

Situational awareness in low-observable distribution grid - exploiting
sparsity and multi-timescale data

by

Shweta Dahale

B.Tech., College of Engineering Pune, 2016

M.Tech., Indian Institute of Technology Gandhinagar, 2018

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Mike Wieggers Department of Electrical and Computer Engineering
Carl R. Ice College of Engineering

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2023

Abstract

The power distribution grid is typically unobservable due to a lack of real-time measurements. While deploying more sensors can alleviate this issue, it also presents new challenges related to data aggregation and the underlying communication infrastructure. Limited real-time measurements hinders the distribution system state estimation (DSSE). DSSE involves estimation of the system states (i.e., voltage magnitude and voltage angle) based on available measurements and system model information. To cope with the unobservability issue, sparsity-based DSSE approaches allow us to recover system state information from a small number of measurements, provided the states of the distribution system exhibit sparsity. However, these approaches perform poorly in the presence of outliers in measurements and errors in system model information. In this dissertation, we first develop robust formulations of sparsity-based DSSE to deal with uncertainties in the system model and measurement data in a low-observable distribution grid. We also combine the advantages of two sparsity-based DSSE approaches to estimate grid states with high fidelity in low observability regions.

In practical distribution systems, information from field sensors and meters are unevenly sampled at different time scales and could be lost during the transmission process. It is critical to effectively aggregate these information sources for DSSE as well as other tasks related to situational awareness. To address this challenge, the second part of this dissertation proposes a Bayesian framework for multi-timescale data aggregation and matrix completion-based state estimation. Specifically, the multi-scale time-series data aggregated from heterogeneous sources are reconciled using a multitask Gaussian process. The resulting consistent time-series alongwith the confidence bound on the imputations are fed into a Bayesian matrix completion method augmented with linearized power-flow constraints for accurate state estimation low-observable distribution system. We also develop a computationally efficient recursive Gaussian process approach that is capable of handling batch-wise

or real-time measurements while leveraging the network connectivity information of the grid. To further enhance the scalability and accuracy, we develop neural network-based approaches (latent neural ordinary differential equation approach and stochastic neural differential equation with recurrent neural network approach) to aggregate irregular time-series data in the distribution grid. The stochastic neural differential equation and recurrent neural network also allows us to quantify the uncertainty in a holistic manner. Simulation results on the different IEEE unbalanced test systems illustrate the high fidelity of the Bayesian and neural network-based methods in aggregating multi-timescale measurements.

Lastly, we develop phase, and outage awareness approaches for power distribution grid. In this regard, we first design a graph signal processing approach that identifies the phase labels in the presence of limited measurements and incorrect phase labeling. The second approach proposes a novel outage detector for identifying all outages in a reconfigurable distribution network. Simulation results on standard IEEE test systems reveal the potential of these methods to improve situational awareness.

Situational awareness in low-observable distribution grid - exploiting
sparsity and multi-timescale data

by

Shweta Dahale

B.Tech., College of Engineering Pune, 2016

M.Tech., Indian Institute of Technology Gandhinagar, 2018

A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Mike Wieggers Department of Electrical and Computer Engineering
Carl R. Ice College of Engineering

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2023

Approved by:

Major Professor
Dr. Balasubramaniam Natarajan

Copyright

© Shweta Dahale 2023.

Abstract

The power distribution grid is typically unobservable due to a lack of real-time measurements. While deploying more sensors can alleviate this issue, it also presents new challenges related to data aggregation and the underlying communication infrastructure. Limited real-time measurements hinders the distribution system state estimation (DSSE). DSSE involves estimation of the system states (i.e., voltage magnitude and voltage angle) based on available measurements and system model information. To cope with the unobservability issue, sparsity-based DSSE approaches allow us to recover system state information from a small number of measurements, provided the states of the distribution system exhibit sparsity. However, these approaches perform poorly in the presence of outliers in measurements and errors in system model information. In this dissertation, we first develop robust formulations of sparsity-based DSSE to deal with uncertainties in the system model and measurement data in a low-observable distribution grid. We also combine the advantages of two sparsity-based DSSE approaches to estimate grid states with high fidelity in low observability regions.

In practical distribution systems, information from field sensors and meters are unevenly sampled at different time scales and could be lost during the transmission process. It is critical to effectively aggregate these information sources for DSSE as well as other tasks related to situational awareness. To address this challenge, the second part of this dissertation proposes a Bayesian framework for multi-timescale data aggregation and matrix completion-based state estimation. Specifically, the multi-scale time-series data aggregated from heterogeneous sources are reconciled using a multitask Gaussian process. The resulting consistent time-series alongwith the confidence bound on the imputations are fed into a Bayesian matrix completion method augmented with linearized power-flow constraints for accurate state estimation low-observable distribution system. We also develop a computationally efficient recursive Gaussian process approach that is capable of handling batch-wise

or real-time measurements while leveraging the network connectivity information of the grid. To further enhance the scalability and accuracy, we develop neural network-based approaches (latent neural ordinary differential equation approach and stochastic neural differential equation with recurrent neural network approach) to aggregate irregular time-series data in the distribution grid. The stochastic neural differential equation and recurrent neural network also allows us to quantify the uncertainty in a holistic manner. Simulation results on the different IEEE unbalanced test systems illustrate the high fidelity of the Bayesian and neural network-based methods in aggregating multi-timescale measurements.

Lastly, we develop phase, and outage awareness approaches for power distribution grid. In this regard, we first design a graph signal processing approach that identifies the phase labels in the presence of limited measurements and incorrect phase labeling. The second approach proposes a novel outage detector for identifying all outages in a reconfigurable distribution network. Simulation results on standard IEEE test systems reveal the potential of these methods to improve situational awareness.

Table of Contents

List of Figures	xii
List of Tables	xvi
List of Nomenclature	xviii
Acknowledgements	xx
Dedication	xxi
Preface	xxii
1 Introduction	1
1.1 Research Questions	4
1.2 Contributions	5
1.3 Organization of the Dissertation	10
2 Literature review	12
2.1 Distribution system state estimation	12
2.2 Integrating multi time-scale measurements	14
2.3 Phase Identification	17
2.4 Sensor placement and outage detection	18
3 Background	21
3.1 State estimation techniques in distribution grid	21
3.1.1 WLS method	23

3.1.2	Compressive Sensing	24
3.1.3	Matrix completion	26
3.1.4	Tensor completion	27
4	Robust and Efficient Sparsity-based state estimators	29
4.1	Robust DSSE approaches	30
4.1.1	Proposed Approach: Robustness to bad data	30
4.1.2	Proposed Approach: Robustness to network parameter uncertainty	32
4.1.3	Simulation Results	33
4.2	Efficient DSSE approaches	38
4.2.1	Proposed Joint MC-CS Approach	39
4.2.2	Proposed CS-MC Approach	44
4.2.3	Computational Complexity	45
4.2.4	Simulation Results	45
4.3	Summary	47
5	Handling multi time-scale measurements using Multi-task GP approach	49
5.1	System model	51
5.2	Proposed Bayesian Approach	51
5.2.1	Stage 1 - Multi-task Gaussian Process	52
5.2.2	Stage II - Bayesian matrix completion based DSSE	55
5.3	Computational complexity	59
5.4	Simulation results	61
5.4.1	IEEE 37 bus test system	62
5.4.2	IEEE 123 bus test system	69
5.5	Summary	72
6	Streaming multi time-scale data using Recursive GP approach	74
6.1	Background	74

