Statistical Analysis of PV penetration impact on Residential Distribution Grids

Harshavardhan Palahalli[c], Paolo Maffezzoni[c], Luca Daniel[d], Giambattista Gruosso[c,*]

[a] *Politecnico di Milano, Dipartimento di Elettronica, Informazione e Bioingegneria, Piazza Leonardo da Vinci, 32 -20133 Milano, Italy.*
[b] *Massachusetts Institute of Technology (MIT), Cambridge, MA 02139, USA (e-mail: luca@mit.edu)*

Abstract

This paper presents a comprehensive apprach to the probabilistic analysis of residential distribution grid injected by distributed PV sources. The approach is data-driven and is able to deal with the general scenario that includes the uncertainty of correlated PV sources and of statistically independent consumer loads. The novel method adopts a Gaussian-Mixture-Model for correctly representing PV sources correlation and a fresh probabilistic analysis techinque employing Multi-Expansion polynomial chaos. Numerical experiments carried out on the Non-Synthetic European low voltage test system, highlight the importance of the comprehensive modeling strategy in order to realistically quantify uncertainty impact on the grid.

Keywords: power grids, Photovoltaic penetration, Polynomial chaos, Probabilistic analysis, Simulations

---

*Corresponding author
Email address: `giambattista.gruosso@polimi.it` (Giambattista Gruosso)

Statistical Analysis of PV penetration impact on Residential Distribution Grids

Harshavardhan Palahalli[c], Paolo Maffezzoni[c], Luca Daniel[d], Giambattista Gruosso[c,*]

[c] *Politecnico di Milano, Dipartimento di Elettronica, Informazione e Bioingegneria, Piazza Leonardo da Vinci, 32 -20133 Milano, Italy.*
[d] *Massachusetts Institute of Technology (MIT), Cambridge, MA 02139, USA (e-mail: luca@mit.edu)*

1. Introduction

It is expected that the massive integration of photovoltaic (PV) distributed generation into residential power distribution grid will have potential adverse effects on network quality, such as over voltage or current exceeding the wire carrying capacity [1], [2]. Evaluating the impact of PV sources is a challenging issue due to the unpredictability of both PV power generation and residential consumers power demand [3]. A detailed analysis requires the adoption of realistic stochastic models for the PV sources and loads as well as the exploitation of advanced probabilistic computational tools [4], [5], [6]. Stochastic models for PV generation are mainly focused on reproducing the temporal and spatial variability of solar irradiance [7]. In fact, spatial correlation between geographically dispersed PV generation sites can play an important role in determining the relevance of their impact on the grid lines. For what concerns residential loads, several statistical models have been proposed in the literature in order to reproduce their uncertainty. Many previous models are based on the assumption that uncertainty in residential power demand can be represented by Gaussian-distributed mutually independent random variables [8], [9].

*Corresponding author
Email address: `giambattista.gruosso@polimi.it` (Giambattista Gruosso)

In this paper, we present a novel data-driven modeling approach to uncertainty quantification in PV penetrated residential grid. The new approach can handle the *general statistical scenario* where a subset of uncertain variables, i.e. PV delivered power, can be strongly correlated among them while another subset, i.e. residential load power demands, are statistically independent. This is achieved by adopting a proper Gaussian Mixture Model (GMM) to represent the joint PDF of correlated PV sources and a parametric statistical model for loads.

Moving on considering probabilistic analysis computation, the basic and reference method is represented by Monte Carlo simulation. However MC method is computationally time consuming and thus several approximate probabilistic analysis techniques have been proposed, which includes Point Estimate Method, Cumulant, Kernel and Surrogate Model [10], [11], [12], [13]. Among all these techniques, Stochastic Response Surface Method (SRSM) based on polynomial chaos expansions has recently gained large attention due to its versatility in handling random variables with nonstandard statistical distributions [14], [15], [16] as well as, in the field of power distribution grid, due to the comprehensive statistical information (i.e. the detailed Probability Density Function) it can provide about observable quantities of interest [17], [18].

In this paper, we extend the usage of SRSM so as to deal with the general uncertainty modeling scenario. To this aim, the SRSM surrogate model is implemented through a novel technique referred to as Multi-Expansion generalized Polynomial Chaos (ME-gPC).

The novel contributions of this paper include:

1. We describe a data-driven modeling of correlated PV sources based on Bayesian estimation and Gaussian Mixture Model (GMM). Our approach relies on the analysis of extensive data set of measurements grouped for different daily *time windows.* As a practical case study, we exploit the data set available for Australian consumer site [19] however, our approach is general and can be applied to any data set.

2. We extend the modeling approach to add the uncertainty of residential consumers power demand. We show how the power demand for a single user (or a small set of users) over a given time window of the day is well represented by a Beta-distributed random variable whose PDF can be extracted from data using a parametric approach.
3. We describe in details the ME-gPC technique able to successfully handle the case of correlated non Gaussian random variables described by GMMs. We remark how the Multi-Expansion method that we first present in this paper is similar but not equal to the method known in the mathematical community and denoted as Multi-Element method [20]. The Multi-Element method has a much more complicated implementation, i.e. it requires to break the statistical space into mutually disjoint regions, than the Multi-expansion technique described in this paper. As a result, Multi-Expansion method is more suitable for probabilistic power grid analysis.
4. We use the (ME-gPC) method for calculating the detailed Probability Density Function (PDF) of a set of observable quantities that are relevant for power system designers. In particular, we consider and compare different PV modeling scenarios, i.e. with or without including PV sources correlation, applied to the Non-Synthetic European low voltage test system [21].

The remainder of this paper is organized as follows: In Sec. II, we illustrate the data-driven GMM modeling strategy for PV sources. In Sec. III, the statistical model for residential loads extracted from measurement data is outlined. Sec. IV details the novel ME-gPC method. Sec. V is used to describe the proposed polynomial chaos based implementation while Sec. VI presents numerical experiments, Finally, the conclusions are outlined in Sec. VII.

2. Modeling Photovoltaic Power Generation

This Section focuses on statistical modeling of correlated PV sources starting from data measured over a sufficiently long time. The method we propose has general validity and can be used to create statistical models of sources (as well as loads) exhibiting correlated uncertainty. As an example to clarify our procedure, we choose the data set of photovoltaic sources [19] provided by the Australian electric distribution company Ausgrid. It contains measured electricity production for rooftop solar systems from 1^{st} July 2012 to 30^{th} June 2013.

In the central hours of the day, the power samples extracted from the data set exhibit a significant variability which introduces a high degree of uncertainty in the generated electric power. This can be due to several factors that include the time-spatial variability in solar irradiation as well as solar cells efficiency, orientation, shadowing effects, and variable interaction with the electrical interface/control. In addition, PV generated power variability has a strong dependence on the time of day which is considered. In order to deal with such a time-dependence, the available measurement are grouped into several sets corresponding to different time windows during the day. For example, we consider windows width of one hour in which the statistical distribution (i.e. the PDF) of generated electrical power is evaluated.

To better clarify this concept, let us consider the electrically delivered power $P_{S_1}(t)$ and $P_{S_2}(t)$ by two close PV plants (in p.u.) placed in the Sydney city over the time window from 01:00 PM to 02:00 PM. Fig. 1 reports the scattered plot of the $P_{S_1}(t)$ and $P_{S_2}(t)$ samples: the two random variables exhibit a strong correlation, in fact samples tend to concentrate along the space diagonal. The calculated correlation coefficient is $\rho_1 = 0.9347$.

