Probabilistic load flow method considering large-scale wind power integration

Xiaoyang DENG[1], Pei ZHANG[1], Kangmeng JIN[1], Jinghan HE[1], Xiaojun WANG[1], Yuwei WANG[2]

Abstract The increasing penetration of wind power brings great uncertainties into power systems, which poses challenges to system planning and operation. This paper proposes a novel probabilistic load flow (PLF) method based on clustering technique to handle large fluctuations from large-scale wind power integration. The traditional cumulant method (CM) for PLF is based on the linearization of load flow equations around the operating point, therefore resulting in significant errors when input random variables have large fluctuations. In the proposed method, the samples of wind power and loads are first generated by the inverse Nataf transformation and then clustered using an improved *K*-means algorithm to obtain input variable samples with small variances in each cluster. With such pre-processing, the cumulant method can be applied within each cluster to calculate cumulants of output random variables with improved accuracy. The results obtained in each cluster are combined according to the law of total probability to calculate the final cumulants of output random variables for the whole samples. The proposed method is validated on modified IEEE 9-bus and 118-bus test systems with additional wind farms. Compared with the traditional CM, $2m+1$ point estimate method (PEM), Monte Carlo simulation (MCS) and Latin hypercube sampling (LHS) based MCS, the proposed method can achieve a better performance with consideration of both computational efficiency and accuracy.

Keywords Cumulant method (CM), Improved *K*-means algorithm, Large-scale wind power integration, Probabilistic load flow (PLF)

✉ Pei ZHANG
2512692577@qq.com

Xiaoyang DENG
14117389@bjtu.edu.cn

Kangmeng JIN
16121467@bjtu.edu.cn

Jinghan HE
jhhe@bjtu.edu.cn

Xiaojun WANG
xjwang1@bjtu.edu.cn

Yuwei WANG
yuwei.wang16@imperial.ac.uk

[1] School of Electrical Engineering, Beijing Jiaotong University, Beijing, China

[2] Department of Electrical and Electronic Engineering, Imperial College London, London, UK

1 Introduction

Load flow study is a vital tool for power system planning and operation. However, there are many uncertainties, which result from changes in load demands, outages of generators and changes of network. Large-scale wind power integration further introduces great uncertainties into power systems. Many researchers have performed researches on applying probabilistic load flow (PLF) methods to handle these uncertainties.

A critical review was provided in [1], where the methods to solve PLF problems were classified into three types, namely simulation methods, approximation methods and analytical methods.

Monte Carlo simulation (MCS) can obtain accurate results after a large number of simulations, which are

generally treated as reference results for comparisons. But, MCS is time-consuming. The application of importance sampling [2], Latin hypercube sampling (LHS) [3, 4] and Latin supercube sampling [5] reduced the computational burden of MCS. For handling correlated input random variables, Nataf transformation [6, 7] and copula function [8] were applied together with LHS. References [9–11] applied a quasi-Monte Carlo approach to solve PLF problems. That approach was more efficient than MCS.

The point estimate method (PEM), one kind of approximation methods, was widely applied to solve PLF problems. Reference [12] first proposed the $2m$ PEM to solve PLF. Reference [13] introduced a modified $2m$ PEM to handle correlated uncertain variables. Compared with $2m$ PEM, $2m+1$ PEM had higher accuracy, but conducted more simulations [14]. For handling correlated uncertain variables, references [15, 16] provided a modified $2m+1$ PEM based on Cholesky decomposition. Reference [17] applied $2m+1$ PEM to solve probabilistic three-phase load flow for unbalanced electrical systems with wind farms. Reference [18] discussed the performance of five point estimate method (5PEM). Reference [19] proposed another approximation method, unscented transformation method (UTM), which could consider correlations of input variables. Approximation methods are generally more efficient than MCS. However, the accuracy and efficiency are sensitive to the number of input random variables.

Analytical methods do not need to run many times of simulations as Monte Carlo method does. In [20], a first-order second-moment method (FOSMM) was applied to obtain the mean and standard deviation of load flow solutions. The sequence operation methodology [21] is one of the analytical methods. It has a great advantage in terms of efficiency by sequence operation. But the sequence operation needs to meet new operation rules, which limits its application. Cumulant method (CM) is another analytical method to solve PLF and has an excellent performance on computational efficiency. In [22, 23], CM and Gram-Charlier expansion were applied to solve PLF. Reference [24] discussed the properties, advantages and deficiencies of three types of series expansions, namely Gram-Charlier, Edgeworth and Cornish-Fisher expansions. Furthermore, Cholesky decomposition [25] and joint cumulants [24, 26] were utilized to deal with correlations of input variables. Reference [27] applied Gaussian mixture approximation method to handle non-normal correlated random variables. References [28, 29] applied the maximum entropy instead of series expansions to calculate probability density functions (PDFs). It could improve the accuracy of PDFs, but required a relatively more complex process.

CM requires less computational effort than other methods. However, it may produce significant errors when input random variables have large fluctuations. The reason is that the linear relationship between input and output variables is estimated based on the linearization of load power equations at the operating point. When input random variables fluctuate away from the operating point, the relationship between input and output variables may change significantly. Reference [30] studied the error resulting from linear model and showed that the error would increase when the varying ranges of input variables increased. The wind power output can change significantly over time due to fluctuations of wind speeds. To solve PLF considering large-scale wind power integration, reference [31] divided PDFs of wind power into multiple intervals, and incorporated these intervals into the integral formulation of calculating cumulants. This method cannot handle the correlation of different input variables and is computationally complex. Our previous work [32] tried to solve probabilistic optimal power flow (P-OPF) with large fluctuations using the method of combined the traditional K-means clustering technique and CM. However, the traditional K-means performs inefficiently for large-scale systems.

To solve PLF considering large-scale wind power integration, this paper proposes a novel PLF method by combining an improved K-means clustering technique and CM. It tackles the problems that the traditional CM cannot handle input random variables with large fluctuations, and that the traditional K-means is not efficient for large-scale systems. Compared with existing methods, such as the traditional CM, $2m+1$ PEM, LHS and MCS, the proposed method can achieve a better performance with consideration of both computational efficiency and accuracy. The proposed method can be used to analyze effects of uncertainties on power systems.

The rest of this paper is organized as follows: Section 2 introduces the CM for PLF formulation. The theoretical framework of the proposed method is described in Section 3. IEEE 9-bus and 118-bus test systems are modified for case studies in Section 4. Finally, conclusions are summarized in Section 5.

2 CM for PLF formulation

CM for PLF is based on the linearization of load flow equations. In the process, a set of equations, which consist of nodal power injections and line power flows, are formulated.

Let X be the vector of nodal power injections, U be the vector consisting of voltage angles at PV and PQ buses and voltage magnitudes at PQ buses, and Z be the vector of power flows in branches. AC power flow equations can be written as:

$$\begin{cases} \boldsymbol{X} = \boldsymbol{f}(\boldsymbol{U}) \\ \boldsymbol{Z} = \boldsymbol{g}(\boldsymbol{U}) \end{cases} \tag{1}$$

where $\boldsymbol{f}(\cdot)$ and $\boldsymbol{g}(\cdot)$ are the corresponding power injection functions and corresponding line flow functions, respectively.