6.1.1	Full GP Approach	77
6.2	Proposed Recursive GP approach	79
6.2.1	RGP-G Approach	80
6.2.2	RGP Approach	86
6.3	Computational Complexity	88
6.4	Simulation results	90
6.4.1	IEEE 37 bus test system	92
6.4.2	IEEE 123 bus test system	96
6.4.3	11,000 bus test system	97
6.5	Summary	98
7	Handling multi time-scale data using neural network based approaches	99
7.1	Background: Neural ODE	99
7.2	Proposed Generative model - Latent Neural ODE	102
7.2.1	Encoder: Recognition network	103
7.2.2	ODE Solver	104
7.2.3	Decoder	104
7.2.4	Training of LODE	104
7.3	Proposed Autoregressive model - SDE-RNN	105
7.3.1	Uncertainty quantification	106
7.4	Simulation results	112
7.4.1	Data processing	112
7.4.2	LODE approach	113
7.4.3	SDE-RNN approach	115
7.5	Summary	120
8	Phase and outage awareness in distribution network	121
8.1	Phase Identification approach	122

8.1.1	System model	123
8.1.2	Proposed phase identification approach	124
8.1.3	Joint Phase Identification and state estimation	129
8.1.4	Simulation results	131
8.2	Sensor placement and outage detection in reconfigurable distribution grid . .	138
8.2.1	System model	138
8.2.2	Proposed sensor placement approach	140
8.2.3	Proposed outage detector	145
8.2.4	Simulation Results	146
8.3	Summary	154
9	Conclusion and future work	155
9.1	Conclusions	155
9.2	Future directions	157
	Bibliography	160
A	Reuse permissions from publishers	177

List of Figures

1.1	Operation of distribution system state estimation	2
1.2	Outline of the dissertation	11
4.1	Case I: Voltage magnitude performance of spatial approaches for different bad data	35
4.2	Case I: Voltage angle performance of spatial approaches for different bad data	35
4.3	Case I: Voltage magnitude performance of spatiotemporal approaches for different bad data	36
4.4	Case I: Voltage angle performance of spatiotemporal approaches for different bad data	36
4.5	Case II: Voltage magnitude performance of spatial approaches in presence of network parameter uncertainty	37
4.6	Case II: Voltage angle performance of spatial approaches in presence of network parameter uncertainty	37
4.7	Framework of the Joint MC-CS approach	40
4.8	Framework of the CS-MC approach	44
4.9	Power recovery at different FADs	46
4.10	Voltage magnitude recovery at different FADs	47
4.11	Voltage angle recovery at different FADs	47
5.1	Relative frequency of data acquisition from different sources in distribution systems	50
5.2	Proposed Bayesian framework	52
5.3	Flowchart of the proposed Bayesian framework in distribution systems . . .	60

5.4	Training dataset for IEEE 37 bus test system	63
5.5	Active power time-series imputation by multi-task GP framework	66
5.6	Reactive power time-series imputation by multi-task GP framework	66
5.7	Active power time-series imputation by multi-task GP framework - 10% Gaussian Error	67
5.8	Measurement configuration at 20% FAD for three phase unbalanced IEEE 37 bus test system.	69
5.9	Power recovery performance at different FADs for IEEE 37 test system corresponding to 1% noise and 10% noise	70
5.10	Voltage magnitude recovery performance at different FADs for IEEE 37 test system corresponding to 1% noise and 10% noise	70
5.11	Voltage angle recovery performance at different FADs for IEEE 37 test system corresponding to 1% noise and 10% noise	71
5.12	Power recovery performance at different FADs for IEEE 123 test system . .	72
5.13	Voltage magnitude recovery performance at different FADs for IEEE 123 test system	73
5.14	Voltage angle recovery performance at different FADs for IEEE 123 test system	73
6.1	Classification of the proposed approaches for processing streaming data . . .	76
6.2	RGP-G prediction approach	86
6.3	RGP-G Interpolation approach of active power injection timeseries at node 11, Phase A	93
6.4	Meter deployment corresponding to 50% FAD for IEEE 37 bus test system .	94
6.5	Comparison of RGP-G interpolation and linear interpolation approach of an active power injection time-series at Node 2 of Phase A	94
7.1	Architecture of proposed Latent ODE approach for smart distribution grid .	103
7.2	Propagation of uncertainty. x_i : Input at irregular intervals	106

7.3	Imputation at node 1 using Latent ODE approach	114
7.4	MSE on the test dataset using Latent ODE approach	115
7.5	Negative ELBO on the test dataset using Latent ODE approach	115
7.6	Predictions at node 12 using Latent ODE approach	116
7.7	Uncertainty quantification in SDE-RNN approach- node 1	118
7.8	Uncertainty quantification in SDE-RNN approach- node 12	118
8.1	Tree representation of network topology with question mark ? representing the unknown phase labels	123
8.2	Different possible configurations for a two-bus connection. (a) three phase end (b) two-phase end (c) one-phase end	124
8.3	Case I: Performance of phase identification approaches in the presence of missing measurements - IEEE 37 bus test system	134
8.4	Case II: Performance of the proposed phase identification approach in the presence of unknown phase labels- IEEE 37 bus test system	134
8.5	Case III: Performance of phase identification approaches in the presence of incorrect phase labels	135
8.6	Case IV: Performance of phase identification approaches in unobservable conditions- IEEE 37 bus test system	135
8.7	Case V: Voltage magnitude recovery for IEEE 37 bus system	136
8.8	Case V: Voltage angle recovery for IEEE 37 bus system	136
8.9	Performance of the proposed phase identification approach under unobserv- able conditions - IEEE 123 bus test system	137
8.10	Performance of the proposed phase identification approach in the presence of unknown phase labels - IEEE 123 bus test system	137
8.11	Meter placement for IEEE 33 bus system - Configuration 1	148
8.12	Meter placement for IEEE 33 bus system - Configuration 2	148

8.13 Probability of outage detection for different load covariances of 33 bus test system	149
8.14 Probability of outage detection for different configurations of 33 bus test system	150
8.15 Probability of outage detection for different configurations of IEEE 119 bus test system	153
8.16 Meter placement for modified IEEE 33 bus system	153

