Research Article

Interval Prediction of Photovoltaic Power Using Improved NARX Network and Density Peak Clustering Based on Kernel Mahalanobis Distance

Wen-He Chen,1 Long-Sheng Cheng,1 Zhi-Peng Chang,2,3 Han-Ting Zhou,1 Qi-Feng Yao,1 Zhai-Ming Peng,1 Li-Qun Fu,1 and Zong-Xiang Chen1,4

1School of Economics & Management, Nanjing University of Science & Technology, Nanjing, China
2School of Business, Anhui University of Technology, Maanshan, China
3Key Laboratory of Multidisciplinary Management and Control of Complex Systems of Anhui Higher Education Institutes, Anhui University of Technology, Maanshan, China
4School of Electrical and Information Engineering, Anhui University of Technology, Maanshan, China

Correspondence should be addressed to Long-Sheng Cheng; cheng_longsheng@163.com and Zong-Xiang Chen; chenzongxiang@ahut.edu.cn

Received 8 December 2021; Revised 23 February 2022; Accepted 24 February 2022; Published 15 March 2022

Academic Editor: Giacomo Fiumara

Copyright © 2022 Wen-He Chen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Photovoltaic (PV) power forecasting can provide strong support for the safe operation of the power system. Existing forecasting methods are ineffective for grid scheduling decisions or risk analysis. The novel multicluster interval prediction method is proposed to consider the volatility and randomness of PV power output. First, this method utilizes the sparse autoencoder (SAE) and Bayesian regularized NARX network (BRNARX) for point forecasting of PV power. Second, density peak clustering improved by kernel Mahalanobis distance (KMDDPC) is applied to classify the dataset into multiple clusters, including forecasting error and meteorological factors. Finally, the joint probability density is established by multivariate kernel density estimation (MKDE) to accomplish the PV power interval prediction. The proposed hybrid method is applied for the interval prediction of PV power at Yulara, Australia. Comparative research of point forecasting is implemented to evaluate the machine learning and deep learning methods, with the proposed SAE-BRNARX under four different periods. Results shows that the average values of nRMSE, MRE, nMAE, and $R^2$ for the four periods are 4.45%, 0.90%, −0.15%, 3.39%, and 95.93%, respectively. Moreover, the results of interval prediction obtained by the other interval prediction approaches are compared with the proposed KMDDPC-MKDE. It shows that the average values of PICP, PINAW, ACE, and nMPICD for four periods are 93.93%, 9.50%, 3.93%, and 7.10% at 90% confidence level, respectively. Outcomes demonstrate that the proposed method can obtain more accuracy, a higher coverage rate, narrower average bandwidth, and a closer distance between the middle of interval and actual value than other methods.

1. Introduction

As an alternative to traditional fossil fuels, PV power is a clean and economical renewable energy. It has become a critical factor in reshaping the power industry through profound technological changes, significantly promoting energy sustainability [1]. The global cumulative installed capacity of PV power in 2020 reached 714 GW and increased by 21.6% over 2019. PV power is easily affected by weather factors, such as GHI, AT, and WS, so it is highly non-scheduled and random [2]. With the large-scale grid-connected PV power, the randomness of PV power output will significantly increase the uncertainty of power grid dispatching and threaten the safe operation of the power system. Therefore, comprehensive and accurate PV power prediction is a necessary tool to reduce the uncertainty of PV power that can provide a reliable basis for power grid dispatching and improve the stability and economic
efficiency of the power system [3]. It is of great significance to enhance the competitiveness of PV power, ensure the safety of the power system, and optimize the power grid operation.

To establish an accurate PV power forecasting model, many scholars have carried out a research series. The prediction model mainly includes point forecast and probability forecast [4]. For the point prediction model, the forecasting methods can be classified into four types: (a) persistence, (b) statistical, (c) machine learning, and (d) hybrid method. The primary methods are ARMA, regression, exponential smoothing, ANN, SVM, etc [2]. Scholars proposed a hybrid method with NARX and a multilayer neural network to avoid the limitations of a single method and improve the performance of the forecasting results. This neural network is a recurrent neural architecture dynamic system with feedback connections enclosing multilayers of the network. The NARX network can evaluate the nonlinear relationship between input and output, predicting the future output time series value [5]. The NARX network has faster convergence and generalizes better than other neural network methods, widely used to solve the nonlinear problem for forecasting the PV power [6, 7]. Meanwhile, the optimized parameter methods are widely applied to improve the forecasting accuracy of the model. The DEPSO-based forecasting model was applied to predict the output of a building-integrated PV system [8]. VanDeventer et al. proposed the SVM optimized by GA for short-term PV power forecasting [9]. Although the above methods can effectively improve the accuracy of PV power forecasting, the forecasting error cannot be avoided due to the inherent randomness and uncertainty of the meteorological system. From the perspective of decision-makers, quantitative analysis of uncertainty can ensure supply reliability with minimized operating costs.

In recent years, interval forecasting, an effective form of a probabilistic forecasting method, has been widely applied to effectively quantify the uncertainty of PV power generation [10]. Interval prediction consists of the maximum and minimum values of point forecasting results based on a data distribution with a certain confidence level, which can provide accurate intervals of PV power fluctuations [11]. Compared with parameter estimation methods affected by specific data distribution assumptions, such as fuzzy inference, Beta distribution function, and Gaussian process, KDE is a popular method to estimate data distribution without prior assumption of datasets [12]. In [13], the distribution of forecasting error was estimated based on the seasonal model and KDE to realize the interval prediction of PV power output. The hybrid models combined with ELM and KDE were established for interval forecasting PV power under weather classification [14]. A prediction intervals estimation method was proposed for solar generation based on GRU neural networks and KDE [15]. Meanwhile, the factors related to the forecasting error have a significant impact on the density function. This article uses the MKDE method to obtain the complete probability density curve of PV power error with meteorological features. Moreover, the upper and lower limits are determined according to their confidence intervals, which is helpful to improve the performance of the interval prediction.

However, the above methods only provide point forecasts and PIs, which are difficult to completely deal with the volatility and uncertainty of PV power output. Some effective clustering methods are introduced into forecasting methods to improve interval prediction accuracy, such as similar days, weather, or seasons [16]. These methods mainly cluster the historical meteorological data, which is hard to ultimately reflect the impact of volatility and uncertainty of the weather on PV power output. Different weather conditions are the significant reason for the error between the forecasting value and the actual value of PV power. Compared with other cluster methods, such as fuzzy c-means, K-means, and hierarchical clustering, DPCA is a novel clustering method. This algorithm can automatically find the cluster center and realize the efficient clustering of arbitrary shape data [17]. It has been widely used in fault detection, image feature, and target monitoring [18–20]. In this article, the dataset containing the errors and weather factors is applied to the cluster using DPCA improved by KMD, which helps to improve the effectiveness of interval prediction.

In this paper, a novel interval prediction model is established for PV power. SAE-BRNARX is applied for point forecasting. We introduce the KDPCA and MKDE to estimate the error distribution. Moreover, we calculate the PIs under different confidence levels. This hybrid model has been validated by using real-world PV power datasets. This article has four main contributions: (1) the point forecasting model is established in this paper, combining with SAE, BR, and NARX network. This method can extract features, select appropriate network parameters to reduce data redundancy and overfitting, and avoid falling into the locally optimal solution. (2) Due to the uncertainty of intraday weather changes, this paper uses 5-minute level data for clustering. According to the difference of forecasting errors and meteorological factors in different periods, the data are clustered into multiple categories to enhance the effectiveness of interval prediction. (3) We utilize KMD to replace Euclidean distance to construct the KPCA, improving the reliability of clustering. KMD can map data from low-dimensional space to high-dimensional space to realize the linear separability of data. (4) MKDE is used to calculate the joint probability density of prediction error and meteorological factors under each cluster to measure upper and lower limits of forecasting intervals under different confidence levels. Compared with other methods, MKDE can better describe the data with unknown distribution to obtain a narrower prediction interval width and higher interval coverage.

