Contents list available at IJRED website
Int. Journal of Renewable Energy Development (IJRED)
Journal homepage: https://ijred.undip.ac.id

Research Article

Probabilistic Assessment of Power Systems with Renewable Energy Sources based on an Improved Analytical Approach

Dinh Duong Le[a*], Van Duong Ngo[b], Nhi Thi Ai Nguyen[a], Van Ky Huynh[b]

[a]*Faculty of Electrical Engineering, The University of Danang - University of Science and Technology, 54 Nguyen Luong Bang, Danang, Vietnam*

[b]*The University of Danang, 41 Le Duan, Danang, Vietnam*

ABSTRACT. The increasing penetration of renewable energy sources has introduced great uncertainties and challenges into computation and analysis of electric power systems. To deal with uncertainties, probabilistic approaches need to be used. In this paper, we propose a new framework for probabilistic assessment of power systems taking into account uncertainties from input random variables such as load demands and renewable energy sources. It is based on the cumulant-based Probabilistic Power Flow (PPF) in combination with an improved clustering technique. The improved clustering technique is also developed in this study by making use of Principal Component Analysis (PCA) and Particle Swarm Optimization (PSO) to reduce the range of variation in the input data, thus increasing the accuracy of the traditional cumulant-based PPF (TCPPF) method. In addition, thanks to adopting PCA for dimensionality reduction, the improved clustering technique can be effectively dealt with a large number of input random variables so that the proposed framework for probabilistic assessment can be applied for large power systems. The IEEE-118 bus test system is modified by adding five wind and eight solar photovoltaic power plants to examine the proposed method. Uncertainties from input random variables are represented by appropriate probabilistic models. Extensive testing on the test system indicates good performance of the proposed approach in comparison to the traditional cumulant-based PPF and Monte Carlo simulation. Extensive testing on the test system, using Matlab (R2015a) on an Intel Core i5 CPU 2.53 GHz/4.00 GB RAM PC, indicates good performance of the proposed approach (PPPF) in comparison to the TCPPF and Monte Carlo simulation (MCS): In terms of computation time, PPPF needs 4.54 seconds, while TCPPF and MCS require 2.63 and 251 seconds, respectively; ARMS errors are calculated for methods using benchmark MCS and their values clearly demonstrate the higher accuracy of PPPF in estimating probability distributions compared to TCPPF, i.e., the maximum (Max) and mean (Mean) values of ARMS errors of all output random variables are: $\mathrm{ARMS}_{\mathrm{PPPF}}^{\mathrm{Max}} = 0.11\%$, $\mathrm{ARMS}_{\mathrm{TCPPF}}^{\mathrm{Max}} = 0.55\%$ and $\mathrm{ARMS}_{\mathrm{PPPF}}^{\mathrm{Mean}} = 0.06\%$, $\mathrm{ARMS}_{\mathrm{TCPPF}}^{\mathrm{Mean}} = 0.35\%$.

Keywords: Renewable energy, uncertainty, Probabilistic Power Flow, clustering technique, power system

Article History: Received: 3rd May 2021; Revised: 5th June 2021; Accepted: 25th June 2021; Available online: 30th June 2021
How to Cite This Article: Le, D.D., Ngo, V.D., Nguyen, N.T.A., and Huynh, V.K. (2021) Probabilistic Assessment of Power Systems with Renewable Energy Sources based on an Improved Analytical Approach. *Int. Journal of Renewable Energy Development*, 10(4),811-818
https://doi.org/10.14710/ijred.2021.38226

1. Introduction

Over the past decades, the deployment of renewable energy in power systems has grown dramatically due to the concerns of carbon dioxide emission as well as the growth in the demand for energy consumption. However, a large number of renewable energy sources has introduced great uncertainties and challenges into the operation and planning of power systems. Uncertainties in power systems result from load demands, renewable energy sources, etc. The conventional Deterministic Power Flow (DPF) is a vital tool for power system planning and operation (Bergen *et al.* 1986). During the computation, it uses specified values of inputs (i.e., loads, power generation, etc.). However, information on loads, power generation (especially wind and solar power generation) and so on in a real power system is not certain so output variables (i.e., state variables including voltages, angles, and line power flows) need to be assessed for a range of loads and generation conditions. Using DPF, it is necessary to carry out for all possible system states requiring an extremely large computational effort (Zhang *et al.* 2004; Morales *et al.* 2007). Therefore, DPF has a significant limitation in the calculation and analysis of power systems. Another way to calculate and analyze the system is to use the worst-case scenario to avoid system risk; however, it does not accurately reflect the state of the system possibly leading to overly pessimistic and non-optimal techno-economic solutions (Hasan *et al.* 2019).

Different from DPF, PPF has been developed to handle the uncertainties and become an crucial tool in power system planning and operation. PPF was first proposed by Borkowska in 1974. Since then, various methods have been developed for PPF. Generally, they could be classified into three main categories: numerical methods, approximate methods, and analytical methods.

MCS is a well-known numerical method for PPF calculation. MCS relies on a significant number of samples and repetitive DPF computations to obtain the results. It

* Corresponding author: ldduong@dut.udn.vn

is recognized as the most accurate method in the field of PPF and its results are usually used as a reference to evaluate the accuracy of other methods. The main drawback of MCS is that it is very time consuming due to a large number of simulations so it is difficult to apply for large power systems in practice. In order to reduce the computational burden of MCS, a number of sampling methods such as Latin hypercube sampling (Liu *et al.* 2016), Latin supercube sampling (Hajian *et al.* 2013) and importance sampling (Huang *et al.* 2011) were developed and applied. MCS using importance sampling techniques via the cross-entropy method was proposed to estimate PPF risk events (Leite da Silva *et al.* 2018). In another attempt to improve computational efficiency, Quasi-MCS was applied (Tao *et al.* 2013; Xie *et al.* 2018).

