

# An Analytic Approach to Probabilistic Load Flow Incorporating Correlation Between Non-Gaussian Random Variables

Yu Huang<sup>1</sup>, Qingshan Xu<sup>1</sup>, Xianqiang Jiang<sup>1</sup>, Yang Yang<sup>1</sup>, Guang Lin<sup>2</sup>

<sup>1</sup>*School of Electrical Engineering, Southeast University,  
Sipailou Campus of Southeast University,  
Nanjing, China*

<sup>2</sup>*Department of Mathematics, Purdue University,  
West Lafayette, IN, USA  
huangyu@seu.edu.cn*

**Abstract**—This paper presents a cumulant-based method for probabilistic load flow (PLF) analysis which incorporates correlation between input random variables. Our approach can approximate non-Gaussian variables of all kinds (e.g. different load profiles or renewable power injections) accurately using the Gaussian mixture model (GMM), which also facilitates the computation of cumulants in a straightforward numerical way. Multiple correlations can be easily handled by transforming correlated variables into a combination of uncorrelated ones. To reduce the deviations introduced by traditional series expansions such as Edgeworth or Cornish-Fisher series, we use C-type Gram-Charlier series instead, which can better predict the probabilistic tail regions and have good convergence property as well. The good performance of the proposed method is verified using the IEEE 30 test system in terms of accuracy and efficiency.

**Index Terms**—Probabilistic load flow; Gaussian mixture model; Correlation; Gram-Charlier expansion.

## I. INTRODUCTION

A surge of grid complexity associated with increasing level of uncertainties, such as high penetration of weather-dependent renewable energies (REs) and plug-in hybrid electric vehicles (PEVs) has brought crucial challenges to system planning and operation [1]–[3]. Moreover, correlation exists among REs in neighbouring geographical areas and loads with the same social factors [4]. In order to accurately evaluate the variation of network parameters such as node voltages or line flows, both the uncertainty and correlation should be taken into account.

Probabilistic load flow, which incorporates dependence between random variables, is an efficient tool in probabilistic analysis as it enables a comprehensive assessment of system working conditions, thus could inform system operators of the weak points and potential crisis under various uncertainties [5]. First proposed by Borkowska in 1974 [6], PLF has gained

its popularity among researchers and much progress has been made in recent years.

In general, PLF methods can be categorized into Monte Carlo simulation (MCS) [7] and analytic methods [8], [9]. MCS is the most straightforward way of solving the problem which involves repeated simulation to obtain probability distribution of output variables, and is widely used in large-sized systems since it proves to be dimensional-independent [10]. Though the excessive CPU time of MCS makes it less attractive, it serves as a benchmark to evaluate the effectiveness of other PLF solutions for its high accuracy and being easy to implement. Recently, with the development of high-performance computing technology [11] and advanced sampling techniques [12], the computational burden of MCS has been reduced dramatically, but it is still an ongoing research area.

Alternatively, analytic methods are more desirable in that they are less computationally intensive and with acceptable accuracy. In [13], the probability density function (PDF) of output variables was obtained by a convolutional-based method using simplified DC load flow. To accelerate the time-consuming process of convolution, Fast Fourier Transformation (FFT) was applied in [14], and Zhang *et al.* [15] presented a cumulant method (CM) by combining the cumulants of variables and Gram-Charlier expansion. Although the use of different series expansions (e.g. Edgeworth series, Cornish-Fisher series) show satisfactory performance when input variables are Gaussian or near-Gaussian, they may have serious convergence problems and large deviations for non-Gaussian distributions [16]. Another prevalent proposal is the point estimate method (PEM), which selects a set of quadrature points to estimate the moments of output variables [17], [18]. However, the computational time of PEM and its various versions is proportional to the number of uncertain variables, making it impractical in large-scale systems.

The dependency of loads and renewable power generation must be considered and revealed in PLF solutions. In [19], linear coefficient was used to measure the degree of linear dependence between different distributed photovoltaic (PV)

Manuscript received 18 August, 2017; accepted 19 January, 2018.

This research was funded by a grant (BY2016076-12) from the Industry-University-Research Collaboration of Jiangsu Province and supported by the Scientific Research Foundation of Graduate School of Southeast University.

units. Rank correlation method, using Spearman or Kendall's correlation coefficients, could keep it invariant under strictly increasing transformations [20]. To give a complete representation of relationship between input variables, the multivariate joint distribution was obtained from Copula theory in [21]. The aforementioned techniques could generate dependent random samples with given correlation, thus have been widely used in MCS and PEM. However, there are few analytic methods that take into account the correlation. Reference [22] proposed a CM considering correlation between load demands, with all load PDFs being Gaussian. Though the Gaussian assumption could largely simplify the modelling difficulty, sometimes the deviations are not negligible due to the variety of load profiles and large renewable injections.

The main contribution of the paper is to develop an improved CM for PLF studies, which can conveniently handle the correlated variables. It has made big progress by solving the subsequent problems in original CM, such as a) calculating the cumulants for non-Gaussian variables, b) considering the dependency without sacrificing computation speed, c) accelerating convergence rate, and d) avoiding negative PDF values. To do this, the non-Gaussian variables are first modelled with GMM, then numerical solution to cumulants is proposed. Correlation matrix is decomposed to modify the transformation matrices, followed by PDF reconstructions using C-type Gram-Charlier series.