The rest of the paper is organized as follows. Section 2 introduces the principles and process of the SAE-BRNARX forecasting model in detail. In Section 3, the KDPCA is proposed for clustering data. Section 4 establishes the uncertainty analysis of the forecasting model. In Section 5, the implementation procedure of the PV power interval forecasting model is given in detail. Section 6 presents the evaluation metrics of point and interval forecasting in detail. Section 7 discusses the interval forecasting results through
2. SAE-BR NARX Point Forecasting Model

Due to the problems of overfitting and falling into the locally optimal solution in forecasting, the model combined with the BR NARX network and SAE is proposed to forecast the PV power output.

2.1. Bayesian Regularized NARX Network. NARX is a dynamic RNN model for time series forecasting. It includes the input layer, hidden layer, output layer, and input-output delay structure. NARX has a memory applied to establish a mapping between inputs and output. Compared with the feedforward network, the NARX model containing exogenous input information has higher degrees of freedom and can require fewer parameters to participate in the calculation. It has more accuracy than other artificial neural networks in describing the characteristics of dynamic time-varying systems, and the formula is as follows:

\[
\hat{y}_t = f(z_t; \theta) + e_t,
\]

\[
z_t = [y_{t-1}, y_{t-2}, \ldots, y_{t-n_y}, x_{t-1}, x_{t-2}, \ldots, x_{t-n_x}],
\]

where \(\hat{y}_t\) is the prediction value of \(y_t\); \(f(\cdot; \theta)\) is the nonlinear mapping function between inputs and output; \(\theta\) is the network parameter to be measured; \(y = [y_1, y_2, \ldots, y_n]^T\) is the observed historical value; \(x = [x_1, x_2, \ldots, x_n]^T\) is the input of the NARX; \(e_t\) is error term and assumes that the mean is 0 and the variance is \(\sigma^2\); \(n_e\) and \(n_x\) are the delay numbers of outputs and inputs, respectively. In this model, time sequences are embedded using two methods: 1st-TDLs, short-term memory filters, and 2nd-feedback signals from the output layer.

The multilayer perceptron is applied to construct the NARX network as a feedforward neural network, reflecting the mapping relationship between inputs and output. NARX network uses feedforward to transmit information through multilayer neurons and optimizes its connection weight to improve the global approximation. It can express complex nonlinear characteristics accurately, such as hysteresis, saturation, and chaos. Therefore, according to equation (1), the NARX network with multiple inputs and single outputs can be expressed as

\[
\hat{y}_t = g_2 \left( \sum_{i=1}^{n_h} w^h_{ji} g_1 \left( \sum_{i=1}^{n_h} w^h_{ji} \phi_{hi} + b^h_j \right) + b^h \right),
\]

where \(n_i\) and \(n_h\) are the numbers of neurons in the input layer and hidden layer of neural network; \(w^h_{ji}\) and \(w^h_{ji}\) are the connection weights between the \(i_{th}\) neuron of input and output and the \(h_{th}\) hidden layer neuron, respectively; \(b^h\) and \(b^h\) are the thresholds of the \(h_{th}\) neuron of hidden layer and output layer, respectively; \(g_1(\cdot)\) and \(g_2(\cdot)\) are the activation functions of the hidden layer and the output layer, respectively; network parameter \(\theta\) denotes the composition of connection weights \(w\) and thresholds \(b\).

BR is used to optimize its parameter \(\theta\) to improve the accuracy network and avoid falling into locally optimal solution in forecasting PV power by NARX network. BR is the method to reduce the negative impact of considerable weight on training results. It can reduce the overfitting and error of the model to adapt to the complexity of the network structure, and the optimal performance function is as follows:

\[
S(\theta) = \beta E_D + \alpha E_\theta,
\]

where \(E_D = \sum_{k=1}^{n_D} (\hat{y}_k - y_k)^2/2 = \sum_{k=1}^{n_D} (\hat{y}_k - y_k)\) is the forecasting ability, \(n_D\) is the number of training data samples; \(E_\theta = \sum_{i=1}^{n_\theta} \theta_i^2/2\) is the regularized penalty term, and \(n_\theta\) is the number of parameters, which can control network parameters to reduce structural redundancy; \(\alpha\) and \(\beta\) is the regularization parameters.

Bayesian theory sets the network parameters as random variables to obtain the posterior probability distribution used to determine the optimal network parameters, as follows:

\[
p(\theta|D, \alpha, \beta) = \frac{p(D|\theta, \beta)p(\theta|\alpha)}{p(D|\alpha, \beta)},
\]

\[
= \frac{\exp(-S(\theta))}{\int \exp(-S(\theta))d\theta}
\]

Therefore, the posterior distribution of regularization parameters of the NARX network is as follows:

\[
p(\alpha, \beta|D) = \frac{p(D|\alpha, \beta)p(\alpha, \beta)}{p(D)}
\]

The optimal regularization parameter based on Gaussian distribution is as follows:

\[
\alpha^* = \frac{\bar{y}}{2E_\theta},
\]

\[
\beta^* = \frac{n_D - \bar{y}}{2E_D},
\]

where \(\bar{y} = n_D - 2\alpha^*\text{trace}((\nabla\nabla S(\theta^*))^{-1})\) is the number of effective network parameters by Bayesian regularization learning; \(E_\theta^*\) and \(E_D^*\) are \(E_\theta\) and \(E_D\) calculated using the optimized network parameter \(\theta^*\).

2.2. Sparse Autoencoder. The SAE neural network is an unsupervised learning algorithm. It uses the back-propagation method to set the outputs equal to the inputs. The SAE tries to realize an approximation function that makes the output is similar to the input. Compared with the autoencoder, the SAE has the sparsity constraint that limits the average activation value of hidden layer neurons. It consists of two layers of neural networks, encoder, and decoder, to reconstruct the input data. The input data \(x\) is encoded using SAE, which is designed to obtain a better representation: \(z = f(W^{(1)}x + b^{(1)})\). Then, the reconstructed data can be expressed as: \(x' = f(W^{(2)}z + b^{(2)})\), where, \(W\) is the weight matrix, and \(b\) is
the deviation matrix. A sparse penalty is added to the loss function of SAE, which can obtain the optimal network weight parameters and avoid overfitting. The objective function is as follows:

$$L = \sum_{n=1}^{N} \| x^{(n)} - x^{*(n)} \|^2 + \eta \rho(Z) + \lambda \| W \|^2,$$

(7)

where $Z = [x^{(1)}, \ldots, x^{(N)}]$ is the encoding of the sample set; $\rho(Z)$ is the sparse metric function, expressed as

$$\rho(Z) = \sum_{j=1}^{p} KL(\rho^* || \tilde{\rho}_j),$$

(8)

where, $\tilde{\rho}_j = 1/N \sum_{n=1}^{N} z_j^{(n)}$ is the average activation value of the $j_{th}$ neuron in the hidden layer, which is similar to the activation probability of the $j_{th}$ neuron. Kullback-Leibler (KL) divergence is used to calculate the heterogeneity between $\tilde{\rho}_j$ and the given value $\rho^*$, which is as follows:

$$KL(\rho^* || \tilde{\rho}_j) = \rho^* \log \frac{\rho^*}{\tilde{\rho}_j} + (1 - \rho^*) \log \frac{1 - \rho^*}{1 - \tilde{\rho}_j},$$

(9)

where, $\rho_j = \rho^*$, $KL(\rho^* || \tilde{\rho}_j) = 0$.

