

Article

Piece-Wise Linear (PWL) Probabilistic Analysis of Power Grid with High Penetration PV Integration

Giambattista Grusso¹ , Luca Daniel² and Paolo Maffezzoni^{1,*} 

¹ Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Piazza Leonardo da Vinci, 32-20133 Milano, Italy; giambattista.grusso@polimi.it

² Massachusetts Institute of Technology (MIT), Cambridge, MA 02139, USA; dluca@mit.edu

* Correspondence: paolo.maffezzoni@polimi.it

Abstract: This paper aims at presenting a novel effective approach to probabilistic analysis of distribution power grid with high penetration of PV sources. The novel method adopts a Gaussian Mixture Model for reproducing the uncertainty of correlated PV sources along with a piece-wise-linear approximation of the voltage-power relationship established by load flow problem. The method allows the handling of scenarios with a large number of uncertain PV sources in an efficient yet accurate way. A distinctive feature of the proposed probabilistic analysis is that of directly providing, in closed-form, the joint probability distribution of the set of observable variables of interest. From such a comprehensive statistical representation, remarkable information about grid uncertainty can be deduced. This includes the probability of violating the safe operation conditions as a function of PV penetration.

Keywords: multivariate piece-wise linear approximation; power distribution grid; photovoltaic; probabilistic load flow; sensitivity analysis



Citation: Grusso, G.; Daniel, L.; Maffezzoni, P. Piece-Wise Linear (PWL) Probabilistic Analysis of Power Grid with High Penetration PV Integration. *Energies* **2022**, *15*, 4752. <https://doi.org/10.3390/en15134752>

Academic Editors: Janaka Ekanayake, Helder Leite and Meghdad Fazeli

Received: 7 June 2022

Accepted: 26 June 2022

Published: 28 June 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The growing penetration of photovoltaic (PV) generators in power distribution grid is expected to greatly increase the statistical uncertainty of power balance problem from the supply side [1]. Such an uncertainty can jeopardize the power grid operation health [2]. For such reasons, there is today great interest in innovative and efficient simulation tools for power grid probabilistic analysis [3–5]. Such novel tools should be able to predict the many potential adverse effects that massive distributed PV injection can bring with it, e.g., over voltage, unacceptable voltage or current variations.

To this aim, simulation tools should own some relevant features. First, they have to rely on realistic statistical models of PV uncertainty and be able to exploit the information supplied by measurements and historic data set [6]. Several approaches to PV modeling have been proposed in the literature [7], some of which are focused on reproducing meteorological conditions variability [8,9] while some others directly exploit PV delivered power data [10].

Evidences gained from such data-driven approaches show that: (a) PV delivered power follows non-standard statistical distributions [10,11], (b) geographically close PV plans tend to exhibit a certain degree of correlation. Techniques for including sources correlations are mostly based on Copula Method (CM) [12,13], or on Gaussian Mixture Models (GMM) [14,15]. These latter are particularly attractive since they provide a closed-form model for approximating the joint statistical distribution of many PV sources with arbitrary precision [16].

A second desirable feature of envisaged simulation tools is they should employ numerically effective probabilistic methods, allowing rapid explorations of several scenarios with power sources connected at different potential injection points. The basic and reference probabilistic method is still the iterative Monte Carlo (MC) simulation for the flexibility and

accuracy it can ensure. The main limitation of MC is the slow convergence of its iterative scheme and the long computational times it can require. Several probabilistic acceleration techniques have been proposed that approximate the input-output relationship between PV injected powers and node voltages variations. Many of such techniques rely on polynomial chaos expansions [17–19] and have been proved to supply excellent results. Unfortunately, they suffer of the curse of dimensionality and lose much of their effectiveness when the number of statistical parameters involved grows too much [20]. This can just be the case for a grid with high penetration PV generators.

To cope with such issues, in this paper we present an original methodology exploiting the evidence that, in the Load Flow (LF) problem, the dependence of node voltages on injected PV powers is almost linear for small relative power variations [21,22]. As a consequence, the multivariate voltage-power relationship can be accurately approximated through local piece-wise linear (PWL) approximations. To improve accuracy, the hyperplanes forming the PWL approximation are centered at those points in the statistical space where the joint Probability Density Function (PDF) of the PV sources is large.

The original contributions of this paper include:

1. A method for modeling with arbitrary precision the joint PDF of many correlated PV sources starting from data sets of PV delivered power. The method relies on a Gaussian Mixture Model (GMM) and produces a clusterization of injected power values. In fact, each Gaussian component identifies a data cluster while the Gaussian mean value vectors determine the centers of the clusters.
2. An accurate PWL approximation of the multivariate voltage-power relationship in the grid. This is determined by exploiting the typical computational flow of a LF simulation tool. More specifically, we show how the voltage sensitivity calculation provides the information required to build the local hyperplane approximation at each cluster center.
3. The demonstration of how the proposed PWL method allows the computing of closed-form expressions, i.e., a GMM, for the joint PDF of a set of observable variables of interest, i.e., a set of node voltages at critical buses in the grid.

The joint PDF can be used to calculate the uncertainty intervals of node voltages as well as the probability of violating safe operation conditions. Such computational capabilities are crucial in order to assess the PV hosting capacity of a given distribution grid.

The methodology is shown only for the case of a PV system, but can be generalized to other renewable energy sources.

The remainder of this paper is organized as follows: Section 1 briefly reviews LF formulation and voltage sensitivity computation. Sections 2–4 illustrate the novel contributions of this paper. In fact, in Section 2, we describe the data-driven GMM of correlated PV sources while, in Section 3, the rationale behind the PWL probabilistic analysis and its implementation details are outlined. Finally, in Section 4, we present several numerical experiments which prove the accuracy and efficiency of the proposed PWL method along with its many potentialities as an uncertainty quantification tool.

