective instance of  $f(\cdot, \delta)$ . Consequently:

$$\begin{aligned} \bar{L}_M(\theta, \delta) &:= \frac{1}{M} \sum_{m=1}^M (\log p(\mathbf{x} | \theta_m) + \log p(\theta_m) \\ &\quad - \log f(\theta_m; \delta)). \end{aligned} \quad (38)$$

1. Equation (38) is an unbiased estimator of  $L(\delta)$ . The suppositional estimation algorithms in [30] serve as a remedy to the problem of improving a function that could be perceived under the mere existence of noise. Assuming that both 1 and  $K-1$  conform to the number of parameters, that is, signified by  $d = \text{card}(\delta)$ .
2. Setting  $t = 0$  and providing particular original standards  $\delta(t)$ .
3. In terms of  $t = 1, 2, \dots$

a. Determining a slope estimation  $\bar{\lambda} = \bar{\lambda}(\theta, \delta^{(t-1)})$  of  $\Delta L$  as follows:

- (i) Feature  $\mathbf{b} = (b_1, \dots, b_d), b_j \sim DU_{\{-1,1\}}$  autonomously for  $j = 1, \dots, d$ .
- (ii) Feature  $\theta_m : f(\cdot; \delta^{(t-1)}), m = 1, \dots, M$  and calculating

$$L^+ := \bar{L}_M(\theta, \delta^{(t-1)} + c_{t-1} \mathbf{b})$$

- (iii) Feature  $\theta_m \sim f(\cdot; \delta^{(t-1)}), m = 1, \dots, M$  and calculating

$$L^- := \bar{L}_M(\theta, \delta^{(t-1)} - c_{t-1} \mathbf{b})$$

- (iv) Setting  $\bar{\lambda}(\theta, \delta^{(t-1)}) = \frac{L^+ - L^-}{2c_{t-1} \mathbf{b}}$

b. Renewing the parameters  $\delta^{(t)} = \delta^{(t-1)} + a_{t-1} \bar{\lambda}$

The Spall conditions [30–32] certify that a difficult combination of  $\delta^{(t)}$  and the best value is  $t \rightarrow \infty$ . The conditions are satisfied with the addition of orders  $\{a_t\}, \{c_t\}, t = 1, 2$ . In the following, the Spall [32] set is provided:

$$a_t = \frac{a}{(t + A)^a}, \text{ and } c_t = \frac{c}{t^\gamma}, t = 1, 2, \dots \quad (39)$$

To calculate the slope  $\nabla L$  by using a Monte Carlo-type estimation for specific  $a, A, c, \gamma > 0$ , Phases of 2(a).i and 2(a).iv suggest the “concurrent disagreement” method of [32]. By utilizing this technique in comparison with a standard finite difference approach (2 vs  $2d$ ), the entire fundamentals of the parameter course  $\delta$  change concurrently. Additionally, in each iteration,

the number of objective function evaluations decrease remarkably. “DU” denotes the distinct unchanging circulations in the range  $\{-1,1\}$ . The fraction in phase 2(a).iv is representative of the nominator being divided by each element of the vector with a slight change in values.

Based on [33], limited sample training of suppositional estimation algorithms cannot be done simply by establishing simple stopping criteria. During our presentation, a favourable achievement was possible by using the following discovery process. The moving average of the parameter standards is calculated for every  $s$  repetitions, as follows:

$$\bar{\delta}^{(\tau)} = \frac{1}{s} \sum_{t=\tau-s+1}^{\tau} \delta^{(t)}, \text{ where } \tau = s, 2s, 3s, 4s, \dots \quad (40)$$

Then, by considering a predetermined progressive digit  $v$ , it could be denoted as follows:

$$\bar{L}^{(\tau)} := \bar{L}_q(\theta, \bar{\delta}^{(\tau)}), \text{ for } \tau = vs, (v+1)s, (v+2)s, \dots \quad (41)$$

The approximation of the unbiased function assessed at  $\bar{\delta}^{(\tau)}$  is as follows:

$$\begin{aligned} V^{(\tau)} &:= \bar{L}^{(\tau-V+j)}, j = 1, \dots, v\}, \\ \text{for } \tau &= vs, (v+1)s, (v+2)s, \dots \end{aligned} \quad (42)$$

Finally, representing the group of the previous  $v$  assessments as follows:

$$R^{(\tau)} := \{\text{sign } V_{j+1}^{(\tau)} - V_j^{(\tau)}, j = 2, \dots, v. \quad (43)$$

To terminate the algorithm, the number of executions at sign vector  $R(\tau)$  should be equal to  $v-1$  at iteration  $\tau \in \{vs, (v+1)s, (v+2)s, \dots\}$ .

## 2.4 | The proposed adjustable MkR method

The risk of losing information during data transformation could be avoided by adding an input into the solar power forecasts within a specified time. This is feasible by adding high-resolution (i.e. captured every 5 s) solar power data, which reproduces certain variation data. From Section 2.1, to repair high-resolution solar power information, it should be processed via a maximum mean discrepancy (MMD) based kernel function. Hence, the consideration of progressed data within a specified time can be performed by the use of a kernel function that is based on radiance. The recent summarized information loss could be moderated by establishing a multikernel regression model.

Nevertheless, modelling the observed complex solar power prediction error is not feasible via the Gaussian error circulations assumption (a counterpart of least-squares loss). To address this problem, it is assumed that the error term of the

suggested MkR model is handled by a GMM for which the quantity of constituents is mechanically designated using the stick breaking structure method.