2.3. Establishment of the SAE-BRNARX Forecasting Model.

The SAE is proposed to extract useful features. This neural network can significantly reduce the redundancy of data structure, making the extracted features more useful for recognition and forecasting. Then, the BRNARX network applies the extracted data for forecasting the PV power output, which can reduce overfitting to improve forecasting accuracy. The system model combined with SAE and BRNARX network is shown in Figure 1. The input features are meteorological factors like GHI, WS, AT, and previous PV power output, and the target is the current PV power output.

The forecasting process has the following steps:

(1) The min-max normalization is applied to normalize the data of inputs and targets. Assume an input vector $X = x_1, x_2, \ldots, x_n$, $i = 1, 2, \ldots, n$. The min-max normalization is formula as follows:

$$x_{nor} = \frac{x_i - X_{min}}{X_{max} - X_{min}}.$$

(10)

(2) The normalized input data is used to train the SAE feature extractor. The input features are encoded using this trained SAE to extract features. The output of SAE is the encoded features from normalized data.

(3) The encoded features are applied as input for training the BRNARX network. The data is divided into three parts, 60% for training, 15% for validation, and 25% for testing.

(4) The PV power output is predicted for the testing set.

(5) The forecasting values of PV power are denormalized to get actual values, and its accuracy is compared with other forecasting models.

3. Density Peak Clustering Algorithm

Improved by Kernel Mahalanobis Distance

PV power output will be affected by meteorological factors. Under different weather conditions, the fluctuation range of PV power output is different and uncertain. At the same time, considering the nonlinear relationship between factors, the KMDDPC algorithm is employed to cluster the samples, including meteorological factors and error between the actual value and forecasting value of PV power. The sample points with similar characteristics are clustered into one cluster for improving the accuracy of interval prediction.

3.1. Density Peak Clustering Algorithm.

DPCA is a clustering method based on density peak. Its primary characteristic is that the density of the cluster center is higher than the adjacent density, and it is far away from the low-density points. Two significant parameters are local density $\rho_i$ and its minimum distance to other high-density points $\delta_i$, respectively, which characterize the point $x_i$. The local density $\rho_i$ is defined as

$$\rho_i = \sum_{j=1}^{n} \exp \left(-\frac{d^2_{ij}}{4\bar{d}^2_{c}}\right),$$

(11)

where, $d_{ij}$ is the Euclidean distance between points $i$ and $j$; $d_c$ is the cutoff distance. $\delta_i$ represents the minimum distance between point $x_i$ and other high-density points, expressed as

$$\delta_i = \begin{cases} \min_{j: \rho_j \geq \rho_i} d_{ij}, & \text{if } \exists j s.t. \rho_j > \rho_i, \\ \max d_{ij}, & \text{otherwise}. \end{cases}$$

(12)

The premise for selecting a cluster center is to establish a decision graph. The decision graph is a point distribution map based on two-dimensional coordinates according to the values of $\rho_i$ and $\delta_i$. The points of high $\rho$ and relatively high $\delta$ can be discerned as the cluster centers. To simplify the selection of cluster centers, DPCA computes the decision value $y_i$ of each point $x_i$, which is defined as

$$y_i = \rho_i \times \delta_i.$$  

(13)

In general, the points with higher decision value $y_i$ are selected as the cluster centers. If selecting the cluster center is finished, the remaining points can be assigned to the same cluster according to their nearest neighbor of high-density points. In addition, the points with high $\rho$ and low $\delta$ can be distinguished as outliers, which can be automatically found and excluded in the clustering process.

3.2. Kernel Mahalanobis Distance.

KMD is composed of KPCA and MD. The basic idea is to first project original patterns into a high-dimensional or infinite-dimensional space by an implicit nonlinear mapping $\phi(x)$. Then, the KMD of the sample data is computed in another low-dimensional kernel principal component space through principal component analysis (PCA), as shown in Figure 2.
The derivation of KMD corresponding to the sample data is as follows:

Let \( X = [x_1, x_2, \ldots, x_n]^T \) be a group of data samples, where \( x_i \in \mathbb{R}^L \), \( L \) is the number of features. Let the group \( X \) map to the feature space by using the implicit nonlinear mapping \( \varphi(x) \), expressed as

\[
\varphi(X) = [\varphi(x_1), \varphi(x_2), \ldots, \varphi(x_n)]^T.
\]  

(14)

The sample data is centralized by

\[
\Phi(x_i) = \varphi(x_i) - \frac{1}{n} \sum_{i=1}^{n} \varphi(x_i).
\]

(15)

The sample data is centralised by

\[
\Phi(x_i) = \varphi(x_i) - \frac{1}{n} \sum_{i=1}^{n} \varphi(x_i).
\]

(15)

The covariance matrix can be calculated by

\[
\varphi(X) = [\varphi(x_1), \varphi(x_2), \ldots, \varphi(x_n)]^T.
\]  

(14)
the covariance matrix as follows:

\[ S_\Phi = \frac{1}{n} \Phi(X)^T \Phi(X) \]

\[ = \frac{1}{n} \sum_{i=1}^{n} \Phi(x_i)^T \Phi(x_i)^T. \]  

(16)

Assume \( \lambda \) and \( p \) are the eigenvalues and weight vectors of the covariance matrix \( S_\Phi \), respectively, which is formulated as follows:

\[ S_\Phi p = \lambda p. \]  

(17)

Substitute equations (16) and (17), as follows:

\[ \frac{1}{n} \Phi(X)^T \Phi(X)p = \frac{1}{n} \sum_{i=1}^{n} \Phi(x_i)^T \Phi(x_i)^T p = \lambda p. \]  

(18)

If \( \lambda \neq 0 \),

\[ p = \frac{1}{\lambda n} \sum_{i=1}^{n} \Phi(x_i) [\Phi(x_i)^T p] \]

\[ = \sum_{i=1}^{n} \alpha_i \Phi(x_i) = \Phi(X)^T \alpha, \]

where \( \alpha = [\alpha_1, \alpha_2, \ldots, \alpha_n]^T \), \( \alpha_i = (1/\lambda n) \sum_{i=1}^{n} \Phi(x_i)p, i = 1, 2, \ldots, n. \)

Substitute equation (19) into equation (18), as follows:

\[ \frac{1}{n} \Phi(X)^T \Phi(X) \Phi(X)^T \alpha = \lambda \Phi(X)^T \alpha, \]  

(20)

\[ \frac{1}{n} \Phi(X) \Phi(X)^T \Phi(X) \Phi(X)^T \alpha = \lambda \Phi(X)^T \alpha. \]

Assume \( K = \Phi(X) \Phi(X)^T \), equation (20) can be transformed into

\[ \frac{1}{n} K^2 \alpha = \lambda K \alpha \Rightarrow K \alpha = n \lambda \alpha. \]  

(21)

3.3. Establishment of KMDDPC Clustering Model. If the distance type is not appropriate, DPCA is difficult to distinguish the differences between samples, resulting in poor classification ability. Considering the nonlinear correlation between features, KMD replaces the Euclidean distance in DPCA. KMDDPC is illustrated in Algorithm 1.

4. Uncertainty Analysis of Forecasting Model

The accurate calculation of PV power forecasting uncertainty is of great significance to support power grid dispatching and reduce the rotating reserve capacity of power generation equipment. This section applies the MKDE and confidence intervals to quantify the error distribution of PV power forecasting for uncertainty analysis for interval prediction.

4.1. Multivariate Kernel Density Estimation. The MKDE is a method to analyze the distribution characteristics of multidimensional data. This method does not require prior knowledge of the relevant data distribution, does not attach any assumptions to data distribution, and has significant practical value. Compared with one-dimensional or two-dimensional KDE, MKDE considers the correlation between errors and other meteorological factors to improve uncertainty analysis accuracy.