The linear equations can be obtained by linearizing (1) at the operating point:

$$\begin{cases} \boldsymbol{U} = \boldsymbol{U}_0 + \boldsymbol{J}_0^{-1}\Delta\boldsymbol{X} = \boldsymbol{U}_0 + \boldsymbol{H}_0\Delta\boldsymbol{X} \\ \boldsymbol{Z} = \boldsymbol{Z}_0 + \boldsymbol{G}_0\Delta\boldsymbol{U} = \boldsymbol{Z}_0 + \boldsymbol{L}_0\Delta\boldsymbol{X} \end{cases} \tag{2}$$

where $\boldsymbol{U}_0$ and $\boldsymbol{Z}_0$ are the values of $\boldsymbol{U}$ and $\boldsymbol{Z}$ at the operating point, respectively; $\Delta\boldsymbol{U}$ and $\Delta\boldsymbol{X}$ are the vectors of the changes in $\boldsymbol{U}$ and $\boldsymbol{X}$; $\boldsymbol{J}_0^{-1}$ is the inverse of Jacobian matrix at the operating point; $\boldsymbol{H}_0 = \boldsymbol{J}_0^{-1}$; $\boldsymbol{G}_0 = (\partial\boldsymbol{Z}/\partial\boldsymbol{U})|_{\boldsymbol{U}=\boldsymbol{U}_0}$; $\boldsymbol{L}_0 = \boldsymbol{G}_0\boldsymbol{J}_0^{-1}$.

For power systems containing wind farms, fluctuations of both wind generation and load can result in uncertainties. The active wind power output can be obtained as follows [23]:

$$P_{\mathrm{w}} = \begin{cases} 0 & 0 < v_i < v_{\mathrm{ci}} \\ (v_i - v_{\mathrm{ci}})/(v_{\mathrm{r}} - v_{\mathrm{ci}})P_{\mathrm{r}} & v_{\mathrm{ci}} \le v_i < v_{\mathrm{r}} \\ P_{\mathrm{r}} & v_{\mathrm{r}} \le v_i < v_{\mathrm{co}} \\ 0 & v_i \ge v_{\mathrm{co}} \end{cases} \tag{3}$$

where P_{w} is the active wind power output; P_{r} is the rated power of the wind farm; v_i is the wind speed of the wind farm; v_{ci}, v_{r} and v_{co} are the cut-in, rated and cut-out speeds of the wind farm, respectively. Wind power output is treated as a negative load, whose power factor is kept constant [13].

Thus, $\Delta\boldsymbol{X}$ can be reformed as follows:

$$\Delta\boldsymbol{X} = \Delta\boldsymbol{W} - \Delta\boldsymbol{L} \tag{4}$$

where $\boldsymbol{W}$ is the vector consisting of active and reactive wind power outputs at corresponding buses; $\boldsymbol{L}$ is the vector consisting of active and reactive load demands at corresponding buses; $\Delta\boldsymbol{W}$ and $\Delta\boldsymbol{L}$ are the vectors of the changes in $\boldsymbol{W}$ and $\boldsymbol{L}$.

From (2), taking a specific variable in $\boldsymbol{U}$ for example, it can be converted to a linear combination as follows:

$$\begin{aligned} u_i &= u_{i0} + \sum_{j=1}^{N_X} h_{0ij}(w_j - w_{j0}) - \sum_{j=1}^{N_X} h_{0ij}(l_j - l_{j0}) \\ &= u_{si0} + \sum_{j=1}^{N_X} h_{0ij}w_j - \sum_{j=1}^{N_X} h_{0ij}l_j \end{aligned} \tag{5}$$

where $u_{si0} = u_{i0} - \sum_{j=1}^{N_X} h_{0ij}w_{j0} + \sum_{j=1}^{N_X} h_{0ij}l_{j0}$; u_i is a specific variable in $\boldsymbol{U}$; w_j is the jth variable in $\boldsymbol{W}$; l_j is the jth variable in $\boldsymbol{L}$; u_{i0}, w_{j0} and l_{j0} are the values of u_j, w_j, and l_j at the operating point, respectively; h_{0ij} is the value at row i and column j of $\boldsymbol{H}_0$; N_X is the number of variables of $\boldsymbol{X}$. For

active and reactive power flows in branches, they can also be expressed as a linear combination of input variables.

According to (5), system variables can be converted to a linear combination of input random variables. Assuming the independence among input random variables, the cumulants of output random variables can be calculated by combining the cumulants of input random variables based on the property of cumulants [22]. In general, there are correlations of input random variables. In this paper, the correlations of input random variables are handled by the Cholesky decomposition algorithm [25].

3 Proposed method

The fundamental reason why the traditional CM has high errors for solving PLF of power systems containing large-scale wind power is that the wind power output can change significantly over time due to fluctuations of wind speeds. Therefore, this paper focuses on how to reduce the fluctuations of input random variables. Given the probability distribution functions and correlation coefficient matrix of input random variables, the samples of input random variables can be generated through the inverse Nataf transformation [33]. The values of input variables at the same position form one point as shown in Fig. 1, where X_i is a column vector of samples for a specific input variable (wind power or load). Then, these points are grouped into several clusters through the K-means algorithm. After clustering, the samples in each cluster have small variances. Furthermore, the proposed method adopts the law of total probability to combine the results obtained using CM for PLF in all clusters.

3.1 Improved K-means algorithm

After generating the whole samples, the K-means clustering is applied to divide the whole samples into several clusters. Each cluster has a cluster center. In fact, the cluster centers form a multi-state model for random variables. For the case with only one input random variable, such as the load at one bus, the obtained cluster centers
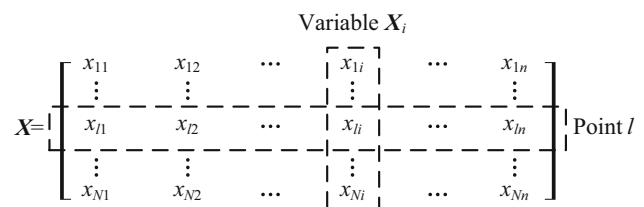

Fig. 1 Samples of input variables

correspond to multiple load levels. The analysis on the cluster centers can represent the analysis on the whole load samples. For the case with two input random variables, such as two wind farm outputs in the same area, the K-means clustering divides their samples into a number of clusters. Each cluster center is a combination of two wind power output levels, which has implied the correlation of these two wind power outputs. For clustering samples of more input variables, the K-means algorithm is conducted in the multi-dimensional Euclidean space. The detailed analyses are introduced in the following subsections.

3.1.1 General steps of K-means algorithm

Step 1: Select initial cluster centers, which is expressed as the matrix M_0.

$$M_0 = \begin{bmatrix} x_{11}^0 & x_{12}^0 & \cdots & x_{1i}^0 & \cdots & x_{1n}^0 \\ \vdots & \vdots & & \vdots & & \vdots \\ x_{j1}^0 & x_{j2}^0 & \cdots & x_{ji}^0 & \cdots & x_{jn}^0 \\ \vdots & \vdots & & \vdots & & \vdots \\ x_{K1}^0 & x_{K2}^0 & \cdots & x_{Ki}^0 & \cdots & x_{Kn}^0 \end{bmatrix} \tag{6}$$

where K is the number of clusters set in advance; x_{ji}^0 is the initial center of the variable i in the jth cluster. The jth cluster center can be expressed as: $\left(x_{j1}^0, x_{j2}^0, \cdots, x_{ji}^0, \cdots, x_{jn}^0\right)$.

Step 2: Calculate the Euclidean distance of all points to each cluster center.

$$E_d(l,j) = \sqrt{\sum_{i=1}^{n}\left(x_{li} - x_{ji}^0\right)^2} \tag{7}$$

where $E_d(l,j)$ denotes the Euclidean distance of point l to the center of the jth cluster.

Step 3: Assign all points to the closest cluster according to the Euclidean distance, and recalculate the cluster centers.

Step 4: Repeat *Steps 2* and *3* until cluster centers don't migrate.

