Implementing Stochastic Response Surface Method and Copula in the presence of data-driven PV source models

Harshavardhan Palahalli, *Student Member, IEEE,* Paolo Maffezzoni, *Senior Member, IEEE,* Pablo Arboleya, *Senior Member, IEEE,* and Giambattista Gruosso, *Senior Member, IEEE.*

***Abstract*—As PV penetration in the power distribution network is growing on an unprecedented scale, probabilistic power flow analysis is becoming crucial to assess the health of network operation. State-of-the-art probabilistic analysis relies on Stochastic Response Surface Method and Gaussian Copula for including PV sources correlation. However, it has been observed how in the presence of complex statistical distributions of injected PV powers, such a standard approach can provide inaccurate evaluations of the output variable distributions. In this article, we carefully investigate the origins of such a drawback and propose a novel implementation flow that overcomes the problem. In order to check the correctness of the proposed methodology, the results obtained with the Stochastic Response Surface Method are compared with Monte Carlo simulations. Several investigations, such as voltage uncertainties, network health and probability of violating quality constraints are conducted on the same test network.**

***Index Terms*—Correlated PV, Gaussian Copula, Photovoltaic generation, Polynomial approximation, Probabilistic load flow, Stochastic Response Surface Method.**

I. Introduction

THE growing penetration of distributed Photovoltaic (PV) sources in the power distribution network is leading to a significant uncertainty in power generation. This can adversely affect the quality/health of the network leading to unwanted malfunctioning, such as node over voltage or phase unbalancing. In order to predict the effects of PV generation uncertainty on network operation, two main ingredients are key: 1) a realistic statistical model of PV sources, 2) an efficient Probabilistic Load Flow methodology [1], [2], [3], [4] able to include the PV model.

In previous works, the geographically dispersed PV systems have been modelled using several techniques [5], [6], [7] that capture the meteorological variability. In many cases, such techniques fit the power output data to known standard distributions such as uniform and beta [8], [9], [10]. Unfortunately, data-driven investigations have shown how PV power generation uncertainty tends to follow non-standard statistical distributions [11], [12]. Besides that, PV systems are commonly correlated among them since they depend on meteorological effects of that day such as the availability of solar irradiation, temperature and wind speed for cooling effect.

In order to model correlation in PV power generation, Copula-based methods [13], [14] are commonly adopted. In [15], Copula method is exploited to model the solar irradiation of PV systems that are spatially correlated, similarly in [16] authors have demonstrated the usage of Gaussian Copula to handle the correlation among the PV sources represented by non-standard statistical distributions in connection with Monte Carlo (MC) simulations. In actual fact, MC simulation, thanks to its accuracy and versatility, still represents the basic method for probabilistic load flow (PLF) analyses and the ultimate reference to check the reliability of other techniques [17].

However, due to its slow convergence rate, MC method commonly requires running tens of thousands load flow simulations becoming unacceptably time consuming for complex networks/scenarios.

An effective approach to speed up MC simulations relies on the adoption of surrogate models of the network response that employ polynomial chaos expansions [18], [19], [20]. Within this frame, Stochastic Response Surface Methods (SRSM) look particularly suitable to power grid analysis since they can directly provide the detailed PDFs of non elementary events/quantities affecting network quality, such as overvoltage or voltage unbalance factor probability. SRSMs have also been extended to handle the correlations among the physical variables by means of Copula method [21]. In fact, Copula allows transforming, i.e. via Nataf transformation, correlated PV powers into a new set of *inner variables* that are statistically independent and Gaussian distributed. A surrogate SRSM model can thus be derived that approximates the relationship among inner variables and output variables of interest via standard Hermite polynomials chaos series expansions.

It has been noted however how in the presence of non-standard complex statistical distributions of uncertain input parameters, such a conventional SRSM implementation with Copula can introduce inaccuracies in evaluating output variables distributions [22].

In this paper, we carefully investigate the reasons of such a relevant implementation drawback. We show how the problem is connected to the high degree of nonlinearity introduced by Nataf transformation when modeling uncertain PV sources. Hence, we propose a novel implementation of SRSM in connection with Copula that fixes the problem. In the novel implementation, Gaussian Copula and SRSM are applied separately in a two-step procedure. In fact, at the first step, Copula is used only as a samples repopulation technique of small PV powers datasets. At the second step, a SRSM surrogate model, built on generalized Polynomials Chaos (gPC), is employed to accurately approximate the mildly (low-order) nonlinear relationship linking PV delivered powers to the output variables of interest.

The novel PLF methodology is applied to a large power distribution network that resembles an European town i.e., the Non-synthetic European low voltage test network (NSELVTN). Network health, in terms of voltage unbalance, is discussed in connection with different penetration levels of the PV sources. In addition, the probability of violating the constraints related to the quality of the power network is evaluated by considering over voltage [23] and voltage unbalance due to different PV source penetration.

The original contributions of this article includes:

- The mathematical details of Copula computational flow when applied to correlated PV sources is reviewed, this has also a tutorial value.
- It is illustrated how conventional implementation of SRSM with Copula can provide inaccurate results due to the (known) limitation of polynomial interpolation in the presence of high-order nonlinearities.
- A novel implementation is thus provided that overcomes the drawback. It is based on the idea of applying Copula and SRSM surrogate model separately in a two-step computational flow.
- The accuracy and the efficiency of the novel method is proved by comparing its predictions with those of MC simulations conducted on the same test network NSELVTN.

The rest of the article is organized as follows: Section II investigates the typical features of PV generation statistical distributions and reviews in details the computational flow of Gaussian copula. In Section III, we illustrate the mathematical implementation details of the novel SRSM method highlighting the differences compared to conventional implementation. Numerical results on the test network are illustrated in Section IV while Conclusions are drawn in Section V.

