10.1016/j.ijepes.2020.106043
Gruosso, G., Maffezzoni, P.
Data-driven uncertainty analysis of distribution networks including photovoltaic generation
(2020) International Journal of Electrical Power and Energy Systems, 121, art. no. 106043,

Data-Driven Uncertainty Analysis of Distribution Networks including Photovoltaic Generation

Giambattista Gruosso[a], Paolo Maffezzoni[a]

[a]*Politecnico di Milano, DEIB,*
Piazza Leonardo da Vinci, 32, I20133 Milan, Italy,

Abstract

This paper investigates residential distribution networks with uncertain loads and photovoltaic distributed generation. An original probabilistic modeling of consumer demand and photovoltaic generation is presented that is based on the analysis of large set of data measurements. It is shown how photovoltaic generation is described by complex non-standard distributions that can be described only numerically. Probabilistic analysis is performed using an enhanced version of the Polynomial Chaos technique that exploits a proper set of polynomial basis functions. It is described how such functions can be generated from the numerically available data. **Compared to other approximate methods for probabilistic analysis, the novel technique has the advantages of modeling accurately truly nonlinear problems and of directly providing the detailed Probability Density Function of relevant observable quantities affecting the quality of service. Compared to standard Monte Carlo method, the proposed technique introduces a simulation speedup that depends on the number of random parameters. Numerical applications to radial and weakly meshed networks are presented where the method is employed to explore overvoltage, unbalance factor and power loss, as a function of photovoltaic penetration and/or network configuration**.

*Giambattista Gruosso
Email address: giambattista.gruosso@polimi.it (Paolo Maffezzoni)

Keywords: Photovoltaic Generation,Uncertainty Analysis, Data-Driven models,Unbalanced Distribution Networks, Probabilistic Load Flow, Polynomial Chaos

1. Introduction

In this paper, we analyze residential Low Voltage Distribution Networks (LVDNs) in the presence of photovoltaic (PV) Distributed Generation (DG). It is well known how the integration of DG into LVDNs can raise many challenges and concerns. This is because traditionally-designed networks are dimensioned for distributing power unidirectionally from the sources towards the passive loads. In DG systems, instead, power is also injected backward into the network leading to many potential troubles, such as overvoltage, phase unbalance and current exceeding wire carrying capacity [1, 2]. **Evaluating the impact that photovoltaic DG could have on residential LVDNs is an important issue. This is particularly challenging due to the remarkable degree of statistical uncertainty exhibited by both PV sources and residential customers power demand.** A detailed analysis requires the adoption of realistic stochastic models for the PV sources and loads[3, 4, 5, 6] as well as the exploitation of advanced probabilistic computational tools able to explore in an efficient way the many operational scenarios.

Several statistical models for residential loads have been adopted in previous studies with many of them being based on the assumption that power demand can be represented by Gaussian-distributed random variables. [7, 8, 9, 10], Different types of models have also been proposed for PV power generation, which include bimodal Gaussian mixture model [11]. For what concerns probabilistic analysis, the basic and reference method still remains Monte Carlo simulation [7], [12]. However, due to the heavy computational times it can require, several approximate probabilistic analysis techniques have also been proposed in the literature [13], [14], [15], which includes Point Estimate, Cumulant methods and Surface Response Method [16, 17, 18, 19, 20]. **Among existing techniques,**

Cumulant method works well for linear (or almost linear) problems and its application to the nonlinear probabilistic power flow relies on approximate linearizations. Similarly, the Point Estimate method can provide approximations of the raw statistical moments of some electrical variables, however further elaborations/approximations are needed in order to derive the Probability Density Function (PDF) of such variables. In this paper, we focus on the Surface Response method based on polynomial chaos expansions and stochastic testing method [21, 22, 23]. The features that make such a method particularly attractive are: (i) it can deal with truly nonlinear problems, as it is the case for load flow problems in unbalanced grids; (ii) it directly provides the detailed PDF of observable quantities that are crucial for determining the quality of service, such as voltage peak, voltage unbalance factor or power loss. In this paper, we present fresh results about the statistical modeling and analysis of distribution networks in the presence of PV distributed generation. Our approach is based on the exploitation of data-driven models for loads and PV generation. Our contributions include among the others:

1. New results about statistical modeling of residential consumers power demand are presented. Our approach is based on the analysis of large data set of measurements grouped for different daily *time windows*. As a practical case study, we exploit the data available for the city of London [24] and classified on the basis of the census of the users. The study highlights how the power demand for a single user (or small set of users) is not distributed as a Gaussian statistical variable. In this case, power demand is better modeled with nonGaussian distributed variables, and in particular with Beta-distributed variables, whose PDF can be determined from data by means of a parametric approach.
2. The issue of the stochastic modeling of photovoltaic sources is investigated. Data analysis shows how the power generated by typical domestic

PV plants have complex distributions. In this case, a parametric modeling is not adequate and delivered power statistical distribution is better described by a numerical approach.

3. We describe and enhanced Probabilistic Load Flow (PLF) technique able to deal with the complex loads/sources modeling scenario. The technique employs a generalized Polynomial Chaos (gPC) method with the set of polynomial basis functions tailored to the non-elementary distributed variables. We show how such gPC basis functions can be accurately generated starting from the numerically available density functions extracted through data analysis.
4. The enhanced gPC method is exploited to calculate the Probability Density Function (PDF) of a set of observable quantities that are relevant for the system design. In particular, we consider both radial and weakly mesh networks and explore the statistical distribution of nodal voltages and branch currents as well as of Voltage Unbalance Factor and total power loss.

The rest of the paper is organized as follows: In Sec. II, we derive realistic data-driven statistical models for users and PV sources. In Sec. III, we show how the gPC basis functions can be derived from the data. Sec. IV is devoted to the implementation of the gPC stochastic analysis for LTVN while Sec. V extends the method to sensitivity analysis. Finally, in Sec. VI, we illustrate the numerical applications of the method to both radial and weakly mesh networks.

2. Modeling power loads and sources

2.1. Residential loads

The first issue of our analysis is finding the typical statistical distributions of residential consumers power demand. To this aim, we exploit the data set of residential consumers daily power profiles measured in the town of London and provided by the UK Power Networks distribution network operator [24]. Power profiles are divided into different sets on the basis of the customers category.

In order to account for the time dependence, measurements data are grouped into several sets corresponding to different time windows during the day. Hence, the statistical distribution of power demand is extracted for each time window. As an example, Fig. 1 shows fifty power profiles over a four-hour time window ranging from 10:00 AM to 02:00 PM for the category of "Affluent customers". **The choice of the window is completely arbitrary and does not affect the method. The choice of a four hour window in this case is due to the low resolution of the available data.** The diversity in the usage of domestic electrical appliances results in a great variability of the values assumed by loads. Indeed, it is seen how a great number of power samples fall within a low-power range (< 1.0 kW) while samples with high power values, close to the peak (≈ 9 kW), are much rare. Qualitatively, this fact suggests how data distribution is non symmetric and non Gaussian. In order to quantitatively determine data distribution, the power values are first normalized to the peak value K_L. In this way, the active power absorbed by a load is written as

$$P_L = K_L\, y \tag{1}$$

where $y \in [0, 1]$ is a random variable distributed accordingly to the PDF $f(y)$. The application of the histogram operator to the normalized power values provides the (approximate) numerical samples $f(y^j)$ of the PDF over a sequence of ordered values $y^j \in [0, 1]$, with $y^{j+1} > y^j$. Fig. 2 shows the PDF data distribution for the considered time window from 10:0 AM to 2:00 PM and the Affluent customers category. In this case, the peak power demand results $K_L = 9.2$ kW. Second, the extracted PDF samples $f(y^j)$ are fitted by trying several standard statistical distributions (among which Weibull, Exponential, and Beta) [25]. We find that, for the available data, the best fitting is achieved by the Beta distribution:

$$f_\beta(y, a, b) = \frac{y^{a-1}(1-y)^{b-1}}{\mathrm{Beta}(a, b)}. \tag{2}$$

where $\mathrm{Beta}(a, b)$ denotes the Beta function while a and b are positive parameters to be determined. The values of parameters a and b are deduced by an

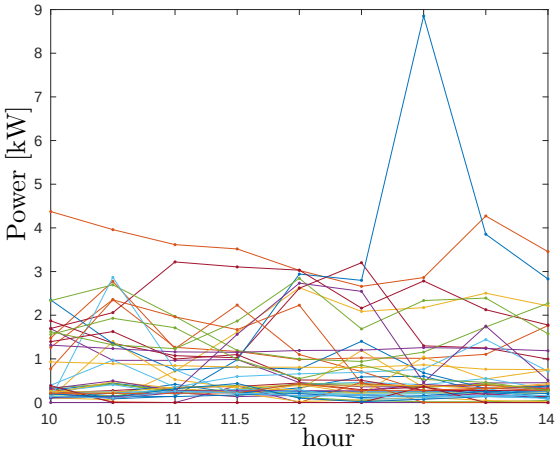

Figure 1: Typical residential power profiles over the time window: 10:00 AM–02:00 PM.

optimization procedure that minimizes the following objective function:

$$\text{Error}(a,b) = \sum_j \left(f_\beta(y^j,a,b) - f(y^j)\right)^2, \tag{3}$$

with the constraints $a > 0$ and $b > 0$. Fig. 2 shows the Beta distribution $f_\beta(y,a,b)$ with parameters $a = 1.1$, $b = 22.8$ that yields data best fitting for the considered case.

