Strathprints Institutional Repository

Clustering methods of wind turbines and its application in short-term wind power forecasts

Yongqian Liu,[1] Xiaoli Gao,[1,a)] Jie Yan,[1] Shuang Han,[1] and David G. Infield[2]
[1]*State Key Laboratory of Alternate Electrical Power System with Renewable Energy Sources, North China Electric Power University, Changping District, Beijing 102206, China*
[2]*Department of Electronic and Electrical Engineering, University of Strathclyde, Glasgow, United Kingdom*

Commonly used wind power forecasts methods choose only one representative wind turbine to forecast the output power of the entire wind farm; however, this approach may reduce the forecasting accuracy. If each wind turbine in a wind farm is forecasted individually, this considerably increases the computational cost, especially for a large wind farm. In this work, a compromise approach is developed where the turbines in the wind farm are clustered and a forecast made for each cluster. Three clustering methods are evaluated: K-means; a self-organizing map (SOM); and spectral clustering (SC). At first, wind turbines in a wind farm are clustered into several groups by identifying similar characteristics of wind speed and output power. Sihouette coefficient and Hopkins statistics indices are adopted to determine the optimal cluster number which is an important parameter in cluster analysis. Next, forecasting models of the selected representative wind turbines for each cluster based on correlation analysis are established separately. A comparative study of the forecast effect is carried to determine the most effective clustering method. Results show that the short-term wind power forecasting on the basis of SOM and SC clustering are effective to forecast the output power of the entire wind farm with better accuracy, respectively, 1.67% and 1.43% than the forecasts using a single wind speed or power to represent the wind farm. Both Hopkins statistics and Sihouette coefficient are effective in choosing the optimal number of clusters. In addition, SOM with its higher forecast accuracy and SC with more efficient calculation when applied into wind power forecasts can provide guidance for the operating and dispatching of wind power. The emphasis of the paper is on the clustering methods and its effect applied in wind power forecasts but not the forecasting algorithms. © 2014 AIP Publishing LLC. [http://dx.doi.org/10.1063/1.4898361]

I. INTRODUCTION

As the increasingly serious problems of fossil energy depletion, environmental pollution, and the demand for electrical energy, it is urgent to exploit the renewable energy. Wind energy is the fastest growing renewable energy source and is now playing an important role in electricity supply.[1,2]

However, the intermittence and variability of wind power will bring great challenges to the stability and power quality of the grid. In order to help solve these problems, accurate and reliable wind power forecasts models should be developed.[3] Many researchers have made significant efforts to develop and optimize algorithms for wind power forecasting. Commonly used forecast methods target a single wind turbine or choose only one representative wind turbine (RWT) to forecast the entire output power of a wind farm. Artificial neural networks in combination with the wavelet transform have been proposed for short-term wind power forecasts.[4] Two neural network algorithms were studied to predict wind power which were radial basis function (RBF) and back propagation.[5]

---

a)Present address: School of Renewable Energy, North China Electric Power University (NCEPU), Beijing 102206, China.

Nevertheless, the characteristics of the output power of the entire wind farm cannot be effectively represented by only one wind turbine. If each wind turbine is modelled separately to forecast wind power, this will lead to huge computational costs.[6] In order to reduce the risk associated with use of a single wind turbine, as well as ensure the high computational efficiency, clustering methods are adopted to cluster wind turbines of a wind farm into different groups by identifying the similarity between them. Forecast models are then established for each cluster, improving forecast accuracy. A novel methodology was proposed to cluster wind turbines of a wind farm into different groups based on a particular distance measure.[7] A grouping mechanism was established by optimized relevance vector machine to divide wind turbines into several groups and tested the validity of the proposed model.[8] However, there is limited research on clustering methods applied in wind power forecasts; especially there is less research for their comparison, let alone the methods to determine the optimal cluster number.

In this paper, a compromise approach of short-term wind power forecasts based on three clustering methods, K-means, self-organizing map (SOM), and spectral clustering (SC), are employed to cluster wind turbines into groups by capturing the similarity of wind speed and wind power, decreasing the forecast error by smoothing effects of multi-clusters. Sihouette coefficient and Hopkins statistics indices are used to determine the optimal cluster number. The forecasting wind information like speed and direction are usually employed to forecast the output power. However, there is no available matured numerical weather prediction (NWP) data. So this work employs the real wind speed in replace of the NWP data and historical output power of each wind turbine in a wind farm for forecasting. Comparing three models, the better model can be used to provide scientific guidance for the operation and dispatch of wind farms.

This paper has a five-section structure. Section I describes the published work and the general content of this paper. Section II describes the principles of the proposed clustering methods. Section III clusters a wind farm into groups with K-means, SOM, and SC, then determines the optimal cluster number by Sihouette coefficient and Hopkins statistics. Section IV uses the clustering results to forecast the output power of the entire wind farm, as well as studies an influence of the cluster numbers on forecast accuracy and compares three clustering methods. Section V makes the final conclusions.

II. PRINCIPLE OF PROPOSED CLUSTERING METHODS

The clustering methods employed for clustering the wind turbines of a wind farm into different groups are introduced below in brief.

