Probabilistic Optimal Power Flow Considering Correlation of Wind Farms via Markov Chain Quasi-Monte Carlo Sampling

Weigao Sun, Student Member, IEEE, Mohsen Zamani, Hai-Tao Zhang, Senior Member, IEEE, Yuanzheng Li

Abstract—The probabilistic characteristics of daily wind speed are not well captured by simple density functions such as Normal or Weibull distributions as suggested by the existing literature. The unmodeled uncertainties can cause unknown influences on the power system operation. In this paper, we develop a new stochastic scheme for the probabilistic optimal power flow (POPF) problem, which can cope with arbitrarily complex wind speed distributions and also take into account the correlation of different wind farms. A multivariate Gaussian mixture model (GMM) is employed to approximate actual wind speed distributions from multiple wind farms. Furthermore, we propose to adopt the Markov Chain Monte Carlo (MCMC) sampling technique to deliver wind speed samples as the input of POPF. We also novelly integrate a Sobol-based quasi-Monte Carlo (QMC) technique into the MCMC sampling process to obtain a faster convergence rate. The IEEE 14- and 118-bus benchmark systems with additional wind farms are used to examine the effectiveness of the proposed POPF scheme.

Index Terms—Probabilistic optimal power flow, Gaussian mixture model, Markov chain Monte Carlo, quasi-Monte Carlo, uncertainty.

I. INTRODUCTION

Wind power generation is experiencing tremendous developments as a clean and renewable energy resource [1]. It considerably contributes to the long-term sustainability of power systems, whereas, introduces significant uncertainties into the overall networks as well. Uncertainty involved with wind power generation may cause operational problems, such as overload of transmission lines, which in turn threaten the reliability and security of power system [2, 3]. It is challenging to study how the uncertainty involved with wind power generation will influence power system operations. Therefore, probabilistic optimal power flow (POPF), as a powerful tool to analysis uncertainties, has attracted considerable attention [4–14]. Instead of calculating the traditional deterministic optimal power flow (DOPF) [15], POPF treats each uncertain variable in power systems as a random variable with certain probabilistic distribution and aims to obtain the statistical information of the optimal solutions. By evaluating the statistical information of output variables, e.g., mean, standard deviation or even probabilistic density function, it is promising to figure out the potential risk and weakness of the power system under investigation.

The existing literature concerning the POPF problem falls into three categories: analytical-based methods [9, 10, 14], the point estimation [4, 12, 13, 16] and Monte Carlo (MC) simulations [7, 17–19]. Considering the large computation burden of solving POPF, analytical methods were developed first. The essential idea of analytical methods is to compute statistical moments of output variables in POPF based on the moments associated with input variables. For instance, [10] developed a cumulant method for POPF problem, which assumed the relationship between linear input and cumulant output. Then, if the input variables follow Gaussian or Gamma distributions, which are analytical known, cumulant method could be deployed to obtain statistical solutions of output variables. In [9], the authors proposed a method that exploits the first-order Taylor series expansion, therein the first two moments of the input load power could be used to obtain the statistical information of the output variables. Analytical methods like [10] and [9] compute swiftly, however, suffer from the issue of accuracy. The implementation of analytical methods in practice also depends heavily on the particular optimal power flow formulations.

Point estimation method has been adopted to solve the POPF problem as well as the probabilistic power flow (PPF) problem [17–22]. In [12] and [13], a two-point estimation method was proposed for POPF problem. It is worth mentioning that in [12], the correlation of input variables was described by the coefficients matrix. The reference [4] addressed the POPF problem by calculating correlation of wind speeds via improving the point estimation method. However, the point estimation scheme only calculates first few statistical moments, which is not accurate enough. Meanwhile, the computational burden is proportional to the uncertain variables numbers, which hinders the further applications in large-scale power systems.

In addition, MC methods have been widely studied for PPF and POPF problems. With the samples from probabilistic density of input variables, the deterministic power flow or optimal power flow is calculated repeatedly which generates samples of output variables. Routine MC method with large enough repeating times (e.g. 10000 times) can give sufficiently accurate results. However, it is computationally expensive. To solve such a dilemma, improved sampling methods were em-
ployed to reduce the computational burden. For instance, Latin hypercube sampling (LHS) [17], Latin supercube sampling [18] and quasi-MC (QMC) methods [7][19] are representative works.

Correlation of the random input variables in PPF or POPF problem has attracted much attention in recent years. Ref. [12] adopted the correlation matrix technique into the point estimation method to address the correlation of wind power generation and loads. Ref. [4] studied correlation of wind speeds with different distributions. Copula function was utilized in [7] to describe the dependent structure of random wind speeds. In [23], a Gaussian mixture model (GMM) was proposed to approximate the probabilistic distribution of loads. In [5] and [24], the multivariate GMM was adopted to describe the wind power uncertainties and their correlation.

