A Data-Driven Nonparametric Approach for Probabilistic Load-Margin Assessment considering Wind Power Penetration

Yijun Xu, *Member, IEEE*, Lamine Mili, *Life Fellow, IEEE*, Mert Korkali, *Senior Member, IEEE*, Kiran Karra, Zongsheng Zheng, Xiao Chen

***Abstract*—A modern power system is characterized by an increasing penetration of wind power, which results in large uncertainties in its states. These uncertainties must be quantified properly; otherwise, the system security may be threatened. Facing this challenge, we propose a cost-effective, data-driven approach to assessing a power system's load margin probabilistically. Using actual wind data, a kernel density estimator is applied to infer the nonparametric wind speed distributions, which are further merged into the framework of a vine copula. The latter enables us to simulate complex multivariate and highly dependent model inputs with a variety of bivariate copulae that precisely represent the tail dependence in the correlated samples. Furthermore, to reduce the prohibitive computational time of traditional Monte-Carlo simulations that process a large amount of samples, we propose to use a nonparametric, Gaussian-process-emulator-based reduced-order model to replace the original complicated continuation power-flow model through a Bayesian-learning framework. To accelerate the convergence rate of this Bayesian algorithm, a truncated polynomial chaos surrogate, which serves as a highly efficient, parametric Bayesian prior, is developed. This emulator allows us to execute the time-consuming continuation power-flow solver at the sampled values with a negligible computational cost. Results of simulations that are performed on several test systems reveal the impressive performance of the proposed method in the probabilistic load-margin assessment.**

***Index Terms*—Probabilistic load margin, data-driven nonparametric model, reduced-order modeling, dependence, vine copula, uncertainty.**

I. Introduction

RECENTLY, uncertainty assessments in power systems have attracted the attention of a number of researchers due to an increasing penetration of renewable generation units. Ignoring these uncertainties will lead to inappropriate planning strategies or control actions, which, in turn, may result in severe system failures. To address this problem, research activities, which include the probabilistic power flow [1], [2], the statistical power system dynamic simulation [3], the stochastic economic dispatch [4], [5], and the probabilistic load-margin formulation [6], [7], have been carried out extensively. Among them, the latter is critical to ensuring the voltage stability of modern power systems with increasing penetration of renewables and, therefore, is chosen as the scope of this paper.

In principle, uncertainty quantification can be conducted under the Monte-Carlo (MC) framework [1], [2], [4], [8], which requires two fundamental steps, i.e., the uncertainty modeling and the uncertainty propagation. The former consists of the marginal distribution modeling for every random input, and the dependent modeling among all the inputs. Traditionally, researchers infer the marginal distributions through some heuristic, parametric distribution functions, such as the Weibull distribution for the wind speed, the beta distribution for the solar irradiance, to cite a few [6], [9]. However, the real-world data may not follow these parametric probability distributions as discussed in [10]. Furthermore, due to geodistance closeness, the dependence among the renewables cannot be ignored. Although several methods (e.g., a whitening transformation [6], [11] and the Gaussian copula [12]) have been proposed in the literature, none of them is able to simulate asymmetric, complex multivariate dependent structures. The Gaussian copula, while versatile, is unable to model tail dependence or nonlinear dependence structures [13]. To overcome these modeling limitations, the vine-copula technique has been explored in several power-system applications, including generating system states for machine learning [13], probabilistic forecast of wind farm generation [10], clustering of residential loads [14], and peak-load estimation [15]. However, these studies do not systematically merge the vine copula into an uncertainty-quantification framework and its corresponding applications, which is one of the contributions of this paper.

Apart from these difficulties in the uncertainty modeling stage, the computational challenge in the uncertainty propagation stage is also well-known [8]. As an example, the continuation power-flow (CPF) algorithm is a typical continuation method based on multiple prediction and correction steps [16]. Although the straightforward MC method based on the evaluations of tens of thousands of samples in the CPF model has been adopted in the published literature

Y. Xu and L. Mili and Z. Zheng are with the Bradley Department of Electrical and Computer Engineering, Virginia Tech, Northern Virginia Center, Falls Church, VA 22043 USA (e-mail:{yijunxu,lmili,zongsheng}@vt.edu).

M. Korkali is with the Computational Engineering Division, Lawrence Livermore National Laboratory, Livermore, CA 94550 USA (e-mail: korkali1@llnl.gov).

X. Chen is with the Center for Applied Scientific Computing, Lawrence Livermore National Laboratory, Livermore, CA 94550 USA (e-mail: chen73@llnl.gov).

K. Karra is with the Applied Physics Lab at Johns Hopkins University Baltimore, MD 21218 USA (e-mail: kiran.karra@gmail.com).

This work was supported, in part, by the U.S. National Science Foundation under Grant 1917308 and by the United States Department of Energy Office of Electricity Advanced Grid Modeling Program, and performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. Document released as LLNL-JRNL-799106.

(see, e.g., [17], [18]), the computational burden is known to be extremely heavy. This is true even for relatively small systems [6]. Therefore, many strategies have been proposed to overcome such a difficulty. Zhao *et al.* [19] propose to use a Latin-hypercube-sampling-based approach to reduce the samples size in the sampling procedure while sacrificing some computing accuracy. Rodrigues *et al.* [17] adopt a dc power-flow-based model to simplify the system model. To overcome the complexity brought by the nonlinearity of the model, Zhang *et al.* [20] propose an analytical method to reduce the computational burden by directly assuming that all the uncertain inputs follow a Gaussian distribution. Haesen *et al.* [6] and Xu *et al.* [7] use some heuristic, parametric distributions to approximate the load margin, but their methods suffer from the *curse of dimensionality* in the high-dimensional case.

To overcome the aforementioned shortcomings, this paper proposes, for the first time, to utilize a data-driven, nonparametric approach based on a novel Gaussian process emulator (GPE) associated with the vine copula to solve the probabilistic CPF problem. The contributions of the paper include the following:

- Starting from the real-world wind data and avoiding the inaccuracy brought by the parametric marginal functions, we propose to adopt a kernel density estimator to obtain more accurate nonparametric input probability density functions (pdfs), which is further merged into the vine copula framework.
- To simulate the high-dimensional dependent samples that represent the uncertainties from the renewables, we adopt a novel vine-copula technique [13], for the first time, in the uncertainty quantification application. This technique outperforms the traditional Gaussian copula for its capability in modeling the high-dimensional, asymmetric, dependent multivariate with a variety of bivariate copulas such as the Frank and the Gumbel copula [7], [21].
- To improve the computing efficiency of the uncertainty-propagation procedure, from Bayesian inference point of view, we propose to use a GPE as a nonparametric, reduced-order model representation for the nonlinear CPF model [22]. This emulator allows us to evaluate the time-consuming CPF solver at the sampled values with a negligible computational cost.
- To further accelerate the convergence rate of the developed Bayesian algorithm, we propose, for the first time, to merge the truncated polynomial chaos expansion (PCE) into the GPE framework by taking the PCE as a highly efficient Bayesian prior in the Bayesian-inference procedure.

The performance of our proposed method has been assessed and analyzed through simulations that are carried out on multiple IEEE standard test systems. These simulations reveal that our method can accurately estimate the probability density function (pdf) of the load margin with two-order-of-magnitude improvement in computing speed compared to the traditional MC method.

This paper is organized as follows: in Section II, problem formulation is presented. In Section III, vine copula is illustrated. Section IV presents the GPE-based reduced-order modeling. Section V summarizes the proposed nonparametric method. Case studies are presented in Section VI, followed by the conclusions and future work in Section VII.

II. Problem Formulation

This section formulates the probabilistic load-margin assessment problem. Let us first formulate the power system forward model as

$$y = f(\boldsymbol{x}). \tag{1}$$

Here, y stands for the quantity of interest (QoI), which, in our case, is the load margin of a bus; $\boldsymbol{x} = [x_1, x_2, \ldots, x_p]$ is a vector of uncertain model parameters described by some probability distribution functions with finite variances. In our work, we consider the wind power generation as the random inputs following some non-heuristic probability distributions; the $f(\cdot)$ is the nonlinear function that represents the continuation power-flow model, which maps the model parameters, $\boldsymbol{x}$, to the QoI, y. To solve this CPF model, we first start from a power-flow solution at initial load values, then we perform multiple prediction and correction steps for the increased load levels as shown in Fig. 1. Using a small step size, the prediction step estimates the voltage magnitude through the tangent vector from the previous solution point. Next, the correction step further fine-tunes the predicted voltage magnitude at a fixed tangent vector. Note that the constraints on the bus voltage as well as on the transmission lines and generator capacity can also be considered. The detailed implementation step has been described by Ajjarapu [16].

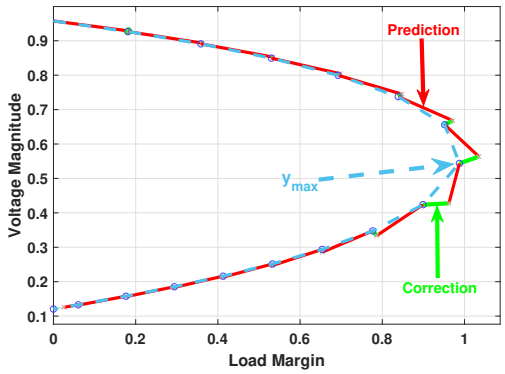


Fig. 1. Continuation power flow with multiple prediction and correction steps.

To obtain the probabilistic description of the load margin y under these uncertain model parameters, a typical MC method draws a large number of N_{sample} samples, $\{\boldsymbol{x}^{(i)}\}_{i=1}^{N_{\text{sample}}}$, that not only reflect the pdfs of the input parameters but also the correlation between them. Then, for each $\boldsymbol{x}^{(i)}$, $i = 1, \ldots, N_{\text{sample}}$,

$$y^{(i)} = f(\boldsymbol{x}^{(i)}) \tag{2}$$

is solved to obtain N_{sample} load-margin solutions, $\{y^{(i)}\}_{i=1}^{N_{\text{sample}}}$, from which the pdf of the load margin is determined. The CPF method is typically employed in power systems despite the fact that even a single evaluation of $f(\cdot)$ will involve multiple prediction and correction steps to obtain the load margin,

y, and, hence, is admittedly a complicated, time-consuming solver—not to mention that N_{sample} is typically required to be a large number in the MC sampling to ensure good computing accuracy. Therefore, the goal of this paper is to greatly reduce the computing time of this method while precisely modeling the dependent data structures of the input samples.