A. K-means method

K-means clustering (MacQueen, 1967) is an unsupervised learning algorithm which is popular in cluster analysis. K-means method aims to partition a given data into k groups.[9,10]

Given a set of n data points $(x_1, x_2,\ldots, x_n)$, where each point is a d-dimensional real vector, K-means clustering is designed to minimize the within-cluster sum of squares. The objective function is written as[11]

$$V = \sum_{j=1}^{k} \sum_{i=1}^{n_j} \|x_i^{(j)} - c_j\|^2,$$ (1)

where $\|x_i^{(j)} - c_j\|^2$ is a distance measure between the data point $x_i^{(j)}$ and the cluster centre c_j; k is the number of clusters, n_j is the number of data points in j cluster. For the numeric features of data from wind farm, in this paper, the Euclidean metric is chosen as the distance measure. K-means selects k initial cluster centres at random from the n data points, so that it may lead to different clustering results. Another issue is how to choose the number of clusters.

B. Self-organizing map

SOM (Kohonen, 1982) is a widely used artificial neural network that is based on unsupervised learning to produce a low-dimensional map from high-dimensional data. SOM includes

an input layer and an output layer which is referred to competitive layer or Kohonen layer, both containing several nodes. Each node of the input layer is linked to that of the output layer by weight.[12,13]

Given a set of n data points $X = (x_1, x_2,\ldots, x_n)$, where each point is a d-dimensional feature vector, the output layer has m nodes organised in a lattice that is commonly one- or two-dimensional. The process of the SOM algorithm is as follows:

(1) Network initialization. The weights $\{w_j, j = 1, 2,\ldots, m\}$ from the n input nodes to the m output nodes are initialized randomly.
(2) Input vector. The input vector $X = (x_1, x_2,\ldots, x_n)^T$ is input to the input layer.
(3) The distance calculation. The distance from w_j to input vector X is calculated by using Euclidean metric. Seek out the winning output node ω with the least distance as

$$\omega = \arg\min_j \|X - w_j\|, \quad j = 1, 2, \cdots, m. \tag{2}$$

(4) The learning of weights. The weight vectors are updated by the following equation:

$$w_j(t + 1) = w_j(t) + \eta(t)\|X(t) - w_j(t)\|, \tag{3}$$

where $\eta(t)$ is the learning rate of the t iteration, it gradually decreases to 0 as t increases. The function is described as

$$\eta(t) = \frac{1}{t} \text{ or } \eta(t) = 0.2\left(1 - \frac{t}{10000}\right). \tag{4}$$

(5) Compute the output. The output o_k is continuously computed according to Eq. (5) by using the updated weights until meeting the requirement of iteration $t \geq t_{max}$,

$$o_k = f(\min_j \|X - w_j\|). \tag{5}$$

When t_{max} is set as 100–200, there is no noticeable change in the feature map. The output of SOM is k winning output nodes ($k \leq m$) which are the clustering results.

C. Spectral clustering

Spectral clustering (Fiedler, 1973), written in SC, is newly developing technique in recent years. Unlike the traditional clustering algorithms, SC is based on spectral graph theory to solve the clustering of non-convex sphere of sample spaces, and converged to global optimal solution.[14]

The basic idea of the SC is to make use of the spectrum of the similarity matrix which can capture the nonlinear and low-dimensional manifold structure of the data to get a clustering result of the data points.[15]

Given a set of n data points $(x_1, x_2,\ldots, x_n)$, where each point is a d-dimensional real vector, these vectors are clustered into k groups. The process of SC algorithm is as follows:[16]

(1) From the similarity matrix $A \in R^{n \times n}$, which is a symmetric matrix, A_{ij} denotes the weights of connecting the i point to the j point. A_{ij} is denoted as

$$A_{ij} = \begin{cases} \exp\left(-\dfrac{\|x_i - x_j\|^2}{\sigma^2}\right) & i \neq j \\ 0 & i = j, \end{cases} \tag{6}$$

where σ is the parameter to control the spread of neighbours.
(2) The Laplacian matrix is defined as $L = D^{-1/2}AD^{-1/2}$, where D is a diagonal matrix as below

$$D_{ii} = \sum_j A_{ij}. \qquad (7)$$

(3) Find the k largest eigenvectors of L, which are orthogonal to each other, to form the matrix $C = (c_1, c_2,\ldots, c_k) \in \mathrm{R}^{n \times k}$ by stacking the eigenvectors in columns.

(4) Form the normalized matrix Y from C according to the following equation:

$$Y_{ij} = \frac{C_{ij}}{\sqrt{\sum_j C_{ij}^1}}. \qquad (8)$$

(5) In the k-dimensional space, taking each row of Y as a vector, K-means or any other algorithm can be employed to cluster these vectors into k clusters.

(6) Assign the data point x_i to cluster j only if row i of the Y is assigned to cluster j.

As SC algorithm captures the main feature of the data rather than the minor one, it has strong robustness that is not sensitive to the irregular data.

III. CLUSTER ANALYSIS OF WIND TURBINES IN A WIND FARM

A. Clustering models structure

The structures of three proposed clustering models (K-means, SOM, and SC) are all illustrated in Fig. 1, containing selection of input samples, normalization processing, clustering of wind turbines by the employed methods and determination of the optimal cluster number.