Consequently, the recommended adjustable rough MkR model, in addition to avoiding information loss, has the advantage of addressing the discrepancy between the assumption of error circulation used in the prediction models and the error circulation of the real solar power prediction. In addition, a differential Bayesian method is a method by which adaptive robust multikernel regression could be optimized and deterministic and probabilistic forecasts could be generated concurrently.

### 3 | ASSESSMENT BENCHMARKS FOR PREDICTION PERFORMANCE

This section introduces the assessment benchmarks for both deterministic and probabilistic predictions.

#### 3.1 | Assessment benchmarks for deterministic predictions

The deterministic prediction performance of various models is based on three benchmarks including standardized mean total error (SMTE), standardized root mean square error (SRMSE), and an improved version of the mean total percentage error (IMTPE) [34,35].

#### 3.2 | Assessment benchmarks for probabilistic predictions

Assessing the constructed prediction intervals (PIs) can be performed by applying the following five criteria to evaluate various models' probabilistic forecasts. The most prominent criterion is PI coverage probability (PICP), which is defined as:

$$PICP = \frac{1}{L} \sum_{l=1}^L I_l, \text{ and } I_l = \begin{cases} 1, & \text{if } R_p^l \in [L_p^l, U_p^l] \\ 0, & \text{otherwise} \end{cases}, \quad (44)$$

where  $L_p^l$  represents the lower boundary and  $U_p^l$  represents the upper boundary of  $l^{\text{th}}$  PI. A higher rate of PICP indicates that existent solar power data are situated in the built PIs. However, assuming a PI nominal confidence level (PINC), higher calculated values of PICP may be inferior [36]. The correlational relationship between smaller digressions of the PINC and PICP and the improved PI structure is positive. This digression is designated by the middling coverage error (MCE) precisely, as follows:

$$ACE = PICP - PINC. \quad (45)$$

Additionally, PI quality could be assessed based on the width criterion. For a given PICP, the PI conveys more useful information with a narrower width. Advantageous decisions

cannot be made with broad PIs. The PI normalized average width (PINAW) is normally applied to the PI breadth [35] as follows:

$$PINAW = \frac{1}{L(Max_p - Min_p)} \sum_{l=1}^L (U_p^l - L_p^l), \quad (46)$$

where  $Max_p$  and  $Min_p$  denote the extreme and least of the prediction objectives, respectively. Simultaneous consideration of both the cover probability and the breadth is essential in weighing the practicality of dissimilar PIs. The cover width-based yardstick (CWY) is a mixed yardstick that is recommended to assess the excellence of the built PIs systematically. The CWY is illustrated as follows:

$$CWC = PINAW [1 + I(PICP)e^{-\rho(PICP-\nu)}]. \quad (47)$$

Consider that at time  $l$ ,  $L$  is the length of the test set,  $C$  signifies the fixed volume and  $R_p^l$  and  $F_p^l$  designate the real solar power and the prognosticated solar power, respectively.

The nominal confidence level is signified via  $\nu$ . In this study,  $\rho$  is adopted to train unacceptable PIs, and it is set to 50 in this investigation. The  $I(PICP)$  is presented as follows:

$$I(PICP) = \begin{cases} 1, & \text{if } PICP < \nu \\ 0, & \text{otherwise} \end{cases}. \quad (48)$$

The quality of the solar energy PIs in terms of sharpness is evaluated by the Winkler score and it can be calculated as follows:

$$WS = \frac{1}{L} \sum_{l=1}^L S_l. \quad (49)$$

In which

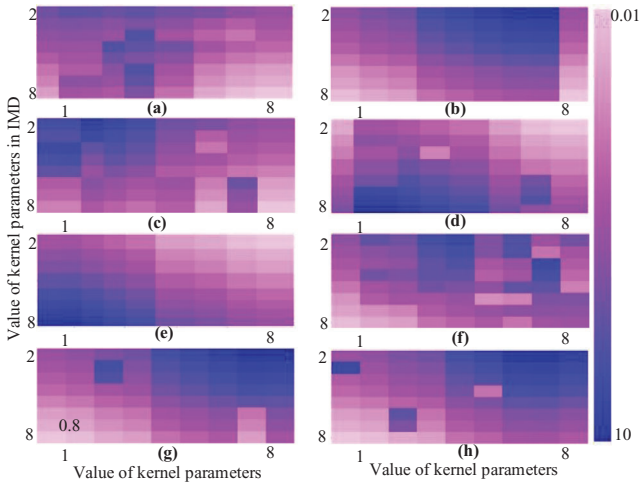
$$S_i = \begin{cases} -2(1 - PINC)W_p^l - 4(L_p^l - R_p^l), & \text{if } R_p^l < L_p^l \\ -2(1 - PINC)W_p^l, & \text{if } R_p^l \in W_p^l \\ -2(1 - PINC)W_p^l - 4(R_p^l - U_p^l), & \text{if } R_p^l > U_p^l \end{cases}, \quad (50)$$

where  $W_p^l = U_p^l - L_p^l$  signifies the breadth of the  $i^{\text{th}}$  built PI and  $R_p^l$  designates the existing solar power at time  $l$ . Larger WS values designate comparatively higher values in the built PIs.