The rest of the paper is organized as follows. Section II introduces the uncertainty modelling of GMM and its reduction algorithm. Section III gives a numerical way of computing cumulants of non-Gaussian variables and approximates output PDFs with C-type Gram-Charlier series. In Section IV, the computational procedure for PLF solution is presented. Section V shows some test results and error analysis, while concluding remarks are made in Section VI.

## II. MODELLING NON-GAUSSIAN POWER INJECTIONS USING GMM

### A. Gaussian Mixture Distribution

A GMM can be viewed as a weighted sum of several Gaussian distribution components, whose PDF, for one-dimensional case, can be defined as [23]

$$f_X(x) = \sum_{i=1}^n \omega_i \times f_{N(\mu_i, \sigma_i^2)}(x), \quad (1)$$

where  $\omega_i$ ,  $\mu_i$  and  $\sigma_i$  are the proportion, mean and standard deviation of the  $i$ th component of the mixture, respectively. The proportional coefficients must subject to the constraint as:

$$\begin{cases} 0 < \omega_i \leq 1, \\ \sum_{i=1}^n \omega_i = 1. \end{cases} \quad (2)$$

The  $i$ th Gaussian component is well-known as

$$f_{N(\mu_i, \sigma_i^2)}(x) = \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-(x-\mu_i)^2/2\sigma_i^2}. \quad (3)$$

For determining the parameters of GMM, expectation maximization (EM) algorithm is commonly used that one may refer to [24]. The biggest advantage of GMM is that it could approximate any PDF with just a few simple forms of Gaussian components, making it convenient for further applications. This is especially useful when a random variable does not fit the typical distribution. Figure 1 illustrates the example of wind power output modelled by GMM with five components, from which the close resemblance of the shape of discrete PDF and that of the GMM is noted.

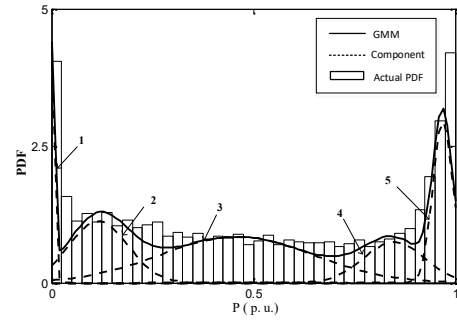


Fig. 1. GMM approximation of wind power output.

### B. Reduction of Components

The accuracy of GMM is dependent on the number of its components. If one starts from the simplest GMM (with only 2 or 3 components) and then increases the number of components by running EM algorithm one at a time, it would be both time-consuming and impractical. Instead, we introduce a reduction method that reduces one component at a time until the desired number of components is reached.

The method identifies a pair of components to be merged that gives the minimal discrepancy from the original GMM. A metric index, the integral square difference (ISD) defined in [25] is used to quantify this difference. All combinations of pairs need to be evaluated and once the pair of components  $i$  and  $j$  are selected, they are merged to form a new Gaussian component  $ij$ , and the parameters are determined as:

$$\bar{\omega}_{ij} = \bar{\omega}_i + \bar{\omega}_j, \quad (4)$$

$$\bar{\mu}_{ij} = \frac{1}{\bar{\omega}_i + \bar{\omega}_j} (\bar{\omega}_i \bar{\mu}_i + \bar{\omega}_j \bar{\mu}_j), \quad (5)$$

$$\bar{\sigma}_{ij}^2 = \frac{1}{\bar{\omega}_i + \bar{\omega}_j} \left( \bar{\omega}_i \bar{\sigma}_i^2 + \bar{\omega}_j \bar{\sigma}_j^2 + \frac{\bar{\omega}_i \bar{\omega}_j}{\bar{\omega}_i + \bar{\omega}_j} (\bar{\mu}_i - \bar{\mu}_j)^2 \right). \quad (6)$$

## III. PROBABILISTIC LOAD FLOW FORMULATION USING IMPROVED CUMULANT METHOD

### A. Cumulants of Non-Gaussian Input Variables

The  $v$ th moment of a random variable  $X$ , whose PDF  $f(X)$  is given, can be defined as

$$\mu_v' = E(X^v) = \int_{-\infty}^{+\infty} x^v f(x) dx, \quad (7)$$

where  $\mu'_1 = m$  is the mean of the random variable and is used to calculate the central moment as follows

$$\mu'_\nu = E\left[(X - m)^\nu\right] = \int_{-\infty}^{+\infty} (x - m)^\nu f(x) dx. \quad (8)$$

However, computing the integral of (7) or (8) is not trivial especially when the random variable follows the non-Gaussian distribution. A common practice is to transform it to an equivalent normal variable, then (7) can be discretized by Gaussian numerical integration as

$$\mu'_\nu = \int_{-\infty}^{+\infty} x^\nu f(x) dx \cong \sum_{k=1}^N w_i^{(k)} \left(\mu + \sigma X_i^{(k)}\right)^\nu, \quad (9)$$

where  $w_i^{(k)}$  and  $X_i^{(k)}$  are determined as Gaussian-Hermite weight and point, respectively.  $\mu$  and  $\sigma$  are the mean and standard deviation of the normal random variable.  $N$  is the quadrature order (equal to the number of Gauss point).

