Basis-adaptive sparse polynomial chaos expansion for probabilistic power flow

Basis-Adaptive Sparse Polynomial Chaos Expansion for Probabilistic Power Flow

F. Ni, P. H. Nguyen, *Member, IEEE,* J. F. G. Cobben, *Member, IEEE*

***Abstract*— This paper introduces the basis-adaptive sparse polynomial chaos (BASPC) expansion to perform the probabilistic power flow (PPF) analysis in power systems. The proposed method takes advantage of three state-of-the-art uncertainty quantification methodologies reasonably: the hyperbolic scheme to truncate the infinite polynomial chaos (PC) series; the least angle regression (LARS) technique to select the optimal degree of each univariate PC series; and the Copula to deal with nonlinear correlations among random input variables. Consequently, the proposed method brings appealing features to PPF, including the ability to handle the large-scale uncertainty sources; to tackle the nonlinear correlation among the random inputs; to analytically calculate representative statistics of the desired outputs; and to dramatically alleviate the computational burden as of traditional methods. The accuracy and efficiency of the proposed method are verified through either quantitative indicators or graphical results of PPF on both the IEEE European Low Voltage Test Feeder and the IEEE 123 Node Test Feeder, in presence of more than one hundred correlated uncertain input variables.**

***Index Terms*— Probabilistic power flow, polynomial chaos, Copula theory, distribution system, photovoltaic generator.**

NOTATION

The primary notation used in this paper is given here, while other definitions are given in the paper as needed.

Constants and Indices:

Ω	Sample space
F	The σ-algebra on Ω
P	Probability measure on (Ω, F)
δ_{ij}	Kronecker delta function
M	Number of random variables
p	Truncation criteria
q	Truncation norm
i_j	The j-th degree of the i-th univariate polynomial
λ	The positive penalty factor
K+1	The total number of gPC expansion terms
N_{ed}	Number of experimental design
N_{sam}	Number of samplers for the LHS-based PPF
T_{total}	The total computational time of PPF
t_{pre}	Time for the pre-process of PPF
t_{post}	Time for the post-process of PPF
t_0	Time for a single round of DPF simulation
t_{ref}	Time required by the reference method
t_{cp}	Time assumed by the method to be compared
PL_i	Original active power demand of the i-th load
PV_i	Original active power production of the i-th PV
r^{ref}	Result obtained from the reference method
r^{cp}	Result obtained from the method to be compared
c_i	Location of the i-th bin center on x-axis
d_i	Height of the i-th rectangular bin on y-axis
N^M	The complete M-dimensional sphere of natural numbers
$\|\mathbf{i}\|_q$	Rank of the multi-index $\mathbf{i}$
PC	Polynomial chaos
gPC	Generalized polynomial chaos
BASPC	Basis-adaptive sparse polynomial chaos
LHS	Latin hypercube sampling

Variables and Sets:

$L^2(\Omega, F, P)$	A finite dimensional random space
ξ	Random input variable
$\boldsymbol{\xi}$	Random input vector
$f(\xi)$	The stochastic response to random variable ξ
$\omega(\xi)$	PDF of random variable ξ
$\omega(\boldsymbol{\xi})$	Joint PDF of random vector $\boldsymbol{\xi}$
$F_{\boldsymbol{\xi}}(\xi)$	Joint CDF of random vector $\boldsymbol{\xi}$
$\{\Phi_i\}$	The i-th univariate polynomial basis
$\{\Psi_i\}$	The i-th multivariate polynomial basis
$\{a_i\}$	Coefficient of the i-th polynomial basis
$\mathbf{a}$	The vector of coefficients$\{a_j\}$
$\hat{\mathbf{a}}$	The calculated coefficient vector via regression
$\mathbf{X}$	A set of experimental design of random vector
$\mathbf{Y}$	The stochastic response vector corresponding to inputs $\mathbf{X}$
$\mathbf{X}\backslash\mathbf{x}_l$	A set of experimental design of random vector without $\mathbf{x}_l$
$f_{-l}^{gPC}(\mathbf{X}\backslash\mathbf{x}_l)$	The gPC expansion of $f(\boldsymbol{\xi})$ with $\mathbf{X}\backslash\mathbf{x}_l$
e_{LOO}	Relative leave-one-out cross-validation error

F. Ni, P. H. Nguyen, J. F. G. Cobben are with the Electrical Energy Systems Group, Eindhoven University of Technology, Eindhoven, 5612 AZ, The Netherlands (e-mail: [f.ni, p.nguyen.hong, j.f.g.cobben]@tue.nl).

J. F. G. Cobben is also with the Alliander, Arnhem 6812 AH, The Netherlands.

I. INTRODUCTION

MODELLING, operating and planning of power systems are inevitably subject to various sources of uncertainties, mostly as fluctuations of load demands, changes of the grid topology, outages of the generator and other network components. Over the last few years, the high penetration of renewable energy sources (RES) such as wind power and photovoltaic (PV), and new forms of energy consumptions such as the heat pump (HP) and electric vehicle (EV), bring extra uncertainties to the power system due to variations in both

weather conditions and stochastic behaviors of the end users. Therefore, the uncertainty quantification (UQ) which is of great importance for both grid operators and energy consumers, has attracted ever-increasing attentions [1]-[2].

As one basic tools of operation and planning of the power system, power flow analysis identifies the steady operation state with node voltages and branch power flow in the power system, based on a specified generating state and network structure [3]. In order to perform the power flow calculation, it is necessary to know conditions of the power system, such as loads and generators in terms of active power, reactive power or voltage. Traditionally, the deterministic power flow (DPF) carries out power flow calculation under specific conditions, resulting in an incomplete understanding of operational states of the power system. Moreover, DPF becomes infeasible when the system conditions are subject to inaccurate evaluations or randomness of natural phenomena [4].

In order to take a deep and overall insight into the power system status affected by various types of uncertain factors, probabilistic power flow (PPF) [5] was found as a new path to evaluate states and desired outputs of the system in terms of probability distribution function (PDF), cumulative distribution function (CDF), probability mass function (PMF) or statistical moments. PPF analysis combines the principle of DPF and the procedure of UQ, yet advances in accounting uncertainties in power systems. During the last two decades, there has been a growing interest in PPF [6]-[8], since the uncertainty sources in power systems grows rapidly in the aspect of dimensionality, especially in distribution systems.

Generally, three main approaches are categorized to perform PPF: simulation method, analytical method, and approximate method [9]. The most straightforward one is the simulation method, which firstly replaces probabilistic inputs with a set of random samples and repeatedly executes a batch of DPF calculations on the designated models. Despite of its popularity, the simulation method propagates uncertainty in the inputs to the outputs of interest without reformulations on the DPF code [10]. To guarantee the convergence of random samplings, such as the brute Monte Carlo (MC) or the MC-like approaches, a large amount of trials are required to represent random inputs, which leads to enormous rounds of the DPF calculation accordingly. Meanwhile, several advanced random sampling methods such as Latin supercube sampling (LSS) Latin hypercube sampling (LHS), and importance sampling have been applied to PPF over the last decade. With the aid of efficient variance reduction techniques, advanced sampling methods are able to achieve the same convergence with a smaller number of trials, hence alleviating the computational effort of the simulation method. In order to avoid cumbersome simulations and to simplify the calculation procedure of PPF, several approaches have been proposed within the scope of the analytical method, among which the multi-linear model based method, the direct convolution method, the Fourier transform method, cumulant method and its extended versions are the most commonly used ones [11]-[12]. Owing to the effective pre-digestion, the analytical method does not need a huge number of repeated simulations; those assumptions conversely limit usages and brings in sources of inaccuracy though. Comparing with the two methods aforementioned, the approximate method is able to reach an ingenious compromise between accuracy and efficiency. This kind of method only needs a small amount of DPF calculations while requiring extra mathematical treatments of random inputs and outputs. In particular, the first-order second-moment method, the point estimate method, and their various alternatives are capable of propagating uncertainty from the inputs to the outputs in the form of statistical moments [13]-[14]. Moreover, several methods that represent and evaluate uncertainties via PDFs, PMFs or CDFs have been employed to address PPF analysis, among which the stochastic collocation interpolation [15] and the polynomial chaos (PC) expansion [16] are impressive ones.

