

# Clustering-Interpolation Method and Its Application to Wind Turbine Generator Curve

Hejun Yang<sup>1</sup>, Bo Hu<sup>1</sup>, Lei Yin<sup>1</sup>, Ya Chen<sup>1</sup>, Qinglong Liao<sup>1</sup>, Yinghao Ma<sup>1</sup>

<sup>1</sup>State Key Laboratory of Power Transmission Equipment & System Security and New Technology, Chongqing University, Chongqing, 400030, P.R. China  
cqyuhj@126.com

**Abstract**—The real-time operating wind turbine power curve (WPC) of a wind turbine generator (WTG) is not completely identical to a WPC provided by the manufacturer because of various factors. In order to obtain an accurate WPC model that can consider various factors, this paper improves a bisecting  $k$ -means clustering algorithm. The improved clustering algorithm is used for partitioning the measured data into a certain number of groups, which can be expressed in their centroids. The interpolation method based on the polynomial is carried out for modelling a WPC of WTG. The modelled WPC is applied to the reliability analysis of the generating systems with a wind farm. The results show that the accuracy of the linear interpolation is higher than that of quadratic interpolation and cubic spline interpolation when there are a relatively large number of clusters.

**Index Terms**—Wind turbine power curve, bisecting  $k$ -means clustering, interpolation, wind farm, reliability analysis.

## I. INTRODUCTION

Wind energy is a clean and renewable energy source, and it is also the world's fastest growing energy resource [1]–[2]. Because of the stochastic and intermittent nature of wind speed, an accurate prediction of wind power is difficult [3]–[4]. The measured wind turbine power curve (WPC) of a wind turbine generator (WTG) in a specific wind farm is different from a WPC provided by a WTG manufacturer [5]–[6]. The possible difference may be because of various factors, such as the wake effect, air density, barometric pressure, temperature variations, clouds, and rain. However, some factors might not be accurately included in a mathematical model for calculating the power output of a WTG [6]. Therefore, in order to calculate the power output of a WTG under a given wind condition and consider the impact of the complex environmental factors, the historical wind data (including wind speed and wind direction) and the corresponding power output data of a WTG in a real wind farm can be used for modelling a new WPC of a WTG.

The difference between a WPC provided by a WTG manufacturer and an empirical WPC measured from a real wind farm has been analysed [5]–[6]. In order to improve the short-term wind power prediction, four direction-dependent

WPC models for different wind direction ranges have been built on the basis of the measured wind data of a real wind farm [5]. A statistical forecasting system is described in [6] for the short-term prediction of wind energy production on the basis of the adaptive combination of alternative dynamic models.

Considerable work has been done for the development of WPC models for the monitoring and forecasting of wind power. In [7], four data-mining approaches for monitoring WPC are compared. This comparison shows that the adaptive neuro-fuzzy-interference system model has the best performance among the analysed approaches. In [8], a probabilistic model of WPC for monitoring purposes based on the copulas theory is developed, and the copula function is used for dealing with the complexity of the relationship between the wind speed and the wind power. Parametric and nonparametric models of WPC have been developed in [9]–[11] aiming to obtain accurate models for the online monitoring and forecasting of wind power. In [12], existing WPC models, such as polynomial power curve, exponential power curve, cubic power curve, and approximate cubic power curve, have been compared to the manufacturer's power curve. In [13], a nonlinear formula for approximating the manufacturer's WPC is proposed on the basis of the interpolation method; the parameters in this formula can be analytically determined. However, the WPC models in [12]–[13] do not consider the environmental factors. In [14], three discrete operational WTG curves (i.e., the power curve, the rotor curve, and the blade pitch curve) are introduced for monitoring a wind farm's performance, and a  $k$ -means clustering is applied. In [15], three different machine learning models are used for estimating the relationship between the wind speed and the power output of a wind farm, and an equivalent power curve model of an entire wind farm under normal operating conditions is built for detecting the anomalous functioning conditions of the wind farm.

The measured power output of WTG in a real wind farm is the real power that synthesizes the impact of complex environmental factors. Therefore, in order to incorporate the impact of complex factors, a WPC model based on an improved bisecting  $k$ -means clustering algorithm and the interpolation method is presented in this paper. As the wind speed is the dominant factor among the factors, the measured wind speed and the power output data as two-dimensional

Manuscript received September 12, 2013; accepted August 19, 2014

This work was supported in part by Natural Science Foundation Project of CQ CSTC (No. CSTC2012JJA90004), and National Key Basic Research Program of China (973 Program) (No. 2012CB215205)

coordinate points will be clustered into several points on the basis of the Euclidean distance. However, the layout of wind farm from which measured data are obtained is constant. Because the wake effect in different layouts is not the same, wake model should be independently considered in order to apply the proposed model to a wind farm with an arbitrary layout. Thus, the wake model will be built for accurately calculating the power output of a wind farm. An analytical WPC formula can be constructed on the basis of the method proposed in this paper. The proposed WPC model will be applied to the reliability analysis of generating systems with the integration of a wind farm.