MKDE is applied to calculate the distribution of the multi-dimensional datasets. Suppose the \( d \)-dimensional dataset \( X_i = [X_{i1}, X_{i2}, \ldots, X_{id}]^T, i = 1, 2, \ldots, n. \) The \( d \)-dimensional vector can be expressed as \( x = [x_1, x_2, \ldots, x_d]^T \), and its joint PDF is defined as \( f(x) = f(x_1, x_2, \ldots, x_d) \), formulated as,

\[ \tilde{f}_n(x) = \frac{1}{n} \sum_{i=1}^{n} K_{h_n}(X_i - x), \]  

(23)

\[ K_{h_n}(u) = h_n^{-d} K(u h_n^{-1}), \]

where, \( K \) and \( K_\theta \) are the unscaled and scaled kernel’s functions, respectively; the smoothing parameter \( h \) is the bandwidth. \( u = (X_i - x)/h \) is the scaled distance. The diagonal bandwidth matrix and multiplicative kernel are applied to reduce the complexity of the calculation. The multiplicative kernel is applied: \( K(x) = k(x_1)k(x_2)\ldots k(x_d) \), where \( k(\cdot) \) is the one-dimensional kernel smoothing function. According to the equation (23), the MKDE is formulated as follows:

\[ \tilde{f}_n(x) = \frac{1}{nh_1 h_2 \ldots h_d} \sum_{i=1}^{n} K \left( \frac{x_1 - X_{i1}}{h_1}, \frac{x_2 - X_{i2}}{h_2}, \ldots, \frac{x_d - X_{id}}{h_d} \right) = \frac{1}{nh_1 h_2 \ldots h_d} \sum_{i=1}^{n} \prod_{j=1}^{d} k \left( \frac{x_j - X_{ij}}{h_j} \right). \]  

(24)
The sample set is represented by \( \mathbf{X} \), and \( x_i \) is the \( i \)th sample of \( \mathbf{X} \), \( i = 1, 2, \ldots, n \). KMD\( n \times n \) denotes the KMD matrix of this sample set. \( \text{kmd}_i \) is the element of the \( i \)th row and \( j \)th column in the KMD matrix, representing the kmd between the \( i \)th sample and \( j \)th sample. \( \delta_i \) is the KMD between point \( i \) and its \( k \) nearest neighbor. \( \mu^k \) is the average mean of \( \delta_i \) in all the points.

1. KMD matrix KMD\(_\text{new} \) is computed
2. for \( i = 1: n \)
3. \( \delta_i^k \), \( \delta_i^\text{new} \) = max kmd\(_i \) is computed
4. end
5. \( \mu^k \) is computed, where \( \mu^k = 1/n \sum_{i=1}^{n} \delta_i \)
6. Cutoff distance \( \text{kmd}_i \), \( \text{kmd}_d = \mu^k + \sqrt{(1/n) \sum_{i=1}^{n} (\delta_i^k - \mu^k)^2} \) is computed
7. for \( i = 1: n \)
8. Local density \( \rho_i \), \( \rho_i = \sum_{j} \exp(-((\text{kmd}_i/\text{kmd}_d))^2) \) is computed
9. end
10. The decision graph is drawn with \( \rho_i \) as the \( x \)-axis and \( \delta_i \) as the \( y \)-axis in the two-dimensional space
11. The points are sorted in ascending order according to the decision value \( y_i \rho_i \times \delta_i \)
12. First \( m \) data points are as cluster centers, and the remaining points will be assigned to the same cluster as their nearest neighbor of high-density points
13. end

**Algorithm 1: KMDDPC.**

The selection of the bandwidth parameter determines the PDF. The AIMSE is introduced to set the bandwidths, and its formula is as follows:

\[
\text{AIMSE}(x) = \frac{1}{4} h^4 \alpha^2 \int \nabla^2 f(x) dx + n^{-1} h^{-2} \beta,
\]

where \( \alpha = \int x^2 K(x) dx \), \( \beta = \int K(x^2) dx \). The optimal bandwidth is as follows:

\[
h_{\text{opt}} = \{ d \beta e^{-2} \left( \int \nabla^2 f(x) dx \right) \}^{1/(d+4)} n^{-1/(d+4)},
\]

To further simplify the calculation of bandwidth, the method of the optimal diagonal bandwidth matrix based on AIMSE removing the minimum is proposed as follows:

\[
h_i = \sigma_i \left( \frac{4}{(d+2)n} \right)^{1/(d+4)},
\]

where \( \sigma_i \) is the standard deviation of \( i \).

4.2. Confidence Interval Based on MKDE. After the joint probability density distribution of the PV power forecasting error is obtained using the MKDE, the CI quantifies the joint probability density distribution.

The PV power forecasting error is the difference between the PV power forecasting value \( P_{\text{fore}} \) and the actual PV power value \( P_{\text{true}} \) at a specified point in time, and its formula is as follows:

\[
e = P_{\text{fore}} - P_{\text{true}}.
\]

Suppose the vector including the errors and meteorological factors \( X = [e, u_1, u_2, \ldots, u_d]^T \), where \( U = [u_1, u_2, \ldots, u_d]^T \) is the vector of meteorological features extracted by SAE. MKDE calculates the joint PDF of vector \( X \), and the CDF of \( X \) is obtained. Its confidence level is calculated as follows:

\[
P(e_{\text{low}} < e < e_{\text{up}}) = 1 - \theta,
\]

where, the interval \([e_{\text{low}}, e_{\text{up}}]\) is supposed as confidence interval under the confidence level of \( 1 - \theta \), the \( e_{\text{low}} \) is the lower limit of the confidence interval, and \( e_{\text{up}} \) is the upper limit. \( P(e_{\text{low}} < e < e_{\text{up}}) \) shows that the probability that \( X \) falls into the interval \([X_{\text{low}}, X_{\text{up}}]\). According to the time points \( X_{\text{low}} \) and \( X_{\text{up}} \) corresponding to the lower and upper limits, point \( e_{\text{low}} \) and \( e_{\text{up}} \) are selected simultaneously. Therefore, the PV power forecasting interval is \([P_{\text{fore}} - |e_{\text{low}}|, P_{\text{fore}} + |e_{\text{up}}|]\). The PIs are as shown in Figure 3.

5. Implementation Procedure of the PV Power Interval Prediction Model

This section proposes a novel PV power interval prediction model that integrates the SAE-BRNARX for point forecasting, the KDP for clustering data, and MKDE for uncertainty analysis and interval prediction. The implementation of the prediction procedure is shown in Figure 4.

The specific calculation process is as follows:

5.1. SAE- BRNARX Point Forecasting Model.

1. The data is preprocessed, including selecting the required data and paradigm, cleaning (removing abnormal data to eliminate the effect of extreme circumstances and filling in missing data during equipment fault), and normalizing the data.
2. To reduce the redundancy and overfitting between indexes, SAE extracts the features by encoding the input data. The encoding data is divided into the training and testing dataset.
3. The training data is used to obtain the parameters of the SAE- BRNARX network for point forecasting of PV power output. Compared with results of other
point forecasting methods, such as BRNARX, NARX, LSTM, BPNN, KELM, and RF, the advancement of SAE-BRNARX is highlighted.

5.2. KMDDPC Clustering Algorithm

(1) The KMDDPC is applied to cluster the training dataset composed of the forecasting error and extracted features in Section 3. Moreover, the specific calculation process is shown in Section 3.3.

5.3. Uncertainty Analysis of PV Power Error

(1) The MKDE is used to calculate the joint distribution
density characteristics of the forecasting error and extracted features under each cluster.

(2) According to the CIs, the upper and lower limits of the joint probability density distribution are calculated to find the point \(x_{\text{low}}\) and \(x_{\text{up}}\) corresponding to the upper and lower limits.