Fig. 1. Distribution of actual daily wind speed.

As we will show later in the paper, the real world daily wind speed data obeys unknown complex distributions, not Weibull distribution which is commonly used to model the probabilistic property of wind speed, see e.g., [4] and [5]. In this paper, we develop a novel stochastic scheme for solving POPF which has no such presumption on the probability density function associated with the collected data. Compared to the existing relevant state-of-the-art works, the proposed scheme can cope with arbitrarily complex wind speed distributions and also take into account the correlation of different wind farms with the help of multivariate GMM. This is the first contribution of this work. Given complications arise from sampling multivariate GMM, the second contribution lies in adopting the powerful Markov chain Monte Carlo (MCMC) sampling technique to obtain inout samples for POPF and also newly integrating Sobol-based QMC method into the MCMC sampling process for a faster convergence rate.

The paper is organized as follows. Section II provides a preliminary of the probabilistic characteristics of power system uncertainties, and thereby presents the POPF problem. Section III introduces the multivariate GMM to represent the joint distribution of wind speeds in multiple wind farms. In Section IV, An MCMC sampler is given and then improved by integrating the sobol-based QMC technique. Section IV-C summarizes the proposed scheme for solving POPF problem. The present scheme is examined by case studies in Section V. Finally, conclusion is drawn in Section VI.

II. PRELIMINARY

We review the sources of uncertainties in power systems and then introduce the POPF problem in this section.

A. Uncertainties in Power Systems

The power system considered in this paper can be sketched as a network represented by a connected undirected graph \((N, E)\), where \(N\) and \(E\) are sets of buses and branches in a power system, respectively. Bus 0 is the slack bus, without loss of generality, its voltage phase angle is taken as reference and set as zero [25]. Other buses are classified into generator or load buses. Here, we consider a power system including both traditional thermal and wind power generations.

The uncertainties of such a power system mostly stems from the randomness associated with load and wind speeds. The load, as an uncertain variable, is often influenced by the usage time, the market electrical price and even the weather condition. It is a common practice to model the probabilistic distribution of load as a normal distribution with parameters obtained from historical data. Thereby, we follow the same remedy and describe the load as a normal distribution and set its mean equal to the base load, and its standard deviation to 5% of its mean [12].

The randomness of wind power generation also plays an important role in the uncertainties of power systems. Wind speeds vary with the time, weather and the location of wind farm and this in turn results in variation of wind power generation. The probabilistic distribution of wind speed is commonly claimed to follow the Weibull distribution in literatures [26]. However, it could also obey the Burr or lognormal distribution, or the combination of Weibull, Burr and lognormal distributions, as claimed in [4]. Indeed, the distribution of real daily wind speed can be arbitrarily complex instead of following several simple distributions, as shown in Fig. 1. Here, the wind speed data is collected from the Measurement and Instrumentation Data Center under the National Renewable Energy Laboratory\(^1\).

A mapping for transforming wind speed to electrical power in a wind turbine is typically described as:

\[
P_i(v) = \begin{cases} 
0, & v \leq v_{in}, \\
f(v), & v_{in} \leq v \leq v_t, \\
P_f, & v_t \leq v \leq v_{out}, \\
0, & v \geq v_{out}, 
\end{cases}
\]  

(1)

where \(v_{in}, v_{out}, v_t\) are the cut-in, cut-out and rated wind speed, respectively. And \(P_i\) is the real power generation of a wind turbine. Here \(f(v)\) represents the generation mechanism of the wind turbine in standard working ranges [7]. The real power \(P^W\) and reactive power \(Q^W\) generated by wind turbines in the \(j\)-th wind farm are shown as follows

\[
P^W_j = P_i \cdot N_i, \\
Q^W_j = \frac{P^W_j}{\cos \varphi} \cdot \sqrt{1 - \cos^2 \varphi},
\]

\(^1\)http://midcdmz.nrel.gov/
where \( N_t \) is the number of wind turbines in a wind farm and \( \cos \phi \) is the power factor.