The use of GMM modelling can avoid the issue of equivalent normalization, which is a burden for the computational procedure in terms of efficiency and accuracy. Applying (9) to each Gaussian component, the weighted sum of which is the statistical moment of non-Gaussian variable. Since the relationship between cumulant and central moment is deduced in [15], once the moment of a random variable is known, cumulants of different orders can be obtained.

### B. Cumulants of Sum of Correlated Random Variables

The non-linear load flow equations can be linearized around a working point as

$$\begin{cases} \mathbf{X} = \mathbf{X}_0 + \mathbf{S}_0(\mathbf{W} - \mathbf{W}_0), \\ \mathbf{Z} = \mathbf{Z}_0 + \mathbf{T}_0(\mathbf{W} - \mathbf{W}_0), \end{cases} \quad (10)$$

where  $\mathbf{W}$  represents node power injections;  $\mathbf{X}$  and  $\mathbf{Z}$  are state vector (bus voltage magnitudes and angles) and output vector (line active and reactive power), respectively; The subscript '0' denotes the linearization point;  $\mathbf{S}_0$  and  $\mathbf{T}_0$  are all sensitivity matrices. From (10), we can see that state variable  $X_i$  (or  $Z_i$ ) is the linear sum of input power injections. For independent random variables, we have the conclusion that the cumulants of a sum are the sum of cumulants [26]. Thus, the cumulants of  $X_i$  (or  $Z_i$ ) can be easily obtained provided that input variables ( $W_i$ ) are uncorrelated. In reality, however, dependency between loads or renewable injections at adjacent locations is obvious. As a consequence, we should first convert the correlated input random variables to a combination of uncorrelated ones. Suppose  $\mathbf{C}_W$  is a correlation coefficient matrix of input random variable  $\mathbf{W} = [W_1, \dots, W_i, \dots, W_n]^T$ . According to [27],  $\mathbf{C}_W$  is symmetric and positive definite so that we could decompose it through Cholesky decomposition

$$\mathbf{C}_W = \mathbf{L}\mathbf{L}^T, \quad (11)$$

where  $\mathbf{L}$  is the lower triangular matrix. If  $\mathbf{A}$  is an orthogonal matrix that meets the requirement of transforming correlated input random variables  $\mathbf{W}$  into the uncorrelated ones, we have

$$\mathbf{Y} = \mathbf{A}\mathbf{W}, \quad (12)$$

where  $\mathbf{Y} = [Y_1, \dots, Y_i, \dots, Y_n]^T$  are uncorrelated variables.

The correlation coefficient matrix of variables  $\mathbf{Y}$ ,  $\mathbf{C}_Y$ , is an identity matrix  $\mathbf{I}$ , which can be written as (13) based on (11) and its definition

$$\begin{aligned} \mathbf{C}_Y &= \rho(\mathbf{Y}, \mathbf{Y}^T) = \rho(\mathbf{A}\mathbf{W}, \mathbf{W}^T \mathbf{A}^T) = \\ &= \mathbf{A}\rho(\mathbf{W}, \mathbf{W}^T) \mathbf{A}^T = \mathbf{A}\mathbf{C}_W \mathbf{A}^T = \\ &= \mathbf{A}\mathbf{L}\mathbf{L}^T \mathbf{A}^T = (\mathbf{A}\mathbf{L})(\mathbf{A}\mathbf{L})^T = \mathbf{I}. \end{aligned} \quad (13)$$

Note that  $\mathbf{A} = \mathbf{L}^{-1}$  is indicated in (13). Then (12) can be rewritten as

$$\mathbf{Y} = \mathbf{L}^{-1}\mathbf{W}. \quad (14)$$

In this way, the correlated input random variables  $\mathbf{W}$  are finally represented by the combination of uncorrelated variables  $\mathbf{Y}$

$$\mathbf{W} = \mathbf{L}\mathbf{Y}. \quad (15)$$

The  $k$ th order cumulant of output variables is computed according to (10) on the independence assumption as:

$$\begin{cases} \Delta \mathbf{X}^{(k)} = \mathbf{S}_0^{(k)} \Delta \mathbf{W}^{(k)}, \\ \Delta \mathbf{Z}^{(k)} = \mathbf{T}_0^{(k)} \Delta \mathbf{W}^{(k)}, \end{cases} \quad (16)$$

where  $\mathbf{S}_0^{(k)}$  and  $\mathbf{T}_0^{(k)}$  are matrices whose elements are  $k$ th power of those in  $\mathbf{S}_0$  and  $\mathbf{T}_0$ .

To incorporate correlation between input variables, the sensitivity matrices  $\mathbf{S}_0$  and  $\mathbf{T}_0$  should be modified as  $\mathbf{S}_1$ ,  $\mathbf{T}_1$  by substituting (15) into (10), and (16) would be changed accordingly as

$$\begin{cases} \Delta \mathbf{X}^{(k)} = (\mathbf{S}_0 \mathbf{G})^{(k)} \Delta \mathbf{Y}^{(k)} = \mathbf{S}_1^{(k)} \Delta \mathbf{Y}^{(k)}, \\ \Delta \mathbf{Z}^{(k)} = (\mathbf{T}_0 \mathbf{G})^{(k)} \Delta \mathbf{Y}^{(k)} = \mathbf{T}_1^{(k)} \Delta \mathbf{Y}^{(k)}. \end{cases} \quad (17)$$

