Short-term Photovoltaic Power Forecasting Using SOM-based Regional Modelling Methods*

*Jun Li** *and Qibo Liu*

(School of Automation and Electrical Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China)

Abstract: The inherent intermittency and uncertainty of photovoltaic (PV) power generation impede the development of grid-connected PV systems. Accurately forecasting PV output power is an effective way to address this problem. A hybrid forecasting model that combines the clustering of a trained self-organizing map (SOM) network and an optimized kernel extreme learning machine (KELM) method to improve the accuracy of short-term PV power generation forecasting are proposed. First, pure SOM is employed to complete the initial partitions of the training dataset; then the fuzzy *c*-means (FCM) algorithm is used to cluster the trained SOM network and the Davies-Bouldin index (DBI) is utilized to determine the optimal size of clusters, simultaneously. Finally, in each data partition, the clusters are combined with the KELM method optimized by differential evolution algorithm to establish a regional KELM model or combined with multiple linear regression (MR) using least squares to complete coefficient evaluation to establish a regional MR model. The proposed models are applied to one-hour-ahead PV power forecasting instances in three different solar power plants provided by GEFCom2014. Compared with other single global models, the root mean square errors (RMSEs) of the proposed regional KELM model are reduced by 52.06% in plant 1, 54.56% in plant 2, and 51.43% in plant 3 on average. Such results demonstrate that the forecasting accuracy has been significantly improved using the proposed models. In addition, the comparisons between the proposed and existing state-of-the-art forecasting methods presented have demonstrated the superiority of the proposed methods. The forecasts of different methods in different seasons revealed the strong robustness of the proposed method. In four seasons, the MAEs and RMSEs of the proposed SF-KELM are generally the smallest. Moreover, the R^2 value exceeds 0.9, which is the closest to 1.

Keywords: Photovoltaic power generation, forecasting, self-organizing map, regional modeling, extreme learning machine

1 Introduction

In terms of replacing traditional energy sources, solar energy has become one of the most popular solutions owing to the advantages of abundant resources, pollution-free, free use, and no need for transportation [1]. Clearly, due to the polytrope of sunshine and meteorological factors, PV power generation exhibits intermittency, fluctuation, and randomness; thus, it is difficult to control and measure. In most cases, the uncertainties in PV power generation directly affect the safety and stability of grid-connected PV systems; however, an accurate forecast is deemed to be an effective way to address this problem [2-3].

Forecasting PV power generation can be divided based on forecasting horizon, meteorological data pattern, and methods used to forecast [4]. Based on the forecasting horizon, the forecasts can be divided into short-term, medium-term, and long-term. The meteorological data pattern and forecasting methods-based classifications are cross-correlated. Persistence models, statistical models, and artificial intelligence (AI) models are commonly used technologies in PV power forecasting.

In the latest research, with the rapid development of AI technology, machine learning and deep learning models are largely adopted. Typically, different neural network (NN) models have been used to forecast solar irradiance or PV power generation [5-7]. Li et al. [8] proposed a hybrid deep learning model, which combined packet decomposition with a long-short-term memory network and had been successfully applied to hourly PV power forecasting.

Manuscript received November 13, 2021; revised June 15, 2022; accepted June 24, 2022. Date of publication March 31, 2023; date of current version August 22, 2022.

* Corresponding Author, E-mail: lijun691201@mail.lzjtu.cn

* Supported by the National Natural Science Foundation of China (51467008) and Gansu Provincial Department of Education Industry Support Program (2021CYZC-32).

Digital Object Identifier: 10.23919/CJEE.2023.000004

Yang et al. [9] employed a heuristic competitive group optimization algorithm to train the linear or nonlinear parameters of radial basis function (RBF) NN and applied a novel RBFNN model to short-term PV power generation forecasting. Additionally, Vandeventer et al. [10] proposed a genetic algorithm-based support vector machine (GASVM) model for short-term power forecasting. Their experimental results have demonstrated the superiority of the proposed GASVM model compared to conventional SVM models. Seyedmahmoudian et al. [11] proposed a mathematical forecasting model that used the differential evolution and particle swarm optimization (DEPSO) technique to address the accuracy and complexity issues associated with PV power forecasting systems.

In the literature, NN-based forecasting models are commonly combined with other data preprocessing techniques to improve forecasting accuracy. Nevertheless, considered an unsupervised and non-parametric extraordinary NN, SOM can be used for both forecasts and data partitions. More importantly, SOM and its extended models are increasingly employed for time series forecasting, dynamic system identification, and control [12-14].

In general terms, a single global model is simple to operate but may not be able to capture the dynamics of the system with high curvatures and smooth the details. Even worse, the forecasting performance of a single global model will be greatly compromised in cases where the data volume is significant.

To eliminate the defects existing in the global model, Júnior et al. [15] proposed a regional modeling method, which combines the clustering of trained SOM with an extreme learning machine (ELM), and achieved excellent performance. Compared with ELM, the kernel ELM (KELM) method is easy to implement and does not require specifying the number of hidden layer nodes in advance. In terms of recent research, different forecasting models combined with KELM have been proposed and applied to wind and PV power generation and multi-step time series forecasts [16-18].

Considering the successful applications of KELM and the advantages of the regional modeling method, in this investigation, we improved the regional modeling processes. Accordingly, the clustering of SOM is accomplished using an evaluable Two-Level algorithm [19] combined with the optimized KELM method [20]. Specifically, a pure SOM is first employed to separate the input space into multiple prototypes [21], followed by the clustering of trained SOM using the fuzzy *c*-means (FCM) algorithm [22]. In the abovementioned process, the DBI is used to evaluate the cluster [23] and determine the optimal size of clustering simultaneously. A given cluster defines a region in the input space formed by merging the prototypes belonging to that cluster. Finally, for each cluster, only data mapped to the corresponding region is combined with optimized KELM or multiple linear regression (MR) to construct a forecasting model. The proposed models are applied to one-hour-ahead PV power generation forecasting using the data derived from the GEFCom2014 [24].

The remainder of this paper is arranged as follows. In Section 2, the SOM algorithm is first introduced, followed by the description of the Two-level algorithm. The implementation of regional modeling methods combined with KELM or MR is described in detail in Section 3. The introduction and preprocessing of experimental data are first presented in Section 4. Moreover, to verify the effectiveness of the proposed methods, the forecasting results of different models are compared and analyzed. Finally, Section 5 draws conclusions.

2 Two-Level algorithms

The concept and implementation of SOM are different from other contiguity-constrained and agglomerative clustering algorithms. SOM attempts to extract and visually display the topological structure of high-dimensional input data, while clustering divides the input data into multiple groups. In this investigation, data partitions are carried out using the proposed Two-Level algorithm, in which SOM is first adopted for initial partitions; then, the output of trained SOM is clustered.

2.1 The SOM networks

SOM generates a set of prototype vectors representing the dataset and maintains the topology by projecting the data from a multi-dimensional input space $\boldsymbol{X}$ to the low-dimensional grid $\boldsymbol{Y}$ [25]. Fig. 1 depicts the

typical structure of the SOM network.

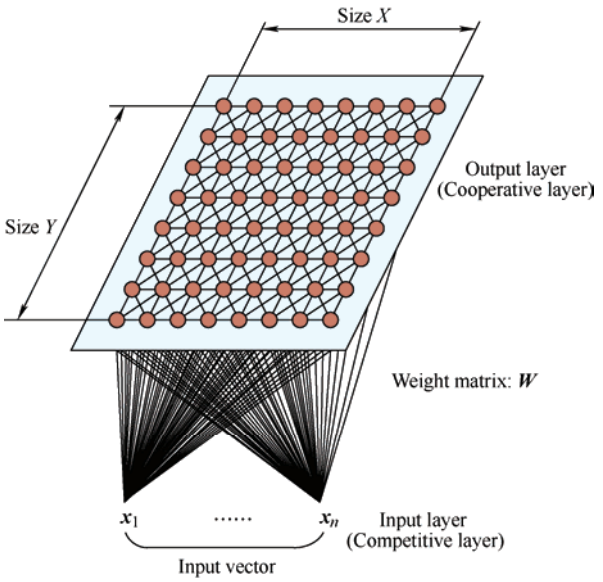


Fig. 1 Typical structure of an SOM network

The SOM algorithm mainly includes two stages, namely, competitive and cooperative. In the first stage, the "winner" neuron is selected. Let $\boldsymbol{W}=\{\boldsymbol{w}_1,\boldsymbol{w}_2,\cdots,\boldsymbol{w}_P\}$, $\boldsymbol{w}_i\in\mathbb{R}^d$ be the weight matrixes of P neurons that are arranged in fixed topological forms, e.g., a hexagonal two-dimensional array. For any input vector $\boldsymbol{x}=[x_1,x_2,\cdots,x_d]\in\mathbb{R}^d\subset\boldsymbol{X}$, a winning neuron $i^*\in\boldsymbol{Y}$ is determined based on the smallest Euclidean distance

$$i^*=\arg\min_i\|\boldsymbol{w}_i-\boldsymbol{x}\| \qquad i=1,2,\cdots,P \tag{1}$$

In the second stage, the weights of the winning neuron and its neighbors are updated. The updating rules are defined as follows

$$\boldsymbol{w}_i(t+1)=\boldsymbol{w}_i(t)+\alpha(t)\cdot\Lambda[i,i^*,\sigma(t)]\cdot[\boldsymbol{x}(t)-\boldsymbol{w}_i(t)] \tag{2}$$

where $0<\alpha(t)<1$ represents the learning rate, $\Lambda[i,i^*,\sigma(t)]$ is the neighborhood function. The update rule of $\Lambda[i,i^*,\sigma(t)]$ is executed by a Gaussian function

$$\Lambda[i,i^*,\sigma(t)]=\exp\left(-\frac{\|\boldsymbol{r}_i(t)-\boldsymbol{r}_{i^*}(t)\|^2}{2\sigma^2(t)}\right) \tag{3}$$

where $\boldsymbol{r}_i(t)$ and $\boldsymbol{r}_{i^*}(t)$ are the coordinates of neurons i and i^* in the output array, respectively, and $\sigma(t)>0$ is the radius of the neighborhood function at step t. The values of $\alpha(t)$ and $\sigma(t)$ should decrease with time to guarantee the convergence of weight vectors. In this study, the updates of $\alpha(t)$ and $\sigma(t)$ use the consider exponential decay method

$$\alpha(t)=\alpha_0\left(\frac{\alpha_{\mathrm{T}}}{\alpha_0}\right)^{(t/\mathrm{T})} \qquad \sigma(t)=\sigma_0\left(\frac{\sigma_{\mathrm{T}}}{\sigma_0}\right)^{(t/\mathrm{T})} \tag{4}$$

where α_0 is the initial learning rate and α_{T} is the final learning rate. Similarly, σ_0 and σ_{T} are the initial and final neighborhood function radii, respectively. The resulting map also preserves the topology of the input samples in the sense that adjacent patterns are mapped into adjacent regions on the map. Owing to this topology-preserving property, SOM can cluster input information and spatial relationships of the data on the map [25].

2.2 Clustering of trained SOM network

Due to the advantages that the FCM algorithm does not depend on previous clustering and implicit assumptions can be made within the form, it is adopted to cluster the output of the trained SOM network.

To implement the Two-Level algorithm, the first step is SOM training, i.e., to obtain the topological mapping units of original data and corresponding Voronoi cells. For each mapping unit, a Voronoi cell, V_i, is defined as

$$V_i=\left\{\boldsymbol{x}\middle|\|\boldsymbol{x}-\boldsymbol{w}_i\|\leqslant\|\boldsymbol{x}-\boldsymbol{w}_j\|,\forall\, j=1,2,\cdots,P,\, j\neq i\right\} \tag{5}$$

where $\boldsymbol{x}$ represents the data vector mapped into V_i. Then, an evaluable FCM is employed to cluster the trained SOM network.