**B. POPF**

We now present the formulation of POPF problem. In a power system, the variables in DOPF fall into two categories: the control variables \( u = [P^G_i, V_i | i \in N'_G] \) and the state variables \( x = [Q^G, \theta, V_j | j \in N] \). The DOPF aims to find the minimum of power generation cost by adjusting the control variables \( u \) subject to the power flow equations and other security constraints. The DOPF problem solves the nonlinear constrained optimization problem:

\[
\min_{u} f(u)
\]

subject to

\[
P^G_i + P^W_i - P^D_i = \sum_{j=1}^{N} V_j (G_{ij} \cos \theta_{ij} + B_{ij} \sin \theta_{ij}),
\]

\[
Q^G_i + Q^W_i - Q^D_i = \sum_{j=1}^{N} V_j (G_{ij} \sin \theta_{ij} - B_{ij} \cos \theta_{ij}),
\]

\[
V_{\min} \leq V_i \leq V_{\max},
\]

\[
P^G_{\min} \leq P_k \leq P^G_{\max},
\]

\[
Q^G_{\min} \leq Q_k \leq Q^G_{\max},
\]

\[
P^W_{\min} \leq P_j \leq P^W_{\max},
\]

\[
Q^W_{\min} \leq Q_j \leq Q^W_{\max},
\]

\[
|S_{cd}| \leq S_{cd,\max},
\]

\[
|P_{cd}| \leq P_{cd,\max},
\]

\[
V_c - V_d \leq \Delta V_{\max},
\]

where \( \forall i \in N'_G, \forall k \in N_G, \forall j \in N_c, \forall (c, d) \in N_I \) and \( f(\cdot) \) is in general a convex polynomial objective function, see e.g. [25] and [27]. In this paper, we set \( f(\cdot) \) as:

\[
f(P^G) = \sum_{k=1}^{n_k} \sum_{j=1}^{d_j} f_k (P^G_k) + \sum_{j=1}^{d_j} d_j
\]

where \( a_k, b_k \) and \( c_k \) are the constant, linear and quadratic coefficient of the cost of \( k \)-th traditional thermal generator, respectively; \( d_j \) is the constant maintenance cost of \( j \)-th wind farm.

The minimization problem (2) and its associated nonlinear constraints can be written into a compact form as

\[
Z = h(W),
\]

with \( W = [P^G, V, P^W, Q^W, P^D, Q^D | i \in N'_G] \). The parameter \( Z \) captures those variables of interest, e.g., the generation cost, bus voltage, active and reactive power flow. In order to capture the uncertainty involved in generation of wind power, we presume that \( P^W, Q^W \) follow certain unknown probability distribution functions. By using sampling techniques, we obtain samples from the underlying distributions associated with those random variables in \( W \). Then the DOPF is performed recursively with the samples of these variables, which in turn yields output samples. One can exploit these samples information to attain estimation about statistical properties of these variables of interest. Afterwards, evaluating these statistical information, it is possible to find the potential risk and the weakness of the power system under investigation. For example, such analysis can provide the probabilities of line overload or bus overvoltage.

**III. MULTIVARIATE GAUSSIAN MIXTURE MODEL**

To approximate wind speed distribution, we will seek assistance from the GMM, which is a probabilistic model that assumes all the data points are generated from a mixture of a finite number of Gaussian distributions with unknown parameters [28]. Each Gaussian distribution is called a Gaussian component. GMM can capture arbitrarily complex distributions by using specific number of Gaussian components with different parameters [29]. Therefore, GMM has been widely used in data classification and machine learning. Meanwhile, it has been verified to be able to model the uncertainties of power systems in [5], [23] and [24].

A GMM is called multivariate if each of its components is multi-dimensional Gaussian distribution. Given a random vector \( x = [x_1, x_2, \cdots, x_D]^T \), the joint probability density function (PDF) of multivariate GMM is

\[
p(x) = \sum_{m=1}^{M} \frac{c_m}{(2\pi)^{D/2}|\Sigma_m|^2} \exp\left[-\frac{1}{2} (x - \mu_m)^T \Sigma_m^{-1} (x - \mu_m) \right]
\]

\[
= \sum_{m=1}^{M} c_m \cdot N(x | \mu_m, \Sigma_m),
\]

where \( M \) is the number of Gaussian components, which is set as a priori according to probabilistic characteristics of data. Random vector \( x \sim \mathcal{N}(\mu_m, \Sigma_m) \in \mathbb{R}^D \), here \( \mathcal{N}(\cdot) \) denotes the \( D \)-dimensional Gaussian distribution function, \( \mu_m \) and \( \Sigma_m \) are the expectation and covariance matrix of \( m \)-th \( D \)-dimensional Gaussian component, respectively. The positive mixture weights \( c_m \) sum to unity, i.e., \( \sum_{m=1}^{M} c_m = 1 \).

