A Generalized Copula-Polynomial Chaos Expansion for Probabilistic Power Flow Considering Nonlinear Correlations of PV Injections

Ketian Ye, Junbo Zhao, *Senior Member, IEEE*, Rui Yang, Yingchen Zhang, *Senior Member, IEEE*, Xiaodong Liu

2020 52nd North American Power Symposium (NAPS) | 978-1-7281-8192-9/21/$31.00 ©2021 IEEE | DOI: 10.1109/NAPS50074.2021.9449783

***Abstract*—This paper develops a generalized Copula-polynomial chaos expansion (PCE) framework for power system probabilistic power flow that can handle both linear and nonlinear correlations of uncertain power injections, such as wind and PVs. A data-driven Copula statistical model is used to capture the correlations of uncertain power injections. This allows us to resort to the Rosenblatt transformation to transform correlated variables into independent ones while preserving the dependence structure. This paves the way of leveraging the PCE for surrogate modeling and uncertainty quantification of power flow results, i.e., achieving the probabilistic distributions of power flows. Simulations carried out on the IEEE 57-bus system show that the proposed framework can get much more accurate results than other alternatives with different linear and nonlinear power injection correlations.**

***Index Terms*—Probabilistic power flow, polynomial chaos, uncertainty quantification, copula, nonlinear correlations**

I. Introduction

Power flow has been widely used in today's energy management system for contingency analysis. With the increased penetration of uncertain distributed energy resources (DERs), such as wind and solar, the voltage magnitudes and angles, and power flows are subject to uncertainties. To deal with that, the probabilistic power flow is proposed [1]. Monte Carlo simulation (MCS) [2] is often employed as a benchmark for probabilistic power flow as it allows to obtain the solutions to the ground-truth via an enormous number of samples. However, it is too time consuming to be applied for practical power systems with large-scales. To this end, other efficient techniques are developed, such as perturbation-based [3] and analytical derivation-based methods [4]. Although they can improve the computational efficiency, their assumptions on probability distributions and liberalizations may not be satisfied under various operating conditions.

Probabilistic power flow actually belongs to the uncertainty quantification that can provide the uncertainty propagation from inputs to the outputs. The latter are quantified by the statistics of the model response. Polynomial chaos expansion (PCE) is a well-developed technique for uncertainty quantification thanks to its capability of capturing the model response statistics given the uncertain inputs. Polynomial Chaos is firstly introduced by Wiener [6] and generalized by Xiu et al. [7]. PCE and its variants have been used for both dynamical simulations and probabilistic power flow [4]-[9]. For example, the sparse PCE is applied for probabilistic load flow calculation. An enhanced PCE for computational efficiency improvement via the ANOVA method is developed [8]. The adaptive sparse PCE for probabilistic power flow is also developed in [9]. It should be noted that PCE typically requires that the uncertain variables are independent with each other. However, the wind and solar exhibits both temporal and spatial correlations. The existing PCE-based probabilistic power flow methods can deal with the linear correlations among uncertain variables. The critical concern is that by investigating the practical solar farm generations, nonlinear correlations are identified [10]. This violates the linear assumption and may lead to large errors, which will be demonstrated in the numerical simulation section.

This paper proposes a generalized Copula-PCE framework for probabilistic power flow that can handle both linear and nonlinear dependencies among the uncertain inputs while achieving high computational efficiency. The data-driven Copula statistics are utilized to model both the linear and nonlinear dependence structure of power injections. This is achieved via the marginal and copula inference. The latter will be taken as the inputs of Rosenblatt transformation to transform correlated variables into independent ones while preserving the dependence structure. As a result, the PCE can be conveniently adopted for probabilistic power flow. The impacts of different nonlinear dependence, copula types and parameters on the proposed method are also investigated.

II. Problem Statement

The probabilistic power flow aims to obtain the statistics of bus voltage magnitudes and angles, i.e., $\boldsymbol{V}$ and $\boldsymbol{\theta}$. Its model can be expressed as follows:

$$\boldsymbol{y} = \mathcal{M}(\boldsymbol{x}) \tag{1}$$

where the input vector is $\boldsymbol{x}$, i.e., uncertain power injections; $\boldsymbol{y}$ is the model response, typically $\boldsymbol{V}$ and $\boldsymbol{\theta}$; $\mathcal{M}$ is the nonlinear mapping function represented by power flow equations.

The above problem belongs to the uncertainty quantification [11] and the surrogate modeling technique is widely used. One

This work was supported by National Science Foundation under grant ECCS 1917308. K. Ye and J. Zhao are with the Department of Electrical and Computer Engineering, Mississippi State University, Starkville, MS 39762 USA. Y. Rui and Y. Zhang are with National Renewable Energy Laboratory, Golden, CO 80401 USA. X. Liu is with Eversource Energy, Boston, MA (e-mail: ky291@msstate.edu, junbo@ece.msstate.edu).

such choice is the PCE method. PCE is a spectral method to express $\boldsymbol{y}$ in terms of a polynomial function of $\boldsymbol{x}$, i.e.,

$$\boldsymbol{y}_{PC} = \sum_{i=0}^{\infty} a_i \Psi_i(\boldsymbol{x}) \tag{2}$$

where Ψ_i is orthogonal basis and a_i is the corresponding coefficient. Replacing original probabilistic power flow model with a series of polynomial expansion $\boldsymbol{y}_{PC}$, a surrogate model $\mathcal{M}_{PCE}$ is then established.