List of Tables

5.1	Root mean square error of imputed time-series data - 1% Gaussian noise (LI: Linear Interpolation approach, GP: multitask Gaussian process approach) . .	64
5.2	Root mean square error of imputed time-series data - 10% Gaussian noise . .	69
5.3	Average time requirement (FAD = 75%)	72
6.1	Case 2: MAPE of active and reactive power imputed time-series data (IEEE 37 bus test system)	93
6.2	MAPE of proposed RGP-G approach for Laplacian and Gaussian noise with standard deviation set as percentage of actual power values	95
6.3	Absolute errors and relative error reductions (%) compared to the actual measurements (LI : Linear Interpolation approach, %: percentage reductions) . .	96
6.4	Case 3: MAPE of imputed time-series data of active and reactive power (IEEE 37 bus test system)	96
6.5	Case 2: MAPE of active and reactive power imputed time-series data (IEEE 123 bus test system)	97
6.6	MAPE of active and reactive power imputed time-series data (11,000 node feeder)	98
7.1	MSE performance of the proposed LODE, linear interpolation and recursive GP approach	114
7.2	Imputation results for different percentages of missing data in test dataset .	119
8.1	Meter placement solution for IEEE 33 bus reconfigurable distribution system	147
8.2	Outage detection using different meter placement approaches	150

8.3	Comparison of the proposed approach with state-of-the-art approaches . . .	151
8.4	Meter placement solution for IEEE 119 bus re-configurable distribution system	152

Nomenclature

DSSE Distribution system state estimation

AMI Advanced metering infrastructure

PMU Phasor measurement unit

CS Compressive Sensing

MC Matrix Completion

SVD Singular value decomposition

DMS Distribution management database

GP Gaussian Process

FAD Fraction of Available Data

NLL Negative log-likelihood

KL KL divergence

WLS Weighted least square

SR Sparsity Ratio

RIP Restricted isometry property

KF Kalman filter

DCT Discrete cosine transform

ODE Ordinary differential equation

SDE Stochastic differential equation

RNN Recurrent neural network

GRU Gated recurrent unit

ML Maximum likelihood

Acknowledgments

My most profound appreciation goes to my advisor Dr. Balasubramaniam Natarajan for his guidance, support, and encouragement during my Ph.D. study. I'm incredibly grateful for his invaluable patience and feedback. I sincerely thank my doctoral committee members, Dr. Anil Pahwa, Dr. Hongyu Wu, and Dr. Weixing Song, for their constructive advice and comments that help improve the quality of my research. I am grateful to Dr. Rui Yang for her mentorship and support during my PhD journey.

I'd like to recognize my colleagues at CPWin Lab, Sai, Hazhar, Kexing, Mohammad, Deepesh, Biswajeet, Dylan, Aabila, Prudhviraaj, and Priyanka, for working alongside and providing constant motivation. I would like to thank my dear and loyal friends at K-STATE, Amogh, Sneha, Gunvant, Laxmi, Prakash, Priya, Snehal, and Ganesh for their support and great care. I will always miss the unforgettable moments and joys you shared.

I'm also grateful to my parents, Dnyaneshwar Dahale, Dr. Shobha Dahale, and my brother Hrushikesh Dahale for their unconditional love, support, and encouragement. Your prayer for me was what sustained me thus far. Finally, I'd like to thank my husband, Biswajeet Rout. Thank you for supporting me for everything, and especially I can't thank you enough for encouraging me throughout this experience.

This material is based upon work supported by the Department of Energy, Office of Energy Efficiency and Renewable Energy (EERE), Solar Energy Technologies Office, under Award Number DE-EE0008767.

Dedication

This thesis is dedicated to my parents Dnyaneshwar Dahale and Dr. Shobha Dahale for their patience and encouragement during my education.

Preface

This dissertation, “Situational awareness in Low-Observable Distribution Grid - Exploiting Sparsity and Multi-timescale Data,” is submitted for the degree of Doctor of Philosophy in the Mike Weigers Department of Electrical and Computer Engineering at Kansas State University. The research was conducted under the supervision of Professor Balasubramaniam Natarajan. This work is original to the best of my knowledge, except where acknowledgments and references are indicated. Most of the work has been presented in the following published peer-reviewed journals and conferences:

1. S. Dahale, H. S. Karimi, K. Lai, and B. Natarajan, “Sparsity Based Approaches for Distribution Grid State Estimation - A Comparative Study,” *IEEE Access*, vol. 8, pp. 198317–198327, 2020, doi: 10.1109/ACCESS.2020.3035378.
2. S. Dahale and B. Natarajan, “Joint Matrix Completion and Compressed Sensing for State Estimation in Low-observable Distribution System,” in *2021 IEEE PES Innovative Smart Grid Technologies Conference - Latin America (ISGT Latin America)*, 2021, pp. 1–5. doi: 10.1109/ISGTLatinAmerica52371.2021.9543006.
3. S. Dahale and B. Natarajan, “Multi Time-scale Imputation aided State Estimation in Distribution System,” in *2021 IEEE Power Energy Society General Meeting (PESGM)*, 2021, pp. 1–5. doi: 10.1109/PESGM46819.2021.9637824.
4. S. Dahale and B. Natarajan, “Bayesian Framework for Multi-Timescale State Estimation in Low-Observable Distribution Systems,” *IEEE Trans. Power Syst.*, vol. 37, no. 6, pp. 4340–4351, Nov. 2022, doi: 10.1109/TPWRS.2022.3155151.
5. S. Dahale and B. Natarajan, “Recursive Gaussian Process over graphs for Integrating Multi-timescale Measurements in Low-Observable Distribution Systems,” *IEEE Trans. Power Syst.*, pp. 1–12, 2022, doi: 10.1109/TPWRS.2022.3204415

6. B. Rout, S. Dahale, and B. Natarajan, “Dynamic Matrix Completion Based State Estimation in Distribution Grids,” *IEEE Trans. Ind. Inform.*, vol. 18, no. 11, pp. 7504–7511, Nov. 2022, doi: 10.1109/TII.2022.3162210.
7. M. Abujubbeh, S. Dahale, and B. Natarajan, “Voltage Violation Prediction in Unobservable Distribution Systems,” in *2022 IEEE Power & Energy Society General Meeting (PESGM), Jul. 2022*, pp. 1–5. doi: 10.1109/PESGM48719.2022.9916805.
8. S. Dahale, A. Pahwa, and B. Natarajan, “Phase Identification in Unobservable Distribution Systems,” *IEEE Trans. Power Deliv.*, 2023 (In press).
9. S. Dahale, and B. Natarajan. “Sensor Placement and Outage Detection for Reconfigurable Distribution System.”, *IEEE Trans. Power Syst.*, (under review).
10. S.Dahale, S. Munikoti, B. Natarajan, R. Yang, “Latent Neural ODE for Integrating Multi-timescale measurements in Smart Distribution Grids,” in *2023 IEEE PES Innovative Smart Grid Technologies Conference - North America (ISGT North America)*, 2023 (manuscript in press).
11. S.Dahale, S. Munikoti, and B. Natarajan, “A General Framework for Uncertainty Quantification in Neural SDE-RNN”, (manuscript to be submitted).