Given a sample set $D=\{\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_N\}$, FCM divides D into a number of clusters C by minimizing the clustering loss function

$$J_f=\sum_{j=1}^{C}\sum_{i=1}^{N}[\mu(\boldsymbol{x}_i)]^b\|\boldsymbol{x}_i-\boldsymbol{m}_j\|^2 \tag{6}$$

where $\boldsymbol{m}_j$ ($j=1,2,\cdots,C$) represents the j-th cluster center, which possesses the same feature dimension as $\boldsymbol{x}$ and $\mu_j(\boldsymbol{x}_i)$ represents the membership function of the i-th sample to the j-th cluster. b is a constant, which can be used to adjust the fuzzy degree of the clustering result. Fig. 2 shows the flowchart of the Two-Level algorithm.

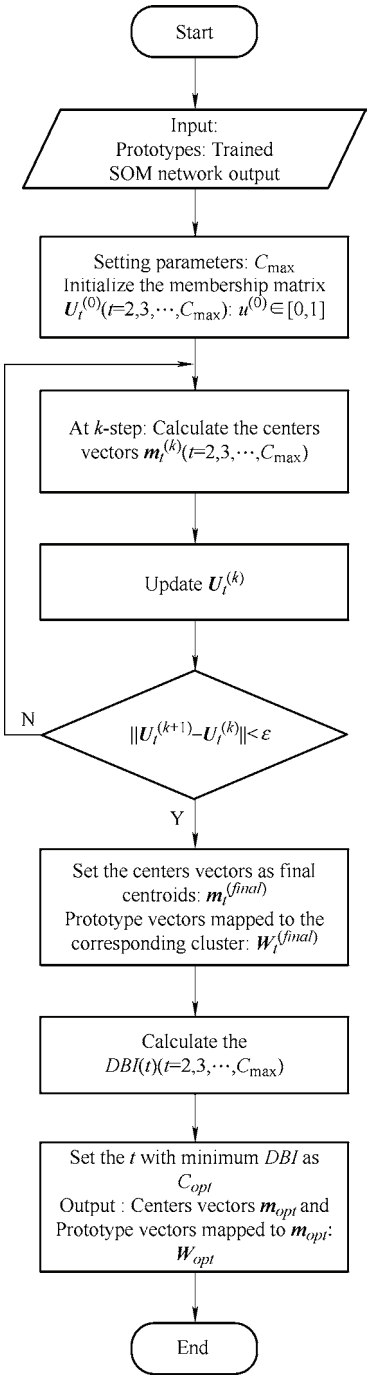


Fig. 2 Flowchart of the Two-Level algorithm

To obtain FCM, the condition that the sum of membership degrees of a sample from each cluster is equal to 1 must be satisfied, which can be expressed as

$$\sum_{j=1}^{C}\mu_j(\boldsymbol{x}_i)=1 \qquad i=1,2,\cdots,N \tag{7}$$

set the partial derivatives of J_f sub $\boldsymbol{m}_j$ and $\mu_j(\boldsymbol{x}_i)$ to 0. The solutions are as follows

$$\boldsymbol{m}_j=\frac{\sum_{i=1}^{N}\left[\mu_j(\boldsymbol{x}_i)\right]^b\boldsymbol{x}_i}{\sum_{i=1}^{N}\left[\mu_j(\boldsymbol{x}_i)\right]^b} \qquad j=1,2,\cdots,C \tag{8}$$

$$\mu_j(\boldsymbol{x}_i)=\frac{\left(1/\left\|\boldsymbol{x}_i-\boldsymbol{m}_j\right\|^2\right)^{\frac{1}{b-1}}}{\sum_{k=1}^{C}\left(1/\left\|\boldsymbol{x}_i-\boldsymbol{m}_k\right\|^2\right)^{\frac{1}{b-1}}} \qquad i=1,2,\cdots,N \tag{9}$$

Contrapose the selection of C, the maximum cluster size of FCM is predefined as $C_{\max}$. Then, the clustering assessment under each C, from 2 to $C_{\max}$, is calculated with DBI. The smaller the DBI is, the better the performance of clustering becomes. The optimal cluster size, C_{opt}, determined by DBI is as follows

$$C_{opt}=\arg\min_{C} DBI(\boldsymbol{W},\boldsymbol{M}_C) \qquad C=2,3,\cdots,C_{\max} \tag{10}$$

where $\boldsymbol{M}_C=\{\boldsymbol{m}_j\}_{j=1}^{C}$, $\boldsymbol{m}_j\in\mathbb{R}^d$, represents a set of C clusters of FCM, as to $\boldsymbol{W}=\{\boldsymbol{w}_i\}_{i=1}^{P}$, $\boldsymbol{w}_i\in\mathbb{R}^d$, refers to the collection of P prototypes in SOM.

3 SOM-based regional modeling methods

3.1 KELM model

Several features differentiate KELM from ELM models. The KELM uses the kernel function to represent the unknown nonlinear feature mapping of the hidden layer, where the number of hidden layer nodes does not need to be provided in advance [26]. Both ELM and KELM can be solved by the regularized least squares algorithm. Under the premise of ensuring the learning speed, the generalization ability of KELM is superior [20].

Moreover, given N sample sets, $(\boldsymbol{x}_i,\boldsymbol{t}_i)\in\mathbb{R}^n\times\mathbb{R}^m$, $i=1,2,\cdots,N$, where $\boldsymbol{x}_i=\left[x_{i,1},x_{i,2},\cdots,x_{i,n}\right]^{\mathrm{T}}\in\mathbb{R}^n$, $\boldsymbol{t}_i=\left[t_{i,1},t_{i,2},\cdots,t_{i,m}\right]^{\mathrm{T}}\in\mathbb{R}^m$, if there are L hidden layer nodes, the SLFN can be embodied as

$$\sum_{i=1}^{L}\beta_i\vartheta\left(\boldsymbol{\omega}_i\cdot\boldsymbol{x}_j+b_i\right)=O_j \qquad j=1,2,\cdots,N \tag{11}$$

where $\vartheta(x)$ is the activation function, $\boldsymbol{\omega}_i=\left[\omega_{i,1},\omega_{i,2},\cdots,\omega_{i,n}\right]^{\mathrm{T}}$ is the input weight vector, β_i denotes the output weight vector between the i-th hidden layer node and the output layer node, b_i is the bias of the i-th hidden layer nodes, $\boldsymbol{\omega}_i\cdot\boldsymbol{x}_j$ represents the scalar product of $\boldsymbol{\omega}_i$ and $\boldsymbol{x}_j$, and O_i is the actual output of the SLFN.

According to the ELM theory, it is assumed that there must be corresponding β_i, $\boldsymbol{\omega}_i$, and b_i that satisfy the following equation

$$\sum_{i=1}^{L}\beta_i \vartheta\left(\boldsymbol{\omega}_i \cdot \boldsymbol{x}_j + b_i\right) = t_j \qquad j = 1,2,\cdots,N \tag{12}$$

the matrix form of Eq. (12) is

$$\boldsymbol{H}\boldsymbol{\beta} = \boldsymbol{T} \tag{13}$$

where, $\boldsymbol{H}$ is the output matrix of the hidden layer node and $\boldsymbol{\beta}$ is the output weight matrix; the desired output is $\boldsymbol{T}$. The concrete expression of earlier mentioned matrixes are as follows

$$\begin{aligned}\boldsymbol{H} &= \left(\boldsymbol{\omega}_1,\boldsymbol{\omega}_2,\cdots,\boldsymbol{\omega}_L;b_1,b_2,\cdots,b_L;\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_L\right) = \\ &\begin{bmatrix} \vartheta\left(\boldsymbol{\omega}_1 \cdot \boldsymbol{x}_1 + b_1\right) & \cdots & \vartheta\left(\boldsymbol{\omega}_L \cdot \boldsymbol{x}_1 + b_L\right) \\ \vdots & & \vdots \\ \vartheta\left(\boldsymbol{\omega}_1 \cdot \boldsymbol{x}_N + b_1\right) & \cdots & \vartheta\left(\boldsymbol{\omega}_L \cdot \boldsymbol{x}_N + b_L\right) \end{bmatrix}_{N\times L}\end{aligned} \tag{14}$$

$$\boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_1^{\mathrm{T}} \\ \vdots \\ \boldsymbol{\beta}_L^{\mathrm{T}} \end{bmatrix}_{L\times m} \qquad \boldsymbol{T} = \begin{bmatrix} \boldsymbol{t}_1^{\mathrm{T}} \\ \vdots \\ \boldsymbol{t}_N^{\mathrm{T}} \end{bmatrix}_{N\times m} \tag{15}$$

The traditional idea based on gradient descent requires all parameters to be adjusted during the iteration. Nevertheless, in the context of ELM, this can be transformed into a problem of solving the linear system, $\boldsymbol{H}\boldsymbol{\beta} = \boldsymbol{T}$. Thus, the least squares solution of the network output weight is

$$\hat{\boldsymbol{\beta}} = \boldsymbol{H}^{\mathrm{T}}\left(\boldsymbol{H}\boldsymbol{H}^{\mathrm{T}} + \eta_k \boldsymbol{I}\right)^{-1}\boldsymbol{T} \tag{16}$$

where η_k is the regularization coefficient, $\boldsymbol{I} \in \mathbb{R}^{N\times N}$ is an identity matrix. Consequently, the output function of ELM can be obtained as follows

$$y(x) = \boldsymbol{h}(\boldsymbol{x})\boldsymbol{\beta} \tag{17}$$

In ELM, if the feature mapping function $\boldsymbol{h}(\boldsymbol{x})$ is unknown, it is an effective solution to introduce a kernel function $K\left(\boldsymbol{x}_i,\boldsymbol{x}_j\right)$ to replace the kernel matrix $\boldsymbol{h}\left(\boldsymbol{x}_i\right)\cdot\boldsymbol{h}\left(\boldsymbol{x}_j\right)$. This is the main idea in the KELM method; accordingly, the kernel function is defined as

$$Q_{\mathrm{ELM}}(i,j) = \boldsymbol{h}\left(\boldsymbol{x}_i\right)\cdot\boldsymbol{h}\left(\boldsymbol{x}_j\right) = K\left(\boldsymbol{x}_i,\boldsymbol{x}_j\right) \tag{18}$$

and the output of KELM is defined as

$$y(x) = \begin{bmatrix} K\left(\boldsymbol{x},\boldsymbol{x}_1\right) \\ \vdots \\ K\left(\boldsymbol{x},\boldsymbol{x}_N\right) \end{bmatrix}\left(Q_{\mathrm{ELM}} + \eta_k \boldsymbol{I}\right)^{-1}\boldsymbol{T} \tag{19}$$

Under the general smoothing assumptions, the Gaussian radial basis function (RBF) tends to perform better. Therefore, the general form of Gaussian RBF adopted in this study can be expressed as follows

$$K\left(\boldsymbol{x}_i,\boldsymbol{x}_j\right) = \exp\left(-\frac{\left\|\boldsymbol{x}_i - \boldsymbol{x}_j\right\|^2}{\gamma^2}\right) \tag{20}$$

where γ represents the radius of Gaussian RBF. γ and η_k are easy to adjust, which significantly improves the performance of KELM. In this study, we employ a differential evolution algorithm [27] to optimize the parameters of ELM and KELM methods.

3.2 Regional KELM

The implementation processes of the SOM-based regional KELM model are discussed step by step.

Step 1: Set the hyper-parameters. In the SOM algorithm, the minimum size of neurons is generally set as $P = 5\sqrt{N}$, where N is the number of training samples. For the purpose of visualization, a two-dimensional hexagonal topology of the network is constructed in this paper. The maximum number of clusters in FCM is set to $C_{\max}=N$.

Step 2: Training of the SOM network. As mentioned in the Two-Level algorithm, the pure SOM completes the preliminary divisions of samples. For each SOM prototype, execute Eq. (5) to obtain the corresponding Voronoi cell.