## II. PROPOSED TECHNIQUE

### A. Improved Bisecting $k$ -means Clustering Algorithm

Data mining is an automatic or semiautomatic process of extracting valuable information from large amounts of data [16]–[17]. In recent years, data mining has attracted a considerable amount of attention from the information industry and the society, because of the wide availability of huge amounts of data and the imminent need for turning such data into useful information and knowledge [18]. Clustering analysis is a data-mining technique used for partitioning data objects into a certain number of groups [17]. The partitioned objects within a group are similar to one another and different from the objects in the other groups. Further, the greater the similarity within a group and the greater the difference between groups, the better or more distinct is the clustering [19].

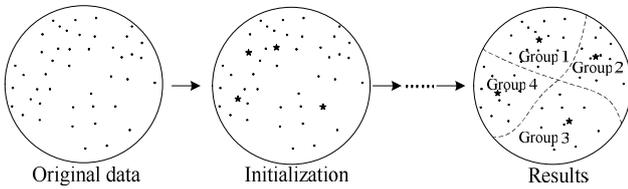


Fig. 1. Clustering processes of SKMC.

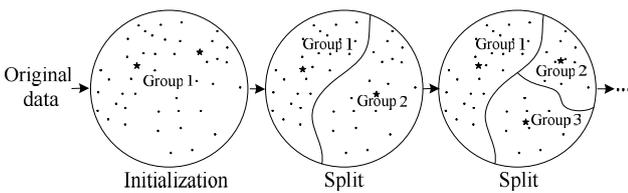


Fig. 2. Clustering processes of BKMC.

The standard  $k$ -means clustering algorithm (SKMC) is one of the best-known and most popular algorithms used in clustering, and it seeks an optimal partition of the data by using different criteria [20]–[21]. However, the results obtained from the SKMC highly depend on the initialization of the clustering parameters; in other words, different initializations may produce different results. In order to overcome the disadvantage of SKMC, a bisecting  $k$ -means clustering algorithm (BKMC) is proposed in [22]. However, there is room for improvement of the BKMC. In order to describe an improved bisecting  $k$ -means clustering algorithm (IBKMC), the clustering processes of SKMC and BKMC are summarized below, and the processes are illustrated in Fig. 1 and Fig. 2, respectively, where the points with circle shape

represent a set of data that need to be clustered and the points with star shape represent the centroids of the groups.

Finding  $k$  groups by using the SKMC algorithm:

Step 1) Select randomly  $k$  points as the initial centroids (initialization).

Step 2) Assign all points to the closest centroid, and recalculate the centroid of each cluster.

Step 3) Repeat Step 2) until the centroid of each group remains constant.

Finding  $k$  groups by using BKMC:

Step 1) Select a group to split.

Step 2) Assign all points that belong to the selected group to two groups using SKMC.

Step 3) Repeat Step 2) until the centroids of two groups remain constant.

Step 4) Repeat Steps 1), 2), and 3) until  $k$  groups are formed.

Assume that all points on a plane need to be partitioned into  $k$  groups. SKMC initially needs to randomly select  $k$  centroids. However, BKMC initially splits all points into only two groups and then, continues to select one group on the basis of a criterion function to split until  $k$  groups are constructed.

Each split based on BKMC assigns only the points from the selected group into two new groups. However, some points in the previously split groups may have a better similarity to a new group, but the points that belong to the previously split groups will not assigned again on the basis of BKMC. Therefore, the paper presents an IBKMC algorithm based on SKMC and BKMC. Assume that a possible result is obtained using BKMC, as shown in Fig. 3(a). However, if IBKMC is conducted, the potential result shown in Fig. 3(b) may be formed. It can be seen from Fig. 3 that three points and one point in group 1 shown in Fig. 3(a) are further assigned to group 2 and group 3, respectively, as shown in Fig. 3(b), and the centroids of the groups are also changed.

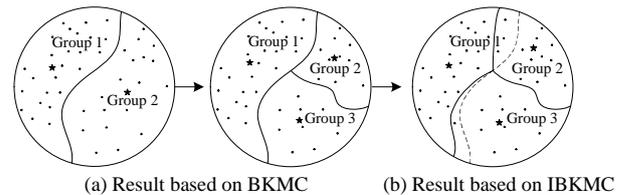


Fig. 3. Comparison of results of BKMC and IBKMC.

In this paper, the wind speed and the corresponding power output of WTG as a coordinate point can always be depicted in a Cartesian coordinate system. The  $x$  axis represents the wind speed, and the  $y$  axis represents the power output. Euclidean distance will be used for representing the distance between a point  $Q_{ji}(v_{ji}, p_{ji})$  and its centroid  $M_j(m_{vj}, m_{pj})$ , where  $j$  represents group  $j$  and  $i$  represents point  $i$  in group  $j$ . The values of  $m_{vj}$  and  $m_{pj}$  can be calculated as follows:

$$m_{vj} = \frac{1}{N_j} \sum_{i=1}^{N_j} v_{ji}, \quad (1)$$

$$m_{pj} = \frac{1}{N_j} \sum_{i=1}^{N_j} p_{ji}, \quad (2)$$

where  $N_j$  denotes the number of points in group  $j$ .