Chapter 1

Introduction

Situational awareness in distribution systems refers to the ability to have a real-time understanding of the system's current state, including the status of equipment, the flow of power, and the system's overall performance. This awareness allows operators and engineers to make informed decisions about the operation of the distribution system and respond quickly to any issues or disturbances that may arise. State estimation (SE) is an effective tool for improving the situational awareness of the grid. Although SE is a well-studied and widely used concept in the transmission system, its use is still a topic of active research in distribution systems. Fig. 1.1 depicts the different components necessary for performing distribution grid state estimation (DSSE). The system states (voltage magnitude and voltage angle) are estimated using the DSSE module by using available measurements and the system model information. The system model comprises of network topology, phase labels, and network parameters. This information is typically stored in the distribution management system (DMS) database. In a distribution network, real-time measurements are limited, and network observability is only achieved if historical data based pseudo-measurements are used. To further aggravate the issue, the presence of outliers in measurements and errors in system model information introduces new sources of uncertainty in DSSE. Since optimal control and operation of the distribution grid is contingent upon the ability to estimate the system states accurately, it is critical to investigate methods to estimate the states accurately with a small

number of measurements even in the presence of uncertainties in both the measurement and system model information. This is one of the fundamental problem we seek to study in the context of three phase unbalanced distribution systems.

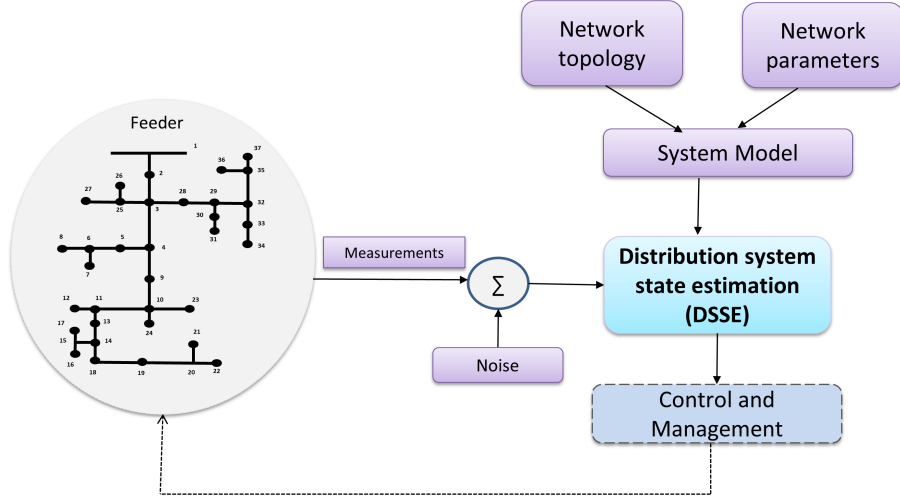


Figure 1.1: *Operation of distribution system state estimation*

Effective situational awareness relies on our ability to analyze and interpret the data provided by monitoring and control systems. In this regard, there has been an increase in the proliferation of advanced metering devices in the smart distribution grid paradigm. According to Energy Information Administration (EIA) data [1], the advanced meter deployments have increased in the United States by approximately 8.3 million meters, representing a 8.8% annual increase in advanced meters from 2019 to 2020. The use of Supervisory control and data acquisition units (SCADA), Phasor Measurement Units (PMUs), Intelligent electronic devices (IEDs), automated feeder switches and voltage regulators, and smart inverters for DERs have provided an opportunity to increase system observability. This increase in sensing results in an increase in the information aggregated at centralized DMS. However, the information aggregated from different sensors presents some important challenges. Firstly, the data from multiple sources are sampled at different rates and are rarely synchronized. The sources of information available at the distribution substations can be broadly classified into two categories: (1) fast rate measurements from RTUs (remote terminal units) sampled at every few seconds, and (2) slow rate measurements from smart meters updated every few

minutes/hours. Secondly, some of the data may be missing or corrupted due to communication network impairments. It is critical to aggregate these heterogeneous measurement sources in order to obtain complete network visibility. Hence, the second fundamental challenge we seek to study is to reconcile multi-timescale measurement time series with noisy, heterogeneous, incomplete and unevenly sampled data that could aid in improving the distribution grid visibility.

Although several methodologies have been continuously developed for addressing the two fundamental challenges (see Chapter 2 for more details), existing methods suffer from various shortcomings, which are categorized into the following five types:

(1) High measurement redundancy requirements: Most of the conventional approaches for state estimation or phase identification approaches require sufficient measurement redundancy to operate. For instance, the conventional weighted least squares estimation approach requires the number of measurements to be higher than the number of states in distribution grid. However, the number of existing telemetered devices that can provide real-time measurements is quite limited, and it is far from being sufficient to provide observability.

(2) Inability to provide uncertainty bounds associated with the estimates: In distribution grids, there are often many sources of uncertainty that can affect the accuracy of state estimates. For example, measurements from sensors can be noisy, and the distribution grid can have complex topologies with varying load conditions. The existing approaches in distribution grid are deterministic in nature, i.e., they do not provide any confidence interval around the estimated values. For instance, multi time-scale imputation approaches in distribution grid, involves simple linear interpolation and exponential moving average methods. The conventional state-of-the-art state estimation approaches also provides point estimates. Uncertainty quantification is an essential tool for improving the accuracy and reliability of state estimation in distribution grids, which is critical for ensuring the safe and efficient operation of the grid.

(3) Poor performance: In order to compensate for the insufficient real-time measurement data, historical data based pseudo-measurements have been used. These pseudo-

measurements obtained from historical load profiles and existing automated meter readings (AMRs) devices have limited accuracy. Hence, the inaccuracy in the measurements results in poor performance of tasks like state estimation, phase identification, etc.

(4) Post-processing methods: The existing DSSE methods deal with bad data and network parameter errors only after the state estimation results are obtained. However, gross errors may still be present after the post-processing stage. For instance, the Chi-squared test is commonly used for detecting bad data in measurement. Unfortunately, this test is not entirely reliable, that is, bad data existing in the measurement set could be missed in certain cases. One reason for this is the approximations used to compute the bad data suspicion threshold, which is set based on an assumed chi-squared distribution for the objective function.

(5) Computational complexity: Grid measurements are also used for outage detection. Existing frameworks for detecting outages in the grid are computationally complex, which hinders the ability to secure the grid as quickly as possible. For instance, the conventional outage detector evaluates the likelihood of the measurements given a discrete set of outage hypotheses. This requires enumerating all possible set of hypothesis, which is computationally expensive.

Based on these perspectives, we seek to address a few fundamental research questions in this dissertation. These questions and the contributions of this dissertation that aim to address them are discussed in the following sections.

1.1 Research Questions

To address the challenges in the distribution system, we ask the following questions:

Question 1: *How can we design distribution system state estimators that are efficient, robust, and practically beneficial for utility companies?*

Question 2: *For enhancing the situational awareness of the grid:*

a: *How can we identify phase labels when there are insufficient measurements and incorrect phase labeling?*

***b:** How can we strategically place sensors to identify outages and develop a low-complexity outage detector for a re-configurable distribution system that can detect all types of outages?*

Question 3: *What methods can be employed to aggregate multi time-scale, missing, and corrupted measurements in a distribution system, to facilitate the state estimation process?*

Question 4: *How can we reliably impute the irregularly sampled measurements even in the presence of aleatoric and epistemic uncertainties inherent in the generative models for the data?*

1.2 Contributions

To address research question 1, Chapter 4 of this dissertation proposes robust formulations of sparsity-based DSSE. Then, we combine the advantages of the two sparsity-based DSSE approach in a single optimization framework to estimate the system states efficiently. The major contributions of these chapters are listed below.