## 4 | CASE STUDIES

This section confirms the predominance of the recommended approach for prediction of hourly global horizontal radiation through studying four locations in China's Tibet zone. The comparison of the prediction accuracy of the global horizontal radiation contains three models.

As mentioned in Section 2, the parameters of the proposed model (i.e.  $a_0$ ,  $b_0$ ,  $c_0$ ,  $d_0$ ,  $e_0$  and  $f_0$ ) should be adjusted optimally

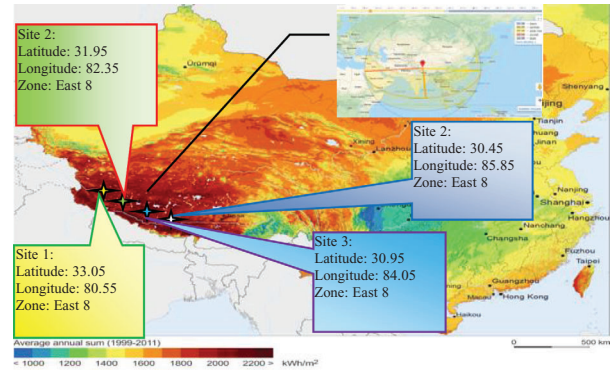


**FIGURE 4** Optimal values of the proposed MkR parameters for the MMAPE error

to obtain the best output from the prediction process. In this work, the partial autocorrelation function (PACF) is used to find the potential inputs of the proposed MkR model. Because the optimal order of the solar signal is three, we can use the proposed model to forecast the previous half hour’s documented five-second solar power output. The optimal values of the proposed MkR parameters for the MMAPE error criteria are presented in Figure 4. The optimal values for the parameters of the proposed model are tuned based on the minimum value of the forecasting error by considering the validation set (Figure 4). In other words, the kernel parameters are optimized by the validation set based forecasting process of the recommended model. The proposed model is formed after determining the input. The proposed model attaches high-resolution information to forecast solar energy. The values of the MMAPE that are being applied to the offered validation set based model are shown in Figure 4. The smallest number of errors in the validation set based prediction by the recommended model achieves optimal parameters.

### 4.1 | Data collection

The four selected locations in China under the labels of Place 1, Place 2, Place 3 and Place 4 are situated in a Tibet self-governing district [37]. Figure 5 indicates the details (i.e. parallels, meridians, and time region). Additionally, it could be inferred from this figure that Tibet is a region in China with richer global horizontal radiation. Therefore, studying the four designated places with diverse parallels (i.e. 4326.99 m, 4737.80 m, 4603.35 m and 4849.95 m) is a substantial effort. Eight groups of mean hourly data files are devoted to each site. These groups are comprised of global horizontal irradiance (W/m<sup>2</sup>), pressure (mbar), rain (cm), hotness (degree centigrade), comparative moisture percent (%), solar peak angle (degree), wind acceleration m/s and wind direction (unit of measurement: degree). These data are obtained from the National Solar Radiation Database (NSRDB) [38]. The study lasted for 365 days (from 1 January to 31 Decem-



**FIGURE 5** Tibet self-governing district test cases in China

ber, 2014) and a 24-h period was considered each day for the solar elevation above 0. The global horizontal radiation hours ranged from approximately 10 to 14 per day because the duration of sunshine varies between seasons. Prediction of global horizontal radiation by using these seven relative meteorological factors is the central goal of this study. By regarding the last seven days of each month as the test data file, the remaining days were treated as the training data. Specifically, 84 days (24-h ahead) were predicted by use of the other 281 days. Each season, one day was selected at random to be prognosticated by three randomly selected days (training data). Generally, 12 days were used to predict one day. To show the effectiveness of the suggested approach in this paper (see Table 2), the last one week of each month are considered as a test dataset, and the remaining days are used for training data while, the Dataset is  $(x,y) = \{(x(t),y_t|t = 1,2,\dots,281)\}$ , Training-set will be  $\{(x(t), y_t | t = 1, 2, \dots, 281)\}$ , and Forecasting test =  $\{(x(t),y_t|t = 282, 283,\dots,365)\}$ . For the pre-processing model we used the Maximum likelihood principal component analysis (PCA) [39] based data imputation. In this model high variance assigns to the missing values prior to PCA allowing to fit PCA model by disregarding the missing points which allows PCA in presence of missing values.

## 5 | NUMERICAL RESULTS AND DISCUSSION

### 5.1 | Deterministic solar energy forecasting

Deterministic solar power prediction is achieved by the qualified recommended model using the MkR. The results are presented in Table 3. In this figure, support vector machine (SVM), least squares SVM (LSSVM), persistence model, autoregressive integrated moving average (ARIMA), quantile regression (QR), wavelet transform plus SVM (WT+SVM), and particle swarm optimization plus ANFIS (PSO+ANFIS) models have been considered to be well-known prediction strategies to compare with the proposed model. Although the running time of proposed algorithm may be higher than of some models, accurate outcome of this model can cover this

**TABLE 2** Training/forecasting sets of proposed test case

	Training sets of 281 days	Prediction sets of 84 days	
One-year data	First 24 days (January)	Last one week (January)	Using 281 days to predict 84 days
	First 21 days (February)	Last one week (February)	
	First 24 days (March)	Last one week (March)	
	First 23 days (April)	Last one week (April)	
	First 24 days (May)	Last one week (May)	
	First 23 days (June)	Last one week (June)	
	First 24 days (July)	Last one week (July)	
	First 24 days (August)	Last one week (August)	
	First 23 days (September)	Last one week (September)	
	First 24 days (October)	Last one week (October)	
	First 23 days (November)	Last one week (November)	
	First 24 days (December)	Last one week (December)	
	Training sets (12 days)	Forecasting sets (1 days)	
Different seasons data	January: 1, 8, 15	January 22	Using 12 days to predict 1 day
	April: 9, 18, 29	May 26	
	August: 1, 14, 27	August 14	
	October: 12, 19, 23	October 25	

