Development of Appropriate PPF Algorithm for Future Planning Process

Version 2.0

Deliverable

Niyam Haque

31 July 2018

ERA-Net Smart Grids Plus | From local trials towards a European Knowledge Community

This project has received funding in the framework of the joint programming initiative ERA-Net Smart Grids Plus, with support from the European Union's Horizon 2020 research and innovation programme.

INTERNAL REFERENCE

- **Deliverable No.:** D 3.1
- **Deliverable Name:** Development of appropriate PPF algorithm for future planning process
- **Lead Partner:** TU/e
- **Work Package No.:** WP3
- **Task No. & Name:** Task 3.1 - Development of appropriate PPF algorithm for future planning process
- **Document (File):** Final version D3.1.docx
- **Issue (Save) Date:** 2018-07-31

DOCUMENT SENSITIVITY

☐	**Not Sensitive**	Contains only factual or background information; contains no new or additional analysis, recommendations or policy-relevant statements
x	**Moderately Sensitive**	Contains some analysis or interpretation of results; contains no recommendations or policy-relevant statements
☐	**Sensitive**	Contains analysis or interpretation of results with policy-relevance and/or recommendations or policy-relevant statements
☐	**Highly Sensitive Confidential**	Contains significant analysis or interpretation of results with major policy-relevance or implications, contains extensive recommendations or policy-relevant statements, and/or contain policy-prescriptive statements. This sensitivity requires SB decision.

DOCUMENT STATUS

	Date	Person(s)	Organisation
Author(s)	31-07-2018	Niyam Haque	TU/e
Verification by	31-07-2018	Phuong H Nguyen	TU/e
Approval by	31-07-2018	Phuong H Nguyen	TU/e

CONTENTS

Disclaimer

The content and views expressed in this material are those of the authors and do not necessarily reflect the views or opinion of the ERA-Net SG+ initiative. Any reference given does not necessarily imply the endorsement by ERA-Net SG+.

About ERA-Net Smart Grids Plus

ERA-Net Smart Grids Plus is an initiative of 21 European countries and regions. The vision for Smart Grids in Europe is to create an electric power system that integrates renewable energies and enables flexible consumer and production technologies. This can help to shape an electricity grid with a high security of supply, coupled with low greenhouse gas emissions, at an affordable price. Our aim is to support the development of the technologies, market designs and customer adoptions that are necessary to reach this goal. The initiative is providing a hub for the collaboration of European member-states. It supports the coordination of funding partners, enabling joint funding of RDD projects. Beyond that ERA-Net SG+ builds up a knowledge community, involving key demo projects and experts from all over Europe, to organise the learning between projects and programs from the local level up to the European level.

www.eranet-smartgridsplus.eu

List of Acronyms

BASPC	Basis-Adaptive Sparse Polynomial Chaos
CDF	Cumulative Distribution Function
DER	Distributed Energy Resource
DPF	Deterministic Power Flow
ED	Experimental Design
EV	Electric Vehicle
gPC	generalized Polynomial Chaos
HP	Heat Pump
LARS	Least Angle Regression
LHS	Lating Hypercube Sampling
LOO	Leave-One Out
LSS	Lating Supercube Sampling
LV	Low Voltage
MC	Monte Carlo
MV	Medium Voltage
PC	Polynomial Chaos
PDF	Probability Distribution Function
PMF	Probability Mass Function
POC	Point of Connection
PPF	Probabilistic Power Flow
PV	Photovoltaic
RE	Relative Error
RES	Renewable Energy Source
SI	Similarity Index
UQ	Uncertainty Quantification
VUF	Voltage Unbalance Factor

List of symbols

Constants and indices

Ω	Sample space
F	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 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 samples for the LHS-based PPF
T_{total}	Total computation time of PPF
t_{pre}	Pre-processing time of PPF
t_{post}	Post-processing time of PPF
t_0	Time for a single round of DPF calculation
t_{ref}	Time required by a reference method
t_{cp}	Time taken by the method to be compared
PL_i	Original active power demand of i-th load
PV_i	Active power generation of the i-th PV system
r^{ref}	Result obtained by the reference method
r^{cp}	Result obtained by the method to be compared
c_i	Location of the i-th bin center on x-axis
d_i	Height of the i-th bin on y-axis
N^M	The complete M-dimensional sphere of natural numbers
$\|\mathbf{i}\|_q$	Rank of the multi-index $\mathbf{i}$.

Variables and sets

$L^2(\Omega, F, P)$	A finite dimensional random space
ξ	Random input vector
$f(\xi)$	The stochastic response to random variable ξ
$\omega(\xi)$	PDF of random variable ξ
$F_{\xi}(\xi)$	Joint CDF of random vector ξ
$\{\Phi_i\}$	i-th univariate polynomial basis
$\{\Psi_i\}$	i-th multivariate polynomial basis
$\{a_i\}$	Coefficient of i-th polynomial basis
$\mathbf{a}$	Vector of coefficients $\{a_j\}$
$\hat{\mathrm{a}}$	Calculated coefficient vector by regression
$\mathbf{X}$	Set of experimental design of random vector
$\mathbf{Y}$	Stochastic response vector of $\mathbf{X}$
$\mathbf{X}\backslash \mathrm{x}_l$	Set of experimental design of random vector except x_l
e_{LOO}	Relative leave-one-out cross-validation error
$f_{-l}^{gPC}(\mathbf{X}\backslash \mathrm{x}_l)$	The gPC expansion of $f(\xi)$ with $\mathbf{X}\backslash \mathrm{x}_l$