### 5.4. PV Power Interval Forecasting Method

(1) According to the time point \(x_{\text{low}}\) and \(x_{\text{up}}\) of the joint probability density, the errors \(e_{\text{low}}\) and \(e_{\text{up}}\) at this time point are selected to establish the forecasting interval of PV power. PICP, PINAW, ACE, and nMPICD validate the effectiveness of interval forecasting. Compared with the results of other interval prediction methods, a novel model combined with SAE-BRNARX, KMDPC, and MKDE can supply a narrower prediction interval width, higher interval coverage and closer distance between the middle of the PIs and the actual value.

### 6. Evaluation Metrics

#### 6.1. Accuracy of Point Forecasting Results

The nRMSE, nMAPE, MRE, nMAE, and \(R^2\) are used to evaluate the point forecasting accuracy of the PV power prediction model [2], which is calculated as follows:

\[
n\text{RMSE} = \frac{1}{P_{\text{cap}}} \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\bar{y}_i - y_i)^2} \times 100%,
\]

\[
n\text{MAPE} = \frac{1}{nP_{\text{cap}}} \sum_{i=1}^{n} \frac{\bar{y}_i - y_i}{y_i} \times 100%,
\]

\[
\text{MRE} = \frac{1}{nP_{\text{cap}}} \sum_{i=1}^{n} \frac{\bar{y}_i - y_i}{P_{\text{cap}}} \times 100%,
\]

\[
n\text{MAE} = \frac{1}{nP_{\text{cap}}} \sum_{i=1}^{n} |\bar{y}_i - y_i| \times 100%,
\]

\[
R^2 = 1 - \frac{\sum_{i=1}^{n} (\bar{y}_i - y_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2},
\]

where \(n\) is the number of the PV power forecasting points; \(P_{\text{cap}}\) is the total installed capacity of the PV power equipment; \(\bar{y}_i\) and \(y_i\) are the forecasting power and actual power at time point \(i\), respectively.

#### 6.2. Performance of Prediction Intervals

PICP, PINAW, ACE and nMPICD are applied to evaluate the performance of prediction intervals. PICP indicates that the percentage of the testing data between the upper and lower limits, and its formula is defined as follows:

\[
\text{PICP} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{[L_i,U_i]}(P_{\text{true}}),
\]

where \(\mathbf{1}_{[L_i,U_i]}(\cdot)\) is the characteristic function, indicating whether the actual PV power output \(P_{\text{true}}\) lies within the range \([L_i,U_i]\). If \(P_{\text{true}}\) lies, \(\mathbf{1}_{[L_i,U_i]}(P_{\text{true}}) = 1\); otherwise, \(\mathbf{1}_{[L_i,U_i]}(P_{\text{true}}) = 0\); \(L_i\) and \(U_i\) are the minimum and maximum of the forecasting at the \(i_{th}\) moment.

In practice, the reliability requires that the PICP should be less than the confidence level of forecasting intervals, and the width of forecasting intervals should be as small as possible. PINAW is an important metric to evaluate the quality of forecasting intervals. And, its formula is defined as

\[
\text{PINAW} = \frac{1}{nR} \sum_{i=1}^{n} (U_i - L_i),
\]

where \(R\) is the difference between maximum and minimum of forecasting, which is applied to normalize the forecasting interval average width in percentage.

ACE denotes the bias between the PICP and PINC, expressed as

\[
\text{ACE} = \text{PICP} - \text{PINC}.
\]

nMPICD is applied to evaluate the distance between the middle of the PIs and the actual value \(P_{\text{true}}\) defined as

\[
n\text{MPICD} = \frac{\sum_{i=1}^{n} |U_i + L_i/2 - P_{\text{true}}|}{\sum_{i=1}^{n} (U_i - L_i)}.
\]

If both PIs cover a point, the PI value is inversely proportional to its performance.

#### 7. Case Study and Discussion

##### 7.1. Data Collection

The installation site of the PV system selected in this paper is Yulara, located in the south of the Northern Territory of Australia, as shown in Figure 5. This area is located in the desert area, with a significant temperature difference between day and night, low precipitation, high solar radiation, and drought. This means that the region has rich solar energy resources and low weather fluctuation of PV power, which can fulfill the local power demand and export the excess energy to other regions. However, the extreme weather conditions in desert locations also lead to the performance degradation of PV systems in many ways, including the overheating of solar photovoltaic panels caused by intense solar radiation, the accumulated dust and soot effect, and the rapid aging and performance degradation of PV panels due to severe weather such as dust storms. The above conditions challenge the construction of an accurate PV interval prediction model.

The data source is DKASC, and the specific installation and specification information is shown in Table 1 the 5-minute level data of each station is collected in a year, including PV power output and meteorological features. The recorded meteorological data include GH, WS, and AT. The study duration is between 6 a.m. and 7 p.m., from March 2019 to November 2020. According to [21], the dataset is
**Table 1: Specification of the PV system.**

| System specification       | Value         |
|----------------------------|---------------|
| Array rating               | 327.6 kW      |
| Panel rating               | 315 W         |
| PV technology              | Poly-si       |
| Number of panels           | 1040          |
| Panel type                 | Jinko solar JKM315PP |
| Inverter size/type         | SMA STP 25000TL-30 |
| Installation completed     | Thu, 3 Mar 2016 |
| Array tilt/azimuth         | Tilt = 20, azi = 0 (solar north) |

**Table 2: Basic statistics of meteorological variables and power production at each station.**

| Parameter                        | Statistic | Autumn     | Winter     | Spring     | Summer     |
|----------------------------------|-----------|------------|------------|------------|------------|
| Global horizontal irradiance (W/m²) | Maximum  | 1204.75    | 915.47     | 1307.22    | 1352.78    |
|                                   | Minimum   | −1.97      | −3.68      | −3.46      | −3.59      |
|                                   | Mean      | 502.49     | 447.13     | 584.57     | 586.09     |
|                                   | Median    | 538.94     | 488.54     | 621.95     | 597.39     |
|                                   | Standard deviation | 305.60     | 254.91     | 351.35     | 356.98     |
| Wind speed (m/s)                  | Maximum   | 6.62       | 7.08       | 8.91       | 12.12      |
|                                   | Minimum   | 0.11       | 0.14       | 0.14       | 0.14       |
|                                   | Mean      | 2.61       | 2.51       | 2.73       | 2.91       |
|                                   | Median    | 2.56       | 2.42       | 2.64       | 2.84       |
|                                   | Standard deviation | 1.06       | 1.12       | 1.17       | 1.16       |

Figure 5: The location and power output of the PV system (information source: DKASC).
divided into four time periods: Autumn (from March to May), Winter (from June to August), Spring (September to November), and Summer (December to February). Furthermore, the basic statistical information is shown in Table 2. The loss and invalid data points of each meteorological parameter and PV power output are eliminated. Furthermore, the data with excessive fluctuation and irregularity are deleted. The dataset is normalized to an interval of [0, 1]. At the same time, 60% of the data are regarded as training sets, 15% as the validation set, and the remaining 25% are regarded as test sets. SAE is used for feature extraction of meteorological information, as shown in Figure 6. Then, the BRNARX network is applied to predict the PV power output. Figure 7 shows the sample time series plots of all models during the last five days of every period, which illustrates the excellent agreement between forecasting value using the SAE-BRNARX network and actual PV power values. Although the actual power curve fluctuates wildly in spring, the SAE- BRNARX network can also accurately forecast the PV power to a certain degree. At the same time, there are considerable deviations between model estimations and actual PV power, during early/late daylight hours and hours of violent fluctuation in NWP. The proposed models can still adapt to these unexpected changes in the PV power profile. Their performance in this period mainly determines the differences in forecasting accuracies in different models. During the last five days of every period, Figure 8 illustrates the changes of AE and APE for the proposed models, which validate the above views.