- New robust formulations of different sparsity-based DSSE approaches to deal with bad data and network parameter uncertainty are derived.
- This work proposes an efficient and unique state estimation formulation that combines matrix completion and compressive sensing based approaches in a single powerflow constrained optimization framework.
- The proposed approach exploits the low rank property and compactness of temporal data in the wavelet transform domain. We validate that both these properties (low-rank and DCT compactness) hold true for practical data.
- The proposed methods are validated against conventional approaches in the standard IEEE 37 bus test system. The robust DSSE formulations offer high fidelity as compared to the conventional approaches in the presence of bad data, parameter errors and limited measurements.

More details on the robust and efficient sparsity-based DSSE formulations can be found in Chapters 4 and in the following published articles:

- S. Dahale, H. S. Karimi, K. Lai, and B. Natarajan, “Sparsity Based Approaches for Distribution Grid State Estimation - A Comparative Study,” *IEEE Access*, vol. 8, pp. 198317–198327, 2020, doi: 10.1109/ACCESS.2020.3035378.
- S. Dahale and B. Natarajan, “Joint Matrix Completion and Compressed Sensing for State Estimation in Low-observable Distribution System,” in *2021 IEEE PES Innovative Smart Grid Technologies Conference - Latin America (ISGT Latin America)*, 2021, pp. 1–5. doi: 10.1109/ISGTLatinAmerica52371.2021.9543006.

Novel Bayesian and neural network-based approaches are proposed to address research questions 3 and 4. Chapter 5 of this dissertation focuses on the multi-task Gaussian process based framework to integrate the heterogeneous measurements of the distribution grid. Chapter 6 further enhances the applicability of the Gaussian process to process streaming data. In this chapter, a novel recursive Gaussian process framework to reconcile real-time measurements is developed. The key contributions of these works are as follows:

- Multiple unevenly sampled time-series data is imputed using multitask Gaussian process by exploiting the underlying spatio-temporal correlations. This approach effectively imputes the time-series data in the presence of missing measurements.
- A hierarchical sparse Bayesian matrix completion approach is proposed for DSSE which leverages the multi time-scale measurements along with its uncertainty. By augmenting the sparse Bayesian matrix completion with linearized power-flow constraints, the proposed method can effectively estimate the states under low observability conditions where standard weighted least squares approaches fail.
- A novel recursive Gaussian process based approach is proposed which is computationally efficient and is flexible to allow for both batch-wise and real time processing of measurements.

- We leverage the graphical structure of the network in the recursive multi-task Gaussian process approach. We prove that exploiting the graph structure of the distribution system leads to a decrease in the variance of the imputed measurements.
- Simulations are carried out on three phase unbalanced IEEE 37, IEEE 123 and 11,000 bus test systems to verify the efficacy of the proposed Bayesian frameworks.
- The underlying computational complexities of all the proposed approaches are also quantified.

More details related to the proposed multi-task GP and the recursive GP based approaches can be found in Chapter 5 and Chapter 6, and in the following articles:

- S. Dahale and B. Natarajan, “Multi Time-scale Imputation aided State Estimation in Distribution System,” in *2021 IEEE Power Energy Society General Meeting (PESGM)*, 2021, pp. 1–5. doi: 10.1109/PESGM46819.2021.9637824.
- S. Dahale and B. Natarajan, “Bayesian Framework for Multi-Timescale State Estimation in Low-Observable Distribution Systems,” *IEEE Trans. Power Syst.*, vol. 37, no. 6, pp. 4340–4351, Nov. 2022, doi: 10.1109/TPWRS.2022.3155151.
- S. Dahale and B. Natarajan, “Recursive Gaussian Process over graphs for Integrating Multi-timescale Measurements in Low-Observable Distribution Systems,” *IEEE Trans. Power Syst.*, pp. 1–12, 2022, doi: 10.1109/TPWRS.2022.3204415
- B. Rout, S. Dahale, and B. Natarajan, “Dynamic Matrix Completion Based State Estimation in Distribution Grids,” *IEEE Trans. Ind. Inform.*, vol. 18, no. 11, pp. 7504–7511, Nov. 2022, doi: 10.1109/TII.2022.3162210.
- M. Abujuubbeh, S. Dahale, and B. Natarajan, “Voltage Violation Prediction in Unobservable Distribution Systems,” in *2022 IEEE Power & Energy Society General Meeting (PESGM)*, Jul. 2022, pp. 1–5. doi: 10.1109/PESGM48719.2022.9916805.

To address research question 4 and further improve the scalability of the GP-based approaches, Chapter 7 develops two neural network-based approaches. These approaches are based on the principles of neural differential equations. The first approach is a generative neural ordinary differential equation (ODE) model while the second approach is auto-regressive in nature, termed as SDE-RNN model. The key contributions of these works are as follows:

- We propose a latent neural ordinary differential equations (LODE) approach to reconcile the multi time-scale measurements in a distribution grid. The proposed approach is capable in performing both imputations and predictions, while being computationally efficient.
- We formulate a novel SDE-RNN which combines the principle of Stochastic differential equations and neural networks to model the irregularly sampled time series and provide a confidence interval around them.
- Derive analytical expressions for quantifying total uncertainty due to aleatoric and epistemic sources in a neural SDE-RNN model.
- Simulation results on the power distribution test system (IEEE 37 network) demonstrate the effectiveness of the proposed approach compared to the state-of-the-art approaches.

More details related to the Latent ODE and SDE-RNN model can be found in Chapter 7 and in the following articles:

- S.Dahale, S. Munikoti, B. Natarajan, R. Yang, “Latent Neural ODE for Integrating Multi-timescale measurements in Smart Distribution Grids,” in *2023 IEEE PES Innovative Smart Grid Technologies Conference - North America (ISGT North America)*, 2023 (manuscript in press).
- S.Dahale, S. Munikoti, and B. Natarajan, “A General Framework for Uncertainty Quantification in Neural SDE-RNN”, (manuscript to be submitted).

To address research question 2a, we propose a phase identification approach where phases are insufficiently labeled and measurements are limited. We use graph signal processing techniques in the optimization formulation to recover the system topology. To address research question 2b, we develop a convex meter placement strategy for outage identification in a reconfigurable distribution network. The placement problem is complemented with a novel outage detection framework that leverages the strategically deployed sensors' measurements at multiple time snapshots and the load estimates via advanced metering infrastructure (AMI) or load forecasting mechanisms. The key contributions of Chapter 8 are as follows:

- The proposed phase identification approach utilizes the nodal voltage magnitude measurements for clustering the graph representing the phase connectivity information. It consists of an optimization formulation that infers the connectivity of the entire multi-phase distribution network in a single shot.
- Simulations are performed on the unbalanced three-phase IEEE 37 and 123 bus test system. Results demonstrate that the proposed approach is insensitive to incorrect phase labels, missing measurements, and unknown phase labels. Results illustrate that the novel approach is over 90% accurate in phase identification and outperforms the state-of-the-art approach by over 70%.
- The novel outage detection approach provides a one-step solution, and it does not require enumerating over all discrete sets of hypotheses. The outage detection algorithm is effective in identifying single and multi-edge outages within the network topology configurations assumed for the meter placement problem.

