

# PROBABILISTIC STEADY-STATE ANALYSIS OF POWER SYSTEMS WITH PHOTOVOLTAIC GENERATIONS

Thesis

Submitted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

by

**B RAJANARAYAN PRUSTY**



DEPARTMENT OF ELECTRICAL AND ELECTRONICS ENGINEERING,  
NATIONAL INSTITUTE OF TECHNOLOGY KARNATAKA,  
SURATHKAL, MANGALORE -575025

MARCH, 2019



# DECLARATION

*by the Ph.D. Research Scholar*

I hereby *declare* that the Research Thesis entitled “**Probabilistic Steady-State Analysis of Power Systems With Photovoltaic Generations**” which is being submitted to the **National Institute of Technology Karnataka, Surathkal** in partial fulfillment of the requirement for the award of the Degree of **Doctor of Philosophy** in **Department of Electrical and Electronics Engineering** is a *bonafide report of the research work carried out by me*. The material contained in this Research Thesis has not been submitted to any University or Institution for the award of any degree.

.....  
B Rajanarayan Prusty, 145080EE14F01  
Department of Electrical and Electronics Engineering

Place: NITK-Surathkal

Date:



# CERTIFICATE

This is to *certify* that the Research Thesis entitled “**Probabilistic Steady-State Analysis of Power Systems With Photovoltaic Generations**” submitted by **B Rajanarayan Prusty** (Register Number: 145080EE14F01) as the record of the research work carried out by him, is *accepted as the Research Thesis submission* in partial fulfillment of the requirements for the award of degree of **Doctor of Philosophy**.

Dr. Debashisha Jena  
(Research Guide)

Dr. B. Venkatesa Perumal  
(Chairman-DRPC, EEE dept.)



## Acknowledgements

It gives me immense pleasure and great sense of satisfaction to express my heartfelt gratitude to those who made this dissertation possible.

I would like to express my sincere gratitude to *Dr. Debashisha Jena* for his guidance, encouragement, and for having been my Ph.D. supervisor. He has been a constant source of inspiration throughout this journey. I feel proud to have worked under his guidance.

I thank National Institute of Technology Karnataka (NITK), Surathkal for giving me an opportunity for doing research and Ministry of Human Resource Development (MHRD), Government of India for awarding research scholarship.

I wish to thank my research progress assessment committee (RPAC) members *Dr. Ashvini Chaturvedi*, and *Dr. P Sam Johnson*, for their constructive feedback and guidance.

Thanks also goes to *Dr. Jora M. Gonda* and *Dr. Vinatha U*, former HODs for providing the necessary resources in the department to carry out my research. Also, I would like to thank HOD, *Dr. B. Venkatesa Perumal* for his thought-provoking ideas and suggestions.

Without my friends in and out of NITK Surathkal, life would have been dull. I am indebted to all of them for their support, valuable inputs and constant encouragement.

I would like to express my deepest gratitude towards my family for their love and patience which kept me going in this journey. Their faith and unconditional love towards me are the reason for whatever I have achieved in my life.

Finally, I thank God Almighty for giving me strength at all times.





## Abstract

Recently, the application of probabilistic methods for power system analyses has become increasingly popular owing to their capability to instill enough confidence in system planner and operator in making more realistic decisions. In the conventional deterministic methods, consideration of a few typically stressed operating conditions are inadequate in solving the present uncertainty problems which are majorly confronted due to the enormous integration of renewable generations along with the conventional load powers. Probabilistic steady-state analysis (PSSA) refers to the adaptation of probabilistic load flow (PLF) to address the aforementioned uncertainties for characterizing the uncertainties in the power system variables referred to as result variables.

Among the many renewable sources, photovoltaics (PVs) have experienced a globally increasing significance as its cost per unit is decreasing day by day. PV generation is intermittent and variable with a higher level of uncertainty; their integration to power system greatly affects the power system variables which is a significant concern in the power system studies. Hence, a study focusing on the various aspects of power systems with integration of such renewable resources is the need of the hour. Therefore, this thesis is dedicated towards the PSSA of PV integrated power systems to examine various uncertainty issues that are likely to be combated in transmission systems.

The primary requirements for PSSA are mainly of threefold which include uncertainty modeling, power system model development, and application of an uncertainty handling method. This thesis aims at the improvement of each of these facets through suitable modifications and eventually resulting in an elegant PSSA.

For the uncertainty modeling, use of the historical record of inputs yields realistic models. For power system expansion and operational planning, such models use the daily time step data corresponding to the time of the year concerning the study of interest. The span of the chosen time series ranges from few months to few years depending on the study requirement or data availability. The daily time series of PSSA inputs such as load

power, PV generation, ambient temperature, etc. exhibit complex patterns that are periodic, encompassing predictable components. It is vital to separate such components from the raw data to characterize the unpredictable residuals referred to as preprocessing. In this regard, methods for preprocessing using multiple linear regression is proposed, and are compared with state of the art methods using the data collected from various places in India and USA. The rationale involved in the development of such models is deliberated in detail. Finally, a scenario-based spatiotemporal probabilistic model is developed by adopting the proposed preprocessing, transformation techniques, principal component analysis, and a suitable time series model capable of accurately modeling the trend in the variance of uncertain inputs.

Risk-based power system studies considering PV generations facilitate in delimiting the permissible penetration by executing essential steps to hedge systems risks. On this line, a risk assessment of PV arrays integrated to New England 39-bus transmission system is carried out. An improved system model is developed by accounting for the effect of environmental conditions, predominantly, the ambient temperature on the branch parameters by considering the electro-thermal coupling effect. The PLF that embodies the above effect in system model is referred to as temperature-augmented PLF (TPLF). It considers uncertainties in PV generation, aggregate load power, and ambient temperature along with their associated correlations for risk assessment. The effect of increased PV penetrations and variation in TPLF model parameters on the statistics of result variables is analyzed in detail. The expected system over-limit risk indices are calculated and are analyzed for different PV penetrations and input correlations.

In general, operational studies require a faster estimation of PSSA. One of the ways to achieve this is through the use of an uncertainty handling method that obtains accurate results in less time. On this line, efforts are made to devise two hybrid methods for PLF and TPLF simulations. Here, “hybrid” refers to the suitable amalgam of two uncertainty handling methods in part or as a whole through suitable modifications. As the thesis focuses on the larger transmission systems, cumulant method is chosen as

one of the potential methods for hybridization. It is seen that, based on a comprehensive result analysis, the proposed hybrid methods exhibit improved performance in the approximation of multimodal probability distributions of the result variables.

For all the above studies, PSSA is carried out on various transmission systems such as New England 39-bus test system, IEEE 14-bus, 57-bus 118-bus test systems, and Indian utility 62-bus test system. MATLAB 7.10 is used to develop the corresponding programming codes for various analyses. Finally, with the aid of the obtained results, the research work in this thesis demonstrates that the proposed models and methods for PSSA are potentially challenging candidates which facilitate in making sensible decisions regarding the planning and operation of PV-integrated power systems.



# Contents

Acknowledgements . . . . .	i
Abstract . . . . .	iii
List of figures . . . . .	xi
List of tables . . . . .	xiv
Abbreviations . . . . .	xvi
<b>1 INTRODUCTION</b>	<b>1</b>
1.1 Background . . . . .	1
1.2 State of the art in Probabilistic load flow . . . . .	3
1.2.1 Modeling of input uncertainties and dependencies . . . . .	3
1.2.1.1 Time-horizons, time-scales and time-steps . . . . .	4
1.2.1.2 Review of uncertainty modeling and dependencies . . . . .	5
1.2.2 Power system models for probabilistic load flow . . . . .	8
1.2.3 Methods to solve probabilistic load flow . . . . .	8
1.2.3.1 Numerical method . . . . .	9
1.2.3.2 Approximate method . . . . .	11
1.2.3.3 Analytical method . . . . .	11
1.2.4 Applications of probabilistic load flow . . . . .	14
1.3 Research gaps . . . . .	15
1.4 Research objectives . . . . .	17
1.5 Structure of the thesis . . . . .	17
<b>2 UNCERTAINTY ANALYSIS AND MODELING</b>	<b>21</b>
2.1 Preamble . . . . .	21
2.2 Uncertainty characterization . . . . .	21
2.2.1 Collection of historical data . . . . .	22
2.2.2 Preprocessing . . . . .	24

2.2.2.1	Data debugging, missing data updation and removal of daylight time shift effect . . . . .	24
2.2.2.2	Estimation of annual predictable variation . . . . .	24
2.2.2.3	Multiple linear regression model for photovoltaic generation . . . . .	27
2.2.2.4	Multiple linear regression model for load power and ambient temperature . . . . .	32
2.2.3	Performance comparison of proposed multiple linear regression models . . . . .	35
2.3	Spatiotemporal forecasting model . . . . .	46
2.3.1	Generalized autoregressive conditionally heteroscedastic model . . . . .	51
2.3.2	Performance comparison of proposed spatiotemporal forecasting model . . . . .	52
2.4	Conclusions . . . . .	54
<b>3</b>	<b>OVER-LIMIT RISK ASSESSMENT</b>	<b>59</b>
3.1	Preamble . . . . .	59
3.2	Temperature-augmented power system model . . . . .	60
3.2.1	Temperature-augmented load flow model . . . . .	61
3.3	Temperature-augmented probabilistic load flow using Monte-Carlo simulation . . . . .	65
3.3.1	Generation of random samples for Monte-Carlo simulation . . . . .	65
3.3.1.1	Random generations from a discrete random variable . . . . .	65
3.3.1.2	Random generations pertaining to correlated continuous random variables . . . . .	66
3.4	System over-limit risk indices . . . . .	67
3.4.1	Over-limit probability calculations . . . . .	67
3.4.2	Over-limit severity calculations . . . . .	68
3.4.3	Calculation of risk indices . . . . .	70
3.5	Case study and discussion of results . . . . .	70
3.5.1	Power system description . . . . .	72
3.5.2	Impact of increased penetration of photovoltaic generations on the statistics of result variables . . . . .	73

3.5.3	Effect of variations of temperature-augmented system model parameter values on the statistics of result variables . . . . .	77
3.5.3.1	Practical implementation of temperature-augmented probabilistic load flow . . . . .	79
3.5.4	Over-limit risk assessment . . . . .	79
3.6	Conclusions . . . . .	83
<b>4</b>	<b>SENSITIVITY MATRIX BASED STEADY-STATE ANALYSIS</b>	<b>85</b>
4.1	Preamble . . . . .	85
4.2	Sensitivity matrix based power system model . . . . .	85
4.2.1	Evaluation of sensitivity matrices of temperature-augmented power system model . . . . .	87
4.3	Proposed hybrid methods . . . . .	89
4.3.1	Extended cumulant method . . . . .	90
4.3.2	Gaussian mixture approximation . . . . .	92
4.3.3	Sequence operation theory . . . . .	93
4.3.4	Combined cumulant and Gaussian mixture approximation . . . . .	95
4.3.4.1	Cluster number selection approach . . . . .	96
4.3.5	Combined cumulant and sequence operation theory . . . . .	98
4.4	Case study and discussion of results . . . . .	101
4.4.1	Measures for accuracy comparison . . . . .	102
4.4.2	Accuracy comparison of probabilistic load flow results using traditional and modified power system models . . . . .	105
4.4.3	Investigation on the accuracy of proposed hybrid methods considering input correlations . . . . .	105
4.4.4	Performance assessment of the proposed hybrid methods in approximating multimodal probability distribution of result variables . . . . .	110
4.4.5	Spatiotemporal model based temperature-augmented probabilistic load flow simulations . . . . .	116
4.5	Conclusions . . . . .	119
<b>5</b>	<b>CONCLUSIONS AND FUTURE SCOPES</b>	<b>123</b>
5.1	Conclusions . . . . .	123
5.2	Future scopes . . . . .	125

A Multiple linear regression model in matrix form	127
B Non-parametric density estimation method	131
C Principal component analysis: an orthogonal linear transformation	133
D Estimation of variance coefficient	135
E Formulation of the temperature-augmented Jacobian matrices	137
F Effect of input correlations on the statistics of the result variables	141
Bibliography	143



# List of Figures

1.1	Research on PLF. . . . .	2
1.2	Time information for uncertainty modeling. . . . .	4
1.3	Classification of PLF methods. . . . .	10
1.4	General performance characteristics of prominent PLF methods. . . . .	17
1.5	The outline of the thesis. . . . .	18
2.1	Uncertainty modeling steps. . . . .	22
2.2	Sites from where PV generation data is collected. . . . .	23
2.3	Weather zones of Texas from where load power data is collected. . . . .	23
2.4	PV generation, load power, and ambient temperature time series for a few cases. . . . .	25
2.5	Pictorial description of the physical relevance of the regressors (a) $\sin \theta_S$ , (b) $\sin \gamma_S$ , and (c) $\cos \theta_I$ . . . . .	26
2.6	Variations of the regressors $T_1$ , $T_2$ and $T_3$ at 12 time instants for a year at Monkton. . . . .	29
2.7	Comparison of models in tracking periodic variation in (i) $PV_2$ at 1 pm, (ii) $PV_1$ at 7 am, and (iii) $PV_3$ at 6 pm. . . . .	36
2.8	Plots of statistical parameters of PV generations before and after pre-processing. . . . .	38
2.9	Plots of PMCCs among the PV generations before and after pre-processing. . . . .	38
2.10	Performance comparison of models (Model-1 and Model-2 refers to (Fan et al., 2014) and (2.1) respectively) in the region of low PV power generation. . . . .	39
2.11	Comparison of probability density plots of PV generation uncertainties for various cases. . . . .	40

2.12	Uncertainty modeling steps as applied to $PV_1$ at 9 am. . . . .	41
2.13	Plot of multiple linear regression model parameters at various time instants for $PV_1$ . . . . .	42
2.14	Comparison of models in tracking periodic variations in few cases of load power and ambient temperature. . . . .	43
2.15	Various steps in the extraction of dominant frequency orders in load power of North Central zone at 7 pm. . . . .	44
2.16	Plots of statistical parameters of load powers from Coast and North weather zones of Texas before and after preprocessing. . . . .	46
2.17	Plots of statistical parameters of ambient temperatures before and after preprocessing. . . . .	47
2.18	Plots of PMCCs between the load powers before and after preprocessing. . . . .	47
2.19	Plots of PMCCs between the ambient temperatures before and after preprocessing. . . . .	48
2.20	Uncertainty modeling steps for load powers of (a) North zone, Texas (2012-2013) at 3 pm and (b) West zone, Texas (2012-2013) at 7 pm. . . . .	49
2.21	Uncertainty modeling steps for ambient temperatures of (a) Lincoln, USA (2012-2013) at 10 am and (b) Mangaluru, India (2013-2014) at 2 pm. . . . .	50
2.22	Daily time stepped observed PV generation time-series. . . . .	52
2.23	Scatter plot of observed PV generations in Figure 2.22. . . . .	52
2.24	Scatter plot of residuals after transformed to standard Gaussian domain. . . . .	53
2.25	Time-series of principal components. . . . .	54
2.26	Comparison of RMSE. . . . .	55
3.1	Electro-thermal coupling effect in a transmission branch. . . . .	61
3.2	Flowchart for temperature-augmented load flow. . . . .	64
3.3	An example case to show the need of a severity function in the risk assessment. . . . .	69
3.4	Plots of severity functions. . . . .	69
3.5	Single line diagram of the PV-integrated New England 39-bus power system. . . . .	71
3.6	Comparison of probability density plots of PV generations for various penetration levels. . . . .	75

3.7	Cumulative probability plots of net load at buses 26, 27 and 28 for various PV penetrations. . . . .	75
3.8	Comparison of probability density plots of (a) $P_{\text{Loss},26-27}$ (5% penetration), (b) $ V_{26} $ (10% penetration), (c) $P_{L,2-3}$ (15% penetration) and (d) $P_{\text{Loss},26-28}$ (20% penetration) using PLF and TPLF. . . . .	78
4.1	Comparison of various sensitivity matrix based power system models.	87
4.2	Methodology for hybridization. . . . .	91
4.3	Steps to obtain cumulants of a linear combination of correlated continuous random variables. . . . .	92
4.4	Conversion of a probability sequence to cumulative probability plot and vice versa. . . . .	94
4.5	Flowchart for establishing the probability distribution of a result variable using univariate-CCGMA. . . . .	97
4.6	Sequence of steps to obtain PMCC between $Y^{\text{eG}}$ and $Y^{\text{eNG}}$ . . . . .	100
4.7	Schematic representation to obtain $u_1$ and $u_2$ of (4.23). . . . .	101
4.8	Comparison of probability density plots of $P_{L,28-29}$ using various PLF models. . . . .	106
4.9	Accuracy comparison of Gaussian Copula based addition type convolution operation to handle different parametric distributions and input correlations. . . . .	108
4.10	Accuracy comparison of Gaussian Copula based addition type convolution and subtraction type convolution operations to handle non-parametric distributions and input correlations. . . . .	108
4.11	Comparison of cumulative probability plots of $P_{L,100-103}$ with and without considering input correlation. . . . .	109
4.12	Comparison of cumulative probability plots of $P_{L,49-54}$ . . . . .	112
4.13	Comparison of cumulative probability plots of $P_{L,46-48}$ , $P_{L,64-61}$ , $Q_{L,45-46}$ and $Q_{L,64-65}$ . . . . .	114
4.14	PV array installation sites in Tamilnadu. . . . .	115
4.15	Single line diagram of PV integrated Indian utility 62-bus system. . .	117
4.16	A set of steps to obtain PV generation time series using the time series of irradiance, ambient temperature, and wind speed. . . . .	118
4.17	Comparison of cumulative probability plots (i) $P_{L,37-38}$ and (ii) $P_{L,41-42}$ . . .	119



# List of Tables

2.1	Technical details of the PV arrays. . . . .	22
2.2	Comparison of coefficient of variations of PV generations before and after preprocessing. . . . .	37
2.3	Estimated values of $t_0$ , and $H$ for various cases in Figure 2.14. . . . .	42
3.1	Use of probability distributions of TPLF result variables. . . . .	65
3.2	Probabilistic description of discrete load powers. . . . .	72
3.3	Base-case PMCC matrix defined among first 12 continuous input random variables. . . . .	73
3.4	Mean values of PV generations in pu for various penetration levels. . . . .	74
3.5	Comparison of PLF and TPLF results for various PV penetration levels. . . . .	76
3.6	Effect of temperature-augmentation on first two statistical moments of branch power flows in a few branches. . . . .	79
3.7	Effect of variations in TPLF model parameter values on average $e_\mu$ , and $e_\sigma$ of branch power flows. . . . .	80
3.8	Over-load probability and thermal over-load probability values of temperature dependent branches for 20% penetration. . . . .	80
3.9	Under-voltage probability and over-voltage probability values of load buses for 20% penetration. . . . .	81
3.10	System over-limit probability values with and without considering input correlation. . . . .	82
3.11	System RSOV, RSOL, and RSTOL values for various PV penetrations. . . . .	82
4.1	Orders of matrices and vectors used in (4.3). . . . .	88
4.2	PV generation locations and statistical descriptions. . . . .	103
4.3	Statistical description of continuous load power distributions. . . . .	103
4.4	Statistical description of discrete load power distributions. . . . .	104

4.5	Details of base-case correlations. . . . .	104
4.6	Details of discrete power generations. . . . .	105
4.7	Accuracy comparison of PLF results using traditional and modified model. . . . .	106
4.8	Different distribution types of input random variables for the accuracy comparison of Gaussian copula based sequence operations. . . . .	107
4.9	Three different input correlation conditions. . . . .	110
4.10	Effect of various input correlation conditions on the statistics of few result variables. . . . .	111
4.11	Comparison of ARMSEs of the proposed hybrid methods considering input correlations. . . . .	113
4.12	Comparison of absolute percentage errors and simulation times for various PLF methods considering base-case correlations. . . . .	113
E.1	Orders of sub-matrices. . . . .	140

## List of Abbreviations

ARCH	Autoregressive conditionally heteroscedastic
ARMSE	Average root mean square error
CCCFM	Combined cumulant and Cornish-Fisher method
CCGCM	Combined cumulant and Gram-Charlier method
CCGMA	Combined cumulant and Gaussian mixture approximation
CCSOT	Combined cumulant and sequence operation theory
CDF	Cumulative distribution function
CFM	Cornish-Fisher method
ECM	Extended cumulant method
EGM	Edgeworth method
GARCH	Generalized autoregressive conditionally heteroscedastic
GCM	Gram-Charlier method
GMA	Gaussian mixture approximation
GMM	Gaussian mixture model
PDF	Probability density function
PLF	Probabilistic load flow
PMCC	Product moment correlation coefficient
PSSA	Probabilistic steady-state analysis
PV	Photovoltaic
RSOL	Risk of system over-load
RSOV	Risk of system over-voltage
RSTOL	Risk of system thermal over-load
TPLF	Temperature-augmented probabilistic load flow





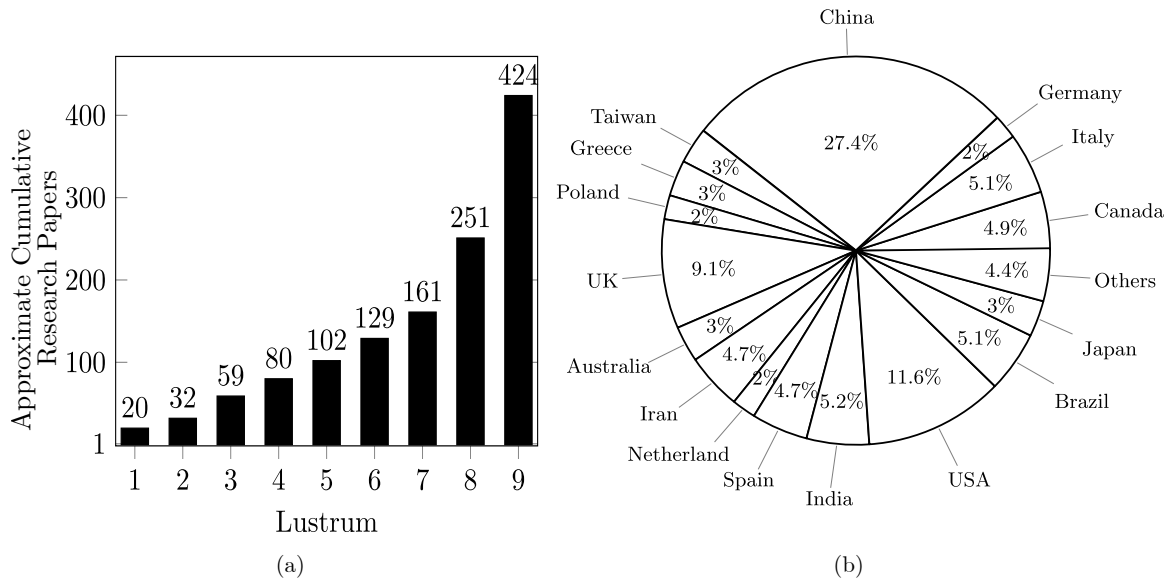
# Chapter 1

## INTRODUCTION

This chapter provides a brief overview of the issues addressed in this thesis. The research scopes stemming out of the critical review on probabilistic steady-state analysis (PSSA) using probabilistic load flow (PLF) followed by the contributions of the thesis are presented.

### 1.1 Background

In recent years, integration of renewable generations with electric power systems has gained much interest owing to the possible environmental and economic benefits. Predominantly, photovoltaic (PV) generations are the most important renewable resources to have experienced a globally increasing significance because of their ability to generate power in varying capacities and different voltage levels (Fan, 2012). Further, PV technology is cost-effective, and its commissioning is easy. However, PV output is uncontrollable as it depends on the Sun's position and other climatic factors. It is site specific, has diurnal and seasonal patterns. In short-term time-horizon, PV generation is rich in daytimes than in nighttimes whereas, in long-term, it is rich in spring than winter. It has the following major characterizing features: (i) predictability (expected change), (ii) uncertainty/randomness (unexpected change), (iii) intermittency (unplanned unavailability), (iv) uncontrollability (power output is not defined by system management), and (v) non-dispatchable due to (ii) and (iii). An increase in uncertainties under higher PV penetrations with their associated spatiotemporal dependencies cause system variables to exceed their allowable limiting values. These pose significant planning and operational challenges in the power systems.



**Figure 1.1:** Research on PLF.

A conventional deterministic steady state analysis ignores the aforesaid uncertainties and dependencies, hence, leads to either overinvestment or insufficient system reliability during planning and operation. Whereas, PSSA using PLF obtains the uncertainties in the result variables as derived from the uncertainties of the input random variables (Borkowska, 1974, Da Silva et al., 1990, Usaola, 2010). The various result variables in a power system include bus voltage magnitudes, branch power flows and losses, generator bus reactive powers, and slack bus powers. The primary requirements to perform PLF include:

- (i) Modeling of power system uncertainties and dependencies.
- (ii) Establishment of a power system model.
- (iii) Application of a method to solve PLF.

An accurate accomplishment of the above three PLF requisites helps in making critical decisions during power system expansion planning, operation (operational planning and real-time operation), and control. There exists vast literature addressing various issues in PLF. Nevertheless, improvements are still possible to make PLF more elegant. The ensuing sections unfold in detail about the state of the art PLF followed by the derived problem formulation.

## 1.2 State of the art in Probabilistic load flow

In this section, an overview of previous work on PLF is provided. A detailed review on specific topics is included in the corresponding chapters.

Ever since the inception of PLF in 1974 (Borkowska, 1974), it has been a rigorous research topic till date. Figure 1.1(a) and (b) respectively provide lustrum wise and country wise approximate research contributions towards PLF in the past 45 years (1974-2018). The following subsections elucidate the PLF requirements and applications in detail.

### 1.2.1 Modeling of input uncertainties and dependencies

One of the primary tasks in PLF is uncertainty modeling. Amongst several uncertainty modeling methods, application of a probabilistic method is most appreciated in the literature (Aien et al., 2016). Probabilistic uncertainty modeling defines the possible distribution of an input random variable with a strict bound based on their historical record. It can be broadly classified as, planning uncertainty and operational uncertainty (Papaefthymiou and Kurowicka, 2009). The former corresponds to the analysis of all possible combination of system inputs whereas; the latter corresponds to the forecast errors. On the one hand, planning uncertainties are quantified to make decisions regarding investments (and their economic merits) in locating adequate generation and transmission assets at suitable places and time; thereby ensuring reliable and efficient delivery of real power when no remedial actions are executed (Zhang and Lee, 2004). The quantification of operational uncertainties, on the other hand, is decisive in recognizing whether system operating limits are within acceptable reliability criteria. Then after, remedial actions are taken to decide on the change of system operating mode which in turn ensures reliability through power balance at all the time instants (Peng et al., 2015). The various remedial actions include load transfer, switching actions, protection/control schemes, temporary reconfigurations, etc.

In PLF, an input random variable  $X$  is expressed as,  $X = \mu_X + \varepsilon_X$  where  $\mu_X$  and  $\varepsilon_X$  respectively are the expected value and uncertainty component of  $X$ . Accurate uncertainty modeling can be developed based on the adequate historical data with acceptable accuracy. In general, the data inadequacy is due to the irregular source database update (missing values), whereas, the data inaccuracy is mainly due to the abnormal data (outliers).

	⊙ Time-horizon	⊙ Time-scale	⊙ Time-step	⊙ Actions/Decisions
Expansion planning	Long-term	Years-to-decade(s)	Yearly	Grid expansion plans, capacity credit assessment, etc.
	Medium-term	Years	Daily	Infrastructural investment plans
	Short-term	Months	Daily	Installation of protection systems, phase shifters, etc.
Operational planning	Long-term	Year(s)	Daily	Release of transmission equipment(s) for maintenance and scheduled repair
	Medium-term	Month(s)	Daily	Delaying or anticipating the commitment of transmission facilities
	Short-term	Day(s)-to-week	Hourly/daily	Modifying the operational strategy, postponing scheduled outages, etc.
Real-time operation	Long-term	Day-to-day(s)	Hourly	Unit commitment
	Medium-term	Minute(s)-to-day	15 min-to-1 hour	Economic dispatch
	Short-term	Second(s)-to-minute(s)	30 sec-to-5 min	Automatic generation control, storage control, assessment of ramping events, etc.

**Figure 1.2:** Time information for uncertainty modeling.

### 1.2.1.1 Time-horizons, time-scales and time-steps

The time-horizons (long-term, mid-term, and short-term), and time-scales (minutes, hours, days, weeks, months, years, several years, decades, etc.) that a study of interest focuses on, plays a significant role in uncertainty modeling. The degree of uncertainty increases significantly from a short time-scale in system operation to a larger time-scale in expansion planning. For example, based on the current operating conditions, a system operator has more confidence in forecasting for the next hour than the next month. The resolution of data in a time series has a significant effect on the uncertainty model. The historical data collected at an interval of one hour or more (e.g., daily, monthly, yearly, etc.) is referred to as a low-resolution data whereas, an ideal time series of one minute time step or less (e.g., half a minute, few seconds, etc.) is referred to as the highest resolution data. Figure 1.2 provides an overall knowledge about the time information for modeling uncertainty in various power system studies.

### 1.2.1.2 Review of uncertainty modeling and dependencies

The modeling of power system uncertainties in conventional and PV-integrated power systems is detailed as under.

- (i) Bus power uncertainties: The uncertainty modeling in conventional as well as PV-integrated power systems is as follows. The uncertainties in conventional power systems are due to load powers. Typically, PLF considers historical time series of inputs to describe the uncertainty components. Either, stationarity of the time series is ensured (Da Silva et al., 1987, 1990, 1991, Coutto Filho et al., 1991) or the expected value and the corresponding confidence interval of the future values are forecasted (Da Silva et al., 1990, 1991, Coutto Filho et al., 1991). The former is generally considered for operational planning whereas; the latter is for expansion planning. The daily peak loads of substations (each day at the same time) for a period of few months (typically, one to three months) corresponding to the period of annual peak are considered for uncertainty modeling. In (Da Silva et al., 1987), restriction on the length of time series to one month is to preserve data stationarity. Whereas, in (Da Silva et al., 1990, 1991, Coutto Filho et al., 1991) the daily peak loads collected for a period of two months have evident trend effects. The stationarity is attained through curve fitting using a set of mathematical functions (Coutto Filho et al., 1991). Since, the data corresponds to a relatively shorter period, the nonlinearities in the load patterns in relation to seasonal change of weather variables is not accounted for; efforts were made only to remove the trend effect from data. The expected value of load power raises by 8% in case of a year ahead operational planning (Da Silva et al., 1987). The short-term (operational) uncertainties are related to short-term factors such as, environmental conditions, social effects, etc. (Coutto Filho et al., 1991). In (Da Silva et al., 1990, Coutto Filho et al., 1991), the uncertainties are due to long-term factors such as demographic growth, economic activities, etc. The annual substation peak loads for several past decades are considered for modeling. The load probability distribution for the time horizon is forecasted using the method of linear regression (Da Silva et al., 1990). In all the above cases, the data pertaining to holidays and weekends are removed as they cause a significant reduction in the peak load values. The obtained stationary samples can reasonably be approximated to a continuous density function

within a certain range if samples are very close to each other (i.e., to have small variance) (Da Silva et al., 1987). This procedure greatly abates computational efforts while applying an analytical method for PLF. In case of samples not very close together (i.e., to have large variance), a possible approximated discrete distribution with definite number of impulses is constructed where the individual probability values sum to unity (Da Silva et al., 1987, Coutto Filho et al., 1991). The dependencies among load powers usually being linear, product moment correlation coefficient (PMCC) is used to measure the strength of linear dependencies (Da Silva et al., 1990). The PMCC values can be directly obtained from the stationary samples in operational planning studies and from the annual peak load series in long-term expansion planning.

For power systems integrated with PV generation, it is very much essential to evaluate the permissible PV penetration using a probability method which is mainly affected by PV generation distribution, seasonal patterns, fluctuation characteristics, and the associated dependencies among nearby PV generations. In system planning, uncertainty associated with PV generations should be quantified for the determination of the variability of the system power flow, which is essential for system dimensioning. In case of an operational study, uncertainty quantification of PV generation is central in the system management, e.g. optimal operation under high penetration (Papaefthymiou and Kurowicka, 2009).

In recent days, a majority of the rooftop installed PV generations located closer to the loads are integrated to the transmission systems. However, such installations are suggested not to provide voltage control as per IEEE standard 1547 (Eftekharnjad et al., 2013), and thus they act as active power sources alone. This leads to an immediate increased concern towards the bus voltage magnitude violations under higher penetrations (Fan, 2012). Further, uncertainties arising from multiple PV generations and other input random variables, as well as their multiple dependencies can cause the result variables to exceed their limiting values. On this line, PLF is carried out to analyze the influence of PV generation uncertainty on the result variables at noon as PV generation has maximum coefficient of variation (calculated as the ratio of standard deviation to mean value) at noon (Fan et al., 2012). PV generation and load power historical data collected each day at noon for two years are considered for uncertainty

modeling. Such a study can be used for planning-based applications like power balance, peak regulation, etc. Unlike, the conventional model where only peak load powers are considered as the worst case condition, the load power uncertainty model as adopted in (Fan et al., 2012) can better assess the possible state of a power system. While peak load values cause low voltage and generation deficiency, light loads give rise to overvoltage conditions (Fan, 2012). The study carried out in (Fan et al., 2012) is an example case to evaluate the influence of PV generation and other input uncertainties at a particular time instant. The risk of steady-state bus voltage violation and branch power flow exceedance under various PV penetrations and correlations are quantified using over-limit probability indices.

In an operational framework, probabilistic forecasting of expected PV generation and other input random variables are required to be accomplished through strategic decisions (Papaefthymiou and Kurowicka, 2009). When such uncertainties at multiple sites are considered in the analysis, spatial and temporal dependencies are required to be modeled for efficient extraction of critical information for reliable system planning and operation (Le et al., 2015).

- (ii) Network uncertainties: The uncertainties in power system network arise either from the outage of any of its components or branch power variations due to environmental factors. Such type of uncertainties is reviewed underneath.
  - (a) Topology uncertainties: Generally, the probability of network element outages are neglected and the probability of basic network topology is assumed to be unity. However, the steady state contingency analysis accounts for the network uncertainties (Da Silva et al., 1987) which are generally characterized by a discrete distribution. A widely used approach to model branch outage uncertainty is to consider an outage being replicated by injecting fictional powers at the connecting buses (Hu and Wang, 2006). Formulations of component outages using optimization methods are deliberated in (Ozdemir et al., 2003, Ceylan et al., 2015).
  - (b) Branch parameter uncertainties: The transmission branch impedance (estimated at a specific temperature) is assumed as constant in most of the PLF studies. However, branch impedance varies with respect to the conductor temperature which in turn is influenced by few probabilistic factors

such as, branch current, ambient temperature, and other environmental aspects (Frank et al., 2013). The branch parameter uncertainty in PLF was first considered in (Su, 2005b). The series and shunt parameters of the branch are assumed to follow uniform and binary distributions respectively. In (Kang et al., 2003), authors have presented a PLF analysis considering the variation of power consumption due to temperature effect. Branch temperature variation due to the variability of branch power flow is presented in (Schlapfer and Mancarella, 2011).

### **1.2.2 Power system models for probabilistic load flow**

PLF uses either the actual algebraic nonlinear equations (non-linear model) or sensitivity matrix-based models (a simplified version of the actual model) for better computational efficiency. Depending on the system type, i.e., transmission or distribution, the modeling formulations differ. PLF was first implemented on a DC model (Borkowska, 1974, Allan et al., 1974). Soon after, four complex linearized formulations were deliberated (Allan and Al-Shakarchi, 1976, 1977). Erroneous estimations of expected values and computational inefficiency are the main pitfalls of these formulations. To preserve both accuracy and computational efficiency, a sensitivity coefficient based linearized formulation using Taylor’s series expansion is developed (Allan et al., 1981). The authors in (Yuan et al., 2011) have analyzed that accuracy of the PLF solution is sensitive to the choice of linearization point, and the linearization corresponding to the expected values of input random variables yields comparatively better results. For a higher level of input uncertainties, such a model results in an inaccurate estimation of the tail of distributions of result variables. A linear model accounting non-linear effects with the help of multiple linearization points is also focused (Allan and Da Silva, 1981). In (Brucoli et al., 1985), a second order term of Taylor series expansion is used for improved accuracy of results. The authors in (Frank et al., 2013) have introduced a temperature-augmented power system model by exploiting the electro-thermal coupling effect of transmission branches.

### **1.2.3 Methods to solve probabilistic load flow**

The PLF obtains the probability distribution of the result variables using an uncertainty handling method. For this, there are numerous methods proposed in the

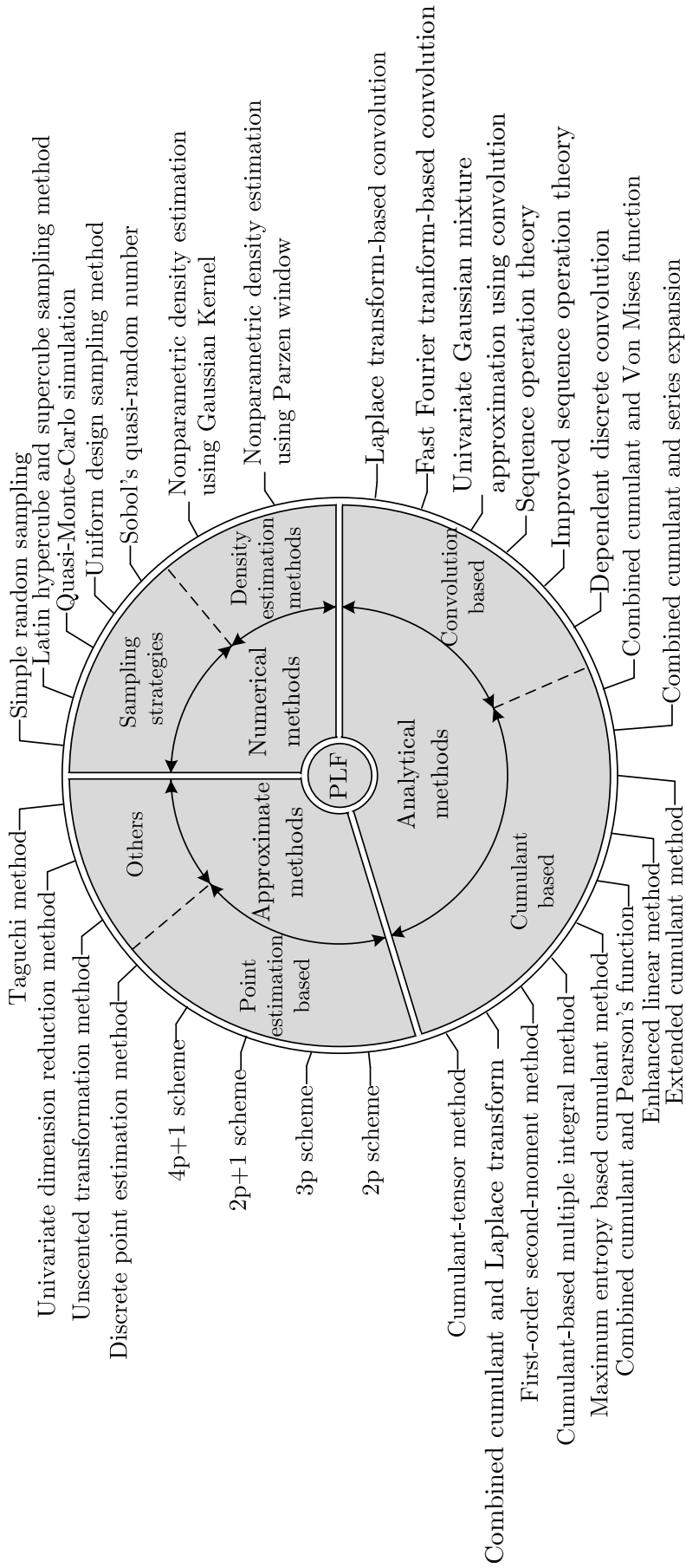


literature. For a quick reference, various PLF evaluation methods are classified in Figure 1.3. Apart from those in the classification, hybrid methods combining more than one of the above methods have gained additional interest.

### 1.2.3.1 Numerical method

A numerical method such as Monte-Carlo simulation is generally used when the model describing input random variables and the result variables are complex. It relies on repetitive random sampling and statistical analysis to compute the probability distributions of the result variables without requiring any simplifications of the original non-linear load flow equations (Aien et al., 2016). It is broadly accepted that Monte-Carlo simulation with simple random sampling is the most accurate method and is used as the benchmark to validate analytical methods and approximate methods. The tremendous computational cost is its main setback. Further, Monte-Carlo simulation results depend on the number of simulations which is independent of the system size (Allan and Da Silva, 1981). Any random selection of sample number may either be large or insufficient for the accuracy of the obtained results. Hence, it is appropriate to evaluate a coefficient of uncertainty to determine the convergence of Monte-Carlo simulation (Fan, 2012). The use of sampling methods such as Latin hypercube (Yu et al., 2009), importance sampling (Huang et al., 2011), Latin supercube (Hajian et al., 2013), uniform design (Cai et al., 2014), and Quasi-Monte-Carlo simulation (Cui and Franchetti, 2013) reduces the computational burden of simple random sampling based Monte-Carlo simulation to some extent. The use of smoothing properties of the linear diffusion process in (Soleimanpour and Mohammadi, 2013) reduces the number of trials required to obtain reasonable results. A graphics processing unit-accelerated Monte-Carlo simulation, by offloading the computational burden to graphic processor, suffices for real-time applications (Zhou et al., 2018).

Depending on whether the associated dependencies are linear or nonlinear, the methods implemented to generate random samples in Monte-Carlo simulation are also different. In the presence of only Gaussian input random variables, Cholesky decomposition based technique is employed to obtain the correlated random samples. For a mixture of Gaussian and non-Gaussian input random variables, two well-established techniques such as, Nataf transformation, polynomial normal transformation are suitable (Cai et al., 2015). Copula function based techniques are appropriate for modeling non-linear dependencies among input random variables. Finally, the underlying dis-



**Figure 1.3:** Classification of PLF methods.

tributions of the result variables can be obtained using Kernel or Parzen window function based nonparametric density estimation methods (Soleimanpour and Mohammadi, 2013, Rouhani et al., 2016).

### **1.2.3.2 Approximate method**

Approximate methods can permit non-linear analysis; they provide approximate statistical information of the result variables. The approach is similar to Monte-Carlo simulation requiring a lesser number of deterministic load flow calculations. The point estimate based approximate methods deal with the statistics of input random variables and does not require the complete knowledge about their probability distributions. Two-point estimation method is the ever simplest point estimation method used in the literature. The variants of point estimation method and their performance comparison is carried out in (Su, 2005b, Morales and Perez-Ruiz, 2007). The computation time of various point estimation methods are directly related with the number of input random variables and is sensitive to system complexity while the estimation of higher order moments is erroneous. The unscented transformation method generates the suitable number of samples (sigma points) deterministically to approximate the probability distributions (Aien et al., 2012). It is computationally efficient than point estimation method and can handle input correlations more effectively. The aforesaid methods obtain the statistical parameters of result variables and use a series expansion method to approximate their probability distributions. Few other approximate methods are discussed in (Zou and Xiao, 2014, Hong et al., 2016).

The point estimation methods proposed by Rosenblueth, Li and Harr can easily handle correlation among input random variables whereas; Hong's point estimation method has a strong assumption of independent inputs (Morales and Perez-Ruiz, 2007). This drawback is subdued by combining point estimation method with Gaussian transformation technique (Morales et al., 2010). The correlation in (Zou and Xiao, 2014) is incorporated by transforming correlated input random variables into independent standard Gaussian domains.

### **1.2.3.3 Analytical method**

An analytical method obtains PLF results quickly by operating either on the probability distributions of input random variables or their cumulants.

- (i) Convolution-based analytical methods and variants: Conventional analytical methods assume simplifications such as linearization of the system model, independence, and Gaussian input random variables (Allan et al., 1974, Allan and Al-Shakarchi, 1976, 1977). For a higher number of input random variables, numerous convolution operations are required to be performed using Laplace transform (Allan et al., 1974). A new discrete frequency domain convolution by applying fast Fourier transform is proposed and found to provide fast and precise results (Allan et al., 1981). The idea of using sequence operations to accomplish arithmetic calculations among the univariate input random variables in PLF is inspired from (Kang et al., 2002). The assumption of independent input random variables is the main disadvantage of the approach. Then after, the concept of multivariate joint probability sequence is introduced in (Liu et al., 2017). Dependent discrete convolution method as proposed in (Zhang et al., 2016) incorporates input dependencies in analysis with the help of copula density functions. A Gaussian mixture model (GMM)-based approach using convolution is applied for PLF under the assumption of independent input random variables (Sirisena and Brown, 1983). It takes a great deal of computational time to accomplish the convolution among input random variables. Then after, the input correlations in GMM-based PLF are accounted in (Valverde et al., 2012).
- (ii) Cumulant-based analytical methods and variants: Cumulant method is a well established analytical method which obtains cumulants of a linear combination of correlated input random variables in a single step (Sanabria and Dillon, 1986, Hoese and Garcés, 1999). The main advantage of this method is its high computational efficiency for larger systems (Fan et al., 2012). Various mechanism to incorporate input correlation in cumulant method is detailed in (Fan et al., 2012, Hoese and Garcés, 1999, Cai et al., 2012). From the obtained cumulants of result variables, probability distributions can be approximated with the help of series expansion methods such as Gram-Charlier method (GCM), Cornish-Fisher method (CFM), and Edgeworth method (EGM) (Fan et al., 2012), Von Mises function (Sanabria and Dillon, 1986), Pearson’s functions (Abdullah et al., 2013), and optimal fitting of distribution based on maximum entropy method (Williams and Crawford, 2013). Few other methods under this category includes, combined cumulant and Laplace transform (Kenari et al.,

2017), cumulant-tensor (Amid and Crawford, 2018) and cumulant based multiple integral method (Wu et al., 2016).