Fig. 3 reports the normalized power demand distribution for another category of customers referred to as "Comfortable customers". In this case the contractual peak power is lower and the measured peak power is $K_L = 5.2$ kW. Also in this second case, the PDF is well approximated by the Beta distribution $f_\beta(y,a,b)$ with parameters $a = 1.2$, $b = 23.1$. **The correlation coefficient [26] calculated on the power demand samples (over the considered time window) of two different customers results very small, i.e.** ≈ 0.01**, thus denoting a certain degree of independence**. Our analysis, repeated for several time windows during the day and different customers categories leads us to the conclusion that residential (normalized) power demand can be realistically reproduced by means of **statistically independent** Beta-distributed random variables whose PDF can be extracted from data through the abovereported

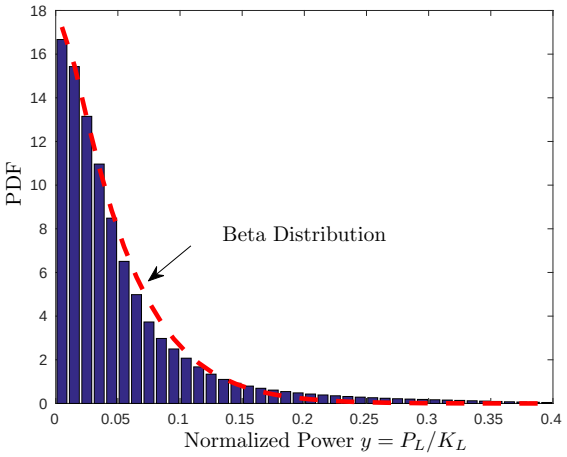

Figure 2: (histogram): normalized power PDF for Affluent customers category ; (dashed line) fitting with the Beta distribution of parameters $a = 1.1$, $b = 22.8$.

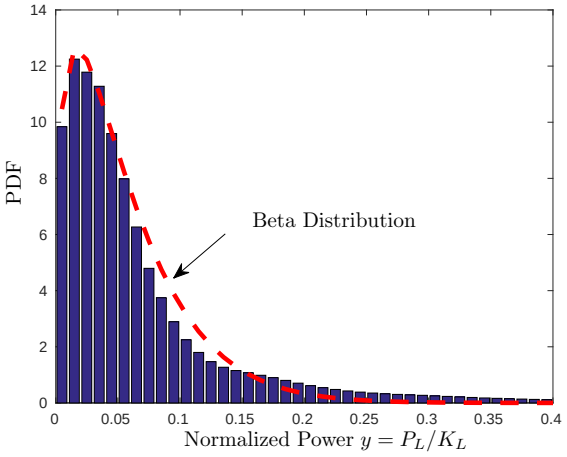

Figure 3: (histogram): normalized power PDF for the Comfortable customers category; (dashed line) fitting with the Beta distribution of parameters $a = 1.2$, $b = 23.1$.

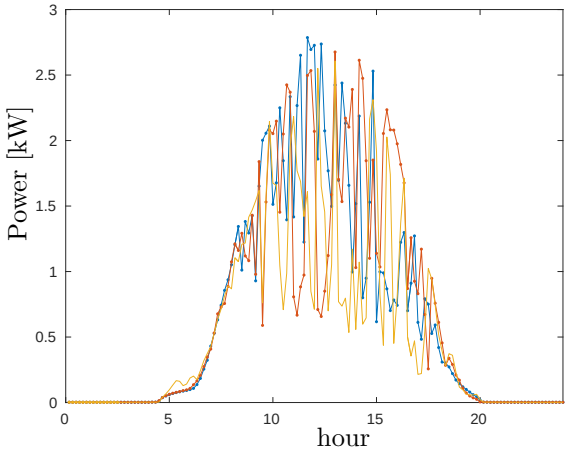

Figure 4: Typical daily power profiles delivered by PV plants.

parametric approach.

2.2. Photovoltaic generation

We pass now to analyze the data set describing the active power delivered by some typical photovoltaic domestic plants placed in the UK Power Networks distribution network. Fig. 4 shows, as an example, three of such daily power profiles. In the central hours of the day, the power samples exhibit a significant variability due to the difference in solar irradiance. Figs. 5 and 6 show the PDF of the PV-delivered power normalized to the peak value $K_S = 3$kW over two time windows from 10:00 AM to 02:00 PM and from 02:00 PM to 06:00 PM.

In both cases, the distributions of the delivered power exhibit quite complex shapes that cannot be easily reconduced to any elementary/standard statistical distribution [27]. Thus, the parametric method used for modeling residential loads is not adequate for modeling PV generation and a numerical approach should be adopted. To this aim, the power delivered by the PV source is written as

$$P_S = K_S\, x, \tag{4}$$

where $x \in [0, 1]$ is a non-elementary-distributed random variable described nu-

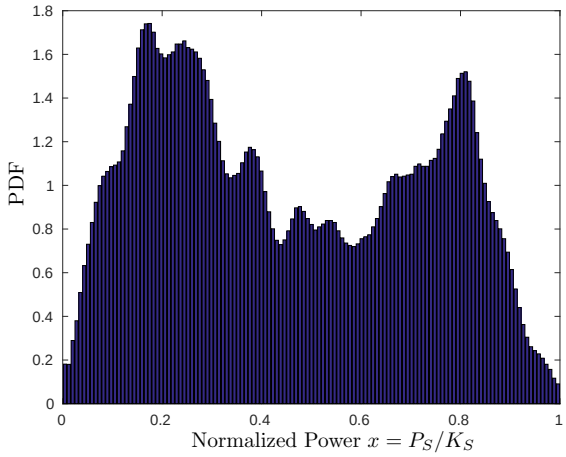

Figure 5: (histogram) PDF of the normalized PV delivered power over the time window: 10:00 AM–02:00 PM.

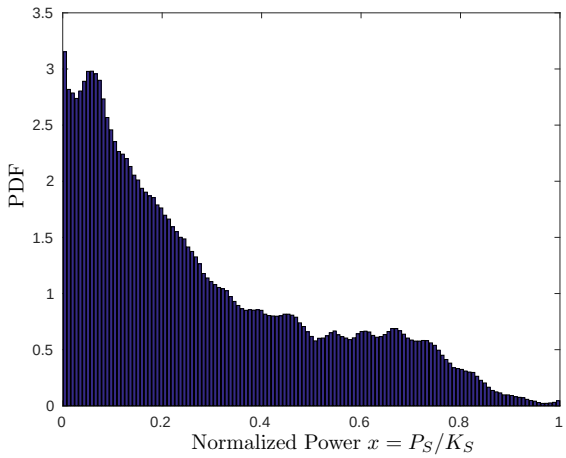

Figure 6: (histogram) PDF of the normalized PV delivered power over the time window: 02:00 PM–06:00 PM.

merically by the samples $f(x^j)$ of its PDF $f(\cdot)$ known over a sequence of discrete values x^j. For each considered daily time window, a sequence of couple values $(x^j, f(x^j))$, along with the associated peak power K_S, is enough to describe statistically the fluctuations of the PV generation. Random realizations of x can be generated numerically by calculating the Cumulative Distribution Function (CDF) $F(x)$ associated to the density $f(x)$. A random value of a variable z is extracted with uniform probability between 0 and 1. It is assigned to the CDF on the vertical axis, i.e. $F(x) = z$ and the corresponding value on the horizontal axis $F^{-1}(z) = x$ provides the wanted realization for x. Since the inverse relationship $F^{-1}(\cdot)$ is known at discrete samples, some form of data interpolation is required.