2. Power Flow Analysis and Sensitivity Computation

We consider a generic power distribution network composed of N nodes. The power flow problem consists in calculating the node voltage values producing the desired power flow at the network terminals. The problem is formulated mathematically by a set of nonlinear equations of the type:

$$\mathbf{f}_n(\vec{\mathbf{V}}) = \mathbf{V}_n^* \mathbf{I}_n - \mathbf{S}_n^* = 0 \quad (1)$$

for $n = 1, \dots, N$, where \mathbf{V}_n , \mathbf{I}_n are voltage and current phasors at node n , vector $\vec{\mathbf{V}} = [\mathbf{V}_1, \dots, \mathbf{V}_N]$ collects all voltages, while $\mathbf{S}_n = P_n + jQ_n$ denotes complex power at node n with P_n and Q_n being the active and reactive power, respectively. Furthermore, node currents are related to node voltages by means of

$$\mathbf{I}_n = \sum_{s=1}^N \mathbf{Y}_{ns} \mathbf{V}_s \tag{2}$$

where \mathbf{Y}_{ns} are the entries of the node admittance matrix. This one is determined by node-interconnection elements, such as grid lines, transformers and phase shifters, which are described by their admittance matrix. Power specification at the terminal nodes are given by loads and generators [22]. In this paper, the complex electrical quantities appearing in the power flow problem (1) are described in Cartesian coordinates (i.e., by their real and imaginary parts) and the resulting nonlinear equations are solved with Newton–Raphson (NR) method. Adopting vector-based notation, the n th voltage phasor $\mathbf{V}_n \in \mathbb{C}$, is represented by the 2×1 vector

$$\vec{V}_n = \text{vect}(\mathbf{V}_n) = \begin{bmatrix} V_n^R \\ V_n^I \end{bmatrix} \tag{3}$$

collecting its real and imaginary parts $V_n^R = \text{Re}\{\mathbf{V}_n\}$ and $V_n^I = \text{Im}\{\mathbf{V}_n\}$, respectively. Similarly, the power flow Equations (1) and (2) can be transformed into the associated vector form

$$\vec{f}_n = \begin{bmatrix} V_n^R & V_n^I \\ -V_n^I & V_n^R \end{bmatrix} \cdot \begin{bmatrix} I_n^R \\ I_n^I \end{bmatrix} + \begin{bmatrix} -P_n \\ Q_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{4}$$

and

$$\begin{bmatrix} I_n^R \\ I_n^I \end{bmatrix} = \sum_{s=1}^N \begin{bmatrix} G_{ns} & -B_{ns} \\ B_{ns} & G_{ns} \end{bmatrix} \cdot \begin{bmatrix} V_s^R \\ V_s^I \end{bmatrix}, \tag{5}$$

respectively, where $\mathbf{Y}_{ns} = G_{ns} + jB_{ns}$ and G_{ns} and B_{ns} are the conductance and the susceptance, respectively. In [23] it has been shown how vector forms (4), (5) and their derivatives allow efficient implementation through complex array operations. In this paper, instead, we describe how the power flow equations can be exploited to compute the sensitivity of some observable electric variables in the power grid (e.g., node voltages at critical buses) versus a set of parameters x_d , for $d = 1, \dots, D$, representing uncertainty. To this aim, we observe that the power flow Equations (4) and (5), collected for $n = 1, \dots, N$, form a system of $2 \times N$ nonlinear real equations that can be denoted in compact form as:

$$F(\vec{V}, \vec{x}) = 0 \tag{6}$$

where $\vec{V} = [V_1^R, V_1^I, \dots, V_N^R, V_N^I]^T$ is the column vector assembling node voltages real and imaginary parts while $\vec{x} = [x_1, \dots, x_D]^T$ denotes the parameters vector. In the subsequent Section 2, we will select as uncertainty parameters the (normalized) active power P_n injected in a set of *actor nodes* where PV generators are expected to be inserted. However, the sensitivity analysis that we are going to present is general and can be applied to any set of electrical parameters in the grid. Let us call \vec{V} the solution of the power flow problem (6) for a given value \vec{x} of parameters vector. A small perturbation $\delta\vec{x}$ of parameters will produce a small perturbation $\delta\vec{V}$ of node voltages such that the power flow problem (6) can be written as:

$$F(\vec{V}, \vec{x}) = F(\vec{V}, \vec{x}) + \frac{\partial F}{\partial \vec{V}} \delta\vec{V} + \frac{\partial F}{\partial \vec{x}} \delta\vec{x} + O(\delta^2) = 0 \tag{7}$$

where $O(\delta^2)$ are second-order terms. In the limit that perturbation goes to zero, neglecting second-order terms, we derive the matrix system:

$$\frac{\partial F}{\partial \vec{V}} \cdot \frac{\partial \vec{V}}{\partial \vec{x}} = - \frac{\partial F}{\partial \vec{x}} \tag{8}$$

The term $\frac{\partial F}{\partial \vec{V}}$ is the $2N \times 2N$ Jacobian matrix, while $\frac{\partial \vec{V}}{\partial \vec{x}}$ is the $2N \times D$ sensitivity matrix that contains the unknown (partial) derivatives of real and imaginary parts of node voltages versus parameters, i.e., $\frac{\partial V_n^R}{\partial x_d}$ and $\frac{\partial V_n^I}{\partial x_d}$. Finally, the $2N \times D$ right-hand side matrix $-\frac{\partial F}{\partial \vec{x}}$ collects the derivatives of power flow equations versus parameters. For instance, when the parameters are the active powers injected at nodes, i.e., $x_d = P_n$, from (4) we see that the columns of $-\frac{\partial F}{\partial \vec{x}}$ are vectors with a single 1 at P_n position and all zero the other terms.

It is worth observing how the solution of the sensitivity problem (8) requires little extra computation compared to the solution of nonlinear problem (6) since from this last one the Jacobian matrix and its LU decomposition are already available.

Finally, from $\frac{\partial V_n^R}{\partial x_d}$ and $\frac{\partial V_n^I}{\partial x_d}$, the sensitivity of voltage magnitudes $V_n = |\mathbf{V}_n|$ can be deduced as follows:

$$\frac{\partial V_n}{x_d} = \frac{1}{V_n} \left(V_n^R \frac{\partial V_n^R}{x_d} + V_n^I \frac{\partial V_n^I}{x_d} \right) \tag{9}$$

3. Gaussian Mixture Model of Correlated PV Sources

We are interested in reproducing, in a realistic way, the statistical uncertainty associated with the power delivered by representative PV installations distributed along the grid. The technique that we propose is general and can be applied to any available data set independently of the way it has been obtained (e.g., through measurements, historic data set or simulations).

The values of the power delivered by D PV sources, during a given daily time window (e.g., during the central hours of the day) are collected over several days and normalized

$$x_d^i = P_d^i / P_d^{Max}, \tag{10}$$

to the maximum installed power P_d^{Max} for the d th source. For such a PV source, the values x_d^i supply the (simultaneously-detected) samples data set from which the (empirical) statistical distribution of random variable x_d can be determined. Variables x_d are distributed over the interval $(0, 1)$.

Common features exhibited by the PV delivered active power are: (a) normalized powers x_d follow non-standard statistical distributions; (b) geographically close power sources show a certain degree of correlation that should be properly reproduced into the model.