The PC expansion was first introduced by Wiener, which exclusively addressed standard Gaussian random variables and the corresponding Hermite polynomial. Later on, Xiu et al. extended the method to the gPC expansion, where several non-Gaussian distributions were handled with a specific family of orthogonal polynomials [17]. If a certain distribution type that is out of the list occurs, the gPC expansion is still applicable with the aid of employing an iso-probabilistic transform [18]-[19]. Moving up with a greater number of random input variables, the gPC expansion encounters a critical barrier in practical usage, i.e. the curse of dimensionality. However, the sparse-adaptive scheme firstly presented by Blatman and Sudret is capable of effectively overcoming this difficulty to a great extent [20]. The previous works of the gPC expansion in this field have investigated effects of uncertain factors on power systems [21]- [22], but with a limited numbers of independent uncertainties (no more than twenty). Therefore, it is important to investigate the applicability and performance of the sparse-adaptive scheme in power system analysis with a greater number of correlated random inputs due to massive integration of RESs, HPs, and EVs recently.

In order to perform PPF analysis of power systems regardless of choices of methods, it is necessary to properly address two critical issues: the large number of random inputs, which leads to a high-dimension UQ problem; and the correlation among random input variables. Hence, methods that are able to handle these two challenges are of great interest. With regard to the first issue, it is difficult for either the analytical method or the approximate method to maintain their advantages. In the presence of massive random inputs, the traditional analytical method needs a large amount of data storage and computational effort to obtain the solution of PPF, while the conventional approximate method may require an even greater number of simulation rounds than the simulation method. When it comes to the correlation issue, a common practice for the approximate methods recently is to firstly de-correlate those dependent input variables, and then adopt existing UQ methods that are suitable for the independent case. In the previous work, the linear dependence was adopted at large, and recently the nonlinear dependence based on the Copula theory has been introduced into the electrical engineering [23]-[24].

Motivated by the two challenges mentioned previously, this paper proposes a novel methodology for PPF, which is able to perform an economical and reliable analysis in the presence of random inputs with both high dimensionality and nonlinear correlation. Specifically, the proposed method belongs to the category of the approximate method, cooperating both the basis-adaptive sparse PC (BASPC) expansion and the Copula theory. The remaining part of this paper is organized as follows:

in Section II, theoretical background related to the BASPC expansion and the Copula theory is briefly introduced; Section III presents detailed computation procedure of the PPF based on BASPC, compared with the PPF that is based on the simulation method; a variety of scenarios are elaborated to verify behaviours of the proposed method and to investigate performance-impact factors in Section IV; the concluding remarks and outlooks come in Section V.

II. Theoretical Background

In this section, we first introduce the BASPC expansion, including the generalized polynomial chaos (gPC) theory and the sparse-adaptive scheme in part A. After that, the Copula theory is presented in part B, which is used to remove the nonlinear dependence between random input variables.

A. Basis-Adaptive Sparse Polynomial Chaos Expansion

1) Generalized polynomial chaos expansion

Given a finite dimensional random space $L^2(\Omega, F, P)$ where Ω is a sample space, F is the σ-algebra on Ω, and P stands for the probability measure on (Ω, F), any stochastic response with finite second moment can be expanded in a convergent series of orthogonal polynomials of the random inputs, according to the Cameron-Martin theorem. By taking the gPC expansion, the relationship between random input variables and the resulting second-order stochastic response can be specified.

Let ξ be a random input variable, then the target stochastic response $f(\xi)$ can be represented as,

$$f(\xi) = \sum_{i=0}^{\infty} a_i \Phi_i(\xi) \tag{1}$$

where $\{a_i\}$ is the coefficient of the i-th polynomial, $\{\Phi_i\}$is the i-th univariate polynomial basis. The orthogonality of basis $\{\Phi_i\}$ with respect to a probability measure can be expressed as,

$$E[\Phi_i \Phi_j] = \int \Phi_i(\xi)\Phi_j(\xi)\,\omega(\xi)\,d\xi = \delta_{ij} E[\Phi_i^2] \tag{2}$$

in which $E[.]$ is the expectation operator, δ_{ij} is the Kronecker delta function, and $\omega(\xi)$ is the PDF of ξ.

As a merit of the gPC expansion compared to the simulation method, statistical moments of the stochastic response can be obtained in closed-form. For instance, the expected value and the variance are expressed by the following,

$$\begin{aligned} &\mu(f(\xi)) = a_0 \\ &\sigma^2(f(\xi)) = \sum_{i=1}^{\infty} a_i^2 E[\Phi_i^2] \end{aligned} \tag{3}$$

In principle, the random inputs can also be a vector $\boldsymbol{\xi}$ with M independent random variables $\{\xi_1, \xi_2, \ldots, \xi_M\}$, the corresponding stochastic response thus is approximated similarly,

$$f(\boldsymbol{\xi}) = \sum_{i=0}^{\infty} a_i \Psi_i(\boldsymbol{\xi}) \tag{4}$$

where $\{\Psi_i\}$ is the multivariate orthogonal polynomial basis that can be constructed as a tensor product of univariate orthogonal polynomials as follows:

$$\Psi_i(\boldsymbol{\xi}) = \Phi_{i_1}(\xi_1) \otimes \Phi_{i_2}(\xi_2) \otimes \cdots \otimes \Phi_{i_M}(\xi_M) \tag{5}$$

where the subscript i_j, j=0,…,M refers to the j-th degree of the i-th univariate polynomial basis. Further, the orthogonality of $\{\Psi_i\}$ still hold with regard to the joint PDF $\omega(\boldsymbol{\xi})$ of the random vector. For practical reasons, the infinite expression either in (1) or (4) has to be truncated to a finite number of terms with a given criteria p. Most often, the gPC expansion of a stochastic response is truncated in such a way that only those polynomial basis in (5) with total degree not higher than p are maintained, namely, $\sum_{j=1}^{M} i_j \le p$. With this truncation strategy, the number of remained terms of the gPC expansion is equal to

$$K = (M+p)!/M!\,p! \tag{6}$$

With known polynomial basis and truncation criteria, the polynomial coefficients $\{a_i\}$ need to be determined in order to complete the gPC expansion. Generally speaking, approaches to evaluate gPC coefficients are classified as the intrusive Galerkin projection and non-intrusive methods. Regarding the intrusive approach, it requires the formulation of a new mathematical problem whose solution is the set of polynomial coefficients. The scale of reformulated problem is remarkably larger than the original one in terms of the number of system states, and is solved by adopting additional solver. Therefore, this approach is not suitable for a system with large-scale random inputs or with a complex mathematical model. On the contrary, non-intrusive approaches treat the original model and solver as a black box, polynomial coefficients can be calculated only by a set of realizations of the original system.

Among non-intrusive approaches, sampling, quadrature and linear regression are widely used. In the first two methods, the polynomial coefficients are calculated by taking advantage of the orthogonality of polynomial bases as follows:

$$a_i = \frac{E[f(\boldsymbol{\xi})\Psi_i(\boldsymbol{\xi})]}{E[\Psi_i^2(\boldsymbol{\xi})]} = \frac{1}{E[\Psi_i^2(\boldsymbol{\xi})]} \int X(\boldsymbol{\xi})\,\Psi_i(\boldsymbol{\xi})\,\omega(\boldsymbol{\xi}) d\boldsymbol{\xi} \tag{7}$$

The linear regression method, also known as the stochastic response surface [19], evaluates the polynomial coefficients by directly solving an optimization problem. Let $\mathbf{X}=\{\mathbf{x}_1,\ldots,\mathbf{x}_N\}$be a set of representatives of the random input vector, also called as the experimental design (ED), and $\mathbf{Y}=\{f(\mathbf{x}_1),\ldots,f(\mathbf{x}_N)\}^{\mathrm{T}}$ be the corresponding stochastic response vector, the set of coefficients can be computed then by minimizing a norm of residuals as,

$$\begin{aligned} S(a_0,\cdots,a_K) &= \sum_{l=1}^{N} [f(\mathbf{x}_l) - \sum_{i=0}^{K} a_i \Psi_i(\mathbf{x}_l)]^2 \\ &= \sum_{l=1}^{N} [f(\mathbf{x}_l) - \mathbf{a}^{\mathrm{T}} \mathbf{H}(\mathbf{x}_l)]^2 \end{aligned} \tag{8}$$

where $\mathbf{a}$ is the vector of coefficients$\{a_i\}$ (i=0,…,K), and the so-called experimental matrix $\mathbf{H}$ has the following form:

$$\mathbf{H} = \begin{bmatrix} \Psi_0(\mathbf{x}_1) & \cdots & \Psi_K(\mathbf{x}_1) \\ \vdots & \ddots & \vdots \\ \Psi_0(\mathbf{x}_N) & \cdots & \Psi_K(\mathbf{x}_N) \end{bmatrix} \tag{9}$$

The method to calculate the coefficient vector $\mathbf{a}$ in (8) could be the ordinary least-squares (OLS), then the solution vector reads,

$$\hat{\mathbf{a}} = (\mathbf{H}^{\mathrm{T}}\mathbf{H})^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{Y} \tag{10}$$

Note that the employed design of experiments $\mathbf{X}$ should contain a sufficient number of points, preferably 2-3 times the number of expansion terms. In addition, the above results are still valid when the stochastic response is a vector.