II. PV uncertainty and Copula Method

In this paper, we are interested in reproducing in simulations the typical statistical features of PV generation and evaluating their effect on the distribution grid. The technique that we propose is general and can be applied to any type of PV delivered power data set, however for exposition purpose we refer to the freely-available data set of PV measurements [24]. Our analysis is done on hourly basis: the active power delivered by typical PV plants over a hourly time window of the day, i.e., from 10:00 AM to 11:00 AM, are extracted for several days. The values of delivered PV (active) power P_{PV} are normalized to the maximum installed power P^M, i.e.

$$x = \frac{P_{PV}}{P^M}. \tag{1}$$

Fig. 1(Top) shows, as an example extracted from [24], the typical statistical distribution of the normalized active power for the morning hour 10:00 to 11:00 AM while Fig. 1(Bottom) reports the distribution for the afternoon hour 3:00 to 4:00 PM. It is apparent how:

- PV power delivery tends to follow non-standard and quite complex statistical distributions.

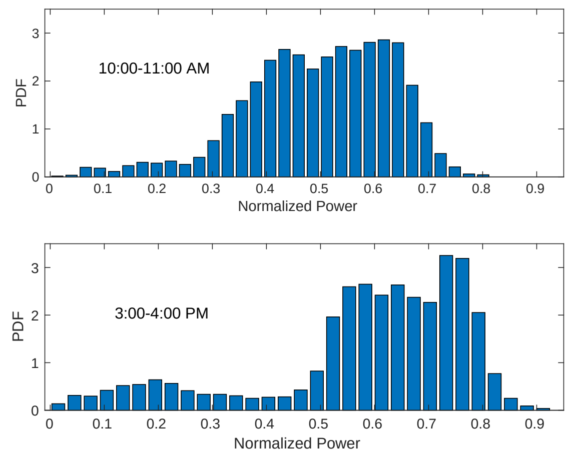


Fig. 1. Typical probability distributions of PV delivered power: (Top) over the morning hour 10:00 to 11:00 AM; (Bottom) over the afternoon hour 3:00 to 4:00 PM.

- PV power delivery is a random *non-stationary* process since its statistics over different hour time windows are different among them. This implies that with PV sources, probabilistic analysis should be repeated for each hour time window. In addition, since available data for each single hour can be quite limited in number, some form of *data repopulation* is indeed needed in order to generate the large data set required by probabilistic analysis (e.g., Monte Carlo).

In addition, even if not seen in Fig. 1, it is well known that when passing to consider many geographically-close PV sources, they exhibit a more or less degree of correlation that should be accounted for in simulations.

A. Transformation of stochastic physical variables

In its original implementation, stochastic response surface method assumes that stochastic variables are normal distributed and independent among them [25], [26]. Since physical variables describing PV power generation do not satisfy such hypotheses, they need to be first transformed into a new set of inner variables. The implementation that we now describe is data-driven since it relies on the availability of an initial set of simultaneous values assumed by the physical variables, i.e. the normalized PV delivered powers. From this initial data set, the Copula model is employed to generate a much wider set of samples to be used in subsequent probabilistic analyses.

Let us denote the physical random variables as x_k, for $k = 1, \ldots, N$ and the associated empirical CDF as $F_k(x_k)$. Such a CDF can be determined/evaluated from data through well known kernel expansion techniques [27].

We have that the u_k random variables defined as

$$u_k = F_k(x_k) \tag{2}$$

are uniformly distributed in $[0, 1]$. Hence, we can define a set of middle variables z_k, that are joint Gaussian, by means of

Nataf's transformation [28].

$$z_k = \Phi^{-1}(u_k)) = \Phi^{-1}(F_k(x_k)) \tag{3}$$

where $\Phi(\cdot)$ denotes the Gaussian cumulative distribution. The samples of z_k derived from the data set of x_k through (3) allow one to compute the empirical correlation matrix C_z among z_k variables. Finally, the z_k variables, that for notation compactness are collected into vector $\vec{z}$, are decorrelated via the following linear transformation

$$\vec{z} = L\,\vec{\xi} \tag{4}$$

where L is the square root of correlation matrix C_z of vector $\vec{\xi}$ calculated with the Cholesky decomposition. While vector $\vec{\xi}$ collects N Gaussian-distributed and independent variables ξ_k. Such normal distributed variables are the inner variables of the model.

B. Generating/Computing correlated samples

The transformation flow described in the previous subsection can be followed backward for generating a large set of samples of physical variables x_k with the prescribed marginal distributions and correlation as described in the Repopulation flowchart shown in Fig. 2. In fact, starting now from a sample (realization) of the normal distributed random variables $\vec{\xi}$ we first determine the related middle variables $\vec{z}$ by means of (4). Second, the associated physical variables x_k are computed through:

$$x_k = F_k^{-1}\left[\Phi(z_k)\right] \tag{5}$$

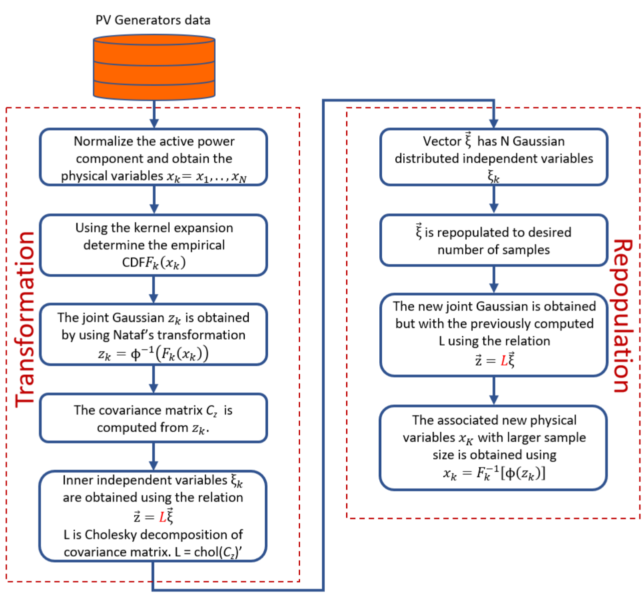


Fig. 2. The flowchart explaining the procedure to generate the large sample set of physical input variables x_k applying Gaussian copula on the PV profiles.