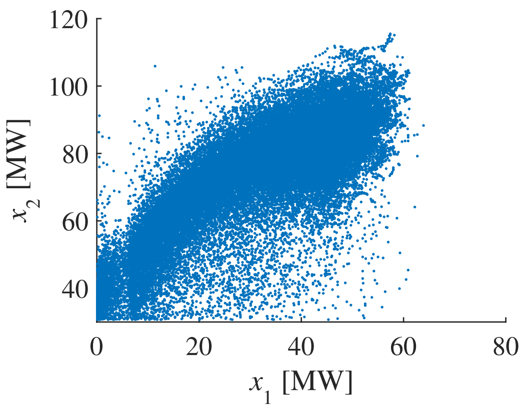

Fig. 1. Joint distribution of two solar farm generations.

PCE model assumes that the input vector $\boldsymbol{x}$ is independent. Although some techniques have been developed to address linear correlations of $\boldsymbol{x}$, this method may not be applicable in the presence of nonlinear correlations. Indeed, the PV station generations can exhibit nonlinear dependencies. For example, by using the realistic data from NREL, the two geographically close solar farm generations are plotted against each other and shown in Fig.1. It is clear from the figure that nonlinear dependence exists. By using the Copula statistics, we find that the nonlinear dependencies may follow $Gumbel$, $Frank$, $Clayton$ types. As a result, the linear assumption-based PCE model may yield significant estimation errors. This motivates us to develop a more general Copula-PCE model that handles both linear and nonlinear correlations of uncertain inputs.

III. Proposed Coupla-PCE Framework

The proposed Coupla-PCE framework is shown in Fig. 2. It contains three key steps, including copula modeling to capture linear and nonlinear dependencies of input variables, inverse Rosenblatt Transformation (inverse RT) to transform dependent variables into independent ones without losing statistical information, and the surrogate modeling for probabilistic power flow calculations. The first step is achieved offline based on inferred copula from historical data. After inverse RT, the independent random vector $\boldsymbol{z}$ is obtained and this allows the PCE construction.

A. Copula Statistics for Nonlinear Dependence Modeling

Copula statistics have been well-known to be able to capture the complicated dependence structures among random variables [12]. The construction of multivariate distribution is general and flexible by defining the copula type and the

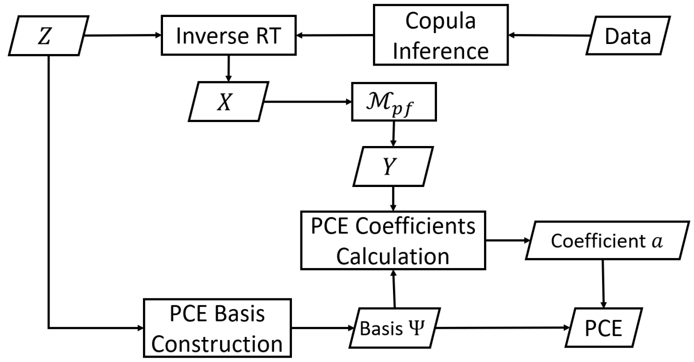

Fig. 2. Framework of copula-based PCE

marginals. According to Sklar's theorem, for a d-dimensional continuous random variable $\boldsymbol{x} = [x_1, .., x_d]$ with marginals $F_1, \ldots, F_d$ and joint cumulative distribution function (CDF), $F_{\boldsymbol{x}}$, there exists a copula function C that satisfies [13]

$$F_{\boldsymbol{x}}(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d)) \tag{3}$$

$$F_{j|\mathcal{A}}(x_j|\boldsymbol{x}_{\mathcal{A}}) = \frac{\partial C_{j,i|\mathcal{A}\setminus\{i\}}(F(x_j|\boldsymbol{x}_{\mathcal{A}\setminus\{i\}}), F(x_i|\boldsymbol{x}_{\mathcal{A}\setminus\{i\}}))}{\partial F(x_i|\boldsymbol{x}_{\mathcal{A}\setminus\{i\}})} \tag{4}$$

where $j \in \mathcal{A}, i \in D\backslash\mathcal{A}$; $\mathcal{A}$ is a subset of indices $D = \{1, \ldots, d\}$; $F_{j|\mathcal{A}}$ denotes the CDF of the random variable x_j conditioned on $x_{\mathcal{A}}$, and $C_{(j|\mathcal{A})}$ is the copula function.

