POLITECNICO

MILANO 1863

DEPARTMENT OF ENERGY
DOCTORAL PROGRAMME IN ELECTRICAL ENGINEERING

Data-driven Quantification of PV Generation and EV Demand Uncertainties to Study their Impact on the Power System Network

Doctoral Dissertation of:
Harshavardhan Palahalli Mallikarjun

Advisor:
Prof. Giambattista Gruosso
Co-advisor:
Prof. Paolo Maffezzoni
Tutor:
Prof. Luigi Piegari
Chair of the Ph.D. Programme:
Prof. Marco Mussetta

Year 2023 - XXXV Cycle

To my mother

Everything I am, you helped me to be....!

Acknowledgements

First and foremost, I like to express my gratitude to my advisor Prof. Giambattista Gruosso, who has been an inspiration, introduced me to the field of research, and guided me all along the Mater thesis period. I will always be thankful for the opportunities he has provided and his expert insights in the execution of my doctoral research. I thank my co-advisor, Prof. Paolo Maffezzoni, who has always intrigued me in solving research problems with his ideas and made my work exciting. I thank Prof. Enrico Ragaini for always supporting me when I'm in need with the opportunities, guidance, and state-of-the-art equipment. I thank my colleagues, Prof. Cesar Eduardo Diaz Londono and Marzieh Hemmati, for collaborating on the research projects and coauthoring many articles. I express my deepest gratitude to Prof. Sbattella Licia and her team for providing all the support when I was in need.

I thank the entire community of Politecnico di Milano for allowing me to study in their esteemed institution and for providing a state-of-the-art research facility. Finally, I thank my parents for their support from the long distance to whom I owe all my achievements.

Abstract

The electric power system network is going through a paradigm shift in power generation and load demand utilisation, with various players acting in power management to boost the economy of the energy market. To incorporate these changes, analysis of the power flow in the network is necessary, predominantly to accommodate the immediate needs of integrating Photovoltaic generating units and Electric vehicle charging infrastructures. The purpose of this thesis is to accurately model photovoltaic generating units and the Electric vehicle charging station uncertainties and to solve critical problems encountered in using these models in probabilistic load flow analysis of the power system network.

The data-driven modeling approach is adopted in this work to reduce the assumptions in modeling and in capturing uncertainties accurately, but it poses many challenges in using them. The key issue in using this approach is the availability of data with large sample sizes, as probabilistic load flow techniques such as Monte Carlo simulation demand a large sample set of data. The other important factor is handling non-elementary statistical distributions obtained from the modeled systems in the load flow analysis.

In this work, modeling the PV systems for use in the probabilistic load flow is facilitated by repopulating the samples using Nataf's transformation technique, and later the correlation among the PV generators is considered and included among the modeled systems using the Gaussian mixture model and Gaussian copula methodology. To model the Electric vehicle demand uncertainty in the network, measurement data obtained from the smart meters present in the vehicle charging stations are used. The vital charging event data are statistically analyzed, and then electric vehicle load profiles are created, including analyzed uncertainties to study the performance of the grid in the presence of such uncertain electric vehicle user behavior in the network.

The stochastic response surface method is widely used in including the correlation among the input variables and accelerating the simulation of probabilistic load flow, but it fails to approximate the results accurately due to high non-linearity involved in the step of copula transformation in approximating the physical input variables and the inner variables of the model. A novel method of probabilistic load flow is proposed in this work by adopting

the inclusion of correlation among the input variables and surrogate modeling of the stochastic response in two separate steps. The repopulation of samples required for the load flow is realized with the utilization of the Gaussian Copula and a surrogate model using polynomial chaos expansion is used to precisely approximate the slightly nonlinear input and output relationship of the probabilistic load flow process.

The proposed modeling approaches and the novel probabilistic load flow methodology is implemented using different standard test networks such as IEEE 13 bus test feeder network, IEEE 69 bus medium voltage test feeder network, IEEE European low voltage test feeder network, and Non-synthetic European low voltage test network. The obtained results are compared against the standard results of Monte Carlo simulation.

The results show the inclusion of correlation among the input random variables of the probabilistic load flow will positively affect improving the accuracy of the simulation. Gaussian copula can handle the non-elementary distributions very well and repopulates the samples by preserving the distribution shape and the correlation among the variables with an error rate $< 1\%$. The proposed novel load flow is delivering a remarkable 50x speed up in the simulation when compared to the Monte Carlo simulation in reaching the same accuracy. Using the tools and techniques developed in this work, an analysis is presented showing the effects of uncertain Photovoltaic generation and Electric vehicle user demand on the low voltage distribution network in three different time windows. The proposed modeling techniques and the novel probabilistic load flow method are general that can be applied to any dataset and any network of interest.

Keywords: Correlated PV generation, Electric vehicles, Gaussian copula, Gaussian mixture model, Polynomial chaos approximation, Probabilistic load flow, Stochastic response surface method.

Abstract in lingua italiana

La rete del sistema elettrico sta attraversando un cambio di paradigma nella generazione di energia e nell'utilizzo della domanda di carico, con vari attori che agiscono nella gestione dell'energia per rilanciare l'economia del mercato energetico. Per recepire questi cambiamenti, è necessaria un'analisi del flusso di energia nella rete, principalmente per soddisfare le esigenze immediate di integrazione delle unità di generazione fotovoltaica e delle infrastrutture di ricarica dei veicoli elettrici. Lo scopo di questa tesi è quello di modellare accuratamente le unità di generazione fotovoltaica e le incertezze delle stazioni di ricarica dei veicoli elettrici e di risolvere i problemi critici incontrati nell'utilizzo di questi modelli nell'analisi probabilistica del flusso di carico della rete del sistema elettrico.

L'approccio di modellazione basata sui dati viene adottato in questo lavoro per ridurre le ipotesi nella modellazione e nel catturare accuratamente le incertezze, ma pone molte sfide nell'utilizzarle. La questione chiave nell'utilizzo di questo approccio è la disponibilità di dati con campioni di grandi dimensioni, poiché le tecniche di flusso di carico probabilistico come la simulazione Monte Carlo richiedono un ampio set di campioni di dati. L'altro fattore importante è la gestione delle distribuzioni statistiche non elementari ottenute dai sistemi modellati nell'analisi del flusso di carico.

In questo lavoro, la modellazione dei sistemi fotovoltaici da utilizzare nel flusso di carico probabilistico è facilitata dal ripopolamento dei campioni utilizzando la tecnica di trasformazione di Nataf, e successivamente viene considerata la correlazione tra i generatori fotovoltaici e inclusa tra i sistemi modellati utilizzando il modello della miscela gaussiana e la copula gaussiana metodologia. Per modellare l'incertezza della domanda dei veicoli elettrici nella rete, vengono utilizzati i dati di misurazione ottenuti dai contatori intelligenti presenti nelle stazioni di ricarica dei veicoli. I dati sugli eventi di carica vitale vengono analizzati statisticamente, quindi vengono creati profili di carico dei veicoli elettrici, comprese le incertezze analizzate per studiare le prestazioni della rete in presenza di tale comportamento incerto degli utenti dei veicoli elettrici nella rete.

Il metodo della superficie di risposta stocastica è ampiamente utilizzato per includere la correlazione tra le variabili di input e accelerare la simulazione del flusso di carico

probabilistico, ma non riesce ad approssimare accuratamente i risultati a causa dell'elevata non linearità coinvolta nella fase di trasformazione della copula nell'approssimare la fisica variabili di input e le variabili interne del modello. In questo lavoro viene proposto un nuovo metodo di flusso di carico probabilistico adottando l'inclusione della correlazione tra le variabili di input e la modellazione surrogata della risposta stocastica in due fasi separate. Il ripopolamento dei campioni necessari per il flusso di carico viene realizzato con l'utilizzo della copula gaussiana e un modello surrogato che utilizza l'espansione del caos polinomiale viene utilizzato per approssimare con precisione la relazione di input e output leggermente non lineare del processo di flusso di carico probabilistico.

Gli approcci di modellazione proposti e la nuova metodologia probabilistica del flusso di carico vengono implementati utilizzando diverse reti di test standard come la rete bus test feeder IEEE 13, la rete bus test feeder media tensione IEEE 69, la rete europea test feeder bassa tensione IEEE e la rete europea non sintetica a bassa tensione rete di prova di tensione. I risultati ottenuti vengono confrontati con i risultati standard della simulazione Monte Carlo.

I risultati mostrano che l'inclusione della correlazione tra le variabili casuali di input del flusso di carico probabilistico influenzerà positivamente il miglioramento dell'accuratezza della simulazione. La copula gaussiana può gestire molto bene le distribuzioni non elementari e ripopolare i campioni preservando la forma della distribuzione e la correlazione tra le variabili con un tasso di errore $< 1\%$. Il nuovo flusso di carico proposto offre una notevole velocità 50x nella simulazione rispetto alla simulazione Monte Carlo nel raggiungere la stessa precisione. Utilizzando gli strumenti e le tecniche sviluppati in questo lavoro, viene presentata un'analisi che mostra gli effetti della generazione fotovoltaico incerta e della domanda degli utenti di veicoli elettrici sulla rete di distribuzione a bassa tensione in tre diverse finestre temporali. Le tecniche di modellazione proposte e il nuovo metodo probabilistico del flusso di carico sono generali che possono essere applicati a qualsiasi set di dati e qualsiasi rete di interesse.

Parole chiave: Approssimazione del caos polinomiale, Copula gaussiana, Flusso di carico probabilistico, fotovoltaica correlata, Generazione Veicoli elettrici, Metodo della superficie di risposta stocastica, Modello di miscela gaussiana.

Contents

List of Figures

List of Tables

1 | Introduction

In this chapter, a brief overview of the challenges in power system analysis and simulation is presented, with the objectives and the original contributions of the research. The vision is to build new computation algorithms and software tools that are helpful in power system analysis using data-driven models of power system elements. The measurement data obtained from smart meters are used in the uncertainty quantification of photovoltaics (PV) and Electric vehicles (EV) to study their impact on the power system network using the novel methodologies and the software tools developed in this work. The organization of this thesis is given in Section 1.6 of this chapter.

1.1. Need for Power System Simulations

Energy security is significant to governing bodies, businesses, and stakeholders as energy plays a crucial factor in the progress of humankind. In recent years, renewable energy lies in the main focus of energy security [1], this change is not just because of market imbalances, resource availability, and political tension in obtaining fossil fuels but also to meet carbon-free goals in building a sustainable and healthy ecosystem [2]. When it comes to renewable energy, electric grid connected solutions such as PV are lucrative [3]. They can be integrated with the existing electric power system infrastructure to exploit and deliver the energy irrespective of geographical restrictions while participating in the demand-driven electricity market [4] for higher profits.

The electrical power system is a complex connection of various electrical components installed to generate, transmit and distribute electrical power. Modern power system network is transnational [5], proving as one of the largest infrastructures on the planet, allowing the exchange of power between countries, and has a vision of becoming transcontinental [6]. Apart from being a complex system, it is always in service, and interruption is not tolerable. Before integrating the renewable energy resources or making any changes to the power system network, testing and analyzing the proposed changes are necessary.

The power system also requires consistent innovation in the field to cope with changing technology, and a good example is the transition of electric grids into smart grids. Simulation helps in this aspect to plan, analyze, optimize and verify such changes incorporated in the power network [7].

In the conventional power system network, power flow is unidirectional, with generation at the top and consumption at the bottom. This paradigm is shifting with distributed generation and exploiting the locally available energy resources making the power flow bidirectional. A few of the hot topics that are under study in support of the modern, sustainable and futuristic power network [8] are high penetration of PV and distributed generation in the power network [9], integration of electric vehicles in the power network [10], impact of PV and EV integration on the power network [11], energy management strategies for optimal power flow in presence of distributed generation [12, 13], protection scheme for power distribution system with distributed energy resources [14, 15] and flexible and secure communications for power systems [16]. All of these studies are conducted by using the simulation techniques.

The power system simulation involves in mathematical modeling of electrical systems, such as generators, transformers, transmission lines, loads, and many other components used in the network [17, 18]. Later, the characteristic equations of those modeled systems are solved in the computers to study their combined behavior in the power network for given input data. The power system simulations are typically applied in long-term generation and expansion of the transmission systems in the planning phase [19] and short-term operational and market analysis [20].

Load flow analysis is the most common type of simulation performed by the power system community, the analysis results in a detailed picture of the power flow from the generators to the loads in the network neglecting the transients, it provides the complex voltages and currents [21, 22] at every node with which the powers and the losses are determined. Load flow analysis is performed in phasor domain using the root mean square (RMS) values, that captures the *Macro* phenomena in the power system network that is of the order in tens of seconds. The Electromagnetic transient (EMT) simulations are a type in which the modeled characteristic equations are time-dependent [23], and the instantaneous value of currents and voltages are determined by solving time-domain circuit equations [24, 25]. EMT simulation can capture the *Micro* phenomena such as harmonics, nonlinearities, and fast transients. It intrigues the power system community to conduct these analyses and observe the power network [26].

1.2.   Challenges in Power System Analysis

The motivation behind this work is to identify the challenges in power system analysis and to provide solutions to as many as possible. The hurdle faced in the power system simulations is the complexity of the power network itself. It is composed of various elements that exhibit nonlinearity and is modeled using complex mathematical equations, which requires high computational capability, resulting in a slower simulation process. Newton Raphson's (NR) [21] method is used in the load flow problem, it is robust in solving nonlinear equations, but its convergence to the solution is not always guaranteed [27], and sometimes it suffers from the divide by zero problems. When the load flow is extended to perform the stochastic analysis of the power system, the challenge is solving the network for a large number of input samples [28] that needs more time to deliver the results and is also tedious to handle the data generated in the run-time of the simulation. This work focus in detail on challenges faced in load flow analysis related to PV, EV, and extendable to other distributed energy resource integration with the power system network.

1.2.1.   Challenges in Probabilistic load flow simulation

In probabilistic load flow simulation (PLF), uncertainty quantification is challenging. It requires precise modeling of the system under study, and many assumptions are made to approximate the behavior of the systems, which might deliver erroneous results. In modeling PV uncertainty, many models are suggested by the research community, starting from solar irradiation and considering the diffusion phenomena, clear sky index, and cell temperature based modeling [29, 30], these are good models but fail to deliver an accurate picture as they ignore the effects of device efficiency, partial shadowing, device failure, and topological uncertainties. Modeling using a measurement data is a good solution but it comes with other challenges such as availability of the historic data for desired period and they tend to follow non-standard statistical distributions [31] that are difficult to handle. After overcoming the hurdles of modeling, running the load flow for large number of samples using Monte Carlo (MC) simulation technique might take more time that some applications can not afford, to overcome this problem there are approximation techniques [32, 33] available but to obtain the results with high accuracy and ability to interface the desired models having non standard statistical distributions using approximation methods is challenging and the challenges of PLF analysis are listed below.

1. Quantification of PV uncertainties using data-driven models is challenging as the power injections from the PV sources results in non-elementary distributions.

2. The measurement samples in the data set to model the PV characteristics are not enough in reaching the convergence of the PLF analysis using MC simulation.

3. The techniques used in the repopulation of the samples in data-driven uncertainty quantification will directly impact the PLF results. As it is necessary to consider the possibility of correlation among the input variables.

4. Once the above-mentioned challenges are vanquished, the PLF using MC simulation suffers from simulation time requirement and memory.

5. Among many methods to accelerate MC simulation, it is necessary to implement PLF that can be used with non-elementary distributed data-driven input models.

6. Generic Polynomial chaos (gPC) is a good alternative in accelerating the PLF, but the challenge in implementing input variables having non-elementary distributions, and using lower order gPC expansions for accurate results to avoid a bad tail end in the resulting distributions are still an open challenge in the field of PLF analysis using gPC.

7. EV uncertainty quantification to analyze the impact on the power distribution network is less explored in the research community, and the challenge is involved in data-driven modeling and the implementation of test cases for scenarios found in the real world.

1.3.   The Research Objectives

The objective of this research is to develop novel techniques and methodologies to integrate the data-driven modeling approach into the stochastic analysis of power systems. It is realized by quantifying the PV and EV uncertainty seen in the power system network and accelerating the process of load flow simulation by using approximation techniques. The specific goals that are met in the research activities are as follows,

1. Data-driven modeling of PV systems to quantify PV uncertainties in the power system network.

2. Stochastic impact analysis of PV using standard test feeder networks.

3. Modeling PV systems with consideration of correlation among the PV generators, using Gaussian Mixture Model and Gaussian Copula.

4. Developing a novel methodology to accelerate probabilistic load flow which is capable of using non-standard distributions in load flow approximation using the Stochastic

Response Surface Method (SRSM).

5. Data-driven modeling of EV charging station (EVCS) behavior to study their impact on the medium voltage (MV) and the low voltage (LV) test network.

1.4. The Research Contributions

Apart from the mathematical formulation and test setup implementation, the critical contribution of this work to the field of Power Systems is listed below.

1. Data-driven modeling of PV might result in non-standard statistical distribution with a lower sample size. Re-populating these samples for use in Monte Carlo simulations using Nataf's transformation that matches the original distribution shape is discussed here.

2. PV systems are spatially correlated, including this correlation in modeling PV characteristics using Gaussian copula, and the Gaussian Mixture model (GMM) is implemented.

3. Impact of the correlated PV generators on the standard test systems are analyzed, which has the importance of practical application of the tools developed.

4. With the use of non-standard statistical distributions, conventional SRSM introduces inaccuracies in approximation. In this work, a novel method is proposed in combination with Gaussian copula. A surrogate model of SRSM is built with generalized polynomial chaos (gPC) to accurately approximate the nonlinear relationship between PV power injection and the voltage outputs of the test network.

5. A novel methodology is introduced to model EVCS behavior using historic data, starting from EV charging events and then moving to EVCS aggregation at the MV node.

6. Stochastic impact analysis of PV generation uncertainty and the EV load demand on the LV distribution network using data-driven modeling approach.

1.5. Bibliographic Disclaimer

This work is the result of my research carried out during my Philosophiae Doctor (Ph.D.) studies, which I have presented at many conferences and published in Journals. The publications relative to this work are summarized below in chronological order.

1. Palahalli, H., Maffezzoni, P. and Gruosso, G., 2020, September. Modeling pho-

tovoltaic generation uncertainties for monte carlo method based probabilistic load flow analysis of distribution network. In 2020 55th International Universities Power Engineering Conference (UPEC) (pp. 1-6). IEEE. [34]

2. Palahalli, H., Maffezzoni, P. and Gruosso, G., 2021. Gaussian copula methodology to model photovoltaic generation uncertainty correlation in power distribution networks. Energies, 14(9), p.2349. [35]

3. Palahalli, H., Maffezzoni, P. and Gruosso, G., 2021, March. Statistical simulation of Electric Vehicle behaviour applied to low voltage distribution network. In 2021 22nd IEEE International Conference on Industrial Technology (ICIT) (Vol. 1, pp. 657-662). IEEE. [36]

4. Palahalli, H., Maffezzoni, P., Arboleya, P., and Gruosso, G., (2022). Implementing Stochastic Response Surface Method and Copula in the Presence of Data-Driven PV Source Models. IEEE Transactions on Sustainable Energy.[37]

5. Palahalli, H., Maffezzoni, P., Daniel, L., and Gruosso, G., (2022). Statistical Analysis of PV penetration impact on Residential Distribution Grids. Sustainable Energy, Grids and Networks, p.100949.[38]

6. Palahalli, H., Diaz, C., Maffezzoni, P. and Gruosso, G., (Accepted). Impact Analysis of Electric Vehicle Charging Stations on the Medium Voltage Distribution Network. In IECON 2022–48th Annual Conference of the IEEE Industrial Electronics Society. IEEE.

1.6.   Dissertation Overview

The organizational overview of the research work presented in this dissertation is given in this section. Introduction to the work carried out in the Ph.D. Studies and the research objectives are presented in Chapter 1.

Chapter 2 describes the background knowledge and state-of-the-art research in the field of PLF. In particular, it introduces the problems that need to be addressed in the stochastic analysis of power systems with PV and EV uncertainty models. Also, the detailed formulation of the complex NR method for use in load flow and the approximation technique of PLF using polynomial chaos are explained.

Chapter 3 details the test networks used in this work to study the impact of PV and EV models and to validate the proposed simulation hypothesis. The mathematical formulation of including uncertainty in load flow of three phase unbalanced systems for stochastic

analysis is given here.

Data-driven Modeling of PV uncertainties is discussed in Chapter 4, primarily using Nataf's transformation, drawback of using Nataf's transformation is noted and updated with GMM and Gaussian copula techniques to preserve the correlation.

Chapter 5 is dedicated to novel SRSM to perform load flow studies with penetration of PV uncertainties, and the detailed mathematical implementation of the proposed methodology is given with an example of implementation on the power network, and the accuracy and efficiency of this methods are compared against Monte Carlo and conventional SRSM methods.

Chapter 6 narrates the Data-driven modeling of EVCS to study the impact of EV integration into the power network at the MV level. A detailed explanation of modeling using EV charging event data of charging stations is presented. The methodology to pass from event-based time-series simulation of EV to the probabilistic simulation is well narrated.

Chapter 7 presents the detailed analysis of the cumulative impact of the PV generation and the EV load demand uncertainty on the LV test network. The analysis is performed in four hour time windows in a day, PLF simulation is on the test network is performed using MC simulation with 19 stochastic input variables. Finally, the conclusion of this work with the scope of future research is discussed in Chapter 8.

2 | Literature Review and State-of-the-Art

This chapter presents the development in the field of power system simulation, with a brief description of the background knowledge and the state-of-the-art research in probabilistic load flow and modeling PV and EV elements integrated in the power system network.

In TS simulations, the positive sequence component or the root mean square (RMS) values of the power system elements are considered, and this type of simulation neglects the harmonics and DC offset by assuming the system works at the fundamental frequency, the minimum time step achieved in this simulation is in the range of few a *ms* in general, load flow studies are classified under TS simulation. Before going into the subject matter a few basics of probability distribution are discussed in the next Section 2.1.

2.1. Probability Distributions

Distributions hold vital information about the event and help to predict the outcome of an event. In general, a distribution is a set of possible values a variable can take, and it also has information regarding the frequency of occurrence of the respective values.

Let $'Y'$ be the actual outcome called a random variable, and $'y'$ is one of the possible outcomes of an event. The cumulative distribution function (CDF) of a random variable Y is defined as probability that Y will achieve, that is less than or equal to y as given in Equation 2.1 [39].

$$F_Y(x) = P(Y \leq x), \forall\, x \in \mathbb{R} \tag{2.1}$$

Consider a measurement device such as a voltmeter having a measurement uncertainty of 0.25%, and such uncertainty follows a normal distribution. The 230V voltage is measured over 10,000 times that represents our samples. The CDF of such measurement obtained by using the considered voltmeter is given in Figure 2.1.
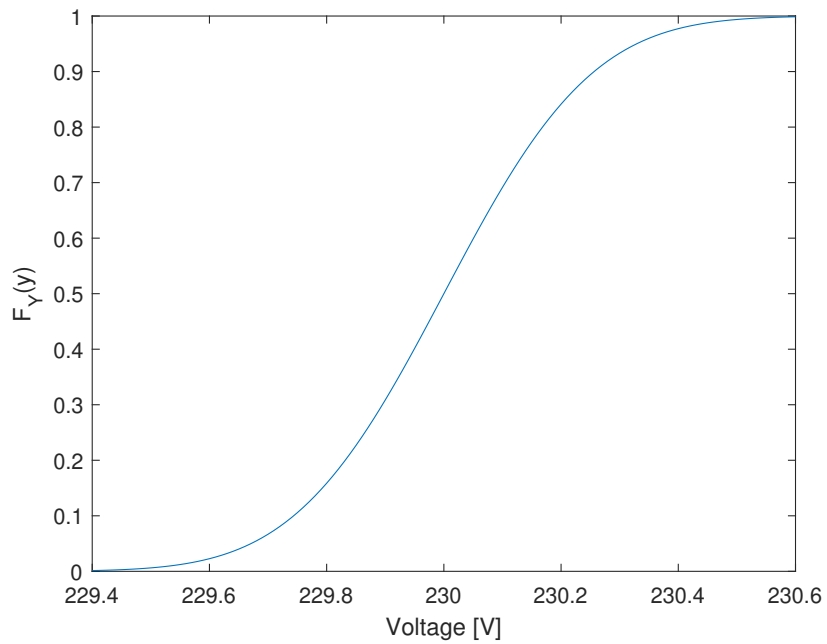
Figure 2.1: CDF of measuring 230 V supply with a voltmeter having an uncertainty of 0.25% full scale voltage.
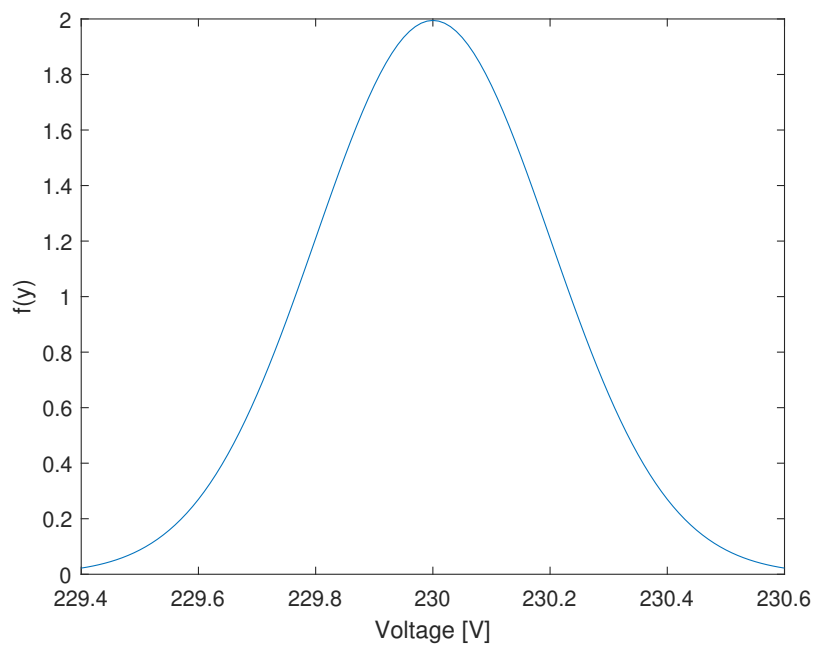
Figure 2.2: PDF of measuring 230 V supply with a voltmeter having an uncertainty of 0.25% full scale voltage.

The probability density function (PDF) expresses the probability of each distinct outcome, and it is the measure of probability per unit of sample space that shows the likely hood of an outcome depending on its occurrence in the sample space [40].

Let $f(y)$ be the PDF of a continues random variable Y, the probability density at any point y is given by Equation 2.2. It is evaluated over the limit of probability in the interval $(y, y + \Delta)$.

$$f(y) = \lim_{\Delta \to 0^+} \frac{P(y < Y \leq y + \Delta)}{\Delta} \tag{2.2}$$

We know that the probability over the interval is given by $P(y < Y \leq y + \Delta) = F_Y(y + \Delta) - F_Y(y)$ hence the Equation 2.2 is rewritten as Equation 2.3 differentiable at y where $F_Y(y)$ is the CDF.

$$\begin{cases} f(y) = \lim_{\Delta \to 0^+} \dfrac{F_Y(y + \Delta) - F_Y(y)}{\Delta} & \text{(2.3a)} \\[2ex] f(y) = \dfrac{\mathrm{d}F_Y(y)}{\mathrm{d}x} = F_Y'(y) & \text{(2.3b)} \end{cases}$$

The PDF of the voltage measurement event with a voltmeter with an uncertainty factor of 0.25% that follows normal distribution is given in Figure 2.2. The PDF holds the information of distribution such as mean, variance, standard deviation, and median. The mean value of the PDF is the expected value in an Event, it defines the axis of symmetry, and it can be considered as an average value. The Median is the mid-value in the distribution that separates the distribution into higher and the lower half of the values equally. The variance of a distribution signifies the variability of the distributed value in sample space from the mean value, it is measured as the mean of the squared distance to the mean value of the distribution, and the positive square root of the variance is the standard deviation.

Using the mean value and the standard deviation, many of the elementary distributions can be constructed, and samples can be reproduced. In many cases the PDF obtained from an event is represented using a *'boxplot'* as shown in Figure 2.3. It summarizes the statistical information to be extracted from the distribution, the central red line in the box plot indicates the median of the distribution, and the points a and b indicate 25th and 75th percentiles respectively. This type of representation is helpful when many PDFs are represented in the same figure space.
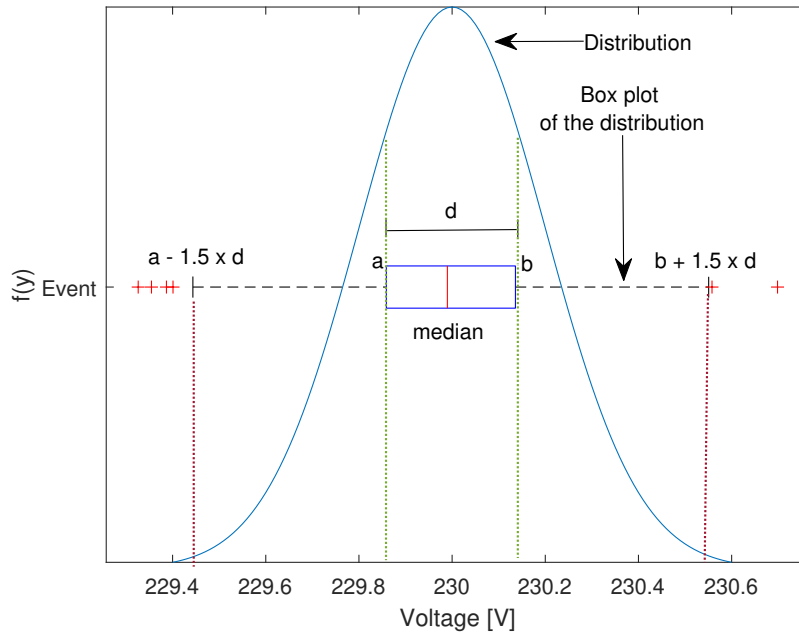
Figure 2.3: Box plot summarizing the statistical information of the PDF

When the measurement results in a non-elementary distribution, the PDF can be determined numerically. The steps to determine the PDF of a non-elementary distribution are given in Algorithm 2.1. The critical factor is the selection of the bin size, the comprehension of the data varies with the bin size hence, it should be selected appropriately.

Algorithm 2.1 Algorithm to determine the PDF numerically

1: Separate the sample space of the distribution into a finite equal number called bins.
2: Count the number of elements e and the median b in each bin.
3: Find the distance between the bins i.e., $b_d = b_2\text{-}b_1$.
4: Determine the sum e of all the bins and divide the elements by the sum to normalize.

5: The numerical PDF can be obtained by $\frac{e}{sum(e)} \times b_d$ and can be plotted verses b as the sample space.

To demonstrate, an example is considered using the measurement data of a smart meter, the active power measurement is normalized and analyzed, it represents a non-elementary distribution, and hence the Algorithm 2.1 is followed to obtain the PDF. The obtained PDF with *8 bins* and *32 bins* are shown in Figure 2.4, showing the importance of selecting appropriate bin size.
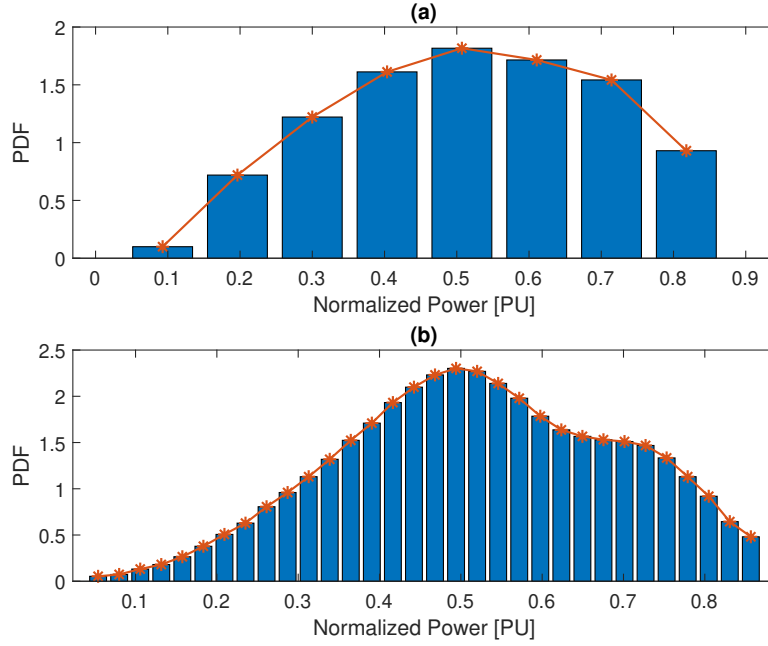
Figure 2.4: Non-elementary PDFs obtained using the normalized active power measurement data, (a) shows the PDF with resulted with 8 equal bin sizes and (b) shows the PDF obtained using a bin size of 32.

2.2. Probabilistic Load flow

Load flow or power flow is a numeric analysis to determine the primary and secondary unknowns of nodes present in the network by using the known entity of the respective nodes. Depending on the type of bus the known and unknown entity varies [41], their details are given in Table 2.1, the inputs are used in load flow to determine the primary and secondary unknowns. $\mathbf{P}$ is active power, $\mathbf{Q}$ is reactive power, $\mathbf{V}$ is voltage, and $\boldsymbol{\delta}$ is the voltage angle.

Type	Input	Primary Unknown	Secondary Unknown
PV	P, V	δ	Q
PQ	P, Q	V, δ	-
Slack	V, δ	-	P, Q

Table 2.1: Types of buses, known and unknown entities of a bus in a power system network

Load flow is employed in solving a system of nonlinear equations with iterative methods. The conventional load flow is often called deterministic load flow (DLF), as it calculates the nodal voltages and line flows at a static operating condition [42]. Consider a power system network composed of N number of buses, connected with N_l number of lines, at each node the voltage phasors are calculated using the load flow as given in Equation 2.4.

$$
\begin{cases}
F_n\left(\vec{V}\right) = S_n - V_n \sum_{i=1}^{N} Y_{ni} V_i^* = 0 & \text{(2.4a)} \\
\\
\qquad S_n = P_n + jQ_n & \text{(2.4b)}
\end{cases}
$$

where,

n	1,2,.....,N
$\vec{V}$	Vector of node voltage phasors
S_n	Complex power
Y_{ni}	Bus admittance matrix
P_n	Active Power
Q_n	Reactive Power

The Equation 2.4 is valid for any time instant t, for a time varying conditions in the network, the above mentioned quantities will be time dependent and can be written as $P_n(t)$, $Q_n(t)$ and $V_n(t)$. The Equation 2.4 can be solved using many methods, among them most used are the Gauss iteration method [43], Gauss-Seidel method [44], Newton-Raphson algorithm [21], DC power flow [45], and Fast decoupled power flow [46]. Each method has its advantages and disadvantages. Among all, the popular one used in the field is the Newton-Raphson algorithm, as it shows fast convergence as long as the initial guess is close to the solution, it exhibits a large region of convergence, and the number of iterations remains the same irrespective of the size of the system.

2.2.1.   Newton Raphson method for power flow solution

In the power flow, it is assumed that the power system network is balanced, and the formulation to derive the solution for the nonlinear Equation 2.4, later it can be extended to the three-phase unbalanced system.

The complex power in any considered bus n is given by,

$$
S_n = V_n \sum_{i=1}^{N} Y_{ni} V_i^* \tag{2.5}
$$

In rectangular coordinates, the elements in Equation 2.5 is given by,

$$
\begin{cases}
V_n = v_n + jf_n & \text{(2.6a)} \\
Y_{ni} = G_{ni} - jB_{ni} & \text{(2.6b)} \\
V_i = v_i + jf_i & \text{(2.6c)} \\
S_n = P_n + jQ_n & \text{(2.6d)}
\end{cases}
$$

Where G and B are conductance and susceptance of the line, P and Q are the active and the reactive powers of the bus n. The Equation 2.5 is now be rewritten as Equation 2.7 and by separating the real and imaginary parts, the active power P_n and the reactive power Q_n is given by Equation 2.8.

$$
P_n + jQ_n = \sum_{i=1}^{N}(v_i + jf_i)(G_{ni} - jB_{ni})(v_i - jf_i) \tag{2.7}
$$

$$
\begin{cases}
P_n = v_n \sum_{i=1}^{N}(G_{ni}v_i - B_{ni}f_i) + f_n \sum_{i=1}^{N}(G_{ni}f_i + B_{ni}v_i) & \text{(2.8a)} \\
Q_n = f_n \sum_{i=1}^{N}(G_{ni}v_i - B_{ni}f_i) + v_n \sum_{i=1}^{N}(G_{ni}f_i + B_{ni}v_i) & \text{(2.8b)}
\end{cases}
$$

In Equation 2.8 the specified electrical quantities (P_n, Q_n) of the buses is a nonlinear simultaneous equations equations of unspecified electrical quantities (v_n, f_n, v_i, f_i). To obtain the solution for such a relation Taylor's series expansion is used. In a balanced three-phase network, there are two specified electrical quantities for every bus hence two equations are required for each bus, for a N bus system, *2N* number of equations are formed, by considering one bus as a slack bus in the network, a total of *2(N-1)* equations needs to be solved.