1 Introduction

Power system analysis is inherently subject to numerous sources of uncertainties. Traditionally, fluctuations in demand, generator outages, faults and failure of network components were prominent sources of power system operations. In recent years, the uncertainties have been further augmented by the proliferation of renewable energy sources (RES) based local generation units such as – wind turbines and solar photovoltaic (PV) systems, as well as by the introduction of various new forms of loads like electric vehicle (EV) and domestic heat pump (HP), leading to new network and market concepts like physical and commercial micro-grids. Consequently, uncertainty quantification (UQ) has attracted major research interest among network operators and market actors for both planning and reliable operation of the future power systems [1]–[4].

Power flow or load flow calculations are conventionally performed to determine the node voltages and branch currents under steady-state conditions. Such analyses are most commonly termed as deterministic power flow (DPF) calculations and require certain network conditions such as active power, reactive power and voltages to be known. However, a DPF calculation cannot provide a complete understanding of the network due to the stochasticity associated with the variables [4].

Probabilistic power flow (PPF) calculations were developed as an alternative to DPF to provide a more comprehensive overview of a network. PPF analyses inherently combine the principles of DPF and UQ to consider the uncertainties. Unlike DPF, PPF calculations asses system states and expected outputs in terms of probability distribution function (PDF), cumulative distribution function (CDF), probability mass function (PMF) and statistical moments [4]–[7].

Three main types of approaches of PPF analysis can be found in the literature: simulation method, analytical method and approximation method. Simulation method represents the most commonly used technique, where uncertain inputs are first converted into a set of random samples, followed by running a batch of DPF calculations on the input for the desired network. In order to ensure the convergence of the random sampling, such as brute-force Monte Carlo (MC) approaches, a large number of trials are required resulting in a huge computational burden and simulation time. A number of advanced sampling methods have been proposed to achieve similar convergence with less number of trials. Such methods include: Latin supercube sampling (LSS), Latin hypercube sampling (LHS) and importance sampling. In order to alleviate the computational burden, analytical methods have been proposed such as- multi-linear model based method, direct convolution method, Fourier transform method, cumulant method etc [8], [9]. These methods run on predefined assumptions and do not require huge number of repeated simulations. However, the assumptions may lead to inaccuracies for UQ. In contrast, approximation methods can strike the balance between the accuracy and efficiency of calculation. They require a small number of DPF calculations and statistical transformations of the input variables. A number of methods have been used to represent and evaluate uncertainties through PDFs, PMFs or CDFs, among which the stochastic collocation interpolation and the polynomial chaos (PC) expansion represent the best results [10]–[12].

Regardless of the methods chosen, two critical issues must be taken into account for PPF analysis – (i) the large number of random inputs that results in a high-dimensional UQ problem, and (ii) the correlation among random input variables. The former is a difficult to issue for both analytical and approximation methods. Conventional analytical methods need large data storage and computational efforts for PPF analysis involving huge number of random inputs. Likewise, approximation methods may need even more simulation rounds than comparable simulation methods. For the latter issue, dependent input variables are first decorrelated before using existing UQ methods [4], [13], [14].

Micro-grids are being considered as the building blocks of future energy system, hosting multitude of distributed energy resources (DERs). Planning of future power system incorporating such micro-grids needs to take numerous sources of uncertainty into account. Such large number of uncertainty sources introduces high-dimensionality to the

classical planning approaches. In this report, a novel methodology for PPF analysis is presented that can handle high dimensionality and nonlinear correlation of the input random variables. The proposed approximation method makes use of basis-adaptive sparse PC (PC) expansion and Copula theory for achieving considerable efficiency and accuracy. The remainder of the report is organized as follows: Section 2 describes the theoretical background of the basis-adaptive sparse PC expansion, Section 3 presents the formulation of the proposed approach, the simulation test setup and the performance indicators are highlighted in Section 4 and the simulation results are discussed in Section 5 before concluding with the brief summary in Section 6.

2 Theoretical background

In this section, first the background of the BASPC expansion is discussed including the generalized PC (gPC) theory and the sparse adaptive scheme. Then, the copula theory is briefly described which is used to address the nonlinear dependence between the input variables.

2.1 Basis-Adaptive Sparse Polynomial Chaos Expansion

2.1.1 Generalized Polynomial Chaos Expansion

According to the Cameron-Martin theorem, for a finite dimensional random space L^2 (Ω, 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 [4].

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,

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

where $E[.]$ is the expectation operator, δ_{ij} is the Kronecker delta function, and $\omega(\xi)$ is the PDF of ξ. Statistical moments of the stochastic response can be obtained in closed-form:

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

If the random input is a vector ξ with M independent random variables $\{\xi_1, \xi_2, \ldots, \xi_M\}$, the corresponding stochastic response is:

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

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. The infinite expression has to be truncated to a finite number of terms with a given criteria, p. Conventionally, the expression is truncated in a way so that only the polynomial basis with total degree not higher than p are kept,

$$I^{M,p} = \{\boldsymbol{\alpha} \in N^M : \| \boldsymbol{\alpha} \|_1 \leq p\} \tag{6}$$

With this truncation strategy, the number of existing terms in the expansion is given by:

$$K = (M + p)!/M\,!\,p! \tag{7}$$

where *M* is the number of independent uncertainty sources while *p* represents the highest order of the one-dimensional polynomial basis.