B. Clustering validity indices

In cluster analysis, the key to the quality of clustering is to choose the optimal cluster number. For K-means, SOM, and SC, the cluster number is a vital parameter. The cluster number should be chosen appropriately, considering not only clustering the wind turbines as effectively as possible but also ensuring higher computational efficiency of the wind power forecasts in Sec. IV. The ideal clustering result has the feature of the smallest distance within clusters and the largest distance among different clusters. There are many clustering validity indices to
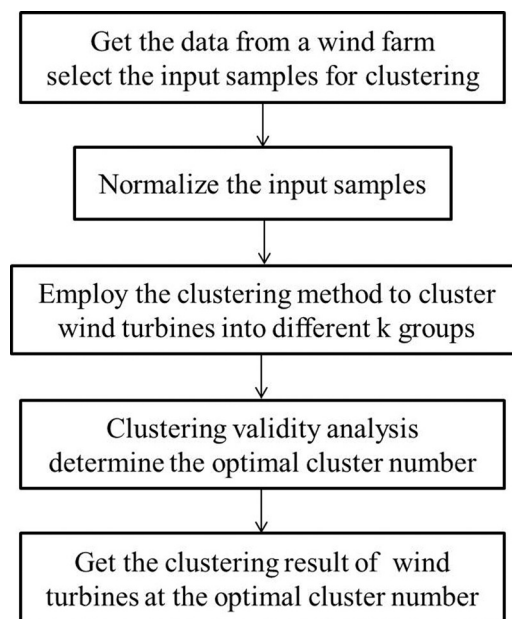

FIG. 1. The structure of three clustering models.

assess the validity of clustering as well as make a choice of the optimal cluster number, such as Calinski-Harabasz index,[17] Davies-Bouldin index,[18] and Sihouette coefficient index.[19] In this paper, two clustering validity indices are used to evaluate the quality of clustering and determine the optimal cluster number, which are Sihouette coefficient (written as S) and Hopkins statistics (written as H).[20,21]

The Sihouette coefficient in d-dimensions is defined as

$$s_i = \frac{b_i - a_i}{\max[a_i, b_i]}, \tag{9}$$

$$a_i = \frac{1}{n_c - 1} \sum_{i,j \in C_c, i \neq j} d(i,j), \tag{10}$$

$$b_i = \min_{p, p \neq c} \left[\frac{1}{n_p} \sum_{i \in C_c, j \in C_p} d(i,j) \right], \tag{11}$$

where n is the number of samples, c is the number of clusters, a_i is the average distance from i sample to j ($j \neq i$) samples within the same C_c cluster, b_i is the minimum average value from i sample within the C_c cluster to j samples with the C_p cluster which i does not belong to. $s_i \in [-1, 1]$, when it is closer to 1, the clustering result is more effective.

Given a set of n data points $X = \{x_i, i = 1 \text{ to } n\}$, where each point is a d-dimensional vector, the Hopkins statistics is defined as

$$H = \frac{\sum_{j=1}^{m} u_j}{\sum_{j=1}^{m} u_j + \sum_{j=1}^{m} w_j}, \tag{12}$$

$$u_j = \min_{v \in X}\{dist(y_i, v)\}, \tag{13}$$

$$w_j = \min_{v \in X, v \neq y_i}\{dist(z_j, v)\}, \tag{14}$$

where $Y = \{y_j, j = 1 \text{ to } m\}$ is m sampling points randomly chosen from the d-dimensional sampling window ($m \ll n$), u_j is the minimum distance from y_j to its nearest point in X, and w_j is the minimum distance from a randomly selected point z_j in X to its nearest point in X, but not Y. When u_j is the same as w_j (H is about 0.5), it implies the data points are evenly distributed. However, if the points are unevenly distributed and can be clustered into different groups, H is supposed to be larger or much less than 0.5. Therefore, the optimal cluster number can be chosen by comparing the average H of each cluster at the k cluster number.

C. Cluster analysis of an example

1. Data

The data are from a wind farm in north China only including mean wind speed and wind power of each wind turbine from SCADA. The wind farm has 33×1.5 MW wind turbines. All the data points are at 2 h intervals covering from January 4th to September 15th in 2011. The total number of the time-series of wind speed and output power is 3004.

2. Data preprocessing

In order to ensure the accuracy and efficiency of the clustering models, the primitive data needs to be preprocessed to guarantee its validity. The improper wind speed which caused by malfunction of the anemoscope and the missing or improper output power which is normal or

abnormal stop of 33 wind turbines is eliminated to make wind speed and output power have a one-to-one correspondence in time-series. After the process of deletion, 10.5% of the primitive data has been removed.

Wind power output from each wind turbine is affected by wind speed, wind direction, temperature, pressure, and other factors. While wind speed is the main factor affecting power output alone it will not allow the best forecasts of power. Therefore, wind power as well as wind speed are taken as the inputs for clustering. It is difficult for clustering methods to capture the similarity and major features of input data with high dimensions and a wide range. In order to diminish the difficulty, the input samples are reduced into low-dimensional ones by the Eqs. (15)–(18), as well as normalized within the range between 0 and 1 via Eqs. (19)–(22),[22] minimizing the difficulty of clustering and maintaining correlation among the data set,

$$p_{mean}(i) = \frac{\sum_{j=1}^{n} p_j(i)}{n}, \tag{15}$$

$$v_{mean}(i) = \frac{\sum_{j=1}^{n} v_j(i)}{n}, \tag{16}$$

$$p_{sd}(i) = \sqrt{\frac{1}{n} \sum_{j=1}^{n} \left[p_j(i) - p_{mean}(i) \right]^2}, \tag{17}$$

$$v_{sd}(i) = \sqrt{\frac{1}{n} \sum_{j=1}^{n} \left[v_j(i) - v_{mean}(i) \right]^2}, \tag{18}$$

where $v_j(i)$ and $p_j(i)$ indicate the j wind speed and power of i wind turbine, n is the number of data points. $p_{mean}(i)$ and $v_{mean}(i)$ express average wind power and speed of i wind turbine, $p_{sd}(i)$ and $v_{sd}(i)$ represent standard deviation of wind power, and speed of i wind turbine,