The Taylor series expansion is given by Equation 2.9, the unknown quantities are $x_1, ...x_n$ and the specified quantities are $y_1, ...y_n$.

$$
y_1 = f_1(x_1, ...x_n)
$$
$$
.
$$
$$
. \tag{2.9}
$$
$$
y_n = f_n(x_1, ...x_n)
$$

In the series expansion the unknown terms x_n can be expressed as sum of initial guess value x_n^o and the unknown residual value Δx_n. After expansion and considering the initial and residual values, the Equation 2.9 is rewritten as Equation 2.10 and further can be simplified as Equation 2.12.

$$
\begin{aligned}
y_1 &= f_1(x_1^o, \ldots x_n^o) + \frac{\partial f_1}{\partial x_1}\big|_{x_1^o}^{\Delta x_1} + \ldots \ldots \frac{\partial f_1}{\partial x_1}\big|_{x_n^o}^{\Delta x_n} \\
&\quad . \\
&\quad . \\
y_n &= f_n(x_1^o, \ldots x_n^o) + \frac{\partial f_n}{\partial x_n}\big|_{x_1^o}^{\Delta x_1} + \ldots \ldots \frac{\partial f_n}{\partial x_n}\big|_{x_n^o}^{\Delta x_n}
\end{aligned}
\tag{2.10}
$$

$$
\begin{bmatrix}
y_1 - f_1(x_1^o, \ldots x_n^o) \\
. \\
. \\
y_n - f_n(x_1^o, \ldots x_n^o)
\end{bmatrix}
=
\begin{bmatrix}
\frac{\partial f_1}{\partial x_1} \ldots \ldots \frac{\partial f_1}{\partial x_n} \\
. \\
. \\
\frac{\partial f_n}{\partial x_1} \ldots \ldots \frac{\partial f_n}{\partial x_n}
\end{bmatrix}
\begin{bmatrix}
\Delta x_1 \\
. \\
. \\
\Delta x_n
\end{bmatrix}
\tag{2.11}
$$

$$
\begin{bmatrix}
\Delta y_1 \\
. \\
. \\
\Delta y_n
\end{bmatrix}
=
\begin{bmatrix}
\frac{\partial f_1}{\partial x_1} \ldots \ldots \frac{\partial f_1}{\partial x_n} \\
. \quad . \quad . \quad . \\
. \quad . \quad . \quad . \\
\frac{\partial f_n}{\partial x_1} \ldots \ldots \frac{\partial f_n}{\partial x_n}
\end{bmatrix}
\begin{bmatrix}
\Delta x_1 \\
. \\
. \\
\Delta x_n
\end{bmatrix}
\tag{2.12}
$$

In Equation 2.12 the known residual matrix is composed of Δy_n elements, the partial derivative matrix composed of $\frac{\partial f}{\partial x}$ elements is a Jacobian matrix, and the unknown residual matrix is composed of Δx_n elements that are obtained by using the inverse Jacobian product with the known residual matrix.

Coming back to the power system, the voltage magnitudes are not directly unknown for the case of generator buses (PV buses) it requires linearizing the Equation 2.13,

$$
V_n^2 = v_n^2 + f_n^2
\tag{2.13}
$$

The Taylor series expansion is used to linearize the Equation 2.8 and Equation 2.13 that results in the below Equation 2.14 with six sub matrices in the Jacobian matrix.

$$\begin{bmatrix} \Delta P \\ \Delta Q \\ \Delta(V^2) \end{bmatrix}_0 = \begin{bmatrix} \frac{\partial P(x)}{\partial v} & \frac{\partial P(x)}{\partial f} \\ \frac{\partial Q(x)}{\partial v} & \frac{\partial Q(x)}{\partial f} \\ \frac{\partial(V^2)}{\partial v} & \frac{\partial(V^2)}{\partial f} \end{bmatrix}_0 + \begin{bmatrix} \Delta v \\ \Delta f \end{bmatrix}_0 \tag{2.14}$$

The iteration is continued until the desired convergence is achieved using the Equation 2.15.

$$\begin{bmatrix} v \\ f \end{bmatrix}_1 = \begin{bmatrix} v \\ f \end{bmatrix}_0 + \begin{bmatrix} \Delta v \\ \Delta f \end{bmatrix}_0 \tag{2.15}$$

Once the complex bus voltages in and the power injected are known, the respective currents are determined which delivers the load flow solution. The DLF neglects [47] the uncertainties present in the network, and these uncertainties include the change in the network configuration, variation of load, and generation entities. The modern power system network is far different from the traditional ones, and it allow bidirectional power flow with the help of distributed energy resources. To include the uncertainties that arises from generation side such as PV generators and the load side such as user behaviour and the EV charging, Probabilistic load flow (PLF) is used [32, 48]. It is a coalition of DLF and stochastic methods characterizing the uncertainties present in the power system network using PDF [34, 49].

Consider a vector $\vec{\xi} = [\xi_1, \xi_2, ..., \xi_u]$ that is made of u number of stochastic parameters which includes the uncertainty. Each element ξ_u in the uncertainty vector is given by a PDF $\rho_u(\xi_u)$. Considering this uncertainty is associated to the power of the candidate nodes, the output of DLF given in Equation 2.4 will become dependent variable of time t and uncertainty vector $\vec{\xi}$, given by $V_n\left(t, \vec{\xi}\right)$ which formulates the PLF problem, the detailed formulation is explained in the next chapters.

2.2.2. Methodologies in PLF evaluation

The goal of the PLF process is to determine the probability distribution of the output variables such as voltages and currents of the power network, using the PDF of input variables. A detailed review of available methods to tackle the PLF problem is given in [30, 32, 50], the most critical factors that show the state-of-the-art in the field are depicted here. The PLF analysis computational method can be classified into three major groups as shown in Figure 2.5, the first numerical methods, the second analytical methods, and the latter are approximation techniques.
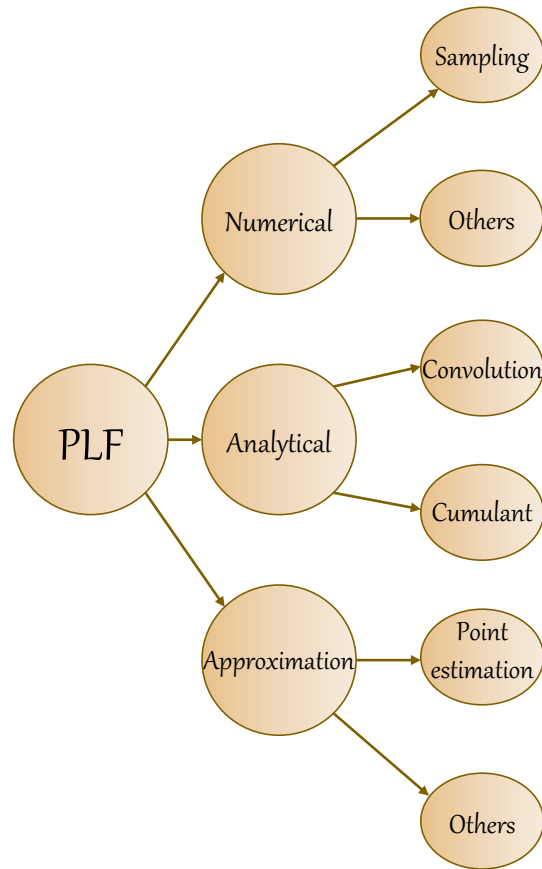
Figure 2.5: Classification of probabilistic load flow evaluation methods.

Numerical methods

In numerical methods of PLF, Monte Carlo (MC) simulations are often used, which is a sampling-based technique that repeatedly performs the DLF by using the samples of input random variables. MC simulations work without any simplification of the input random variables, and hence they can deliver highly accurate results. MC is used as the standard reference method to compare with other PLF evaluation methods.

As MC simulations run for a large number of samples, the stopping criteria could be a large number or a limit of a coefficient variation. Despite all its advantages, it is not efficient concerning the simulation time. As the sample size is high, it troubles the memory of the processor in run-time while simulating large power system networks. To increase the simulation speed of MC several methods are proposed. The first type is by using different sampling techniques such as uniform design sampling [51, 52] and Latin hypercube [53–55]. Another method is to adopt a different approach than sampling, such as quasi-random number-based MC [56, 57], using nonparametric density estimators [58, 59] to decrease

the number of samples to use in MC. This work uses simple random sampling based MC simulation, with large samples to evaluate the impact of PV and EV uncertainties on power system network and to compare the accuracy with the proposed methods.

Analytical Methods

The analytical method of PLF establishes the relationship between the input and output variables of the power flow problem by performing the arithmetic operations on the PDFs of the input variables. The analytical methods are further classified into convolution type and cumulant type. In convolution type, as the name suggests, the PLF is obtained by performing the convolution using the Laplace transform [60] or Fast Fourier transform in the case of discrete convolution in the frequency domain that promises to deliver results faster. In [13], discrete convolution technique is used to calculate the joint distribution of multiple independent variables in PLF.

The convolution process is cumbersome in a large system with multiple input and output variables, which led to a new method of PLF using cumulants [61]. The idea is to replace the moments in the distribution of input variables with their cumulants and to identify the relationship between input and output variables using the sensitivity of the system obtained by running a few DLF. Several series expansion techniques such as the Cornier Fisher [62], Gram Charlier [63], and Edgeworth method are used to determine the distribution shape of the output variables, in [64] an effort is made to compare the results using the above mentioned series expansion and they summarized that all the above discussed expansions are accurate methods as long as the distribution of the interested variable is close to the normal distribution, and the Cornier Fisher and Edgeworth method suffers from bad tail-end behavior.

Approximation Methods

In this type, the mainly used techniques are the point estimation method (PEM), the statistical information about the input PDFs are extracted, and the minimum number of samples containing this information, such as mean and variance are created to use in the power flow [65]. Depending on the type of the approximation method, the number of runs required to perform DLF is determined. For m input random variables 2m, 2m+1, 4m+1 DLF runs are required in case of 2m scheme, 3m scheme, and 4m+1 scheme [66]. It drastically reduces the time required to evaluate PLF, but the major setback is that it cannot handle non-Gaussian input random variables and suffers from accuracy degradation.

The other approximation method is Unscented transformation method (UTM) [67, 68],

in which the samples for the power flow are generated by approximating the PDFs of the input random variables to equivalent probabilistic mass functions (PMF) such that both the input and the newly generated samples have same moments. This idea is not explored much by the research community.

Hybrid methods

This is an unconventional classification of PLF methodology. To exploit the advantages of the above-mentioned methods, researchers have combined them to obtain exceptional results. In [69] a method that combines the polynomial chaos expansion and the non-parametric kernel density estimation technique is used capturing the correlation among the PV generators. Another nonparametric method is used in PLF by using saddle point approximation in [70] with first-order Taylor series expansion. This method requires m+1 DLF runs for m number of input random variables, and the obtained results are better when compared with other cumulant methods such as 2PEM and UTM. A recent study in comparison of PLF using parametric and non-parametric methods is depicted in [71] and showed that using nonparametric approaches such as kernel density estimation delivers better results than the parametric approaches such as PEM and UTM.

2.3.   Uncertainties

According to [50], uncertainties associated with the Power system are classified into two types, i.e., input uncertainties which are randomness associated with the generation and consumption of electric power, and the second one is network uncertainties. Network uncertainties are due to changes in the network configuration due to switching operations, device failures and changes in the line parameters due to environmental factors such as temperature and humidity. This work focus only on the input uncertainties associated with the power system network.

2.3.1.   Load power uncertainties

The load profile in the electrical network is highly dependent on time due to the user behavior, and it varies according to the season, which constitutes the deterministic element of the load behavior modeling. The uncertainty element is caused due to variations in the load connected, equipment failures, and changes in climatic conditions. The load profiles are aggregated at the bus level, and it varies depending on the type of the network, such as low voltage and medium voltage.

The PDF of the load powers are obtained for the concerned time window using historical data, these PDFs can be used in two ways, one is to fit the standard statistical distribution to the historical data and generate the samples for use in PLF, and the other is to use the empirical data directly. In [72–74] the load PDFs are modeled using Gaussian distribution, in [75] authors have modeled load uncertainty using Beta distribution, and they have considered 4 hours of time window due to the low resolution of the samples in the data.

2.3.2. PV power uncertainties

Many PV generation modeling techniques are available to capture the PV uncertainties and to use in PLF to investigate PV impact on the power system network. PV power generation mainly depends on solar radiation available to harvest energy on the surface of the PV panels. The amount of radiation that falls on the surface depends on many factors such as the latitude, orientation of the panels, angle of inclination, and other meteorological factors such as clouds, temperature, and the diffusion index. A few of the modeling methodologies that are widely used are explained below.

Clearness index based modeling

Clearness index is the ratio of global solar radiation on any horizontal surface H of unit length to the extraterrestrial solar radiation on a horizontal surface of same unit length H_o [76, 77], it is calculated using the Equation 2.16.

$$k_t = \frac{H}{H_o}$$
$$H_o = \frac{24 \times 3600}{\pi} G_{sc} \left(1 + 0.033 \cdot cos\frac{360n}{365} \right) \times \left(cos\phi\, cos\delta\, sinw_s + \frac{\pi w_s}{180} sin\phi\, sin\delta \right)$$

$$(2.16a)$$

$$(2.16b)$$

where,

n	1,2,.....,N
k_t	Clearness index
ϕ	Location latitude
n	Days in a year considered
δ	Daily solar declination
w_s	Solar sunset hour angle

The PV power output can be obtained using the clearness index using the Equation 2.17.

$$P_{pv} = s_a \eta \left(T k_t - T' k_t^2 \right) \tag{2.17}$$

s_a is surface area, η is the efficiency of the PV panels, P_{pv} is output PV power that represents a random variable, T and T' [78] are parameters that depend on δ, ϕ, n, and w_s and the relation obtained in the article [77].

Diffuse fraction and clearness index based modeling

In this method, the diffused irradiation fraction (cloudiness index) is considered along with the clearness index, in [76] the cloudiness is considered as the function of the clearness index as given in Equation 2.18.

$$\left(k_{df} = \frac{H_d}{H} \right) \approx f \left(k_t = \frac{H}{H_o} \right) \tag{2.18}$$

A Similar model is proposed in [79] in which the diffused irradiation factor k_{df}t is expressed as the ratio of the modeled direct beam irradiance to the measured irradiance on the surface using a Pyranometer.

$$k_{df} = \frac{I_{measured}}{I_{model}} \tag{2.19a}$$

$$I_{model} = I_o \, \varepsilon \, exp \left(-0.8662 \, T_{LK} \, m \, \delta_r(m) \, sin(\gamma_s) \right) \tag{2.19b}$$

where,

$I_{measured}$	Measured output from Pyranometer
I_o	Solar constant
ε	Solar distance correction
m	Correction of air mass
δ_r	Rayleigh optical depth
γ_s	Solar altitude angle
T_{LK}	Linke turbidity factor

Finally the out PV power is obtained using the Equation 2.17 by replacing the clearness index factor k_t by diffuse fraction clearness index factor k_{df}.

Global irradiance and cell temperature based modeling

This model is based on PV power curve, that the active power generated by PV is directly proportional to the global solar irradiance with the consideration of cell operating temperature as given in Equation 2.20. In this model, the global irradiance G and the cell operating temperature ΔT are the random variables that can be represented by Beta and Gaussian distributions respectively [80].

$$P_{pv} = s_a\, \eta\, G\, (1 - K_T\, \Delta T) \tag{2.20}$$

where,

s_a	Surface area
η	Efficiency of the PV panels
G	Global solar irradiance
K_T	Temperature coefficient
ΔT	Cell temperature

Depending on the number of cells stacked in the PV module, the output power is given as an algebraic sum of individual cell outputs.

PV power curve based modeling

In this method PV characteristics equation to deliver the active power is considered that use the irradiation as the input variable, rest of the parameters can be obtained in the manufacturer data sheet. The PV output power P_{pv} is the product of current delivered I_{pv} by the PV panel and V_{pv} the terminal voltage of the panel.

$$P_{pv} = I_{pv} \times V_{pv} \tag{2.21}$$

The cell current I_{pv} can be obtained by the Equation 2.22 [81], The input parameter λ is a random variable that follows the beta distribution in the Equation 2.23 [50].

$$I_{pv} = N_p I_{ph} - N_p I_o \left[\exp\left(\frac{\frac{V_{pv}}{N_{cell} N_s} + \frac{I_{pv} R_s}{N_p}}{n V_t} \right) - 1 \right] - I_{sh} \tag{2.22}$$

where,

I_{ph} Photo current
I_o Module saturation current
I_{sh} Current in shunt element
N_s Series modules connected in string
N_p Parallel strings in array

$$I_{ph} = [I_{sc} + K_i(T_{cell} - T_{ref})] \lambda \tag{2.23}$$

where,

I_{sc} Short-circuit current at 25^0C and $1kW/m^2$
K_i Short-circuit current temperature coefficient
T_{cell} Cell temperature in Kelvin
T_{ref} Reference temperature in Kelvin
λ Ratio of solar irradiation to global reference $1kW/m^2$

$$I_{rs} = \frac{I_{sc}}{\left[\exp\left(\frac{qV_{oc}}{N_{cell}KnT_{cell}} \right) - 1 \right]} \tag{2.24}$$

where,

I_{rs} Reverse saturation current
q Electron Charge
N_{cell} Series cells to form module
n Ideal factor of diode
k Boltzmann's constant
E_g Energy band gap of PV material

$$I_o = I_{rs} \left[\frac{T_{cell}}{T_{ref}} \right]^3 \exp\left\{ \frac{qE_g}{nK} \left(\frac{1}{T_{ref}} - \frac{1}{T_{cell}} \right) \right\} \tag{2.25}$$

$$V_t = \frac{KT_{cell}}{q} \tag{2.26}$$

$$I_{sh} = \frac{V_{pv} \frac{N_p}{N_{cell}N_s} + I_{pv}R_s}{R_{sh}} \tag{2.27}$$

Where,

V_t Diode thermal voltage

V_{oc} Open circuit voltage

R_s, R_{sh} Series and Shunt resistance

The data of active power delivery of PV are gathered by any of the above methods, then the standard statistical distributions are used to fit the data. The best fit distribution is selected to generate the samples for use in PLF studies. To model PV uncertainty in [82] log-normal and bimodal distributions are used, in [64, 78, 80, 83] beta distribution is used. Only a few articles use the empirical data obtained from the measurement, in [84] historical empirical data is directly used in the PLF by using moments and cumulants of the input PDFs. In [75] PDF of the PV active power is used to generate the samples for use in PLF without using fitting methods. In this work, historical PV measurement data is used to model the PV uncertainty, and the obtained non-standard statistical distributions are used to regenerate the samples without losing their original characteristics, which will be explained in the next chapters.

2.3.3. EV uncertainties

The uncertainties in the EV interaction with the electric power network arise from many phenomena such as the type of charger used (level-1, level-2), EV charging station infrastructure, and the human behavior that makes it more random. Modeling EV uncertainty is highly explored these days due to its immediate need in adapting the power system network and its popularity, yet it is not very well established. There are many modeling methods proposed by the research community, and a few of them are explained here.

One of the methods to model EV behavior is the Spatial-Temporal model [85, 86], which evaluates the impact of large-scale EV integration into the power network, this method considers the power system and transportation system analysis. The origin and destination of EV with charging event as a load to power network are considered in this method. In [87] the EV behaviour is modeled by also considering the initial state of charge (SOC) of the batteries in the EV, the start time of charging and the SOC of the battery behaviour are considered as a Gaussian distribution in this work whereas, uniform distribution is used to represent the same characteristics in [88].

In [89] the demand active power of EV is modeled considering various parameters such as arrival time, charging time, range for single and multiple EVs, and the active power demanded is fitted with weibull distribution in case of single EV and normal distribution in case of multiple EVs.

A data-driven modeling approach of EV charging demand is proposed in [90–93], by pro-

cessing heterogeneous EV demand using measurement data obtained from smart meters. It works for the specific selected data set with assumptions, a general method to model the EV behavior from the measurement or EV charging events that replicate the real-life scenarios is still missing.

2.3.4. Correlation in uncertainty modeling

The power system input variables might be dependent among each other, the PV systems that are geographically near are spatially correlated due to the similarity in weather patterns [29]. The load behaviour [94] are correlated due to similar human activity with respect to the time. In case of EV, the correlation might exits due to geographic location and spatial availability at the charging stations. In some cases the correlation can also exits between PV-EV [95], PV-loads [96], and loads-EV [91].

The correlation among the random variables is determined by the use of covariance. The measure of joint variability among two random variables is divided by their standard deviations to obtain the correlation coefficient among them. A higher correlation coefficient means the higher value of one variable coincides with the higher value of the other variable. For any two random variables X and Y, having the sample size n the covariance among them is given by Equation 2.28, where $\bar{x}$ and $\bar{y}$ are the mean values, σ_x and σy are the standard deviation. The correlation coefficient among them is determined by using Equation 2.29.

$$Cov(X,Y) = \sum_{i=1}^{n} \frac{(x_i - \bar{x})(y_i - \bar{y})}{n} \tag{2.28}$$

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sigma_x, \sigma y} \tag{2.29}$$

In the analytical methods of PLF, the correlation among the input variables is directly used while determining the relationship between the input and the output variables [97, 98]. Other techniques to model the uncertainty including correlation among the random variables correlation is by using Gaussian mixture model (GMM) and Copulas. In GMM multiple Gaussian distributions are used to match the sub population of the input data, i.e., with a mixture of finite number of Gaussian distributions all the data points of the inputs can be grouped [99]. Copulas are the entity or the functions that contains the dependence information about the correlated random variables [29, 100], they are used to model the dependence between two or more random variables where the marginal

distributions of the input random variables are combined to arrive at a joined distribution, the correlation coefficient ρ plays an important role while using copulas that is explained in detail in the next chapters. In [69] the correlation among wind and PV input variables is included in PLF evaluation, polynomial chaos expansion, and the stochastic response surface method is used in which the correlation inclusion is with the help of Cholesky decomposition.

2.4.   Uncertainty quantification using generalized Polynomial Chaos

Monte Carlo is a non-intrusive method that takes a sample from the PDF of desired input variable and delivers the solution by running the power flow for every sample, and it takes a large number of sample runs to converge to an accurate solution. To accelerate this process, generalized Polynomial Chaos (gPC) expansion can be used and approximate the input and output relationship by running fewer load flows for the selected number of test points.

Consider l number uncertainty elements present in the uncertainty vector $\vec{\xi}$ i.e.,

$$\vec{\xi} = [\xi_1, \xi_2, ...\xi_l]$$

The desired output of interest in the power flow solution such as a node voltage that depends on time t and the uncertainty vector $\vec{\xi}$ i.e., $V[t_m, \vec{\xi}]$ having a finite variance is approximated using a β-order truncated series.

$$V\left[t_m, \vec{\xi}\right] \approx \sum_{i=1}^{N_b} c_i(t) H_i[\vec{\xi}] \qquad (2.30)$$

The Equation 2.30 has $c_i(t)$ polynomial chaos coefficients that weighs N_b multi-variate basis function $H_i[\vec{\xi}]$, and these basis functions are the product of uni-variate orthogonal polynomials that depends on the PDF of the uncertainty vector elements ξ_r [101] as given in Equation 2.31.

$$H_i[\vec{\xi}] = \prod_{r=1}^{l} \phi_{i_r}(\xi_r) \qquad (2.31)$$

where,

$\phi_{ir}(\xi_r)$ uni-variate orthogonal polynomial

i_r degree of the uni-variate orthogonal polynomial

l Number of elements present in the uncertainty vector $\vec{\xi}$

The type of these uni-variate orthogonal polynomials depends on the distribution type of the uncertainty vector element ξ_r, as given in Table 2.2.

Si.No.	Distribution type	$\phi_{ir}(\xi_r)$
1	Gaussian	Hermite polynomials
2	Gamma	Laguerre polynomials
3	Beta	Jacobi polynomial
4	Poisson	Charlier polynomial
5	Binomial	Krawtchouk polynomial
6	Negative Binomial	Meixner polynomials
7	Hyper geometric	Hahn polynomials
8	Uniform	Legendre polynomials

Table 2.2: The type of uni-variate orthogonal polynomials depends on the input distribution type of the uncertainty vector element.

For any given number of uncertainty elements l, and the series expansion truncation order-β the sum of the degrees of the uni-variate orthogonal polynomials i_r used should be less than or equal to the series expansion truncation order-β as given in Equation 2.32.

$$\beta \geq \sum_{r=1}^{l} i_r \tag{2.32}$$

The number of basis functions to be used in the expansion depends on the truncation order-β, and the number of stochastic parameters l used in the analysis is given in Equation 2.33.

$$N_b = \frac{(\beta + l)!}{\beta! \, l!} \tag{2.33}$$

For example, two ($l{=}2$) independent Gaussian distributions parameters ξ_1 and ξ_2 are used as input stochastic parameters with the series truncation order $\beta = 3$, 10 basis functions are formed which are the product of Hermite polynomials given in Table 2.3.

Uni-variate Hermite polynomials
$\phi_0(\xi) = 1,\ \phi_1(\xi) = \xi,\ \phi_2(\xi) = \xi^2 - 1,\ \phi_3(\xi) = \xi^3 - 3\xi$
Multi-variate gPC basis functions
$\phi_0(\xi_1)\phi_0(\xi_2),\ \phi_0(\xi_1)\phi_1(\xi_2),\ \phi_0(\xi_1)\phi_2(\xi_2),\ \phi_0(\xi_1)\phi_3(\xi_2),$ $\phi_1(\xi_1)\phi_0(\xi_2),\ \phi_1(\xi_1)\phi_1(\xi_2),\ \phi_1(\xi_1)\phi_2(\xi_2),$ $\phi_2(\xi_1)\phi_0(\xi_2),\ \phi_2(\xi_1)\phi_1(\xi_2)$ $\phi_3(\xi_1)\phi_0(\xi_2)$

Table 2.3: Uni-variate Hermite polynomials and the Multi-variate gPC basis functions for two independent Gaussian distributed variable $\vec{\xi} = [\xi_1, \xi_2]$ and the truncation order $\beta = 3$.

The gPC coefficients c_i should be determined to use in Equation 2.30, it can be obtained using two methods the first one is an intrusive one called Galerkin Projection and the second one is the stochastic collocation method.

2.4.1.   Galerkin Projection to find the gPC coefficients

In this method, Equation 2.5 has to be modified for the integration of the gPC expansion. The Equation 2.30 is used in evaluation of every nodal voltage $V_n(t_m)$ resulting in Equation 2.34.

$$V_n\left[t_m, \vec{\xi}\right] \approx \sum_{i=1}^{N_b} c_i^n(t_m) H_i[\vec{\xi}] \tag{2.34}$$

It leads to finding $N_b X N$ number of gPC coefficients where N is the total number of nodes to be evaluated in the test network, which is represented by a non-linear system given by Equation 2.35.

$$\left\langle F_n\left\{\vec{V}\left[\vec{\xi}\right], H_i[\vec{\xi}]\right\}\right\rangle_\Omega = 0 \tag{2.35}$$

$\langle \cdot \rangle_\Omega$ represents the stochastic space inner product, that requires a huge computational effort to realize the solution. For a case with the order $\beta = 3$, the stochastic uncertainty parameter $l=3$ will result in *20* basis functions and let the number of buses in the network is *N=100*, it will result in 2000 nonlinear systems that requires the solution. Due to its

complexity in the implementation, this technique of finding gPC coefficients is not much used in power system applications.

2.4.2. Stochastic Collocation to find the gPC coefficients

In this method, several test points are selected to determine the gPC coefficients of the series given in Equation 2.30. This method does not demand the change of load flow equations as seen in the previous method, and it is called collocation-based Stochastic Testing (ST) [102]. According to this method, we need to find the N_b number of gPC coefficients $c_i(t)$ testing points $N_s = N_b$ are selected from the uncertainty vector $\vec{\xi}^k$ for $k = 1, ..N_s$ in the stochastic space in which the output parameter for the selected testing point is determined using the traditional load flow analysis i.e., $V_k(t) = V\left[t, \vec{\xi}^k\right]$ using Equation 2.15. With this, at every testing point the Equation 2.30 is forced to exactly fit the values of $V_k(t)$ resulting in a linear system given in Equation 2.36.

$$M\vec{c}(t) = \vec{V}(t) \tag{2.36}$$

The column vector $\vec{c}(t) = [c_1(t), ...c_{N_b}(t)]^T$ collects the N_b unknown gPC coefficients and the column vector $\vec{V}(t) = [V_1(t), ...V_{N_s}(t)]^T$ collects the N_s node voltages. The transitional square matrix $M = \{a_{k,i}\} = \left\{H_i\left[\vec{\xi}^k\right]\right\}$ of the size $N_b X N_s$ gathers the gPC basis functions to be determined at the testing points as given in Equation 2.37. The testing points and the selected basis functions decides the characteristics of the matrix M and hence it can be predetermined and can be used for any instants of time $t = t_m$.

$$M = \begin{bmatrix} H_1(\vec{\xi}^1) & \cdots & H_{N_b}(\vec{\xi}^1) \\ \vdots & \ddots & \vdots \\ H_1(\vec{\xi}^{N_s}) & \cdots & H_{N_b}(\vec{\xi}^{N_s}) \end{bmatrix} \tag{2.37}$$

The ST method of determining the gPC coefficients is very efficient, for the parameters $\beta = 3$, and the stochastic input parameter $l=3$, it needs to run only *20* runs of deterministic load flow irrespective of the size of the network.

The accuracy of the gPC interpolation scheme used in the PLF is highly dependent on the selection of the testing points in the stochastic space. Good accuracy can be achieved by considering the uni-variate polynomial $\phi(\xi_r)$ of highest order β that describes the rth uncertainty parameter ξ_r with the PDF $\sigma_r(\xi_r)$. The selection is done differently based on the type, in the uni-variate case, the testing points related to (P+1) Gauss

quadrature nodes are used in the numerical integration of ξ_r^k. In case of multivariate with l number of stochastic input parameters the testing point vector $\overrightarrow{\xi}^k = \left[\xi_1^k, \xi_2^k, ... \xi_l^k\right]$ are determined by considering a multidimensional grid of all the possible combination of uni-variate quadrature nodes.

The number of quadrature nodes in the multidimensional grid is given by $(\beta+1)^l$, which is greater than the number of basis functions N_b. For example, $l=2$ and $\beta = 3$ the number of Gauss nodes is 16, whereas the number of basis function $N_b = 10$. To solve this problem, a subset formed by $N_s = N_b$ quadrature nodes are selected as testing points. The selection of those points is made using criteria, and the first one is to select the statistically important quadrature nodes i.e., the nodes with larger weight values. The second criterion is to make the M matrix full ranked and well conditioned.

2.4.3.   gPC basis function for data-driven stochastic variables

The gPC basis functions are obtained using the uni-variate orthogonal polynomials, depending on the PDF and the distribution type of the stochastic variable. Table 2.2 details the resulting uni-variate orthogonal polynomials using different types of distribution. In data-driven modeling of PV and other elements of the power system network, the candidate of interest will result in a non-elementary distributed random variable whose PDF is known numerically. In such a case, it is mandatory to compute the valid gPC basis functions [75] instead of using the ones mentioned in Table 2.2.

Consider a random variable x having its PDF $f(x)$ which is non-elementary distributed over $x \in I \subseteq \mathbb{R}$ the PDF samples $f(x^j)$ are known only numerically in the sequence $x^j \in I$. The gPC basis functions of x are a set of polynomials $q_i(x)$ orthonormal in regards to the inner products as given in Equation 2.38.

$$\langle q_i(x), q_j(x)\rangle = \int_I q_i(x)q_j(x)f(x)dx = \delta_{ij} \tag{2.38}$$

where,

i and j   polynomial degrees

δ_{ij} Kronecker delta operator

These polynomials are obtained by using the three-term recurrence formula on a set of monic orthogonal polynomials $\pi_i(x)$. The three-term recurrence relation [103] is given in Equation 2.39.

$$
\begin{cases}
\pi_{i+1}(x) = (x - \alpha_i)\pi_i(x) - \beta_i\pi_{i-1}(x) & \text{(2.39a)} \\
\pi_{-1}(x) = 0 & \text{(2.39b)} \\
\pi_0(x) = 1 & \text{(2.39c)}
\end{cases}
$$

where,

i   0,1,2,....

α_i and β_i positive constants

In the three term recurrence relation, the real positive constants α_i and β_i depends on the distribution shape of $f(x)$, to compute these constants a iterative Darboux's formula [75, 104] is used as given in Equation 2.40.

$$
\begin{cases}
\alpha_i = \dfrac{\int_I x\pi_i^2(x)f(x)dx}{\int_I \pi_i^2(x)f(x)dx} & \text{(2.40a)} \\[4mm]
\beta_{i+1} = \dfrac{\int_I \pi_{i+1}^2(x)f(x)dx}{\int_I \pi_i^2(x)f(x)dx} & \text{(2.40b)}
\end{cases}
$$

From the Three Term Recurrence relation given in Equation 2.39 and the initialisation $\beta_0 = 1$, the first $\hat{n}$ gPC basis functions are deduced as,

$$
q_i(x) = \frac{\pi_i(x)}{\sqrt{\beta_0\beta_1....\beta_i}}, \forall \ i = 0, 1, ...\hat{n} \tag{2.41}
$$

In this implementation, the stochastic variable x is non-elementary distributed over the interval $I = [0,1]$. The samples of the PDF $f(x^j)$ are known over a sequence of N_s+1 values that are equidistant j from each other. The samples can be represented using $x^j = j\,\Delta x$ in I with $\Delta x = 1/N_s$ and $j = 0, 1, ...N_s$. With this considerations, the Darboux's formula given in Equation 2.40 is be simplified into the integral form given in Equation 2.42.

$$
\int_I p(x)f(x)dx \tag{2.42}
$$

The term $p(x)$ is the known polynomial through their coefficients, the PDF f(x) is available in the numerical form having $N_s + 1$ sample points x^j. The integral type given in Equation 2.42 is solved by using Simpson's integration formula with the number of Simpson's integration interval N_{int} as the exact half of the number of samples N_s i.e., $N_s = 2N_{int}$ and the evaluation results in the Equation 2.43.

$$\int_I p(x)f(x)dx \approx \frac{\Delta x}{3} \sum_{s=1}^{N_{int}} p(x^{2s-1}) \cdot f(x^{2s-1}) + s4 \cdot p(x^{2s}) \cdot f(x^{2s}) + p(x^{2s+1}) \cdot f(x^{2s+1})$$

(2.43)

Using the three-term recurrence relation with Simpson's integration formula delivers the orthogonal polynomials for the non-standard stochastic variables obtained from the data-driven modeling approach. The obtained orthogonal polynomials are used in the gPC expansion in approximating the input-output relationship of the PLF. The accuracy of this method is verified by applying it to the known standard distributions, in [101] a beta distributed stochastic variable is considered, and the resulting polynomial sequence represents a Jacobi polynomial proving the accuracy to use the technique for non-standard statistical distributions. The testing points are selected among the Gaussian quadrature points $x^j \in \mathbb{R}$ with $j = 1, 2, ...\hat{n} + 1$, enabling the evaluation of the integral given in Equation 2.44.

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

(2.44)

The element w^j are the Gaussian weights of the polynomial $g(x)$ of degree $\leq 2\hat{n} + 1$. The detailed implementation in determining the Gaussian weights using Darboux's formula given in Equation 2.40 is described in [75].