On the contrary, approximate methods need much less computational time in comparison to that for MCS. Point estimate method (PEM) is considered as the representative method in this category. In this approach, input random variables are represented by a number of pairs of values and weights and then the moments of the output random variables are computed via functions characterizing input-output relationship. Su (Su 2005) first proposed the $2m$ PEM for PPF in which input random variables were assumed to be uncorrelated. It was then modified by Aien (Aien *et al.* 2014) to deal with correlated random variables. 2m+1 PEM was also introduced to improve the accuracy (Morales *et al.* 2007); however, it needs more simulations compared to $2m$ PEM. Gupta (Gupta 2016) pointed out that both three PEM (3PEM) and five PEM (5PEM) methods can solve PPF accurately in the presence of load and wind generation uncertainties. Generally, the efficiency of point estimate method decreases as the number of random variables increases, thus, it also faces great challenges in applying for large power systems.

In analytical methods, power flow equations are linearized so that arithmetic algorithms (such as convolution and cumulant techniques) can be adopted to obtain probability distributions (i.e., probability density functions and cumulative distribution functions, denoted as PDFs and CDFs, respectively) of output random variables based on probability distributions of input random variables. In the early stages of analytical methods, convolution techniques (Allan *et al.* 1977; Borkowska *et al.* 1974) were commonly used. Fast Fourier transform was employed by Allan in 1981 (Allan *et al.* 1981). However, computation efficiency of these techniques was still low. A first-order second-moment method was also adopted to obtain the mean and standard deviation of output random variables (Wan *et al.* 2012). Among analytical methods, cumulant method is one of the fastest ones for PPF computation (Fan *et al.* 2012), so it is suitable for large power systems with a large number of input random variables. Cumulant method executes power flow calculation only once and obtains cumulants of output variables from the cumulant of input variables through a simple linear transformation (Zhang *et al.* 2004). This method needs a reconstruction techniques such as the Gram Charlier series expansion (Zhang *et al.* 2004) and Cornish Fisher series expansion (Ruiz-Rodriguez *et al.* 2012) to recover the probability distributions of output variables. Cumulant method is based on the linearization of power flow equations, thus when the input random variables have large variations, the traditional cumulant method for PPF could result in significant errors.

Based on the advantages and disadvantages of each group of above-mentioned methods, cumulant method can be applied to the planning and operation of power systems with different time frames in practice if the accuracy of the method is improved. This is why, in the present paper, the focus is on improving the accuracy of the TCPPF method. Great efforts are made to address the issue by proposing to use a clustering technique to enhance the TCPPF.

Application of a clustering technique aims to obtain the samples in each cluster with smaller variation, compared to large range of variation of the whole samples, then TCPPF is implemented for each cluster instead of the whole samples to improve the accuracy of the cumulant method. Among many existing clustering algorithms, K-means is one of the most popular one (Gan *et al.* 2007). In solving PPF and also probabilistic optimal power flow, Deng (Deng *et al.* 2017; Deng *et al.* 2019) and Zhou (Zhou *et al.* 2020) tried to handle large variations of input random variables using the method of combined cumulant and the conventional K-means. K-means is easy to execute; however, it has a number of drawbacks: it only converges to arbitrary local optima and does not guarantee to find the global optimum solution for clustering; it is difficult to predict the number of clusters; random selection initial cluster centers has a strong impact on the final results. Moreover, for a large power system with large number of random variables associated with loads, renewable energy sources, etc., PPF problem needs to manage high dimensional input dataset that makes the task of clustering more challenging.

In order to overcome the above-mentioned limitations of TCPPF in general and of existing techniques for clustering as well in particular, an improved clustering technique based on the combination of PCA and PSO is proposed. The proposed PPF approach making use of the improved clustering technique is tested on the modified IEEE-118 bus test system to demonstrate its performance in comparison to MCS and the TCPPF.

The remainder of the paper is organized as follows. Section 2 presents the formulation of the TCPPF. In Section 3, the improved clustering technique is described, while the proposed PPF approach and framework for probabilistic assessment of power systems are given in Section 4. Section 5 describes the testing of the proposed approach on the modified IEEE-118 bus test system and the results are discussed. Section 6 gives further discussions on the application of methods for probabilistic assessment of power systems. Finally, Section 7 summarizes the main conclusions of the paper.

2. Cumulant-based Probabilistic Power Flow Formulation

In this section, a TCPPF method is presented (Zhang *et al.* 2004). It is based on linear relationship between output random variables (i.e., state variables including voltages, angles, real and reactive power flows) and input random variables (i.e., nodal power injections such as load demands, renewable power generation).

The basic power flow equations can be represented by a matrix form as in Eq. (1) and Eq. (2).

$$w = g(x) \tag{1}$$

$$z = h(x) \quad (2)$$

where:
- w : vector of nodal power injections;
- x: vector of state variables;
- z: vector of line power flows;
- $g(x)$: power flow equations;
- $h(x)$: functions to compute line power flows.