### C. Approximation of Output Probability Distribution

A key problem in PLF study is how to effectively estimate the distribution of output variables (bus voltage or line flow). The A-type Gram-Charlier (AGC) series, given in (18), is a prevalent approach to approximating output PDFs

$$f_X(x) = f_0(x) \left[ 1 + \sum_{i=1}^{\infty} \frac{1}{i!} c_i H_i(x) \right], \quad (18)$$

where  $c_i$  is the  $i$ th series expansion coefficient, which can be

determined by standard deviation and central moments of  $X$ .  $H_i(\cdot)$  represents the  $i$ th Hermite orthogonal polynomial.  $f_0(x)$  is the base function of standard normal PDF. Despite the handiness of AGC method, it has known limitations such as poor convergence, and might yield negative probability densities in the tail region when skewness and kurtosis of  $X$  are outside the set range. This is especially true with large-scale renewable injections, which brings about more uncertainties. Herein, we introduce the CGC series for approximation, defined as

$$f_X(x) = \frac{\exp\left[\sum_{i=1}^{\infty} \frac{1}{i} \kappa_i H_i(x)\right]}{\int \exp\left[\sum_{i=1}^{\infty} \frac{1}{i} \kappa_i H_i(x)\right] dx}, \quad (19)$$

where  $\kappa_i$  is the  $i$ th series expansion coefficient. The exponential form of CGC series in (19) could ensure the probability densities of  $X$  to be non-negative and range from 0 to 1. Compared with AGC series, CGC is considered superior not only in the convergence properties, but it gives better approximation for PDFs that largely deviate from Gaussian.

For a given order  $k$  of CGC series expansion, the approximation coefficients vector  $\boldsymbol{\kappa} = [\kappa_1, \kappa_2, \dots, \kappa_k]$  can be obtained by solving the linear equation

$$\mathbf{G}\boldsymbol{\kappa} = \mathbf{B}, \quad (20)$$

where  $\mathbf{B}$  is a  $k \times 1$  vector with  $b_1 = 0$ ,  $b_m = -(m-1)\psi_{m-2}$  ( $m = 2, 3, \dots, k$ ).  $\psi_m$  denotes the  $m$ th order Hermit moment as

$$\psi_m = \int H_m(x) f_X(x) dx = E[H_m(x)]. \quad (21)$$

$\mathbf{G}$  is a  $k \times k$  symmetric matrix with the elements being

$$g_{ij} = \sum_{m=0}^{i+j-2} \frac{1}{m!} \Delta_{i-1, j-1, m} \psi_m, \quad (22)$$

where  $\Delta_{i-1, j-1, m}$  is expressed as

$$\Delta_{i-1, j-1, m} = \begin{cases} \prod_{p=i, j, m} \frac{\Gamma(p+1)}{\Gamma(q-p+1)}, & i+j+m \in \text{even}, \\ & q = \frac{i+j+m}{2} \geq i, j, m, \\ 0, & \text{elsewhere.} \end{cases} \quad (23)$$

If we arbitrary set the expansion order  $k = 4$ , the matrix  $\mathbf{G}$  and vector  $\mathbf{B}$  could be derived as:

$$\mathbf{G} = \begin{bmatrix} 1 & m & \mu'_2 - 1 & \mu'_3 - 3m \\ \mu'_2 & \mu'_3 - m & \mu'_4 - 3\mu'_2 & \\ & \mu'_4 - 2\mu'_2 + 1 & \mu'_5 - 4\mu'_3 + 3m & \\ & & \mu'_6 - 6\mu'_4 + 9\mu'_2 & \end{bmatrix}_{4 \times 4}, \quad (24)$$

$$\mathbf{B} = \begin{bmatrix} 0 & -1 & -2m & -3(\mu'_2 - 1) \end{bmatrix}_{4 \times 1}^T. \quad (25)$$

Since  $\mathbf{G}$  is symmetric, only upper diagonal elements are displayed. Then the CDF of output variables can be calculated by a numerical integral of (19).

#### IV. COMPUTATIONAL PROCEDURE

The computational procedure of the proposed method is illustrated through flow chart in Fig. 2. Detailed steps involved are discussed below.

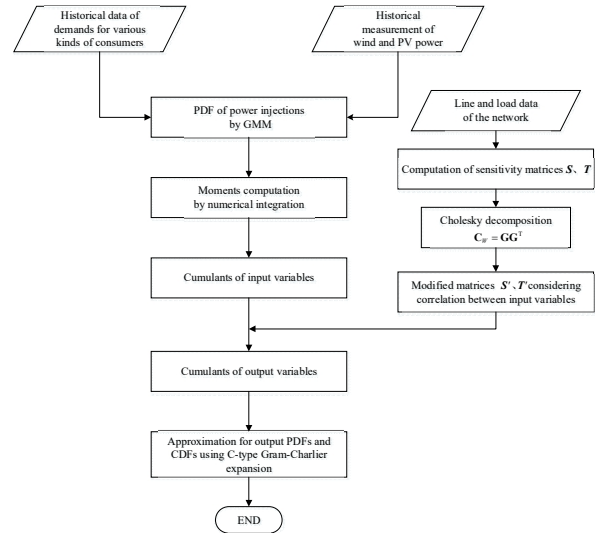


Fig. 2. Flow chart of computational procedure for the proposed method.

Step 1: Model the probabilistic distributions of all the load demands and renewable energy output using GMM with availability of historical measurement and statistical analysis. Component reduction is performed if required.