### Table 2: Continued.

| Parameter | Statistic | Autumn | Winter | Spring | Summer |
|-----------|-----------|--------|--------|--------|--------|
| Ambient temperature (°C) | Maximum | 44.06 | 31.06 | 42.54 | 46.58 |
| | Minimum | 4.40 | 0.94 | 4.41 | 15.11 |
| | Mean | 26.67 | 17.52 | 28.26 | 33.99 |
| | Median | 26.62 | 18.22 | 29.67 | 34.82 |
| | Standard deviation | 7.95 | 5.77 | 7.03 | 5.96 |
| Power (kW) | Maximum | 291.64 | 268.96 | 296.84 | 296.06 |
| | Minimum | 0.00 | 0.00 | 0.01 | 0.00 |
| | Mean | 148.98 | 145.95 | 149.60 | 141.62 |
| | Median | 166.43 | 164.63 | 163.68 | 146.17 |
| | Standard deviation | 85.50 | 78.62 | 90.17 | 88.40 |
| Normalized power (W/Wp) | Maximum | 1.00 | 1.00 | 1.00 | 1.00 |
| | Minimum | 0.00 | 0.00 | 0.00 | 0.00 |
| | Mean | 0.51 | 0.54 | 0.50 | 0.48 |
| | Median | 0.57 | 0.61 | 0.55 | 0.49 |
| | Standard deviation | 0.29 | 0.29 | 0.30 | 0.30 |

### Table 3: Parameters in the SAE-BRNARX network.

| Method | Main parameter | Autumn | Winter | Spring | Summer |
|--------|----------------|--------|--------|--------|--------|
| SAE | Numbers of neurons | 3 | 3 | 3 | 3 |
| | Encoder transfer function | Stalin | Stalin | Stalin | Stalin |
| | Decoder transfer function | Purelin | Purelin | Purelin | Purelin |
| | L2 weight regularization | 0.10 | 0.10 | 0.10 | 0.10 |
| | Sparsity regularization | 4 | 4 | 4 | 4 |
| | Sparsity proportion | 0.50 | 0.50 | 0.50 | 0.50 |
| | Epochs | 100 | 100 | 100 | 100 |
| | Numbers of neurons | 6 | 15 | 8 | 6 |
| | Input delay | 50 | 44 | 47 | 46 |
| | Feedback delay | 2 | 9 | 7 | 2 |
| | Train function | Trainbr | Trainbr | Trainbr | Trainbr |
| | Epochs | 100 | 100 | 100 | 100 |

7.2. Effect of Point Forecasting. An ordinary personal computer implements the interval prediction scheme with AMD R7-3700K and 16.00 GB of RAM. The data samples are divided into four time periods from March 2019 to February 2020 to test the performance of the proposed model. The SAE-BRNARX network is used for point forecasting of PV power in four time periods, and its parameters are shown in Table 3. Furthermore, the parameters of comparative methods are shown in Table 4.
Table 4: Parameters in the comparative methods.

| Parameters                  | BRNARX         | BPNN         | LSTM         | NARX         | KELM         |
|-----------------------------|----------------|--------------|--------------|--------------|--------------|
|                             | Autumn Winter | Spring Summer|              |              |              |
| Initial hidden units        | 7 14          | 9 10         | 50           | 10           | 4            |
| Input delay                 | 53 48         | 51 49        |              |              |              |
| Feedback delay              | 3 11          | 8 3          |              |              |              |
| Train function              | Trainbr       | Trainbr      | Trainbr      | Trainbr      | Trainbr      |
| Epochs                      | 100 100       | 100          | 100          | 100          |              |
|                             |               |              |              |              |              |
| Parameters                  | Value         | Parameters   | Value        | Parameters   | Value        |
| Initial hidden units        | 25            | 50           |              |              |              |
| Initial learn rate          | 0.01          | 0.0001       |              |              |              |
| Minimum error of the goal   | 500           |              |              |              |              |
| Epochs                      | 500           |              |              |              |              |
| Hidden layer function       | Tansig        |              |              |              |              |
| Output layer function       | Purelin       |              |              |              |              |
| Gradient descent learning function | Trangdx |              |              |              |              |
|                             |               |              |              |              |              |
| Parameters                  | Value         | Parameters   | Value        | Parameters   | Value        |
| Initial hidden units        | 10            |              |              |              |              |
| Input delay                 | 50            |              |              |              |              |
| Feedback delay              | 5             |              |              |              |              |
| Train function              | Trainbr       |              |              |              |              |
| Epochs                      | 100           |              |              |              |              |
|                             |               |              |              |              |              |

Figure 6: Feature extraction of meteorological factors using SAE for four periods.
BRNARX, KELM and RF have similar prediction effectiveness, worse than the LSTM and BANARX networks. BPNN has the worst prediction effectiveness. Thus, the effectiveness of the combined forecasting model is confirmed that proposed in this paper. The main advantage of the NARX network model is considering the simultaneous effects of historical observations and exogenous inputs on the data. SAE and BR are applied to improve the NARX network. The former reduces the redundancy of the externalities input data of the NARX network, and the latter can well solve the overfitting problem and find the local global solution. Thus, the fluctuation range of predicted power is closer to the actual power fluctuation. From Table 6, the numbers of predicted errors (normalized form) in four time periods within $\pm 0.02$, $\pm 0.05$, and $\pm 0.1$ are more than several other methods, respectively. It indicates that the errors of the proposed method tend to the normal distribution.

7.3. Effect of Interval Prediction. KMDDPC is used to cluster the samples, including forecasting errors and meteorological features extracted by SAE. The KMDs between every two samples in the four periods are shown in Figure 9. Figure 10 shows the clustering results when the number of clusters is $4(K = 4)$. The original linear inseparability samples are mapped to high-dimensional space through KMD to realize linear separability and sample clustering. Figure 11 shows the error fluctuation trend and its clustering by sorting the errors according to the original time series. It can be seen that $K = 4$ can avoid overclassification and better describe the distribution characteristics of errors. It does not produce clusters with too few samples due to too many clusters. At the same time, the errors are effectively clustered according to their value and extracted features. The reason is that the category of PV power forecast errors is vital for determining the reserve capacity required for grid
dispatch. From the economic view, it is impossible to configure a large reserve margin for a tiny probability scenario during grid operation and dispatch.

Figure 12 represents the joint cumulative probability distributions of forecasting errors and extracted features calculated by MKDE in different clusters. Figure 13 demonstrates interval prediction of PV power under multiple clusters using MKDE at different confidence levels, such as 95%, 90%, and 80%, in the five last days of each period. To further validate the availability and universality of the proposed model, we calculate the PICP, PINAW, ACE and nMPICD of the four periods at different confidence levels, as shown in Table 7. ACE denotes that the PICP is significantly higher than its confidence level. PINAW shows that the widths of the PI depend on the levels of the CI. The larger prediction intervals have a higher probability of covering the actual values. If the confidence level decreases, the average distance between the middle of the prediction interval and the actual value is closer, as shown in nMPICD. From the evaluation results in four time periods, the performances of the Interval predictions in summer and winter are lower than theirs in the Autumn and Spring.

As shown in Table 8, compared with the K-means, FCM, and nonclustering model, the proposed model combined...
with SAE and KMDDPC can successfully track the variation of PV power output at 90% confidence level. It has a narrower average bandwidth and a closer distance between the middle of PI and actual value. Table 9 also illustrates that the interval prediction using SAE data is better than that using NWP data. In four time periods, the predicted performance of MKDE is greater than other models, such as KDE, Monte Carlo, and Quantile Regression. PICP and ACE of the proposed model are almost higher than other models. Furthermore, PINAW and nMPICD are almost lower than other models. It shows that the multiple clusters determined by MKDE are helpful for interval prediction, which can supply a narrower prediction interval width and higher interval coverage.