To handle such a general case, we adopt a Gaussian Mixture Model (GMM), i.e., the joint PDF of random variables x_d is approximated by a model of the type [16,24]:

$$PDF(\vec{x}) = \sum_{k=1}^K \pi_k \mathcal{N}(\vec{x} | \vec{\mu}_k, \Sigma_k). \tag{11}$$

Each component $\mathcal{N}(\vec{x} | \vec{\mu}_k, \Sigma_k)$ is a multi-variate Gaussian density defined by its mean $\vec{\mu}_k$ and covariance matrix Σ_k [25]. The weighting parameters π_k are the mixing coefficients satisfying the conditions to be probabilities:

$$0 \leq \pi_k \leq 1, \quad \sum_{k=1}^K \pi_k = 1 \tag{12}$$

As a result, in the expression (11), the mixing coefficients can be seen as the prior probability of picking the k th component. i.e., $\pi_k = p(k)$, while each component $\mathcal{N}(\vec{x} | \vec{\mu}_k, \Sigma_k)$

reduces to the probability of \vec{x} conditioned to k , i.e., $p(\vec{x}|k)$. With this interpretation, (11) can be rewritten as

$$PDF(\vec{x}) = \sum_{k=1}^K p(k) p(\vec{x}|k) \quad (13)$$

For a given data set \vec{x}^i , with $i = 1, \dots, N_s$, formed by N_s samples, the parameters π_k , $\vec{\mu}_k$ and Σ_k in (11) can be extracted with an elegant and powerful iterative technique referred to as Expectation Maximization (EM) [24]. The main computational steps of EM algorithm are reported in the Appendix A, while an efficient implementation of EM can be found in the Matlab routine `fitgmdist.m`.

In the rest of this Section 2, we illustrate some features of the EM algorithm by exploiting a freely-available data set of PV measurements [26]. This dataset presents data from several PV plants installed in the same site, so the ground irradiance and weather conditions are the same. The only difference is in the panel technologies, their exposure, and the type of power control. The active power delivered by ten PV plants over the considered daily time window, i.e., from 12:00 p.m. to 1:00 p.m., after normalization to the maximum installed power, provide the samples of the random variables x_d , with $d = 1, \dots, D = 10$ of type (10).

The EM algorithm is then applied to this data set assuming a growing number K of Gaussian components. Figure 1 shows the scattered plot of variables x_5, x_6 samples along with the *projection* of the mean values $\vec{\mu}_k$ onto the x_5, x_6 subspace. Such variables exhibit a high degree of correlation since their samples tend to concentrate along the plane diagonal.

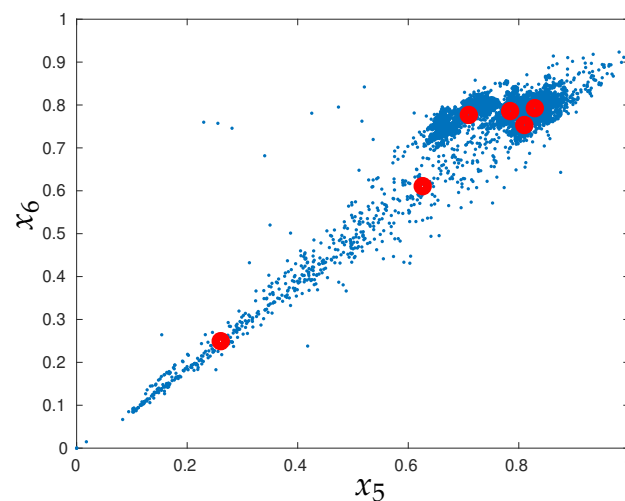


Figure 1. (Blue Dot): Scattered plot of variables x_5, x_6 samples; (Red Circles) the *projection* of the mean values $\vec{\mu}_k$ calculated by EM algorithm.

Figure 2 reports, as a second example, the scattered plot of variables x_5, x_1 samples and the projection of mean values $\vec{\mu}_k$ on the subspace they define. Variables x_5, x_1 are less correlated among them and their samples are distributed over a larger portion of the plane. Some observations are in order.

1. As it is well known in the literature, the EM works as a clustering algorithm where the mean values $\vec{\mu}_k$ represent the centers of the clusters [24].
2. EM tends to place the cluster centers where the samples are more dense that is where the joint $PDF(\vec{x})$ is large. As a result, in the regions with high samples density the centers are close among them while they are spaced apart in those regions with low sample density.

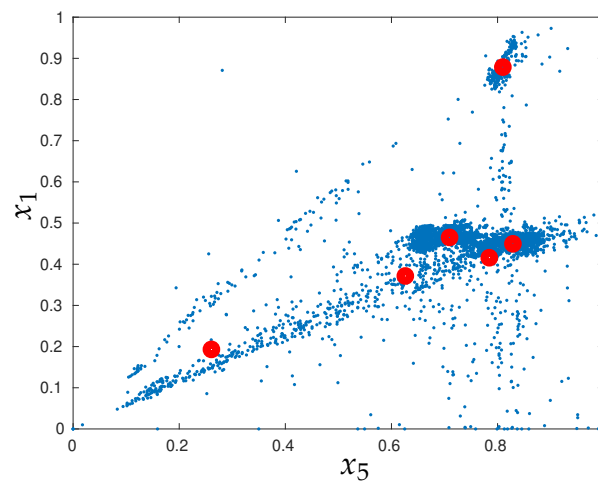


Figure 2. (Blue Dot): Scattered plot of variables x_5, x_1 samples; (Red Circles) the *projection* of the mean values $\bar{\mu}_k$ calculated by EM algorithm.

The aforementioned observations will be exploited in the PWL probabilistic analysis implementation presented in next Section 3.

In the rest of this section, it remains to see how, given a data set \bar{x}^i , an adequate number K of Gaussian components can be fixed. Even though some techniques for selecting K are available in the literature [24], we preferred to adopt the following eurism that revealed to work well with PV data. The eurism consists in increasing gradually the cluster number K in EM algorithm till it appears one cluster containing a too small fraction of the N_s samples. In fact, the *membership degree* of a given sample \bar{x}^i to the k th cluster can be calculated as:

$$m_{ik} = \mathcal{N}(\bar{x}^i | \bar{\mu}_k, \Sigma_k). \quad (14)$$

The sample \bar{x}^i is assigned to the cluster \hat{k} with the highest membership degree $m_{i\hat{k}}$, $m_{i\hat{k}} \geq m_{ik}$ for $k \neq \hat{k}$. In this way, all the samples are distributed among the clusters so that the k th cluster will contain n_k samples or equivalently the fraction of samples n_k/N_s . When the fraction of samples n_k/N_s within the smallest cluster goes down a given threshold (e.g., 2%) the algorithm is stopped. Figure 3 reports the fraction of samples n_k/N_s in the smallest cluster versus the number K of Gaussian components for the example herein considered.