$$p_{mean1}(i) = \frac{p_{mean}(i) - p_{mean}(\min)}{p_{mean}(\max) - p_{mean}(\min)}, \tag{19}$$

$$p_{sd2}(i) = \frac{p_{sd}(i) - p_{sd}(\min)}{p_{sd}(\max) - p_{sd}(\min)}, \tag{20}$$

$$v_{mean3}(i) = \frac{v_{mean}(i) - v_{mean}(\min)}{v_{mean}(\max) - v_{mean}(\min)}, \tag{21}$$

$$v_{sd4}(i) = \frac{v_{sd}(i) - v_{sd}(\min)}{v_{sd}(\max) - v_{sd}(\min)}, \tag{22}$$

where $p_{mean1}(i)$, $p_{sd2}(i)$, $v_{mean3}(i)$, and $v_{sd4}(i)$ are the normalized input samples, $p_{mean}(\min)$, $p_{mean}(\max)$ and $p_{sd}(\min)$, $p_{sd}(\max)$ are the minimum and maximum of p_{mean} and p_{sd}, so are $v_{mean}(\min)$ $v_{means}(\max)$ and $v_{sd}(\min)$ $v_{sd}(\max)$.

Then, the input samples matrix $X \in R^{33 \times 4}$ is assigned as

$$X = \begin{bmatrix} p_{mean1}(1) & p_{sd2}(1) & v_{mean3}(1) & v_{sd4}(1) \\ p_{mean1}(2) & p_{sd2}(2) & v_{mean3}(2) & v_{sd4}(2) \\ \cdots & \cdots & \cdots & \cdots \\ p_{mean1}(33) & p_{sd2}(33) & v_{mean3}(33) & v_{sd4}(33) \end{bmatrix}.$$
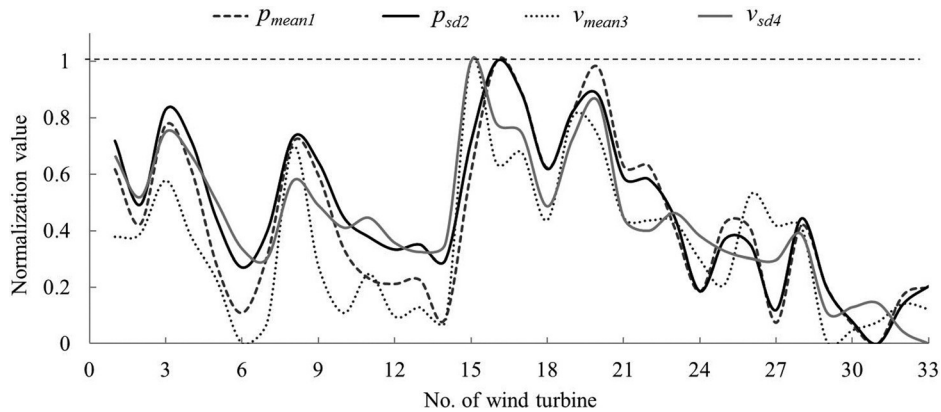
FIG. 2. The preprocessed input data of 33 wind turbines.

The input data points of 33 wind turbines after preprocessing are given in Fig. 2. Four groups of data have the similar trend from 1 to 33 wind turbine. This implies the wind speed and output power samples with large mean value have large standard deviation.

3. Clustering validity results

Due to the variability of K-means and SOM clustering results, the clustering result with the maximum S in 10 iterative computations is chosen at the k number of clusters. The proposed clustering models are carried out to cluster the wind turbines, respectively, at 2 to 8 cluster numbers.

The average Sihouette coefficient (S) and Hopkins statistics (H) plotted against cluster number are shown in Figs. 3(a) and 3(b). As can be seen in Fig. 3(a), the average S of three clustering methods is basically increasing with number of clusters. When k is more than 5, the average S grows very slowly or even decreases a bit, mainly because the real cluster number is less than the set value. For one clustering method and the same samples, clustering results with larger cluster number have the higher average S, which means the clustering results are more reasonable. However, too high cluster number will cause more computational costs during the cluster and forecast modelling. Hence, as for this index, balancing clustering effect and computational efficiency, $k = 5$ may be the optimal cluster number.

Fig. 3(b) shows the average H falls as cluster number (k) rises. When k varies from 1 to 5, the decline from 0.72 to 0.5 or so of the average H is because the distribution of the data within the same cluster becomes more symmetrical and random. When k is more than 5, average H decreases less than 0.5. As all clusters with the average H of 0.5 at k cluster number are evenly distributed, $H = 0.5$ is taken as the requirement of selecting the optimal cluster number. For K-means, $k = 5$ (average $H = 0.53$) is closer to the requirement of $H = 0.5$ than $k = 6$ (average $H = 0.41$). For SOM, $k = 5$ (average $H = 0.50$) corresponds to the requirement. For SC, $k = 5$ (average $H = 0.51$) is closer to the requirement than $k = 6$ (average $H = 0.46$). Hence, $k = 5$ could be taken as the partition to identify the natural clusters in the data sets.

It can be seen that it is much easier to determine the optimal cluster number quantitatively by Hopkins statistics index than by Sihouette coefficient one. Combining the two indices, the optimal cluster number of 33 wind turbines is chosen as 5.