3.1.2 Methods for improving performance of K-means

1) Selection of the initial cluster centers

The clustering performance is sensitive to the initial cluster centers, so that it is important to select them. In this paper, 10% of samples are randomly selected for clustering first. The obtained cluster centers through the first clustering can reflect the locations of cluster centers for the whole samples to some extent. Then, the obtained cluster centers are used as initial cluster centers to perform K-means clustering for the whole samples.

2) Determination of the appropriate value of K

It is necessary to determine the number of clusters before performing K-means clustering. Therefore, the weighted average radius (WAR) is proposed to evaluate the clustering performance. The WAR can be calculated as follows:

$$R = \sum_{j=1}^{K} p_j r_j \tag{8}$$

where R is the WAR; p_j is the ratio of the number of points in the jth cluster to the number of all points; r_j is the radius of the jth cluster [34].

In general, the value of WAR decreases with the increase of the number of clusters. Furthermore, the value of WAR decreases slowly once the number of clusters exceeds one value, which indicates that the quality of clustering doesn't improve significantly once the number of clusters is larger than that value. Therefore, that value is suggested as the appropriate value of K.

3) Dimensionality reduction

For improving the efficiency of K-means to handle high-dimensional samples, the singular value decomposition (SVD) can be used when the number of input variables is high. X consisting of the samples of input random variables is an $N \times n$ matrix. Carry on SVD to X:

$$X = U_x \Sigma_x V_x^T \tag{9}$$

where Σ_x is a diagonal matrix with singular values along the main diagonal; U_x and V_x^T are the left and right singular matrices derived by performing SVD on X, respectively.

The high-dimensional samples X can be converted to low-dimensional samples X' as follows:

$$X' = X V_x(1:r) \tag{10}$$

where $V_x(1:r)$ is the first r columns of V_x. The value of r is the number of singular values, whose quadratic sum exceeds 90% of the quadratic sum of all singular values [34]. It is more efficient to perform K-means on the low-dimensional samples X' than on the high-dimensional samples X.

3.1.3 Overall procedure of improved K-means algorithm

According to the methods described in above subsections, the overall procedure of the improved K-means algorithm is shown in Fig. 2, where N_{max} is the number of input variables to perform dimensionality reduction.
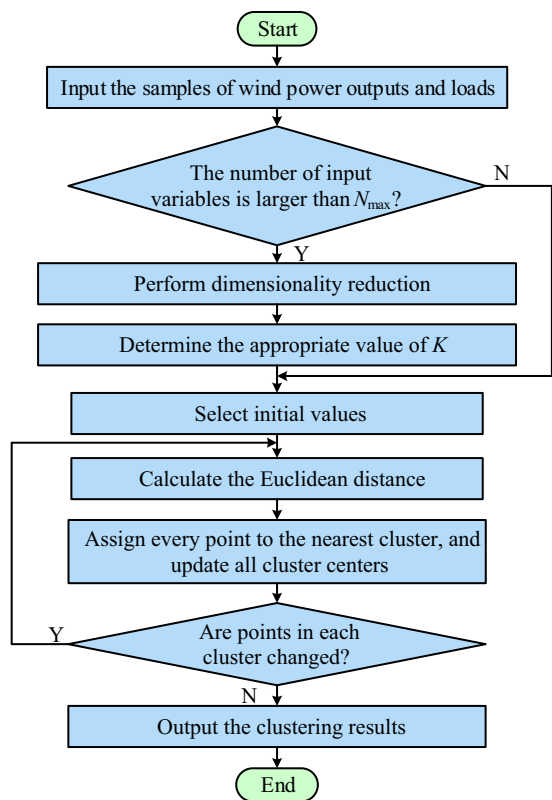
Fig. 2 Overall procedure of improved K-means algorithm

3.2 Computation of final cumulants

After the K-means clustering, a number of clusters are identified. In each cluster, the cumulant method is utilized to solve PLF. Once the computation for all clusters is completed, the law of total probability is applied to combine the moments obtained in all clusters to obtain the final cumulants of output random variables for the whole samples.

Assuming y to be one of output random variables, its final cumulants can be calculated as follows:

Step 1: The cumulants obtained using CM in each cluster can be converted to the corresponding moments.

$$\mu_r^i = \begin{cases} k_1^i & r = 1 \\ k_r^i + \sum_{j=1}^{r-1} C_{r-1}^j \mu_j^i k_{r-j}^i & r > 1 \end{cases} \tag{11}$$

where μ_r^i is the rth moment of y for the ith cluster; k_r^i is the rth cumulant of y for the ith cluster; C_{r-1}^j is the binomial coefficient, which is equal to the number of subsets of j distinct elements of $r-1$ elements.

Step 2: The final moments for the whole samples can be calculated according to the law of total probability.

$$\mu_r^y = \sum_{i=1}^{K} p_i \mu_r^i \tag{12}$$

where μ_r^y is the rth moment for the whole samples.

Step 3: The final cumulants for the original whole samples can be calculated as (13).

$$k_r^y = \begin{cases} \mu_1^y & r = 1 \\ \mu_r^y - \sum_{j=1}^{r-1} C_{r-1}^j \mu_j^y k_{r-j}^y & r > 1 \end{cases} \tag{13}$$

where k_r^y is the rth cumulant of the output variable y for the whole samples.

3.3 Procedure of solving PLF using proposed method

Figure 3 shows the flow chart of the proposed method, where k is the current cluster. A five-step procedure is described as follows.

Step 1: Apply the inverse Nataf transformation to generate wind speed and load samples. The wind power samples can be obtained according to (3).

Step 2: Apply the improved K-means to cluster the wind power and load samples into a number of clusters.

Step 3: In each cluster, the CM is used to solve probabilistic load flow considering correlations of wind power outputs and loads. The correlated samples are first transformed to uncorrelated samples using the Cholesky
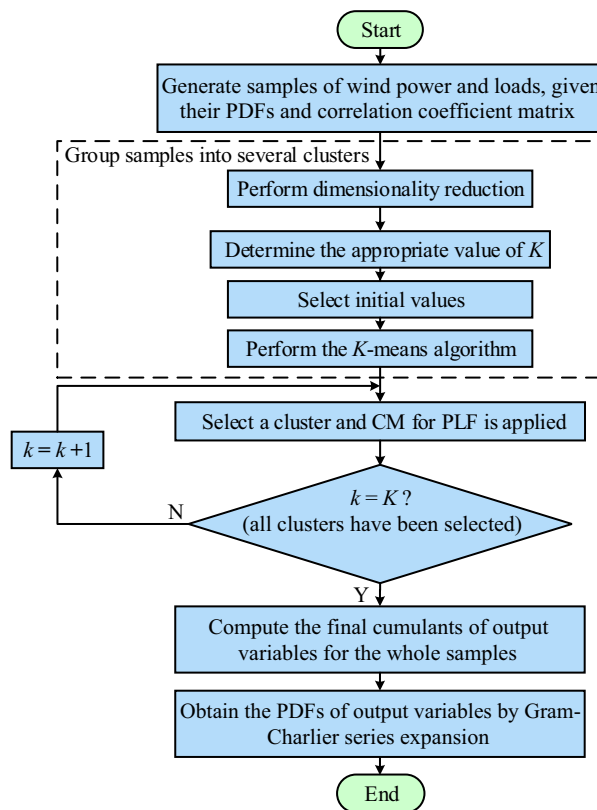

Fig. 3 Flow chart of proposed method

decomposition. Then, calculate the cumulants of uncorrelated samples [25]. Finally, the CM introduced in Section 2 are executed to calculate the cumulants of all output random variables.

Step 4: Calculate the final cumulants of output random variables using the method introduced in Section 3.2.