The input correlation can be included in cumulant method by various formulations as described below.

- (a) Formulation based on joint distribution: This method necessitates the availability of joint probability density function or joint probability mass function of input random variables. In reality, it is cumbersome to obtain joint distribution of input random variables (Fan et al., 2012). Nataf model and Morgenstern model are used to construct the joint distribution from the prescribed marginals and covariances (Liu and Der Kiureghian, 1986). Morgenstern model parameter is either constant or dependent on the shape parameters of the marginal distributions. Nataf model provides a much wider range of PMCC whereas; a limit imposed on the permitted value of PMCC is the main disadvantage of Morgenstern model. Efforts have been made (Papaefthymiou and Kurowicka, 2009) to construct joint distribution from the marginal distributions using copula method. However, the method is convenient for a bivariate case with increasing intricacy for a multivariate case.
- (b) Formulation-based on orthogonal transformation: Cholesky factorization based orthogonal transformation method accurately transforms Gaussian correlated random variables into uncorrelated sets, but the significant error is introduced when applied to multivariate non-Gaussian cases (Chen et al., 2015). Rotational linear orthogonal transformation technique accurately estimates the cumulants of correlated non-Gaussian input random variables for orders higher than two (Hoese and Garcés, 1999).
- (c) Extended correlation handling methods: There are two approaches discussed in the literature to incorporate input correlation in cumulant method (Hoese and Garcés, 1999, Cai et al., 2012). Although the correlations among Gaussian and non-Gaussian input random variables can be considered, the methods are not free from the use of a series expansion while approximating the probability distributions of the result variables.

Convolution-based analytical methods are computationally less efficient, and most of them assume input random variables are independent (Allan et al., 1974, 1981,

Kang et al., 2002). The dependent discrete convolution method that adopts copula density function to incorporate input dependency is computationally less efficient for an increased number of input random variables (Zhang et al., 2016). Further, the chosen small sequence intervals to reduce discretization error increases the computational time drastically. The cumulant-based analytical method in (Sanabria and Dillon, 1986) that separately deals with continuous and discrete parts of result variables assumes the continuous input random variables to follow Gaussian distributions. The analytical methods in (Fan et al., 2012) use series expansion methods. Although series expansion methods are computationally more efficient than Von Mises function, the performance of series expansion based cumulant method worsens in approximating multimodal probability distributions of result variables. Maximum entropy algorithm optimally fits the distribution based on its entropy. It overcomes the demerits of series expansion methods but for the reconstruction of multimodal distributions, a higher number of moments are required to be estimated (Williams and Crawford, 2013). Both Lagrange multiplier calculation method using Newton solver and MATLAB minimization functions to obtain maximum entropy has the limitations for the higher number of moments. The former is sensitive to the chosen initial conditions whereas the latter suffers from the computational limitations of MATLAB. The analytical method-based PLFs as discussed in (Fan et al., 2012, Hoese and Garcés, 1999, Cai et al., 2012, Abdullah et al., 2013) applies various mechanisms for modeling input correlations. There are no clear indications in the literature about the applicability of combined cumulant and Laplace transform, cumulant-tensor, cumulant based multiple integral method in approximating multimodal probability distributions of result variables.

#### **1.2.4 Applications of probabilistic load flow**

PLF is a vital tool for PSSA of power systems during the planning and operational studies. Various other applications of PLF that has not been covered in sections 1.2.1 through 1.2.3 are briefly reviewed underneath.

With the help of PLF, the dimensioning and setting of voltage controlled devices (switched capacitors, static reactive compensators, and tap-changing transformers) achieve a better voltage control in (Hatziargyriou and Karakatsanis, 1994, 1997, Hatziargyriou et al., 2005, Su, 2005a).

Three-phase PLF evaluates the steady-state condition of a power system considering unbalance in load and other components. The interaction between unbalanced three-phase source and high voltage DC transmission system is studied in (Caramia and Varilone, 1998). PLF in a system with unbalanced load is discussed in (Caramia et al., 1999). The deterministic and stochastic components of three-phase real and reactive powers are modeled separately in (Wang and Pierrat, 2001) for the probabilistic assessment of voltage unbalance factor. Voltage regulation problem due to the inclusion of PV generation is addressed in (Gomez-Gonzalez et al., 2013, Alam et al., 2013). A PLF considering asymmetrical line parameters and the unbalanced load is studied in (Ran and Miao, 2016). A probabilistic model to describe the uncertainties associated with harmonics is developed to obtain the steady-state network voltages at harmonic frequencies (Mohammadi, 2015).

PLF is also applied for the maintenance, operation, and expansion of electric railway systems (Ho et al., 2003, 2004, 2005). Starting with a simplified model based on feeding current, return conductor current and train current (Ho et al., 2003), few other studies considering train position as random variable (Ho et al., 2004) and train voltage model under various feeding systems (Ho et al., 2005) are carried out for PLF in electric railway system.

### 1.3 Research gaps

Based on the critical review carried out in the previous section, the following research gaps are worth highlighting.

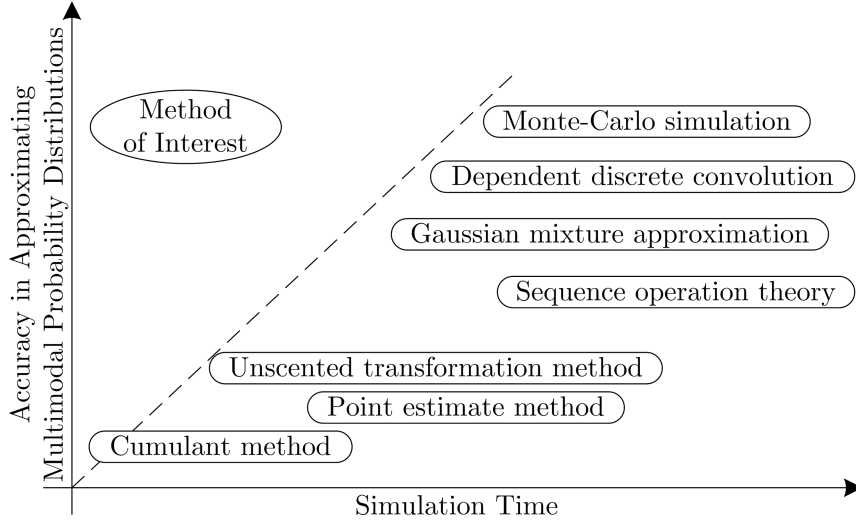
Firstly, it is learnt that the existing methods for characterizing the input uncertainties have not adopted an adequate preprocessing to filter out the deterministic components from the historical records of various inputs such as load power, PV generation, ambient temperature, etc. As geographical, environmental and social factors profoundly influence the inputs mentioned above, a thorough understanding of these physical phenomena to ascertain the actual influencing factors and their level of influence is essential to predict the complex deterministic pattern in data. Further, a multi-time instant preprocessing is necessary to ensure the extraction of actual unpredictable component of uncertainty at various time instants. Such a study considering the data of various places by developing generic models for the characterization of unpredictable uncertainty is worth investing. Further, a realistic spatiotemporal model

capable of extracting the statistical properties from historically observed data for the generation of future scenarios is also essential.

Secondly, none of the temperature related PLF studies have explored the effect of ambient temperature uncertainty on branch parameters. However, it is found that the temperature-augmented load flow accounting for the temperature effect has a dominant impact on branch power flows and losses. On this line, it would be interesting to analyze the uncertainty influence of ambient temperature on PLF result variables. Further, the over-limit risks due to the uncertainties in ambient temperature and PV generation in case of a PV-integrated power system considering input dependencies can help in providing a better insight into the system reliability. Except that the PV generation has maximum coefficient of variation at noon, the other input random variables have maximum coefficient of variation values at different other time instants. Hence, a more realistic risk assessment can only be accomplished by performing PLF simulations not just at noon but also at the other time instants.

Thirdly, the research attention towards proposing a new PLF method to accurately approximate multimodal probability distributions of result variables is highly essential. Although simple random sampling-based Monte-Carlo simulation along with a nonparametric density estimation approach is suitable in the present context, for an operational study requiring a timely decision, its application is not feasible as it is computationally burdensome and thus, use of an analytical method is an appropriate choice. There is no single analytical method in the literature that can accurately approximate multimodal probability distributions with reduced computational effort while considering input correlations. The available analytical methods either fail to accurately approximate multimodal probability distributions of result variables by including the input correlation effect or require higher computational time to approximate probability distributions of result variables. Further, it is observed that none of the PLF studies have considered the correlation effects among a mixture of Gaussian, non-Gaussian and arbitrary distributions. One of the ways to address the above concerns is through hybridization. Wherein, the chosen methods for hybridization not only should yield the desired overall performance but also subdue the shortcomings of the individual techniques. In a nutshell, an improved hybrid method is expected to fit into the place indicated as “Method of Interest” [refer Figure 1.4] which is of interest in this thesis.





**Figure 1.4:** General performance characteristics of prominent PLF methods.

## 1.4 Research objectives

Based on the research gaps, the aim of the thesis is to “develop models and methods for PSSA of PV integrated power systems”. As a part of it, the following issues are addressed:

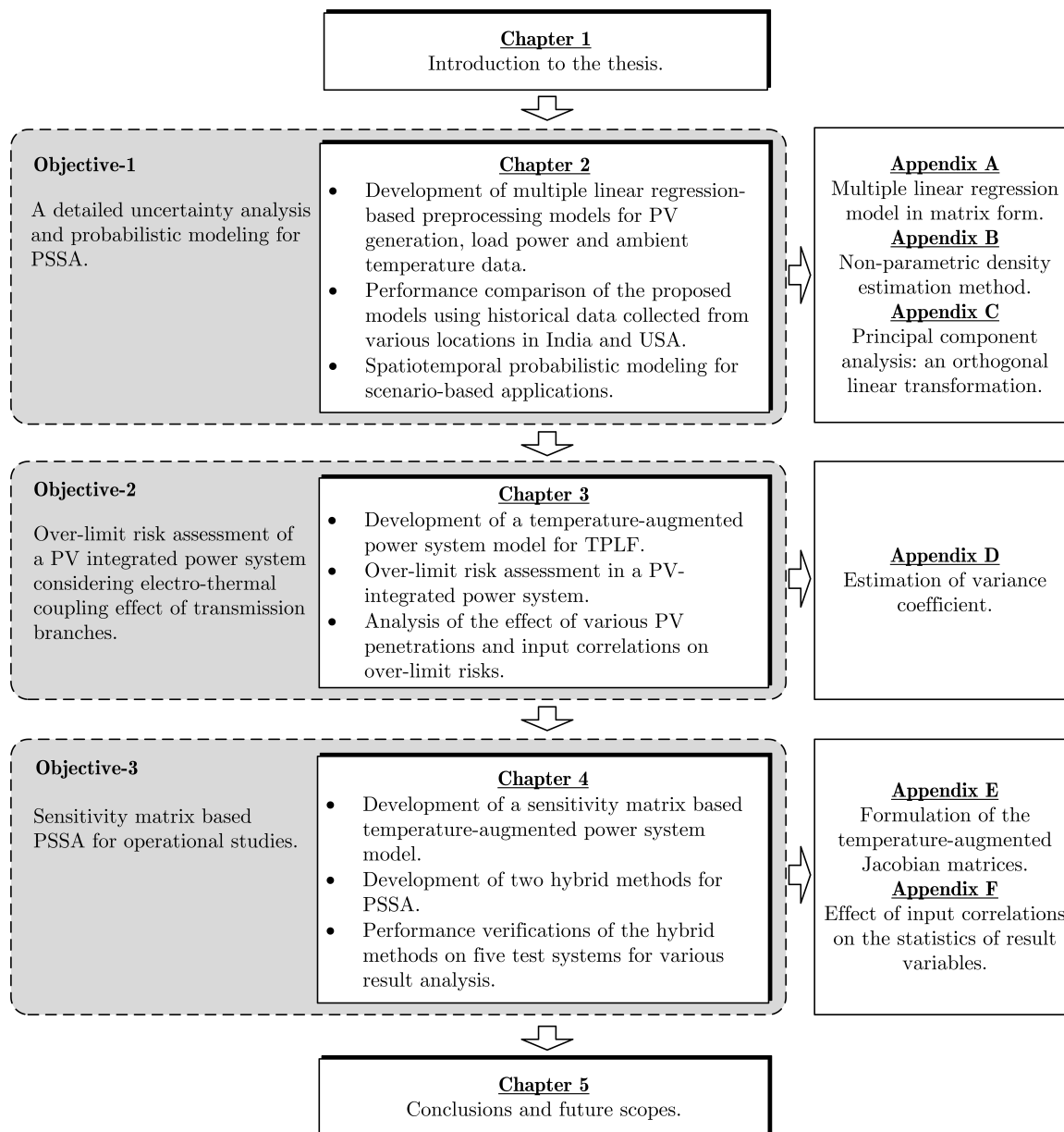
- (i) A detailed uncertainty analysis and probabilistic modeling for PSSA.
- (ii) Over-limit risk assessment of power systems under various PV penetrations and input correlations considering electro-thermal coupling effect of transmission branches.
- (iii) Sensitivity matrix based PSSA for operational studies.

## 1.5 Structure of the thesis

The thesis includes five chapters and six appendices. The outline of the thesis is briefly highlighted in Figure 1.5. A detailed organization is elucidated underneath.

Chapter 1 presents the background and motivation of the thesis along with the relevant literature review. Following the research gaps, the contributions and thesis outline are provided.

Chapter 2 extensively analyzes the uncertainty characteristics and correlation effects in historical PV generation, load power, and ambient temperature data of five



**Figure 1.5:** The outline of the thesis.

years duration. Based on the various observations of the complex periodic patterns in the time series of aforesaid inputs and through a thorough understanding of their physical process, potential regressors for multiple linear regression model to trace the predictable components are ascertained. A multi-time instant preprocessing using the above regression models is carried out resulting in the unpredictable components of these uncertain inputs. Finally, a scenario-based multivariate spatiotemporal model is developed by adopting multi-time instant preprocessing, transformation techniques, principal component analysis and a suitable time series model.

Chapter 3 develops a temperature-augmented power system model to perform PLF. A detailed calculation of the model parameters is elucidated. The system over-limit risk indices are formulated, and risk assessment is performed in PV-integrated New England 39-bus power system using temperature-augmented PLF (TPLF). The effect of variations of TPLF model parameters on the statistics of result variables is also discussed. The probability distributions of PV generations under various PV penetrations are compared. Finally, the effects of PV penetrations and input correlations on over-limit risks of result variables are discussed through detailed result analyses.

Chapter 4 analyzes various results such as (i) comparison of the accuracy of different sensitivity matrix based power system models, (ii) investigation on the accurate handling of input correlations by the proposed hybrid methods, and (iii) accuracy of the proposed hybrid methods in approximating multimodal probability distributions of result variables using PLF and TPLF. The above verifications are carried out on five different test systems considering a variety of distribution types and correlations.

Chapter 5 summarizes the conclusions drawn from Chapter 2 through 4. Finally, few possible extensions of this research work are suggested.

## Publication based on the chapter

### Refereed journal publications

1. Prusty, B. R. and Jena D. (2017). 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*, 69:6677-6683. (Elsevier)

# Chapter 2

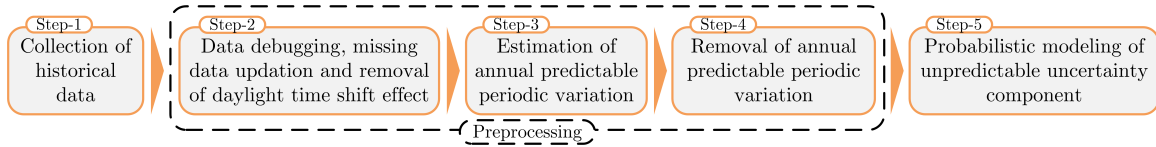
## UNCERTAINTY ANALYSIS AND MODELING

### 2.1 Preamble

Uncertainty analysis is an essential step in probabilistic load flow (PLF) to quantify the unpredictable components of input random variables based on their historical data whereas; a suitable probabilistic model defines their possible distributions. This chapter aims at devising suitable preprocessing approaches to characterize the uncertainties of PLF inputs. The suggested multi-time instant preprocessing approaches based on multiple linear regression models are helpful in studies where information regarding the time of the year is vital. Further, the proposed preprocessing is utilized for the analytical characterization of the spatiotemporal information to obtain the most likely photovoltaic (PV) generation time series for a future time-horizon.

### 2.2 Uncertainty characterization

The probabilistic characterization of uncertain variables is an essential step before any probabilistic power system analysis. Unlike studies that consider specific parametric distributions (either assumed or decided by the analyst) to quantify uncertainties, in this study, based on historical observations, realistic uncertainty models are developed. A set of steps as described in Figure 2.1 is applied to historical observations of uncertain variables collected at different time instants and from various locations.



**Figure 2.1:** Uncertainty modeling steps.

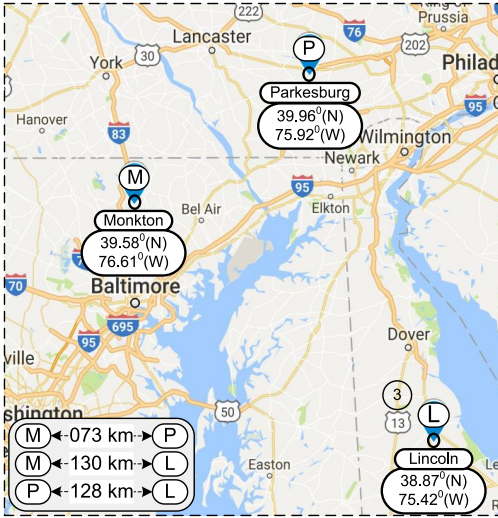
## 2.2.1 Collection of historical data

The PV generation data are collected from rooftop PV installations of three nearby locations [refer Figure 2.2] in the USA (Monkton (M), Parkesburg (P) and Lincoln (L)) (PVOUT, 2016). They are respectively denoted as  $PV_1$ ,  $PV_2$  and  $PV_3$ . The load power data are collected from eight weather zones of Texas [refer Figure 2.3], i.e., Coast (C), East (E), Far West (FW), North (N), North Central (NC), South (S), South Central (SC) and West (W) (ERCOT, 2016). The ambient temperature data are collected from Lincoln of USA (WUNDER-GROUND, 2016) along with three locations in India (Berhampur (B), Coimbatore (CO), Mangaluru (MA)) (NREL, 2017). The PV generation, load power and ambient temperature data corresponding to the years 2012 to 2016 are collected from the USA whereas, the Indian ambient temperature data corresponds to the years 2010 to 2014. The data are collected at an hourly time step excluding the leap days. The technical details of the three PV arrays are provided in Table 2.1.

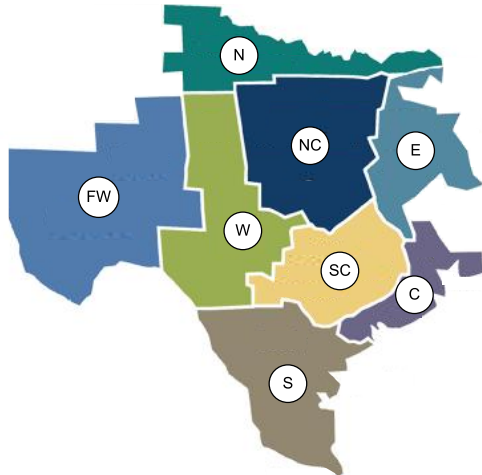
**Table 2.1:** Technical details of the PV arrays.

Details	PV Array 1	PV Array 2	PV Array 3
Place	Monkton	Parkesburg	Lincoln
Standard test conditions	$T_{\text{Ref}}^{\text{P}} = 25^{\circ}\text{C}$	$T_{\text{Ref}}^{\text{P}} = 25^{\circ}\text{C}$	$T_{\text{Ref}}^{\text{P}} = 25^{\circ}\text{C}$
	$\beta_{\text{Ref}} = 0.0045^{\circ}\text{C}^{-1}$	$\beta_{\text{Ref}} = 0.0045^{\circ}\text{C}^{-1}$	$\beta_{\text{Ref}} = 0.0045^{\circ}\text{C}^{-1}$
	$\eta_{\text{Ref}} = 15.27\%$	$\eta_{\text{Ref}} = 13.60\%$	$\eta_{\text{Ref}} = 14.69\%$
Capacity (kW)	10	7.65	8.88
Module dimension (mm <sup>3</sup> )	1650 × 992 × 40	1580 × 808 × 40	1665 × 991 × 50
No of modules	40	48	34
Module maximum power (W)	250	185	225
Orientation	South ( $\gamma = 180^{\circ}$ )	South ( $\gamma = 180^{\circ}$ )	South-West ( $\gamma = 225^{\circ}$ )
Tilt angle (deg.)	30	25	30

Note:  $T_{\text{Ref}}^{\text{P}}$  = Reference cell temperature,  $\beta_{\text{Ref}}$  = Temperature coefficient for cell efficiency,  $\eta_{\text{Ref}}$  = Reference PV generation efficiency, and  $\gamma$  = Array azimuth angle.



**Figure 2.2:** Sites from where PV generation data is collected.



**Figure 2.3:** Weather zones of Texas from where load power data is collected.

## 2.2.2 Preprocessing

Preprocessing of historical data is an important step before uncertainty modeling. It consists of three significant steps (step-2, step-3 and step-4) as indicated in Figure 2.1. These steps are applied to the historically observed PV generation, load power, and ambient temperature data.

### 2.2.2.1 Data debugging, missing data updation and removal of daylight time shift effect

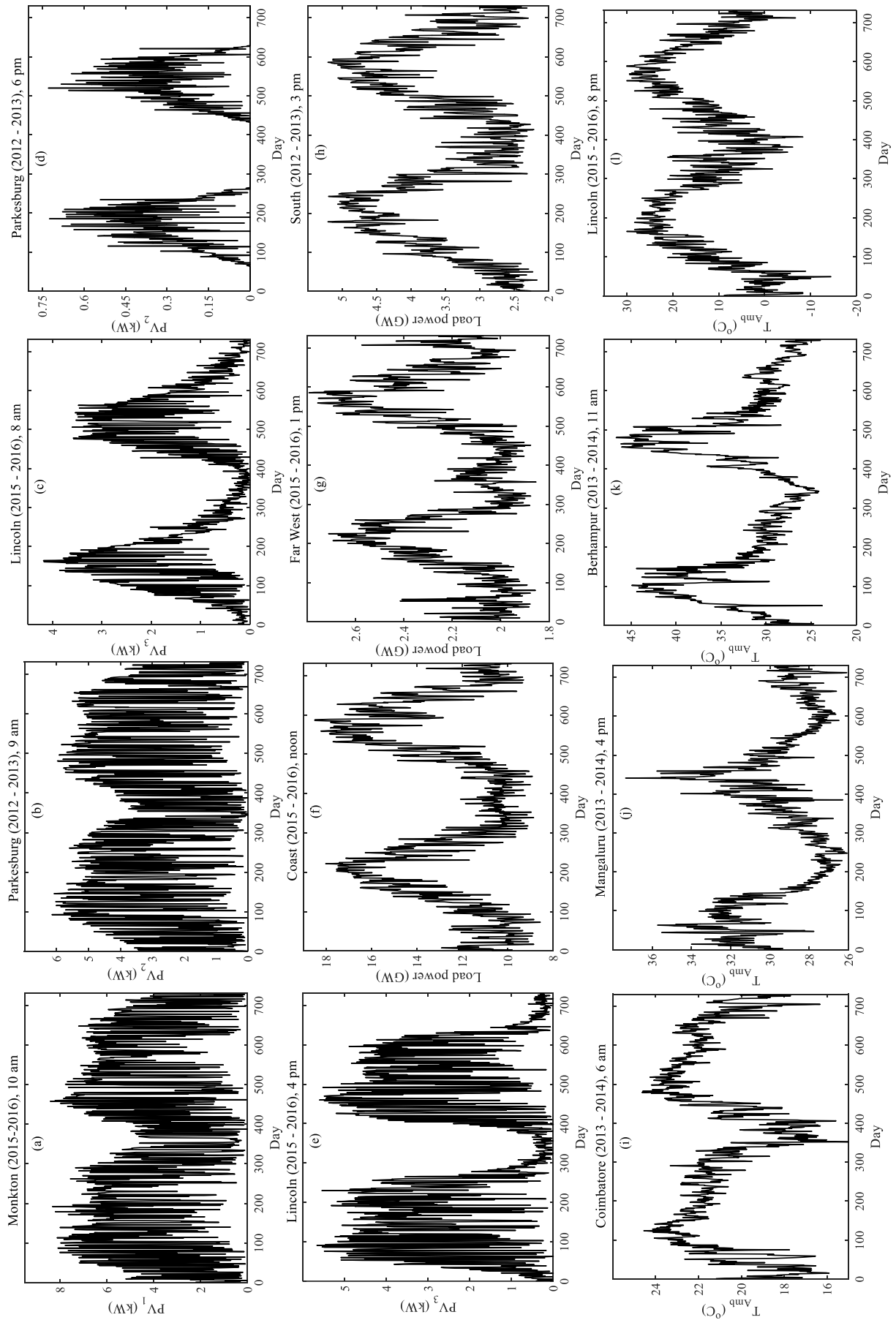
This step is imperative in preprocessing as historical observations are often vulnerable to missing values, and outliers. In few cases of the historical observations, missing values are noticed to be less than 2% of the total database. Such data are updated with the recent previous values. The effect of daylight time shifting (popularly known as daylight saving time) is removed from the data.

### 2.2.2.2 Estimation of annual predictable variation

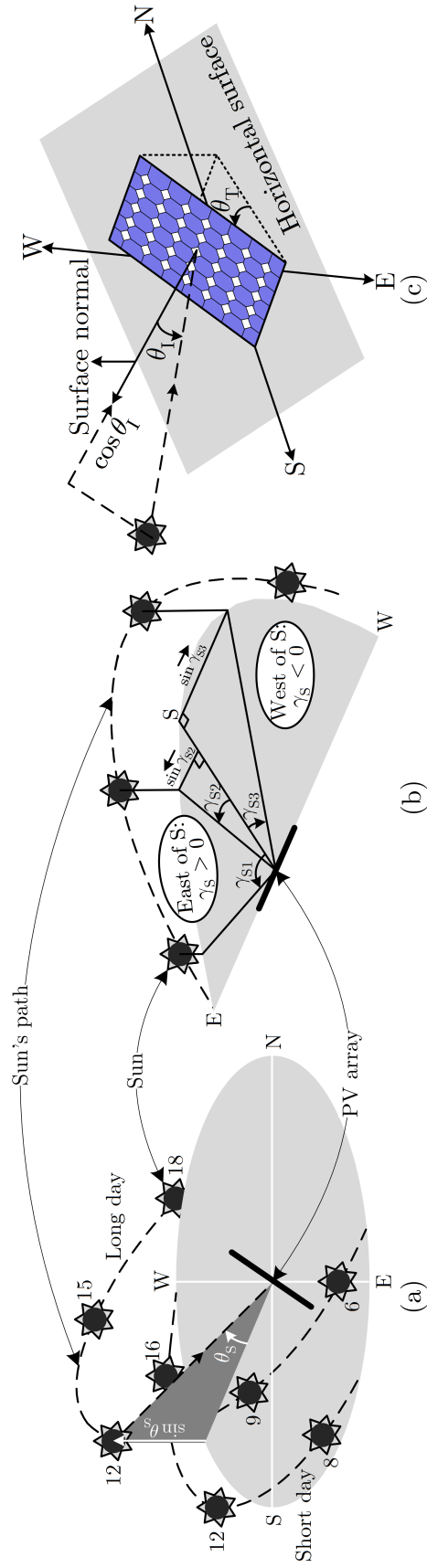
The historical daily time step data that correspond to a specific time of the year conflates deterministic and unpredictable components. The deterministic component refers to the annual periodic variation whereas; unpredictable component refers to the uncertainty. The presence of the deterministic components influence the probability distributions and correlations of the uncertainty components. Hence, an accurate estimation of the deterministic component is an essential task. In the literature, multiple linear regression-based models [refer Appendix A] are used to estimate the periodic variations. In such a model, the right choice of regressors is important for accurate predictions. The chosen regressors should have some theoretical relevance to the model. The potential regressors are chosen by analyzing the daily time step historical record of the variables at different time instants for any possible clues for the existence of predictable variations. The time series of a few cases are plotted in Figure 2.4. The duration and time of the year for each case are highlighted.

In case of PV generation, the changing solar position and the PV installation location affect the production pattern whereas, the complex patterns of load power at various places are due to social, environmental, and economic factors and the same for ambient temperature are due to latitude, altitude, seasonal change, vegetation,





**Figure 2.4:** PV generation, load power, and ambient temperature time series for a few cases.



**Figure 2.5:** Pictorial description of the physical relevance of the regressors (a)  $\sin \theta_s$ , (b)  $\sin \gamma_s$ , and (c)  $\cos \theta_T$ .

urbanization, etc. The historical data show strong yearly periodic patterns. Two distinct set of observations are made and are as follows:

*Observations from PV generation time series:*

- (i) At different locations and time instants, production patterns are extremely different [refer Figure 2.4(a)-(e)].
- (ii) The production patterns are multimodal [refer Figure 2.4(a) and (b)], skewed [refer Figure 2.4(c) and (d)] or both [refer Figure 2.4(e)].
- (iii) The production patterns have an evident decreasing trend [refer Figure 2.4(a)-(e)].

*Observations from load power and ambient temperature time series:*

- (i) At a particular place, but, at different time instants, the variation patterns are nearly the same. At different places, the variation patterns are different [refer Figure 2.4(f)-(l)].
- (ii) The periodic variations are highly complex [refer Figure 2.4(f)-(k)].
- (iii) The trend effects and additive seasonalities are dissimilar [refer Figure 2.4(f)-(l)]. For all the cases of load power, there exist increasing trend characteristics. However, the trend characteristics of ambient temperature are highly complex.

Based on the above observations, two different multiple linear regression models are proposed and for the same, modeling steps are discussed underneath.

### **2.2.2.3 Multiple linear regression model for photovoltaic generation**

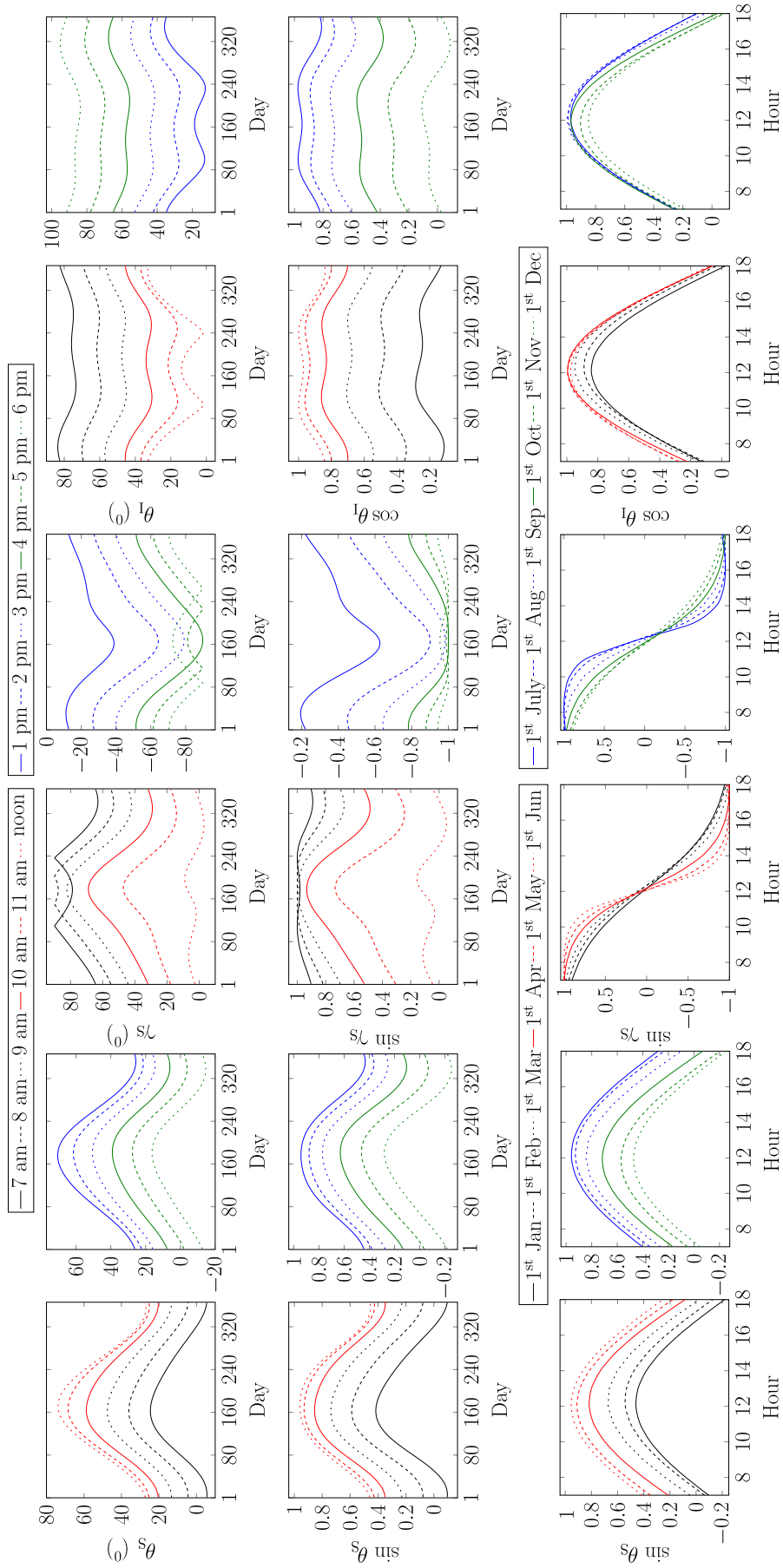
PV generation being greatly influenced by Earth-Sun geometry, a clear sky model using multiple linear regression suffices to trace the production pattern. The chosen regressors are selective trigonometric functions over solar elevation angle ( $\theta_S$ ), solar azimuth angle ( $\gamma_S$ ), and incident angle of solar radiation ( $\theta_I$ ). The ensuing explanation reveals the appropriate choice of trigonometric functions.

The PV generation at a particular location and time instant with a fixed tilt angle  $\theta_T$  depends on the following major factors:

- (i) Factors that describe the apparent position of the Sun in the sky as viewed from the location are elevation and azimuth angles.

- (a) The  $\theta_S$  being a function of the Sun's height, is an important factor that measures the intensity of solar irradiance. For higher  $\theta_S$ , solar intensity is more concentrated i.e., more light strike a given area and vice versa. At noon, i.e., for  $\theta_S = 90^\circ$  direct solar irradiance is maximum whereas, an oblique  $\theta_S$  gives less direct solar irradiance because of the factor  $\sin \theta_S$  [refer Figure 2.5(a)].
- (b) For a given orientation of panel,  $\gamma_S$  specifies the angle by which the Sun deviates from the oriented direction. The  $\gamma_S$  is measured on the horizontal plane and it completely neglects the Sun's elevation. The  $\sin \gamma_S$  indicate the actual horizontal distance of the Sun's position from the oriented direction; their positive and negative values respectively correspond to the position of the Sun at pre-noon and post-noon time instants [refer Figure 2.5(b)].
- (ii) The  $\theta_I$  defined as the angle between the line that points to the Sun and the line that is normal to the module surface depends on location (changes place to place), the Sun's position (changes in every instant of time and day), and plane direction (changes if plane is moved, i.e. affected by array tilt angle  $\theta_T$ ). The cosine component of  $\theta_I$  is that component of direct solar radiation which can be collected by the plane. If  $\theta_I$  is  $0^\circ$ ,  $\cos \theta_I$  becomes maximum i.e., maximum solar radiation is collected by the plane. Smaller the  $\theta_I$ , radiation capturing is better. Intensity on a tilted surface is the product of normal intensity and  $\cos \theta_I$ . Hence, the maximum amount of intensity that could reach a collector is reduced by a factor  $\cos \theta_I$  [refer Figure 2.5(c)].

The proposed model embodies three significant regressors:  $T_1 = \sin \theta_S$ ,  $T_2 = \sin \gamma_S$ , and  $T_3 = \cos \theta_I$ . The annual and daily variations of the three regressors at 12 time instants (7 am to 6 pm at an hourly interval) during the day time for one year at Monkton is shown in Figure 2.6. The negative values of  $\theta_S$  indicates that the Sun is under the horizon, which generally happens before sun rise and after sun set. In addition, the trend effect (e.g., decrease in production pattern due to aging effect) is accounted in the model by adding two extra terms:  $T_4 = d_N$ , and  $T_5 = (d_N)^2$  where,  $d_N$  is a number mapped as the day number of the year. The inclusion of  $T_5$  in addition to  $T_4$ , is to capture the possible nonlinearity in trend effect. The proposed model is suggested to be of the form:



**Figure 2.6:** Variations of the regressors  $T_1$ ,  $T_2$  and  $T_3$  at 12 time instants for a year at Monkton.

$$P_{PV} = a_M T_1 + b_M T_2 + c_M T_3 + d_M T_4 + e_M T_5 + f_M + \varepsilon_{PV} = \hat{P}_{PV} + \varepsilon_{PV} \quad (2.1)$$

where  $a_M, b_M, \dots, f_M$  are the unknown model parameters whose values are calculated using the method of least squares;  $P_{PV}$  is the observed PV generation;  $\hat{P}_{PV}$  is the predictable PV generation;  $\varepsilon_{PV}$  is the error term or residual of the model which refers to the uncertainty component of PV generation. The residual sum of squares is defined as,

$$RS(a_M, b_M, c_M, d_M, e_M, f_M) = \sum_{k=1}^{ss} \left\{ P_{PV}(k) - \hat{P}_{PV}(k) \right\}^2 \quad (2.2)$$

where “ss” is the total number of data samples collected at a particular time instant for five years. The values of the model parameters are obtained by solving the following equations:

$$\frac{\partial(RS)}{\partial a_P} = 0, \frac{\partial(RS)}{\partial b_P} = 0, \frac{\partial(RS)}{\partial c_P} = 0, \frac{\partial(RS)}{\partial d_P} = 0, \frac{\partial(RS)}{\partial e_P} = 0, \text{ and } \frac{\partial(RS)}{\partial f_P} = 0 \quad (2.3)$$

*Formulations to evaluate  $T_1, T_2$  and  $T_3$ :*

The expressions for  $\theta_S$  and  $\gamma_S$  are respectively given as,

$$\theta_S = \sin^{-1} (\sin \theta_{LA} \sin \theta_D + \cos \theta_H \cos \theta_{LA} \cos \theta_D) \quad (2.4)$$

$$\gamma_S = \sin^{-1} \left( \frac{-\sin \theta_H \cos \theta_D}{\cos \theta_S} \right) \quad (2.5)$$

where  $\theta_{LA}$  is the latitude,  $\theta_D$  is the declination angle,  $\theta_H$  is the hour angle. For the Northern hemisphere,  $\theta_D$  is calculated as,

$$\theta_D = 23.45^\circ \sin \left( \frac{d_N + 284}{365} \times 360^\circ \right). \quad (2.6)$$

The value of  $\theta_H$  can be calculated by converting clock time to solar time (Holbert and Srinivasan, 2011). A step by step formulation for the same is provided underneath.

(i) Step-1: Equation of time is calculated as,

$$t_E = 9.87 \sin(2D) - 7.53 \cos(D) - 1.5 \sin(D), \text{ where, } D = \left(\frac{360}{365}\right) (d_N - 81). \quad (2.7)$$

(ii) Step-2: Local solar time is calculated as,

$$t_{LS} = t_{LC} \pm (4 \text{ min.} / \text{deg.}) (\theta_{LO}^{\text{STM}} - \theta_{LO}) + t_E \quad (2.8)$$

where  $t_{LC}$  is the local clock time or standard time,  $\theta_{LO}$  is the local longitude of the location,  $\theta_{LO}^{\text{STM}}$  is the local longitude of standard time meridian which is calculated as,

$$\theta_{LO}^{\text{STM}} = 15^0 \left[ \frac{\theta_{LO}}{15^0} \right] \quad (2.9)$$

where  $\left[ \frac{\theta_{LO}}{15^0} \right]$  returns the nearest integer to  $\frac{\theta_{LO}}{15^0}$ . The sign of second term in (2.8) is positive (+) for Western longitudes and negative (-) for Eastern longitudes.

(iii) Step-3: Finally,  $\theta_H$  is calculated as,

$$\theta_H = 15^0 (t_{LS} - 12). \quad (2.10)$$

The range of  $\gamma_S$  as calculated using (2.5) is  $+90^0$  to  $-90^0$ . Using (2.5), (2.6) and (2.10), the criteria to obtain the required value of  $\gamma_S$  with reference to the true North (Dave et al., 1975) is given as,

$$\gamma_S^R = \begin{cases} 180^0 - \gamma_S, & \cos \theta_H \geq (\tan \theta_D / \tan \theta_{LA}) \\ 360^0 + \gamma_S, & \cos \theta_H < (\tan \theta_D / \tan \theta_{LA}) \end{cases}. \quad (2.11)$$

Using (2.11), the expression for  $\theta_I$  of a non-tracking type tilted PV array for a given orientation is given as,

$$\theta_I = \cos^{-1} \{ \sin \theta_S \cos \theta_T + \cos \theta_S \sin \theta_T \cos (\gamma - \gamma_S^R) \}. \quad (2.12)$$

For the South and the South-West facing PV arrays, the values of  $\gamma$  are  $180^0$  and  $225^0$  respectively. Hence, the first three regressors can be obtained using (2.4), (2.5) and (2.12).

Finally, the residual i.e., the uncertainty of the PV generation is obtained as,

$$\varepsilon_{\text{PV}} = P_{\text{PV}} - \hat{P}_{\text{PV}} \quad (2.13)$$

The probability density of the uncertainty component as obtained using (2.13) can be approximated by using a non-parametric density estimation method [refer Appendix B] as suggested in (Soleimanpour and Mohammadi, 2013, Rouhani et al., 2016).

#### 2.2.2.4 Multiple linear regression model for load power and ambient temperature

A multiple linear regression model for preprocessing depends significantly on the type of seasonalities in the time series. The seasonalities in load power and ambient temperature time series as evident from Figure 2.4 are additive. The time series is split into distinct yearly datasets to obtain the seasonality orders. The fast Fourier transform method applied to the data set of each year, yields in a set  $S$  with elements representing the dominant frequency orders. The preceding approach leads to a reduced number of candidate frequency components from which the optimal frequency numbers are selected using a penalizing strategy. Unlike (McLeod and Gweon, 2013), where starting with the fundamental component, the immediate succeeding orders are sequentially appended until the optimal order is reached, the proposed approach on the other hand, considers only those frequency components (in a decreased order of their dominance) that are deemed fit in modeling the concealed seasonalities. Having obtained the set  $H$  (a subset of  $S$ ) with  $h_0$  elements and the optimal time trend order  $t_0$ , the proposed regression model is suggested to be of the form:

$$Y = \sum_{t=1}^{t_0} a_{\text{M},t} (d_{\text{N}})^t + \sum_{h \in H} \left( b_{\text{M},h}^s \sin \frac{2\pi h d_{\text{N}}}{\text{T}_{\text{S}}} + b_{\text{M},h}^c \cos \frac{2\pi h d_{\text{N}}}{\text{T}_{\text{S}}} \right) + c'_{\text{M}} + \varepsilon_{\text{Y}} = \hat{Y} + \varepsilon_{\text{Y}} \quad (2.14)$$

where  $\hat{Y}$  is the predictable component of  $Y$ ;  $d_{\text{N}} = 1$  corresponds to the first sample of  $Y$  (load power/ambient temperature time series). The first summation in (2.14) accounts for the trend effect whereas; the Fourier terms in the second summation trace the seasonalities in the data. The residual  $\varepsilon_{\text{Y}}$  refers to the uncertainty of the regressand  $Y$ . The seasonal period “ $\text{T}_{\text{S}}$ ” is taken as 365 for a daily time step time



series excluding leap days.