III. Uncertainty Modeling

In this section, we present the way to generate dependent high-dimensional samples as model inputs via vine copula.

1) Copula: Recently, the copulas have been proven to be successful in many industrial and financial applications for modeling the dependency between random inputs [13], [23]. According to Sklar's theorem, any joint multivariate cumulative distribution function $F_{\mathbf{X}}$ of a p-dimensional random vector can be expressed in terms of its marginal distributions and a copula to represent their dependence. Formally, we have

$$F_{\mathbf{X}}(\mathbf{x}) = C(F_{X_1}(x_1), F_{X_2}(x_2), \ldots, F_{X_p}(x_p)). \tag{3}$$

Here, $F_{X_i}(x_i)$ is the ith input marginal and $C(\cdot)$ is a copula that describes the dependence structure between the p-dimensional input variables [24]. Accordingly, its joint multivariate density function, $f_{\mathbf{X}}$, can be obtained via

$$f_{\mathbf{X}}(\mathbf{x}) = c(F_{X_1}(x_1), \ldots, F_{X_p}(x_p)) \prod_{i=1}^{p} f_i(x_i). \tag{4}$$

Here, c is the p-variate copula density and $f_i(x_i)$ is the marginal density for ith variable. Since there exist different copula families, the choice of the copula function will influence the accuracy of the dependence modeling. The Gaussian copula is advantageous in certain applications thanks to its ability to generate high-dimensional correlated samples [7], but it also lacks the ability to generate asymmetric, tail-dependent samples. On the other hand, the Archimedean copulas are more useful in scenarios which require nonlinear tail-dependence modeling. However, they are generally not scalable due to being limited to the bivariate case [10], [25]. To overcome these shortcomings, we resort to vine copula next.

2) Vine Copula: Being a powerful tool in simulating high-dimensional correlated samples with various types of tail-dependence structures involved, vine copula is known for its capability of decomposing a multivariate density function into a cascade of bivariate pair copulas [25], [26]. Starting from the factorization on the joint density function, we get

$$\begin{aligned} f_{\mathbf{X}}(\mathbf{x}) =& f_p(x_p) \cdot f_{p-1|p}(x_{p-1}|x_p) \cdot f_{p-2|p-1,p}(x_{p-2}|x_{p-1}, x_p) \\ & \cdots f_{1|2,\ldots,p}(x_1|x_2, \ldots, x_p) \\ =& \prod_{i=1}^{p} f_{i|i+1,\ldots,p}(x_i|x_{i+1}, \ldots, x_p). \end{aligned} \tag{5}$$

Based on the property that each term in (5) can be decomposed into the appropriate pair-copula times a conditional marginal density via

$$f(x|\boldsymbol{v}) = c_{x,v_j|\boldsymbol{v}_{-j}}\{F(x|\boldsymbol{v}_{-j}), F(v_j|\boldsymbol{v}_{-j})\} f(x|\boldsymbol{v}_{-j}), \tag{6}$$

for a d-dimensional vector $\boldsymbol{v}$. Here, v_j is one arbitrarily chosen component of $\boldsymbol{v}$ and $\boldsymbol{v}_{-j}$ denotes the vector of $\boldsymbol{v}$, excluding this component [27]. By using (6) iteratively in (5), as Mai has presented in [25], the joint multivariate pdf in (5) can be further transformed into the product of only bivariate copulas and one-dimensional pdf, e.g.,

$$\begin{aligned} & f_{2|1}(x_2|x_1) = c_{2|1}(F_2, F_1) \cdot f_2(x_2) \\ & f_{3|1,2}(x_3|x_1, x_2) = c_{3,2|1}(F_{3|1}, F_{2|1}) \cdot c_{3,1}(F_3, F_1) \cdot f_3(x_3) \\ & f_{4|1,2,3}(x_4|x_1, x_2, x_3) = c_{4,2|1,3} \cdot c_{4,1|3} \cdot c_{4|3} \cdot f_4(x_4), \end{aligned} \tag{7}$$

and so on for the higher-dimensional cases.

3) C-Vine and D-Vine: It is worth pointing out that the order of pairwise conditioning on (5) is not unique [25], [27]. Thus, we need a systematic way to decompose it and to provide a unique solution. Two popular choices are the canonical vine (C-vine) and the drawable vine (D-vine). Both of them make use of a graphical tool to facilitate their decomposition into a cascade of copula density functions forming $p-1$ trees. For the C-vine copula, a p-dimensional joint pdf is decomposed as

$$f_{\mathbf{X}}(x_1, \ldots, x_p) = \prod_{i=1}^{p} f_i(x_i) \prod_{j=1}^{p-1} \prod_{i=1}^{p-j} c_{j,j+i|1,\ldots,j-1}. \tag{8}$$

Similarly, $f_{\mathbf{X}}$ is decomposed via the D-vine copula as

$$f_{\mathbf{X}}(x_1, \ldots, x_p) = \prod_{i=1}^{p} f_i(x_i) \prod_{j=1}^{p-1} \prod_{i=1}^{p-j} c_{i,j+i|i+1,\ldots,i+j-1}. \tag{9}$$

Here, $c_{j,j+i|1,\ldots,j-1}$ is short for $c_{j,j+i|1,\ldots,j-1}(F(x_j|x_1,\ldots,x_{j-1}), F(x_{i+j}|x_1,\ldots,x_{j-1}))$ and $c_{i,j+i|i+1,\ldots,i+j-1}$ is short for $c_{i,j+i|i+1,\ldots,i+j-1}(F(x_i|x_{i+1},\ldots,x_{i+j-1}), F(x_{i+j}|x_{i+1,\ldots,x_{i+j-1}}))$. Let us take a 4-dimensional case as an example, it is clear that $f_{\mathbf{X}}(x_1, x_2, x_3, x_4)$ can be decomposed via the C-vine as

$$\begin{aligned} & f_{\mathbf{X}}(x_1, x_2, x_3, x_4) = f_1(x_1) \cdot f_2(x_2) \cdot f_3(x_3) \cdot f_4(x_4) \\ & \cdot c_{1,2}(F_1(x_1), F_2(x_2)) \cdot c_{1,3}(F_1(x_1), F_3(x_3)) \cdot c_{1,4}(F_1(x_1), F_4(x_4)) \\ & \cdot c_{2,3|1}(F_{2|1}(x_2|x_1), F_{3|1}(x_3|x_1)) \cdot c_{2,4|1}(F_{2|1}(x_2|x_1), F_{4|1}(x_4|x_1)) \\ & \cdot c_{3,4|1,2}(F_{3|1,2}(x_3|x_1, x_2), F_{4|1,2}(x_4|x_1, x_2)), \end{aligned} \tag{10}$$

or via the D-vine as

$$\begin{aligned} & f_{\mathbf{X}}(x_1, x_2, x_3, x_4) = f_1(x_1) \cdot f_2(x_2) \cdot f_3(x_3) \cdot f_4(x_4) \\ & \cdot c_{1,2}(F_1(x_1), F_2(x_2)) \cdot c_{2,3}(F_2(x_2), F_3(x_3)) \cdot c_{3,4}(F_3(x_3), F_4(x_4)) \\ & \cdot c_{1,3|2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)) \cdot c_{2,4|3}(F_{2|3}(x_2|x_3), F_{4|3}(x_4|x_3)) \\ & \cdot c_{1,4|2,3}(F_{1|2,3}(x_1|x_2, x_3), F_{4|2,3}(x_4|x_2, x_3)). \end{aligned} \tag{11}$$

4) Sampling Strategy: As (6) shows that the evaluation of the density of the vine copula involves the evaluation of the conditional distributions. For every j in (6), the corresponding conditional cumulative distribution function (cdf) has been shown by Joe [26] to be given by

$$F(x|\boldsymbol{v}) = \frac{\partial C_{x,v_j|\boldsymbol{v}_{-j}}\{F(x|\boldsymbol{v}_{-j}), F(v_j|\boldsymbol{v}_{-j})\}}{\partial F(v_j|\boldsymbol{v}_{-j})}. \tag{12}$$

Following Aas in [27], let us introduce an h function, denoted as $h(u_1; u_2, \Theta)$, to represent the above conditional cdf in the bivariate case when $X_1 = U_1$ and $X_2 = U_2$ are uniform random variables. Then, we have

$$h(u_1; u_2, \Theta) = F(u_1|u_2) = \frac{\partial C_{u_1,u_2}(u_1; u_2, \Theta)}{\partial u_2}. \tag{13}$$

Here, the u_2 denotes the conditioning variable and Θ denotes the set of parameters for the bivariate copula C_{u_1,u_2}. Using

this conditional cdf, the p-dimensional dependent, uniformly distributed variables can be sampled iteratively via

$$\begin{aligned} U_1 &= W_1, \\ U_2 &= F_{2|1}^{-1}(W_2|U_1), \\ U_3 &= F_{3|2,1}^{-1}(W_3|U_1,U_2), \\ \cdots &= \cdots, \\ U_p &= F_{p|p-1,\ldots,1}^{-1}(W_p|U_1,\ldots,U_{p-1}), \end{aligned} \tag{14}$$

where samples $W_i \overset{i.i.d.}{\sim} U[0,1], i = 1,\ldots,p$. Similarly to the relationship between conditional distribution F and $h(u_1;u_2,\Theta)$, the inverse conditional distribution F^{-1} is represented by an inverse h function, $h^{-1}(u_1;u_2,\Theta)$ for the bivariate case. More details on $h^{-1}(u_1;u_2,\Theta)$ for the bivariate copulas have been provided in [25, ch. 5].

Remark 1. *It is worth noting that the random variables $\{X_1,X_2,\ldots,X_p\}$ are not guaranteed to follow a uniform probability distribution in practice. To circumvent this difficulty, the transformation between $\{X_1,X_2,\ldots,X_p\}$ and $\{U_1,U_2,\ldots,U_p\}$ can be fulfilled via an inverse cdf mapping as*

$$X_1 = F_1^{-1}(U_1), X_2 = F_2^{-1}(U_2), \cdots, X_p = F_p^{-1}(U_p). \tag{15}$$

This inverse cdf mapping enables the vine copula to be applicable to all the closed-form probability distributions [13], [25], [27].