It is seen how passing from $K = 6$ to $K = 7$ the fraction of samples in the smallest cluster has a knee and reduces to about 2%. This means that the Gaussian components in excess of six account for very small fractions of samples and thus are statistically negligible. For the considered example, $K = 6$ is an adequate number of Gaussian components.

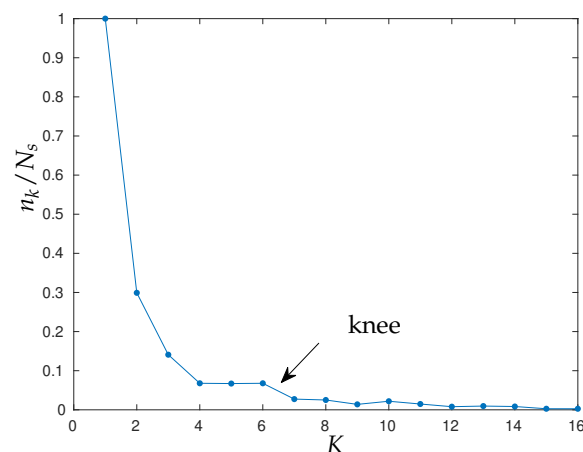


Figure 3. Fraction of samples n_k/N_s in the smallest cluster versus the number K of clusters.

4. Piece-Wise Linear Probabilistic Analysis

4.1. The Probabilistic Problem

We consider a distribution power grid containing D PV generators whose delivered active powers are represented by D random variables x_d having a GMM joint PDF (11).

We are interested in calculating the PDF of a set of observable electrical variables y_j , with $j = 1, \dots, O$ in the grid, for instance the module of the node voltage at some critical buses. Observable variables are assembled in the vector $\vec{y} = [y_1, \dots, y_O]^T$.

For any sample (i.e., any statistical realization) of vector \vec{x} its elements x_k , scaled by the associated installed power P_k^{Max} , provide the values of the PV powers to be injected into the grid. A deterministic load flow analysis of the grid, with such injected PV powers, allows the determination of the corresponding value assumed by the observable variables in the vector \vec{y} . Such a computational flow is the elementary mechanism which is at the basis of Monte Carlo PLF method. A very large number of samples of \vec{x} are generated according to the joint PDF (11) and the related \vec{y} are computed performing a deterministic load flow analysis for each sample. A sufficiently large set of computed sample of \vec{y} allows the estimation of the statistical distribution, that is the marginal $PDF(y_j)$ of each desired observable variable. Unfortunately, due to the slow convergence rate, MC can require several ten of thousands samples (i.e., repeated load flow analyses) to achieve satisfactory $PDF(y_j)$ approximations. Techniques for accelerating MC rely on the observation that the load flow analysis establishes a *deterministic vector valued relationship*

$$\vec{y} = l(\vec{x}) \quad (15)$$

between the parameters (e.g., PV injected powers) and the observable variables (e.g., the node voltages). In other terms, the element of vector \vec{y} are multi-variate functions $y_j(\vec{x})$ of the parameters. Acceleration techniques thus employ compact models, referred to as surrogate models, for approximating such multi-variate functions $y_j(\vec{x})$.

Stochastic Response Surface Methods (SRSM) or techniques based on generalized Polynomial Chaos (gPC) adopt series expansions of multi-variate polynomial basis functions *spanning the whole statistical/parameter space* to approximate $y_j(\vec{x})$ [17–19]. The success of such techniques in PLF analysis relies on the fact that load-flow-established relationships $y_j(\vec{x})$ are commonly almost linear and thus they can be accurately approximated by series expansions of low order (commonly order ≤ 2) polynomials. For problems of small size, a few load flow simulations are enough to build the surrogate model, i.e., to calculate the coefficients that weight the polynomials in the series. However, such techniques suffer of the curse of dimensionality since the number of weighting coefficients in the gPC expansion, corresponding to the minimum number of required load flow simulations, is given by $\frac{(D+\beta)!}{D! \beta!}$, with D being the number of statistical parameters and β the expansion order [17]. As a result, the number of required LF simulations grows super-linearly with D and β . For instance, an order-2 gPC expansion requires (at least) 66 simulations for $D = 10$ parameters whereas it requires (at least) 1326 simulations for $D = 50$ parameters. If an order-3 gPC model is adopted, the required minimum number of simulations to perform grows to 286 for $D = 10$ and 23,426 for $D = 50$.

4.2. Piece-Wise Linear Approximation

The idea behind the novel method consists in: *replacing the multi-variate polynomials spanning the whole domain/space with piece-wise linear approximations built around those points in the parameter space where the joint probability PDF(\vec{x}) is large.*

When the joint probability is given by model (11), the means $\vec{\mu}_k$ of the Gaussian components can be used to build the piece-wise approximation. In fact, in view of the analysis presented in Section 2, we know that the Gaussian means are the centers of the clusters where samples density is large.

Thus, in a neighborhood of the k th center $\vec{\mu}_k$, the vector valued relationship (15) can be linearized as:

$$\vec{y} = l(\vec{\mu}_k) + \left. \frac{\partial l}{\partial \vec{x}} \right|_{\vec{x}=\vec{\mu}_k} \cdot (\vec{x} - \vec{\mu}_k). \tag{16}$$

The term

$$S_{\vec{x}}(\vec{\mu}_k) = \left. \frac{\partial l}{\partial \vec{x}} \right|_{\vec{x}=\vec{\mu}_k} \tag{17}$$

is a $O \times D$ matrix containing the sensitivities of observable variables y_j versus parameters x_d computed at $\vec{\mu}_k$. Once, the load flow problem $\vec{y}_k = l(\vec{\mu}_k)$ has been solved, the computation of sensitivities requires little extra time. The rationale behind piece-wise linear approximation can be qualitatively illustrated with the help of Figure 4 showing the case of a single parameter x problem.

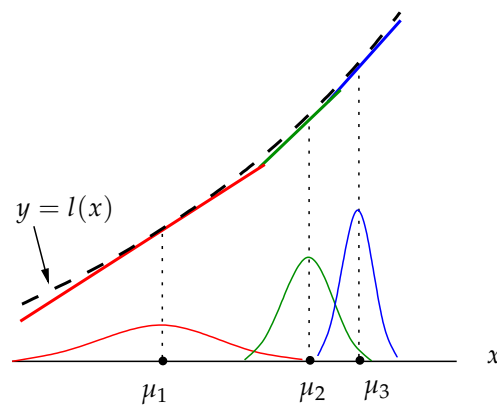


Figure 4. Piece-wise linear approximation of the almost linear relationship $y = l(x)$ around the GMM centers. Approximation accuracy is greater where $PDF(x)$ is large.