4. Clustering of wind turbines results

The clustering results of three clustering methods are shown in Table I. It can be seen that every cluster of three methods has more than four wind turbines, avoiding the situation of the cluster with one isolated wind turbine. The wind turbines in the first and fifth cluster of three methods are identical. While in other clusters, there are some differences among the methods, indicating the similarities of the corresponding wind turbines are a little obscure and not easy to identify.
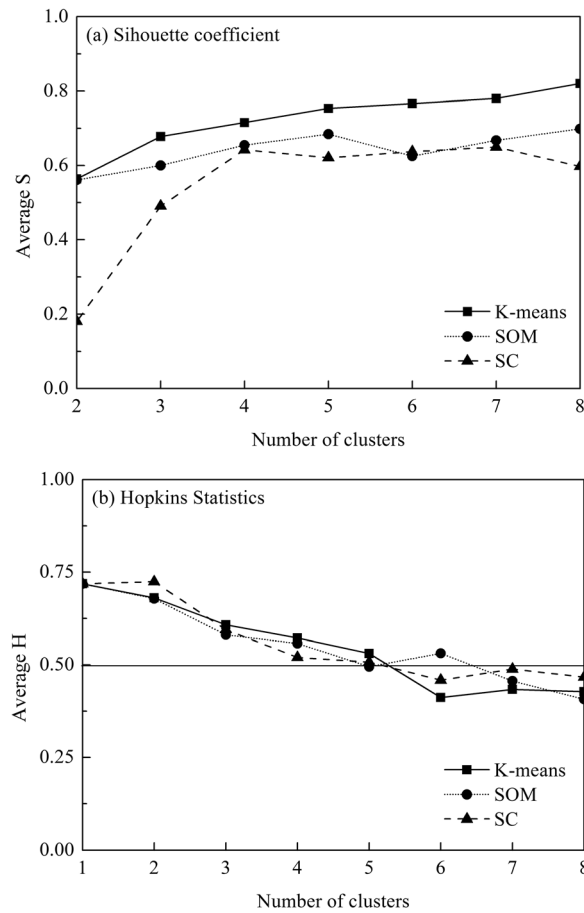
FIG. 3. Average *S* and *H* plotted against number of clusters.

For each cluster, the RWT is selected by wind power correlation coefficients of wind turbines in the same cluster. For example, one cluster has five wind turbines which are 29#, 30#, 31#, 32#, and 33#. The correlation coefficients between five turbines' output power are shown in Table II. 32# wind turbine with the highest average correlation coefficients (it is 0.966) is chosen to represent the cluster.

IV. SHORT-TERM WIND POWER FORECASTS WITH CLUSTERING METHODS

A. The fundamentals of radial basis function

In this study, the RBF neural network is used to conduct the short-term wind power forecasts 72 h in advance. RBF neural network is a three-layer feed forward network, including the input layer, hidden layer, and output layer.[23] RBF can give a reasonable solution for such a

TABLE I. Clustering results of K-means, SOM, and SC.

Cluster	K-means		SOM		SC	
	WT in the cluster	RWT	WT in the cluster	RWT	WT in the cluster	RWT
1	1,4,9,18,21,22	1	1,4,9,18,21,22	1	1,4,9,18,21,22	1
2	2,5,23,26,28	23	2,5,23,25,26,28	25	2,23,24,26–28	23
3	6,7,10–14,24,25	10	7,10–13,24,27	10	5–7,10–14,25	7
4	27,29–33	32	6,14,29–33	29	29–33	32
5	3,8,15–17,19,20	16	3,8,15–17,19,20	16	3,8,15–17,19,20	16

TABLE II. Wind power correlation coefficients in the cluster. The boldface value shows the highest average correlation coefficient.

WT	29#	30#	31#	**32#**	33#
29#	1				
30#	0.974	1			
31#	0.968	0.961	1		
32#	0.962	0.964	0.965	1	
33#	0.945	0.944	0.949	0.972	1
Average	0.962	0.961	0.961	**0.966**	0.952

highly nonlinear problem as wind power forecasts due to its structure characterized by a combination of non-supervised (in the hidden layer) and supervised (in the output layer) training.[24]

Given an input vector $x \in R^n$, the output of the network $f(x)$ is given by

$$f(x) = \sum_{i=1}^{N} w_i G(\|x - c_i\|),$$

(23)

where N is the number of neurons in the hidden layer, c_i is the cluster centre for neuron i, and w_i is the weight between hidden layer and output layer. $G(\|x - c_i\|)$ is Gaussian function, it is expressed by

$$G(\|x - c_i\|) = \exp\left(-\frac{\|x - c_i\|^2}{2\sigma_i^2}\right),$$

(24)

where σ_i is the width of nonlinear transforming unit.

B. Evaluation indices

Two commonly used statistical error evaluation indices are employed to judge the performance of the wind power forecasts based on K-means, SOM, and SC methods. Root mean square error (RMSE) in Eq. (25) is calculated for all validation period, which can give a better evaluation of prediction error.[8,25] Mean absolute error (MAR) is another used error criteria in Eq. (26),

$$RMSE = \frac{\sqrt{\sum_{i=1}^{n} (P_{ai} - P_{pi})^2}}{Cap \times \sqrt{n}},$$

(25)

$$MAE = \frac{\sum_{i=1}^{n} |P_{ai} - P_{pi}|}{Cap \times n},$$

(26)

where P_{ai} and P_{pi}, respectively, signify the actual and forecast value of wind power at i time; Cap indicates the installed capacity of a wind farm; n is the number of forecast samples involved.