2.5. Concluding Remarks

PLF studies, it is well understood that the MC method is accurate and takes more time to deliver the results, hence other PLF techniques are recommended to overcome the problem. When modeling uncertainties of power system elements, the data-driven approach delivers good results as the approximations in modeling are minimal and can also capture the topology uncertainties in the network. The problem with this modeling approach is that the quantified uncertainties might lead to non-standard (non-elementary) statistical distributions that should be used as input random variables in the PLF process.

Only a few techniques, such as the approximation methods, can use these non-standard statistical distributions, but they suffer from the problem of accuracy. In [61] a case of using bi-modal correlated input random variable used in PLF with IEEE 14 bus system

is evaluated, and the proposed approximation methods such as 2PEM and 3PEM could not deliver accurate results compared to MC simulation. Polynomial chaos expansion can be used efficiently with stochastic variables having elementary statistical distributions resulting in the utilization of standard uni-variate orthogonal polynomials, but in presence of nonelementary distributed variables, gPC basis functions should be determined numerically. In [69] polynomial chaos expansion is used with Cholesky decomposition to include uncertainty in PLF. This method suffers from a bad tail end due to the use of higher-order polynomials in the case of using the nonelementary distributions obtained through a data-driven modeling approach.

These gaps in the PLF simulation studies are addressed in this work, and the use of nonelementary statistical distributions in PLF is facilitated. The case of correlated PV generators is considered to demonstrate the use of nonstandard statistical distributions in multiple PLF methodologies. The use of Gaussian copula for modeling uncertainty in PLF that can be evaluated with lower order generic polynomial chaos expansion is also presented. Among all the EV modeling approaches, data-driven modeling of EV charging behavior from the heterogeneous EV charging event delivers a detailed picture, and moving from the charging station to the aggregation is necessary to study the impact analysis of EV charging on the MV distribution network, and an effort is made to realize this objective in this work.

3 | Test Networks and the PLF Formulation

This Chapter describes the test networks used in this work to evaluate the impact of the PV and EV uncertainties and to prove the proposed uncertainty modeling and the PLF methodology.

3.1. IEEE 13 Node Test Feeder Network

IEEE 13 node test feeder network [105] is a radial distribution feeder that operates at *50 Hz, 4.16 kV* with an unbalanced distribution of loads resulting in unbalanced power distribution and different operating voltage of the phases in the lines and the buses. The single-line diagram of the test network is shown in Figure 3.1.
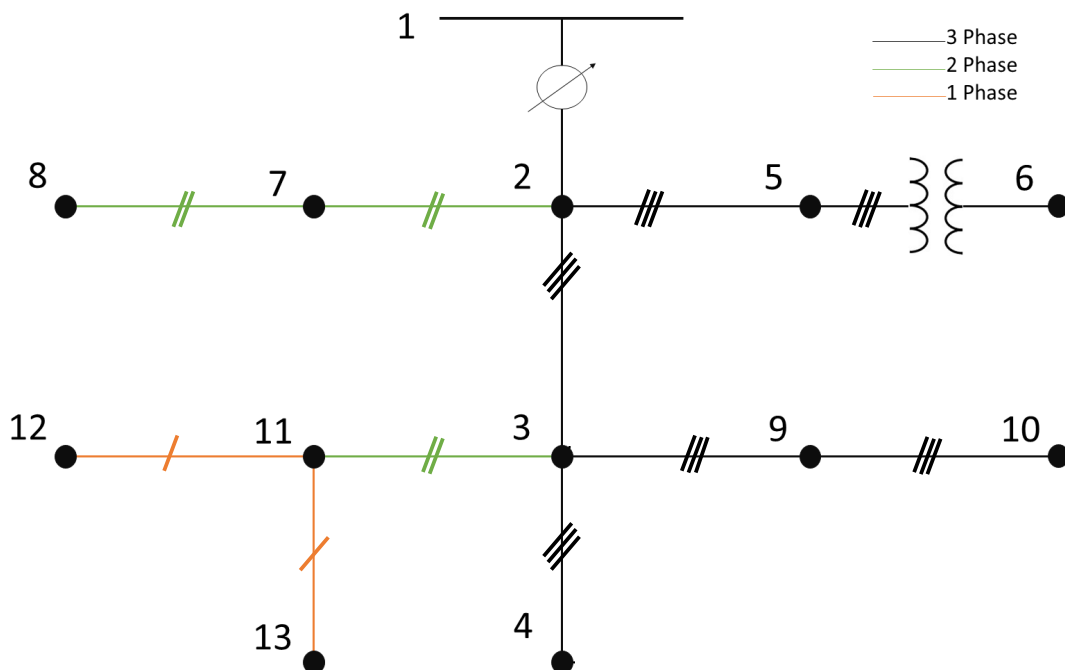
Figure 3.1: IEEE 13 node test feeder network

The test network has a total of 13 buses, it is highly loaded for a *4.16 kV* feeder, it has an inline transformer between bus 5 and 6 to study the *400 V* distribution at bus 6, a substation voltage regulator, and shunt capacitor banks of *200 kvar* and *100 kvar* connected to the bus 10 and 12 respectively. The network constitutes various phasing at different sections, and the line connected between bus 2 - bus 8, bus 3 - bus 11 is of two-phase, the line connected between the buses 11 - 12 and 11 - 13 are of single phase while the rest of the lines are of three phases. The loads connected at the buses are called spot loads whereas, in one case, the distributed loads are connected between buses 2 and 3. The total loading of the network according to the respective phases is given in the Table 3.1.

Si.No.	Parameter	Active Power (kW)	Reactive Power (kvar)
1	Phase A	1175	616
2	Phase B	1039	665
3	Phase C	1252	821
4	Total	3466	2102

Table 3.1: IEEE 13 node test feeder network load data according to the phases.

The main reason to select this network is it is very popular in the research community because of its properties that are exploited to test various test scenarios. It has fewer buses which help in a faster simulation process.

3.2.   Non-synthetic European Low Voltage Test Network

Non-synthetic European low voltage test network (NSELVTN) [106] is a distribution network test system that resembles an actual European town. It is made using real measurement data using smart meters, and it operates at *50 Hz*, three-phase *416 V* voltage rating (line to line). It is an isolated four-wire system with neutral on customer grounds. The network diagram is given in Figure 3.2.
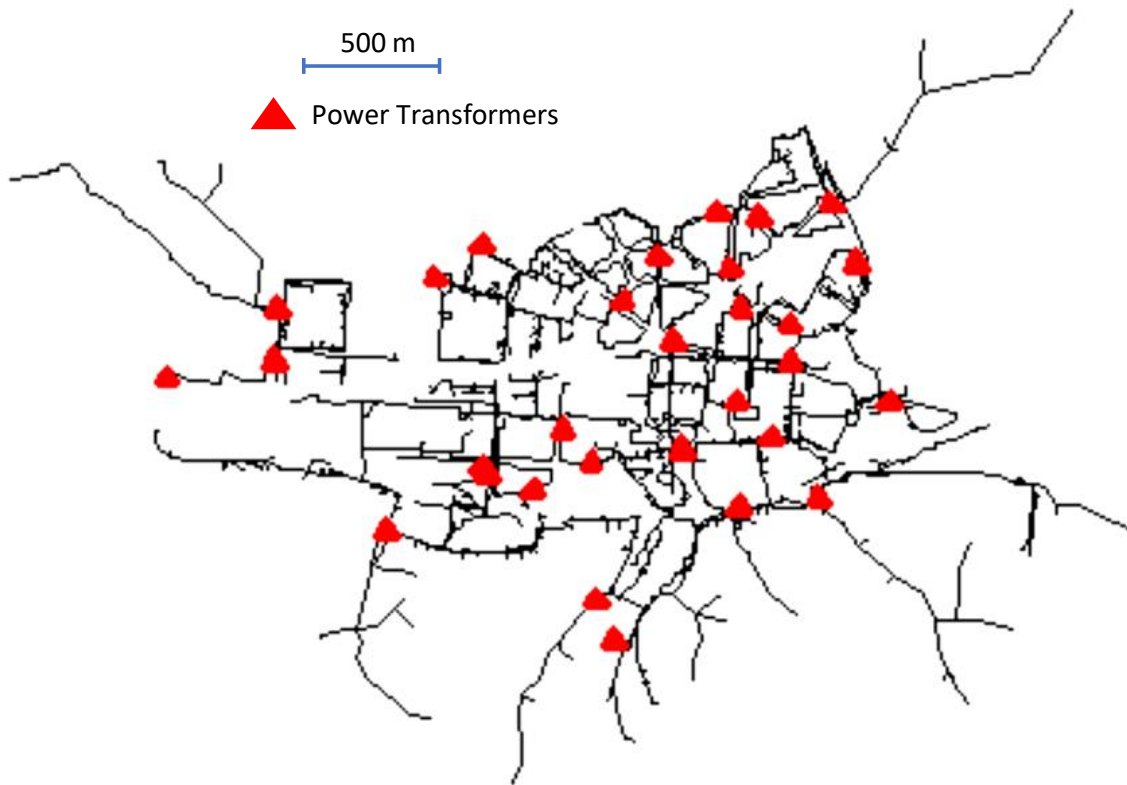
Figure 3.2: Non-synthetic European low voltage test network diagram with the power transformer locations.

It consists of 10,290 buses connected with 8087 loads, the distribution network is connected to the medium voltage line with the help of power transformers at the locations shown in Figure 3.2, and the medium voltage line is hidden. Each load is represented by its load shape that is captured using smart meters, and the load shapes are present for 20 calendar days with hourly samples.

The NSELVTN network is very complex concerning the size and meshing, to analyze in detail, the part of the network is selected by isolating the network supplied by the substation transformer 13. The sub-network of NSELVTN that is extracted and used is given in Figure 3.3. This network has a total of 7 feeder lines with 29 buses. The active power demand of the sub-network is given in Table 3.2 a total of *480 kW* peak load demand is served.
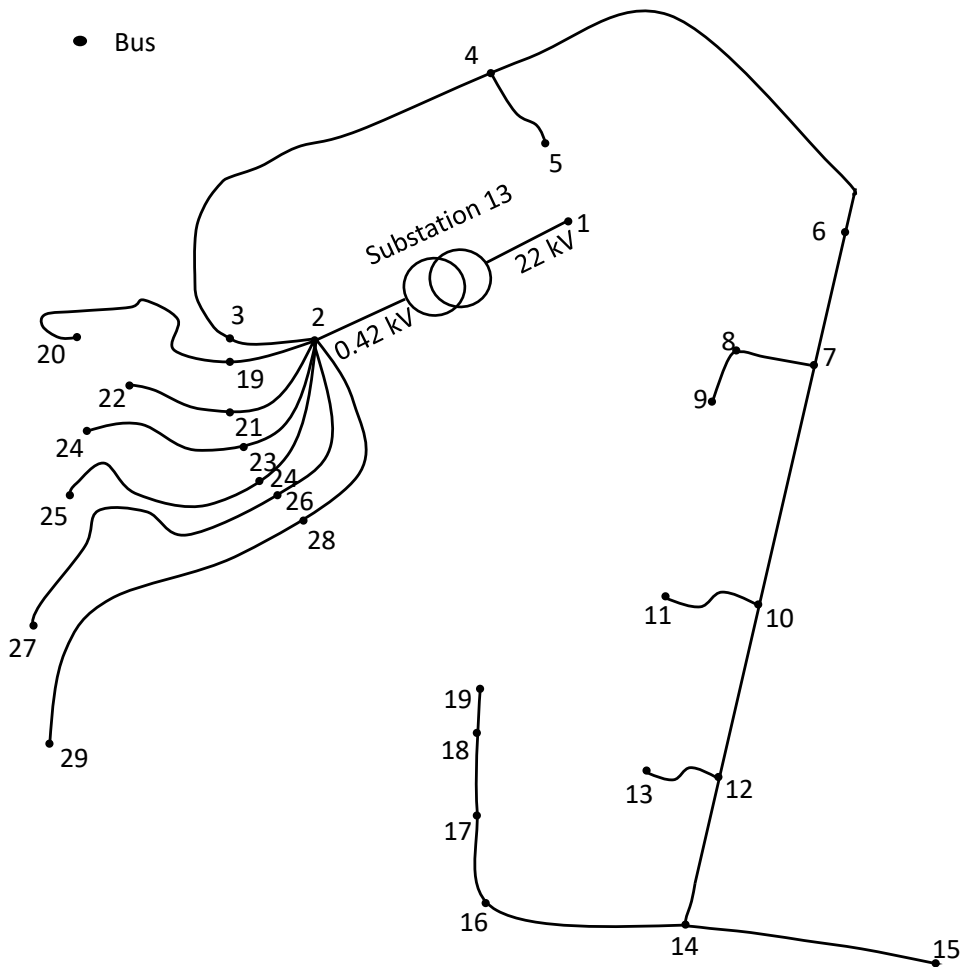
Figure 3.3: extracted Network of NSELVTN isolated and supplied with the power transformer 13.

Si.No.	Parameter	Active Power (kW)
1	Phase A	225
2	Phase B	135
3	Phase C	120
4	Total	480

Table 3.2: NSELVTN sub-network load data according to the phases.

3.3. IEEE European Low Voltage Test Network

IEEE European low voltage test network [107] is a test standard of a power distribution system network representing a real world network. It is a meshed network with a total area of *22,500 m²* and the loads are distributed randomly among the phases creating an unbalance in the system. It is power by a single power transformer of *800 kVA* rating, the operating line to line voltage of the network is *416 kV* at *50 Hz* frequency. It has a total of 906 buses, with 55 residential loads having an installed load capacity of 328 kW distributed among all three phases of the network [108, 109].
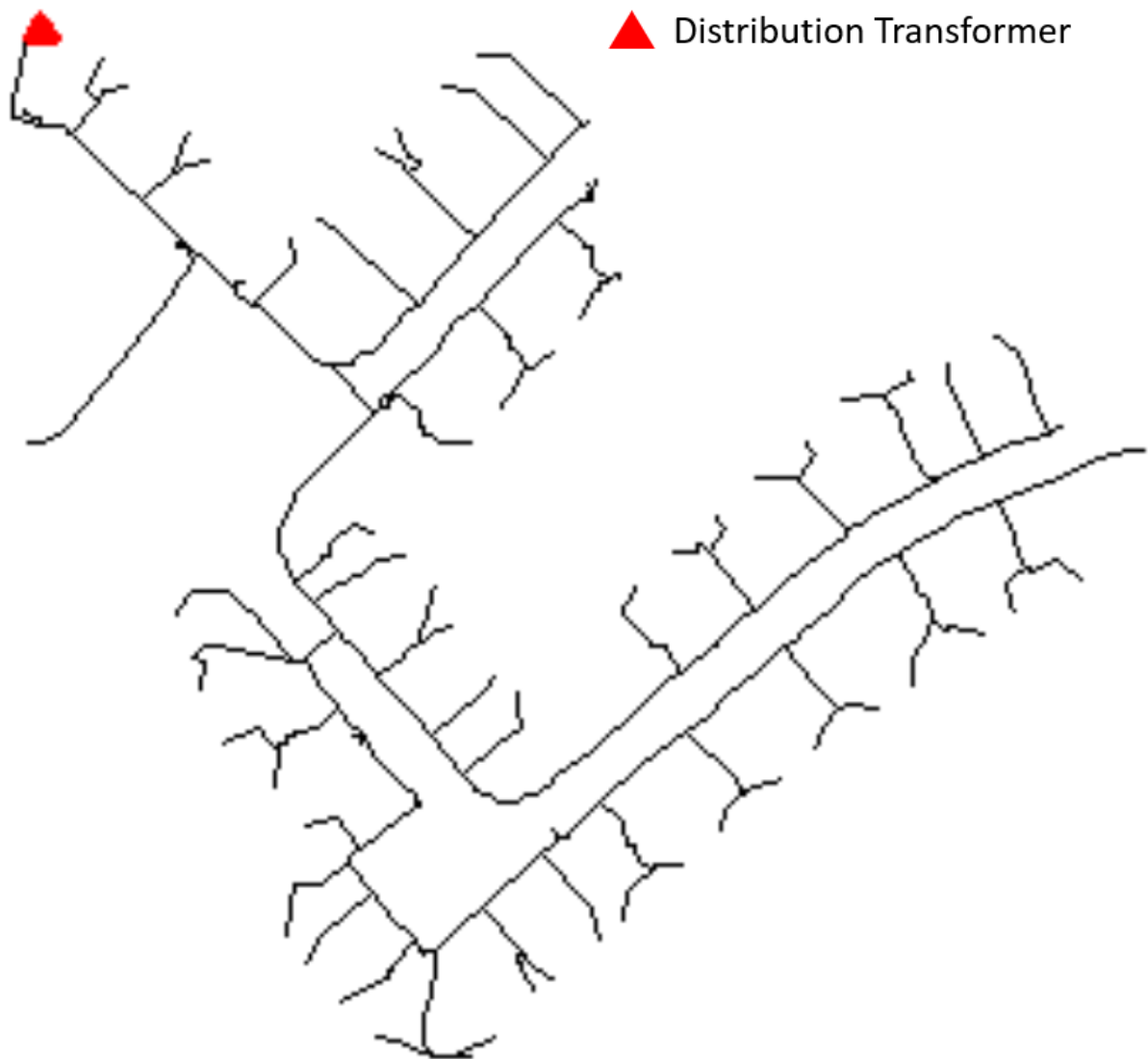
Figure 3.4: Single line diagram of the IEEE European low voltage test network.

3.4. 69 Bus Medium Voltage Test Network

69 bus is a three-phase balanced radial MV test network that operates at *50 Hz 12.7 kV* voltage rating, and as the name suggests, it has 69 buses and has a total of *11.4 MW 8.08 Mvar* load connected to it. The single-line diagram of the test network is shown in Figure 3.5. More details on the loading parameters of the network can be found in [110, 111].
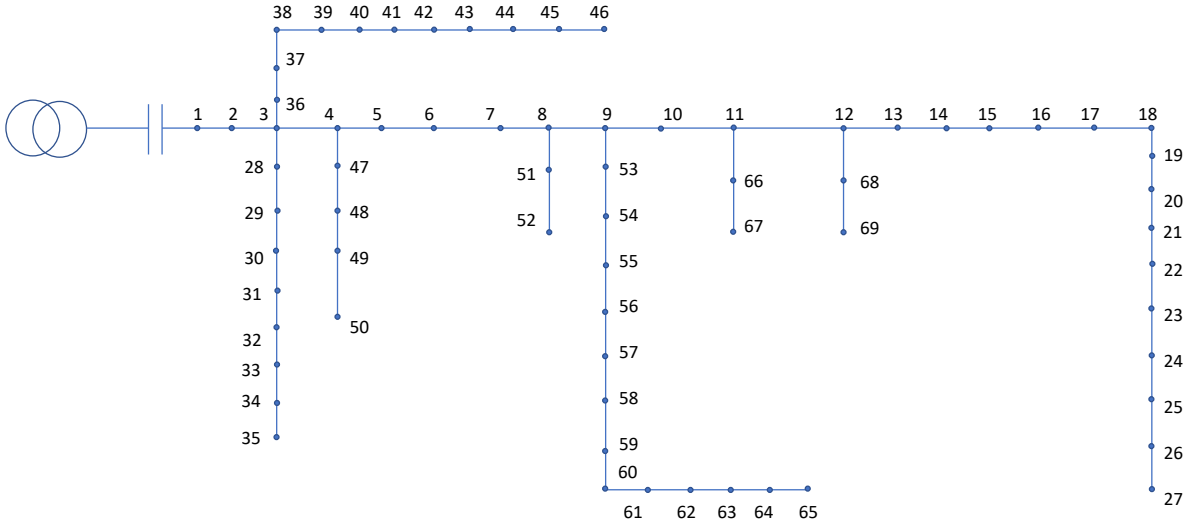
Figure 3.5: Single line diagram of the 69 bus MV test feeder network.

3.5. Probabilistic Load Flow formulation for unbalanced three phase systems

As explained in section 2.2, the PLF aims to determine the current injections at the buses with their complex voltages. The equation 2.4 is valid for a balanced three-phase network, the determined voltage at any bus is the voltage of all three phases. In the case of the unbalanced systems, the voltage among the phases is different, therefore a new formulation for an unbalanced case is given in this section.

Consider an electrical network with N number of buses with l number of lines. The phase current relationship with its respective phase voltage is given by the Equation 3.1.

$$\begin{bmatrix} I_1^{abc} \\ \vdots \\ I_N^{abc} \end{bmatrix} = \begin{bmatrix} Y_{11}^{abc} & \cdots & Y_{1N}^{abc} \\ \vdots & \ddots & \vdots \\ Y_{N1}^{abc} & \cdots & Y_{NN}^{abc} \end{bmatrix} \begin{bmatrix} V_1^{abc} \\ \vdots \\ V_N^{abc} \end{bmatrix} \tag{3.1}$$

where,

a, b, c	Phase representation
I_i^{abc}	Three phase current
Y_{ij}^{abc}	Admittance matrix
V_j^{abc}	Three phase voltage

$$I_i^{abc} = \begin{bmatrix} I_i^a \\ I_i^b \\ I_i^c \end{bmatrix}$$

I_i^a, I_i^b, I_i^c Current injected in i^{th} bus

$$Y_{ij}^{abc} = \begin{bmatrix} Y_{ij}^{aa} & Y_{ij}^{ab} & Y_{ij}^{ac} \\ Y_{ij}^{ba} & Y_{ij}^{bb} & Y_{ij}^{bc} \\ Y_{ij}^{ca} & Y_{ij}^{cb} & Y_{ij}^{cc} \end{bmatrix}$$

Y_{ij}^{pq}	Element of admittance matrix
i, j	Combination of buses
p, q	Combination of phases

$$V_j^{abc} = \begin{bmatrix} V_j^a \\ V_j^b \\ V_j^c \end{bmatrix}$$

V_j^a, V_j^b, V_j^c Complex voltage of jth bus

For any bus i and phase p, the current I_i^p of the bus is given by the Equation 3.2.

$$I_i^p = \sum_{k=1}^{N} \sum_{q=a,b,c} Y_{ik}^{pq} V_k^q \tag{3.2}$$

The unbalanced three phase system complex power is determined using Equation 3.3,

$$S_i^p = V_i^p \sum_{k=1}^{N} \sum_{q=a,b,c} (Y_{ik}^{pq})^* (V_k^q)^* \tag{3.3}$$

Where,

S_i^p Complex power injected

$(Y_{ik}^{pq})^*$ Admittance conjugate

$(V_k^q)^*$ Voltage conjugate

The voltage quality of the power system network in a three-phase system is measured by the inequality of the phase voltages in the selected node. This inequality in the voltage is called as voltage unbalanced factor (VUF), it is defined as the ratio of the negative voltage sequence component to the positive voltage sequence component, and it is expressed in percentage as given in Equation 3.4.

$$VUF = \frac{|V_n|}{|V_p|} * 100 \tag{3.4}$$

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

$$V_p = \frac{V_{AB} + \alpha \cdot V_{BC} + \alpha^2 \cdot V_{CA}}{3}$$

where,

V_n Negative voltage sequence component

V_p Positive voltage sequence component

V_{AB}, V_{BC}, V_{CA} Line voltages

α $exp(j120°)$

α^2 $exp(j240°)$

The unbalanced complex power Equation 3.2 is of DLF, to move from DLF to PLF probabilistic uncertainty factor should be included. The uncertainty is represented using a set of stochastic parameters $\vec{\xi} = \{\xi_1, \xi_2, ...\xi_l\}$, each ξ_l represents a random variable with the PDF $f(\xi_l)$,this stochastic parameters can be voltage, current or power injections at any nodes. In this work the stochastic parameter considered is the active power injections in the power system network by PV and EV elements. The reactive power is not considered due to IEEE standard 1547 [112] as the PV elements should not inject reactive power into the grid.

The active power injected or absorbed at any node in the grid varies with respect to the time t_m, hence the modeled PDF of the uncertainty is time dependent i.e., $f[t_m, \xi_l]$. With the dependence of time t_m and the uncertainty vector $\vec{\xi}$ the voltage of nth node p^{th} phase is given by $V_n^p[t_m, \vec{\xi}]$. For every time step an uncertainty vector is constructed $\vec{\xi^m}$ and load flow is conducted for each vector element to determine $V_n^p[t_m, \vec{\xi^m}]$, this process formulated

the PLF of a three phase unbalanced system.

3.5.1. PLF simulation Framework

The framework to conduct the PLF simulation using the load flow simulation software is given in Figure 3.6, the stochastic modeling and the uncertainty vector generation is performed in the MATLAB software, then the desired network configuration is loaded in the load flow simulator, the active power in the desired nodes are changed, and DLF is performed, the network variables are stored for further analysis. This process is repeated for $\vec{\xi^m}$ of times to perform the PLF.
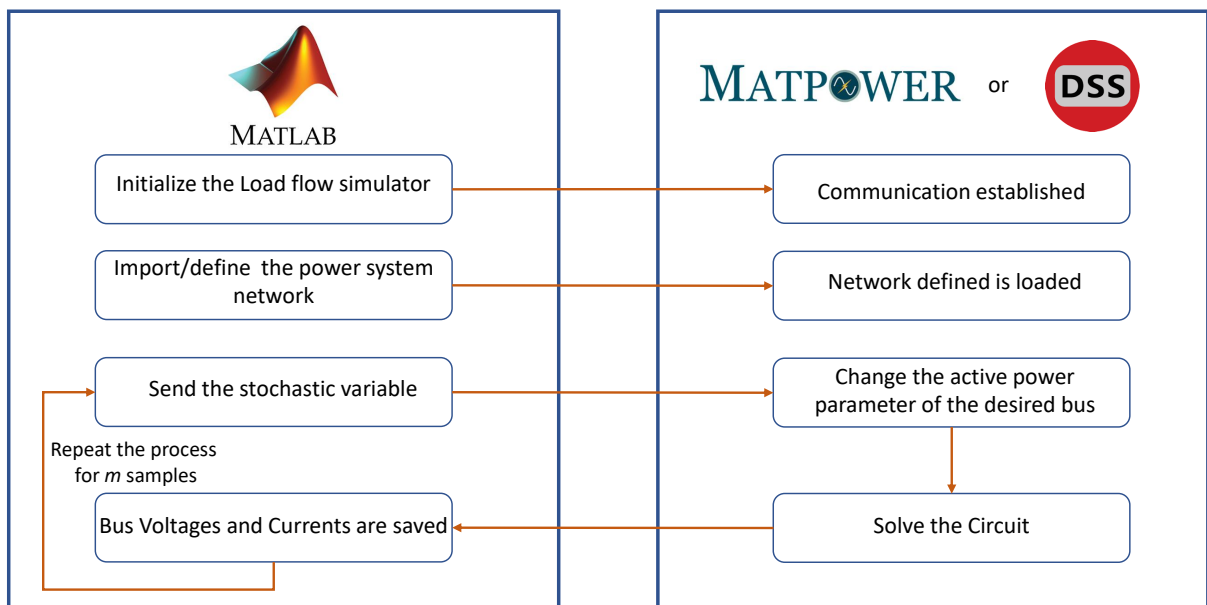
Figure 3.6: Simulation framework to conduct the PLF using MATLAB and the transient simulation software such as MATPOWER or OpenDSS.

4 | Data-Driven Modeling of PV Generation Uncertainty

In this Chapter modeling PV uncertainties using power output measurement data of PV generators is explained for use in the PLF simulation. Initially to repopulate the samples for use in PLF, only the shape of the PDF is considered and modeled using Nataf's transformation. Later, the inclusion of correlation among the PV generators is made possible with the use of the Gaussian Mixture Model and Gaussian Copula.

4.1. PV Power Generation Data

Obtaining PV power generation data is the first step in modeling PV generation uncertainty. As explained earlier in Section 2.3.2, uncertainty in PV power generation is caused by meteorological effects and system uncertainties such as device failure. The modeling techniques such as Clearness sky index, diffused fraction modeling, and PV power curve-based modeling can include meteorological uncertainties, but it fails to capture uncertainties caused due to partial shading, device failure, power network, and topology uncertainties.

To overcome such limitations in modeling, a data-driven approach that uses historical measurement data of PV plants output power is considered. This way, it can include both meteorological and system-dependent uncertainties in PV power generation. The proposed method is developed for historical measurement data but can also be applied to other methodologies. For demonstration purposes here, the recorded measured data of active power delivered by domestic PV systems present in the distribution network of Sydney city is considered [113].

The active power delivered by PV plants in a day by three different PV plants that are spatially separated are given in Figure 4.1, and significant variations of power are seen in the mid-day hours.
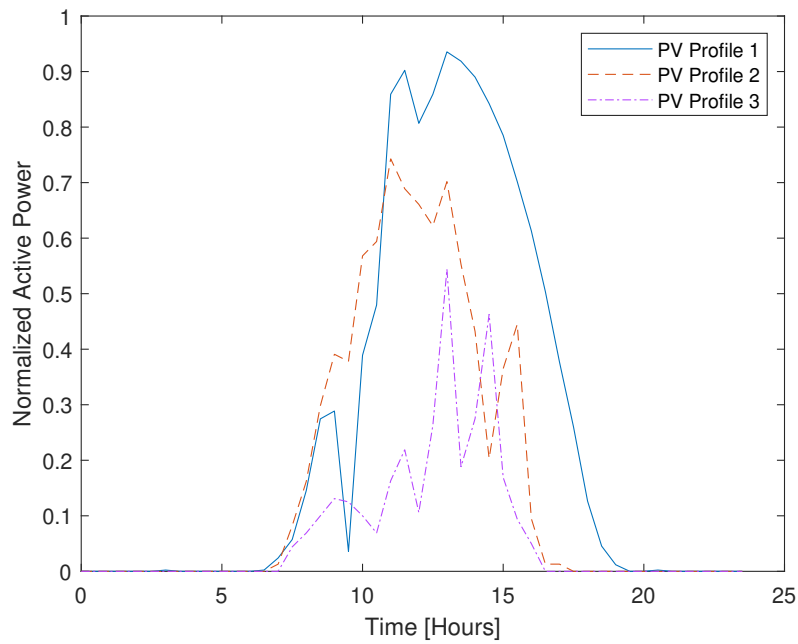
Figure 4.1: Normalized active power delivered by three spatially separated PV plants on the same day.
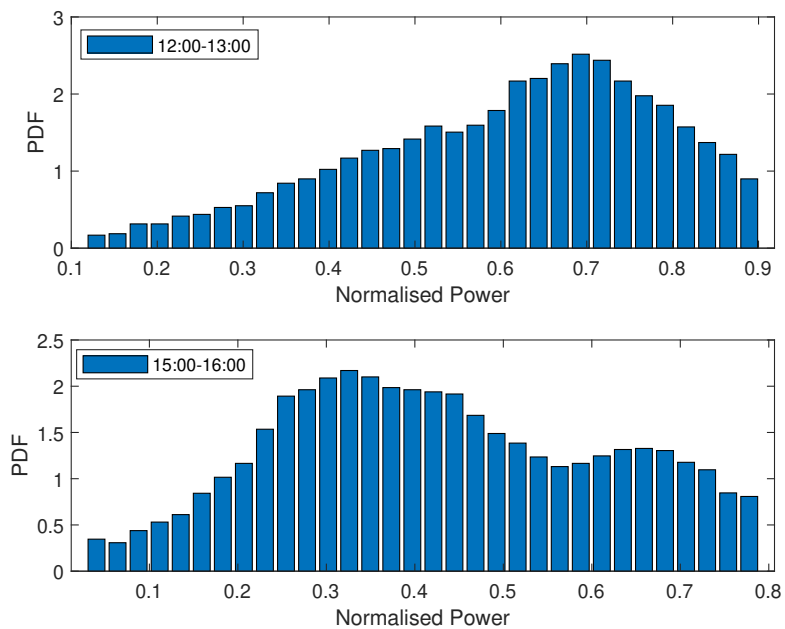
Figure 4.2: PDFs of a normalized active power delivered by a PV plant for one hour time window. A clear change in the shape of the distribution is observed between 12:00-13:00 and 15:00-16:00 time windows.

The data collected is for one calendar year from $1^{st} July$ 2012 to $30^{th} June$ 2013, the measurement samples are recorded every 30 minutes. The recorded data is grouped into an hourly time window representing the sample time t_m, and for a one-hour time window, there are 730 samples. The PDF of such hourly time window obtained from the samples is shown in Figure 4.2, the change in the distribution shape is seen between the 12:00-13:00 hour time window and the 15:00-16:00 hour time window. Such change in the distribution shape is observed for every hour time window starting from 09:00 to 17:00. For PLF evaluation, 730 samples are not enough, a large number of samples are required for every time window Hence, the repopulation of samples is necessary.

To generate large sample sets required for MC simulation according to the non-standard statistical distribution shape as shown in Figure 4.2, different modeling approaches are followed. The first one is immediate that PV power samples of individual PV generators are considered independent of each other, and the PDF shape is preserved in the repopulated samples using cumulative distribution (CDF). In reality, the PV generators that are geographically close to each other and separated by minimal distance will be correlated in power generation. In this case, considering such correlation among the repopulated samples is imperative.

4.2. Inverse CDF method

In the inverse CDF method, the active power samples of the PV generators are considered as independent random variables, and the samples are repopulated independently, hence this method is also called as *Independent method*. The normalized active power grouped into hourly time window represented by a random variable X given in Equation 4.1 is considered for repopulation.

$$X = \frac{P_s}{k_s} \tag{4.1}$$

Where,

X   Normalized active power
P_s Measured active power in kW
k_s Installed capacity of the PV generator in kW

As discussed in Section 3.5, the statistical input parameter of the PLF is the PV active power uncertainty, the random variable X represents the uncertainty element ξ i.e., $X = \xi_1$, as the input random variable X grouped into time, the sample time t_m is considered into picture. The distribution of X is obtained from the data is non-elementary, and

its PDF is given by $f[t_m, x]$. The CDF $F_X(t_M)$ is deduced numerically using the PDF $f[t_m, x]$, random samples are drawn from a uniformly distributed variable $u \in [0, 1]$ and given as input to the inverse CDF $F_X^{-1}(\cdot)$ which will provide the repopulated sample values of x of a random variable X as given in Equation 4.2 called as Nataf's transformation [55].

$$X = F_X^{-1}(u) \Leftrightarrow u = F_X(x) \tag{4.2}$$

For each realisation $X=x$, the PV active power involving the uncertainty is computed using $P_n[t_m, \xi_1] = k_s \cdot x$. The same process is repeated for second statistical variable $\xi_2 = Y$, that represents the active power samples second PV generator, and so on for all the considered PV generators for the PLF. The active power of PV employed is used an input to the load flow simulation to get the associated $V_n^p[t_m, \vec{\xi}]$.
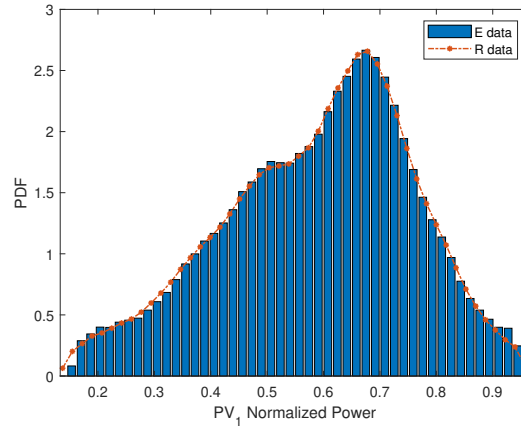
(a) PDF of PV generator 1
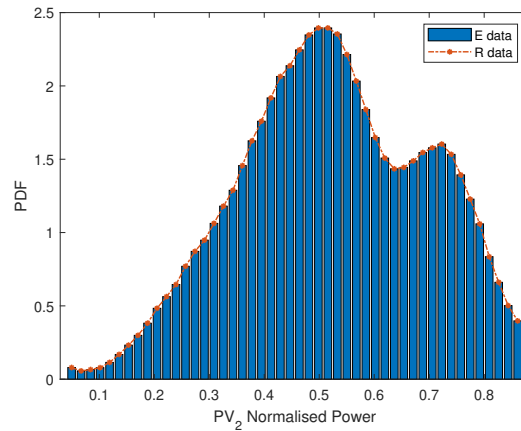
(b) PDF of PV generator 2

Figure 4.3: Marginal PDFs of the PV generators computed with empirical data (E data) and with repopulated samples (R data) using independent method.

The marginal PDFs of two PV generators obtained by grouping measurement data for the 12:00-13:00 time window are given in Figure 4.3, and the PDF of the repopulated samples obtained using the independent method fits the empirical data with relative errors $< 1\%$.