The commonly used copulas include Gaussian, t and Clayton types. Note that most bivariate copulas only require one parameter to describe the correlation between random variables. However, these copulas face challenges when being applied to random vectors with high dimensions. To deal with that, vine copula is proposed [13]. For regular vine (R-vine) copulas, the dependency structure is determined by a product of bivariate copulas and a nested set of trees. Two special cases of R-vine, i.e., the Canonical vine (C-vine) and Drawable vine (D-vine), are often utilized to simplify vine structure. In this paper, the C-vine copula is used and it is expressed as

$$c(\boldsymbol{x}) = \prod_{j=1}^{d-1} \prod_{i=1}^{d-j} c_{j,j+i|\{1,\ldots,j-1\}}(x_{j|\{1,\ldots,j-1\}}, x_{j+i|\{1,\ldots,j-1\}}) \tag{5}$$

where $c_{j,j+i|1,\ldots,j-1}(x_{j|1,\ldots,j-1}, x_{j+i|1,\ldots,j-1})$ is the bivariate copula between x_j, x_{j+i} conditioned on $x_1, \ldots, x_{j-1}$. The copula structure is inferred from data. Firstly, bivariate copulas are inferred for each pair of random variables. Based on that, vine copula is deduced, where the parameters are estimated via the maximum likelihood estimator.

B. Inverse Rosenblatt Transformation

Rosenblatt transformation (RT) allows us to transform dependent random variables into independent random ones. RT is general and it is able to handle both linear and nonlinear dependencies. Given a random vector $\boldsymbol{x} \in \mathbb{R}^n$ and its marginal CDF F_i and copula C, RT maps $\boldsymbol{x}$ to an independent random vector $\boldsymbol{u}$ with $u_i \sim U[0,1], (i = 1, \ldots, n)$ [21]. After that, the independent random vector $\boldsymbol{u}$ can be further transformed into another random vector $\boldsymbol{z}$ by corresponding inverse probability integral transform (PIT) [16]. For example, $\boldsymbol{z}$ can be an independent standard normal distributed random vector if the inverse CDF of a standard normal distribution is applied. Accordingly, the inverse operation of transforming $\boldsymbol{z}$ into $\boldsymbol{x}$ with dependent structure $C_{\boldsymbol{x}}$ is provided as follows:

$$\boldsymbol{x} = T_{RT}^{-1} \circ T_{PIT}(\boldsymbol{z}) \tag{6}$$

$$T_{PIT} : \boldsymbol{z} \rightarrow \boldsymbol{u} = [\Phi(z_1), \Phi(z_2), \ldots, \Phi(z_n)] \tag{7}$$

$$T_{RT}^{-1} : \boldsymbol{u} \rightarrow \boldsymbol{x} = \begin{cases} F_1^{-1}(u_1) \\ \ldots \\ F_{k|1,\ldots,k-1}^{-1}(u_n|u_1,\ldots,u_{k-1}) \\ \ldots \\ F_{n|1,\ldots,n-1}^{-1}(u_n|u_1,\ldots,u_{n-1}) \end{cases} \tag{8}$$

where $F_{k|1,\ldots,k-1}^{-1}$ is the inverse CDF of the conditioned random variable $u_n|u_1,\ldots,u_{k-1}$ and Φ is the CDF of the standard normal distribution. When the random variable is Gaussian copula type, RT reduces to the Nataf Transformation.

C. PCE with Nonlinear Dependence of Inputs

As discussed in section II, PCE can be constructed once the basis Ψ_i and the corresponding coefficients a_i are determined. The construction of basis can be achieved via the Stiltjes procedures [14]. For independent random vector $\boldsymbol{x} = [x_1, x_2, \ldots, x_n]$ with more than one random variable, the orthogonal basis is then built based on one-dimensional PC through the tensor product [16]. For related inner product calculations, Gaussian quadrature rule can be used. As for coefficients a_i, several methods are available, such as projection method and least-squares minimization method using a set of collocation points. Gaussian quadrature is a commonly used projection method with good accuracy and efficiency [16]. In particular, Smolyak's sparse quadrature can be used to overcome the curse of dimensionality when dealing with high-dimensional integration [11].

In practice, truncation is performed and (2) becomes

$$\boldsymbol{y}_{PC}^t = \sum_{i=0}^{N_P} a_i \Psi_i(\boldsymbol{x}) \tag{9}$$

where $N_P = (n+P)!/(n!P!) - 1, n$ is the number of components in $\boldsymbol{x}$; P is the degree of the truncated polynomial function. Note that the adaptive sparse PCE allows one to build a PCE with growing number of components [16].