In this example, the $PDF(x)$ is given by the superposition of the three scaled Gaussian components reported in Figure 4. According to EM algorithm described in Section 2, the two centers μ_2 and μ_3 are placed where the $PDF(x)$ is large and are close one each other. On the contrary, the center μ_1 is placed at a greater distance apart where the $PDF(x)$ is much smaller. As a result, the linearizations around μ_2 and μ_3 , where the majority of x samples will indeed occur, match tightly the load flow relationship $y = l(x)$ while for $x \ll \mu_1$, the linearization accuracy can be relaxed. In this way, the inaccuracy due to local linearization involves a few samples eventually producing a negligible effects on $PDF(y)$ calculation.

Thanks to linearization (16), the $PDF(\vec{y}|k)$, i.e., conditioned to the fact that the k th Gaussian component has been selected, is a multi-variate Gaussian distribution too whose mean vector $\vec{\mu}_k^y$ and covariance matrix Σ_k^y are easily computed [25]. In fact,

$$\vec{\mu}_k^y = \langle \vec{y} \rangle = l(\vec{\mu}_k) \tag{18}$$

and

$$\Sigma_k^y = \langle \vec{y} - \vec{\mu}_k^y, \vec{y} - \vec{\mu}_k^y \rangle = S_{\vec{x}}(\vec{\mu}_k) \cdot \Sigma_k \cdot S_{\vec{x}}(\vec{\mu}_k)^T \tag{19}$$

where $\langle \cdot \rangle$ denotes the expectation operator. Finally, exploiting conditional probability properties, we derive the GMM for the observable variables vector \vec{y} :

$$PDF(\vec{y}) = \sum_{k=1}^K p(k) PDF(\vec{y}|k) = \sum_{k=1}^K \pi_k \mathcal{N}(\vec{y} | \vec{\mu}_k^y, \Sigma_k^y). \tag{20}$$

In conclusion, the GMM describing the joint statistics of uncertain PV powers \vec{x} is propagated through the load flow problem to describe the joint statistics of uncertain observable variables \vec{y} .

5. Numerical Experiments

The simulation experiments are performed on a modified three-phase version of the IEEE 69 bus test case [27]. This case is a single phase equivalent that we make three-phase in our simulation software [23]. For our simulation we use the same loads of the original case and the distribution grid topology, with the buses numbering reported in Figure 5.

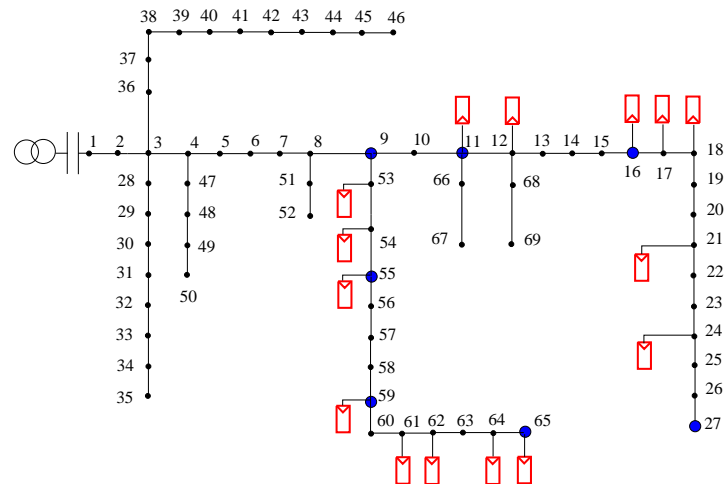


Figure 5. The IEEE 69 bus distribution grid. In red the PV and in Blue the point of observation.

The single-phase IEEE 69, bus grid is first extended to a three-phase network by replicating the deterministic single-phase loads (i.e., the nominal active and reactive powers they absorb) at the three phases of each bus. Second, uncertainty is reproduced into the three-phase grid by injecting random PV generators distributed along the feeders. PV generators are connected to the three phases of fifteen buses as highlighted in Figure 5 and this results in an unbalanced three-phase network. Table 1 lists the injection bus numbers along with their nominal active power P_n .

Table 1. Injection Buses.

Bus N.	P_n [p.u.]
11	0.145
12	0.145
16	0.045
17	0.060
18	0.060
21	0.114
24	0.028
53	0.0043
54	0.0264
55	0.0240
59	0.1000
61	1.2440
62	0.0320
64	0.2270
65	0.0590

PV injected active powers PV_n are determined from nominal active power P_n as follows:

$$PV_n = \alpha P_n x_n, \quad (21)$$

where α here denotes the PV *penetration level*, i.e., the ratio between the installed PV power and the nominal active one, while x_n are normalized random variables defined as (10)

extracted from a representative wide measurement data set [26]. On the whole, the number of random parameters x_d is $D = 15 \times 3 = 45$.

We want to quantify the impact that PV power injection uncertainty, with different penetration α , can have on node voltages at some critical bus in the grid.

In the first experiment, in order to verify the effectiveness of the proposed method, we calculate the marginal PDF of the voltage nodes in the grid by using: (1) the proposed PWL methodology; (2) the MC method (with 10,000) used here as a reference. To this aim, the joint $PDF(\vec{x})$ of the $D = 45$ random normalized powers x_d is approximated by a GMM of the type (11). The GMM parameters (π_k , $\vec{\mu}_k$ and Σ_k) are determined with EM algorithm as seen in Section 2 (and in the Appendix A) and an adequate number of Gaussian components for the considered data set is found to be $K = 18$. Figure 6 shows the statistical distributions $PDF(V_n)$ for some node voltages module (Phase-A) along the feeder from Bus-9 to Bus-65 as computed with PWL and MC methods when a PV penetration level $\alpha = 50\%$ is assumed. Similarly, Figure 7 reports the $PDF(V_n)$ for some node voltages along the feeder from Bus-11 to Bus-27 for the same penetration level.