By contrast, we consider now the two powers $P_{S_3}(t)$ and $P_{S_4}(t)$ delivered by two PV plants placed at geographically distant locations. The scattered plot of such power data is reported in Fig. 2. In this case, the data set shows a weak statistical correlation given by $\rho_2 = 0.2219$.

As a result, in order to be realistic we need to develop a general modeling framework able to properly reproduce the joint statistical behavior (i.e. the joinf PDF) of many PV power sources while accounting for their actual correlation. The general method that we present in this paper relies on the adoption of

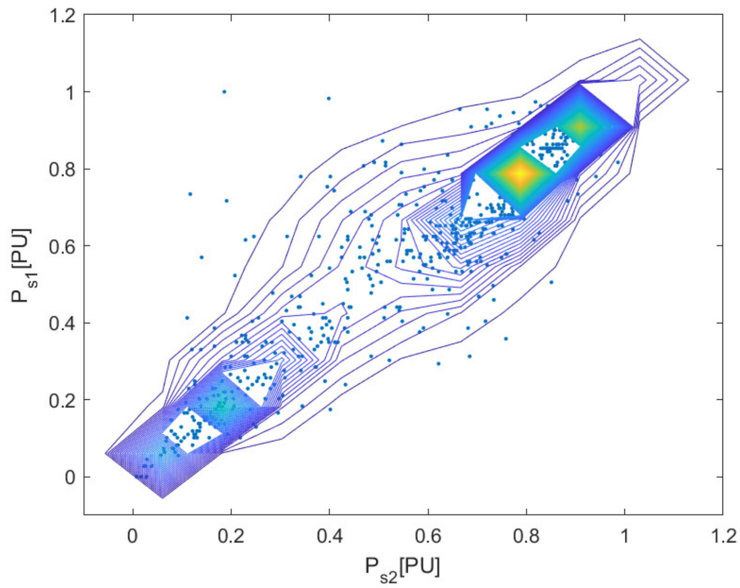

Figure 1: Case of two PV systems with strong correlation: scattered plot of generated powers $P_{S_1}(t)$ $P_{S_2}(t)$ and the contour of the GMM model fitting to the PV data.

Gaussian-Mixture-Models (GMM) [22], [23]. For instance, the combined effects of P_{S_1} and P_{S_2} can be modeled by means of two random variables x_1 and x_2, collected into column vector $\vec{x} = [x_1, x_2]^T$, having joint probability density function given by:

$$f(\vec{x}) = \sum_{n=1}^{N_G} \alpha_n \, g(\vec{x}; \vec{\mu}_n, \Sigma_n). \tag{1}$$

Each $g(\cdot)$ function in the series is a multi-variate Gaussian described by its mean-value vector $\vec{\mu}_n$ and covariance matrix Σ_n, i.e.:

$$g(\vec{x}; \vec{\mu}_n, \Sigma_n) = \frac{1}{((2\pi)^2 |\Sigma_n|)^{0.5}} \, e^{-\frac{1}{2}(\vec{x}-\vec{\mu}_n)^T \Sigma_n^{-1} (\vec{x}-\vec{\mu}_n)}. \tag{2}$$

The scalar coefficients $0 < \alpha_n \leq 1$ are the weights and are such that $\sum_{n=1}^{N_G} \alpha_n = 1$ while "T" denotes the transposition operator [24]. The GMM model (1) can approximate any PDF using a sum of weighted multi-variate Gaussians and this approximation converges to any PDF when the number N_G of components is increased [22]. The joint PDF of scattered data $x_1 \approx P_{S_1}$ versus $x_2 \approx$

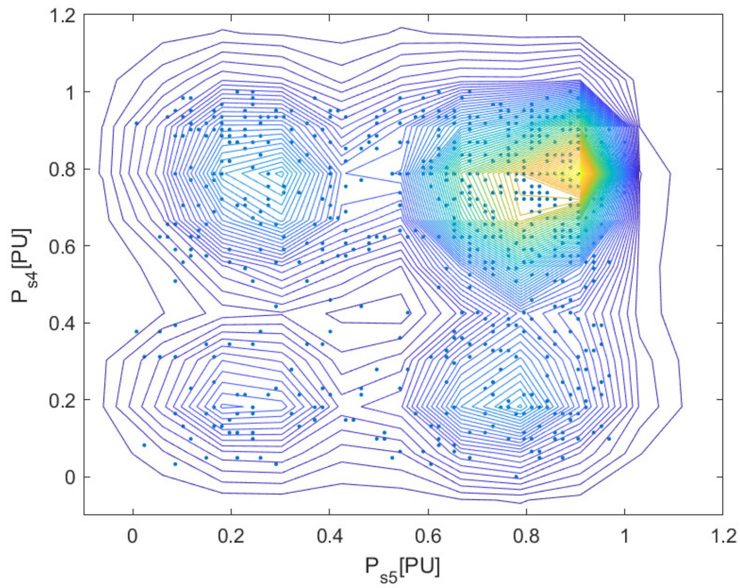

Figure 2: Case of two PV systems with weak correlation: scattered plot of generated powers $P_{S_3}(t)$ $P_{S_4}(t)$ and the contour of the GMM model fitting to the PV data.

P_{S_2} is fitted with model (1) by means of an iterative technique referred to as Expectation Maximization (EM) [24]. The EM algorithm is implemented in the function 'fitgmdist' of Matlab. For a given number N_G of Gaussian components, the EM algorithm provides α_n, $\vec{\mu}_n$ and Σ_n, for $n = 1, \ldots, N_G$. In Figs. 1 and 2, we report the contour plot of the PDFs generated by fitting the scattered data shown in the same figures with a GMM model of the type (1) made of $N_G = 2$ Gaussian components in Fig. 1 and $N_G = 4$ Gaussian components in Fig. 2.

The EM algorithm requires specifying the number N_G of Gaussian terms. To assess a proper value for N_G, we use a technique where the value of N_G is gradually increased, and the GMM components' variances are checked. More specifically, the considered criteria is the minimum of the sum of variances of nth component (variances are the diagonal elements of co-variance matrices), i.e.,

$$\sigma_{min} = \min_{n} \text{sum}[\text{diag}(\Sigma_n)]. \tag{3}$$

Fig. 3 shows the values of the parameter σ_{min} versus the number of Gaussian

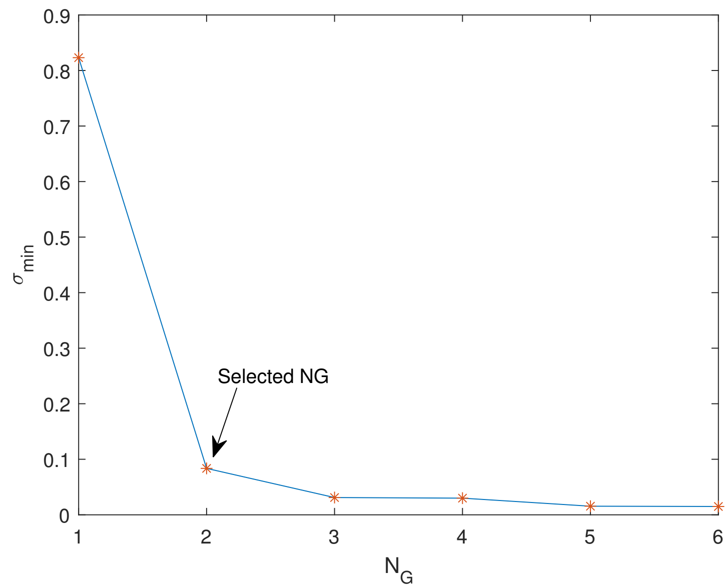

Figure 3: The dependence of parameter σ_{min} versus number of Gaussian N_G while fitting the model to the PV data.

components N_G in GMM required for the data set shown in Fig. 1, the curve presents a *knee* in correspondence of $N_G = 2$, that is a rapid decrease. This means that further increasing N_G yields Gaussian components with relatively much smaller variances. Gaussian terms with too slight variances are negligible at first approximation and thus $N_G = 2$ is selected as the appropriate number of terms.