With known polynomial basis and truncation criteria, the polynomial coefficients $\{a_i\}$ need to be determined in order to complete the expansion. Approaches to evaluate PC coefficients in general are classified as the intrusive Galerkin projection and non-intrusive methods. The intrusive approaches need the formulation of another mathematical problem that provides the set of polynomial coefficients. Non-intrusive approaches, on the other hand, take the actual model and the solver as a black box and the coefficients are then calculated by a set of realizations of the original system. Sampling, quadrature and linear regression are some of the most widely used non-intrusive approaches. The linear regression method estimates the polynomial coefficients by means of optimization [4].

Let $\mathbf{X} = \{\mathrm{x}_1, \ldots, \mathrm{x}_\mathrm{N}\}$ be a set of representatives of the random input vector, also called as the experimental design (ED), and $\mathbf{Y} = \{f(\mathrm{x}_1), \ldots, f(\mathrm{x}_\mathrm{N})\}^T$ be the corresponding stochastic response vector, the set of coefficients can be calculated 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 **a** is the vector of coefficients$\{a_i\}$ (i=0,…,K), and the so-called experimental matrix **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 a 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 experimental design, **X** should contain a sufficient number of points, preferably 2-3 times the number of expansion terms.

2.1.2 Sparse-adaptive scheme

As expressed in (7), the number of expansion terms $K+1$ increases rapidly with M and p, which extends the number of ED. For instance, taking $p = 4$ as fixed, *K* becomes 1001 to 10626 if *M* is raised from 10 to 20. The sparse-adaptive scheme is proposed to deal with this issue of high-dimensionality. Following sections present the scheme more in details.

Revised truncation strategy

Any truncation strategy aims at specifying a nonempty 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 favouring interactions among low-degree terms, through the revised truncation strategy [15]:

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

where **i** is the multi-index. Note that the conventional strategy corresponds to the **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. The penalized optimization problem in (11) allows to select the most important polynomials from the initially defined set. A number of algorithms are available to solve such problems including, least absolute shrinkage and selection operator, the forward stage-wise regression and least angle regression (LARS). In reality, the algorithms are similar to each other in terms of optimization mechanism, and differs mostly in the greedy level, the precision target and the search speed.

In statistical sciences, LARS is used for fitting linear models to high-dimensional data, and is efficient even when the number of regressors is much higher than the available data. For gPC expansion, LARS has been successfully applied to construct the sparse PC expansion [15]. In a nutshell, LARS formulates a regression problem of a reduced size by selecting the most relevant polynomials with respect to the given model evaluations **Y**.

Basis-adaptive procedure

Assigning the truncation criteria p, is strongly problem dependent, although gPC expansions with $p = 2$ has been reported to be accurate for estimating the first two statistical moments of a stochastic response [16]. Instead of fixing a value beforehand, the basis-adaptive procedure starts from a range of candidates, and automatically chooses the best option through an 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. Thus, it is possible to find a polynomial function that exactly satisfies the relationship between distinct ED $\mathbf{X} = \{\mathrm{x}_1,\ldots,\mathrm{x}_\mathrm{N}\}$ and the resulting model evaluations $\mathbf{Y} = \{f(\mathrm{x}_1),\ldots,f(\mathrm{x}_\mathrm{N})\}^T$. However, the function that maximizes its efficacy on a specific ED, may render unsatisfactory 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

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

where $f_{-l}^{gPC}(\mathbf{X} \setminus \mathrm{x}_l)$ stands for the gPC expansion of $f(\xi)$ with specific ED $\mathbf{X} \setminus \mathrm{x}_l = \{\mathrm{x}_k, k = 1, \ldots, N, k \neq l\}$. For 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}.

2.2 Copula theory

In principle, the gPC expansion is used to approximate the stochastic response of independent random input variables. A number of techniques have been proposed, for example- linear transformations, proper orthogonal decomposition or K- L expansion, in order to apply the gPC expansion for correlated inputs [17], [18]. However, these methods address the correlation by transformations to eliminate the linear correlations. Although linear correlation can effectively describe the linear or monotonic dependence structure, yet cannot provide a satisfactory measure of the nonlinear or non-monotonic dependence. In such a case, the Copula theory allows a more generalized modelling of the dependence structure [19]. According to the Sklar's theorem, a joint CDF can be represented by univariate functions and an additional function that describes their dependences,

$$F_{\xi}(\xi) = C(F_{\xi_1}(\xi_1), \ldots, F_{\xi_M}(\xi_M)) \quad (14)$$

where $\xi = \{\xi_1, \ldots, \xi_M\}$ is the *M*-dimensional random input vector, $F_{\xi}(\xi)$ is the marginal distribution of the *i*-th random variable, and *C*(.) is the parametric copula function. Various types of copula functions have been described in literature, while the Gaussian copula has been used in this work.

3 Proposed approach

This section presents the detailed procedure of the proposed PPF analysis. First, an overview of the DPF is provided. Next, the formulations of the BASPC expansion along with conventional PPF based on LHS are discussed. Finally, key points are provided to compare both methods.