The model parameters  $a_M$ ,  $b_M^s$ ,  $b_M^c$  and  $c'_M$  are estimated using the least squares method. For multiplicative seasonalities, the developed model can be easily extended by estimating the changing seasonal variances as discussed in (Mcleod and Gweon, 2013). The extraction of dominant frequency orders and the criteria for model selection are elaborated as under:

*Extraction of dominant frequency orders:*

A dominant frequency component carries the maximum energy among all frequencies in the amplitude spectrum. There can be, however, more than one dominant frequencies present in a time series. They can be ordered by their energies, referred to as second dominant, third dominant, etc. Extraction of dominant frequencies is the process of their identification, and sometimes their removal from the time series. The best way to find out the dominant frequencies is via amplitude spectrum obtained using fast Fourier transform. An amplitude spectrum reveals the hidden dominant frequencies in the time series. It is also important to rank the dominant frequencies. The various steps proposed to extract the dominant frequency orders are as follows:

- (i) Step-1: The daily time-series  $[\text{TS}]_{365N_y \times 1}$  of  $N_y$  years is split into distinct year data sets:  $[\text{TS}_1]_{365 \times 1}$ ,  $[\text{TS}_2]_{365 \times 1}$ ,  $\dots$ ,  $[\text{TS}_{N_y}]_{365 \times 1}$ .
- (ii) Step-2: The dominant frequencies in the distinct year data sets are calculated using the following sub-steps:
  - (a) Apply fast Fourier transform to obtain the amplitude spectrum in frequency domain.
  - (b) The frequency numbers are arranged in descending order of their amplitudes in the spectrum.
  - (c) First 30 frequency orders pertaining to  $N_y$  years are stored in matrix  $[\text{DF}_1]_{30 \times N_y}$ .
- (iii) Step-3: The unique frequency orders (say, “ufc” in number) common in any two amongst  $N_y$  years are stored in vector  $[\text{DF}_2]_{\text{ufc} \times 1}$ .
- (iv) Step-4: The position of elements of vector  $[\text{DF}_2]$  in the columns of matrix  $[\text{DF}_1]$  are stored in matrix  $[\text{DF}_3]_{\text{ufc} \times N_y}$ . The corresponding elements of vector  $[\text{DF}_2]$  not present in the columns of matrix  $[\text{DF}_1]$  are set to zero in matrix  $[\text{DF}_3]$ .

(v) Step-5: Finally, the dominant frequency orders of the given time-series are the list of frequency orders ranked using the following sub-steps:

- (a) An intermediate matrix  $[\text{IM}]_{\text{ufc} \times 2}$  is formed with column values, respectively the total number, and sum of corresponding row-wise non-zero elements in matrix  $[\text{DF}_3]$ .
- (b) The matrix  $[\text{IM}]$  is partitioned into sub matrices  $[\text{IM}_1], [\text{IM}_2], \dots, [\text{IM}_{N_y-1}]$  of same number of columns and contains rows of  $[\text{IM}]$  with first column values same as  $N_y, N_y - 1, \dots, 2$  respectively. The rows in the sub-matrices are rearranged as per the increasing order of their second column values. The first column values are updated with their corresponding frequency orders.
- (c) Lastly, the sub matrices are row-wise concatenated and the first column values indicating the first “ufc” dominant frequencies (in the decreasing order of their energy) are stored in vector  $[\text{DF}_4]_{\text{ufc} \times 1}$ .

*Criteria for model selection:*

The adopted penalizing strategy for obtaining  $t_0$  and  $h_0$  to ensure a balance between good-fit and parsimony is discussed herein. Among the available criteria for model selection, Bayesian information criterion induces higher penalization for models and being parsimonious; it is suitable for relatively larger data sets (Dziak et al., 2017) as considered in this work. In general, the Bayesian information criterion scores (Mcleod and Gweon, 2013) are calculated as,

$$\text{BIC} = -2 \ln(\ell_{\max}) + (\text{mp}) \cdot \ln(\text{ss}) \quad (2.15)$$

where “mp” is known as the number of model parameters. For the regression model in (2.14) under the assumption that  $\varepsilon$  is Gaussian distributed with mean as zero, variance as  $\sigma_\varepsilon^2$  and since, each  $\varepsilon_i$  ( $i = 1, 2, \dots, \text{ss}$ ) is independent and identically distributed, the likelihood of  $\varepsilon$  is the product of likelihood contributions by each observation. It is given as,

$$\begin{aligned}
\ell &= f(\varepsilon/\sigma_\varepsilon) = \prod_{i=1}^{\text{ss}} f(\varepsilon_i/\sigma_\varepsilon) \\
&= \prod_{i=1}^{\text{ss}} \frac{1}{\sigma_\varepsilon \sqrt{2\pi}} e^{-\frac{(\varepsilon_i)^2}{2\sigma_\varepsilon^2}} = (2\pi\sigma_\varepsilon^2)^{-\frac{\text{ss}}{2}} \cdot e^{-\frac{\sum_{i=1}^{\text{ss}} (\varepsilon_i)^2}{2\sigma_\varepsilon^2}} = (2\pi\sigma_\varepsilon^2)^{-\frac{\text{ss}}{2}} \cdot e^{-\frac{\text{RS}}{2\sigma_\varepsilon^2}}
\end{aligned} \tag{2.16}$$

where residual sum of squares,  $\text{RS} = \sum_{i=1}^{\text{ss}} (\varepsilon_i)^2$ , and  $f(\cdot)$  is the probability density function. Maximum likelihood ( $\ell_{\max}$ ) in (2.15) can be obtained by maximizing (2.16) i.e., by solving  $\frac{d\ell}{d\sigma_\varepsilon} = 0$  as discussed under:

$$\begin{aligned}
&\left\{ (2\pi\sigma_\varepsilon^2)^{-\frac{\text{ss}}{2}} \cdot e^{-\frac{\text{RS}}{2\sigma_\varepsilon^2}} \cdot \frac{\text{RS}}{\sigma_\varepsilon^3} \right\} + \left\{ e^{-\frac{\text{RS}}{2\sigma_\varepsilon^2}} \cdot (2\pi)^{-\frac{\text{ss}}{2}} \cdot (-\text{ss}) \cdot \sigma_\varepsilon^{-\text{ss}-1} \right\} = 0 \\
&\Rightarrow \sigma_\varepsilon^{-\text{ss}-3} \cdot e^{-\frac{\text{RS}}{2\sigma_\varepsilon^2}} \cdot (\text{RS} - \text{ss}\sigma_\varepsilon^2) = 0 \Rightarrow \sigma_\varepsilon^2 = 0, \frac{\text{RS}}{\text{ss}}
\end{aligned} \tag{2.17}$$

Substituting  $\sigma_\varepsilon^2 = \frac{\text{RS}}{\text{ss}}$  of (2.17) in (2.16) yields,

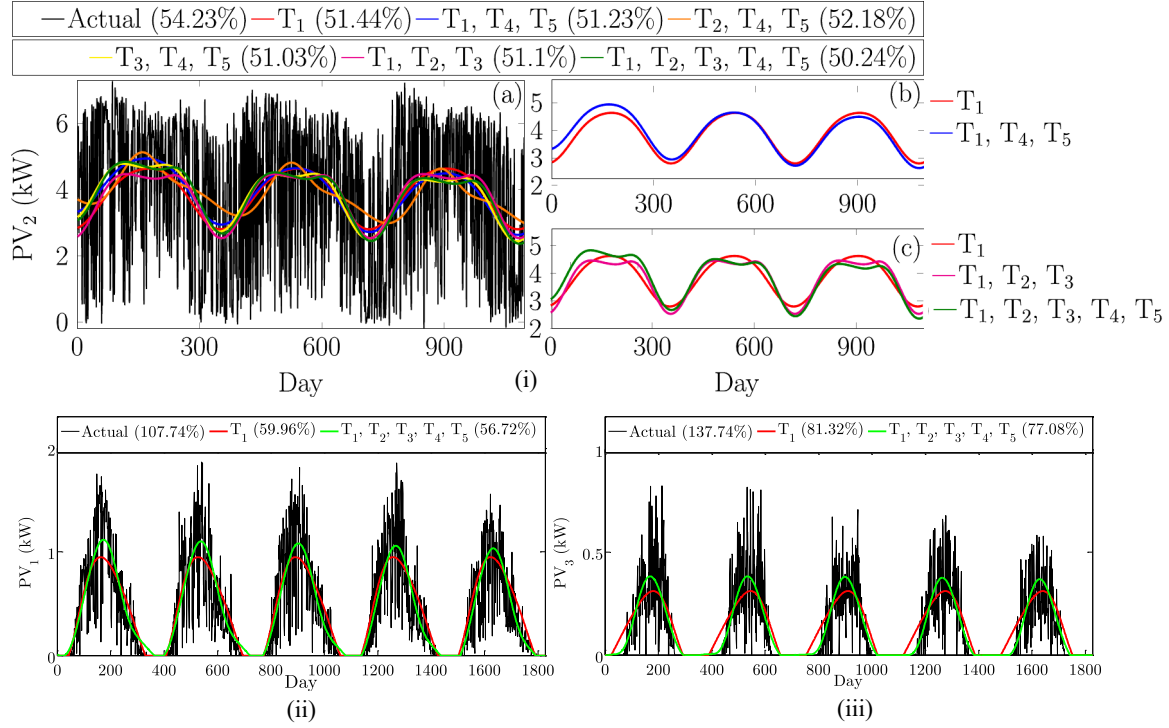
$$\begin{aligned}
\ell_{\max} &= \left( \frac{2\pi \text{RS}}{\text{ss}} \right)^{-\frac{\text{ss}}{2}} \cdot e^{-\frac{\text{ss}}{2}} \\
&\Rightarrow \ln(\ell_{\max}) = \left( -\frac{\text{ss}}{2} \right) \left\{ \ln \left( \frac{2\pi \text{RS}}{\text{ss}} \right) + 1 \right\} \\
&\Rightarrow -2 \ln(\ell_{\max}) = \text{ss} \left\{ \ln \left( \frac{2\pi \text{RS}}{\text{ss}} \right) + 1 \right\}
\end{aligned} \tag{2.18}$$

Using (2.18) in (2.15) yields,

$$\text{BIC} = \text{ss} \left\{ \ln \left( \frac{2\pi \cdot \text{RS}}{\text{ss}} \right) + 1 \right\} + (\text{mp}) \cdot \ln(\text{ss}) \tag{2.19}$$

### 2.2.3 Performance comparison of proposed multiple linear regression models

In this section, the performance of the proposed models in (2.1), and (2.14) are compared with that of the model(s) in the literature. The suitability of the proposed model in (2.1) for improved preprocessing as compared to (Fan et al., 2014) is demonstrated in light of the following subsequent major observations. The observed data of



**Figure 2.7:** Comparison of models in tracking periodic variation in (i) PV<sub>2</sub> at 1 pm, (ii) PV<sub>1</sub> at 7 am, and (iii) PV<sub>3</sub> at 6 pm.

PV<sub>1</sub>, PV<sub>2</sub> and PV<sub>3</sub> respectively at 7 pm, 1 pm and 6 pm (arbitrarily chosen) are taken into consideration to show the capability of (2.1) in predicting the periodic variation. The corresponding results of different possible models of PV<sub>2</sub> are compared in Figure 2.7 (i)-(a). The various terms used in the models are highlighted. It is evident that the observed data has a multimodal periodic variation. The models considering the term(s): T<sub>1</sub> (Fan et al., 2014), (T<sub>1</sub>, T<sub>4</sub>, T<sub>5</sub>), and (T<sub>2</sub>, T<sub>4</sub>, T<sub>5</sub>) fail to trace multimodality in data. However, the inclusion of the terms: T<sub>4</sub>, and T<sub>5</sub> in latter two cases track the trend in data which (Fan et al., 2014) fails as indicated in Figure 2.7(i)-(b). Although the terms T<sub>1</sub>, T<sub>2</sub>, and T<sub>3</sub> are partially dependent, a model which includes only a single term (say T<sub>3</sub>) is less accurate as compared to the model with all the three terms taken into consideration. In Figure 2.7(i)-(c), comparison of three plots concludes that (2.1) traces the multimodality and trend effect which (Fan et al., 2014) fails. Further, the cases in Figure 2.7(ii) and (iii) reveals the capability of the proposed model in tracing skewness and trend effect present in data. The performance comparison of various models in Figure 2.7 is based on the reduction of coefficient of variation with respect to the coefficient of variation of the actual data.

The percentage coefficient of variation is calculated as,

$$\% CV = \frac{\text{Standard deviation value}}{\text{Mean value}} \times 100. \quad (2.20)$$

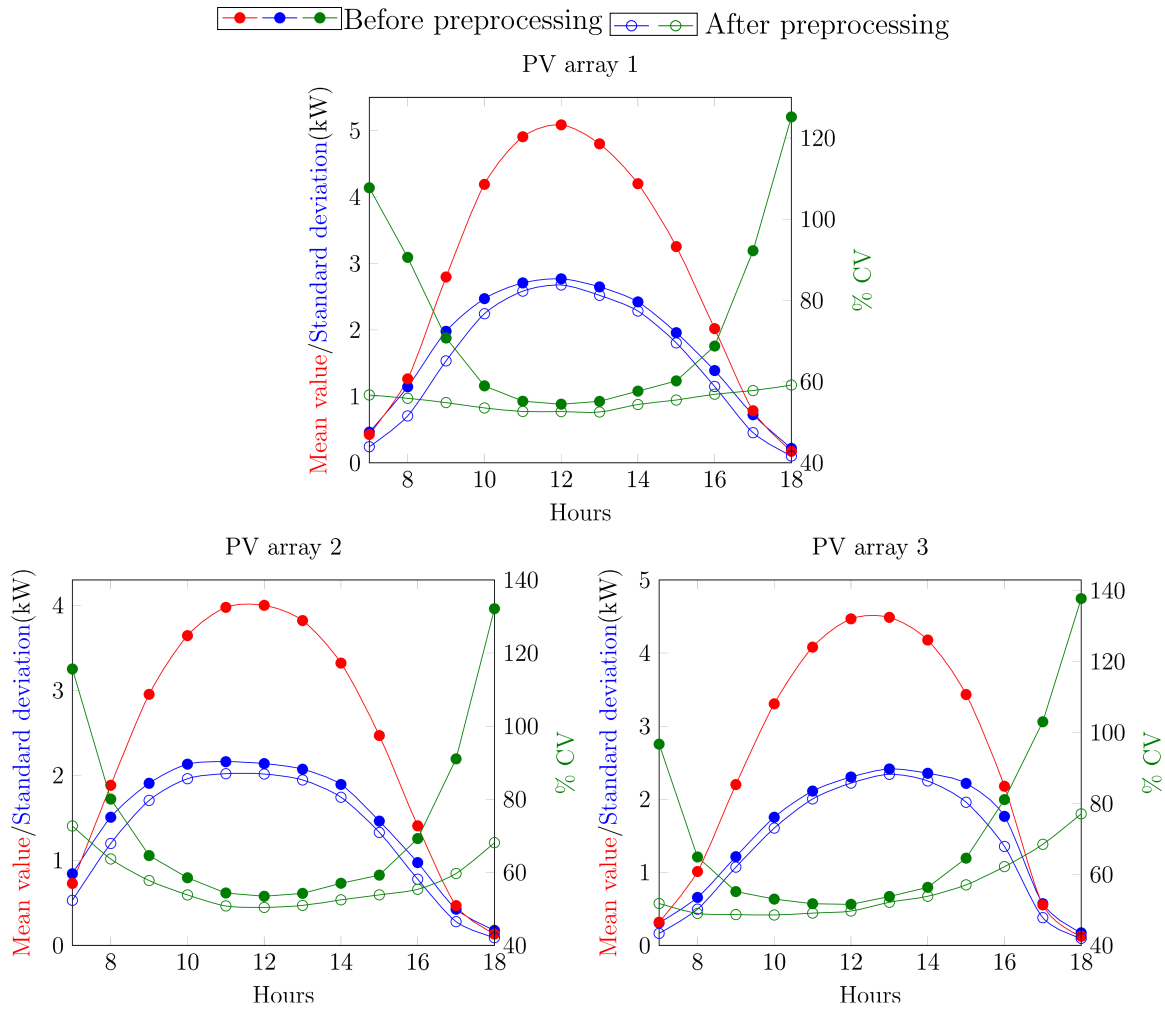
There is a reduction in coefficient of variation from 54.23% (actual data) to 50.24% by (2.1) i.e., 3.99% whereas; the reduction is 2.79% by (Fan et al., 2014) in Figure 2.7(i). In the other two case of Figure 2.7, the reduction in coefficient of variation are 51.02% and 60.66% by (2.1) whereas, 47.78% and 56.42% by (Fan et al., 2014). A set of similar inferences can be drawn in the remaining cases [refer Table 2.2].

**Table 2.2:** Comparison of coefficient of variations of PV generations before and after preprocessing.

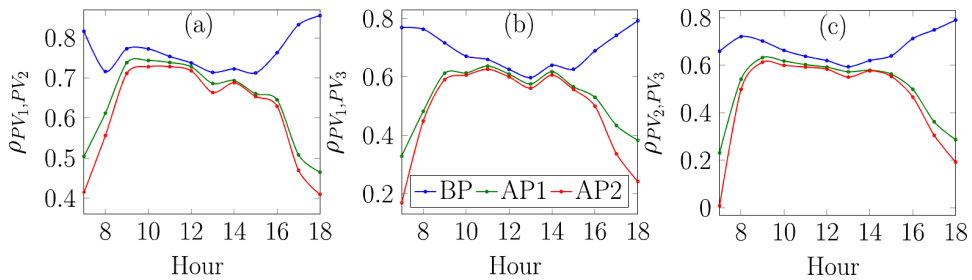
Time instant	% CV								
	PV array 1			PV array 2			PV array 3		
	BP	AP1	AP2	BP	AP1	AP2	BP	AP1	AP2
7:00 am	107.74	59.96	56.72	115.63	74.31	71.68	96.72	53.99	51.81
8:00 am	90.60	57.47	55.89	80.03	64.89	63.65	64.92	51.49	49.01
9:00 am	70.74	56.07	54.84	64.57	58.32	57.75	55.16	49.58	48.69
10:00 am	58.96	54.13	53.52	58.50	54.30	53.81	53.05	49.15	48.56
11:00 am	55.18	52.90	52.65	54.34	51.07	50.79	51.74	49.42	49.10
noon	54.44	52.82	52.62	53.45	50.71	50.39	51.57	49.88	49.68
1:00 pm	55.15	53.10	52.51	54.23	51.44	50.24	53.78	52.27	52.15
2:00 pm	57.66	55.26	54.37	57.00	53.32	52.46	56.36	54.01	53.85
3:00 pm	60.18	56.19	55.46	59.22	54.46	53.83	64.58	57.65	57.06
4:00 pm	68.76	57.99	56.90	69.23	56.40	55.35	81.08	64.48	62.20
5:00 pm	92.24	60.47	57.80	91.03	61.95	59.67	103.02	70.34	68.48
6:00 pm	125.20	64.37	59.16	132.12	73.34	68.11	137.74	81.32	77.08

Note: BP stands for before preprocessing. AP1 and AP2 stands for the values after preprocessing using (Fan et al., 2014) and (2.1) respectively.

To facilitate a better understanding of the obtained results, the time axis of plots in Figure 2.8 is divided into three zones: zone-1 (7 am to 10 am), zone-2 (10 am to 4 pm) and zone-3 (4 pm to 6 pm). From Figure 2.8, it is observed that the reduction in coefficient of variation is less at the time instants of zone-2 whereas the reduction is more in the zone-1 and zone-3 time instants. The maximum reduction occurs at extreme time instants (i.e., 7 am and 6 pm) for all the three places. At 7 am, the reductions using (Fan et al., 2014) compared to actual data are 47.78%, 41.32% and

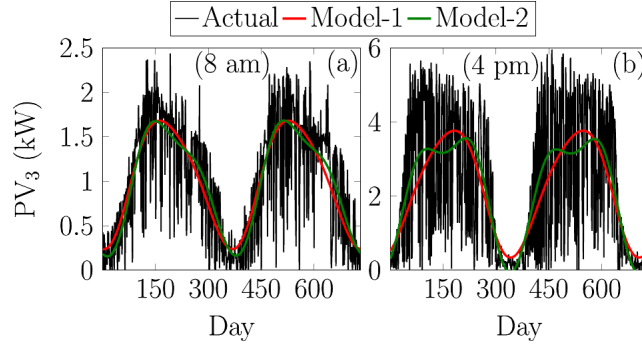


**Figure 2.8:** Plots of statistical parameters of PV generations before and after preprocessing.



Note: BP stands for before preprocessing. AP1 and AP2 stands for the values after preprocessing using (Fan et al., 2014) and (2.1) respectively.

**Figure 2.9:** Plots of PMCCs among the PV generations before and after preprocessing.

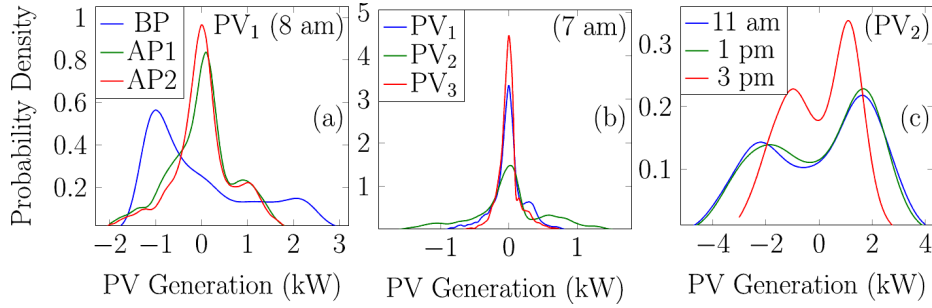


**Figure 2.10:** Performance comparison of models (Model-1 and Model-2 refers to (Fan et al., 2014) and (2.1) respectively) in the region of low PV power generation.

42.73% respectively for the three arrays whereas, the corresponding reductions are 51.02%, 43.95% and 44.91% for (2.1). At 6 pm, the reductions using (Fan et al., 2014) compared to actual data are 60.83%, 58.78%, and 56.42% respectively for the three arrays whereas; the corresponding reductions are 66.04%, 64.01%, and 60.66% for (2.1). Further, product moment correlation coefficient (PMCC) represented as  $\rho$  between the pair of PV generations, respectively given as,  $\rho_{PV_1, PV_2}$ ,  $\rho_{PV_1, PV_3}$  and  $\rho_{PV_2, PV_3}$  are plotted in Figure 2.9. A significant reduction in PMCC values is observed using (2.1). In all the cases, the PMCC values are positive due to nearly the common effects of solar irradiance, ambient temperature, and other environmental factors. It is evident from Figure 2.9 that PMCC values are reduced with increasing distance. The geographical distance between PV arrays 1 and 3 as well as PV arrays 2 and 3 are nearly the same; hence, the PMCC plots closely resemble. On the other hand, a lesser distance between PV arrays 1 and 2 leads to higher values of PMCC at all the time instants.

From the above result analysis, it is clear that the model in (2.1) performs better over (Fan et al., 2014) for zone-1 and zone-3 time instants. However, for zone-2 time instants, especially at noon, a significant distinction in the models performance is imperceptible. Since in a probabilistic assessment, both input uncertainties and the associated correlations affect the probability distributions of result variables, the model in (2.1) would be more selective for accurate estimation of results. The results demonstrated in Figure 2.10 for a few cases reveal that in the region of low-power production, (Fan et al., 2014) fails to perform as desired in comparison to the model in (2.1).

From Figure 2.11(a), it is noteworthy that the density plots after preprocessing



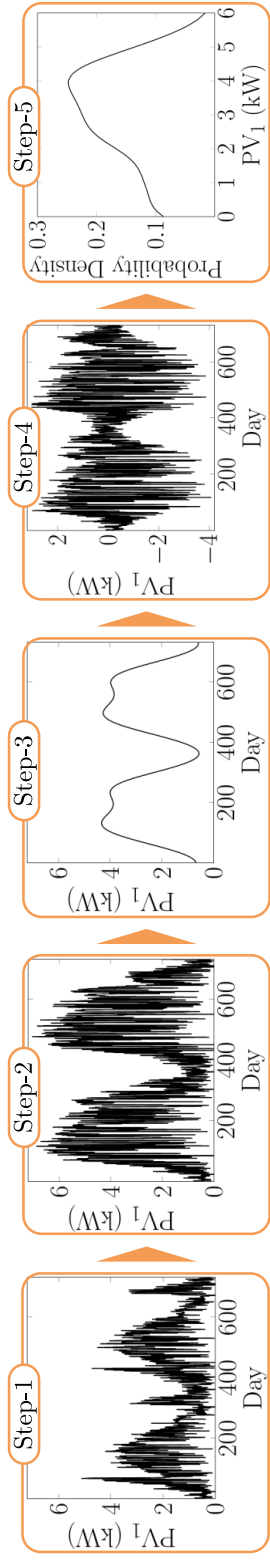
Note: BP stands for before preprocessing. AP1 and AP2 stands for the values after preprocessing using (Fan et al., 2014) and (2.1) respectively.

**Figure 2.11:** Comparison of probability density plots of PV generation uncertainties for various cases.

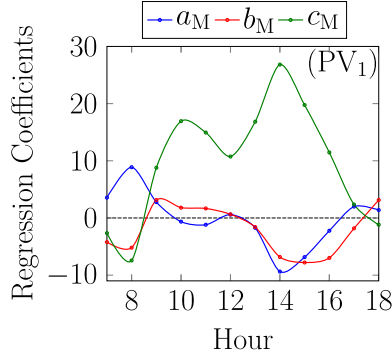
are significantly different from the raw data. Further, there is a slight difference in the density plots after preprocessing by the models. For time instants in zone-1 and zone-3, varying geographical locations has a remarkable impact on the density plots, and there is no certainty regarding the number of modes that may be present in the distributions. Figure 2.11(b) serves as an example case supporting the above-claimed observation. However, for time instants in zone-2, the density plots are observed to be bimodal. In Figure 2.11(c), at 11 am, 1 pm and 3 pm the probability density plots are bimodal. Similar inferences can be drawn for other time instants in zone-2 for all the three arrays. Various steps in uncertainty modeling as described in Figure 2.1 for  $PV_1$  at 9 am is shown in Figure 2.12. Although stationarity in mean value is achieved, standard deviation value is still changing. Similar observations are also noticed in the remaining cases. The non-Gaussian arbitrary probability distributions of residuals indicate that PV generation time series at various places and time instants are extremely random.

The plots of multiple linear regression parameters at various time instants is shown in Figure 2.13. Among the first three regressors of the model,  $T_2$  has a lower value of correlations with  $T_1$  and  $T_3$  whereas,  $T_1$  and  $T_3$  are highly correlated. The regressors  $T_1$  and  $T_3$  differs from each other only by the inclusion of  $\theta_T$  in  $T_3$  (for  $\theta_T = 0^0$ ,  $T_1 = T_3$ ). The plot of  $a_M$  is approximately a mirror image of plot of  $c_M$  (with disproportionate magnitudes) as the effect of the corresponding regressor is masked by the other. Since, all the three terms, directly or indirectly affect the PV production patterns and also has better predictions as evident through Figure 2.7 to Figure 2.10, it is imperative to consider all of them in the multiple linear regression model for





**Figure 2.12:** Uncertainty modeling steps as applied to  $PV_1$  at 9 am.



**Figure 2.13:** Plot of multiple linear regression model parameters at various time instants for  $PV_1$ .

multi-time instant preprocessing.

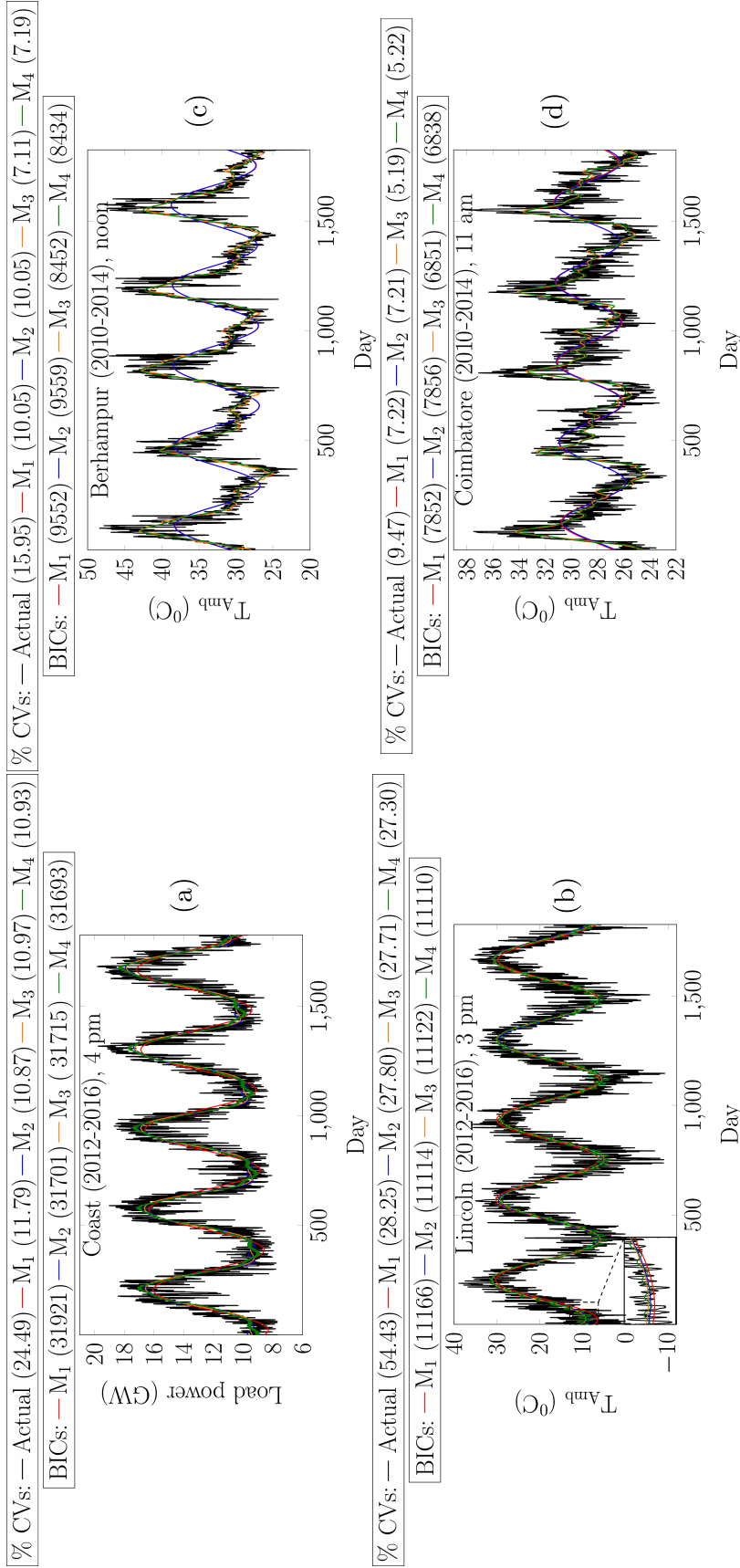
Now, the suitability of the proposed model in (2.14) is evaluated through comparison of results with that of the following models.

- (i)  $M_1$ : Model that considers seasonal and annual variations (Fan et al., 2012).
- (ii)  $M_2$ : Model with fixed order time trend and Fourier terms (e.g.,  $t_0=3$ ,  $H = \{1, 2, 3, 4, 5\}$  for load power, and  $t_0=2$ ,  $H = \{1\}$  for ambient temperature).
- (iii)  $M_3$ : Model wherein starting with the fundamental component, the immediate succeeding orders are sequentially appended (McLeod and Gweon, 2013).

**Table 2.3:** Estimated values of  $t_0$ , and  $H$  for various cases in Figure 2.14.

Figure 2.14	$M_3$		$M_4$	
	$t_0$	$H$	$t_0$	$H$
(a)	2	$\{1, 2, 3\}$	2	$\{1, 52, 2, 6\}$
(b)	3	$\{1\}$	3	$\{1, 18, 28\}$
(c)	7	$\{1, 2, 3, 4, 5, 6\}$	7	$\{1, 2, 8, 16, 5, 6\}$
(d)	10	$\{1, 2, 3, 4, 5, 6, 7, 8\}$	10	$\{1, 2, 4, 3, 15, 11, 8\}$

Henceforth, the model in (2.14) is referred to as  $M_4$ . Unlike in (Fan et al., 2012, 2014), a reduced coefficient of variation value after preprocessing cannot be used as a model performance evaluation index as the regressors in  $M_1$  and  $M_2$ , being fixed for any arbitrary data yields possibly an under-fitted or over-fitted model. It is worth mentioning that an over-fitted model has a low coefficient of variation compared to a best-fit model which together has lower coefficient of variation than an under-fitted

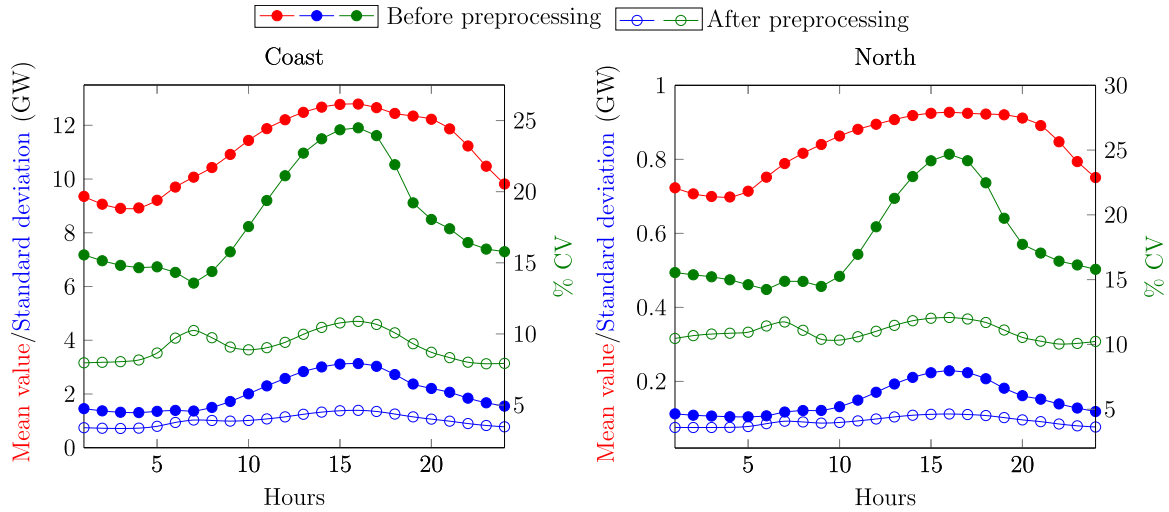


**Figure 2.14:** Comparison of models in tracking periodic variations in few cases of load power and ambient temperature.



model. Therefore, a lower BIC score that confirms the potential regressors for a best-fit model is considered as the performance evaluation criterion. The periodic variations captured by the models for a few randomly chosen cases are compared in Figure 2.14; the %CV values and BIC scores for each case are highlighted. The results are elucidated on a case to case basis. For all the cases, BIC scores are less for  $M_4$ , but %CV after preprocessing is not the least. It is quite evident from Figure 2.14 that, at different places, load power and ambient temperature time series have dissimilar trends and seasonalities. It is difficult to capture such characteristics with a fixed number of regressors as in  $M_1$  and  $M_2$ . Frequency orders: 3, 4 and 5 being not dominant, their inclusion in  $M_2$  overfits the model leading to a reduced %CV. However, in  $M_3$  and  $M_4$ , the flexibility in choosing  $t_0$  and  $h_0$  facilitates in identifying the true variations. In order to clearly demonstrate the competency of  $M_4$  over  $M_3$ , the corresponding estimated values of  $t_0$  and  $H$  are presented in Table 2.3. The values of  $t_0$ , unlike  $H$ , are same for both the models. Further, there is no definite trend behavior in Figure 2.14(b), (c), and (d) leading to corresponding higher values. On the other hand, each regressor included in  $M_3$  and  $M_4$  being penalized, the obtained  $H$  in case of  $M_3$  has every chances of missing the dominant frequency component(s). For instance, 52<sup>nd</sup> order frequency component (that represents weekly variation in load power) being the second dominant frequency, is not present in  $H$  of  $M_3$  [refer Table 2.3]. Hence, it is important to select beforehand the set of dominant frequency components that truly characterize the concealed seasonalities.

A set of steps for the extraction of dominant frequency orders as discussed in section 2.2.2.4 are applied to the load power data of North Central zone during 2012-2016 at 7 pm [refer Figure 2.15]. The plots of statistical parameters at various time instants before and after preprocessing for a few cases of load powers and ambient temperatures are shown in Figure 2.16 and Figure 2.17 respectively. It is evident that the plots of statistical parameters of load power at various places are almost same. However, the same is not true in case of ambient temperature. Further, the plots of PMCCs at various time instants before and after preprocessing for a few cases of load powers and ambient temperatures are shown in Figure 2.18 and Figure 2.19 respectively. The comparison of few cases in Figure 2.18 indicates that, before and after preprocessing, the PMCC plots are nearly opposite to each other. The same applies to other cases. On the other hand, the PMCC values of ambient temperatures are significantly reduced after preprocessing as compared to the values



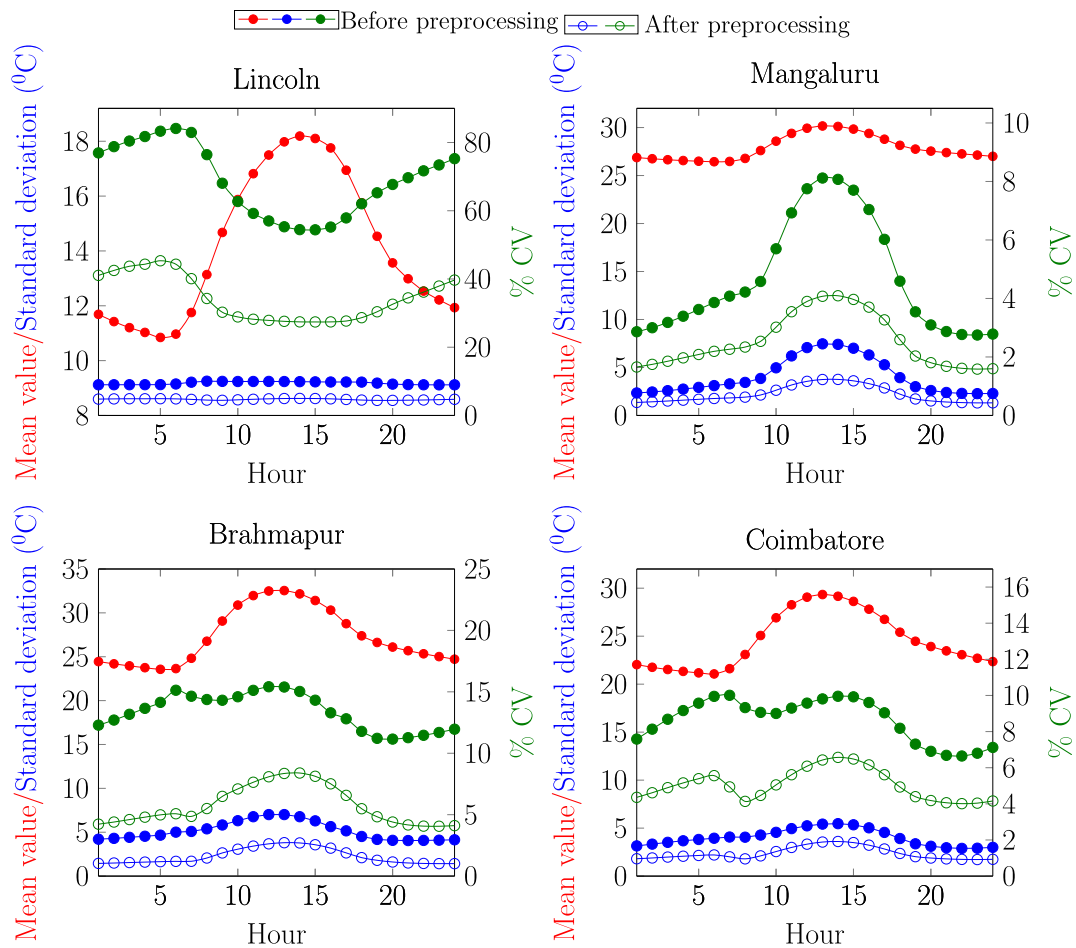
**Figure 2.16:** Plots of statistical parameters of load powers from Coast and North weather zones of Texas before and after preprocessing.

before preprocessing.

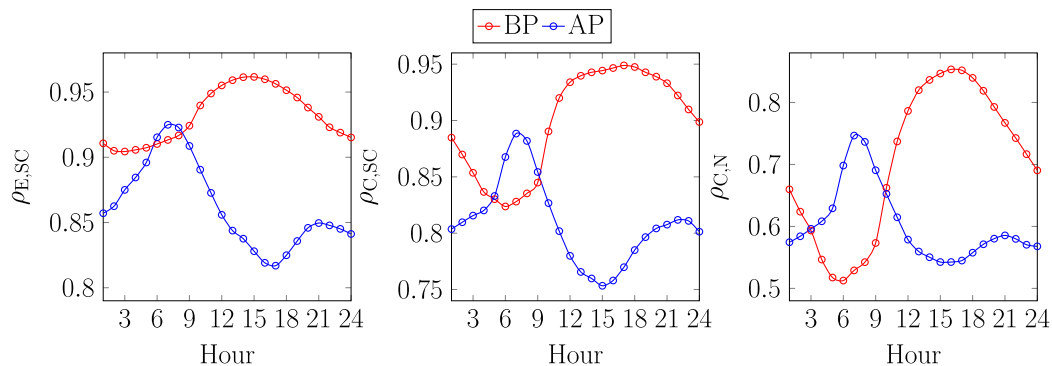
Various steps in uncertainty modeling as described in Figure 2.1 is applied to a few cases of load powers and ambient temperatures and are shown in Figure 2.20 and Figure 2.21 respectively. In both the cases, stationarity in mean value is achieved but, standard deviation value is not strictly stationary. Similar observations are also noticed in the remaining cases. Unlike PV generation, where there is no specific distribution type to approximate the uncertainty components, in case of load powers and ambient temperatures, Gaussian distribution approximately fits their uncertainty components at most of the time instants. In few cases, the distributions with heavier tails are approximated by  $t$ -distributions.

## 2.3 Spatiotemporal forecasting model

The uncertainty modeling procedures (based on time of the year) as detailed in section 2.2 are suitable for planning studies. Whereas a preminent spatiotemporal model that efficiently extracts the statistical properties from the observed data is ideal for operational studies. In general, the development of a multivariate spatiotemporal probabilistic model is strenuous as it has to deal with multivariate probability distribution functions. To address this, a multidimensional characterization based on a spatiotemporal scenario set can be employed as a viable substitute. It is achieved by

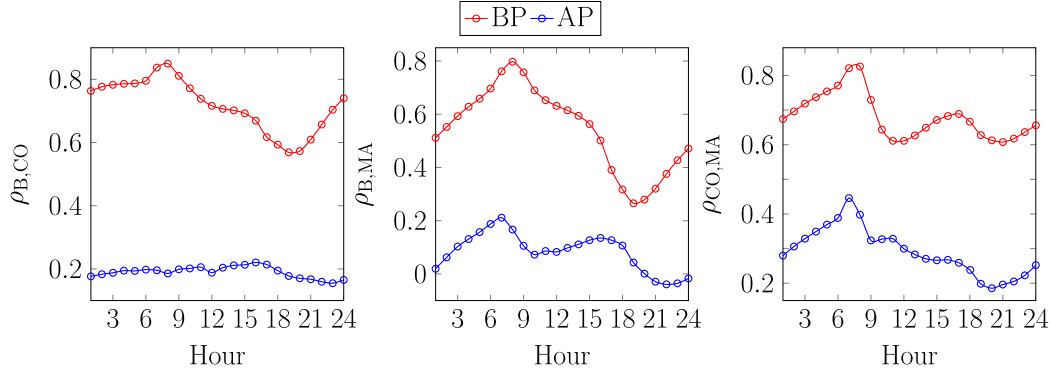


**Figure 2.17:** Plots of statistical parameters of ambient temperatures before and after preprocessing.



Note: BP stands for before preprocessing, AP stands for after preprocessing.

**Figure 2.18:** Plots of PMCCs between the load powers before and after preprocessing.



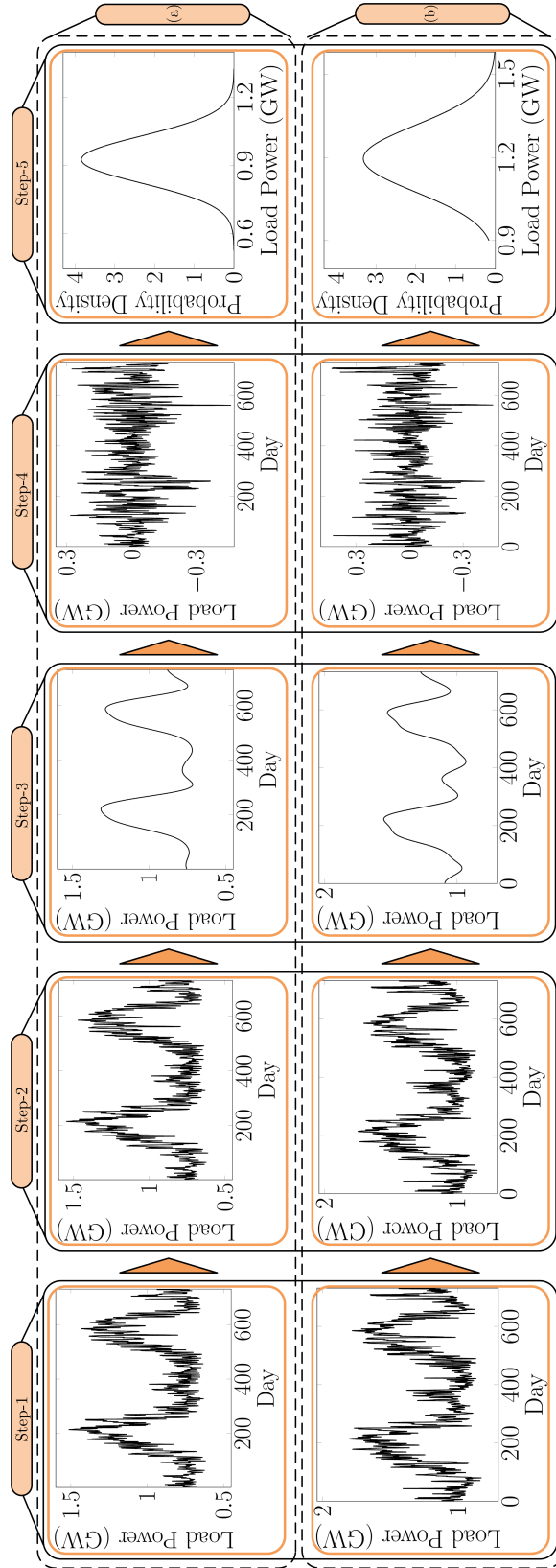
Note: BP stands for before preprocessing, AP stands for after preprocessing.