3. Modeling residential load

The uncertainty of residential loads power consumption can also be included into our probabilistic model starting from a statistical analysis of available residential active power consumption data. More specifically, in this paper we will rely on measurement data available for the test network adopted in our analysis, i.e., the Nonsynthetic European low voltage test network [21]. This network contains a set of smart meter data measured for different loads.

As an instance, Fig. 4 reports the scattered plot of two load powers data

P_{L1} and P_{L2} measured at two different buses in the test network. Load powers are scattered in a way that suggests they are rather independent. In fact, the correlation coefficient computed with these load data results $\rho = 0.072$. By repeating such a statistical data analysis for several loads extracted from the data set and for several time windows during the day, we concluded that the typical uncertainty of residential power demand can be realistically reproduced by independent random variables.

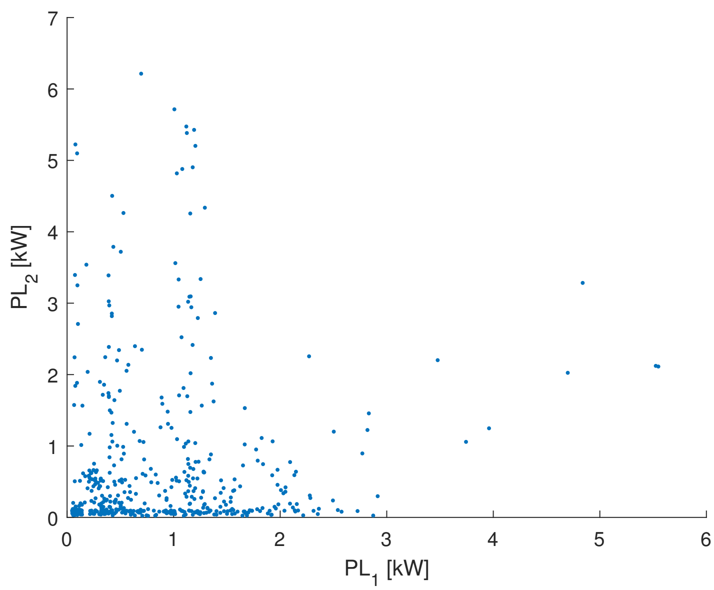

Figure 4: Scattered plot of powers P_{L1} and P_{L2}, these powers are absorbed by the loads connected to Phase A of the bus 1279 and 1283 in the test network.

The univariate PDF of such variables can be extracted from the data through a fitting/parametric approach. To this aim, the power data are first normalized to the peak value K_L. In this way, the active power absorbed by a load is written as

$$P_L = K_L \, y \tag{4}$$

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. 5 shows the PDF data distribution for the Normalised power P_{L1}.

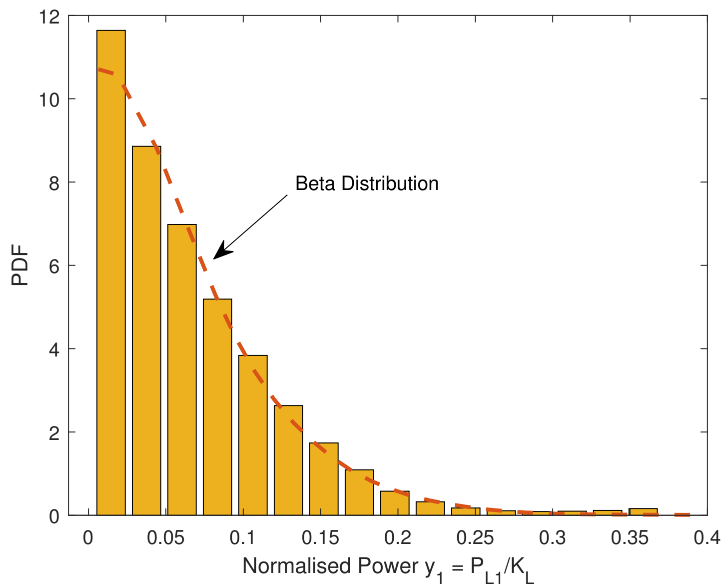

Figure 5: (Histogram) Normalized power PDF of P_{L1} which is Phase A of bus 1279. Figure shows the PDF of random variables y_1 after scaling P_{L1} by K_L which is 15kW. (dashed line) fitting with the Beta distribution of parameters $a = 1.2448$, $b = 19.4836$.

The extracted PDF samples $f(y^j)$ are fitted by trying several standard statistical distributions (among which Weibull, Exponential, and Beta). We determined that, for the available data, the best fitting is achieved by the Beta distribution:

$$f_\beta(y, a, b) = \frac{y^{a-1}(1-y)^{b-1}}{\mathrm{Beta}(a, b)}. \tag{5}$$

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 optimization procedure that minimizes the following objective function:

$$\mathrm{Error}(a, b) = \sum_j \left[f_\beta(y^j, a, b) - f(y^j)\right]^2, \tag{6}$$

with the constraints $a > 0$ and $b > 0$. Fig. 5 shows the Beta distribution $f_\beta(y, a, b)$ with parameters $a = 1.2448$, $b = 19.4836$ that yields data best fitting for the considered case.

4. Multi-Expansion gPC Probabilistic Analysis

The probabilistic analysis methodology that we describe in this paper is able to handle the general scenario: the network contains a first set of N_S random variables x_k strongly correlated among them plus a second set of N_L random variables y_k that can be realistically modeled as being statistically independent. The correlated random variables are collected into the vector $\vec{x} = [x_1, \ldots, x_{N_S}]^T$ whereas the independent variables are collected into the vector $\vec{y} = [y_1, \ldots, y_{N_L}]^T$. Please note that such a general arrangement includes the special case where all variables are correlated, i.e., $N_L = 0$ as well as the case where all variables are independent, i.e., $N_S = 0$. In light of the analysis presented in Sec. 2, for the adopted test network vector $\vec{x}$ should contain the power delivered by PV plants. The joint PDF of correlated variables is approximated by a GMM model (1) extracted from the available data set and for the one considered *time window* of the day. Besides, in light of the analysis presented in Sec. 3, for the adopted test network the vector $\vec{y}$ of independent variables should include the network loads described by their univariate PDF. More specifically, load uncertainty is described by statistically independent Beta-distributed variables $y_r \in [0, 1]$. Each y_r represents the normalized power demand of one residential customer over the same time window.

Hence, we focus on one electrical *observable variable* that can affect the quality of the network and denote it generically as V. Such a variable, which can be one node voltage as well as a line current, is seen as the output of the probabilistic problem.