3.1 DPF calculation

The power flow analysis is the most commonly used tool in power systems. Power flow analyses are performed for a wide range of applications including checking of operational limits for diverse loading conditions and network planning purposes. For single-phase systems, classical DPF problems are defined by the following set of equations,

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

where, P_i and Q_i are 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 between bus i and j respectively. The total number of buses available in the network is n, while the voltage phase angle difference is represented by $\theta_{ij} = \theta_i - \theta_j$.

3.2 PPF using BASPC

The DPF analysis is mostly used to determine steady-state operating parameters of the power system consisting of state variables, inputs and outputs. The state variables 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. On the other hand, the outputs can be a vector of interest, e.g. power flows and currents through the lines. In most of the real cases, inputs of DPF analyses consist of random variables, which requires PPF analysis. In this report, a novel method of PPF is proposed by combining the BASPC expansion with LHS sampling.

3.3 Comparison between classical PPF and BASPC-based PPF

The performance of a new PPF method is usually compared with the simulation based counterparts, such as MC approaches. In recent years, smarter random sampling methods are being used as more efficient techniques to generate random trials. For instance, LHS has been reported as a more efficient method than MC for power system applications [20]. In this work, the reference PPF method has been designed using LHS, and the proposed method based on BASPC is compared with the reference method in terms of the accuracy and efficiency.

Procedure 1 and Procedure 2 in Table 1 depict the steps required by the proposed approach and the conventional PPF method based on LHS respectively. Both of the methods involve pre and post process stages, although the purpose and associated computational burdens are 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 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 N_{sam} samplers on the available surrogate models. Unlike the first procedure, LHS based PPF executes N_{sam} rounds of DPFs at the pre-process stage.

The total computational time T_{total} required for the entire procedure of PPF is the summation of times consumed in both pre- process and post-process, represented by t_{pre} and t_{post}. In case of the BASPC-LHS based PPF, t_{pre} is calculated by, $t_{sam} + N_{ed} \times t_0 + t_{cal}$, where t_{sam} is the for generating $\mathbf{H}_2$, t_0 is the time taken by a single round of DPF calculation and t_{cal} denotes the time required for generating $\mathbf{H}_1$, building surrogate models and executing $\mathbf{H}_1$ through the surrogate models. In contrast, for LHS-based PPF, $t_{pre} = t_{sam} + N_{sam} \times t_0$. It needs to be noted that, the experimental design is much less than the number of elements in $\mathbf{H}_1$. Thus, the proposed method based on BASPC-LHS consumes much less time for the DPF calculations compared to the conventional approach.

Table 1. Procedures of the PPF algorithms with BASPC-LHS and conventional PPF using LHS

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

4 Simulation setup

In this section, a brief overview of the simulation test case is provided along with the indicators of evaluating the performances. The simulations are performed in MATLAB R2017a in computer equipped with an Intel Core i7 processor with four cores at 3.40GHz and 8GB of RAM. OpenDSS [21] has been used to run the DPF calculations and UQLab from ETH Zurich [22] has been used to construct the BASPC expansion.

4.1 Test networks

In this work, performance of the proposed method is tested on the IEEE European Low Voltage Test Feeder [23]. This typical European LV feeder is operated radially with a base frequency of 50Hz. The feeder consists of 906 buses, 55 of which represent residential point of connections (POCs). As shown in Figure 1, each load is equipped with a single-phase connection.

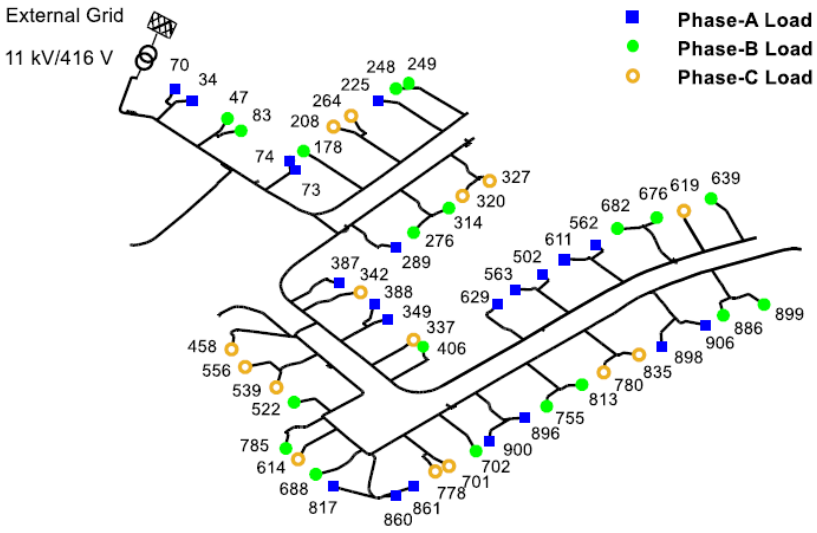

Figure 1. Single-line diagram of the IEEE European LV test feeder.

The external medium-voltage (MV) network has been modelled as a voltage source with an impedance. The network parameters have been adopted from [23]. In addition to the loads, simulation test cases have been designed with solar PV systems in all the residential POCs. The PV systems work in unity power factor with output active power in ranging between 1-5 kW. A simplified approach is adopted in this work, where all the loads are modelled to operate in constant power mode. Since both load and PV generation are subject to uncertainties, they have been represented by random variables with particular type of distributions. It is important to note that, the PV outputs can be correlated due to the same irradiance in a certain geographical region.