DPF is performed for the system and then using Taylor series expansion to linearize the above equations around the solution point $\overline{x}$ gives as in Eq. (3) and Eq. (4).

$$\Delta w = \mathrm{G}|_{\overline{x}}\, \Delta w \quad (3)$$

$$\Delta z = \mathrm{H}|_{\overline{x}}\, \Delta w \quad (4)$$

where: $\mathrm{G}|_{\overline{x}}$ is the inverse of the Jacobian matrix and $\mathrm{H}|_{\overline{x}}$ is the sensitivity matrix of power flows with respect to nodal power injections; $\mathrm{G}|_{\overline{x}}$ and $\mathrm{H}|_{\overline{x}}$ are computed at the solution point $\overline{x}$.

In PPF computation, each element of w, x and z is considered as the realization of a random variable associated with each nodal power injection, state variable and power flow, respectively. Based on the linearized relationships in (3) and (4), cumulant-based PPF can be adopted.

The procedure for TCPPF method is briefly described as follows:
- Solve DPF for the system to obtain the expected value of random state variables $\overline{x}$ and the sensitivity matrices $\mathrm{G}|_{\overline{x}}$ and $\mathrm{H}|_{\overline{x}}$ computed at $\overline{x}$;
- Calculate cumulants of state variables and line power flows based on (3) and (4) and using cumulants of input random variables;
- Obtain PDFs and CDFs of the output random variables of interest using a series expansion technique.

Popular series expansion techniques, such as the Gram Charlier series expansion (Zhang *et al.* 2004) and Cornish Fisher series expansion (Ruiz-Rodriguez *et al.* 2012) could be adopted to approximate probability distributions for output random variables. The Cornish Fisher series expansion is chosen to use in the current paper because of its better performance for non-Gaussian distributions (Ruiz-Rodriguez *et al.* 2012).

3. Improved Clustering Technique

3.1 Data clustering

Clustering is the task of partitioning data points in a dataset into a number of clusters such that data points in the same clusters are more similar to other data points in the same cluster than those in other clusters (Gan *et al.* 2007). In other words, clustering algorithms maximize the similarity among members within the same clusters, while minimizing the dissimilarities between different clusters. Therefore, clustering problem can be treated as an optimization problem so that optimization algorithms, such as GA (Genetic Algorithms) (Bandyopadhyay *et al.* 2002; Falkenauer *et al.* 1998), PSO (Abdmouleh *et al.* 2017; Van der Merwe *et al.* 2003), etc., can be applied. Different from K-means, such optimization algorithms give global optimum solution for clustering and provide a good performance for clustering. Among them, PSO has main advantages as follows (Abdmouleh *et al.* 2017): it is quite simple to implement; it needs few parameters to adjust; it can converge fast and give robust result. Thus, PSO is applied for clustering in this paper.

For high-dimensional datasets, the task of clustering faces several difficulties, resulting in obtained results. To get over this problem, in this paper an effective way is suggested to support the clustering task by applying PCA to reduce dimensions of the dataset before using PSO algorithm.

3.2 Principal Component Analysis

PCA is a dimensionality reduction method (Jackson *et al.* 1991). PCA is often applied to reduce the dimensionality of large datasets. It transforms a large set of variables into a smaller one, remaining most of the information in the original set (Jolliffe *et al.* 2002). PCA is carried out on a square symmetric matrix, i.e., the correlation or the covariance matrix of the dataset. The covariance matrix is used if the scales of the considered variables are not different. Otherwise, the data need to be standardized before using PCA or the correlation matrix is applied (Le *et al.* 2015).

Suppose n-by-m matrix A contains a dataset in which n rows and m columns correspond to observations or samples and variables, respectively. PCA is carried out on the dat aset as follows (Le *et al.* 2015):

- Center the data by subtracting the mean of all data points from each individual data point to obtain centered matrix $\mathrm{A_c}$;
- Compute eigenvalue $(\lambda_i, i = 1 \div m)$ and eigenvector $(\mathrm{e}_i, i = 1 \div m)$ pairs of the covariance matrix or correlation matrix of the dataset;
- Sort eigenvalues in descending order (i.e., $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m$);
- Form the projection matrix: $\mathrm{E} = [\mathrm{e}_i \mathrm{e}_2 \cdots \mathrm{e}_m]$;
- Use E to transform $\mathrm{A_c}$ into a n-by-m matrix $\mathrm{B} = (\mathrm{E^T A_c^T})^\mathrm{T}$, where each column of B is called a PC (Principal Component).

The variance of the i^{th} PC $(i = 1 \div m)$ is equal to the eigenvalue λ_i associated with that PC. The 1st column of B (the 1st PC) corresponding to the largest eigenvalue λ_1 is the most important component that contains most of the variance (information) in the dataset A, followed by the 2nd component, and so on.

Each PC contributes to total variance of the data as in Eq. (5) (Le *et al.* 2015).

$$v_i = \frac{\lambda_i}{\sum_{j=1}^{m} \lambda_j} \quad (5)$$

If the first k eigenvectors that correspond to the k $(k \ll m)$ largest eigenvalues are selected, a reduced matrix $\mathrm{B_k} = (\mathrm{E_k^T A_c^T})^\mathrm{T}$ could be obtained, where $\mathrm{E_k} = [\mathrm{e}_i \mathrm{e}_2 \cdots \mathrm{e}_k]$.