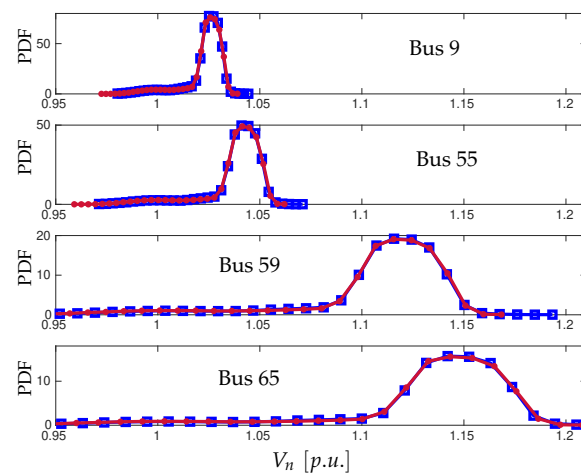


Figure 6. $PDF(V_n)$ at Buses {9, 55, 59, 65} along the feeder: (Square blue marker) computed with MC; (Dot red marker) computed with PWL method.

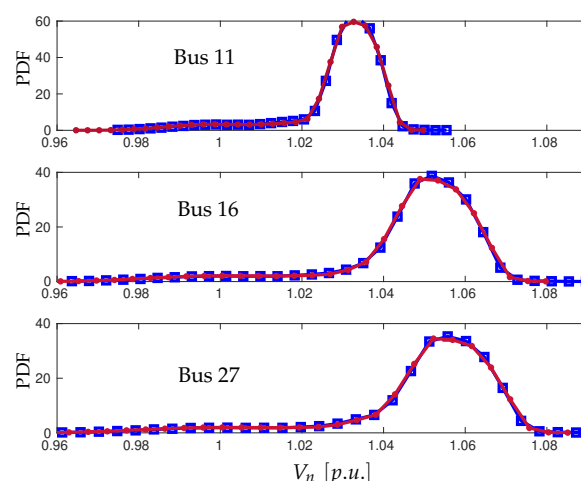


Figure 7. $PDF(V_n)$ at Buses {11, 16, 27} along the feeder: (Square blue marker) computed with MC; (Dot red marker) computed with PWL method.

Some observations are in order.

- 1 In all of the nodes of the grid, the PDFs calculated by PWL and MC reference method match with excellent accuracy, the relative difference between them being always $<0.5\%$.

- 2 Marginal PDFs represent the a priori probability distributions of node voltages and are calculated from PV uncertainty model with no further information available (e.g., with no measurement information). In this condition, from Figures 6 and 7, it is seen how voltage magnitude uncertainty interval grows moving from the internal Bus-9 towards external Bus-65 as well as from Bus-11 to Bus-27. Thus, Bus-65 and Bus-27 are the critical Buses in this grid, i.e., the ones with large voltage uncertainty.
- 3 Figure 8 shows the sensitivities of voltage magnitude at Bus-27 and Bus-65 (Phase-A) versus PV power at injection Buses (Phase-A). It is seen how voltage at Bus-27 exhibits large sensitivity with respect to injection Buses from 11 to 24 whereas voltage at Bus-65 mainly depends on power injection at Buses from 53 to 65.

It is worth observing how the sensitivity curves in Figure 8 are multi-valued functions: the sensitivity of a given voltage versus a given injection Bus number can take several values within an interval. This is due to the fact that sensitivity values reported in Figure 8 are computed at different cluster centers in the parameter space. Such space-varying sensitivity values determine the local piece-wise linear (PWL) approximations of voltage-power relationship.

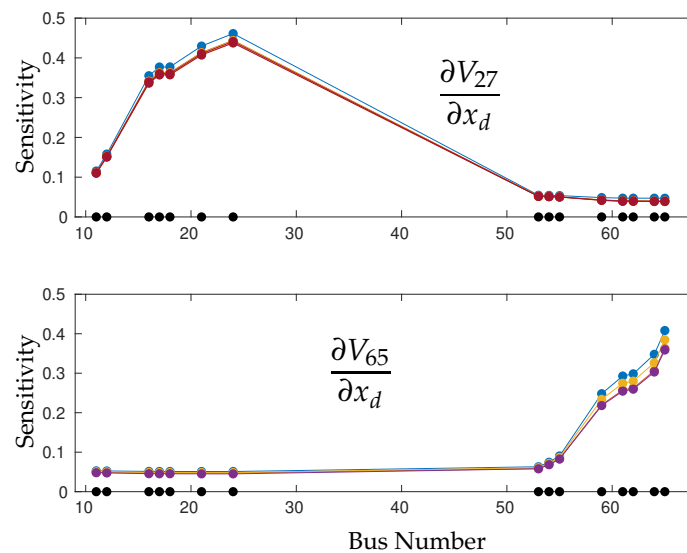


Figure 8. Sensitivities of voltage at Bus-27 and Bus-65 versus PV power at injection Buses.

- 4 The knowledge of marginal PDFs allows the evaluation of the probability that voltage magnitudes violate safe operating conditions. In fact, given $PDF(V_n)$ and the Cumulative Distribution Function (CDF), the probability that voltage V_n violates an upper threshold V can be calculated as:

$$p(V_n \geq V) = \int_V^{\infty} PDF(v_n) dv_n = 1 - CDF(V). \quad (22)$$

Thanks to the numerical efficiency of PWL simulation, we can rapidly explore several scenarios with different PV penetrations α and evaluate threshold violation occurrence. Figure 9 reports the probabilities $p(V_{65} \geq V)$ for growing penetration levels. For instance, if an upper threshold voltage $V = 1.10$ p.u. is fixed, voltage V_{65} violates the constraint with probability $\approx 85\%$ for PV penetration $\alpha = 50\%$. The probability of violating the upper threshold limit $V = 1.10$ p.u. is of $\approx 92\%$ for PV penetration $\alpha = 80\%$ whereas it is decisively zero for $\alpha = 40\%$. This means that, for the assumed upper threshold voltage, a PV hosting capacity corresponding to $\alpha = 40\%$ PV penetration is a safe limit.

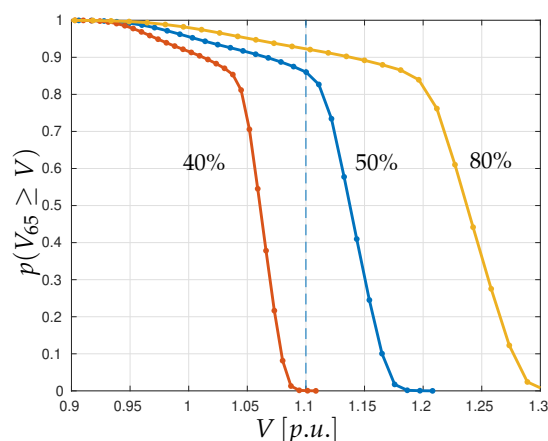


Figure 9. Probability of violating upper value V for growing penetration levels α computed with PWL method.