With calculated orthogonal basis and the corresponding coefficients, the mean and variance can be estimated via

$$E[\boldsymbol{y}_{PC}^t] = a_0 \qquad var[\boldsymbol{y}_{PC}^t] = \textstyle\sum_{i=1}^{N_P} a_i^2 E[\psi_i^2] \tag{10}$$

where $E[\psi_i^2] = 1$ for an orthonormal basis. It is worth pointing out that the PCE can work as a surrogate model as well to provide pointwise predictions of model response $\boldsymbol{y}$. The surrogate model is more computationally efficient when using MCS as compared to the original power flow model.

As shown in Fig.1, the practical PV power injections have nonlinear dependence structure. The traditional PCE using (9) and (10) assume that the random inputs are independent. Specifically, in the presence of nonlinear dependency structure, the orthogonality property does not hold for true [17]. As a result, some modifications are required to deal with that. In this paper, two main strategies are identified, Strategy I) reconstruction of PCE that preserves the orthogonality property even the random inputs have dependencies and Strategy II) preprocessing on inputs to transform the dependent variables into independent ones.

Strategy I: the first strategy is to develop a more general method for constructing a PCE basis. Gram-Schmidt orthonormalization is an alternative way to construct basis when tensor product fails to preserve orthogonality in the presence of dependent inputs [17]. However, applying additional complicated procedures is conflicting with the initial purpose of simplifying the probabilistic power flow model. Computational demand will also increase under this circumstance [17].

Strategy II: the second strategy is to transform dependent input random vector $\boldsymbol{x}$ into an independent one $\boldsymbol{z}$. Then, the PCE can be directly applied to $\boldsymbol{z}$ and obtain the following form:

$$\boldsymbol{y}_{PC}^t = \sum_{i=0}^{N_P} a_i' \Psi_i'(\boldsymbol{z}) \tag{11}$$

In this paper, the second strategy is adopted to achieve reduced complexity and easy implementations.

D. Algorithm Implementation

Fig. 2 illustrates the general procedure of constructing PCE. For PCE in (11), the basis Ψ and the coefficients a_i are the two fundamental elements. For input $\boldsymbol{x}$ with dependent structure $C_{\boldsymbol{x}}$, additional step for data transformation is required. In summary, the main steps for implementing the proposed method are as follows:

- Step 1: Modeling the nonlinear dependent inputs using copula statistics. Specifically, the CDF (4) is obtained from inferred copula $C_{\boldsymbol{x}}$ and its inverse function. Consequently, the inverse RT (8) is then built on inverse CDF. Note that this step is done offline and data-driven as long as the dependence structure does not change, which is the case for practical power systems;
- Step 2: Constructing the PCE basis Ψ based on independent random vector $\boldsymbol{z}$. As discussed in section III-C, this step involves the Stieltjes procedure [14], tensor product [16], and parameter estimations.
- Step 3: Calculating coefficients of PCE. Coefficients are evaluated on a set of observations $\{(z_i, y_i)\}_{(i=1)}^{N_c}$, where z_i are representative samples of the random vector $\boldsymbol{z}$ and y_i are the corresponding responses of the original model $\mathcal{M}_{pf}$. Since the original model requires input $\boldsymbol{x}$, the collocation points vector $\boldsymbol{z}$ are first transformed to $\boldsymbol{x}$ via (7) and (8) before being fed into the original model. Note that the number of collocation points N_c must be no smaller than the number of basis N_P. In addition, the PCE coefficients estimation process is the most time-consuming parts of the proposed method but it naturally fits the parallel computing structure. In the future work, we will explore the benefits of doing that to further speed up the computing.

TABLE I
COMPARISON RESULTS OF DIFFERENT PCE MODELS

Model	$\mathbf{e}_{\mu}(\times 10^{-2}\%)$		$\mathbf{e}_{\sigma^2}(\%)$		**CPU time** (s)
	V	θ	V	θ	
$\mathcal{M}_{pf}$	—	—	—	—	302.63
$\mathcal{M}_{indep}$	0.582	13	39.50	55.03	$(9.14 + 6.57)$
$\mathcal{M}_{lnr}$	0.224	16	2.14	2.03	$(10.64 + 7.66)$
$\mathcal{M}_{nonlnr}$	0.01	0.473	0.46	0.11	$(12.52 + 9.07)$

IV. NUMERICAL RESULTS

In this section, the performances of the proposed method are tested on the modified IEEE 57-bus system with integration of PV and wind generations. PV injections are assumed to have nonlinear dependencies and modeled by Coupla statistics while the power injections from wind generations are assumed to be linearly correlated Weibull distribution [23]. Specifically, using the data set from SRML [19] that has annual year 2019 historical data of 5-minute time interval, three stations are selected and copula dependence structures of PV generations $C_{\boldsymbol{x}}$ are inferred. $C_{\boldsymbol{x}}$ consists of five random variables $\boldsymbol{x} = [x_1, x_2, x_3, x_4, x_5]$, where x_1, x_2, x_3 are power outputs of PV generations while x_4, x_5 are those from wind generations. These five generations are assumed to be at buses 14, 44, 45, 41 and 43, respectively as shown in Fig.3. For PV power outputs x_1, x_2, x_3, the marginals are estimated through kernel density estimation. Vine copula structure $C_{\boldsymbol{x}}$ for $\boldsymbol{x}$=$[x_1, x_2, x_3, x_4, x_5]$ is inferred by maximum likelihood estimator. Note that the wind power outputs x_4, x_5 are independent of PV power outputs x_1, x_2, x_3. The approximated dependence structure model $C_{\boldsymbol{x}}$ is used for actual power output to generate input $\boldsymbol{x}$.