Step 5: Approximate the PDFs of output random variables using Gram-Charlier series expansion due to its good tail behavior [22].

This paper solves PLF problems for a determined network and does not consider equipment contingencies such as $N-1$ contingency. If equipment contingencies are required, the proposed method can be performed for each contingency. Then, the results can be combined according to the law of total probability.

4 Case study

The proposed method, namely the improved K-means based cumulant method (IKCM), is tested on modified IEEE 9-bus and 118-bus test systems [35], which are integrated with additional wind farms. Table 1 lists the particulars of wind farms. In addition, $v_{ci} = 3$ m/s, $v_r = 13$ m/s, and $v_{co} = 25$ m/s [33]. The wind farms are assumed to be PQ buses, whose power factors are kept constant at 0.85 lag [13]. In these two cases, MCS with 20000 samples is applied to solve PLF, and its results are treated as the benchmark to assess the accuracy and efficiency of the proposed method. In addition, the uncorrelated CM (UCM), the correlated CM (CCM), the $2m+1$ PEM and LHS-based MCS are conducted for comparison purpose. The UCM does not consider correlations of input random variables. The CCM handles correlated input random variables using the Cholesky decomposition. The $2m+1$ PEM is proposed in [15]. Reference [6] proposed an LHS-based PLF method and proved that it could obtain accurate results by hundreds of simulations. In this paper, the LHS-based MCS is conducted with 500 samples. The errors of cumulants and PDFs obtained using IKCM, UCM, CCM, $2m+1$ PEM and LHS are measured by the indices of absolute percent error (APE) and average root mean square (ARMS), as shown in (14) and (15), respectively. The

programs are developed using MATLAB and are executed on a PC with 2.6 GHz Intel (R) Core (TM) i5 duo processor and 8 GB DDR3 RAM.

$$APE = |(r_o - r_{MCS})/r_{MCS}| \times 100\% \tag{14}$$

$$ARMS = \frac{1}{N_p} \sqrt{\sum_{i=1}^{N_p} (OM_i - MCS_i)^2} \tag{15}$$

where r_o is the cumulant value obtained using different methods except MCS; r_{MCS} is the cumulant value obtained using MCS; OM_i denotes the value of the ith point on the PDFs obtained using different methods except MCS; MCS_i denotes the value of the ith point on the PDFs obtained using MCS; N_p is the number of points on PDFs.

4.1 Case 1: modified IEEE 9-bus test system

4.1.1 Basic information

In modified IEEE 9-bus test system, all loads have constant power factors. The active load demand at each bus is modeled as a Gussian distribution, whose mean is provided in MATPOWER [35] and standard deviation is equal to 10% of its mean. Weibull distributions are used to model wind speeds. Table 2 lists the shape and scale parameters of wind speeds [36]. The correlation coefficient between loads is assumed to be 0.8, the correlation coefficient between wind speeds is assumed to be 0.76, and the correlation coefficient between the wind speed and load at the same bus is assumed to be 0.2 [33]. The PDFs of active load power and wind power at bus 7 are depicted by histograms as shown in Figs. 4 and 5.

4.1.2 Performance of improved K-means clustering

According to (12), the relationship between WAR and the number of clusters can be obtained, as shown in Fig. 6.

It can be observed that the WAR declines slowly after the number of clusters is more than 40. This implies that the clustering performance will not significantly improve when the number of clusters is above 40. Therefore, the K value is suggested to be 40.

Table 1 Particulars of wind farms

Test system	Bus No.	Rated capacity (MW)
IEEE 9-bus system	7, 9	60
IEEE 118-bus system	17, 30, 59, 80, 92, 100	250

Table 2 Parameters of wind speeds (case 1)

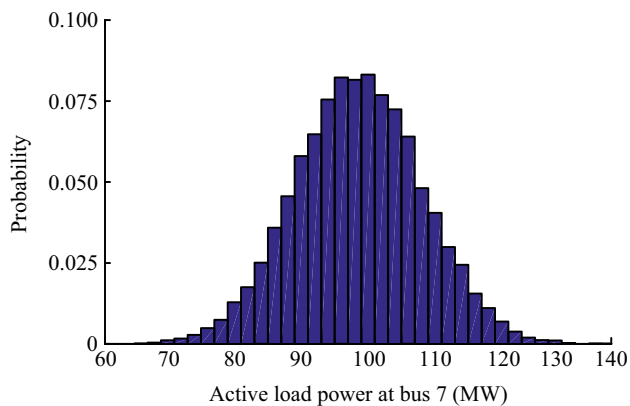| Wind speed	Shape parameter	Scale parameter
W1 (bus 7)	1.732	6.611
W2 (bus 9)	2.036	7.933

Fig. 4 PDF of active load power at bus 7
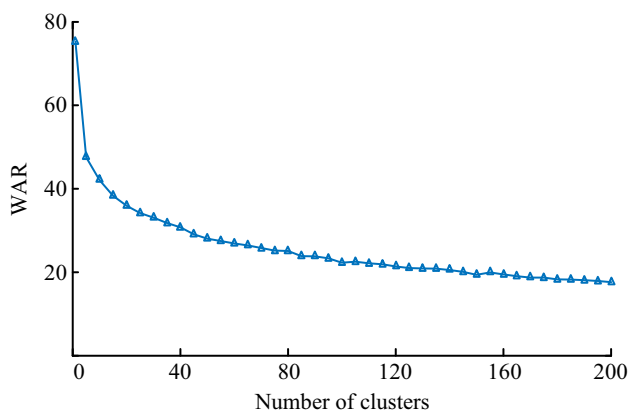

Fig. 5 PDF of wind power at bus 7

Fig. 6 Relationship between WAR and number of clusters

Table 3 Comparison of variances of input variables

Input variable	S_{mean} (MW^2)	S_{min} (MW^2)	S_{max} (MW^2)	S for whole samples (MW^2)
Wind power at bus 7	32.45	7.48	57.79	320.25
Wind power at bus 9	34.68	15.44	49.88	349.94
Load at bus 5	42.78	16.14	66.02	80.04
Load at bus 7	39.65	17.83	61.91	100.48
Load at bus 9	40.03	20.13	66.37	157.09

level of each input random variable in each cluster. The chosen values are labelled as S_{min}, S_{max} and S_{mean}, respectively. The column labelled S is the variance of a specific input random variable for the original total samples. It can be observed that the variances after the improved K-means clustering are much smaller than those for the original whole samples.

4.1.3 Probabilistic results

Table 4 lists the results of different methods used to solve PLF problems for this test system. The results are aggregated into: VA which stands for voltage angles, VM which stands for voltage magnitudes, PL which stands for line active power flows, and QL which stands for line reactive power flows, since it is difficult to present all output variables individually. The columns labelled "ε_{r1}", "ε_{r2}", "ε_{r3}" and "ε_{r4}" are APE values of the first four cumulants compared with those obtained using MCS. The mean and maximum values of the APE values are shown to demonstrate the scope of APE values for a class of variables.