2) The sparse-adaptive scheme

As expressed in (6), the total number of the gPC expansion terms K+1 grows rapidly along with increase of both M and p, which directly increases the number of ED. For example, by taking p=4, K grows from 1001 to 10626 when M increases from 10 to 20, thus blowing up efforts on both calculation and simulation, significantly. To work out the high dimensionality difficulty, the sparse-adaptive scheme was proposed. In what follows, the sparse-adaptive scheme is introduced through three different steps.

Revised truncation strategy

In principal, any truncation strategy aims at specifying a non-empty nested subset of the complete M-dimensional sphere of natural numbers. The truncation strategy mentioned above is a special case, which evenly treats the degree of each univariate polynomial in the tensor product as shown in (5). Nevertheless, according to the sparsity-of-effects principle, a model response is always dominated by the main effects. In other words, the interactions among low-degree univariate polynomial basis will be of more statistical significance, relative to the interactions among high-degree terms. Thus, one of the straightforward ways to reduce the value of K+1 comes out by favoring interactions among low-degree terms, through the revised truncation strategy [20]:

$$\|\mathbf{i}\|_q = \sqrt[q]{\sum_{j=1}^{M}(i_j)^q} \le p, \quad \mathbf{i}\in\mathbb{N}^M \ \&\ q\in(0,1) \tag{11}$$

where $\mathbf{i}$ is the multi-index. Note that the conventional strategy corresponds to the $\mathbf{1}$-norm condition by setting q=1. If q<1, the total number K can be dramatically reduced as the high-degree interactions are penalized.

Advanced regression algorithm

The revised truncation strategy presents a novel way to build the gPC expansion in a sparse form. Besides, the high-degree interactions can be further reduced by adding a penalty term on the least-square minimization problem in (8) as,

$$S(a_0,\cdots,a_K)=\sum_{l=1}^{N}[f(\mathbf{x}_l)-\mathbf{a}^{\mathrm{T}}\mathbf{H}(\mathbf{x}_l)]^2+\lambda\sum_{l=1}^{N}a_l \tag{12}$$

where λ is the positive penalty factor. In (12), the penalized optimization problem enables a selection of the most important polynomials among an initially defined set. In order to solve the problem in the format of (12), several algorithms are available, such as the least absolute shrinkage and selection operator, the forward stage-wise regression and the least angle regression (LARS). In fact, these algorithms are similar to each other in the sense of optimization mechanism, but differences mainly lie in the greedy level, the precision target and the search speed.

In statistics, LARS is a regression tool for fitting the linear model to high-dimensional data, thus is effective even when the number of regressors is much bigger than the available data. In the context of gPC expansion, LARS has been applied to build the sparse PC expansion, with main steps explained in [20]. In short, by selecting the most relevant polynomials with respect to the given model evaluations $\mathbf{Y}$, LARS forms a regression problem with a reduced size, which consequently decreases the number of terms of the gPC expansion K+1.

Basis-adaptive procedure

So far, truncation strategies for the gPC expansion require a predestined criteria p, but no general conclusions have been stated on how to properly assign a value to the strongly problem-dependent value p, although it was claimed by some authors that the gPC expansion with p=2 is usually accurate for estimating the first two statistical moments of a stochastic response [25]. Instead of setting a fixed value beforehand, the basis-adaptive procedure intends to start from a range of candidates, and then automatically choose the best one via an accuracy assessment of the current gPC expansion. As one of the merits of regression methods, the posterior error can be readily evaluated without extra model evaluations, which is a sound option to assess the approximation accuracy of the gPC expansion.

Among those regression methods, it is possible to find a polynomial function that exactly satisfies the relationship between distinct ED $\mathbf{X}=\{\mathbf{x}_1,\ldots,\mathbf{x}_N\}$ and the resulting model evaluations $\mathbf{Y}=\{f(\mathbf{x}_1),\ldots,f(\mathbf{x}_N)\}^{\mathrm{T}}$. However, the function which tries to maximize its efficacy on a specific ED, is likely to have a poor performance on a new set of data. To make a regression model having an acceptable accuracy of capturing the structure of available data on one hand, and being smart enough to avoid over-fitting on the other hand, the relative leave-one-out (LOO) cross-validation error e_{LOO} has been adopted in the machine learning widely. In the context of the gPC expansion, the e_{LOO} is given by

$$e_{LOO}=\frac{\sum_{l=1}^{N}(f(\mathbf{x}_l)-f_{-l}^{gPC}(\mathbf{X}\setminus\mathbf{x}_l))^2}{\sum_{l=1}^{N}(f(\mathbf{x}_l)-\frac{1}{N}\sum_{l=1}^{N}f(\mathbf{x}_l))^2} \tag{13}$$

in which $f_{-l}^{gPC}(\mathbf{X}\setminus\mathbf{x}_l)$ stands for the gPC expansion of $f(\boldsymbol{\xi})$ with specific ED $\mathbf{X}\setminus\mathbf{x}_l=\{\mathbf{x}_k,\ k=1,\ldots,N,\ k\neq l\}$. Given a target threshold and a range of possible values of criteria p, the basis-adaptive procedure selects the best candidate in terms of the lowest e_{LOO}.

B. The Copula Theory

In principle, the gPC expansion can be used to approximate the stochastic response of independent random input variables only. To apply the gPC expansion in case of correlated inputs, several methods have been proposed in the literature, such as linear transformations, proper orthogonal decomposition or K-L expansion [26]-[27]. However, all of these methods tackle the correlation problem by using transformations to remove the linear correlations.

Linear correlation is able to describe the linear or monotonic dependence structure well, but not a satisfactory measure of the nonlinear or non-monotonic dependence. However, the Copula theory enables a more general and flexible modelling of the dependence structure [28]. According to the Sklar's theorem, any joint CDF can be presented by univariate functions and an extra function that characterizes their dependences,

$$F_{\boldsymbol{\xi}}(\xi)=C(F_{\xi_1}(\xi_1),\cdots,F_{\xi_M}(\xi_M)) \tag{14}$$

where $\boldsymbol{\xi}=\{\xi_1,\ldots,\xi_M\}$is the M-dimensional random input vector, left-hand side of the above equation is the marginal distribution of the i-th random variable, and $C(.)$ is called the parametric copula function. Various copula functions have been proposed in literature, and the Gaussian copula is used in this work.

III. THE PROPOSED APPROACH FOR PROBABILISTIC THREE-PHASE POWER FLOW

In this section, we intend to explain the whole procedure of the proposed method for PPF analysis. In addition to the PPF based on the BASPC expansion, basic formulations of the DPF and the procedure of PPF based on LHS are presented as well.