**Figure 2.19:** Plots of PMCCs between the ambient temperatures before and after preprocessing.

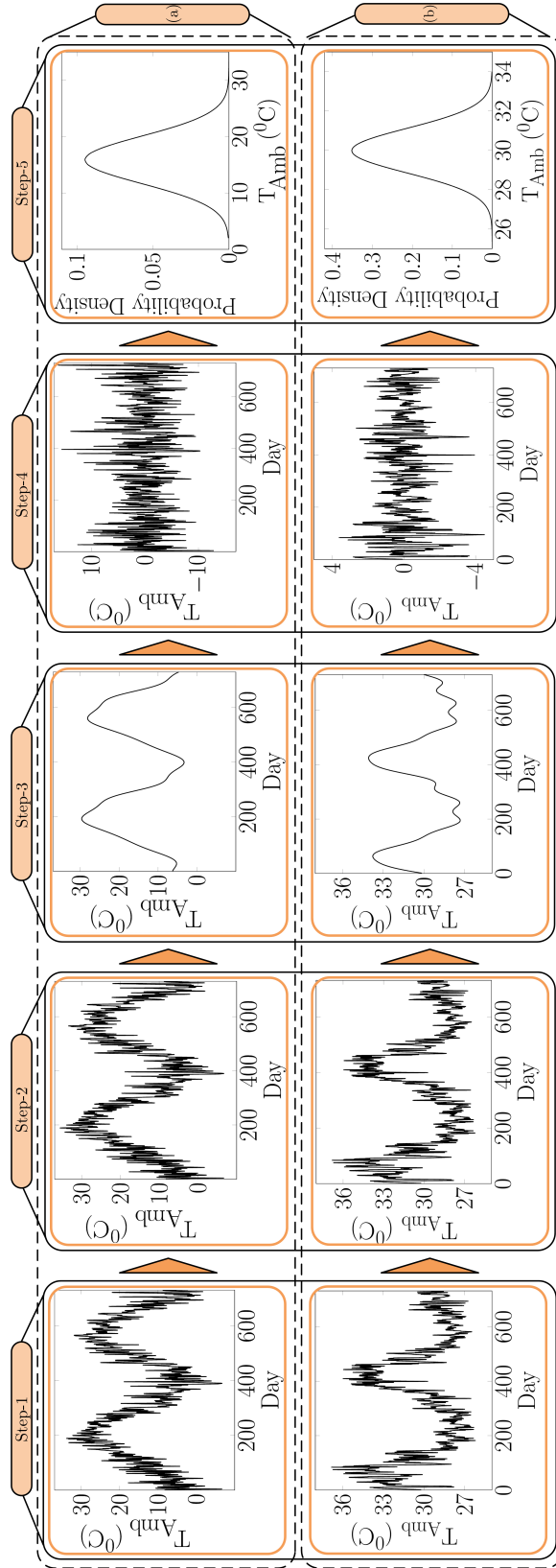
approximating the continuous stochastic process by a discrete equivalent and thereby embedding temporal dependency information that cannot be assimilated using interval/quantile representation (Le et al., 2015). The authors in (Le et al., 2015) applied the principal component analysis and a time-series technique to analytically characterize the statistical facets of the multi-site multi-variable stochastic process. Gaussianity and stationarity respectively being the primary requirements for principal component analysis and time-series models, they are managed through the modeling framework using transformation and preprocessing techniques respectively. Although the transformation of non-Gaussian time series to Gaussian domain is straightforward, two primary concerns with (Le et al., 2015) are the inadequacy in adoption of the preprocessing approach mainly emerging due to the difficulty in proper identification of epochs and the use of a time series model such as autoregressive or autoregressive moving average cannot adequately model the time-dependent variability of the residual after preprocessing. To this concern, the preprocessing techniques as discussed in section 2.2 are employed and the concern of modeling the changing standard deviation of input random variables are better characterized with the help of an autoregressive conditionally heteroscedastic (ARCH) or generalized autoregressive conditionally heteroscedastic (GARCH) time series model. The complete modeling steps are discussed underneath.

- (i) Step-1: Arrange the historically observed time-series data in the form as given in (C.1) [refer Appendix C].
- (ii) Step-2: Obtain the mean stationary residuals of inputs of step-1 using prepro-





**Figure 2.20:** Uncertainty modeling steps for load powers of (a) North zone, Texas (2012-2013) at 3 pm and (b) West zone, Texas (2012-2013) at 7 pm.



**Figure 2.21:** Uncertainty modeling steps for ambient temperatures of (a) Lincoln, USA (2012-2013) at 10 am and (b) Mangaluru, India (2013-2014) at 2 pm.

cessing approaches described in section 2.2.

- (iii) Step-3: The non-Gaussian mean stationary data of step-2 is transformed into standard Gaussian domain using  $\Phi^{-1} \left[ \hat{F}(\bullet) \right]$  where  $\hat{F}(\bullet)$  is the estimated empirical cumulative distribution function (CDF) of the mean stationary inputs and  $\Phi^{-1}(\bullet)$  is the inverse of the standard Gaussian CDF.
- (iv) Step-4: Apply the principal component analysis [refer Appendix C] to the transformed time-series of step-3.
- (v) Step-5: The temporal correlations in the principal components of step-4 is modelled using GARCH or the simpler form ARCH whichever is suitable [refer section 2.3.1].
- (vi) Step-6: Use the model in step-5 for each principal component to generate future time series as per the requirement of the study.
- (vii) Step-7: The obtained data in step-6 are back transformed to their original domain using the reverse transformation of step-3.
- (viii) Step-8: The items removed in the preprocessing step are added back to obtain scenarios obeying the complete characteristics.

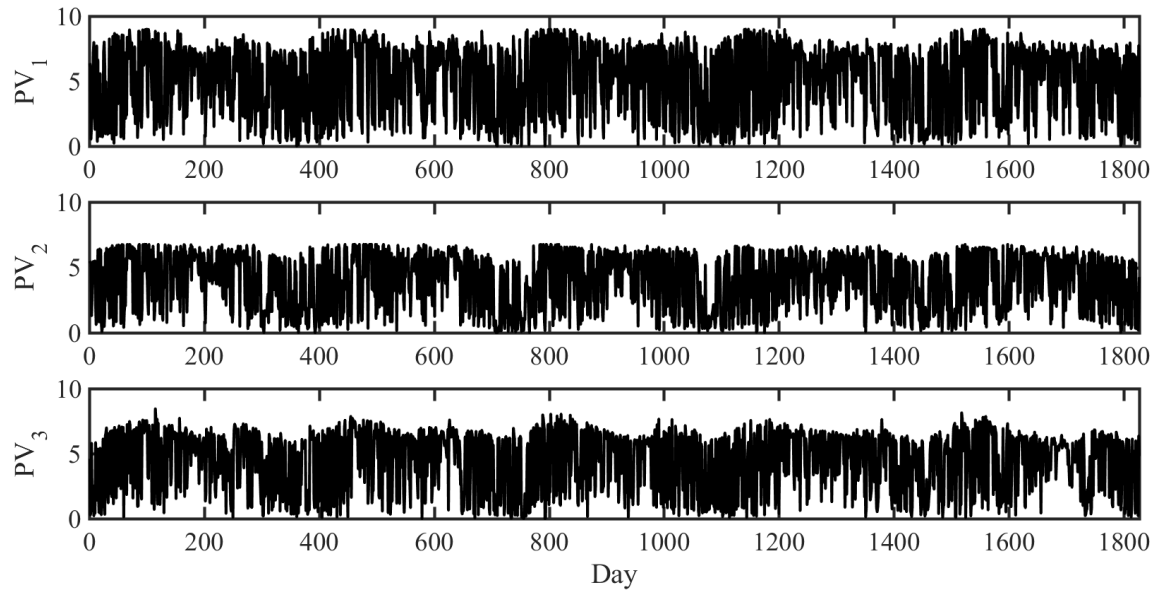
### 2.3.1 Generalized autoregressive conditionally heteroscedastic model

GARCH model is able to describe the variation in time of the standard deviation of the residuals. A GARCH(p,q) model with appropriate values of model parameters “p” and “q” fits variance non-stationarity or volatility of the residual  $\varepsilon_t$  such that  $\varepsilon_t = \sigma_t z_t$  where  $z_t$  is a strong white noise process which is assumed as independent and identically distributed  $\sim N(0, 1)$ , and  $\sigma_t$  is the time dependent standard deviation given as,

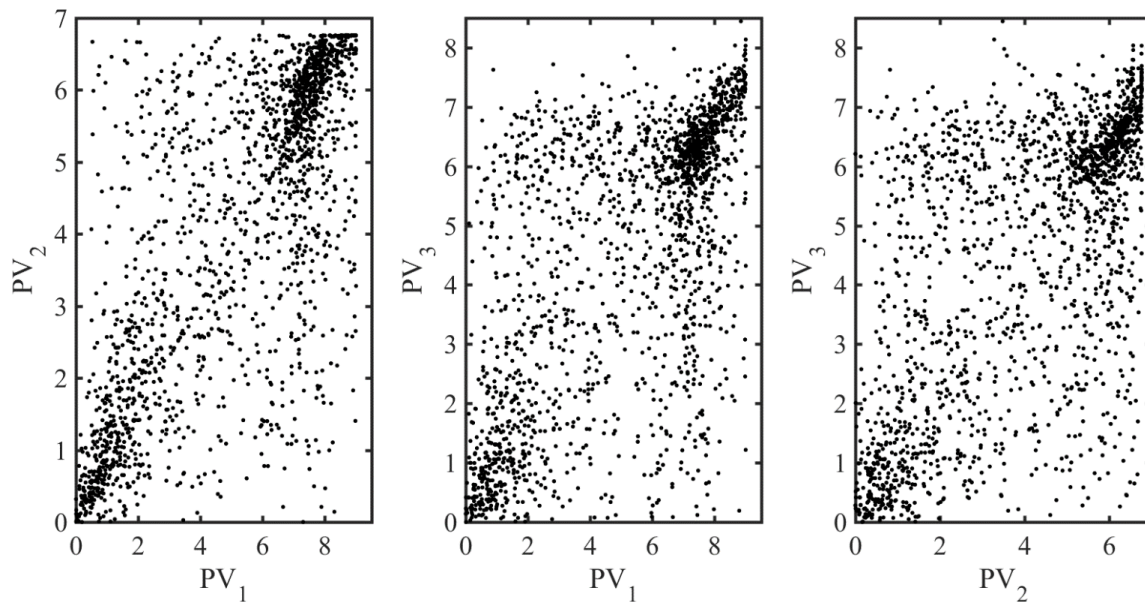
$$\sigma_t^2 = \alpha_0 + \alpha_1 \sigma_{t-1}^2 + \dots + \alpha_p \sigma_{t-p}^2 + \alpha'_1 \varepsilon_{t-1}^2 + \dots + \alpha'_q \varepsilon_{t-q}^2 \quad (2.21)$$

where “p” and “q” are respectively referred to as the orders of the GARCH terms  $\sigma^2$  and ARCH terms  $\varepsilon^2$ ;  $\alpha_0$ ,  $\alpha_i (i = 1, 2, \dots, p)$ , and  $\alpha'_i (i = 1, 2, \dots, q)$  are the model parameters.

The necessary and sufficient conditions for coefficients is given as,  $\alpha_0 > 0$ ,  $\alpha_i \geq 0$ ;  $i = 1, 2, \dots, p$ , and  $\alpha'_i \geq 0$ ;  $i = 1, 2, \dots, q$ . ARCH model is a special case of GARCH model with  $\alpha_i = 0, \forall i$ .



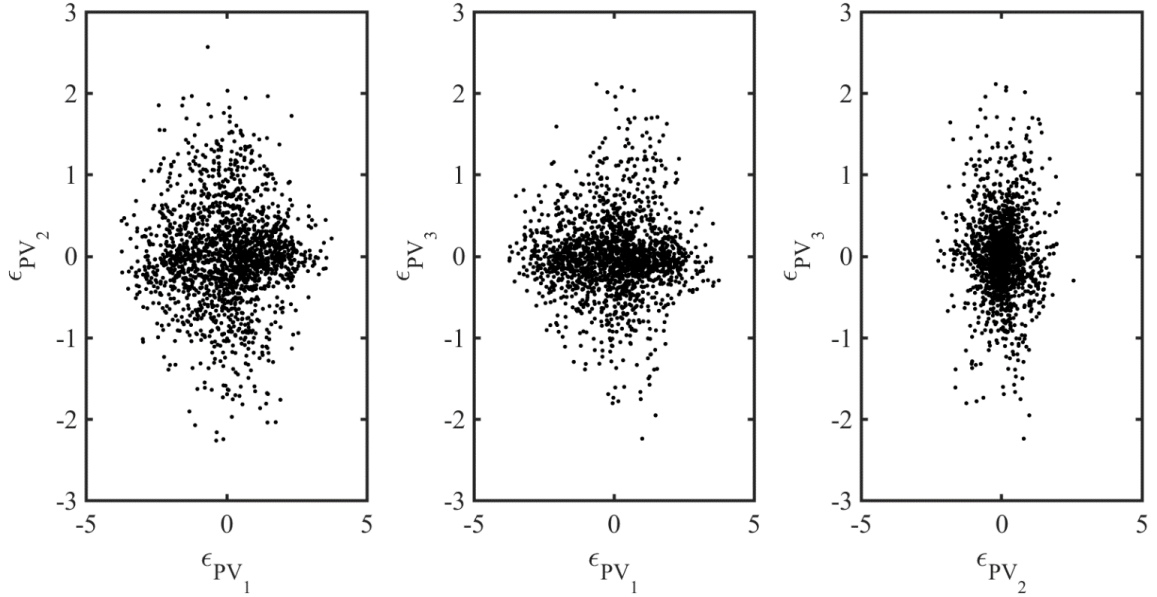
**Figure 2.22:** Daily time stepped observed PV generation time-series.



**Figure 2.23:** Scatter plot of observed PV generations in Figure 2.22.

### 2.3.2 Performance comparison of proposed spatiotemporal forecasting model

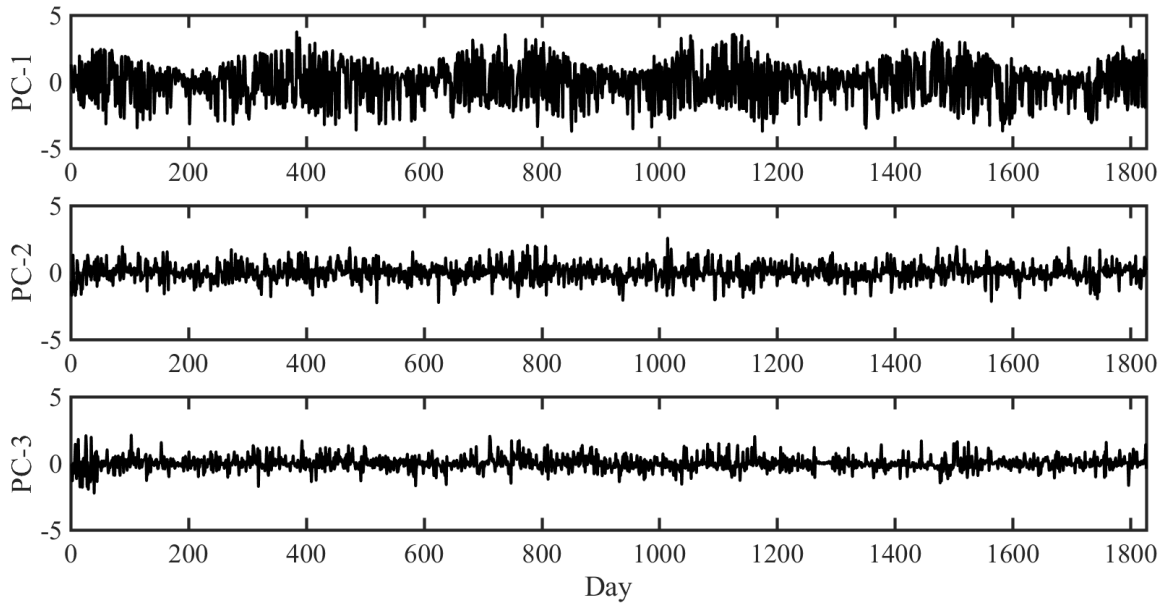
In this section, the suitability of the proposed spatiotemporal model for improved forecasting as compared to (Le et al., 2015) is demonstrated. Daily time-stepped



**Figure 2.24:** Scatter plot of residuals after transformed to standard Gaussian domain.

observed data of  $PV_1$ ,  $PV_2$ , and  $PV_3$  corresponding to the noon from 2012 to 2016 is taken into consideration (refer Figure 2.22). The scatter plots of the above PV generation time series are shown in Figure 2.23. Firstly, the observed data are preprocessed using (2.1). Although the residuals after preprocessing are mean stationary, the standard deviation is still time changing (volatile standard deviation). The obtained non-Gaussian mean stationary data is transformed into a standard Gaussian domain which is the primary requirement for principal component analysis. The scatter plot of residuals after transformed to standard Gaussian domain is shown in Figure 2.24. Principal component analysis as applied to the above three transformed mean stationary data yields principal components PC-1, PC-2, and PC-3. The time series relevant to each principal component is shown in Figure 2.25. It is evident from the plot that the principal components are different in terms of magnitude. PC-1 contains the largest percentage of information in the data; PC-2 contains the second largest and so on. The PMCC values  $\rho_{PC-1, PC-2}=0$ ,  $\rho_{PC-1, PC-3}=0$ , and  $\rho_{PC-2, PC-3}=0$  show that principal components are uncorrelated.

Finally, the ARCH model is built on the basis of principal components with regressors as decided from the sample partial autocorrelation plots. The sample partial autocorrelation values are obtained using the data corresponding to the first four years i.e., 2012-2015 as training sample. The developed models of principal compo-



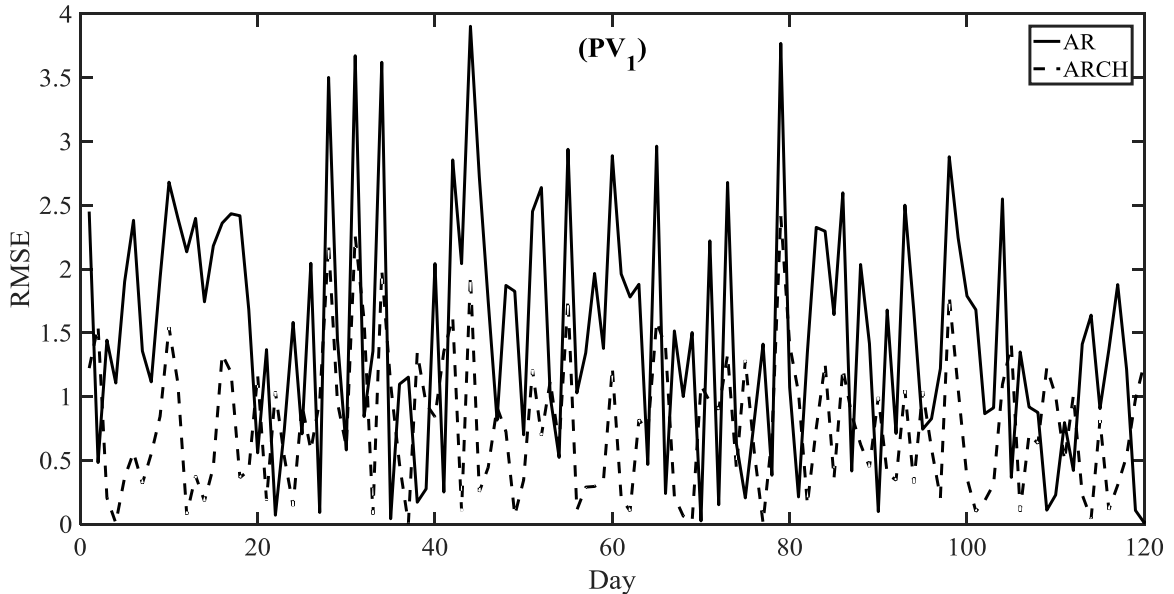
**Figure 2.25:** Time-series of principal components.

nents are then used to forecast future samples. The data pertaining to the year 2016 is considered as the testing sample and the accuracy of the proposed spatiotemporal model as compared to (Le et al., 2015) is demonstrated in light of the forecasting accuracy. Root mean square error (RMSE) is used as the index for forecasting accuracy. The RMSE plot for forecasted  $PV_1$  for the first 120 days of the year 2016 is shown in Figure 2.26. It can be clearly observed from the figure that, the model in (Le et al., 2015) i.e., the AR model fails to model the volatility effect in standard deviation hence leads to higher values of RMSE value.

## 2.4 Conclusions

This chapter has analyzed the uncertainty characteristics of historical PV generation, load power and ambient temperature data using the proposed ingenious way of determining the potential regressors for multiple linear regression model based multi-time instant preprocessing techniques. In summary, the following observations from the result analysis are worth noting.

- (i) In case of PV generation, the yearly periodic deterministic component is mainly influenced by solar position and the PV installation location. Hence, the possible generic model with a fixed number of regressors suitable for various places



**Figure 2.26:** Comparison of RMSE.

and time instants is proposed. On the other hand, load power and ambient temperature are profoundly influenced by social and environmental factors respectively; it is challenging to develop possible generic models with a fixed number of regressors. Hence, for each case, the regressors representing the proper time trend terms and seasonalities that deemed fit are ascertained.

- (ii) The proposed models for preprocessing were effective in removing the predictable components from the data as compared to the literature thereby justifying their improved performance. In case of PV generation, the proposed preprocessing is more effective in the time instants of zone-1 and zone-3.
- (iii) The residuals after preprocessing have reduced values of standard deviations as compared to before preprocessing in all the cases. The reduction is not uniform at various time instants. In case of PV generation, the reduction is more in the zone-1 and zone-3 time instants than zone-2 time instants. It is because, during zone-2 time instants, PV generation is high, so is the uncertainty which dominates over the predictable variation. In case of load power, the reduction is extremely less during 1 am to 7 am for all the studied cases. However, at the remaining time instants, the reduction is significant. It is due to the fact that, aggregate load up till 7 am from the early morning is comparatively low, and so is the standard deviation. In case of ambient temperature, except for

Mangaluru where the reduction is more in zone-2 time instants, all other places of the studied cases the reduction is reasonably constant.

- (iv) The residuals after preprocessing are stationary as far as mean value is concerned; however, the standard deviation is dependent on time. This effect is more prominent in case of PV generation as compared to the other two cases.
- (v) There are also some critical observations from PMCC plots. There is a reduction in PMCC values among PV generations as well as among ambient temperatures after preprocessing as compared to the corresponding values before preprocessing. The PMCC values between PV generations are significantly reduced at zone-1 and zone-3 time instants whereas; the reduction is less for zone-2 time instants whereas; the decrease of PMCCs is nearly uniform in case of ambient temperature. There is a peculiar observation of an opposite characteristics plots of PMCCs between aggregate load powers before and after processing for all pairs of the studied weather zones.
- (vi) Finally, a spatiotemporal model using proposed preprocessing, transformation techniques, principal component analysis, and ARCH/GARCH model is proposed.

The developed multi-time instant uncertainty modelling can be applicable for (i) long-term planning for the network reinforcement of PV-integrated power systems, (ii) scheduling of daily PV generation energy, (iii) calculation of operational reliability indices for power systems integrated with PV generations, (iv) chronological probabilistic analysis of power systems, etc.



## Publications based on the chapter

### Refereed journal publications

1. Prusty, B. R. and Jena D. (2018). Preprocessing of multi-time instant PV generation data. *IEEE Transactions on Power Systems*, 33(3):3189-3191. (IEEE)

### Book series

1. Prusty, B. R. and Jena D. (2019). Uncertainty modeling steps for probabilistic steady state analysis. In *Applications of Computing, Automation and Wireless Systems in Electrical Engineering* (Springer)



# Chapter 3

## OVER-LIMIT RISK ASSESSMENT

### 3.1 Preamble

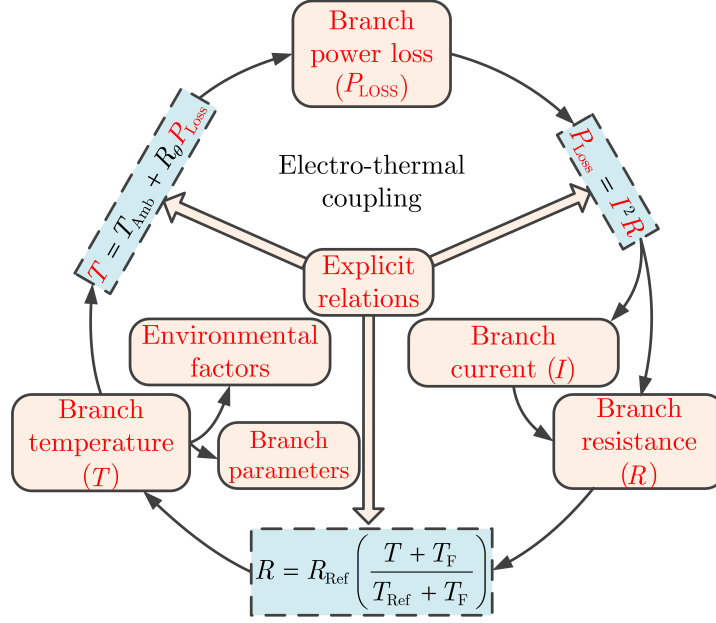
The structural changes in power systems due to the enormous integration of photovoltaic (PV) generations necessitate the development of suitable models and methods to quantify system risks accurately. Over-limit risk assessment is the central issue in planning under uncertainty and probabilistic load flow (PLF) is the fundamental tool. The planning engineers involved with decision making process prefer to handle accurate risk indices. Generally, peak values of the uncertain inputs are used as worst case condition in planning studies. However, different uncertain input random variables have peak values at different time instants as observed in chapter 2. Further, time instants pertaining to highest variance and mean value need not necessarily be the same. Hence, it is essential to carry out the risk assessment for various time instants (say, at an hourly interval) during the planning stage to maintain system risk at an acceptable level. This chapter extends the conventional power system model by incorporating the temperature effect to perform a more realistic risk assessment. The PLF study that uses temperature-augmented system model by considering ambient temperatures of the transmission branches as uncertain inputs is referred to as temperature-augmented PLF (TPLF). Monte-Carlo simulation method is used for both PLF and TPLF simulations for various result analyses including risk assessment.

## 3.2 Temperature-augmented power system model

PLF is carried out by propagating input uncertainties through a power system model that describes the functional relationships among the input random variables and result variables (Papaefthymiou, 2007). In the conventional power system model for PLF, transmission branch resistances are presumed temperature independent. In reality, there exists a strong coupling among resistance, temperature and real power loss of a transmission branch. This electro-thermal coupling can be better understood from Figure 3.1. As evident from the figure, besides the ohmic losses, environmental factors (ambient temperature, wind speed, wind direction, etc.) and branch parameters (diameter, emissivity, etc.) dictate the equilibrium temperature of the branch. Among the various factors, dominant is the ambient temperature. By introducing an extra mismatch equation related to temperature, temperature-augmented load flow carried out in (Frank et al., 2013) reduces the temperature-related error that is inherent in conventional load flow. The parameters describing the thermal characteristics are generally being not specified in the datasheet of transmission branches, thermal resistance model is used for the development of temperature-augmented load flow model instead of the thermal balance equation (the former model do not fundamentally alter the latter) under the following assumptions:

- (i) The system is under electrical and thermal steady-state.
- (ii) The three-phase system is assumed balanced.
- (iii) The branch conductor is an ideal uniform material, i.e., its temperature is equal to the ambient temperature in the absence of current flow.
- (iv) The variation in branch reactance due to temperature variation is negligible.

The electrical dynamics are neglected in load flow so long as the system is assumed free from fault or any other abnormal conditions. The above assumptions are often reasonable as transmission branches do not typically have a vast change in power flow within the time-scales of classical steady-state. Further, the thermal dynamics of the transmission branches can be assumed short as compared to the changes in branch loading over time. The changes in the power flow for most of the part are roughly the same from minute-to-minute. Moreover, changes those occur, say a reduction in power flow due to load dropping off as the night sets in, are on the time-scale of minutes. Because of this, the thermal dynamics can be neglected.



**Figure 3.1:** Electro-thermal coupling effect in a transmission branch.

With a motivation to analyze the effect of ambient temperature uncertainty on PLF result, temperature-augmented load flow is first revisited, and its possible application for PLF is discussed in section 3.3.

### 3.2.1 Temperature-augmented load flow model

The resistance of a transmission branch  $i - j$  (branch connecting  $i^{\text{th}}$  bus and  $j^{\text{th}}$  bus) is expressed as,

$$R_{i-j} = R_{\text{Ref}, i-j} \left( \frac{T_{i-j} + T_{\text{F}, i-j}}{T_{\text{Ref}, i-j} + T_{\text{F}, i-j}} \right) \quad (3.1)$$

where  $T_{i-j}$  is the conductor temperature of the branch  $i - j$ ,  $T_{\text{F}}$  is the temperature constant,  $R_{\text{Ref}, i-j}$  and  $T_{\text{Ref}, i-j}$  are the reference values of  $R_{i-j}$  and  $T_{i-j}$  respectively. According to thermal resistance model,  $T_{i-j}$  is expressed as,

$$T_{i-j} = T_{\text{Amb}, i-j} + T_{\text{Rise}, i-j} = T_{\text{Amb}, i-j} + R_{\theta, i-j} P_{\text{Loss}, i-j} \quad (3.2)$$

where  $T_{\text{Amb}}$  and  $T_{\text{Rise}}$ , are the ambient temperature and branch temperature rise above  $T_{\text{Amb}}$  respectively,  $R_{\theta}$  is the thermal resistance,  $P_{\text{Loss}}$  is the branch power loss. Using (3.2) a temperature related expression is introduced as,

$$T'_{i-j} = T_{i-j} - R_{\theta, i-j} P_{\text{Loss}, i-j} = T_{\text{Amb}, i-j}. \quad (3.3)$$

As the real and reactive bus power injections ( $P$ , and  $Q$  respectively) are specified, the mismatch equations  $\Delta P$ , and  $\Delta Q$  respectively can be explicitly expressed (Wang et al., 2010). However, it is difficult to define the branch temperature mismatch equations since their values are unknown a priori. Nevertheless, an equation for  $T'$  can be defined for the temperature dependent branches where  $T'$  is the measure of the difference between the present value of  $T$  and the calculated temperature by using the state variables i.e., bus voltage angle  $\delta$  and bus voltage magnitude  $|V|$  (Frank et al., 2013). From (3.3), it is clear that, the value of  $T'$  is equal to  $T_{\text{Amb}}$ . As transmission branches are rated according to maximum temperature,  $R_{\theta, i-j}$  is calculated as,

$$R_{\theta, i-j} = \frac{T_{\text{Rated rise}, i-j}}{P_{\text{Rated loss}, i-j}} \quad (3.4)$$

where  $T_{\text{Rated rise}, i-j}$  is the expected temperature rise at rated power level, typically dictated by line sag considerations, allowable operating temperature of insulating materials, etc. The value of  $P_{\text{Rated loss}, i-j}$  is either specified or calculated as,

$$P_{\text{Rated loss}, i-j} = I_{\text{Rated}, i-j}^2 R_{\text{Hot}, i-j} \quad (3.5)$$

where  $I_{\text{Rated}, i-j}$  is the MVA rating of the branch  $i-j$  assuming voltage magnitudes of the connecting buses at 1 pu;  $R_{\text{Hot}, i-j}$  is the conductor resistance at  $T_{\text{Rated rise}, i-j}$ .

The value of  $P_{\text{Rated loss}}$  is either specified or is calculated using the fully loaded resistance  $R_{\text{Hot}}$  measured at worst-case  $T_{\text{Amb}}$  i.e.,  $T_{\text{Amb-wc}}$ . The expression of  $R_{\text{Hot}}$  for a temperature dependent branch using (3.1) is given as,

$$R_{\text{Hot}, i-j} = R_{\text{Ref}, i-j} \left( \frac{T_{\text{Amb-wc}, i-j} + T_{\text{Rated rise}, i-j} + T_{\text{F}, i-j}}{T_{\text{Ref}, i-j} + T_{\text{F}, i-j}} \right). \quad (3.6)$$

The expression for  $P_{\text{Loss}, i-j}$  (Frank et al., 2013) is given as,

$$P_{\text{Loss}, i-j} = g_{i-j} (|V_i|^2 + |V_j|^2 - 2|V_i||V_j|\cos\delta_{ij}) \quad (3.7)$$

where  $g_{i-j}$  is the conductance of branch  $i-j$ ;  $\delta_{ij} = \delta_i - \delta_j$ . Substituting (3.7) in (3.3) yields,

$$T'_{i-j} = T_{i-j} - R_{\theta, i-j} g_{i-j} (|V_i|^2 + |V_j|^2 - 2|V_i||V_j|\cos\delta_{ij}). \quad (3.8)$$

Finally, all the mismatch equations expressed as the difference between the specified and calculated values are given as,

$$\begin{aligned}\Delta P_i &= P_{\text{Sp},i} - P_i, \\ \Delta Q_i &= Q_{\text{Sp},i} - Q_i, \\ \Delta T'_{i-j} &= T_{\text{Amb},i-j} - T'_{i-j},\end{aligned}\tag{3.9}$$

where  $P_{\text{Sp},i}$  and  $Q_{\text{Sp},i}$ , respectively are the specified real and reactive power injections at an  $i^{\text{th}}$  bus;  $P_i$  and  $Q_i$  at an  $i^{\text{th}}$  bus are calculated by using the steady-state load flow equations (Wang et al., 2010).

In terms of temperature-augmented Jacobian matrix, the mismatch equations in (3.9) are expressed as,

$$\begin{pmatrix} \Delta P \\ \Delta Q \\ \Delta T' \end{pmatrix} = \begin{pmatrix} \frac{\partial P}{\partial \delta} & \frac{\partial P}{\partial |V|} & \frac{\partial P}{\partial T} \\ \frac{\partial Q}{\partial \delta} & \frac{\partial Q}{\partial |V|} & \frac{\partial Q}{\partial T} \\ \frac{\partial T'}{\partial \delta} & \frac{\partial T'}{\partial |V|} & \frac{\partial T'}{\partial T} \end{pmatrix} \begin{pmatrix} \Delta \delta \\ \Delta |V| \\ \Delta T \end{pmatrix} = (\text{JM}_1) \begin{pmatrix} \Delta \delta \\ \Delta |V| \\ \Delta T \end{pmatrix}.\tag{3.10}$$

The matrix equation (3.10) relates the mismatch vector of  $P$  and  $Q$  at all buses and  $T'$  of all temperature dependent branches with that of the error vector of  $\delta$ , and  $|V|$  of all buses and  $T$  of all temperature dependent branches. The error component of state vectors using (3.10) is obtained as,

$$\begin{pmatrix} \Delta \delta \\ \Delta |V| \\ \Delta T \end{pmatrix} = (\text{JM}_1^{-1}) \begin{pmatrix} \Delta P \\ \Delta Q \\ \Delta T' \end{pmatrix}\tag{3.11}$$

where  $\text{JM}_1$  is the state vector Jacobian matrix.

The model so developed is referred to as a single slack bus temperature-augmented load flow model. The main advantage of this model is that the transmission branch temperature can be directly calculated from the load flow solution using the Newton-Raphson method. In temperature-augmented load flow, the elements in the state vector are updated using the errors estimated in (3.11) in each iteration. This continues until all mismatches fall within a presumed tolerance. The flowchart for temperature-augmented load flow adopting bus-type switching logic for checking reactive power limit specifications at a P |V| bus is shown in Figure 3.2.

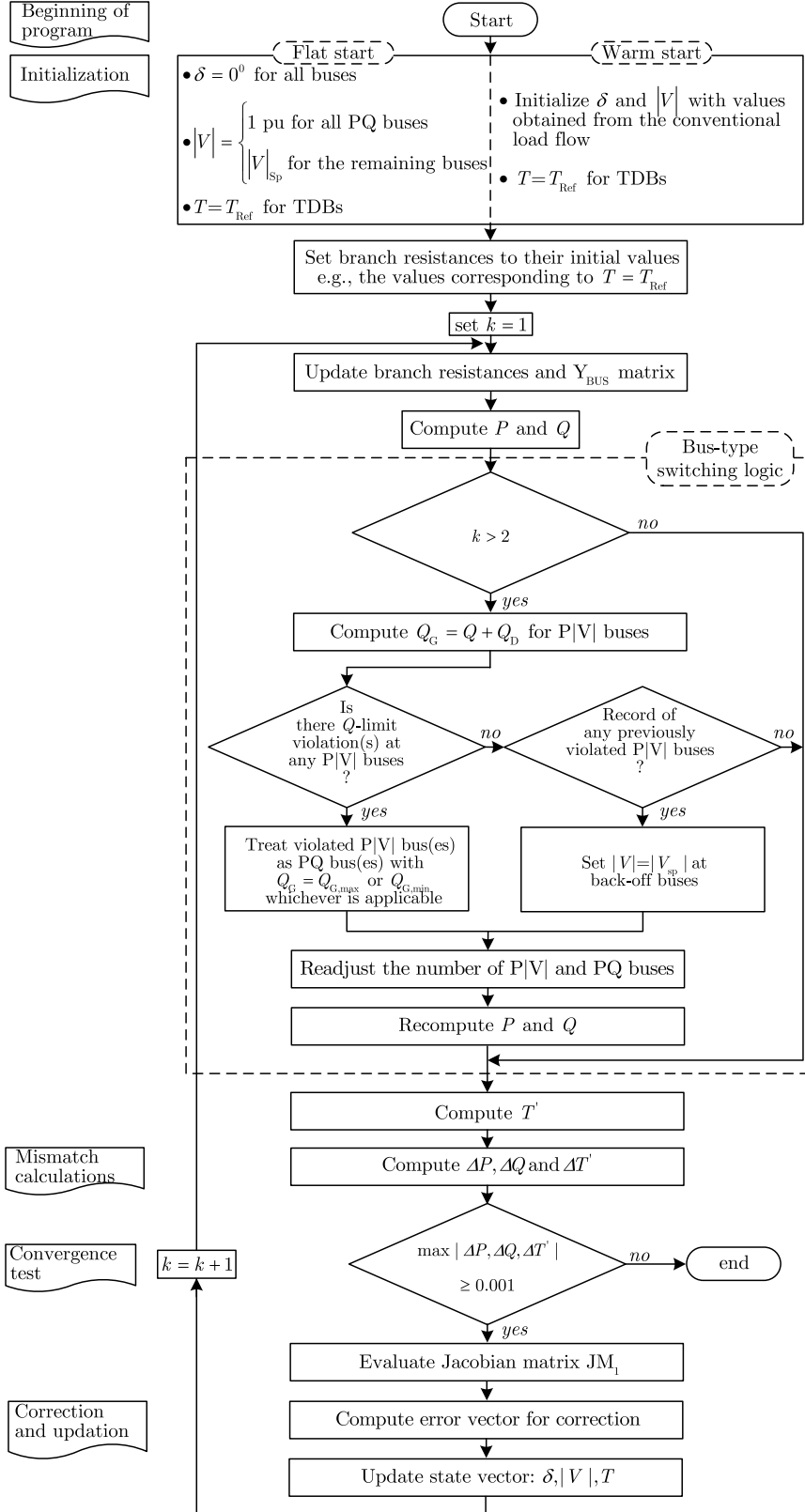


Figure 3.2: Flowchart for temperature-augmented load flow.



### 3.3 Temperature-augmented probabilistic load flow using Monte-Carlo simulation

For a TPLF simulation, the major requirements are (i) temperature-augmented power system model, (ii) probability distributions of the input random variables, and (iii) correlation informations among the input random variables.

Monte-Carlo simulation method solves the TPLF by a series of  $N_S$  temperature-augmented load flow simulations, with each time considering a set of values for input random variables. Simultaneously,  $N_S$  samples as required for the result variables are obtained. The value of  $N_S$  is generally obtained by setting variance coefficient value [refer Appendix D] less than 1% for all the result variables (Fan, 2012). The probability distributions of the TPLF result variables enable the calculation of adequacy indices as discussed in Table 3.1 which are extremely useful for over-limit risk assessment.

**Table 3.1:** Use of probability distributions of TPLF result variables.

Result variable	Adequacy index
Bus voltage magnitude	Steady-state under voltage and overvoltage probabilities.
Branch temperature	Probability of branch temperature above the allowable limit.
P V  bus reactive power	Probability of generator violating the reactive power limit.
Branch power flow	Steady-state overload probability of transmission branch.
Slack bus power	Probability of slack bus power exceeding the limit.
Branch power loss	Accurate probabilistic monitoring of sag and thermal limits.

#### 3.3.1 Generation of random samples for Monte-Carlo simulation

Under the assumption that, discrete and continuous input random variables are independent to each other, the mechanisms to generate  $N_S$  samples of discrete and correlated continuous input random variables are discussed as under.

##### 3.3.1.1 Random generations from a discrete random variable

For a given discrete random variable, define vectors  $d$  and  $p$  each of length  $ld$  that respectively stores the discrete values of the random variable and the associated prob-

ability values. Then carry out the following steps:

- (i) Develop uniform random generation vector  $u$ .
- (ii) Obtain cumulative sum vector  $c$  for the elements of  $p$ .
- (iii) For  $i = 1, 2, \dots, ld$ 
  - (a) Create vector  $lk$  that stores the linear indices of the logical expression  $(u > c(i)) \& (u \leq c(i + 1))$ .
  - (b) Develop a random generation vector  $rg$  that satisfies the condition  $rg(lk) = i$ .

### 3.3.1.2 Random generations pertaining to correlated continuous random variables

In particular, the random number generation techniques produce uniformly distributed samples whereas; the inclusion of the desired correlation is carried out in the standard Gaussian domain. So a step by step algorithm to obtain the desired  $N_S$  samples of “nc” correlated continuous input random variables is detailed underneath.

- (i) Step-1: Define matrix  $[I_D]_{1825 \times nc}$  that stores the historical observations of “nc” random variables for past five years (1825 data samples) column wise.
- (ii) Step-2: Develop product moment correlation coefficient (PMCC) matrix  $[\rho]_{nc \times nc}$  with an off-diagonal element  $\rho_{ij}$  representing the PMCC between  $i^{\text{th}}$  and  $j^{\text{th}}$  random variables as calculated from their historical observations.
- (iii) Step-3: Obtain matrix  $[u]_{N_S \times nc}$  that combines column wise the  $N_S$  samples pertaining to “nc” independent uniform random variables.
- (iv) Step-4: Transform matrix  $[u]_{N_S \times nc}$  to a new matrix  $[u^C]_{N_S \times nc}$  that combines  $N_S$  samples pertaining to “nc” correlated uniform random variables. The transformation requires a set of following steps.
  - (a) Convert the column vectors of  $[u]_{N_S \times nc}$  to standard Gaussian domain using the inverse cumulative distribution function (CDF) transformation and unite in matrix  $[v]_{N_S \times nc}$ . Note here that the transformed samples are still not correlated.
  - (b) To account for the transformation from Gaussian to uniform domain the off-diagonal elements  $\rho_{ij}$  of the desired PMCC matrix is corrected as,  $\rho_{ij}^C = 2 \sin\left(\frac{\pi}{6}\rho_{ij}\right)$  and thus leading to a new corrected PMCC matrix  $[\rho^C]_{nc \times nc}$ .

- (c) The upper triangular matrix obtained from the Cholesky decomposition of  $[\rho^C]_{nc \times nc}$  multiplied with  $[v]_{N_S \times nc}$  yields the column wise correlated samples concatenated in matrix  $[v^C]_{N_S \times nc}$ .
- (d) The obtained correlated samples in step-4(c) are transformed back to the uniform domain using CDF of the standard Gaussian distributions yields in matrix  $[u^C]_{N_S \times nc}$ .
- (v) Step-5: Establishment of the marginal distributions of the continuous random variables using a set of following sub steps.
  - (a) Transform the column vectors of matrix  $[I_D]_{1825 \times nc}$  to Gaussian domain using  $\Phi^{-1} \left[ \hat{F}(\bullet) \right]$  where  $\hat{F}(\bullet)$  is the estimated empirical CDF of the columns of  $[I_D]_{1825 \times nc}$  and  $\Phi^{-1}(\bullet)$  is the inverse of the standard Gaussian CDF.
  - (b) The principal component analysis [refer Appendix C] applied to the transformed data in step-5(a) yields  $[I_D^T]_{1825 \times nc}$ , the columns of which are independent of each other.
  - (c) The obtained data in step-5(b) are back-transformed to their original domains.
  - (d) The estimated empirical distributions of the samples of step-5(c) pertaining to the continuous random variables are referred to as their marginal distributions.
- (vi) Step-6: Finally, the samples of the correlated continuous random variables are obtained through the inverse marginal CDF transformation on  $[u^C]_{N_S \times nc}$ .