**TABLE 3** Result of a deterministic comparison

Prediction Approaches	Site 1			Site 2			Site 3			Site 4		
	NMAE	NRMSE	MMAPE	NMAE	NRMSE	MMAPE	NMAE	NRMSE	MMAPE	NMAE	NRMSE	MMAPE
SVM	6.34	9.65	12.65	6.68	9.99	12.99	5.58	9.11	12.11	7.11	10.21	13.42
ARIMA	5.98	9.06	12.16	6.32	9.93	13.03	5.22	8.52	11.62	6.75	10.15	13.9
Persistence	5.65	8.94	11.76	6.3	9.59	12.63	4.99	8.28	10.78	6.73	9.81	13.5
QR	5.24	8.67	11.34	5.89	9.32	12.21	4.58	8.01	10.36	6.32	9.75	13.08
PSO+ANFIS	4.98	8.55	11.02	5.14	9.14	11.98	4.14	7.65	9.54	6.07	9.34	12.47
WT+SVM	4.35	8.12	10.32	4.75	8.97	11.63	3.72	7.23	9.32	5.57	8.93	12.22
LSSVM	4.18	8.07	10.45	4.83	8.72	11.32	3.52	7.41	9.47	4.96	8.85	12.19
Proposed	<b>2.13</b>	<b>4.47</b>	<b>7.24</b>	<b>2.47</b>	<b>4.81</b>	<b>7.58</b>	<b>1.37</b>	<b>3.71</b>	<b>6.26</b>	<b>2.6</b>	<b>4.94</b>	<b>7.71</b>

point in different error criteria and sites. In this comparison, three well-known error criteria have been considered; the proposed model could provide lower values for all criteria at all solar sites.

The results of several models' prediction in terms of the three error indices, that is, NMAE Equation (51), NRMSE Equation (52) and MMAPE Equation (53) are as follows:

$$\frac{1}{LC} \sum_{l=1}^L \left| R_p^l - F_p^l \right| \times 100\%, \quad (51)$$

$$\sqrt{\frac{1}{L} \sum_{l=1}^L \left( \frac{R_p^l - F_p^l}{C} \right)^2} \times 100\%, \quad (52)$$

$$\sum_{l=1}^L \left( \frac{R_p^l - F_p^l}{\sum_{l=1}^L R_p^l} \right) \times 100\%. \quad (53)$$

It could be inferred that among the compared five prediction models, the LSSVM's achievement was the best and the SVM's performance was the worst. However, the proposed model could provide better results in comparison with LSSVM, where  $(4.18 - 2.13)/4.18 = 49\%$  improvement can be considered for NMAE error at Site 1. The improvement at this site for NRMSE and MMAPE are  $(8.07 - 4.47)/8.07 = 47\%$  and  $(10.45 - 7.24)/10.45 = 31\%$ , respectively. At the second site, the values of improvement in comparison with LSSVM (as the best compared model) are 49%, 45%, and 33% for NMAE, NRMSE and MMAPE, respectively. Furthermore, these types of improvement for the third and fourth sites are: 61%, 50% and 34% for Site 3 and 47%, 44%, and 37% for Site 4. In the proposed model, the prediction error is directed by AGM, not an alternative of a lone Gaussian circulation. The predominance of the prediction model over the others is due to the following reasons. First, the AGM has a much longer tail in comparison with a lone Gaussian circulation. This causes the proposed model to be richer than the other standard models. Second, as the AGMs are outstandingly correct for every

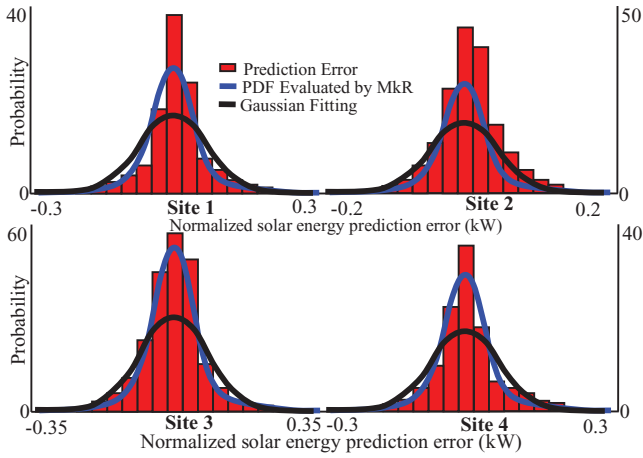


FIGURE 6 Existing solar energy prediction error and related PDF evaluated by the proposed MkR