A. DPF Calculation in Power Systems

The power flow analysis is one of the most commonly used tools in power systems to diagnose whether the state variables within the normal operational scope, and strategically planning for the grid. Concerning the case of single phase, the DPF problem can be introduced by the following equations as [29],

$$\begin{cases} P_i = V_i \sum_{i=1}^{n} V_j (G_{ij} \cos\theta_{ij} + B_{ij} \sin\theta_{ij}), \ i = 1, \cdots, n \\ Q_i = V_i \sum_{i=1}^{n} V_j (G_{ij} \sin\theta_{ij} - B_{ij} \cos\theta_{ij}), \ i = 1, \cdots, n \end{cases} \quad (15)$$

where P_i and Q_i are the active and reactive power injections at Bus i; V_i and θ_i are the magnitude and angle of the voltage at Bus i; G_{ij} and B_{ij} denote the conductance and susceptance of the branch connecting Bus i and Bus j, respectively, while the angle difference between the two involved buses is represented in the form of $\theta_{ij} = \theta_i - \theta_j$. In addition, n represents the number of buses in the analyzed power system.

B. The Procedure of PPF based on BASPC-LHS

The DPF captures steady-state operating characteristics of the power system consisting of state variables, inputs and outputs. In our case, state variable include the voltage at each bus; operating parameters of the system are regarded as the inputs, such as network parameters, power demand of the loads, and power injection of generators. Further, the outputs could be a vector of interest, e.g. power flows and currents through the lines. In most real cases, inputs of power flow are random, thus motivating the PPF analysis.

In this work, a novel method of PPF combining the BASPC expansion and LHS sampling, i.e., the BASPC-LHS based PPF is introduced. The detailed procedure is shown below:

Procedure 1. PPF with BASPC-LHS

Pre-Process

1. Read the parameter of random inputs, including the number, distribution types and parameters;
2. Generate a $\boldsymbol{M \times N_{sam}}$ sampling matrix $\mathbf{H_1}$ for the random inputs by LHS;
3. De-correlate the random inputs by Copula if correlated;
4. Generate a $\boldsymbol{M \times N_{ed}}$ matrix $\mathbf{H_2}$ of inputs as the ED;
5. Execute a batch of DPFs, with deterministic values of the experimental matrix $\mathbf{H_2}$ substituting for the random inputs column by column;
6. Build the sparse gPC representation, i.e., surrogate model of each output via the BASPC expansion;
7. Execute sampling matrix $\mathbf{H_1}$ through each surrogate model;

Post-Process

8. Get post-process results of each output.

C. Comparison of PPFs based on BASPC-LHS and LHS

Regarding comparison between two different PPF methods, it is common to involve simulation methods (such as MC) in evaluating the other PPF methods. In recent years, an attempt has been made on the usage of new random sampling methods that provide more efficient approaches to generate random trials. As a more straightforward sampling technique, LHS has been found to perform more efficiently than MC with applications to power system analysis [30]. In this paper, PPF based on LHS is regarded as the reference method, and then comparisons are made between BASPC-LHS and LHS in terms of accuracy and efficiency. Instead of eight steps in the BASPC-LHS based method, the procedure of applying LHS to PPF requires less steps as given below:

Procedure 2. PPF with LHS

Pre-Process

1. Read the parameters of random inputs, including the number, distribution types and parameters;
2. Generate a $\boldsymbol{M \times N_{sam}}$ sampling matrix $\mathbf{H_1}$ for the random inputs by LHS;
3. Execute a batch of DPFs, with deterministic values of the sampling matrix $\mathbf{H_1}$ substituting for the random inputs column by column;

Post-Process

4. Get post-process results of each output.

Concerning **Procedure 1** and **2**, both of them comprise two stages: pre-process and post-process. However, the purposes and computational effort of these two stages in the two methods are quite different. In the BASPC-LHS based PPF, the main objective is to build the surrogate model of each output by means of the BASPC expansion with results of $\boldsymbol{N_{ed}}$ rounds of DPF at the pre-process stage. Then, target outputs of PPF are calculated via statistical inference on outcomes obtained by executing $\boldsymbol{N_{sam}}$ samplers on the available surrogate models. Unlike the first procedure, LHS based PPF straightly executes $\boldsymbol{N_{sam}}$ rounds of DPFs at the pre-process stage. Note that representative statistics of each output can be captured as byproducts in **Procedure 1** as mentioned in (3). The block diagram of these two methods is shown in Fig.1.

The total computational time $\boldsymbol{T_{total}}$ required for the entire procedure of PPF is the summation of times consumed in both pre-process and post-process, represented by $\boldsymbol{t_{pre}}$ and $\boldsymbol{t_{post}}$. In the BASPC-LHS based PPF, $\boldsymbol{t_{pre}}$ equals to $\boldsymbol{t_{sam} + N_{ed} \times t_0 + t_{cal}}$, in which $\boldsymbol{t_{sam}}$ is the time for generating $\mathbf{H_2}$, $\boldsymbol{t_0}$ is the simulation time of a single round DPF; $\boldsymbol{t_{cal}}$ stands for the time of generating $\mathbf{H_1}$, building surrogate models, and executing $\mathbf{H_1}$ through surrogate models. In the LHS based PPF, $\boldsymbol{t_{pre} = t_{sam} + N_{sam} \times t_0}$. Note that $\boldsymbol{N_{ed}}$ is much less than $\boldsymbol{N_{sam}}$, thus, the BASPC-LHS based method is able to alleviate the computational burden of the LHS-based PPF dramatically, considering $\boldsymbol{t_{cal}}$ and $\boldsymbol{t_{post}}$ are much less than the simulation time of repeated DPFs.

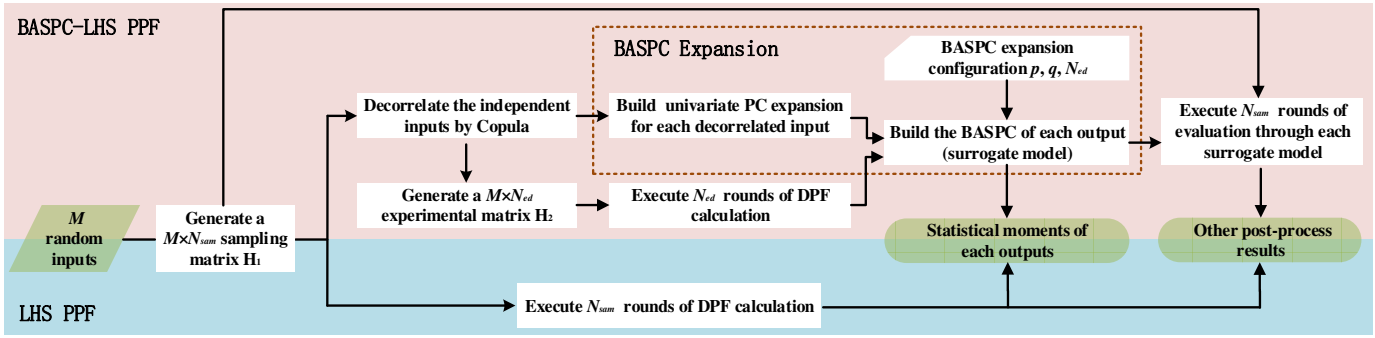

Fig.1. Block diagram of PPFs based on BASPC-LHS and LHS

IV. Test Results and Discussions

In this section the proposed method is applied in solving the PPF problem on two different power systems, in view of the procedure as detailed explained in Section III. The whole work is carried out in MATLAB R2014a; the computer used for simulation is equipped with Intel Core 4 Quad CPU at 3.40 GHz with 8 GB RAM. Simulation tool OpenDSS [31] is used to calculate the DPF, and toolbox UQLab [32] from ETH Zurich is adopted for construction of the BASPC expansion.

A. Test Setups

1) Systems under study

In this work, performance of the proposed method is firstly tested on the IEEE European Low Voltage Test Feeder. It is a typical three-phase low-voltage (LV) feeder in Europe, which is a radial distribution system with a base frequency of 50 Hz. Through a transformer at substation, voltage (phase-to-phase) level of the main feeder and laterals is stepped down from 11 kV to 416 V. On this feeder, there are 906 buses and 55 residential loads in total. Each load is connected to the grid by means of single-phase connection as shown in Fig.2.

The external medium-voltage (MV) grid is modelled as a voltage source with an impedance; and data for the transformer, lines and loads are given in [33]. Moreover, in order to analyze the influence of RES, 30 roof-top PVs are randomly installed at buses with loads on this feeder as one test scenario. All of the PVs work in unity power factor with output active power in the range of 1-5 kW. Besides, the connection of each PV in terms of phase follows the same one as the existing base load.