It can be observed from Table 4 that all APE values obtained using IKCM are small, which indicates that the cumulants obtained using IKCM are approximately the same as those of MCS. The worst APE value of the proposed method's results is 59.29%, which occurs at the ε_{r4} of VM at bus 5. However, the actual error for this VM is only -3.66×10^{-11} p.u. This variable may mislead the comparison on APE and should not be applied to assess the performance of different methods. Compared with IKCM, the $2m+1$ PEM has smaller values of ε_{r1}, but has much larger values of ε_{r3} and ε_{r4}. The CCM has much larger values of ε_{r1}, ε_{r2}, ε_{r3} and ε_{r4} than IKCM. There are two points to be pointed out about the CCM. First, the values of ε_{r3} and ε_{r4} are much larger than the values of ε_{r1} and ε_{r2}. Second, the results of reactive quantities (VM and QL) are worse than those of active quantities (VA and PL), which can be significantly observed from the values of ε_{r3}. The UCM has large values of ε_{r2}, ε_{r3} and ε_{r4}. Compared with

Table 3 shows the clustering results of the K-means algorithm. After clustering, input random variable samples are grouped into 40 clusters. The variance can reflect the fluctuation of one random variable. For each cluster, the variance of the random variable is calculated. As a result, 40 variance values corresponding to 40 clusters are obtained. Among these 40 values, we choose the minimum, the maximum and the mean value to present the fluctuation

Table 4 Comparison of first four cumulants (case 1)

Variable	Method	ε_{r1} (%)		ε_{r2} (%)		ε_{r3} (%)		ε_{r4} (%)	
		Mean	Max	Mean	Max	Mean	Max	Mean	Max
VA	IKCM	0.13	0.34	0.04	0.06	0.32	0.70	6.06	10.68
	CCM	0.89	2.26	2.59	3.47	44.84	50.62	39.07	56.61
	UCM	0.89	2.26	21.88	23.90	46.72	58.50	76.78	80.17
	2m+1 PEM	0.01	0.03	0.49	0.55	52.30	60.14	182.98	330.88
	LHS	0.28	0.84	0.51	0.90	9.29	16.40	14.48	20.49
VM	IKCM	0.01	0.01	0.22	0.34	6.44	14.99	13.59	59.29
	CCM	0.05	0.08	4.42	6.22	157.13	305.96	64.70	93.03
	UCM	0.05	0.08	28.34	61.27	165.10	477.64	85.65	190.82
	2m+1 PEM	0.00	0.00	2.42	4.42	64.47	192.03	289.63	892.59
	LHS	0.00	0.01	0.51	0.95	43.66	128.37	145.17	750.09
PL	IKCM	0.05	0.13	0.02	0.03	0.57	2.49	17.50	36.40
	CCM	0.28	0.69	5.34	10.97	35.42	55.30	58.22	140.90
	UCM	0.28	0.69	29.48	69.38	61.25	139.98	170.57	470.08
	2m+1 PEM	0.00	0.00	0.17	0.33	43.45	64.05	537.64	1088.21
	LHS	0.40	1.56	0.65	1.23	9.06	22.38	29.70	71.33
QL	IKCM	0.71	1.51	0.20	0.41	5.83	27.29	10.51	48.69
	CCM	3.67	9.15	4.04	6.49	113.03	345.44	61.54	100.08
	UCM	3.67	9.15	36.14	107.65	170.13	1034.79	125.36	349.20
	2m+1 PEM	0.01	0.03	1.69	6.87	77.73	278.20	394.61	1308.17
	LHS	0.69	2.01	0.67	1.92	28.09	127.52	36.37	149.98

IKCM, LHS produces slightly larger errors. However, LHS has much smaller ε_{r1} and ε_{r2} than CCM and UCM, and has significantly smaller ε_{r3} and ε_{r4} than 2m+1 PEM. The second, third and fourth cumulants can reflect the variance, skewness and kurtosis of an output random variable, respectively. Therefore, the large values of ε_{r2}, ε_{r3} and ε_{r4} can result in distortions on PDF curves. The PDF curves of output variables (VA at PV and PQ buses, VM at PQ buses, PL and QL) are approximated using 7-order Gram-Charlier series expansion.

Figures 7 and 8 show the PDFs of the VM at bus 9 and the active power flow in line 7–8, respectively. From the comparison in Figs. 7 and 8, the PDFs of IKCM can better match MCS histograms than CCM, UCM and 2m+1 PEM. The PDFs of LHS are close to those of IKCM.

Figure 9 shows the ARMS results of PDF curves of all output random variables. The box plots corresponding to IKCM are all below those corresponding to CCM, UCM, 2m+1 PEM and LHS, which indicates that the PDFs produced by IKCM can approximate those of MCS better. Comparison between LHS and IKCM shows that the ARMS values of the PDFs of LHS are also small, and that the PDFs of LHS are close to those of IKCM. It should be pointed out that in Fig. 9d, there are two exception values in the box plots of IKCM. However, the actual ARMS values are only 0.0995% and 0.1170%. It can be seen that
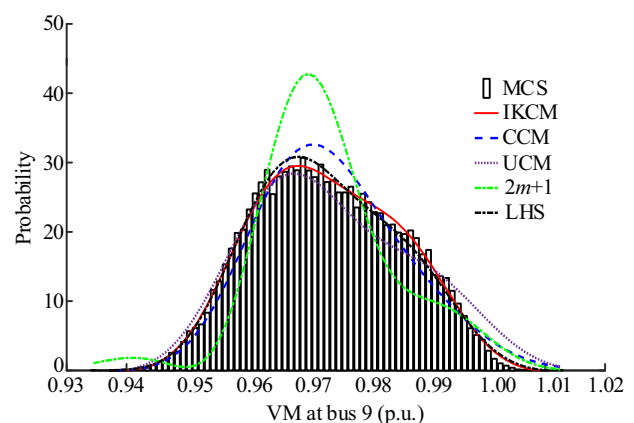
Fig. 7 PDFs of VM at bus 9

the CCM perform worse on reactive quantities than active quantities. This characteristic can also be observed from the results of cumulants and PDFs. The reason is that reactive quantities generally have higher degree of nonlinearity than active quantities. Moreover, the computation time of each method consumes and the number of deterministic power flow (DPF) calculations conducted by each method are shown in Table 5. It can be seen that the proposed method spends much less time than LHS and MCS.
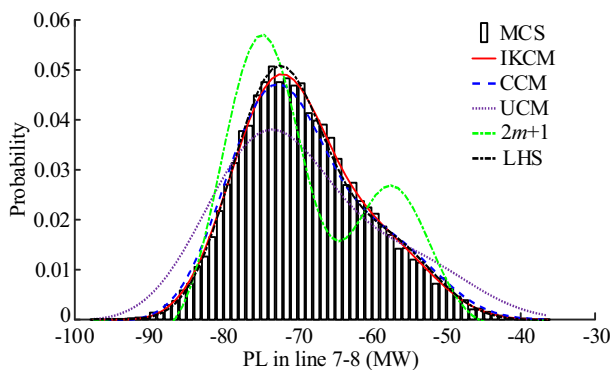
Fig. 8 PDFs of PL in line 7-8

An additional experiment is conducted to examine the performance of the proposed method with more clusters, where the proposed method with 60 clusters is implemented on this test system. The results indicate that the proposed method with 60 clusters is more accurate than 40 clusters. For example, the ε_{r1}, ε_{r2}, ε_{r3} and ε_{r4} of QL in line 1–4, which are obtained using 40 clusters, are 1.45%, 0.41%, 8.64% and 18.18%, respectively. These values obtained using 60 clusters are 1.22%, 0.30%, 5.96% and 15.97%, respectively. It can be seen that more clusters will produce more accurate results. Obviously, more clusters will require more computation time.

It can be concluded that the proposed method has higher computational accuracy than CCM, UCM, 2m+1 PEM and LHS, and is more efficient than LHS and MCS. In addition, more clusters can achieve higher accuracy at the expense of efficiency.

4.2 Case 2: modified IEEE 118-bus test system

The modified IEEE 118-bus test system is used to examine the feasibility of the proposed method for a large system with multiple wind farms. Weibull distributions are used to model wind speeds. Table 6 lists the shape and scale parameters of wind speed distributions [36]. The correlations of wind speeds at buses 17 and 30, buses 59 and 80, and buses 92 and 100 are set to be 0.88, and others are set to be 0.48. All loads have constant power factors. The active load demand at each bus is modeled as Gussian distribution, whose mean is provided in MATPOWER [35] and standard deviation is equal to 10% of its mean.