Step 2: Compute the raw and central moments of 1st to 4th order for different types of power injections. The corresponded cumulants can be obtained as well.

Step 3: Input line and load data of the network, then determine the linearized point of load flow equation and initial sensitivity matrices by Newton method.

Step 4: Transform correlated input random variables into the combination of uncorrelated ones using Cholesky decomposition, and modify the sensitivity matrices.

Step 5: Approximate the PDF and CDF of each bus voltage and line flow using CGC series, whose expansion coefficients are computed from linear algebraic operation of moments.

#### V. CASE STUDY

The proposed analytic PLF method has been applied to the IEEE 30 test system, in order to evaluate its general performance. The main results were compared to those of MCS with 10000 samples and corresponding error analysis was also carried out.

Most of the data involved were provided in [28], and we modified the system by adding two wind farms at buses 25 and 29 (100 MW and 80 MW, respectively). It is assumed that each wind farm is controlled such that the power factor is kept constant at 0.95 p.u. GMM parameters of PDFs used to model the wind power output are presented in Table I, all the data given are in p.u.

TABLE I. GMM PARAMETERS OF ACTIVE WIND POWER INJECTIONS (P.U.).

	$\mu_1$	$\mu_2$	$\mu_3$	$\mu_4$	$\mu_5$	$\sigma_1$	$\sigma_2$	$\sigma_3$	$\sigma_4$	$\sigma_5$	$\omega_1$	$\omega_2$	$\omega_3$	$\omega_4$	$\omega_5$
$P_{25}$	0.22	0.36	0.45	0.73	0.95	0.07	0.14	0.12	0.11	0.05	0.3	0.2	0.2	0.2	0.1
$P_{29}$	0.16	0.23	0.35	0.54	0.76	0.04	0.09	0.10	0.10	0.03	0.2	0.2	0.2	0.2	0.2

TABLE II. GMM PARAMETERS OF THE LOAD ACTIVE POWER (P.U.).

	$\mu_1$	$\mu_2$	$\mu_3$	$\sigma_1$	$\sigma_2$	$\sigma_3$	$\omega_1$	$\omega_2$	$\omega_3$
$P_4$	-0.072	-0.080	-	0.058	0.046	-	0.6	0.4	-
$P_7$	-0.250	-0.198	-	0.025	0.030	-	0.5	0.5	-
$P_{21}$	-0.172	-0.183	-0.160	0.016	0.023	0.008	0.4	0.3	0.3

Three typical profiles of load demands are also approximated by GMM at buses 4, 7 and 21, shown in Table II. The demand of remaining buses (not presented in Table II) are assumed to be Gaussian random variables, with standard deviation fixed at 5 % of the mean value. The loads are generally classified into two groups. Group1 consists of real power demands  $P_{15}$ ,  $P_{23}$  and  $P_{24}$ , which has the correlation coefficient  $\rho = 0.7$ , and the others (group 2) including  $P_{16}$  to  $P_{22}$ , has a correlation factor of  $\rho = 0.5$ . The correlation between outputs of two wind farms are strong ( $\rho = 0.8$ ), while they are considered independent of the load demand. The programs were implemented with Matlab 2014b on a PC with Intel core i5 3.0 GHz and 3 GB of RAM.

The proposed method, original CM and MCS are applied to estimate the PDFs of bus voltage and line flow. The original CM mentioned here was explicitly explained in [29], with AGC series and not considering the correlation. Taking bus 25 and line 10-17 as an example, we compared the results ( $U_{25}$  and  $Q_{10-17}$ ) of the three methods in Fig. 3.

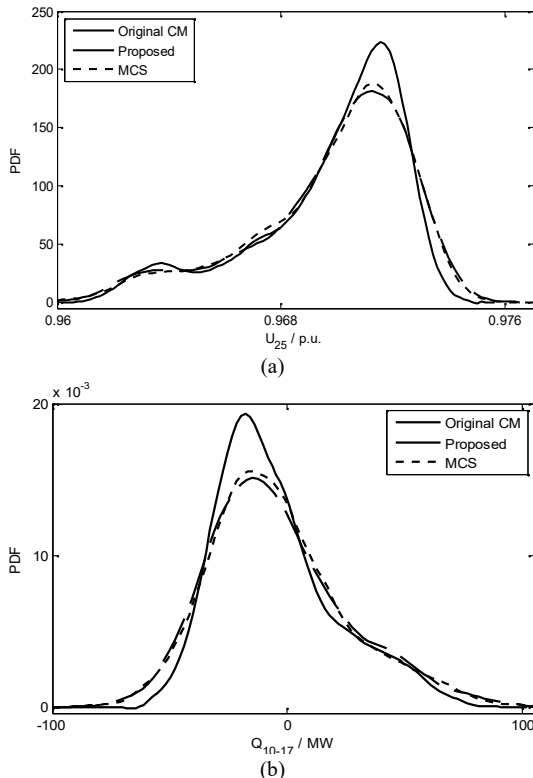


Fig. 3. PDF comparison of bus voltage and line flow: a) voltage magnitude of bus 25; b) reactive power of line 10-17.

It can be seen that some negative PDF values appear in the

tail regions of both pictures using the original CM. This is often the case when large-scale renewable energy injected, increasing skewness and kurtosis of the output variables. Large deviations could also be found in original CM, if MCS is taken as the best reference, while the error of the proposed method is negligible.