The Repopulation computational flow establishes a deterministic relationship linking the inner variables to the physical ones, i.e.:

$$x_k = h_k(\xi_1, \ldots, \xi_N) \tag{6}$$

It is worth noticing how such a relationship $h_k(\cdot)$ can exhibit a *high degree of non linearity* since it incorporates the application of normal CDF $\Phi(\cdot)$ and of the inverse CDF $F_k^{-1}(\cdot)$. *As a result, the numerical approximation of $h_k(\cdot)$ with polynomials can introduce inaccuracy even when high-order polynomial degrees are adopted.*

C. PLF with MC

The Copula computational flow can be exploited to perform PLF simulations of the power grid with Monte Carlo (MC). To this aim, we denote with y a quantity of interest in the power grid or observable variable that we want to describe statistically. Relevant observable variables y are node voltage module, node phase, line current or more complicate figure of merit describing grid power quality such as Voltage Unbalance Factor (VUF). The generic observable variable is related to the statistical physical ones via the deterministic relationship:

$$y = l(x_1, \ldots, x_N) \tag{7}$$

determined by grid power flow balance and Kirchhoff's laws. It is worth noting that: *commonly, the function $l(\cdot)$ relating generated powers to observable variables (e.g., node voltages) is almost linear and is well approximated by polynomials in x_k of order ≤ 2* [1].

III. PLF with Stochastic Response Surface Method (SRSM)

A. Conventional SRSM

The most straightforward way to apply conventional SRSM [25], [26] in connection with Copula is that of using the surrogate model to approximate the multi-variate relationship between Copula inner parameters ξ_k and grid observable variable y. Such a relationship is the result of a function composition. In fact, due to Copula transformation, physical variables x_k can be seen as a *nonlinear* function of inner parameters ξ_k via (6). Since observable variable y depends on injected normalized powers x_k, it turns out that output y ultimately depends on Copula inner variables ξ_k through the following input-output relationship:

$$y = l(x_1, \ldots, x_N) = \widetilde{h}(\xi_1, \ldots, \xi_N) \tag{8}$$

In its conventional implementation with Copula, the surrogate SRSM model approximates the $\widetilde{h}(\cdot)$ relationship with the polynomial chaos expansion:

$$y = \widetilde{h}(\vec{\xi}) \approx \sum_{i=0}^{N_b-1} \widetilde{c}_i\, H_i(\vec{\xi}), \tag{9}$$

formed by N_b multi-variate basis functions $H_i(\vec{\xi})$. The advantage of this approach is the ease of implementation: the inner stochastic parameters ξ_k are standard normal distributed and thus the associated polynomial basis functions $H_i(\vec{\xi})$ are standard Hermite polynomials [29].

The drawback of such an implementation is due to the nonlinear nature of the $x_k = h_k(\xi_1, \ldots, \xi_N)$ relationship determined by Copula transformation when dealing with PV delivered powers. In this case, polynomial approximation (9) can provide a quite poor approximation. To better explain this concept, we now present and example where we employ Hermite polynomials, of growing order, to interpolate the relationship of type (5) established by the Copula computational flow:

$$h(\xi) = F^{-1}\left[\Phi(\xi)\right] \tag{10}$$

where ξ is a normal distributed random variable, $\Phi(\cdot)$ denotes the Gaussian cumulative distribution while $F^{-1}(\cdot)$ is the inverse CDF of normalized PV power whose PDF is shown in Fig. 1(Bottom). Fig. 3 shows function $h(\xi)$ interpolation

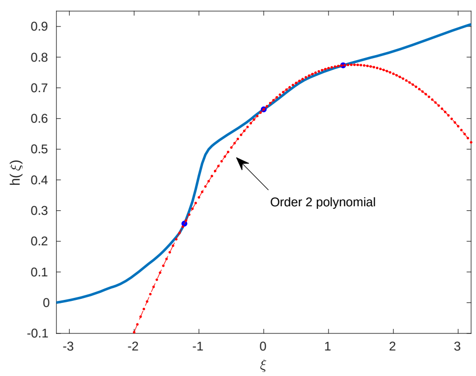


Fig. 3. (Continuous line) the nonlinear relationship $h(\xi)$; (Dashed line) Order-2 Hermite polynomial interpolation.

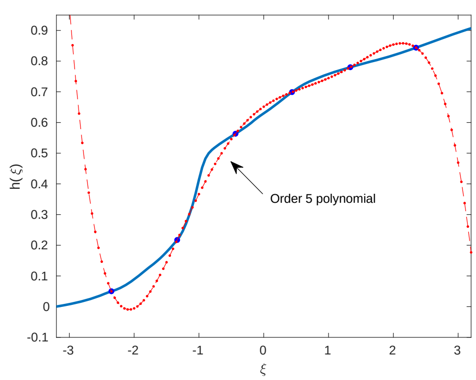


Fig. 4. (Continuous line) the nonlinear relationship $h(\xi)$; (Dashed line) Order-5 Hermite polynomial interpolation.

with an order-2 Hermite polynomial while Fig. 4 shows the order-5 polynomial interpolation. According to Theory, as interpolation points, we select the Gaussian quadrature nodes (the number of nodes is equal to polynomial Order plus 1) [29]. We can see how polynomial interpolation results to be rather inaccurate far from the interpolation points and that when polynomial order is increased to 5, interpolation oscillations appear at the borders of the domain (well known problem of over-fitting) [30].

B. The idea behind novel SRSM

The novel technique we propose overcomes the drawback of conventional implementation by applying Gaussian Copula and SRSM surrogate model in two separate steps. At the first step, the Repopulation computational flow of Copula sketched in Fig. 2 is used to enlarge the small data sets of PV delivered powers available over each considered window time. At the second step, a polynomial SRSM surrogate model is employed to accurately approximate the mildly nonlinear relationship linking normalized PV delivered powers to observable output y:

$$y = l(x_1, \ldots, x_N) \tag{11}$$

The SRSM can then be used as an efficient way of evaluating $y = l(x_1, \ldots, x_N)$ in place of running a large number of deterministic Load Flow. *The advantage of SRSM novel implementation compared to conventional one relies on the usage of low order polynomials, which results in much more accurate approximations*. The price to pay with the proposed implementation is that now surrogate SRSM cannot be built with standard polynomial chaos bases since variables x_k are not standard distributed. Instead, proper generalized basis functions associated to x_k statistical variables have to be precalculated. This can be achieved via a three-term recurrence relation and iterative Darboux's formula as illustrated in [31], [32].