The independent method is good and efficient in modeling PV generation uncertainties, but the major drawback is, that it is assumed the PV generation is independent. In reality, PV generators are spatially correlated, and the degree of correlation in this type is proportional to the distance among the generators. It is necessary to include this correlation among the input random variables in modeling and can be done using GMM and Gaussian Copula.

4.3.   Gaussian Mixture Model

In Gaussian mixture modeling method [114], a mixture of a finite number of Gaussian distributions are used to fit the empirical data of the measured active power delivered by PV generators. This method is general, it can be used to model correlation between any sources, loads and other parameters of interest, for demonstration PV generators are considered. Also here, the normalized active power of the generators in the hourly time window is considered as given in Equation 4.1. Active power measurement of two PV generators in the time window 12:00-13:00 hour, whose marginal PDFs are given in Figure 4.3 is considered.

The combined effects of two PV generators PV_1 and PV_2 is modeled with the two random variables X_1 and X_2, garnered into a column vector $\vec{X} = [X_1, X_2]^T$ which has the joint probability function $f(\vec{X})$ given in Equation 4.3.

$$
\begin{cases}
f(\vec{x}) = \sum_{n=1}^{N_G} \alpha_n \, g(\vec{x}, \vec{\mu_n}, \Sigma_n) & \text{(4.3a)} \\[2ex]
g(\vec{x}, \vec{\mu_n}, \Sigma_n) = \dfrac{1}{((2\pi)^n |\Sigma_n|)^{0.5}} \, e^{-\frac{1}{2}(\vec{x}-\vec{\mu}_n)^T \Sigma_n^{-1} (\vec{x}-\vec{\mu}_n)} & \text{(4.3b)}
\end{cases}
$$

where,

n	1,2,.....,N
N_G	Number of Gaussian distributions
$\vec{X}$	Vector of random variables
α_n	scalar coefficients
$\vec{\mu_n}$	Mean value vector
Σ_n	Covariance matrix

Function $g(\cdot)$ is a multivariate Gaussian described by its mean value vector $\vec{\mu_n}$ and covariance matrix Σ_n. The coefficients $0 < \alpha_n \leq 1$ are the weights so that $\sum_{n=1}^{N_G} \alpha_n = 1$. The Equation 4.3 is the GMM model that can approximate the PDF using a sum of weighted multivariate Gaussians whose approximation converges to the PDF when the number of N_G components is increased [115].

The joint PDF of the PV sources using empirical data is fitted with the GMM model given in Equation 4.3 using Expectation Maximization (EM) iterative algorithm [116], the EM technique is defined in the MATLAB function *"fitgmdist"*, for a given number of Gaussian components N_G EM algorithm delivers α_n, $\vec{\mu_n}$ and Σ_n.

EM algorithm requires a number of Gaussian components N_G as an input to determine how many Gaussians to be used to fit the model, the variance of the GMM components is verified by gradually increasing the value of N_G, the criteria to stop is to reach the minimum sum of the variances of nth component σ_{min} given in Equation 4.4.

$$\sigma_{min} = \min_{n} \; sum[diag(\Sigma_n)] \tag{4.4}$$
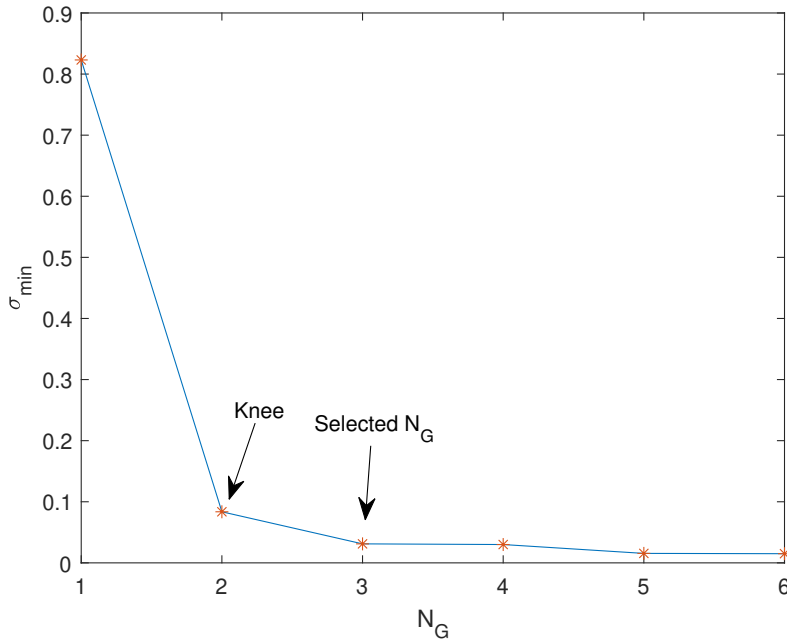
Figure 4.4: The dependence of minimum sum of diagonal elements of the covariance matrix σ_{min} with the Number of Gaussians N_G is used.

The empirical PDFs of the PV generators shown in Figure 4.3 are considered, and the GMM model is used to fit the scattered empirical data. To determine the number of Gaussians N_G to be used in fitting, the minimum sum of diagonal elements of the covariance matrix is determined as shown in Figure 4.4. Once the Knee point is crossed, the increase in the number of Gaussians will result in relatively smaller variances, here the Knee is observed with $N_G = 2$ so, $N_G = 3$ is selected as input to the GMM model. Figure 4.5 shows the PDFs obtained after passing the correlated random variables i.e., empirical data of normalized active power two PV generators considered for a time window of 12:00-13:00 having correlation coefficient $\rho = 0.826$ through the GMM model, the scattered plot shows the empirical data while the contour lines show the PDFs obtained after fitting.
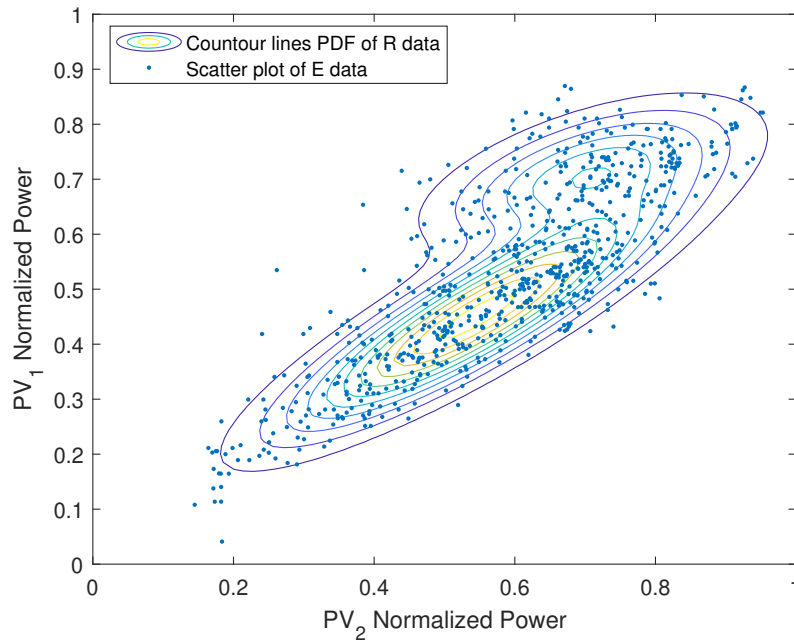
Figure 4.5: The PDFs were obtained by fitting the scattered empirical data of the PV generators using the GMM model.

4.4.   Gaussian Copula

A copula is a function that contains the dependence information about the random variables, and it couples the uni-dimensional marginal distributions. The marginal distributions of random variables are combined to arrive at a joined distribution using copulas [117]. The marginal dependency is investigated in a common domain, and once the joined distribution is obtained, repopulated samples of random variables are mapped to their original form using their CDFs without losing any information associated.

Consider physical random variables as x_k, for $k = 1, ..., N$ and its empirical CDF is given by $F_k(x_k)$, to arrive at joined distribution from the marginals Equation 4.5 is used. Repopulation of samples using copulas involves two steps, the first is the transformation of physical variables, and the second step is the regeneration of correlated samples.

$$F_{1k}(x_1, x_k) = C(F_1(x_1), F_k(x_k)) \tag{4.5}$$

Where,

k	1,2,.....,N
F_{1k}	Joint CDF of variables x_1, x_k
C	Copula function carrying dependence
$F_1(x_1)$	CDF of variable x_1
$F_k(x_k)$	CDF of variable x_k

4.4.1.   Transforming physical variables

As defined earlier, physical random variables x_k and associated CDF $F_k(x_k)$ is selected. Now consider a uniform random variables u_k that are distributed in $u \in [0, 1]$ the random variables u_k and x_k are related as given in Equation 4.6. A set of middle variables z_k is obtained by using Nataf's transformation [55] given in Equation 4.7, $\Phi(\cdot)$ represents the Gaussian CDF hence the name Gaussian Copula.

$$u_k = F_k(x_k) \tag{4.6}$$

$$z_k = \Phi^{-1}(u_k) = \Phi^{-1}(F_k(x_k)) \tag{4.7}$$

In Gaussian Copula, the middle variables z_k are assumed to be joint Gaussian. As given in Equation 4.7 the samples of z_k are derived from the measurement data set x_k, it facilitates to compute the correlation matrix C_z among the z_k variables. These z_k variables are collected into a vector $\vec{z}$ that can be decorrelated using a liner transformation relationship given in Equation 4.8.

$$\vec{z} = L\,\vec{\xi} \tag{4.8}$$

Parameter L in the Equation 4.8 is the square root of correlation matrix C_z of vector $\vec{\xi}$ which is determined using Cholesky decomposition method [54]. The vector $\vec{\xi}$ holds N Gaussian distributed and independent variables ξ_k, and these normal distributed variables are the inner variables of the model. The detailed transformation of physical variables to arrive at inner independent variables is given in Figure 4.6.
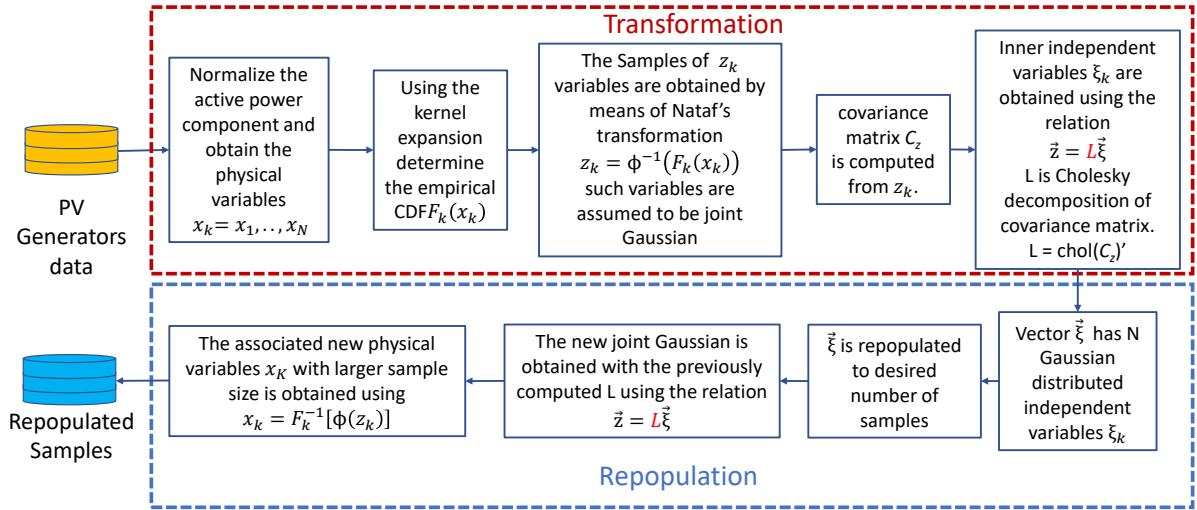
Figure 4.6: Flowchart to implement Gaussian copula for sample repopulation, showing from transformation of physical stochastic variables ξ_k obtained from the PV profiles to regeneration of desired number of samples for the use in PLF.

4.4.2. Generating correlated samples

The transformation of physical input stochastic variables as explained in previous Section 4.4.1 is followed backward to generate a large set of correlated samples of physical variable x_k while preserving the shape of the marginal distributions and the correlation among them. The desired number of samples are generated using the normal distributed inner independent variables $\vec{\xi}$, the related middle variables are obtained using $\vec{z}$ as given

in Equation 4.8, later the associated physical variables x_k are determined by taking the inverse CDF of z_k as given in Equation 4.9. The computational flow of repopulation establishes a deterministic relationship between the physical stochastic input variables x_k and the inner independent variables ξ_k.

$$x_k = F_k^{-1}[\Phi(z_k)] \tag{4.9}$$

As previously shown in Figure 4.3, the marginal PDFs computed with the empirical data of the active power delivered by PV generators for the time window 12:00-13:00 are considered. The correlation coefficient among the PV generator data for these selected random variables is $\rho_E = 0.826$, the samples for the PLF are repopulated using Gaussian Copula, and the correlation among the repopulated variables is $\rho_R = 0.817$, the marginal PDFs of empirical data and the repopulated samples are given in Figure 4.7.
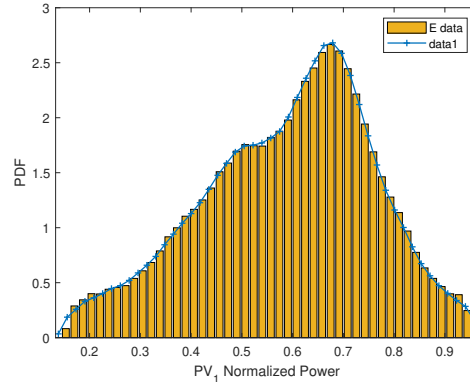
(a) PDF of PV generator 1
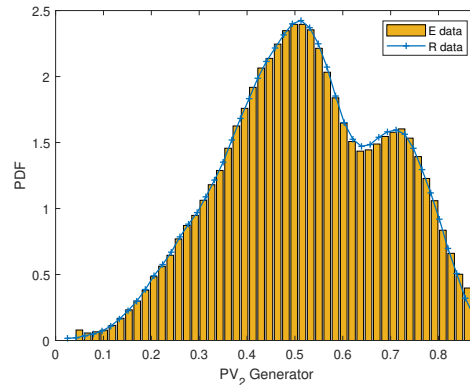
(b) PDF of PV generator 2

Figure 4.7: Marginal PDFs of the PV generators computed with empirical data (E data) and with repopulated samples (R data) using Gaussian Copula.

It is confirmed that using Gaussian copula, non-standard statistical distributions can be handled effectively for the use in PLF, the repopulated variables resulted has correlation coefficient with a relative error of around $\approx 1\%$, and the PDF shape exactly matches the PDF of the empirical data with a relative error rate of $< 1\%$.

4.5.   Comparison among modeling methods

All the methods discussed in this Chapter proved to handle non-standard statistical distribution of PV generation uncertainty. In this section, all the described methods are compared to see conclusive results.

The important parameters interested in modeling are, the PDF shape, with which the characteristics of the distribution such as mean and the standard deviation are obtained, and then the correlation among the variables. The scatter plots of the two PV generators data and the repopulated samples of the same using all the modeling methodologies described above are shown in Figure 4.8.
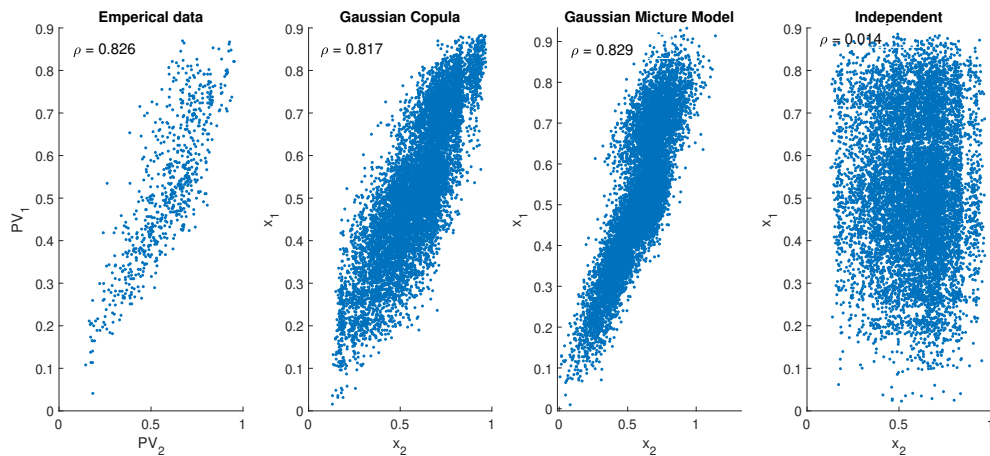
Figure 4.8: Scatter plot of the normalised active power of PV generation, is plotted with empirical data, and repopulated samples obtained from Gaussian Copula, GMM, and Independent method.

It is expected that the independent method will not reproduce samples considering correlation, but the resulting PDFs effectively fit the distribution of empirical data. The GMM model preserves correlation more accurately, but the resulting PDF shape is more distorted than the original, and the relative error is much higher. The Gaussian copula is effective as it preserves the Shape and correlation efficiently, the details of comparison are given in Table 4.1.

Method	Correlation	PDF Shape	Simulation time
	relative error	relative error	10k samples
Independent	-	$\leq 1\%$	1.09 s
GMM	$< 1\%$	$\leq 5\%$	0.5 s
Gaussian Copula	$\approx 1\%$	$\leq 1\%$	1.13 s

Table 4.1: Comparison among proposed modeling techniques to capture PV uncertainty.

4.6. Impact of Correlated PV generators on the Power System Network

In this section, the proposed modeling techniques of PV generation uncertainty, such as the Independent method, GMM, and Gaussian Copula are included in PLF using MC simulation methodology to study their impact on the power system network. Four correlated PV generators are considered, their correlation among each other is given in Table 4.2.

	PV_1	PV_2	PV_3	PV_4
PV_1	1	0.826	0.732	0.618
PV_2	0.826	1	0.694	0.588
PV_3	0.732	0.694	1	0.730
PV_4	0.618	0.588	0.730	1

Table 4.2: Correlation coefficient among PV generators for the 12:00-13:00 hour time window.

To evaluate the impact of correlated PV on the power system network, IEEE 13 node test network is modified with the inclusion of four correlated PV generating sources having correlation as mentioned in Table 4.2, the modification details of the test network are given in Table 4.3 and shown in Figure 4.9.

Element	Bus name	Phase	P [kW]
1ϕ PV generator 1	4	B	250
1ϕ PV generator 2	9	B	250
1ϕ PV generator 3	4	C	300
1ϕ PV generator 4	9	C	300

Table 4.3: Modifications in IEEE 13 node test feeder network to include correlated PV generators.
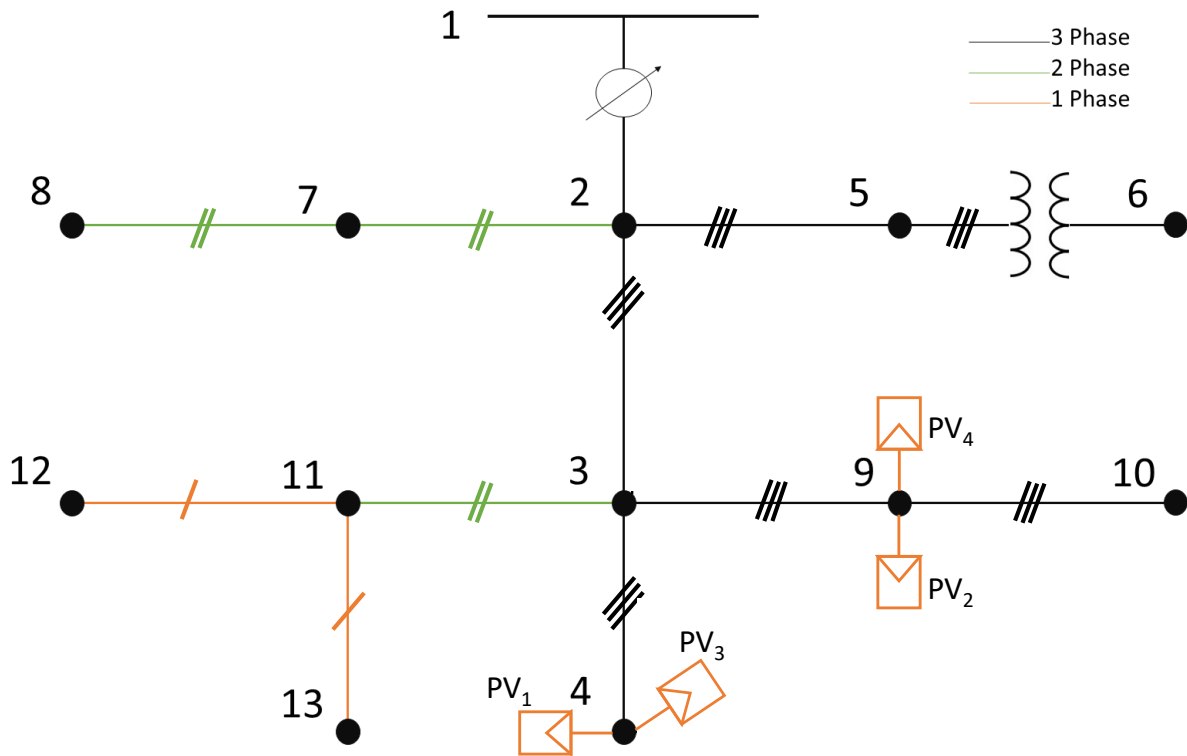
Figure 4.9: Modified 13 node test feeder network with four correlated PV generators, the voltages are measured at Bus 3 to study the impact of PV generators on the network.

The PV generators connected to *Phase B* are having installed capacity of *250 kW*, and connected to *Phase C* are having installed capacity of *300 kW*, the installed capacity is considered according to the load demand present in the respective phases. The uncertainty vector of the PV generators ξ_1, ξ_2, ξ_3 & ξ_4 are gathered using the modeling methodologies in hourly time windows from 09:00 to 18:00, then the distributions each having 10,000 samples are scaled to the respective installed capacity to realize the active power stochastic variable $P_1[t_m, \xi_1]$, $P_2[t_m, \xi_2]$, $P_3[t_m, \xi_3]$ and $P_4[t_m, \xi_4]$ to use in MC simulation.

MC simulation is performed with independent, GMM, and Gaussian copula to model the PV characteristics. Figure 4.10 shows the Phase C voltage distribution measured at Bus 3 for the time window 14:00-15:00 for different modeling techniques. It can be observed the effect of considering correlation among the PV sources results in a wider variability in the node voltage i.e., the standard deviation of *23.16 V* in the case of Gaussian copula, *23.23 V* of deviation in case of GMM whereas, *16.28 V* in case of independent method.
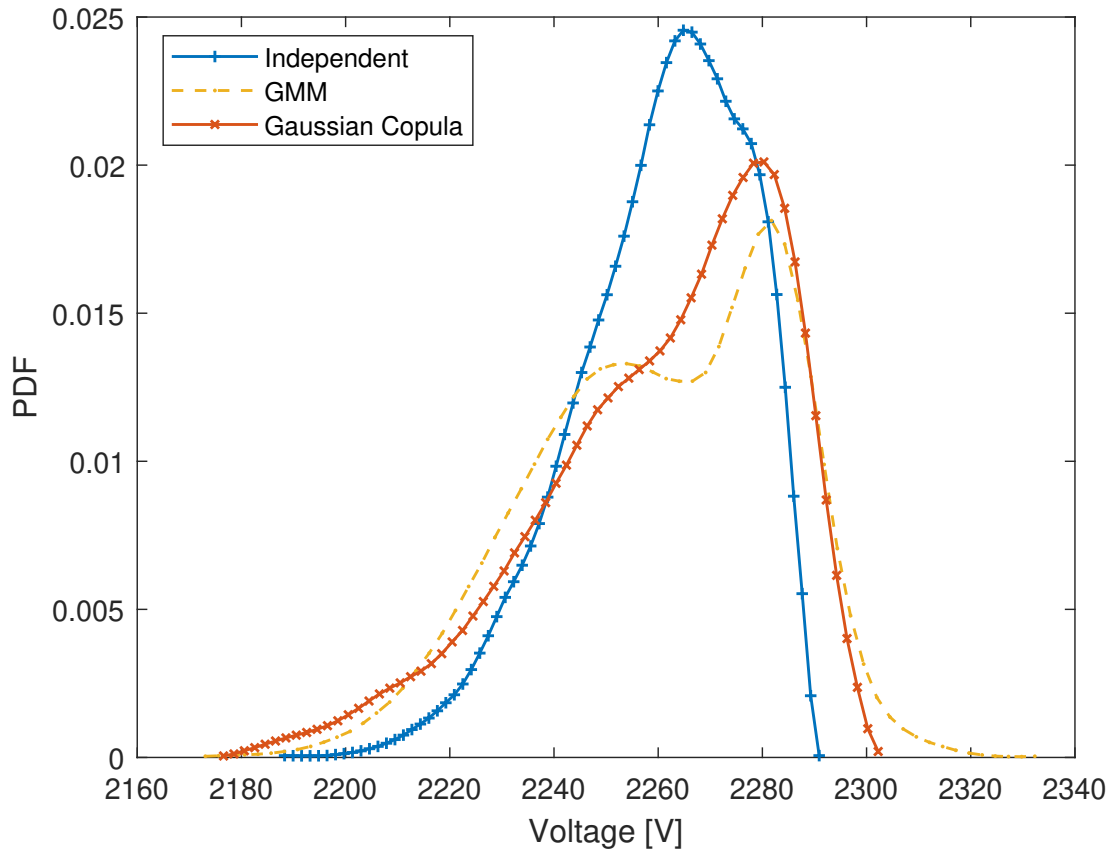
Figure 4.10: The Phase C voltage distribution of Bus 3 for the time window 14:00-15:00, the active power stochastic variables are modeled using Independent method, GMM and Gaussian Copula.

The Phase B (ϕ_b) and Phase C (ϕ_c) voltage distributions of Bus 3 resulted from MC simulation in hourly time window is given in Figure 4.11. The mean value of the distribution results in approximately the same in all modeling methods. The variation is seen in the standard deviation between independent and correlated methods.
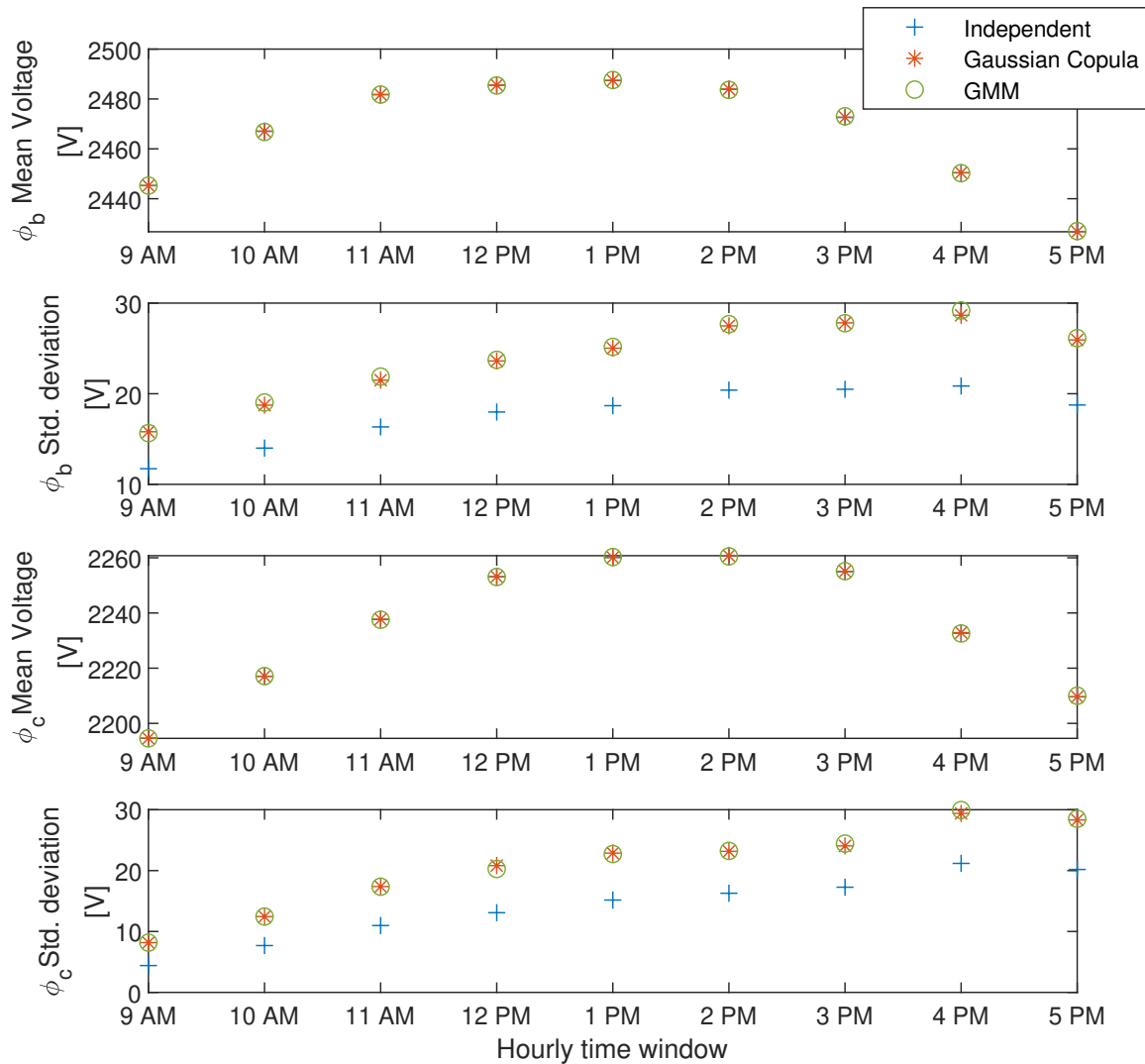
Figure 4.11: Comparison of mean and standard deviation of the Bus 3 Phase B (ϕ_b) and Phase C (ϕ_c) voltage distributions using different modeling approaches of PV generation uncertainty.

From Figure 4.11 it is evident to consider correlation among the sources present in the power system network, hence the Independent method can be ruled out, as it fails to consider such correlation among the variables. GMM and Gaussian copula effectively handle the correlation approximately with the same accuracy, but from Figure 4.10 it is hard to determine which shape of the PDF is accurate.

To determine which method can deliver better results, the distribution shape of empirical data of the PV generators PV_3 and PV_4 are compared with repopulated samples using

GMM and Gaussian copula as shown in Figure 4.12. The Gaussian copula can fit the distribution of empirical data more accurately than the GMM. Therefore, results from Gaussian copula are used to evaluate the impact of PV generators on the network.
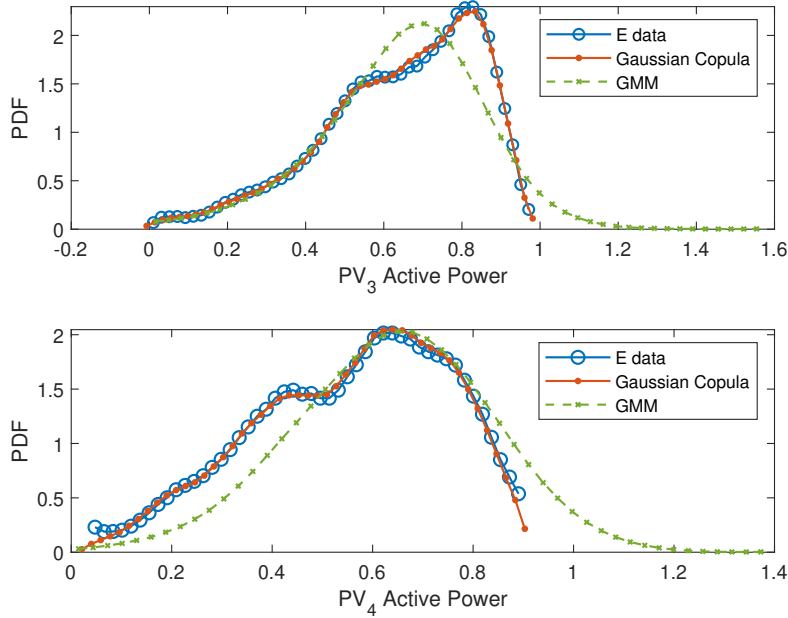
Figure 4.12: Comparison of marginal PDFs of PV generators PV_3 and PV_4, and repopulated samples using GMM and Gaussian Copula.

The hourly voltage distributions of Phase B and Phase C due to change in PV active power delivery resulted in PLF using Gaussian Copula is given in Figure 4.13 and Figure 4.14. The network health due to injection of uncertain active power into the network is determined using the voltage unbalance factor in percentages given by the Equation 3.4, and the results in hourly time windowed distribution of $VUF\%$ due to the PV generators are given in Figure 4.15.

The obtained results clearly shows that the considering correlation among the PV generators data significantly varies the numerically determined statistical distribution of the network. Specifically, MC simulation with correlation resulted in wider variability of node voltages compared to independent method where correlation is ignored.
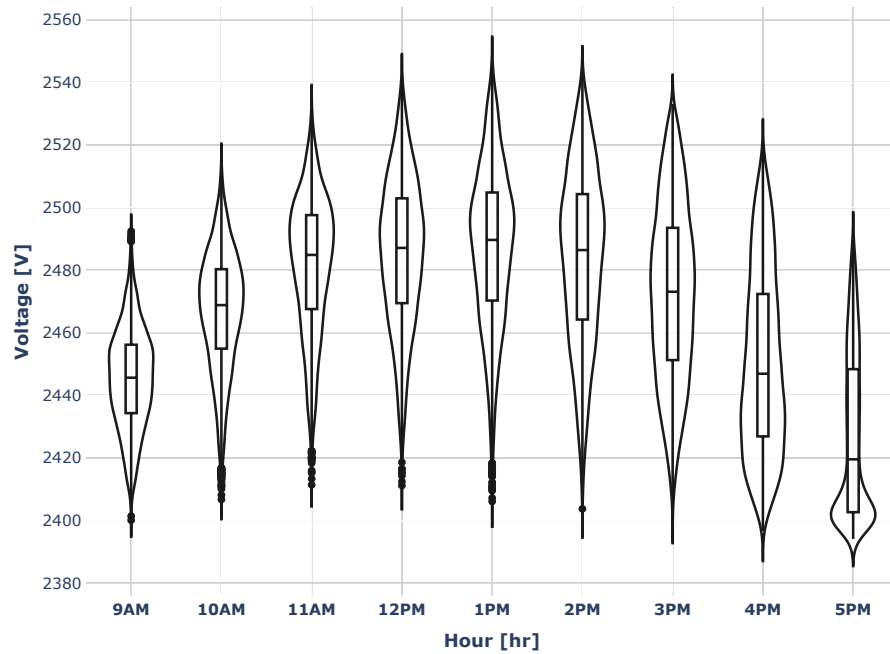
Figure 4.13: Bus 3 phase B voltage distributions of IEEE 13 node test feeder network considered in hourly time window from 09:00 to 18:00.
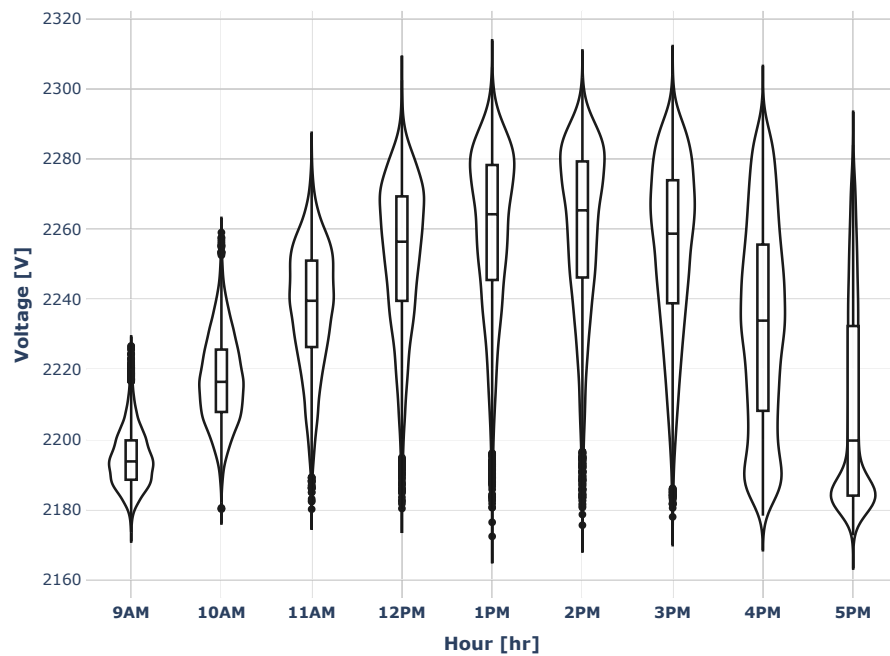
Figure 4.14: Bus 3 phase C voltage distributions of IEEE 13 node test feeder network considered in hourly time window from 09:00 to 18:00.
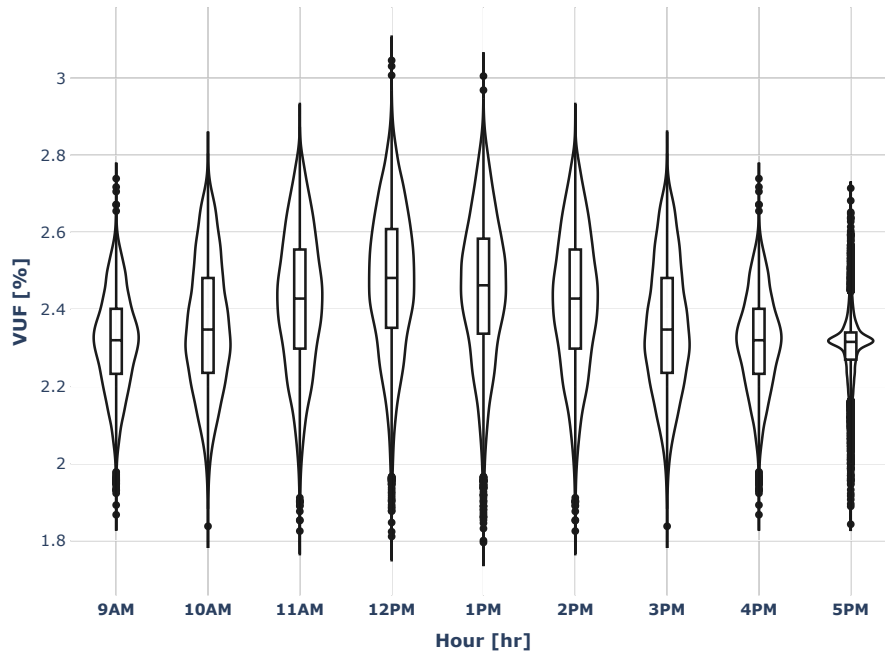
Figure 4.15: Distribution of the voltage unbalance factor measured in percentage of Bus 3 IEEE 13 node test feeder network considered in hourly time window from 09:00 to 18:00.