3. Building the gPC basis functions

Probabilistic analysis with gPC method (described in Sec. IV) requires determining the proper polynomial chaos basis functions associated to each random variable. For random variables with known elementary distributions (e.g. Gaussian, Uniform, Beta) such polynomials are known in closed form [21]. Instead, for non-elementary-distributed random variables or for variables whose PDF is known only numerically the correct gPC basis functions should be computed. To this aim, we consider a generic random variable x non-elementary-distributed over the domain $x \in I \subseteq \mathbb{R}$ and described by the associated probability density function (PDF) $f(x)$. This includes the case where the samples $f(x^j)$ of the PDF are known only numerically over a sequence of values $x^j \in I$ as it is the case for PV sources. The gPC basis functions associated to such a variable x are a set of polynomials $q_i(x)$ that are orthonormal with respect to the inner product

$$\langle q_i(x), q_j(x) \rangle = \int_I q_i(x)\, q_j(x)\, f(x) dx = \delta_{ij}, \tag{5}$$

where i and j are the polynomial degrees and δ_{ij} denotes the Kronecker delta operator. Such polynomials can be obtained by first calculating a set of monic

orthogonal polynomials $\pi_i(x)$ through the three-term recurrence relation

$$\begin{aligned}\pi_{i+1}(x) &= (x-\alpha_i)\,\pi_i(x) - \beta_i \pi_{i-1}(x), \quad i = 0,\ 1, \cdots \\ \pi_{-1}(x) &= 0, \quad \pi_0(x) = 1,\end{aligned} \tag{6}$$

where α_i and β_i are real positive constants that depend on the PDF $f(x)$. A numerically stable algorithm for computing such coefficients is achieved by combining the iterative Darboux's formula

$$\alpha_i = \frac{\int\limits_I x\pi_i^2(x)f(x)dx}{\int\limits_I \pi_i^2(x)f(x)dx},\ \beta_{i+1} = \frac{\int\limits_I \pi_{i+1}^2(x)f(x)dx}{\int\limits_I \pi_i^2(x)f(x)dx},\ \ i = 0, 1, \cdots, \tag{7}$$

with the recurrence relation (6) and initialization $\beta_0 = 1$. Hence, the first $\hat{n}$ gPC basis functions are deduced as follows:

$$q_i(x) = \frac{\pi_i(x)}{\sqrt{\beta_0\beta_1\cdots\beta_i}},\ \text{for } i = 0, 1, \cdots, \hat{n}. \tag{8}$$

In our numerical implementation, without loss of generality, we suppose that the non-elementary-distributed variables x are defined over the normalized interval $I = [0, 1]$. The samples $f(x^j)$ of the PDF, are known over a sequence of $N_s + 1$ equally-spaced values $x^j = j\,\Delta x$ in I with $j = 0, 1, \ldots, N_s$ and $\Delta x = 1/N_s$.

The computations in (7) always reduces to calculating integrals of the form

$$\int\limits_I p(x)f(x)dx \tag{9}$$

where $p(x)$ is a polynomial known in closed form (i.e., known through its coefficients) while the PDF $f(x)$ is available in numerical form at the $N_s + 1$ sample points x^j. In our implementation, the integral (9) is evaluated numerically through the Simpson's integration formula. To this aim, we require that N_s is even, i.e. $N_s = 2\,N_{int}$ where N_{int} is the number of intervals of Simpson integration. Integral (9) is then calculated as

$$\begin{aligned}\frac{\Delta x}{3}\sum_{s=1}^{N_{int}} & p(x^{2s-1})\cdot f(x^{2s-1}) \quad + \\ & 4\cdot p(x^{2s})\cdot f(x^{2s}) \quad + p(x^{2s+1})\cdot f(x^{2s+1}).\end{aligned} \tag{10}$$

Under the mild hypothesis $\hat{n} << N_s$, we checked that the recurrence relation (6) implemented with the Simpson integration formula (10) provides a quite stable procedure for generating orthogonal polynomial functions. Consider for example, the Beta-distributed variable $y \in [0, 1]$ describing residential power demand through the PDF $f_\beta(y)$ in (2) with parameters $a = 1.1$ and $b = 22.8$, as shown in Fig. 2. Using $N_s = 256$ samples, we obtain the following polynomials sequence

$$\begin{aligned} q_0(y) &= 1 \\ q_1(y) &= 10.8 \quad +11.9y \\ q_2(y) &= 56.3 \quad +136.8y \quad +81.6y^2 \\ q_3(y) &= 182 \quad +742y \quad +977.8y^2 \quad +419y^3 \\ &\vdots \end{aligned} \tag{11}$$

It is easy verified that such polynomials correspond to the normalized Jacobi-chaos polynomials, as it should be for Beta distributed variables [21].

If, instead, the non-elementary distributed variable x describing PV generation with the PDF samples shown in Fig. 5 (and available only numerically) is considered, the associated polynomials chaos are computed to be

$$\begin{aligned} q_0(x) &= 1 \\ q_1(x) &= 0.170 \quad +1.893x \\ q_2(x) &= -1.239 \quad +0.322x \quad +4.414x^2 \\ q_3(x) &= -0.252 \quad -4.072x \quad +1.025x^2 \quad +8.826x^3 \\ &\vdots \end{aligned} \tag{12}$$

In this way, the correct polynomials are associated to each random variable describing either PV generation or customers power demand.

We conclude this Section by addressing another crucial issue connected with gPC methods. In fact, the implementation of gPC method with stochastic testing (ST) (that we will describe in the next Section) requires that polynomial basis (8) be evaluated in correspondence of a set of a few testing points. The number $\hat{n} + 1$ of the testing points for each variable is such $\hat{n} << N_s$. The selection of the testing points has a strong impact on the accuracy of the gPC

(19) approximation. In our implementation, the testing points for each random variable x are selected in order to preserve the accuracy when calculating the polynomial inner product (5). To this aim, testing points are selected among Gauss quadrature points (or nodes) $\hat{x}^j \in \mathbb{R}$, with $j = 1, \ldots, \hat{n}+1$. Gauss quadrature in fact guarantees that the calculation of integrals of the type

$$\int_{\mathbb{R}} g(x)f(x)dx \approx \sum_{j=1}^{\hat{n}+1} g(\hat{x}^j)w^j, \tag{13}$$

where w^j denote the Gauss weights, is exact for any polynomial $g(x)$ of degree $\leq 2\hat{n}+1$ [22, 23].

Gauss nodes and weights can be derived by ordering the coefficients in (7) in the following symmetric tridiagonal matrix

$$J = \begin{bmatrix} \alpha_0 & \sqrt{\beta_1} & & & \\ \sqrt{\beta_1} & \alpha_1 & \ddots & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \alpha_{\hat{n}-1} & \sqrt{\beta_{\hat{n}}} \\ & & & \sqrt{\beta_{\hat{n}}} & \alpha_{\hat{n}} \end{bmatrix} \tag{14}$$

The eigenvalue decomposition of $\mathbf{J}$ results

$$\mathbf{J} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{U}^T \tag{15}$$

where $\boldsymbol{\Sigma}$ is a diagonal matrix while $\mathbf{U}$ is a unitary matrix. The jth diagonal element of Σ provides the Gauss node $\hat{x}^j$ while the jth element of the first row of $\mathbf{U}$, i.e. $u_{1,j}$ is such that $w^j = u_{1,j}^2$.

As an example, for the Beta-distributed variable $y \in [0\,1]$ with PDF shown in (2), and adopting polynomial numbers $\hat{n} = 3$, the Gauss nodes are found to be

$$\hat{y}^1 = 0.0136, \quad \hat{y}^2 = 0.0671, \quad \hat{y}^3 = 0.1624, \quad \hat{y}^4 = 0.3054 \tag{16}$$

Similarly, for the non-elementary-distributed variable x with PDF shown in Fig. 5, the Gauss nodes are

$$\hat{x}^1 = 0.1008, \quad \hat{x}^2 = 0.3132, \quad \hat{x}^3 = 0.6587, \quad \hat{x}^4 = 0.8783 \tag{17}$$

Intuitively, it is seen how the Gauss nodes tend to be concentrated where the PDF is large, i.e. around the random variable values that have the largest occurrence probability. Finally, for multi-variable problems the grid of Gauss nodes is obtained by performing the tensor product of the univariate nodes. For l random variables and polynomial order $\hat{n}$, the tensor product results in a grid of $\hat{n}^l$ multi-variable Gauss nodes.

4. Probabilistic Analysis with gPC Method

We consider a LVDN containing N_L loads of the type (1) described by the Beta-distributed variables $y_r \in [0,1]$ with $r = 1, \ldots, N_L$. Each y_r represents the power demand of one residential customer over a given considered *time window* of the day. The network also contains PV sources described by N_S non-elementary distributed variables $x_s \in [0,1]$ of the type (4), with $s = 1, \ldots, N_s$. In this case, a single x_s variable represents the statistical uncertainty of solar irradiance in a given geographic area and thus it scales all the PV sources that are geographically close. The random variables can be collected into the vector of size $N_L + N_S$

$$\vec{\xi} = [y_1, \ldots, y_{N_L}, x_1, \ldots, x_{N_s}]. \tag{18}$$

Such vector variable represents the input of the probabilistic problem.

Hence, we focus on a set of N_{out} *observable variables* of the LVDN that can affect the quality of the network and denote them as V^j, with $j = 1, \ldots, N_{out}$. Such variables include some node voltages or line currents in the LVDN. They are seen as the output of the probabilistic problem.

Each realization of the random variables y_r and x_s in $\vec{\xi}$ corresponds to well determined load and source power values and thus to well determined node voltage values calculated by solving the deterministic load flow problem. As a result, the jth observation variable $V^j(\vec{\xi})$ (e.g. a node voltage) is a nonlinear function of the random variables $\vec{\xi}$ and thus it is a random variable as well.