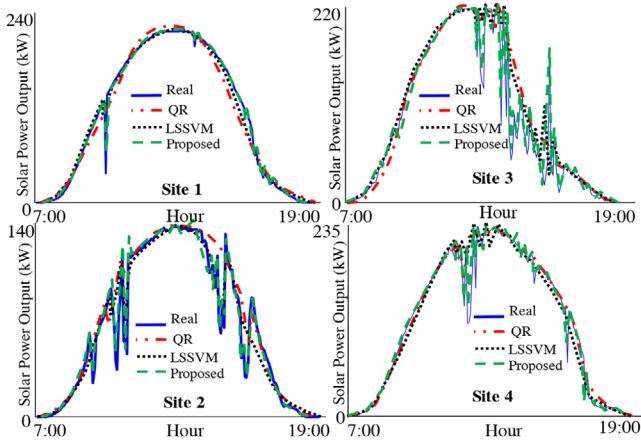


FIGURE 7 Day-ahead prediction in comparison with other models for May 26

constant circulation, the solar power prediction error of the AGM models was estimated more accurately. Figure 6 is representative of the real solar power prediction error and the conforming PDF values from the MkR model. Furthermore, the deterministic prediction model in comparison with LSSVM and QR are presented in Figure 7 for all sites on May 26. As shown in this figure, the proposed method could fit the real solar signal in comparison with the other models and obtain better accuracy by following the signal’s volatility.

### 5.1.1 | Step-by-step implementation of proposed model

In this section, the proposed forecasting approach is applied step-by-step to show the effects of proposed improvements and novelties in forecasting approach. For this purpose, the proposed model is applied on site 1, through NMAE, NRMSE and MMAPE criteria. As shown in Table 4, the suggested model is tested by classic version of MMD, classic GM and classic VBM.

TABLE 4 Result of a improvements comparison

Prediction Approaches	Site 1		
	NMAE	NRMSE	MMAPE
Proposed + classic MMD and GM	4.98	6.95	9.07
Proposed + classic GM and VBM	3.77	6.63	8.78
Proposed with classic MMD	3.07	5.72	8.24
Proposed with classic GM	2.76	5.63	7.93
Proposed with classic VBM	2.49	5.38	7.72
Proposed with FR Smirnov	3.12	5.87	8.87
Proposed	<b>2.13</b>	<b>4.47</b>	<b>7.24</b>

As presented in this paper, it is clear that simple MMD and VBM cannot support suitable and accurate prediction outcome. In addition, the classic versions trap in local point and provide big value in different error criteria. Also, to show the efficiency of proposed approach, the IMD model is compared with multivariate Kolmogorov–Smirnov test (FR Smirnov). In this model, accurate setting of proposed model’s parameter is a critical point to get best prediction results. Therefore, fine setting of existing parameters can be the main point of proposed approach.

As shown in this table, the effects of each improvement have been indicated step-by-step while, the improvement of MMD, GM and VBM are 44%, 29%, and 17% respectively, in comparison with classic versions for NMAE criteria.

Additionally, the improvement of proposed method in comparison with FR Smirnov and improved IMD model is about 46%, 31% and 22% for NMA, NRMSE AND MMAPE, respectively, which proof the validity and superiority of suggested approach.

## 5.2 | Probabilistic solar energy forecasting

In this section, the proposed forecasting method is applied to the mentioned test case to obtain the probabilistic outcomes. According to the probability density figure (Figure 7), the proposed AGM approach is better than the other single Gaussian methods.

Consequently, deriving built PIs from the AGM may result in more precise data than deriving them from a single Gaussian prediction circulation. The PIs built via the proposed MkR approach are evaluated against the PIs produced by other well-known models (i.e. persistence, QR, and ARIMA) to obtain the proposed MkR efficiency. The outcomes of the other three prediction approaches in comparison with the proposed model for different PINCs (90%, 96%, and 99%) are depicted in Figures 8 and 9. The PIs made by the proposed MkR model under dissimilar PINCs are shown in Figure 10. As shown in Figures 8 and 9, the PIs with the highest coverage probability and smallest gap according to the ACE standards are produced by the suggested MkR approach under all PINCs. Note that the applied submission of the built PIs is vitally affected by high coverage of the PIs. Regarding the values of PINAW, the PIs

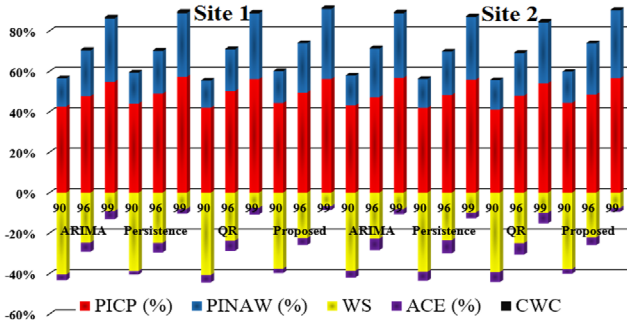


FIGURE 8 Results of probabilistic comparison (Sites 1 and 2)

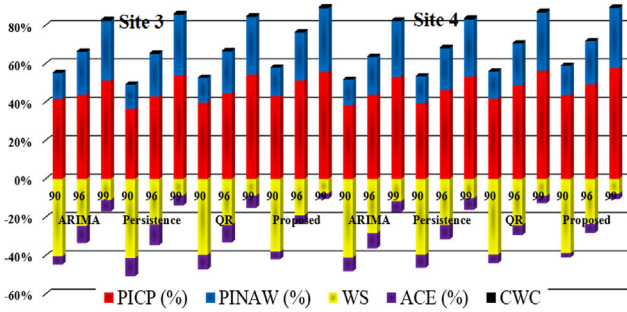


FIGURE 9 Results of deterministic comparison (Sites 3 and 4)

produced by the proposed MkR are wide, while the PIs produced by ARIMA are the narrowest among the models. As a result, being restricted to merely three pointers (i.e. PICP, ACE, and PINAW) makes determining the best probabilistic prediction approach a challenging task. The CWC index is the best choice for evaluating the constructed PIs since it taps into both coverage probability and the PI width, which must be considered in the evaluation. We can stand by the CWC value information in Figures 8 and 9 and the Winkler score criterion to prove the recommended model outperformed the other models. In summary, if the problems of the PI size are removed, the recommended MkR model is the best in terms of performance among other models.