Step 3: Clustering of the trained SOM network. An evaluable FCM executes the clustering of the trained SOM network. As mentioned before, the DBI evaluates the effectiveness of clustering, simultaneously, and finds the optimal size of clustering, C_{opt}, the selection of C_{opt} follows Eq. (10). Once the optimal cluster size C_{opt} is determined, the optimal regional divisions are equivalent to the clustering of the trained SOM network. The *r*-th cluster of SOM prototypes is composed of all neuron vectors $\boldsymbol{w}_i$, which are mapped onto the cluster $\boldsymbol{C}_r$, $r = 1,2,\cdots,C_{opt}$. The set of SOM prototypes associated with the *r*-th cluster of FCM is defined as

$$\begin{aligned}\boldsymbol{C}_r = \Big\{\boldsymbol{w}_i \in \mathbb{R}^d \,\Big|\, &\left\|\boldsymbol{w}_i - \boldsymbol{m}_r\right\| < \left\|\boldsymbol{w}_i - \boldsymbol{m}_j\right\| \\ &\forall\, j = 1,2,\cdots,C_{opt}, j \neq r\Big\}\end{aligned} \tag{21}$$

Step 4: Data partitions. After the above processes, the input vectors are still stored in the Voronoi cells. To further divide the data into determined C_{opt} data partitions. Let $\boldsymbol{F}_i$ be the sample set mapped to the *i*-th region, N_i is the number of samples in $\boldsymbol{F}_i$. $\boldsymbol{F}_i$ is composed of $\boldsymbol{x}_1$, $\boldsymbol{x}_2$, $\cdots$, $\boldsymbol{x}_{N_i}$, and the SOM

prototypes to which these input vectors belong are mapped to the *i*-th region.

Step 5: Construct different KELM models in different regions. Finally, the original data is divided into C_{opt} subsets. For each subset, data mapped to this region is used to train the KELM model optimized by the DE algorithm, then, a SOM-based SF-KELM model for short-term PV power generation forecasting is constructed.

In step 5, combining SOM's regional partitions with an optimized MR, we can obtain a SOM-based Regional MR model, i.e., the SF-MR model.

4 Case analysis

In this section, the forecasting of PV power generation is carried out on 12 subsets (divided by month) of three solar power plants. The data for 2012 and 2013 collected in each plant are employed to construct a training dataset, and their counterparts in 2014 are used for forecasts. The forecasts obtained with single ELM and KELM methods are used to evaluate the performance of the proposed models.

4.1 Performance evaluation metric

To evaluate the performance of the proposed SF-KELM and SF-MR models, the mean absolute error (MAE), root means square error (RMSE), and coefficient of determination (R^2) are adopted. The concrete expressions of these evaluation metrics are as follows.

$$\mathrm{MAE}=\frac{1}{N}\sum_{i=1}^{N}\left|y(i)-f\left(x_i\right)\right| \tag{22}$$

$$\mathrm{RMSE}=\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left[y(i)-f(\boldsymbol{x}_i)\right]^2} \tag{23}$$

$$R^2=1-\frac{\frac{1}{N}\sum_{i=1}^{N}\left[y(i)-f(\boldsymbol{x}_i)\right]^2}{\frac{1}{N}\sum_{i=1}^{N}\left[y(i)-\bar{y}\right]^2} \tag{24}$$

where N is the size of the test set, $y(i)$ represents the actual output, $f\left(\boldsymbol{x}_i\right)$ represents the predicted value, and $\bar{y}$ represents the average of predicted power. The index, $R^2\,(0<R^2<1)$, reflects the fitting degree of the forecasting model; the closer R^2 is to 1, the higher the accuracy of the model is.

4.2 Experimental results and analysis

The experimental data, provided by GEFCom2014, are derived from three solar power plants located in a certain region of Australia. The raw data supplied by the institution consists of two parts: The first part records the daily PV power generation (from 00:00 to 23:00) of each solar power plant from 2012 to 2014. The second part is numerical records of 12 independent meteorological variables provided by the European Centre for Medium-range Weather Forecasts (ECMWF) with a sampling interval of 1 h [24]. The descriptions of meteorological variables are reported in Tab. 1.

Tab. 1 The description information of 12 independent meteorological factors

Label	Description	Unit
VAR78	Total column liquid water (TCLW)	kg/m^2
VAR79	Total column ice water (TCIW)	kg/m^2
VAR134	Surface pressure (SP)	Pa
VAR157	Relative humidity at 1 000 mbar (r)	%
VAR164	Total cloud cover (TCC)	0-1
VAR165	10-metre U wind component (10U)	m/s
VAR166	10-metre V wind component (10V)	m/s
VAR167	2-metre temperature (2T)	K
VAR169	Surface solar rad down (SSRD)	J/m^2
VAR175	Surface thermal rad down (STRD)	J/m^2
VAR178	Top net solar rad (TSR)	J/m^2
VAR228	Total precipitation (TP)	m

Because different meteorological factors correspond to different ranges and metrics, in this investigation, all selected inputs are mapped to the interval [0,1]. Das et al. [4] pointed out that a remarkably strong correlation exists between PV power generation and solar irradiance. Such results are ubiquitous, regardless of the geographic location of PV power plants and weather conditions. Based on the above conclusions, it is clear to choose VAR169 and VAR178 as the inputs of the proposed models. Accordingly, the filter-based relief and wrapper-based sequential forward selection (SFS) methods are used to select the remaining features; the results of the two methods are comprehensively considered to determine the final inputs [28]. Tabs. 2-3 report the screening results of meteorological factors in solar power plant #1.

Tab. 2 Feature selection results using the SFS method

Stage	Hit feature	Score	Total features
1	VAR175	0.980 8	VAR169/VAR175/VAR178
2	VAR228	0.979 4	VAR169/VAR175/VAR178/VAR228

Tab. 3 Feature selection results using the ReliefF method

Rank	Hit feature	Weight
1	VAR175	0.041 4
2	VAR166	0.003 3
3	VAR157	0.003 1
4	VAR167	0.001 1
5	VAR165	$-4.013\ 0\times10^{-4}$
6	VAR79	–0.004 1
7	VAR78	–0.004 7
8	VAR134	–0.010 1
9	VAR228	–0.010 5
10	VAR164	–0.012 0

Finally, VAR157, 166, 167, 169, 175, 178, and 228 are determined as the inputs of the forecasting model. The screening results of plants #2 and #3 are the same as plant #1.

4.3 Experimental results and analysis

The inputs of forecasting models are composed of different meteorological factors, and the output is PV power. For each solar power plant, the PV power generation forecasting is conducted one hour ahead. The data for 2012 and 2013 are used for model training, while the data for 2014 are used for testing. If the whole dataset is divided into training and test sets according to a ratio of 0.7 to 0.3, the forecasting performance will be greatly compromised. This is because a large amount of data is generated in each plant, with a size of 19 704. More importantly, in some cases, the underlying laws will be missing. To avoid such deficiencies, the data is divided into different subsets monthly to establish multiple forecasting models.

Tab. 4 reports the parameter settings of SOM for regional models, where N represents the size of the training dataset. The neuron weights of the proposed SF-KELM and SF-MR models are randomly initialized using training samples.

Tab. 4 SOM parameter settings of SF-KELM and SF-MR models

Parameter	Regional models
Prototype size	$5\sqrt{N}$
Network structure	Two-dimensional hexagonal topology
Training algorithm	Sequential
Batch size	50
Training epoch	100
Initial learning rate α_0	0.8
Final learning rate α_T	0.01
Initial radius σ_0	$5\sqrt{N}$ /2
Final radius σ_T	0.01
Clustering of trained SOM	Evaluable FCM
Maximum clusters	$5\sqrt{N}$

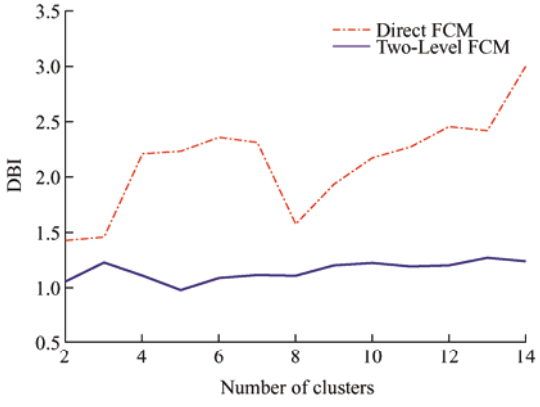
The weights and biases of inputs to the hidden layer in ELM are randomly initialized. Sigmoid is selected as the activation function of the hidden layer neurons. As mentioned previously, the kernel function of KELM assumes the general form of the Gaussian kernel function; accordingly, the parameters γ, η_k, and L are determined by the DE algorithm. Tab. 5 reports the optimized parameters of ELM and KELM in plant #1.

Tab. 5 Optimized parameters of ELM and KELM in plant #1

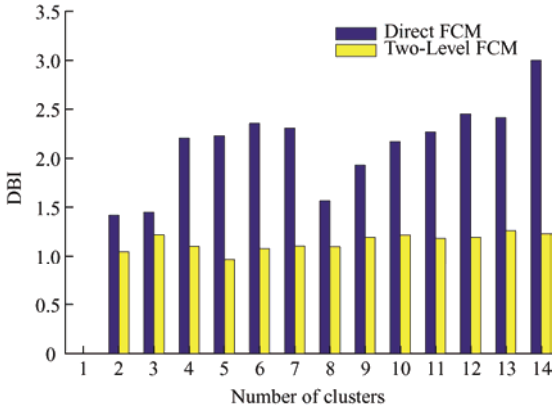
Month	ELM (L)	KELM (η_k)	KELM (γ)
Jan.	38	82.429 0	0.610 1
Feb.	34	84.832 6	1.554 4
Mar.	29	77.512 3	1.045 6
Apr.	37	70.817 3	0.538 1
May	38	5.888 4	0.169 3
Jun.	38	47.591 1	0.724 0
Jul.	36	64.891 8	1.311 3
Aug.	38	25.773 1	0.831 8
Sept.	36	33.355 6	1.240 0
Oct.	38	78.757 4	1.296 4
Nov.	33	52.596 8	0.335 0
Dec.	34	62.683 7	0.655 3

Consider the training dataset in plant#1, April, 2012 and 2013 as an example. Fig. 3 shows the DBIs of direct FCM and Two-Level FCM under different clustering sizes. Note that the optimal clustering size

of the proposed Two-Level algorithm is 5, which yields superior performance.



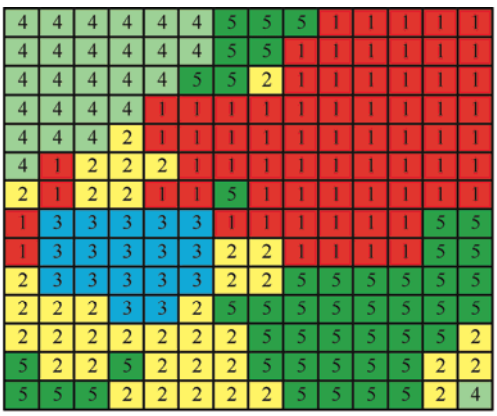
(a) Cluster validity statistics



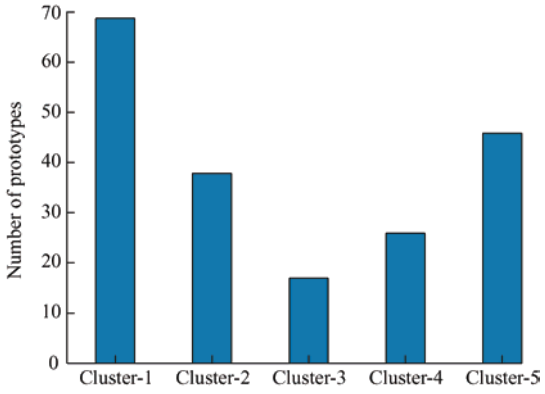
(b) Cluster validity statistics

Fig. 3 Results of DBI under different cluster sizes in Plant#1, April, 2012 and 2013

Fig. 4 displays the clustering results of the trained SOM network. In Fig. 4a, the number in each small square indicates the cluster to which the prototype belongs. Moreover, the adjacent prototypes arranged in the output layer are divided into the same cluster, which echoes the neighborhood update strategy of the SOM network and verifies the effectiveness of the proposed Two-Level algorithm. Fig. 4b presents the number of prototypes in each cluster.