The generalized polynomial chaos (gPC) method consists in approximating the deterministic relationship that exists between vector $\vec{\xi}$ and each observation

variable with an order-γ truncated series expansion of the type [21]

$$V^j(\vec{\xi}) \approx \sum_{i=0}^{N_b-1} c_i^j \, H_i(\vec{\xi}), \tag{19}$$

formed by N_b multi-variate basis functions $H_i(\vec{\xi})$ weighted by unknown coefficients c_i^j.

Each multi-variate basis function is given by the product

$$H_i(\vec{\xi}) = \prod_{r=1}^{N_L} q_{i_r}(y_r) \cdot \prod_{s=1}^{N_S} q_{i'_s}(x_s) \tag{20}$$

where $q_{i_r}(y_r)$ are the orthogonal polynomials of degree i_r associated to residential consumer power demand while $q_{i'_s}(x_s)$ are the orthogonal polynomial of degree i'_s associated to PV generation.

The multi-variate basis functions (20) can also be referred to by means of the index vector $\vec{i} = [i_1, \ldots, i_{N_L}, i'_1, \ldots, i'_{N_S}]$ collecting the univariate polynomial degrees i_r and i'_s. For a given number of parameters $N_L + N_S$ and series expansion truncation order γ, the index degrees i_r and i'_s of the univariate polynomials in (20) forming $H_i(\vec{\xi})$ should satisfy the following constraint

$$|\vec{i}| = \sum_{r=1}^{N_L} i_r + \sum_{s=1}^{N_S} i_s \leq \gamma. \tag{21}$$

There is a one to one correspondence between the scalar index $0 \leq i \leq N_b - 1$ and the index vector $\vec{i}$. As a consequence, the gPC expansion (19) can be rewritten in vector index notation as follows [22]:

$$V^j(\vec{\xi}) \approx \sum_{|\vec{i}| \leq \gamma} c_{\vec{i}}^j \, H_{\vec{i}}(\vec{\xi}), \tag{22}$$

Without loosing generality, we can select $i = 0$ to correspond to $|\vec{i}| = 0$, meaning that $H_0(\vec{\xi}) = 1$.

For given truncation order γ and number of parameters $N_L + N_S$, the number of gPC basis functions in (19) is given by

$$N_b = \frac{(\gamma + N_L + N_S)!}{\gamma! \, (N_L + N_S)!}. \tag{23}$$

The expansion coefficients c_i^j in the series (19) [or equivalently $c_{\vec{i}}^j$ in (22)] can be calculated with a technique referred to as Stochastic Testing (ST) method. According to ST method the N_b unknown coefficients c_j in the series (19) are calculated by selecting $N_{sam} = N_b$ testing points $\vec{\xi}^k$, for $k = 1, \ldots, N_{sam}$ among the multi-variable Gauss quadrature nodes. In each one of the testing points, the observation variable $V_k(t) = V(\vec{\xi}^k)$ is evaluated by running a deterministic LF analysis. Hence, the series expansions (19) are enforced to fit *exactly* (i.e., the polynomials interpolate the samples) the values V_k^j at the testing points. For the jth observation variable, this results in the following linear system

$$\mathbf{M}\,\vec{c}^{\,j} = \vec{V}^j, \tag{24}$$

where $\vec{c}^{\,j} = [c_0^j, \ldots, c_{N_b-1}^j]^T$ and $\vec{V}^j = [V_1^j, \ldots, V_{N_s}^j]^T$ are the column vectors collecting the unknown coefficients and observation variable values respectively.

The $N_b \times N_b$ square matrix $\mathbf{M} = \{a_{k,i}\} = \{H_i(\vec{\xi}^k)\}$ collects the gPC basis functions evaluated at the testing points, i.e.

$$\mathbf{M} = \begin{bmatrix} H_0(\vec{\xi}^1) & \ldots & H_{N_b-1}(\vec{\xi}^1) \\ \vdots & \ddots & \vdots \\ H_0(\vec{\xi}^{N_s}) & \ldots & H_{N_b-1}(\vec{\xi}^{N_s}) \end{bmatrix}. \tag{25}$$

It is worth observing that matrix $\mathbf{M}$, sometimes referred to as the experiment matrix, remains the same for each observation variable, so it is precalculated, inverted and used for any j as follows:

$$\vec{c}^{\,j} = \mathbf{M}^{-1}\,\vec{V}^j. \tag{26}$$

Once the coefficients c_i^j are computed, the mean value and variance of the jth observation variable $V^j(\vec{\xi})$ can easily be deduced as follows:

$$\begin{aligned} \mu_j &= \langle V^j(\vec{\xi})\rangle = c_0^j \\ \sigma_j^2 &= \langle V^j(\vec{\xi}) \cdot V^j(\vec{\xi})\rangle = \sum_{i=1}^{N_b-1} (c_i^j)^2. \end{aligned} \tag{27}$$

Similarly, in view of orthogonality (5), the covariance of two observation variable

$V^j(\vec{\xi})$ and $V^k(\vec{\xi})$ result:

$$\langle (V^j(\vec{\xi}) - \mu_j) \cdot (V^k(\vec{\xi}) - \mu_k) \rangle = \sum_{i=1}^{N_b - 1} c_i^j \, c_i^k. \tag{28}$$

In addition, and even more importantly, the gPC expansion (19) provides a surrogate compact model for the multi-dimensional relationship $V^j(\vec{\xi})$ that links observation variables to random parameters.

The compact gPC model (19) can then be used in connection with the MC method in order to determine the detailed PDF shape of $V^j(\vec{\xi})$. This is achieved by generating a very large number N_{mc} of uncertainty vectors $\vec{\xi}^k = [y_1^k, \ldots, y_{N_L}^k, x_1^k, \ldots, x_{N_s}^k]$, i.e. the realizations, accordingly to the joint probability distribution of variables in $\vec{\xi}$. For each realization $\vec{\xi}^k$, the corresponding realization of the observation variable $V^j(\vec{\xi})$ is evaluated by means of (19) in a numerically efficient way (much more efficiently than running a LF analysis). As the number N_{mc} of evaluations grows, at limit tending to infinity, the distribution of values calculated with the gPC model tends to the statistical distribution of $V^j(\vec{\xi})$. As a result, the detailed PDF shape of $V^j(\vec{\xi})$ can be determined in very short times, i.e. one million evaluations take a few seconds on a quad-core computer. In Algorithm 1, we summarize the main computational steps of the proposed data-driven uncertainty analysis method.

5. Sensitivity Analysis using gPC

From the gPC expansion (19) it is possible to extract another useful piece of information, i.e. the sensitivity of the observable variable $V^j(\vec{\xi})$ with respect to parameters y_r and x_s. As an instance, let us focus on the rth parameter y_r, the associated sensitivity is defined as:

$$S_r^j = \left. \frac{\partial V(\vec{\xi})}{\partial y_r} \right| \vec{\xi} = \vec{0} \tag{29}$$

Algorithm 1 Data-Driven Uncertainty Analysis

Data Analysis:
Fix a *time window*, extract data for users y and PV sources x
Compute PDFs $f_x(x)$ and $f_y(y)$
Build univariate gPC $q(x)$ and $q(y)$ with (5)–(7)
Determine univariate Gauss quadrature nodes $\hat{x}^j$, $\hat{y}^j$,
Network Analysis:
In the LVDN: fix N_L loads y_r and N_S sources x_s and gPC order γ
Form N_b multi-variate basis $H_i(\vec{\xi})$ (20)
Select N_b testing points $\vec{\xi}^k$ among tensor product of Gauss quadrature nodes
Form and invert experiment matrix (25)
For each testing point $\vec{\xi}^k$ run a Deterministic Load Flow and store observable variables $V^j(\vec{\xi}^k)$
Output
for $j = 1 : N_{out}$ **do** (for each observable variable V^j)
    find gPC coefficients c_i^j (26)
    calculate μ_j, σ_j, PDF, sensitivity S_r^j

In order to compute the sensitivity, the product terms in the expansion (22) are reordered in the following way:

$$V^j(\vec{\xi}) \approx \sum_{\vec{i}=[0,0,i_r,0,\ldots,0]} c^j_{\vec{i}}\, H_{\vec{i}}(\vec{\xi}) + \sum_{\vec{i}=[\ldots,i_{r-1},0,i_{r+1},\ldots]} c^j_{\vec{i}}\, H_{\vec{i}}(\vec{\xi}), \tag{30}$$

where the first sum includes those products containing polynomial $q_{i_r}(y_r)$ of degree $i_r > 0$ and all of the other polynomials of degree zero, i.e. $H_{\vec{i}}(\vec{\xi}) = q_{i_r}(y_r)$, while the second sum includes the remainder. Only the first term in (30) contributes to sensitivity which thus results:

$$S^j_r = \sum_{\vec{i}=[0,0,i_r,0,\ldots,0]} c^j_{\vec{i}}\, \frac{\partial q_{i_r}(y_r)}{\partial y_r} \Bigg|_{\vec{\xi}=\vec{0}} \tag{31}$$

The general form of polynomial $q_{i_r}(y_r)$ is

$$q_{i_r}(y_r) = 1 + a^{i_r}_1 y_r + a^{i_r}_2 y_r^2 + \cdots, \tag{32}$$

so that the derivative in (31) reduces to

$$\frac{\partial q_{i_r}(y_r)}{\partial y_r} \Bigg|_{\vec{\xi}=\vec{0}} = a^{i_r}_1 \tag{33}$$

i.e. it corresponds to the coefficient of $q_{i_r}(y_r)$ that multiplies y_r. A similar derivation holds for computing the sensitivity of the observable variables with respect to parameters x_s.