More details related to the phase identification and meter placement approach can be found in Chapter 8 and in the following patent disclosure and articles:

- S. Dahale, "Method for Determination of Phase Labels in a Three Phase Electric Power Distribution Network", US Patent 63/347,643, June 1, 2022.
- S. Dahale, A. Pahwa, and B. Natarajan, "Phase Identification in Unobservable Distribution Systems," *IEEE Trans. Power Deliv.*, 2023 (In press).

- S. Dahale, and B. Natarajan. “Sensor Placement and Outage Detection for Reconfigurable Distribution System.”, *IEEE Trans. Power Syst.*, (under review).

1.3 Organization of the Dissertation

Figure 1.2 outlines the structure of this dissertation. The remainder of this dissertation is organized as follows. Chapter 2 describes the related work and their limitations in addressing the challenges of a real distribution grid. Chapter 3 discusses the fundamental idea of weighted least-squares and sparsity-based state estimation in three phase distribution grids. In chapter 4, we exploit the different optimization formulations for implementing a robust DSSE. Then, we develop joint formulations of the sparsity-based DSSE methods. In chapter 5, we develop a Bayesian framework for integrating the multi time-scale measurements in smart distribution grid. In chapter 6, we develop a novel recursive GP framework for handling the streaming data. Then, chapter 7 provides two approaches which leverages the principles of neural differential equations for reconciling the multi time-scale measurements and quantifying the uncertainties. Chapter 8 provides phase identification, meter placement and outage detection approaches. The summary of this dissertation and future ideas constitute chapter 9.

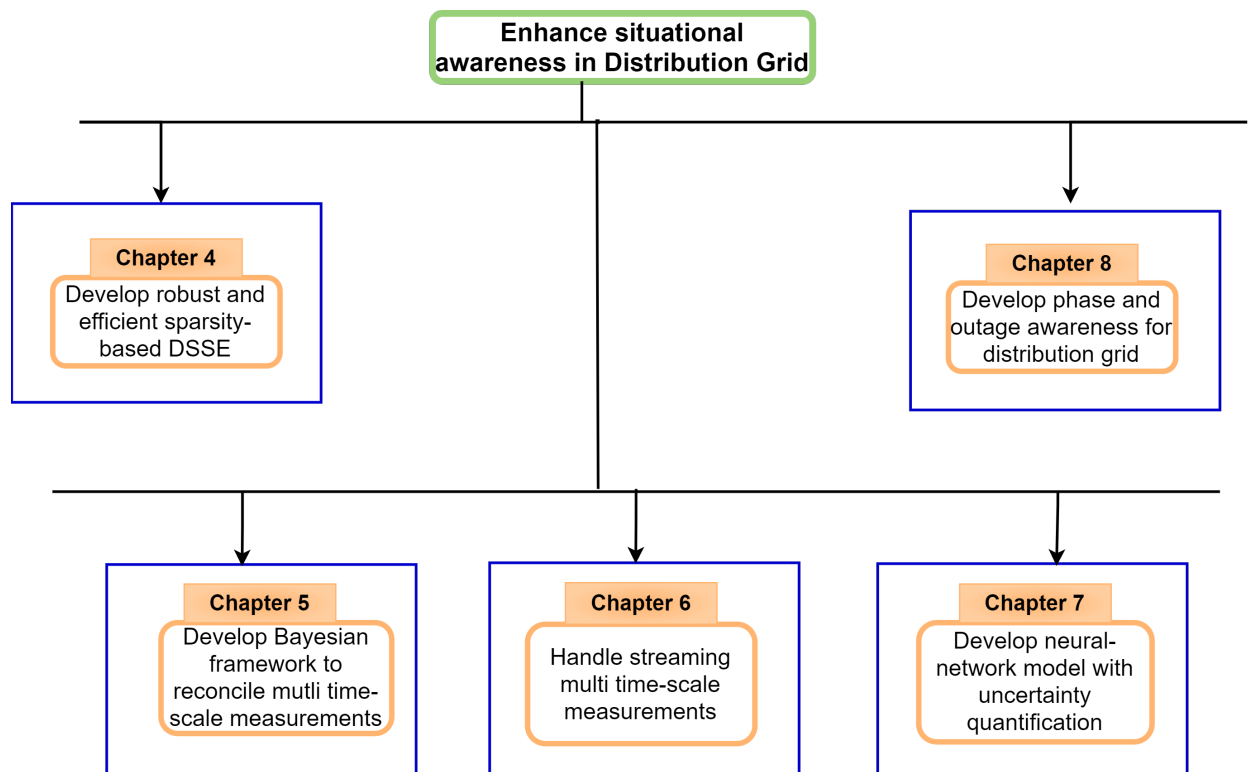


Figure 1.2: *Outline of the dissertation*

Chapter 2

Literature review

2.1 Distribution system state estimation

Weighted Least Squares (WLS) has been the traditional approach for DSSE. In order to guarantee full observability for the WLS-based state estimation, historical data based pseudo-measurements have been used to artificially compensate for insufficient data [2]. However, the WLS + pseudo-measurement state estimation paradigm suffers from huge data requirements [2] and poor estimation performance [3].

Recently the challenges posed by unobservability and limited measurement availability at the grid edge have been addressed by sparsity-based DSSE methods. All these methods exploit the underlying smoothness or sparsity of the raw or linearly transformed measurements/system states. These methods exploit the network structure to perform state estimation at current levels of data availability and observability. Thus, the requirement of creating pseudo-measurements for ensuring observability is eliminated. Compressive Sensing (CS) based DSSE was the first class of solutions proposed where the sparsity of measured data in a linear transformation basis was exploited to compress measurements. The spatio-temporal correlation between loads and distributed generation in a single-phase distribution system is exploited in the estimation strategies proposed in [4]. A noiseless CS approach is applied to reconstruct the real and reactive power measurements across the grid and estimate

the voltage magnitude and angle. [5] extends this DSSE approach to three-phase distribution system. A dynamic compression scheme is employed in [6] on three publicly available datasets for spatial and spatio-temporal compression. Recursive dynamic state estimation is implemented in [7] by exploiting the sparsity in distribution grid data. An efficient dynamic solution for online smart grid topology identification using CS is presented in [8]. Recently, [9] and [10] introduced the idea of implementing DSSE based on matrix and tensor completion algorithms, respectively. Both these methods utilize the sparsity or smoothness of raw measurements. Sparsity in CS indicates the signal of interest is sparse in a specific transform domain. However, in matrix/tensor completion, it indicates that the singular vector of the original matrix/tensor is sparse [11]. Matrix/tensor completion methods impute missing elements in a matrix/tensor by obtaining a suitable low rank approximation of the incomplete matrix/tensor. The matrix completion algorithm exploits the underlying spatial correlation in the data. Recently in [9], matrix completion along with noise-resilient power flow constraints was employed to estimate states in a distribution grid. A matrix completion formulation that is robust to bad data is presented in [12]. An algorithm for dealing with data loss in PMU-based power systems using matrix completion and compressive sensing is proposed in [13] and [14] respectively. Apart from exploiting spatial correlation, the existence of inherent spatio-temporal correlation in states and measurements can be leveraged using tensor completion algorithms. Tensor completion based approaches for DSSE estimation is proposed in [10] and is demonstrated to provide accurate state estimation in low-observable systems. Using low-rank canonical polyadic decomposition, [15] presents a model-free state estimation and energy forecasting framework for distribution systems.