4.2 Performance indicators

To assess the accuracy of the BASPC-LHS based method, the results obtained from LHS based PPF with 10^5 samples are regarded as the reference in the analyses. A number of indices have been used to analyse the performance of the proposed approach. The quantitative indicators, mean value, variance, skewness and kurtosis of the desired outcomes are first compared in terms of the relative error (*RE*) given by,

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

where 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.

Next, graphical results, such as frequency histograms after the post-processing are analysed through similarity index (*SI*). In this case, *SI* is defined by modifying the city block distance,

$$SI(r^{ref}, r^{cp}) = 1 - \frac{1}{2}\sum_{i=1}^{n}\sqrt{[(c_i^{ref} - c_i^{cp})^2 + (d_i^{ref} - d_i^{cp})^2]} \qquad (16)$$

where n is the number of bins, c_i corresponds to the location of the i-th bin centre 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. Intuitively, a greater value of the index indicates a better fit between the two frequency histograms. For the sake of conciseness, *SI* in (16) can be further simplified by locating bin centres at the same position. Therefore, $c_i^{ref} = c_i^{cf}$ and $SI(r^{ref}, r^{cp})$ ranges between 0 and 1. By means of this metric, the similarity of two frequency histograms can be qualitatively detected. Apart from the accuracy, the efficiency of the proposed approach is another concern of this work. The gain in efficiency is quantified through the percentage of reduction in computation time, *PD* as defined in (17),

$$PD(t^{ref}, t^{cp}) = (t^{ref} - t^{cp}) / t^{ref} \qquad (17)$$

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

5 Numerical results and discussion

In this section, the simulation results are discussed in details. First, the performance of the surrogate model is evaluated for different design parameters. Then, the post-process results are compared with LHS-based PPF followed by a brief overview of the required computational overhead.

5.1 Evaluation of the surrogate model

The performance of the surrogate model has been evaluated using the load profiles provided with the IEEE European LV test feeder [23]. PPF calculations were performed using the values of the peak load at 09:26.

At the base case, each of the 55 POCs was considered as an uncertainty source with active power modelled as Gaussian distribution $N\left(PL_i, 0.1^2\right)$, where PL_i denotes the active power consumption of i-th POC at 09:26. As discussed earlier, BASPC expansion depends on a number of parameters, such as the truncation norm q, range of truncation criteria p and number of samples in ED N_{ed}. The impact of these three parameters are investigated with reference to the results obtained through LHS based PPF with 10^5 samples. A base configuration of the BASPC expansion is formulated considering $q = 0.8, p = 1-6$ and $N_{ed} = 800$.

Table 2 presents the influence of the design parameters of the BASPC expansion on the relative error, e_{LOO} of the magnitude of three-phase voltage at bus 899 and total calculation time, T_{total}. The behaviour of the surrogate model is subject to changes upon variation in either of q, p or N_{ed}. Lower values of truncation norm, ED and truncation criteria lead to a higher value of e_{LOO}. This is due to the weakening of the BASPC expansion for approximating the outputs. However, the time taken for simulation is reduced due to the lower computation burden of building the surrogate model. Comparing both efficiency and accuracy, an efficient configuration of the surrogate model is chosen with truncation norm 0.8, range of truncation criteria 1-6 and number of ED 250.

Table 2. Impact of the BASPC configuration on the surrogate model

BASPC configuration			e_{LOO} (p.u.)			T_{total} **(s)**
q	p	N_{ed}	V_A	V_B	V_C	
0.8	1-6	800	7.37×10^{-06}	9.70×10^{-06}	6.30×10^{-06}	143.97
0.4	1-6	800	8.91×10^{-06}	2.14×10^{-05}	1.00×10^{-05}	31.33
0.8	**1-2**	800	1.36×10^{-05}	2.31×10^{-05}	1.04×10^{-05}	28.34
0.8	1-6	**250**	2.19×10^{-05}	5.19×10^{-05}	1.83×10^{-05}	10.53

The performance of the surrogate model has also been assessed in terms of number of random input variables and their correlation. In addition to the active power at 55 buses, PV generation at these buses have also been considered. The PV generation of the buses has been modelled as Uniform distribution $U(0.9\times PV_i, 1.1\times PV_i)$, where PV_i denotes the PV generation at each POC at 09:26. The solar irradiation data has been adopted from [24]. As discussed earlier, the PV generation has been considered as correlated in one of the simulation cases using copula. Gaussian copula has been adopted in this case with

correlation coefficient being 0.8. Table 3 shows the effects of the random variables on three-phase voltage magnitude at bus 899 and the total simulation time. Increasing number of random variables and consideration of correlation among them lead to slight decrease in the accuracy and increment of the calculation time of the model. However, the magnitude of the error, e_{LOO} are in the same order in all cases, while the simulation time increases by roughly 6s.