The recommended model is of large width because AGM has a longer tail in comparison with the single Gaussian distribution. As a result, possessing a large PINC leads to the width increasing. If the PINC is low, then there will be a distribution with a sharp peak, which leads to the PIs having a narrower width.

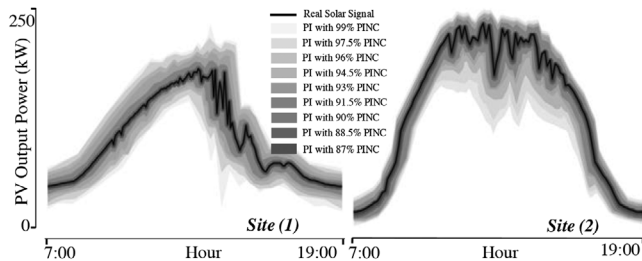


FIGURE 10 Results of the PIs based on the proposed MkR under different PINCs

### 5.3 | Diebold–Mariano test

This test model is suggested by Diebold and Mariano [40] which is described as:

- Let  $\{y_t\}$  indicates the real data series. And  $\{\hat{y}_{i,t}^b\}$  describes the  $i^{\text{th}}$  competing  $b$ -step prediction signal.
- Assume the prediction errors from the  $i^{\text{th}}$  competing approach are,  $e_{i,t}^b = (i, 1, 2, 3, \dots, m)$  while  $m$  shows the number of prediction methods. So,  $b$ -step prediction errors  $e_{i,t}^b$  is:

$$e_{i,t}^b = y_t^b - \hat{y}_{i,t}^b \quad (i = 1, 2, 3, \dots, m). \quad (54)$$

And the accuracy of prediction can be evaluated by:

$$L(y_t^b, \hat{y}_{i,t}^b) = L(e_{i,t}^b). \quad (55)$$

The value of  $b$  is considered as 1 in this work, and the superscript  $b$  is deleted in the following context. In this paper, the following squared-error loss is considered:

$$L_2(y_t, \hat{y}_{i,t}) = L_2(e_{i,t}) = \sum_{t=1}^T (e_{i,t})^2. \quad (56)$$

And absolute-error loss function:

$$L_1(y_t, \hat{y}_{i,t}) = L_1(e_{i,t}) = \sum_{t=1}^T |e_{i,t}|. \quad (57)$$

In these equations, the squared-error loss one penalizes the larger errors more severely. To compare the better outcome based prediction errors, that is, A and B models, the equal accuracy hypothesis is tested based on null hypothesis as:

$$H_0 : E[L(e_{1,t})] = E[L(e_{2,t})]. \quad (58)$$

While the alternative hypothesis can be defined as:

$$H_1 : E[L(e_{1,t})] \neq E[L(e_{2,t})]. \quad (59)$$

So, Diebold–Mariano test is described based on the loss differentials  $d_t$ :

$$d_t = L(e_{1,t}) - L(e_{2,t}). \quad (60)$$

Also, if the null hypothesis is defined as  $H_0: E[d_t] = 0$ . So, the sample mean loss differential can be presented as:

$$\bar{d} = \frac{1}{T} \sum_{t=1}^T d_t = \frac{1}{T} \sum_{t=1}^T [L(e_{1,t}) - L(e_{2,t})]. \quad (61)$$

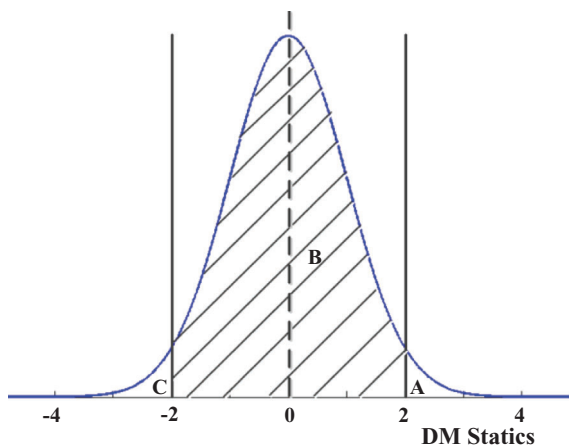


FIGURE 11 Normal distribution of DM

The statistic test of DM is:

$$DM = \frac{\bar{d}}{\sqrt{\frac{2\pi \hat{f}_d(0)}{T}}} \xrightarrow{d} N(0, 1). \tag{62}$$

In which  $2\pi \hat{f}_d(0)$  is a consistent predictor of the asymptotic variance of  $\sqrt{T}\bar{d}$ . The variance is considered in the statistic due to the sample of loss differentials  $d_t$  are serially correlated for  $b > 1$ . Accordingly, DM statistics converge to a normal distribution; the null hypothesis can be rejected if  $|DM| > 1.96$  at the 5% level. The mentioned status correlated to the zone A and zone C in Figure 11. Else, if  $|DM| \leq 1.96$ , the null hypothesis  $H^0$  cannot be rejected, that is shown in zone B of Figure 11.