In this paper, the AC power flow model is taken as the original model $\mathcal{M}_{pf}$. The selected responses are voltage magnitude V_i and voltage angle θ_i on bus i. The model outputs are $\boldsymbol{y} = [V_1, \ldots, V_{57}, \theta_1, \ldots, \theta_{57}]$. For multi-output systems, each component in $\boldsymbol{y}$ is modeled by an independent polynomial expansion P_i on the same collocation points. Three comparison models are considered when the inputs $\boldsymbol{x} = [x_1, \ldots, x_5]$ have vine copula structure. Model I: All the variables $x_i (i = 1, \ldots, 5)$ are independent, that is, constructing PCE model $\mathcal{M}_{indep}$ ignoring the dependence of input vector $\boldsymbol{x}$. Model II: a linear relationship between random variables is assumed as an approximation. The correlation coefficients can be employed to describe the dependence structure of the input. Following that, PCE model $\mathcal{M}_{lnr}$ is constructed. In this case, the transformation of correlated input $\boldsymbol{x}$ into uncorrelated input $\boldsymbol{z}$ can be either RT or inverse Cholesky Transformation. Model III: The proposed method that uses a vine copula structure and RT to help construct PCE model $\mathcal{M}_{nonlnr}$ considering both linear and nonlinear dependent inputs. A total number of 114 polynomial expansions are included in each PCE model. The performance is evaluated on the evaluation data set $\boldsymbol{x}_{val}$, which is sampled from copula $C_{\boldsymbol{x}}$. MCS results $\boldsymbol{y}_{val} = \mathcal{M}_{pf}(\boldsymbol{x}_{val})$ are used as the baseline to assess the performance of each PCE models. The mean and variance of the output of each PCE model are used to calculated and used to assess the accuracy of each one. Specifically, the percentage relative absolute error (%RAE) is used as the index and is shown as follows:

$$\%RMAE = \frac{1}{N} \sum \left| \frac{y^* - \hat{y}}{y^*} \right| \tag{12}$$

where $\hat{y}$ are the estimated results from method while y represents the true values.

For measuring the distance between two different distributions, the Kullback–Leibler divergence characterizes the difference with respect to probability density function [17]. Note that voltage magnitudes and angles on reference bus and the voltage magnitudes on PV generator buses are excluded from the evaluation index as they are constant. Since the statistics of the estimation result are the main focuses, %RAE of mean and variance on each measuring response, $\{V_i,\theta_j\}$, are calculated, and the averages of %RAEs are denoted as $\{e_{\mu}, e_{\sigma^2}\}$. All simulations are conducted in MATLAB with 3.00 GHz Intel Core i7-9700 PC.

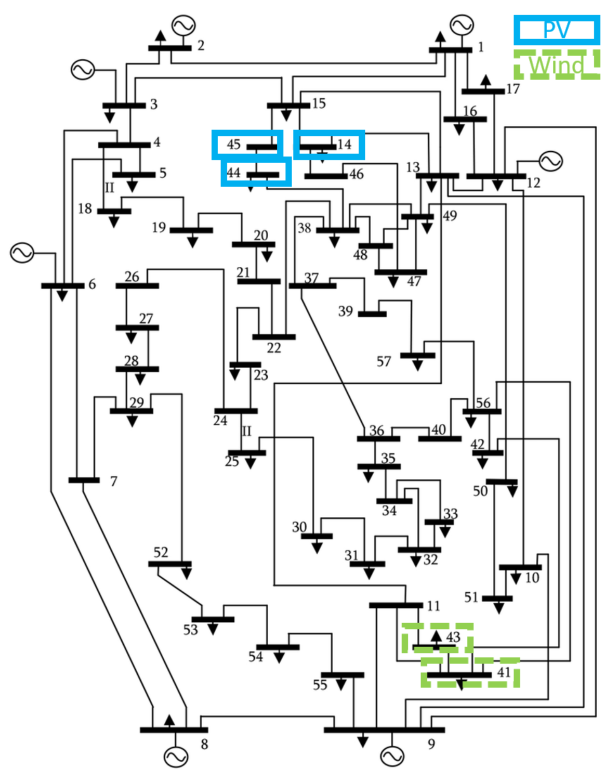

Fig. 3. Modified IEEE 57-bus system with PV and wind integration

A. Validation of the Proposed Method

Due to space limitation, only the results of voltage magnitude and angle at bus 45 are used for illustrations. The evaluation results for all four models $\{\mathcal{M}_{pf}, \mathcal{M}_{indep}, \mathcal{M}_{lnr}, \mathcal{M}_{nonlnr}\}$ are provided as demonstrated in Fig.4. The experiment is based on the data set from

TABLE II

COMPARISON RESULTS OF DIFFERENT PCE MODEL PERFORMANCES WITH DIFFERENT COPULA TYPES AND PARAMETERS.