Table 3. Influence of the types and number of random input variables on the surrogate model

Random inputs			e_{LOO} (p.u.)			T_{total} **(s)**
Type	**Number**	**Correlation**	V_A	V_B	V_C	
Gaussian	55	No	2.19×10^{-05}	5.19×10^{-05}	1.83×10^{-05}	10.53
Mixed	110	No	2.77×10^{-05}	6.14×10^{-05}	2.53×10^{-05}	11.04
Mixed	110	**Yes**	6.79×10^{-05}	1.50×10^{-04}	3.99×10^{-05}	16.38

In addition to the design parameters and random input variables, the impact of different outputs have also been taken into consideration for the surrogate model. Table 4 presents the influence of different outputs on the performance of the model. The efficient configuration of the surrogate model has been used to estimate the three-phase voltage magnitude and voltage unbalance factor (VUF) at bus 899 along with the apparent power flow at the transformer, S_{tran}. The BASPC expansion is seen to perform better for the voltage magnitudes compared to the power flow and voltage unbalance factor. While the bus voltages are direct output of the PPF, extra calculations are performed in order to quantify the line flows and VUF. This additional computational effort is reflected in terms of the increased error estimates and simulation time.

Table 4. Impact of different outputs on the surrogate model

Indicators	V_A	V_B	V_C	S_{tran}	VUF
e_{LOO} (p.u.)	2.19×10^{-05}	5.19×10^{-05}	1.83×10^{-05}	6.01×10^{-03}	8.25×10^{-03}
T_{total} (s)		10.53		11.05	16.19

5.2 Comparison with classical PPF

In this section, post-process results of the proposed BASPC expansion model are compared to the results obtained through classical PPF method based on LHS. The efficient surrogate model has been used to obtain the BASPC-LHS results of the 10^5 samples originally used for the PPF calculations through LHS based PPF.

Figure 2 illustrates the histograms of the voltage magnitude in each phase at bus 899 for conventional and proposed PPF methods. From visual point of view, both of the approaches represent similar results of the unbalanced voltages at the POC. The differences in voltage magnitudes are expected due to the unequal demand of power in each phase. *SI* has been used to quantify the resemblance of the frequency histograms. As shown in Table 5, *SI* values for each pair of phase voltages are well above 99%.

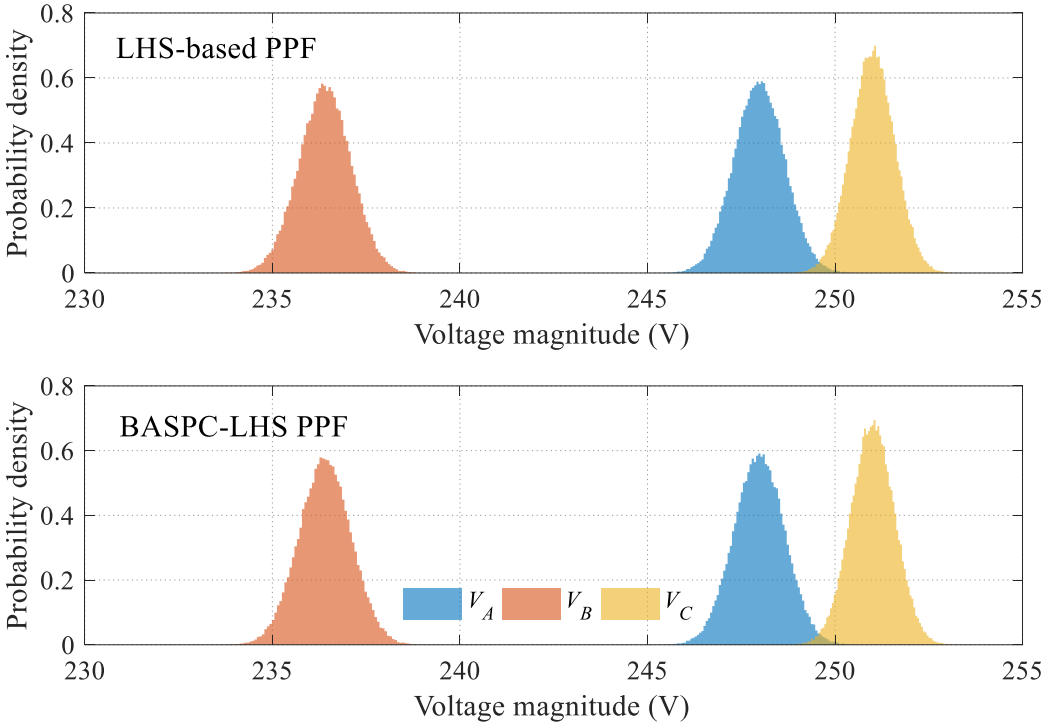

Figure 2. Frequency histogram of three-phase voltage magnitude at bus 899.

Table 5. SI for the phase voltages at bus 899

	V_A	V_B	V_C
SI (%)	99.82	99.70	99.86

Factors that influence the performance of the surrogate model also render impacts on the post-process results. For instance, different number of ED will result in changes in the statistical properties of the output variables. Figure 3 illustrates the PDFs of the output voltage magnitude of phase A at bus 899 for different number ED in BASPC expansion keeping the truncation norm and truncation criteria fixed. For comparison, the histogram of output voltage magnitude obtained through LHS-based has also been plotted in the same figure. It can be observed that the proposed approach with $N_{ed} = 50$ does not yield results that follow the reference voltage. However, the proposed and reference methods coincide when the N_{ed} reaches up to 250.