The root mean square error (RMSE) for group  $j$ , designated as  $RMSE_{Gj}$ , as a criterion function is used for selecting a group that needs to be split

$$RMSE_{Gj} = \sqrt{\frac{1}{N_j} \sum_{i=1}^{N_j} \left[ (v_{ji} - m_{vj})^2 + (p_{ji} - m_{pj})^2 \right]}. \quad (3)$$

The entire clustering error for a different cluster group  $k$ , designated as  $RMS_{CE}$ , can be calculated as follows

$$RMSE_{CE} = \sqrt{\frac{1}{N_m} \sum_{j=1}^k \sum_{i=1}^{N_j} \left[ (v_{ji} - m_{vj})^2 + (p_{ji} - m_{pj})^2 \right]}. \quad (4)$$

where  $N_m$  denotes the amount of measured wind data.

SKMC is sensitive to the selection of the initial centroids. BKMC can overcome the sensitivity of the selection of the initial centroids, but it does not re-assign the points that belong to the previously split groups. Therefore, an IBKMC combining SKMC and BKMC is presented in this paper. In IBKMC, BKMC is used for splitting a group into two groups and producing two centroids that are used as the initial centroids of SKMC. The clustering processes of IBKMC are described in detail as follows:

Step 1) Initially, all points construct a group. Let  $j=1$ .

Step 2) Select randomly two points from a selected group as the initial centroids.

Step 3) Calculate the Euclidean distances between all points in the selected group and the two centroids, and assign all points to the closest centroid.

Step 4) Recalculate the centroids of the two groups.

Step 5) Repeat Step 3) and Step 4) until the two centroids remain constant.  $j=j+1$ .

Step 6) Use the  $j$  centroids as the initial centroids of SKMC for reassigning all points that belong to the  $j$  centroids, and recalculate the centroids of the groups.

Step 7) Go to Step 2) to split the next group with the maximum  $RMSE_{Gj}$  until  $j=k$ .

Step 8) Check all centroids. For any  $i$  and  $j \in (1, 2, \dots, N_j)$ , and  $i \neq j$ , if  $m_{vi} = m_{vj}$ , the centroid of the  $j$ th group will be replaced by a closest point to  $(m_{vi}, m_{pj})$ , and go to Step 6) until all  $m_{vi}$  are unequal.

If there is no requirement of  $m_{vi} = m_{vj}$  ( $i \neq j, i, j \in (1, 2, \dots, N_j)$ ), implement the above Steps 1) to 7); otherwise, implement the Steps 1) to 8).

### B. WPC Modelling Based on Interpolation

In mathematics, a polynomial equation of degree  $n$  can be expressed as (5)

$$p(v) = \sum_{i=0}^n a_i v^i, \quad (5)$$

where  $v$  denotes a variable (it denotes wind speed in this paper);  $a_0, a_1, \dots, a_n$  represent the coefficients of the  $n$ th degree polynomial equation; and  $p(v)$  denotes the functional value of the polynomial equation (it denotes the power output of a WTG with a wind speed  $v$  in this paper).

According to the proposed IBKMC from Steps 1) to 8),  $k$  clustering centroids  $(M_1(m_{v1}, m_{p1}), M_2(m_{v2}, m_{p2}), \dots, M_k(m_{vk}, m_{pk}))$  are obtained. A function that matches the values at the clustering centroids is developed. A  $(k-1)$ th degree polynomial equation that passes through all the clustering centroids can be formulated analytically for obtaining the algebraic equation of WPC. Equation (5) can be rewritten as follows:

$$\begin{cases} m_{p1} = a_{k-1}m_{v1}^{k-1} + a_{k-2}m_{v1}^{k-2} + \dots + a_2m_{v1}^2 + a_1m_{v1} + a_0, \\ m_{p2} = a_{k-1}m_{v2}^{k-1} + a_{k-2}m_{v2}^{k-2} + \dots + a_2m_{v2}^2 + a_1m_{v2} + a_0, \\ \vdots \\ m_{pk} = a_{k-1}m_{vk}^{k-1} + a_{k-2}m_{vk}^{k-2} + \dots + a_2m_{vk}^2 + a_1m_{vk} + a_0, \end{cases} \quad (6)$$