### 5.3.1 | DM test evaluation

In this section the proposed forecasting approach is compared with other two closest models in deterministic and probabilistic models. Obtained results of this comparison are presented in Table 5. In this comparison, the zero hypothesis means that the obtained performance differences among the two compared models is insignificant, and alternative hypothesis indicates significant differences in this comparison.

As shown in this table, test results in all comparisons are  $H_1$  which means significant differences between proposed method and other approaches are higher than 1.96. So, by this comparison we can claim that the forecasting accuracy of the suggested model is better than other methods.

## 6 | CONCLUSION

In this paper, a new model for the prediction of solar energy in both deterministic and probabilistic cases is presented. The recommended approach is highly capable of fitting an extensive series of unceasing compound circulations. The effectiveness of the proposed model was tested in different solar

TABLE 5 Obtained results of DM test

Comparison	Criteriaa	Deterministic	
		Test value	Test result
Proposed versus SVM	AE	-3.1460	$H_1$
	SE	3.4214	$H_1$
Proposed versus ARIMA	AE	-2.9832	$H_1$
	SE	3.2658	$H_1$
Proposed versus persistence	AE	2.8745	$H_1$
	SE	-3.1745	$H_1$
Proposed versus QR	AE	-2.6284	$H_1$
	SE	-2.9875	$H_1$
Proposed versus PSO+ANFIS	AE	2.3485	$H_1$
	SE	2.7562	$H_1$
Proposed versus WT+SVM	AE	-2.1468	$H_1$
	SE	2.5627	$H_1$
Proposed versus LSSVM	AE	-2.0156	$H_1$
	SE	2.4750	$H_1$

<sup>a</sup>AE indicates the absolute-error loss; SE indicates squared-error loss.

photovoltaic power plants at four sites. The proposed approach was compared with different well-known forecasting models of both deterministic and probabilistic types. For this problem, three deterministic error criteria and five probabilistic criteria have been considered to provide accurate prediction report as numerical analysis outcome. In all obtained results, the proposed model could outperform other models. Additionally, different PI assessment benchmarks under different PINCs have been considered in the probabilistic model. All results demonstrated the superiority and validity of the proposed approach in comparison with the other models.

## PERMISSION TO REPRODUCE MATERIALS FROM OTHER SOURCES

None.

### ORCID

Oveis Abedinia  <https://orcid.org/0000-0001-9327-8738>  
 Mehdi Bagheri  <https://orcid.org/0000-0002-8078-9173>

### REFERENCES

1. Abedinia, O., et al.: Solar energy forecasting based on hybrid neural network and improved metaheuristic algorithm. *Comput. Intell.* 34(1), 241–260 (2018)
2. Ren, Y., et al.: Ensemble methods for wind and solar power forecasting—A state-of-the-art review. *Renew and Sustain. Energy Reviews* 50, 82–91 (2015)
3. Perez, R., et al.: Short term irradiance variability: Preliminary estimation of station pair correlation as a function of distance. *Sol. Energy* 86(8), 2170–2176 (2012)
4. Perez, R., et al.: Parameterization of site-specific short-term irradiance variability. *Sol. Energy* 85(7), 1343–1353 (2011)