Five quantitative indicators for phase A voltage magnitude are tabulated in Table 6. The RE and SI in different indicate the importance of a reasonable number of samples in the ED.

Table 6. Influence of ED on SI and RE

N_{ed}	SI (%)	***RE***			
		Mean	**Variance**	**Skewness**	**Kurtosis**
50	83.83	2.45×10^{-04}	4.25×10^{-01}	1.08	0.25
250	99.82	2.65×10^{-07}	8.16×10^{-04}	0.76	2.26×10^{-03}
800	99.85	1.57×10^{-07}	1.77×10^{-03}	0.45	1.57×10^{-03}

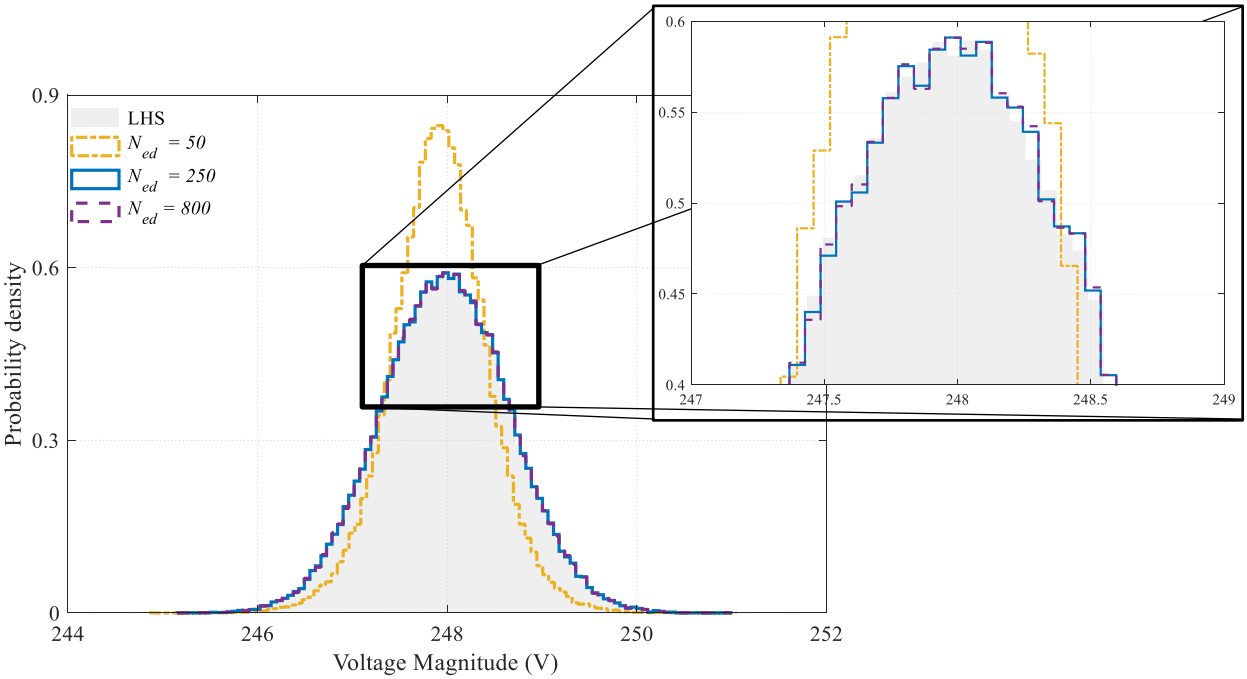

Figure 3. Frequency histogram of phase A voltage magnitude.

5.3 Computational overhead

As discussed in Section 5.2, the proposed BASPC expansion based PPF algorithm can provide results with a high degree of accuracy. A major strength of the proposed approach originates from the efficiency of calculation, i.e. the relatively short time of calculation. Table 7 summarizes the computational overhead of the proposed approach compared to LHS-based PPF. The proposed approach needs to perform a fraction of DPF calculations compared to the conventional methods, resulting in up to 99.68% saving in the computational time. It needs to be noted that the calculations have been performed using parallel computing. Thus, the savings in time will be even higher if parallel computing is not available.

Table 7. Computational burden of PPF

	PCE-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)	9.59	10.22	141.25	141.25	0.07	0.55
t_{post} (s)	0.04	0.24	0.08	0.62	331.37	3273.37
t_{total} (s)	9.63	10.46	141.33	141.87	331.44	3273.92
PD (%)	97.09	99.68	57.36	95.66	-	-

6 Conclusions

With the energy transition, an increased amount of uncertainty is expected in the future power systems with the rapid proliferation of various DERs and their integration in micro-grids. Planning of the future power systems involving numerous physical and commercial micro-grids my therefore be computationally demanding and extremely time consuming. Efficient PPF techniques are therefore significant for UQ and to estimate the propagation of uncertainty. The objective of this report has been to present a novel method of PPF analysis based on BASPC expansion that can handle a large number of random input variables. The correlation among the inputs has also been addressed using Gaussian Copula. The proposed method has been tested and compared to conventional technique with LHS through simulation in IEEE European LV test feeder.

Various configurations of the BASPC expansion have been assessed in terms of the truncation criteria, truncation norm and experimental design. The effects have also been analysed for different numbers and types of inputs, outputs, number of samples and the presence of correlation among the input variables. The accuracy and efficiency of the proposed method have also been quantified with respect to the conventional method. The simulation results show that the proposed technique can be used as an efficient tool for the PPF analysis providing a considerable degree of accuracy at a low computational overhead.