The multivariate GMM in (4) has a parameter set \( \Theta = \{c_m, \mu_m, \Sigma_m | m = 1, 2, \cdots, M\} \) to be determined. Estimating the parameter set \( \Theta \) based on historical data is known as a learning process. Here, we focus on the expectation maximization (EM) algorithm for GMM parameter estimation. EM is an iterative procedure for maximum likelihood parameter estimation from dataset with latent variables. To estimate the parameter set \( \Theta \), we write the log-likelihood function as

\[
l(\Theta) = \log \left( \sum_{m=1}^{M} c_m \cdot N(x, z | \mu_m, \Sigma_m) \right),
\]

where \( x \) is the observed data and \( z \) is the unobserved latent data. EM algorithm contains two main steps: E-step, it guesses the values of latent data \( z \); M-step, it assumes the gaussses of \( z \) are correct, and applies the maximum likelihood estimation to update \( \Theta \). The EM algorithm which is widely used and proved to be effective in practice has been embedded in many commercial softwares.
IV. SAMPLING METHOD

To implement the POPF calculation, efficient sampling technique should be designed to obtain sufficient wind power samples as inputs of problem (3). Due to the complications arise from sampling a multivariate GMM, we propose to exploit the MCMC sampling technique to yield sufficient wind speed samples, then transform them to wind power samples by the mapping Eq. (1). Furthermore, a sobol-based QMC technique is integrated into the MCMC sampling to obtain faster convergence rate.

A. MCMC Sampler

MCMC is a powerful sampling technique which can provide samples from arbitrary probability density \( p(x) \), which has been adopted for wind power simulation in [30]. It works in two stages: proposal and acceptance. Given \( x_k \), a candidate point \( \xi_k \) is drawn from a proposal distribution \( q(\xi_k|x_{k-1}) \), where \( \xi_k \sim q(\cdot|x_{k-1}) \) is a possible realization for \( x_k \). Then compute the acceptance probability

\[
\alpha(\xi_k|x_{k-1}) = \min \left\{ 1, \frac{p(\xi_k)q(x_{k-1}|\xi_k)}{p(x_{k-1})q(\xi_k|x_{k-1})} \right\},
\]

and draw a random variable \( \bar{z} \) from uniform distribution \( \mathcal{U}(0,1) \). If \( \bar{z} < \alpha(\xi_k|x_{k-1}) \), accept \( \xi_k \) and set \( x_k = \xi_k \). Otherwise, reject \( \xi_k \) and set \( x_k = x_{k-1} \).

MCMC generates a Markov chain \((x_0, x_1, \ldots, x_k)\), as the transition probabilities from \( x_k \) to \( x_{k+1} \) depends only on \( x_k \). After a sufficient burn-in period, for example, \( k \) steps, the Markov chain approaches its stationary distribution. Then, the samples in \((x_{k+1}, \ldots, x_{k+n})\) are the samples from \( p(x) \), where \( n \) is the number of samples.

Choice of the proposal distribution \( q(\xi_k|x_{k-1}) \) has a significant influence on the performance of the MCMC sampler. A widely used proposal distribution is obtained from the random walk below:

\[
\xi_k = x_{k-1} + v_k,
\]

where \( v_k \) is a random perturbation which is commonly a white noise. Hence the proposal distribution \( \xi_k \) is symmetric

\[
q(\xi_k|x_{k-1}) = q(x_{k-1}|\xi_k).
\]

In this specific situation, the acceptance probability in (5) becomes

\[
\alpha(\xi_k|x_{k-1}) = \min \left\{ 1, \frac{p(\xi_k)p(x_{k-1})}{p(x_{k-1})p(\xi_k)} \right\}.
\]

The Eq. (6) provides an intuitive explanation of the MCMC sampler. By this means, \( x_k \) converges to the target distribution \( p(x) \) by accepting or rejecting a candidate proposal \( \xi_k \). If the proposal \( \xi_k \) is more likely to be a realization of the density \( p(x_k) \) than the previous iteration \( x_{k-1} \), i.e., \( p(\xi_k) > p(x_{k-1}) \), then the proposal \( \xi_k \) is accepted as a realization \( x_k = \xi_k \). Otherwise, it still has a chance to be retained. In fact, the proposal \( \xi_k \) less likely than \( x_{k-1} \) is retained with a probability \( p(\xi_k)/p(x_{k-1}) \leq 1 \).