Using this vine-copula technique, we are able to generate the correlated samples that reflect the precise dependent structures of the model inputs such as loads and renewables.

IV. Reduced-Order Modeling

In this section, we present a nonparametric, reduced-order modeling technique using GPE.

A. Problem Description

Let us first formulate the probabilistic load-margin assessment problem in the GPE framework. Here, the CPF model is denoted by the aforementioned $f(\cdot)$. Its corresponding vector-valued random input of p dimensions is denoted as $\mathbf{x}$, which accounts for the uncertainties from the variations of the wind generation. Due to the randomness of $\mathbf{x}$, we may observe n samples as a finite collection of the model input as $\{\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_n\}$. Accordingly, its evaluated model output $f(\mathbf{x})$, i.e., load margin, also becomes random, and has its corresponding n realizations denoted by $\{f(\mathbf{x}_1),f(\mathbf{x}_2),\ldots,f(\mathbf{x}_n)\}$. If we assume that the model output is a realization of a Gaussian process, then the finite collection, $\{f(\mathbf{x}_1),f(\mathbf{x}_2),\ldots,f(\mathbf{x}_n)\}$, of the random variables, $f(\mathbf{x})$, will follow a joint multivariate normal probability distribution as

$$\begin{bmatrix} f(\mathbf{x}_1) \\ \vdots \\ f(\mathbf{x}_n) \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} m(\mathbf{x}_1) \\ \vdots \\ m(\mathbf{x}_n) \end{bmatrix}, \begin{bmatrix} k(\mathbf{x}_1,\mathbf{x}_1) & \cdots & k(\mathbf{x}_1,\mathbf{x}_n) \\ \vdots & \ddots & \vdots \\ k(\mathbf{x}_n,\mathbf{x}_1) & \cdots & k(\mathbf{x}_n,\mathbf{x}_n) \end{bmatrix} \right). \tag{16}$$

Here, let us denote $\boldsymbol{m}(\cdot)$ as the mean function and $\boldsymbol{k}(\cdot,\cdot)$ as a kernel function that represents the covariance function. Then, (16) can be simplified as

$$\boldsymbol{f}(\mathbf{X})\,|\mathbf{X} \sim \mathcal{N}\left(\boldsymbol{m}(\mathbf{X}), \boldsymbol{k}(\mathbf{X},\mathbf{X})\right), \tag{17}$$

where $\mathbf{X}$ is an $n \times p$ matrix, denoted by $[\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_n]^\intercal$; $\boldsymbol{f}(\mathbf{X})$ stands for $[f(\mathbf{x}_1),f(\mathbf{x}_2),\ldots,f(\mathbf{x}_n)]^\intercal$; and $\boldsymbol{m}(\mathbf{X})$ represents $[m(\mathbf{x}_1),m(\mathbf{x}_2),\ldots,m(\mathbf{x}_n)]^\intercal$.

Now, if an independent and identically distributed (i.i.d.) Gaussian noise $\boldsymbol{\varepsilon} \sim \mathcal{N}(0,\sigma^2\mathbf{I}_n)$ (where $\mathbf{I}_n$ and σ^2 are an n-dimensional identity matrix and the variance, respectively) is considered in the system output, $\boldsymbol{f}(\mathbf{X})$, the observations, $\mathbf{Y}$, will be expressed as

$$\mathbf{Y}|\mathbf{X} \sim \mathcal{N}\left(\boldsymbol{m}(\mathbf{X}), \boldsymbol{k}(\mathbf{X},\mathbf{X}) + \sigma^2\mathbf{I}_n\right). \tag{18}$$

Note that $\boldsymbol{\varepsilon}$ is also called a "nugget". If $\sigma^2 = 0$, then $f(x)$ is observed without noise. However, in practical implementation, the nugget is suggested to be added for numerical stability [28].

B. Bayesian Inference

Here, we present the way to use the aforementioned finite collection of n samples, $(\mathbf{Y},\mathbf{X})$, to infer the unknown system output, $\mathbf{y}(\mathbf{x})$, on the sample space of $\mathbf{x} \in \mathbb{R}^p$ in a Bayesian-inference framework.

It is well-known that a Bayesian posterior distribution of the unknown system output can be inferred from a Bayesian prior distribution of $\mathbf{y}(\mathbf{x})$ and the likelihoods obtained from the observations. Let us first assume a Bayesian prior distribution of $\mathbf{y}(\mathbf{x})|\mathbf{x}$, expressed as

$$\mathbf{y}(\mathbf{x})|\mathbf{x} \sim \mathcal{N}\left(\boldsymbol{m}(\mathbf{x}), \boldsymbol{k}(\mathbf{x},\mathbf{x}) + \sigma^2\mathbf{I}_{n_x}\right). \tag{19}$$

Combined with the observations provided by the finite collection of the samples $\{\mathbf{Y},\mathbf{X}\}$, we can formulate the joint distribution of $\mathbf{Y}$ and $\mathbf{y}(\mathbf{x})|\mathbf{x}$ as

$$\begin{bmatrix} \mathbf{Y} \\ \mathbf{y}(\mathbf{x})|\mathbf{x} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{m}(\mathbf{X}) \\ \boldsymbol{m}(\mathbf{x}) \end{bmatrix}, \begin{bmatrix} \mathbf{K}_{11} & \mathbf{K}_{12} \\ \mathbf{K}_{21} & \mathbf{K}_{22} \end{bmatrix} \right), \tag{20}$$

where $\mathbf{K}_{11} = \boldsymbol{k}(\mathbf{X},\mathbf{X}) + \sigma^2\mathbf{I}_n$; $\mathbf{K}_{12} = \boldsymbol{k}(\mathbf{X},\mathbf{x})$; $\mathbf{K}_{21} = \boldsymbol{k}(\mathbf{x},\mathbf{X})$; and $\mathbf{K}_{22} = \boldsymbol{k}(\mathbf{x},\mathbf{x}) + \sigma^2\mathbf{I}_{n_x}$.

Now, using the rules of the conditional Gaussian distribution [29], we can infer the Bayesian posterior distribution of the system output $\mathbf{y}(\mathbf{x})$ conditioned upon the observations $(\mathbf{Y},\mathbf{X})$. It follows a Gaussian distribution given by

$$\mathbf{y}(\mathbf{x})|\mathbf{x},\mathbf{Y},\mathbf{X} \sim \mathcal{N}\left(\boldsymbol{\mu}(\mathbf{x}), \boldsymbol{\Sigma}(\mathbf{x})\right), \tag{21}$$

where

$$\boldsymbol{\mu}(\mathbf{x}) = \boldsymbol{m}(\mathbf{x}) + \mathbf{K}_{21}\mathbf{K}_{11}^{-1}(\mathbf{Y} - \boldsymbol{m}(\mathbf{X})), \tag{22}$$

$$\boldsymbol{\Sigma}(\mathbf{x}) = \mathbf{K}_{22} - \mathbf{K}_{21}\mathbf{K}_{11}^{-1}\mathbf{K}_{12}. \tag{23}$$

To this point, the general form of the GPE has been derived. Here, (22) serves as a reduced-order model (a.k.a. the response surface or surrogate model) to very closely capture the behavior of the nonlinear CPF model while being computationally inexpensive to evaluate.

C. Mean and Covariance Functions

Let us describe the mean function $\boldsymbol{m}(\cdot)$ and the covariance function represented via the kernel $\boldsymbol{k}(\cdot,\cdot)$ that characterizes the GPE. The mean function models the prior belief about the existence of a systematic trend expressed as

$$\boldsymbol{m}(\mathbf{x},\boldsymbol{\beta}) = \mathbf{H}(\mathbf{x})\boldsymbol{\beta}. \tag{24}$$

Here, $\mathbf{H}(\mathbf{x})$ can be any set of basis functions. For example, let $\mathbf{x}_i = [x_{i1},\ldots,x_{ip}]$ be the ith sample, where $i = 1,2,\ldots,n$, wherein x_{ik} represents its kth element, where $k = 1,2,\ldots,p$. For instance, $\mathbf{H}(\mathbf{x}_i) = 1$ is a constant basis; $\mathbf{H}(\mathbf{x}_i) = [1, x_{i1},\ldots,x_{ip}]$ is a linear basis and $\boldsymbol{\beta}$ is a vector of hyperparameters.

Here, the covariance function is represented by a kernel function expressed as

$$k\left(\mathbf{x}_i,\mathbf{x}_j|\boldsymbol{\theta}\right) = \mathrm{Cov}(\mathbf{x}_i,\mathbf{x}_j|\boldsymbol{\theta}). \tag{25}$$

The parameters of a kernel function are defined as follows: τ and ℓ_k are the hyperparameters defined in the positive real line; σ^2 and ℓ_k correspond to the order of the magnitude and the speed of variation in the kth input dimension, respectively. Popular choices for the kernel functions are listed as

$$\text{Square Exponential}: k_{\text{SE}} = \tau^2 \exp\left(-\sum_{k=1}^{p}\frac{r_k^2}{2\ell_k^2}\right), \tag{26a}$$

$$\text{Exponential}: k_{\text{E}} = \tau^2 \exp\left(-\sum_{k=1}^{p}\frac{|r_k|}{\ell_k}\right), \tag{26b}$$

$$\text{Rational Quadratic}: k_{\text{RQ}} = \tau^2\left(1+\sum_{k=1}^{p}\frac{r_k^2}{2\alpha\ell_k^2}\right)^{-\alpha}, \tag{26c}$$

$$\text{Martin 3/2}: k_{3/2} = \tau^2\left(1+\sum_{k=1}^{p}\frac{\sqrt{3}r_k}{\ell_k}\right)\exp\left(-\sum_{k=1}^{p}\frac{\sqrt{3}r_k}{\ell_k}\right), \tag{26d}$$

where $r_k = |x_{ik} - x_{jk}|$. Until now, the model structure of the GPE has been fully defined. For simplicity, let $\boldsymbol{\theta} = [\tau,\ell_1,\ldots,\ell_p]$ contain the hyperparameters of the covariance function. Then, we write $\boldsymbol{\eta} = (\sigma^2,\boldsymbol{\beta},\boldsymbol{\theta})$ to represent all the hyperparameters in the GPE model.