The relationship between WAR and the number of clusters for case 2 can be obtained using (12). The number of clusters is suggested to be 40. In this test system, there are 105 input random variables, including 6 wind power outputs and 99 load demands. Therefore, the dimensionality reduction based on SVD is applied in the K-means process, where the first six singular values are selected and
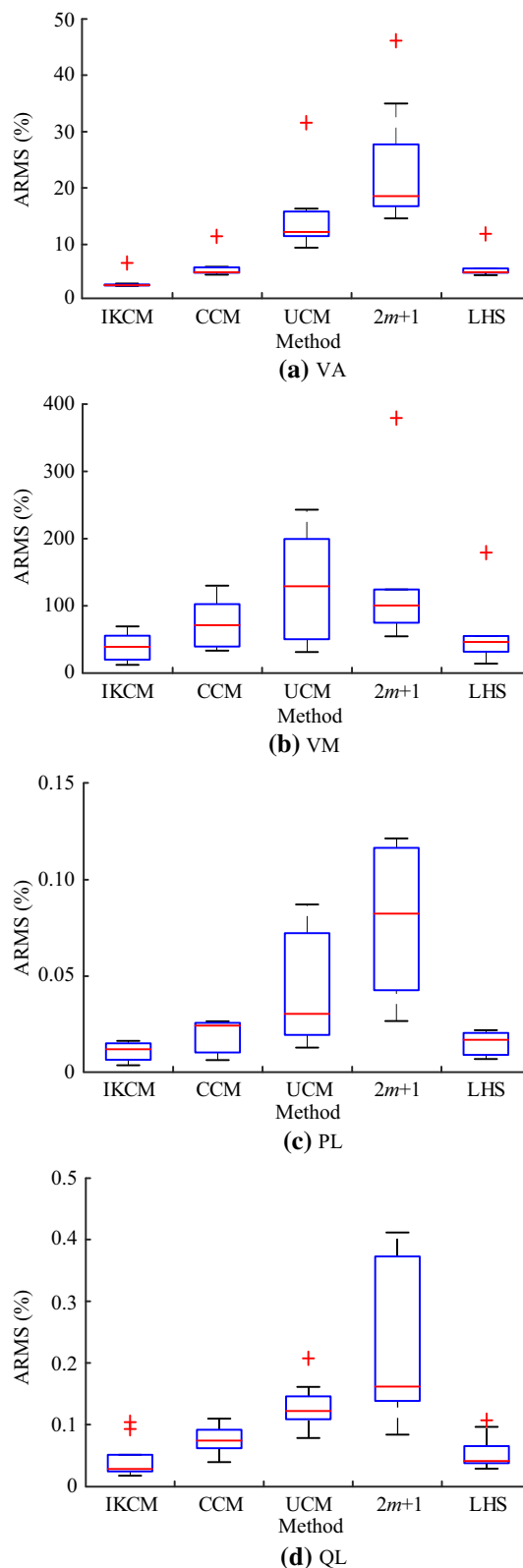

Fig. 9 Comparisons of ARMS of PDF curves for case 1

their sum is equal to 92.13% of the quadratic sum of all singular values. The computation times of the traditional

K-means and the improved *K*-means with SVD are 3.08 s and 0.92 s, respectively. It can be seen that the *K*-means algorithm achieves an efficiency improvement through the dimensionality reduction based on SVD.

Table 5 Comparison of computation time (case 1)

Method	Number of DPF	Computation time (s)
MCS	20000	120.07
IKCM	40	2.23
CCM	1	0.88
UCM	1	0.77
$2m+1$ PEM	11	0.94
LHS	500	4.46

Table 6 Parameters of wind speeds (case 2)

Wind speed	Shape parameter	Scale parameter
W1 (bus 17)	2.027	7.647
W2 (bus 30)	1.513	8.135
W3 (bus 59)	1.526	5.612
W4 (bus 80)	1.946	9.872
W5 (bus 92)	1.619	6.124
W6 (bus 100)	1.690	8.522

Table 7 presents the results of cumulants. Figures 10 and 11 show the PDFs of the PL in line 100–101 and the QL in line 79–80. It is of note that the proposed method has slightly large ε_{r3} and ε_{r4} values for very few system variables. The reason is that the Cholesky decomposition algorithm used to handle correlations has some errors for high-order cumulants when input random variables are non-normal distributions. The final PDFs of these output random variables obtained using the proposed method still have low ARMS values as shown in Fig. 12. The comparison between Table 8 and Table 5 shows that the $2m+1$ PEM does not have the obvious advantage of computational efficiency over the proposed method when solving PLF problems for this large test system. This is because the $2m+1$ PEM conducts 211 load flow simulations for this large test system due to 105 input random variables. According to the results for this IEEE 118-bus test system, the same conclusion that the proposed method has higher computational accuracy than CCM, UCM, $2m+1$ PEM and LHS, and spends much less time than LHS and MCS, can be achieved.

4.3 Discussion about stability of proposed method

The proposed method is based on the clustering algorithm. Theoretically, the result of clustering is the local

Table 7 Comparison of first four cumulants (case 2)

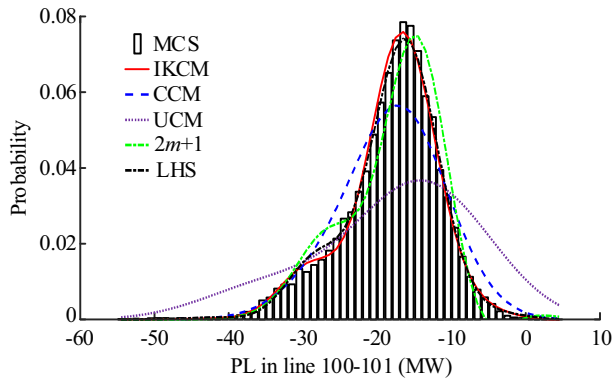| Variable	Method	ε_{r1} (%)		ε_{r2} (%)		ε_{r3} (%)		ε_{r4} (%)	
		Mean	Max	Mean	Max	Mean	Max	Mean	Max
VA	IKCM	0.09	0.17	0.05	0.11	0.15	0.66	0.43	2.16
	CCM	0.60	1.18	8.82	12.03	23.14	47.21	64.93	96.58
	UCM	0.60	1.18	58.18	68.88	79.40	92.00	93.89	106.11
	$2m+1$ PEM	0.00	0.01	0.23	0.64	15.39	51.18	97.75	338.40
	LHS	0.10	0.21	0.81	2.72	5.01	13.36	2.39	11.04
VM	IKCM	0.01	0.01	0.59	1.31	6.06	40.35	17.66	163.22
	CCM	0.07	0.14	18.40	40.71	81.79	221.30	82.96	120.81
	UCM	0.07	0.14	39.53	54.11	84.11	190.21	89.87	116.30
	$2m+1$ PEM	0.00	0.00	4.05	15.55	107.40	841.54	1681.23	13432.02
	LHS	0.00	0.00	1.00	2.53	39.89	287.79	216.51	1878.25
PL	IKCM	0.20	4.51	0.05	0.21	4.45	106.31	13.64	394.90
	CCM	1.37	28.63	7.69	26.27	32.16	87.26	74.51	354.55
	UCM	1.37	28.63	38.92	248.66	59.43	293.62	124.98	2404.25
	$2m+1$ PEM	0.00	0.07	0.16	1.20	23.17	114.68	881.86	25466.93
	LHS	0.35	7.11	2.61	9.56	31.15	1125.82	53.84	1509.69
QL	IKCM	0.72	14.04	0.29	1.76	3.10	47.44	5.93	55.66
	CCM	5.65	96.93	9.76	59.06	91.56	2547.54	101.59	1580.87
	UCM	5.65	96.93	38.21	243.48	80.55	606.22	137.32	2420.75
	$2m+1$ PEM	0.01	0.18	1.32	26.49	59.89	2169.29	593.86	6721.76
	LHS	0.33	6.13	2.59	8.45	15.44	270.66	56.91	1335.39