## 3.4 System over-limit risk indices

The over-limit risk indices are calculated as the product of event's over-limit probability and the corresponding severity (Li et al., 2015). The various events are under-voltage, over voltage, overload, thermal overload, etc. A detailed formulation of the same is discussed underneath.

### 3.4.1 Over-limit probability calculations

The over-limit probability of a result variable is defined as the probability of exceeding a predefined limiting value (Fan, 2012). For a result variable  $X$ , the probability of

exceeding the higher limit  $X_{H,Limit}$ ,

$$\text{prob}(X > X_{H,Limit}) = L(X_{H,Limit}) = 1 - F(X_{H,Limit}) \quad (3.12)$$

where  $L(\bullet)$  and  $F(\bullet)$ , respectively are the complementary distribution function and CDF. Now the probability of falling below the lower limit,  $X_{L,Limit}$  is denoted as,

$$\text{prob}(X < X_{L,Limit}) = F(X_{L,Limit}). \quad (3.13)$$

From the complementary distribution functions of bus voltage magnitudes, the values of under-voltage probability and over-voltage probability at the  $h^{\text{th}}$  hour are calculated as,

$$\text{UVP}_h = 1 - \prod_{i=1}^n (1 - \text{UVP}_{ih}), \text{OVP}_h = 1 - \prod_{i=1}^n (1 - \text{OVP}_{ih}) \quad (3.14)$$

where “n” is the total number of buses in the system;  $\text{UVP}_{ih}$  and  $\text{OVP}_{ih}$  respectively are the under-voltage probability and over-voltage probability values of  $i^{\text{th}}$  bus at  $h^{\text{th}}$  hour.

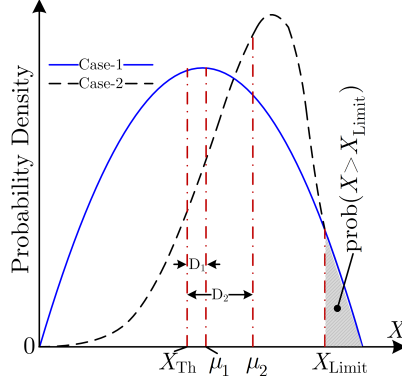
Similarly from the complementary distribution functions of branch apparent power flows and branch temperatures, the over-load probability and thermal over-load probability at the  $h^{\text{th}}$  hour are calculated as,

$$\text{OLP}_h = 1 - \prod_{k=1}^{\ell} (1 - \text{OLP}_{kh}), \text{TOLP}_h = 1 - \prod_{k=1}^{\ell} (1 - \text{TOLP}_{kh}) \quad (3.15)$$

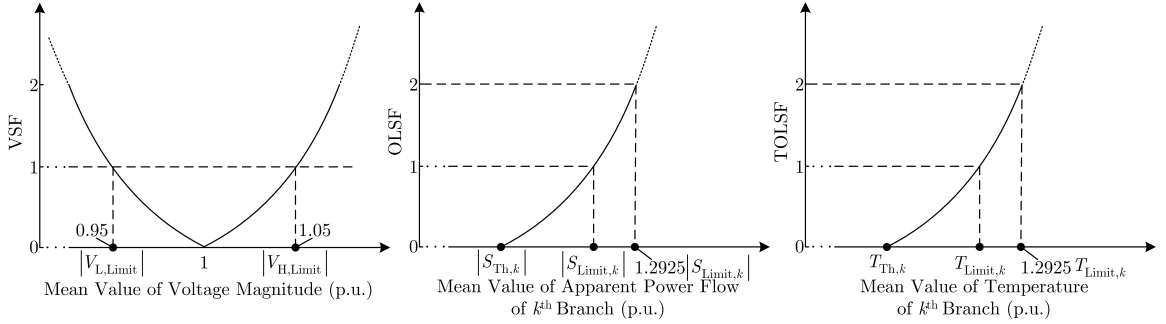
where “ $\ell$ ” is the total number of branches in the system,  $\text{OLP}_{kh}$  and  $\text{TOLP}_{kh}$  respectively are the over-load probability and thermal over-load probability values of  $k^{\text{th}}$  branch at the  $h^{\text{th}}$  hour.

### 3.4.2 Over-limit severity calculations

The severity of over-limit quantifies the deviation of mean value of the result variable from the reference or threshold value. It signifies the extent of the over-limit violation. In risk assessment, there is a need to quantify the severity of over-limit as it can be observed from Figure 3.3, the two probability density plots of the same result variable (respectively corresponding to different input uncertainties and correlations) may have the same over-limit probability values but deviations of their mean values



**Figure 3.3:** An example case to show the need of a severity function in the risk assessment.



**Figure 3.4:** Plots of severity functions.

from a prespecified threshold may be different. The severity functions used in this study are continuous (exponential functions). The voltage severity function of  $i^{\text{th}}$  bus at  $h^{\text{th}}$  hour is expressed as,

$$\text{VSF}_{ih} = e^{k_V \text{VD}_{ih}} - 1 \quad (3.16)$$

where  $k_V = \ln(2) / ||V_{\text{Limit}}| - 1|$ ; voltage deviation of  $i^{\text{th}}$  bus at  $h^{\text{th}}$  hour is calculated as,  $\text{VD}_{ih} = |1 - \mu_{|V_{ih}|}|$ ,  $\mu_{|V_{ih}|}$  is the mean value of the voltage of  $i^{\text{th}}$  bus at  $h^{\text{th}}$  hour. When  $\mu_{|V_{ih}|} = 1$  pu, voltage severity function is zero and its value equals to unity for  $\mu_{|V_{ih}|} = |V_{\text{Limit}}|$ . Similarly, over-load severity function and thermal over-load severity function of  $k^{\text{th}}$  branch at  $h^{\text{th}}$  hour are respectively given as,

$$\text{OLSF}_{kh} = e^{k_{\text{OL}} \text{OLD}_{kh}} - 1, \quad \text{TOLSF}_{kh} = e^{k_{\text{TOL}} \text{TOLD}_{kh}} - 1 \quad (3.17)$$

where  $k_{\text{OL}} = 2 \ln(2) / |S_{\text{Limit},k}|$ ;  $k_{\text{TOL}} = 2 \ln(2) / T_{\text{Limit},k}$ ;  $|S_{\text{Limit},k}|$ , and  $T_{\text{Limit},k}$  respectively are the apparent power flow limit and branch temperature limit of the  $k^{\text{th}}$

branch.

The severity functions as defined in (3.16) and (3.17) are plotted in Figure 3.4. In (3.17), over-load deviation and thermal over-load deviation of  $k^{\text{th}}$  branch at  $h^{\text{th}}$  hour are respectively calculated as,

$$\text{OLD}_{kh} = \left| |S_{\text{Th},k}| - \mu_{|S_{kh}|} \right|, \text{TOLD}_{kh} = |T_{\text{Th},k} - \mu_{T_{kh}}| \quad (3.18)$$

where  $\mu_{|S_{kh}|}$  and  $\mu_{T_{kh}}$ , respectively are the mean values of apparent power flow and temperature of the  $k^{\text{th}}$  branch at the  $h^{\text{th}}$  hour,  $|S_{\text{Th},k}|$  and  $T_{\text{Th},k}$  respectively are the threshold values of apparent power flow and temperature of  $k^{\text{th}}$  branch (taken as 50% of  $|S_{\text{Limit},k}|$  and  $T_{\text{Limit},k}$  respectively).

### 3.4.3 Calculation of risk indices

The risk of system over-voltage (RSOV) at the  $h^{\text{th}}$  hour is calculated as,

$$\text{RSOV}_h = \sum_{i=1}^n (\text{OVP}_{ih}) \cdot (\text{VSF}_{ih}). \quad (3.19)$$

Finally, the risk of system over-load (RSOL) and risk of system thermal over-load (RSTOL) at the  $h^{\text{th}}$  hour are respectively calculated as,

$$\text{RSOL}_h = \sum_{k=1}^{\ell} (\text{OLP}_{kh}) \cdot (\text{OLSF}_{kh}), \text{RSTOL}_h = \sum_{k=1}^{\ell} (\text{TOLP}_{kh}) \cdot (\text{TOLSF}_{kh}) \quad (3.20)$$

## 3.5 Case study and discussion of results

The risk assessment is carried out on modified New England 39-bus power system. The modification is the integration of the three PV arrays as discussed in section 2.2.1. The PV arrays are assumed free from providing any voltage support to the system; hence, the reactive power generations are zero, and the corresponding buses are treated as PQ type. The PV system connected buses can also be modeled as P|V| type, provided, the inverters connected to the systems have inherent reactive power capability. Suitable reactive power limits are to be set to maintain the bus voltages at the specified values. This limit corresponds to the nameplate reactive limit of the

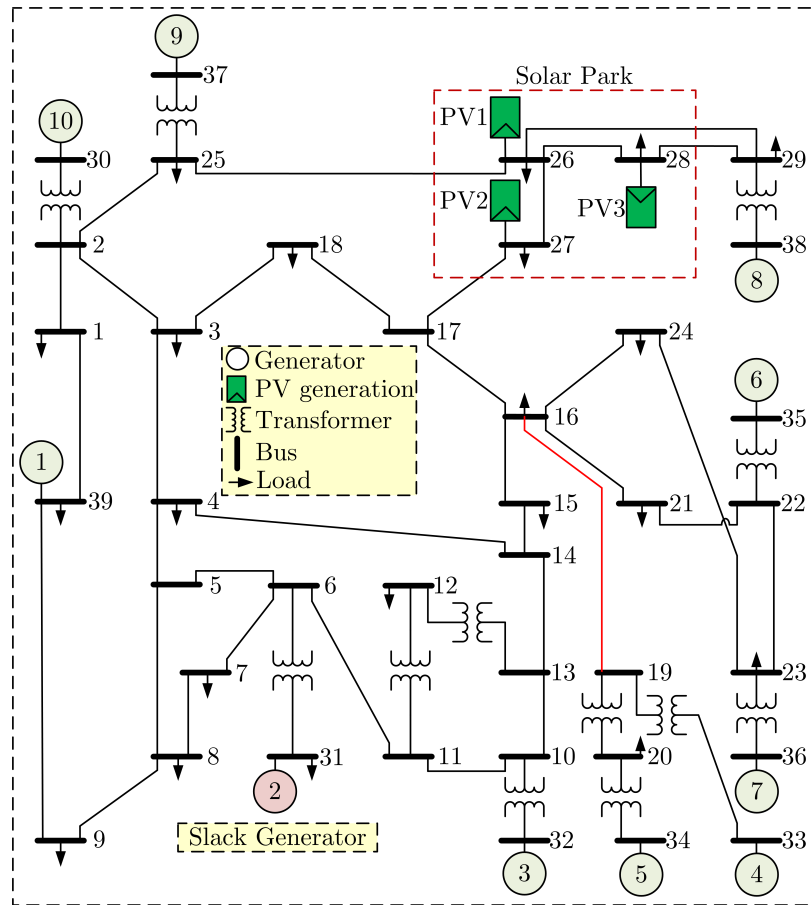


Figure 3.5: Single line diagram of the PV-integrated New England 39-bus power system.

inverter. The single line diagram of the modified New England 39-bus power system is shown in Figure 3.5.

**Table 3.2:** Probabilistic description of discrete load powers.

Bus 9		Bus 18		Probability value
Real (pu)	Reactive (pu)	Real (pu)	Reactive (pu)	
0.04	-0.61	1.32	0.15	0.10
0.05	-0.64	1.46	0.20	0.15
0.06	-0.67	1.58	0.28	0.30
0.07	-0.68	1.66	0.34	0.25
0.09	-0.69	1.70	0.43	0.20

### 3.5.1 Power system description

The power system data used in this study is adopted from (Frank et al., 2016). The system base power is set to 100 MVA, while the base value of branch temperature is set to 100 °C for computational convenience. The temperature base does not bear any relationship with voltage and power base values. The bus power data specified for the test system are deterministic, hence, the probabilistic nature of the input quantities has to be devised. The three PV arrays as discussed in section 2.2.1 are included at buses 26, 27, and 28. The loads connected at buses 9, 18, 21, 23, 24, 25, 26, 27, 28 and 29 are assumed as random variables. The discrete load instants and their corresponding probability values for real and reactive load powers at buses 9 and 18 are specified in Table 3.2. The real load powers at buses 21, 23, 24, 25, 26, 27, 28, and 29 follow Gaussian distributions with coefficient of variations same as that of the values after preprocessing of the load data of eight weather zones as discussed in section 2.2.1. The mean values correspond to the specified deterministic data. The load power factors at these buses are presumed constant. In practice, it is difficult to collect ambient temperature data for all the temperature dependent branches for possible probability distributions. Hence, for simplicity, the statistical parameters of ambient temperatures for all the temperature dependent branches are assumed same.

The PMCC matrix for TPLF is constituted among 20 continuous input random variables which include generations of the three PV arrays, ambient temperature ( $T_{Amb}$ ), real and reactive load powers at buses 21, 23, 24, 25, 26, 27, 28, and 29.



Since, the collected load power data and PV generation data belongs to different places, the PMCC between PV generations and load power is assumed to be 0.3. The other PMCC values are calculated from the historical data after preprocessing. For various result analysis, Monte-Carlo simulation with 30000 samples are used for performing PLF and TPLF simulations at noon. The programming codes for Monte-Carlo simulation are developed using MATLAB 7.10 and are executed on a computer with *i7* processor with RAM size of 8 GB. The base-case PMCC matrix defined among the first 12 continuous input random variables is indicated in Table 3.3. The PMCC values between reactive load powers and the first four input random variables [refer Table 3.3] are same as that of the real load powers and first four input random variables due to the assumptions of constant load power factors.

**Table 3.3:** Base-case PMCC matrix defined among first 12 continuous input random variables.

	PV <sub>1</sub>	PV <sub>2</sub>	PV <sub>3</sub>	T <sub>Amb</sub>	P <sub>D,21</sub>	P <sub>D,23</sub>	P <sub>D,24</sub>	P <sub>D,25</sub>	P <sub>D,26</sub>	P <sub>D,27</sub>	P <sub>D,28</sub>	P <sub>D,29</sub>
PV <sub>1</sub>	1	0.7183	0.6005	0.0518	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3
PV <sub>2</sub>	0.7183	1	0.5841	0.0538	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3
PV <sub>3</sub>	0.6005	0.5841	1	0.1356	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3
T <sub>Amb</sub>	0.0518	0.0538	0.1356	1	0	0	0	0	0	0	0	0
P <sub>D,21</sub>	0.3	0.3	0.3	0	1	0.7868	0.5172	0.5787	0.6534	0.7336	0.7799	0.5865
P <sub>D,23</sub>	0.3	0.3	0.3	0	0.7868	1	0.6205	0.8146	0.8871	0.6655	0.8559	0.7680
P <sub>D,24</sub>	0.3	0.3	0.3	0	0.5172	0.6205	1	0.7057	0.6905	0.4899	0.6486	0.7969
P <sub>D,25</sub>	0.3	0.3	0.3	0	0.5787	0.8146	0.7057	1	0.9421	0.5160	0.7334	0.8386
P <sub>D,26</sub>	0.3	0.3	0.3	0	0.6534	0.8871	0.6905	0.9421	1	0.5909	0.8222	0.8590
P <sub>D,27</sub>	0.3	0.3	0.3	0	0.7336	0.6655	0.4899	0.5160	0.5909	1	0.8071	0.6488
P <sub>D,28</sub>	0.3	0.3	0.3	0	0.7799	0.8559	0.6486	0.7334	0.8222	0.8071	1	0.8484
P <sub>D,29</sub>	0.3	0.3	0.3	0	0.5865	0.7680	0.7969	0.8386	0.8590	0.6488	0.8484	1

### 3.5.2 Impact of increased penetration of photovoltaic generations on the statistics of result variables

The percentage penetration level of PV generation (percentage of total system real load power) is expressed as,

$$\% \text{ PL} = \frac{\text{Mean value of total PV generation}}{\text{Mean value of total system load}} \times 100. \quad (3.21)$$

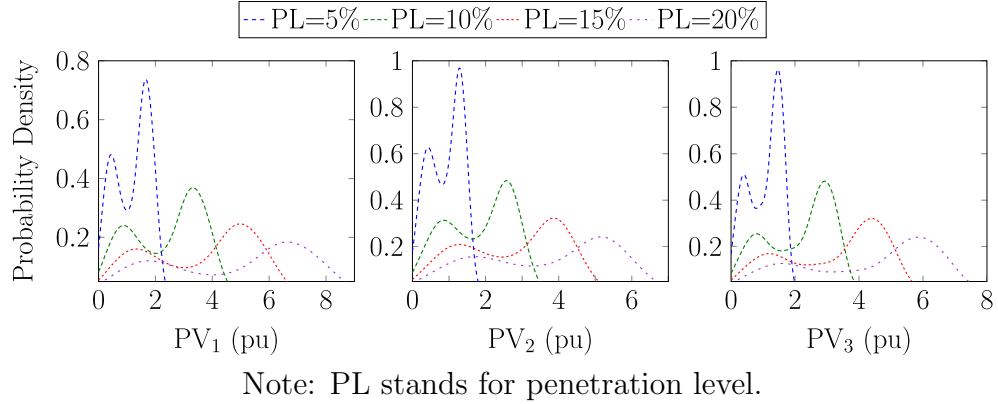
The mean value of total PV generation for a given penetration level is obtained by using (3.21). It is assumed that the obtained mean value corresponds to the

time of maximum PV generation (usually occurs at noon) and is shared among the three PV arrays. Let,  $\mu_1^{12}$ ,  $\mu_2^{12}$ , and  $\mu_3^{12}$  (the superscript “12” denotes noon) be the assigned mean values of the PV generations of three arrays based on their capacities as specified in Table 2.1. By adopting this nomenclature, a method to compute the mean values of individual PV generations is explained by considering 5% penetration level. The mean value of the total PV generation using (3.21) is obtained as 3.1271 pu (total system real load power is 62.5423 pu). The ratio of ratings of PV<sub>2</sub> to PV<sub>1</sub>, and PV<sub>3</sub> to PV<sub>1</sub> is calculated as 0.765 and 0.888 respectively. Hence, the value 3.1271 pu is shared among the PV arrays as,  $\mu_1^{12} = 1.1787$  pu,  $\mu_2^{12} = 0.9017$  pu, and  $\mu_3^{12} = 1.0467$  pu. The PV generation for the  $p^{\text{th}}$  array i.e., PV <sub>$p$</sub> <sup>12</sup> corresponding to a given penetration level is calculated by multiplying a factor (obtained as a ratio of  $\mu_p$  to the mean of actual PV generation of the  $p^{\text{th}}$  array) with actual PV generation of the array. These set of steps are applied to various other penetration levels for obtaining the mean values of PV generations at noon as indicated in Table 3.4.

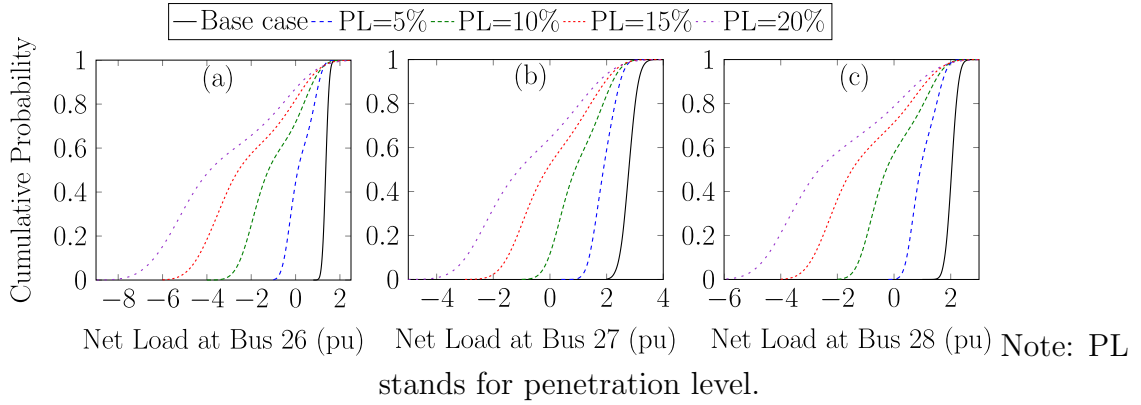
**Table 3.4:** Mean values of PV generations in pu for various penetration levels.

PV array	Base-case	Penetration level			
		5%	10%	15%	20%
PV <sub>1</sub>	0.0283	1.1787	2.3574	3.5361	4.7148
PV <sub>2</sub>	0.0222	0.9017	1.8034	2.7051	3.6068
PV <sub>3</sub>	0.0238	1.0467	2.0934	3.1401	4.1867

To analyze the effect of an increase in PV penetrations on the statistics of result variables, various penetration levels such as base-case, 5%, 10%, 15%, and 20% are considered. The probability density plots of PV generations pertaining to various penetration levels at noon for all the three PV arrays are shown in Figure 3.6. It is observed that, with an increase in penetration, the variance of PV generation increases. Further, the probability density plots are observed to be magnified while the shape remains the same. The cumulative probability plots of net real power load at buses 26, 27 and 28 for various penetration levels are plotted in Figure 3.7. It is observed that the net load power variability in all the three cases increases with an increase in PV penetration leading to a subsequent decrease in the mean values. Further, the lower tails of the distributions are shifted towards the negative axis. As it is evident from Figure 3.7, in the base-case, low PV penetration does not cause multimodality in the cumulative probability of net load power. For the remaining



**Figure 3.6:** Comparison of probability density plots of PV generations for various penetration levels.



**Figure 3.7:** Cumulative probability plots of net load at buses 26, 27 and 28 for various PV penetrations.

four PV penetration cases, cumulative probability plots of net load are multimodal at buses 26, 27 and 28. It can be further observed that for penetration above 10%, bidirectional power injections are evident at buses 27 and 28 whereas the same is noticed above 5% at bus 26.

Since the buses and branches in the vicinity to the PV array locations are subjected to more uncertainty influences (Fan et al., 2012), the probability distributions of bus voltage magnitudes of buses 26, 27 and 28; branch temperatures and power flows in the branches 26-27 and 26-28 are analyzed. Both PLF and TPLF simulations are performed at noon considering various PV penetration cases as described in Table 3.4 and the results are compared in Table 3.5. Henceforth,  $P_{L,i-j}$ ,  $Q_{L,i-j}$ , and  $|S_{L,i-j}|$  are used to represent respectively the real, reactive, and apparent power flows in the

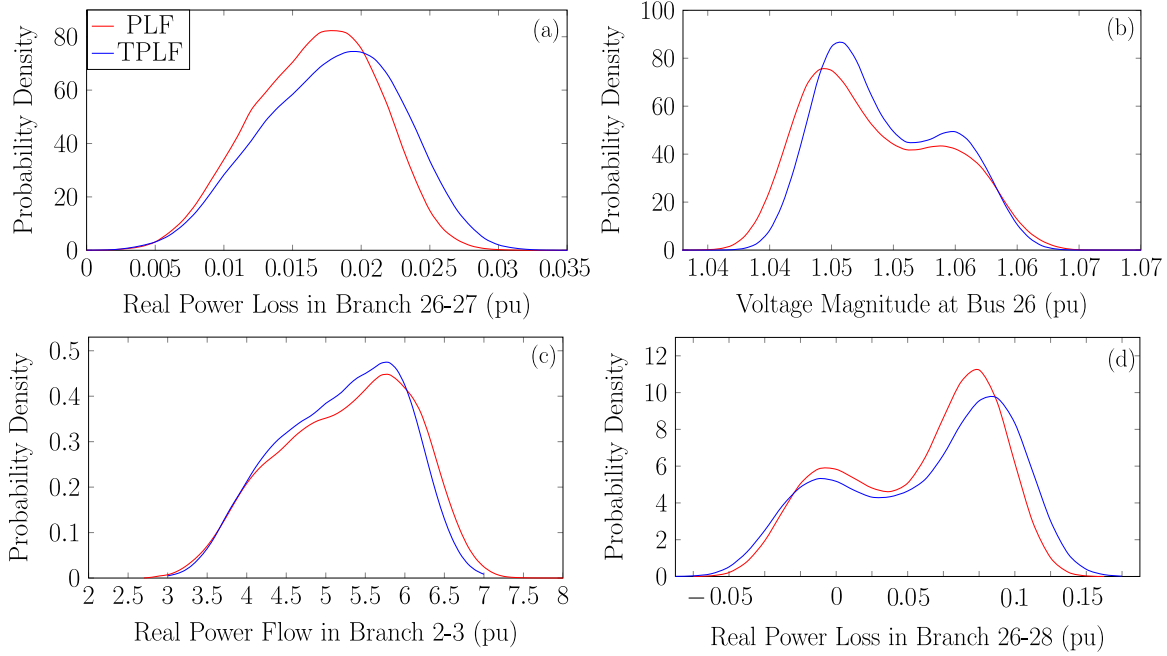
**Table 3.5:** Comparison of PLF and TPLF results for various PV penetration levels.

Result variable	PLF				TPLF						
	Base-case	PL=5%	PL=10%	PL=15%	PL=20%	Base-case	PL=5%	PL=10%	PL=15%	PL=20%	
$ V_{26} $	$\mu$	1.0520	1.0524	1.0486	1.0403	1.0267	1.0519	1.0526	1.0493	1.0419	1.0302
	$\sigma$	0.0022	0.0016	0.0056	0.0142	0.0285	0.0023	0.0016	0.0050	0.0126	0.0247
$ V_{27} $	$\mu$	1.0378	1.0375	1.0324	1.0223	1.0063	1.0374	1.0372	1.0324	1.0228	1.0080
	$\sigma$	0.0026	0.0022	0.0069	0.0168	0.0331	0.0026	0.0022	0.0066	0.0159	0.0306
$ V_{28} $	$\mu$	1.0501	1.0512	1.0496	1.0453	1.0376	1.0500	1.0513	1.0502	1.0465	1.0400
	$\sigma$	0.0018	0.0010	0.0028	0.0078	0.0165	0.0019	0.0010	0.0024	0.0067	0.0140
$T_{26-27}$	$\mu$	0.1	0.1	0.1	0.1	0.1	0.2185	0.2510	0.2963	0.3559	0.4349
	$\sigma$	0	0	0	0	0	0.0498	0.0551	0.0751	0.1169	0.1859
$T_{26-28}$	$\mu$	0.1	0.1	0.1	0.1	0.1	0.1897	0.2020	0.2186	0.2406	0.2698
	$\sigma$	0	0	0	0	0	0.0486	0.0500	0.0543	0.0654	0.0846
$P_{L,26-27}$	$\mu$	2.5941	3.5371	4.4996	5.4299	6.3454	2.5975	3.5518	4.5194	5.4645	6.3937
	$\sigma$	0.3041	0.4963	0.9299	1.3667	1.7890	0.3063	0.4985	0.9353	1.3846	1.8226
$P_{L,26-28}$	$\mu$	-1.4228	-2.0524	-2.6951	-3.3272	-3.9718	-1.4200	-2.0514	-2.6925	-3.3267	-3.9537
	$\sigma$	0.2702	0.3536	0.6269	0.9270	1.2288	0.2696	0.3525	0.6246	0.9227	1.2251
$Q_{L,26-27}$	$\mu$	0.6792	0.6794	0.7288	0.8263	0.9712	0.6840	0.6873	0.7431	0.8502	1.0131
	$\sigma$	0.0427	0.0418	0.0886	0.1872	0.3405	0.0421	0.0424	0.0935	0.2007	0.3679
$Q_{L,26-28}$	$\mu$	-0.2159	-0.1253	-0.0459	0.0194	0.0744	-0.2117	-0.1171	-0.0315	0.0451	0.1146
	$\sigma$	0.0457	0.0475	0.0780	0.1182	0.1698	0.0470	0.0500	0.0839	0.1296	0.1888
$ S_{L,26-27} $	$\mu$	2.6834	3.6029	4.5595	5.4939	6.4223	2.6879	3.6187	4.5814	5.5318	6.4768
	$\sigma$	0.2904	0.4898	0.9276	1.3733	1.8104	0.2926	0.4923	0.9336	1.3929	1.8476
$ S_{L,26-28} $	$\mu$	1.4418	2.0575	2.6973	3.3298	3.9764	1.4385	2.0561	2.6946	3.3298	3.9597
	$\sigma$	0.2589	0.3493	0.6239	0.9254	1.2280	0.2583	0.3484	0.6224	0.9217	1.2257
$P_{31}$	$\mu$	6.6174	3.6556	0.6932	-2.1857	-4.9783	6.6461	3.6737	0.7318	-2.1042	-4.8050
	$\sigma$	1.6653	1.6686	2.5176	3.5503	4.6210	1.6563	1.6504	2.4843	3.4861	4.4277
$P_{31}$	$\mu$	2.1702	1.6881	1.6305	1.9856	2.7494	2.1845	1.7048	1.6552	2.0165	2.7717
	$\sigma$	0.3973	0.1568	0.1317	0.7033	1.6449	0.3950	0.1538	0.1344	0.7010	1.5991

branch  $i-j$ .  $P_{31}$  and  $Q_{31}$ , respectively are the slack bus real and reactive powers. From the comparison of results in Table 3.5, it is clear that the increase in penetration has a noticeable effect on branch temperatures and branch loadings. The PLF and TPLF results are nearly the same for bus voltage magnitudes indicating that temperature-augmentation barely affects bus voltage magnitudes. However, on the other hand, it has a significant effect on the variance of real and reactive power flows; the effect becomes more prominent for increased penetrations. In all the above three cases for bus voltage magnitudes with the increase in penetration beyond 5%, mean values are decreased whereas the standard deviation values are increased due to the uncertainty influence of PV generations at these buses. However, in the base-case, the patterns of mean values and standard deviations are significantly biased because of bidirectional power injections at these buses. TPLF provides the statistical information of branch temperatures of temperature dependent branches which PLF fails to accomplish. Due to temperature augmentation, real power flow variability of the branch 26-27 using TPLF is significantly increased as compared to PLF. The probability density plots of few result variables using PLF and TPLF are compared in Figure 3.8. The differences in probability density plots are noticeable in all the cases. Similar inferences are obtained in case of other result variables as well. The slack bus powers are radically changed especially for 15% and 20% penetrations. The slack bus absorbs the excess real power, resembling the real-time scenario of power export to the nearby systems. The analysis of other branch power flow distributions indicates that, in most of the cases, the increase in penetration leads to bi-directional power flows indicated by the probability distributions extended to both positive and negative axes.

### 3.5.3 Effect of variations of temperature-augmented system model parameter values on the statistics of result variables

The base-case values of the model parameters  $T_{\text{Ref}}$ ,  $T_{\text{Rated rise}}$ , and  $T_{\text{Amb-wc}}$  respectively are 10 °C, 25 °C, and 40 °C. The above three parameters cannot always be considered constant and the effect of their variation on the statistics of the result variables would be of interest. The analysis is carried out for 5% penetration considering a few result variables associated with the branches where the effect of temperature dependence is the highest. As the temperature effect is taken into account, resistances of all the



**Figure 3.8:** Comparison of probability density plots of (a)  $P_{\text{Loss}, 26-27}$  (5% penetration), (b)  $|V_{26}|$  (10% penetration), (c)  $P_{L, 2-3}$  (15% penetration) and (d)  $P_{\text{Loss}, 26-28}$  (20% penetration) using PLF and TPLF.

temperature dependent branches are increased. Two relative percentage error indices are defined to quantify the effect of a change in first two statistical parameters due to temperature-augmentation, and are given as,

$$e_{\mu} = \left| \frac{\mu_{\text{WT}} - \mu_{\text{T}}}{\mu_{\text{WT}}} \right| \times 100, \quad e_{\sigma} = \left| \frac{\sigma_{\text{WT}} - \sigma_{\text{T}}}{\sigma_{\text{WT}}} \right| \times 100 \quad (3.22)$$

where  $\mu_{\text{WT}}$  and  $\mu_{\text{T}}$ , respectively are the mean values of a particular result variable as obtained using PLF and TPLF;  $\sigma_{\text{WT}}$  and  $\sigma_{\text{T}}$ , respectively are the standard deviation values of a particular result variable as obtained using PLF and TPLF.

There is a significant increase in the value of  $e_{\mu}$  with an average value that amounts to 123.84% for branch temperature due to the inclusion of temperature effect. On the other hand, the average  $e_{\mu}$  of branch power flow and branch power loss respectively amounts to 5.65% and 6.93% whereas, average  $e_{\sigma}$  of those respectively amounts to 26.29% and 33.39%. The effect of temperature-augmentation on other result variables is significantly less. The values of  $e_{\mu}$  and  $e_{\sigma}$  of branch power flow in case of a few branches where the effect of temperature dependence is prominent are indicated in Table 3.6. It is observed that the temperature augmentation leads to an increase

**Table 3.6:** Effect of temperature-augmentation on first two statistical moments of branch power flows in a few branches.

Branch	% Relative change in		Real power		Reactive power		Apparent power	
	Branch resistance	Branch temperature	$e_\mu$	$e_\sigma$	$e_\mu$	$e_\sigma$	$e_\mu$	$e_\sigma$
02-03	08.73	207.91	00.36	02.17	02.42	18.58	00.43	02.64
02-25	08.30	197.72	00.48	01.73	05.54	15.13	00.06	00.56
06-11	10.45	248.70	00.13	01.13	05.62	03.77	00.13	01.13
10-11	07.73	184.11	00.08	00.99	01.56	12.40	00.10	01.09
15-16	07.34	174.80	00.04	00.89	00.67	04.49	00.11	01.01
16-19	08.27	196.83	00.06	05.46	00.97	00.04	00.07	02.16
21-22	07.26	172.78	00.03	00.74	00.23	01.57	00.04	00.85
23-24	06.21	147.98	00.01	00.93	10.49	02.77	00.01	00.94
26-27	06.34	150.91	00.42	00.45	01.16	01.62	00.44	00.50

in power flow variability. The effect of variations of model parameter values on the average of error indices of power flows using (3.22) are provided in Table 3.7. In a majority of the cases, the increase in model parameter values either increases or decreases the values of error indices. However, in a few cases, though the effect is prominent, no fixed pattern justifies the change in error indices.

### 3.5.3.1 Practical implementation of temperature-augmented probabilistic load flow

In reality, the ambient temperature of different branches have diverse values of coefficient of variations, and the PMCCs between them are also different. Hence, a practical implementation of TPLF necessitates the collection of data through the sensors installed at different locations in the power system. Further, the ignorance of correlation effect between the ambient temperatures at various locations may degrade the aptitude of TPLF for the accurate approximation of branch power flow and power loss probability distributions.

### 3.5.4 Over-limit risk assessment

The probability distributions of result variables as obtained by TPLF are useful in determining the system over-limit risk indices under various PV penetrations and input correlations. All the system buses and branches are considered for evaluating

**Table 3.7:** Effect of variations in TPLF model parameter values on average  $e_\mu$ , and  $e_\sigma$  of branch power flows.

Result variable	Average of $e_\mu$				Average of $e_\sigma$			
	$T_{\text{Ref}}$ =10 °C	$T_{\text{Ref}}$ =15 °C	$T_{\text{Ref}}$ =20 °C	$T_{\text{Ref}}$ =25 °C	$T_{\text{Ref}}$ =10 °C	$T_{\text{Ref}}$ =15 °C	$T_{\text{Ref}}$ =20 °C	$T_{\text{Ref}}$ =25 °C
Power flow	1.6932	4.0455	2.9946	4.0364	4.0603	3.6370	3.5356	3.4557
Result variable	$T_{\text{Rated rise}}$ =25 °C	$T_{\text{Rated rise}}$ =30 °C	$T_{\text{Rated rise}}$ =35 °C	$T_{\text{Rated rise}}$ =40 °C	$T_{\text{Rated rise}}$ =25 °C	$T_{\text{Rated rise}}$ =30 °C	$T_{\text{Rated rise}}$ =35 °C	$T_{\text{Rated rise}}$ =40 °C
Power flow	1.6932	1.9870	2.0921	2.2813	4.0603	4.4332	4.7729	5.0793
Result variable	$T_{\text{Amb-wc}}$ =40 °C	$T_{\text{Amb-wc}}$ =50 °C	$T_{\text{Amb-wc}}$ =60 °C	$T_{\text{Amb-wc}}$ =70 °C	$T_{\text{Amb-wc}}$ =40 °C	$T_{\text{Amb-wc}}$ =50 °C	$T_{\text{Amb-wc}}$ =60 °C	$T_{\text{Amb-wc}}$ =70 °C
Power flow	1.6932	1.6509	1.6259	1.5471	4.0603	4.2205	3.9791	3.8785

**Table 3.8:** Over-load probability and thermal over-load probability values of temperature dependent branches for 20% penetration.

Branch	OLP	TOLP	Branch	OLP	TOLP	Branch	OLP	TOLP	Branch	OLP	TOLP
01-02	0	0	05-08	0	0	14-15	0.0930	0.0086	21-22	0	0
01-39	0	0	06-07	0	0	15-16	0.7282	0.5633	22-23	0	0
02-03	0.7734	0.5212	06-11	0.9145	0.7439	16-17	0.0184	0	23-24	0	0
02-25	0.8536	0.6998	07-08	0	0	16-19	0	0	25-26	0.3174	0.0591
03-04	0.6536	0.5421	08-09	0	0	16-21	0	0	26-27	0.6110	0.4132
03-18	0.0580	0.0055	09-39	0	0	16-24	0	0	26-28	0.0118	0.0016
04-05	0.2141	0.0453	10-11	0.7022	0.4937	17-18	0.2140	0.0426	26-29	0	0
04-14	0	0	10-13	0	0	17-27	0.6360	0.5397	28-29	0	0
05-06	0	0	13-14	0	0						

Note: OLP: Over-load probability, TOLP: Thermal over-load probability.



over-limit risks. Table 3.8 shows over-load probability ( $> |S_{\text{Limit}}|$ ) and thermal over-load probability ( $> T_{\text{Limit}}$ ) values of temperature dependent branches at noon for 20% penetration. Table 3.9 indicates the under-voltage probability ( $< 0.95$  pu) and over-voltage probability ( $> 1.05$  pu) at noon for 20% penetration at all the load buses. The value of  $|S_{\text{Limit}}|$  is taken as the MVA rating of the branch and that of  $T_{\text{Limit}}$  is set to 0.5 pu, since the maximum allowable temperature of value 50 °C is typically selected to avoid loss of strength, sag and branch losses, etc. The over-limit probability values are calculated from the respective complementary distribution functions. The system over-limit probability values using (3.14) and (3.15) for various penetrations are compared in Table 3.10 with and without considering input correlations. The over-limit probability values differ significantly considering input correlation. The error becomes more prominent with the increase in penetration. It is evident from the comparison of results that with an increase in penetration, system under-voltage probability, over-load probability, and thermal over-load probability values increase whereas, system over-voltage probability values decreases.

**Table 3.9:** Under-voltage probability and over-voltage probability values of load buses for 20% penetration.

Bus	UVP	OVP	Bus	UVP	OVP	Bus	UVP	OVP
1	0	0.1069	11	0.1406	0.0001	21	0	0.1896
2	0.0774	0.0437	12	0.0142	0.0033	22	0	0.0741
3	0.2860	0.0028	13	0.1043	0.0053	23	0	0.0595
4	0.1436	0.0025	14	0.0525	0.0045	24	0	0.7212
5	0.1039	0.0035	15	0	0.0344	25	0	0.2459
6	0.1966	0.0002	16	0.0093	0.0723	26	0.0055	0.1180
7	0.1891	0.0001	17	0.0418	0.0545	27	0	0.2692
8	0	0.0343	18	0	0.1863	28	0	0.2467
9	0.0014	0.0035	19	0	0	29	0	0.0082
10	0.0207	0.0028	20	0	0.0095			

Note: UVP: Under-voltage probability, OVP: Over-voltage probability.

Finally, the impact of input correlation on the calculation of system risk indices using (3.19) and (3.20) at noon is compared in Table 3.11 for various PV penetrations. It is interesting to note that, in most of the cases, unaccounted input correlations in analysis leads to an underestimation of risk indices. Calculation of risk indices considering input correlations reflect the system security more precisely and thus provides highly useful information for decision-makers. The system RSOV increases

**Table 3.10:** System over-limit probability values with and without considering input correlation.

Penetration level	System UVP		System OVP		System OLP		System TOLP	
	WC	BC	WC	BC	WC	BC	WC	BC
Base-case	0	0	1	1	0	0.0027	0	0.0002
5%	0	0	1	1	0.0619	0.1707	0.0098	0.0235
10%	0	0	1	1	0.9203	0.9297	0.4049	0.5029
15%	0	0.0010	0.9998	1	0.9995	0.9998	0.9539	0.9660
20%	0.4765	0.7814	0.9351	0.9609	1	1	0.9991	0.9994

Note: UVP: Under-voltage probability, OVP: Over-voltage probability, OLP: Over-load probability, TOLP: Thermal over-load probability, WC: Without correlation, BC: Base-case correlation.

**Table 3.11:** System RSOV, RSOL, and RSTOL values for various PV penetrations.

Risk index	Penetration level	Without correlation	Base-case correlation
RSOV	Base-case	04.1905	04.0715
	5%	04.6464	04.7135
	10%	02.8328	02.9559
	15%	02.1462	02.4382
	20%	01.5556	01.8008
RSOL	Base-case	0	00.0008
	5%	00.0422	00.1138
	10%	01.7250	01.8098
	15%	06.0273	05.8387
	20%	12.2257	11.5308
RSTOL	Base-case	0	0
	5%	00.0029	00.0069
	10%	00.2850	00.3600
	15%	02.1760	02.2598
	20%	07.5092	07.2365

with increase in penetration from base-case to 5% but for penetration beyond 5%, system RSOV decreases. The over-limit risk assessment at the remaining time instants of the day follows the similar set of steps as adopted for noon. It is to be noted that, as the influence of input uncertainty at a specific instant of time is evaluated, conventional generation dispatch strategies are not considered. However, it is an essential aspect of PV integrated power systems to adjust the total conventional generation to match the changes in PV generation with automatic generation control or day-ahead generation scheduling (Fan et al., 2013).

### 3.6 Conclusions

This chapter has presented a risk-based power system planning with large-scale integration of PV generations. The effect of PV penetrations and change in values of TPLF model parameters on the statistics of result variables were analyzed in detail. Further, the over-limit risk indices were calculated for various penetrations and input correlations. In summary, the following observations from the result analysis are worth noting.

- (i) The change in values of TPLF model parameters has a significant effect on statistics of result variables.
- (ii) A TPLF study considering higher PV penetrations lead to a higher variability in branch power flows and branch temperatures as compared to the other result variables.
- (iii) TPLF simulations under various PV penetrations and input correlations help in identifying the critical buses and branches by quantifying system risk indices. It assists in providing the theoretical support for the assessment of system reinforcement and its reliable operation.
- (iv) The calculated risk indices reflecting the system security level can be used as an indicator for evaluating the power system security.

As compared to PLF, an improved accuracy of TPLF results qualifies as a suitable choice for applications including (i) probabilistic assessment of line sag and thermal capacity, (ii) probabilistic voltage stability study, (iii) probabilistic optimal power flow, etc.

## Publication based on the chapter

### Refereed journal publications

1. Prusty, B. R. and Jena D. (2018). An over-limit risk assessment of PV integrated power system using probabilistic load flow based on multi-time instant uncertainty modeling. *Renewable Energy*, 116:367-383. (Elsevier)

# Chapter 4

## SENSITIVITY MATRIX BASED STEADY-STATE ANALYSIS

### 4.1 Preamble

In this chapter, an upgrade to the sensitivity matrix based power system model is proposed by considering the electro-thermal coupling effect of transmission branches. Two hybrid methods are proposed to accurately approximate the multimodal probability distributions of result variables with reduced computational effort while considering input correlations. Then after, probabilistic steady-state analysis (PSSA) using probabilistic load flow (PLF) and temperature-augmented PLF (TPLF) is carried out using the proposed hybrid methods aiding the operational study of power systems.

### 4.2 Sensitivity matrix based power system model

The primary step in PSSA for an operational study is the development of a sensitivity matrix based power system model. The sensitivity matrix is a relation matrix between input and result variables. Unlike the Monte-Carlo simulation, PLF/TPLF using sensitivity matrix based power system model facilitates in obtaining faster results in operational studies as the elements of the sensitivity matrix is calculated only once.