6. Numerical Results

Power distribution networks have a radial, or weakly meshed, topology made of several feeders. Commonly, remote controlled switches are deployed along each feeder in order to isolate small subfeeder areas in case of fault or with the aim of minimizing power loss [28]. For such reasons, it makes sense to investigate loads uncertainty effects and PV penetration by focusing on small subnetworks arranged in a single or a few feeders. In our numerical examples, we present four different Cases. In the first two Cases, we focus on the single-feeder three-phase symmetrical distribution network shown in Fig. 7. The network can

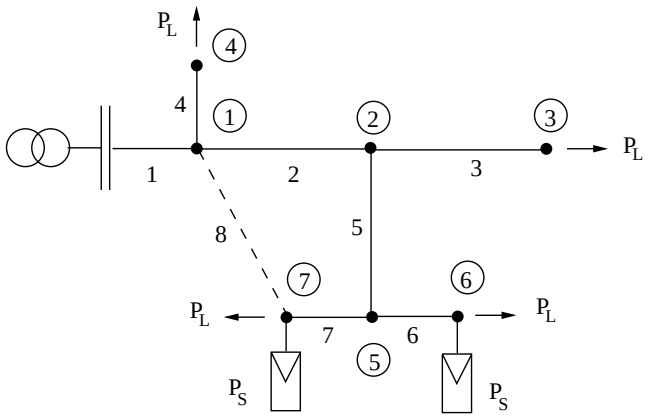

Figure 7: Single-feeder network. The network topology is radial or weakly meshed. At nodes 6 and 7 it is possible to have both load and PV generation.

Table 1: The line parameter used in the example. R_1 and X_1 are the series parameter

	Area $[mm^2]$	N Phase [-]	length [m]	R_1 $[\Omega/km]$	X_1 $[\Omega/km]$
Branches 1, 4, 6, 7	3· 50 + 25Cu	3+N	100	0.391	0.078
Branches 2, 3, 5, 8	3· 50 + 25Cu	3+N	200	0.391	0.078

be reconfigured in a weakly meshed one by switching-on the branch number 8 (shown in the figure with a dashed line). In the first Case, four users are connected to the network and no PV generation is considered. In the second Case, in addition to the four loads, two PV sources are connected to the terminal nodes of the net. In the third Case, the analysis is extended to the three-feeder unbalanced topology shown in Fig. 17. In particular, we investigate the effect that single-phase PV source penetration can have on branch current and three-phase voltage unbalance. **Finally, in the fourth example, we consider a larger distribution network and evaluate the scalability of the gPC method and its performance compared to reference MC method.**

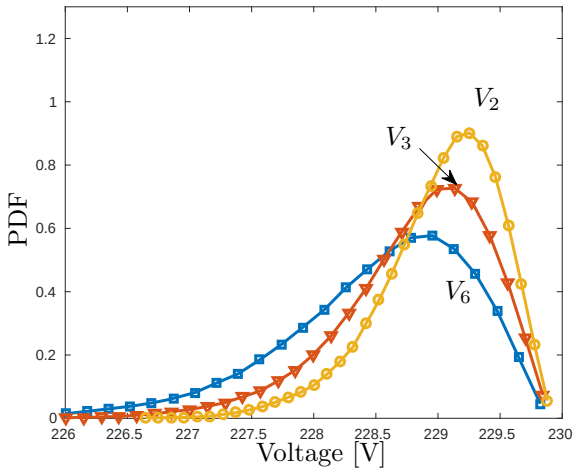

Figure 8: PDFs of some nodal voltages magnitude in the single-feeder radial network (with no PV source).

6.1. Case I: Single-feeder Network with only loads

The network in Fig. 7 is formed of seven branches connected to seven nodes. Node 0 is the slack bus and its voltage is fixed to $V_0 = 230$ V. A weakly mesh network is simply obtained from the radial one by adding the branch number 8 that forms a loop. The parameters of the lines are reported in Tab. 1. In the first scenario, four users are connected to the network at nodes 3, 4, 6 and 7. Each user absorbs a random power P_{L_r}, with $r = 1, \ldots, 4$, of the type (1) with peak value $K_L = 10$ kW scaled by the Beta-distributed variables y_r. In simulations, the Beta-distributed variables have the parameter values $a = 1.1$ and $b = 22.8$ extracted in Sec. II for residential consumers over the time window 10:00 AM–02:00 PM. We apply the gPC method outlined in Algorithm 1 with $N_L = 4$ random variables and adopting truncation order $\gamma = 3$. More specifically, the univariate polynomials $q_{i_r}(y_r)$ forming gPC expansion (19) are the univariate Jacobi-chaos polynomial (11) associated to Beta-distributed variables. Fig. 8 shows the computed statistical distribution of the magnitude of some nodal voltages in the radial network. Uncertain power consumption by loads results in a reduction of voltages compared to slack bus. In particular, voltage V_6 at

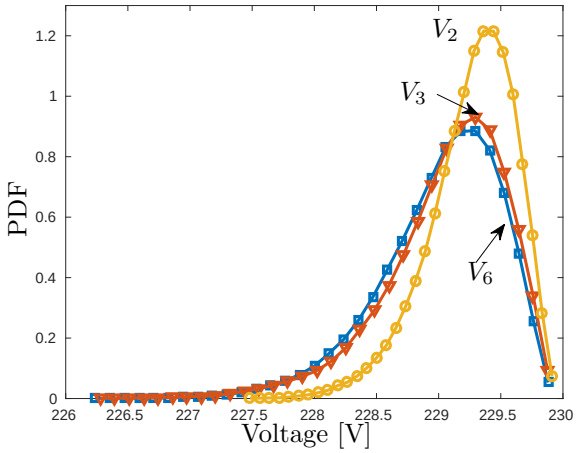

Figure 9: PDFs of some nodal voltages magnitude in the single-feeder meshed network (with no PV source).

terminal node 6 exhibits the largest variability, i.e. it ranges within the interval $\approx (226, 230)V$, compared to the voltages at terminal node 3 and conjunction node 2. Fig. 9 shows the statistical distribution of the same nodal voltages in the weakly mesh network. Compared to the radial network, the interval of variability of voltages reduce. For instance, in the mesh network, V_6 ranges within $\approx (228, 230)V$. Fig. 10 illustrates the statistical distribution of the magnitude of some branch currents in the radial network. Interestingly, current distribution in the terminal branch 6 closely resembles the Beta-like distribution of power load. Currents in the internal branches 5 and 2 have larger variability intervals and, in addition their PDFs differ from the Beta distribution of the single load tending to a sort of Gaussian.

In Fig. 11, we report the statistical distribution of some branch currents in the mesh network. Compared to radial network, terminal current I_6 preserves its distribution while currents I_5 and I_2 in the internal branch exhibit a reduction of their variability interval. The analysis of this first Case leads us to this interesting conclusion: the presence of the loop in the mesh network reduces the interval of variability of both nodal voltages and branch currents induced

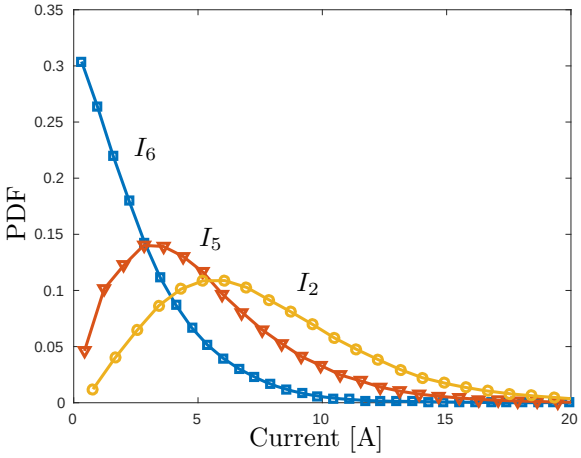

Figure 10: PDFs of some branch current in the single-feeder radial network (with no PV source).