The probabilistic analysis of the network can be achieved with the reference Monte Carlo method. This implies generating a huge number N_{mc} of uncertainty variables vector $[\vec{x}^T, \vec{y}^T]^T$ according to the joint probability distribution of variables. In order to generate samples for the correlated variables $\vec{x}$, a two step procedure is required: (i) a Gaussian component of index n is selected randomly on the basis of the prior probabilities α_n; (ii) a realization of variables

$\vec{x}$ is obtained randomly from the selected multi-variate Gaussian distribution $g(\vec{x}; \vec{\mu}_n, \Sigma_n)$. The random generation of samples for variables forming $\vec{y}$ is simpler and can be directly obtained from the univariate PDF $f(y_r)$.

For each realization of the PV power sources $\vec{x}$ and loads power demand $\vec{y}$, the corresponding realization of the observable variable V is calculated by running a deterministic Load Flow analysis of the distribution grid. An observable variable is in fact a deterministic function of uncertainty variables $\vec{x}$ and $\vec{y}$. As the number, N_{mc} of evaluations grows, at limit tending to infinity, the distribution of the calculated values of V describes its detailed statistical distribution. Since MC converges slowly, with an error that reduces as $1/N_{mc}^{0.5}$, the number of the required LF simulations tends to be large and probabilistic analysis time-consuming.

A way for accelerating MC methods while preserving high accuracy is the adoption of a surrogate model for the observable variable V based on gPC expansions. In what follows, we describe how the conventional gPC method can be extended to handle the case of nonGaussian distributed random variables described by the GMM (1). Such an extension relies on the fact that once the n^{th} Gaussian component has been (randomly) selected, variables $\vec{x}$ are completely described by the multi-variate Gaussian distribution $g(\vec{x}; \vec{\mu}_n, \Sigma_n)$. It is thus possible to introduce a new set of variables z_n univocally defined by the following linear transformation

$$\vec{x} = L_n \vec{z}_n + \vec{\mu}_n. \tag{7}$$

Matrix L_n is the square root of the nth covariance matrix, i.e.

$$\Sigma_n = L_n \, L_n^T \tag{8}$$

Matrix L_n is simply calculated through the Cholesky decomposition of (symmetric) Σ_n. It is worth noticing how the newly defined variables in the vector $\vec{z}_n$ are Normal distributed $\mathcal{N}(\mathbf{0}, \mathbf{1})$, i.e. they are Gaussian-distributed variables with zero mean and unitary variance. In addition, such Gaussian variables are decorrelated and thus statistically independent.

In view of (7), the observable variable V can now seen as a function of variables $\vec{z}_n$, $\vec{y}$ and approximated by a gPC series expansion. Using the compact notation

$$\vec{\xi_n} = [\vec{z}_n^{\,T}, \vec{y}^{\,T}]^T, \tag{9}$$

The relationship $V(\vec{\xi_n})$ is approximated with an order-γ truncated series expansion of the type

$$V(\vec{\xi_n}) \approx \sum_{i=0}^{N_b-1} c_i^n \, H_i(\vec{\xi_n}), \quad n = 1, \ldots, N_G \tag{10}$$

formed by N_b multi-variate basis functions $H_i(\vec{\xi_n})$ weighted by unknown coefficients c_i^n. For each Gaussian component forming (1), a different gPC series expansion of the type (10) is adopted. In this sense, the method may be referred to as a *Multi-Expansion* gPC method.

An interesting interpretation of such a technique is that of seeing the relationship $V(\vec{\xi_n})$ as a discrete-type random variable that can assume N_G different expressions (10) with probabilities α_n. Since the nth expansion series (10) holds only for $\vec{x}$ belonging to a subregion of the parameter space centered around $\vec{\mu_n}$, its truncation order γ can be selected smaller compared to conventional Single-Expansion gPC method. It is worth noting how the Multi-Expansion method we employ in this paper follows an idea similar to what has been presented in [25] and referred to as the Multi-element method. However, the Multi-element method requires breaking the statistical space into mutually disjoint regions over which the joint PDF is sharply fragmented. By contrast, the Multi-expansion method approximates the joint PDF as the sum of smooth partially-overlapped Gaussian kernel functions, making implementation more straightforward while accounting for variables correlation.

5. gPC Method and Stochastic Collocation

The Multi-Expansion gPC in connection with GMM models reduces the probabilistic analysis to n single-expansion gPC subproblems. The single-expansion problem is well documented in the literature [14], [15], [17]. In our case, each subproblem has N_S independent and Normally-distributed variables z_s, with $s = 1, \ldots, N_S$ and N_L independent variables y_r, with $r = 1, \ldots, N_L$. As a consequence, the multivariate basis functions forming gPC expansion (10) are all of the type:

$$H_i(\vec{\xi_n}) = \prod_{s=1}^{N_S} q_{i_s}(z_s) \cdot \prod_{r=1}^{N_L} q_{i'_r}(y_r), \tag{11}$$

where $q_{i_s}(z_s)$ are the univariate Hermite polynomials of degree i_s while $q_{i'_r}(y_r)$ are the univariate Jacobi-chaos polynomials of degree i'_r [14]. For a given number of parameters $N_S + N_L$ and series truncation order γ, the index degrees i_s and i'_r should satisfy the following constraint

$$\sum_{s=1}^{N_S} i_s + \sum_{r=1}^{N_L} i'_r \leq \gamma. \tag{12}$$

The number of gPC basis functions in (10) is given by

$$N_b = \frac{(\gamma + N_S + N_L)!}{\gamma!\,(N_S + N_L)!}. \tag{13}$$

The expansion coefficients c_i^n in the series (10) can be calculated with a known technique referred to as Stochastic Testing (ST) method [15]. For each Gaussian component n, the N_b unknown coefficients c_i^n are calculated by selecting $N_{sam} = N_b$ testing points $\vec{\xi}_n^k$, for $k = 1, \ldots, N_{sam}$ among the multivariate Gauss quadrature nodes. Univariate Gauss quadrature nodes for Normally-distributed variables z_s and Beta-distributed variables y_r are known in closed form or can be easily computed [16]. Multivariate nodes are then obtained from the tensor product (i.e. all possible combinations) of univariate ones.

Each testing point (9) corresponds to given values of variables $\vec{z_n} = \mathbf{Z^k}$ and $\vec{y} = \mathbf{Y^k}$. Variable values $\vec{z_n} = \mathbf{Z^k}$ are then transformed into PV delivered powers $\vec{x}$ by means of the linear transformation (7). Variable values $\vec{y} = \mathbf{Y^k}$, after scaling provide load absorbed powers. As a result, for each testing point, the observable variable $V_k^n = V(\vec{\xi}_n^k)$ can be evaluated by running a deterministic LF analysis. Hence, the series expansions (10) are enforced to fit *exactly* (i.e., the polynomials interpolate the samples) the values V_k^n at the testing points.

This results in the following linear system

$$\mathbf{M}\,\vec{c}^{n} = \vec{V}^{n}, \tag{14}$$

where $\vec{c}^n = [c_0^n, \dots, c_{N_b-1}^n]^T$ and $\vec{V}^n = [V_1^n, \dots, V_{N_{sam}}^n]^T$ are the column vectors collecting the unknown coefficients and observable 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.

$$\mathbf{M} = \begin{bmatrix} H_0(\vec{\xi}^1) & \dots & H_{N_b-1}(\vec{\xi}^1) \\ \vdots & \ddots & \vdots \\ H_0(\vec{\xi}^{N_{sam}}) & \dots & H_{N_b-1}(\vec{\xi}^{N_{sam}}) \end{bmatrix}. \tag{15}$$

Interestingly matrix $\mathbf{M}$ remains the same for each Gaussian component, so it can be precalculated, inverted and used for any n as follows:

$$\vec{c}^n = \mathbf{M}^{-1}\,\vec{V}^n \quad \text{for} \;\; n = 1, \dots, N_G \tag{16}$$

Once, the c_i^n coefficients have been computed, the surrogate models (10) can be exploited within MC method in a significantly efficient way. In fact, for each randomly selected Gaussian component samples of the independent variables $\vec{\xi}_n = [\vec{z}_n, \vec{y}]$ are easily generated and then the nth model (10) provides the realization of observable variable V without having to run a LF simulation. In this way millions of realizations of V can be obtained in a few seconds on a quad-core computer allowing a detailed determination of the PDF.