			Copula Parameter t				**Copula Type**				**PCE Degree** n			
			1	2	3	4	$Frank$	$Gumbel$	t	$Clayton$	2	3	4	5
$\mathbf{e}_{\mu}$ ($\times 10^{-3}\%$)	V	$\mathcal{M}_{nonlnr}$	0.088	0.192	0.078	0.149	0.036	0.177	0.091	0.100	0.218	0.100	0.053	0.013
		$\mathcal{M}_{lnr}$	0.115	0.811	0.189	0.609	0.125	0.206	0.609	2.242	0.587	2.242	1.224	0.557
	θ	$\mathcal{M}_{nonlnr}$	4.394	7.418	5.313	7.881	3.147	9.748	6.142	4.735	8.354	4.735	3.14	0.633
		$\mathcal{M}_{lnr}$	25.19	59.99	10.28	41.37	26.72	22.95	44.86	162.4	39.68	162.4	88.86	31.14
$\mathbf{e}_{\sigma^2}$ (%)	V	$\mathcal{M}_{nonlnr}$	0.693	0.853	0.725	0.328	0.328	0.340	0.573	0.462	3.339	0.462	0.469	0.278
		$\mathcal{M}_{lnr}$	2.561	2.439	2.233	1.653	1.828	0.756	1.139	2.141	3.825	2.141	1.576	1.523
	θ	$\mathcal{M}_{nonlnr}$	0.239	0.306	0.374	0.112	0.145	0.082	0.402	0.109	1.741	0.109	0.250	0.127
		$\mathcal{M}_{lnr}$	2.274	1.827	1.561	1.087	2.491	1.204	1.366	2.033	3.260	2.033	1.140	1.355

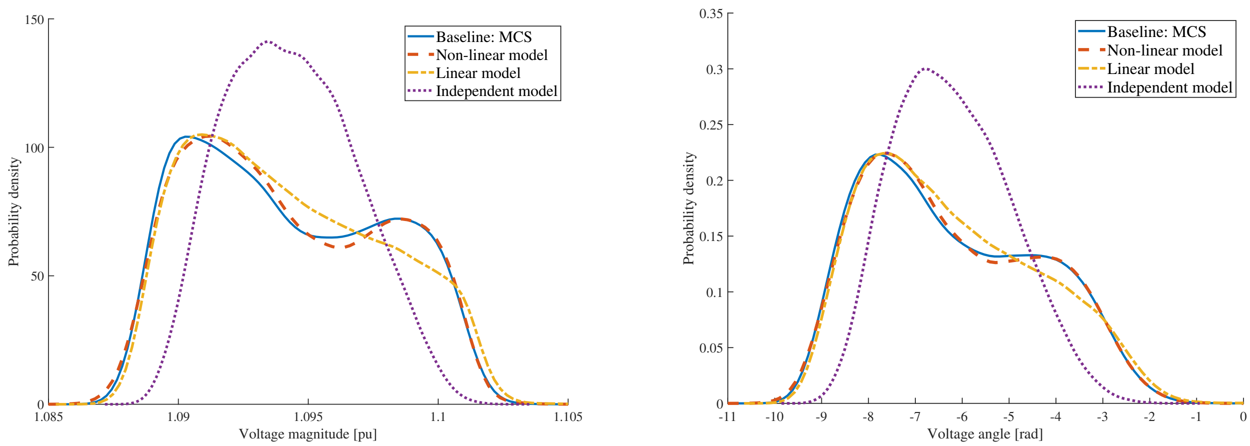

Fig. 4. Comparison results: PDFs of voltage magnitude and angle on bus 45 from $\{\mathcal{M}_{pf}, \mathcal{M}_{indep}, \mathcal{M}_{lnr}, \mathcal{M}_{nonlnr}\}$

TABLE III

COPULA STRUCTURE OF $C_{\boldsymbol{x}}$ AND PARAMETER SETTINGS

Bivariate copula	**Family**	**Rotation**	**Parameter** t
$C_{3,2}$	Clayton	0	3.7173
$C_{3,1}$	Gumbel	180	3.0480
$C_{2,1\|3}$	Gumbel	0	1.3364
$C_{4,5\|3,2,1}$	Gaussian	0	0.6012
others	Independent	0	—

SRML introduced in section IV. The validation input data set $\boldsymbol{x}_{val}$ is sampled from $C_{\boldsymbol{x}}$ with a sampling number of $N=10^5$, where $C_{\boldsymbol{x}}$ is inferred from data set. The degree of PCE $n=3$. The PDF curve is obtained by kernel density estimation.