C. Mathematical implementation

A truncated series expansion of an order-γ as given in (12) is formed by N_b multi-variate generalized basis functions $\Psi_i(\vec{x})$ weighted by unknown coefficients c_i.

$$y = l(\vec{x}) \approx \sum_{i=0}^{N_b-1} c_i\, \Psi_i(\vec{x}), \tag{12}$$

Each multi-variate basis function $\Psi_i(\vec{x})$ is given by the product

$$\Psi_i(\vec{x}) = \prod_{k=1}^{N} \psi_{i_k}(x_k) \tag{13}$$

where $\psi_{i_k}(x_k)$ are the uni-variate polynomials of degree i_k associated to the random variable x_k having non-standard marginal probability density function (PDF) $f_k(x_k)$.

The uni-variate polynomials for non-standard distributed physical variables x_k are obtained through a three-term recurrence relation applied to the empirical (i.e., data-driven) PDFs [31].

At this stage, the following observation is in order. When variables x_k are mutually independent, the multi-variate polynomials of (13), given by the product of uni-variate ones, provide a set of orthogonal basis functions that allow approximating the smooth $y(\vec{x})$ with a high accuracy and in addition to extract the statistical moments of observable y analytically. When variables x_k are not independent, as it is our case, the multi-variate polynomials in (13) are no more orthogonal among them so the expansion given in (12) cannot be used for analytical computations. To this aim, a new set of orthogonal polynomial basis might be derived by applying a Gram-Smith orthogonalization process [33].

However, when (12) is simply used as a surrogate model for accelerating MC method, polynomials orthogonalization is

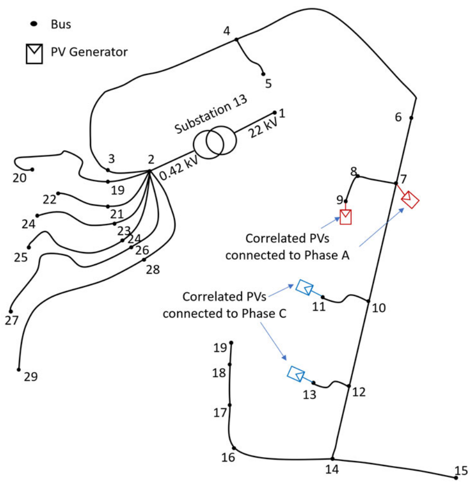


Fig. 5. Detail of the NSELVTN feeder where correlated PV generators are injected into Phase A and Phase C.

not required anymore. In fact, multi-variate functions of (13) continue to preserve excellent approximation capabilities of the smooth relationship $y(\vec{x})$ even in the presence of correlated variables. Thus they can still be adopted as basis functions in the proposed response surface model given by (12).

The expansion coefficients c_i in (12) can be calculated with a Least Square regression technique in which $N_s \geq N_b$ samples are generated for $\vec{x}$. In each sample $\vec{x}_j$ the observable variable value $y_j = l(\vec{x}_j)$ is calculated by running a deterministic load flow analysis. The coefficients c_i minimize the squared difference between the observable variables and the gPC evaluated at the samples, i.e.:

$$c_i \approx \arg\min_{\tilde{c}_i} \frac{1}{N_s} \sum_{j=1}^{N_s} \left(y_j - \sum_{i=0}^{N_b-1} \tilde{c}_i \, \Psi_i(\vec{x}_j) \right)^2 \qquad (14)$$

(14) is solved by introducing the experiment matrix as given in (15), it collects the N_b multi-variate polynomials evaluated at the N_s samples.

$$\mathbf{M} = \begin{bmatrix} \Psi_0(\vec{x}_1) & \dots & \Psi_{N_b-1}(\vec{x}_1) \\ \vdots & \ddots & \vdots \\ \Psi_0(\vec{x}_{N_s}) & \dots & \Psi_{N_b-1}(\vec{x}_{N_s}) \end{bmatrix} \qquad (15)$$

In vector form, it results in the linear system as given in (16).

$$(\mathbf{M}^T\mathbf{M}) \begin{bmatrix} c_0 \\ \vdots \\ c_{N_b-1} \end{bmatrix} = \mathbf{M}^T \begin{bmatrix} y_1 \\ \vdots \\ y_{N_s} \end{bmatrix} \qquad (16)$$

IV. Results

A. Definitions and Test network

A power distribution network with deterministic loads and uncertain PV generators is employed for validating the method. The network is the NSELVTN [34] that represents a real distribution network of an European town, it operates at 230V, 50Hz frequency. In order to investigate the impact of PV generation, PV sources representing PV power plants are inserted with different penetration factors. The penetration factor is defined as the ratio between the installed peak photovoltaic P_{PV} power and the total peak power of the loads P_L ([35]). The P_L is 350kW for each phase.

First, in the following sub-section IV-B the method is applied considering four PV sources with realistic degrees of mutual correlations. Later in the sub-section IV-C, the impact of PV penetration on grid power quality is studied.

To define grid power quality, several indicators go into defining the metrics that characterize network quality [23], [36], [37]. We have chosen in this article to look at only two of these indicators: Voltage Deviation (VD) and Voltage Unbalance Factor (VUF) [38] evaluated at some critical buses into the network.

B. Validation of the proposed method

In order to validate the proposed method, one feeder of the NSELVTN is modified with the injection of PV sources. This feeder is made by 90 buses and we focus the observations only on 29 of them as shown in the detail of Fig. 5. This feeder represents a small portion of the larger network that we simulate. The PV sources are connected to Phase A and Phase C as described in Table I and shown in the Fig. 5. These PVs will inject power into the lines assuming PV penetration of 5.7% in phase A and of 3.4% in the phase C. The source are correlated and the correlation coefficients are shown in Table II.