6. Numerical Results

In this Section, we exploit the proposed ME-gPC method to analyze/quantify the uncertainty of nodal voltages in the power grid due to the presence of distributed PV generation. To this aim, we consider several scenarios/investigations. In subsection 6.1, we compare the case of PV systems having high statistical correlation among them with the opposite case of uncorrelated PV systems.

Simulations are applied to a power distribution network exhibiting the typical features of European distribution networks and referred to as the Nonsyn-

Table 1: Subsections of numerical results

Subsection	Modelling	Correlation type	Case study
6.1	GMM and Independent	Highly correlated and uncorrelated	Modelling comparision
6.2	GMM with varying loads	Highly correlated and uncorrelated	Voltage unbalance
6.3	GMM	Highly correlated	Monte Carlo check

thetic European low voltage test network (NSELVTN) [21]. The NSELVTN contains several loads with prescribed power consumption.

For the purposes of this article, the network is modified by injecting several PV generators into its subsection shown in Figure 6. In doing that, two different scenarios are considered. In the first one, presented in subsection 6.1, loads are fixed to their nominal values (i.e., loads uncertainty is not included). In the second one, illustrated in subsection 6.2, the statistical uncertainty of a subset of loads is included via a set of independent random variables. This second example thus corresponds to the general scenario described in Section 4, where network uncertainty is represented by a set $\vec{x}$ of (strongly) correlated random variables along with a set $\vec{y}$ of independent ones. Finally, in subsection 6.3, in order to validate the results of the proposed ME-gPC method, Monte Carlo simulations are conducted to compare the results.

The analysis and numerical results conducted in the subsequent subsections are summarized in Table. 1. All the simulations have been done using a joint co-simulation [17] between Matlab for the part of distributions and sampling points generation and Opendss[26] for the network simulation. The machine used is an Intel I7 with 16GB of ram and 2.3 GHz clock frequency.

6.1. NSELVTN with correlated PV sources

In this subsection, we investigate the importance of accounting for the actual correlation degree of PV sources in probabilistic load flow analysis.

To this aim, the original NSELVTN is injected by two PV systems connected to Phase C of bus 1279 and to Phase C of bus 1283. The effects of these PV

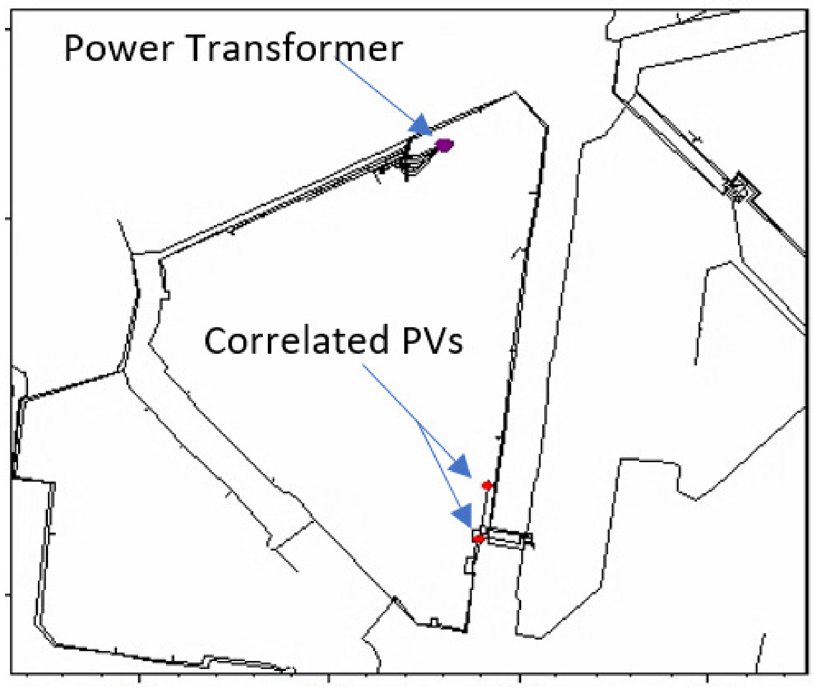

Figure 6: The detail of NSELVTN subsection where PV sources are inserted.

Table 2: Modified parameters for NSELVTN

Parameter	Phasing	Node	P [kW]	Q [kVAR]
1ϕ PV Generator	C	1279	53	26.5
1ϕ PV Generator	C	1283	60	30
1ϕ Constant load	A	1279	26	13
1ϕ Constant load	A	1283	05	2.5
1ϕ Constant load	C	1279	20	10
1ϕ Constant load	C	1283	19	9.5

sources are observed at bus 1284 which is placed in between injecting buses 1279 and 1283. In this first example, power absorbed by loads are assumed to be constant as reported in Table. 2. The two powers sources are statistically correlated and are represented by the two random variables x_1, x_2 whose joint probability is described mathematically by the GMM model (1). In the first case, GMM model fits the power data, i.e. $x_1 \approx P_{S_1}$ and $x_2 \approx P_{S_2}$ for the strongly correlated case reported in Fig. 1. Power values are scaled in order to simulate $\approx$ 53kW and $\approx$ 60kW of maximum power respectively.

In the second case, instead, random variables x_1, x_2 are assumed to be statistically independent and described by their marginal distribution probabilities. This latter modeling approach is referred to as the Independent method [27].

Fig. 7 reports the voltage distribution at bus 1284, Phase C, over the hourly time window 11:00AM – 12:00PM, computed using the GMM method and Independent method modeling. We see how a voltage variability range of $\approx$ 32V is observed at Phase C when correlation among PV sources is included into the model. By contrast, when the same PV sources are assumed statistically independent, the variability range reduces to $\approx$ 22V, for the considered hourly time window.

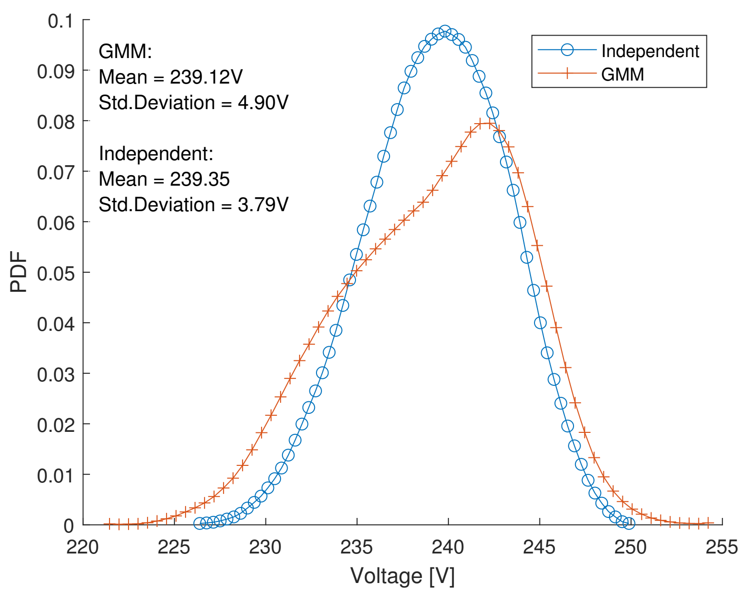

Figure 7: Phase C voltage distribution at node 1284 due to power injection by PV sources at node 1279 and 1283 connected to Phase C. The distribution is given for the hourly time window considered for 11:00AM – 12:00PM. The data used are of same PV sources with considering correlation using GMM and without considering the correlation using the independent method.