D. Parametric Surrogate versus Nonparametric Surrogate

Recently, the generalized polynomial chaos expansion (PCE) has also been advocated in literature as a parametric response surface [2], [11], [12]. By using it, the stochastic output, y, is represented as a weighted sum of a given set of orthogonal polynomial chaos basis functions constructed from the probability distribution of the input random variables via

$$y = \sum_{i=0}^{N_m} a_i\Phi_i(\boldsymbol{x}). \tag{27}$$

Here, $\Phi_i(x_1,x_2,\ldots,x_p)$ denote the corresponding polynomial chaos bases; a_i denotes the ith polynomial chaos coefficient; N_m+1 is the number of the PCE bases under the maximum truncated order, t, of the polynomial chaos basis functions, where $N_m = (p+t)!/(p!t!) - 1$. For more details about the PCE, the reader is referred to [11].

Here, it is worth pointing out that the parametric PCE model can be perfectly merged into the nonparametric GPE model under the Bayesian-inference framework. This can be fulfilled simply by using PCE as the GPE prior's systematic trend, i.e., the mean function. Then, for n samples, the $\mathbf{H}$ matrix in (24) is directly equal to

$$\mathbf{H} = \begin{bmatrix} \Phi_0(\boldsymbol{x_1}) & \Phi_1(\boldsymbol{x_1}) & \ldots & \Phi_{N_P}(\boldsymbol{x_1}) \\ \Phi_0(\boldsymbol{x_2}) & \Phi_1(\boldsymbol{x_2}) & \ldots & \Phi_{N_P}(\boldsymbol{x_2}) \\ \vdots & \vdots & \ddots & \vdots \\ \Phi_0(\boldsymbol{x_n}) & \Phi_1(\boldsymbol{x_n}) & \ldots & \Phi_{N_P}(\boldsymbol{x_n}) \end{bmatrix}. \tag{28}$$

All the other settings in the GPE remain unchanged. This operation makes sense because it is well-known that a good Bayesian prior facilitates a success of a Bayesian inference. With a good Bayesian prior, the Bayesian posterior can converge fast and, therefore, be inferred more effectively. In this view, the widely recognized PCE serves as an excellent candidate for the Bayesian prior. This enables the GPE to enjoy the efficiency brought the PCE surrogate and the accuracy brought by the nonparametric kernel functions at the same time.

Remark 2. *It is well-known that the PCE suffers from the "curse of dimensionality" under a high-dimensional case. Then, a proper truncation strategy needs to be applied. Here, it is suggested to ignore the coupling effect as proposed in [30]. This truncation strategy can reduce the number of the PCE terms from a combinatorial number to a linear number with respect to the dimension. Then, the accuracy of the surrogate model will be further improved by fine-tuning the nonparametric kernel functions in the Bayesian posteriors as (22) indicates.*

V. The Proposed Method

Let us illustrate the steps for conducting the probabilistic load-margin assessment using the GPE.

A. Wind-Speed Inference

First, let us infer the marginal density functions of the wind speed from the historical data. Unlike the parametric probability distribution, such as the Weibull or the lognormal probability distribution, etc., that have been used in the wind speed modeling [9], [11], this article would like to select a more general, nonparametric inference based on a kernel density estimator [10], [31]. This estimator can infer the closed-form univariate probability density function directly from data via

$$\hat{f}_X(x) = \frac{1}{Nh}\sum_{i=1}^{N} K\left(\frac{x-x_i}{h}\right). \tag{29}$$

Here, N is the number of the historical samples. h is the bandwidth parameter, typically obtained via $h = \left(4\hat{\sigma}^5/3N\right)^{1/5} \approx 1.06\hat{\sigma}N^{-1/5}$, where $\hat{\sigma}$ is the estimated sample standard deviation and K is a nonnegative kernel function, which is chosen as the standard Gaussian kernel in this article. Note that the random input x is specifically referred to the wind speed data in this article. It is also worth pointing out that the real-world wind data may contain some outliers that may require a few preprocessing steps [32]. However, the detection and the detailed analysis of the outliers are outside the scope of this article.

B. Vine-Copula Inference

Now, we need to estimate the vine-copula structure from the data. This procedure involves the selection of a vine, the selection of bivariate copula families for each edge and the estimation of their corresponding parameters.

For the selection of a vine, we would like to select the D-vine as Becker has advocated in [23] to model the wind. As for the bivariate families, the selection will be made among the following: the Gaussian copula, the Student's t-copula, the Clayton copula, the Gumbel copula, and the Frank copula as well as the rotated version of the Archimedean copulas.

Here, the parameters, Θ, for each copula family can be estimated through a sequential estimation advocated in the literature [10], [23], [27]. Proposed by Aas *et al.* [27], this sequential estimation is very straightforward for a high-dimensional case. Once the copula families are selected, the likelihood of each copula pair is maximized sequentially. Starting from the first tree of a vine using historical data, the estimation is conducted sequentially to the higher-order trees that involve conditional distributions using the generated samples obtained via (14). Note that the estimation for each copula family is very easy to perform since the dimension is only 2. More details have been presented in [27].

C. Training-Sample Generation

In order to acquire the GPE-based surrogate described in (22), we need to obtain the observation sets contained in $(\mathbf{Y}, \mathbf{X})$. To obtain the system realization, $\mathbf{Y}$, we must generate n samples, $\mathbf{X}$. As a popular computer design tool, Latin hypercube sampling (LHS) is chosen to generate $i.i.d.$ uniformly distributed samples [33]. Then, these $i.i.d.$ uniformly distributed samples can be mapped to follow the target marginal distributions via the inverse cdf mapping as described in (15). Note that the closed-form expressions for the target marginals given by (15) are obtained through the abovementioned kernel density estimator. Then, we use the aforementioned vine copula to further transform these $i.i.d.$ samples into the dependent ones to improve the training performances. Note that the wind speed cannot be directly evaluated in the CPF model, following literature [6], [34], [35], the wind speed, x, is mapped into the wind power, P_w, through a piecewise relation described as

$$P_w(x) = \begin{cases} 0 & x \le v_{\text{in}} \text{ or } x > v_{\text{out}} \ , \\ \frac{x - v_{\text{in}}}{v_{\text{rated}} - v_{\text{in}}} \cdot P_r & v_{\text{in}} < x < v_{\text{rated}} \ , \\ P_r & v_{\text{rated}} < x < v_{\text{out}} \ . \end{cases} \tag{30}$$

Here, v_{in}, v_{out}, and v_{rated} are the cut-in, cut-out, and rated wind speed, respectively; and P_r represents the rated wind power. After mapping the samples of wind speed, $\mathbf{X}$, into the samples of the wind power, the latter are run through the full CPF model to obtain a small number of observations, $\mathbf{Y}$, that will be used next.

D. GPE Construction

With $(\mathbf{Y}, \mathbf{X})$, we can estimate the hyperparameters $\boldsymbol{\eta}$ in the GPE. Following Gelman *et al.* [36], we choose to adopt the Gaussian maximum likelihood estimator since it is the most efficient estimator under a Gaussian distribution, which is followed by the calculated residuals, and it is easy to compute. First, to indicate the hyperparameters, let us rewrite (18) as

$$\mathbf{Y}|\mathbf{X}, \boldsymbol{\eta} \sim \mathcal{N}\left(\boldsymbol{m}\left(\mathbf{X}\right), \boldsymbol{k}\left(\mathbf{X}, \mathbf{X}\right) + \sigma^2 \mathbf{I}_n\right). \tag{31}$$

Then, using the Gaussian maximum likelihood estimator, we obtain

$$\widehat{\boldsymbol{\eta}} = \left(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\theta}}, \widehat{\sigma}^2\right) = \underset{\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2}{\arg\max} \log P\left(\mathbf{Y}|\mathbf{X}, \boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2\right). \tag{32}$$

Using (24)–(31), the marginal log-likelihood can be expressed as

$$\begin{aligned} &\log P\left(\mathbf{Y}|\mathbf{X}, \boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2\right) \\ &= -\frac{1}{2}(\mathbf{Y} - \mathbf{H}\boldsymbol{\beta})^{\intercal}\left[\boldsymbol{k}(\mathbf{X}, \mathbf{X}|\boldsymbol{\theta}) + \sigma^2 \mathbf{I}_n\right]^{-1}(\mathbf{Y} - \mathbf{H}\boldsymbol{\beta}) \\ &\quad - \frac{n}{2}\log 2\pi - \frac{1}{2}\log\left|\boldsymbol{k}(\mathbf{X}, \mathbf{X}|\boldsymbol{\theta}) + \sigma^2 \mathbf{I}_n\right|, \end{aligned} \tag{33}$$

which implies that the Gaussian maximum likelihood estimator of $\boldsymbol{\beta}$ conditioned on $\boldsymbol{\theta}$ and σ^2 is a weighted least-squares estimator given by

$$\hat{\boldsymbol{\beta}}\left(\boldsymbol{\theta}, \sigma^2\right) = \left[\mathbf{H}^{\intercal}\left[\boldsymbol{k}(\mathbf{X}, \mathbf{X}|\boldsymbol{\theta}) + \sigma^2 \mathbf{I}_n\right]^{-1}\mathbf{H}\right]^{-1}\mathbf{H}^{\intercal}\left[\boldsymbol{k}(\mathbf{X}, \mathbf{X}|\boldsymbol{\theta}) + \sigma^2 \mathbf{I}_n\right]^{-1}\mathbf{Y}. \tag{34}$$

Since $\widehat{\boldsymbol{\beta}}$ is a function of $\left(\widehat{\boldsymbol{\theta}}, \widehat{\sigma}^2\right)$, let us insert (34) into (33) to reduce the number of the hyperparameters. Then, (32) is further simplified as

$$\left(\widehat{\boldsymbol{\theta}}, \widehat{\sigma}^2\right) = \underset{\boldsymbol{\theta}, \sigma^2}{\arg\max} \log P\left(\mathbf{Y}|\mathbf{X}, \hat{\boldsymbol{\beta}}\left(\boldsymbol{\theta}, \sigma^2\right), \boldsymbol{\theta}, \sigma^2\right). \tag{35}$$

Now, we only need to estimate the hyperparameters $\left(\widehat{\boldsymbol{\theta}}, \widehat{\sigma}^2\right)$ instead of $\left(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\theta}}, \widehat{\sigma}^2\right)$. Then, we utilize a gradient-based optimizer to solve this optimization as described in [37]. Once $\widehat{\boldsymbol{\eta}}$ is obtained, the GPE model is fully constructed. More details can be found in [36].