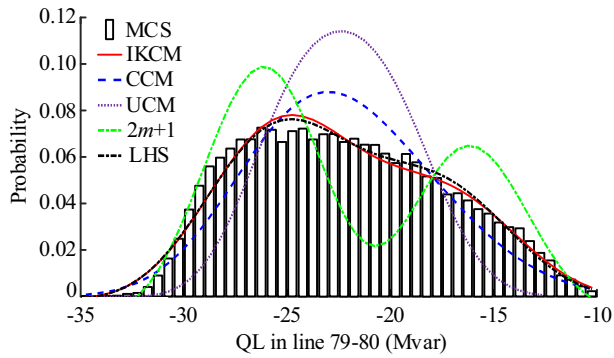Fig. 10 PDFs of PL in line 100-101

Fig. 11 PDFs of QL in line 79-80

optimal solution, which is influenced by the initial cluster centers. In order to examine the stability of the proposed method, probabilistic power flow for the modified IEEE 118-bus test system is conducted 100 times with random initial cluster centers. The APE values of the first four cumulants of each type of variables obtained in each simulation are summed and averaged. In Table 9, the columns labelled with $\varepsilon_{r1,mean}$, $\varepsilon_{r2,mean}$, $\varepsilon_{r3,mean}$ and $\varepsilon_{r4,mean}$ are the mean values of APE values of the first four cumulants for 100 simulations. It can be seen that the errors in Table 9 are approximately equal to the corresponding values of the proposed method in Table 7, which demonstrates that the proposed method can achieve stable and accurate results.

5 Conclusion

A novel PLF method considering large-scale wind power integration is proposed in this paper. In the process of the proposed method, an improved K-means algorithm is used to cluster the samples of input random variables, and the law of total probability is applied to combine the results obtained in each cluster. From the case studies on modified
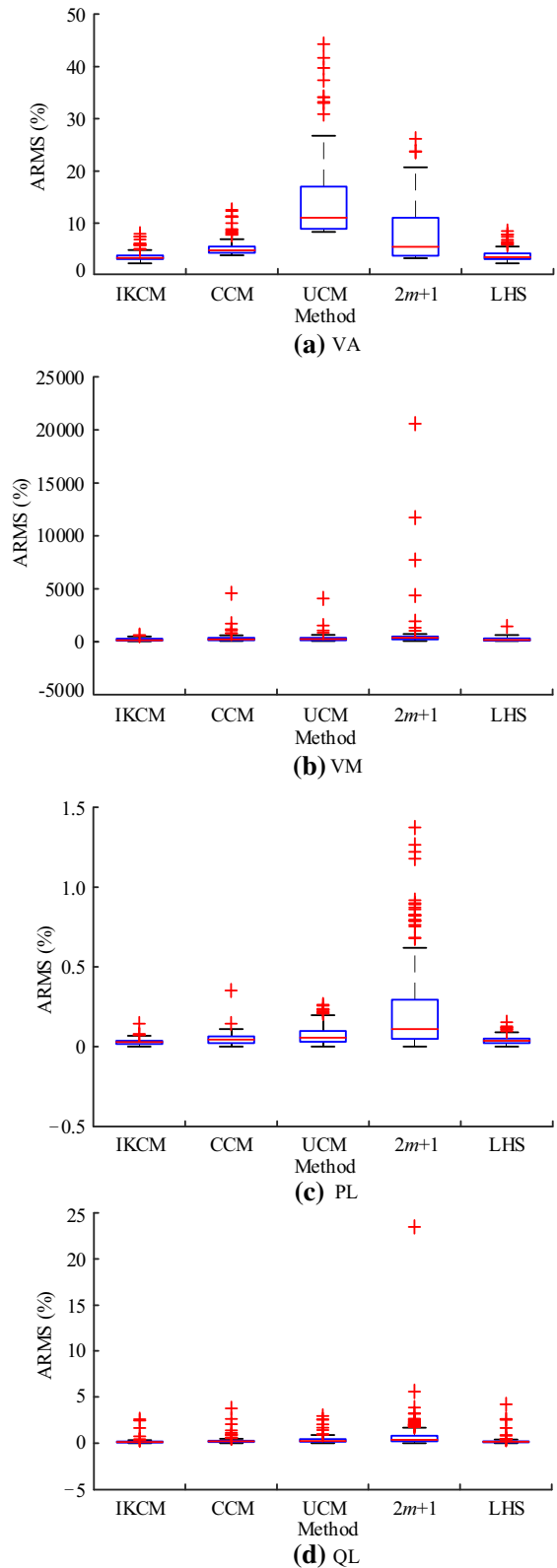

Fig. 12 Comparisons of ARMS of PDF curves for case 2

Table 8 Comparison of computation time (case 2)

Method	Number of DPF	Computation time (s)
MCS	20000	191.36
IKCM	40	13.60
CCM	1	9.24
UCM	1	8.83
$2m+1$ PEM	211	11.12
LHS	500	22.55

Table 9 Errors of first four cumulants for 100 simulations

Variable	$\varepsilon_{r1,mean}$ (%)	$\varepsilon_{r2,mean}$ (%)	$\varepsilon_{r3,mean}$ (%)	$\varepsilon_{r4,mean}$ (%)
VA	0.09	0.05	0.19	0.42
VM	0.01	0.55	6.53	18.15
PL	0.20	0.05	4.77	14.50
QL	0.71	0.28	3.14	6.69

IEEE 9-bus and 118 bus test systems, some conclusions are drawn as follows:

1) To solve PLF considering large-scale wind power integration, the proposed method can achieve higher accuracy than traditional CM, $2m+1$ PEM and LHS, and higher efficiency than LHS and MCS. In other words, the proposed method can achieve a better performance with consideration of both computational efficiency and accuracy.
2) More clusters will produce more accurate results at the expense of time. The suggested number of clusters should be determined in advance.
3) The traditional CM considering the correlation of input random variables generally has significant errors for reactive quantities.
4) The $2m+1$ PEM has accurate results for the first two cumulants but not for the third and fourth cumulants, which results in significant errors in PDFs of output variables.

In conclusion, as the proposed method has been tested on the small and large test systems, it can provide an accurate and efficient tool for power system planning and operation with large-scale wind power.