5 | Novel PLF using SRSM and Gaussian Copula

In the previous chapters, it is well demonstrated how the data-driven modeling of PV generators results in non-standard statistical distributions and how correlation among such generators will influence the simulation results. MC simulations stands as an ultimate reference to compare against any other PLF methods due to the simulation results accuracy. However, its slow convergence rate makes the simulation run with a large sample set of data, resulting in higher simulation time that are not acceptable.

As explained earlier, there are many approaches to speed up the PLF process, using analytical or approximation approaches. An effective way of doing PLF is to adopt a surrogate model of the network response that employs polynomial chaos expansion. In this adoption, Stochastic Response Surface Method is preferable as they directly deliver PDFs of non-elementary quantities of the power network such as voltage and VUF in the PLF process. In [69] SRSM is extended with copula for PLF analysis, the transformation using Natf's allowed PV powers into a new set of inner variables $\vec{\xi}$ which is statistically independent. A surrogate SRSM can be derived approximating the relationship with the inner variables $\vec{\xi}$ and the outer variables containing the uncertainty x_k using standard Hermite polynomial.

5.1. Conventional SRSM

Let the observable variable of the grid in the PLF process such as voltage, current, and $VUF\%$ be y, a direct method to implement conventional SRSM along with copula is by using a surrogate model in the approximation of the multi-variate relationship between inner independent parameters ξ_k of copula and the observed variable y. To remind, in copula functions the non-linear function of the inner variables ξ_k connects the physical variables x_k. As the observed variable y depends on the input physical variables x_k carrying the uncertainty i.e., normalized active power delivered by PV, it turns out that the output observed variable y is ultimately depends on the inner variables ξ_k as given in

Equation 5.1.

$$\begin{cases} x_k = h_k(\xi_1, ..., \xi_N) & \text{(5.1a)} \\ y = l(x_1, ..., x_N) = \tilde{h}(\xi_1, ..., \xi_N) & \text{(5.1b)} \end{cases}$$

In the conventional implementation of SRSM with copula, the surrogate model approximates the non linear relationship $\tilde{h}(\cdot)$ using polynomial chaos expansion formed by N_b multi-variate basis functions $H_i(\vec{\xi})$ given in Equation 5.2. The main benefit of this method is ease of implementation, the inner stochastic parameters are standard normal distributions and hence the associated polynomial basis functions $H_i(\vec{\xi})$ are standard Hermite polynomials.

$$y = \tilde{h}(\vec{\xi}) \approx \sum_{i=0}^{N_b-1} \tilde{c}_i H_i(\vec{\xi}) \tag{5.2}$$

5.1.1.   Drawback of Conventional SRSM

The hindrance of this method is due to the fact that the relationship between the physical variables x_k and the inner variables $\xi_1, ..., \xi_N$ is highly nonlinear, which is defined during the transformation step of Copula functions. In such case, the polynomial chaos will deliver poor approximations [118]. The high non-linearity observed in copula transformation step is because of using CDF $\Phi(\cdot)$ and inverse CDF $F_k^{-1}(\cdot)$ as explained in the flowchart given in Figure 4.6.

An example is considered to understand this problem, a Hermite polynomial of growing order is employed to interpolate the relationship 5.3 obtained in Copula implementation.

$$h(\xi) = F^{-1}[\Phi(\xi)] \tag{5.3}$$

Where,

ξ Normally distributed random variable
$\Phi(\cdot)$ Gaussian CDF
$F^{-1}(\cdot)$ Inverse CDF

The interpolation of function $h(\xi)$ with 2nd and 5th order Hermite polynomial interpolation is shown in Figure 5.1, Gaussian quadrature nodes are selected as interpolation points, i.e., the number of nodes is equal to the polynomial order+1. In Figure 5.1, it can be seen

how the interpolation results are inaccurate from the interpolation points. As the order is increased to 5, interpolation points oscillate at the domain borders, which is a problem of over fitting.
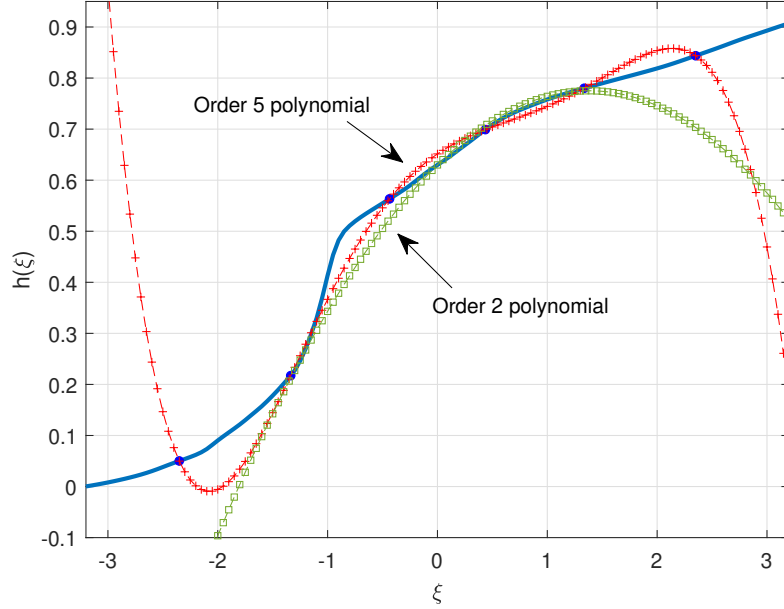
Figure 5.1: The continues line in blue shows the nonlinear relationship $h(\xi)$, the green line is the 2nd order Hermite polynomial interpolation and the red line is 5th order Hermite polynomial interpolation.

To avoid these inaccuracies in the PLF evaluation of output variable distributions using conventional SRSM in presence of non-standard statistical distributions of uncertain input parameters, a novel technique is used. The idea for the novel implementation is to apply the SRSM surrogate model and the Copula in two separate computational flows, and it overcomes the above-discussed drawback. To prove the accuracy and efficiency of the proposed novel PLF methodology, the obtained results are compared against the results of MC simulations for the same cases.

5.2. Novel SRSM with Gaussian Copula

The novel technique of PLF is implemented using Gaussian Copula and SRSM surrogate modeling in two separate steps. In the initial step, the repopulation of samples required for the PLF is realized with the utilization of the Gaussian Copula technique as described in detail in Section 4.4. The repopulated data contains the active power delivered by PV considered in an hourly time window. In the second step, a polynomial SRSM surrogate

model is used to precisely approximate the slightly nonlinear input and output relationship of the PLF process i.e., the relationship between the observable output variable of the grid y and the physical input variables of PV active power delivered x_k as given in Equation 5.4.

$$y = l(x_1, ..., x_N) \tag{5.4}$$

The main advantage of the proposed novel SRSM is that it uses lower order polynomial to approximate the relation 5.4, which results in more accurate approximations. Earlier the standard Hermite polynomials $H_i(\vec{\xi})$ were used as the variable to approximate ξ_k which are standard normal distributed variables. In the novel approach, the physical variables x_k which are non-standard statistical distributions are approximated, the surrogate SRSM cannot be built with the standard polynomial chaos bases. Instead, generalized base functions associated with statistical variables x_k are predetermined using a Three-Term Recurrence relation [103] and iterative Darboux's formula [75].

5.2.1.   Mathematical formulation of Novel SRSM

A truncated series expansion of an order-γ as given in Equation 5.5 is formed using N_b multi-variate generalized basis functions $\Psi_i(\vec{x})$ weighted by unknown coefficients c_i.

$$
\begin{cases}
y = l(\vec{x}) \approx \displaystyle\sum_{i=0}^{N_b-1} c_i \Psi_i(\vec{x}) & \text{(5.5a)} \\[2ex]
\Psi_i(\vec{x}) = \displaystyle\prod_{k=1}^{N} \psi_{ik}(x_k) & \text{(5.5b)}
\end{cases}
$$

Every multi-variate basis function $\Psi_i(\vec{x})$ is the product of the uni-variate polynomials $\psi_{ik}(x_k)$ of degree i_k associated to the physical random variable x_k that has non-standard marginal probability density function $f_k(x_k)$. As mentioned earlier, the uni-variate polynomials $\psi_{ik}(x_k)$ for non-standard distributed physical random variables are determined through a Three-Term Recurrence relation applied to the empirical PDFs obtained in the modeling step.

It is observed that if the physical random variables x_k are mutually independent, the multi-variate polynomials in Equation 5.5b are given by the product of the uni-variate polynomials, that provide a set of orthogonal basis allowing the approximation of smooth $y(\vec{x})$ relationship with high accuracy and also it helps in extraction of the statistical moments of the observed variable y. When the physical variables x_k are not independent,

as in the proposed data-driven modeling of correlated PV generation, the multi-variate polynomials are not orthogonal, so the expansion is not used for the analytical computations. To solve this issue, a new set of orthogonal polynomial bases can be derived using the Gram-Smith orthogonalization process.

However, when the relation 5.5a is simply used to accelerate the process of MC simulation, polynomial orthogonalization is not required, multi-variate functions of 5.5b preserves the approximation capability of the relationship $y(\vec{x})$ in presence of the correlated variables. Hence they are adopted as basis functions in the proposed response surface model given in Equation 5.5a.

In the Equation 5.5a the expansion coefficient c_i is determined using the least square regression technique in which the samples $N_s \geq N_b$ are generated for $\vec{x}$. In each sample $\vec{x_j}$ the observable variable value $y_j = l(\vec{x_j})$ is determined by running a deterministic load flow. The coefficients c_i minimizes the squared difference between the observable variables and the gPC evaluated at the samples i.e.,

$$c_i \approx \arg\min_{\tilde{c}_i} \frac{1}{N_s} \sum_{j=1}^{N_s} \left(y_j - \sum_{i=0}^{N_b-1} \tilde{c}_i \, \Psi_i(\vec{x_j}) \right)^2 \tag{5.6}$$

The Equation 5.6 is solved by the introduction of the matrix given in Equation 5.7, that collects the N_s samples in vector form, and it results in the linear system as given in Equation 5.8.

$$\mathbf{M} = \begin{bmatrix} \Psi_0(\vec{x}_1) & \dots & \Psi_{N_b-1}(\vec{x}_1) \\ \vdots & \ddots & \vdots \\ \Psi_0(\vec{x}_{N_s}) & \dots & \Psi_{N_b-1}(\vec{x}_{N_s}) \end{bmatrix} \tag{5.7}$$

$$(\mathbf{M}^T\mathbf{M}) \begin{bmatrix} c_0 \\ \vdots \\ c_{N_b-1} \end{bmatrix} = \mathbf{M}^T \begin{bmatrix} y_1 \\ \vdots \\ y_{N_s} \end{bmatrix} \tag{5.8}$$

Computational flow to implement the novel SRSM with Gaussian Copula are as follows;

1. From the initial data set having lesser samples repopulate large sample set of desired quantity of x_k as explained in Section 4.4.

2. Determine the generalized polynomial chaos related to the marginals having PDF $f(x_k)$ using a Three-Term Recurrence Relation and gather the sample points for

DLF.

3. Approximate the relationship between input and output, i.e., the power and voltage in this case with the generic polynomial chaos expansion given in Equation 5.5.

4. The series coefficients c_i are obtained by solving Equation 5.8.

5. The gPC expansion is used to estimate the resulting output PDF $f(y)$ instead of running simulation for a large sample set.

5.3.   Validation of the Novel SRSM

To validate the proposed novel PLF method, the distribution network modified by adding the PV generating sources is evaluated using the novel SRSM approach. Two test networks are considered to evaluate the performance, one is the subsection of the NSELVTN, and the other is the 69-bus medium voltage test network as explained in Section 3.2 and Section 3.4 respectively. The reason for selecting the two test networks is to study the system behavior at lower and medium voltage levels, PV sources are sized according to the selected network which will impact the simulation.
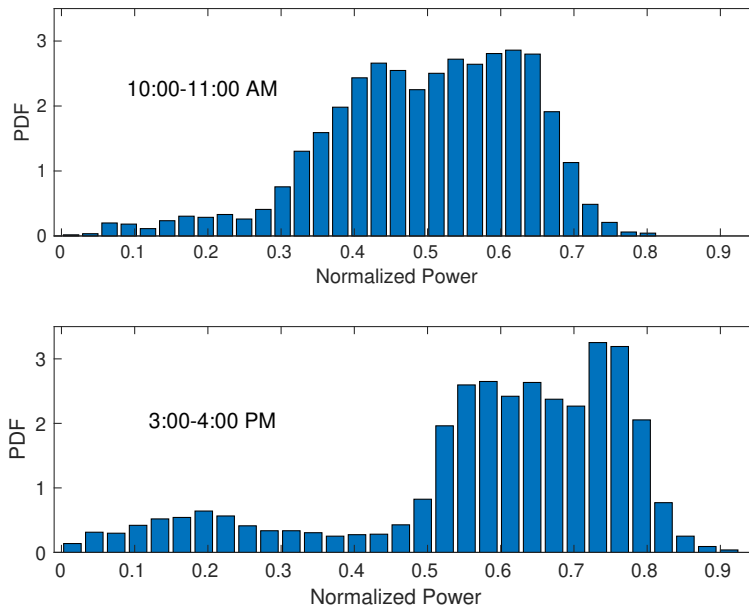
Figure 5.2: Probability density of normalized PV power delivered in two time windows of one hour each obtained from the new set of measurement data.

The measurement data to model the PV generation uncertainty are obtained from [119],

the reason for changing the data is to prove that the proposed methodology of Gaussian copula is general and any data set can be used. As explained earlier, the analysis is done on an hourly basis and the PDF of normalized active power delivered by PV at two hourly time windows is given in Figure 5.2, and the PDFs are non-standard and complex statistical distributions.

5.3.1. Impact analysis of correlated PV generators on NSELVTN

A part of the NSELVTN is used to study the impact of the correlated PV generators and to evaluate the performance of the proposed PLF methodology in comparison with others. NSELVTN operates at 50Hz, 416 V (phase-phase) in an unbalanced load distribution configuration. Four correlated PV generators of the size that are present in the distribution network, such as solar rooftops are considered, the size and the phasing of the considered PV generators are given in Table 5.1 and the respective single line diagram of the modified NSELVTN network is given in Figure 5.3.
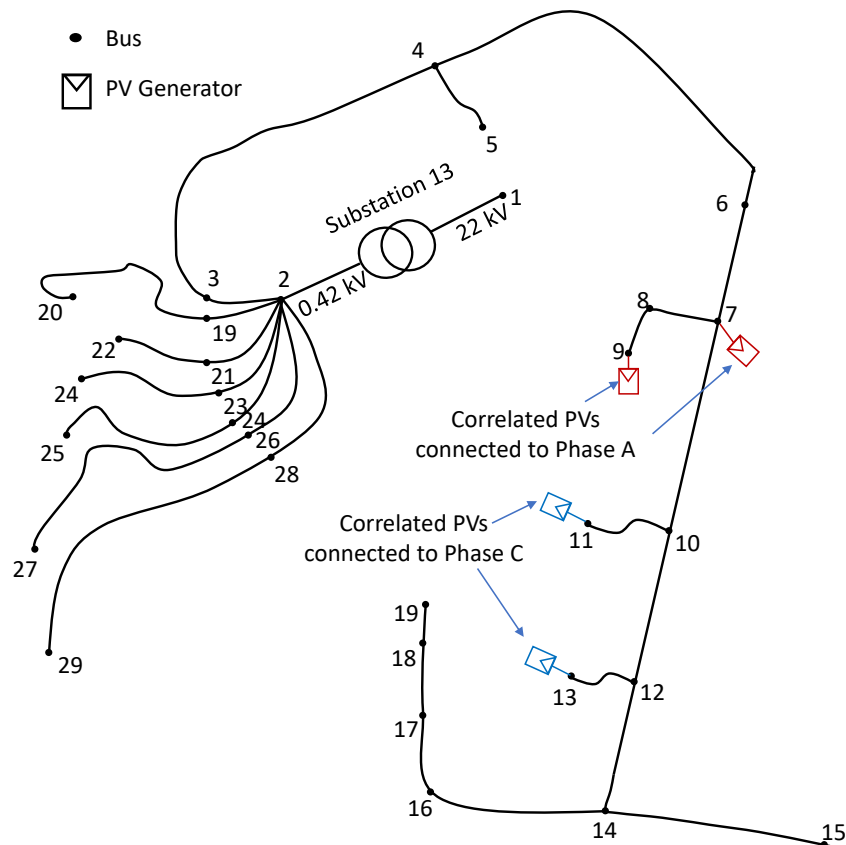
Figure 5.3: NSELVTN feeder with additional PV generators injecting power in Phase A and Phase C.

Element	Bus name	Phase	P [kW]
1ϕ PV generator 1	7	A	10
1ϕ PV generator 2	9	A	10
1ϕ PV generator 3	11	C	6
1ϕ PV generator 4	13	C	6

Table 5.1: Installed capacity of the considered PV generators in the NSELVTN and their respective phasing connections.

	PV_1	PV_2	PV_3	PV_4
PV_1	1	0.68	0.72	0.48
PV_2	0.68	1	0.94	0.87
PV_3	0.72	0.94	1	0.78
PV_4	0.48	0.87	0.78	1

Table 5.2: The correlation coefficient among PV generators empirical data for the 13:00-14:00 hour time window.

	PV_1	PV_2	PV_3	PV_4
PV_1	1	0.66	0.73	0.47
PV_2	0.66	1	0.94	0.88
PV_3	0.73	0.96	1	0.77
PV_4	0.47	0.88	0.77	1

Table 5.3: The Correlation coefficient among PV generators using the repopulated samples for the 13:00-14:00 hour time window.

The first step in the novel PLF is to repopulate the samples using Gaussian Copula, the analysis is conducted for every hour time window from 09:00-17:00, as an example hourly time window of 13:00-14:00 is discussed here. The four PV generator data are repopulated using Gaussian Copula for 10,000 samples. The correlation among the PV generators empirical data and the repopulated samples are given in Table 5.2 and Table 5.3 respectively.

The high accuracy of the proposed Gaussian Copula is proved in Figure 5.4, the marginal PDF shape of PV generators considered for a given time window of 13:00-14:00 is plotted using empirical data, and the repopulated samples, the PDFs match with the relative error $< 1\%$ also from the Tables 5.2 and 5.3, the correlation among the PV generators are well preserved with less error percentage.
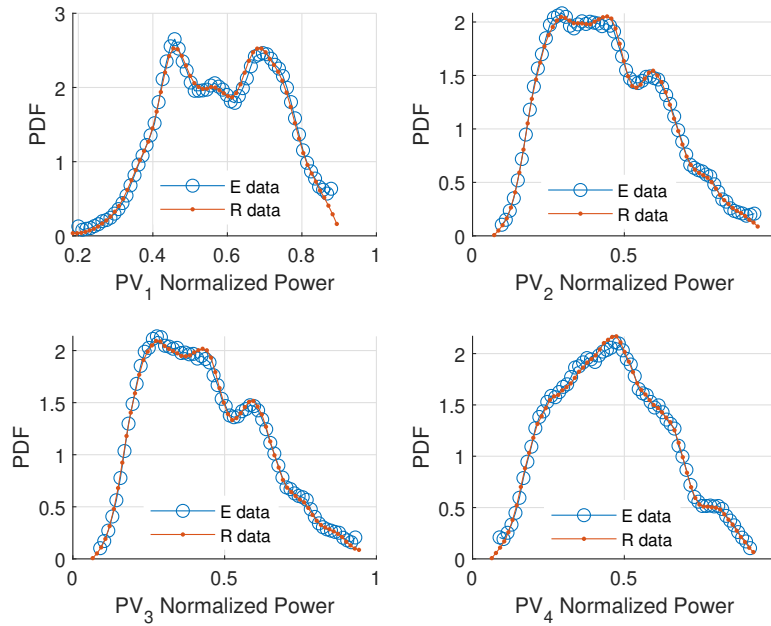
Figure 5.4: Marginal PDF of the PV generators obtained using empirical data (E-data) and the repopulated samples (R-data).

The novel SRSM proposed in this work is compared against the conventional SRSM method and the reference MC simulation with 10,000 samples. The number of samples for MC simulation standard reference is fixed by knowing the convergence of the result after running multiple simulations starting from 4,000 to 10,000 samples. Once the MC simulation has reached convergence, the standard deviation of the resulting PDF will have a slight variation. The comparison of Phase C voltage distribution of bus 12 evaluated using MC simulation with different sample sizes is given in Figure 5.5. Passing from 8,000 samples to 10,000 samples in MC simulation, the shape of the PDF had a small change, with the standard deviation of 0.411 V and 0.409 V respectively, showing the convergence of the simulation for 10,000 samples, hence this result is used as a standard reference to compare the proposed SRSM technique.
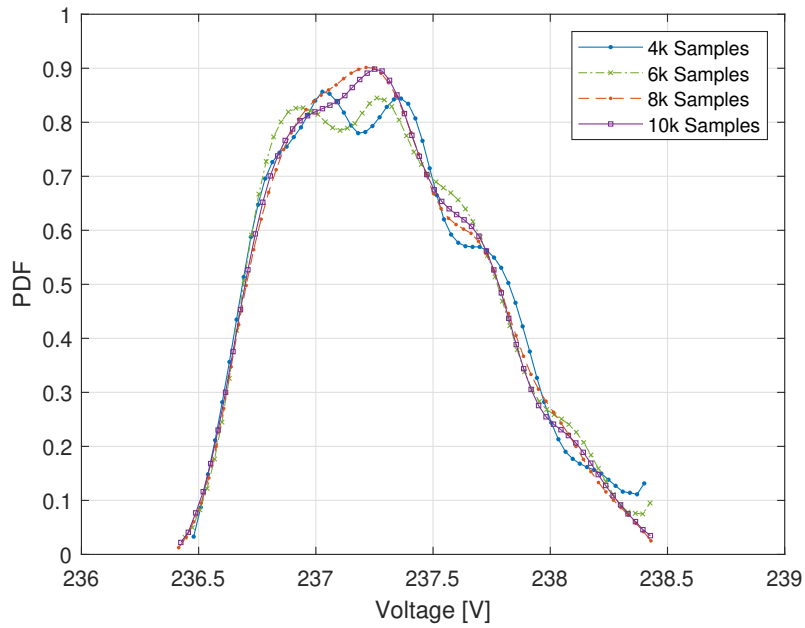
Figure 5.5: Phase C voltage distribution of bus 12 of the NSELVTN test network, evaluated using MC simulation with variable sample sizes from 4000 to 10,000 samples.
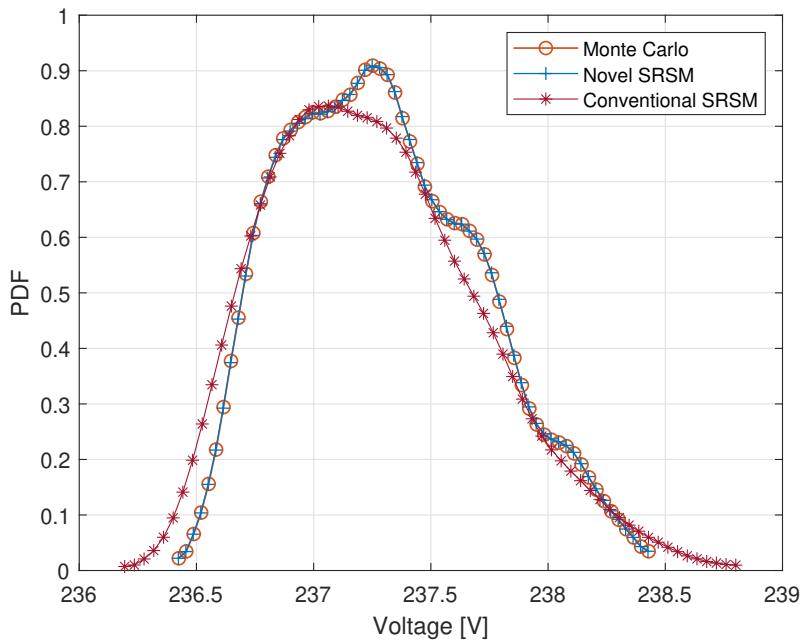
Figure 5.6: PLF evaluation Comparison of Phase C voltage distribution of bus 12 of the NSELVTN test network: a) MC simulation with 10,000 samples; b) Conventional SRSM having 5th order Hermite polynomials; c) Novel SRSM using 2nd order generalized polynomial chaos.

The comparison among all the three considered PLF evaluation techniques are given in Figure 5.6. The conventional SRSM is realized using 5th order Hermite polynomials, yet it failed to approximate the input-output relation of power and voltage accurately when the non-standard statistical distributions are used. The proposed novel SRSM method is implemented using 2nd order generalized polynomial chaos, which fits the MC simulation standard with 10,000 samples with great accuracy. The evaluation is performed on a personal computer having an Intel i7 processor of 3.3 GHz clock, the comparison of simulation characteristics among the considered methodologies is given in Table 5.4.

Si.No.	MC Simulation	Conventional SRSM	Novel SRSM
Total DLF runs	10,000	36	15
Simulation Time	291 s	16.5 s	5.9 s
Relative Error	-	$\approx 10\%$	$< 0.4\%$

Table 5.4: Comparison of simulation characteristics among 10,000 MC, Conventional SRSM and the Novel SRSM.

The PDF resulting in novel SRSM has a relative error of $< 0.4\%$ in all the points when compared to the standard MC simulation result, and it only takes 15 DLF runs. In this example, the novel SRSM introduces a remarkable $\approx 50\text{x}$ speed up in PLF evaluation compared to the MC simulation for the same accuracy.

The PDF of the voltages observed at Phase A and Phase C of bus 6 and bus 14 are given in Figure 5.7, and the voltage distributions of bus 12 for all the time windows from 09:00-17:00 are given in Figure 5.8 and Figure 5.9, the corresponding distributions of $VUF\%$ determined are given Figure 5.10. It can be noticed that the voltage magnitude is increasing with the injection of PV power, high variablity are observed in the mid day hours. Most significantly, $VUF\%$ is decreasing with an increase in PV power injection, this is because the considered network has unbalanced distribution of loads, hence the power injection by PV compensates for such an unbalance.
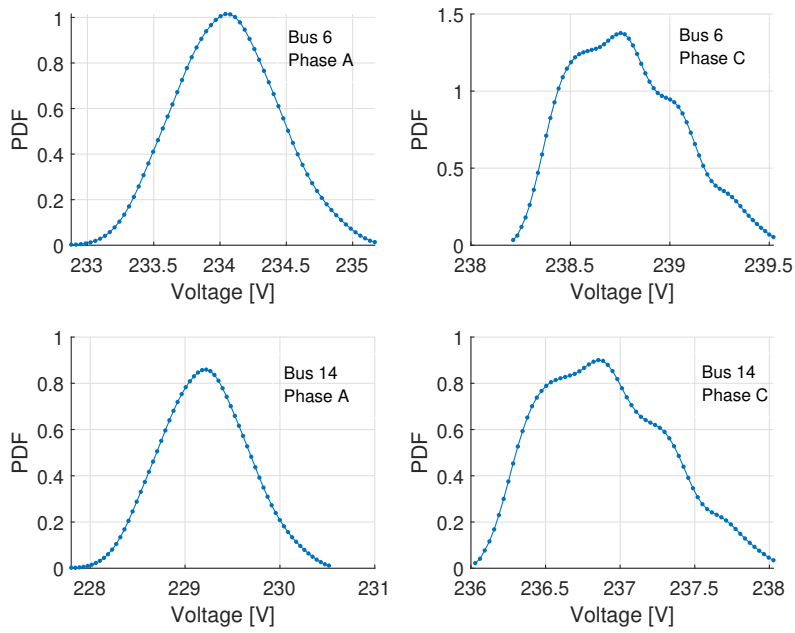
Figure 5.7: PDF of Phase A and Phase C Voltage distribution in bus 6 and bus 14 due to the presence of the correlated PV generators in the network.
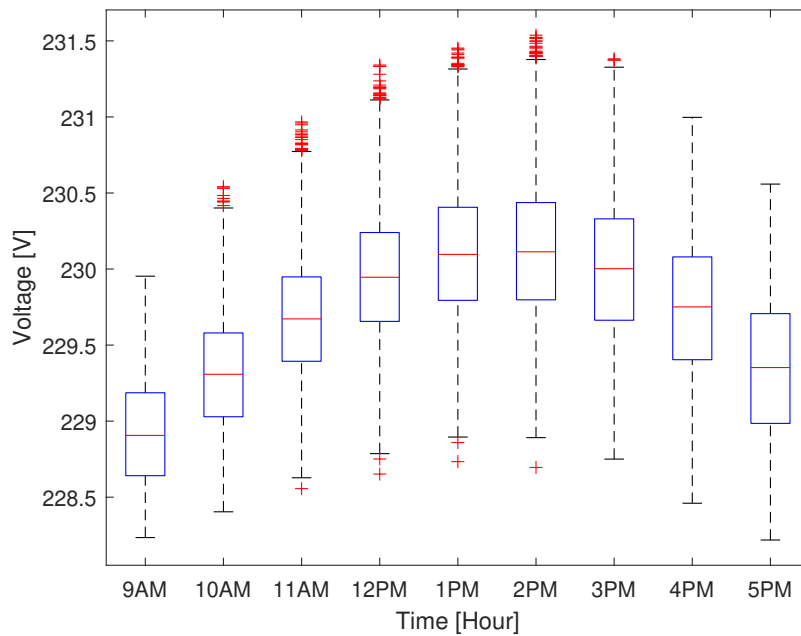
Figure 5.8: Box plot of the hourly distribution of bus 12 Phase A voltage in NSELVTN.
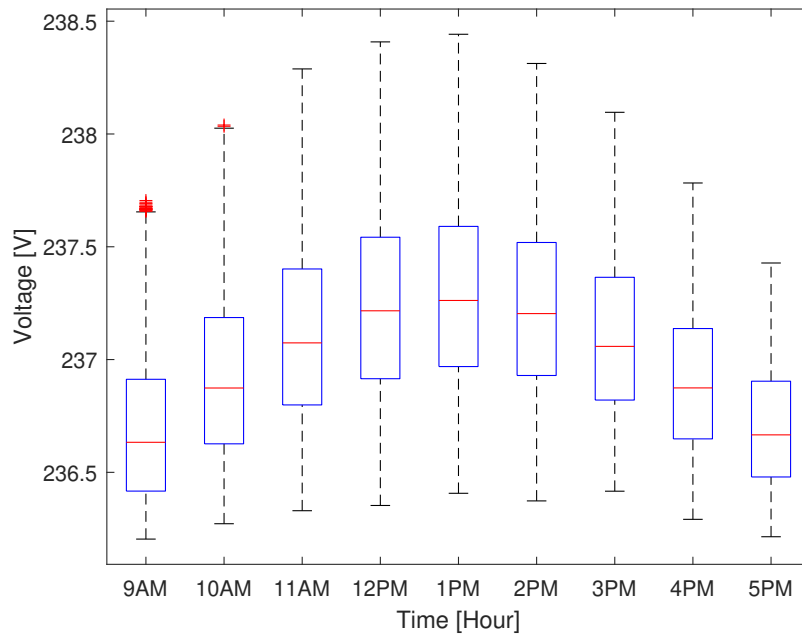
Figure 5.9: Box plot of the hourly distribution of bus 12 Phase C voltage in NSELVTN.
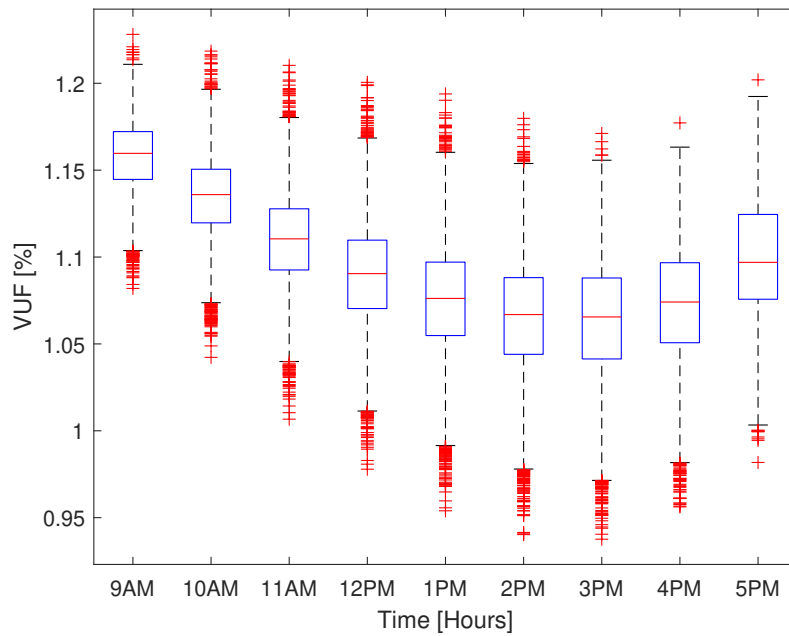
Figure 5.10: Box plot of the hourly distribution of bus 12 $VUF\%$ in NSELVTN.