In (6),  $m_{vj}$  and  $m_{pj}$  for any  $j$  are known values, whereas  $a_0, a_1, \dots, a_n$  are the unknown values. Therefore, in order to obtain the deterministic expression of a polynomial of degree  $k-1$ , equation (6) needs to be solved for obtaining the values of  $a_0, a_1, \dots, a_n$ . Equation (6) can be expressed using a matrix as follows:

$$\begin{bmatrix} 1 & m_{v1} & m_{v1}^2 & \dots & m_{v1}^{k-2} & m_{v1}^{k-1} \\ 1 & m_{v2} & m_{v2}^2 & \dots & m_{v2}^{k-2} & m_{v2}^{k-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & m_{vk} & m_{vk}^2 & \dots & m_{vk}^{k-2} & m_{vk}^{k-1} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_{k-2} \\ a_{k-1} \end{bmatrix} = \begin{bmatrix} m_{p1} \\ m_{p2} \\ \vdots \\ m_{pk} \end{bmatrix}, \quad (7)$$

Equation (7) can be simplified using  $\mathbf{M}_v \times \mathbf{A} = \mathbf{M}_p$ . Matrix  $\mathbf{M}_v$  is a Vandermonde matrix [23-24]. Thus, the determinant of matrix  $\mathbf{M}_v$  can be calculated using (8):

$$\begin{vmatrix} 1 & m_{v1} & m_{v1}^2 & \dots & m_{v1}^{k-2} & m_{v1}^{k-1} \\ 1 & m_{v2} & m_{v2}^2 & \dots & m_{v2}^{k-2} & m_{v2}^{k-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & m_{vk} & m_{vk}^2 & \dots & m_{vk}^{k-2} & m_{vk}^{k-1} \end{vmatrix} = \prod_{1 \leq j < i \leq k} (m_{vi} - m_{vj}), \quad (8)$$

Due to  $m_{vi} \neq m_{vj}$  for any  $i$  and  $j$  ( $i \neq j$ ). Thus, matrix  $\mathbf{M}_v$  is invertible. Therefore,  $\mathbf{A}$  can be calculated using (9)

$$\mathbf{A} = \mathbf{M}_v^{-1} \mathbf{M}_p = \frac{1}{\prod_{1 \leq j < i \leq k} (m_{vi} - m_{vj})} \mathbf{M}_v^* \mathbf{M}_p. \quad (9)$$

where  $\mathbf{M}_v^*$  denotes the algebraic cofactor of matrix  $\mathbf{M}_v$ .

In this paper, the linear interpolation and quadratic interpolation based on the above analysis is used for modelling the WPCs of a WTG. The cubic spline interpolation introduced in [25] is also used for modelling a WPC.  $RMSE_{WPC}$  from (10) is used for calculating the error of the different interpolation methods

$$RMSE_{WPC} = \sqrt{\frac{1}{N_m} \sum_{t=1}^{N_m} \left[ (P_m(v) - P_{WPC}(v))^2 \right]}. \quad (10)$$

where  $P_m(v)$  denotes the measured power output of a WTG having the wind speed  $v$ ;  $P_{WPC}(v)$  represents the power output calculated by the WPC having the wind speed  $v$ .

### III. APPLICATION OF PROPOSED WPC TO RELIABILITY ANALYSIS OF GENERATING SYSTEMS

#### A. Wake Model

As the wake effect has a significant influence on the energy production of a wind farm [26], it should be incorporated into the analysis of a power system with the integration of a wind farm. References [27] and [28] introduced the Jensen wake model used for a flat terrain and the Lissaman model used for a complex terrain.

#### B. Reliability Analysis

A wind farm is incorporated into a conventional generating system, and an analytic method [29], [30] is used for evaluating the reliability of the combined generating system (CGS). In the wind farm, the wind speed data as the input variable are used for calculating the wind power output using the WPC model proposed in this paper.

In this paper, reliability indices, such as the loss of load probability (LOLP), loss of load expectation (LOLE, hours/year), and expected energy not supplied (EENS, MWh/year), are evaluated. These indices can be expressed as follows:

$$LOLP = \sum_{i \in S_f} p_i, \quad (11)$$

$$LOLE = \sum_{i \in S_f} p_i T, \quad (12)$$

$$EENS = \sum_{i \in S_f} C_i p_i T. \quad (13)$$

where  $S_f$  denotes a set of all system failure states;  $p_i$  and  $C_i$  indicate the probability and the capacity of failure state  $i$ , respectively;  $T$  denotes the total time length.

An index for describing the energy production of a wind farm, designated as  $EPWF$  (MWh/year), is calculated for comparing the accuracy of the different WPC models. The error of  $EPWF$  for the different WPC models, designated as  $EPER$ , is calculated for directly reflecting the accuracy.

$$EPER = \frac{|EPWF_m - EPWF|}{EPWF_m}, \quad (14)$$

where  $EPWF_m$  denotes the energy production of the measured data in a real wind farm.  $EPWF$  indicates the energy production obtained from a WPC model.

### IV. CASE STUDIES

WTGs with a rated power of 850 kW are installed in a real wind farm, and the cut-in speed, rated speed, and cut-out speed of the WTG are 3 m/s, 11 m/s, and 20 m/s, respectively. A nonlinear relationship between the wind speed and the power output of a WTG measured in the real wind farm is illustrated in Fig. 4. A WPC provided by the wind turbine manufacturer, designated as MWPC, and a typical WPC [31],

designated as TWPC, are also illustrated in Fig. 4. The power output data in failure states of WTG are removed from data which was used to model WPC.