For the sake of simplicity, all of the loads are modelled as constant power in this LV feeder. As mentioned in Section I and II, either load demands or PV generations are subject to uncertainties, thus, they are represented by random variables. Further, some random variables are likely to be independent, since outputs of PV on this LV feeder vary in a correlated manner due to quite similar geographical and environmental conditions, such as solar irradiation and wind speed.

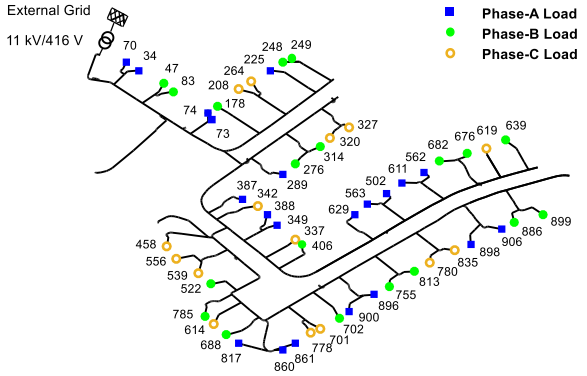

Fig.2. Single-line diagram of the IEEE European LV Test Feeder

Aiming at verifying the feasibility of the proposed method in other power systems, IEEE 123 Node Test Feeder [33] is adopted in this paper as well. This feeder is an unbalanced 4.16 kV distribution system which contains both overhead and underground line segments. In addition, different types of load (constant impedance, constant power and constant current) are dispersed on all three-phase, two-phase or single-phase laterals.

For these two distribution systems, five test scenarios have been designed to verify the BASPC-LHS based PPF, details of uncertain inputs can be found from TABLE I.

TABLE I
Uncertain Inputs of PPF on Two Test Systems

System	Number	Source	Type
IEEE European LV Test Feeder	**55**	Active power of 55 loads	Gaussian
	110	Active power of 55 loads	Gaussian
		Power factor of 55 loads	Uniform
	140	Active power of 55 loads	Gaussian
		Power factor of 55 loads	Uniform
		Active power of 30 PVs	Uniform
IEEE 123 Node Test Feeder	**140**	Active power of 70 loads	Gaussian
		Reactive power of 70 loads	Uniform
	182	Active power of 91 loads	Gaussian
		Reactive power of 91 loads	Uniform

2) Evaluation indexes

To assess the accuracy of the BASPC-LHS based method, the results obtained from LHS based PPF with 10^5 samplers are regarded as the reference herein. Different evaluation indexes are used to compare two kinds of PPF output. The first type is quantitative results, namely, mean value, variance, skewness and kurtosis of the desired outcomes. In this case, the precision level is evaluated via the relative error (*RE*) given by,

$$RE(r^{ref}, r^{cp}) = \left| 1 - r^{cp} \middle/ r^{ref} \right| \tag{16}$$

in which r^{ref} is the result of LHS based PPF while r^{cp} is the outcome of BASPC-LHS based method. *RE* is able to give an indication of how good a result (method) is, relative to the given reference.

The second type is the graphical results, such as frequency histograms after the post-processing. In this case, the similarity index (*SI*) is proposed by modifying the city block distance,

$$SI(r^{ref}, r^{cp}) = 1 - \frac{1}{2}\sum_{i=1}^{n} \sqrt{[(c_i^{ref} - c_i^{cp})^2 + (d_i^{ref} - d_i^{cp})^2]} \tag{17}$$

in which n is the number of bins, c_i corresponds to the location of the i-th bin center on the x-axis, and d_i is the height of the i-th rectangular bin on the y-axis that indicates percentage of the number of elements in the i-th bin. Note that this index makes sense only if the number of samples and bins in the two frequency histograms are identical to each other. Obviously, a greater value of the index indicates a better fit between the two frequency histograms. For the sake of conciseness, similarity index in (17) can be further simplified by locating bin centers at the same position. Therefore, $c_i^{ref} = c_i^{ref}$ and $SI(r^{ref}, r^{cp})$ ranges between 0 and 1. By means of this metric, the similarity degree of two frequency histograms can be qualitatively detected.

Apart from the accuracy, the efficiency is another concern of this work. With the percentage reduction in computation time *PD* as defined in (18), merits of BASPC-LHS based PPF with respect to the efficiency become visible,

$$PD(t^{ref}, t^{cp}) = (t^{ref} - t^{cp}) / t^{ref} \tag{18}$$

where t^{ref}, t^{cp} refer to the time required to execute PPF by means of LHS and BASPC-LHS, respectively.

B. Evaluation on Surrogate Model

Concerning the IEEE European LV Test Feeder, load shapes with one-minute resolution over one day are provided in [33]. In the subsection *B* and *C* of Section IV, PPFs at the on-peak moment, i.e., t=09:26:00 are implemented. In this subsection,

accuracy of the surrogate model in BASPC-LHS based method is evaluated.

As the ***base scenario***, there are 55 independent uncertainty sources, and all loads maintain a constant power factor 0.95. For simulating the stochastic behavior of all the 55 loads, their active powers are assumed to follow Gaussian distribution $\boldsymbol{N}(PL_i, 0.1^2)$, where PL_i is the original active power demand of the i-th (i=1,…,50) load at t=09:26:00. As introduced in Section II, configuration of the BASPC expansion depends on several parameters, mainly on truncation norm q, range of truncation criteria p, and number of ED N_{ed}. In order to investigate impacts of the configuration on performances of the surrogate model which influences the result of BASPC-LHS based PPF, these three parameters are set to different values. Moreover, results of the LHS based PPF with 10^5 trials are set as the reference; and the BASPC expansion is regarded as the ***base configuration***, with truncation norm 0.8, range of truncation criteria 1-6 and number of ED 800.

Taking the voltage magnitude of phase A at Bus 899 for illustration, TABLE II summarizes the role of each parameter plays on the proposed method's behavior in terms of relative e_{LOO} and t_{pre}. From TABLE II, it can be concluded that behavior of the surrogate model in terms of both two indicators changes if there is any variation of q, p or N_{ed}. The value of e_{LOO} will be increased in the wake of reduction of the truncation norm, or shrinkage of the range of truncation criteria, or abatement of the number of ED. The reason is that the surrogate model obtained from the BASPC expansion is weakened to approximate the real behavior of voltage magnitude, due to changes in the configuration. On the contrary, t_{pre} will decrease since less computational efforts are required to get the surrogate model. In view of both accuracy and efficiency, the BASPC expansion with truncation norm 0.8, range of truncation criteria 1-6, number of ED 250 is named as the ***efficient configuration*** which will be adopted to implement the majority of subsequent tests. The surrogate model obtained by expansion with the ***efficient configuration*** is noted as ***efficient surrogate model***.

TABLE II
INFLUENCE OF BASPC CONFIGURATION ON SURROGATE MODEL

BSAPC Configuration			e_{LOO}	t_{pre}
q	p	N_{ed}	(p.u.)	(s)
0.8	1-6	800	8.81×10^{-7}	58.76
0.4	1-6	800	1.45×10^{-6}	29.16
0.8	**1-2**	800	1.45×10^{-6}	28.17
0.8	1-6	**250**	2.36×10^{-6}	9.13

Except for internal configuration of the BASPC expansion, there are several external factors also exert influences on the performance, which will be discussed in the next subsection, by taking the phase-A voltage magnitude at Bus 899 for instance as well. The first external factor is the probability distribution type of uncertain inputs. Apart from the Gaussian distribution, demand of active power may follow some other distributions such as Uniform and Triangle distribution, depending on the amount of available information. Therefore, Uniform random variables $\boldsymbol{U}(0.9\times PL_i, 1.1\times PL_i)$ (i=1,…,50) is analyzed as the second type while the mixture of 30 Gaussian random inputs $\boldsymbol{N}(PL_i, 0.1^2)$ (i=1,…,30) and 25 Uniform random inputs $\boldsymbol{U}(0.9\times PL_i, 1.1\times PL_i)$ (i=31,…,55) as the third type. With the ***efficient configuration*** of the BASPC expansion, TABLE III shows that evaluations on surrogate models in the presence of three types of uncertain inputs. It indicates that the accuracy of three surrogate models vary on the same order of magnitude; and differences in calculation time are less than 5 seconds.