Given that sparsity-based DSSE methods work with limited data, it is important to understand the impact of bad data within this limited set. Conventionally, statistical tests, such as the χ^2 -test and the largest normalized residual test are employed for bad data detection and identification, respectively [16]. Both these tests are post-SE processing techniques which rely on least-squares estimated residuals. Least absolute value (LAV) [17] [18], least median of squares (LMS) and least trimmed squares (LTS) [19] estimators have also been used to detect the bad measurements. However, these methods have huge computational

cost and high measurement redundancy requirements, which limit their applications in distribution grids with low observability. Therefore, there is a need to integrate and understand the effect of bad data in sparsity-based DSSE methods.

It is important to note that DSSE assumes perfect knowledge of the network parameters. However, the underlying network parameters may be erroneous because of inaccurate manufacturing data, human data entry error, unreported device upgrade or ambient/operating condition variations [20]. Several approaches for parameter error identification have been developed, which are mainly based on residual sensitivity analysis [21] and state vector augmentation [22]. However, these methods are either post-processing methods or run into the observability problem with the increasing scale of the system. The impact of network parameter uncertainties on sparsity-based DSSE methods is currently unknown.

2.2 Integrating multi time-scale measurements

Weighted least squares (WLS) has been the traditional approach for distribution system state estimation (DSSE). In order to guarantee full observability for the WLS-based state estimation, historical data based pseudo-measurements have been used to artificially compensate for insufficient data. These pseudo-measurements do not capture the real-time spatio-temporal correlations underlying the stochastic environment (e.g., a distribution grid with rooftop PV's). In [23], the WLS estimation exploits simple linear interpolation and extrapolation technique to reconcile time-series data collected at two different time-scales. However, this interpolation approach does not exploit the spatio-temporal relationships across the network to impute the time-series data. Secondly, the approach is simplistic and limited to two time-scale data. Thirdly, the imputed time-series data is still limited only to specific spatial locations in the network where measurements are collected. In practice, there may not be adequate number of measurements to guarantee observability. Hence, the approach in [23] is limited in its practical applicability to distribution systems. The issue of non-synchronized measurements in DSSE is addressed in [24] based on the credibility of each available measurement and appropriately adjusting the variance of measurements. The authors in [25]

propose an extended Kalman filter based approach to deal with the irregular sensor sampling and random communication delays. The problem of incorporating limited real-time measurements as well as the delayed measurements from smart meters for DSSE was addressed in [26]. A Bayesian estimator proposed in [27] exploits the probabilistic dependencies of the measurement variables on the system states. However, in this approach, it is difficult to know the joint probability distribution apriori. Authors in [28] use PMU and SCADA measurements for DSSE. This approach performs DSSE by incorporating a subset of these measurements available at time t along with the predicted SCADA measurements obtained using the information from the previous state estimates. It suffers from large measurement redundancy requirements (around 1.7), which makes it impractical for low-observable distribution systems. Furthermore, [28] does not consider any missing measurements scenario that could occur while aggregating measurements over finite bandwidth communication networks. A load evolution model for the slow-rate measurements is proposed in [29] for performing day-ahead forecasting. This approach relies on recursive Kalman filter (KF) updates for dynamic DSSE. However, KF typically needs the Hessian inverse computations at every step, which can be computationally burdensome. Also, the approach is not demonstrated for unbalanced systems. A first-order prediction-correction approach using PMU and smart meter data is performed in [30]. The main limitation of the approaches in [29] and [30] approaches is that they assume smart meter measurements are available at all load bus (i.e., system is fully observable). Neural network models are a popular technique for integrating irregularly sampled measurements. Recurrent neural networks (RNN) form the first choice for modeling high dimensional, regularly sampled time series data. The success of RNNs is due to its memory and cell state modules that can capture long range dependencies. A natural extension of RNNs to unevenly sampled time series data is to divide the timeline into equally-sized intervals [31], [32], and impute or aggregate observations using averages. This pre-processing stage maligns the information, particularly about the timing of measurements, as they constitute a piece of important information about the data. Another approach uses a simple exponential decay between observations, and updates the hidden states accordingly [32]. But the states of RNN-decay model approach zero if the time gaps between observations are

high. However, Neural ODE, which is a novel framework combining deep learning principles and differential equations has been found to be suitable for modeling irregularly sampled time series [33]. Neural ODEs can systematically accommodate continuous time series data using a technique that captures the continuous nature of hidden states. Recently, [34] combines neural ODEs with RNN, leading to a method referred to as ODE-RNN. The hidden states in these ODE-RNN models obey an ODE between consecutive observations and are only updated at observed time instances. The ODE-RNN models are suitable for sparse observations and thus very effective for imputation and one-step prediction. Authors in [35] propose a Latent SDE with an adjoint method for continuous time series modeling. Neural SDEs driven by stochastic processes with jumps is proposed in [36] to learn continuous and discrete dynamic behavior. Generative models built from SDEs whose drift and diffusion coefficients are samples from a Gaussian process are introduced in [37]. A neural Jump ODE framework for modeling the conditional expectation of a stochastic process in an irregularly observed time series is proposed in [38]. However, all neural ODE driven methodologies only offer a point estimate and fail to quantify the uncertainty associated with the estimates.

UQ approaches in general neural networks are mainly categorized into two classes: Bayesian and Ensemble methods. Bayesian approaches, such as Bayesian neural nets proposed in [39] quantify the uncertainty by imposing probability distributions over model parameters. Though, these approaches provide a principled way of uncertainty quantification, the exact inference of the parameter posteriors is intractable. Also, specifying the priors for the parameters of deep neural networks becomes challenging when size of the network increases. To deal with these challenges, approximation methods like variational inference [40], Laplace approximation [41], Assumed density filtering [42] and stochastic gradient MCMC [43] are used. Authors in [44] proposed to use Monte-Carlo Dropout (MC-dropout) during inference to quantify the uncertainty in neural networks. However, MC dropout is computationally expensive for larger network and only capture epistemic uncertainty. Non-Bayesian approaches have also been used for UQ but they also demand a large computational effort [45]. A different route has been taken in [46] where SDEs are used to quantify both aleatoric and epistemic uncertainty by training the SDE model based on out-of-distribution (OOD) training data.