5. Chu, Y., et al.: Short-term reforecasting of power output from a 48 MWe solar PV plant. *Sol. Energy* 112, 68–77 (2015)
6. Golestaneh, F., et al.: Very short-term nonparametric probabilistic forecasting of renewable energy generation—With application to solar energy. *IEEE Trans. Power Syst.* 31(5), 3850–3863 (2016)
7. Khodayar, M., et al.: Convolutional Graph Autoencoder: A Generative Deep Neural Network for Probabilistic Spatio-temporal Solar Irradiance Forecasting. *IEEE Trans. Sustainable Energy* 11(2), 571–583 (2019)
8. Zhang, Y., Jianxue, W.: GEFCom2014 probabilistic solar power forecasting based on k-nearest neighbor and kernel density estimator.” In: 2015 IEEE Power & Energy Society General Meeting, pp. 1–5. IEEE, Piscataway (2015)
9. Iversen, E.B., et al.: Probabilistic forecasts of solar irradiance using stochastic differential equations. *Environmetrics* 25(3), 152–164 (2014)
10. Bessa, R.J., et al.: Probabilistic solar power forecasting in smart grids using distributed information. *Int. Journal of Elect. Power & Energy Syst.* 72, 16–23 (2015)
11. Ranjan, R., Gneiting, T.: Combining probability forecasts. *J. R. Stat. Soc.: Series B (Statistical Methodology)* 72(1), 71–91 (2010)
12. Mellit, A., et al.: A simplified model for generating sequences of global solar radiation data for isolated sites: Using artificial neural network and a library of Markov transition matrices approach. *Sol. Energy* 79(5), 469–482 (2005)
13. Feng, C., Zhang, J.: Hourly-similarity based solar forecasting using multi-model machine learning blending. In: 2018 IEEE Power & Energy Society General Meeting (PESGM), pp. 1–5. IEEE, Piscataway (2018)
14. Mohammadi, K., et al.: Horizontal global solar radiation estimation using hybrid SVM-firefly and SVMwavelet algorithms: A case study. *Nat. Hazards* 1–18 (2015b)
15. Hussain, S., Al-Alili, A.: A new approach for model validation in solar radiation using wavelet, phase and frequency coherence analysis. *Appl. Energy* 164, 639–649 (2016a) <https://doi.org/10.1016/j.apenergy.2015.12.038>
16. Wu, J., Chan, C.K.: Prediction of hourly solar radiation with multi-model framework. *Energy Convers. Manag.* 76, 347–355 (2013)
17. Benmouiza, K., Chekneane, A.: Forecasting hourly global solar radiation using hybrid k-means and nonlinear autoregressive neural network models. *Energy Convers. Manag.* 75, 561–569 (2013)
18. Azimi, R., Ghayekhloo, M., Ghofrani, M.: A hybrid method based on a new clustering technique and multilayer perceptron neural networks for hourly solar radiation forecasting. *Energy Convers. Manag.* 118, 331–344 (2016) <https://doi.org/10.1016/j.enconman.2016.04.009>
19. Ghayekhloo, M., et al.: A novel clustering approach for short-term solar radiation forecasting. *Sol. Energy* 122, 1371–1383 (2015) <https://doi.org/10.1016/j.solener.2015.10.053>
20. Sun, S., et al.: A decomposition-clustering-ensemble learning approach for solar radiation forecasting. *Sol. Energy* 163, 189–199 (2018) <https://doi.org/10.1016/j.solener.2018.02.006>
21. Olatomiwa, L., et al.: A support vector machinefirefly algorithm-based model for global solar radiation prediction. *Sol. Energy* 115, 632–644 (2015) <https://doi.org/10.1016/j.solener.2015.03.015>
22. Halabi, L.M., et al.: Performance evaluation of hybrid adaptive neuro-fuzzy inference system models for predicting monthly global solar radiation. *Appl. Energy* 213, 247–261 (2018) <https://doi.org/10.1016/j.apenergy.2018.01.035>
23. Sun, H., et al.: Empirical investigation on modeling solar radiation series with ARMAeGARCH models. *Energy Convers. Manag.* 92, 385e395 (2015) <https://doi.org/10.1016/j.enconman.2014.12.072>
24. Benmouiza, K., Chekneane, A.: Small-scale solar radiation forecasting using ARMA and nonlinear autoregressive neural network models. *Theor. Appl. Climatol.* 124(3e4), 945–958 (2016) <https://doi.org/10.1007/s00704-015-1469-z>
25. Voyant, C., et al.: Numerical weather prediction (NWP) and hybrid ARMA/ANN model to predict global radiation. *Energy* 39(1), 341–355 (2012) <https://doi.org/10.1016/j.energy.2012.01.006>
26. Gretton, A., et al.: A kernel two-sample test. *JMLR* 13(Mar), 723–773 (2012)
27. Hong, T., et al.: Probabilistic energy forecasting: Global Energy Forecasting Competition 2014 and beyond. *Int. J. Forecasting* 32(3), 896–913 (2016)
28. Glaus, P., Honkela, A., Rattray, M.: Identifying differentially expressed transcripts from RNA-Seq data with biological variation. *Bioinformatics* 28, 1721–1728 (2012)
29. McGrory, C., Titterton, D.: Variational approximations in Bayesian model selection for finite mixture distributions. *Comput. Stat. Data Anal.* 51, 5352–5367 (2007)
30. Spall, J.: Multivariate stochastic approximation using a simultaneous perturbation gradient approximation. *IEEE Trans. Autom. Control* 37, 332–341 (1992)
31. McGrory, C.A.: Variational approximations in Bayesian model selection. Doctoral dissertation, University of Glasgow (2005)
32. Spall, J.: Adaptive stochastic approximation by the simultaneous perturbation method. *IEEE Trans. Autom. Control* 45, 1839–1853 (2000)
33. Hutchison, D., Spall, J.: Stochastic approximation in finite samples using surrogate processes. In: Proceedings of 43rd IEEE Conference on Decision and Control. IEEE, Piscataway (2004)
34. Abedinia, O., Amjady, N., Zareipour, H.: A new feature selection technique for load and price forecast of electrical power systems. *IEEE Trans. Power Syst.* 32(1), 62–74 (2016)
35. Abedinia, O., Raisz, D., Amjady, N.: Effective prediction model for Hungarian small-scale solar power output. *IET Renewable Power Gener.* 11(13), 1648–1658 (2017)
36. Pinson, P., Kariniotakis, G.: Conditional prediction intervals of wind power generation. *IEEE Trans. Power Syst.* 29(3), 1166–74 (2014)
37. Solar zoom in China. Available from: <https://www.solarzoom.com/thread-246224-1-1.html>
38. National Solar Radiation Database (NSRDB) in National Renewable Energy Laboratory (NREL). <https://nsrdb.nrel.gov/internationaldatasets>. Accessed 2013
39. Folch-Fortuny, A., Artega, F., Ferrer, A.: Assessment of maximum likelihood PCA missing data imputation. *J. Chemom.* 30(7), 386–393 (2016)
40. Diebold, F.X.: *Element of Forecasting*, 4 ed., Cincinnati, OH, USA: Thomson South-Western, pp. 257–287 (2007)

**How to cite this article:** Abedinia, O., Bagheri, M., Agelidis, V.G.: Application of an adaptive Bayesian-based model for probabilistic and deterministic PV forecasting. *IET Renew. Power Gener.* 15, 2699–2714 (2021). <https://doi.org/10.1049/rpg2.12194>