TABLE III
INFLUENCE OF INPUTS' TYPE ON SURROGATE MODEL

Uncertain Inputs			e_{LOO}	t_{pre}
Type	Number	Correlation	(p.u.)	(s)
Gaussian	55	No	2.36×10^{-6}	9.13
Uniform			4.11×10^{-6}	13.55
Mixed			3.72×10^{-6}	10.25

Besides the distribution type of random inputs, the second external factor is the dimension, i.e., the number of uncertain sources. The ability of dealing with the high dimensionality of random input variables is a considerable merit of the BASPC expansion compared to the traditional PC expansion, thus it is important to discuss the effect of dimensionality. By taking into account 55 Gaussian distributed active power demands and 55 Uniform distributed power factors $\boldsymbol{U}(0.85, 0.95)$, there are 110 random inputs. Moreover, the total number of uncertainties comes up to 140 if there are 30 PV generations penetrated into this feeder, and all of the active power outputs follow the Uniform distribution $\boldsymbol{U}(0.9\times PV_i, 1.1\times PV_i)$, where PV_i stands for the original active power production of the i-th (i=1,…,30) PV at t=09:26:00, reffering to [34]. In these two new scenarios, the first four rows of TABLE IV lists the comparative results, which shows the accuracy of the surrogate model decreases briefly, and the calculation time increases slightly, along with the increase of dimensionality. But even so, there is room to improve the behavior of the surrogate model by adjusting the configuration of BASPC expansion if the number of uncertain inputs grows to a great extent.

As mentioned in Section II, statistical correlation among random inputs can be tackled via the Copula, thus impacts of the correlation condition on the surrogate model will also be considered. Assuming correlations among outputs of 30 PVs are 0.8, the bottom row of TABLE IV lists the two evaluation indicators. It is noticed that the correlation condition does not decrease the accuracy of the surrogate model but extends the computation time.

TABLE IV
INFLUENCE OF INPUTS' NUMBER AND CORRELATION ON SURROGATE MODEL

Uncertain Inputs			e_{LOO}	t_{pre}
Type	Number	Correlation	(p.u.)	(s)
Mixed	**55**	No	3.72×10^{-6}	10.25
	110	No	3.67×10^{-3}	10.38
	140	No	3.90×10^{-3}	11.77
		Yes	3.90×10^{-3}	13.09

Until now, only the voltage magnitude has been discussed; there are other desired quantities in practice though. Take the voltage angle and power loss for illustration, BASPC expansion with the ***efficient configuration*** is obtained to investigate the impact of output's type in the ***base scenario***. TABLE V shows values of indicators with regard to phase-A voltage magnitude at Bus 899, current of line 78 and active power loss of the whole feeder. It illustrates that both accuracy and efficiency of the surrogate model changes when different outputs are considered. The BASPC expansion performs better on the bus voltage than the line current and the active power loss, which reveals that the performance of the surrogate model will be influenced by the computational complexity of the output of interest. In this case,

extra computation effort is required to get the value of power loss after the PPF calculation while the voltage and current are the direct output of running PPF.

TABLE V
INFLUENCE OF OUTPUTS' TYPE ON SURROGATE MODEL

	Voltage	Current	Power Loss
e_{LOO} (p.u.)	2.36×10^{-6}	2.08×10^{-6}	3.97×10^{-4}
T_{pre} (s)	9.13	9.70	10.03

C. Evaluations on Post-process Results

As explained in Section III, several post-process results can be obtained either through the LHS based PPF or the BASPC-LHS based method by means of LHS samplers. In this subsection, post-process results of the voltage magnitude at Bus 899 are used to demonstrate the comparative tests. In the ***base scenario*** mentioned before, the LHS based PPF adopts 10^5 trials to execute a batch of DPFs on the original feeder, thus the BASPC-LHS based method is going to use the same trials for post-processing on the ***efficient surrogate model***. Fig.3 shows frequency histograms of three-phase voltage magnitudes at Bus 899. On one hand, it demonstrates that voltage magnitudes are different among phase A, B and C, due to unbalanced power demands among three phases. On the other hand, the two groups of results from two different approaches coincide with each other well from visual point of view. In view of these three pairs of frequency histogram, TABLE VI shows that *SI* values for all of the three phases are above 99%.

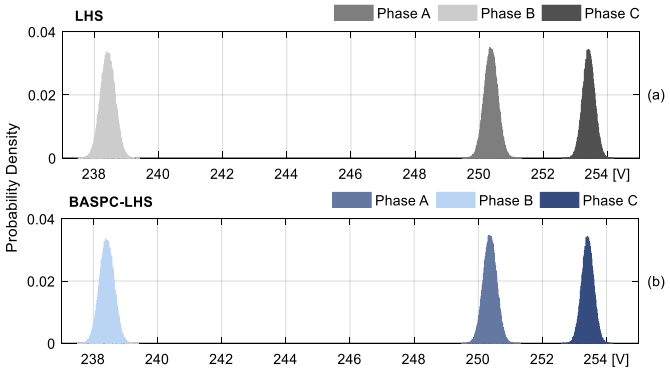

Fig.3. Frequency histogram of three-phase voltage magnitudes at Bus 899

TABLE VI
SI EVALUATION OF THREE-PHASE VOLTAGE MAGNITUDES AT BUS 899

	Phase A	Phase B	Phase C
SI (%)	99.84	99.83	99.87

Indeed, factors that influence performance of the surrogate model in the BASPC-LHS based PPF will bring impacts on the post-process results as well. Consider different numbers of ED, comparative results of the PDFs, CDFs and statistical moments will come to the same conclusions. For illustration, phase-A voltage magnitude at Bus 899 is selected as the target variable. Fig. 4 illustrates four frequency histograms obtained by LHS based PPF and BASPC-LHS based method with truncation norm 0.8, range of truncation criteria 1-6 and different numbers of ED. It is noticed that the frequency histogram obtained by the BASPC-LHS based PPF with 50 N_{ed} does not follow the reference. However, the results of BASPC-LHS based method coincide with the reference result well when N_{ed} comes up to 250. Further, TABLE VII presents five quantitative indicators to reveal the influence of N_{ed} on the post-process results of the BASPC-LHS based PPF.

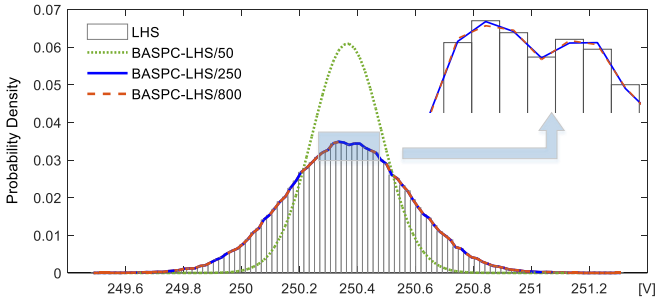

Fig. 4. Frequency histogram of phase-A voltage magnitude at Bus 899

TABLE VII
INFLUENCE OF THE NUMBER OF ED ON SI AND REs

N_{ed}	*SI* (%)	*RE*			
		Mean	Variance	Skewness	Kurtosis
50	73.76	3.69×10^{-7}	6.73×10^{-1}	9.91×10^{-1}	5.00×10^{-2}
250	99.84	3.67×10^{-8}	1.85×10^{-4}	7.50×10^{-1}	8.59×10^{-6}
800	99.85	3.31×10^{-8}	3.00×10^{-4}	4.56×10^{-1}	7.59×10^{-6}

In statistics, the cumulative distribution gives the probability that a random variable is less than or equal to a predefined threshold, or lies in a semi-closed interval. Fig. 5 plots the cumulative distribution curves of phase-A voltage magnitude at Bus 899 which are obtained by different methods. Besides, the values of $P(U_A \le 250)$, $P(250.4 < U_A \le 250.8)$ and $P(U_A > 250.8)$ are compared in TABLE VIII. It is obvious that BASPC-LHS based method with the ***efficient configuration*** performs great while the results are the same as the reference with the ***base configuration***.