The cumulative contribution of the first k PCs is calculated as in Eq. (6).

$$\mathrm{C}_k = \sum_{l=1}^{k} v_l \quad (6)$$

Eventually, PCA projects the data in matrix A (size n-by-m) into lower dimension subspace (size n-by-k) by picking up a first few numbers of PCs (i.e., k) with the largest variances.

3.3 Improved clustering technique

The improved clustering technique, based on combination of PCA and PSO clustering algorithm, developed in this paper is applied to deal with input random variables in PPF problem. The technique is implemented as follows:

- Construct data matrix A with size of n-by-m in which n correspond to number of samples and m correspond to the variables;
- Carry out PCA and transform data into the reduced projected data B_k (size of n-by-k, $k \ll m$);
- Perform PSO algorithm to partition data into distinct clusters.

4. Proposed Framework for Probabilistic Assessment

This section introduces the proposed framework for the probabilistic assessment of power systems taking into account uncertainties from input random variables (i.e., load, renewable power generation, etc.) with large range of variation resulting in significant errors for the TCPPF. In order to tackle this problem, the clustering technique, presented in Section 3, is developed to partition sample data of input random variables into distinct clusters in which the obtained data in each cluster have smaller range of variation compared to that in the original data. The TCPPF is then carried out for each cluster.

Procedure for the proposed approach is described as follows:

- *Step 1*: Represent uncertainties from the input random variables by probabilistic models.
- *Step 2*: Generate samples of the inputs based on their probabilistic models.
- *Step 3*: Perform the improved clustering technique (Section 3) to partition sample data into distinct clusters.
- *Step 4*: Run TCPPF for each cluster to obtain cumulants of output random variables (i.e., bus voltage, line power flow, etc.) based on cumulants of input random variables. In case the input variables are correlated, the decomposition technique presented by Cai (Cai *et al.* 2012) is adopted to obtain uncorrelated data prior to running TCPPF.
- *Step 5*: Compute the entire cumulants of output random variables by adopting the law of total probability to combine all results obtained from all clusters.
- *Step 6*: Approximate the PDFs and/or CDFs of output random variables using a series expansion technique.
- *Step 7*: Perform probabilistic assessment for the system such as probability of line overloading, probability of over-/under-voltage, etc.

5. Tests and Results

In this section, the IEEE 118-bus test system is adopted to evaluate the performance of the proposed probabilistic assessment approach. The diagram and necessary information of branches, buses, and generators of the system are given by Christie (Christie 1993). The system contains 19 generators, 35 synchronous condensers, 177 lines, 9 transformers, 189 power injections from loads (Only non-zero active and reactive power injections are considered). In this test, the focus is on the use of the improved data clustering technique while considering uncertainties from both loads and renewable energy sources (i.e., wind and solar). Hence, the system is modified by adding five wind power plants and eight solar photovoltaic power plants to buses as shown in Table 1 and Table 2, respectively. Uncertainties from loads and renewable energy sources are assumed to be provided and for the sake of simplicity and without loss of generality: Load at each bus is represented by a normal distribution, whose mean is the base value and standard deviation is equal to 10% of its mean; the uncertainties of wind power are assumed to have Weibull distributions with their parameters as shown in Table 1 and they are correlated with correlation coefficient equal to 0.7; the uncertainties of solar photovoltaic power are assumed to follow Beta distributions with their parameters as given in Table 2 and they are correlated with correlation coefficient equal to 0.8.

In the test system, there are 202 input random variables, including 5 from wind power plants, 8 from solar photovoltaic power plants and 189 from active and reactive powers of loads. Based on probability distributions of input random variables and their correlation information, 10000 samples are generated. Taken as examples, the PDF of active load power at bus 11 and wind power at bus 14 are plotted by histograms as shown in Fig. 1 and Fig. 2, respectively.

Matrix A of the samples (its form as described in Section 3) has the size of 10000×202. With such a high-dimensional dataset, the improved clustering technique can effectively deal with. The remaining steps of the procedure for the proposed approach are then performed to achieve the results. At the same time, both a MCS with 10000 samples and the TCPPF are used to solve PPF to assess the performance of the proposed approach. The obtained results from MCS are treated as the benchmark. The numerical studies are implemented in Matlab (R2015a) on an Intel Core i5 CPU 2.53 GHz/4.00 GB RAM PC.

Table 1
Wind power generation information

Bus	Rated power (MW)	Parameters of Weibull distributions of wind power generation	
		Scale parameter	Shape parameter
2	90	12	1.4
14	100	30	1.6
50	60	21	2.1
51	80	15	1.5
84	50	10	2.5

Table 2
Solar photovoltaic power generation information

Bus	Rated power (MW)	Parameters of Beta distributions of solar photovoltaic power generation	
		Parameter 1 (α)	Parameter 2 (β)
16	100	3	8
21	90	1.5	9
23	60	9	14
28	50	12	19
93	35	2.5	9
94	70	1.6	7
95	85	1.2	5
98	65	4	12