Remark 3. *Note that since the hyperparameters $\left(\widehat{\boldsymbol{\theta}}, \widehat{\sigma}^2\right)$ are estimated thorough a simple and straightforward gradient-based optimization method under the Gaussian assumption, the global optimal solution is not guaranteed here. Although a full Bayesian posterior can be approximated via some heuristic method (e.g., MCMC), the computational burden will increase accordingly. Fortunately, Kennedy and O'Hagan [22] have shown that the accuracy of the GPE method is not sensitive to these hyperparameters and, therefore, name them "roughness parameters". Thus, the abovementioned two-stage maximum likelihood estimation has been widely used in the literature for the Gaussian-process regression [22], [36], [37].*

E. Sample Evaluation

Now, we can execute an MC sampling procedure to generate a large amount of samples and transform them into the dependent ones through the inferred vine-copula structures. These large amounts of samples can be evaluated through the GPE-based surrogate expressed in (22) at almost no computational cost. Finally, the pdf of the load margin can be obtained.

Here, we provide the summarized steps for the proposed method in Algorithm 1 and its corresponding flowchart in Fig. 2 for the readers's convenience.

Algorithm 1 Data-Driven Probabilistic Load-Margin Assessment Algorithm

1: Prepare the power system CPF model and the historical wind data;
2: Infer the nonparametric closed-form marginals using KDE via (29) for the wind speed data at different wind sites;
3: Select a vine-copula structure (C-vine or D-vine) and estimate the corresponding bivariate copulae;
4: Generate dependent samples, $\{x^{(i)}\}_{i=1}^{N_{\text{sample}}}$, to represent wind speeds for different wind sites through the inferred vine copula;
5: Generate a small amount of dependent training samples, $\mathbf{X}$, for the GPE model;
6: Map the training samples to wind power samples via (30);
7: Evaluate the CPF model at the wind power samples to obtain the realizations, $\mathbf{Y}$;
8: Select the kernel and the trend in the GPE model;
9: Estimate the hyperparameters in the GPE model;
10: Evaluate the GPE model at generated dependent samples, $\{x^{(i)}\}_{i=1}^{N_{\text{sample}}}$, to obtain system responses, $\{y^{(i)}\}_{i=1}^{N_{\text{sample}}}$;
11: Plot the pdf for $\{y^{(i)}\}_{i=1}^{N_{\text{sample}}}$ to represent load-margin uncertainty.

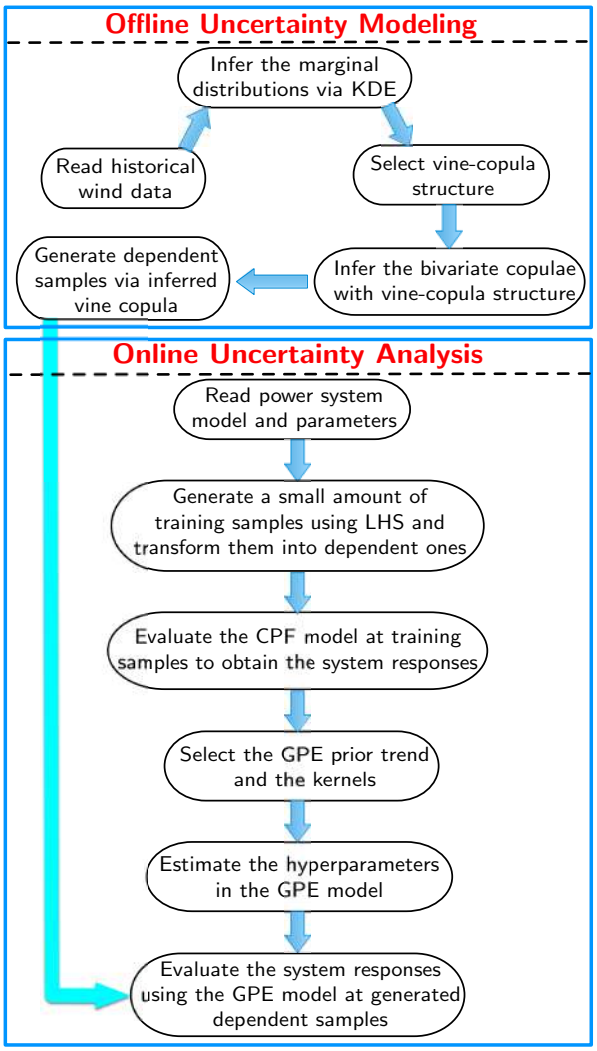


Fig. 2. Flowchart of the proposed method.

VI. Simulation Results

Various case studies are conducted under different IEEE test systems [38]. The real-world wind speed data are obtained via the NREL's Western Wind Data Set [39].

A. Case Study One

This case study is conducted on the IEEE 57-bus test system. Four 50 MW wind farms are added at Buses 16, 17, 47, and 48, respectively, to introduce randomness in the CPF model. The parameters set for wind generators are $v_{\text{in}} = 4$ [m/s], $v_{\text{out}} = 25$ [m/s], and $v_{\text{rated}} = 15$ [m/s] [34]. The real-world wind speed data for these four wind farms are collected from Sites $\#1116, \#9246, \#9435,$ and $\#9386$, from NREL's Data Set in the first season of 2004, respectively [39].

1) Nonparametric Marginal Inference: The kernel density estimator is first used to estimate the marginal densities for the wind speed data set of each site. Note that in the KDE, popular choice of the kernel, $K(\cdot)$, includes Gaussian kernel, box kernel, Epanechnikov kernel, and triangular kernel, etc. [40]. Besides, it is also worth pointing out that the key point for the success of the KDE relies on the choice of a proper bandwidth, instead of the type of the kernel. Thus, we highly encourage readers to set the default bandwidth using the rule-of-thumb as described in Section V. Let us take the wind speed data for Site $\#1116$ as an example. The KDE-inferred marginals with different kernels and bandwidth are plotted in Fig. 3. It is shown that under the default bandwidth, all of the kernels can provide good marginal approximations while the marginal approximated by a smaller bandwidth shows multiple locally spiking structures. It can also be seen that the marginal density of Site $\#1116$ does not strictly follow some heuristic parametric distributions (e.g., Weibull and Gamma). Therefore, it is necessary to use the nonparametric KDE to obtain its marginal.

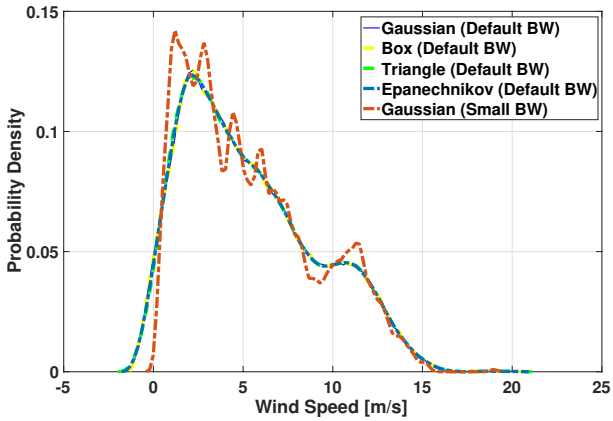


Fig. 3. KDE-inferred marginals of wind speed for Site $\#1116$ with different kernels and different bandwidth.

2) Vine-Copula Inference: Now, we use the D-vine copula to model the dependence structure of the data as Becker suggested [23]. The averaged half-hourly wind speed data for these four wind sites are used to infer the vine-copula structure. Here, we have $p(p-1)/2 = 6$ pair-copulas inferred from the four wind sites as listed in Table I. It shows that for Sites $\#9246, \#9435,$ and $\#9386$, using a symmetric copula (e.g., the Gaussian and Student's t-copulas) to describe their

dependence would not be sufficiently accurate, but use of the copulas from the Archimedean copula family (e.g., the Frank and Clayton copulas) is deemed more appropriate. This demonstrates the rationality and the necessity of using the vine copula to model the dependence structure of the wind generation instead of the traditional symmetric Gaussian copula.

TABLE I
PAIR COPULAE FOR WIND FARMS

Index	Pair	Family	Rotation	Θ
1	$C_{1,2}$	Student's t	0	$\{0.727, 21.1\}$
2	$C_{2,3}$	Student's t	0	$\{0.980, 1.93\}$
3	$C_{3,4}$	Frank	270°	$\{-1.11\}$
4	$C_{1,3\|2}$	Student's t	0	$\{-0.096, 13.31\}$
5	$C_{2,4\|3}$	Clayton	0	$\{0.116\}$
6	$C_{1,4\|2,3}$	Independent	0	–

3) Sample Comparison: Here, we use the inferred vine-copula structure to generate $3,000$ samples and compare them to the historical data. The marginal plots and scatter plots are shown in Fig. 4. It shows that the vine copulas provide convincing simulated samples, which match the asymmetric tail-dependent historical samples very well.

To quantitatively measure the accuracy of the vine copula, two indices suggested in the literature [10], [41] were included: (i) the log-likelihood (LL) and (ii) the Akaike information criterion (AIC). Their quantitative values obtained via different copulae have been provided in Table II. From these two indices, it can be clearly seen that both the D- and C-vine copulae exhibit very high accuracy, whereas the traditional Gaussian copula exhibits the lowest accuracy.

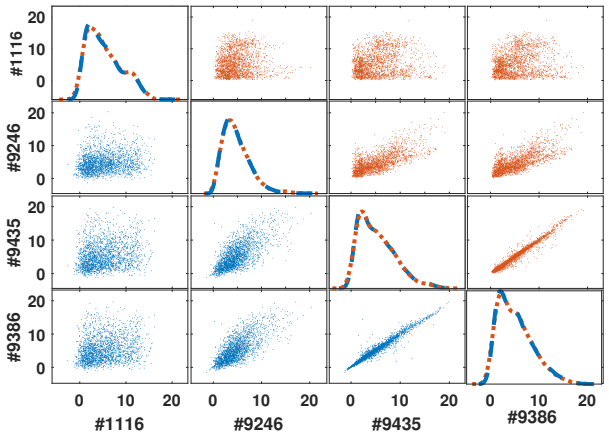


Fig. 4. Marginal and scatter plots of the simulated data (blue) and historical data (red) for the IEEE 57-bus test case.