Table 6: Statistical errors (normalized form) of different point forecasting models in four periods.

| Methods         | Autumn | Winter |
|-----------------|--------|--------|
|                 | Range of forecasting errors | Range of forecasting errors |
|                 | ±0.02  | ±0.05  | ±0.1  | ±0.02  | ±0.05  | ±0.1  |
| SAE-BRNARX      | 510    | 1235   | 2166  | 393    | 990    | 1874  |
| BRNARX          | 224    | 551    | 1051  | 397    | 905    | 1716  |
| NARX            | 156    | 245    | 400   | 237    | 572    | 1212  |
| BPNN            | 181    | 286    | 590   | 372    | 930    | 1628  |
| LSTM            | 379    | 973    | 1804  | 308    | 803    | 1729  |
| KELM            | 392    | 1005   | 1933  | 391    | 987    | 1873  |
| RF              | 308    | 711    | 1312  | 242    | 596    | 1140  |

Table 9: Statistical errors (normalized form) of different point forecasting models in four periods.

| Methods         | Spring | Summer |
|-----------------|--------|--------|
|                 | Range of forecasting errors | Range of forecasting errors |
|                 | ±0.02  | ±0.05  | ±0.1  | ±0.02  | ±0.05  | ±0.1  |
| SAE-BRNARX      | 382    | 1330   | 1989  | 769    | 1824   | 2801  |
| BRNARX          | 281    | 703    | 1351  | 634    | 1396   | 2203  |
| NARX            | 277    | 725    | 1449  | 322    | 862    | 1835  |
| BPNN            | 168    | 453    | 928   | 336    | 853    | 1579  |
| LSTM            | 368    | 1002   | 1850  | 717    | 1769   | 2693  |
| KELM            | 255    | 709    | 1616  | 548    | 1341   | 2265  |
| RF              | 247    | 640    | 1318  | 571    | 1394   | 2329  |

Figure 9: KMD matrix in four periods.
Figure 10: Classification results using KMDDPC.

Figure 11: Error fluctuation trend and its cluster result according to the original time series.
Figure 12: Joint cumulative probability distributions of errors using MKDE for four periods.

Figure 13: PV power prediction interval four periods.
### Table 7: Interval prediction performance for four periods.

|                | Autumn     | Winter     | Spring     | Summer     |
|----------------|------------|------------|------------|------------|
| **95% CI**     |            |            |            |            |
| PICP (%)       | 95.6       | 98.69      | 94.91      | 97.15      |
| PINAW (%)      | 12.27      | 6.54       | 14.44      | 8.74       |
| ACE (%)        | 0.60       | 3.69       | -0.09      | 2.15       |
| nMPICD (%)     | 6.81       | 2.55       | 11.92      | 8.76       |
| **90% CI**     |            |            |            |            |
| PICP (%)       | 92.1       | 94.94      | 93.16      | 95.5       |
| PINAW (%)      | 11.16      | 5.5        | 13.44      | 7.9        |
| ACE (%)        | 2.10       | 4.94       | 3.16       | 5.50       |
| nMPICD (%)     | 6.12       | 2.47       | 11.56      | 8.26       |
| **80% CI**     |            |            |            |            |
| PICP (%)       | 82.95      | 84.19      | 85.29      | 89.3       |
| PINAW (%)      | 8.81       | 4.01       | 9.31       | 6.44       |
| ACE (%)        | 2.95       | 4.19       | 5.29       | 9.30       |
| nMPICD (%)     | 5.62       | 2.23       | 11.18      | 7.34       |
Table 8: Effectiveness evaluation of interval prediction models under different clustering methods at 90% CI.

|        | SAE + KMDDPC + MKDE | NWP + KMDDPC + MKDE | SAE + MKDE | SAE + DPCA + MKDE | SAE + FCM + MKDE | SAE + K-means + MKDE |
|--------|---------------------|----------------------|------------|-------------------|-----------------|----------------------|
| **Autumn** |                   |                      |            |                   |                 |                      |
| PICP (%) | 92.10              | 91.16                | 91.07      | 92.16             | 87.08           | 92.05                |
| PINAW (%) | 11.16              | 16.58                | 17.79      | 17.95             | 10.80           | 18.81                |
| ACE (%)  | 2.10               | 1.16                 | 1.07       | 2.16              | -2.92           | 2.05                 |
| nMPICD (%) | 6.12               | 6.93                 | 7.83       | 6.36              | 8.67            | 6.49                 |
| **Winter** |                   |                      |            |                   |                 |                      |
| PICP (%) | 94.94              | 91.93                | 91.44      | 89.46             | 83.87           | 87.58                |
| PINAW (%) | 5.50               | 13.61                | 13.24      | 13.55             | 4.88            | 10.25                |
| ACE (%)  | 4.94               | 1.93                 | 1.44       | -0.54             | -6.13           | -2.42                |
| nMPICD (%) | 2.47               | 2.67                 | 3.02       | 4.06              | 10.09           | 3.97                 |
| **Spring** |                   |                      |            |                   |                 |                      |
| PICP (%) | 93.16              | 92.76                | 90.96      | 91.04             | 88.83           | 91.67                |
| PINAW (%) | 13.44              | 22.56                | 17.54      | 16.75             | 9.63            | 21.85                |
| ACE (%)  | 3.16               | 2.76                 | 0.96       | 1.04              | -1.17           | 1.67                 |
| nMPICD (%) | 11.56              | 11.92                | 12.07      | 11.87             | 12.98           | 11.69                |
| **Summer** |                   |                      |            |                   |                 |                      |
| PICP (%) | 95.50              | 93.77                | 94.17      | 94.91             | 94.60           | 94.34                |
| PINAW (%) | 7.90               | 19.70                | 18.50      | 20.79             | 8.13            | 18.13                |
| ACE (%)  | 5.50               | 3.77                 | 4.17       | 4.91              | 4.60            | 4.34                 |
| nMPICD (%) | 8.26               | 10.87                | 9.54       | 8.57              | 9.07            | 10.16                |
8. Conclusion

PV power has considerable fluctuations in different seasons and days, and large errors cannot be avoided in the point forecast. For obtaining comprehensive and practical forecasting information to deal with the uncertainty of PV power generation, a novel interval forecasting method based on SAE-BARNARX and KMDDPC-MKDE is proposed in this paper.

By analyzing the PV power datasets from Yulara in four periods, the conclusions can be summarized as follows: (1) From the results of point prediction using SAE-BRNARX, the average values of nRMSE, nMAPE, MRE, nMAE, and $R^2$ for the four periods are 4.45%, 0.90%, −0.15%, 3.39%, and 95.93%, respectively. The accuracy of prediction is higher than other deep learning and machine learning methods. (2) Compared with other classification methods, KMDDPC can better reflect the nonlinear relationship between features, which is more reasonable to classify the samples into multiple clusters. (3) MKDE can effectively simulate the prediction error distribution, which takes into account the influence of meteorological factors on PV power output. Its joint probability density curves at different confidence levels are generated to measure the uncertainty associated with point forecasting and PIs. (4) From the results of interval prediction using KMDDPC-MKDE, the average values of PICP, PINAW, ACE, and nMPICD for four time periods are 93.93%, 9.50%, 3.93%, and 7.10% at 90% confidence level, respectively. Compared with other methods, this method has narrower bandwidth, higher coverage, and a closer distance between the middle of PI and actual value. It supplies a new way to realize the interval prediction, which is helpful for system reliability assessment of PV power plants and dispatch of the smart grid.