(a) Clustered prototypes



(b) The number of prototypes in each cluster

Fig. 4 Clustering results of the trained SOM network in Plant#1, April, 2012 and 2013

To evaluate the performance of the proposed SF-KELM and SF-MR models, considering the forecasts of each plant in April, 2014 an example, the error metrics of all forecasting models are reported in Tabs. 6-8.

Tab. 6 Performance comparisons of different models in Plant #1, April, 2014

Methods	Global models			Direct clustering model
	MR	ELM	KELM	DKELM
MAE/kW	0.115 6	0.090 9	0.076 4	0.054 8
RMSE/kW	0.160 9	0.153 4	0.113 9	0.090 3
R^2(%)	0.530 7	0.573 5	0.764 6	0.852 0

Methods	Regional models	
	SF-MR	SF-KELM
MAE/kW	0.044 7	0.030 5
RMSE/kW	0.072 0	0.060 6
R^2(%)	0.905 9	0.933 3

Tab. 7 Performance comparisons of different models in Plant #2, April, 2014

Methods	Global models			Direct clustering model
	MR	ELM	KELM	DKELM
MAE/kW	0.106 7	0.096 4	0.065 9	0.059 7
RMSE/kW	0.144 7	0.138 1	0.108 9	0.103 7
R^2(%)	0.639 0	0.670 8	0.795 6	0.814 5

Methods	Regional models	
	SF-MR	SF-KELM
MAE/kW	0.044 3	0.030 8
RMSE/kW	0.073 6	0.056 0
R^2(%)	0.906 7	0.945 8

Tab. 8 Performance comparisons of different models in Plant #3, April, 2014

Methods	Global models			Direct clustering model
	MR	ELM	KELM	DKELM
MAE/kW	0.114 7	0.098 4	0.070 5	0.064 6
RMSE/kW	0.154 5	0.144 8	0.113 6	0.110 0
R^2(%)	0.593 0	0.639 8	0.778 3	0.792 2
Methods	Regional models			
	SF-MR		SF-KELM	
MAE/kW	0.042 5		0.034 6	
RMSE/kW	0.073 2		0.063 6	
R^2(%)	0.907 9		0.930 5	

Observe from Tabs. 6-8 that in any plant, the proposed SF-MR and SF-KELM models have reached their highest accuracy. Compared with a single MR, in three plants, the MAEs of the proposed SF-MR model are reduced by 61.33%, 58.48%, and 54.93%, RMSEs are reduced by 55.25%, 49.14%, and 41.81%, and accuracy of R^2 is improved by 70.70%, 41.89%, and 45.21%. Significantly, the proposed SF-KELM model outstands other contrast models and has achieved the highest accuracy. The average values of MAEs, RMSEs, and R^2 generated by SF-KELM are 0.032 0, 0.060 1, and 0.936 5, respectively. The accuracy of SF-KELM is slightly higher than that of SF-MR, mainly because the learning ability of NNs is superior in processing nonlinear systems, especially for meteorological factors with strong uncertainty and nonlinearity. In addition, the FCM-KELM method based on direct clustering achieves satisfactory forecasting performance. However, its accuracy is inferior compared with the proposed SF-MR and SF-KELM models.

In addition, to further evaluate the forecasting performance of the proposed SF-KELM and SF-MR models, the forecasting methods mentioned in the introduction are also adopted for comparative analysis. For simplicity, the forecasting method proposed in Ref. [8] is abbreviated as WPD-LSTM, competitive swarm optimization algorithm optimized RBFNN in Ref. [9] is abbreviated as CSO-RBFNN, and methods proposed in Refs. [10-11] are abbreviated as GASVM and MP, respectively. Considering the forecasts of each plant in April, 2014 an example, the error metrics of all proposed and other reference methods are reported in Tabs. 9-11.

Tab. 9 Forecasting comparisons between proposed and other reference methods of Plant #1, April, 2014

Methods	Contrast methods			
	WPD-LSTM	CSO-RBFNN	GASVM	MP
MAE/kW	0.033 6	0.031 8	0.037 0	0.041 8
RMSE/kW	0.060 0	0.065 0	0.064 6	0.074 3
R^2(%)	0.934 7	0.923 4	0.924 2	0.897 2
Methods	Proposed methods			
	SF-MR		SF-KELM	
MAE/kW	0.044 7		0.030 5	
RMSE/kW	0.072 0		0.060 6	
R^2(%)	0.905 9		0.933 3	

Tab. 10 Forecasting comparisons between proposed and other reference methods of Plant #2, April, 2014

Methods	Contrast methods			
	WPD-LSTM	CSO-RBFNN	GASVM	MP
MAE/kW	0.030 5	0.037 3	0.043 1	0.061 4
RMSE/kW	0.057 8	0.075 5	0.071 3	0.089 5
R^2(%)	0.942 4	0.901 7	0.912 3	0.861 7
Methods	Proposed methods			
	SF-MR		SF-KELM	
MAE/kW	0.044 3		0.030 8	
RMSE/kW	0.073 6		0.056 0	
R^2(%)	0.906 7		0.945 8	

Tab. 11 Forecasting comparisons between proposed and other reference methods of Plant #3, April, 2014

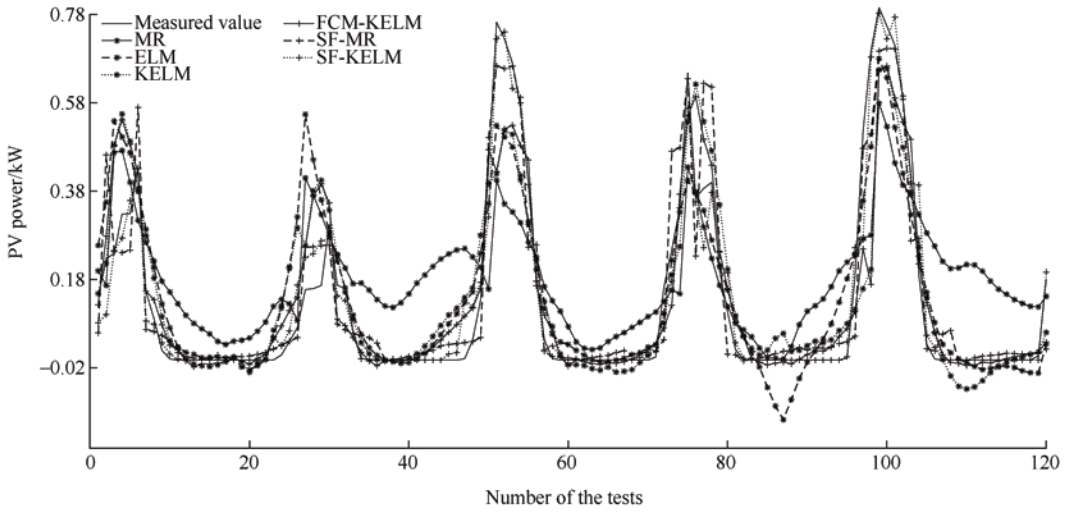
Methods	Contrast methods			
	WPD-LSTM	CSO-RBFNN	GASVM	MP
MAE/kW	0.037 1	0.033 0	0.043 0	0.056 4
RMSE/kW	0.064 3	0.069 7	0.070 0	0.082 0
R^2(%)	0.929 0	0.916 5	0.915 9	0.886 4
Methods	Proposed methods			
	SF-MR		SF-KELM	
MAE/kW	0.042 5		0.034 6	
RMSE/kW	0.073 2		0.063 6	
R^2(%)	0.907 9		0.930 5	

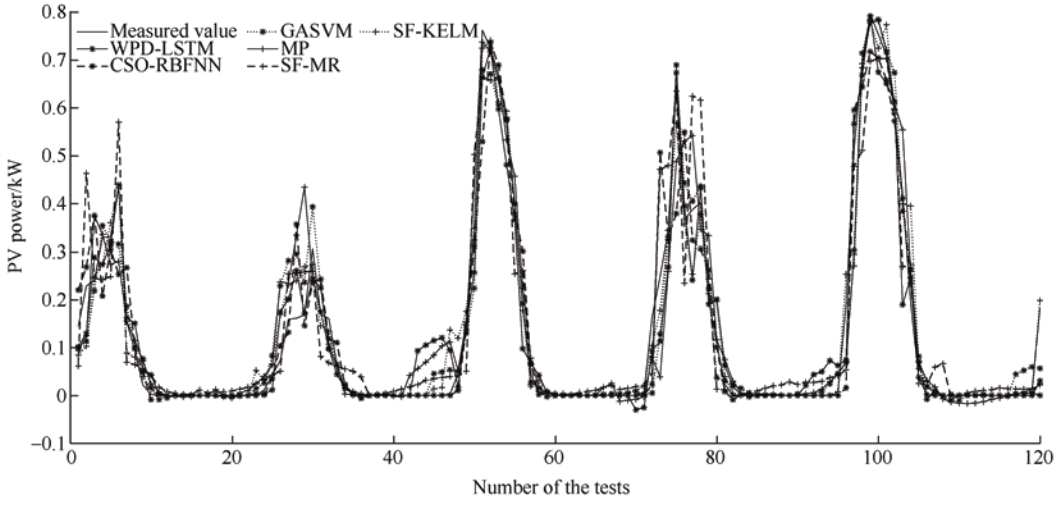
In the three plants, although the forecasting accuracy of the proposed SF-MR method is inferior to those of the WPD-LSTM, CSO-RBFNN, and GASVM methods, it is generally better than that of the MP method. The MAEs and RMSEs of the SF-KELM method are superior to those of the CSO-RBFNN, GASVM, and MP methods, but are comparable to that of the WPD-LSTM method. In addition, the model matching index R^2 of the proposed SF-KELM method is generally the highest among all methods, except for forecasts in plant #1. The other contrast methods also

achieved relatively high R^2, which manifests that all forecasting methods have been well trained.

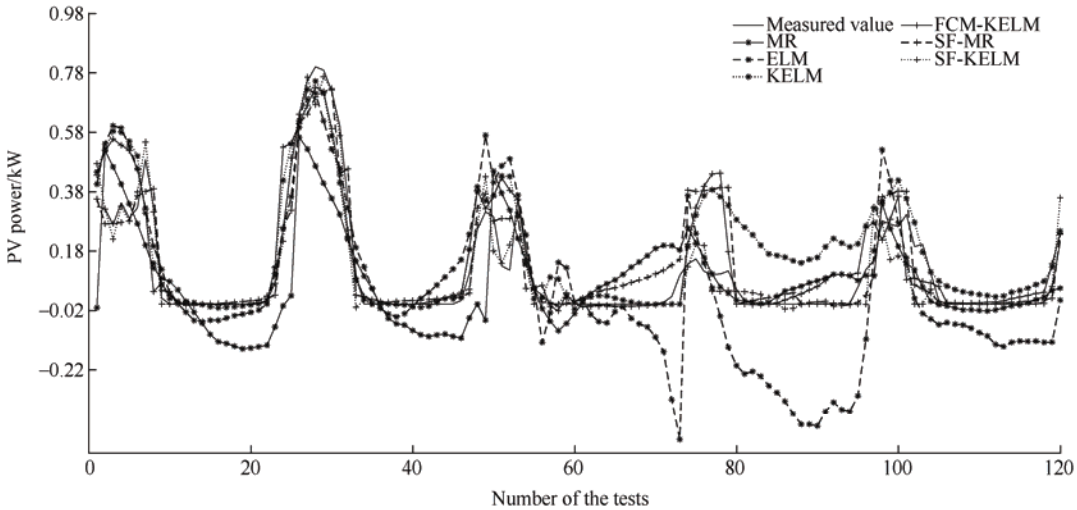
From the above analysis, we conclude that the forecasts of the proposed SF-EKTSK method are better than those obtained using the single ELM, KELM, and DKELM methods, as well as those of the reference forecasting methods, CSO-RBFNN, GASVM, and MP. However, they are comparable to that obtained with the WPD-LSTM method.