References

[1] Prusty BR, Jena D (2016) A critical review on probabilistic load flow studies in uncertainty constrained power systems with photovoltaic generation and a new approach. Renew Sust Energ Rev 69:1286–1302
[2] Glynn PW, Iglehart DL (1989) Importance sampling for stochastic simulations. Management Science 35(11):1367–1392
[3] Yu H, Chung CY, Wong KP et al (2009) Probabilistic load flow evaluation with hybrid Latin hypercube sampling and Cholesky decomposition. IEEE Trans Power Syst 24:661–667
[4] Luo G, Chen JF, Cai DF et al (2013) Probabilistic assessment of available transfer capability considering spatial correlation in wind power integrated system. IET Gener Transm Distrib 7:1527–1535
[5] Hajian M, Rosehart WD, Zareipour H (2013) Probabilistic power flow by Monte Carlo simulation with Latin supercube sampling. IEEE Trans Power Syst 28:1550–1559
[6] Chen Y, Wen J, Cheng S (2013) Probabilistic load flow method based on Nataf transformation and Latin hypercube sampling. IEEE Trans Sustain Energy 4(2):294–301
[7] Liu Y, Gao S, Cui H et al (2016) Probabilistic load flow considering correlations of input variables following arbitrary distributions. Electr Power Syst Res 140:354–362
[8] Li B, Shahzad M, Qi B et al (2018) Probabilistic computational model for correlated wind farms using copula theory. IEEE Access 6:14179–14187
[9] Cui T, Franchetti F (2013) A quasi-Monte Carlo approach for radial distribution system probabilistic load flow. In: Proceedings of IEEE PES innovative smart grid technologies conference (ISGT), Washington DC, USA, 1-6 February 2013, pp 1–6
[10] Chen W, Yan HQ, Pei XP et al (2016) A quasi Monte Carlo probabilistic load flow method of distribution system containing distributed generation and electric vehicle charging load based on Sobol sequence. In: Proceedings of China international conference on electricity distribution (CICED), Xi'an, China, 10–13 August 2016, pp 1–7
[11] Xu X, Yan Z (2017) Probabilistic load flow calculation with quasi-Monte Carlo and multiple linear regression. Int J Electr Power Energy Syst 88:1–12
[12] Su CL (2005) Probabilistic load-flow computation using point estimate method. IEEE Trans Power Syst 20:1843–1851
[13] Aien M, Khajeh MG, Rashidinejad M et al (2014) Probabilistic power flow of correlated hybrid wind-photovoltaic power systems. IET Renew Power Gen 8:649–658
[14] Morales JM, Perez-Ruiz J (2007) Point estimate schemes to solve the probabilistic power flow. IEEE Trans Power Syst 22:1594–1601
[15] Morales JM, Baringo L, Conejo AJ et al (2010) Probabilistic power flow with correlated wind sources. IET Gener Transm Distrib 4:641–651
[16] Wang X, Gong Y, Jiang C (2015) Regional carbon emission management based on probabilistic power flow with correlated stochastic variables. IEEE Trans Power Syst 30:1094–1103
[17] Gupta N, Daratha N (2017) Probabilistic three-phase load flow for unbalanced electrical systems with wind farms. Int J Electr Power Energy Syst 87:154–165
[18] Gupta N (2016) Probabilistic load flow with detailed wind generator models considering correlated wind generation and correlated loads. Renew Energy 94:96–105
[19] Aien M, Fotuhi-Firuzabad M, Aminifar F (2012) Probabilistic load flow in correlated uncertain environment using unscented transformation. IEEE Trans Power Syst 27:2233–2241
[20] Wan C, Xu Z, Dong ZY et al (2012) Probabilistic load flow computation using first-order second-moment method. In:

Proceeding of the 2012 IEEE power and energy society general meeting, San Diego, USA, 22–26 July 2012, pp 1--6

[21] Liu H, Tang C, Han J et al (2017) Probabilistic load flow analysis of active distribution network adopting improved sequence operation methodology. IET Gener Transm Distrib 11(9):2147–2153

[22] Zhang P, Lee ST (2004) Probabilistic load flow computation using the method of combined cumulants and Gram-Charlier expansion. IEEE Trans Power Syst 19:676–682

[23] Yuan Y, Zhou J, Ju P et al (2011) Probabilistic load flow computation of a power system containing wind farms using the method of combined cumulants and Gram-Charlier expansion. IET Renew Power Gen 5:448–454

[24] Fan M, Vittal V, Heydt GT et al (2012) Probabilistic power flow studies for transmission systems with photovoltaic generation using cumulants. IEEE Trans Power Syst 27:2251–2261

[25] Cai D, Chen J, Shi D et al (2012) Enhancements to the cumulant method for probabilistic load flow studies. In: Proceedings of 2012 IEEE power and energy society general meeting, San Diego, USA, 22–26 July 2012, pp 1–8

[26] Ran X, Miao S (2016) Three-phase probabilistic load flow for power system with correlated wind, photovoltaic and load. IET Gener Transm Distrib 10:3093–3101

[27] Prusty BR, Jena D (2016) Combined cumulant and Gaussian mixture approximation for correlated probabilistic load flow studies: a new approach. CSEE J Power Energy Syst 2(2):71–78

[28] Sui BY, Hou K, Jia HJ et al (2018) Maximum entropy based probabilistic load flow calculation for power system integrated with wind power generation. J Mod Power Syst Clean Energy 6(5):1042–1054

[29] Amid P, Crawford C (2018) A cumulant-tensor based probabilistic load flow method. IEEE Trans Power Syst 33(5):5648–5656

[30] Allan RN, Da Silva AL, Burchett RC (1981) Evaluation methods and accuracy in probabilistic load flow solutions. IEEE Trans Power App Syst 100(5):2539–2546

[31] Zhu X, Liu X, Zhang JH (2013) Probabilistic load flow method considering large-scale wind power integration. Proceedings CSEE 33:77–85

[32] Deng X, He J, Zhang P (2017) A novel probabilistic optimal power flow method to handle large fluctuations of stochastic variables. Energies 10(10):1–21

[33] Li Y, Li W, Yan W et al (2014) Probabilistic optimal power flow considering correlations of wind speeds following different distributions. IEEE Trans Power Syst 29:1847–1854

[34] Leskovec J, Ullman JD (2014) Mining of massive datasets. Cambridge University Press, Cambridge

[35] MATPOWER homepage. http://www.pserc.cornell.edu/matpower. Accessed 20 January 2017

[36] Masseran N, Razali AM, Ibrahim K (2012) An analysis of wind power density derived from several wind speed density functions: the regional assessment on wind power in Malaysia. Rebew Sust Energ Rev 16:6476–6487

Xiaoyang DENG received his B.Sc. degree from Hebei University, Hebei Province, China, in 2013. He is currently pursuing his Ph.D. degree at Beijing Jiaotong University, Beijing, China. His research interests include reliability evaluation of power system with renewable energies, reliability evaluation of integrated energy systems, and probabilistic load flow.

Pei ZHANG is Vice Dean of Hanergy Renewable School of Beijing Jiaotong University, Beijing, China. He was the Director of Smart Grid in Resource Group of Accenture and the Program Manager leading Power System Analysis, Operation and Planning group at Electric Power Research Institute (EPRI) in USA. He obtained his Ph.D. degree from Imperial College of Science, Technology and Medicine, University of London, United Kingdom. His research interests include power system stability analysis and application of advanced IT technologies to power systems.

Kangmeng JIN received her B.Sc. degree from Yanshan University, Hebei Province, China, in 2016. She is currently pursuing her M.Sc. degree at Beijing Jiaotong University, Beijing, China. Her research interests include probabilistic load flow of integrated energy systems.

Jinghan HE received the B.Sc. degree and the M.Sc. degree in automation from the Tianjin University, Tianjin, China, in 1987 and 1994, respectively. She is currently a Professor at Beijing Jiaotong University, Beijing, China. Her main research interests include protective relaying, fault distance measurement, and location in power systems.

Xiaojun WANG received the B.Sc., M.Sc., and Ph.D. degrees in electrical engineering from North China Electric Power University, Beijing, China, in 2001, 2004, and 2008, respectively. He is currently an Associate Professor with Beijing Jiaotong University, Beijing, China. His main research interests include protection relaying and fault location in active distribute system as well as analysis and control in integrated energy system.

Yuwei WANG is currently pursuing her B.Eng. degree in electrical and electronic engineering from Imperial College London, London, UK. Her main research interests include FPGA hardware acceleration of image processing and wearable intelligent devices.