Figure 4 gives the CDFs of  $U_{25}$  and  $Q_{10-17}$ , which corresponds to the PDFs in Fig. 3. As is expected, the Y-coordinate values can go beyond the range of 0-1. To demonstrate the accuracy of the proposed method quantitatively, we used the average root mean square (ARMS) metric [30] as a comparative check. ARMS is a measure of the deviation & error of the model, and is defined as

$$\text{ARMS} = \sqrt{\frac{\sum_{i=1}^N (\text{Candidate}_i - \text{MCS}_i)^2}{N}}, \quad (26)$$

where  $\text{Candidate}_i$  is the  $i$ th CDF value using the candidate method, and  $\text{MCS}_i$  is the corresponding Monte Carlo point.  $N$  is the total number of selected points,  $N = 100$  was chosen in this paper.

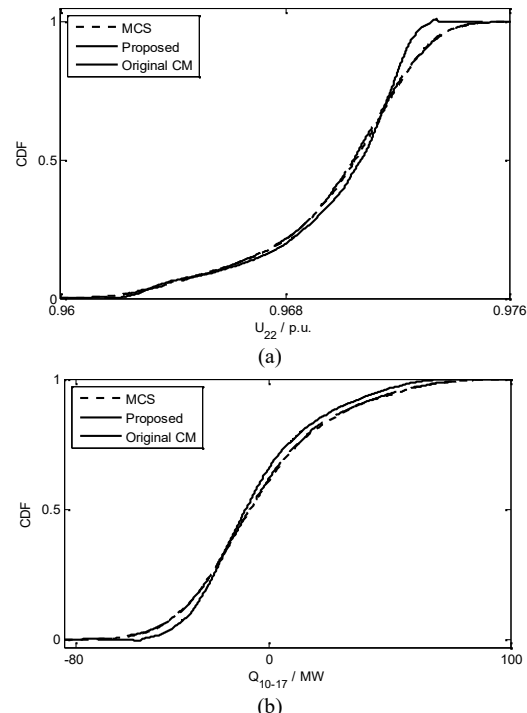


Fig. 4. CDF comparison of bus voltage and line flow: a) voltage magnitude of bus 25; b) reactive power of line 10-17.

Table III shows the mean and maximum ARMS of all kinds

of output variables obtained from the proposed method and original CM. Not surprisingly, The ARMS deviations with the proposed method are comparatively smaller, most of which are less than 1 %, even for the maximum values, showing large improvement compared with original CM.

TABLE III. ARMS OF OUTPUT VARIABLES IN IEEE 30 SYSTEM.

Output variables	ARMS/% (Mean/Max)	
	Original CM	Proposed
Voltage magnitude	1.32/2.55	0.37/0.64
Voltage angle	1.93/2.65	0.54/0.88
Line flow (real)	0.92/1.36	0.24/0.59
Line flow (reactive)	4.70/6.25	1.07/1.48

To evaluate the influence of correlation between the wind outputs that may have on the system operation, the average of mean and standard deviation of bus voltage and line flow were investigated with the correlation factor  $\rho$  ranging from 0 (independent) to 1 (completely correlated). As shown in Fig. 5, the average expected values are hardly influenced by the correlation of wind power, while there is a good linearity for standard deviation and the correlation factor. The fluctuation of system operation status would thus be intensified as the increase of  $\rho$ , making it prone to violating the limit and diminishing the system security.

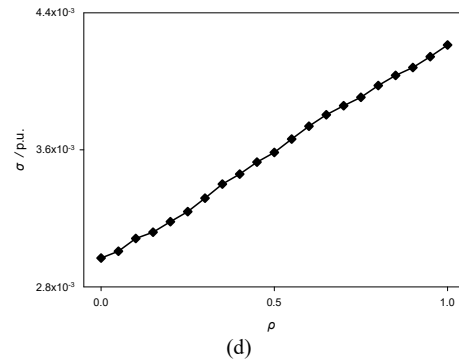
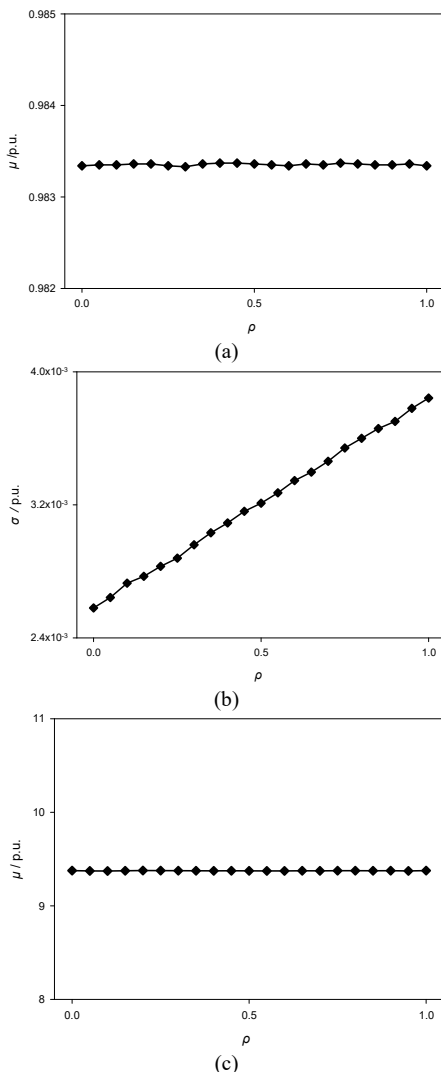


Fig. 5. Relationship between system operation parameters and wind power correlation coefficient: a) the average of expected value of bus voltage; b) the average of standard deviation of bus voltage; c) the average of expected value of line real power; d) the average of standard deviation of line real power.