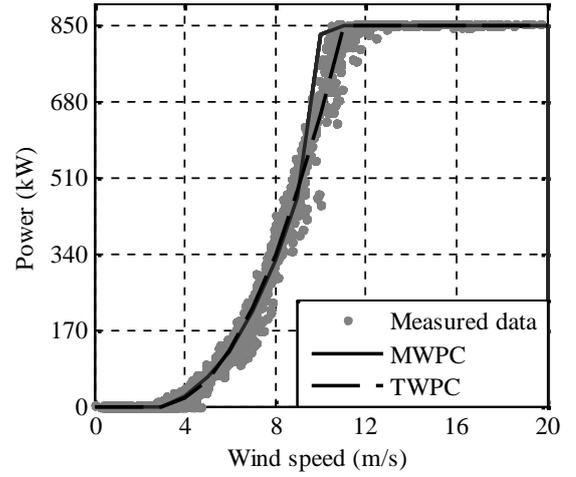


Fig. 4. WPC of a WTG.

Because of the impact of various factors, the real relationship between the wind speed and the power output of a WTG is not always completely in accordance with the WPC provided by the manufacturer. Therefore, it is essential to model a WPC of the WTG for precisely reflecting the relationship by using the method proposed in this paper.

#### A. WPC Modelling based on Proposed Method

##### 1) Clustering based on IBKMC.

The measured wind speed and power output data as points in a two-dimensional Euclidean space are clustered into a certain number of groups ( $k$ ) on the basis of the proposed IBKMC. The percentage of  $RMSE_{CE}$  shown in Fig. 5 is used for assessing the clustering error.

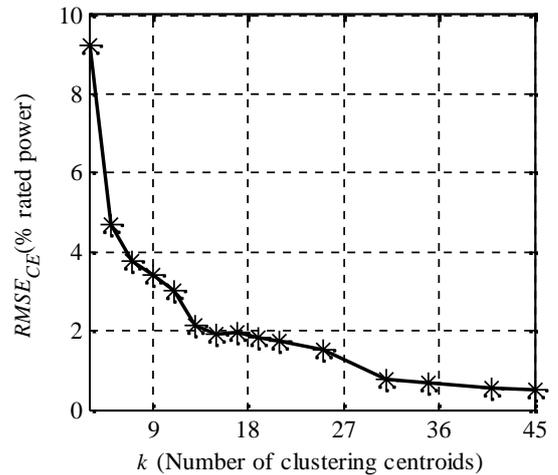
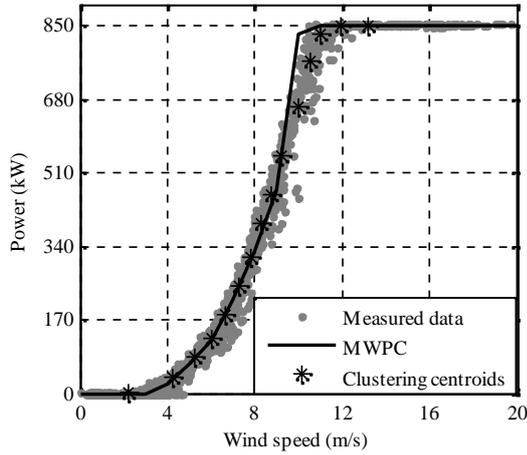


Fig. 5.  $RMSE_{CE}$  with different numbers of clusters.

It can be seen from Fig. 5 that  $RMSE_{CE}$  gradually decreases with an increase in the number of clusters, but it starts to decline slowly at  $k = 15$ .

The clustering centroids with  $k = 15$  are shown in Table I, and for a comparative analysis, the centroids are also illustrated in Fig. 6.

It can be seen from Fig. 6 that in the vicinity of the rated wind speed, there is a relatively large difference between MWPC and the clustering centroids.

Fig. 6. Clustering centroids with  $k = 15$ .TABLE I. CLUSTERING CENTROIDS WITH  $k = 15$ .

Number	Centroids	Number	Centroids
1	(3, 0)	9	(8.79, 460.6)
2	(4.27, 37.1)	10	(9.20, 549.1)
3	(5.28, 82.3)	11	(9.97, 664.1)
4	(6.05, 127.0)	12	(10.57, 767.8)
5	(6.67, 182.5)	13	(11.02, 830.3)
6	(7.27, 248.3)	14	(11.96, 850)
7	(7.82, 316.8)	15	(13.21, 850)
8	(8.32, 393.2)	-	-

## 2) WPC modelling based on interpolation method.

According to the clustering centroids, linear interpolation, quadratic interpolation, and cubic spline interpolation are used for modelling the WPC of a WTG. The  $RMSE_{WPC}$  of different interpolation methods is shown in Fig. 7.