In the second experiment, we compare the PWL method to another state-of-the-art technique employed in probability load flow problems. The technique relies on Copula method [28] for modeling correlated PV sources combined with generalized-Polynomial-Chaos (gPC) approximation [17–19]. Figure 10 reports the $PDF(V_{65})$ at Bus-65, in the IEEE 69 bus grid with $D = 45$ PV random sources and penetration level $\alpha = 50\%$, as computed with reference MC (10,000 simulations), with PWL method and with Copula+gPC method for expansion order $\beta = 2$. It is seen how the PDF shape predicted by both Copula+gPC and PWL method fit with great accuracy that given by reference MC method.

In Table 2, we report the number of Load Flow simulations required by the three methods, their simulation times on the same Quad-core computer, and the relative error (in PDF calculation) with respect to reference MC. It is seen how, in this experiment, PWL method introduces a remarkable $200\times$ speed-up compared to MC and a $50\times$ speed-up compared to Copula+gPC method.

Table 2. Methods Comparison.

	Reference MC	PWL	Copula+gPC
LF Numb.	10,000	18	1081
Sim. time [s]	161	0.8	43
Rel. Error	—	<0.5%	<0.5%

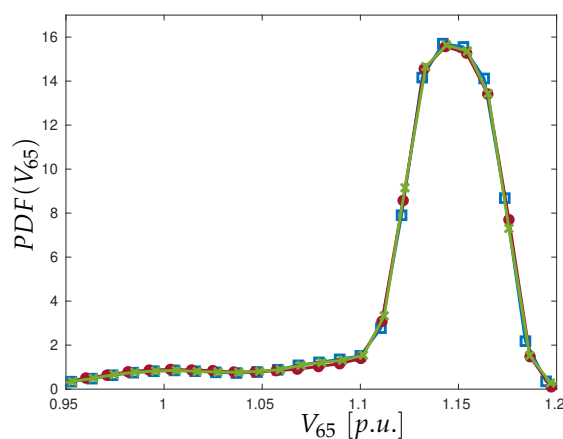


Figure 10. Detail of the $PDF(V_{65})$ at Bus-65: (Square blue marker) computed with reference MC; (Dot red marker) computed with PWL method; (Green cross marker) computed with Copula and gPC (for expansion order $\beta = 2$).

6. Discussion

In this paper an innovative approach to the probabilistic analysis of power grid with high PV penetration has been illustrated. The proposed approach is able to reproduce the uncertainty of correlated PV sources through a GMM model. The GMM method is general and can be applied to any type of data no matter if obtained with meter measurements, historic data set or simulations.

The proposed probabilistic analysis adopts a PWL approximation of the voltage-power LF relationship built around the GMM centers. PWL approximation exploits the multivariate information provided by voltage sensitivity calculations and allows the overcoming of the curse of dimensionality limitation exhibited by other surrogate models. For such reasons, it can be applied in those cases where PV injection is distributed in a large number of nodes.

In addition, it has been shown how the PWL method allows the deriving of closed-form expressions, for the joint PDF of a set of observable variables of interest by propagating the input GMM to the output. Via numerical examples, it has been shown how GMM model can be employed to identify critical buses in the grid, evaluate their a priori uncertainty intervals as well as the probability of violating safe operation conditions as a function of PV penetration. As a result, the numerically-efficient PWL analysis described in this paper is expected to become a worthwhile tool in the complex problem of estimating PV hosting capacity [29].

Author Contributions: Conceptualization, G.G. and P.M.; Data curation, G.G.; Formal analysis, L.D.; Investigation, G.G.; Methodology, P.M.; Software, P.M.; Validation, L.D.; Writing—review & editing, G.G., L.D. and P.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: <https://dkasolarcentre.com.au/> accessed on 1 January 2022.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

In this Appendix, we sketch the main computational steps of the Expectation Maximization (EM) algorithm that, for a given data set \vec{x}^i , with $i = 1, \dots, N_s$, of N_s samples and a given number K of Gaussian components (i.e., of clusters), calculates the parameters π_k , $\vec{\mu}_k$ and Σ_k in GMM (11). The algorithm relies on computing the probability that sample i belongs to cluster k as follows:

$$\gamma_{ik} = \frac{\pi_k \mathcal{N}(\vec{x}^i | \vec{\mu}_k, \Sigma_k)}{\sum_{j=1}^K \pi_j \mathcal{N}(\vec{x}^i | \vec{\mu}_j, \Sigma_j)}. \quad (\text{A1})$$

The GMM parameters are derived as:

$$\vec{\mu}_k = \frac{\sum_{i=1}^{N_s} \gamma_{ik} \vec{x}^i}{\sum_{i=1}^{N_s} \gamma_{ik}}, \quad (\text{A2})$$

$$\Sigma_k = \frac{\sum_{i=1}^{N_s} \gamma_{ik} (\bar{x}^i - \bar{\mu}_k) (\bar{x}^i - \bar{\mu}_k)^T}{\sum_{i=1}^{N_s} \gamma_{ik}} \quad (\text{A3})$$

and

$$\pi_k = \frac{1}{N_s} \sum_{i=1}^{N_s} \gamma_{ik} \quad (\text{A4})$$

Expressions (A2)–(A4) are not a closed-form solution for the parameters because the terms γ_{ik} depend themselves on parameters. However, the above equations can be evaluated in an iterative way. Starting from an initial guess for π_k , $\bar{\mu}_k$ and Σ_k , expression (A1) is used to estimate the terms γ_{ik} that, in turn, after substitution in (A2)–(A4) provide a better estimation for parameters. The algorithm has been proved to converge. More details can be found in [24], Ch. 9.