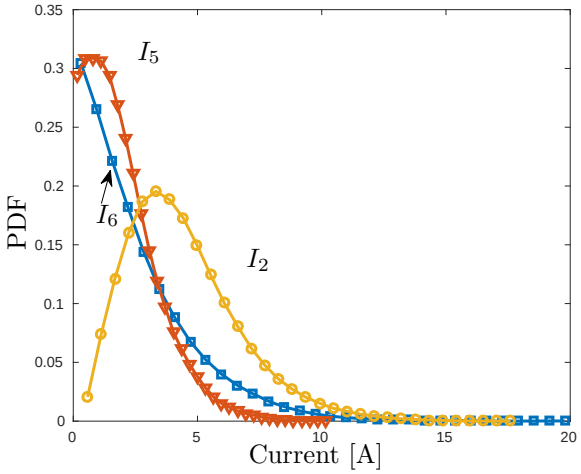

Figure 11: PDFs of some branch current in the single-feeder meshed network (with no PV source).

by uncertain loads.

6.2. Case II: Single-feeder Network with loads and PV generation

In the second Case, in addition to the four loads, two PV sources are connected to nodes 6 and 7. The two PV sources are described by the model (4) where the peak delivered power K_S is scaled by a random variable x_1 describing solar irradiance uncertainty. Variable x_1 is assumed to be distributed as shown in Figs. 5 for the time window 10:00 AM–02:00 PM. Growing values of the peak source power K_S are considered in order to evaluate the effect of PV penetration.

In this case, the gPC model has $N_L + N_S = 4+1 = 5$ random variables. The univariate polynomials $q_{i_r}(y_r)$ univariate Jacobi-chaos polynomial (11) are associated to the four Beta-distributed variables while the orthogonal polynomials $q_{i'_s}(x_s)$ (12) are associated to the PV variable x_s with $s = 1$.

Fig. 12 illustrates the computed statistical distribution of the voltage magnitude V_6 in the radial network for PV power peak values $K_S = 3$kW and $K_S = 6$kW. The distributed PV generation introduces several changes with respect to the scenario with only loads. In fact, the probability distribution of V_6 moves towards voltage values greater than 230 V (interval of variability $\approx 228 - 242$ V) giving rise to potential over voltage. Interestingly, in the mesh network, the effect of PV generation is less relevant compared to the radial net as shown in Fig. 13.

Similar observations hold for currents. Fig. 14 shows the effect that PV generation has on current I_2 in the radial network. For $K_S = 6$kW, the current module has a wider interval of variability that ranges from zero to 50A.

The effect that PV sources has on currents is less relevant in the mesh network as shown in Fig. 15. In this case, current I_2 has zero probability to exceed 22A. In view of this second Case study, we come to the conclusion that the presence of a loop in the mesh network mitigates the effects due to the penetration of distributed PV generation. A confirmation of this is also obtained by computing the sensitivity of the observable variables versus the

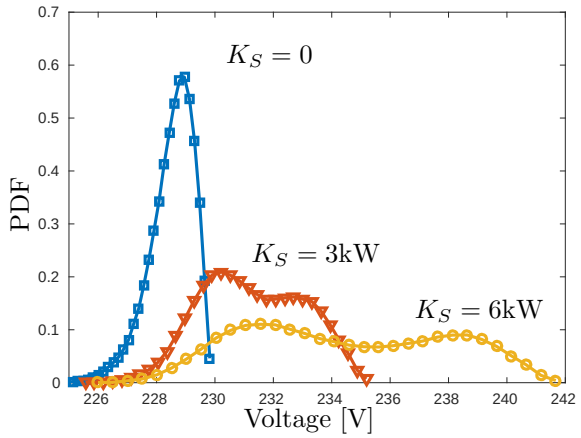

Figure 12: PDFs of nodal voltage V_6 in the single-feeder radial network in the presence of PV sources.

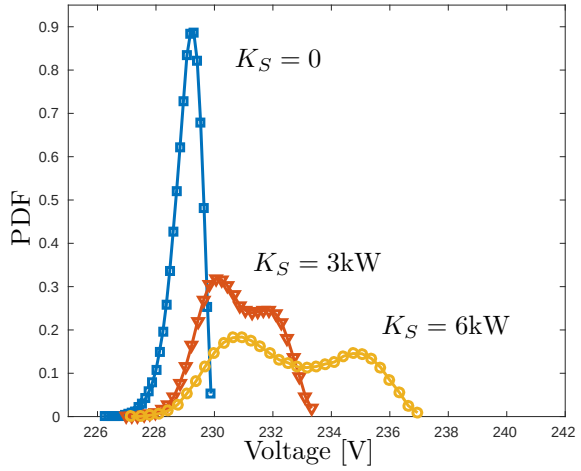

Figure 13: PDFs of V_6 in the single-feeder meshed network in the presence of PV sources.

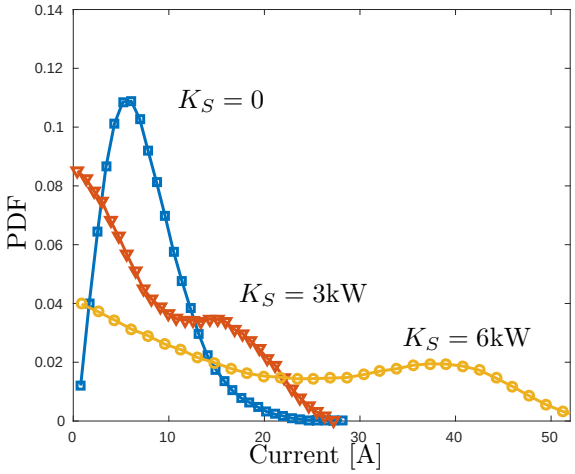

Figure 14: PDFs of current I_2 in the single-feeder radial network in the presence of PV sources.

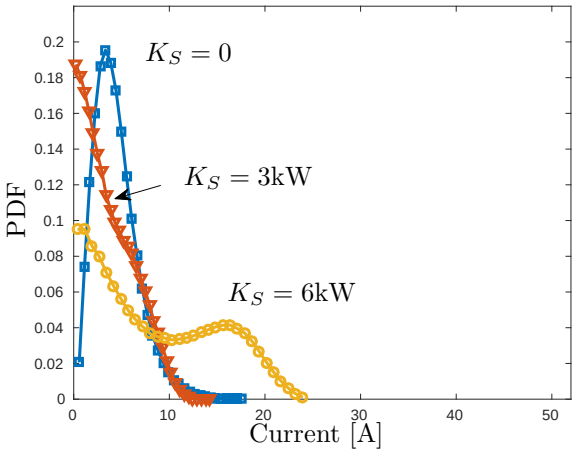

Figure 15: PDFs of I_2 in the single-feeder meshed network in the presence of PV sources.

Table 2: Table of sensitivity coefficients.

	Statistical Variables				
	y_1	y_2	y_3	y_4	x_1
	Radial Network				
V_6	-0.26	-0.08	-0.51	-0.42	3.38
I_2	-2.2	0.00	-2.15	-2.15	15.6
	Mesh Network				
V_6	-0.17	-0.12	-0.34	-0.21	2.04
I_2	-1.13	-0.27	-1.06	-0.82	6.97

stochastic variables as defined in (31). The sensitivity of variables V_6 and I_2 are reported in Tab. 2 for the radial and mesh networks. It is seen how a positive variation (increase) in the power absorbed by loads, represented by variables y_r, induces negative variations (reduction) in voltage and current. By contrast, an increase in the power delivered by PV sources represented by x_1, induces positive variations (increase) in voltage and current. The sensitivity versus PV generation dominates over sensitivity to loads and in the mesh network the module of such coefficients is smaller than in the radial one.

We end Case II by checking the accuracy of the gPC method by comparing the calculated PDFs with those obtained with the reference Monte Carlo method. Fig. 16 reports the PDFs of nodal voltage V_6 in the radial net with PV generation $K_S = 6$kW computed with the MC method with $2,000$, $5,000$ and $10,000$ iterations. With the graphical detail of Fig. 16, the PDF curve provided by MC method with $10,000$ iteration is indistinguishable from the PDF curve calculated with the proposed gPC method. For this simple network, the simulation times of gPC and MC (with $10,000$ iterations) are 3.3s and 315s respectively. The gPC method introduces a two orders of magnitude speed up compared to MC simulation for the same accuracy.

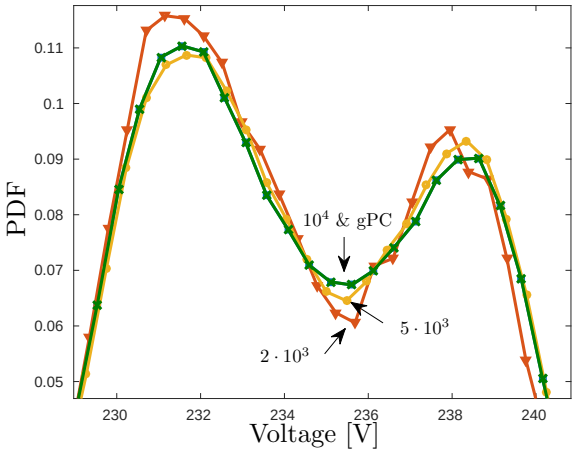

Figure 16: Detail of the PDFs of V_6 in the radial net with $K_S = 6$kW computed with Monte Carlo (for different numbers of iterations) and gPC method.