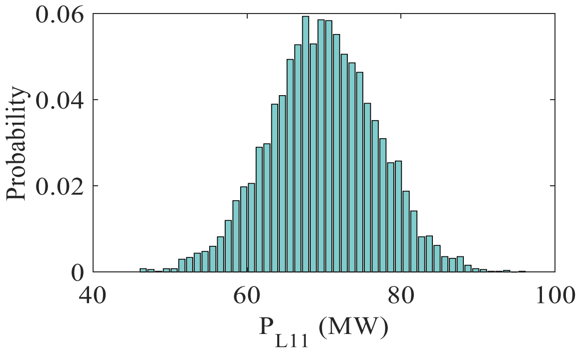

Fig. 1 Histograms of active load power at bus 11

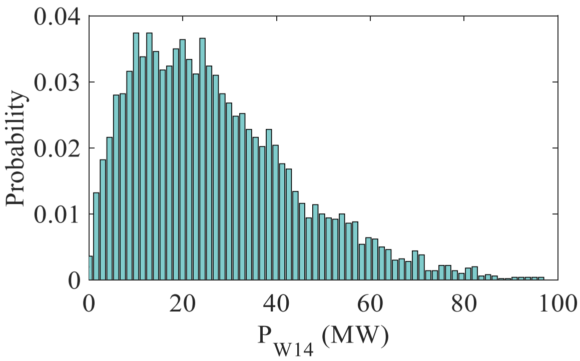

Fig. 2 Histograms of wind power generation at bus 14

In order to evaluate the accuracy of results obtained from a PPF method in comparison with that from MCS, ARMS (Average Root Mean Square) error (Le *et al.* 2016; Zhang *et al.* 2004) is calculated as in Eq. (7).

$$ARMS = \frac{1}{N}\sqrt{\sum_{1}^{N}(F_{MCSi} - F_{PPFi})^2} \tag{7}$$

where: F_{MCSi} and F_{PPFi} are the i^{th} values on CDF curves obtained by MCS and the PPF method, respectively; N is the number of points on CDFs (Le *et al.* 2016). In this test, the base power of 100 MVA is adopted.

PPF computation is performed for all three methods to obtain results of all output random variables in terms of PDFs and/or CDFs; however, for illustration, the PDFs and CDFs of selected output random variables are shown. Fig. 3 and Fig. 4 depict PDFs and CDFs of real power flow through line 63-59 (i.e., denoted as $P_{63\text{-}59}$), respectively, while PDFs and CDFs of voltage at bus 51 (i.e., denoted as V_{51}) are shown in Fig. 5 and Fig. 6, respectively.

As observed from the figures, the curves obtained by the proposed approach (denoted as PPPF) can better match the curves from MCS than those from the TCPPF, showing the good performance of the proposed approach in estimations of probability distributions. It is worth noting that the accuracy of PPPF is significantly improved compared to TCPPF at the left and right boundary regions (i.e., left and right tails) of the probability distributions. It is because the cumulant method is based on the linearization of power flow equations and when the input random variables have large variations (probability distributions with long left and right tails), the linearization is significantly affected causing remarkable errors in the left and right tails of probability distributions of output random variables. In this paper, thanks to adopting clustering technique, large range of variation of the whole samples of input random variables is divided into various smaller range of variation in PPPF so its accuracy is greatly improved compared to TCPPF.

The results of ARMS errors for $P_{63\text{-}59}$ and V_{51} are shown in Table 3. They again indicate the very good accuracy of the proposed aprroach and the good performance of applying the improved clustering technique that is developed in this paper. ARMS errors help to quantitatively evaluate the accuracy of methods in comparison to the benchmark MCS. The smaller values of ARMS errors demonstrate the higher accuracy of PPPF in estimating probability distributions compared to TCPPF. For showing an overview of the accuracy in terms of ARMS errors in this test, the maximum (Max) and mean (Mean) values of ARMS errors of all output random variables (including bus voltages and angles, active and reactive power flows) are also given as follows: $\text{ARMS}_{\text{PPPF}}^{\text{Max}} = 0.11\%$, $\text{ARMS}_{\text{TCPPF}}^{\text{Max}} = 0.55\%$ and $\text{ARMS}_{\text{PPPF}}^{\text{Mean}} = 0.06\%$, $\text{ARMS}_{\text{TCPPF}}^{\text{Mean}} = 0.35\%$.

For demonstrating the performance of all considered methods in computation time, Table 4 is shown. It clearly indicates that all the considered cumulant methods just needs a few seconds for computation, compared to a hundred of seconds required by MCS. Among the two cumulant methods, PPPF requires more computation time since it has to perform the computation for all clusters while TCPPF does it only once. However, the computational burden does not increase significantly. Therefore, PPPF method has high accuracy and gives calculation results in a short time, so it is suitable for calculation and analysis of power systems, especially large power systems in practice where MCS is likely not feasible to implement.

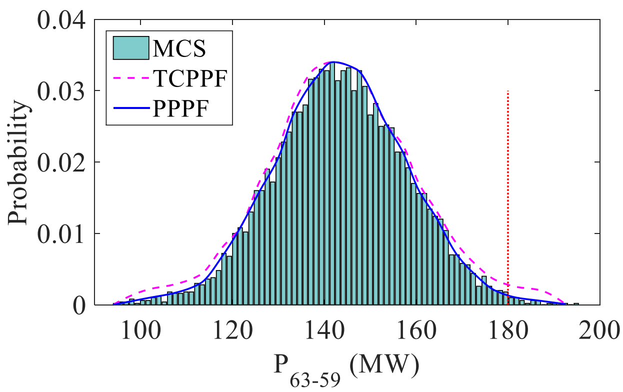

Fig. 3 PDFs of real power flow through line 63-59 ($P_{63\text{-}59}$)