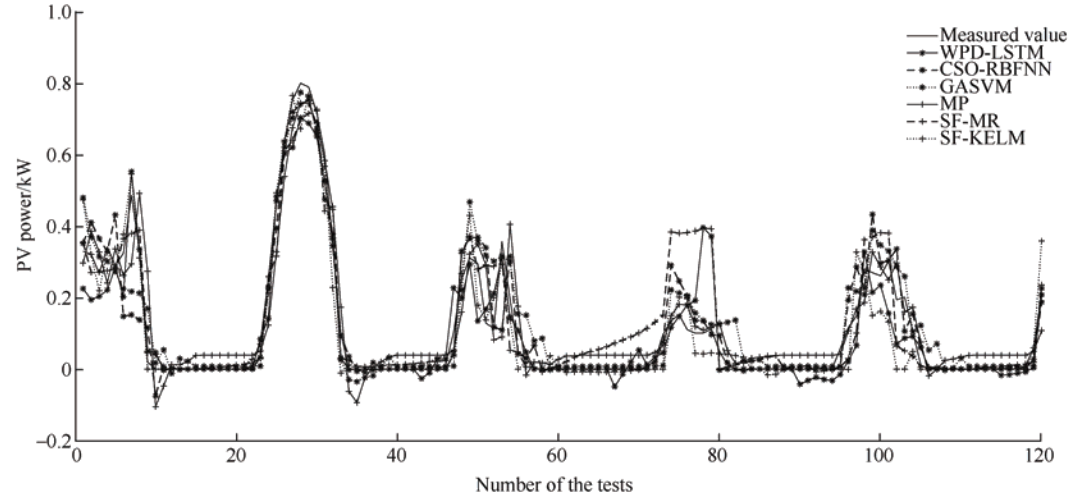
Fig. 5 shows the comparisons between the predicted and actual outputs of solar power plants #1-3, April, 2014. Overall, both the proposed SF-KELM and referenced WPD-LSTM, CSO-RBFNN, and GASVM methods can well fit the actual power curve.



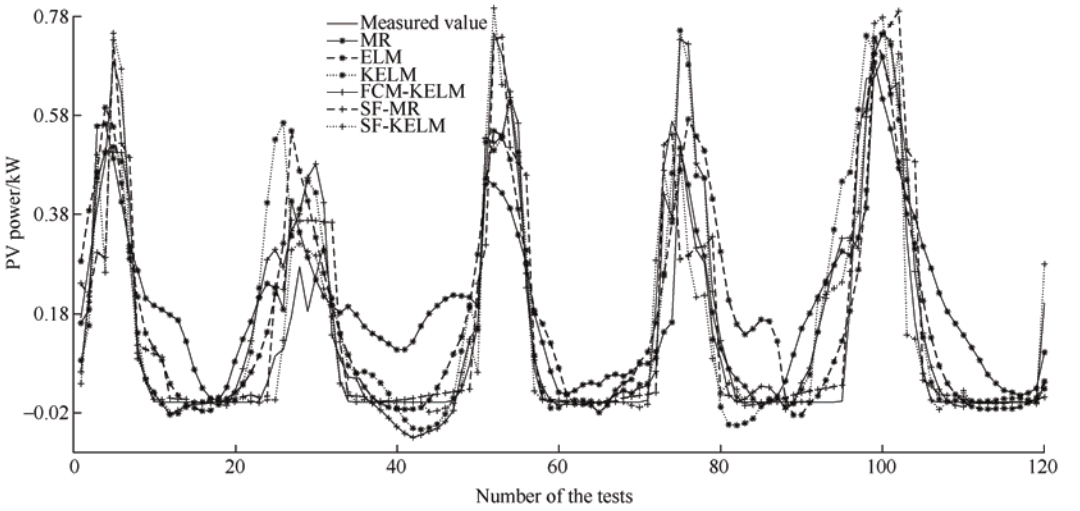


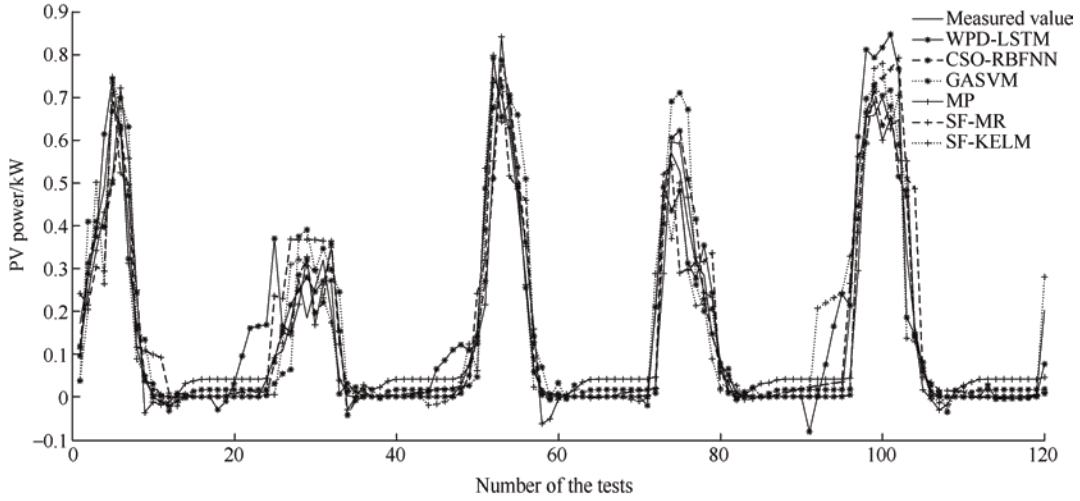
(a) Comparisons in plant#1, April, 2014





(b) Comparisons in plant#2, April, 2014





(c) Comparisons in plant#3, April, 2014

Fig. 5 Contrast results between the predicted and actual outputs in plant#1-3, April, 2014

The predicted outputs of SF-MR and SF-KELM methods exhibit only a small deviation at a few peak power points. In contrast, single global ELM, KELM, and combined FCM-KELM methods can not accurately approximate the actual outputs, especially at zero-power-point and peak-power-point. Even the fitting degree of the proposed SF-MR method is not as good as that of the referenced WPD-LSTM, CSO-RBFNN, and GASVM methods. However, it still achieves high forecasting accuracy. The forecasting curves generated by the proposed SF-KELM method are the most desired. Moreover, compared with most

referenced methods and the proposed SF-MR method, it is closer to the actual power curves.

Such results indicate that the SF-KELM method can quickly capture and respond to the existing power changes, which have also verified the superiority of the proposed Two-Level algorithm.

Fig. 6 shows the error diagrams generated by different forecasting methods of plants #1-3, April, 2014, which proves the superiority of the proposed SF-KELM method again.

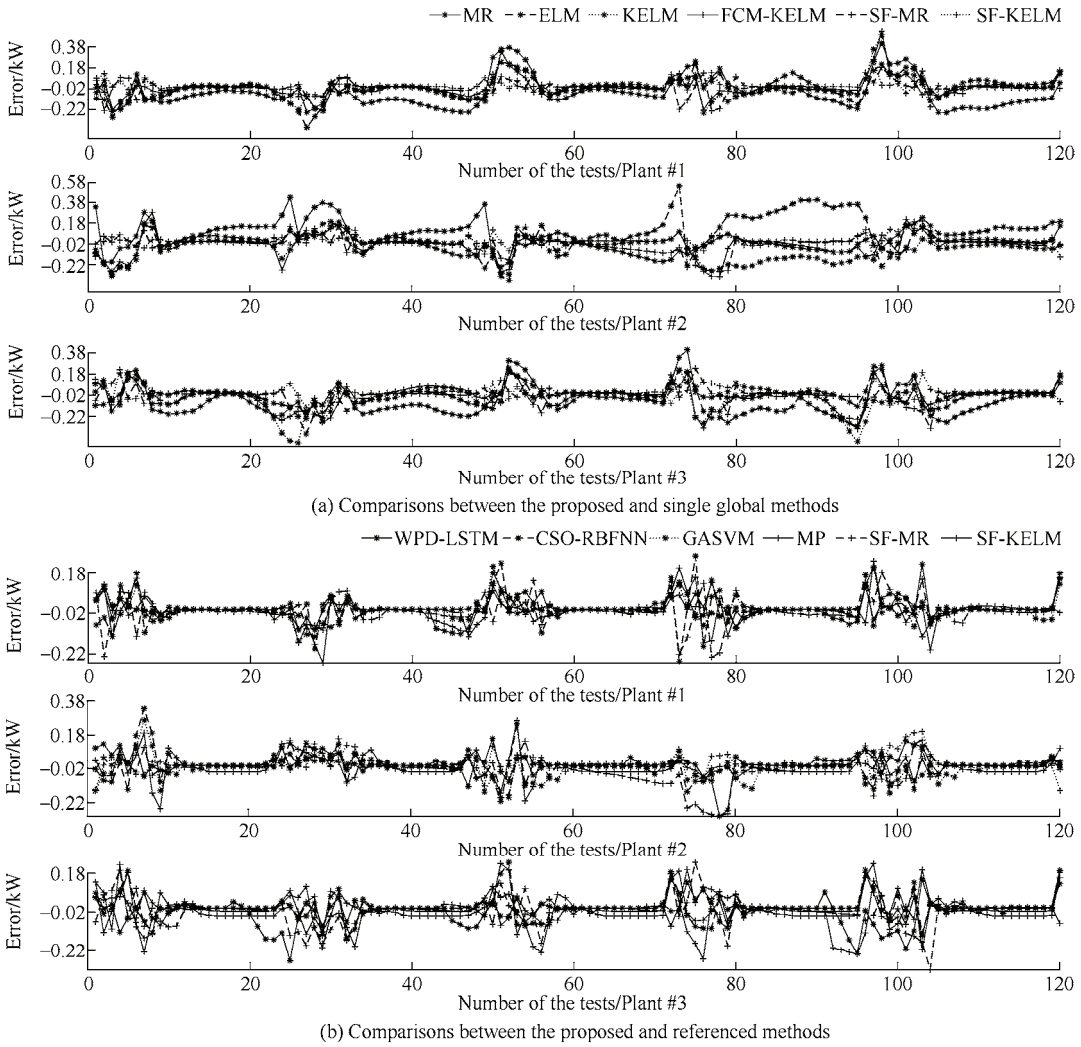


Fig. 6 Errors between the predicted and actual outputs in plant#1-3, April, 2014

To investigate the influence of seasonal factors on short-term PV power generation forecasting, taking the solar power plant#1 as an example, the forecasts of different methods in different seasons are compared. October, January, April, and July of 2012 and 2013 are selected as four training data sets. A random sample of five days in the corresponding month of solar power plant#1 from 2014 is employed as a test dataset.

Tab. 12 reports the error metrics of all forecasts in different seasons, to underline the indicators more clearly. Fig. 7 displays the statistical results of MAE, RMSE, and R^2 for all methods in all seasons. According to Fig. 7 and Tab. 9, the SF-KELM method has the highest accuracy among all forecasting methods, followed by the SF-MR method. In four seasons, the MAEs and RMSEs of the proposed SF-KELM are the smallest; additionally, the R^2 exceeds 0.9, which is the closest to 1. The performance of other methods is relatively poor, and various indices exhibit different degrees of deviation. The forecasts in different seasons have verified the effectiveness of the proposed SF-MR and SF-KELM models.