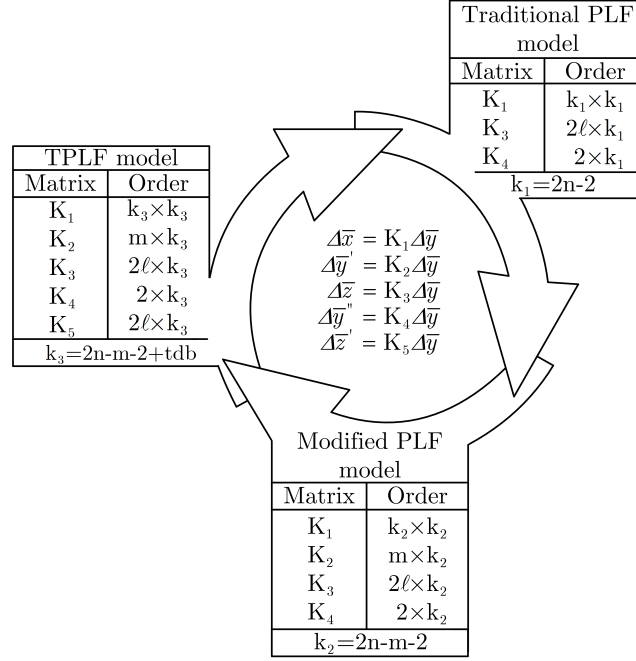
For an “ $n$ ” bus power system comprising of “ $m$ ” P|V| buses and “ $\ell$ ” branches, various sensitivity matrix based power system models are summarized in Figure 4.1. The sensitivity matrices of uncertainty components of various result variables used in

Figure 4.1 are as defined under:

- (i)  $K_1$ : State vector ( $\bar{x}$ ) sensitivity matrix.
- (ii)  $K_2$ : P |V| bus reactive power vector ( $\bar{y}'$ ) sensitivity matrix.
- (iii)  $K_3$ : Branch power flow vector ( $\bar{z}$ ) sensitivity matrix.
- (iv)  $K_4$ : Slack bus power vector ( $\bar{y}''$ ) sensitivity matrix.
- (v)  $K_5$ : Branch power loss vector ( $\bar{z}'$ ) sensitivity matrix.

It is to mention here that, the superscripts “ $\prime$ ” and “ $''$ ” used for few result variables [refer Figure 4.1] is to have a clear distinction among the result variables and it does not imply the derivative operation. The sensitivity matrices are obtained from the converged load flow solutions considering the expected values of input random variables as the point of linearization. In Figure 4.1,  $\Delta\bar{y}$  is the uncertainty component vector of input random variables and “tdb” is the total number of temperature dependent branches. The transmission branches having non-zero series resistances are referred to as temperature dependent branches. The various models are discussed as under:

- (i) Traditional PLF model: In this model, the case of generator hitting the reactive power limit is not considered, and all the P |V| buses are treated as PQ type (Fan et al., 2012). The bus power injections at the PQ buses are modeled probabilistically and the combined effect of input uncertainties on the probability distributions of bus voltages, branch power flows, and slack bus powers are obtained. Since, P |V| buses are treated as PQ type, matrix “ $K_2$ ” is absent in this model.
- (ii) Modified PLF model: In this model, the orders of the sensitivity matrices are reduced by eliminating the reactive power equations of P |V| buses. The reactive power demands at these buses may be uncertain, but the solution to their incremental change is irrelevant so long as the reactive power injections at these buses stay well within the limit (Wang and McDonald, 1994). However, in a case where solution convergence is achieved but Q-limit violates at certain P |V| buses, the respective buses are changed to PQ type. Accordingly, the orders of the sensitivity matrices will change.
- (iii) TPLF model: It is the extension of modified PLF model that accounts for the electro-thermal coupling effect by including the ambient temperatures of the transmission branches in the input vector.



**Figure 4.1:** Comparison of various sensitivity matrix based power system models.

In modified PLF model and TPLF model, P |V| bus voltage magnitude uncertainty can also be included by modifying the dimensions of the Jacobian matrices. For steady-state power flows, the voltage magnitude of the P |V| buses does not change over the specified value as long as the system has sufficient reactive power support.

#### 4.2.1 Evaluation of sensitivity matrices of temperature-augmented power system model

In compressed form (3.11) is represented as,

$$\Delta \bar{x} = \mathbf{J} \mathbf{M}_1^{-1} \Delta \bar{y} = \mathbf{K}_1 \Delta \bar{y} \quad (4.1)$$

Using (4.1), the uncertainty component vectors of  $\bar{y}'$ ,  $\bar{z}$ ,  $\bar{y}''$  and  $\bar{z}'$  (Sauer and Hoveida, 1982, Fan et al., 2012) are respectively given as,

$$\begin{aligned} \Delta \bar{y}' &= \mathbf{J} \mathbf{M}_2 \Delta \bar{x} = \mathbf{J} \mathbf{M}_2 \mathbf{K}_1 \Delta \bar{y} = \mathbf{K}_2 \Delta \bar{y} \\ \Delta \bar{z} &= \mathbf{J} \mathbf{M}_3 \Delta \bar{x} = \mathbf{J} \mathbf{M}_3 \mathbf{K}_1 \Delta \bar{y} = \mathbf{K}_3 \Delta \bar{y} \\ \Delta \bar{y}'' &= \mathbf{J} \mathbf{M}_4 \Delta \bar{x} = \mathbf{J} \mathbf{M}_4 \mathbf{K}_1 \Delta \bar{y} = \mathbf{K}_4 \Delta \bar{y} \\ \Delta \bar{z}' &= \mathbf{J} \mathbf{M}_5 \Delta \bar{x} = \mathbf{J} \mathbf{M}_5 \mathbf{K}_1 \Delta \bar{y} = \mathbf{K}_5 \Delta \bar{y} \end{aligned} \quad (4.2)$$

Formulations of the Jacobian matrices  $\text{JM}_1$ ,  $\text{JM}_2$ ,  $\text{JM}_3$ ,  $\text{JM}_4$  and  $\text{JM}_5$  are detailed in Appendix E. Using (4.1) and (4.2), the vectors  $\bar{x}$ ,  $\bar{y}'$ ,  $\bar{z}$ ,  $\bar{y}''$  and  $\bar{z}'$  can be written as,

$$\begin{aligned}
\bar{x} &= \text{K}_1 \bar{y} + \bar{x}_0, \bar{x}_0 = \bar{x}^0 - \text{K}_1 \bar{y}^0 \\
\bar{y}' &= \text{K}_2 \bar{y} + \bar{y}'_0, \bar{y}'_0 = \bar{y}'^0 - \text{K}_2 \bar{y}^0 \\
\bar{z} &= \text{K}_3 \bar{y} + \bar{z}_0, \bar{z}_0 = \bar{z}^0 - \text{K}_3 \bar{y}^0 \\
\bar{y}'' &= \text{K}_4 \bar{y} + \bar{y}''_0, \bar{y}''_0 = \bar{y}''^0 - \text{K}_4 \bar{y}^0 \\
\bar{z}' &= \text{K}_5 \bar{y} + \bar{z}'_0, \bar{z}'_0 = \bar{z}'^0 - \text{K}_5 \bar{y}^0
\end{aligned} \tag{4.3}$$

In (4.3)  $\bar{x}^0$ ,  $\bar{y}^0$ ,  $\bar{y}'^0$ ,  $\bar{y}''^0$ ,  $\bar{z}^0$  and  $\bar{z}'^0$  are the expected values of  $\bar{x}$ ,  $\bar{y}$ ,  $\bar{y}'$ ,  $\bar{y}''$ ,  $\bar{z}$  and  $\bar{z}'$  respectively. The values of  $\bar{x}^0$ ,  $\bar{y}^0$ ,  $\bar{y}'^0$ ,  $\bar{y}''^0$ ,  $\bar{z}^0$ ,  $\bar{z}'^0$ , “ $\text{K}_1$ ”, “ $\text{K}_2$ ”, “ $\text{K}_3$ ”, “ $\text{K}_4$ ” and “ $\text{K}_5$ ” are obtained from converged load flow solution. Separating the discrete and continuous parts of  $\bar{y}$  in (4.3) yields,

$$\begin{aligned}
\bar{x} &= \text{K}_1 \bar{y}^{\text{D}} + \text{K}_1 \bar{y}^{\text{C}} + \bar{x}_0 = \bar{x}^{\text{D}} + \bar{x}^{\text{C}} + \bar{x}_0 \\
\bar{y}' &= \text{K}_2 \bar{y}^{\text{D}} + \text{K}_2 \bar{y}^{\text{C}} + \bar{y}'_0 = \bar{y}'^{\text{D}} + \bar{y}'^{\text{C}} + \bar{y}'_0 \\
\bar{z} &= \text{K}_3 \bar{y}^{\text{D}} + \text{K}_3 \bar{y}^{\text{C}} + \bar{z}_0 = \bar{z}^{\text{D}} + \bar{z}^{\text{C}} + \bar{z}_0 \\
\bar{y}'' &= \text{K}_4 \bar{y}^{\text{D}} + \text{K}_4 \bar{y}^{\text{C}} + \bar{y}''_0 = \bar{y}''^{\text{D}} + \bar{y}''^{\text{C}} + \bar{y}''_0 \\
\bar{z}' &= \text{K}_5 \bar{y}^{\text{D}} + \text{K}_5 \bar{y}^{\text{C}} + \bar{z}'_0 = \bar{z}'^{\text{D}} + \bar{z}'^{\text{C}} + \bar{z}'_0
\end{aligned} \tag{4.4}$$

where  $\bar{x}^{\text{C}} = \bar{x}^{\text{G}} + \bar{x}^{\text{NG}}$ ,  $\bar{y}'^{\text{C}} = \bar{y}'^{\text{G}} + \bar{y}'^{\text{NG}}$ ,  $\bar{y}''^{\text{C}} = \bar{y}''^{\text{G}} + \bar{y}''^{\text{NG}}$ ,  $\bar{z}^{\text{C}} = \bar{z}^{\text{G}} + \bar{z}^{\text{NG}}$ , and  $\bar{z}'^{\text{C}} = \bar{z}'^{\text{G}} + \bar{z}'^{\text{NG}}$ .

**Table 4.1:** Orders of matrices and vectors used in (4.3).

Vector	Order	Vector	Order	Vector	Order	Matrix	Order	Matrix	Order
$\bar{x}$	$k \times 1$	$\bar{y}'$	$m \times 1$	$\bar{y}'_0$	$m \times 1$	$\text{K}_1$	$k \times k$	$\text{K}_4$	$2 \times k$
$\bar{y}$	$k \times 1$	$\bar{y}''$	$2 \times 1$	$\bar{z}_0$	$2\ell \times 1$	$\text{K}_2$	$m \times k$	$\text{K}_5$	$2\ell \times k$
$\bar{z}$	$2\ell \times 1$	$\bar{x}_0$	$k \times 1$	$\bar{y}''_0$	$2 \times 1$	$\text{K}_3$	$2\ell \times k$		

Note:  $k = 2n - m - 2 + \text{tdb}$ .

In (4.4) the superscripts “D”, “C”, “G” and “NG” stand for discrete, continuous, Gaussian and non-Gaussian respectively. The components  $y^{\text{D}}$ ,  $y^{\text{G}}$  and  $y^{\text{NG}}$  may not be present at all the buses. Further, these components at all the buses are either correlated or a few are correlated while remaining statistically independent with the rest. The presence of discrete and non-Gaussian input random variables in the linear combination of (4.4) may result in multimodal probability distributions of the



result variables depending mainly on the sensitivity coefficients. In (4.4), the discrete and continuous components are assumed independent (Usaola, 2009). The orders of different vectors/matrices involved in (4.3) are indicated in Table 4.1. The model formulated in (4.4) is the proposed sensitivity matrix based model where the orders of the sensitivity matrices depend on the model type as described in Figure 4.1.

### 4.3 Proposed hybrid methods

For an operational study the development of a suitable analytical method or hybridization of the existing analytical methods is an equally important task for the approximation of multimodal probability distributions of result variables. The proposed methods aim at hybridizing extended cumulant method (ECM) with Gaussian mixture approximation (GMA) and then with copula-based sequence operation theory [refer Figure 4.2]. The corresponding advantages and disadvantages of each of the methods chosen for hybridization are highlighted in Figure 4.2. First, the methods are elaborated, followed by the motivations for hybridization is discussed. Then after, each of the hybrid methods is explained in detail. The proposed hybrid methods necessitate linear relationships among the result variables and the input random variables. The generalized expression of a result variable in (4.4) can be represented as,

$$\begin{aligned}
Y = & \left\{ sc_1 \cdot y_1^D + sc_2 \cdot y_2^D + \dots + sc_{2n-m-2+tdb} \cdot y_{2n-m-2+tdb}^D \right\} \\
& + \left\{ sc_1 \cdot y_1^G + sc_2 \cdot y_2^G + \dots + sc_{2n-m-2+tdb} \cdot y_{2n-m-2+tdb}^G \right\} \\
& + \left\{ sc_1 \cdot y_1^{NG} + sc_2 \cdot y_2^{NG} + \dots + sc_{2n-m-2+tdb} \cdot y_{2n-m-2+tdb}^{NG} \right\} + Y_0
\end{aligned} \tag{4.5}$$

where “sc” is sensitivity coefficient, the value of which is either positive or negative;  $Y_0$  is a constant term that corresponds to the last summand of (4.4).

For non-zero sensitivity coefficients, considering each summand of (4.5) as an equivalent random variable, a simplified representation is given as,

$$Y = (Y_1^D \pm Y_2^D \pm \dots \pm Y_{nd}^D) + (Y_1^C \pm Y_2^C \pm \dots \pm Y_{nc}^C) \pm Y_0. \tag{4.6}$$

where subscripts “nd” and “nc” are respectively the total number of discrete and continuous input random variables.

Separating the Gaussian and non-Gaussian components of (4.6) yields,

$$\begin{aligned}
Y &= (Y_1^D \pm Y_2^D \pm \dots \pm Y_{nd}^D) + (Y_1^G \pm Y_2^G \pm \dots \pm Y_{ng}^G) \\
&\quad + (Y_1^{NG} \pm Y_2^{NG} \pm \dots \pm Y_{nng}^{NG}) \pm Y_0 \\
&= Y^{eD} \pm Y^{eG} \pm Y^{eNG} \pm Y_0
\end{aligned} \tag{4.7}$$

where subscripts “ng” and “nng” are respectively the total number of Gaussian and non-Gaussian input random variables;  $Y^{eD}$ ,  $Y^{eG}$  and  $Y^{eNG}$  respectively are the equivalent discrete, equivalent Gaussian, and equivalent non-Gaussian components.

### 4.3.1 Extended cumulant method

ECM is an extended version of cumulant method which is capable of estimating the cumulants (of any order) of a linear combination of correlated Gaussian and non-Gaussian random variables (Hoese and Garcés, 1999). The linear combination of correlated continuous input random variables as expressed in (4.6) is given as,

$$Y^{eC} = Y_1^C \pm Y_2^C \pm \dots \pm Y_{nc}^C. \tag{4.8}$$

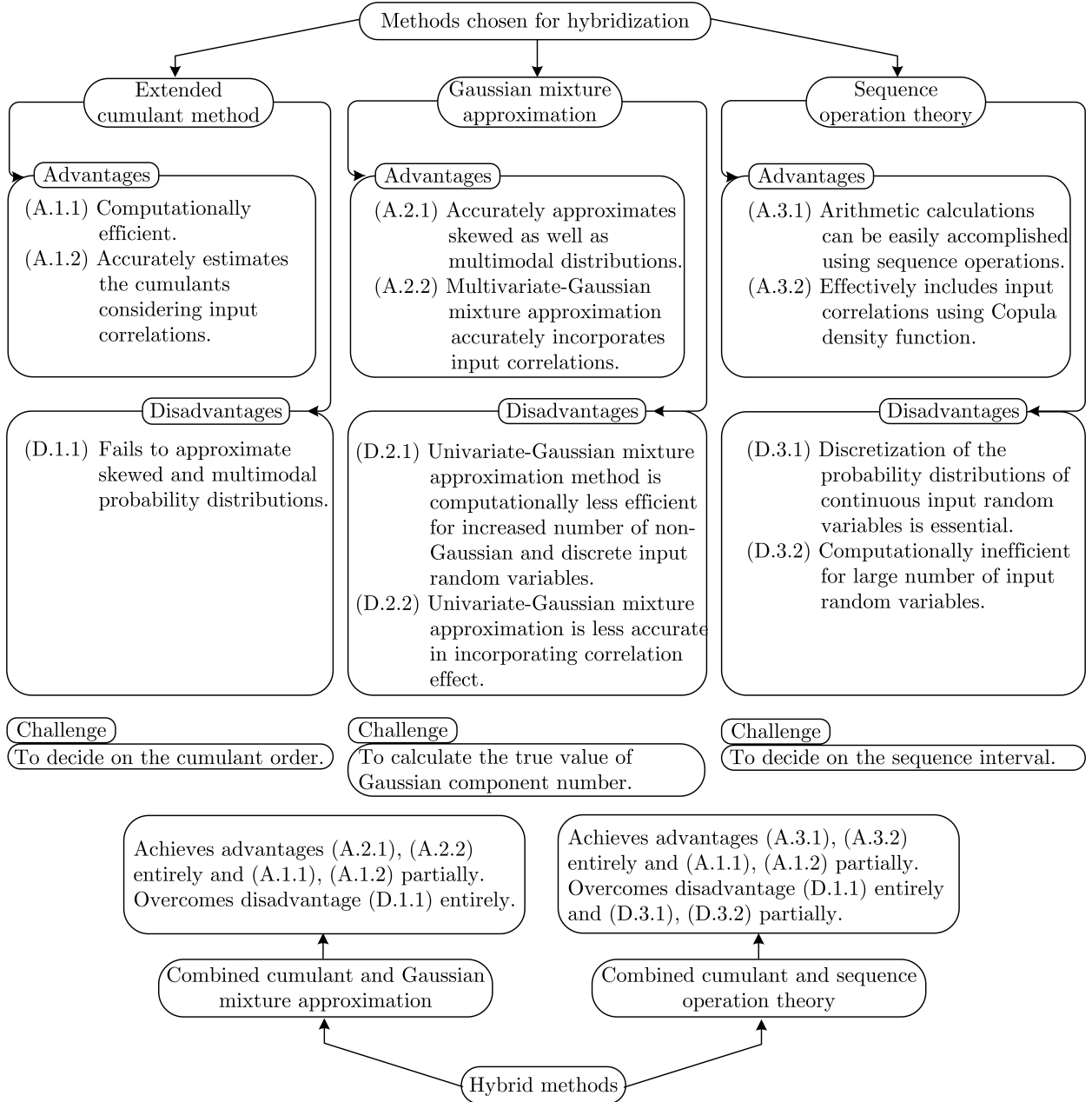
In compact form, (4.8) can be written as  $Y^{eC} = W_{nc-2} \pm Y_{nc}^C$  where  $W_{nc-2}$  is evaluated in  $(nc - 2)^{th}$  step. At any  $i^{th}$  step, as can be seen from Figure 4.3 that,  $W_{i-1}$  uses the statistical parameters of  $W_0$  to  $W_{i-2}$  e.g.,  $(nc - 1)^{th}$  step uses the statistical parameters of  $W_0, W_1, \dots, W_{nc-2}$  where  $W_0 = Y_1^C$ .

In step-1,  $W_1 = Y_1^C \pm Y_2^C$ . The cumulants of  $W_1$  are obtained as,

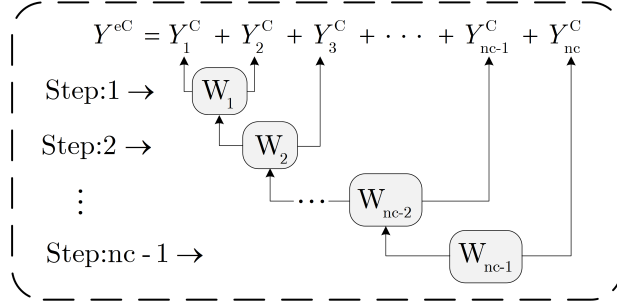
$$C_{W_1, k} = \begin{cases} A(k) C_{Y_1^C, k} + (\pm 1)^k C_{Y_2^C, k}, & \sigma_{Y_2^C} \geq \sigma_{Y_1^C} \\ C_{Y_1^C, k} + (\pm 1)^k A(k) C_{Y_2^C, k}, & \sigma_{Y_1^C} \geq \sigma_{Y_2^C} \end{cases} \tag{4.9}$$

where the function  $A(k) = (1 + \rho)^k - \rho^k$ ,  $\rho = \begin{cases} \pm \rho_{Y_1^C, Y_2^C} \left( \frac{\sigma_{Y_2^C}}{\sigma_{Y_1^C}} \right), & \sigma_{Y_2^C} \geq \sigma_{Y_1^C} \\ \pm \rho_{Y_1^C, Y_2^C} \left( \frac{\sigma_{Y_1^C}}{\sigma_{Y_2^C}} \right), & \sigma_{Y_1^C} \geq \sigma_{Y_2^C} \end{cases}$ .

In (4.9),  $C_{Y_1^C, k}$ ,  $C_{Y_2^C, k}$ , and  $C_{W_1, k}$  are the  $k^{th}$  order cumulants of  $Y_1^C$ ,  $Y_2^C$ , and  $W_1$  respectively;  $\sigma_{Y_1^C}$  and  $\sigma_{Y_2^C}$  are the standard deviations of  $Y_1^C$  and  $Y_2^C$  respectively;  $\rho_{Y_1^C, Y_2^C}$  is the product moment correlation coefficient (PMCC) between  $Y_1^C$  and  $Y_2^C$ . For  $W_i = W_{i-1} \pm Y_{i+1}^C$ ,  $1 \leq i \leq nc - 2$ , the parameters  $\sigma_{W_i}$  and  $\rho_{W_i, Y_{i+2}^C}$  are respectively calculated as,



**Figure 4.2:** Methodology for hybridization.



**Figure 4.3:** Steps to obtain cumulants of a linear combination of correlated continuous random variables.

$$\sigma_{W_i} = \sqrt{\sigma_{W_{i-1}}^2 \pm 2\rho_{W_{i-1}, Y_{i+1}^C} \sigma_{W_{i-1}} \sigma_{Y_{i+1}^C} + \sigma_{Y_{i+1}^C}^2}, \quad (4.10)$$

$$\rho_{W_i, Y_{i+2}^C} = \frac{\rho_{W_{i-1}, Y_{i+2}^C} \sigma_{W_{i-1}} \pm \rho_{Y_{i+1}^C, Y_{i+2}^C} \sigma_{Y_{i+1}^C}}{\sigma_{W_i}}. \quad (4.11)$$

### 4.3.2 Gaussian mixture approximation

GMA method approximates the probability distributions of both non-Gaussian and discrete input random variables through a convex summation of Gaussian distributions. GMA as applied to individual non-Gaussian one-dimensional random variables is referred to as univariate-GMA (Sirisena and Brown, 1983, Valverde et al., 2012). In multivariate-GMA, the correlated non-Gaussian input random variables at various buses are augmented together forming a multivariate structure. Both the types are discussed underneath.

- (i) Univariate-GMA: The probability density function (PDF) of a non-Gaussian bus power  $y_j^{\text{NG}}$  at  $j^{\text{th}}$  bus can be approximated by a  $g_j^{\text{th}}$  order Gaussian sum given as,

$$f(y_j^{\text{NG}}) = \sum_{k=1}^{g_j} w_{jk} f_{N_k}(\mu_{jk}, \sigma_{jk}^2)(y_j^{\text{NG}}); \sum_{k=1}^{g_j} w_{jk} = 1 \quad (4.12)$$

where  $w_{jk}$  is the weight factor;  $f_{N_k}$  is Gaussian PDF of the  $k^{\text{th}}$  component of  $y_j^{\text{NG}}$ ;  $\mu_{jk}$  and  $\sigma_{jk}^2$  are respectively the mean and variance of the  $k^{\text{th}}$  component.

A discrete bus power  $y_j^{\text{D}}$  at  $j^{\text{th}}$  bus is approximated as a mixture of Gaussian components given as,

$$f(y_j^D) = \sum_{k=1}^{ld_j} w'_{jk} f_{N_k(d_{jk},0)}(y_j^D); \sum_{k=1}^{ld_j} w'_{jk} = 1 \quad (4.13)$$

where  $d_{jk}$  and  $w'_{jk}$  are the discrete values and the corresponding probability weights of  $y_j^D$  respectively;  $ld$  is the length of the discrete sequence.

- (ii) Multivariate-GMA: In case of multivariate-GMA, the probability distribution of an  $nr$ -dimensional augmented structure  $y_a$  is given as,

$$f(y_a) = \sum_{k=1}^g w_k f_{N_k}(y_a); \sum_{k=1}^g w_k = 1. \quad (4.14)$$

The PDF of  $nr$ -dimensional  $k^{\text{th}}$  Gaussian PDF is expressed as,

$$f_{N_k}(y_a) = \frac{1}{(2\pi)^{\frac{nr}{2}} |\Sigma_k|^{\frac{1}{2}}} e^{-\frac{1}{2}(y_a - \mu_k)^T (\Sigma_k)^{-1} (y_a - \mu_k)} \quad (4.15)$$

where  $w_k$ ,  $\mu_k$  and  $\Sigma_k$  are the weight factor, mean vector and covariance matrix of the  $nr$ -dimensional  $k^{\text{th}}$  Gaussian PDF respectively;  $|\Sigma_k|$  and  $(\Sigma_k)^{-1}$  are the determinant and inverse of  $\Sigma_k$  respectively.

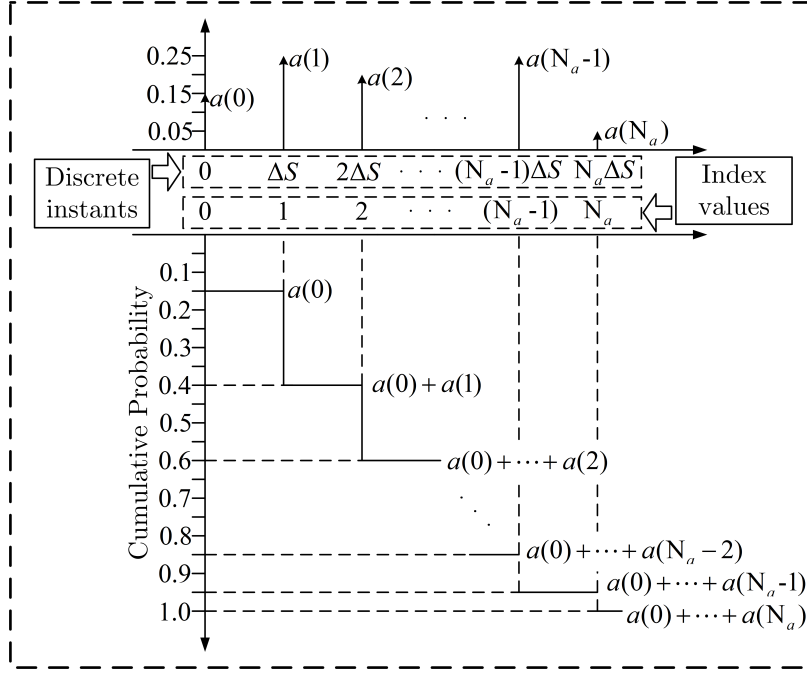
The probability distributions of discrete bus powers are modeled in a similar way as in the case of univariate-GMA. The weight factors/probability weights satisfy the condition:  $0 < w, w' \leq 1$ .

The Gaussian mixture model (GMM) parameters of each Gaussian component for a specific “g” are obtained using the expectation maximization algorithm (Carmona-Delgado et al., 2015).

### 4.3.3 Sequence operation theory

In sequence operation theory-based framework, knowledge of the following concepts is essential.

- (i) Discrete sequence: A discrete sequence  $a(i)$  is defined as a sequence comprising of numeric values at discrete instants  $i = 0, 1, \dots, N_a$  satisfying the conditions:  $a(i) = 0, i > N_a$  and  $a(i) \neq 0, i = N_a$  where  $N_a$  is the sequence length.



**Figure 4.4:** Conversion of a probability sequence to cumulative probability plot and vice versa.

- (ii) Probability sequence: A discrete sequence  $a(i)$  is said to be a probability sequence if it satisfies,

$$\sum_{i=0}^{N_a} a(i) = 1. \quad (4.16)$$

An example of probability sequence with sequence interval  $\Delta S$  is depicted in Figure 4.4. The cumulative sum of probabilities obtains its cumulative probability plot. Similarly, for a given cumulative probability plot, the corresponding probability sequence can be obtained using the reverse process. A probability distribution, as well as a scalar, can be represented as a probability sequence. The latter is referred to as an one-point distribution.

- (iii) Sequence operations: Sequence operations are used to accomplish arithmetic operations (Kang et al., 2002); they are addition type convolution, subtraction type convolution, and sequence multiplication operation. For two probability

sequences  $a(i_a)$  and  $b(i_b)$ , they are respectively expressed as,

$$\begin{aligned}
u(i) &= \sum_{i_a+i_b=i} \text{CD}(i_a, i_b) \cdot a(i_a) \cdot b(i_b), \quad i = 0, 1, \dots, N_a + N_b \\
v(i) &= \begin{cases} \sum_{i_a \leq i_b} \text{CD}(i_a, i_b) \cdot a(i_a) \cdot b(i_b), & i = 0 \\ \sum_{i_a - i_b = i} \text{CD}(i_a, i_b) \cdot a(i_a) \cdot b(i_b), & 1 \leq i \leq N_a \end{cases} \\
s(i) &= \sum_{i_a i_b = i} a(i_a) \cdot b(i_b), \quad i = 0, 1, \dots, N_a N_b
\end{aligned} \tag{4.17}$$

The copula density function  $\text{CD}(i_a, i_b)$  in (4.17) is expressed as,

$$\text{CD}(i_a, i_b) = \text{cou} \left( \sum_{k_1=0}^{i_a} a(k_1), \sum_{k_2=0}^{i_b} b(k_2), \rho_{ab} \right) = \text{cou}(U_1, U_2, \rho_{ab}) \tag{4.18}$$

where  $i_a = 0, 1, \dots, N_a$ ,  $i_b = 0, 1, \dots, N_b$  and  $\rho_{ab}$  is the PMCC between those two continuous random variables from where the sequences  $a$  and  $b$  are originated.

The addition type convolution and subtraction type convolution are strictly applicable to sequences having the same interval. The resultant sequence interval is same as that of the input. On the other hand, sequence multiplication operation is applied to sequences (at least one must correspond to a scalar) with different intervals and the resultant sequence interval is the product of individual intervals.

#### 4.3.4 Combined cumulant and Gaussian mixture approximation

The hybridization of cumulant method and GMA referred to as combined cumulant and Gaussian mixture approximation (CCGMA) does not necessitates the convolution operations of conventional GMA. The same is achieved with cumulant calculations, thereby saving substantial computational time.

Since a GMM approximates probability distributions of non-Gaussian input random variables by convex summation of Gaussian distributions and each Gaussian distribution can be adequately described by first two cumulants, cumulant-based rep-

representations of (4.12) and (4.13) are given as,

$$C_{y_j^{NG}} = \sum_{k=1}^{g_j} w_{jk} N_k \left( \begin{array}{c} \mu_{jk} \\ \sigma_{jk}^2 \end{array} \right), C_{y_j^D} = \sum_{k=1}^{ld} w'_{jk} N_k \left( \begin{array}{c} d_{jk} \\ 0 \end{array} \right). \quad (4.19)$$

In CCGMA, each non-Gaussian input random variable or an augmented structure are approximated by a  $g^{\text{th}}$  order Gaussian sum. In case of a discrete input random variable, the approximated Gaussian sum number equals to the number of discrete impulses. The total number of Gaussian components ( $N_r$ ) essential to approximate the probability distribution of a result variable is calculated as the product of the Gaussian components pertaining to all non-Gaussian and discrete input random variables. Finally, the distribution of a result variable can be established from the weighted sum of distributions of Gaussian components obtained in  $N_r$  evaluations. In each evaluation, equivalent probability weight is obtained as the product of probability weights of all the Gaussian components in the respective evaluation. The accuracy of the estimated multimodal probability distributions of result variables mainly depends on the determination of true number of mixture components and a proper initialization strategy. The true number of mixture components are obtained using cluster distortion function based approach (Pham et al., 2005), and the results of k-means clustering algorithm pertaining to the optimal cluster number are used for the initialization of expectation-maximization algorithm to obtain GMM parameters. A flowchart for the implementation of univariate-CCGMA is provided in Figure 4.5.

#### 4.3.4.1 Cluster number selection approach

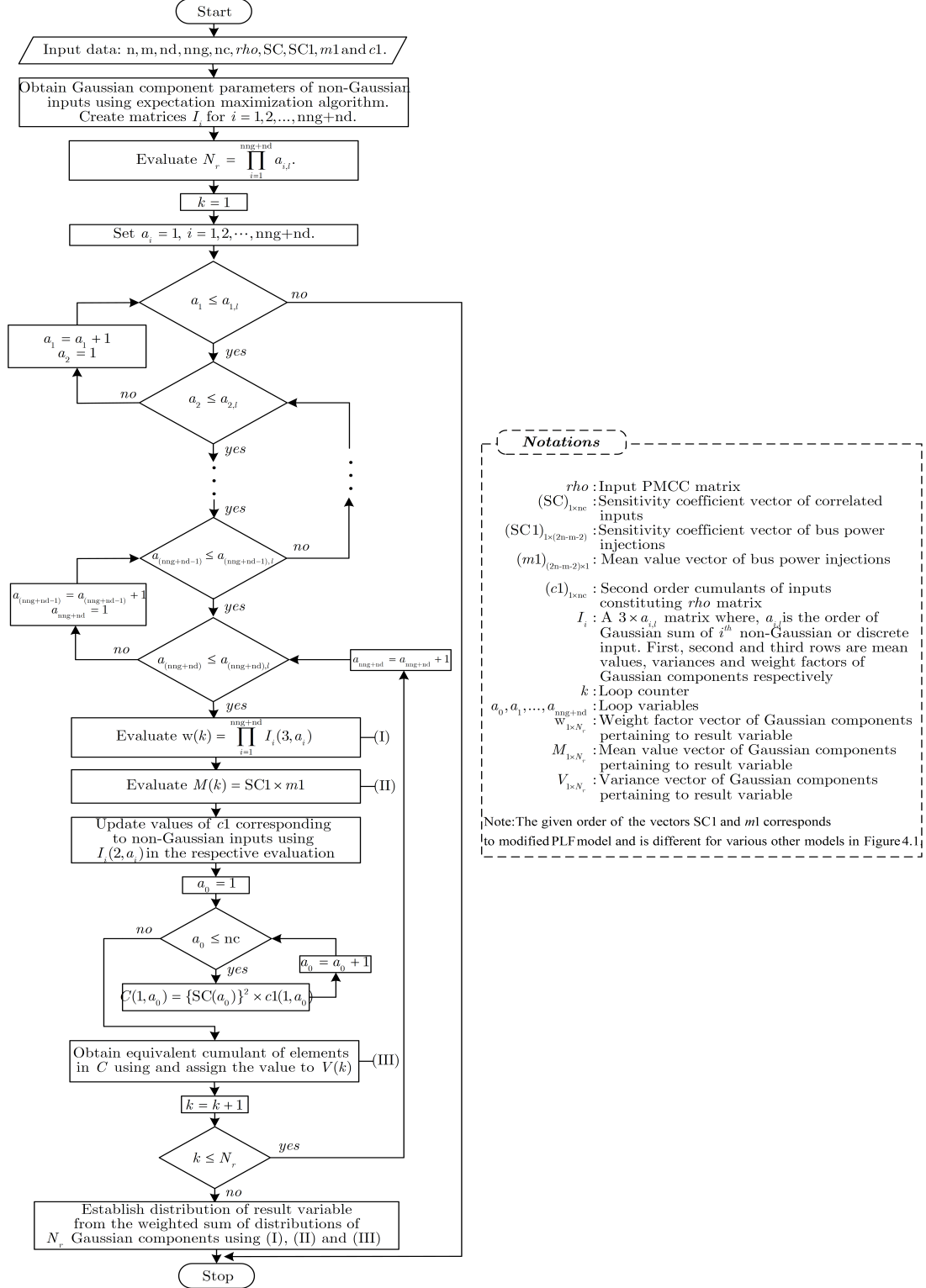
A cluster distortion function-based approach is applied for the estimation of an optimal number of data clusters for a given data. Here, the idea is to identify the concentrated regions of objects in the data, i.e., the data distortion. The sum of cluster distortion for a given value of “ $g$ ” is given as,

$$S_g = \sum_{j=1}^g I_j \quad (4.20)$$

and the function  $f(g)$  for cluster number selection is given as,

$$f(g) = \begin{cases} 1 & , g = 1 \\ \frac{S_g}{\alpha_g S_{g-1}}, S_{g-1} \neq 0 \ \& \ g > 1 \end{cases} \quad (4.21)$$





**Figure 4.5:** Flowchart for establishing the probability distribution of a result variable using univariate-CCGMA.

where,  $I_j$  is the distortion of  $j^{\text{th}}$  cluster.

If data dimension  $N_d > 1$ , then  $\alpha_g$  in (4.21) is expressed as,

$$\alpha_g = \begin{cases} 1 - \frac{3}{4N_d} & , g = 2 \\ \alpha_{g-1} + \frac{1-\alpha_{g-1}}{6} & , g > 2 \end{cases} \quad (4.22)$$

and must satisfy the condition  $0 < \alpha_g \leq 1$ . For a smaller value of  $f(g)$ , the data is more concentrated. Hence, the value of “g” that yields smaller value of  $f(g)$  is regarded as optimal.

### 4.3.5 Combined cumulant and sequence operation theory

The hybridization of cumulant method and sequence operation theory referred to as combined cumulant and sequence operation theory (CCSOT) separately deals with the discrete, Gaussian and non-Gaussian input random variables to achieve the desired higher accuracy with the reduced computational burden. The overall procedure to obtain the probability distribution of a result variable using CCSOT is as follows:

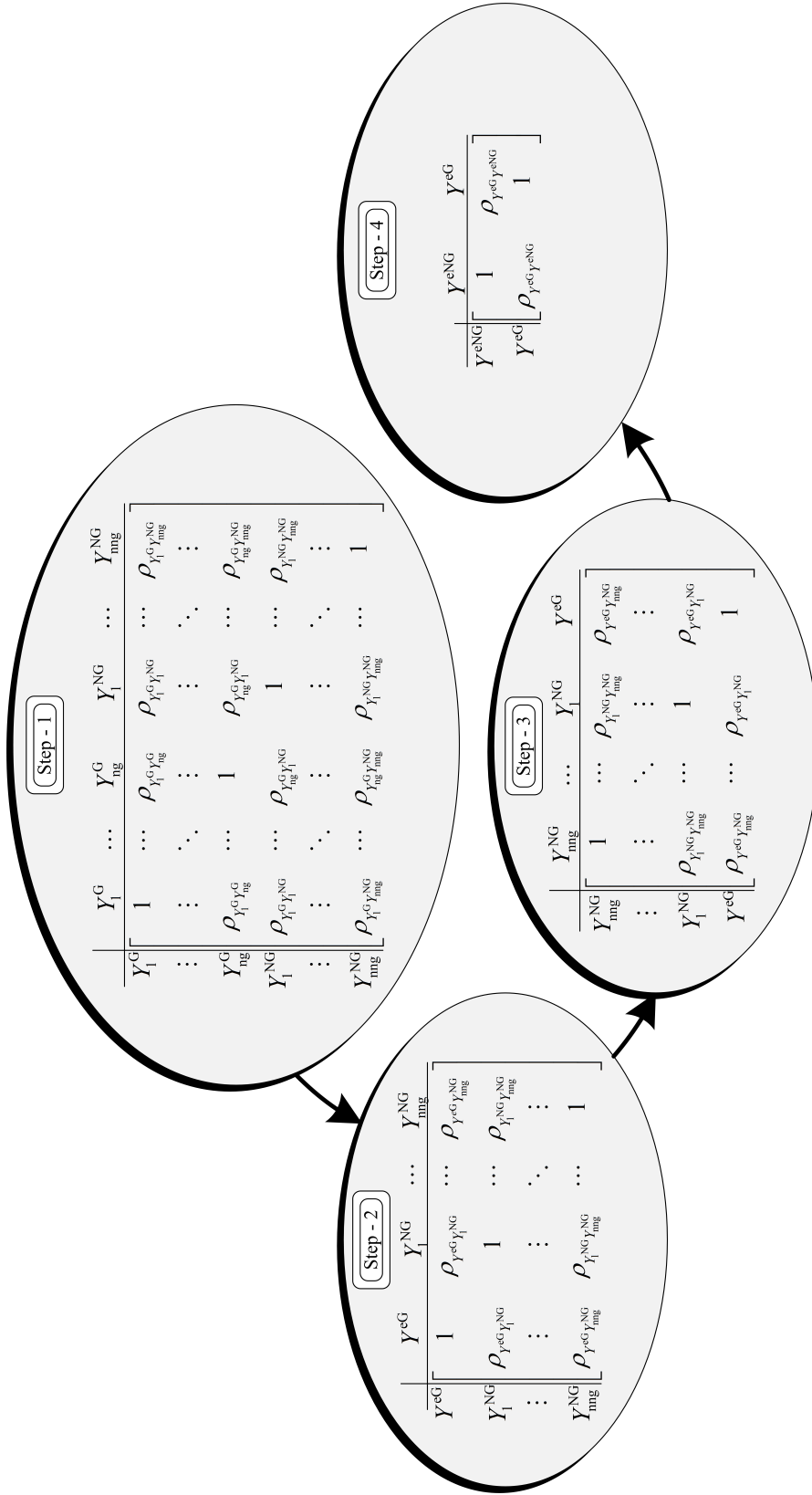
- (i) Discretization: The sensitivity coefficients and probability distributions of input random variables other than Gaussian are discretized.
  - (a) A sensitivity coefficient “sc” is converted to a probability sequence of interval  $\Delta S_1$ . Sequence length “N” is obtained as  $N = \text{sc}/\Delta S_1$ . The probability value at the  $N^{\text{th}}$  discrete instant is unity whereas, at the remaining “N” instants, its value is zero.
  - (b) A discrete distribution of load power with a maximum value of “ $L_M$ ” is converted to a new probability sequence of interval  $\Delta S_2$ . Sequence length “N” is obtained as  $N = L_M/\Delta S_2$ . The load values  $L_i, i = 1, 2, \dots, L_M$  correspond to the instants  $L_i/\Delta S_2$  in the new probability sequence. At the remaining “ $N + 1 - M$ ” instants, the probability value is zero.
  - (c) For a continuous non-Gaussian input random variable having a known empirical marginal cumulative distribution function (CDF) with the cumulative probability values estimated at an interval of  $\Delta S_2$ ,  $N = E/\Delta S_2$ . “N” is the upper bound of the random variable. The probability values at the “N+1” instants can be obtained by the reverse process of Figure 4.4.

- (ii) Calculation of PMCC between  $Y^{eG}$  and  $Y^{eNG}$ : Figure 4.6 indicates various intermediate steps involved in obtaining the PMCC between  $Y^{eG}$  and  $Y^{eNG}$ . In step-1, the given PMCC matrix is rearranged as per the sequence of input random variables as indicated in (4.8). The reduced dimension PMCC matrix in step-2 is obtained from step-1 PMCC matrix using (4.9) to (4.11). Repositioning the elements of step-2 PMCC matrix yields step-3 PMCC matrix. Finally, in step-4, the PMCC between  $Y^{eG}$  and  $Y^{eNG}$  is obtained using (4.9) to (4.11).
- (iii) Probability sequences pertaining to discrete and non-Gaussian components of (4.7): The product terms of discrete as well as non-Gaussian components of (4.7) are evaluated using sequence multiplication operation yields probability sequence of interval  $\Delta S = \Delta S_1 \Delta S_2$ . Finally, the addition type convolution or subtraction type convolution (depending on the sign of sensitivity coefficients) among the probability sequences obtained using sequence multiplication operation yield  $Y^{eD}$  and  $Y^{eNG}$ . For the evaluation of addition type convolution/subtraction type convolution operation, Gaussian copula density function is used. It extracts the dependency structure from correlated standard Gaussian distributions and has the benefit that, the marginal distributions can be arbitrary. Copula helps in modeling the marginal distributions and dependence structure separately. Gaussian copula density function is expressed as,

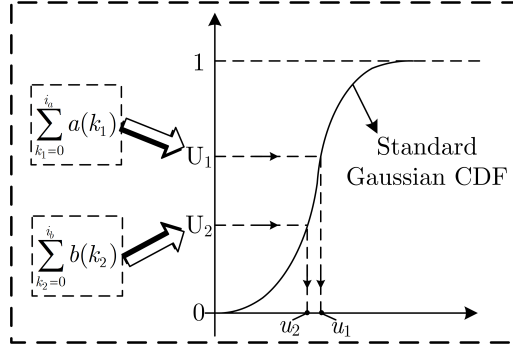
$$\text{cou}(U_1, U_2, \rho_{ab}) = \frac{1}{\sqrt{1 - \rho_{ab}^2}} \cdot e^{-\frac{1}{2} \left( \frac{\rho_{ab}^2 u_1^2 - 2\rho_{ab} u_1 u_2 + \rho_{ab}^2 u_2^2}{1 - \rho_{ab}^2} \right)} \quad (4.23)$$

where  $u_1$  and  $u_2$  are obtained from the inverse cumulative probability transformation [refer to Figure 4.7]; the values of “ $U_1$ ” and “ $U_2$ ” lies in the range  $[0,1]$ .