5.3.2.   Study of distribution grid quality violation in presence of PV generators

The exploitation of the novel SRSM in application to rapidly evaluate the probability of the violation of network constraints due to penetrating PV sources are explained here. Several matrices of the grid quality variables are defined, and the critical value not to be violated for maintaining the good quality of the grid can be determined. To serve this purpose, two cases are considered in which the 6 PV sources are connected to Phase C of the NSELVTN as shown in Figure 5.11. In the first case, the PV penetration ratio i.e., the ratio of installed active power to the total demand power is fixed to 10.3%, and in case 2, it is increased to 17.2% the corresponding amount of PV powers are given in Table 5.5.
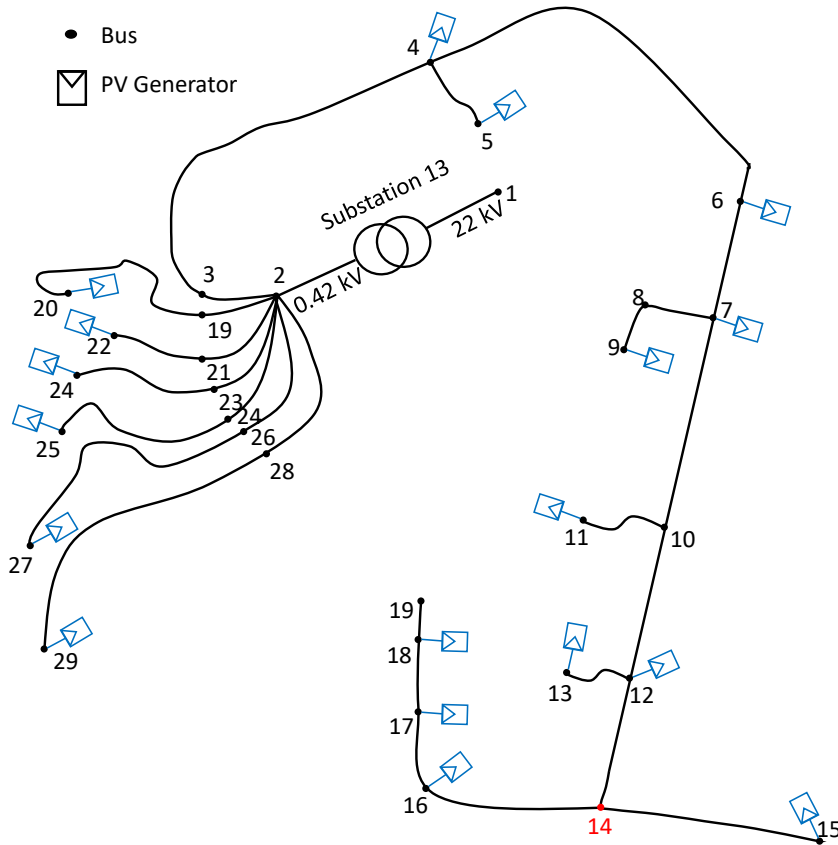
Figure 5.11: PV generators are connected to Phase C of the network for grid quality evaluation. Bus 14 marked in red is the critical bus which is an observable node in this case.

Case.No.	PV Penetration Ratio	Active Power	Total generators
Case I	10.3%	36 kW	6 units
Case II	17.4%	60 kW	6 units

Table 5.5: Pv generators and their corresponding penetration ratio in Phase C to evaluate the grid quality.

Bus 14 is considered as the critical bus in the network, the node voltages of this bus are observable to check the network quality with voltage deviation and $VUF\%$ as the indicators. The threshold considered for the voltage deviation is *0.03 pu* and the $VUF\%$ is 3%. The Figure 5.12 shows the PDF of the voltage deviation indicator, evaluated using novel SRSM considering two cases of PV penetration. The area of the two PDFs that corresponds to the voltage deviation *>0.03 pu* is determined, and the probability of voltage deviation violation are 51% and 62.6% in Case I and Case II respectively. The violation indicator of $VUF\%$ in both cases are given in Figure 5.13, the probability of violation in Case I is 45% and 52% in Case II.
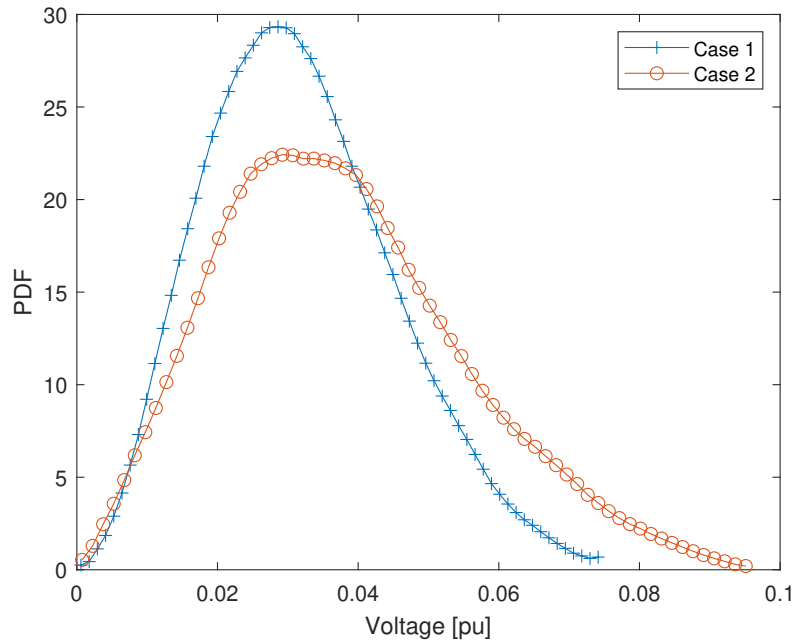
Figure 5.12: The PDF of voltage deviation in bus 14 due to the PV penetration analyzed in two cases.
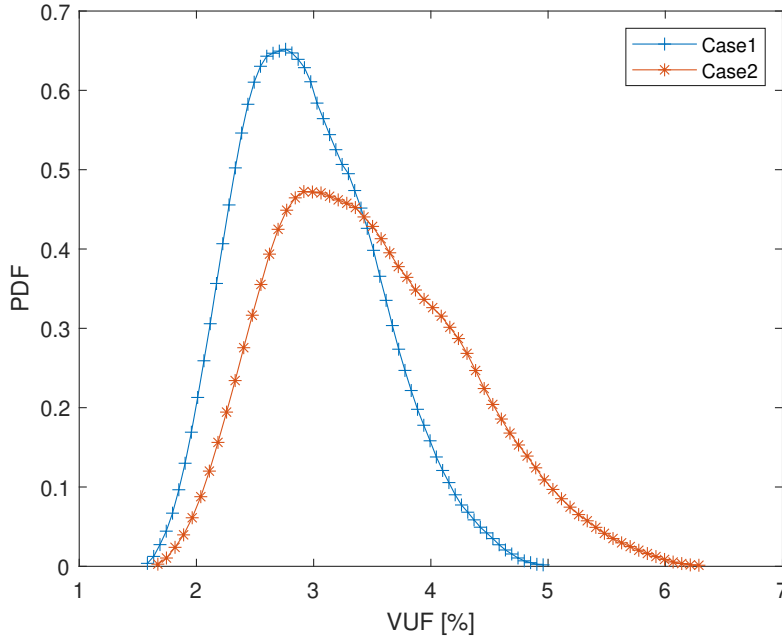
Figure 5.13: The PDF of $VUF\%$ in bus 14 due to the PV penetration analyzed in two cases.

5.3.3.　Validation using 69-bus Test Network

To further validate the proposed novel SRSM another test network operating in the medium voltage range IEEE 69-bus test network is selected. It is a balanced 3ϕ network operating at *12.7 kV, 50 Hz* and has a total load demand of *11.4 MW, 8.08 Mvar*. Four correlated PV generators with correlation coefficients as in Table 5.3 are considered at the locations shown in Figure 5.14, and the measurements are taken at bus 18 and bus 27 for the impact evaluation. The connection details and the installed capacity of the PV generators are given in Table 5.6.

The reason for selecting the bus 18 and 27 is due to their sensitivity to variation of active power injected into the other buses. Moreover, in the selected test network, bus 27 is at a longer distance from the substation transformer and tends to have high fluctuations in voltage.

Element	Bus name	Phase	P [kW]
1ϕ PV generator 1	16	A	200
1ϕ PV generator 2	20	A	200
1ϕ PV generator 3	23	C	350
1ϕ PV generator 4	24	C	350

Table 5.6: Installed capacity of the considered PV generators in the 69-bus MV test network and their respective phasing connections.
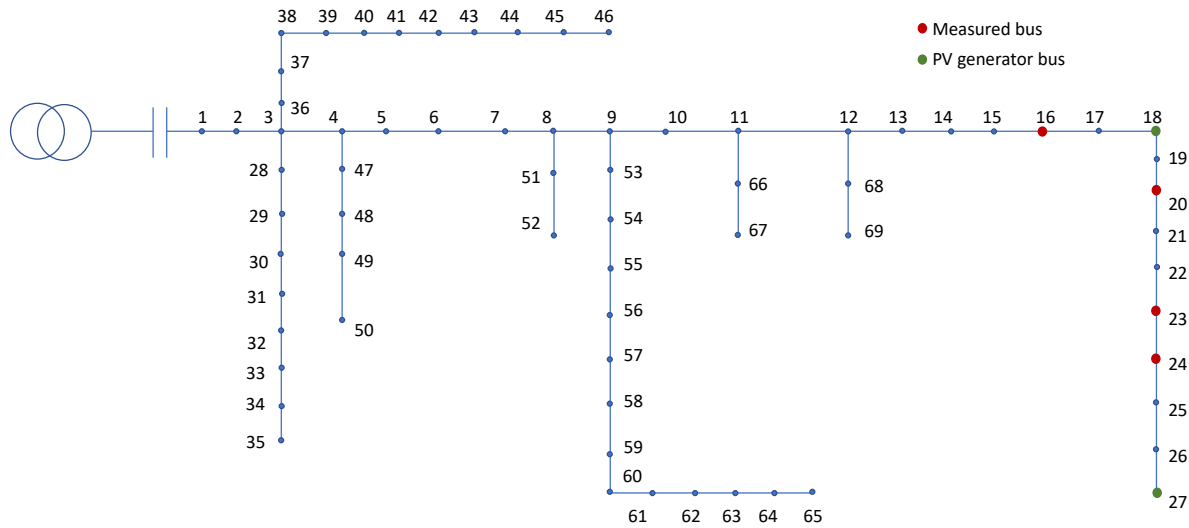
Figure 5.14: 69-bus MV test network with additional PV generators injecting power in phase A and phase C.

The comparison of the novel SRSM results with the MC simulation of 10,000 samples and the conventional SRSM with 5th order Hermite polynomial is given in Figure 5.15. The proposed method matched the MC simulation, whereas the conventional SRSM is not accurate enough to deliver results. The phase A and phase C voltages of buses 18 and 27 due to the presence of the PV generators for one hour time window 13:00-14:00 are given in Figure 5.16, and the related $VUF\%$ of both the buses are given in Figure 5.17.
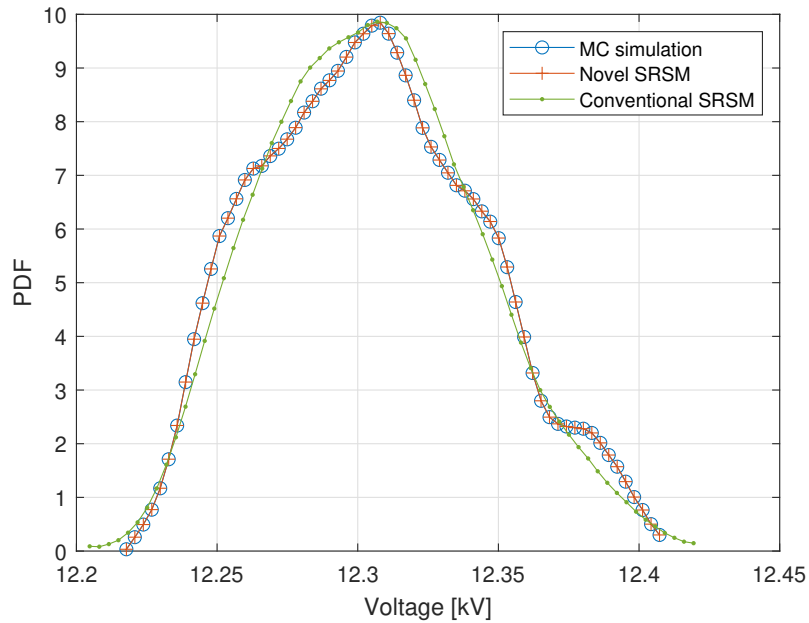
Figure 5.15: PDF of bus 27 phase C voltage, Novel SRSM compared with MC simulation of 10,000 samples, and conventional SRSM with 5th order Hermite polynomial.
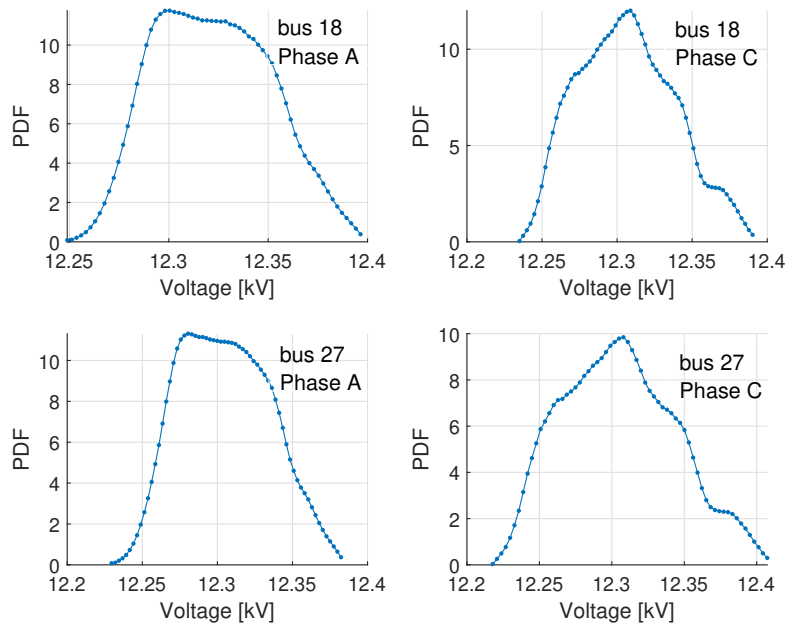
Figure 5.16: PDF of phase A and phase C voltages of bus 18 and 27 due to PV power injections.
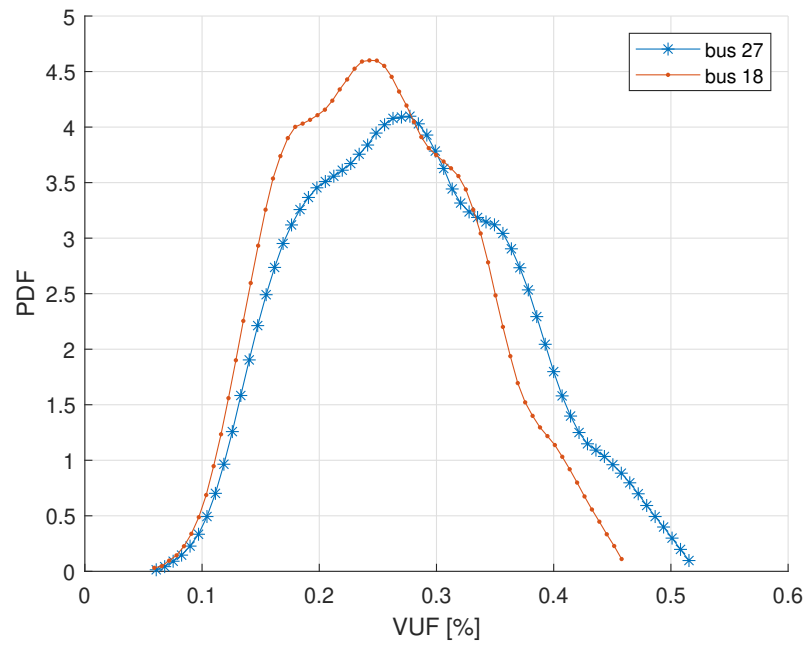
Figure 5.17: PDF of $VUF\%$ in bus 18 and 27 due to PV power injections.

6 | Data-Driven Modeling of EV Demand Uncertainty

In this Chapter, modeling of the EV charging behavior using the measurement data obtained from the charging events is described, in which the EV aggregation is integrated into the MV network to study the impact of growing EV demand in the power system network at medium voltage level. The impact analysis is performed using MC probabilistic load flow simulation using the IEEE 69-bus test network.

The motivation to adopt the proposed modeling approach is to replicate the real-life EV charging scenarios, starting from the EV charging event and moving to the charging station and aggregation of such stations connected to the MV network. By this approach, the dependence on the type of EV is eliminated but mainly depends on the level of EV charger used in the charging event.

6.1. Modeling EV Charging Station Demand

The historic charging events registered in the smart meters of the public EV charging stations (EVCS) are used, and the proposed modeling method is general that can be used with any charging event data. In this work, the public charging data set of two different providers are used. The first is from the California Institute of Technology (Caltech) [120], which contains the EV charging event data of 54 EVCS present on the campus. The second set of data is obtained from 28 public EVCS that are dispersed in Boulder city of Colorado [121]. The reason for the selected second set of data is because of diversified EV user behavior. in the first case, only academic personnel uses the EVCS and is biased toward the usage of university opening hours, whereas in the second case, the EVCS is present in public areas such as parks, cinema halls, and shopping streets. Data-driven modeling of EVCS demand is performed in two steps, one analyzing EV charging event data and the second moving to repopulation and aggregation of EVCS behavior.

6.1.1. Analyzing the EV charging events data

An EV charging event is defined as the process of the EV connecting to the EVCS and disconnecting after charging. From every data set [120, 121] more than 10,000 valid events are filtered after the removal of bad data and outliers. From every charging event vital information such as, *Start time*, and *Duration* are extracted. *Start time* is the time in a day at which the EV is connected to EVCS for charging, and the *Duration* is the total length of time the vehicle is connected to the EVCS only for charging purposes. The resulting *Start time* and *Duration* from the event data are statistically analyzed by knowing the probability distribution as given in Figure 6.1 and Figure 6.2.
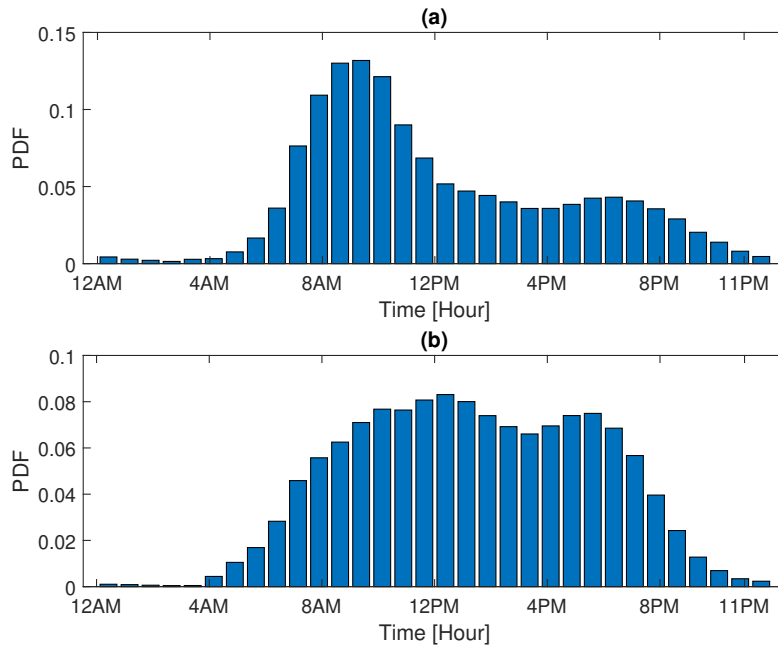
Figure 6.1: The PDF of EV connection time instants for charging, Figure (a) is obtained from Caltech data, and (b) resulted from analyzing Boulder city data.

The duration PDF of EV charging is approximately the same in both data sets, whereas distinct variation can be seen in the starting time of the EV charging event. The peak charging is expected around 09:00 hours in the first set, while it is observed twice at 12:00 and 15:00 hours in the second set of data. The variables *Start time* and *Duration* might depend on each other which is analyzed using the scatter plot shown in Figure 6.3. From Figure 6.3(a) of Caltech data, it is evident that certain hours of starting time, such as 07:30-11:30 shows the Independence of the duration as vehicles connected at this duration are charged from minimum to maximum duration of time. While the other instants of starting time are dependent on the duration. Figure 6.3(b) shows no dependence among

the variables in the case of Boulder city-data.
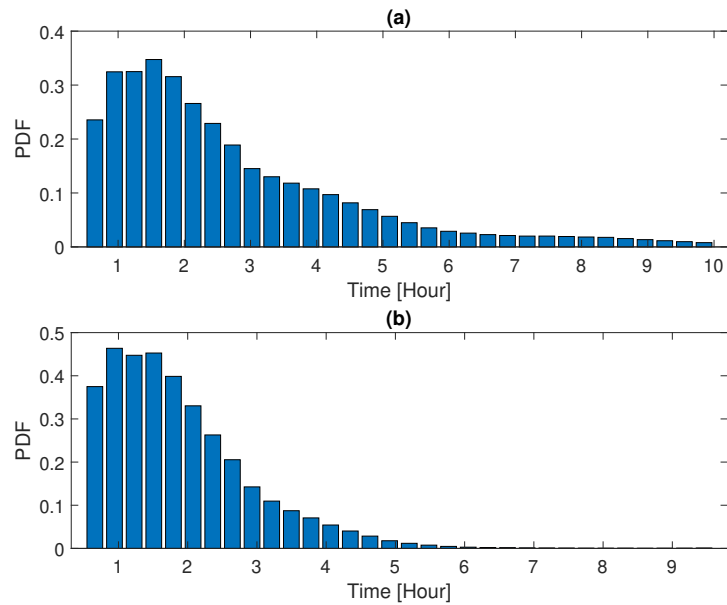
Figure 6.2: The PDF of EV charging duration, (a) obtained using Caltech data and (b) is obtained using Boulder city data.
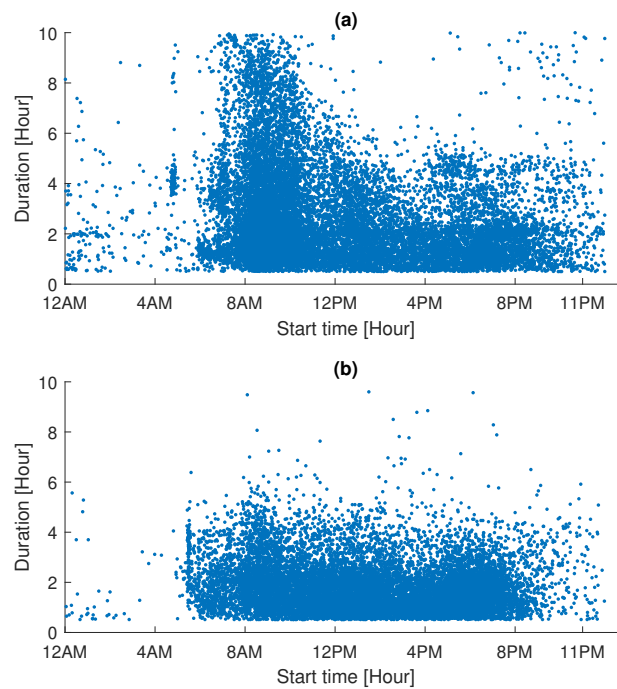
Figure 6.3: Scatter plots of the duration of charging versus the start time of charging (a) obtained using Caltech data, and (b) obtained using Boulder city data.

Using the energy delivered by EVCS and the *Duration*, the power level of the charging event is extracted, it is important in deciding the power level of the charging event as EV is integrated into the network as additional load demand. Figure 6.4 shows the PDF of active power delivered by EVCS to the EV during charging events.
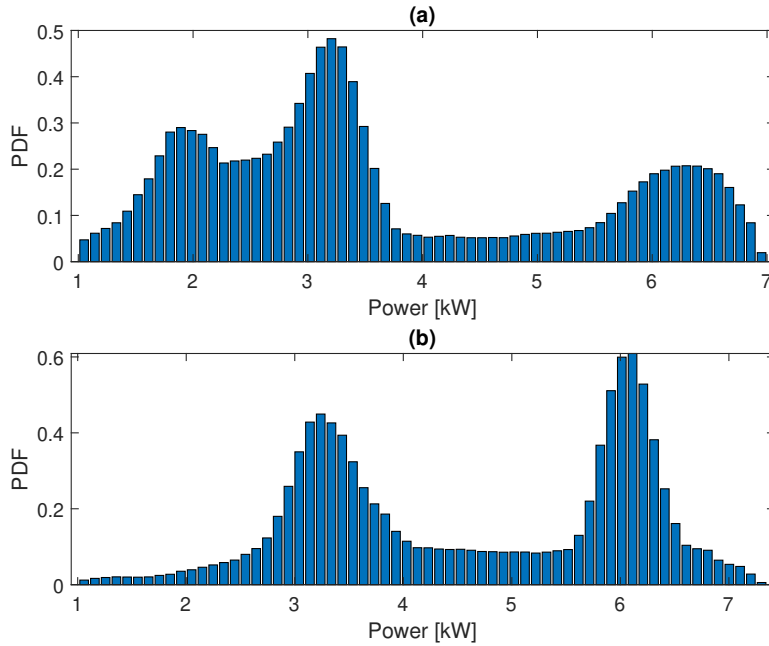
Figure 6.4: PDF of active power delivered consumed by PV in the charging events (a) obtained using Caltech data, and (b) obtained using Boulder city data.

6.1.2. Re-population: EV charging profiles to EVCS aggregation

The data analyzed is of EV charging events obtained from the EVCS ideally, EVCS are connected to the MV distribution network for power consumption during charging. The goal is to model this behavior of EVCS to study its impact on the distribution network at the MV level. The idea is to aggregate the EVCS behavior connected to the same MV node of the distribution network so that the EV behavioral uncertainty is grouped, making it easy to implement in the PLF process, the aggregation configuration employed in this work is shown in Figure 6.5.

In the modeling framework, EV charging events are repopulated, and the EV charging profiles for each charging station is generated. Later, these EVCS are aggregated by summing the load demands of each EV for respective time instant to draw the power from the MV network for the modeled charging event. To include any dependency among the

variables such as *Start time, duration* and *power level* an expeditious method is followed by forming a lookup table from the event data that carries the information associated, and a random number is called the information associated to the resulted random number is assigned to the charging event.
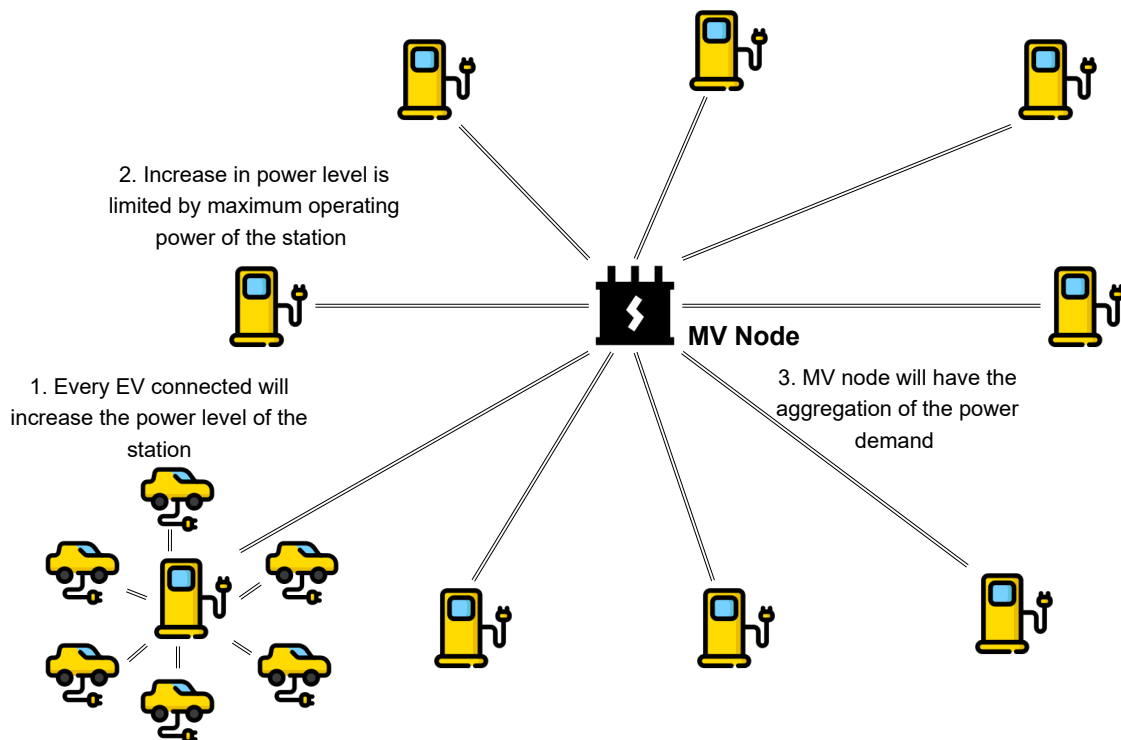
Figure 6.5: Aggregation of EVCS at MV node, as the EV numbers in the area increase, EVCS increases thereby increasing the power demand level at the MV node.

To generate the EV charging profiles associated to a charging station, following assumptions are made;

1. EVCS will deliver the charge to the EV in constant power level in an event. This power level might change in the next event and will follow the distribution given in Figure 6.4.

2. For one EV aggregator a maximum number of 'J' EVs are operated.

3. Each EVCS has a limit of *50 kW* power capacity rating, any EV previously connected requesting the power exceeds the limit of EVCS rating should wait until the EVCS is available to deliver the energy at set power.
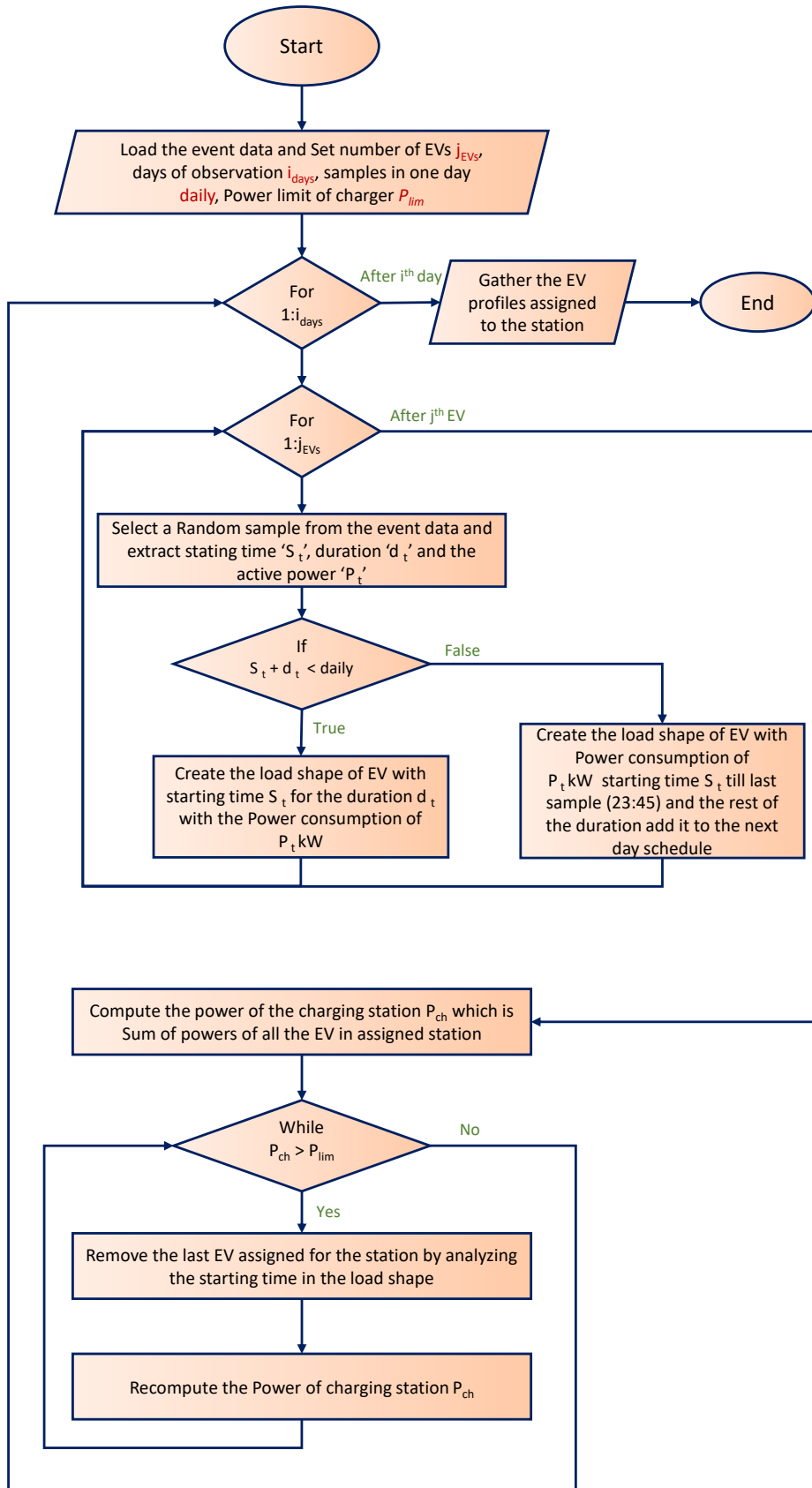
Figure 6.6: Algorithm to generate the EV loadshape from the EV charging event measurement data.

A total of $'J'_C$ EVs are assigned for each aggregator dispersed over $'C'$ charging stations, the EV profiles are created as daily load shapes with a sample period of *15 minutes*. Hence, for one day, 96 samples are present, and the detailed algorithm for this implementation is given in Figure 6.6.

A random number among the index of the lookup table is called, *Start time*, *duration*, and the *power level* corresponding to the indexed random number details are extracted to generate the new EV profile. This process is repeated for the desired number of samples required for MC simulation, the samples are limited by using the repopulation for *'i'* number of days.
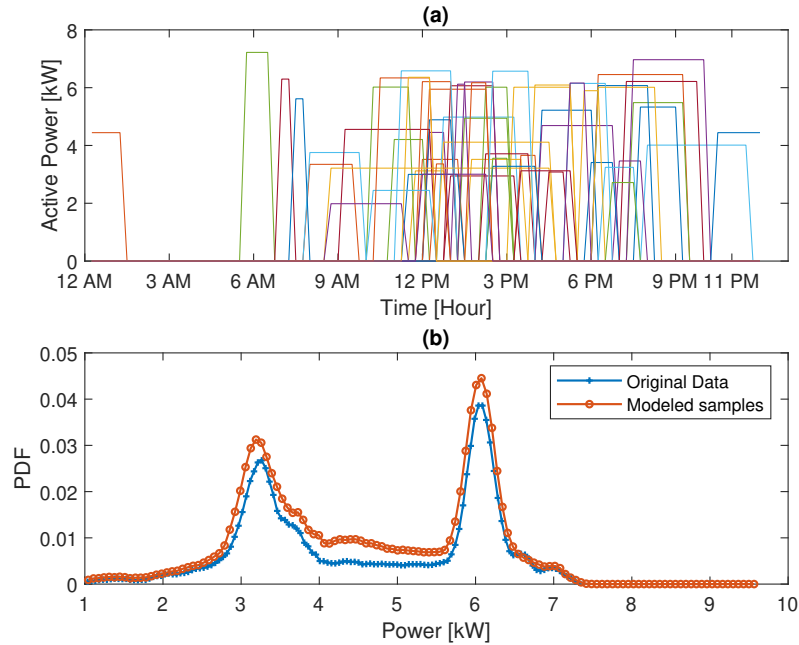
Figure 6.7: (a) Shows the load profiles of 6 EVs connected to a charging station, (b) shows the PDF of average active power of the EV charging event data of the empirical data compared to the modeled data.

Figure 6.7(a) shows the load profiles of 6 EVs connected to a charging station over a daily time period generated for *10 days* using the proposed modeling methodology. Figure 6.7(b) shows the PDF of the average power rating of the EV charging event in an EVCS from the point of view of a single EVCS, the modeled EV charging behavior is in good agreement with the measured charging event data.
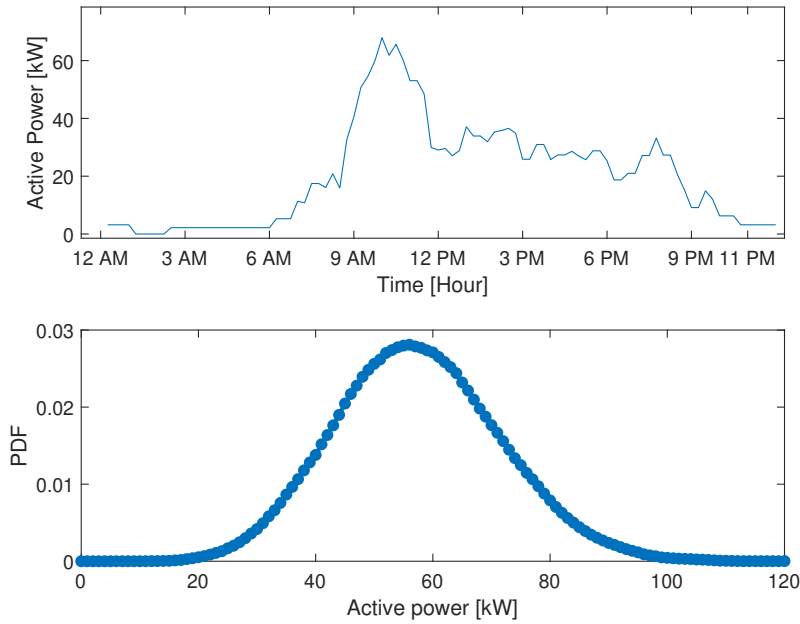
Figure 6.8: (a) Shows the daily load profile of a aggregated EVCS demand at a MV node, (b) shows the PDF of aggregated EVCS demand for time window 09:00-11:00 seen at MV node.