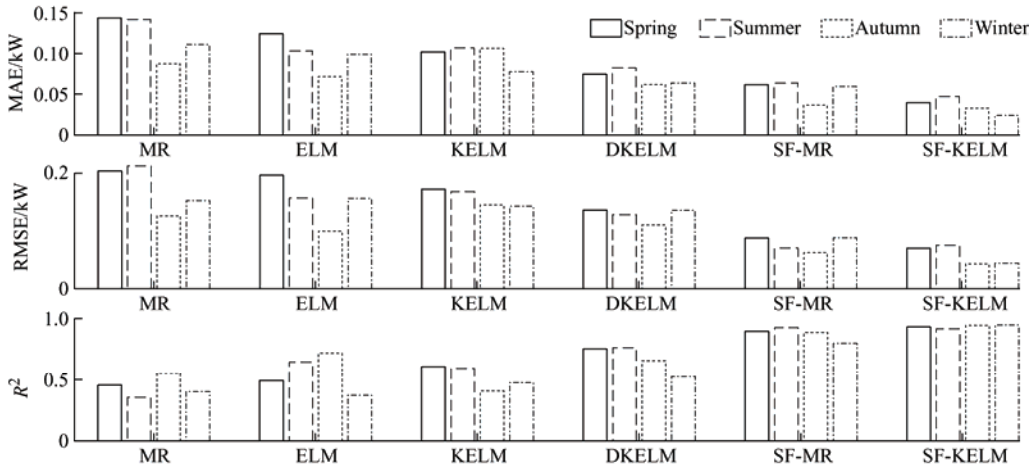


Fig. 7 Statistical results of MAE, RMSE and R^2

Tab. 12 The error metrics of different methods in different seasons

Season	Indexes	MR	ELM	KELM
Spring	MAE	0.144 0	0.124 6	0.102 1
	RMSE	0.203 7	0.196 7	0.172 3
	R^2	0.452 1	0.489 1	0.608 3
Summer	MAE	0.141 9	0.103 4	0.107 1
	RMSE	0.212 6	0.157 2	0.168 2
	R^2	0.350 8	0.645 2	0.593 7
Autumn	MAE	0.087 9	0.071 8	0.106 7
	RMSE	0.126 1	0.099 7	0.145 2
	R^2	0.549 8	0.718 8	0.403 9
Winter	MAE	0.111 3	0.099 3	0.077 7
	RMSE	0.152 4	0.156 1	0.143 0
	R^2	0.399 4	0.369 7	0.471 1

Season	Indexes	DKELM	SF-MR	SF-KELM
Spring	MAE	0.074 9	0.062 0	0.040 0
	RMSE	0.136 1	0.087 6	0.070 2
	R^2	0.755 5	0.898 6	0.934 9
Summer	MAE	0.082 6	0.064 0	0.047 4
	RMSE	0.128 0	0.070 2	0.075 6
	R^2	0.764 7	0.929 2	0.917 9
Autumn	MAE	0.062 3	0.036 8	0.032 9
	RMSE	0.110 2	0.062 5	0.043 0
	R^2	0.656 4	0.889 5	0.947 6
Winter	MAE	0.064 1	0.059 6	0.024 3
	RMSE	0.136 0	0.087 9	0.044 1
	R^2	0.521 6	0.800 2	0.949 7

The comparisons between the proposed and abovementioned reference methods in different seasons of solar power plant #1 are further reported in Tab. 13.

It is clear that in four seasons, the forecasts of the proposed Sf-KELM method are generally better than proposed SF-MR and referenced CSO-RBFNN, GASVM, and MP methods, and comparable with referenced WPD-LSTM method. The results presented in Tabs. 9-13 demonstrate the excellent performance of the proposed SF-KELM method in short-term PV power generation forecasting and strong robustness.

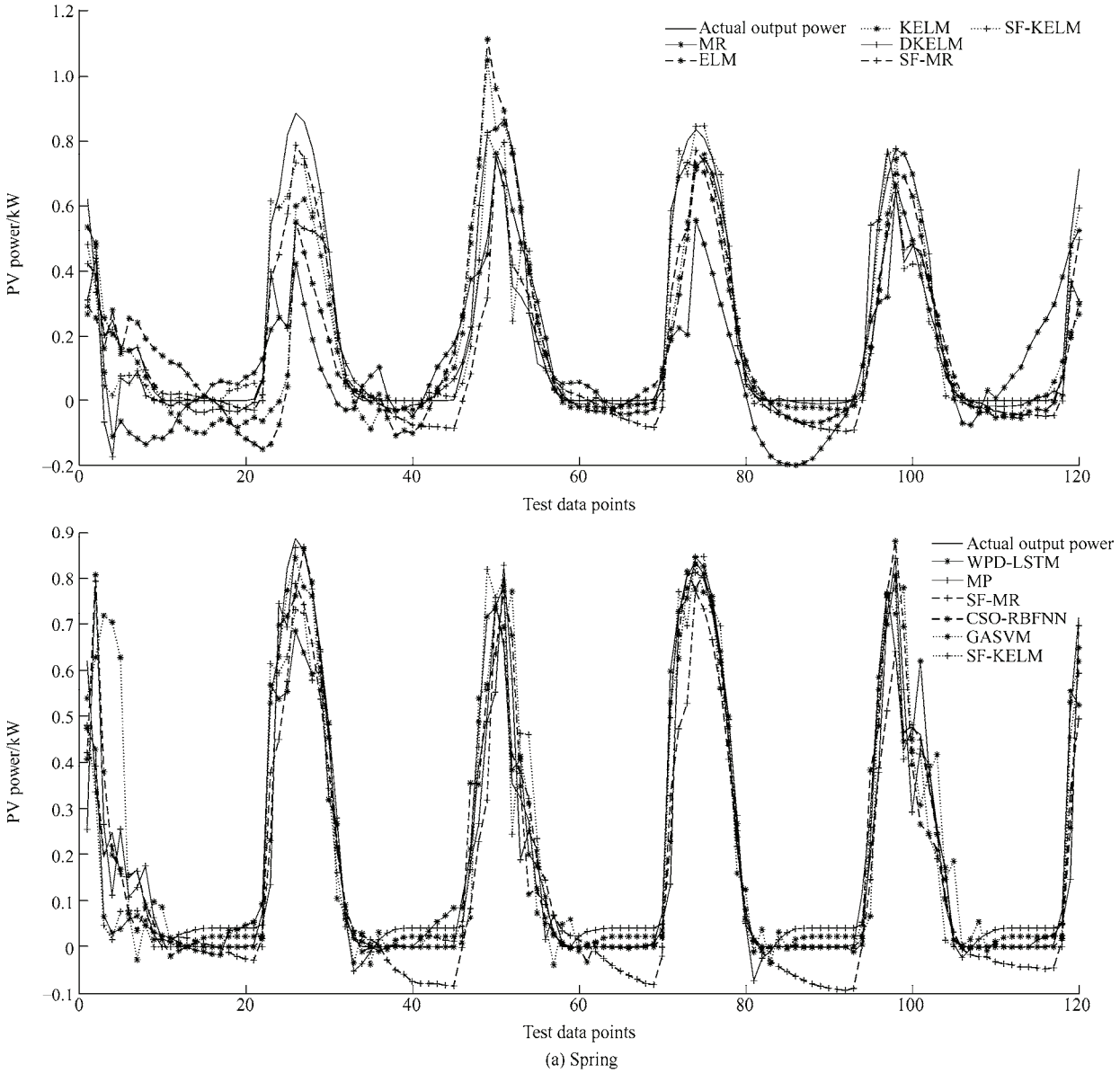
Tab. 13 Forecasting comparisons between proposed and other reference methods in different seasons

Season	Indexes	WPD-LSTM	CSO-RBFNN	GASVM
Spring	MAE	0.048 0	0.037 1	0.053 9
	RMSE	0.071 1	0.080 6	0.086 1
	R^2	0.933 3	0.929 2	0.919 2
Summer	MAE	0.042 1	0.031 7	0.048 4
	RMSE	0.064 3	0.061 2	0.072 3
	R^2	0.940 6	0.946 1	0.933 8
Autumn	MAE	0.030 5	0.032 6	0.045 2
	RMSE	0.058 4	0.065 3	0.070 9
	R^2	0.938 0	0.922 5	0.908 6
Winter	MAE	0.027 9	0.038 6	0.047 2
	RMSE	0.053 3	0.086 8	0.080 5
	R^2	0.926 5	0.817 0	0.842 6

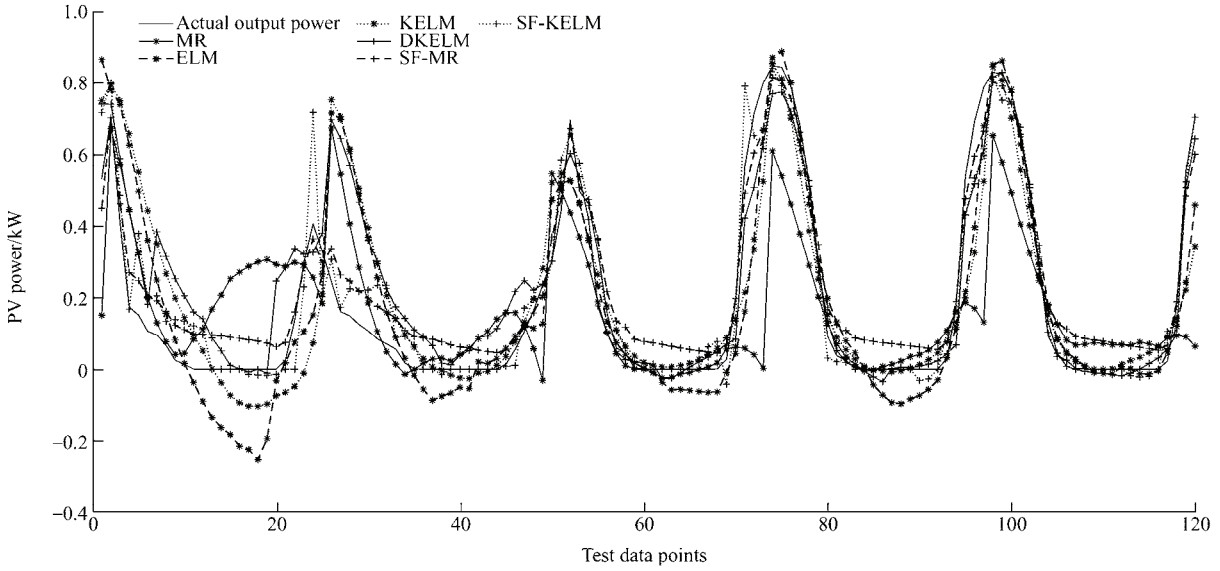
Season	Indexes	MP	SF-MR	SF-KELM
Spring	MAE	0.062 8	0.062 0	0.040 0
	RMSE	0.103 4	0.087 6	0.070 2
	R^2	0.859 0	0.898 6	0.934 9
Summer	MAE	0.049 3	0.064 0	0.047 4
	RMSE	0.077 9	0.070 2	0.075 6
	R^2	0.912 9	0.929 2	0.917 9
Autumn	MAE	0.046 2	0.036 8	0.032 9
	RMSE	0.064 6	0.062 5	0.043 0
	R^2	0.881 8	0.889 5	0.947 6
Winter	MAE	0.053 9	0.059 6	0.024 3
	RMSE	0.084 7	0.087 9	0.044 1
	R^2	0.814 7	0.800 2	0.949 7