TABLE I
Peak value of the PV sources active power and total value of the distributed load for each phase. The $\cos\phi$ is equal to 0.9

Parameter	Phasing	Bus name	P [kW]
$1-ph$ PV Generator 1	A	7	10
$1-ph$ PV Generator 2	A	9	10
$1-ph$ PV Generator 3	C	11	6
$1-ph$ PV Generator 4	C	13	6
$1-ph$ Loads	A	Distributed	125
$1-ph$ Loads	B	Distributed	70
$1-ph$ Loads	C	Distributed	60

TABLE II
Correlation coefficient among PV generators

PV Generator	PV1	PV2	PV3	PV4
PV1	1.00	0.68	0.72	0.48
PV2	0.68	1.00	0.94	0.87
PV3	0.72	0.94	1.00	0.72
PV4	0.48	0.87	0.78	1.00

As observable variable, we select the node voltage at bus 12, Phase C. We compute the PDF of such a variable with:

1) The reference Monte Carlo (MC) method conducted with 10,000 samples.
2) the conventional SRSM implementation;
3) the novel SRSM presented in this paper.

The Fig. 6 shows a comparison among the results provided by the methods. When compared to the reference MC, we see how the conventional SRSM, even though was implemented with order-5 Hermite polynomials, fails to approximate the results accurately in case of non standard distributions as

TABLE III
METHODS COMPARISON

	reference MC	Conventional SRSM	New SRSM
LF Numb.	10,000	36	15
Sim. time [s]	291	16.5	5.9
Rel. Error	—	≈ 10 %	< 0.4 %

inputs. By contrast, the proposed novel SRSM, implemented with generalized order-2 polynomials chaos, fits with great accuracy the MC simulation result that runs $10,000$ samples. Table III compares the three methods in terms of number of Load Flow simulations required, simulation time and PDF accuracy (with respect to reference MC method). The PDF computed with novel SRSM implementation has a relative error versus the one computed with MC that is < 0.004 in all points. Furthermore, in this example, the novel SRSM introduces a remarkable $\approx 50\times$ speed up factor compared to MC simulation for the same accuracy.

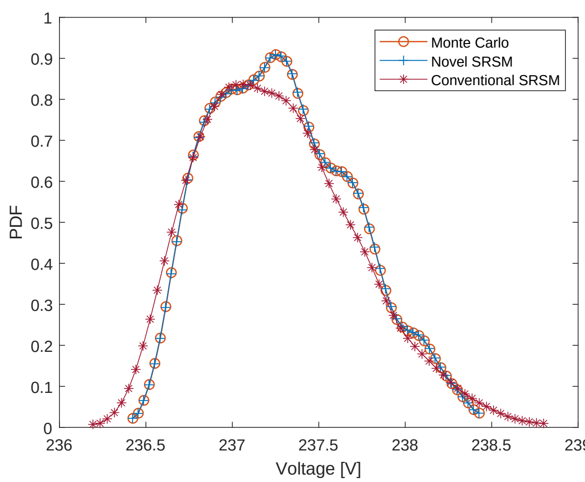


Fig. 6. Comparison of Phase C voltage distribution at bus 12 computed with the methods: a) Monte Carlo simulation with 10,000 samples (reference method); b) conventional SRSM using order-5 Hermite polynomials; c) novel SRSM using order-2 generalized polynomial chaos.

Thanks to its numerical efficiency as well as accuracy, the novel SRSM method can thus be exploited for comprehensive and reliable investigations about several observable variables/figures of merit of interest and for different hourly time windows of the day. For instance, the voltage distributions at phase A and C at the same bus are shown in Fig. 7 and in Fig. 8, respectively, for different time slots. An increase in voltage magnitude is seen with a maximum in the middle hours of the day due to the increase of the injected PV power.

As a second example, we calculate the statistical distribution of VUF. VUF is defined as the ratio of the negative voltage sequence component to the positive voltage sequence component and is known to be a key figure of merit in order to assess network health. Fig. 9 reports VUF calculated at bus 12: interestingly it is seen how VUF mean value reduces in the middle hours of the day where PV power injection is larger. This is because, in the considered scenario, the network with

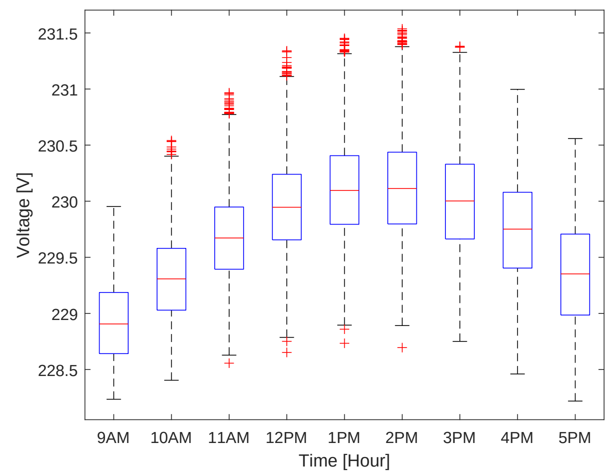


Fig. 7. Box diagram for the hourly voltage distribution of Phase A at bus 12.

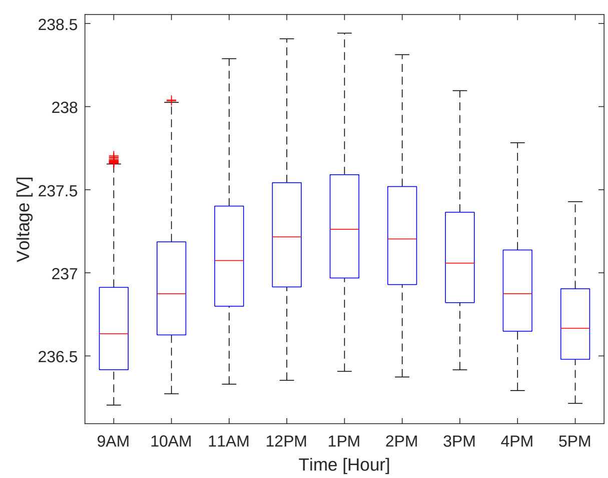


Fig. 8. Box diagram for the hourly voltage distribution of Phase C at bus 12.

no PV injection is unbalanced by loads while PV injection tends to compensate such an unbalance.