With the increase in EV, the power level of EVCS increases, cumulating the power levels of all the EVCS connected to a single MV bus, which will result in the charging station demand as a continuous random variable. Figure 6.8(a) shows the daily load profile of such aggregation of EVCS seen at an MV bus. Whereas, Figure 6.8(b) shows the PDF of the active power demand of the aggregated EVCS at the MV node for the time window of 09:00-11:00 obtained using the proposed modeling methodology.

Depending on the scenario of investigation, as many stations as required can be aggregated per phase, the power demand distribution as shown in Figure 6.8(b) will be the vector of uncertainty represented by $P_n[t_m, \vec{\xi}]$.

6.2.   Impact analysis of EVCS on the MV distribution network

To study the impact of the EVCS present in the power system network, IEEE 69-bus MV test feeder network is selected. As explained earlier, it is a three phase system operates at *50 Hz* frequency and *12.7 kV* medium voltage rating with a total load demand of *11.4 MW* and *8.08 Mvar* in all the three phases.

10 EVCS load demand is aggregated at an MV node with the aggregation configurations as shown in Figure 6.5, the aggregated EVCS demand are connected to the bus 16, 19, 21, 24, 26, and 27. The reason for selecting these buses is the voltage drop at the tail of the network is higher, and they are sensitive to power fluctuations, more importantly, bus 27 is the critical bus present in the network. To study the network unbalance, all the connected to Phase A of the MV node. The voltage measurements are recorded at bus 18 to study the effects due to EVCS aggregation present in the rest of the network. The detailed Figure 6.9 shows the EVCS aggregation and the measurement bus location in the 69-bus MV test network.
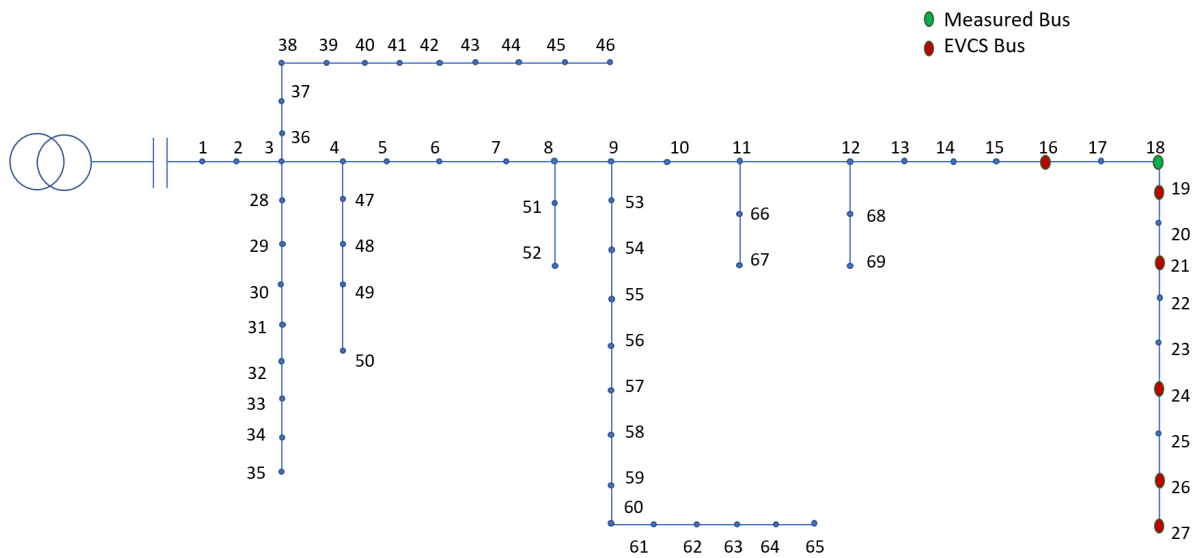
Figure 6.9: 69-bus medium voltage test network with the aggregated EVCS load and the measurement bus for impact analysis using MC simulation.

The stochastic impact analysis is performed by selecting three different time windows, the first window is between 07:30-11:30, the second window is at 12:00-16:00, and the last window is at 16:00-20:00. The selection of the time windows depends on the characteristics of the *start time* of EV charging event data, the first and last window represents the peak charging time and second window presents the valley of the *start time* shown in Figure 6.1.

In the MV network, it is also necessary to incorporate the load behaviour [122], the static loads present in the MV test network are multiplied with the scaling factor of *0.65 PU, 0.8 PU, and 1.0 PU* in the first, second and the third window respectively.
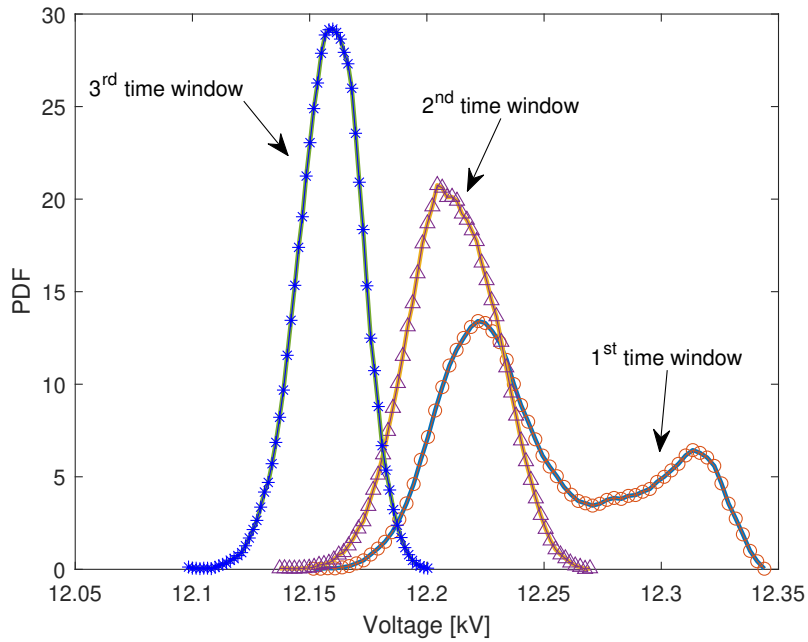
Figure 6.10: PDF of phase A voltage distribution of bus 18 in 69-bus MV distribution test network at three different time windows. 1st window is at 07:30-11:30 and 65% installed load capacity, 2nd window at 12:00-16:00 and 80% installed load capacity and the last window is at 16:00-20:00 with 100% installed load capacity.
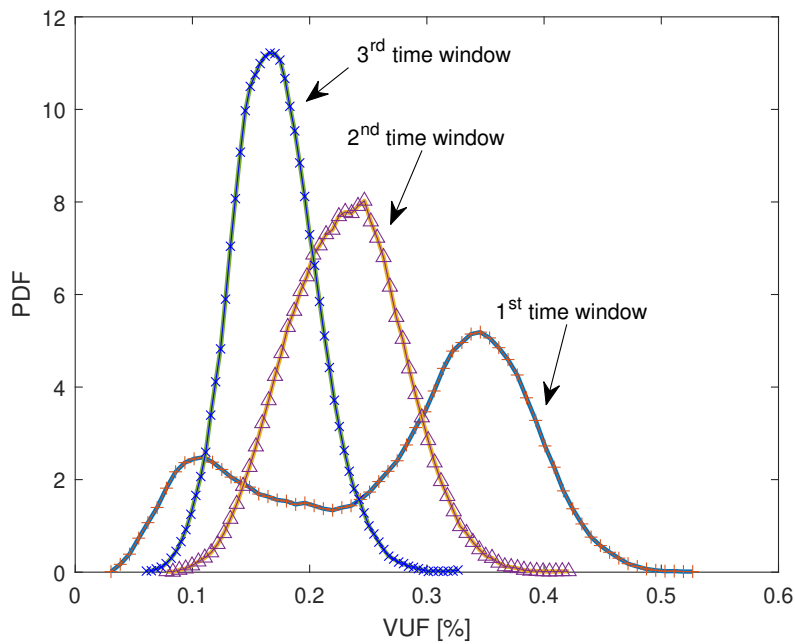
Figure 6.11: PDF of $VUF\%$ distribution of bus 18 in 69-bus MV distribution test network at three different time windows.

MC simulation is performed with 10,000 samples, the resulting voltage distribution PDF measured at bus 18 at three different time windows are given in Figure 6.10. It can be noted that in the 1st time window the installed load capacity is minimum (0.65 PU of installed capacity) and, the EV charging demand is maximum, hence the variability of the voltage is maximum in this time window causing greater impact on the grid voltage deviation with maximum deviation of *200V*.

The impact of the EVCS on the network unbalance at bus 18 is found by determining the $VUF\%$ as given in Figure 6.11, maximum unbalance is seen in first time window with wide interval, the unbalance of other two time windows lies inside the interval of the first one signifying the first time window is critical in managing the EV behaviour in the MV network.

7 | Impact Analysis of PV Generation and EV Demand Uncertainty

In the previous Chapters, the impact on the distribution system due to uncertainties in PV generation and the EV load demand is analyzed. Whereas in this Chapter, both the uncertainties are considered together, and the collective impact of these uncertainties on the LV distribution network is analyzed. The PV power injection uncertainty should be mitigated by the EV demand present in the network, Data-driven modeling approach is adopted to model PV and EVCS load demand, and the PLF analysis is performed using the MC simulation technique.

7.1. IEEE European Low Voltage Test Network

The test system selected for the analysis is the IEEE European low voltage test network [108, 109], it is a moderately meshed unbalanced system that represents a real distribution network. It is powered using an *800 KVA 11 kV/0.416 kV* transformer operating at *50 Hz* frequency. It has a total of 55 loads with the installed capacity of *328.14 kW* distributed in the network composed of 906 buses, Phase A has a loading of *112 kW*, Phase B with *127.66 kW* and followed by Phase C with *89.18 kW*.

The PV sources are integrated with the PV penetration ratio, which is the ratio of the total installed capacity of the PV to the total load capacity. A PV penetration of *40%* in a modern distribution system is reasonably acceptable, as the solar rooftops are considered, this PV of total installed capacity of *135 kW≈40%* PV penetration ratio is split into *9* different correlated PV generators with an installed capacity of *15 kW* each. Considering the area of the distribution network and the residential load clusters, it is reasonable to assume the presence of *60* EVs dispersed over *10* EVCS in the network. The modified test network with the addition of correlated PV generators and the EVCS is shown in Figure 7.1 with modifications as mentioned in Table 7.1.
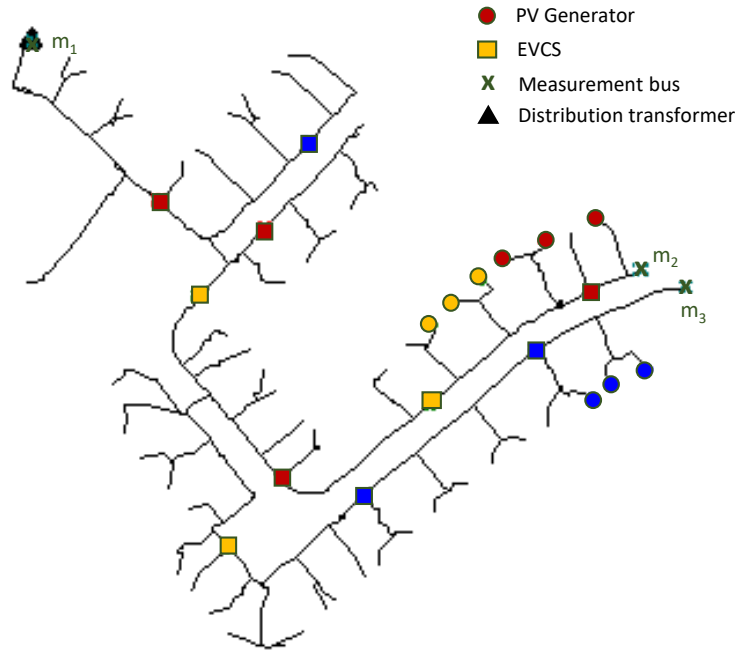
Figure 7.1: Modified IEEE low voltage test network with *9* correlated PV generators and *10* EV charging stations. The colors red, yellow, and blue represent the three phases of the electric network.

Element	Bus no.	Phase	Element	Bus no.	Phase
EVCS 1	59	A	EVCS 8	596	A
EVCS 2	123	B	PV Source 3	611	B
EVCS 3	133	A	PV Source 4	639	A
EVCS 4	187	C	PV Source 5	676	A
EVCS 5	332	A	PV Source 6	682	A
EVCS 6	460	B	EVCS 9	707	C
PV Source 1	502	B	EVCS 10	794	C
PV Source 2	562	B	PV Source 7	886	C
EVCS 7	594	B	PV Source 8	899	C
PV Source 9	906	C			

Table 7.1: Integration of the PV sources and the EVCS in the IEEE LV test network. Installed capacity of each PV source is of *15 kW* and the EVCS is of *27 kW*.

7.2. Data-Driven modeling of PV generation uncertainty

The impact analysis of EV and PV uncertainties is performed using MC simulation with suitable time windows. The measurement data [119] of the PV generators present in the close vicinity is used in the uncertainty quantification. The data gathered is for one calendar year and sampled every *5 minutes*, with this data set an appreciable quantity of samples are available, yet it is not enough for MC simulation. The Gaussian copula is used to repopulate the samples preserving the correlation. The flowchart for repopulating the samples using Gaussian copula is given in Figure 4.6. The stochastic analysis is performed in hourly time windows. As an example, the PDF obtained for two PV generators for the time window of 12:00-13:00 hour is shown in Figure 7.2. The PDF of the repopulated samples matches the PDF of the empirical data showing the accuracy of the repopulation methodology. The correlation coefficient among the generators for the same time window after repopulation is given in Table 7.2, it is verified that the correlation coefficient is preserved with an error $< 0.5\%$.
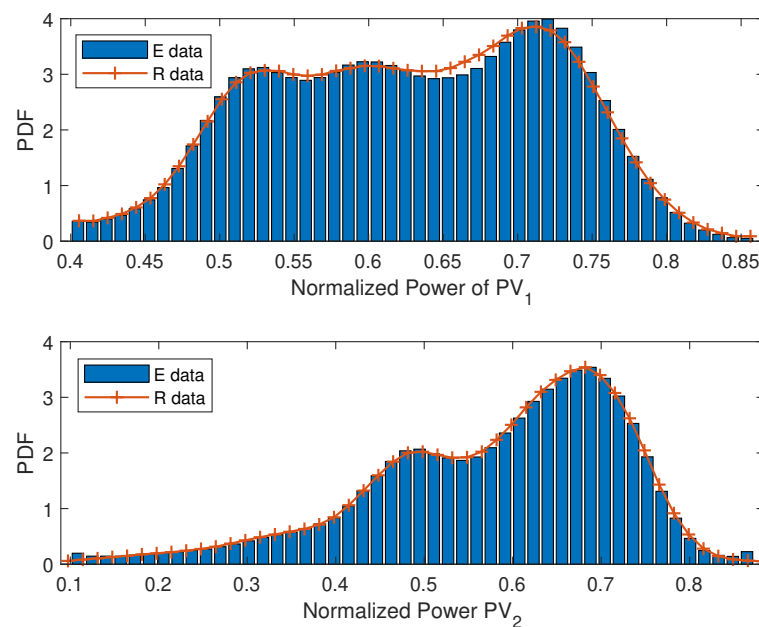
Figure 7.2: Normalised active power PDF of two PV generators analyzed for an hourly time window of 12:00-13:00.

	PV_1	PV_2	PV_3	PV_4	PV_5	PV_6	PV_7	PV_8	PV_9
PV_1	1	0.4236	0.6933	0.7235	0.8724	0.5167	0.7824	0.7784	0.9349
PV_2	0.4236	1	0.7355	0.7064	0.5980	0.7600	0.7139	0.7025	0.5069
PV_3	0.6933	0.7355	1	0.9783	0.7420	0.9404	0.9673	0.9674	0.6940
PV_4	0.7235	0.7064	0.9783	1	0.7678	0.9105	0.9716	0.9810	0.7203
PV_5	0.8724	0.5980	0.7420	0.7678	1	0.5622	0.8380	0.8235	0.9276
PV_6	0.5167	0.7600	0.9404	0.9105	0.5622	1	0.8691	0.8701	0.5118
PV_7	0.7824	0.7139	0.9673	0.9716	0.8380	0.8691	1	0.9876	0.7893
PV_8	0.7784	0.7025	0.9674	0.9810	0.8235	0.8701	0.9876	1	0.7780
PV_9	0.9349	0.5069	0.6940	0.7203	0.9276	0.5118	0.7893	0.7780	1

Table 7.2: The correlation coefficient among PV generators repopulated samples for the 13:00-14:00 hour time window.

7.3. Data-Driven modeling of EVCS demand

The IEEE European LV test system is modified with the addition of *10* EVCS, and the load demand of these stations is modeled using the EV charging event data obtained from the smart meters of the charging station.
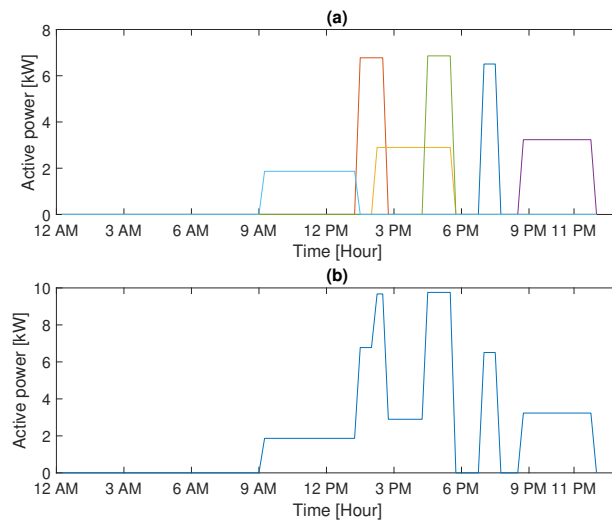
Figure 7.3: (a) is the day EV charging profiles of *6* EVs scheduled to charge at EVCS1, (b) is the time series active power demand of EVCS1 in one day.

Two different data sets are used, one from Caltech [120] and the other from Boulder city Colorado [121]. The vital information such as *Start time, Duration* and the *Energy consumed* is extracted and analysed as explained in Section 6.1.1. The charging profiles of 60 EVs distributed over 10 EVCS are produced using the methodology as explained in the flowchart of Figure 6.6. The EV charging profiles for *6* EVs scheduled to charge in *EVCS1* is given in Figure 7.3.

The analysis is conducted in a four-hour time window, hence the active power demand of the single EVCS for a time window of 07:30-11:30, 12:00-16:00, and 16:00-20:00 are given in Figure 7.4. The other EVCS considered tends to follow the same behavior.
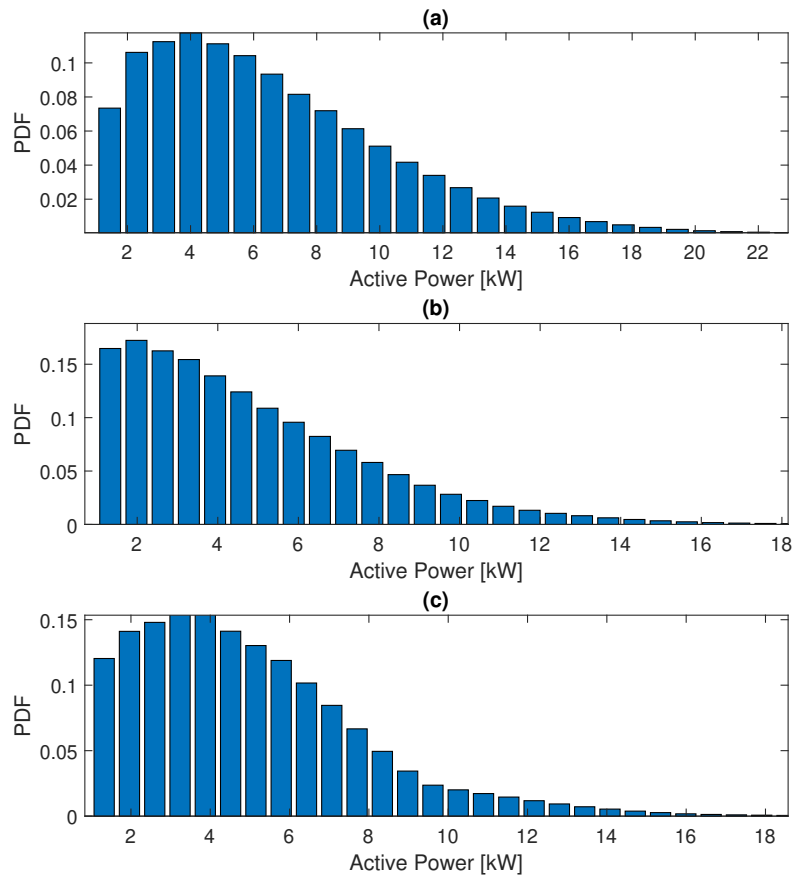
Figure 7.4: (a) is the PDF of a single EVCS active power demand in kW for the first time window 07:30-11:30, (b) is the active power PDF in kW for the second time window 12:00-16:00 while (c) shows the active power demand in the third time window 16:00-20:00.

7.4.   Impact analysis on the LV network

To study the impact on the LV network due to the injection of PV power and the EV charging loads, PLF analysis using MC simulation is performed. The analysis is concerned with three-time windows as mentioned in Figure 7.4, the installed load capacity of the network is varied according to the selected time window. In the first time window of 07:30-11:30, the loads are maintained at *0.65 pu* times of installed capacity. In the second time window of 12:00-16:00, the loads are at *0.8 pu* times the installed load whereas, in the last time window of 16:00-20:00 the loads are at full capacity i.e., *1.0 pu* times the installed capacity. The EV and PV uncertainties are extracted for these time windows and applied as additional elements of the network. Since the Voltages are subjected to high variability at the end of the network, the measurements are taken at *bus 617* point m_2 and *bus 881* m_3, and also at the substation bus m_1 as shown in the Figure 7.1.

7.4.1.   Analysis using Caltech data

Initially, the impact analysis was performed using Caltech data presented. The voltage measured at point m_2 of the LV network, the voltage variations are due to uncertain EV and PV power injections and is given in Figure 7.5.

Figure 7.5: Three phase voltage distributions and the $VUF\%$ measured at the point m_2 of the LV network. (a), (b) and (c) shows the voltage distributions measured in 1st, 2nd and 3rd time windows respectively. (d) shows the $VUF\%$ measured in all the time windows.

Phases A and B show wide variability than Phase C, it is because the PV generators of Phase A and B are nearer to the measuring point and the PV generators connected to the Phase C are at the far end of the electrical network from the measuring point m_2. The same phenomena can be observed in the case of measurements obtained at point m_3, a high variability is observed in Phase C than in the other phases.

The most impact on the voltage quality of the network is caused in the 3rd time window i.e., at 16:00-20:00 operating at 1.0 PU of installed capacity. The loading is higher, the EV demand is higher and less injection of power from the PV source, and the mean $VUF\%$ is seen up to 1.5% highest compared to 1.3% and 1.01% in the first and second-time windows respectively.

A significant improvement in the voltage variations and the $VUF\%$ is seen in the second time window operating with 0.8 pu of installed load capacity, and it is due to the fact that the EV load demands are compensated by the PV power injections, whose uncertainties are mitigating each other and causing a lesser burden on the distribution network.

Figure 7.6: Three phase voltage distributions and the $VUF\%$ measured at the point m_3 of the LV network. (a), (b) and (c) shows the voltage distributions measured in 1st, 2nd and 3rd time windows respectively. (d) shows the $VUF\%$ measured in all the time windows.

The point m_3 of the LV network suffers from an under voltage problem that is seen in Figure 7.6, it is due to its farthest position from the substation. Much improvement in

voltage variations is seen in the third time window, where the effect of EV and PV is minimal, and the network is operating at maximum installed load capacity.

In the substation bus i.e., at the measurement point m_1 of the LV network the voltage variation seen is significantly less and it is in the order of 2.5 V, but an import factor to observe is the current distribution, the amount of power that is delivered from the MV network to this LV distribution network can be obtained. The three phase currents and the $VUF\%$ measured at m_1 in all the time windows is given in Figure 7.7.

Figure 7.7: Three phase current distributions and the $VUF\%$ measured at the point m_1 of the LV network. (a), (b) and (c) shows the current distributions measured in 1st, 2nd and 3rd time windows respectively. (d) shows the $VUF\%$ measured in all the time windows.

The Figure 7.7 shows the complete detail of current distributions in three time windows, most of the power is utilized from the MV network in the third time window where the loading is maximum and very less power generation from the PV sources. Though the loading in second time window is increased from 0.65 pu to 0.8 pu, due to injection of power from PV sources and the less EV activity in this time window, lesser current flow is measured in this window than in the first time window.

The above analysis is conducted using Caltech data, where the EV user behaviour is biased according to university working hours. Hence a Boulder city data is used in the next analysis to verify how the change in EV data set will impact on the results.

7.4.2.   Analysis using Boulder city data

In this analysis, Boulder city data is used where the EV user behavior is not confined to one profile such as a university. The analysis is performed in the same three-time windows, with 0.65 pu, 0.8 pu, and 1.0 pu of installed load capacity in the first, second and third-time windows respectively. The voltage distributions measured in all time windows at the measuring point m_2 are given in Figure 7.8.
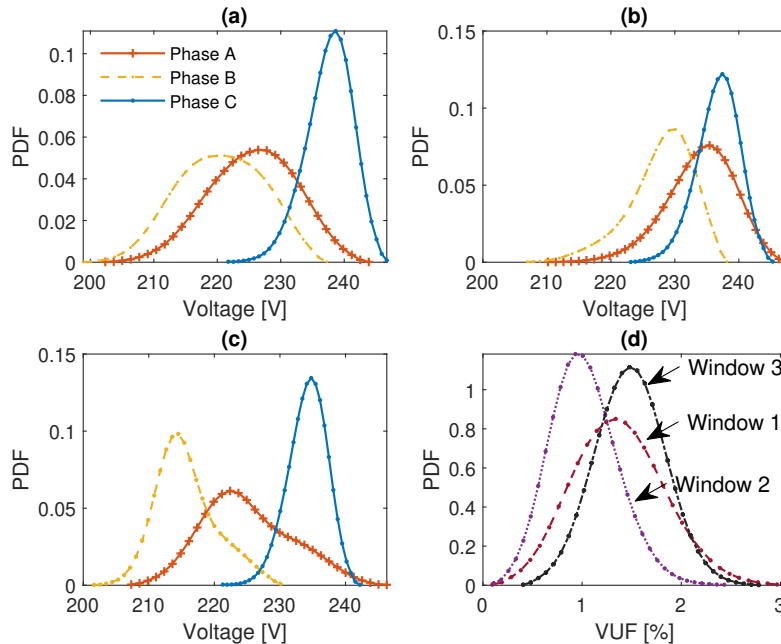

Figure 7.8: Three phase voltage distributions and the $VUF\%$ measured at the point m_2 of the LV network. (a), (b) and (c) shows the voltage distributions measured in 1st, 2nd and 3rd time windows respectively. (d) shows the $VUF\%$ measured in all the time windows.

The change in the EVCS user behavior did not impact much on the LV network measured at the point m_2, except for a slight increase in the variability of voltages in the 2nd time window, similar impact characteristics are observed here. The 3rd time window has the higher $VUF\%$ with a mean value of 1.52 V while, the 1st and 2nd time windows have a mean value of 1.31 V and 1.04 V respectively.

The most vulnerable section of the selected test network is the measuring point m_3, the voltage distributions and the $VUF\%$ analyzed using the boulder city data are given in Figure 7.9, wider voltage variability is seen in Phase C in all the time windows. The current distributions measured in point m_1 are given in Figure 7.10 whose results are similar to the previous case.
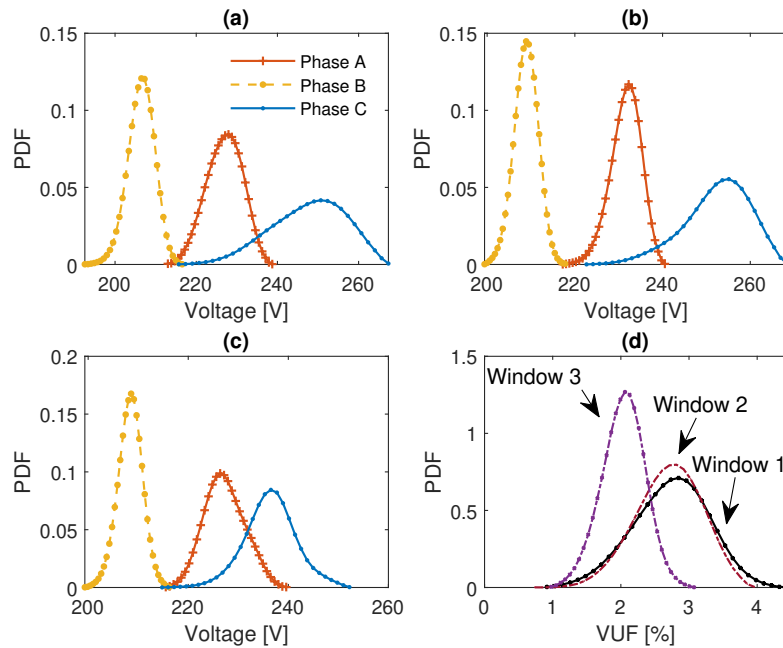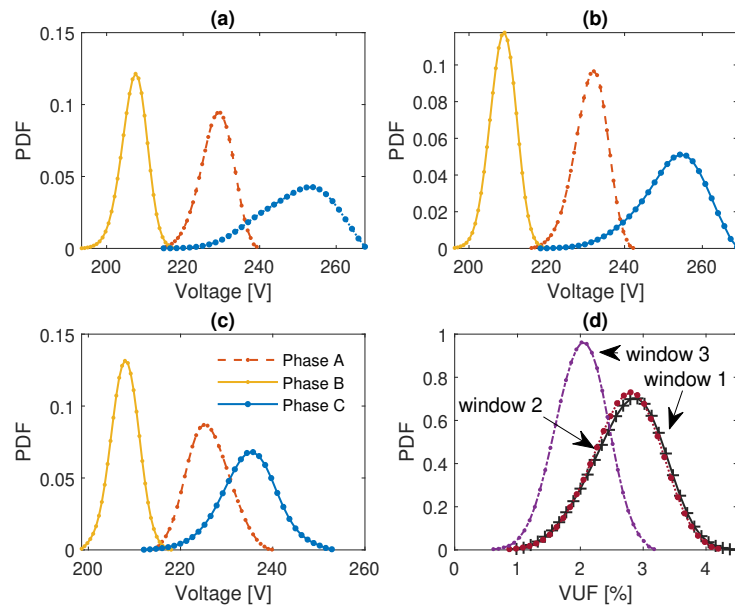
Figure 7.9: Three phase voltage distributions and the $VUF\%$ measured at the point m_3 of the LV network. (a), (b) and (c) shows the voltage distributions measured in 1st, 2nd and 3rd time windows respectively. (d) shows the $VUF\%$ measured in all the time windows.
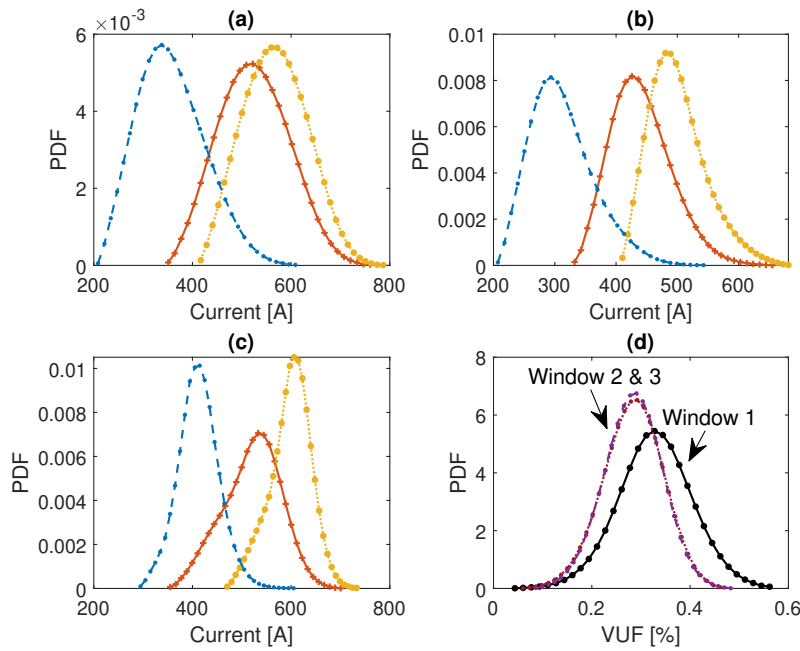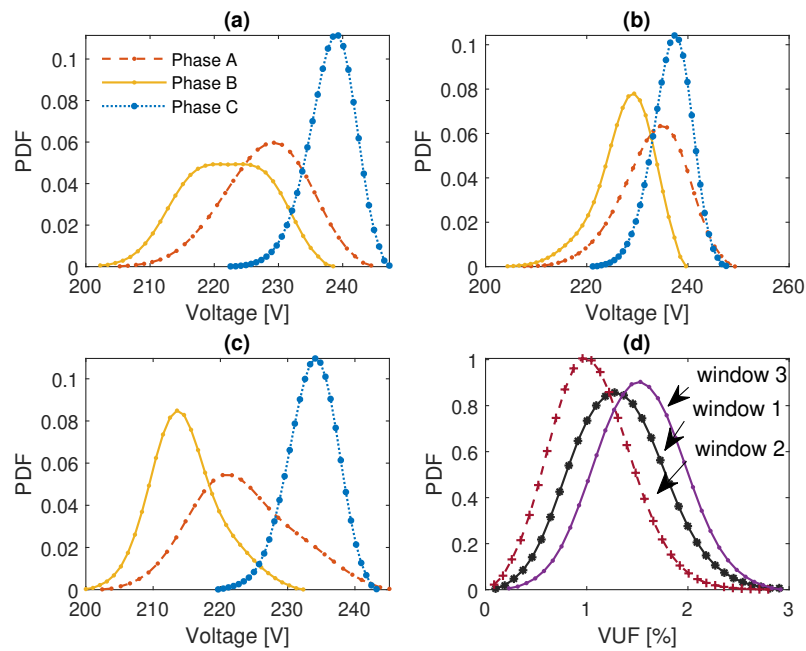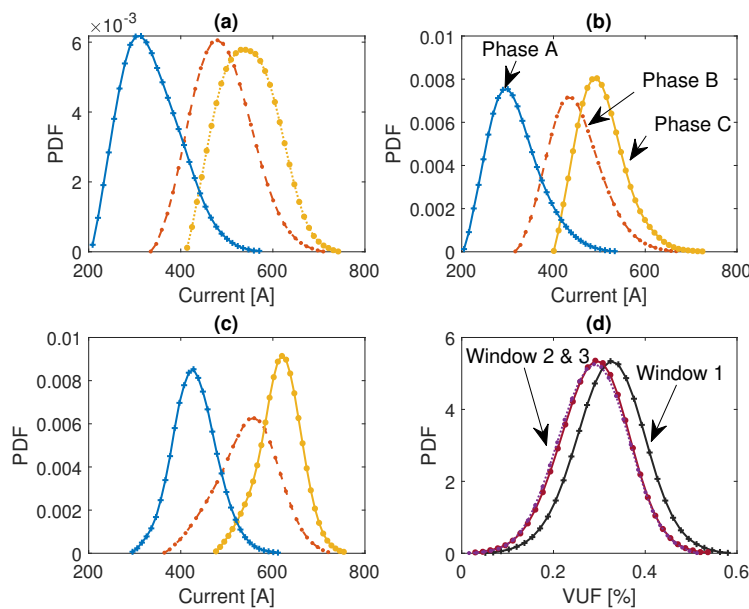

Figure 7.10: Three phase current distributions and the $VUF\%$ measured at the point m_1 of the LV network. (a), (b) and (c) shows the current distributions measured in 1st, 2nd and 3rd time windows respectively. (d) shows the $VUF\%$ measured in all the time windows.