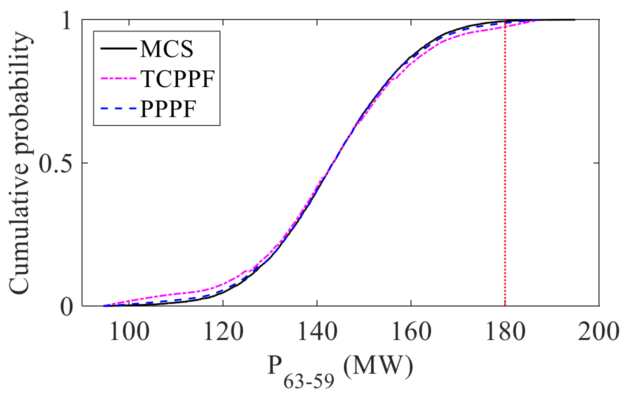

Fig. 4 CDFs of real power flow through line 63-59 ($P_{63\text{-}59}$)

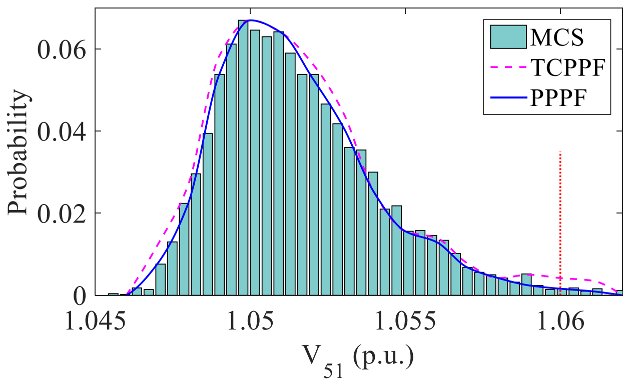

Fig. 5 PDFs of voltage at bus 51 (V_{51})

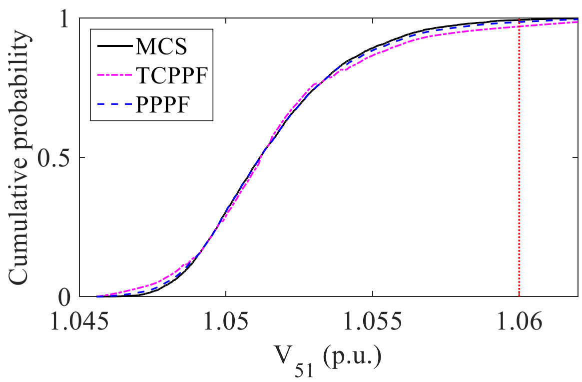

Fig. 6 CDFs of voltage at bus 51 (V_{51})

Different from conventional DPF that uses specific values of input variables of a power flow computation problem to obtain fixed values for output variables, PPF can provide a range of all possible values and statistical information for output variables that are very helpful for probabilistic assessment of the power system accounting for uncertainties from input variables. The probability of line overloading, over-/under-voltage, and so on can be estimated.

Taken as an example, suppose that the upper limit of the real power flow of line 63-59 is 180 MW (the vertical line in Fig. 3 and Fig. 4), the probability so that power through the line is over its limit can be calculated as in Eq. (8):

$$\mathrm{P}\{P_{63-59} > 180\} = 1.6\% \tag{8}$$

Similarly, probability so that voltage at a specific bus is out of the operating range can be evaluated. Assume that the operating range of voltage is [0.94, 1.06] p.u., the probability being greater than the upper limit is determined in Eq. (9).

$$\mathrm{P}\{V_{51} > 1.06\} = 1.5\% \tag{9}$$

Table 3
ARMS errors

Output	ARMS (%)	
	TCPPF	Proposed approach
$P_{63\text{-}59}$	0.31	0.05
V_{51}	0.45	0.09

Table 4
Comparison of computation time

Method	Time (s)
MCS	251
TCPPF	2.63
PPPF	4.54

Similarly, any output random variable of interest in the system could be assessed. The results obtained by the above probabilistic assessment framework can assist the system operator in analyzing the operating states of the system to make appropriate decisions as well as propose suitable solutions for the system.

6. Further Discussions on the Application of Methods for Probabilistic Assessment of Power Systems

In this paper, a new framework for the probabilistic assessment of power systems has been proposed. In this section, a discussion on possible applications of methods is presented.

Each group of PPF methods has its own advantages and disadvantages. As discussed above, MCS method gives accurate and reliable results; however, it is very time consuming. Both approximation and analytical methods need much less computational time compared to that for MCS. However, the efficiency of approximation method (typically PEM) decreases as the number of random variables increases so it is difficult to apply to large power systems with many random variables. On the contrary, analytical method, especially cumulant method, is suitable for computation of large power systems. Due to adopting the linearization of power flow equations, it encounters difficulties when the input random variables have large variations. This disadvantage is overcome by the approach proposed in this paper.

Basically, a PPF method is appropriate to be selected or not depending on the actual application needed:

- When solving problems in power systems in planning domain (e.g., many years, a year, seasons, months, weeks) or in operation planning domain (up to a few days), the users may not necessarily need fast tools for assessment of the system. In this case, MCS is suitable to be used. However, for a very large power system with too many inputs, MCS is sometimes not feasible because of its computational complexity. In such a case, PPPF is recommended as an alternative. Data on load at each bus and renewable energy sources such as wind and solar collected over a long period of time from several months to a year or many years are used to analyze and construct probability distributions for input random variables.
- For dealing with problems in operation framework from a few minutes (very short-term operation) to a few hours (short-term operation) and within 24 hours (day ahead), approximation and analytical methods are suitable to be applied. However, for a large power system, cumulant method, especially the PPPF, is recommended. Information associated with probability distributions of loads and power generation from wind and solar resources for each

time-step can be provided by a probabilistic forecasting technique (Botterud *et al.* 2011) or a scenario-based forecasting technique (Le *et al.* 2015), then PPF is performed using these distributions.