C. Study of violation of power grid quality due to PV penetration

Finally, in this last section, we are going to show how the proposed numerically-efficient SRSM can be exploited to rapidly evaluate the probability of violation of some network constraints versus the penetration rate of PV sources. In fact, by means of joint evaluations of several metrics it is possible to define the quality of the grid variables and determine the critical value that should not be reached in order to keep a good quality of the distribution network. In our example, we consider two cases in which we insert six PV sources on phase C line. In the case 1, the penetration index for the phase C is fixed to 10.3% while, in the case 2, it is enhanced to 17.2%. The value and the total amount of PV power sources is shown in Table IV while the correlation coefficients assumed in simulation are given in Table V. The positions of the PV generators are highlighted in the Fig. 10 In this example, we

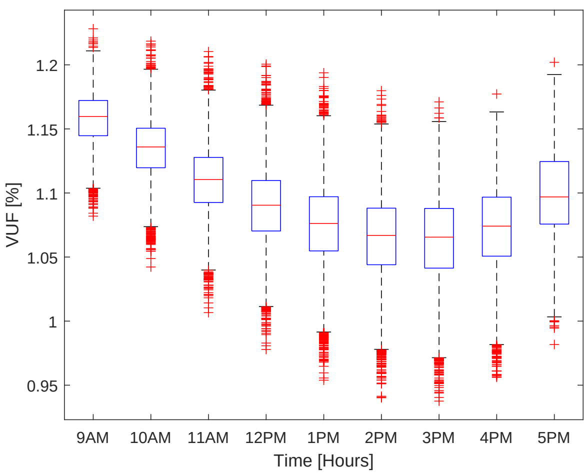


Fig. 9. Box diagram for the Voltage Unbalance Factor at bus 12 in percentage.

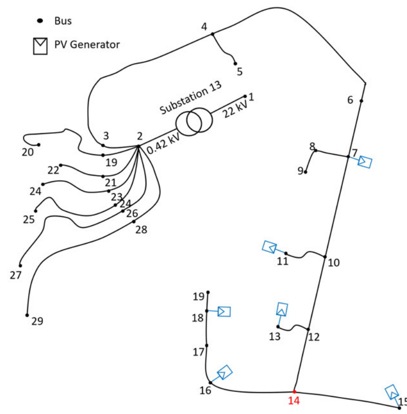


Fig. 10. PV generators connected to Phase C of the network with the penetration ratio as in case1. Buses with the PV generators are in blue, the bus without PV generator but critically violating the set thresholds than all other buses is in Red and called as critical bus.

assume as critical bus the bus number 14 in Fig. 10 where the node voltages are employed as observable variables to check the network quality in terms of voltage deviations (VD) and voltage unbalance factor (VUF) indicators. In simulations, the VD threshold is assumed $0.03\% p.u.$ while VUF threshold is 3%.

TABLE IV
PV GENERATORS AND ITS PENETRATION RATIO FOR PHASE C

	PV Penetration for phase C Ratio [%]	Active power [kW]	Total number of PV generators
Case I	10.3%	36	6 units
Case II	17.14%	60	6 units

TABLE V
CORRELATION COEFFICIENT AMONG 6 PV GENERATORS

PV Generator	PV1	PV2	PV3	PV4	PV5	PV6
PV1	1.00	0.19	0.82	0.74	0.18	0.82
PV2	0.19	1.00	0.47	0.63	0.90	0.49
PV3	0.82	0.47	1.00	0.97	0.46	0.95
PV4	0.74	0.63	0.97	1.00	0.62	0.98
PV5	0.18	0.90	0.46	0.62	1.00	0.48
PV6	0.82	0.49	0.95	0.98	0.48	1.00

Fig. 11 reports the statistical distribution (i.e., the PDF) of the VD indicator, as computed with the novel SRSM, for the two different levels of PV penetration considered in case 1 and 2. Hence, by calculating the portion of the areas of the two PDFs corresponding to Voltage > 0.03 p.u. we determine that the probability of VD indicator violation are 51.1% and 62.6% in case 1 and 2, respectively.

Similarly, Fig. 12 shows the statistical distribution of VUF% indicator as computed with the numerically-efficient SRSM. The likelihood of VUF violation is 11% in case 1 while it increases to more than and 52% in case 2.

Exploiting the SRSM, we also calculate the joint probability of VD and VUF simultaneous violations that results in 11% in case 1 and 52% in case 2. This confirms that VUF indicator is strictly correlated to VD so that when the first indicator violation occurs also VD is over the threshold.

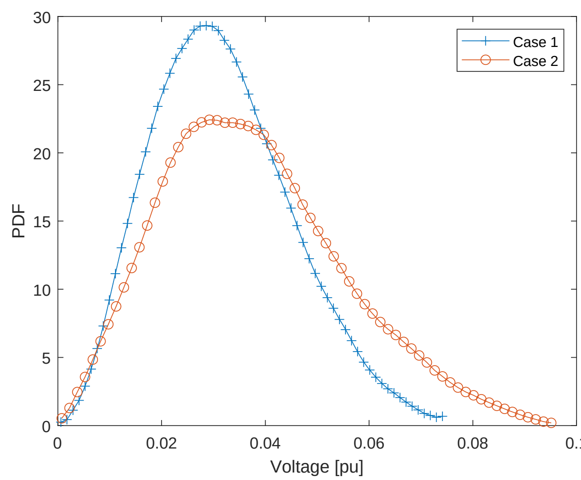


Fig. 11. The voltage deviation in bus 14 for the two analyzed cases

V. Conclusion

In this paper, we have provided an accurate and numerically-efficient technique for evaluating the statistical uncertainty of power grid electrical variables due to the penetration of PV distributed sources. The proposed technique combines the Gaussian Copula method, for handling correlation among PV sources, with an original implementation of the stochastic response surface method (SRSM) able to deal with non-standard statistical distributions. The main feature of the novel approach is a two-step implementation of Copula and SRSM that allows the employment of low-order generalized polynomial chaos in place of high-order standard Hermite polynomials (as it is the case for conventional SRSM implementations). The accuracy/efficiency of the novel SRSM methodology have been checked via probabilistic simulations of the Non-synthetic European low voltage test network and validated through comparisons with the reference Monte Carlo (MC) method. For the examples presented in the paper, the novel SRSM introduces an almost two-orders of magnitude speed-up factor compared to MC, for the same accuracy. This