Algorithm 1 QMC-MCMC Sampler

Step 1. Initialize \( x_0 \) satisfying \( p(x_0) > 0 \), and set \( k = 1 \).
Step 2. At iteration \( k \), draw a candidate point \( \xi_k \) by Sobol-based QMC method from a proposal distribution \( q(\xi_k|x_{k-1}) \), where \( \xi_k \sim q(\cdot|x_{k-1}) \) is a possible realization for \( x_k \).
Step 3. Compute the acceptance probability

\[
\alpha(\xi_k|x_{k-1}) = \min \left\{ 1, \frac{p(\xi_k)q(x_{k-1}|\xi_k)}{p(x_{k-1})q(\xi_k|x_{k-1})} \right\}.
\]

Step 4. Draw a random variable \( \bar{z} \) by Sobol-based QMC method from uniform distribution \( \mathcal{U}(0,1) \). If \( \bar{z} < \alpha(\xi_k|x_{k-1}) \), accept \( \xi_k \) and set \( x_k = \xi_k \). Otherwise, reject \( \xi_k \) and set \( x_k = x_{k-1} \).
Step 5. Set \( k = k+1 \) and return to step 2.

B. Improving MCMC by Integrating QMC Method

MCMC sampler works well for sampling from arbitrary probability density functions, however, it suffers from the costly computation burden and slow convergence rate [31]. This motivates us to integrate QMC into MCMC to obtain faster convergence rate [31]. Consider a quantity \( \mu \), which is of interest, can be expressed as \( E(f(X)) \) for a real valued function \( f(\cdot) \) and random vector \( X \) with probability density \( p(\cdot) \) on \( \mathbb{R}^d \). Then, \( \mu \) can be expressed as \( \int_{\mathbb{R}^d} f(x)p(x)dx \). In simple MC, one estimates \( \mu \) by

\[
\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^{n} f(x_i),
\]

where \( x_i \) for \( i = 1, \ldots, n \) are independent random samples obtained from \( p(\cdot) \). A pseudo-random number generator is commonly used to simulate the \( x_i \) values, which is known as the simple random sampling (SRS). QMC is a variant of simple MC to obtain a higher rate of convergence by using low-discrepancy sequences [32]. In QMC, the \( x_i \) values chosen deterministically are more uniformly distributed than the pseudo-random numbers in SRS. There exist several different ways to generate quasi-random low-discrepancy sequences which result in different instances of QMC method, such as the Sobol sequence [7] and Latin hypercube sample [17]. Here, we focus on the Sobol sequence whose implementation is introduced in [33]. Fig. 2 depicts 10000 uniform random numbers generated by Sobol technique, Latin hypercube sample and pseudo-random number technique. It is observed that Sobol sequence results in more uniformly distributed (or low-discrepancy) samples than other ones. We now discuss the convergence rate of QMC compare to MC method. To this end, we recall the follow definition.

**Definition 4.1:** (Star discrepancy [34]) Let \( \delta(a) = \text{Vol}([0,a]^d) - \frac{1}{n} \sum_{i=1}^{n} I_{x_i \in [0,a]^d} \) be the local discrepancy function at point \( a \in [0,1]^d \). Here \( \text{Vol}(S) \) is the \( d \)-dimensional volume of set \( S \), and \([0,a] \) denotes a \( d \)-dimensional box with 0 and \( a \) at opposite corners. \( I_{x_i \in [0,a]^d} \) is the indicator function defined as \( I = 1 \) for \( x_i \in [0,a] \). Otherwise, \( I = 0 \). The star discrepancy is

\[
D_n^* = D_n^*(x_1, \ldots, x_n) = \sup_{a \in [0,1]^d} |\delta(a)|,
\]
when \( D^*_n \to 0 \), then \( \hat{\mu}_n \to \mu \).

For random numbers generated by QMC, their star discrepancy satisfies \( D^*_n = \mathcal{O}(n^{-1} \log(n)^{d-1}) \) as \( n \to \infty \). Thus, the convergence rate of QMC is \( \mathcal{O}(n^{-1+\epsilon}) \) for any \( \epsilon > 0 \), which is faster than that of MC which is \( \mathcal{O}(n^{-0.5}) \) [35]. Empirical comparisons demonstrate that QMC often outperforms MC for a reasonable sample number \( n \).

Due to the computational advantage of QMC, it is of interest to integrate it into MCMC in order to obtain a faster convergence rate. We call the resulted method as QMC-MCMC. It is easy to implement the QMC-MCMC by replacing the MC points by QMC points to generate the proposals and acceptances in an MCMC sampler. Details of the QMC-MCMC sampler is presented in Algorithm 1. We provide the theoretically consistency guarantee of QMC-MCMC as below.

**Definition 4.2:** (CUD) If for every integer \( d \geq 1 \), the points \( z_i = (u_1, \cdots, u_{d+i-1}) \in [0,1]^d \) satisfy \( \lim_{n \to \infty} D^*_n(z_1, \cdots, z_n) = 0 \). Then the sequence \( u_1, u_2, \cdots \in [0,1] \) is CUD.