In Fig. 7,  $RMSE_{WPC}$  based on these three interpolation methods first decreases rapidly with an increase in  $k$  and then tends to stabilize. However, when the value of  $k$  is greater than 9, cubic spline interpolation has a greater error than linear interpolation and quadratic interpolation.

In this paper, WPC obtained by linear interpolation is designated as LWPC, whereas that obtained by quadratic

interpolation is designated as QWPC. On the basis of the numerical analysis theory, the coefficients of piecewise LWPC and QWPC are obtained, as shown in Table II and Table III, where LB and UB denote the lower bound and the upper bound, respectively. When the wind speed is less than 2.24 m/s, the power output is zero, and when the wind speed is higher than 13.21 m/s, the power output is 850 kW. LWPC and QWPC are illustrated in Fig. 8.

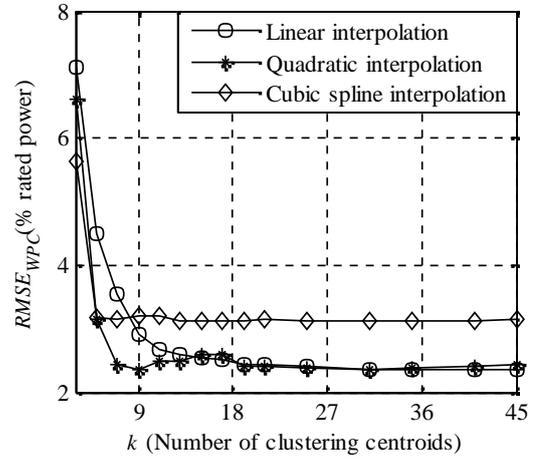
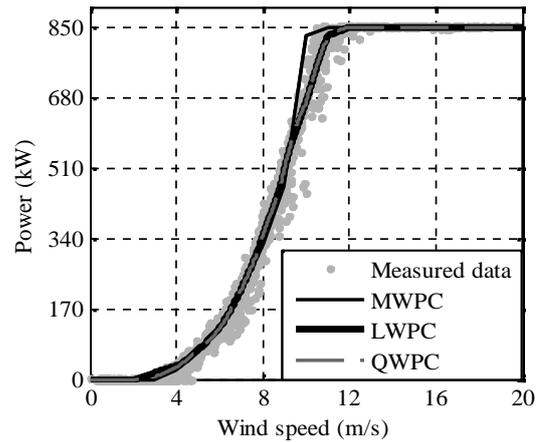
Fig. 7.  $RMSE_{WPC}$  of different interpolation methods.

Fig. 8. WPCs based on different methods.

TABLE II. EQUATION COEFFICIENTS BASED ON LINEAR INTERPOLATION.

Wind speed range		Equation coefficients		Wind speed range		Equation coefficients	
LB	UP	$a_0$	$a_1$	LB	UP	$a_0$	$a_1$
3.0	4.27	-87.78	29.26	8.32	8.79	-810.14	144.58
4.27	5.28	-153.51	44.66	8.79	9.20	-1454.37	217.87
5.28	6.05	-222.96	57.82	9.20	9.97	-820.07	148.89
6.05	6.67	-414.47	89.45	9.97	10.57	-1058.39	172.80
6.67	7.27	-548.30	109.51	10.57	11.02	-709.14	139.76
7.27	7.82	-668.65	126.05	11.02	11.96	600.90	20.82
7.82	8.32	-864.56	151.12	11.96	13.21	850.00	0.00

TABLE III. EQUATION COEFFICIENTS BASED ON QUADRATIC INTERPOLATION.

Wind speed range		Equation coefficients		
LB	UB	$a_0$	$a_1$	$a_2$
3.0	5.28	-1.26	-19.85	6.76
5.28	6.67	502.59	-199.47	22.70
6.67	7.82	153.50	-92.14	14.46
7.82	8.79	-1302.22	259.69	-6.73
8.79	9.97	-6183.89	1270.30	-58.52
9.97	11.02	-4384.02	821.12	-31.57
11.02	13.21	-646.88	238.41	-9.47

### A. Application of WPC Model to Reliability Analysis

The wind farm with 20 identical WTGs is incorporated into the Roy Billinton Test System (RBTS), which has an annual peak load of 185 MW and an installed capacity of 240 MW. The detailed basic reliability data are presented in [32].

In the wind farm, the WTGs are distributed in four rows and five columns. The distance between any two adjacent rows is 400 m, and the distance between two adjacent columns is 300 m. The rotor diameter ( $2r$ ) is 56 m, and the hub height is 70 m. Force outage rate of a WTG is 0.04. It is

assumed that the failed upstream WTG will not produce wake.

In order to compare the accuracy of the proposed WPC model, the measured wind speed data are used for evaluating the reliability of the CGS. As LWPC has a smaller  $RMSE_{WPC}$  when  $k$  has a relatively large value, it will be analysed in this section. LWPC with  $k = 15, 20, 30, 40, 50$  is used for evaluating the reliability of the CGS and for calculating  $EPWF$  and  $EPER$ . The results with or without considering the wake effect are shown in Table IV and Table V, respectively.