8 | Conclusions

8.1. Final Remarks

This thesis is developed with a clear objective of quantifying the PV generation uncertainties and EV demand uncertainties using a data-driven modeling approach. The modeled electrical elements have been integrated with the test network to perform the stochastic impact analysis of such uncertain elements on the power system network operating at different voltage levels. The thesis covers various aspects of probabilistic load flow analysis, from data analysis and stochastic modeling to advanced numerical simulation proposing the novel accelerated load flow with correlated input variables.

The reason to adopt a data-driven modeling approach is to accurately model the system characteristics as present in the real world with minimal assumptions, thereby capturing the whole behavior of the system. Using this approach, an accurate analysis is conducted to verify the performance of the modeled systems, and then the impact of such systems on the power network is studied. The uncertainty is included in the deterministic load flow, and the analysis is conducted in suitable time windows. With the uncertainty vector and time as a variable, the deterministic load flow problem is converted into probabilistic load flow, the detailed formulation is narrated in Section 3.5.

The historical measurement data of PV systems and EV charging events are collected. To prove the generality of the proposed techniques the analysis is performed with multiple data sets that are reliable and used by the research community. To model PV generation uncertainty, data sets obtained from the Australian utility network and United Kingdom power network are used. The EV demand modeling is done with the data sets obtained from the California institute of technology and Boulder city of Colorado is used.

Initially, the probabilistic load flow analysis is performed with Monte Carlo simulation as it is the standard reference method, and the first challenge faced is the availability of the large sample set to use in MC simulation as it requires a large sample set to reach the convergence of the solution. In the conventional methods, the practice is to fit the data to a elementary statistical distribution and generate the required number

of samples from that distribution, but in data-driven analysis, the main threat posed is the resulting distributions obtained were non-elementary, as demonstrated in Section 4.1, hence a statistical method is necessary to repopulate samples.

At this stage, the goal of repopulating the samples is to obtain the same distribution shape of the input random variables with a large sample set. This is achieved using the Inverse CDF method, the complex distribution shape of the input random variables and the distribution shape of the regenerated samples are highly accurate with an error rate $< 1\%$. Later it is realized that the input random variables might be correlated, and this correlation will affect the accuracy of the results hence modeling techniques such as the Gaussian Mixture Model and Gaussian Copula are used in repopulating samples and used in PLF. Using these approaches not only delivers the accurate distribution shape, but also preserves the correlation among the variables.

In Chapter 4 of this work, the detailed analysis is conducted using all three above-mentioned methods in modeling PV generation uncertainty, and the modeled PV system is tested with IEEE 13 node test feeder network. It clearly shows the consideration of correlation among the input variables will have a significant improvement in the accuracy of the results. Among all the methods the Gaussian Copula is further exploited in the PLF because of its simplicity and high accuracy.

MC simulation is very expensive in terms of time and computational resources. The various PLF accelerating methods are followed by the research community, and in our case, well suited is using polynomial chaos expansion. In the power system, PLF analysis using polynomial chaos is still in an early development stage, and it is used to include the correlation among the input variables in accelerated PLF. Conventional SRSM uses polynomial chaos expansion and copula the surrogate model approximates the non-linear relationship between the physical input variables x_k and the inner variables $\xi_1, ..., \xi_N$ which is highly nonlinear defined in the transformation step of copula function. This method suffers from the drawbacks of inaccurate results, the bad tail end of the distributions, and numerical ill-conditioning. Along with these issues, higher-order polynomial expansion has to be used with more sample points. Although it has inaccuracies, it is one of the widely used methodologies in PLF analysis.

To overcome the drawbacks of the conventional SRSM, a novel SRSM is developed in this work. A simple idea is followed to apply the method of SRSM surrogate modeling and the copula in two separate computational flows, thereby reducing the non-linearity observed in approximating the input-output relationship. A big threat faced in implementing this novel approach is as it requires generalized base functions in building surrogate SRSM,

as non-elementary statistical distributions are used. The problem is solved successfully thanks to Darboux's formula and the three-term recurrence relation. The novel approach uses lower-order polynomials and results in more accurate approximations.

A test case is implemented to prove the proposed methodology. Four correlated PV generators resulting in non-elementary distributions are selected to integrate with NSELVTN. The obtained results are compared with conventional SRSM and the standard reference MC simulation. The obtained results are more accurate and take less time than the conventional SRSM, and novel SRSM delivers the result approximately *50*x faster than the MC simulation with the same accuracy as MC, which is a remarkable achievement. An effort is made in this work to extend the theory to the practical application to find the PV hosting capacity for a small section of the NSELVTN.

To model the EV load demand uncertainty, EV charging event data is used, vital information of the charging events such as *Start time*, *Duration*, and the *Energy delivered* are extracted. In the case of EV, the input variables are dependent among each other and hence a primitive but equally effective method is used in repopulating samples for use in MC simulation. A random number is called and the EV charging event data related to that number is used in EV profile generation, and it is found that this method is effective in preserving the characteristics of the charging event and the same has been demonstrated in Section 6.1.2. The EV load demand is created starting from an EV charging event and moving to the daily load profile of an EVCS and then the aggregation of EVCS. The proposed modeling methodology can be applied to the analysis in both LV and MV distribution networks by selecting the appropriate aggregation. The analysis is conducted in a four-hour time window with variable installed load capacity. The analysis conducted on 69 bus MV test network with aggregation of *10 EVCS* at every selected node, shows the EVCS aggregation daily profiles have a higher impact in the morning hours i.e., 07:30-11:30 the load in the test network is operated with 65% percent of installed load capacity the impact of EV is much higher compared to other time windows with voltage variability of about *210 V* and the *VUF%* variability of 0.45%.

In the last chapter, the analysis is performed including both the PV generation uncertainty and EV load demand uncertainty in the IEEE European LV test network. The test network is modified by including *9* PV generators of *15 kW* each that corresponds to *40%* PV penetration ratio that are modeled using Gaussian copula methodology preserving the correlation among the variables, *10 EVCS* are setup in the network, each EVCS having an installed capacity of *27 kW*. MC simulation is performed in a four hour time windows to analyze the impact of PV and EV demand uncertainty, higher impacts are seen in the third time window i.e., 16:00-20:00 when the PV generation less, load is operating

at maximum installed capacity. With this it is evident that with proper planing, EV demand uncertainty can be mitigated with PV generation uncertainty, the burden on the distribution network can be reduced, the tools and techniques developed in this thesis are very much useful in performing such analysis using data-driven methods that delivers results with high accuracy.

8.2. Future work

The future developments and the improvements suggested are as follows:

1. The proposed novel SRSM is a very good methodology in PLF, the selection of the samples points in the 3D space using stochastic collocation is very complex, and it is observed that the proposed method is failing to deliver sample points for more than *19* variables that are having non-elementary distributions. Further investigation is required in the selection of sample points among the quadrature nodes resulting in good accuracy with shorter computational time.

2. The NSELVTN used in this work is a good test network that is made using measurement data, but it is very complex to handle it for PLF analysis. Much of the work is needed in the direction to extract a smaller section of the network that is highly interesting to perform various types of analysis, such as hosting capacity and EV integration scenarios.

3. When the test network with several buses is used, statistical variation of load using measurement data is not easy, and it is due to the complexity of handling a large number of random variables as an input of PLF analysis that needs a solution.

4. Reliable residential load data for multiple loads in an LV distribution with large observation time intervals are not available. A solution has to be figured out to synthesize this data using any start-of-the-art techniques with which more accurate results in PLF uncertainty analysis can be achieved.

5. The tools and techniques of PLF developed in this work can be used as a predicting tool by the distributed energy management system in estimating the power flow in presence of PV and EV elements. The integration of load flow analysis into a real-time electromagnetic transient simulation test bench has to be developed.

Bibliography

[1] Beng Wah Ang, Wei Lim Choong, and Tsan Sheng Ng. Energy security: Definitions, dimensions and indexes. *Renewable and sustainable energy reviews*, 42:1077–1093, 2015.

[2] Fazıl Gökgöz and Mustafa Taylan Güvercin. Energy security and renewable energy efficiency in eu. *Renewable and Sustainable Energy Reviews*, 96:226–239, 2018.

[3] Sifan Jiang, Can Wan, Chen Chen, Erbao Cao, and Yonghua Song. Distributed photovoltaic generation in the electricity market: status, mode and strategy. *CSEE Journal of Power and Energy Systems*, 4(3):263–272, 2018.

[4] Zhong Zhang, Ran Li, and Furong Li. A novel peer-to-peer local electricity market for joint trading of energy and uncertainty. *IEEE Transactions on Smart Grid*, 11 (2):1205–1215, 2019.

[5] Tobias Aigner, Stefan Jaehnert, Gerard L Doorman, and Terje Gjengedal. The effect of large-scale wind power on system balancing in northern europe. *IEEE Transactions on Sustainable Energy*, 3(4):751–759, 2012.

[6] Arturs Purvins, Luigi Sereno, Mircea Ardelean, Catalin-Felix Covrig, Tilemahos Efthimiadis, and Philip Minnebo. Submarine power cable between europe and north america: A techno-economic analysis. *Journal of Cleaner Production*, 186:131–145, 2018.

[7] Panayiotis Demetriou, Markos Asprou, Jairo Quiros-Tortos, and Elias Kyriakides. Dynamic ieee test systems for transient analysis. *IEEE Systems Journal*, 11(4): 2108–2117, 2015.

[8] Iqra Javid, Aditya Chauhan, Sahil Thappa, SK Verma, Y Anand, A Sawhney, VV Tyagi, and S Anand. Futuristic decentralized clean energy networks in view of inclusive-economic growth and sustainable society. *Journal of Cleaner Production*, 309:127304, 2021.

[9] Murray Thomson and David G Infield. Network power-flow analysis for a high

penetration of distributed generation. *IEEE Transactions on power systems*, 22(3): 1157–1162, 2007.

[10] Jun Su, Tek Tjing Lie, and Ramon Zamora. Integration of electric vehicles in distribution network considering dynamic power imbalance issue. *IEEE Transactions on Industry Applications*, 56(5):5913–5923, 2020.

[11] Khaled Hajar, Baoling Guo, Ahmad Hably, and Seddik Bacha. Smart charging impact on electric vehicles in presence of photovoltaics. In *2021 22nd IEEE International Conference on Industrial Technology (ICIT)*, volume 1, pages 643–648. IEEE, 2021.

[12] Sergio Bruno, Silvia Lamonaca, Giuseppe Rotondo, Ugo Stecchi, and Massimo La Scala. Unbalanced three-phase optimal power flow for smart grids. *IEEE Transactions on Industrial Electronics*, 58(10):4504–4513, 2011.

[13] Jing Wang, Michael Blonsky, Fei Ding, Seth C Drew, Harsha Padullaparti, Shibani Ghosh, Ismael Mendoza, Soumya Tiwari, Jorge E Martinez, Jose JD Dahdah, et al. Performance evaluation of distributed energy resource management via advanced hardware-in-the-loop simulation. In *2020 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, pages 1–5. IEEE, 2020.

[14] Ferry A Viawan, Daniel Karlsson, Ambra Sannino, and Jaap Daalder. Protection scheme for meshed distribution systems with high penetration of distributed generation. In *2006 Power Systems Conference: Advanced Metering, Protection, Control, Communication, and Distributed Resources*, pages 99–104. IEEE, 2006.

[15] Hun-Chul Seo. New protection scheme in loop distribution system with distributed generation. *Energies*, 13(22):5897, 2020.

[16] Thomas Pfeiffenberger, Jia Lei Du, Pedro Bittencourt Arruda, and Alessandro Anzaloni. Reliable and flexible communications for power systems: Fault-tolerant multicast with sdn/openflow. In *2015 7th International Conference on New Technologies, Mobility and Security (NTMS)*, pages 1–6. IEEE, 2015.

[17] Min Chen, Lasse A Rosendahl, Thomas J Condra, and John K Pedersen. Numerical modeling of thermoelectric generators with varing material properties in a circuit simulator. *IEEE Transactions on Energy Conversion*, 24(1):112–124, 2009.

[18] Himanshu Jain, Abhineet Parchure, Robert P Broadwater, Murat Dilek, and Jeremy Woyak. Three-phase dynamic simulation of power systems using combined trans-

mission and distribution system models. *IEEE Transactions on Power Systems*, 31 (6):4517–4524, 2016.

[19] Jun Hua Zhao, John Foster, Zhao Yang Dong, and Kit Po Wong. Flexible transmission network planning considering distributed generation impacts. *IEEE Transactions on Power Systems*, 26(3):1434–1443, 2010.

[20] Rajkumar Kaushik, Om Prakash Mahela, Pramod Kumar Bhatt, Baseem Khan, Akhil Ranjan Garg, Hassan Haes Alhelou, and Pierluigi Siano. Recognition of islanding and operational events in power system with renewable energy penetration using a stockwell transform-based method. *IEEE Systems Journal*, 2020.

[21] Hieu Le Nguyen. Newton-raphson method in complex form [power system load flow analysis]. *IEEE transactions on power systems*, 12(3):1355–1359, 1997.

[22] PW Sauer and MA Pai. Power system steady-state stability and the load-flow jacobian. *IEEE Transactions on power systems*, 5(4):1374–1383, 1990.

[23] Sunil Subedi, Manisha Rauniyar, Saima Ishaq, Timothy M Hansen, Reinaldo Tonkoski, Mariko Shirazi, Richard Wies, and Phylicia Cicilio. Review of methods to accelerate electromagnetic transient simulation of power systems. *IEEE Access*, 9:89714–89731, 2021.

[24] Arjen A van der Meer, Madeleine Gibescu, Mart AMM van der Meijden, Wil L Kling, and Jan A Ferreira. Advanced hybrid transient stability and emt simulation for vsc-hvdc systems. *IEEE Transactions on Power Delivery*, 30(3):1057–1066, 2014.

[25] Hani Saad, Albane Schwob, and Yannick Vernay. Study of resonance issues between hvdc link and power system components using emt simulations. In *2018 Power Systems Computation Conference (PSCC)*, pages 1–8. IEEE, 2018.

[26] Heng Nian and Liang Kong. Transient modeling and analysis of vsc based dc microgrid during short circuit fault. *IEEE Access*, 7:170604–170614, 2019.

[27] Lutz Janicke and Arnulf Kost. Convergence properties of the newton-raphson method for nonlinear problems. *IEEE transactions on magnetics*, 34(5):2505–2508, 1998.

[28] Amirsaman Arabali, Mahmoud Ghofrani, Mehdi Etezadi-Amoli, and Mohammed Sami Fadali. Stochastic performance assessment and sizing for a hybrid power system of solar/wind/energy storage. *IEEE Transactions on Sustainable Energy*, 5(2):363–371, 2013.

[29] Joakim Widén, Mahmoud Shepero, and Joakim Munkhammar. Probabilistic load flow for power grids with high pv penetrations using copula-based modeling of spatially correlated solar irradiance. *IEEE Journal of Photovoltaics*, 7(6):1740–1745, 2017.

[30] Umar Hanif Ramadhani, Mahmoud Shepero, Joakim Munkhammar, Joakim Widen, and Nicholas 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*, 120:106003, 2020.

[31] Gonzalo Esteban Constante-Flores and Mahesh S Illindala. Data-driven probabilistic power flow analysis for a distribution system with renewable energy sources using monte carlo simulation. *IEEE Transactions on Industry Applications*, 55(1): 174–181, 2018.

[32] Peiyuan Chen, Zhe Chen, and Birgitte Bak-Jensen. Probabilistic load flow: A review. In *2008 Third International Conference on Electric Utility Deregulation and Restructuring and Power Technologies*, pages 1586–1591. IEEE, 2008.

[33] Giambattista Gruosso, Roberto S Netto, Luca Daniel, and Paolo Maffezzoni. Joined probabilistic load flow and sensitivity analysis of distribution networks based on polynomial chaos method. *IEEE Transactions on Power Systems*, 35(1):618–627, 2019.

[34] Harshavardhan Palahalli, Paolo Maffezzoni, and Giambattista 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)*, pages 1–6. IEEE, 2020.

[35] Harshavardhan Palahalli, Paolo Maffezzoni, and Giambattista Gruosso. Gaussian copula methodology to model photovoltaic generation uncertainty correlation in power distribution networks. *Energies*, 14(9):2349, 2021.

[36] Harshavardhan Palahalli, Paolo Maffezzoni, and Giambattista Gruosso. Statistical simulation of electric vehicle behaviour applied to low voltage distribution network. In *2021 22nd IEEE International Conference on Industrial Technology (ICIT)*, volume 1, pages 657–662. IEEE, 2021.

[37] Harshavardhan Palahalli, Paolo Maffezzoni, Pablo Arboleya, and Giambattista Gruosso. implementing stochastic response surface method and copula in the presence of data-driven pv source models. *IEEE Transactions on Sustainable Energy*, 2022.

[38] Harshavardhan Palahalli, Paolo Maffezzoni, Luca Daniel, and Giambattista Gruosso. Statistical analysis of pv penetration impact on residential distribution grids. *Sustainable Energy, Grids and Networks*, page 100949, 2022.

[39] John H Drew, Andrew G Glen, and Lawrence M Leemis. Computing the cumulative distribution function of the kolmogorov–smirnov statistic. *Computational statistics & data analysis*, 34(1):1–15, 2000.

[40] Edwin T Jaynes. *Probability theory: The logic of science.* Cambridge university press, 2003.

[41] Ye Guo, Boming Zhang, Wenchuan Wu, Qinglai Guo, and Hongbin Sun. Solvability and solutions for bus-type extended load flow. *International Journal of Electrical Power & Energy Systems*, 51:89–97, 2013.

[42] Brian Stott. Review of load-flow calculation methods. *Proceedings of the IEEE*, 62 (7):916–929, 1974.

[43] Newton G Bretas and Arturo S Bretas. The extension of the gauss approach for the solution of an overdetermined set of algebraic non linear equations. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 65(9):1269–1273, 2018.

[44] José Manuel Gutiérrez, Ángel Alberto Magreñán, and Juan Luis Varona. The "gauss-seidelization" of iterative methods for solving nonlinear equations in the complex plane. *Applied Mathematics and Computation*, 218(6):2467–2479, 2011.

[45] Thomas J Overbye, Xu Cheng, and Yan Sun. A comparison of the ac and dc power flow models for lmp calculations. In *37th Annual Hawaii International Conference on System Sciences, 2004. Proceedings of the*, pages 9–pp. IEEE, 2004.

[46] Ray D Zimmerman and Hsiao-Dong Chiang. Fast decoupled power flow for unbalanced radial distribution systems. *IEEE Transactions on Power systems*, 10(4): 2045–2052, 1995.

[47] Enrico Zio, Roberta Piccinelli, Maurizio Delfanti, Valeria Olivieri, and Mauro Pozzi. Application of the load flow and random flow models for the analysis of power transmission networks. *Reliability Engineering & System Safety*, 103:102–109, 2012.

[48] Giambattista Gruosso, Giancarlo Storti Gajani, Zheng Zhang, Luca Daniel, and Paolo Maffezzoni. Uncertainty-aware computational tools for power distribution networks including electrical vehicle charging and load profiles. *IEEE Access*, 7: 9357–9367, 2019.

[49] AM Leite Da Silva, S_M P_ Ribeiro, VL Arienti, RN Allan, and MB Do Coutto Filho. Probabilistic load flow techniques applied to power system expansion planning. *IEEE Transactions on Power Systems*, 5(4):1047–1053, 1990.

[50] B Rajanarayan Prusty and Debashisha Jena. A critical review on probabilistic load flow studies in uncertainty constrained power systems with photovoltaic generation and a new approach. *Renewable and Sustainable Energy Reviews*, 69:1286–1302, 2017.

[51] Hans Janssen. Monte-carlo based uncertainty analysis: Sampling efficiency and sampling convergence. *Reliability Engineering & System Safety*, 109:123–132, 2013.

[52] Attoti Bharath Krishna and Abhijit R Abhyankar. Uniform experimental design-based nonparametric quasi-monte carlo for efficient probabilistic power flow. *IEEE Transactions on Power Systems*, 2022.

[53] Zhen Shu and Panida Jirutitijaroen. Latin hypercube sampling techniques for power systems reliability analysis with renewable energy sources. *IEEE Transactions on Power Systems*, 26(4):2066–2073, 2011.

[54] Han Yu, CY Chung, KP Wong, HW Lee, and JH Zhang. Probabilistic load flow evaluation with hybrid latin hypercube sampling and cholesky decomposition. *IEEE Transactions on Power Systems*, 24(2):661–667, 2009.

[55] Yan Chen, Jinyu Wen, and Shijie Cheng. Probabilistic load flow method based on nataf transformation and latin hypercube sampling. *IEEE Transactions on Sustainable Energy*, 4(2):294–301, 2012.

[56] Bin Zou and Qing Xiao. Solving probabilistic optimal power flow problem using quasi monte carlo method and ninth-order polynomial normal transformation. *IEEE Transactions on Power Systems*, 29(1):300–306, 2013.

[57] Tao Cui and Franz Franchetti. A quasi-monte carlo approach for radial distribution system probabilistic load flow. In *2013 IEEE PES Innovative Smart Grid Technologies Conference (ISGT)*, pages 1–6. IEEE, 2013.

[58] N Soleimanpour and M Mohammadi. Probabilistic load flow by using nonparametric density estimators. *IEEE Transactions on Power systems*, 28(4):3747–3755, 2013.

[59] Hossein Nosratabadi, Mohammad Mohammadi, and Amin Kargarian. Nonparametric probabilistic unbalanced power flow with adaptive kernel density estimator. *IEEE Transactions on Smart Grid*, 10(3):3292–3300, 2018.

[60] Meghdad Tourandaz Kenari, Mohammad Sadegh Sepasian, Mehrdad Setayesh Nazar, and Hossein Ali Mohammadpour. Combined cumulants and laplace transform method for probabilistic load flow analysis. *IET Generation, Transmission & Distribution*, 11(14):3548–3556, 2017.

[61] Yulong Che, Xiaoru Wang, Xiaoqin Lv, and Yi Hu. Probabilistic load flow using improved three point estimate method. *International Journal of Electrical Power & Energy Systems*, 117:105618, 2020.

[62] Julio Usaola. Probabilistic load flow with wind production uncertainty using cumulants and cornish–fisher expansion. *International Journal of Electrical Power & Energy Systems*, 31(9):474–481, 2009.

[63] Trevor Williams and Curran Crawford. Probabilistic load flow modeling comparing maximum entropy and gram-charlier probability density function reconstructions. *IEEE Transactions on Power Systems*, 28(1):272–280, 2012.

[64] Miao Fan, Vijay Vittal, Gerald Thomas Heydt, and Raja Ayyanar. Probabilistic power flow studies for transmission systems with photovoltaic generation using cumulants. *IEEE Transactions on Power Systems*, 27(4):2251–2261, 2012.

[65] Chun-Lien Su. Probabilistic load-flow computation using point estimate method. *IEEE Transactions on Power Systems*, 20(4):1843–1851, 2005.

[66] Xiaomeng Ai, Jinyu Wen, Tong Wu, and Wei-Jen Lee. A discrete point estimate method for probabilistic load flow based on the measured data of wind power. *IEEE Transactions on Industry Applications*, 49(5):2244–2252, 2013.

[67] Morteza Aien, Mahmud Fotuhi-Firuzabad, and Farrokh Aminifar. Probabilistic load flow in correlated uncertain environment using unscented transformation. *IEEE Transactions on Power systems*, 27(4):2233–2241, 2012.

[68] Sui Peng, Xi Pang, Huixiang Chen, Xingyu Lin, Yong Lin, Zhengmin Zuo, Hao Yu, and Junjie Tang. Comparison on three unscented transformation methods for solving probabilistic load flow. In *2019 IEEE Innovative Smart Grid Technologies-Asia (ISGT Asia)*, pages 3941–3946. IEEE, 2019.

[69] Zhouyang Ren, Wenyuan Li, Roy Billinton, and Wei Yan. Probabilistic power flow analysis based on the stochastic response surface method. *IEEE Transactions on Power Systems*, 31(3):2307–2315, 2015.

[70] Mohammad Mohammadi, Hooman Basirat, and Amin Kargarian. Nonparametric

probabilistic load flow with saddle point approximation. *IEEE Transactions on Smart Grid*, 9(5):4796–4804, 2017.

[71] Ali Reza Abbasi. Comparison parametric and non-parametric methods in probabilistic load flow studies for power distribution networks. *Electrical Engineering*, pages 1–12, 2022.

[72] Alireza Soroudi, Morteza Aien, and Mehdi Ehsan. A probabilistic modeling of photo voltaic modules and wind power generation impact on distribution networks. *IEEE Systems Journal*, 6(2):254–259, 2011.

[73] S Surender Reddy. Optimal scheduling of thermal-wind-solar power system with storage. *Renewable energy*, 101:1357–1368, 2017.

[74] Reza Hemmati, Rahmat-Allah Hooshmand, and Amin Khodabakhshian. Market based transmission expansion and reactive power planning with consideration of wind and load uncertainties. *Renewable and Sustainable Energy Reviews*, 29:1–10, 2014.

[75] Giambattista Gruosso and Paolo Maffezzoni. Data-driven uncertainty analysis of distribution networks including photovoltaic generation. *International Journal of Electrical Power & Energy Systems*, 121:106043, 2020.

[76] Hossein Khorasanizadeh, Kasra Mohammadi, and Navid Goudarzi. Prediction of horizontal diffuse solar radiation using clearness index based empirical models; a case study. *International Journal of Hydrogen Energy*, 41(47):21888–21898, 2016.

[77] Guido Carpinelli, Pierluigi Caramia, and Pietro Varilone. Multi-linear monte carlo simulation method for probabilistic load flow of distribution systems with wind and photovoltaic generation systems. *Renewable Energy*, 76:283–295, 2015.

[78] Stefania Conti and Salvatore Raiti. Probabilistic load flow using monte carlo techniques for distribution networks with photovoltaic generators. *Solar Energy*, 81(12): 1473–1481, 2007.

[79] Chun Sing Lai, Youwei Jia, Malcolm D McCulloch, and Zhao Xu. Daily clearness index profiles cluster analysis for photovoltaic system. *IEEE Transactions on Industrial Informatics*, 13(5):2322–2332, 2017.

[80] B Rajanarayan Prusty and Debashisha Jena. Combined cumulant and gaussian mixture approximation for correlated probabilistic load flow studies: a new approach. *CSEE Journal of Power and Energy Systems*, 2(2):71–78, 2016.

[81] Harshavardhan Palahalli, Yujia Huo, and Giambattista Gruosso. Modelling of pho-
tovoltaic systems for real-time hardware simulation. In *ELECTRIMACS 2019*, pages
3–15. Springer, 2020.

[82] Julian J Warren, Michael Negnevitsky, and Thanh Nguyen. Probabilistic load flow
analysis in distribution networks with distributed solar generation. In *2016 IEEE
Power and Energy Society General Meeting (PESGM)*, pages 1–5. IEEE, 2016.

[83] Neeraj Gupta. Gauss-quadrature-based probabilistic load flow method with voltage-
dependent loads including wtgs, pv, and ev charging uncertainties. *IEEE Transac-
tions on Industry Applications*, 54(6):6485–6497, 2018.

[84] Miao Fan, Vijay Vittal, Gerald T Heydt, and Raja Ayyanar. Probabilistic power flow
analysis with generation dispatch including photovoltaic resources. *IEEE Transac-
tions on Power Systems*, 28(2):1797–1805, 2012.

[85] Yunfei Mu, Jianzhong Wu, Nick Jenkins, Hongjie Jia, and Chengshan Wang. A
spatial–temporal model for grid impact analysis of plug-in electric vehicles. *Applied
Energy*, 114:456–465, 2014.

[86] Sungwoo Bae and Alexis Kwasinski. Spatial and temporal model of electric vehicle
charging demand. *IEEE Transactions on Smart Grid*, 3(1):394–403, 2011.

[87] Kejun Qian, Chengke Zhou, Malcolm Allan, and Yue Yuan. Modeling of load
demand due to ev battery charging in distribution systems. *IEEE transactions on
power systems*, 26(2):802–810, 2010.

[88] Rong-Ceng Leou, Chun-Lien Su, and Chan-Nan Lu. Impact analysis of electric
vehicles on distribution systems considering uncertainties. In *IECON 2013-39th
Annual Conference of the IEEE Industrial Electronics Society*, pages 2063–2068.
IEEE, 2013.

[89] Gan Li and Xiao-Ping Zhang. Modeling of plug-in hybrid electric vehicle charging
demand in probabilistic power flow calculations. *IEEE Transactions on Smart Grid*,
3(1):492–499, 2012.

[90] Xiaochen Zhang and Santiago Grijalva. An advanced data driven model for resi-
dential electric vehicle charging demand. In *2015 IEEE Power & Energy Society
General Meeting*, pages 1–5. IEEE, 2015.

[91] Erotokritos Xydas, Charalampos Marmaras, Liana M Cipcigan, Nick Jenkins, Steve
Carroll, and Myles Barker. A data-driven approach for characterising the charging
demand of electric vehicles: A uk case study. *Applied energy*, 162:763–771, 2016.

[92] Oliver Frendo, Jérôme Graf, Nadine Gärtner, and Heiner Stuckenschmidt. Data-driven smart charging for heterogeneous electric vehicle fleets. *Energy and AI*, 1: 100007, 2020.

[93] Ioannis Antonopoulos, Valentin Robu, Benoit Couraud, and David Flynn. Data-driven modelling of energy demand response behaviour based on a large-scale residential trial. *Energy and AI*, 4:100071, 2021.

[94] AM Leite Da Silva, VL Arienti, and RN Allan. Probabilistic load flow considering dependence between input nodal powers. *IEEE Transactions on Power Apparatus and Systems*, 4(6):1524–1530, 1984.

[95] Peerapat Vithayasrichareon, Graham Mills, and Iain F MacGill. Impact of electric vehicles and solar pv on future generation portfolio investment. *IEEE Transactions on sustainable energy*, 6(3):899–908, 2015.

[96] Aaron Lei Liu, Mehdi Shafiei, Gerard Ledwich, Wendy Miller, and Ghavameddin Nourbakhsh. Correlation study of residential community demand with high pv penetration. In *2017 Australasian Universities Power Engineering Conference (AUPEC)*, pages 1–6. IEEE, 2017.

[97] B Rajanarayan Prusty and Debashisha Jena. Modeling of correlated photovoltaic generations and load demands in probabilistic load flow. In *2015 Annual IEEE India Conference (INDICON)*, pages 1–6. IEEE, 2015.

[98] Guido Carpinelli, Renato Rizzo, Pierluigi Caramia, and Pietro Varilone. Taguchi's method for probabilistic three-phase power flow of unbalanced distribution systems with correlated wind and photovoltaic generation systems. *Renewable Energy*, 117: 227–241, 2018.

[99] Ravindra Singh, Bikash C Pal, and Rabih A Jabr. Statistical representation of distribution system loads using gaussian mixture model. *IEEE Transactions on Power Systems*, 25(1):29–37, 2009.

[100] H Valizadeh Haghi, M Tavakoli Bina, MA Golkar, and SM Moghaddas-Tafreshi. Using copulas for analysis of large datasets in renewable distributed generation: Pv and wind power integration in iran. *Renewable Energy*, 35(9):1991–2000, 2010.

[101] Dongbin Xiu and George Em Karniadakis. The wiener–askey polynomial chaos for stochastic differential equations. *SIAM journal on scientific computing*, 24(2): 619–644, 2002.

[102] Zheng Zhang, Tarek A El-Moselhy, Ibrahim M Elfadel, and Luca 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):1533–1545, 2013.

[103] RJ Yanez, JS Dehesa, and AF Nikiforov. The three-term recurrence relation and the differentiation formulas for hypergeometric-type functions. *Journal of mathematical analysis and applications*, 188(3):855–866, 1994.

[104] Q-Han Park and HJ Shin. Darboux transformation and crum's formula for multi-component integrable equations. *Physica D: Nonlinear Phenomena*, 157(1-2):1–15, 2001.

[105] Ieee 13 node test feeder data, 1992. URL `https://site.ieee.org/pes-testfeeders/resources/`. (Date last accessed 29-August-2022).

[106] Arpan Koirala, Lucía Suárez-Ramón, Bassam Mohamed, and Pablo Arboleya. Non-synthetic european low voltage test system. *International Journal of Electrical Power & Energy Systems*, 118:105712, 2020.

[107] Ieee european low voltage test feeder, 2022. URL `https://cmte.ieee.org/pes-testfeeders/resources/`. (Date last accessed 05-October-2022).

[108] Yasin Zabihinia Gerdroodbari, Reza Razzaghi, and Farhad Shahnia. Decentralized control strategy to improve fairness in active power curtailment of pv inverters in low-voltage distribution networks. *IEEE Transactions on Sustainable Energy*, 12 (4):2282–2292, 2021.

[109] Umar Hanif Ramadhani, Reza Fachrizal, Mahmoud Shepero, Joakim Munkhammar, and Joakim Widén. Probabilistic load flow analysis of electric vehicle smart charging in unbalanced lv distribution systems with residential photovoltaic generation. *Sustainable Cities and Society*, 72:103043, 2021.

[110] Md Shahin Alam and Seyed Ali Arefifar. Optimal allocation of ev charging stations in distribution systems considering discharging economy and system reliability. In *2021 IEEE International Conference on Electro Information Technology (EIT)*, pages 1–8. IEEE, 2021.

[111] Cesar Diaz-Londono, Luigi Colangelo, Fredy Ruiz, Diego Patino, Carlo Novara, and Gianfranco Chicco. Optimal strategy to exploit the flexibility of an electric vehicle charging station. *Energies*, 12(20):3834, 2019.

[112] Thomas S Basso and Richard DeBlasio. Ieee 1547 series of standards: interconnection issues. *IEEE Transactions on Power Electronics*, 19(5):1159–1162, 2004.

[113] Australian grid solar home electricity data sampled half hourly 1 july 2012 to 30 june 2013., 2013. URL `https://www.ausgrid.com.au/Industry/Our-Research/Data-to-share/Solar-home-electricity-data`. (Date last accessed 29-August-2022).

[114] Yingchao Dong, Hongli Zhang, Cong Wang, and Xiaojun Zhou. A novel hybrid model based on bernstein polynomial with mixture of gaussians for wind power forecasting. *Applied Energy*, 286:116545, 2021.

[115] Daniel Alspach and Harold Sorenson. Nonlinear bayesian estimation using gaussian sum approximations. *IEEE transactions on automatic control*, 17(4):439–448, 1972.

[116] Todd K Moon. The expectation-maximization algorithm. *IEEE Signal processing magazine*, 13(6):47–60, 1996.

[117] Jared S Murray, David B Dunson, Lawrence Carin, and Joseph E Lucas. Bayesian gaussian copula factor models for mixed data. *Journal of the American Statistical Association*, 108(502):656–665, 2013.

[118] Guanzhong Wang, Huanhai Xin, Di Wu, Ping Ju, and Xichen Jiang. Data-driven arbitrary polynomial chaos-based probabilistic load flow considering correlated uncertainties. *IEEE Transactions on Power Systems*, 34(4):3274–3276, 2019.

[119] Dka solar centre, alice springs project, 2022. URL `https://dkasolarcentre.com.au/`. (Date last accessed 29-August-2022).

[120] Zachary J Lee, Tongxin Li, and Steven H Low. Acn-data: Analysis and applications of an open ev charging dataset. In *Proceedings of the Tenth ACM International Conference on Future Energy Systems*, pages 139–149, 2019.

[121] Electric vehicle charging station energy consumption, 2022. URL `https://open-data.bouldercolorado.gov/datasets`. (Date last accessed 29-August-2022).

[122] Ali Al-Wakeel, Jianzhong Wu, and Nick Jenkins. State estimation of medium voltage distribution networks using smart meter measurements. *Applied energy*, 184:207–218, 2016.