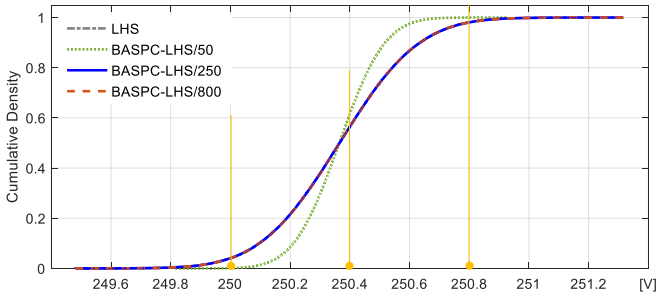

Fig. 5. Cumulative Distribution of phase-A voltage magnitude at Bus 899

TABLE VIII
INFLUENCE OF THE NUMBER OF ED ON CUMULATIVE DENSITY

N_{ed}	*Cumulative Density* (%)		
	$U_A \le 250$	$250.4 < U_A \le 250.8$	$U_A > 250.8$
50	0.12	38.41	0.01
250	4.07	41.60	1.92
800	4.07	41.58	1.93
LHS	**4.06**	**41.57**	**1.94**

* U_A is the random variable, referring to phase-A voltage magnitude.

Up to now, scalar-valued target outputs have been discussed. In case that multi-component outputs are needed, BASPC-LHS based PPF will build a specific surrogate model for each output via independent BASPC expansions on the shared ED.

D. Discussion on Computational Effort

By comparative tests on both quantitative indicators and graphic results, it is proved that the BASPC-LHS based method is very accurate relative to the LHS based PPF. Moreover, it should be emphasized that the proposed approach is much more efficient than the LHS based one, which relieves computational

effort and saves the total simulation time. Considering BASPC-LHS based method with the ***efficient configuration*** and LHS based PPF with 10^5 trials, only 250 rounds of DPF are required to get the quantitative results in the former method, while 10^5 rounds of DPF are needed in the latter one. Regarding post-process results, although the same amount of trials are adopted in the former method, but those samplers are only used to make statistical inference of the PPF output via its surrogate model, which is constructed by the BASPC expansion with 250 ED. Compared to the computational effort for executing $\mathbf{10^5}$ rounds of DPF calculation, the effort paid to calculate statistical values of the surrogate model of target output with $\mathbf{10^5}$ samplers is much less and close to negligible.

As mentioned in Section III, the total time T_{total} consumed by BASPC-LHS based method includes the time for pre-process and post-process. TABLE IX lists the total computational time of PPF in the ***base scenario*** for both two approaches. It is found that BASPC-LHS method with the ***efficient configuration*** only consumed 9.84 s while the LHS based PPF required 3562.03 s. That is, at the on-peak moment, BASPC-LHS based PPF is able to save 3552.19 s while remaining high accuracy. Nevertheless, the advantage of the proposed method in term of efficiency becomes more significant if time-series PPFs are carried out. By implementing BASPC-LHS based method with the ***efficient configuration*** and LHS based PPF with 10^5 trials, Fig. 6 shows the time-series simulation results of phase-A voltage magnitude at Bus 899 with one-hour resolution over one day. It can be found that the results obtained from these two methods are quite close over the whole horizon. However, the total time (without parallel computing) required in LHS based PPF is 82617 s while 202.692 s in the proposed method.

TABLE IX
COMPUTATIONAL BURDEN OF PPFS ON THE IEEE EUROPEAN LV TEST FEEDER

	BASPC-LHS				**LHS**	
N_{ed}	250		800		0	0
N_{sam}	10^4	10^5	10^4	10^5	10^4	10^5
t_{pre} (s)	9.13	9.13	58.76	58.76	0.04	0.17
t_{post} (s)	0.36	0.71	0.47	0.87	342.23	3561.86
T_{total} (s)	9.49	**9.84**	59.23	59.63	342.27	**3562.03**

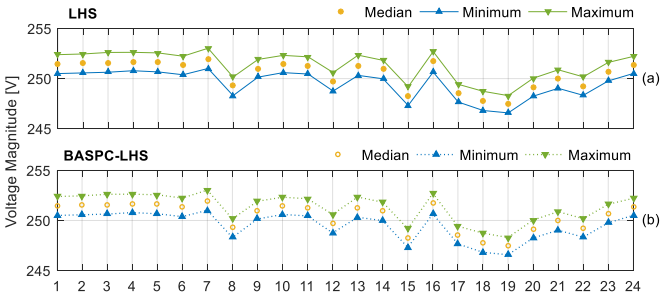

Fig. 6. Time-series phase-A voltage magnitudes at Bus 899

In addition to the IEEE European LV Test Feeder, IEEE 123 Node Test Feeder is also adopted to test the proposed method. TABLE X shows performance of the BASPC-LHS based PPF on the IEEE 123 Node Test Feeder in presence of 140 and 182 independent random inputs, with the ***efficient surrogate model***. In case of 140 uncertainties, it is noticed that the BASPC-LHS based method requires more computational effort in the IEEE European LV Test Feeder than the IEEE 123 Node Test Feeder. The reason lies in that complexity of the target power system significantly influences t_0. Moreover, it is found that the size and topology of target power system makes a bigger difference to T_{total}, relative to the number of uncertainties.

TABLE X
COMPUTATIONAL BURDEN OF PPFS ON THE IEEE EUROPEAN LV TEST FEEDER AND THE IEEE 123 NODE TEST FEEDER

System	M	**Method**	N_{ed}	N_{sam}	T_{total} (s)
IEEE 123 Node Test Feeder	**140**	BASPC-LHS	250	10^5	**6.25**
		LHS	--		1206.31
	182	BASPC-LHS	250		7.26
		LHS	--		1235.78
IEEE European LV Test Feeder	55	BASPC-LHS	250		9.84
		LHS	--		3562.03
	140	BASPC-LHS	250		**13.81**
		LHS	--		3597.36

TABLE XI gives the overall comparison of performances between the two methods on both test systems. It can also be concluded that the more complex power system is, the more computational effort can be saved by the proposed method. In subsection B, it has been found that correlation condition decreases accuracy of the surrogate model. With respect to the overall performance, the correlation condition will also reduce values of *PD* and *SI*, as listed in the last row of TABLE XI.

TABLE XI
OVERALL COMPARISON OF PPFS BASED ON BASPC-LHS AND LHS

System	M	**Correlation**	***PD* (%)**	***SI* (%)**
IEEE 123 Node Test Feeder	**140**	No	**99.48**	**99.52**
	182	No	99.41	99.43
IEEE European LV Test Feeder	55	No	99.72	99.84
	140	No	**99.62**	**99.25**
		Yes	**99.56**	**98.95**

In practice, the computational effort of PPF relies on several other aspects beyond the scope of this work, such as the concision of DPF algorithm, the efficiency of programming language and the level of integration among each participated session. Nevertheless, comparisons between two different PPF methods presented in this work are implemented in the same environment.

V. CONCLUSIONS

This paper presents a novel method of PPF based on the BASPC expansion to tackle a large number of either dependent or correlated random input variables. The main objective of this work is to verify accuracy and efficiency of the BASPC-LHS based PPF, compared to the LHS based PPF.

Different BASPC configurations are tested to solve the PPF problem, with respect to truncation norm q, range of truncation degree p, number of ED N_{ed}. Besides, number of random inputs M, form of inputs' PDF, type of outputs, and number of samplers for post-processing N_{sam} were discussed as well. In the presence of nonlinear correlation among random inputs, the Copula is applied to remove the correlation. Performances of the proposed method were tested on IEEE European LV Test Feeder and IEEE 123 Node Test Feeder. Either quantitative indicators or graphical results prove that, the BASPC-LHS based PPF improves the computational efficiency dramatically whilst maintaining high accuracy, relative to traditional PPFs.

Merits of the proposed method will be of benefit to studies of complex power systems in presence of uncertainties, especially for online usages. Future work will broaden applications of this method, such as network planning, load management and so on.

VI. References

[1] R. Doherty, M. O. Malley, "A new approach to quantify reserve demand in systems with significant installed wind capacity," *IEEE Trans. on Power Systems*, vol.20, no.2, pp.587-595, 2005.

[2] H. Bludszuweit, J. A. Domínguez-Navarro, "A probabilistic method for energy storage sizing based on wind power forecast uncertainty," *IEEE Trans. on Power Systems*, vol.26, no.3, pp.1651-1658, 2011.