TABLE IV. RESULTS WITHOUT CONSIDERING WAKE EFFECT.

WPC	k	LOLP	LOLE	EENS	EPWF	EPER (%)
Measured data	-	8.934E-05	0.7826	6.759	28238.6	-
LWPC	15	8.886E-05	0.7784	6.729	28578.8	1.20
	20	8.891E-05	0.7789	6.731	28561.6	1.14
	30	8.908E-05	0.7803	6.739	28468.1	0.81
	40	8.917E-05	0.7811	6.743	28421.3	0.65
	50	8.926E-05	0.7819	6.750	28339.4	0.36
TWPC	-	9.027E-05	0.7907	6.796	27780.1	1.62

TABLE V. RESULTS CONSIDERING WAKE EFFECT.

WPC	k	LOLP	LOLE	EENS	EPWF
LWPC	15	9.593E-05	0.8403	7.209	23367.4
	20	9.578E-05	0.8390	7.215	23302.7
	30	9.581E-05	0.8393	7.229	23167.4
	40	9.58E-05	0.8396	7.235	23105.1
	50	9.61E-05	0.8415	7.245	23008.1
TWPC	-	9.701E-05	0.8498	7.300	22434.3

From Table IV, the indices of LWPC and TWPC can be compared with those of the measured data. It can be seen that LWPC with any value of  $k$  given in Table IV has a higher accuracy than TWPC, and  $EPER$  gradually decreases with an increase in  $k$ . According to the comparison of Table IV and Table V, the wake effect decreases the reliability of the generating system and the wind energy production.

comparing the impact of the wake effect. The results show that the accuracy of the WPC models first decreases rapidly with an increase in the number of clusters and then tends to stabilize. Moreover, LIM has a superior performance in comparison with QIM and CSIM when there are a relatively large number of clusters.

### V. CONCLUSIONS

The energy production of a wind farm is affected by various factors. However, some factors might not be expressed using an accurate mathematical model for calculating the wind energy production. Therefore, it is essential to model a wind turbine generator (WTG) according to the measured wind speed and the power output data for including the impacts of various factors.

In this paper, we present an improved bisecting  $k$ -means clustering algorithm (IBKMC). IBKMC is used for partitioning the measured data into a certain number of groups that can be expressed using their centroids. The interpolation method based on polynomials is applied to the clustering centroids for modelling the wind power curve (WPC) of a WTG. The linear interpolation method (LIM), quadratic interpolation method (QIM), and cubic spline interpolation method (CSIM) are used for analyses based on the root mean square error (RMSE) in this paper. The WPC model is applied to the reliability analysis of a combined generating system (CGS) with the integration of a wind farm. The CGS with and without considering the wake effect is evaluated for

### REFERENCES

- [1] K. Xie, R. Billinton, "Considering wind speed correlation of WECS in reliability evaluation using the time-shifting technique", *Electr. Power Syst. Res.*, vol. 79, no. 4, pp. 687–693, 2009. [Online]. Available: <http://dx.doi.org/10.1016/j.epr.2008.10.013>
- [2] D. Komarov, S. Stupar, A. Simonovic, "Prospects of wind energy sector development in Serbia with relevant regulatory framework overview", *Renew. Sust. Energ. Rev.*, vol. 16, no. 5, pp. 2618–2630, 2012. [Online]. Available: <http://dx.doi.org/10.1016/j.rser.2012.01.067>
- [3] G. Sideratos, N. D. Hatziargyriou, "An advanced statistical method for wind power forecasting", *IEEE Trans. Power Syst.*, vol. 22, no. 1, pp. 258–265, 2007. [Online]. Available: <http://dx.doi.org/10.1109/TPWRS.2006.889078>
- [4] K. Xie, Z. Jiang, W. Li, "Effect of wind speed on wind turbine power converter reliability", *IEEE Trans. Energy Convers.*, vol. 27, no. 1, pp. 96–104, 2012. [Online]. Available: <http://dx.doi.org/10.1109/TEC.2011.2179656>
- [5] M. Khalid, A. V. Savkin, "A method for short-term wind power prediction with multiple observation points", *IEEE Trans. Power Syst.*, vol. 27, no. 2, pp. 579–586, 2012. [Online]. Available: <http://dx.doi.org/10.1109/TPWRS.2011.2160295>
- [6] I. Sanchez, "Short-term prediction of wind energy production", *Int. J. Forecasting*, vol. 22, no. 1, pp. 43–56, 2006. [Online]. Available: <http://dx.doi.org/10.1016/j.ijforecast.2005.05.003>
- [7] M. Schlechtingen, I. F. Santos, S. Achiche, "Using data-mining approaches for wind turbine power curve monitoring: a comparative study", *IEEE Trans. Sustainable Energy*, vol. 4, no. 3, pp. 671–679,