7. Conclusions

In this paper, a new framework is proposed for probabilistic assessment of power systems accounting for uncertainties from input random variables (i.e., load, renewable power generation, etc.) based on the cumulant-based PPF in combination with an improved clustering technique. The improved clustering technique, developed in this study, helps to enhance the cumulant PPF method to deal with large range of variation in the input random variables. In addition, the combination of Principal Component Analysis and Particle Swarm Optimization in the improved clustering technique makes the cumulant PPF method able to apply for a large power system with large number of random variables associated with loads, renewable energy sources and so on.

The proposed method is tested on the modified IEEE-118 bus test system and its results are compared with the results obtained by MCS and the traditional cumulant-based PPF.

From theoretical analysis and case study on the modified IEEE-118 bus test system, some conclusions are summarized as follows:

- When the input random variables of PPF problem have large variations, the TCPPF could result in significant errors, especially at the tails of the probability distributions, so it is not suitable to solve PPF for a system with large variations of input variables.
- The proposed method can deal with large variations of input variables thanks to adopting the improved clustering technique. The case study clearly point out that the proposed method can achieve higher accuracy than TCPPF while remaining higher efficiency compared to MCS.
- From a practical point of view, PPPF method is suitable to apply for both planning and operation of power systems in practice. It is able to apply for large power systems with large number of random variables.

In this paper, we mostly focus on development of the improved clustering technique and the proposed framework for probabilistic assessment of power systems considering uncertainties from loads and renewable energy sources. However, in a real power system, there exist several sources of uncertainty following various types of probability distribution. In future work, further study on the proposed method should be directed to incorporate those factors into the PPF problem and the assessment of power systems. The improved clustering technique proposed in this paper can be considered for application to manage input data in solving other problems in power systems in particular as well as other fields if needed.

Acknowledgments

This research was funded by the Ministry of Education and Training, Vietnam under project number B2020-DNA-02.

References

Abdmouleh, Z., Gastli, A., Ben-Brahim, L., Haouari, M., & Ahmed Al-Emadi, N.A. (2017). Review of optimization techniques applied for the integration of distributed generation from renewable energy sources. *Renewable Energy*, 113, 266-280; http://dx.doi.org/10.1016/j.renene.2017.05.087

Aien, M., Khajeh, M.G., Rashidinejad, M., & Fotuhi-Firuzabad, M. (2014). Probabilistic power flow of correlated hybrid wind-photovoltaic power systems. *IET Renew Power Gen.*, 8, 649-658; https://doi.org/10.1049/iet-rpg.2013.0120

Allan, R.N., & Al-Shakarchi, M.R.G. (1997). Probabilistic techniques in A.C. load-flow analysis. *Proc. Inst. Elect. Eng.*, 124, 154-160; https://doi.org/10.1049/piee.1977.0027

Allan, R.N., Leite da Silva, A.M., & Burchett, R.C. (1981). Evaluation Methods and Accuracy in Probabilistic Load Flow Solutions. *IEEE Transactions on Power Apparatus and Systems*, PAS-100, 2539-2546; https://doi.org/10.1109/TPAS.1981.316721

Bandyopadhyay, S., & Maulik, U. (2002). Genetic clustering for automatic evolution of clusters and application to image classification. *Pattern Recognit*, 35(6), 1197-1208; https://doi.org/10.1016/S0031-3203(01)00108-X

Bergen, A.R., & Vittal, V. (1986) *Power Systems Analysis*. Prentice-Hall: NJ, USA.

Borkowska, B. (1974). Probabilistic load flow. *IEEE Trans. Power App. Syst.*, PAS-93, 752-759; https://doi.org/10.1109/TPAS.1974.293973

Botterud, A., Zhi, Z., Wang, J., Bessa, R.J., Keko, H., Mendes, J., Sumaili, J., & Miranda, V. (2011). Use of wind power forecasting in operations decisions. Tech. Rep. ANL/DIS-11-8, Argonne Nat. Lab.

Cai, D., Chen, J., Shi, D., Duan, X., Li, H., & Yao, M. (2012). Enhancements to the cumulant method for probabilistic load flow studies. *Proceedings of 2012 IEEE power and energy society general meeting*, San Diego, USA, pp. 1-8; https://doi.org/10.1109/PESGM.2012.6343972

Christie, R., Power System Test Case Archive. (1993). Available online: https://www2.ee.washington.edu/research/pstca/pf118/pg_tca118bus.htm (accessed on 21 April 2021)

Deng, X., He, J., Zhang, P., & Sciubba, E. (2017). A Novel Probabilistic Optimal Power Flow Method to Handle Large Fluctuations of Stochastic Variables. *Energies*, 10(10), 1623; https://doi.org/10.3390/en10101623

Deng, X., Zhang, P., Jin, K., He, J., Wang, X., & Wang, Y. (2019). Probabilistic load flow method considering large-scale wind power integration. *Journal of Modern Power Systems and Clean Energy*, 7(4), 813-825; https://doi.org/10.1007/s40565-019-0502-0

Falkenauer, E. (1998) *Genetic Algorithms and Grouping Problems*. John Wiley & Sons, Inc.: New York, NY, USA.