The PDFs of the nodal voltage considered and computed for several hourly time windows exhibit different shapes: the mean value and standard deviation of voltage distribution are much varied in the morning and evening hours but not in the afternoon. Fig. 8 shows the mean value of the Phase C voltage distributions of bus 1284 and the standard deviations analyzed for every hour from 09:00AM to 06:00PM using the GMM and Independent methods.

It is worth noticing how the standard deviations computed with GMM methods, i.e., properly considering PV correlation, are always greater than those estimated assuming PV sources independence.

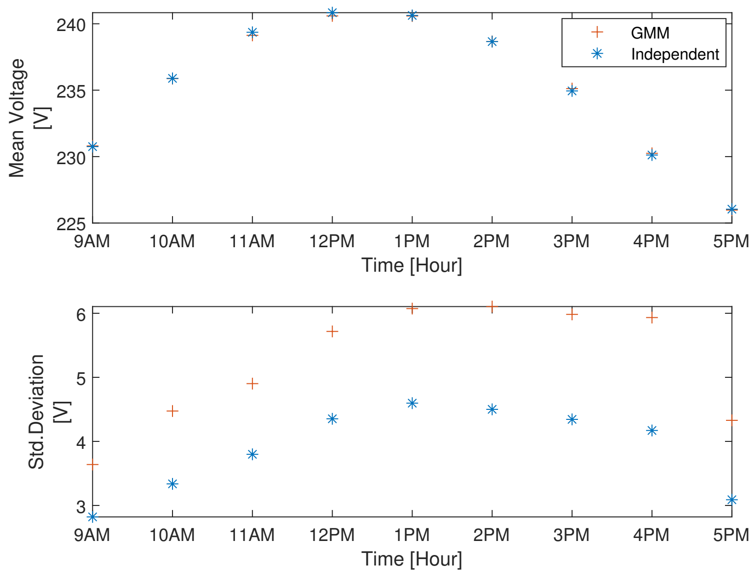

Figure 8: Mean and standard deviation of Phase C voltage distribution at node 1284 analyzed for every hour from 09:00AM to 06:00PM while comparing the modelling technique that consider the correlation using GMM and discards such correlation using Independent method considering the same data set of PV.

PV power injection at Phase C induces variations in the voltage at phase C while the other 2 phases, Phase A and B, remain almost unchanged. This leads to voltage unbalance in the network. Voltage unbalance is commonly defined as the ratio of the negative sequence voltage component to the positive sequence voltage component, and it is usually expressed in percentage as $VUF\%$. The voltage distributions obtained in the hourly analysis is classified to three groups for easy understanding and visualization. The groups are morning 09:00AM–12:00PM, afternoon 12:00PM–03:00PM and evening 03:00PM–06:00PM.

The $VUF\%$ is determined for both the voltage distribution data sets of correlated PV sources, and the uncorrelated PV sources is shown in the Fig. 9. Also in this case, it is evident how the precise determination of $VUF\%$ can only be obtained by considering the correct correlation among PV sources. Correlation in fact tends to produce a wider range of $VUF\%$ at all time instants.

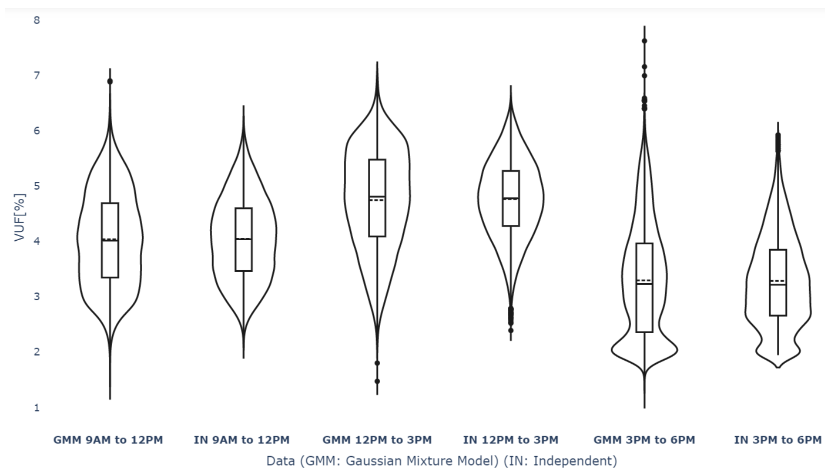

Figure 9: Violin plots representing the distribution and box plot representing the median and mean(–) of Percentage voltage unbalance factor (V.U.F) of the network at bus 1284 due to PV sources connected to phase C of nodes 1279 and 1283. The two curves are obtained former modelling the PV with GMM and latter considering them independent.

6.2. *NSELVTN with PV sources and uncertain loads*

In this subsection, we study the case where the statistical uncertainty of loads is included in the model. In fact, now loads are modeled as random variables, whose active power consumption is a statistical parameter modeled using the measurement data set available for the NSELVTN. The active power data is fitted to a beta distribution, and the respective distribution is scaled back to the desired rated power as given in Table. 3 and used in simulations.

Fig. 10 shows the simulated Phase C voltage distribution of bus 1284 for hour window of 11:00AM – 12:00PM . We note how the shape and variability range of the distribution are significantly varied compared to the case investigated in previous subsection with constant loads.

With uncertain loads, the variability range of Phase C voltage distribution of Bus 1284 grows to $\approx$ 44V while the same variability range but computed assuming PV sources are independent is $\approx$ 37V.

We conclude that including loads uncertainty and considering PV sources correlation provides the modeling scenario with the greatest variability range of the considered observation variable.

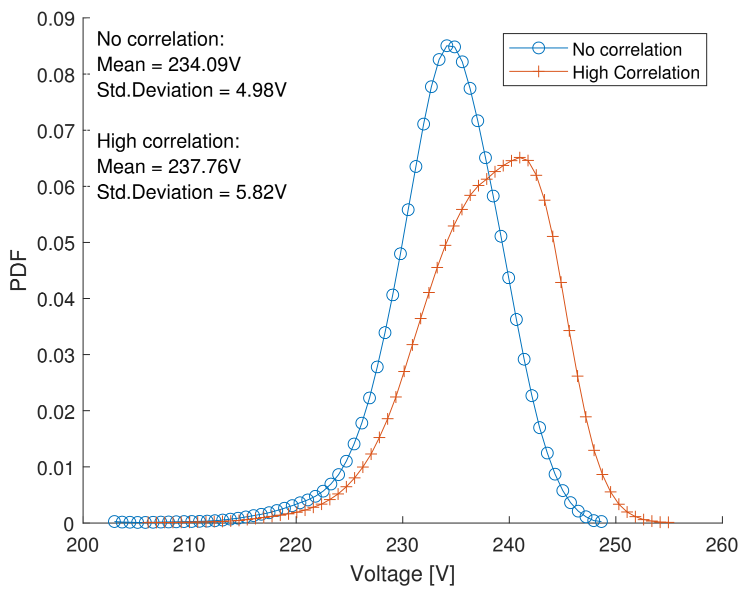

Figure 10: Phase C voltage distribution at node 1284 due to power injection by PV sources and variable load connected to node 1279 and 1283 in Phase C. The distribution is given for the hourly time window considered for 11:00AM – 12:00PM.