2013. [Online]. Available: <http://dx.doi.org/10.1109/TSTE.2013.2241797>
- [8] S. Gill, B. Stephen, S. Galloway, "Wind turbine condition assessment through power curve copula modelling", *IEEE Trans. Sustainable Energy*, vol. 3, no. 1, pp. 94–101, 2012. [Online]. Available: <http://dx.doi.org/10.1109/TSTE.2011.2167164>
- [9] M. Lydia, A. I. Selvakumar, S. S. Kumar, "Advanced algorithms for wind turbine power curve modelling", *IEEE Trans. Sustainable Energy*, vol. 4, no. 3, pp. 827–835, 2013. [Online]. Available: <http://dx.doi.org/10.1109/TSTE.2013.2247641>
- [10] A. Kusiak, H. Zheng, Z. Song, "On-line monitoring of power curves", *Renew. Energy*, vol. 34, no. 6, pp. 1487–1493, 2009. [Online]. Available: <http://dx.doi.org/10.1016/j.renene.2008.10.022>
- [11] A. Kusiak, H. Zheng, Z. Song, "Models for monitoring wind farm power", *Renew. Energy*, vol. 34, no. 3, pp. 583–590, 2009. [Online]. Available: <http://dx.doi.org/10.1016/j.renene.2008.05.032>
- [12] C. Carrillo, A. F. O. Montano, J. Cidra, "Review of power curve modelling for wind turbines", *Renew. Sust. Energ. Rev.*, vol. 21, pp. 572–581, 2013. [Online]. Available: <http://dx.doi.org/10.1016/j.rser.2013.01.012>
- [13] X. Liu, "An improved interpolation method for wind power curves", *IEEE Trans. Sustainable Energy*, vol. 3, no. 3, pp. 528–534, 2012. [Online]. Available: <http://dx.doi.org/10.1109/TSTE.2012.2191582>
- [14] A. Kusiak, A. Verma, "Monitoring wind farms with performance curves", *IEEE Trans. Sustainable Energy*, vol. 4, no. 1, pp. 192–199, 2013. [Online]. Available: <http://dx.doi.org/10.1109/TSTE.2012.2212470>
- [15] A. Marvuglia, A. Messineo, "Monitoring of wind farms' power curves using machine learning techniques", *Appl. Energy*, vol. 98, pp. 574–583, 2012. [Online]. Available: <http://dx.doi.org/10.1016/j.apenergy.2012.04.037>
- [16] D. J. Hand, H. Mannila, P. Smyth, *Principles of Data Mining*. MIT press, 2001.
- [17] I. H. Witten, E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann, 2005.
- [18] J. Han, M. Kamber, J. Pei, *Data Mining: Concepts and Techniques*. Morgan Kaufmann, 2006.
- [19] P. N. Tan, *Introduction to Data Mining*. Pearson Education India, 2007.
- [20] M. F. Eltibi, W. M. Ashour, "Initializing k-means clustering algorithm using statistical information", *Int. J. Comput. Appl. Technol.*, vol. 29, no. 7, pp. 51–55, 2011.
- [21] R. Xu, D. Wunsch, *Clustering*, Wiley-IEEE Press, 2008. [Online]. Available: <http://dx.doi.org/10.1002/9780470382776>
- [22] M. Steinbach, G. Karypis, V. Kumar, *A Comparison of Document Clustering Techniques*. KDD Workshop on Text Mining, 2000.
- [23] W. Cheney, D. Kincaid, *Linear Algebra: Theory and Applications*, the Australian Mathematical Society, 2009.
- [24] G. Strang, *Linear Algebra and Its Applications*, Brooks Cole India, 2005.
- [25] W. Gautschi, *Numerical Analysis*, Springer Science + Business Media, 2012.
- [26] H. Kim, C. Singh, A. Sprintson, "Simulation and estimation of reliability in a wind farm considering the wake effect", *IEEE Trans. Sustainable Energy*, vol. 3, no. 2, pp. 274–282, 2012. [Online]. Available: <http://dx.doi.org/10.1109/TSTE.2011.2174260>
- [27] N. O. Jensen, "A note of wind generator interaction", Denmark: Riso National Laboratory, 1993, pp. 1–17.
- [28] A. Spera David, *Wind Turbine Technology: Fundamental Concepts of Wind Turbine Engineering*. New York: American Society of Mechanical Engineers, 2009.
- [29] R. Billinton, R. N. Alan, *Reliability Evaluation of Power Systems*, New York: Plenum Press, 1996. [Online]. Available: <http://dx.doi.org/10.1007/978-1-4899-1860-4>
- [30] R. Billinton, A. A. Chowdhury, "Incorporation of wind energy conversion systems in conventional generating capacity adequacy assessment", *IEE Proc. Part C Gener. Transm. Distrib.*, vol. 139, no. 1, pp. 47–56, 1992.
- [31] R. Billinton, S. Kumar, N. Chowdhury, "A reliability test system for educational purposes-basic data", *IEEE Trans. Power Syst.*, vol. 4, no. 3, pp. 1238–1244, 1989. [Online]. Available: <http://dx.doi.org/10.1109/59.32623>