- (iv) Determination of probability sequence pertaining to Gaussian component of (4.7): Linear combination of correlated Gaussian random variables is also a Gaussian random variable and its probability distribution can be described by first two cumulants (Villanueva et al., 2014). The first two cumulants  $C_1$  and  $C_2$  of Gaussian component of (4.7) is calculated using the formulations as discussed in section 4.3.1. Its cumulative probability values are obtained in the range  $[C_1 - 4\sqrt{C_2}, C_1 + 4\sqrt{C_2}]$  at intervals of  $\Delta S$ . Finally, the probability sequence is recovered using reverse process of Figure 4.4.
- (v) Distribution of result variable: The probability sequences pertaining to discrete,



**Figure 4.6:** Sequence of steps to obtain PMCC between  $Y^{\text{eG}}$  and  $Y^{\text{eNG}}$ .



**Figure 4.7:** Schematic representation to obtain  $u_1$  and  $u_2$  of (4.23).

Gaussian, non-Gaussian components of (4.7) have same sequence interval of  $\Delta S$ . The PMCC values between  $Y^{eD}$  and  $Y^{eG}$  as well as between  $Y^{eD}$  and  $Y^{eNG}$  are zero whereas, between  $Y^{eG}$  and  $Y^{eNG}$  is obtained using step (iii). Using the obtained probability sequences of  $Y^{eD}$ ,  $Y^{eG}$  and  $Y^{eNG}$  and the associated correlations, the probability sequences of the result variables can be obtained using addition type convolution or subtraction type convolution. Finally, the cumulative probability plots can be obtained.

## 4.4 Case study and discussion of results

The performance of the proposed hybrid methods is verified on various test systems such as New England 39-bus test system (Frank et al., 2016), IEEE 14, 57, 118-bus test systems (UWEE, 2015) and Indian utility 62-bus system (PWTEB, 2017) by performing PLF and TPLF using sensitivity matrix based power system models as described in Figure 4.1. The programming codes are developed using MATLAB 7.10 and are executed on a computer with *i7* processor and RAM size of 8 GB.

PLF is carried out to accomplish the following investigations.

- (i) Accuracy comparison of results using traditional PLF model and modified PLF model.
- (ii) Investigations on the accuracy of proposed hybrid methods considering input correlations.
- (iii) Performance assessment of the proposed hybrid methods in approximating multimodal probability distributions of result variables.

Using TPLF model, the following investigation is carried out.

- (i) The effectiveness of the proposed spatiotemporal model in capturing both spatial and temporal correlations among the input random variables of TPLF is examined. Multivariate-CCGMA is applied for the approximation of the probability distributions of result variables.

For various result analysis, PV arrays are considered to be installed at certain buses of the above test systems. The PV arrays inject only real power to the system, and the reactive power generations of the PV arrays are assumed zero.

Since the motivation behind proposing hybrid methods is to approximate the multimodal probability distributions of result variables accurately, a mixture of discrete, Gaussian and non-Gaussian input random variables of different statistical parameters are considered. The statistical parameters and correlation details of the input random variables in case of all the test systems are described in Table 4.2 through 4.6. In Table 4.2, the PV array installation buses are randomly chosen. The assumed PMCC values for IEEE 14-bus, 57-bus and 118-bus systems in Table 4.5 is adopted from (Fan et al., 2014). The probability mass functions of real power generation of conventional generators in Table 4.6 is assumed to follow Bernoulli distribution.

#### 4.4.1 Measures for accuracy comparison

Accuracy of the proposed methods is ascertained by calculating two types of errors such as average root mean square error (ARMSE) and absolute percentage error. The ARMSE is calculated as,

$$\frac{\sqrt{\sum_{i=1}^{N_p} (\text{CDF}_{\text{MCS}_i} - \text{CDF}_{\text{COM}_i})^2}}{N_p} \quad (4.24)$$

where  $\text{CDF}_{\text{MCS}_i}$  and  $\text{CDF}_{\text{COM}_i}$  respectively are the  $i^{\text{th}}$  values on the CDF plots using Monte-Carlo simulation and the comparing method;  $N_p$  is the selected number of points in the CDF plots whose value is set to 100 in this thesis.

Absolute percentage error in standard deviation ( $\sigma$ ) is calculated as,

$$e_\sigma = \left| \frac{\sigma_{X,\text{MCS}} - \sigma_{X,\text{COM}}}{\sigma_{X,\text{MCS}}} \right| \times 100 \quad (4.25)$$

where  $\sigma_{X,\text{MCS}}$ , and  $\sigma_{X,\text{COM}}$  respectively are the standard deviation values of  $X$  using MCS and the method under comparison.

**Table 4.2:** PV generation locations and statistical descriptions.

Parameters	New England 39-bus	IEEE 14-bus	IEEE 57-bus	IEEE 118-bus	Indian utility 62-bus
Bus no.	26 27 28	13 14 16	17	19 77 92 103 105	18 38 41 53 62
No. of arrays	1 1 1	2 2 2	2	1 1 1 1 1	1 1 1 1 1
Distribution	Non-parametric	Beta (CV=30%)	Beta (CV=46.72%)	Non-parametric	Non-parametric
Penetration level	5%	20% (Array-1), 25% (Array-2)	75%	10%	
Penetration based on	System load power	Local bus load power	System load power	System load power	

Note: CV: Coefficient of variation.

**Table 4.3:** Statistical description of continuous load power distributions.

Parameters	New England 39-bus	IEEE 14-bus	IEEE 57-bus	IEEE 118-bus	Indian utility 62-bus
Bus no.	21,23,24,25,26,27,28,29	5,13,14	13,14,15,16,17	19,77,92,103,104,105,106,107	18,38,41,53,62
Distribution	Gaussian	Gaussian (CV=7%)	Gaussian (CV=5%)	Gaussian (CV=10%)	Gaussian (CV=10%)

Note: CV: Coefficient of variation.

**Table 4.4:** Statistical description of discrete load power distributions.

Parameters	New England 39-bus	IEEE 14-bus	IEEE 57-bus	IEEE 118-bus											
Bus no.	9, 18	9	47	47											
Real power	0.134	0.196	0.302	0.348	0.373	0.134	0.196	0.308	0.349	0.373	0.23	0.24	0.36	0.38	0.39
Reactive power	Refer Table 3.2	0.11	0.17	0.196	0.21	0.07	0.09	0.12	0.15	0.11	-	-	-	-	-
Probability		0.1	0.15	0.3	0.25	0.2	0.1	0.15	0.25	0.2	0.1	0.15	0.3	0.25	0.2

**Table 4.5:** Details of base-case correlations.

Parameters	New England 39-bus	IEEE 14-bus	IEEE 57-bus	IEEE 118-bus	Indian utility 62-bus
Among PV generations	Obtained from data	0.7	0.7	0.5	Obtained from data
PV generation and local real load power	0.3	0.3	0.3	0.3	0.3
Among real load powers	Obtained from data	0.2	0.2	0.5	Obtained from data
Among reactive load powers	Obtained from data	0.2	0.2	0.5	Obtained from data
Local real and reactive load powers	Obtained from data	0.5	0.5	1	Obtained from data
PV generation and ambient temperature	Obtained from data	-	-	-	Obtained from data



**Table 4.6:** Details of discrete power generations.

Parameters	IEEE 14-bus		IEEE 57-bus					
	Bus 2		Bus 3		Bus 12			
Unit no.	1	2	1	2	1	2	3	4
Capacity (pu)	0.22	0.22	0.22	0.22	0.84	0.84	0.84	0.84
Forced outage rate	0.1	0.08	0.1	0.08	0.09	0.09	0.07	0.07

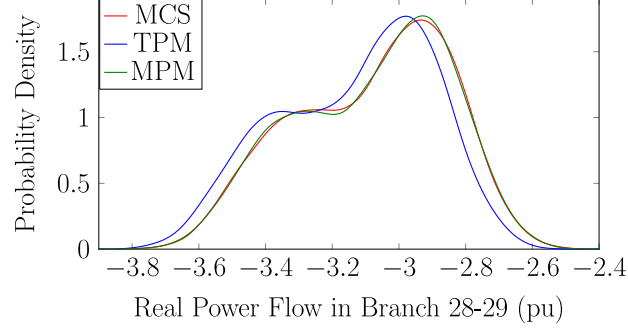
Aggregate error is calculated by summing up the values of error for all the result variables. The average error is obtained as the ratio of aggregate error and the total number of result variables.

#### 4.4.2 Accuracy comparison of probabilistic load flow results using traditional and modified power system models

The competency of modified PLF model in approximating probability distributions of result variables using CCSOT as compared to the traditional model is verified in Table 4.7. The PV-integrated New England 39-bus system is used for the result analysis. The average absolute percentage error in the standard deviation of result variables using both the models are calculated by considering Monte-Carlo simulation results as reference. It is observed that PLF results so obtained using traditional model are more erroneous as compared to the modified PLF model. Among the result variables, the error in the bus voltage magnitude is the highest as traditional model treats P |V| buses as PQ type leading to erroneous estimation of sensitivity coefficients. The probability distributions of  $P_{L,28-29}$  using both the models are compared in Figure 4.8 as an example case. The higher accuracy of modified model as compared to the traditional one is evident.

#### 4.4.3 Investigation on the accuracy of proposed hybrid methods considering input correlations

This section investigates the capability of the proposed hybrid methods in including input correlations in the analysis. At first, the accuracy of Gaussian copula based sequence operations used in CCSOT is investigated by considering different distribution types and correlations as defined in Table 4.8 for various cases of arbitrarily



Note: MCS: Monte-Carlo simulation, TPM: Traditional PLF model, MPM: Modified PLF model.

**Figure 4.8:** Comparison of probability density plots of  $P_{L,28-29}$  using various PLF models.

**Table 4.7:** Accuracy comparison of PLF results using traditional and modified model.

PLF result variables		Average $e_\sigma$	
		Traditional model	Modified model
$\bar{x}$	$\delta$	01.62	00.72
	$ V $	778.02	25.77
$\bar{y}'$	$Q$	-	18.07
$\bar{z}$	$P_L$	25.23	03.21
	$Q_L$	110.12	11.21
$\bar{y}''$	$P_{31}$	00.84	00.25
	$Q_{31}$	182.72	16.79
$\bar{z}'$	$P_{\text{Loss}}$	29.87	07.41
	$Q_{\text{Loss}}$	103.38	09.53

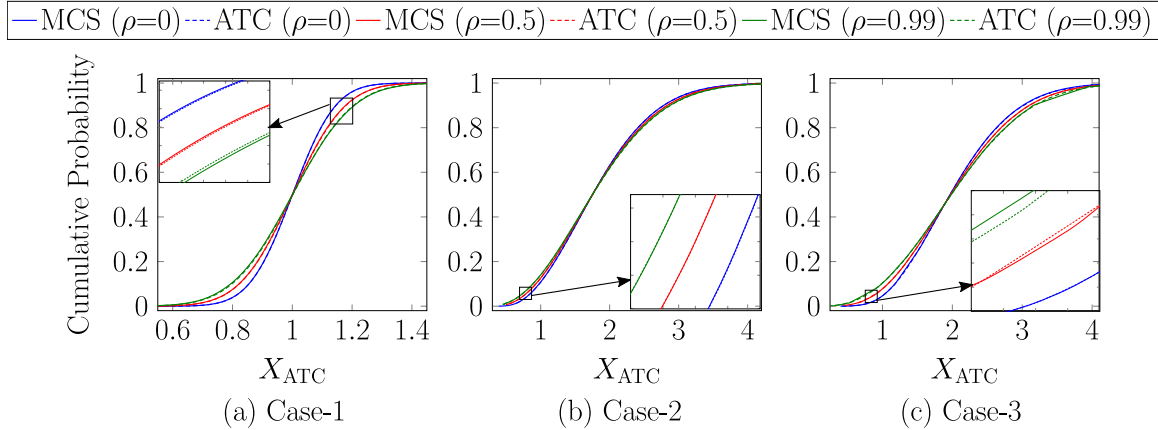
chosen input random variables denoted as  $X_1$  and  $X_2$ .

**Table 4.8:** Different distribution types of input random variables for the accuracy comparison of Gaussian copula based sequence operations.

Case	Inputs	Distribution type	Distribution parameters
1	$X_1$ $X_2$	Gaussian	$\mu = 0.5, \sigma = 0.08$
2	$X_1$ $X_2$	Gaussian Weibull	$\mu = 0.5, \sigma = 0.08$ shape parameter=1.5, scale parameter=2
3	$X_1$ $X_2$	Weibull Beta	shape parameter=1.5, scale parameter=2 shape parameters of 5 and 2
4	$X_1$ $PV_1$ after preprocessing at noon $X_2$ $PV_2$ after preprocessing at noon		Non-parametric

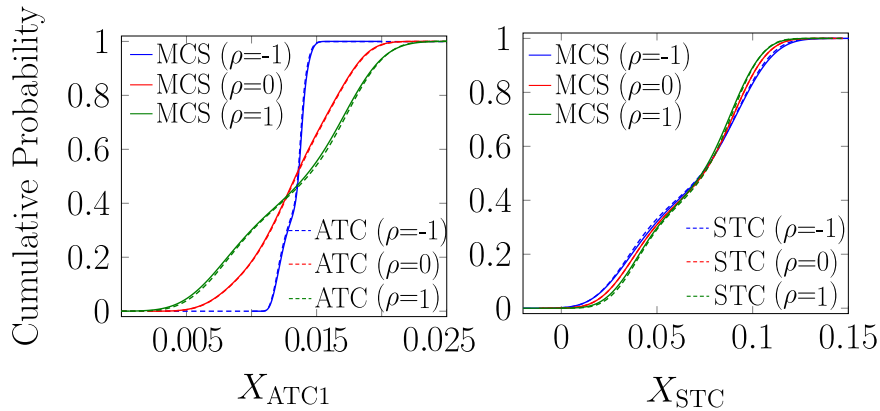
The accuracy of addition type convolution operation in handling different PMCCs is shown in Figure 4.9 considering the first three cases of Table 4.8. The results of an addition type convolution operated sequence  $X_{ATC} = X_1 + X_2$  so obtained for the first three cases are compared with that of Monte-Carlo simulation to ascertain their accuracy. In order to further investigate the accuracy of addition type convolution and subtraction type convolution operations in incorporating both positive and negative values of correlation between two non-parametric multimodal probability distributions, case-4 in Table 4.8 is taken into consideration. The results for addition type convolution operated sequence  $X_{ATC1} = X_1 + X_2$  and subtraction type convolution operated sequence  $X_{STC} = 10X_1 - X_2$  are compared with that of Monte-Carlo simulation as shown in Figure 4.10. In both the figures, the results of the sequence operations closely follow the Monte-Carlo simulation plots indicating their accuracy in handling correlated input random variables of any distributions.

Further, the capability of both the hybrid methods in handling correlated input random variables precisely in PLF is evaluated by considering IEEE 118-bus test system. In case of CCGMA, cluster distortion function based approach estimates the value of “g” as 4. For CCSOT the values of  $\Delta S_1$  and  $\Delta S_2$  are taken as 0.01. The cumulative probability plots of  $P_{L,100-103}$  using univariate-CCGMA and CCSOT are compared with Monte-Carlo simulation in Figure 4.11 with and without considering base-case correlation. It is evident that the plots obtained using both the hybrid methods are closer to Monte-Carlo simulation. Various other PMCCs are also con-



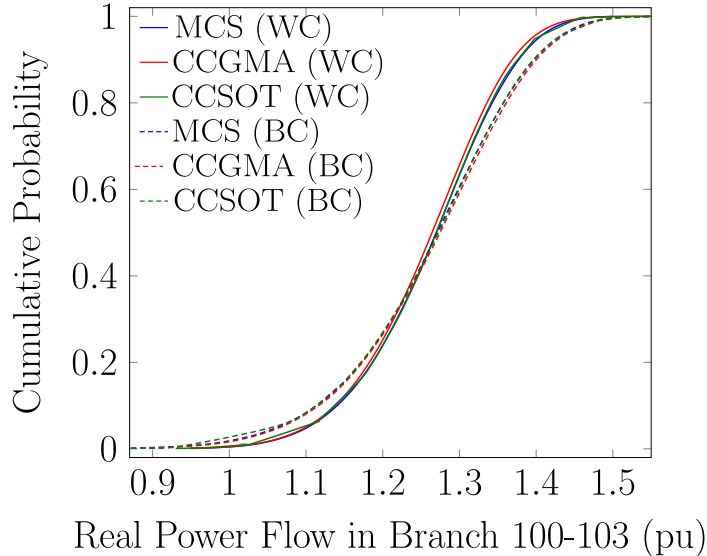
Note: MCS: Monte-Carlo simulation, ATC: Addition type convolution.

**Figure 4.9:** Accuracy comparison of Gaussian Copula based addition type convolution operation to handle different parametric distributions and input correlations.



Note: MCS: Monte-Carlo simulation, ATC: Addition type convolution, STC: Subtraction type convolution.

**Figure 4.10:** Accuracy comparison of Gaussian Copula based addition type convolution and subtraction type convolution operations to handle non-parametric distributions and input correlations.



Note: MCS: Monte-Carlo simulation, WC: Without correlation, BC: Base-case correlation.

**Figure 4.11:** Comparison of cumulative probability plots of  $P_{L,100-103}$  with and without considering input correlation.

sidered, and it is found that for any value of PMCC, in case of all the result variables, plots obtained using CCGMA, CCSOT, and MCS are nearly the same. This indicates that the proposed hybrid methods are capable of handling input correlations satisfactorily. Due to the consideration of base-case correlation, the standard deviation of  $P_{L,100-103}$  is increased.

Finally, the effect of multiple input correlations, i.e. among the PV generations, load powers and PV generations with their local loads on the probability distributions of result variables is studied by considering three correlation conditions as defined in Table 4.9. The analysis is carried out on New England 39-bus, IEEE 14 and 118-bus test systems by performing PLF and TPLF simulations. Multivariate-CCGMA is applied to obtain the probability distribution plots of the result variables. The effect of various values of PMCC on the first two statistical moments pertaining to each correlation type is compared in Table 4.10 for a few cases. It is to note that the impact of input correlation has a noticeable effect on the standard deviation of the result variables and the effect is more prominent in the tail regions of the distributions. Depending on the sign of the associated sensitivity coefficients, the effect of input correlation either elongates or shortens the tail region of the distributions. It can be noticed from Table 4.10 that an increase in PMCC value decreases the standard

deviation values of  $P_{L,103-105}$  and  $Q_{L,25-26}$ . However, the reverse is noticed in the remaining cases.

**Table 4.9:** Three different input correlation conditions.

Correlation type	C-1	C-2	C-3
PV-PV	0, 0.5, 1	BC	BC
PV-Local load	BC	0, 0.5, 1	BC
Load Load	BC	BC	0, 0.5, 1

Note: BC: Base-case correlation.

From the above analysis of results, it is clear that correlation does not have any effect on the mean value and has an effect on standard deviation [refer Table 4.10]. A fixed change in the value of PMCC results in nearly a fixed change in the value of standard deviation. The reason for such observations is due to the linearization of non-linear load flow equations. In the sensitivity matrix based power system models, result variables are linearly related with the input random variables. The correlation between the input random variables does not affect the expected value of result variables whereas; the variance of result variables is a linear function of input correlation [refer Appendix F]. In case of non-linear models, input correlation affects the expected values of result variables and also the higher order moments of result variables could be affected even more. Refer (F.3) and (F.4) of Appendix F.

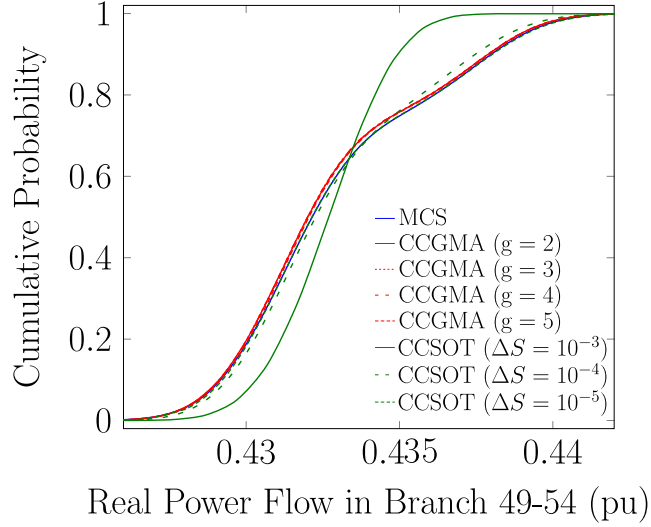
#### 4.4.4 Performance assessment of the proposed hybrid methods in approximating multimodal probability distribution of result variables

Solution accuracy and computational efficiency are the two performance criteria considered. The performance of the proposed hybrid methods is investigated on IEEE 14, 57 and 118-bus test systems considering the statistics of input random variables as defined in Table 4.2 through 4.4 and Table 4.6 as well as the base-case correlations as in Table 4.5. The results are compared with that of ECM using series expansion methods, dependent discrete convolution, and Monte-Carlo simulation. First six cumulants of result variables are considered while applying a series expansion based ECM.

The accuracy and efficiency of CCGMA are related to “g”, and those of CCSOT

**Table 4.10:** Effect of various input correlation conditions on the statistics of few result variables.

Test system/model type	Correlation type	Result variable	PMCC	$\mu$	$\sigma$
New England 39-bus/TPLF	C-1	$\delta_{27}$	0	-0.1948	0.0233
			0.5	-0.1948	0.0251
			1	-0.1948	0.0269
IEEE 118-bus/PLF	C-1	$P_{L, 103-105}$	0	0.3473	0.0298
			0.5	0.3473	0.0278
			1	0.3473	0.0257
New England 39-bus/TPLF	C-2	$Q_{L, 25-26}$	0	-0.1668	0.0068
			0.5	-0.1668	0.0065
			1	-0.1668	0.0062
IEEE 14-bus/PLF	C-3	$ V_{13} $	0	1.0548	0.00114
			0.5	1.0548	0.00120
			1	1.0548	0.00126
IEEE 118-bus/PLF	C-3	$P_{L, 105-106}$	0	0.1145	0.0287
			0.5	0.1145	0.0292
			1	0.1145	0.0297



Note: MCS: Monte-Carlo simulation.

**Figure 4.12:** Comparison of cumulative probability plots of  $P_{L,49-54}$ .

are related to  $\Delta S$ . In the previous section, a cluster distortion function based approach was applied for determining the value of “g”, and the value of  $\Delta S$  was set to 0.0001. In order to examine their performance, four different values of “g” such as 2, 3, 4, 5 and three different values of  $\Delta S$  such as 0.001 ( $\Delta S_1 = 0.1$ ,  $\Delta S_2 = 0.01$ ), 0.0001 ( $\Delta S_1 = 0.01$ ,  $\Delta S_2 = 0.01$ ) and 0.00001 ( $\Delta S_1 = 0.001$ ,  $\Delta S_2 = 0.01$ ) are now considered. The cumulative probability plots for  $P_{L,49-54}$  with various values of “g” and  $\Delta S$  are compared in Figure 4.12. The CCSOT result with  $\Delta S = 0.00001$  is more close to Monte-Carlo simulation (ARMSE = 0.0316) followed by CCGMA plots for various “g”. CCSOT result with  $\Delta S = 0.001$  approximates the cumulative probability plot with maximum error i.e., ARMSE = 0.8931. The aggregate and average ARMSE for the considered “g” and  $\Delta S$  using univariate-CCGMA, multivariate-CCGMA and CCSOT are compared in Table 4.11 for IEEE 118-bus system (558 result variables). Unlike, CCSOT where the reduction in  $\Delta S$  reduces the ARMSE value, the increase in the value of “g” does not yield reduced ARMSE in case of CCGMA. The peculiar observation might be because of the increase in the value of “g” yields data overfitting.

The average absolute percentage errors in standard deviation using CCGMA and CCSOT for all the 558 result variables are compared with combined cumulant and Gram-Charlier method (CCGCM), combined cumulant and Cornish-Fisher method (CCCFM), dependent discrete convolution and Monte-Carlo simulation in Table 4.12. The first six cumulants of the result variables are calculated in case of CCGCM and



**Table 4.11:** Comparison of ARMSEs of the proposed hybrid methods considering input correlations.

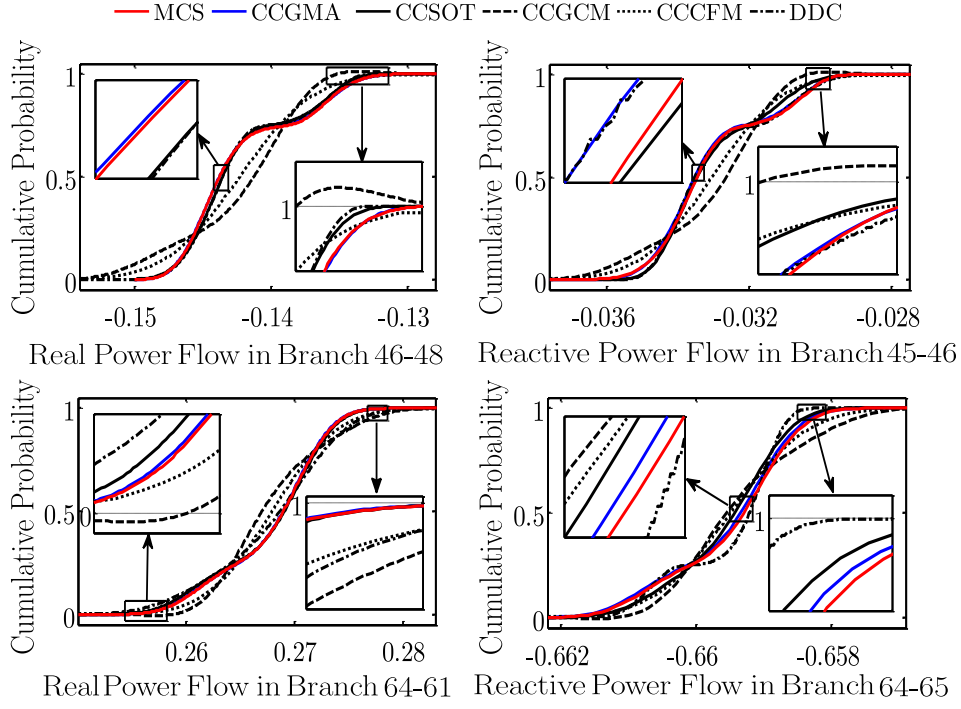
PLF method	Aggregate ARMSE	Average ARMSE
Univariate-CCGMA	$g = 2, N_r = 160$	78.2696
	$g = 3, N_r = 1215$	77.9332
	$g = 4, N_r = 5120$	77.4433
	$g = 5, N_r = 15625$	79.1356
Multivariate-CCGMA	$g = 2, N_r = 10$	81.7975
	$g = 3, N_r = 15$	75.9182
	$g = 4, N_r = 20$	71.4296
	$g = 5, N_r = 25$	80.6712
CCSOT	$\Delta S = 0.001$	425.9772
	$\Delta S = 0.0001$	70.5312
	$\Delta S = 0.00001$	59.4270

CCCFM. The value of  $\Delta S$  is set to 0.0001 in case of dependent discrete convolution. The value of “g” is obtained using cluster distortion function based approach.

**Table 4.12:** Comparison of absolute percentage errors and simulation times for various PLF methods considering base-case correlations.

PLF method	IEEE 14-bus		IEEE 57-bus		IEEE 118-bus	
	Average $e_\sigma$	Time (sec.)	Average $e_\sigma$	Time (sec.)	Average $e_\sigma$	Time (sec.)
Univariate-CCGMA	2.58 ( $N_r = 1600$ )	8.15	4.21 ( $N_r = 25600$ )	308.85	3.96 ( $N_r = 5100$ )	398.88
Multivariate-CCGMA	2.47 ( $N_r = 200$ )	1.02	3.64 ( $N_r = 4800$ )	48.64	3.73 ( $N_r = 20$ )	16.38
CCSOT	2.38	24.37	3.59	73.43	3.63	117.81
CCGCM	2.64	5.72	4.43	55.86	4.30	60.75
CCCFM	2.50	6.88	3.78	57.14	3.78	79.30
Dependent discrete convolution	2.30	37.69	3.48	196.84	3.53	974.55
Monte-Carlo simulation	-	20.06	-	100.30	-	1842.15

Cumulative probability plots for  $P_{L,46-48}$ ,  $P_{L,64-61}$ ,  $Q_{L,45-46}$ , and  $Q_{L,64-65}$  considering base-case correlations using all the methods are compared in Figure 4.13. In all the cases, CCGMA, CCSOT and dependent discrete convolution plots are closer to Monte-Carlo simulation plots as compared to CCGCM and CCCFM plots. CCGCM and CCCFM are found incapable of approximating multimodal distributions of the result variables. It is to further notice that, the upper tails of cumulative probabil-



**Figure 4.13:** Comparison of cumulative probability plots of  $P_{L,46-48}$ ,  $P_{L,64-61}$ ,  $Q_{L,45-46}$  and  $Q_{L,64-65}$ .

ity plots pertaining to  $P_{L,46-48}$ , and  $Q_{L,45-46}$  using CCGCM exceed unity whereas; lower tail of  $P_{L,64-61}$  goes below the zero value. Similar observations are also noticed in case of few other result variables. The reduced accuracy of the series expansion methods is because they provide an approximation around the Gaussian distribution and accuracy mainly depends on how much the distribution is close to the Gaussian distribution. The total number of input random variables being more, the simulation time for dependent discrete convolution exceeds Monte-Carlo simulation in case of IEEE 14-bus and 57-bus systems. It is to note that, the simulation time in case of multivariate-CCGMA is least as compared to other methods, the average  $e_\sigma$  in case of CCSOT is the lowest compared to all the methods. It is further to mention that, for approximating unimodal probability distributions, CCGCM and CCCFM are the better choices.

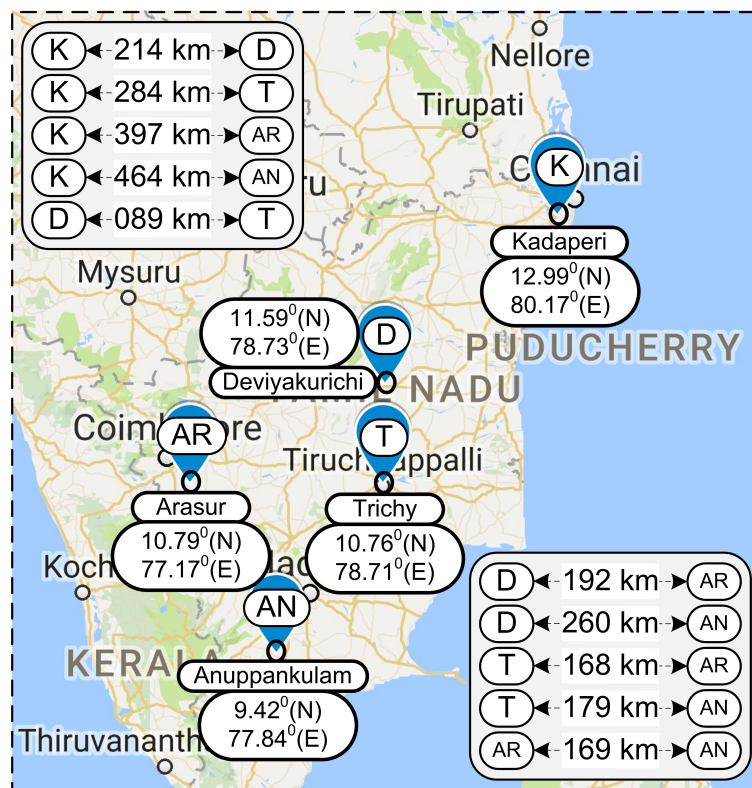


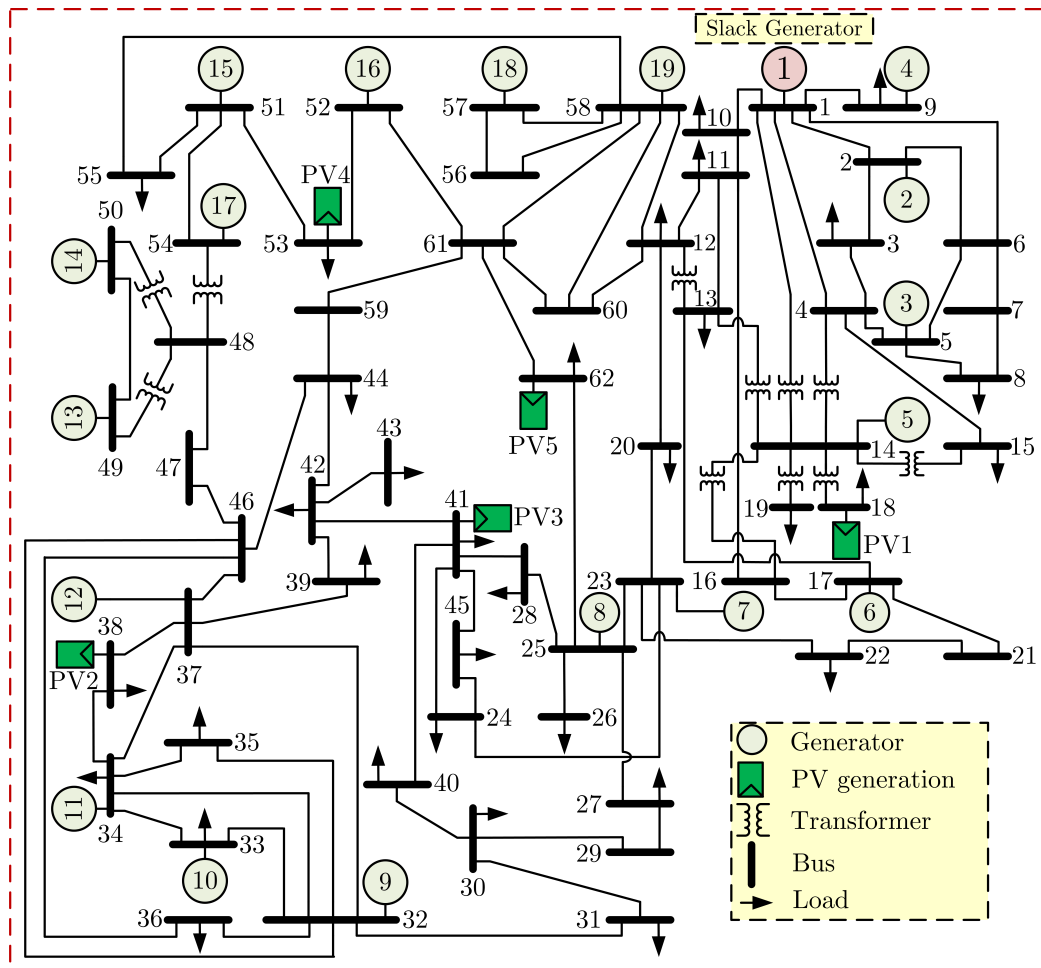
Figure 4.14: PV array installation sites in Tamilnadu.

#### 4.4.5 Spatiotemporal model based temperature-augmented probabilistic load flow simulations

In this section, TPLF is carried out on modified Indian utility 62-bus system by using the spatiotemporal model as discussed in section 2.3. The modification is the integration of five PV arrays, geographical locations of which are as shown in Figure 4.14.

For TPLF simulations the system base power and base temperature are set to 100 MVA and 100 °C respectively. CCGMA method is applied for approximating the multimodal probability distributions of result variables. For the given values of transmission branch MVA ratings, temperature-augmented load flow has convergence problem. Since Newton-Raphson method does not guarantee convergence, if the starting values are far from the true solution, load flow algorithm is initialized using warm start, i.e. started from the solution to a conventional load flow solution to facilitate algorithm stability and speed convergence. Hourly irradiance, ambient temperature, and wind speed data are collected for the past 15 years (2000-2014). The chosen sites for PV array installation corresponds to the bus numbers 18, 38, 41, 53 and 62 of Indian utility 62-bus system. The single line diagram of the system under study is shown in Figure 4.15. Taking irradiance data (direct normal irradiance  $r_{DN}$ , diffuse horizontal irradiance  $r_{DH}$ , and global horizontal irradiance  $r_{GH}$ ), ambient temperature  $T_{Amb}$ , and wind speed  $v_W$  of the past 15 years, the PV generation time series for the considered period is obtained using a set of steps as discussed in Figure 4.16.

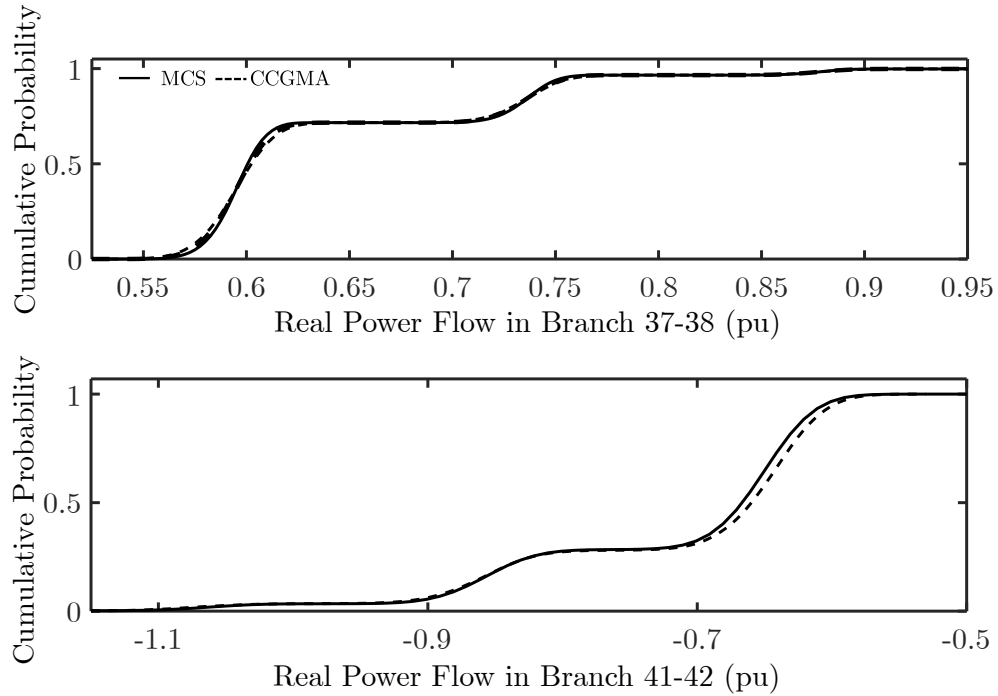
The ambient temperature of all the transmission branches are presumed constant and corresponds to the location Trichy. Now, the TPLF inputs include the obtained PV generation time series using Figure 4.16, ambient temperature, and load real and reactive powers. The above inputs form a multivariate structure and a set of steps as discussed in section 2.3 is applied for the development of a spatiotemporal model to forecast their future time series. Autoregressive conditionally heteroscedastic (ARCH) model built with orders 4 to 8 fits all the principal components. The developed ARCH models for principal components is used for the generation of future time series. Finally, the obtained scenarios obey all the characteristics of the historical observations of inputs and their corresponding probability distributions are approximated using a non-parametric density estimation method. The cumulative



**Figure 4.15:** Single line diagram of PV integrated Indian utility 62-bus system.

Specified data			Measured data	
Local coordinates and time	Array details	Standard test conditions	Basic solar irradiance	Environmental data
$\theta_{LA}, \theta_{LO}, \theta_{LO}^{STM}, t_{LC}, d_N$	$A, \theta_T, \gamma$	$T_{Ref}^P, \beta_{Ref}, \eta_{Ref}$	$r_{DN}, r_{DH}, r_{GH}$	$T_{Amb}, v_W$
Calculation steps		Variables/Formulations		
① Basic solar position angles	$\theta_D, \theta_H$	Equations (2.6), (2.10)		
② Derived solar position angles	$\theta_S, \gamma_S$	Equations (2.4), (2.5)		
③ Solar incident angle	$\theta_I$	Equation (2.12)		
④ Derived solar irradiance at titled PV array	$r_{DT} = r_{DN} \cos \theta_I$ $r_{RT} = 0.1r_{GH}(1 - \cos \theta_T)$ $r_{DIT} = 0.5[r_{DH}(1 + \cos \theta_T) - r_{GH}(0.012\theta_S + 0.0212)(1 - \cos \theta_T)]$			
⑤ Total broadband irradiance at titled PV array	$r_T = r_{DT} + r_{RT} + r_{DIT}$			
⑥ PV cell temperature	$T_C = r_T e^{-2.81 - 0.0455v_W} + T_{Amb}$			
⑦ Temperature dependent electrical efficiency	$\eta_{PV} = \eta_{Ref}[1 - \beta_{Ref}(T_C - T_{Ref}^P)]$			
⑧ PV generation output	$P_{PV} = r_T A \eta_{PV}$			

**Figure 4.16:** A set of steps to obtain PV generation time series using the time series of irradiance, ambient temperature, and wind speed.



**Figure 4.17:** Comparison of cumulative probability plots (i)  $P_{L,37-38}$  and (ii)  $P_{L,41-42}$ .

probability plots of a few result variables (real power flow in the branches 37-38 and 41-42) for the system under consideration using CCGMA is compared with that of Monte-Carlo simulation [refer Figure 4.17]. Both the plots are nearly the same. The similar observations are also noted in case of remaining result variables indicating the accuracy of the CCGMA used for TPLF.

## 4.5 Conclusions

This chapter has presented the PSSA of power systems using the proposed hybrid methods such as CCGMA and CCSOT. Firstly, the traditional PLF model is modified, and the resulting reduced order model is further upgraded by accounting for the electro-thermal coupling effect of transmission branches. With regard to the accurate approximation of multimodal probability distributions of result variables, the performance of the proposed hybrid methods were verified as applied to various test systems. In summary, the following observations from the result analysis are worth noting.

- (i) The accuracy of modified PLF model is highly appreciable in comparison to

traditional model in the studied system.

- (ii) The direct calculation of probability distributions of branch temperatures using sensitivity matrices without solving the thermal balance differential equation is the main advantage of the TPLF. Further, it is free from the requirement of the value of the parameters describing the thermal characteristics of branch conductors.
- (iii) The proposed hybrid methods effectively incorporates the input correlations and thereby accurately approximates the multi-modal probability distributions of result variables.
- (iv) The assumption of the same value of PMCC between the components of two different Gaussian mixtures as that between the Gaussian mixtures in univariate-CCGMA degrades its accuracy as compared to multivariate-CCGMA.
- (v) For sensitivity matrix-based linear models, input correlation does affect the second and higher order moments of the result variables whereas the mean value is unaffected.



## Publications based on the chapter

### Refereed journal publications

1. Prusty, B. R. and Jena D. (2019). A Spatiotemporal probabilistic model-based temperature-augmented probabilistic load flow considering PV generations. *International Transactions on Electrical Energy Systems*, DOI: 10.1002/2050-7038. 2819 (Wiley)
2. Prusty, B. R. and Jena D. (2017). A sensitivity matrix-based temperature-augmented probabilistic load flow study. *IEEE Transactions on Industry Applications*, 53(3):2506-2516. (IEEE)
3. Prusty, B. R. and Jena D. (2016). 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. (CSEE)

### Refereed papers in conference proceedings

1. Prusty, B. R. and Jena D. (2017). A detailed formulation of sensitivity matrices for probabilistic load flow assessment considering electro-thermal coupling effect. In *Asia-Pacific Power and Energy Engineering Conference (APPEEC), 2017 IEEE PES*, pages 1-6. (IEEE)
2. Prusty, B. R. and Jena D. (2016). Estimation of optimal number of components in Gaussian mixture model-based probabilistic load flow study. In *India Conference (INDICON), 2016 IEEE Annual*, pages 1-6. (IEEE)
3. Prusty, B. R. and Jena D. (2016). An efficient hybrid technique for correlated probabilistic load flow study with photovoltaic generations. In *Power Systems Conference (NPSC), 2016 National*, pages 1-6. (IEEE)
4. Prusty, B. R. and Jena D. (2015). Modeling of correlated photovoltaic generations and load demands in probabilistic load flow. In *India Conference (INDICON), 2015 Annual IEEE*, pages 1-6. (IEEE)

## Book series

1. Prusty, B. R. and Jena D. (2018). Probabilistic load flow in a transmission system integrated with photovoltaic generations. In *Applications of Computing, Automation and Wireless Systems in Electrical Engineering* (Springer)

# Chapter 5

## CONCLUSIONS AND FUTURE SCOPES

This chapter summarizes the significant results emanated as an outcome of the research work carried out by providing general conclusions and discussion on the key findings, followed by suggestions for the possible extensions of this research work.

### 5.1 Conclusions

The increased uncertainties in the present era power systems due to the large-scale integration of renewable generations in conjunction with the conventional uncertainties has significantly motivated the researchers towards the adoption of probabilistic methods for various power system studies. Among which, PLF reveals the steady-state characteristics of the power system thereby providing valuable information for power system planning and operation. The primary goal of this thesis was to propose various models and methods for PSSA. Following the effort in accomplishing the set research objectives, the general conclusions that are worth highlighting are as follows.

The second chapter was devoted to the probabilistic modeling of various uncertain inputs such as PV generation, aggregate load power, and ambient temperature. Preprocessing approaches were applied to the daily time step data of aforesaid inputs to filter out the predictable variations in data. It was shown that, in case of PV generation, the applied preprocessing approach of developing a generic multiple linear regression model with regressors possessing a theoretically formulated physical relevance was successful in tracing the periodic changing patterns in data collected at

multi-time instants from various locations. On the other, for load power and ambient temperature, as the periodic patterns are much complicated, extraction of their concealed seasonalities is strenuous. Therefore, the formulation of regressors exhibiting physical relevance is not straightforward. To combat this issue, a methodology to select a reduced but sensible candidate set of frequency components for capturing seasonalities was deliberated. The developed spatiotemporal model using proposed preprocessing along with transformation techniques, principal component analysis and a suitable time series model effectively characterizes the statistical facets of uncertain inputs.