Similar conclusions can be drawn when looking at the mean value and standard deviations computed over several hourly windows of the day. Fig. 11 shows such statistical quantities computed for the Phase C voltage: standard devia-

Table 3: Beta parameters to fit the load measurement data

Active Power [kW]	K_L [kW]	a	b
P_{L1}	15	1.24	19.48
P_{L2}	15	0.59	11.44
P_{L3}	15	1.29	29.72
P_{L2}	15	0.833	4.74

tions computed with GMM method are always greater than those estimated assuming that PV sources are independent.

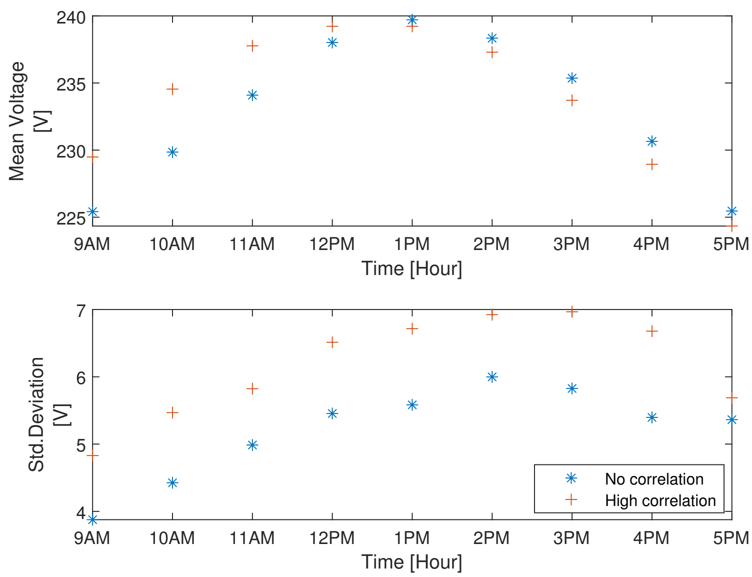

Figure 11: Mean and standard deviation of Phase C voltage distribution at node 1284 analyzed for every hour from 09:00AM to 06:00PM with varying load, comparing the consideration of correlated and uncorrelated PV sources.

Finally, loads uncertainty contributes to increase voltage unbalance in the network as shown in in Fig. 12. Here, the $VUF\%$ is presented for three groups of time, i.e., morning, afternoon, and evening. PV sources correlation and loads uncertainty is the modeling scenario with the greatest variability range of VUF.

6.3. Comparison with Monte Carlo Simulations

Monte Carlo (MC) simulation is the primary and reference method to compare other analytical and approximation probabilistic load flow analysis methods. In this subsection, for accuracy check, the proposed ME-gPC method is compared with the results obtained with Monte Carlo simulations. The simulation times of ME-gpPC and Monte Carlo methods are also compared.

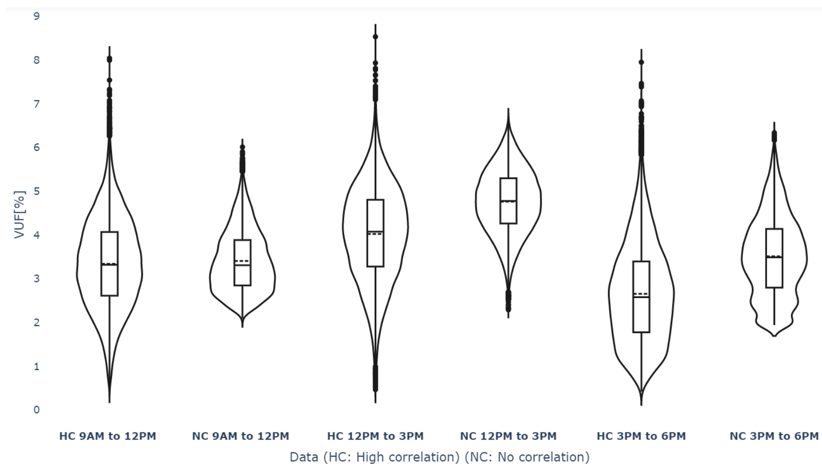

Figure 12: Violin plots representing the distribution and box plot representing the median and mean(–) of Percentage voltage unbalance factor of the network at bus 1284 due to PV sources connected to phase C and varying load connected to phase A and C of nodes 1279 and 1283. The comparison among the highly correlated and uncorrelated case is given.

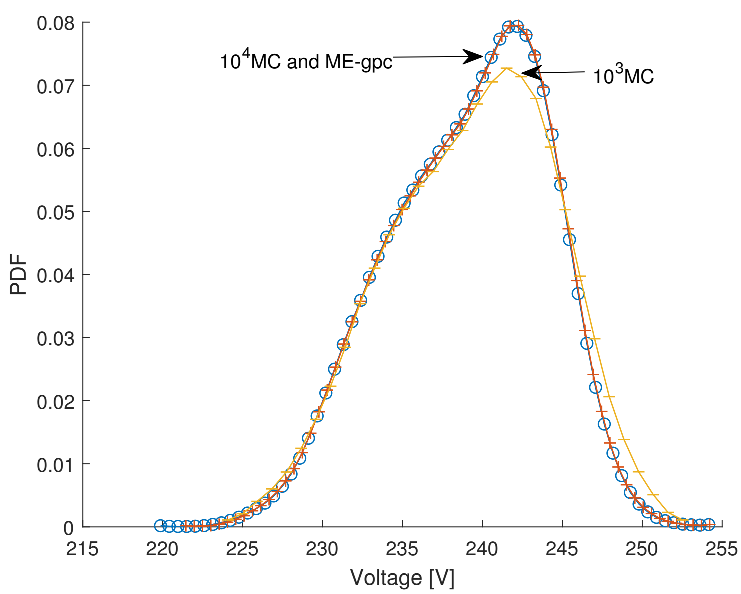

Figure 13: PDF of Bus 1284 Phase C voltage of the test network, with highly correlated PV and constant power load, computed with Monte Carlo for different number of iterations and ME-gPC method for highly correlated Ausgrid data set analyzed for 11:00AM–12:00PM time window.