The Sobol sequence we employed in this paper is CUD. It can be integrated into the Metropolis Hastings algorithm to generate a consistent Markov chain as given in Lemma 4.1. The detailed consistency proof is further explained in [35] and [36].

**Lemma 4.1:** Consider Markov chains with finite state spaces \( \Omega = \{ \omega_1, \cdots, \omega_K \} \), let \( x_i \in \Omega \) for \( i \geq 1 \) be sampled from the standard construction for Markov chains, using a CUD sequence \( u_i \). Assume that all \( K^2 \) transition probabilities are positive. Then \( \hat{\mu}_n \to \mu \) holds as \( n \to \infty \) for all bounded \( f \).

**C. Proposed POPF Scheme**

The proposed scheme for solving POPF with considering correlation of wind speeds with arbitrary distribution is summarized in Algorithm 2. Note that in the last step of Algorithm 2, mean \( \mu \) or standard deviation (STD) \( \sigma \) is just rough analysis results with the output variable samples. If necessary, we can further establish the probability density function of output variable.

**Algorithm 2 Proposed POPF Scheme**

**Step 1.** Given \( D \) wind farms in a power system, collect wind speed data \( x = [x_1, x_2, \cdots, x_D]^T \) where \( x_i \) with \( i = 1, 2, \cdots, D \), represents the wind speed data of \( i \)-th wind farm.

**Step 2.** Obtain the joint PDF of the multivariate GMM

\[
p(x) = \sum_{m=1}^{M} c_m \mathcal{N}(x|\mu_m, \Sigma_m),
\]

with considering the correlation of \( D \) wind farms. Where the parameter set \( \Theta = \{ c_m, \mu_m, \Sigma_m | m = 1, 2, \cdots, M \} \) is determined by EM algorithm.

**Step 3.** Draw wind speed samples from \( p(x) \) via the QMC-MCMC sampler in Algorithm 1.

**Step 4.** Transform wind speed samples into wind power, then recursively calculate the DOPF problem (3) with each wind power sample.

**Step 5.** Collect the output variable samples and compute their statistical information such as mean \( \mu \) or STD \( \sigma \).

V. CASE STUDIES

Case studies will be conducted on two benchmark systems, the modified IEEE 14-bus and 118-bus systems. Eight wind turbines are integrated in one wind farm, the rated power of each wind turbine is 5 MW. The rated, cut-in and cut-out wind speeds are set as 12 m/s, 2 m/s and 18 m/s [4]. The simulation was conducted on an MacBook Pro with 64-bit Intel i5 CPU at 2.3GHz and 8GB of RAM. MATPOWER, an MATLAB power system simulation toolbox, was adopted to solve the deterministic optimal power flow [37].

The accuracy of POPF solving in this paper is estimated by calculating the errors of mean and standard deviation compared with the accuracy reference values. POPF solutions using SRS with large enough sample size, e.g., \( N = 10000 \), is set as the accurate reference. The error index is defined as

\[
E^*_f = \left| \frac{I^*_a - I^*_s}{I^*_a} \right| \times 100\%.
\]

Here \( I \) is the statistical property such as the mean \( \mu \) or STD \( \sigma \) associated with POPF output variables. \( I_a \) is the accurate reference value obtained from SRS, while \( I_s \) is the simulated results using a certain sampling method with \( N \) samples, here \( N \leq 10000 \). The symbol * can be any output variables of the POPF computation such as the optimal cost, bus voltage or power flow.

A. Modeling and Sampling Wind Speed

The wind speed data of six wind farms is collected from the Measurement and Instrumentation Data Center under the National Renewable Energy Laboratory. The wind speed data on minutely basis for one day at six different wind farms is
Fig. 3. Target Gaussian mixture models of six wind farms and their samples by the proposed QMC-MCMC sampler.

| Variables | Methods | Mean $\mu$ (%) | STD $\sigma$ (%) |
|-----------|---------|----------------|-----------------|
| Cost ($)  | Ref     | 12270          | 1225.1          |
|           | SRS-MCMC| 12295          | 0.2054          | 1286.1          | 4.9645          |
|           | LHS-MCMC| 12303          | 0.2306          | 1242.1          | 1.8076          |
|           | QMC-MCMC| 12270          | 0               | 1198.2          | 2.1940          |
| V(p.u.)   | Ref     | 1.0225         | 0.0220          |
|           | SRS-MCMC| 1.0232         | 0.0901          | 0.0212          | 3.0764          |
|           | LHS-MCMC| 1.0219         | 0.0654          | 0.0225          | 2.3378          |
|           | QMC-MCMC| 1.0222         | 0.0374          | 0.0226          | 2.5422          |

Fig. 4. Performance comparisons of three sampling methods on the modified IEEE 14-bus system along increasing sampling size $N$.