Chapter three performed an over-limit risk assessment on the New England 39-bus power system with large-scale integration of PV generations considering electro-thermal coupling effect of transmission branches. From the results, it was observed that the change in values of TPLF model parameters has a significant impact on the statistics of result variables. In a PV-integrated power system, PSSA applied to a temperature-augmented system model offers an effective way of examining the effects of PV penetration and ambient temperature uncertainties on the steady-state values of the result variables. A TPLF study considering higher PV penetrations lead to a higher variability of branch temperatures, branch power flows and losses as compared to the other result variables. TPLF simulations under various PV penetrations and input correlations aided in identifying the critical buses and branches by quantifying system risk indices. It also assisted in providing theoretical support for the assessment of system reinforcement and reliable operation. The calculated risk indices reflecting the system security level can be used as an indicator for evaluating the power system security.

The fourth chapter proposed sensitivity matrix based power system models and hybrid methods for PSSA. The performance of the above models and methods were verified on New England 39-bus test system, IEEE 14, 57, 118-bus test systems, and Indian utility 62-bus system by performing PLF and TPLF simulations. The accuracy of modified PLF method is highly appreciable in comparison to traditional PLF model in the studied systems. The direct calculation of probability distributions of branch temperatures using sensitivity matrices without the need for solving the thermal balance differential equation is the main advantage of the TPLF. Further, it is free from the requirement of the value of the parameters describing the thermal characteristics of branch conductors. The proposed hybrid methods effectively in-

incorporated the input correlations and also accurately approximated the multi-modal probability distributions of result variables. The assumption of the same value of PMCC between the components of two different Gaussian mixtures as that between the Gaussian mixtures in univariate-CCGMA is found to degrade accuracy as compared to multivariate-CCGMA. For sensitivity matrix-based linear models, input correlation does affect the second and the higher order moments of the result variables whereas the mean value remains unaffected.

## 5.2 Future scopes

Based on the research carried out in this thesis, the recommendations for future work involve the following possible extensions.

- (i) *Selection of the true number of mixture components in GMM based PSSA:* In general, the techniques used for the selection of mixture component number including the one applied in this thesis impose each candidate for the optimal number to fit some criteria for all the possible candidates. Further, the difficulty lies in the proper initialization and selection of the true number of mixture components for multivariate data with clusters being not well separated.
- (ii) *Modeling of non-linear dependency:* In practice, the dependence between various input random variables in any probabilistic analysis is rarely linear and therefore, ignorance of this dependency will cause significant error in the study. Although the proposed scenario-based spatiotemporal model analytically characterizes the correlation effect accurately, fails to capture the actual dependence structure. Capturing such dependencies in a multivariate case with the help of a suitable Copula-based model is worth investigating.
- (iii) *Further improvement of sensitivity matrix based power system model:* The TPLF model so developed includes only the ambient temperatures of temperature dependent branches in the input vector to account for the electro-thermal coupling effect. The inclusion of other influencing factors such as wind speed, solar irradiance, etc., is expected to result in further improvement in the estimated TPLF results.
- (iv) *Storage requirements in PV integrated Power systems:* In this thesis, PSSA is performed under the assumption that, the power system is memory-less, i.e., power system without storage. Determination of storage requirements for the

large-scale integration of PV generations into the power system to manage the emanating uncertainties would also be an exciting area of research.

# Appendix A

## Multiple linear regression model in matrix form

The linear regression is a statistical data analysis technique that captures the extent to which there is a linear relationship between the response variable or regressand and the regressors. The regressors are known as predictors or explanatory variables as they predict or explain the factors that influence the regressand. The estimated regression coefficients can calculate the degree of the impact. The negative coefficient value dictates the inverse relation with the regressand, and the positive value can be said to have a positive influence. A generic multiple linear regression model with “k” regressors is given as,

$$y_i = a_{M,0} + a_{M,1}T_{1,i} + a_{M,2}T_{2,i} + \cdots + a_{M,k}T_{k,i} + \varepsilon_i, i = 1, 2, \cdots, ss \quad (\text{A.1})$$

where  $y$  is the regressand,  $T_1$  through  $T_k$  are the regressors,  $a_{M,0}$  through  $a_{M,k}$  are the regression coefficients and  $\varepsilon$  is the residual of the model.

The model (A.1) is a form of parametric regression where the relationship between regressand and regressors has a predetermined form. It models the conditional mean of the regressand under the assumptions of (i) assigning equal weights to all the data points and (ii) minimization of error sum of squares.

In linear regression, linear refers to the linearity in parameters and not the linearity in regressors, i.e. parameters appear with power one; it is not multiplied or divided by any other parameters. While the equation is linear in parameters, one can transform

the predictor variable in ways that produce curvature. The model is still linear in the parameters even though the predictor variables are non-linear.

The model (A.1) in matrix form is given as,

$$[Y] = [T][a_M] + [\varepsilon] \quad (\text{A.2})$$

$$\text{where } [Y] = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_{ss} \end{pmatrix}_{ss \times 1}, [T] = \begin{pmatrix} 1 & T_{1,1} & \cdots & T_{k,1} \\ 1 & T_{1,2} & \cdots & T_{k,2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & T_{1,ss} & \cdots & T_{k,ss} \end{pmatrix}_{ss \times (k+1)},$$

$$[a_M] = \begin{pmatrix} a_{M,0} \\ a_{M,1} \\ \vdots \\ a_{M,k} \end{pmatrix}_{(k+1) \times 1}, [\varepsilon] = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_{ss} \end{pmatrix}_{ss \times 1}.$$

If  $\hat{a}_M$  is a vector of estimates of  $a_M$ , the estimated multiple linear regression model is given as,

$$\hat{Y} = T\hat{a}_M + e \quad (\text{A.3})$$

where  $e$  denotes the vector of residuals which is computed as,

$$e = \hat{Y} - T\hat{a}_M \quad (\text{A.4})$$

Now the sum of square of residuals as a function of  $\hat{a}_M$  is given as,

$$\begin{aligned} \text{RS}(\hat{a}_M) &= \sum e_i^2 = e^T e = (\hat{Y} - T\hat{a}_M)^T (\hat{Y} - T\hat{a}_M) \\ &= (\hat{Y})^T \hat{Y} - (\hat{Y})^T T\hat{a}_M - (\hat{a}_M)^T T^T \hat{Y} + (\hat{a}_M)^T T^T T\hat{a}_M \end{aligned} \quad (\text{A.5})$$

where the superscript ‘‘T’’ is the transpose operation.

The least square estimator is obtained by minimizing (A.5). Therefore the derivatives are set equal to zero, i.e.  $\frac{\partial(\text{RS})}{\partial \hat{a}_M} = 0$ . This results in

$$- 2T^T \hat{Y} + 2T^T T\hat{a}_M = 0$$



$$\Rightarrow \hat{a}_M = (\mathbf{T}^T \mathbf{T})^{-1} \mathbf{T}^T \hat{Y} \quad (\text{A.6})$$

A solution for  $\hat{a}_M$  can be obtained using (A.6) if the inverse of  $\mathbf{T}^T \mathbf{T}$  exists i.e., the matrix  $\mathbf{T}$  should have rank  $k + 1$ . As the matrix  $\mathbf{T}$  is a  $ss \times (k + 1)$  matrix, this requires that  $ss \geq (k + 1)$ . In practice, it is required that “ $k$ ” is considerably smaller than “ $ss$ ”.

If the matrix  $\mathbf{T}$  has rank “ $k$ ”, it follows that the Hessian matrix  $\frac{\partial^2(\text{RS})}{\partial \hat{a}_M \partial (\hat{a}_M)^T} = 2\mathbf{T}^T \mathbf{T}$  is a positive definite matrix. Hence, (A.6) is indeed the minimum of (A.5).



# Appendix B

## Non-parametric density estimation method

In general, a non-parametric density estimation method is opted when the probability distribution cannot be defined with the help of any parametric distribution form. Non-parametric density estimation based on Kernel function with fixed bandwidth for univariate case is discussed underneath.

For a given data of “N” samples, i.e.  $x = \{x_1, x_2, \dots, x_N\}$ , an estimate of its probability density function  $f(x)$  using Kernel function is given as,

$$\hat{f}(x) = \frac{1}{N} \sum_{i=1}^N \frac{1}{h} \varphi\left(\frac{x - x_i}{h}\right) \quad (\text{B.1})$$

where the smoothing parameter  $h$  is the bandwidth;  $\varphi(\bullet)$  is the Kernel function with choices to include Gaussian window, Parzen window, etc.

A proper choice of bandwidth and Kernel function are the important aspects in non-parametric density estimation. The most popular approach to decide on the value of  $h$  is the adoption of Silvermans thumb rule which minimizes the mean integrated square error and is defined as,

$$h = \left(\frac{4\hat{\sigma}^5}{3N}\right)^{1/5} \quad (\text{B.2})$$

where  $\hat{\sigma}$  is the standard deviation of the data.

Having decided an appropriate bandwidth, the Gaussian Kernel, the most common

choice is defined as,

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}. \quad (\text{B.3})$$

The Parzen window function on the other hand is given as,

$$\varphi(z) = \begin{cases} 1, & |z| \leq 0.5 \\ 0, & |z| > 0.5 \end{cases}. \quad (\text{B.4})$$

# Appendix C

## Principal component analysis: an orthogonal linear transformation

Principal component analysis is a statistical technique that uses orthogonal transformation to convert a set of observations of correlated random variables into an uncorrelated set referred to as the principal components. Each principal component is a univariate time series. Application of principal component analysis to correlated Gaussian data, results in independent principal components. In case of a set of non-Gaussian data, independent principal components are obtainable only if a suitable transformation is used to convert the non-Gaussian data into Gaussian before applying principal component analysis. The application of principal component analysis follows a set of steps as discussed under:

- (i) Step-1: For “nc” correlated random variables each of “ss” samples, define a matrix  $W$  of the form:

$$W = \begin{pmatrix} w_1^1 & w_1^2 & \cdots & w_1^{ss} \\ w_2^1 & w_2^2 & \cdots & w_2^{ss} \\ \vdots & \vdots & \ddots & \vdots \\ w_{nc}^1 & w_{nc}^2 & \cdots & w_{nc}^{ss} \end{pmatrix}_{nc \times ss} \quad (C.1)$$

where an element  $w_i^j$  corresponds to the  $j^{\text{th}}$  element of  $i^{\text{th}}$  random variable.

- (ii) Step-2: Obtain mean subtracted matrix  $W_C$  as,

$$W_C = W - [\text{diag}(\mu_1, \mu_2, \dots, \mu_{nc})]_{nc \times nc} \begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 1 \end{pmatrix}_{nc \times ss} \quad (\text{C.2})$$

where  $\text{diag}(\bullet)$  is the diagonal matrix and  $\mu_i$  is calculated as,

$$\mu_i = \frac{1}{ss} \sum_{j=1}^{ss} w_i^j. \quad (\text{C.3})$$

- (iii) Step-3: Form correlation coefficient matrix  $(\rho)_{nc \times nc}$  and check its positive definiteness. In case of a non-positive definite, repair for positive semi-definite matrix.
- (iv) Step-4: Obtain eigenvalues of  $(\rho)_{nc \times nc}$  and arrange the eigenvalues in the order of  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{nc}$ . Obtain the corresponding eigenvectors  $u_i$ ,  $i = 1, 2, \dots, nc$  and form matrix  $U = \begin{pmatrix} u_1 | & u_2 | & \dots | & u_{nc} \end{pmatrix}$ .
- (v) Step-5: Finally, the principal components are obtained as,

$$Z = U^T W_C \quad (\text{C.4})$$

where each row in  $Z$  represents a principal component i.e., a time series which is univariate and uncorrelated with other principal components.

# Appendix D

## Estimation of variance coefficient

For an “nc” correlated input random variables (each of  $N_S$  samples) represented as  $X_{ij}; i = 1, 2, \dots, nc; j = 1, 2, \dots, N_S$  and “dnr” result variables, represented as,  $D_k; k = 1, 2, \dots, dnr$  bearing a functional relationship  $D = F(X)$ , the estimates of the expected values of result variables is given as,

$$\hat{E}(D_j) = \frac{1}{N_S} \sum_{j=1}^{N_S} F(X_j) \quad (\text{D.1})$$

where  $X_j$  in the  $j^{\text{th}}$  sampled value vector of “nc” input random variables and  $D_j$  is the  $j^{\text{th}}$  value vector of “dnr” result variables.

The uncertainty of the estimate in (D.1) is given by its variance which is calculated as,

$$V(\hat{E}(D)) = \frac{V(D)}{N_S} \quad (\text{D.2})$$

where  $V(D)$  is the variance of  $D$ , estimated as,

$$\hat{V}(D) = \frac{1}{N_S - 1} \sum_{j=1}^{N_S} \left\{ F(X_j) - \hat{E}(D_j) \right\}^2.$$

Finally, the variance coefficient  $\beta$  using (D.1) and (D.2) is estimated as,

$$\beta = \frac{\sqrt{V(\hat{E}(D))}}{\hat{E}(D)}. \quad (\text{D.3})$$





# Appendix E

## Formulation of the temperature-augmented Jacobian matrices

The formulations of the Jacobian matrices are as described under:

- (i) Formulation of  $JM_1$ : Since  $P$  at all buses except for the slack and  $Q$  at all the load buses are known, corresponding mismatch equations can be explicitly expressed. By eliminating the rows and columns corresponding to mismatch equations of  $P$  and  $Q$  of the slack bus as well as  $Q$  of the P|V| buses from (3.10) yields,

$$\begin{pmatrix} \Delta P^{NS} \\ \Delta Q^L \\ \Delta T' \end{pmatrix} = \begin{pmatrix} \frac{\partial P^{NS}}{\partial \delta} & \frac{\partial P^{NS}}{\partial |V|} & \frac{\partial P^{NS}}{\partial T} \\ \frac{\partial Q^L}{\partial \delta} & \frac{\partial Q^L}{\partial |V|} & \frac{\partial Q^L}{\partial T} \\ \frac{\partial T'}{\partial \delta} & \frac{\partial T'}{\partial |V|} & \frac{\partial T'}{\partial T} \end{pmatrix} \begin{pmatrix} \Delta \delta \\ \Delta |V| \\ \Delta T \end{pmatrix}. \quad (E.1)$$

In a more simplified form, (E.1) is expressed as,

$$\begin{pmatrix} \Delta P^{NS} \\ \Delta Q^L \\ \Delta T' \end{pmatrix} = \begin{pmatrix} J_1 & J_2 & J_3 \\ J_4 & J_5 & J_6 \\ J_7 & J_8 & J_9 \end{pmatrix} \begin{pmatrix} \Delta \delta \\ \Delta |V| \\ \Delta T \end{pmatrix} = (JM_1) \begin{pmatrix} \Delta \delta \\ \Delta |V| \\ \Delta T \end{pmatrix} \quad (E.2)$$

where  $P^{NS}$  is the real power injection vector at the non-slack buses and  $Q^L$  is the

reactive power injection vector at the PQ buses. The uncertainty component of state vector using (E.2) is obtained as,

$$\begin{pmatrix} \Delta\delta \\ \Delta|V| \\ \Delta T \end{pmatrix} = (\mathbf{JM}_1^{-1}) \begin{pmatrix} \Delta P^{\text{NS}} \\ \Delta Q^{\text{L}} \\ \Delta T' \end{pmatrix} = (\mathbf{JM}_1^{-1}) \begin{pmatrix} \Delta P^{\text{NS}} \\ \Delta Q^{\text{L}} \\ \Delta T_{\text{Amb}} \end{pmatrix}. \quad (\text{E.3})$$

The sub-matrices  $\mathbf{J}_1$ ,  $\mathbf{J}_2$ ,  $\mathbf{J}_4$ , and  $\mathbf{J}_5$  are same as that of the sub-matrices of conventional load flow Jacobian matrix (Wang et al., 2010). The formulation of sub-matrices  $\mathbf{J}_3$  and  $\mathbf{J}_6$  is detailed in (Frank et al., 2013). The partial differentiation of (3.8) with respect to the state variables yields the elements of the sub-matrices  $\mathbf{J}_7$ ,  $\mathbf{J}_8$ , and  $\mathbf{J}_9$ . Although the temperature related mismatch equation as developed in (Frank et al., 2013) is slightly different from that developed in (3.8), the sub-matrices of  $\mathbf{JM}_1$  in both the cases are identical.

- (ii) Formulation of  $\mathbf{JM}_2$ : By considering the rows corresponding to the mismatch equations of  $Q$  of P|V| buses and the columns corresponding to  $P^{\text{NS}}$  and  $Q^{\text{L}}$  of (E.1) yields,

$$\begin{aligned} (\Delta Q^{\text{P|V|}}) &= \begin{pmatrix} \frac{\partial Q^{\text{P|V|}}}{\partial \delta} & \frac{\partial Q^{\text{P|V|}}}{\partial |V|} & \frac{\partial Q^{\text{P|V|}}}{\partial T} \end{pmatrix} \begin{pmatrix} \Delta\delta \\ \Delta|V| \\ \Delta T \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{J}_{10} & \mathbf{J}_{11} & \mathbf{J}_{12} \end{pmatrix} \begin{pmatrix} \Delta\delta \\ \Delta|V| \\ \Delta T \end{pmatrix} = (\mathbf{JM}_2) \begin{pmatrix} \Delta\delta \\ \Delta|V| \\ \Delta T \end{pmatrix}. \end{aligned} \quad (\text{E.4})$$

In (E.4),  $Q^{\text{P|V|}}$  is the reactive power injection vector at the P|V| buses. The sub matrices  $\mathbf{J}_{10}$ ,  $\mathbf{J}_{11}$ , and  $\mathbf{J}_{12}$  for  $Q^{\text{P|V|}}$  are formulated in the same way as that of  $\mathbf{J}_4$ ,  $\mathbf{J}_5$ , and  $\mathbf{J}_6$  for  $Q^{\text{L}}$ .

- (iii) Formulation of  $\mathbf{JM}_3$ : The uncertainty component of branch power flow vector is expressed as,

$$\begin{aligned}
\begin{pmatrix} \Delta P_L \\ \Delta Q_L \end{pmatrix} &= \begin{pmatrix} \frac{\partial P_L}{\partial \delta} & \frac{\partial P_L}{\partial |V|} & \frac{\partial P_L}{\partial T} \\ \frac{\partial Q_L}{\partial \delta} & \frac{\partial Q_L}{\partial |V|} & \frac{\partial Q_L}{\partial T} \end{pmatrix} \begin{pmatrix} \Delta \delta \\ \Delta |V| \\ \Delta T \end{pmatrix} \\
&= \begin{pmatrix} J_{13} & J_{14} & J_{15} \\ J_{16} & J_{17} & J_{18} \end{pmatrix} \begin{pmatrix} \Delta \delta \\ \Delta |V| \\ \Delta T \end{pmatrix} = (\text{JM}_3) \begin{pmatrix} \Delta \delta \\ \Delta |V| \\ \Delta T \end{pmatrix}.
\end{aligned} \tag{E.5}$$

In (E.5), an element of the left side vector is the power flow of branch  $i - j$  at the  $i^{\text{th}}$  bus (i.e., the power flow from  $i^{\text{th}}$  bus to  $j^{\text{th}}$  bus). The formulation of the sub-matrices  $J_{13}$ ,  $J_{14}$ ,  $J_{16}$ , and  $J_{17}$  is detailed in (Wang et al., 2010). The formulation of sub-matrices  $J_{15}$  and  $J_{18}$  is detailed in (Frank et al., 2013).

- (iv) Formulation of  $\text{JM}_4$ : By considering the rows corresponding to the mismatch equations of  $P$  and  $Q$  of slack bus and the columns corresponding to  $P^{\text{NS}}$  and  $Q^{\text{L}}$  of (E.1) yields,

$$\begin{aligned}
\begin{pmatrix} \Delta P^{\text{S}} \\ \Delta Q^{\text{S}} \end{pmatrix} &= \begin{pmatrix} \frac{\partial P^{\text{S}}}{\partial \delta} & \frac{\partial P^{\text{S}}}{\partial |V|} & \frac{\partial P^{\text{S}}}{\partial T} \\ \frac{\partial Q^{\text{S}}}{\partial \delta} & \frac{\partial Q^{\text{S}}}{\partial |V|} & \frac{\partial Q^{\text{S}}}{\partial T} \end{pmatrix} \begin{pmatrix} \Delta \delta \\ \Delta |V| \\ \Delta T \end{pmatrix} \\
&= \begin{pmatrix} J_{19} & J_{20} & J_{21} \\ J_{22} & J_{23} & J_{24} \end{pmatrix} \begin{pmatrix} \Delta \delta \\ \Delta |V| \\ \Delta T \end{pmatrix} = (\text{JM}_4) \begin{pmatrix} \Delta \delta \\ \Delta |V| \\ \Delta T \end{pmatrix}.
\end{aligned} \tag{E.6}$$

In (E.6),  $P^{\text{S}}$  and  $Q^{\text{S}}$  are the real and reactive power injection vectors at the slack bus. The formulation of sub-matrices of  $\text{JM}_4$  is carried out in the same way as that of  $\text{JM}_1$ .

- (v) Formulation of  $\text{JM}_5$ : Real and reactive power flow vector from  $j^{\text{th}}$  bus to  $i^{\text{th}}$  bus is developed by using a set of similar steps as used for (E.5). It is given as,

$$\begin{pmatrix} \Delta P'_L \\ \Delta Q'_L \end{pmatrix} = (\text{JM}'_3) \begin{pmatrix} \Delta \delta \\ \Delta |V| \\ \Delta T \end{pmatrix} \tag{E.7}$$

The branch power loss vector is obtained as the sum of (E.5) and (E.7). It is

given as,

$$\begin{aligned} \begin{pmatrix} \Delta P_{\text{Loss}} \\ \Delta Q_{\text{Loss}} \end{pmatrix} &= \begin{pmatrix} \Delta P_L + \Delta P'_L \\ \Delta Q_L + \Delta Q'_L \end{pmatrix} \\ &= (\text{JM}_3 + \text{JM}'_3) \begin{pmatrix} \Delta \delta \\ \Delta |V| \\ \Delta T \end{pmatrix} = (\text{JM}_5) \begin{pmatrix} \Delta \delta \\ \Delta |V| \\ \Delta T \end{pmatrix}. \end{aligned} \quad (\text{E.8})$$

In (E.8)  $\text{JM}_5$  can also be obtained by the partial differentiation of branch power loss with respect to the state variables. The orders of sub-matrices are indicated in Table E.1.

**Table E.1:** Orders of sub-matrices.

SM	Order	SM	Order	SM	Order
$\text{J}_1$	$(n-1) \times (n-1)$	$\text{J}_9$	tdb $\times$ tdb	$\text{J}_{17}$	$\ell \times (n-m-1)$
$\text{J}_2$	$(n-1) \times (n-m-1)$	$\text{J}_{10}$	$m \times (n-1)$	$\text{J}_{18}$	$\ell \times$ tdb
$\text{J}_3$	$(n-1) \times$ tdb	$\text{J}_{11}$	$m \times (n-m-1)$	$\text{J}_{19}$	$1 \times (n-1)$
$\text{J}_4$	$(n-m-1) \times (n-1)$	$\text{J}_{12}$	$m \times$ tdb	$\text{J}_{20}$	$1 \times (n-m-1)$
$\text{J}_5$	$(n-m-1) \times (n-m-1)$	$\text{J}_{13}$	$\ell \times (n-1)$	$\text{J}_{21}$	$1 \times$ tdb
$\text{J}_6$	$(n-m-1) \times$ tdb	$\text{J}_{14}$	$\ell \times (n-m-1)$	$\text{J}_{22}$	$1 \times (n-1)$
$\text{J}_7$	tdb $\times (n-1)$	$\text{J}_{15}$	$\ell \times$ tdb	$\text{J}_{23}$	$1 \times (n-m-1)$
$\text{J}_8$	tdb $\times (n-m-1)$	$\text{J}_{16}$	$\ell \times (n-1)$	$\text{J}_{24}$	$1 \times$ tdb

# Appendix F

## Effect of input correlations on the statistics of the result variables

For “nr” random variables,  $X_1, X_2, \dots, X_{nr}$  with known expected values  $\mu_{X_i}$ , and variances  $\sigma_{X_i}^2$ ,  $i = 1, 2, \dots, nr$ , the expected value and variance of a linearized model  $Y = a_1X_1 + a_2X_2 + \dots + a_{nr}X_{nr}$  is given as,

$$\mu_Y = \sum_{i=1}^{nr} a_i \mu_{X_i}, \quad (\text{F.1})$$

$$\sigma_Y^2 = \sum_{i=1}^{nr} a_i^2 \sigma_{X_i}^2 + 2 \sum_{i=1}^{nr} \sum_{j=1}^{i-1} a_i a_j \text{cov} [X_i, X_j], \quad (\text{F.2})$$

where  $a_1, a_2, \dots, a_{nr}$  are the constant coefficients;  $\text{cov} [X_i, X_j]$  is the covariance between the random variables  $X_i$  and  $X_j$ .

It can be noted from (F.1) and (F.2) that the expected value of  $Y$  is not affected by input correlation whereas; the variance is influenced by input correlation. Interestingly, the variance is a linear function of input correlation. In case of a non-linear model, input correlation also affects the expected value. Consider a non-linear model of the form:  $Y = X_1X_2$ . The expected value and variance of  $Y$  are given as,

$$\mu_Y = \mu_{X_1}\mu_{X_2} + \text{cov} [X_1, X_2], \quad (\text{F.3})$$

$$\sigma_Y^2 = \mu_{X_1^2}\mu_{X_2^2} + \text{cov} [X_1^2, X_2^2] - (\mu_{X_1}\mu_{X_2} + \text{cov} [X_1, X_2])^2. \quad (\text{F.4})$$



# Bibliography

- Abdullah, M. A., Agalgaonkar, A., and Muttaqi, K. M. (2013). Probabilistic load flow incorporating correlation between time-varying electricity demand and renewable power generation. *Renewable energy*, 55:532–543.
- Aien, M., Fotuhi-Firuzabad, M., and Aminifar, F. (2012). Probabilistic load flow in correlated uncertain environment using unscented transformation. *IEEE Transactions on Power systems*, 27(4):2233–2241.
- Aien, M., Hajebrahimi, A., and Fotuhi-Firuzabad, M. (2016). A comprehensive review on uncertainty modeling techniques in power system studies. *Renewable and Sustainable Energy Reviews*, 57:1077–1089.
- Alam, M. J., Muttaqi, K., and Sutanto, D. (2013). A three-phase power flow approach for integrated 3-wire mv and 4-wire multigrounded lv networks with rooftop solar pv. *IEEE Transactions on Power Systems*, 28(2):1728–1737.
- Allan, R. and Al-Shakarchi, M. (1976). Probabilistic ac load flow. In *Proceedings of the Institution of Electrical Engineers*, volume 123, pages 531–536. IET.
- Allan, R. and Al-Shakarchi, M. (1977). Probabilistic techniques in ac load-flow analysis. In *Proceedings of the Institution of Electrical Engineers*, volume 124, pages 154–160. IET.
- Allan, R., Borkowska, B., and Grigg, C. (1974). Probabilistic analysis of power flows. In *Proceedings of the Institution of Electrical Engineers*, volume 121, pages 1551–1556. IET.
- Allan, R. and Da Silva, A. L. (1981). Probabilistic load flow using multilinearisation. In *IEE Proceedings C (Generation, Transmission and Distribution)*, volume 128, pages 280–287. IET.

- Allan, R., Da Silva, A. L., and Burchett, R. (1981). Evaluation methods and accuracy in probabilistic load flow solutions. *IEEE Transactions on Power Apparatus and Systems*, (5):2539–2546.
- Amid, P. and Crawford, C. (2018). A cumulant-tensor based probabilistic load flow method. *IEEE Transactions on Power Systems*.
- Borkowska, B. (1974). Probabilistic load flow. *IEEE Transactions on Power Apparatus and Systems*, (3):752–759.
- Brucoli, M., Torelli, F., and Napoli, R. (1985). Quadratic probabilistic load flow with linearly modelled dispatch. *International Journal of Electrical Power & Energy Systems*, 7(3):138–146.
- Cai, D., Chen, J., Shi, D., Duan, X., Li, H., and Yao, M. (2012). Enhancements to the cumulant method for probabilistic load flow studies. In *Power and Energy Society General Meeting, 2012 IEEE*, pages 1–8. IEEE.
- Cai, D., Li, X., Zhou, K., Xin, J., and Cao, K. (2015). Probabilistic load flow algorithms considering correlation between input random variables: A review. In *Industrial Electronics and Applications (ICIEA), 2015 IEEE 10th Conference on*, pages 1139–1144. IEEE.
- Cai, D., Shi, D., and Chen, J. (2014). Probabilistic load flow with correlated input random variables using uniform design sampling. *International Journal of Electrical Power & Energy Systems*, 63:105–112.
- Caramia, P., Carpinelli, G., Varilone, P., and Verde, P. (1999). Probabilistic three-phase load flow. *International Journal of Electrical Power & Energy Systems*, 21(1):55–69.
- Caramia, P. and Varilone, P. (1998). Probabilistic ac/dc 3-phase load flow. In *Harmonics and Quality of Power Proceedings, 1998. Proceedings. 8th International Conference On*, volume 2, pages 1018–1028. IEEE.
- Carmona-Delgado, C., Romero-Ramos, E., and Riquelme-Santos, J. (2015). Probabilistic load flow with versatile non-gaussian power injections. *Electric Power Systems Research*, 119:266–277.



- Ceylan, O., Ozdemir, A., and Dag, H. (2015). Double branch outage modeling and simulation: Bounded network approach. *International Journal of Electrical Power & Energy Systems*, 73:369–376.
- Chen, C., Wu, W., Zhang, B., and Sun, H. (2015). Correlated probabilistic load flow using a point estimate method with nataf transformation. *International Journal of Electrical Power & Energy Systems*, 65:325–333.
- Coutto Filho, M. D., Da Silva, A. L., Arienti, V., and Ribeiro, S. (1991). Probabilistic load modelling for power system expansion planning. In *Probabilistic Methods Applied to Electric Power Systems, 1991., Third International Conference on*, pages 203–207. IET.
- Cui, T. and Franchetti, F. (2013). A quasi-monte carlo approach for radial distribution system probabilistic load flow. In *Innovative Smart Grid Technologies (ISGT), 2013 IEEE PES*, pages 1–6. IEEE.
- Da Silva, A. L., Arienti, V., and Barbosa, M. (1987). Probabilistic techniques in load flow analysis a practical application. In *Probabilistic Methods Applied to Electric Power Systems*, pages 643–651. Elsevier.
- Da Silva, A. L., Lima, J. M., Ribeiro, S., Arienti, V., Schilling, M. T., Vieira Filho, X., and Soares, S. (1991). Operational and expansion planning of brazilian systems based on probabilistic load flow. In *Probabilistic Methods Applied to Electric Power Systems, 1991., Third International Conference on*, pages 197–202. IET.
- Da Silva, A. L., Ribeiro, S. P., Arienti, V., Allan, R., and Do Coutto Filho, M. (1990). Probabilistic load flow techniques applied to power system expansion planning. *IEEE Transactions on Power Systems*, 5(4):1047–1053.
- Dave, J., Halpern, P., and Myers, H. (1975). Computation of incident solar energy. *IBM Journal of Research and Development*, 19(6):539–549.
- Dziak, J. J., Coffman, D. L., Lanza, S. T., and Li, R. (2017). Sensitivity and specificity of information criteria. *PeerJ PrePrints*.
- Eftekharnajad, S., Vittal, V., Heydt, G. T., Keel, B., and Loehr, J. (2013). Impact of increased penetration of photovoltaic generation on power systems. *IEEE transactions on power systems*, 28(2):893–901.

- ERCOT (2016). Hourly load data. [http://www.ercot.com/gridinfo/load/load\\_hist](http://www.ercot.com/gridinfo/load/load_hist). [Online].
- Fan, M. (2012). *Probabilistic power flow studies to examine the influence of photovoltaic generation on transmission system reliability*. Arizona State University.
- Fan, M., Vittal, V., Heydt, G. T., and Ayyanar, R. (2012). Probabilistic power flow studies for transmission systems with photovoltaic generation using cumulants. *IEEE Transactions on Power Systems*, 27(4):2251–2261.
- Fan, M., Vittal, V., Heydt, G. T., and Ayyanar, R. (2013). Probabilistic power flow analysis with generation dispatch including photovoltaic resources. *IEEE Transactions on Power Systems*, 28(2):1797–1805.
- Fan, M., Vittal, V., Heydt, G. T., and Ayyanar, R. (2014). Preprocessing uncertain photovoltaic data. *IEEE Transactions on Sustainable Energy*, 5(1):351–352.
- Frank, S., Sexauer, J., and Mohagheghi, S. (2013). Temperature-dependent power flow. *IEEE Transactions on Power Systems*, 28(4):4007–4018.
- Frank, S., Sexauer, J., and Mohagheghi, S. (2016). Temperature dependent power flow in MATLAB. [https://github.com/TDPF/TDPF/blob/master/39bus\\_data.mat](https://github.com/TDPF/TDPF/blob/master/39bus_data.mat). [Online].
- Gomez-Gonzalez, M., Ruiz-Rodriguez, F., and Jurado, F. (2013). A binary sfla for probabilistic three-phase load flow in unbalanced distribution systems with technical constraints. *International Journal of Electrical Power & Energy Systems*, 48:48–57.
- Hajian, M., Rosehart, W. D., and Zareipour, H. (2013). Probabilistic power flow by monte carlo simulation with latin supercube sampling. *IEEE Transactions on Power Systems*, 28(2):1550–1559.
- Hatziargyriou, N. and Karakatsanis, T. (1994). A probabilistic approach to control variable adjustment for power system planning applications. In *Control, 1994. Control'94. International Conference on*, volume 1, pages 733–738. IET.
- Hatziargyriou, N. and Karakatsanis, T. (1997). Distribution system voltage and reactive power control based on probabilistic load flow analysis. *IEE Proceedings-Generation, Transmission and Distribution*, 144(4):363–369.

- Hatziargyriou, N., Karakatsanis, T., and Lorentzou, M. (2005). Voltage control settings to increase wind power based on probabilistic load flow. *International Journal of Electrical Power & Energy Systems*, 27(9-10):656–661.
- Ho, T., Chi, Y., Wang, J., and Leung, K. (2004). Load flow in electrified railway. In *Power Electronics, Machines and Drives, 2004.(PEMD 2004). Second International Conference on (Conf. Publ. No. 498)*, volume 2, pages 498–503. IET.
- Ho, T., Chi, Y., Wang, J., Leung, K., Siu, L., and Tse, C. (2005). Probabilistic load flow in ac electrified railways. *IEE Proceedings-Electric Power Applications*, 152(4):1003–1013.
- Ho, T., Wang, J., Leung, K., Tse, C., and Siu, L. (2003). Probabilistic load flow calculation for ac traction supplies with at feeding.
- Hoese, A. and Garcés, F. (1999). Stochastic correlated simulation: an extension of the cumulant method to include time-dependent energy sources. *International Journal of Electrical Power & Energy Systems*, 21(1):13–22.
- Holbert, K. E. and Srinivasan, D. (2011). Solar energy calculations. In *Handbook Of Renewable Energy Technology*, pages 189–204. World Scientific.
- Hong, Y.-Y., Lin, F.-J., and Yu, T.-H. (2016). Taguchi method-based probabilistic load flow studies considering uncertain renewables and loads. *IET Renewable Power Generation*, 10(2):221–227.
- Hu, Z. and Wang, X. (2006). A probabilistic load flow method considering branch outages. *IEEE Transactions on Power Systems*, 21(2):507–514.
- Huang, J., Xue, Y., Dong, Z., and Wong, K. (2011). An adaptive importance sampling method for probabilistic optimal power flow. In *Power and Energy Society General Meeting, 2011 IEEE*, pages 1–6. IEEE.
- Kang, C., Xia, Q., and Xiang, N. (2002). Sequence operation theory and its application in power system reliability evaluation. *Reliability Engineering & System Safety*, 78(2):101–109.
- Kang, M., Chen, C., Ke, Y., and Lee, T. (2003). Stochastic load flow analysis by considering temperature sensitivity of customer power consumption. In *Power Tech Conference Proceedings, 2003 IEEE Bologna*, volume 3, pages 6–pp. IEEE.

- Kenari, M. T., Sepasian, M. S., Nazar, M. S., and Mohammadpour, H. A. (2017). Combined cumulants and laplace transform method for probabilistic load flow analysis. *IET Generation, Transmission & Distribution*, 11(14):3548–3556.
- Le, D. D., Gross, G., and Berizzi, A. (2015). Probabilistic modeling of multisite wind farm production for scenario-based applications. *IEEE Transactions on Sustainable Energy*, 6(3):748–758.
- Li, X., Zhang, X., Wu, L., Lu, P., and Zhang, S. (2015). Transmission line overload risk assessment for power systems with wind and load-power generation correlation. *IEEE Transactions on Smart Grid*, 6(3):1233–1242.
- Liu, H., Tang, C., Han, J., Li, T., Li, J., and Zhang, K. (2017). Probabilistic load flow analysis of active distribution network adopting improved sequence operation methodology. *IET Generation, Transmission & Distribution*, 11(9):2147–2153.
- Liu, P.-L. and Der Kiureghian, A. (1986). Multivariate distribution models with prescribed marginals and covariances. *Probabilistic Engineering Mechanics*, 1(2):105–112.
- McLeod, A. I. and Gweon, H. (2013). Optimal deseasonalization for monthly and daily geophysical time series.
- Mohammadi, M. (2015). Probabilistic harmonic load flow using fast point estimate method. *IET Generation, Transmission & Distribution*, 9(13):1790–1799.
- Morales, J. M., Baringo, L., Conejo, A. J., and Mínguez, R. (2010). Probabilistic power flow with correlated wind sources. *IET generation, transmission & distribution*, 4(5):641–651.
- Morales, J. M. and Perez-Ruiz, J. (2007). Point estimate schemes to solve the probabilistic power flow. *IEEE Transactions on power systems*, 22(4):1594–1601.
- NREL (2017). Hourly temperature data. <https://maps.nrel.gov/nsrdb-viewer/>. [Online].
- Ozdemir, A., Lim, J. Y., and Singh, C. (2003). Branch outage simulation for mvar flows: bounded network solution. *IEEE Transactions on Power Systems*, 18(4):1523–1528.

- Papaefthymiou, G. (2007). *Integration of stochastic generation in power systems*. Delft University of Technology.
- Papaefthymiou, G. and Kurowicka, D. (2009). Using copulas for modeling stochastic dependence in power system uncertainty analysis. *IEEE Transactions on Power Systems*, 24(1):40–49.
- Peng, C., Lei, S., Hou, Y., and Wu, F. (2015). Uncertainty management in power system operation. *CSEE Journal of Power and Energy Systems*, 1(1):28–35.
- Pham, D. T., Dimov, S. S., and Nguyen, C. D. (2005). Selection of k in k-means clustering. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, 219(1):103–119.
- PVOUT (2016). Hourly PV generation data. <https://www.pvoutput.org/ladder>. [Online].
- PWTEB (2017). Tamilnadu electricity board statistics at a glance 1999-2000. [www.ee.washington.edu/research/pstca](http://www.ee.washington.edu/research/pstca). [Online].
- Ran, X. and Miao, S. (2016). Three-phase probabilistic load flow for power system with correlated wind, photovoltaic and load. *IET Generation, Transmission & Distribution*, 10(12):3093–3101.
- Rouhani, M., Mohammadi, M., and Kargarian, A. (2016). Parzen window density estimator-based probabilistic power flow with correlated uncertainties. *IEEE Transactions on Sustainable Energy*, 7(3):1170–1181.
- Sanabria, L. and Dillon, T. (1986). Stochastic power flow using cumulants and von mises functions. *International Journal of Electrical Power & Energy Systems*, 8(1):47–60.
- Sansavini, G., Piccinelli, R., Golea, L., and Zio, E. (2014). A stochastic framework for uncertainty analysis in electric power transmission systems with wind generation. *Renewable Energy*, 64:71–81.
- Sauer, P. and Hoveida, B. (1982). Constrained stochastic power flow analysis. *Electric Power Systems Research*, 5(2):87–95.

- Schlapfer, M. and Mancarella, P. (2011). Probabilistic modeling and simulation of transmission line temperatures under fluctuating power flows. *IEEE Transactions on Power Delivery*, 26(4):2235–2243.
- Sirisena, H. and Brown, E. (1983). Representation of non-gaussian probability distributions in stochastic load-flow studies by the method of gaussian sum approximations. In *IEE Proceedings C (Generation, Transmission and Distribution)*, volume 130, pages 165–171. IET.
- Soleimanpour, N. and Mohammadi, M. (2013). Probabilistic load flow by using nonparametric density estimators. *IEEE Transactions on Power systems*, 28(4):3747–3755.
- Su, C.-L. (2005a). Distribution probabilistic load flow solution considering network reconfiguration and voltage control devices. In *15th Power Systems Computation Conference*. Citeseer.
- Su, C.-L. (2005b). Probabilistic load-flow computation using point estimate method. *IEEE Transactions on Power Systems*, 20(4):1843–1851.
- Usaola, J. (2009). Probabilistic load flow in systems with wind generation. *IET generation, transmission & distribution*, 3(12):1031–1041.
- Usaola, J. (2010). Probabilistic load flow with correlated wind power injections. *Electric Power Systems Research*, 80(5):528–536.
- UWEE (2015). Power system test cases archive. [www.ee.washington.edu/research/pstca](http://www.ee.washington.edu/research/pstca). [Online].
- Valverde, G., Saric, A., and Terzija, V. (2012). Probabilistic load flow with non-gaussian correlated random variables using gaussian mixture models. *IET generation, transmission & distribution*, 6(7):701–709.
- Villanueva, D., Feijóo, A. E., and Pazos, J. L. (2014). An analytical method to solve the probabilistic load flow considering load demand correlation using the dc load flow. *Electric Power Systems Research*, 110:1–8.
- Wang, X. and McDonald, J. (1994). *Modern power system planning*. McGraw-Hill International (UK) Limited.

- Wang, X.-F., Song, Y., and Irving, M. (2010). *Modern power systems analysis*. Springer Science & Business Media.
- Wang, Y.-J. and Pierrat, L. (2001). A method integrating deterministic and stochastic approaches for the simulation of voltage unbalance in electric power distribution systems. *IEEE Transactions on Power Systems*, 16(2):241–246.
- Williams, T. and Crawford, C. (2013). Probabilistic load flow modeling comparing maximum entropy and gram-charlier probability density function reconstructions. *IEEE Transactions on Power Systems*, 28(1):272–280.
- Wu, W., Wang, K., Li, G., Jiang, X., and Wang, Z. (2016). Probabilistic load flow calculation using cumulants and multiple integrals. *IET Generation, Transmission & Distribution*, 10(7):1703–1709.
- WUNDER-GROUND (2016). Hourly temperature data. <https://www.wunderground.com/history>. [Online].
- Yu, H., Chung, C., Wong, K., Lee, H., and Zhang, J. (2009). Probabilistic load flow evaluation with hybrid latin hypercube sampling and cholesky decomposition. *IEEE Transactions on Power Systems*, 24(2):661–667.
- Yuan, Y., Zhou, J., Ju, P., and Feuchtwang, J. (2011). Probabilistic load flow computation of a power system containing wind farms using the method of combined cumulants and gram-charlier expansion. *IET renewable power generation*, 5(6):448–454.
- Zhang, N., Kang, C., Singh, C., and Xia, Q. (2016). Copula based dependent discrete convolution for power system uncertainty analysis. *IEEE Transactions on Power Systems*, 31(6):5204–5205.
- Zhang, P. and Lee, S. T. (2004). Probabilistic load flow computation using the method of combined cumulants and gram-charlier expansion. *IEEE transactions on power systems*, 19(1):676–682.
- Zhou, G., Bo, R., Chien, L., Zhang, X., Yang, S., and Su, D. (2018). Gpu-accelerated algorithm for online probabilistic power flow. *IEEE Transactions on Power Systems*, 33(1):1132–1135.

Zou, B. and Xiao, Q. (2014). Probabilistic load flow computation using univariate dimension reduction method. *International Transactions on Electrical Energy Systems*, 24(12):1700–1714.



## Bio-data

Name: B Rajanarayan Prusty

Date of birth: 05-12-1985

Marital status: Married

E-mail: b.r.prusty@ieee.org, openeye.engg@gmail.com

Mobile: +91-6371257662, +91-9861845744

### Address

S/o: Dr. B. Venketeswar Prusty

Bank colony-6th lane,

Near All India Radio (AIR),

Lochapada road,

Berhampur, Odisha-760001, India.

### Education

- **M.Tech.** in Electrical Engineering (2011)  
Institute: National Institute of Science and Technology (NIST), Berhampur, Odisha  
University: Biju Patnaik University of Technology (BPUT), Odisha  
CGPA: 8.68
- **B.Tech.** in Electrical and Electronics Engineering (2007)  
Institute: National Institute of Science and Technology (NIST), Berhampur, Odisha  
University: Biju Patnaik University of Technology (BPUT), Odisha  
CGPA: 7.24

### Award

- POSOCO Power System Awards (PPSA) - 2019