TABLE II
QUANTITATIVE TEST FOR DIFFERENT COPULAE

Copula	D-Vine	C-Vine	Gaussian
LL	$4.215 \cdot 10^3$	$4.214 \cdot 10^3$	$3.892 \cdot 10^3$
AIC	$-8.4296 \cdot 10^3$	$-8.429 \cdot 10^3$	$-7.803 \cdot 10^3$

Remark 4. *It is worth pointing out that although the vine-copula algorithm takes more computing time to analyze the data structure than the straightforward multivariate Gaussian copula, it does not influence the computational efficiency of this uncertainty quantification application. This is because the uncertainty modeling can be conducted offline as demonstrated in Fig. 2. This makes sense since the historical wind data can be viewed as unchanged for a certain time period, for which the vine-copula algorithm does not need to be trained repeatedly. For an online probabilistic load-margin assessment application, once the current system topology as well as the control and operating states are updated, we can directly apply the offline, well-trained vine copula on the online application. Therefore, there is no need to consider the extra training time required for the vine copula when compared to the Gaussian copula.*

4) Uncertainty Propagation Validation: Now, let us validate the performance of the proposed GPE method in the uncertainty propagation stage. The simulation results obtained with the MC and GPE methods are provided in Fig. 5. The simulation results obtained using the MC method with $10,000$ samples are used as a benchmark to validate the GPE-based method (cf. [42]). In order to demonstrate that the result of the MC method can serve as a credible benchmark, we plot the convergence rate of the MC method in Fig. 6. It shows that the mean and variance of the MC method converge asymptotically given $10,000$ realizations. This demonstrates the credibility of the results obtained with the MC method.

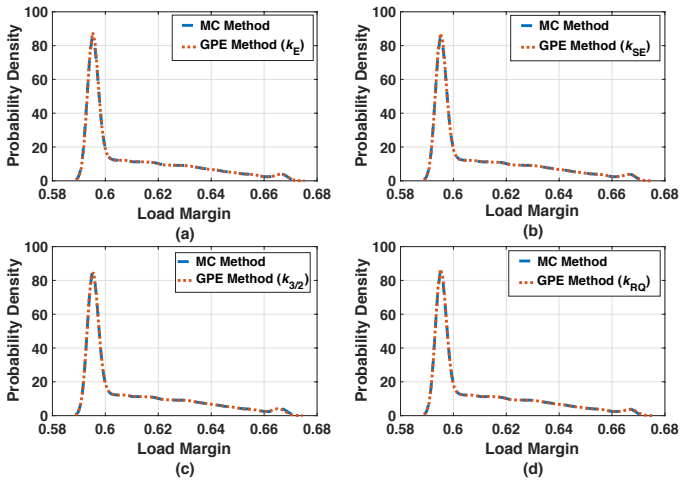


Fig. 5. Probability density plots for load margin using GPE with (a) k_{E} kernel, (b) k_{SE} kernel, (c) $k_{3/2}$ kernel, and (d) k_{RQ} kernel.

It can also be seen that with only 15 training samples, the GPE method with a 2nd-order truncated PCE functions can provide highly accurate simulation results under different kernel functions. To further quantitatively verify the accuracy of the proposed method, we choose to use the well-known Kullback-Leibler divergence (KLD) as the index [43], which measures the difference between the GPE-approximated pdf, π_{GPE}, and the MC-approximated pdf, π_{MC}, via

$$D(\pi_{\mathrm{GPE}}||\pi_{\mathrm{MC}}) = \int \pi_{\mathrm{GPE}} \log \frac{\pi_{\mathrm{GPE}}}{\pi_{\mathrm{MC}}}. \tag{36}$$

It is obvious that $D(\pi_{\mathrm{GPE}}||\pi_{\mathrm{MC}})$ equals zero when $\pi_{\mathrm{GPE}} = \pi_{\mathrm{MC}}$. Here, π_{MC} is obtained from the pdf obtained from evaluating the aforementioned $10,000$ samples directly through the full CPF model. Then, we use the KLD to quantitatively test the accuracy of the proposed method with different kernels.

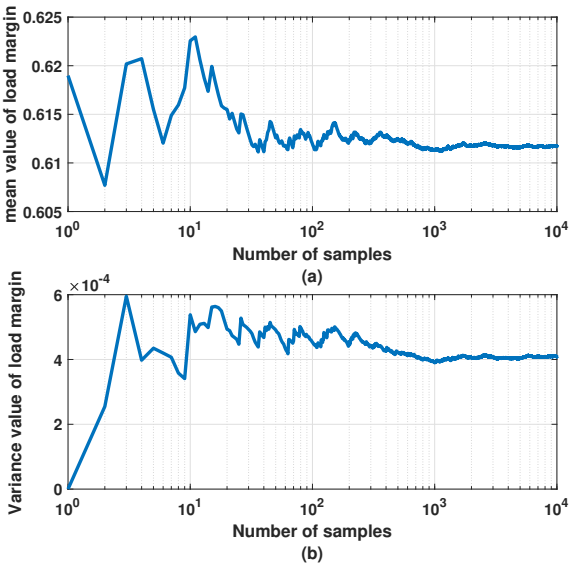


Fig. 6. Convergence test of the MC method with (a) mean and (b) variance of load margin.

The calculation results are shown in Table III. It can be seen that the proposed method provides quite similar estimation accuracy using all four types of kernels, with each calculated $D(\pi_{\text{GPE}}||\pi_{\text{MC}})$ tending to zero. Therefore, the proposed GPE-based method significantly reduces the computational burden over the traditional MC method without any loss of accuracy.

TABLE III
VALIDATION OF THE PROPOSED METHOD USING KLD

Kernel	k_{E}	k_{SE}	$k_{3/2}$	k_{RQ}
KLD	0.0019	0.0017	0.0030	0.0016

B. Case Study Two

This case study is conducted on the IEEE 118-bus test system to validate the performance of the proposed method on a larger system. In this experiment, 11 wind farms with rated power of 45 MW, 60 MW, 50 MW, 50 MW, 90 MW, 75 MW, 30 MW, 30 MW, 30 MW, 30 MW, and 90 MW are added at Buses 3, 7, 13, 16, 37, 38, 45, 50, 93, 94, and 114, respectively, to introduce the randomness in the CPF model.

The parameters set for the wind generators are kept the same as in Case Study One. The real-world wind speed data for these 11 wind farms are collected from Sites #12160, #12094, #11083, #10547, #9708, #9639, #9434, #9204, #1550, #1491, and #1032 of the NREL's wind data of 2004, respectively [39]. The samples obtained by the inferred D-vine are provided in Fig. 7, which demonstrates the existence of quite complicated dependence structures among these wind generation data, which further requires the validation of the proposed method.

1) Validation of The Proposed Method: Here, we test the performance of the proposed method using different kernels and different training samples. The KLD is used to quantify the accuracy of the proposed method. The simulation results obtained with the MC method with $10,000$ samples are used as a benchmark to validate the GPE-based method. It is worth pointing out that it takes more than 6 h to evaluate these

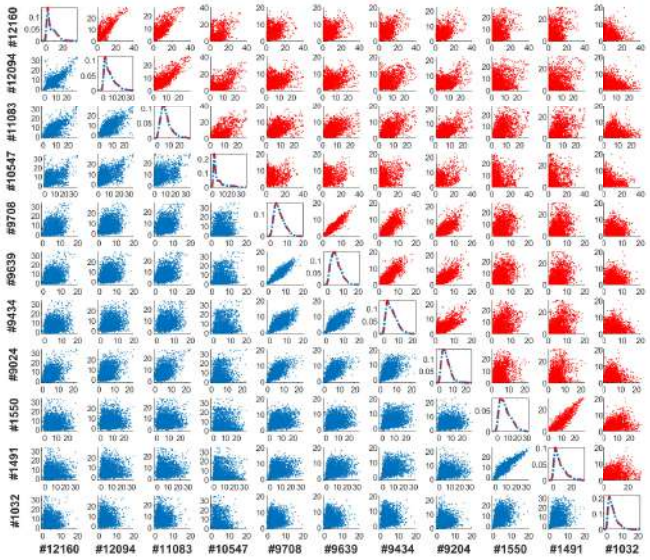


Fig. 7. Marginal and scatter plots of the simulated data (blue) and historical data (red) for the IEEE 118-bus test case.

$10,000$ samples using the CPF model in the MC method, which is very time-consuming. In contrast, the proposed GPE method takes much less time to complete the computing as shown in Table IV. With only 50 samples, the GPE method can complete the training stage in 1.5 min and the evaluation stage in around 5 s while maintaining a very high accuracy. Apart from the KLD, it can be seen from Fig. 8 that the pdf obtained by the proposed method matches the simulation results of the MC method very well. We can also see from Table IV that there is practically no difference in the computing efficiency and accuracy of the proposed method under different kernels. Furthermore, an increase in the training samples will not bring obvious improvement in the performance of the proposed method.

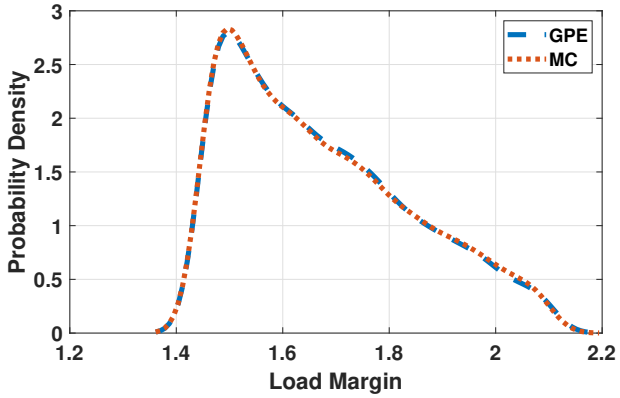


Fig. 8. Probability density plots for load margin using the GPE with k_{SE} kernel and 50 training samples, and the MC method.