The conventional generator on bus 3 is removed and replaced by the wind farm 1. The rated capability of these three wind farms is 40 MW. We assume the three wind farms in the modified IEEE 14-bus system are mutually influential, the correlation among them is considered.

We calculated the POPF results provided by SRS with 10000 samples as the accurate reference values. The MCMC sampling method is improved by the Latin hypercube sampling and Sobol-based QMC technique. Note that here we abbre-
violate the MCMC method with simple random sampling as SRS-MCMC, and MCMC improved by the Latin hypercube sampling and Sobol-based QMC technique as LHS-MCMC and QMC-MCMC, respectively. Performances of these three sampling methods with sampling size $N = 2000$ are compared against the accurate reference values in Table I. Note that the voltage magnitude error index $\epsilon_\mu^V$ in Table I is measured at bus 12 in the IEEE 14-bus system. It is observed that the proposed QMC-MCMC method almost achieves relatively small mean error index $\epsilon_\mu$ and STD error index $\epsilon_\sigma$. To further compare their performance, we plot the mean error index of optimal cost $\epsilon_{cost}^\mu$ and voltage magnitude $\epsilon_\mu^V$ along increasing sampling sizes in Fig. 4. The error index associated with QMC-MCMC method always keeps smaller than those of other two methods which verifies the virtue of the former.

### C. IEEE 118-bus System

We now investigate the proposed POPF scheme on the IEEE 118-bus system [4][7]. As shown in Table II, the 118-bus system is divided into three areas with each area has been integrated into two wind farms. We consider the correlation between every two wind farms in each area. Specially, wind farms 1, 2 are integrated to bus 11, 17, wind farms 3, 4 are integrated to bus 37, 51, wind farms 5, 6 are integrated to bus 83, 96. The rated capability of these six wind farms is 40MW.

Table III provides the performance comparisons of three sampling methods. Here, the Voltage angle $\theta$ at bus 98 and power flow $P_{line}$ at line 69-70 is measured. Mean error indexes of voltage angle $\epsilon_\theta$ and power flow $\epsilon_{P_{line}}$ with increasing sampling sizes are shown in Fig.5. As can be seen, the QMC-MCMC method outperforms the other two ones.

### VI. Conclusions

In this paper, we developed a new stochastic scheme for POPF problem based on the multivariate GMM and Markov chain quasi-Monte Carlo sampling technique. The scheme approximates arbitrarily complex wind speed distributions from multiple wind farms with considering their correlation. A MCMC sampler is adopted in our scheme to generate the wind speed samples for POPF solving. We novelly integrated the Sobol-based QMC technique into the MCMC sampling process to obtain a faster convergence rate. Two case studies on modified IEEE 14- and 118-bus systems with additional wind farms are conducted to verify the effectiveness of the proposed POPF scheme.

### REFERENCES

[1] S. Wang, X. Zhang, L. Ge, and L. Wu. 2-D wind speed statistical model for reliability assessment of microgrid. *IEEE Transactions on Sustainable Energy*, 7(3):1159–1169, 2016.

[2] Z. Chen, L. Wu, and M. Shahidehpour. Effective load carrying capability evaluation of renewable energy via stochastic long-term hourly based SCUC. *IEEE Transactions on Sustainable Energy*, 6(1):188–197, 2015.

[3] S. Wang, X. Liu, K. Wang, L. Wu, and Y. Zhang. Tracing harmonic contributions of multiple distributed generations in distribution systems with uncertainty. *International Journal of Electrical Power & Energy Systems*, 95:585–591, 2018.

[4] Y. Li, W. Li, W. Yan, J. Yu, and X. Zhao. Probabilistic optimal power flow considering correlations of wind speeds following different distributions. *IEEE Transactions on Power Systems*, 29(4):1847–1854, 2014.

[5] D. Ke, C. Y. Chung, and Y. Sun. A novel probabilistic optimal power flow model with uncertain wind power generation described by customized Gaussian mixture model. *IEEE Transactions on Sustainable Energy*, 7(1):200–212, 2016.

[6] Y. Zhang, S. Shen, and J. L. Mathieu. Distributionally robust chance-constrained optimal power flow with uncertain renewables and uncertain reserves provided by loads. *IEEE Transactions on Power Systems*, 32(2):1378–1388, 2017.

[7] Z. Q. Xie, T. Y. Ji, M. S. Li, and Q. H. Wu. Quasi-Monte Carlo based probabilistic optimal power flow considering the correlation of wind speeds using copula function. *IEEE Transactions on Power Systems*, 33(2):2239–2247, 2018.