Table I shows the comparison results of each method, including the estimation errors as well as the computing times. It can be found that if no correlations are modeled for the inputs, $\mathcal{M}_{indep}$ yields significant errors on both the bus voltage magnitudes and angles. When the linear correlation is used to approximate the nonlinear dependence structure among the inputs, $\mathcal{M}_{lnr}$ achieves significant improvement as compared to that of $\mathcal{M}_{indep}$. By contrast, the proposed $\mathcal{M}_{nonlnr}$ is able to effectively capture both linear and nonlinear correlations of the inputs and yield much more accurate results than the other two methods. In terms of computational efficiency of each method, the CPU times are the average values of PCE construction by implementing the algorithm 100 times. For the MCS, $N = 10^5$ samples are used. It is worth pointing out that the CPU time for PCE models consists of two main parts, the collocation points determination via regressions and PCE model basis, coefficients and voltage magnitudes and angles calculations. In Table I, the CPU time is divided to reflect that. It has been elaborated in the algorithm implementation section that the first part of CPU time can be significantly reduced via the parallel computing. While for the second part, more powerful computers and the ANOVA technique can be leveraged to reduce that. These will be our future works. It is observed that since the proposed method involves more complicated process of dealing with nonlinear dependencies, it takes a little bit more time than the $\mathcal{M}_{lnr}$. The MCS is of 15 times higher order complexity of the proposed method.

B. Effects of Degree of PCE Orders

It is well-known that the degree of PCE order can impact both accuracy and computational cost. It is usually set to be 2 in the literature as it allows us to achieve sufficiently high accuracy. To investigate if this choice works for the scenario, where nonlinear dependence occurs, different degrees of PCE orders are tested. Table.II shows the testing results. Note that the initial experimental settings are the same as those in the

Section IV-A. It can be found that the overall performance improves when the order n increases as expected. Both $\mathcal{M}_{lnr}$ and $\mathcal{M}_{nonlnr}$ achieve better results on e_μ than e_{σ^2}. This is also clearly demonstrated in Fig. 4. However, there are exceptions such as e_{σ^2} of $n = 3$ and $n = 4$ for $\mathcal{M}_{nonlnr}$. This indicates that the nonlinear model is complicated and n can be a hyper-parameter for tuning. According to (9), model construction time will also increase with more terms in PCE. For this power system application, to balance computational efficiency without loss of accuracy, the number of degrees is chosen to be $n = 3$.

C. Effects of Copula Types and Parameters

The inferred copula $C_{\boldsymbol{x}}$ describes the linear/nonlinear relationship among the uncertain inputs. As a result, different bivariate copula types and parameters will affect the results. On other hand, due to weather and geographical changes, the dependence structure of PV power outputs may vary. To investigate their impacts on the proposed method, different copula types are considered. Specifically, the bivariate copula type of (x_3, x_2) is set to be $\{Frank, Gumbel, t\}$, respectively. While the bivariate copula parameter t of (x_3, x_2) is altered to $\{2, 3, 4\}$ respectively to investigate the impacts of copula parameter. The results are presented in Table II. It can be observed that different dependence structures of the random variables indeed affect PCE model performance. In general, under various conditions, our proposed method still outperforms the $\mathcal{M}_{lnr}$. Specifically, the performances of $\{\mathcal{M}_{lnr}, \mathcal{M}_{nonlnr}\}$ on e_μ are close. But there are notable differences on e_{σ^2}. For different copula types, $\mathcal{M}_{nonlnr}$ significantly outperforms $\mathcal{M}_{lnr}$ for all indices. This indicates that the nonlinear model has outstanding advantages on these three types. On the other hand, the performances of both models are close for some indices and copula types, e.g., e_μ on V. Note that for different copula parameters t, the nonlinear dependencies affect the degree of nonlinearity. When t=4, $\mathcal{M}_{nonlnr}$ achieves much better performance than the $\mathcal{M}_{lnr}$. It is interesting to note that if the nonlinearity is weak, their performances are close while $\mathcal{M}_{lnr}$ has better computational efficiency. Therefore, depending on the degree of nonlinearity, the $\mathcal{M}_{nonlnr}$ and $\mathcal{M}_{lnr}$ can be effectively combined together to obtain high accuracy while being computationally attractive. This will be investigated in the future.

V. Conclusion

In this paper, a generalized Copula-PCE framework is proposed to deal with both linear and nonlinear dependent uncertain inputs for power system probabilistic power flow. The critical step is to infer the dependence structure from the data via copula statistics and leverage the RT to transform dependent variables into independent ones. The latter allows us to conveniently adopt PCE for quantifying power flow results. Simulation results on the IEEE 57-bus system show that the proposed method can deal with different nonlinear dependence structures modeled by different copula types and parameters. Its performance is also much better than the method that only considers linear correlations. It is also found that the widely used 2 degrees of order for PCE does not work well for nonlinear dependent variables, instead a degree of 3 is a good trade-off between computational efficiency and accuracy.