2) Comparison Studies: Here, we conduct comparison studies with another existing method, the Latin hypercube sampling (LHS), which is a stratified sampling technique that can be used to reduce the number of runs necessary for an MC simulation to achieve a reasonably accurate random distribution [33]. This has been further verified and applied

TABLE IV
VALIDATION OF THE GPE METHOD WITH DIFFERENT TRAINING SAMPLES UNDER DIFFERENT KERNELS ON THE IEEE 118-BUS SYSTEM

Samples	Training Time (s)	Realization Time (s)	Total Time (s)	KLD
Group 1 using GPE with k_{E} Kernel				
50	88.07	4.78	92.85	$2.92 \cdot 10^{-4}$
75	134.23	4.81	139.04	$3.31 \cdot 10^{-4}$
100	166.23	5.14	171.37	$2.63 \cdot 10^{-4}$
Group 2 using GPE with k_{SE} Kernel				
50	87.39	4.84	92.23	$2.57 \cdot 10^{-4}$
75	134.95	4.92	139.87	$2.60 \cdot 10^{-4}$
100	167.61	5.66	173.27	$2.38 \cdot 10^{-4}$
Group 3 using GPE with $k_{3/2}$ Kernel				
50	88.07	4.78	92.85	$2.63 \cdot 10^{-4}$
75	130.90	4.86	135.76	$2.71 \cdot 10^{-4}$
100	160.75	5.09	165.84	$2.93 \cdot 10^{-4}$
Group 4 using GPE with k_{RQ} Kernel				
50	90.20	5	95.20	$3.59 \cdot 10^{-4}$
75	131.86	5.03	136.89	$2.62 \cdot 10^{-4}$
100	166.27	5.22	171.41	$3.38 \cdot 10^{-4}$

in some power-system applications such as probabilistic load-margin and probabilistic power-flow analyses [19], [44], [45]. Researchers advocate this method for its capability to provide a good statistical approximation of power system states by using a small number of "near-random" samples, e.g., 200 and 400. Here, we choose different sample sizes, N_{LHS}, for the LHS method to make comparison studies with the proposed method considering the computing accuracy and the computational efficiency. The simulation results are shown in Table V. The pdfs are plotted in Fig. 9. It can be seen that although the LHS method with 500 samples can provide a fairly reasonable pdf approximation, the GPE outperforms the LHS method in both the computing accuracy and the computational efficiency.

TABLE V
COMPARISON STUDIES FOR THE ACCURACY AND EFFICIENCY

N_{LHS}	100	200	300	400	500
KLD	0.1324	0.0980	0.0808	0.0671	0.0495
Time (s)	165.1	333.4	487.6	654.7	842.7

3) Further Discussion: Let us now discuss the possible generalization of the proposed method to much larger power systems. Apart from the regional-scale power systems that we have used in this article, the scale of the real-world power systems can increase to a much larger size. This means that even a single calculation of the CPF model, which involves multiple prediction and correction steps based on the power-flow solver, can be very time-consuming. Therefore, even if our proposed GPE-based method can greatly reduce the computation time of the MC simulation, the training period will still be time-consuming, reducing the overall performance of the proposed method.

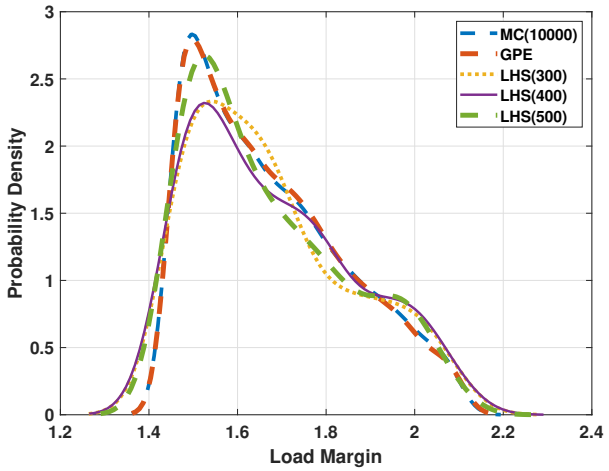


Fig. 9. Comparison studies between the GPE method and the LHS method.

Fortunately, the proposed method can be perfectly merged into the parallel-computing framework. Unlike some sequential method that might not be easily computed in parallel, the training period of the proposed method can be parallelized. This renders the proposed method amenable to the emerging GPU-based computing method (cf. [42], [46]) or some hybrid GPU/CPU-based computing method (cf. [47]). By this way, the proposed method can not only encompass the advantages of the GPE method, but also harness the power of modern-computing technologies. Undoubtedly, the parallelization will help the proposed method be generalized to much larger power systems. We would like to combine the GPE method with a GPU-based method for large-scale power-system applications as part of our future work.

VII. CONCLUSIONS AND FUTURE WORK

In this paper, we propose a novel data-driven GPE-based method for the probabilistic load-margin analysis. The high-dimensional, complicated dependence structure of the wind generation data are modeled with the vine copula. Then, the GPE serves as a nonparametric, reduced-order model for the nonlinear CPF that enables an evaluation of the time-consuming CPF solver at the sampled values at a negligible computational cost. Simulation results reveal that the proposed method exhibits an impressive performance as compared to the traditional MC method.

In the future, we would like to further explore the following aspects:

- The GPE method developed in this paper is not robust to outliers. However, as shown in the literature (cf. [32]), there may exist some outliers in the raw wind data. We will explore the ideas of making our GPE method robust to outliers (cf. [48]) or adding a raw-data preprocessing step (as done in [32]) prior to applying the GPE method.
- As mentioned earlier, the proposed method is currently applied in a regional-scale power grid. In the future, it would be meaningful to combine the current GPE method with the power of modern-computing technologies, such as a GPU-based method, to extend our method's applicability to a much larger power grid.

- Although the current CPF model is not a time-series problem, it would be interesting to expand the GPE method to a related uncertainty quantification problem involving time series such as the stochastic economic dispatch. Then, it would be interesting to combine the GPE method with state-of-the-art neutral-network-based methods that can recover full stochastic scenarios [49].

REFERENCES

[1] M. Fan, V. Vittal, G. T. Heydt, and R. Ayyanar, "Probabilistic power flow analysis with generation dispatch including photovoltaic resources," *IEEE Trans. Power Syst.*, vol. 28, no. 2, pp. 1797–1805, May 2013.

[2] Y. Xu, M. Korkali, L. Mili, X. Chen, and L. Min, "Risk assessment of rare events in probabilistic power flow via hybrid multi-surrogate method," *IEEE Trans. Smart Grid*, vol. 11, no. 2, pp. 1593–1603, Mar. 2020.

[3] Y. Xu, L. Mili, A. Sandu, M. R. von Spakovsky, and J. Zhao, "Propagating uncertainty in power system dynamic simulations using polynomial chaos," *IEEE Trans. Power Syst.*, vol. 34, no. 1, pp. 338–348, Jan. 2019.

[4] C. Safta, R. L.-Y. Chen, H. N. Najm, A. Pinar, and J.-P. Watson, "Efficient uncertainty quantification in stochastic economic dispatch," *IEEE Trans. Power Syst.*, vol. 32, no. 4, pp. 2535–2546, Jul. 2017.

[5] Z. Hu *et al.*, "Uncertainty quantification in stochastic economic dispatch using Gaussian process emulation," *arXiv:1909.09266*, Sep. 2019.

[6] E. Haesen, C. Bastiaensen, J. Driesen, and R. Belmans, "A probabilistic formulation of load margins in power systems with stochastic generation," *IEEE Trans. Power Syst.*, vol. 24, no. 2, pp. 951–958, May 2009.

[7] X. Xu, Z. Yan, M. Shahidehpour, H. Wang, and S. Chen, "Power system voltage stability evaluation considering renewable energy with correlated variabilities," *IEEE Trans. Power Syst.*, vol. 33, no. 3, pp. 3236–3245, May 2018.

[8] A. M. Leite da Silva, I. Coutinho, A. Z. De Souza, R. Prada, and A. Rei, "Voltage collapse risk assessment," *Electr. Power Syst. Res.*, vol. 54, no. 3, pp. 221–227, Jun. 2000.

[9] Z. Qin, W. Li, and X. Xiong, "Incorporating multiple correlations among wind speeds, photovoltaic powers and bus loads in composite system reliability evaluation," *Appl. Energy*, vol. 110, pp. 285–294, Oct. 2013.

[10] Z. Wang, W. Wang, C. Liu, Z. Wang, and Y. Hou, "Probabilistic forecast for multiple wind farms based on regular vine copulas," *IEEE Trans. Power Syst.*, vol. 33, no. 1, pp. 578–589, Jan. 2018.

[11] Z. Ren, W. Li, R. Billinton, and W. Yan, "Probabilistic power flow analysis based on the stochastic response surface method," *IEEE Trans. Power Syst.*, vol. 31, no. 3, pp. 2307–2315, May 2016.

[12] F. Ni, P. H. Nguyen, and J. F. G. Cobben, "Basis-adaptive sparse polynomial chaos expansion for probabilistic power flow," *IEEE Trans. Power Syst.*, vol. 32, no. 1, pp. 694–704, Jan. 2017.

[13] I. Konstantelos *et al.*, "Using vine copulas to generate representative system states for machine learning," *IEEE Trans. Power Syst.*, vol. 34, no. 1, pp. 225–235, Jan. 2019.

[14] M. Sun, I. Konstantelos, and G. Strbac, "C-vine copula mixture model for clustering of residential electrical load pattern data," *IEEE Trans. Power Syst.*, vol. 32, no. 3, pp. 2382–2393, May 2017.

[15] M. Sun, Y. Wang, G. Strbac, and C. Kang, "Probabilistic peak load estimation in smart cities using smart meter data," *IEEE Trans. Ind. Electron.*, vol. 66, no. 2, pp. 1608–1618, Feb. 2019.

[16] V. Ajjarapu, *Computational Techniques for Voltage Stability Assessment and Control*. Springer Science & Business Media, 2007.

[17] A. B. Rodrigues, R. B. Prada, and M. D. G. da Silva, "Voltage stability probabilistic assessment in composite systems: Modeling unsolvability and controllability loss," *IEEE Trans. Power Syst.*, vol. 25, no. 3, pp. 1575–1588, Aug. 2010.

[18] B. Qi, K. N. Hasan, and J. V. Milanović, "Identification of critical parameters affecting voltage and angular stability considering load-renewable generation correlations," *IEEE Trans. Power Syst.*, vol. 34, no. 4, pp. 2859–2869, Jul. 2019.