TABLE IV. COMPUTATION TIME COMPARISON OF THE THREE METHODS.

	CPU time/s		
	IEEE 30	IEEE 57	IEEE 118
MCS	386.2	402.8	434.1
Original CM	1.49	2.35	6.76
Proposed	8.40	12.13	18.42

Finally, we compared the performance of the methods (MCS, original CM and proposed) in terms of the run time. Table IV lists the results tested in three different-sized systems. It is concluded that both the proposed method and original CM can significantly reduce the computational burden in PLF analysis compared with MCS, while their execution time goes up steadily as the size of network grows. Note that additional time is needed for proposed method than original CM due to the GMM modeling and correlation processing. However, this process is so fast that the overall time consumption is still reasonable.

## VI. CONCLUSIONS

In this paper, an improved analytic approach to PLF studies based on CM was proposed. This method could overcome the deficiencies of the original CM such as cumbersome computation of cumulants with non-Gaussian random variables, and not being able to manipulate the correlation. The IEEE 30 test system was used to verify the correctness and effectiveness of the proposed method, and the following features can be summarized as

- 1) The GMM used in this paper can accurately model the intermittent and multimodal nature of wind power distribution with specific parameters estimated iteratively for each mixture component. This model can also simplify the problem of cumulant calculation in a numerical way with its Gaussian components.
- 2) The correlation of input power injections is taken into account by representing correlated random variables in a linear sum of the uncorrelated ones so that the precondition of cumulant arithmetic could be satisfied. The influence of wind correlation on the expected value of output variables remains nearly unchanged but is proportional to the standard deviation.
- 3) The PDFs (or CDFs) of outputs reconstructed by CGC series expansion are very close to those obtained from MCS

with 10000 samples, particularly in the tail regions, with the average of ARMS less than 1 %.

4) Compared with MCS, the computational efficiency is enhanced by up to 40 times without sacrificing much of the accuracy, making it possible for some practical applications such as online power flow analysis, risk assessment, and so forth.