Fan, M., Vittal, V., Heydt, G.T., & Ayyanar, R. (2012). Probabilistic Power Flow Studies for Transmission Systems with Photovoltaic Generation Using Cumulants. *IEEE Trans. Power Syst.*, 27(4), 2251-2261; https://doi.org/10.1109/TPWRS.2012.2190533

Gan, G., Ma, C., & Wu, J. (2007) *Data Clustering: Theory, Algorithms, and Applications*. SIAM: Philadelphia, PA, USA,.

Gupta, N. (2016). Probabilistic load flow with detailed wind generator models considering correlated wind generation and correlated loads. *Renewable Energy*, 94, 96-105; https://doi.org/10.1016/j.renene.2016.03.030

Hajian, M., Rosehart, W.D., & Zareipour, H. (2013). Probabilistic power flow by Monte Carlo simulation with Latin Supercube sampling. *IEEE Trans. Power Syst.*, 28(2), 1550-1559; https://doi.org/10.1109/TPWRS.2012.2214447

Hasan, K.N., Preece, R., & Milanović, J.V. (2019). Existing approaches and trends in uncertainty modelling and probabilistic stability analysis of power systems with renewable generation. *Renewable and Sustainable Energy Reviews*,101, 168-180; https://doi.org/10.1016/j.rser.2018.10.027

Huang, J., Xue, Y., Dong, Z.Y., & Wong, K.P. (2011). An adaptive importance sampling method for probabilistic optimal power flow. *Proceedings of the 2011 IEEE Power and Energy Society General Meeting*, Detroit, MI, USA, pp. 1-6; https://doi.org/10.1109/PES.2011.6039167

Jackson, J.E. (1991) *A Users Guide to Principal Component Analysis*. Hoboken, Wiley: NJ, USA.

Jolliffe, I.T. (2002) *Principal Component Analysis*. Springer: New York, NY, USA.

Le, D.D., Gross, G., & Berizzi, A. (2015). Probabilistic Modeling of Multisite Wind Farm Production for Scenario-based Applications. *IEEE Transactions on Sustainable Energy*, 6(3), 748-758; https://doi.org/10.1109/TSTE.2015.2411252

Le, D.D., Berizzi, A., & Bovo, C. (2016). A probabilistic security assessment approach to power systems with integrated wind resources. *Renewable Energy*, 85, 114-123; https://doi.org/10.1016/j.renene.2015.06.035

Leite da Silva, A.M., & Milhorance de Castro, A. (2018). Risk Assessment in Probabilistic Load Flow via Monte Carlo Simulation and Cross-Entropy Method. *IEEE Trans. Power Syst.*, 34(2), 1193 - 1202; https://doi.org/10.1109/TPWRS.2018.2869769

Liu, Y., Gao, S., Cui, H., & Yu, L. (2016). Probabilistic load flow considering correlations of input variables following arbitrary distributions. *Electr. Power Syst. Res.*, 140, 354-362; https://doi.org/10.1016/j.epsr.2016.06.005

Morales, J.M., & Perez-Ruiz, J. (2007). Point estimate schemes to solve the probabilistic power flow. *IEEE Trans Power Syst*, 22(4), 1594-1601; https://doi.org/10.1109/TPWRS.2007.907515

Ruiz-Rodriguez, F.J., Hernandez, J.C., & Jurado, F. (2012). Probabilistic load flow for photovoltaic distributed generation using the Cornish-Fisher expansion. *Electr. Power Syst. Res.*, 89, 129-138; https://doi.org/10.1016/j.epsr.2012.03.009

Su, C.L. (2005). Probabilistic load-flow computation using point estimate method. *IEEE Trans. Power Syst.,* 20, 1843-1851; https://doi.org/10.1109/TPWRS.2005.857921

Tao, C., & Franchetti, F. (2013). A Quasi-Monte Carlo approach for radial distribution system probabilistic load flow. *In Proceedings of the IEEE PES Innovative Smart Grid Technologies (ISGT)*, Washington, DC, USA, pp. 1-6; https://doi.org/10.1109/ISGT.2013.6497894

Van der Merwe, D.W., & Engelbrecht, A.P. (2003). Data clustering using particle swarm optimization. *Proc. 2003 Congress on Evolutionary Computation*, 1, 215-220; https://doi.org/10.1109/CEC.2003.1299577

Xie, Z.Q., Ji, T.Y., Li, M.S., & Wu, Q.H. (2018). Quasi-Monte Carlo Based Probabilistic Optimal Power Flow Considering the Correlation of Wind Speeds Using Copula Function. *IEEE Trans. Power Syst.*, 33, 2239-2247; https://doi.org/10.1109/TPWRS.2017.2737580

Zhang, P., & Lee, S.T. (2004). Probabilistic load flow computation using the method of combined cumulants and Gram-Charlier expansion. *IEEE Trans. Power Syst.*, 19(1), 676-682; https://doi.org/10.1109/TPWRS.2003.818743

Zhou, Z., Tang, F., Liu, D., Wang, C., & Gao, X. (2020). Probabilistic Assessment of Distribution Network with High Penetration of Distributed Generator. *Sustainability*, 12, 1709; https://doi.org/10.3390/su12051709