The proposed algorithm will be a suitable tool for the planning purposes of the future distribution network including DERs, micro-grids and demand response methodologies. Future research will be directed at employing the algorithm for the enhancement network planning process by clustering of the micro-grids.

References

[1] R. Doherty and M. O' Malley, "A new approach to quantify reserve demand in systems with significant installed wind capacity," *IEEE Trans. Power Syst.*, 2005.

[2] Y. V. Makarov, P. V. Etingov, J. Ma, Z. Huang, and K. Subbarao, "Incorporating uncertainty of wind power generation forecast into power system operation, dispatch, and unit commitment procedures," *IEEE Trans. Sustain. Energy*, 2011.

[3] M. Waite and V. Modi, "Potential for increased wind-generated electricity utilization using heat pumps in urban areas," *Appl. Energy*, 2014.

[4] F. Ni, P. H. Nguyen, and J. F. G. Cobben, "Basis-Adaptive Sparse Polynomial Chaos Expansion for Probabilistic Power Flow," *IEEE Trans. Power Syst.*, vol. 32, no. 1, pp. 694–704, 2017.

[5] M. Fan, V. Vittal, G. T. Heydt, and R. Ayyanar, "Probabilistic power flow analysis with generation dispatch including photovoltaic resources," *IEEE Trans. Power Syst.*, 2013.

[6] A. M. Leite da Silva, S. M. P. Ribeiro, V. L. Arienti, R. N. Allan, and M. B. Do Coutto Filho, "Probabilistic load flow techniques applied to power system expansion planning," *Power Syst. IEEE Trans.*, 1990.

[7] Z. Hu and X. Wang, "A probabilistic load flow method considering branch outages," *IEEE Trans. Power Syst.*, 2006.

[8] P. Zhang and S. T. Lee, "Probabilistic Load Flow Computation Using the Method of Combined Cumulants and Gram-Charlier Expansion," *IEEE Trans. Power Syst.*, 2004.

[9] M. Fan, V. Vittal, G. T. Heydt, and R. Ayyanar, "Probabilistic power flow studies for transmission systems with photovoltaic generation using cumulants," *IEEE Trans. Power Syst.*, 2012.

[10] J. Tang, F. Ni, F. Ponci, and A. Monti, "Dimension-Adaptive Sparse Grid Interpolation for Uncertainty Quantification in Modern Power Systems: Probabilistic Power Flow," *IEEE Trans. Power Syst.*, 2016.

[11] Z. Ren, W. Li, R. Billinton, and W. Yan, "Probabilistic Power Flow Analysis Based on the Stochastic Response Surface Method," *IEEE Trans. Power Syst.*, 2016.

[12] J. M. Morales and J. Pérez-Ruiz, "Point estimate schemes to solve the probabilistic power flow," *IEEE Trans. Power Syst.*, 2007.

[13] G. Papaefthymiou and D. Kurowicka, "Using copulas for modeling stochastic dependence in power system uncertainty analysis," *IEEE Trans. Power Syst.*, 2009.

[14] D. Cai, D. Shi, and J. Chen, "Probabilistic load flow computation using Copula and Latin hypercube sampling," *Transm. Distrib. IET Gener.*, 2014.

[15] G. Blatman, "Adaptive sparse polynomial chaos expansions for uncertainty propagation and sensitivity analysis," Universit ´e Blaise Pascal, Clermont-Ferrand, France, 2009.

[16] T. Lovett, F. Ponci, and A. Monti, "A polynomial chaos approach to measurement uncertainty," *Instrum. Meas. …*, vol. 55, no. 3, pp. 729–736, 2006.

[17] M. Aien, M. Fotuhi-Firuzabad, and F. Aminifar, "Probabilistic load flow in correlated uncertain environment using unscented transformation," *IEEE Trans. Power Syst.*, 2012.

[18] Y. Chen, J. Wen, and S. Cheng, "Probabilistic load flow method based on nataf transformation and latin hypercube sampling," *IEEE Trans. Sustain. Energy*, 2013.

[19] F. Durante and C. Sempi, "Copula Theory : an Introduction," *Copula Theory Its Appl.*, 2010.

[20] Zhen Shu and P. Jirutitijaroen, "Latin Hypercube Sampling Techniques for Power Systems Reliability Analysis With Renewable Energy Sources," *IEEE Trans. Power Syst.*, 2011.

[21] EPRI, "OpenDSS, distribution system simulator." [Online]. Available: http://sourceforge.net/projects/electricdss. [Accessed: 26-Jun-2018].

[22] S. Marelli and B. Sudret, "UQLab: A Framework for Uncertainty Quantification in Matlab," in *Vulnerability, Uncertainty, and Risk*, 2014.

[23] "IEEE PES Distribution Test Feeder Working Group," "Distribution test feeders." [Online]. Available: https://ewh.ieee.org/soc/pes/dsacom/%0Atestfeeders/. [Accessed: 26-Jun-2018].

[24] A. Haque, M. Nijhuis, G. Ye, P. Nguyen, F. Bliek, and J. Slootweg, "Integrating Direct and Indirect Load Control for Congestion Management in LV Networks," *IEEE Trans. Smart Grid*, 2017.