6.3. Case III: Unbalanced Three-feeder Network

In this example, we study the three-feeder topology shown in Fig. 17. Such a topology is a standard when studding network reconfiguration and loss reduction problems [28], [29]. In our implementation, the network is radial with 22 branches connected to 22 nodes, however it can be reconfigured into a weakly meshed topology by switching-on the branches shown with a dashed line. For simplicity all branches have the same length of 100m with per-unit series resistance and admittance $R_1 = 0.391$ $[\Omega/km]$, $X_1 = 0.078$ $[\Omega/km]$ respectively. Nine residential consumer loads are connected to the nodes 4, 6, 8 of the first feeder, 11, 12, 14 of the second feeder and 16, 21, 22 of the third feeder. The loads are three-phase with an active power P_{L_r}, with $r = 1, \ldots, 9$, of the type (1) with peak value $K_L = 10$ kW scaled by the Beta-distributed variables y_r. The three-phase power is equally divided among the three phase thus leading to a balanced network. In addition, two single-phase PV sources are connected to nodes 21 and 22 at phase C. The two PV sources are described by the model (4) where the peak delivered power $K_S = 6$kW is scaled by a random variable x_1 describing solar irradiance uncertainty distributed as shown in Figs. 5. The

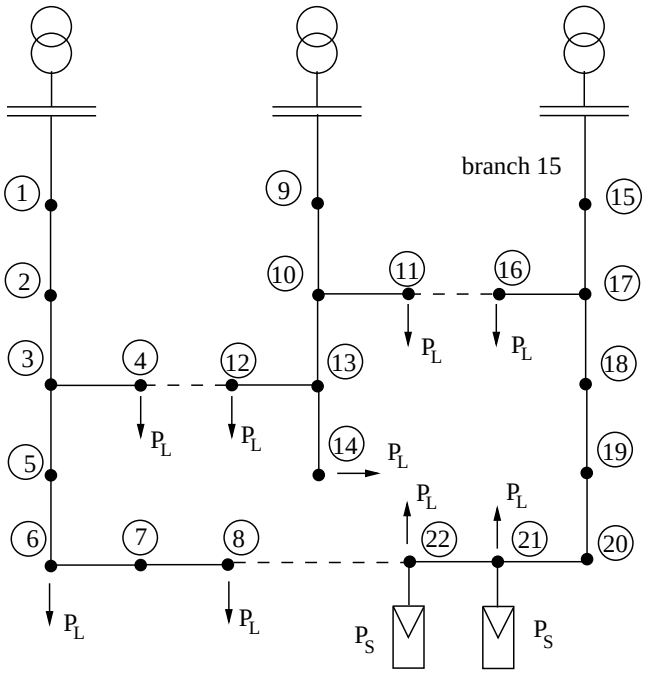

Figure 17: Three-feeder network. The network topology is radial or weakly meshed.

PV sources introduce unbalancing into the network.

We apply the gPC method explained in Algorithm 1 with $N_L + N_S = 9 + 1 = 10$ random variables. Fig. 18 reports the computed statistical distribution of the magnitude for the three phases of the internal branch 15 (highlighted in Fig. 17) in the radial network. It can be seen how the module of current into Phase C line (i.e. with PV generation) has a much wider variability interval than Phase A and B with only residential consumers loads. Fig. 19, instead, reports the same current distribution in the meshed topology: all currents reduce their variability intervals.

Finally, Figs. 20, 21 show the calculated statistical distribution of two remarkable figures of merit: the Voltage Unbalance Factor (VUF) (in percentage) computed at terminal node 22 and the total power loss in the three-feeder network. **The VUF is defined as the ratio of the negative voltage sequence**

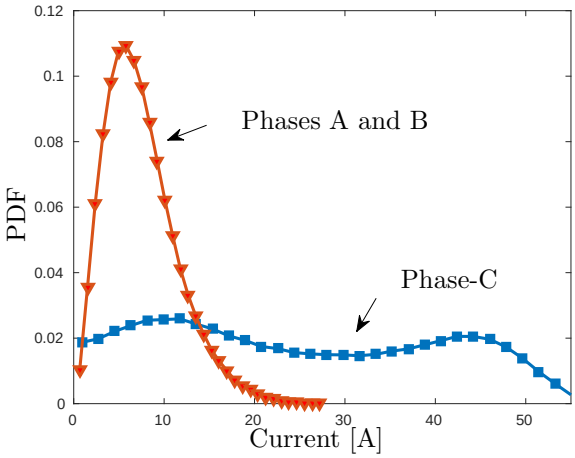

Figure 18: PDFs of phases A, B, C of branch 15 current in the three-feeder radial network.

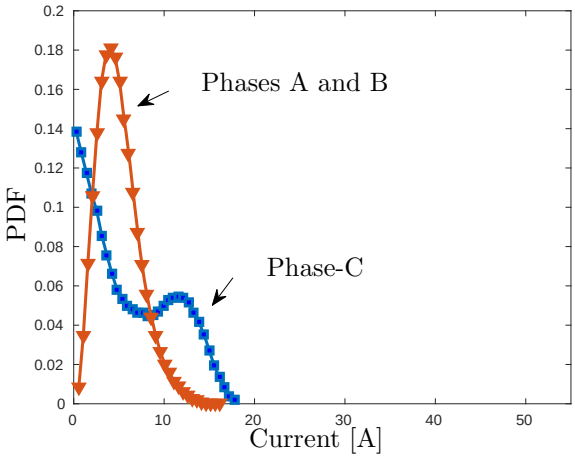

Figure 19: PDFs of phases A, B, C of branch 15 current in the three-feeder meshed network.

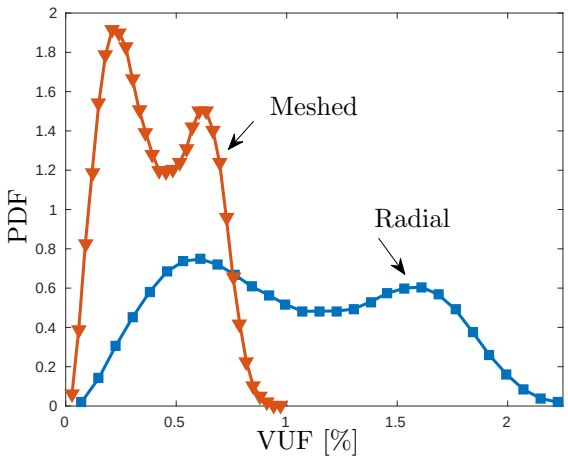

Figure 20: PDFs of VUF at node 22 in the three-feeder radial and meshed network.

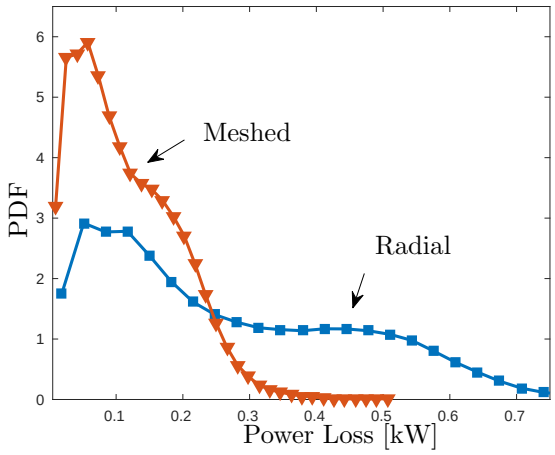

Figure 21: PDFs of total power loss in the three-feeder radial and (reconfigured) meshed network.

component V_n to the positive voltage sequence component V_p, i.e.

$$\mathbf{VUF} = \frac{|V_n|}{|V_p|} \cdot 100, \tag{34}$$

with

$$V_n = \frac{V_{AB} + a^2 \cdot V_{BC} + a \cdot V_{CA}}{3} \tag{35}$$

and

$$V_p = \frac{V_{AB} + a \cdot V_{BC} + a^2 \cdot V_{CA}}{3}, \tag{36}$$

where V_{AB}, V_{BC}, V_{CA} are the phasors of the unbalanced line voltages while $a = \exp(j\,120^o)$ and $a^2 = \exp(j\,240^o)$.

It can be seen how the reconfigured meshed topology mitigates significantly the unbalancing effect induced by uncertain PV generation (i.e., zero probability of VUF $> 1\,\%$) as well as it reduces the total power loss (i.e., zero probability of total loss > 0.3 kW).