[8] A. Kazemdehdash, M. Mohammadi, and A. R. Seifi. The generalized cross-entropy method in probabilistic optimal power flow. *IEEE Transactions on Power Systems*, 2018.
[9] X. Li, Y. Li, and S. Zhang. Analysis of probabilistic optimal power flow taking account of the variation of load power. *IEEE Transactions on Power Systems*, 23(3):992–999, 2008.

[10] A. Schellenberg, W. Rosehart, and J. Aguado. Cumulant-based probabilistic optimal power flow (P-OPF) with Gaussian and gamma distributions. *IEEE Transactions on Power Systems*, 20(2):773–781, 2005.

[11] B. Zou and Q. Xiao. Solving probabilistic optimal power flow problem using quasi Monte Carlo method and ninth-order polynomial normal transformation. *IEEE Transactions on Power Systems*, 29(1):300–306, 2014.

[12] M. Aien, M. Fotuhi-Firuzabad, and M. Rashidinejad. Probabilistic optimal power flow in correlated hybrid wind-photovoltaic power systems. *IEEE Transactions on Smart Grid*, 5(1):130–138, 2014.

[13] G. Verbic and A. Canizares. Probabilistic optimal power flow in electricity markets based on a two-point estimate method. *IEEE Transactions on Power Systems*, 21(4):1183–1893, 2006.

[14] A. Schellenberg, W. Rosehart, and J. A. Guado. Introduction to cumulant-based probabilistic optimal power flow (P-OPF). *IEEE Transactions on Power Systems*, 20(2):1184–1186, 2005.

[15] W. Wei, J. Wang, and L. Wu. Distribution optimal power flow with real-time price elasticity. *IEEE Transactions on Power Systems*, 33(1):1097–1098, 2018.

[16] C. L. Su. Probabilistic load-flow computation using point estimate method. *IEEE Transactions on Power Systems*, 20(4):1843–1851, 2005.

[17] H. Yu, C. Y. Chung, K. P. Wong, H. W. Lee, and J. H. Zhang. Probabilistic load flow evaluation with hybrid Latin hypercube sampling and Cholesky decomposition. *IEEE Transactions on Power Systems*, 24(2):661–667, 2009.

[18] M. Hajian, W. D. Rosehart, and H. Zareipour. Probabilistic power flow by Monte Carlo simulation with Latin supercube sampling. *IEEE Transactions on Power Systems*, 28(2):1550–1559, 2013.

[19] Y. Y. Xu and Z. Yan. Probabilistic load flow calculation with quasi-Monte Carlo and multiple linear regression. *International Journal of Electrical Power & Energy Systems*, 88:1–12, 2017.

[20] Z. Wang, C. Shen, F. Liu, and F. Gao. Analytical expressions for joint distributions in probabilistic load flow. *IEEE Transactions on Power Systems*, 32(3):2473–2474, 2017.

[21] Y. Tian, V. Vittal, G. T. Heydt, and R. Ayyanar. Probabilistic power flow studies for transmission systems with photovoltaic generation using cumulants. *IEEE Transactions on Power Systems*, 27(4):2251–2261, 2012.

[22] T. Williams and C. Crawford. Probabilistic load flow modeling comparing maximum entropy and Gram-Charlier probability density function reconstructions. *IEEE Transactions on Power Systems*, 28(1):272–280, 2013.

[23] R. Singh, B. C. Pal, and R. A. Jabr. Statistical representation of distribution system loads using Gaussian mixture model. *IEEE Transactions on Power Systems*, 25(1):29–37, 2010.

[24] Z. Wang, C. Shen, F. Liu, X. Wu, C. C. Liu, and F. Gao. Chance-constrained economic dispatch with non-Gaussian correlated wind power uncertainty. *IEEE Transactions on Power Systems*, 32(6):4880–4893, 2017.

[25] Y. Tang, K. Divijotham, and S. Low. Real-time optimal power flow. *IEEE Transactions on Smart Grid*, 8(6):2963–2973, 2017.

[26] M. R. Patel. *Wind and solar power systems: design, analysis, and operation*. CRC press, 2005.

[27] L. Gan and H. L. Steven. Optimal power flow in direct current networks. *IEEE Transactions on Power Systems*, 29(6):2892–2904, 2014.

[28] S. M. Ross. *Introduction to Probability Models*. Academic press, 2014.

[29] C. Robert. *Machine Learning: A Probabilistic Perspective*. 2014.

[30] G. Papaefthymiou and B. Klockl. MCMC for wind power simulation. *IEEE Transactions on Energy Conversion*, 23(1):234–240, 2008.