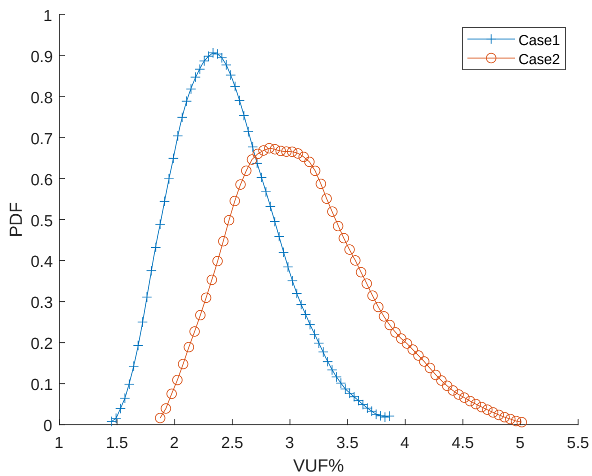


Fig. 12. The VUF% distribution in bus 14 for the two analyzed cases

reduces significantly the computational time needed to explore several scenarios for assessing the quality of the grid due to an increase of PV sources installed. More specifically, it has been shown how SRSM allows fast probabilistic calculations of important grid quality indicators, such as voltage deviation or voltage unbalance, with the determination of the probability that such indicators violate prescribed quality thresholds.

References

[1] J. Tang, F. Ni, F. Ponci, and A. Monti, "Dimension-adaptive sparse grid interpolation for uncertainty quantification in modern power systems: Probabilistic power flow," *IEEE Transactions on Power Systems*, vol. 31, no. 2, pp. 907–919, March 2016.

[2] S. Chevalier, L. Schenato, and L. Daniel, "Accelerated probabilistic power flow in electrical distribution networks via model order reduction and neumann series expansion," *IEEE Transactions on Power Systems*, pp. 1–1, 2021.

[3] Z. A. Memon, R. Trinchero, Y. Xie, F. G. Canavero, and I. S. Stievano, "An iterative scheme for the power-flow analysis of distribution networks based on decoupled circuit equivalents in the phasor domain," *Energies*, vol. 13, no. 2, 2020. [Online]. Available: https://www.mdpi.com/1996-1073/13/2/386

[4] A. Pandey, M. Jereminov, M. R. Wagner, D. M. Bromberg, G. Hug, and L. Pileggi, "Robust power flow and three-phase power flow analyses," *IEEE Transactions on Power Systems*, vol. 34, no. 1, pp. 616–626, Jan 2019.

[5] C. S. Lai, Y. Jia, M. D. McCulloch, and Z. Xu, "Daily clearness index profiles cluster analysis for photovoltaic system," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 5, pp. 2322–2332, 2017.

[6] F. Li, Y. Lin, J. Guo, Y. Wang, L. Mao, Y. Cui, and Y. Bai, "Novel models to estimate hourly diffuse radiation fraction for global radiation based on weather type classification," *Renewable Energy*, vol. 157, pp. 1222–1232, 2020.

[7] A. K. Barnes, J. C. Balda, and A. Escobar-Mejía, "A semi-markov model for control of energy storage in utility grids and microgrids with pv generation," *IEEE Transactions on Sustainable Energy*, vol. 6, no. 2, pp. 546–556, 2015.

[8] M. Bashir and J. Sadeh, "Optimal sizing of hybrid wind/photovoltaic/battery considering the uncertainty of wind and photovoltaic power using monte carlo," in *2012 11th International Conference on Environment and Electrical Engineering*, 2012, pp. 1081–1086.

[9] B. Zhou, X. Yang, D. Yang, Z. Yang, T. Littler, and H. Li, "Probabilistic load flow algorithm of distribution networks with distributed generators and electric vehicles integration," *Energies*, vol. 12, no. 22, p. 4234, 2019.

[10] U. H. Ramadhani, M. Shepero, J. Munkhammar, J. Widén, and N. Etherden, "Review of probabilistic load flow approaches for power distribution systems with photovoltaic generation and electric vehicle charging," *International Journal of Electrical Power & Energy Systems*, vol. 120, p. 106003, 2020.

[11] H. Palahalli, P. Maffezzoni, and 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.

[12] M. Fan, V. Vittal, G. T. Heydt, and R. Ayyanar, "Probabilistic power flow studies for transmission systems with photovoltaic generation using cumulants," *IEEE Transactions on Power Systems*, vol. 27, no. 4, pp. 2251–2261, 2012.

[13] W. Wu, K. Wang, B. Han, G. Li, X. Jiang, and M. L. Crow, "A versatile probability model of photovoltaic generation using pair copula construction," *IEEE Transactions on Sustainable Energy*, vol. 6, no. 4, pp. 1337–1345, 2015.

[14] Y. Wang, N. Zhang, Q. Chen, J. Yang, C. Kang, and J. Huang, "Dependent discrete convolution based probabilistic load flow for the active distribution system," *IEEE Transactions on Sustainable Energy*, vol. 8, no. 3, pp. 1000–1009, 2017.

[15] J. Widén, M. Shepero, and 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*, vol. 7, no. 6, pp. 1740–1745, 2017.

[16] H. Palahalli, P. Maffezzoni, and G. Gruosso, "Gaussian copula methodology to model photovoltaic generation uncertainty correlation in power distribution networks," *Energies*, vol. 14, no. 9, p. 2349, 2021.

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

[18] A. Koirala, T. V. Acker, D. Van Hertem, J. Camargo, and R. D'hulst, "General polynomial chaos vs crude monte carlo for probabilistic evaluation of distribution systems," in *2020 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, 2020, pp. 1–6.

[19] H. Su, X. Dong, and X. Yu, "Probabilistic load flow analysis based on sparse polynomial chaotic expansion," *Journal of Electrical Engineering & Technology*, vol. 15, no. 2, pp. 527–538, 2020.

[20] G. Gruosso, R. S. Netto, L. Daniel, and P. Maffezzoni, "Joined probabilistic load flow and sensitivity analysis of distribution networks based on polynomial chaos method," *IEEE Transactions on Power Systems*, vol. 35, no. 1, pp. 618–627, 2020.