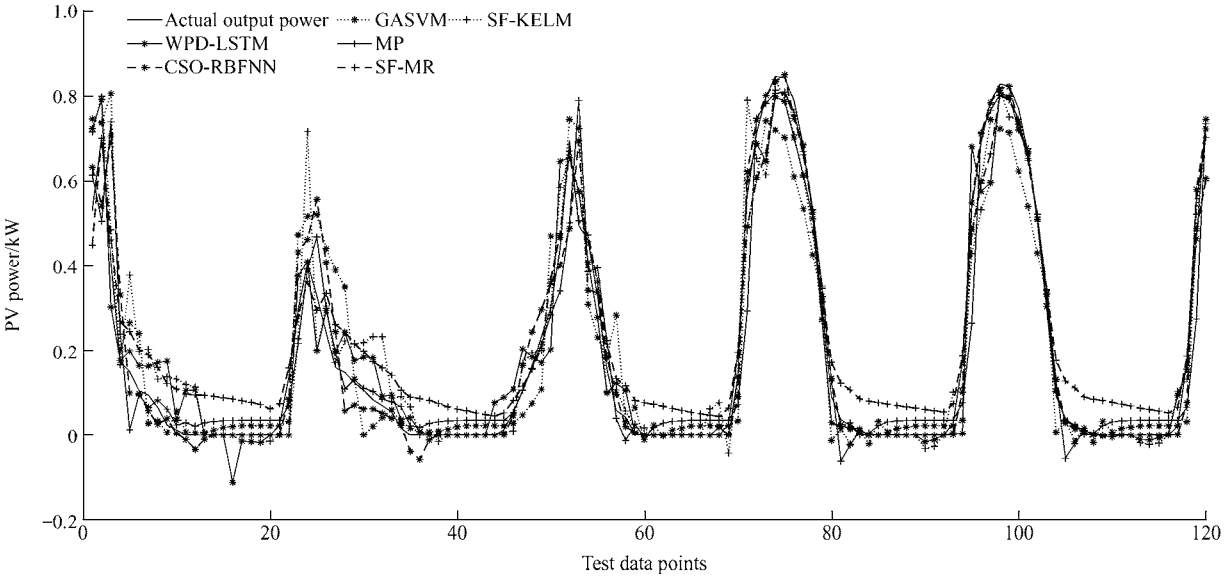
The main reasons for the proposed SF-KELM method to generate such good forecasts can be summarized in two points: First, it benefits from the nonlinear learning ability of NN. Second, it benefits from the data partitions of the regional partition method combining SOM and FCM algorithms.

Fig. 8 shows the comparisons between the predicted and measured power values for different models in different seasons, as well as the errors. The power generation of each season contains different types. On some days, the generated power changes rapidly, with obtrusive spikes and sudden drops. In such cases, the forecasts of the models used for comparisons deviate far from the actual values and cannot capture the existing changes. Conversely, the proposed SF-KELM and SF-MR models can quickly respond to power changes and accurately approximate the outputs. Over the four seasons, the errors generated by SF-KELM fluctuate around zero, and the corresponding change ranges are generally the narrowest among all methods.

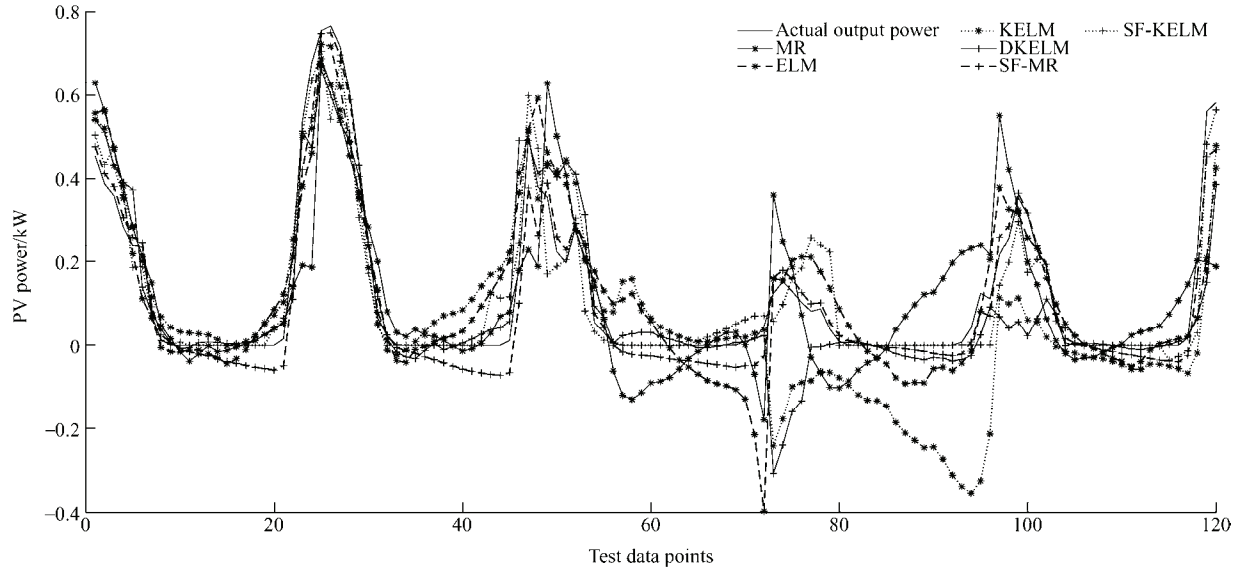


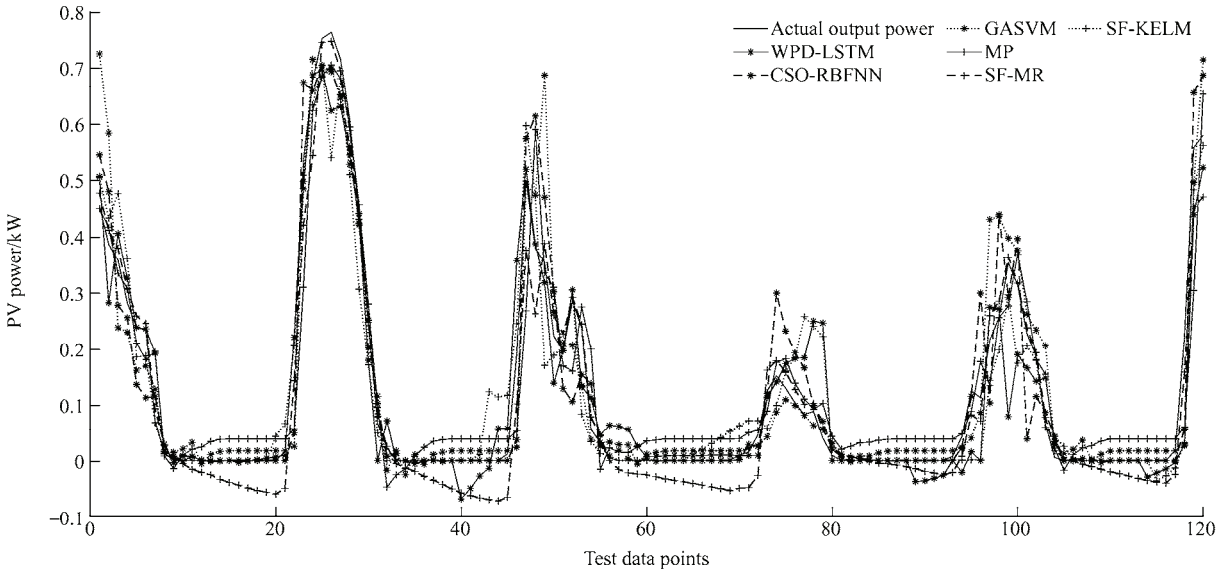
(a) Spring


Actual output power
KELM
SF-KELM
MR
DKELM
ELM
SF-MR
PV power/kW
Test data points

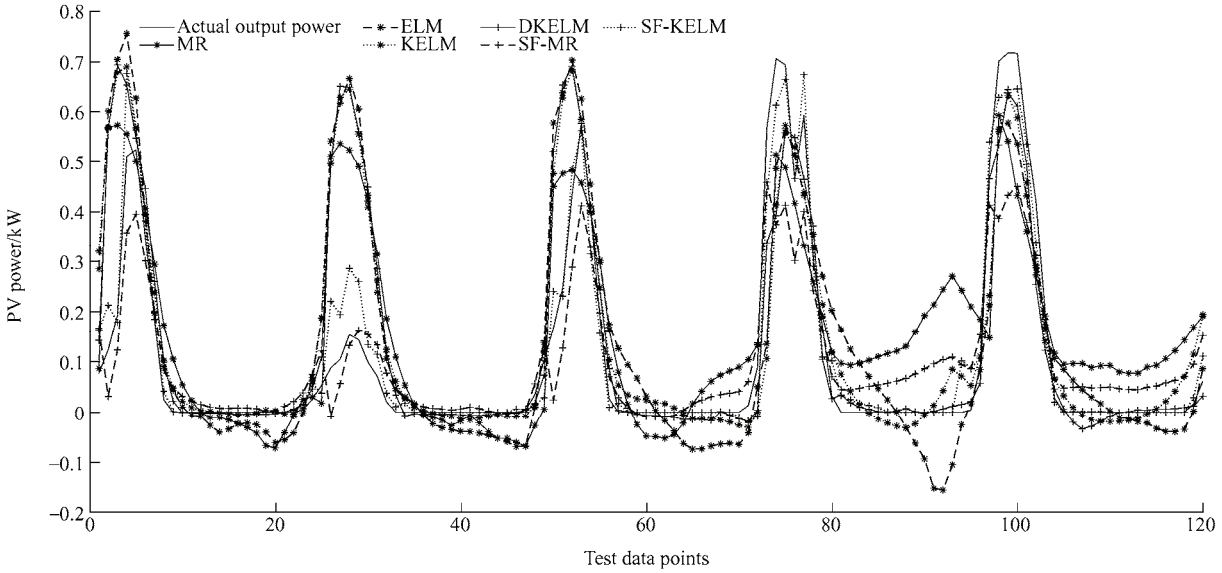

Actual output power
GASVM
SF-KELM
WPD-LSTM
MP
CSO-RBFNN
SF-MR
PV power/kW
Test data points