C. Case study

1. Selection of training samples

With the same data in Sec. III, considering lack of the numerical weather prediction information at the spot of each wind turbine as well as wind direction plus other factors of each wind turbine collected from SCADA, historical wind speed data of the representative wind turbine in each cluster is the main input data to the forecasting network, while the total output

power of wind turbines in the same cluster is taken as training target. 80% of the available wind speed and power data are considered as training samples, and 20% as test samples, to evaluate the accuracy of short-term wind power forecasts with the proposed clustering methods.

2. Verification results of optimal cluster number

The wind power forecast errors (RMSE and MAE) on the basis of three clustering methods plotted against cluster number are shown in Figs. 4(a) and 4(b). The red lines in Figs. 4(a) and 4(b) represent the RMSE and MAE values of forecast power without clustering methods. The black curves are beneath the red lines, shown that the proposed clustering methods are effective to improve wind power forecasts. Both RMSE and MAE values basically reduce with the increase of cluster number, it proves that cluster number is an important parameter to forecast accuracy. At 2 to 5 cluster numbers, the downward trend of RMSE and MAE is obvious, while it gets to 6, 7, or even larger, the improvements of forecast accuracy is much smaller.

As a result, the cluster number of 5 in this case can ensure better forecast accuracy as well as lower computational costs caused by too large cluster number. Meanwhile, it verifies the optimal cluster number got by Sihouette coefficient and Hopkins statistics is feasible.

3. Comparison results of three clustering methods

The wind power forecast errors of three clustering methods at optimal cluster number of 5 are shown in Table III. Non-clustering represents wind power forecasts by using an optimal wind turbine of all the ones. The results show that owing to the smoothing effects of multiple
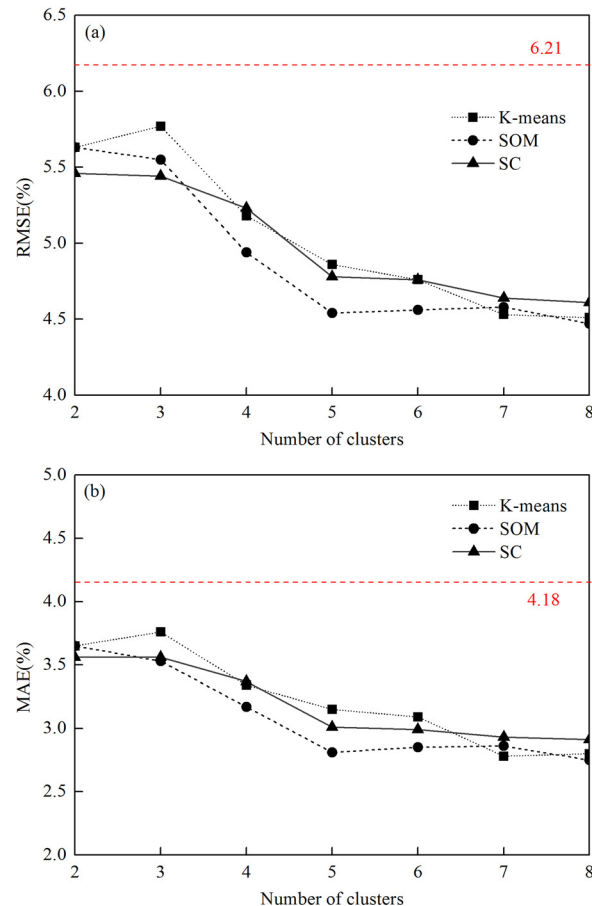

FIG. 4. Forecast errors plotted against number of clusters.

TABLE III. Forecast errors comparison of non-clustering and clustering methods.

Methods	RMSE(%)	MAE(%)	Runtime(s)
Non-clustering	6.21	4.18	—
K-means	4.86	3.15	18.98
SOM	4.54	2.81	47.20
SC	4.78	3.01	0.11

clusters than only one wind turbine, the forecasts with the proposed clustering methods all have better forecast accuracy than non-clustering, improving the RMSE of forecast values by $(6.21\%{-}4.86\%) = 1.35\%$ with K-means, $(6.21\%{-}4.54\%) = 1.67\%$ with SOM, and $(6.21\%{-}4.78\%) = 1.43\%$ with SC.

Comparing the K-means, SOM, and SC models, the forecast with SOM by the RMSE 4.54% performs better than K-means 4.86% and SC 4.78% because of its strong capacity resistant to the interference of input data by neural nodes.

The real and forecast values of output power for 5 days are shown in Fig. 5. Forecasts with three clustering models have more accurate forecast value which is more approximate to real value by smoothing effect than non-clustering. Figs. 6(a)–6(d) show the relative error probability distribution of forecast value and real value of output power. The shapes of power relative error probability distribution of three clustering models are tighter and higher than non-clustering model. It shows K-means with 80.1% power relative error between −5% and 5%, SOM with 84.1% and SC with 81.1% have effectively decreased the forecast errors in contrast to the commonly used forecast method (non-clustering) with 69.2%. In addition, the improvement of SOM is more significant than the others.

The 4th row of Table II shows the runtime of the clustering process of three models. The runtime of K-means and SOM clustering is more than SC, because SC clustering result is constant, while K-means and SOM clustering results are changeable. It is necessary for K-means and SOM to iterate 10 times to select the best clustering result, that leading to longer runtime. Hence, SC is more suitable for clustering large-scale wind farms because of its stability of clustering result. While the iteration clustering of K-means and SOM may run for a longer time, largely reducing the effectiveness.