#### REFERENCES

- [1] P. Eser, A. Singh, N. Chokani, "Effect of increased renewables generation on operation of thermal power plants", *Applied Energy*, vol. 164, pp. 723–732, 2016. DOI: 10.1016/j.apenergy.2015.12.017.
- [2] J. C. Hernandez, F. J. Ruiz-Rodriguez, F. Jurado, "Modelling and assessment of the combined technical impact of electric vehicles and photovoltaic generation in radial distribution systems", *Energy*, vol. 141, pp. 316–332, 2017. DOI: 10.1016/j.energy.2017.09.025.
- [3] T. Brown, "Transmission network loading in Europe with high shares of renewables", *IET Renewable Power Generation*, vol. 9, no. 1, pp. 57–65, 2015. DOI: 10.1049/iet-rpg.2014.0114.
- [4] A. Botterud, Z. Zhou, J. Wang, "Wind power trading under uncertainty in LMP markets", *IEEE Trans. power systems*, vol. 27, no. 2, pp. 894–903, 2012. DOI: 10.1109/TPWRS.2011.2170442.
- [5] G. Carpinelli, P. Caramia, P. Varilone, "Multi-linear Monte Carlo simulation method for probabilistic load flow of distribution systems with wind and photovoltaic generation systems", *Renewable Energy*, vol. 76, pp. 283–295, 2015. DOI: 10.1016/j.renene.2014.11.028.
- [6] B. Borkowska, "Probabilistic load flow", *IEEE Trans. Power Apparatus and Systems*, vol. 93, no. 3, pp. 752–759, 1974. DOI: 10.1109/TPAS.1974.293973.
- [7] Z. Q. Xie, T. Y. Ji, M. S. Li, "Quasi-Monte Carlo based probabilistic optimal power flow considering the correlation of wind speeds using copula function", *IEEE Trans. Power Systems*, 2018, to be published. DOI: 10.1109/TPWRS.2017.2737580.
- [8] M. Fan, V. Vittal, G. T. Heydt, "Probabilistic power flow studies for transmission systems with photovoltaic generation using cumulants", *IEEE Trans. Power Systems*, vol. 27, no. 4, pp. 2251–2261, 2012. DOI: 10.1109/TPWRS.2012.2190533.
- [9] A. R. Malekpour, T. Niknam, A. Pahwa, "Multi-objective stochastic distribution feeder reconfiguration in systems with wind power generators and fuel cells using the point estimate method", *IEEE Trans. Power Systems*, vol. 28, no. 2, pp. 1483–1492, 2013. DOI: 10.1109/TPWRS.2012.2218261.
- [10] I. Papaioannou, W. Betz, K. Zwirgmaier, "MCMC algorithms for subset simulation", *Probabilistic Engineering Mechanics*, vol. 41, pp. 89–103, 2015. DOI: 10.1016/j.probengmech.2015.06.006.
- [11] Z. K. Wei, J. F. Xu, X. K. Dai, "Research on coarse-grained parallel algorithm of the Monte-Carlo simulation for probabilistic load flow calculation", in *Proc. Int. Conf. Power and Energy (CPE 2014)*, Shanghai, China, 2014, pp. 77–81. DOI: 10.1201/b18409-15.
- [12] Y. Chen, J. Wen, S. Cheng, "Probabilistic load flow method based on Nataf transformation and Latin hypercube sampling", *IEEE Trans. Sustainable Energy*, vol. 4, no. 2, pp. 294–301, 2013. DOI: 10.1109/TSTE.2012.2222680.
- [13] R. N. Allan, B. Borkowska, C. H. Grigg, "Probabilistic analysis of power flows", *Proceedings of the Institution of Electrical Engineers*, vol. 121, no. 12, pp. 1551–1556, 1974. DOI: 10.1049/piee.1974.0320.
- [14] R. N. Allan, A. M. Da Silva, R. C. Burchett, "Evaluation methods and accuracy in probabilistic load flow solutions", *IEEE Trans. Power Apparatus and Systems*, vol. 100, no. 5, pp. 2539–2546, 1981. DOI: 10.1109/TPAS.1981.316721.
- [15] P. Zhang, S. T. Lee, "Probabilistic load flow computation using the method of combined cumulants and Gram-Charlier expansion", *IEEE Trans. Power Systems*, vol. 19, pp. 676–682, 2004. DOI: 10.1109/TPWRS.2003.818743.
- [16] T. Williams, C. Crawford, "Probabilistic load flow modeling comparing maximum entropy and Gram-Charlier probability density function reconstructions", *IEEE Trans. Power Systems*, vol. 28, no. 1, pp. 272–280, 2013. DOI: 10.1109/TPWRS.2012.2205714.
- [17] B. R. Prusty, D. Jena D, "A critical review on probabilistic load flow studies in uncertainty constrained power systems with photovoltaic generation and a new approach", *Renewable and Sustainable Energy Reviews*, vol. 69, pp. 1286–1302, 2017. DOI: 10.1016/j.rser.2016.12.044.
- [18] C. Chen, W. Wu, B. Zhang, "Correlated probabilistic load flow using a point estimate method with Nataf transformation", *International Journal of Electrical Power & Energy Systems*, vol. 65, pp. 325–333, 2015. DOI: 10.1016/j.ijepes.2014.10.035.
- [19] J. Usaola, "Probabilistic load flow with correlated wind power injections", *Electric Power Systems Research*, vol. 80, no. 5, pp. 528–536, 2010. DOI: 10.1016/j.eprsr.2009.10.023.
- [20] M. T. Puth, M. Neuhauser, G. D. Ruxton, "Effective use of Spearman's and Kendall's correlation coefficients for association between two measured traits", *Animal Behaviour*, vol. 102, pp. 77–84, 2015. DOI: 10.1016/j.anbehav.2015.01.010.
- [21] R. Becker, "Generation of time-coupled wind power infeed scenarios using pair-copula construction", *IEEE Trans. Sustainable Energy*, 2017, to be published. DOI: 10.1109/TSTE.2017.2782089.
- [22] D. Villanueva, A. E. Feijoo, J. L. Pazos, "An analytical method to solve the probabilistic load flow considering load demand correlation using the DC load flow", *Electric Power Systems Research*, vol. 110, pp. 1–8, 2014. DOI: 10.1016/j.eprsr.2014.01.003.
- [23] D. Reynolds, "Gaussian mixture models", *Encyclopedia of biometrics*, pp. 827–832, 2015. DOI: 10.1007/978-1-4899-7488-4\_196.
- [24] J. P. Vila, P. Schniter, "Expectation-maximization Gaussian-mixture approximate message passing", *IEEE Trans. Signal Processing*, vol. 61, no. 19, pp. 4658–4672, 2013. DOI: 10.1109/TSP.2013.2272287.
- [25] D. J. Salmond, "Mixture reduction algorithms for target tracking in clutter", *SPIE signal and data processing of small targets*, vol. 1305, no. 1, pp. 434–445, 1990.
- [26] M. Vijay, D. Jena, "PSO based neuro fuzzy sliding mode control for a robot manipulator", *Journal of Electrical Systems and Information Technology*, vol. 4, no. 1, pp. 243–256, 2017. DOI: 10.1016/j.jesit.2016.08.006.
- [27] D. Kurowicka, R. M. Cooke, *Uncertainty analysis with high dimensional dependence modelling*. London: John Wiley & Sons, 2006, ch. 4. DOI: 10.1002/0470863072.
- [28] R. Christie, "Power system test archive", University of Washington, 1999. [Online]. Available: <http://www.ee.washington.edu/research/pstca>
- [29] Y. Yuan, J. Zhou, P. Ju, "Probabilistic load flow computation of a power system containing wind farms using the method of combined cumulants and Gram-Charlier expansion", *IET renewable power generation*, vol. 5, no. 6, pp. 448–454, 2011. DOI: 10.1049/iet-rpg.2010.0218.
- [30] M. H. Alobaidi, F. Chebana, M. A. Meguid, "Robust ensemble learning framework for day-ahead forecasting of household based energy consumption", *Applied Energy*, vol. 212, pp. 997–1012, 2018. DOI: 10.1016/j.apenergy.2017.12.054.