References

[1] B. Borkowska, "Probabilistic load flow," *IEEE Trans. Power App. Syst.*, vol. PAS-93, no. 3, pp. 752-755, May-Jun, 1974.

[2] P. Jorgensen, J. S. Christensen and J. O. Tande, "Probabilistic load flow calculation using Monte Carlo techniques for distribution network with wind turbines," *8th International Conference on Harmonics and Quality of Power*, Athens, Greece, pp. 1146-1151, 1998.

[3] Y. Xu, *et al*., "Propagating uncertainty in power system dynamic simulations using polynomial chaos," *IEEE Trans. Power Syst.*, vol. 34, no. 1, pp. 338-348, Jan. 2019.

[4] X. Sun, *et al*., "Probabilistic load flow calculation based on sparse polynomial chaos expansion," *IET Gener. Transm. Distrib.*, vol. 12, no. 11, pp. 2735-2744, Jun. 2018.

[5] M. E. El-Hawary, G. A. N. Mbamalu, "A comparison of probabilistic perturbation and deterministic based optimal power flow solutions," *IEEE Trans. Power Syst.*, vol. 6, no. 3, pp. 1099-1105, Aug. 1991.

[6] N. Wiener, "The homogeneous chaos", *Amer. J. Math.*, vol. 60, No. 4, pp. 897-936, Oct. 1938.

[7] Xiu, D. and Karniadakis, G.E.. "The Weiner-Askey polynomial chaos for stochastic differential equations," *SIAM J. Sci. Comput.*. vol. 24, no. 2, pp. 619-644, 2002.

[8] Y. Xu, L. Mili and J. Zhao, "Probabilistic power flow calculation and variance analysis based on hierarchical adaptive polynomial chaos-ANOVA method," *IEEE Trans. Power Syst.*, vol. 34, no. 5, pp. 3316-3325, Sept. 2019.

[9] F. Ni, P. H. Nguyen, 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.

[10] W. Wu, *et al*., "A versatile probability model of photovoltaic generation using pair copula construction," *IEEE Trans. Sustainable Enery*, vol. 6, no. 4, pp. 1337-1345, Oct. 2015.

[11] E. Torre, *et al*, "A general framework for data-driven uncertainty quantification under complex input dependencies using vine copulas," *Probabilistic Eng. Mech.*, vol 55, pp. 1-16, Jan. 2019.

[12] P. Krupskii, H. Joe, "Factor copula models for multivariate data," *J. Multivariate Anal.*, vol. 120, pp. 85-101, 2013.

[13] H. Joe, "Families of m-variate distributions with given margins and $m(m-1)/2$ bivariate dependence parameters," *L. Ruschendorf, B. Schweizer, and M. D. Taylor (Eds.), Distributions with fixed marginals and related topics*, vol. 28 of Lecture Notes–Monograph Series, pp. 120–141. Institute of Mathematical Statistics, 1996.

[14] W. Gautschi, "Orthogonal polynomials-constructive theory and applications," *J. Comput. Appl. Math.*, vol. 12-13, pp. 61-76, 1985.

[15] Golub, G. H. and J. H. Welsch, "Calculation of gauss quadrature rules," *Math. Comput.*, vol. 23, pp. 221–230, 1969.

[16] S. Marelli, B. Sudret, "UQLab user manual – Polynomial chaos expansions," Report # UQLab-V1.3-104, Chair of Risk, Safety and Uncertainty Quantification, ETH Zurich, Switzerland, 2019.

[17] Emiliano Torre, *et al.*, "Data-driven polynomial chaos expansion for machine learning regression," *J. Comput. Phys.*. vol. 388, pp. 601-623, Jul. 2019.

[18] I. Abouzahr and R. Ramakumar, "Loss of power supply probability of stand-alone photovoltaic systems: a closed form solution approach," *IEEE Trans. Energy Convers.*, vol. 6, no. 1, pp. 1–11, Mar. 1991.

[19] University of Oregon, Solar Radiation Monitoring Laboratory, dataset available at: http://solardat.uoregon.edu/SolarData.html

[20] National Renewable Energy Laboratory, dataset available at: https://www.nrel.gov/grid/solar-power-data.html

[21] M. Rosenblatt, "Remarks on a multivariate transformation," *Ann. Math. Stat.*, vol. 23, pp. 470–472, 1952.

[22] S. Marelli, B. Sudret, "UQLab: A framework for uncertainty quantification in MATLAB," ICVRAM 2014, University of Liverpool, United Kingdom, vol. 13-16, pp. 2554–2563, Jul. 2014.

[23] X. Xu, Z. Yan, "Probabilistic load flow calculation with quasi-Monte Carlo and multiple linear regression," *Int. J. Elect. Power Energy Syst.*, vol. 88, pp. 1-12, 2017.