[19] J. Zhao, Y. Bao, and G. Chen, "Probabilistic voltage stability assessment considering stochastic load growth direction and renewable energy generation," in *Proc. IEEE Power Energy Soc. Gen. Meet.*, 2018.

[20] J. Zhang, C. Tse, K. Wang, and C. Chung, "Voltage stability analysis considering the uncertainties of dynamic load parameters," *IET Gener. Transm. Distrib.*, vol. 3, no. 10, pp. 941–948, Oct. 2009.

[21] G. Papaefthymiou and D. Kurowicka, "Using copulas for modeling stochastic dependence in power system uncertainty analysis," *IEEE Trans. Power Syst.*, vol. 24, no. 1, pp. 40–49, Feb. 2009.

[22] M. C. Kennedy and A. O'Hagan, "Bayesian calibration of computer models," *J. R. Stat. Soc. Ser. B (Stat. Method.)*, vol. 63, no. 3, pp. 425–464, 2001.

[23] R. Becker, "Generation of time-coupled wind power infeed scenarios using pair-copula construction," *IEEE Trans. Sustainable Energy*, vol. 9, no. 3, pp. 1298–1306, Jul. 2018.

[24] R. B. Nelsen, *An Introduction to Copulas*. Springer Science & Business Media, 2007.

[25] M. Jan-Frederik and M. Scherer, *Simulating Copulas: Stochastic Models, Sampling Algorithms, and Applications*, 2nd ed. Singapore: World Scientific, 2017.

[26] H. Joe, "Families of m-variate distributions with given margins and $m(m-1)/2$ bivariate dependence parameters," *Distributions with Fixed Marginals and Related Topics*, vol. 28, pp. 120–141, 1996.

[27] K. Aas, C. Czado, A. Frigessi, and H. Bakken, "Pair-copula constructions of multiple dependence," *Insur. Math. Econ.*, vol. 44, no. 2, pp. 182–198, Apr. 2009.

[28] A. Pepelyshev, "The role of the nugget term in the Gaussian process method," in *mODa 9–Advances in Model-Oriented Design and Analysis*. Springer, 2010, pp. 149–156.

[29] M. L. Eaton, *Multivariate Statistics: A Vector Space Approach*. New York, NY, USA: John Wiley & Sons, 1983.

[30] Y. Xu, L. Mili, and J. Zhao, "A novel polynomial-chaos-based Kalman filter," *IEEE Signal Process. Lett.*, vol. 26, no. 1, pp. 9–13, Jan. 2019.

[31] R. J. Bessa, V. Miranda, A. Botterud, J. Wang, and E. M. Constantinescu, "Time adaptive conditional kernel density estimation for wind power forecasting," *IEEE Trans. Sustainable Energy*, vol. 3, no. 4, pp. 660–669, Oct. 2012.

[32] L. Zheng, W. Hu, and Y. Min, "Raw wind data preprocessing: a data-mining approach," *IEEE Trans. Sustainable Energy*, vol. 6, no. 1, pp. 11–19, Jan. 2015.

[33] T. J. Santner, B. J. Williams, W. Notz, and B. J. Williams, *The Design and Analysis of Computer Experiments*. Springer, 2003.

[34] H. Sheng and X. Wang, "Probabilistic power flow calculation using non-intrusive low-rank approximation method," *IEEE Trans. Power Syst.*, vol. 34, no. 4, pp. 3014–3025, Jul. 2019.

[35] S. Roy, "Market constrained optimal planning for wind energy conversion systems over multiple installation sites," *IEEE Trans. Energy Convers.*, vol. 17, no. 1, pp. 124–129, Mar. 2002.

[36] A. Gelman *et al.*, *Bayesian Data Analysis*, 3rd ed. Boca Raton, FL, USA: Chapman & Hall, 2014.

[37] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. Cambridge, MA, USA: MIT Press, 2006.

[38] University of Washington, Power Systems Test Case Archive. (accessed: 10/2019). [Online]. Available: https://labs.ece.uw.edu/pstca/

[39] NREL Western Wind Data Set. (accessed on 10/30/2019). [Online]. Available: https://www.nrel.gov/grid/western-wind-data.html

[40] B. Silverman, *Density Estimation for Statistics and Data Analysis*. Chapman & Hall/CRC, 1998.

[41] Z. Wang, W. Wang, C. Liu, and B. Wang, "Forecasted scenarios of regional wind farms based on regular vine copulas," *J. Mod. Power Syst. Clean Energy*, vol. 8, no. 1, pp. 77–85, Jan. 2020.

[42] G. Zhou *et al.*, "GPU-accelerated algorithm for online probabilistic power flow," *IEEE Trans. Power Syst.*, vol. 33, no. 1, pp. 1132–1135, Jan. 2018.

[43] Y. Marzouk and D. Xiu, "A stochastic collocation approach to Bayesian inference in inverse problems," *Commun. Comput. Phys.*, vol. 6, no. 4, pp. 826–847, Oct. 2009.

[44] H. Yu, C. Chung, K. Wong, H. Lee, and J. Zhang, "Probabilistic load flow evaluation with hybrid Latin hypercube sampling and Cholesky decomposition," *IEEE Trans. Power Syst.*, vol. 24, no. 2, pp. 661–667, May 2009.

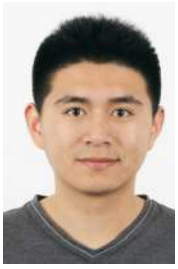
[45] Y. Chen, J. Wen, and S. Cheng, "Probabilistic load flow method based on Nataf transformation and Latin hypercube sampling," *IEEE Trans. Sustainable Energy*, vol. 4, no. 2, pp. 294–301, Apr. 2013.

[46] M. M. A. Abdelaziz, "OpenCL-accelerated probabilistic power flow for active distribution networks," *IEEE Trans. Sustainable Energy*, vol. 9, no. 3, pp. 1255–1264, Jul. 2018.

[47] G. Zhou *et al.*, "GPU-accelerated batch-ACPF solution for N-1 static security analysis," *IEEE Trans. Smart Grid*, vol. 8, no. 3, pp. 1406–1416, May 2017.

[48] M. Gu, X. Wang, and J. O. Berger, "Robust Gaussian stochastic process emulation," *Ann. Stat.*, vol. 46, no. 6A, pp. 3038–3066, 2018.

[49] Y. Chen, Y. Wang, D. Kirschen, and B. Zhang, "Model-free renewable scenario generation using generative adversarial networks," *IEEE Trans. Power Syst.*, vol. 33, no. 3, pp. 3265–3275, May 2018.



Yijun Xu (M'19) received the Ph.D. degree from Bradley Department of Electrical and Computer Engineering at Virginia Tech, Falls Church, VA, on December, 2018. He is currently a postdoc associate at Virginia Tech-Northern Virginia Center, Falls Church, VA. He did the computation internship at Lawrence Livermore National Laboratory, Livermore, CA, and power engineer internship at ETAP – Operation Technology, Inc., Irvine, California, in 2018 and 2015, respectively. His research interests include power system uncertainty quantification, uncertainty inversion, and decision-making under uncertainty.

Lamine Mili (LF'17) received the Ph.D. degree from the University of Liège, Belgium, in 1987. He is a Professor of Electrical and Computer Engineering, Virginia Tech, Blacksburg. He has five years of industrial experience with the Tunisian electric utility, STEG. At STEG, he worked in the planning department from 1976 to 1979 and then at the Test and Meter Laboratory from 1979 till 1981. He was a Visiting Professor with the Swiss Federal Institute of Technology in Lausanne, the Grenoble Institute of Technology, the École Supérieure D'électricité in France and the École Polytechnique de Tunisie in Tunisia, and did consulting work for the French Power Transmission company, RTE.

His research has focused on power system planning for enhanced resiliency and sustainability, risk management of complex systems to catastrophic failures, robust estimation and control, nonlinear dynamics, and bifurcation theory. He is the co-founder and co-editor of the *International Journal of Critical Infrastructure*. He is the chairman of the IEEE Working Group on State Estimation Algorithms. He is a recipient of several awards including the US National Science Foundation (NSF) Research Initiation Award and the NSF Young Investigation Award.

Mert Korkali (SM'18) received the Ph.D. degree in electrical engineering from Northeastern University, Boston, MA, in 2013. He is currently a Research Staff Member at Lawrence Livermore National Laboratory, Livermore, CA. From 2013 to 2014, he was a Postdoctoral Research Associate at the University of Vermont, Burlington, VT.

His current research interests lie at the broad interface of robust state estimation and fault location in power systems, extreme event modeling, cascading failures, uncertainty quantification, and probabilistic grid planning. He is the co-chair of the IEEE Task Force on Standard Test Cases for Power System State Estimation. Dr. Korkali is currently serving as an Editor of the IEEE OPEN ACCESS JOURNAL OF POWER AND ENERGY and of the IEEE POWER ENGINEERING LETTERS, and an Associate Editor of *Journal of Modern Power Systems and Clean Energy*.

Kiran Karra received his Ph.D. degree in Electrical Engineering from Virginia Tech in Falls Church, VA on July 2018. He is currently a Research Staff Member at the Johns Hopkins University Applied Physics Lab. From 2013 to 2018, he was a research faculty member at Virginia Tech's Hume Center in Arlington, VA. His current interests lie at the intersection of statistical signal processing and machine learning.

Zongsheng Zheng (S'18) received the B.S. degree in bioinformatics from Southwest Jiaotong University, Chengdu, China, in 2013. He is currently pursuing the Ph.D. degree in electrical engineering with Southwest Jiaotong University, Chengdu, China. During 2018-2019, he was a visiting scholar at Bradley Department of Electrical and Computer Engineering at Virginia Tech-Northern Virginia Center, Falls Church, VA, USA. His research interests include uncertainty quantification, parameter and state estimation.

Xiao Chen received his Ph.D. degree in Applied Mathematics from Florida State University, Tallahassee, FL in 2011. He is a Computational Scientist and a Project Leader in the Center for Applied Scientific Computing at Lawrence Livermore National Laboratory, Livermore, CA. He works primarily on the development and application of advanced computational and statistical methods and techniques to power engineering, reservoir simulation, subsurface engineering, and seismic inversion. His research interests include uncertainty quantification, data assimilation, and machine learning.