References

1. Impram, S.; Varbak Nese, S.; Oral, B. Challenges of renewable energy penetration on power system flexibility: A survey. *Energy Strategy Rev.* **2020**, *31*, 100539. [[CrossRef](#)]
2. Mortazavi, H.; Mehrjerdi, H.; Saad, M.; Lefebvre, S.; Asber, D.; Lenoir, L. A Monitoring Technique for Reversed Power Flow Detection With High PV Penetration Level. *IEEE Trans. Smart Grid* **2015**, *6*, 2221–2232. [[CrossRef](#)]
3. Chevalier, S.; Schenato, L.; Daniel, L. Accelerated Probabilistic Power Flow in Electrical Distribution Networks via Model Order Reduction and Neumann Series Expansion. *IEEE Trans. Power Syst.* **2021**, *37*, 2151–2163. [[CrossRef](#)]
4. Memon, Z.A.; Trincherio, R.; Xie, Y.; Canavero, F.G.; Stievano, I.S. An Iterative Scheme for the Power-Flow Analysis of Distribution Networks based on Decoupled Circuit Equivalents in the Phasor Domain. *Energies* **2020**, *13*, 386. [[CrossRef](#)]
5. Tang, J.; Ni, F.; Ponci, F.; Monti, A. Dimension-Adaptive Sparse Grid Interpolation for Uncertainty Quantification in Modern Power Systems: Probabilistic Power Flow. *IEEE Trans. Power Syst.* **2016**, *31*, 907–919. [[CrossRef](#)]
6. Vallée, F.; Klonari, V.; Lisiecki, T.; Durieux, O.; Moyné, F.; Lobry, J. Development of a probabilistic tool using Monte Carlo simulation and smart meters measurements for the long term analysis of low voltage distribution grids with photovoltaic generation. *Int. J. Electr. Power Energy Syst.* **2013**, *53*, 468–477. [[CrossRef](#)]
7. Ramadhani, U.H.; Shepero, M.; Munkhammar, J.; Widén, J.; Etherden, N. Review of probabilistic load flow approaches for power distribution systems with photovoltaic generation and electric vehicle charging. *Int. J. Electr. Power Energy Syst.* **2020**, *120*, 106003. [[CrossRef](#)]
8. Lai, C.S.; Jia, Y.; McCulloch, M.D.; Xu, Z. Daily clearness index profiles cluster analysis for photovoltaic system. *IEEE Trans. Ind. Informatics* **2017**, *13*, 2322–2332. [[CrossRef](#)]
9. Li, F.; Lin, Y.; Guo, J.; Wang, Y.; Mao, L.; Cui, Y.; Bai, Y. Novel models to estimate hourly diffuse radiation fraction for global radiation based on weather type classification. *Renew. Energy* **2020**, *157*, 1222–1232. [[CrossRef](#)]
10. Gruosso, G.; Maffezzoni, P. Data-driven uncertainty analysis of distribution networks including photovoltaic generation. *Int. J. Electr. Power Energy Syst.* **2020**, *121*, 106043. [[CrossRef](#)]
11. Fan, M.; Vittal, V.; Heydt, G.T.; Ayyanar, R. Probabilistic power flow studies for transmission systems with photovoltaic generation using cumulants. *IEEE Trans. Power Syst.* **2012**, *27*, 2251–2261. [[CrossRef](#)]
12. Widén, J.; Shepero, M.; Munkhammar, J. Probabilistic load flow for power grids with high PV penetrations using copula-based modeling of spatially correlated solar irradiance. *IEEE J. Photovoltaics* **2017**, *7*, 1740–1745. [[CrossRef](#)]
13. Palahalli, H.; Maffezzoni, P.; Gruosso, G. Gaussian Copula Methodology to Model Photovoltaic Generation Uncertainty Correlation in Power Distribution Networks. *Energies* **2021**, *14*, 2349. [[CrossRef](#)]
14. Wang, Z.; Kang, J.; Cheng, L.; Pei, Z.; Dong, C.; Liang, Z. Mixed Gaussian Models for Modeling Fluctuation Process Characteristics of Photovoltaic Outputs. *Front. Energy Res.* **2019**, *7*, 76. [[CrossRef](#)]
15. Ke, D.; Chung, C.; Sun, Y. A novel probabilistic optimal power flow model with uncertain wind power generation described by customized Gaussian mixture model. *IEEE Trans. Sustain. Energy* **2015**, *7*, 200–212. [[CrossRef](#)]
16. Alspach, D.; Sorenson, H. Nonlinear Bayesian estimation using Gaussian sum approximations. *IEEE Trans. Autom. Control* **1972**, *17*, 439–448. [[CrossRef](#)]
17. Xiu, D.; Karniadakis, G.E. The Wiener–Askey polynomial chaos for stochastic differential equations. *SIAM J. Sci. Comput.* **2002**, *24*, 619–644. [[CrossRef](#)]
18. Ren, Z.; Li, W.; Billinton, R.; Yan, W. Probabilistic power flow analysis based on the stochastic response surface method. *IEEE Trans. Power Syst.* **2015**, *31*, 2307–2315. [[CrossRef](#)]
19. Gruosso, G.; Netto, R.S.; Daniel, L.; Maffezzoni, P. Joined Probabilistic Load Flow and Sensitivity Analysis of Distribution Networks Based on Polynomial Chaos Method. *IEEE Trans. Power Syst.* **2020**, *35*, 618–627. [[CrossRef](#)]

20. Munikoti, S.; Natarajan, B.; Jhala, K.; Lai, K. Probabilistic Voltage Sensitivity Analysis to Quantify Impact of High PV Penetration on Unbalanced Distribution System. *IEEE Trans. Power Syst.* **2021**, *36*, 3080–3092. [[CrossRef](#)]
21. Fan, M.; Vittal, V.; Heydt, G.T.; Ayyanar, R. Probabilistic Power Flow Analysis With Generation Dispatch Including Photovoltaic Resources. *IEEE Trans. Power Syst.* **2013**, *28*, 1797–1805. [[CrossRef](#)]
22. Powell, L. *Power System Load Flow Analysis*; Electrical Engineering; McGraw-Hill: New York, NY, USA, 2004.
23. Maffezzoni, P.; Gruosso, G. Complex-Array-Operation Newton Solver for Power Grids Simulations. *IEEE Access* **2020**, *8*, 47984–47992. [[CrossRef](#)]
24. Bishop, C.M. *Pattern Recognition and Machine Learning*, 1st ed.; Springer: Berlin/Heidelberg, Germany, 2006.
25. Papoulis, A.; Pillai, U. *Probability, Random Variables and Stochastic Processes*, 4th ed.; McGraw-Hill: New York, NY, USA, 2001.
26. DKA Solar Centre. Alice Springs Project. Available online: <https://dkasolarcentre.com.au/> (accessed on 1 February 2022).
27. Das, D. Optimal placement of capacitors in radial distribution system using a Fuzzy-GA method. *Int. J. Electr. Power Energy Syst.* **2008**, *30*, 361–367. [[CrossRef](#)]
28. Munkhammar, J.; Widén, J.; Hinkelman, L.M. A copula method for simulating correlated instantaneous solar irradiance in spatial networks. *Sol. Energy* **2017**, *143*, 10–21. [[CrossRef](#)]
29. Abad, M.S.S.; Ma, J.; Zhang, D.; Ahmadyar, A.S.; Marzooghi, H. Probabilistic Assessment of Hosting Capacity in Radial Distribution Systems. *IEEE Trans. Sustain. Energy* **2018**, *9*, 1935–1947. [[CrossRef](#)]