As a result, short-term wind power forecasts on the basis of SOM and SC methods are both effective to lower the risk of forecast errors and can supply more scientific guidance for
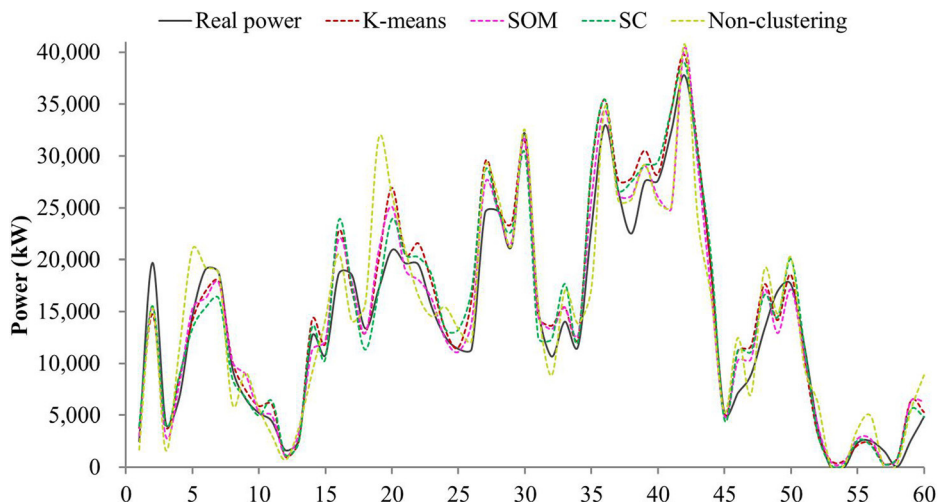

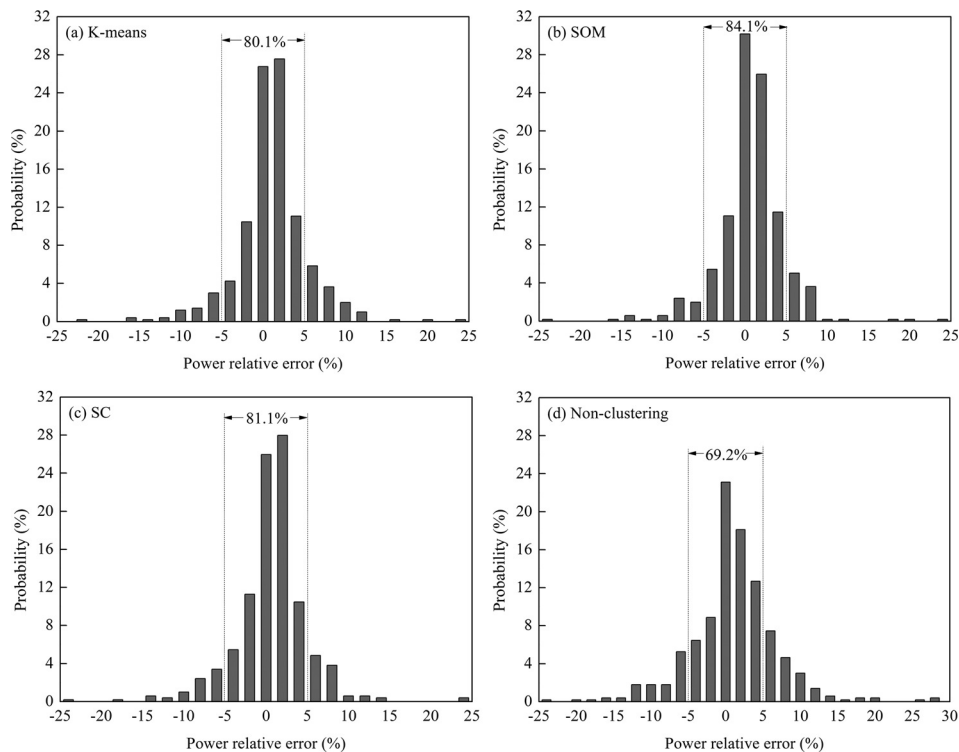FIG. 5. The real and forecast values of output power for 5 days.

FIG. 6. Relative error probability distribution of forecast and real power of (a) K-means, (b) SOM, (c) SC, and (d) non-clustering models.

the operating and dispatching of wind power. It is effective and feasible to use Sihouette coefficient and Hopkins statistics to select the optimal cluster number which is important to power forecast effect, especially Hopkins statistics, which can quantitatively make a choice of the number.

V. CONCLUSION

This paper employs three clustering methods (K-means, SOM, and SC) for improving short-term wind power forecasts. The following conclusions can be drawn:

(1) The clustering models of K-means, SOM, and SC are effective in generating clusters of wind turbines within a wind farm. Using the clusters can improve the short-term wind power forecast accuracy, respectively, by 1.35%, 1.67%, and 1.43% in relation to the non-clustering analysis. Among the proposed methods, SOM, and SC methods are the most effective in terms of power forecast accuracy. For forecast accuracy, SOM model is best, while for computational efficiency, SC model is best.
(2) For the adaptability of clustering models, SOM is more suitable for clustering the medium-scale wind farms, since it provides the highest forecast accuracy. While for large scale wind farms, the computing efficiency of SC can make it more useful.
(3) Cluster number should be chosen reasonably which is important to wind power forecast accuracy. It is demonstrated that both the Hopkins statistics index and the Sihouette coefficient index are effective in selecting the optimal number of clusters, especially Hopkins statistics, which can quantitatively make a decision of the number.