[3] X.F. Wang, Y.H. Song, and M. Irving, *Modern power systems analysis*. Springer Science & Business Media, 2010.

[4] H.X. Yang, L. Lu, and J. Burnett, "Weather data and probability analysis of hybrid photovoltaic–wind power generation systems in Hong Kong." *Renewable Energy*, vol.28, no.11, pp.1813-1824, 2003.

[5] B. Borkowska, "Probabilistic load flow," *IEEE Trans. Power App. Syst*., vol. PAS-93, no. 3, pp. 752–759, May 1974.

[6] S.M.P. Ribeiro, et al., "Probabilistic load flow techniques applied to power system expansion planning," *IEEE Trans. on Power Systems*, vol.5, no.4, pp.1047-1053, 1990.

[7] Z.C. Hu, X.F. Wang, "A probabilistic load flow method considering branch outages," *IEEE Trans. on Power Systems*, vol.21, no.2 pp.507-514, 2006.

[8] M. Fan, et al., "Probabilistic power flow analysis with generation dispatch including photovoltaic resources," *IEEE Trans. on Power Systems*, vol.28, no.2, pp.1797-1805, 2013.

[9] P.Y. Chen, C. Zhe, and B. Bak-Jensen, "Probabilistic load flow: A review," *Third International Conference on Electric Utility Deregulation and Restructuring and Power Technologies, IEEE*, pp.1586-1591, 2008.

[10] M. Hajian, W.D. Rosehart, & H. Zareipour, "Probabilistic power flow by Monte Carlo simulation with latin supercube sampling," *IEEE Trans. on Power Systems*, vol.28, no.2, pp.1550-1559, 2013.

[11] P. Zhang, S.T Lee, "Probabilistic load flow computation using the method of combined cumulants and Gram-Charlier expansion," *IEEE Trans. on Power Systems*, vol.19, no.1, pp.676-682, 2004.

[12] M. Fan, et al., "Probabilistic power flow studies for transmission systems with photovoltaic generation using cumulants," *IEEE Trans. on Power Systems*, vol.27, no.4, pp.2251-2261, 2012.

[13] J.M. Morales, J. Perez-Ruiz, "Point estimate schemes to solve the probabilistic power flow." *IEEE Trans. on Power Systems*, vol.22, no.4, pp.1594-1601, 2007.

[14] X.M. Ai, et al., "A discrete point estimate method for probabilistic load flow based on the measured data of wind power," *IEEE Trans. on Industry Applications*, vol.49, no.5, pp. 2244-2252, 2013.

[15] J. Tang, F. Ni; F. Ponci; A. Monti, "Dimension-Adaptive Sparse Grid Interpolation for Uncertainty Quantification in Modern Power Systems: Probabilistic Power Flow," *IEEE Trans. on Power Systems*, vol.31 no.2, pp.907-919, 2015.

[16] Z. Ren, W. Li, R. Billinton, W. Yan, "Probabilistic Power Flow Analysis Based on the Stochastic Response Surface Method," *IEEE Trans. on Power Systems*, vol.31, no.3, PP. 2307-2315, 2016.

[17] D.B. Xiu, G.E. Karniadakis, "The Wiener--Askey polynomial chaos for stochastic differential equations," *SIAM journal on scientific computing*, vol.24, no.2, pp.619-644, 2002.

[18] C. Hu, B.D. Youn, "Adaptive-sparse polynomial chaos expansion for reliability analysis and design of complex engineering systems," *Struct. Multidiscipl. Optim.*, vol.43, no.3, pp.419-442, 2011.

[19] S. S. Isukapalli, A. Roy, and P.G. Georgopoulos, "Stochastic response surface methods (SRSMs) for uncertainty propagation: application to environmental and biological systems," *Risk analysis*, vol.18, no.3, pp.351-363, 1998.

[20] G. Blatman, "Adaptive sparse polynomial chaos expansions for uncertainty propagation and sensitivity analysis," Doctoral dissertation, Clermont-Ferrand 2, 2009.

[21] E. Haesen, et al., "A probabilistic formulation of load margins in power systems with stochastic generation," *IEEE Trans. on Power Systems*, vol.24, no.2, pp.951-958, 2009.

[22] K.Y. Wang, G.J. Li, X.C. Jiang, "Applying probabilistic collocation method to power flow analysis in networks with wind farms," *Power and Energy Society General Meeting (PES), 2013 IEEE*, vol., no., pp.1-5, 21-25 July 2013.

[23] G. Papaefthymiou, D. Kurowicka, "Using copulas for modeling stochastic dependence in power system uncertainty analysis," *IEEE Trans. on Power Systems*, vol.24, no.1, pp.40-49, 2009.

[24] D.F. Cai, D.Y. Shi, J.F. Chen, "Probabilistic load flow computation using Copula and Latin hypercube sampling," *IET Generation, Transmission & Distribution*, vol.8, no.9, pp.1539-1549, 2014.

[25] T.E. Lovett, F. Ponci, A. Monti, "A polynomial chaos approach to measurement uncertainty," *IEEE Trans. on Instrumentation and Measurement*, vol.55, no.3, pp.729-736, 2006.

[26] M. Aien, M. Fotuhi-Firuzabad, F. Aminifar, "Probabilistic load flow in correlated uncertain environment using unscented transformation," *IEEE Trans. on Power Systems*, vol.27, no.4, pp.2233-2241, 2012.

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

[28] P. Jaworski, F. Durante, W.K. Hardle, T. Rychlik, *Copula theory and its applications*. New York, Springer, 2010.

[29] X. Li, D.J., Du, J.X. Pei and MI. Menhas, "Probabilistic load flow calculation with Latin hypercube sampling applied to grid-connected induction wind power system," *Transactions of the Institute of Measurement and Control*, vol. 35, pp. 56-65, 2011.

[30] Z. Shu, P. Jirutitijaroen, "Latin hypercube sampling techniques for power systems reliability analysis with renewable energy sources," *IEEE Trans. on Power Systems*, vol.26, no.4, pp.2066-2073, 2011.

[31] Electric Power Research Institute, OpenDSS, Distribution System Simulator [Online], Available: http://sourceforge.net/projects/electricdss.

[32] S. Marelli, and B. Sudret, "UQLab: A framework for uncertainty quantification in Matlab," Vulnerability, Risk Analysis and Management, Proc. 2nd Int. Conf. on, Liverpool, 2014, 2554-2563.

[33] IEEE PES. (2016, Feb.) Distribution test feeders. [Online]. Available: http://www.ewh.ieee.org/soc/pes/dsacom/testfeeders/index.html.

[34] Electricity North West Ltd, LCT profiles [Online], Available: http://www.enwl.co.uk/about-us/the-future/lcnf-tier-1-nia/low-voltage-network-solutions

Biographies

Fei Ni received her Ph.D. degree in natural science at the Institute for Automation of Complex Power Systems of E.ON Energy Research Center at RWTH Aachen University, Aachen, Germany, in 2015.

In 2015, she joined the Electrical Energy Systems group at Eindhoven University of Technology, Eindhoven, the Netherlands, as a postdoc researcher. Her research interests include uncertainty modelling and simulation, sensitivity and reliability analysis, data mining and predictive control of electrical systems.

Phuong H. Nguyen (M'06) received his Ph.D. at the Eindhoven University of Technology, the Netherlands, in 2010.

He is an Assistant Professor in the Electrical Energy Systems group at the Eindhoven University of Technology, the Netherlands. He was a Visiting Researcher with the Real-Time Power and Intelligent Systems (RTPIS) Laboratory, Clemson University, USA, in 2012 and 2013. His research interests include distributed state estimation, control and operation of power systems and multi-agent systems, and their applications in the future power delivery system.

Joseph F.G. Cobben (M'11) received the Ph.D. degree in Electrical Engineering from Eindhoven University of Technology (TU/e), Netherlands, in 2007.

He works as a research scientist in Alliander and specialized in safety of the LV- and MV-networks, power quality and safety of installations connected to the networks. He is member of several national and international standardization committees about requirements for low and high voltage installations and characteristics of the supply voltage. Dr. Cobben is a part-time professor at TU/e with research area of smart grids and power quality, and the author of several technical books.