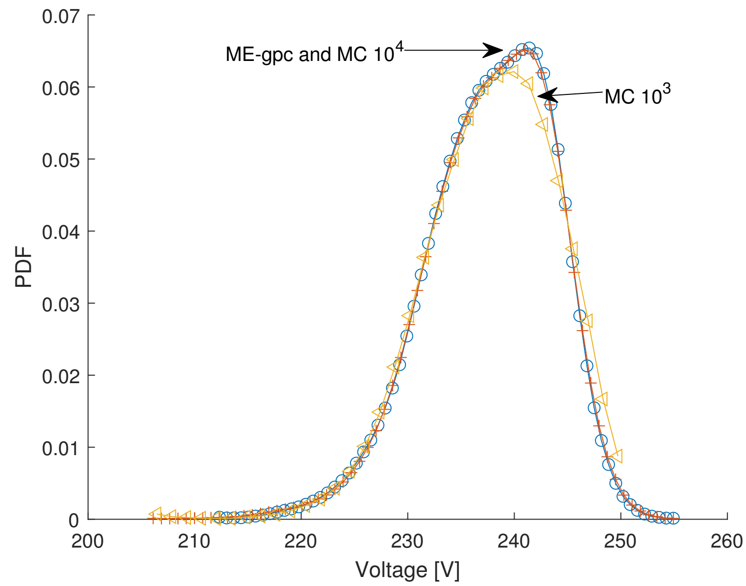

Figure 14: PDF of Bus 1284 Phase C voltage of the test network with highly correlated PV and statistically varying load. computed with Monte Carlo for different number of iterations and ME-gPC method for highly correlated Ausgrid data set analyzed for 11:00AM–12:00PM time window

Fig. 13 and Fig. 14 show the PDFs of Phase C Voltages of bus 1284 computed with the Monte Carlo method running 1000 and 10,000 iterations along with the same PDFs calculated with the proposed ME-gPC method. The PDF curve obtained with Monte Carlo method (10,000 samples) is almost indistinguishable from the PDF curve computed with ME-gPC method, i.e. their relative difference being always < 0.003, thus confirming the accuracy of ME-gPC method. On the other hand, the probabilistic load flow simulation of NSELVTN using ME-gPC with 2 correlated PV sources requires a simulation time of about 28s, whereas MC running 10,000 iterations requires more than 48minutes. ME-gPC simulation with 6 parameters (2 correlated PV and 4 loads) runs in about 55s whereas, for the same accuracy, MC simulation requires about 65minutes. As a result, in both cases the proposed ME-gPC method introduces a two orders of magnitude speedup factor compared to MC simulation for the same accuracy.

7. Conclusion

In this paper, we have illustrated an original technique for modeling and simulating PV-penetrated residential power distribution grid. The method proposed in the paper is data driven since it relies on the availability of measurements data sets of PV-generated and consumer-absorbed power. It has been shown how the statistical uncertainty of correlated PV sources can be properly reproduced in simulations by adopting a Gaussian-Mixture-Model (GMM). We have described how the parameters of the GMM can be extracted from the available PV data measurements grouped for different daily time windows. It has been observed how for residential consumer loads, instead, the uncertainty is better represented by independent random variables.

In view of that, a novel probabilistic analysis technique based on Multi-Expansion generalized Polynomial Chaos (ME-gPC) has been described in details that is able to handle the general modeling scenario. Several numerical experiments have been discussed comparing different modeling cases where correlation of PV sources and loads uncertainty have been either included into the model i.e. thanks to ME-gPC method, or they have been disregarded. Numerical results highlight the importance of including both the above-mentioned uncertainty aspects. In fact, neglecting PV correlation and/or loads uncertainty always results in the under-quantification of uncertainty impact on the grid.

References

[1] S. Impram, S. Varbak Nese, B. Oral, Challenges of renewable energy penetration on power system flexibility: A survey, Energy Strategy Reviews 31 (2020) 100539.

[2] H. Mortazavi, H. Mehrjerdi, M. Saad, S. Lefebvre, D. Asber, L. Lenoir, A monitoring technique for reversed power flow detection with high pv penetration level, IEEE Transactions on Smart Grid 6 (5) (2015) 2221–2232.

[3] 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.

[4] S. Chevalier, L. Schenato, L. Daniel, Accelerated probabilistic power flow in electrical distribution networks via model order reduction and neumann series expansion, IEEE Transactions on Power Systems (2021) 1–1doi:10.1109/TPWRS.2021.3120911.

[5] J. Tang, F. Ni, F. Ponci, A. Monti, Dimension-adaptive sparse grid interpolation for uncertainty quantification in modern power systems: Probabilistic power flow, IEEE Transactions on Power Systems 31 (2) (2016) 907–919.

[6] Z. A. Memon, R. Trinchero, Y. Xie, F. G. Canavero, I. S. Stievano, An iterative scheme for the power-flow analysis of distribution networks based on decoupled circuit equivalents in the phasor domain, Energies 13 (2) (2020).
URL https://www.mdpi.com/1996-1073/13/2/386

[7] M. N. Kabir, Y. Mishra, R. Bansal, Probabilistic load flow for distribution systems with uncertain pv generation, Applied Energy 163 (2016) 343–351.

[8] Q. Hou, N. Zhang, E. Du, M. Miao, F. Peng, C. Kang, Probabilistic duck curve in high pv penetration power system: Concept, modeling, and empirical analysis in china, Applied Energy 242 (2019) 205–215.

[9] 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.

[10] A. G. Zamani, A. Zakariazadeh, S. Jadid, Day-ahead resource scheduling of a renewable energy based virtual power plant, Applied Energy 169 (2016) 324–340.

[11] 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.

[12] 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.

[13] 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.

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

[15] 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.

[16] Z. Zhang, X. Yang, G. Marucci, P. Maffezzoni, I. A. M. Elfadel, G. Karniadakis, L. Daniel, Stochastic testing simulator for integrated circuits and mems: Hierarchical and sparse techniques, Proceedings of the IEEE 2014 Custom Integrated Circuits Conference (2014) 1–8.

[17] G. Gruosso, P. Maffezzoni, Z. Zhang, L. Daniel, Probabilistic load flow methodology for distribution networks including loads uncertainty, International Journal of Electrical Power & Energy Systems 106 (2019) 392 – 400.

[18] G. Gruosso, G. S. 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.

[19] Ausgrid, Solar home half-hour data - 1 july 2012 to 30 june 2013. URL https://www.ausgrid.com.au/Industry/Our-Research/Data-to-share/

[20] X. Wan, G. E. Karniadakis, An adaptive multi-element generalized polynomial chaos method for stochastic differential equations, Journal of Computational Physics 209 (2) (2005) 617–642. doi:https://doi.org/10.1016/j.jcp.2005.03.023.

[21] A. Koirala, L. Suárez-Ramón, B. Mohamed, P. Arboleya, Non-synthetic european low voltage test system, International Journal of Electrical Power & Energy Systems 118 (2020) 105712.

[22] D. Alspach, H. Sorenson, Nonlinear bayesian estimation using gaussian sum approximations, IEEE Transactions on Automatic Control 17 (4) (1972) 439–448. doi:10.1109/TAC.1972.1100034.

[23] Y. Dong, H. Zhang, C. Wang, X. Zhou, A novel hybrid model based on bernstein polynomial with mixture of gaussians for wind power forecasting, Applied Energy 286 (2021) 116545.

[24] C. M. Bishop, Pattern Recognition and Machine Learning, 1st Edition, Springer, 2006.

[25] Y. Xu, L. Mili, A. Sandu, M. R. v. Spakovsky, J. Zhao, Propagating uncertainty in power system dynamic simulations using polynomial chaos, IEEE Transactions on Power Systems 34 (1) (2019) 338–348. doi:10.1109/TPWRS.2018.2865548.

[26] EPRI, Opendss. URL http://smartgrid.epri.com/SimulationTool.aspx

[27] H. Palahalli, P. Maffezzoni, G. Gruosso, Modeling photovoltaic generation uncertainties for monte carlo method based probabilistic load flow analysis of distribution network, in: 2020 55th International Universities Power Engineering Conference (UPEC), IEEE, 2020, pp. 1–6.