Although the proposed method is evaluated to be effective for interval prediction of PV power, some limitations are still worth further study. The predicted model needs more meteorological factors, such as wind direction and rainfall, to deal with the fluctuation and randomness in the PV power. Some hyperparameters are determined by experience. The model lacks methods for detecting outliers and irrelevant variables. Further research can introduce some methods to enhance its performance. For example, a metaheuristic algorithm should be employed to optimize the BR, for selecting the appropriate parameters to improve the prediction accuracy. The recognition algorithm is applied for the adaptive detection of outliers. Moreover, the relevant variables can be selected to establish joint probability density curves with errors. The PI with a narrower width, higher coverage, and closer distance between the middle of the interval and the actual point can be applied for smart grid operation.

**Abbreviations**

PV: Photovoltaic  
NWP: Numerical weather prediction  
ANN: Artificial neural network  
SVM: Support vector machine  
GA: Genetic algorithm  
NARX: Nonlinear autoregressive exogenous model  
DEPSO: Particle swarm optimization improved by differential evolution  
BR: Bayesian regularization  
SAE: Sparse autoencoder  
PI: Prediction interval  
CI: Confidence interval  
KDE: Kernel density estimation  
GRU: Gated recurrent unit  
MKDE: Multivariable kernel density estimation  
PDC: Density peaks clustering  
KMD: Kernel Mahalanobis distance  
MD: Mahalanobis distance  
PDF: Probability density function  
LSTM: Long short-term memory  
RF: Random forest  
RNN: Recurrent neural network  
nRMSE: Normalized root mean square error  
CDF: Cumulative distribution function  
PDF: Probability density function
nMAPE: Normalized mean absolute percentage error
MRE: Mean relative error
nMAE: Normalized mean absolute error
$R^2$: Decision coefficient
PICP: Predict interval coverage probability
nPINC: Normalized mean predict interval center deviation
PINC: Predict interval nominal confidence
PINAW: Predict interval normalized average width
ACE: Average coverage error
DKASC: Desert knowledge Australia solar center
GHI: Global horizontal irradiance
WS: Wind speed
AT: Ambient temperature
AE: Absolute error
APE: Absolute percentage error
TDL: Tapped delay line
KPCA: Kernel principal component analysis
PCA: Principal component analysis
CDF: Cumulative distribution function
KELM: Kernel based extreme learning machine
BPNN: Back propagation neural network
AIMSE: Asymptotic integrated mean square error.

Data Availability
The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest
The authors declare that they have no conflicts of interest.

Acknowledgments
This work was supported by the National Natural Science Foundation of China (project no. 71673001), the Postgraduate Research and Practice Innovation Program of Jiangsu Province (project no. KYCX21_0356), Key Grant Project of Humanities and Social Science Foundation of the Higher Education Institutions of Anhui Province (project no. SK2021ZD0034), and the Open Fund of Key Laboratory of Anhui Higher Education Institute (project no. CS2020-ZD02).

References
[1] D. Gielen, F. Boshell, D. Saygin, M. D. Bazilian, N. Wagner, and R. Gorini, “The role of renewable energy in the global energy transformation,” Energy Strategy Reviews, vol. 24, pp. 38–50, 2019.
[2] U. K. Das, K. S. Tey, M. Seyedmahmoudian et al., “Forecasting of photovoltaic power generation and model optimization: A review,” Renewable and Sustainable Energy Reviews, vol. 81, pp. 912–928, 2018.
[3] B. R. Prusty and D. Jena, “A critical review on probabilistic load flow studies in uncertainty constrained power systems with photovoltaic generation and a new approach,” Renewable and Sustainable Energy Reviews, vol. 69, pp. 1286–1302, 2017.
[4] J. Antonanzas, N. Osorio, R. Escobar, R. Urraca, F. J. Martinez-de-Piñon, and F. Antonanzas-Torres, “Review of photovoltaic power forecasting,” Solar Energy, vol. 136, pp. 78–111, 2016.
[5] S. Billings, “Nonlinear system identification: NARMAX methods in the time, frequency, and spatio-temporal domains,” p. 607, 2013, https://www.wiley.com/en-cn/Nonlinear+System+Identification%26;+NARMAX+Methods+in+the+Time%26;Frequency%26;+2C%26;+and+Spatio%26;Temporal+Domains-p-9781118535554.
[6] M. A. Hassan, N. Bailek, K. Bouchoucha, and S. C. Nwokolo, “Ultra-short-term exogenous forecasting of photovoltaic power production using genetically optimized non-linear auto-regressive recurrent neural networks,” Renewable Energy, vol. 171, pp. 191–209, 2021.
[7] M. Massaoudi, I. Chihi, L. Sidhom et al., “An effective hybrid NARX-LSTM model for point and interval PV power forecasting,” IEEE Access, vol. 9, pp. 36571–36588, 2021.
[8] M. Seyedmahmoudian, E. Jamei, G. Thirunavukkarasu et al., “Short-term forecasting of the output power of a building-integrated photovoltaic system using a metaheuristic approach,” Energies, vol. 11, no. 5, p. 1260, 2018.
[9] W. VanDeventer, E. Jamei, G. S. Thirunavukkarasu et al., “Short-term PV power forecasting using hybrid GASVM technique,” Renewable Energy, vol. 140, pp. 367–379, 2019.
[10] Y. He and H. Li, “Probability density forecasting of wind power using quantile regression neural network and kernel density estimation,” Energy Conversion and Management, vol. 164, pp. 374–384, 2018.
[11] B. Gu, H. Shen, X. Lei, H. Hu, and X. Liu, “Forecasting and uncertainty analysis of day-ahead photovoltaic power using a novel forecasting method,” Applied Energy, vol. 299, Article ID 117291, 2021.
[12] B. Zhou, X. Ma, Y. Luo, and D. Yang, “Wind power prediction based on LSTM networks and nonparametric kernel density estimation,” IEEE Access, vol. 7, pp. 165279–165292, 2019.
[13] Y. Han, N. Wang, M. Ma, H. Zhou, S. Dai, and H. Zhu, “A PV power interval forecasting based on seasonal model and nonparametric estimation algorithm,” Solar Energy, vol. 184, pp. 515–526, 2019.
[14] J. Guo, X. Xu, W. Lian, and H. L. Zhu, “A new approach for interval forecasting of photovoltaic power based on generalized weather classification,” International Transactions on Electrical Energy Systems, vol. 29, no. 4, Article ID e2802, 2019.
[15] C. Pan, J. Tan, and D. Feng, “Prediction intervals estimation of solar generation based on gated recurrent unit and kernel density estimation,” Neurocomputing, vol. 453, pp. 552–562, 2021.
[16] P. F. Jiménez-Pérez and L. Mora-López, “Modeling and forecasting hourly global solar radiation using clustering and classification techniques,” Solar Energy, vol. 135, pp. 682–691, 2016.
[17] A. Laio and L. Alessandro, “Clustering by fast search and find of density peaks,” Science, vol. 344, no. 6191, pp. 1492–1496, 2014.
[18] C. Zhang, Q. Guo, and Y. Li, “Fault detection method based on principal component difference associated with DPCA,” Journal of Chemometrics, vol. 33, no. 1, pp. e3082, 2019.
[19] F. Zhang, J. Yang, J. Qian, and Y. Xu, “Nuclear norm-based 2-DPCA for extracting features from images,” IEEE Transactions on Neural Networks and Learning Systems, vol. 26, no. 10, pp. 2247–2260, 2015.
[20] Q. Zhang, J. Wang, P. Lin, and L. Hou, “Application of an improved DPCA algorithm in rock-fall monitoring based on
synthetic aperture radar, “Natural Hazards, vol. 94, no. 3, pp. 1043–1055, 2018.

[21] Y. Zhou, N. Zhou, L. Gong, and M. Jiang, “Prediction of photovoltaic power output based on similar day analysis, genetic algorithm and extreme learning machine,” Energy, vol. 204, Article ID 117894, 2020.