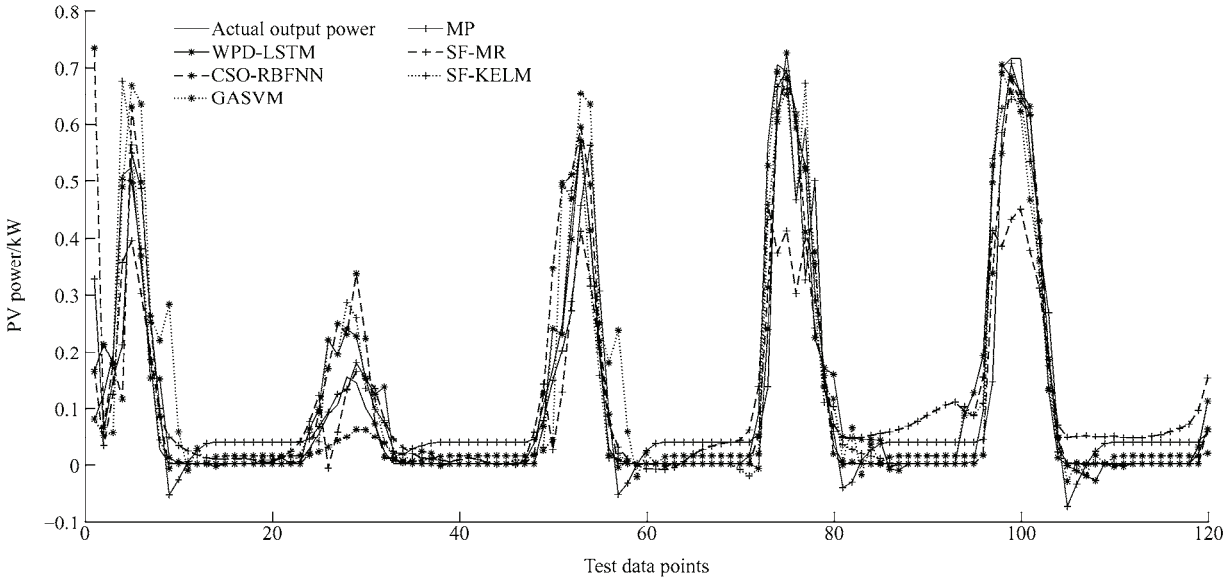
(b) Summer


Actual output power
KELM
SF-KELM
MR
DKELM
ELM
SF-MR
PV power/kW
Test data points

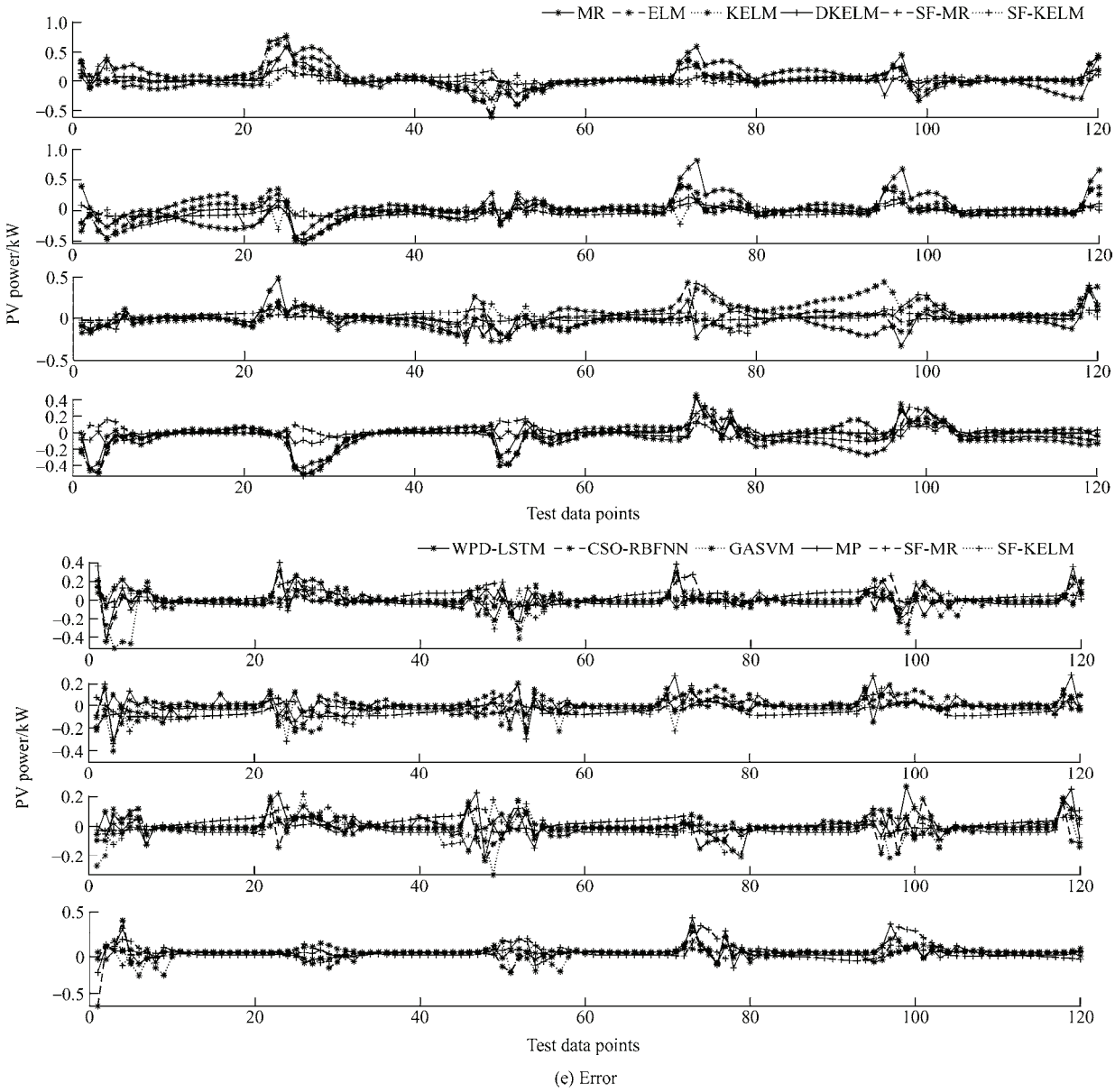


(c) Autumn





(d) Winter



(e) Error

Fig. 8 Forecasting and error curves of different methods collected in different seasons

5 Conclusions

In this study, to improve the forecasting accuracy of short-term PV power generation, regional models based on an evaluable Two-Level algorithm are proposed. Under the unified framework of regional modeling, firstly, the SOM is adopted to complete the initial partitions of the training dataset, followed by the clustering of the trained SOM network with an evaluable FCM. Finally, in each data partition, regional SF-KELM or SF-MR models are constructed.

The proposed models are applied to one-hour-ahead PV power generation forecasting for different plants derived from the GEFCom2014. Moreover, MR, ELM, and KELM are used as benchmark methods. Beyond that, to highlight the effectiveness of the Two-Level algorithm and direct clustering method, the forecasts based on direct FCM are conducted. From the results obtained with comparison charts and error diagrams, it is concluded that the SF-KELM method has achieved the highest accuracy. For future research directions, we will focus on combining multiple kernel SOM and other more competitive forecasting methods and processing the uncertainty in PV power generation.

References

[1] S Sobri, S Koohi-Kamali, N A Rahim. Solar photovoltaic generation forecasting methods: A review. *Energy*

Conversion and Management, 2018, 156(1): 459-497.

[2] X Yang, M Xu, S Xu, et al. Day-ahead forecasting of photovoltaic output power with similar cloud space fusion based on incomplete historical data mining. *Applied Energy*, 2017, 206: 683-696.

[3] M Pan, C Li, R Gao, et al. Photovoltaic power forecasting based on a support vector machine with improved ant colony optimization. *Journal of Cleaner Production*, 2020, 277: 123948.

[4] 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*, 2018, 81(1): 912-928.

[5] L Wang, O Kisi, M Zounemat-Kermani, et al. Solar radiation prediction using different techniques: Model evaluation and comparison. *Renewable and Sustainable Energy Reviews*, 2016, 61: 384-397.

[6] F Almonacid, P J Pérez-Higueras, E F Fernández, et al. A methodology based on dynamic artificial neural network for short-term forecasting of the power output of a PV generator. *Energy Conversion and Management*, 2014, 85: 389-398.

[7] A Dolara, F Grimaccia, S Leva, et al. A physical hybrid artificial neural network for short term forecasting of PV plant power output. *Energies*, 2015, 8(2): 1138-1153.

[8] P Li, K Zhou, X Lu, et al. A hybrid deep learning model for short-term PV power forecasting. *Applied Energy*, 2020, 259: 114216.

[9] Z Yang, M Mourshed, K Liu, et al. A novel competitive swarm optimized RBF neural network model for short-term solar power generation forecasting. *Neurocomputing*, 2020, 397: 415-421.

[10] W Vandeventer, E Jamei, G S Thirunavukkarasu, et al. Short-term PV power forecasting using hybrid GASVM technique. *Renewable Energy*, 2019, 140: 367-379.

[11] M Seyedmahmoudian, E Jamei, G S Thirunavukkarasu, et al. Short-term forecasting of the output power of a building-integrated photovoltaic system using a metaheuristic approach. *Energies*, 2018, 11(5): 1260.

[12] A Lendasse, J Lee, V Wertz, et al. Forecasting electricity consumption using nonlinear projection and self-organizing maps. *Neurocomputing*, 2002, 48(1-4): 299-311.

[13] G A Barreto, A F R Araujo. Identification and control of dynamical systems using the self-organizing map. *IEEE Transactions on Neural Networks*, 2004, 15(5): 1244-1259.

[14] G Simon, A Lendasse, M Cottrell, et al. Time series forecasting: Obtaining long term trends with self-organizing maps. *Pattern Recognition Letters*, 2005, 26(12): 1795-1808.

[15] A H S Júnior, G A Barreto, F Corona. Regional models: A new approach for nonlinear system identification via clustering of the self-organizing map. *Neurocomputing*, 2015, 147: 31-46.

[16] W Fu, K Zhang, K Wang, et al. A hybrid approach for multi-step wind speed forecasting based on two-layer decomposition, improved hybrid DE-HHO optimization and KELM. *Renewable Energy*, 2021, 164: 211-229.

[17] R Ye, Q Dai. MultiTL-KELM: A multi-task learning algorithm for multi-step-ahead time series prediction. *Applied Soft Computing*, 2019, 79: 227-253.

[18] L Xiao, W Shao, F Jin, et al. A self-adaptive kernel extreme learning machine for short-term wind speed forecasting. *Applied Soft Computing*, 2021, 99: 106917.

[19] J Vesanto, E Alhoniemi. Clustering of the self-organizing map. *IEEE Transactions on neural networks*, 2000, 11(3): 586-600.

[20] G B Huang, H Zhou, X Ding, et al. Extreme learning machine for regression and multiclass classification. *IEEE Transactions on Systems*, *Man*, *and Cybernetics*, *Part B (Cybernetics)*, 2011, 42(2): 513-529.

[21] P Andras. Kernel-kohonen networks. *International Journal of Neural Systems*, 2002, 12(2): 117-135.

[22] B Yan, N Xu, L P Xu, et al. An improved partitioning algorithm based on FCM algorithm for extended target tracking in PHD filter. *Digital Signal Processing*, 2019, 90: 54-70.

[23] M Halkidi, Y Batistakis, M Vazirgiannis. On clustering validation techniques. *Journal of Intelligent Information Systems*, 2001, 17(2): 107-145.

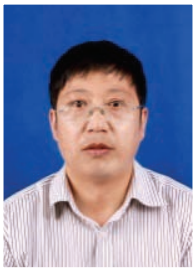
[24] T Hong, P Pinson, S Fan, et al. Probabilistic energy forecasting: Global energy forecasting competition 2014 and beyond. *International Journal of Forecasting*, 2016, 32(3): 896-913.

[25] G A Barreto. Time series prediction with the self-organizing map: A review. *Perspectives of Neural-symbolic Integration*, 2007: 135-158.

[26] J Li, M Li. Prediction of ultra-short-term wind power based on BBO-KELM method. *Journal of Renewable and Sustainable Energy*, 2019, 11(5): 056104.

[27] T Matias, F Souza, R Araújo, et al. Learning of a single-hidden layer feedforward neural network using an optimized extreme learning machine. *Neurocomputing*, 2014, 129: 428-436.

[28] P Zhu, W Zhu, Q Hu, et al. Subspace clustering guided unsupervised feature selection. *Pattern Recognition*, 2017, 66: 364-374.



Jun Li received his B.S. and M.S. degrees in Traffic Information Engineering and Control from the Lanzhou Jiaotong University, Lanzhou, China in 1991, 1999, respectively. Later, he received a Ph.D. in Instrument Science and Technology from Xi'an Jiaotong University, Xi'an, China, in 2006. He is currently a Professor at the School of Automation and Electrical Engineering at Lanzhou Jiaotong University. His current research interests include theory and applications of nonlinear system modeling and prediction and renewable engineering forecasting based on artificial intelligence and machine learning methods.

Qibo Liu received his B.S. and M.S. degrees in Automation and Control Engineering from Lanzhou Jiaotong University, Lanzhou, China, in 2018, 2022, respectively. He is currently pursuing a Ph.D. in Control Engineering at Shanghai Jiaotong University. His current research interests include the theory and applications of nonlinear system modeling and model predictive control based on artificial intelligence and machine learning methods.