6.4. Case IV: 141 bus distribution system

We end the numerical experiments by evaluating the performance of gPC method versus MC method in the probabilistic analysis of the benchmark 141bus distribution grid. Such a network is provided within the MatPower software suite and its description can be found in [30]. To this aim, we replace a number N_L of loads (active) power demand with random variables of the type (1) and perform probabilistic analysis with the MC method ($10,000$ iterations) and the enhanced gPC method. For both methods we use MatPower as the deterministic load flow solver. The simulation time required by MC method is $\approx 2,900$ s on a i7 Quad core computer and is almost independent of the number of parameters. By contrast, the simulation times taken by the gPC method grows with the number N_L of random parameters as reported in table 3. For the case $N_L = 50$ parameters, the speedup factor introduced by gPC method remains significant being $\approx 7\times$ for the same accuracy.

Table 3: gPC simulation times [s]

$N_L = 5$	$N_L = 10$	$N_L = 20$	$N_L = 50$
6	19	71	398

7. Conclusion

In this paper, we have presented a statistical methodology for the analysis of distribution networks in the presence of uncertain loads and PV generation. One key feature of our method is that of being based on realistic data-driven models for both PV generation and residential consumers power demand. In the paper, for illustrative reasons, we have employed a data set available for the city of London, however the method can be applied to any other data set. The statistical analysis that we have presented relies on the generalized Polynomial Chaos (gPC) method. However, we have shown how residential consumers power demand and power delivered by PV sources can follow non-standard/non-elementary statistical distributions. As a consequence, the known gPC method should be enhanced in order to handle such complex nonstandard scenarios. Method extensions include the computation of new sets of basis functions as well as of properly selected testing points. The enhanced method has been applied to both radial and weakly mesh networks in order to explore the probability distribution of nodal voltages and branch currents as a function of PV penetration. Other important figures of merit measuring network unbalance and power loss have been investigated. Simulation results highlight how meshed networks are more resilient than radial topologies against the adverse effects of loads/sources variability and PV penetration.

References

[1] D. Santos-Martin, S. Lemon, Simplified modeling of low voltage distribution networks for pv voltage impact studies, IEEE Transactions on Smart Grid 7 (2015) 1–8. doi:10.1109/TSG.2015.2500620.

[2] R. Tonkoski, D. Turcotte, T. H. M. EL-Fouly, Impact of high pv penetration on voltage profiles in residential neighborhoods, IEEE Transactions on Sustainable Energy 3 (3) (2012) 518–527. doi:10.1109/TSTE.2012.2191425.

[3] R. Markovič, M. Gosak, V. Grubelnik, M. Marhl, P. Virtič, Data-driven classification of residential energy consumption patterns by means of functional connectivity networks, Applied Energy 242 (2019) 506 – 515.

[4] N. Mohan, K. Soman, S. S. Kumar, A data-driven strategy for short-term electric load forecasting using dynamic mode decomposition model, Applied Energy 232 (2018) 229 – 244.

[5] J. Hernández, F. Ruiz-Rodriguez, F. Jurado, Modelling and assessment of the combined technical impact of electric vehicles and photovoltaic generation in radial distribution systems, Energy 141 (2017) 316–332.

[6] F. Ruiz-Rodriguez, J. Hernández, F. Jurado, Voltage behaviour in radial distribution systems under the uncertainties of photovoltaic systems and electric vehicle charging loads, International Transactions on Electrical Energy Systems 28 (2).

[7] S. Conti, S. Raiti, Probabilistic load flow using monte carlo techniques for distribution networks with photovoltaic generators, Solar Energy 81 (2007) 1473–1481. doi:10.1016/j.solener.2007.02.007.

[8] J. Jardini, C. M.V. Tahan, M. Gouvea, S. Ahn, F. Figueiredo, Daily load profiles for residential, commercial and industrial low voltage consumers, IEEE Transactions on Power Delivery 15 (2000) 375 – 380. doi:10.1109/61.847276.

[9] G. Gruosso, G. Gajani, Z. Zhang, L. Daniel, P. Maffezzoni, Uncertainty-aware computational tools for power distribution networks including electrical vehicle charging and load profiles, IEEE Access 7 (2019) 9357–9367.

[10] G. Gruosso, P. Maffezzoni, Z. Zhang, L. Daniel, Probabilistic load flow methodology for distribution networks including loads uncertainty, Inter-

national Journal of Electrical Power and Energy Systems 106 (2019) 392–400.

[11] J. Widén, M. Shepero, J. Munkhammar, Probabilistic load flow for power grids with high pv penetrations using copula-based modeling of spatially correlated solar irradiance, IEEE Journal of Photovoltaics PP (99) (2017) 1–6.

[12] G. Carpinelli, P. Caramia, P. Varilone, Multi-linear Monte Carlo simulation method for probabilistic load flow of distribution systems with wind and photovoltaic generation systems, Renewable Energy 76 (C) (2015) 283–295. doi:10.1016/j.renene.2014.11.

[13] C. Wu, F. Wen, Y. Lou, F. Xin, Probabilistic load flow analysis of photovoltaic generation system with plug-in electric vehicles, International Journal of Electrical Power & Energy Systems 64 (2015) 1221–1228. doi:10.1016/j.ijepes.2014.09.014.

[14] M. Kabir, Y. Mishra, R. Bansal, Probabilistic load flow for distribution systems with uncertain PV generation, Applied Energy 163 (C) (2016) 343–351. doi:10.1016/j.apenergy.2015.1.

[15] H. Zhang, Y. Xu, Probabilistic load flow calculation by using probability density evolution method, International Journal of Electrical Power & Energy Systems 99 (2018) 447 – 453. doi:https://doi.org/10.1016/j.ijepes.2018.01.043.

[16] J. M. Morales, J. Perez-Ruiz, Point estimate schemes to solve the probabilistic power flow, IEEE Transactions on Power Systems 22 (4) (2007) 1594–1601.

[17] G. Plattner, H. F. Semlali, N. Kong, Analysis of probabilistic load flow using point estimation method to evaluate the quantiles of electrical networks state variables, CIRED - Open Access Proceedings Journal 2017 (1) (2017) 2087–2091. doi:10.1049/oap-cired.2017.0179.

[18] P. Amid, C. Crawford, A cumulant-tensor-based probabilistic load flow method, IEEE Transactions on Power Systems 33 (5) (2018) 5648–5656. doi:10.1109/TPWRS.2018.2811707.

[19] M. Fan, V. Vittal, G. T. Heydt, R. Ayyanar, Probabilistic power flow studies for transmission systems with photovoltaic generation using cumulants, IEEE Transactions on Power Systems 27 (4) (2012) 2251–2261. doi:10.1109/TPWRS.2012.2190533.

[20] Z. Ren, W. Li, R. Billinton, W. Yan, Probabilistic power flow analysis based on the stochastic response surface method, IEEE Transactions on Power Systems 31 (3) (2016) 2307–2315.

[21] D. Xiu, G. E. Karniadakis, The wiener–askey polynomial chaos for stochastic differential equations, SIAM Journal on Scientific Computing 24 (2) (2002) 619–644.

[22] Z. Zhang, T. A. El-Moselhy, I. M. Elfadel, L. Daniel, Stochastic testing method for transistor-level uncertainty quantification based on generalized polynomial chaos, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems 32 (10) (2013) 1533–1545.

[23] Z. Zhang, T. A. El-Moselhy, I. A. Elfadel, L. Daniel, Calculation of generalized polynomial-chaos basis functions and gauss quadrature rules in hierarchical uncertainty quantification, Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on 33 (2014) 728–740. doi:10.1109/TCAD.2013.2295818.

[24] London, Smartmeter energy consumption data in london households. URL https://data.london.gov.uk/

[25] B. Stephen, A. J. Mutanen, S. Galloway, G. Burt, P. Järventausta, Enhanced load profiling for residential network customers, IEEE Transactions on Power Delivery 29 (1) (2014) 88–96.

[26] A. Papoulis, U. Pillai, Probability, random variables and stochastic processes, 4th Edition, McGraw-Hill, 2001.

[27] M. J. Sanjari, H. B. Gooi, Probabilistic forecast of pv power generation based on higher order markov chain, IEEE Transactions on Power Systems 32 (4) (2017) 2942–2952.

[28] C. Ten, Y. Tang, Electric Power: Distribution Emergency Operation, Taylor & Francis, a CRC title, part of the Taylor & Francis imprint, a member of the Taylor & Francis Group, the academic division of T&F Informa, plc, 2018.

[29] S. Civanlar, J. J. Grainger, H. Yin, S. S. H. Lee, Distribution feeder reconfiguration for loss reduction, IEEE Transactions on Power Delivery 3 (3) (1988) 1217–1223.

[30] H. Khodr, F. Olsina, P. D. O.-D. Jesus, J. Yusta, Maximum savings approach for location and sizing of capacitors in distribution systems, Electric Power Systems Research 78 (7) (2008) 1192 – 1203. doi:https://doi.org/10.1016/j.epsr.2007.10.002.