[1]S. Fan, J. R. Liao, R. Yokoyama *et al.*, "Forecasting the wind generation using a two-stage network based on meteorological information," IEEE Trans. Energy Convers. **24**, 474–482 (2009).

[2]N. Bigdeli, K. Afshar, A. S. Gazafroudi *et al.*, "A comparative study of optimal hybrid methods for wind power prediction in wind farm of Alberta, Canada," Renewable Sustainable Energy Rev. **27**, 20–29 (2013).

[3]Y. Q. Liu, J. Shi, Y. P. Yang *et al.*, "Short-term wind-power prediction based on wavelet transform-support vector machine and statistics-characteristics analysis," IEEE Trans. Ind. Appl. **48**, 1136–1141 (2012).

[4]J. P. S. Catalao, H. M. I. Pousinho, and V. M. F. Mendes, "Short-term wind power forecasting in Portugal by neural networks and wavelet transform," Renewable Energy **36**, 1245–1251 (2011).

[5]Y. Q. Liu, J. J. Piao, and S. Han, "Study on two neural network algorithms to predict wind power," Mod. Electr. Power **28**, 49–52 (2011) (in Chinese), available online at http://d.wanfangdata.com.cn/Periodical_xddl201102010.aspx.

[6]M. Ali, I. S. Ilie, J. V. Milanovic *et al.*, "Probabilistic clustering of wind generators," in *IEEE Power and Energy Society General Meeting* (2010), pp. 25–29.

[7]Y. Ma, J. N. Jiang, and T. Runolfsson, "Cluster analysis of wind turbines of large wind farm," in *Power Systems Conference and Exposition* (2009), pp. 15–18.

[8]J. Yan, Y. Q. Liu, S. Han *et al.*, "Wind power grouping forecasts and its uncertainty analysis using optimized relevance vector machine," Renewable Sustainable Energy Rev. **27**, 613–621 (2013).

[9]T. Kanungo, D. M. Mount, N. S. Netanyahu *et al.*, "An efficient k-means clustering algorithm: Analysis and implementation," IEEE Trans. Pattern Anal. Mach. Intell. **24**, 881–892 (2002).

[10]K. Wagstaff, C. Cardie, S. Rogers *et al.*, "Constrained K-means clustering with background knowledge," in *Proceedings of the 18th International Conference on Machine Learning* (2001), pp. 577–584.

[11]A. M. Fahim, A. M. Salem, F. A. Torkey *et al.*, "An efficient enhanced k-means clustering algorithm," J. Zhejiang Univ., Sci., A **7**, 1626–1633 (2006).

[12]V. V. Srinivas, S. Tripathi, A. R. Rao *et al.*, "Regional flood frequency analysis by combining self-organizing feature map and fuzzy clustering," J. Hydrol. **348**, 148–166 (2008).

[13]W. S. Wang, J. Wang, and K. W. Wang, "Application of SOM neural network and C means method in load classification," in *Proceedings of the CSU-EPSA* (2011), Vol. 23, pp. 36–39.

[14]X. Y. Cai, G. Z. Dai, and L. B. Yang, "Survey on spectral clustering algorithms," Comput. Sci. **35**, 14–18 (2008) (in Chinese), available online at http://d.wanfangdata.com.cn/Periodical_jsjkx200807004.aspx.

[15]F. P. Nie, Z. N. Zeng, I. W. Tsang *et al.*, "Spectral embedded clustering: A framework for in-sample and out-of-sample spectral clustering," IEEE Trans. Neural Network **22**, 1796–1808 (2011).

[16]A. N. Ng, M. I. Jordan, and Y. Weiss, "On spectral clustering: analysis and an algorithm," in *Proceedings of the Advances in Neural Information Processing Systems* (2001), Vol. 14, pp. 849–856.

[17]R. B. Calinski and J. Harabasz, "A dendrite method for cluster Analysis," Commun. Stat. **3**, 1–27 (1974).

[18]D. L. Davies and D. W. Bouldin, "A cluster separation measure," IEEE Trans. Pattern Anal. Mach. Intell. **PAMI-1**, 224–227 (1979).

[19]L. Kaufman and P. J. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis* (John Wiley & Sons, New York, 1990).

[20]A. Banerjee and R. N. Dave, "Validating clusters using the Hopkins statistic," in *Proceedings of the 2004 IEEE International Conference on Fuzzy Systems* (IEEE, 2004), pp. 149–153.

[21]J. W. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques* (Morgan Kaufmann Publishers, USA, 2001).

[22]S. Jie, W. J. Lee, Y. Q. Liu *et al.*, "Forecasting power output of photovoltaic system based on weather classification and support vector machine," IEEE Trans. Ind. Appl. **48**, 1064–1069 (2012).

[23]G. Y. Zuo, Y. F. Fan, and J. F. Qiao, "Doorplate recognition for a mobile robot based on PSO and RBF neural network," in *2010 International Conference on Measuring Technology and Mechatronics Automation* (2010), pp. 13–14.

[24]G. Sideratos and N. Hatziargyriou, "Using radial basis neural networks to estimate wind power production," in IEEE Power Engineering Society General Meeting (IEEE, 2007).

[25]N. Amjady, F. Keynia, and H. Zareipour, "Short-term wind power forecasting using ridgelet neural network," Electr. Power Syst. Res. **81**, 2099–2107 (2011).