[21] Z. Ren, W. Li, R. Billinton, and W. Yan, "Probabilistic power flow analysis based on the stochastic response surface method," *IEEE Transactions on Power Systems*, vol. 31, no. 3, pp. 2307–2315, 2015.

[22] G. Wang, H. Xin, D. Wu, P. Ju, and X. Jiang, "Data-driven arbitrary polynomial chaos-based probabilistic load flow considering correlated uncertainties," *IEEE Transactions on Power Systems*, vol. 34, no. 4, pp. 3274–3276, 2019.

[23] M. S. S. Abad, J. Ma, D. Zhang, A. S. Ahmadyar, and H. Marzooghi, "Probabilistic assessment of hosting capacity in radial distribution systems," *IEEE Transactions on Sustainable Energy*, vol. 9, no. 4, pp. 1935–1947, 2018.

[24] "Dka solar centre, alice springs project." [Online]. Available: https://dkasolarcentre.com.au/

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

[26] S. S. Isukapalli, *Uncertainty analysis of transport-transformation models*. Rutgers The State University of New Jersey-New Brunswick, 1999.

[27] A. W. Bowman and A. Azzalini, *Applied smoothing techniques for data analysis: the kernel approach with S-Plus illustrations*. OUP Oxford, 1997, vol. 18.

[28] Y. Chen, J. Wen, and S. Cheng, "Probabilistic load flow method based on nataf transformation and latin hypercube sampling," *IEEE Transactions on Sustainable Energy*, vol. 4, no. 2, pp. 294–301, 2013.

[29] D. Xiu and G. E. Karniadakis, "The wiener–askey polynomial chaos for stochastic differential equations," *SIAM journal on scientific computing*, vol. 24, no. 2, pp. 619–644, 2002.

[30] C. M. Bishop, *Pattern Recognition and Machine Learning*, 1st ed. Springer, 1 2006.

[31] G. Gruosso and P. Maffezzoni, "Data-driven uncertainty analysis of distribution networks including photovoltaic generation," *International*

Journal of Electrical Power & Energy Systems, vol. 121, p. 106043, 2020.

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

[33] C. Cui and Z. Zhang, "Stochastic collocation with non-gaussian correlated process variations: Theory, algorithms, and applications," *IEEE Transactions on Components, Packaging and Manufacturing Technology*, vol. 9, no. 7, pp. 1362–1375, 2018.

[34] A. Koirala, L. Suárez-Ramón, B. Mohamed, and P. Arboleya, "Non-synthetic european low voltage test system," *International Journal of Electrical Power & Energy Systems*, vol. 118, p. 105712, 2020.

[35] Q. Wu, B. Han, M. Shahidehpour, C. Qi, G. Li, K. Wang, and L. Luo, "Extreme bifurcation interval with deep pv penetration considering correlated net bus load forecast errors," *IEEE Transactions on Sustainable Energy*, vol. 11, no. 4, pp. 2063–2076, 2019.

[36] M. Karimi, H. Mokhlis, K. Naidu, S. Uddin, and A. Bakar, "Photovoltaic penetration issues and impacts in distribution network – a review," *Renewable and Sustainable Energy Reviews*, vol. 53, pp. 594–605, 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S136403211500903X

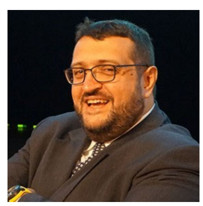
[37] F. Ding and B. Mather, "On distributed pv hosting capacity estimation, sensitivity study, and improvement," *IEEE Transactions on Sustainable Energy*, vol. 8, no. 3, pp. 1010–1020, 2017.

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

HARSHAVARDHAN PALAHALLI M. (Student Member, IEEE) received Bachelor degree in Electrical and Electronics Engineering from Visvesvaraya Technological University, India, in 2014, and a Masters degree in Electrical Engineering from Politecnico Di Milano, Italy, in 2018. Currently, he is in Politecnico di Milano as a Ph.D. candidate. His research interests include Probabilistic load flow, uncertainty quantification of PV and EV, Real-time simulation of electric systems, digital twins and Hardware-in-the-loop simulation.

Paolo Maffezzoni (Senior Member, IEEE) received the Laurea degree (summa cum laude) in electrical engineering from the Politecnico di Milano, Italy, in 1991, and the Ph.D. degree in electronic instrumentation from the Università di Brescia, Italy, in 1996. He is currently a Full Professor of Electrical Engineering with the Politecnico di Milano. His research interests include the development of advanced computational methods for the modeling, simulation and design of nonlinear circuits and systems, with applications to electronics and power systems. He has published more than 140 research papers in international journals and conferences. He has served as an Associate Editor for the IEEE Transactions on Computer- Aided Design of Integrated Circuits and Systems and the IEEE Transactions on Circuits and Systems I: Regular Papers.

Pablo Arboleya (Senior Member, IEEE) received the M.Sc. and Ph.D. (with distinction) degrees from the University of Oviedo, Gijón, Spain, in 2001 and 2005, respectively, both in electrical engineering. He is currently working as an Associate Professor with the Department of Electrical Engineering, University of Oviedo. He is the Managing Editor of the International Journal of Electrical Power and Energy Systems and is holder of the Gijón Smart Cities Chair with the University of Oviedo. His main research interests are focused in the micro-grid and smart-grid modeling and operation techniques, Internet of the energy applications, railway traction networks simulation and combined ac/dc power flow algorithms.



Giambattista Gruosso (Senior Member, IEEE) was born in 1973. He received the B.S. and M.S. degrees in electrical engineering and the Ph.D. degree in electrical engineering from the Politecnico di Torino, Italy, in 1999 and 2003, respectively. From 2002 to 2011, he was an Assistant Professor with the Department of Electronics and Informatics, Politecnico di Milano, where he has been an Associate Professor, since 2011. He is the author of more than 80 papers on Journals and conferences on the topics. He does research in electrical engineering, electronic engineering, and industrial engineering. His main research interests include electrical vehicles transportation electrification, electrical power systems optimization, simulation of electrical systems, digital twins for smart mobility, factory and city, and how they can be obtained from data.