However, this approach fails to evaluate uncertainty in a principled manner.

2.3 Phase Identification

Existing phase identification algorithms presented in the literature fall under three categories: approaches that (1) employ additional hardware; (2) primarily use power measurements, and (3) use voltage measurements. The hardware-based methods [47] - [48] employ signal-injection methodology where distinct signals are injected in each phase of the customers. These methods are expensive as it requires extensive work by field personnel. Power measurement-based approaches leverage the fact that the sum of all the customer loads on a particular phase is equal to the power measured at the substation transformer, less the losses [49]. Authors in [50] propose a novel phase and topology identification based on the load summing concept, principal component analysis, and graph theory. A tabu-search method is used to assign lateral phases in [51] while [52] performs spectral saliency analysis on load profiles. The disadvantage of the load summing approach is the need to model the technical losses, clock synchronization errors, and other losses in the smart meter readings. This results in inaccurate phase identification in real distribution systems when the time-series measurements per node are less than the number of total consumers in the network. The power-based measurement approaches are accurate in a case where the measurements are assumed noiseless. However, measurement errors due to synchronization issues, clock skew, and line losses deteriorate the performance of these approaches. To overcome these limitations, the availability of these noisy measurements needs to be increased to identify phase labels. For acceptable performance, the measurement redundancy required grows to three times the total number of customers. Thus, these approaches are not scalable with the size of the network and the number of customers. Furthermore, these power-based approaches are sensitive to unmetered loads. Another popular phase-identification method employs voltage-based approaches [53]. These approaches leverage the voltage time-series measurements and are based on the idea that customers on the same phase will have voltage times-series that are more correlated than the customers on different phases. Supervised Machine learning

techniques proposed in [54] use a subset of correct phase labels to train the model. Cross-correlation between customer voltage profiles and known voltage profiles for each phase is performed in [55]. A constrained multi-tree approach proposed in [56] incorporates known topology information as constraints to the phase identification algorithm. Spectral clustering using voltage time-series data to identify phase labels is employed in [57], [58]. However, this approach requires extensive training data (about a one-year duration), which is impractical as the results will not reflect the frequently occurring dynamic changes in the distribution system.

The phase identification approaches discussed earlier uses historical-based voltage/power time-series measurements ranging from months to a year. Such large data requirements are impractical to meet as (1) the distribution grid is generally unobservable and data is limited; and (2) ground-truth data amidst changes in grid topology may be unavailable. Therefore, these approaches may prove to be inaccurate under dynamic grid changes. The main challenge of the phase identification approaches is to provide high accuracy even with the limited measurements available in the distribution grid.

2.4 Sensor placement and outage detection

Sensor placement in distribution systems is motivated by various applications such as theft detection [59], communication infrastructure cost minimization [60], outage detection [61], observability enhancement [62, 63] and state estimation [64, 65]. However, meter placement is often challenging due to the conflicting nature of sensor placement objectives. Several sensor placement approaches have been proposed to optimally place sensors to achieve one or more of the following objectives: namely, minimizing the number of sensors installed, minimizing the cost of sensors [64],[66], optimizing performance under a resource constraint [67], minimizing state estimation errors [68], [69] and improving observability. A meter placement scheme using heuristic rules to reduce the variance of voltage magnitude and angle estimates at non-metered nodes is proposed in [70]. As meters are placed using a heuristic approach, they may not guarantee an optimal solution. Authors in [71] minimize the number of meter

locations and maximize the information from each meter location. A fisher information-based measurement placement strategy is proposed in [72]. These approaches focus on the meter placement problem to obtain low estimation errors.

Authors in [73] formulate a mixed-integer linear program for placing the PMUs to ensure system observability with minimum load loss during a single branch outage. However, the sensor placement approaches consider a limited set of outages. An optimal PMU strategy for ensuring observability is proposed in [74], in which the authors also consider the scenario of single-line outages and PMU outages. However, this approach assumes predefined bus locations where outages are likely to occur. In [75] and [76], the authors proposed an outage detection approach by combining the optimally deployed real-time sensors and load estimates obtained via AMI sensors or a forecasting mechanism. They formulated outage detection as a hypothesis testing problem, for which the maximum a-posteriori probability (MAP) detector is applied, and maximum misdetection probability is used as the performance metric for placing sensors at the optimal locations. An optimal meter placement strategy for outage identification for a radial distribution network is presented in [77]. The meter placement approaches proposed in [75, 76, 77] deal with a single distribution network configuration. Furthermore, the outage detector in these approaches evaluates the likelihood of the measurements given a discrete set of outage hypotheses and chooses the hypothesis which provides the maximum likelihood. Thus, the outage detector requires enumerating a possible set of outage hypotheses, which is computationally expensive and hinders the goal of detecting all possible outages as quickly as possible.

Previous research that considers reconfigurable distribution networks primarily focusses on meter placement to ensure observability and not on outage detection. Authors in [68] propose a robust meter placement algorithm to optimize the worst-case estimation accuracy among all possible system configurations. In [78], PMU placement via greedy search in reconfigurable DN is considered. However, this approach places PMU at every node to avoid losing observability in reconfiguring the DN. [79] investigates PMU placement in a multi-configuration DN by increasing the node reliability. This approach places two PMUs at each node to make it observable. In [80], an integer linear programming formulation is proposed

for reconfigurable DN, which uses a union operator among different configurations as the guide to final meter placement. However, meter placement for outage identification in a reconfigurable DN is still not investigated. This paper addresses this limitation by developing a convex meter placement approach for reconfigurable DN. Furthermore, a computationally efficient outage detector is required, which avoids enumerating all possible sets of hypotheses for detecting the most likely outage scenario.

Chapter 3

Background

The power distribution system is typically unobservable due to insufficient real-time measurements. While deploying more sensors can alleviate this issue, it also presents new challenges related to data aggregation and the underlying communication infrastructure. Therefore, developing state estimation methods that enhance situational awareness at the grid edge with compressed measurements is critical. In this chapter, we focus on the different state estimation approaches used in distribution grids.

3.1 State estimation techniques in distribution grid

Distribution system state estimation (DSSE) refers to the procedure of obtaining the correct voltage phasors at all nodes based on real-time measurements and the power network model. Consider a power distribution system with $|\mathcal{P}|$ three phase non-slack buses. The measurements in a distribution system includes power injections at each bus, power flow, and current over each distribution line, along with bus voltages which are related to the states (voltage magnitude and voltage angle) by the non-linear equation (3.1),

$$\mathbf{z} = h(\mathbf{x}) + \boldsymbol{\epsilon} \tag{3.1}$$

where, $\mathbf{z} \in \mathbb{C}^m$ represents the vector of measurements, $\mathbf{x} \in \mathbb{C}^n$ denotes the vector of states, and the function $h(\mathbf{x})$ capture the relationship between states and measurements. $\boldsymbol{\epsilon}$ represents the measurement noise vector. The state estimation task is to estimate $\mathbf{x}$ based on $\mathbf{z}$ and knowledge of $h(\cdot)$. This non-linear function $h(\cdot)$ relating the voltage states $\mathbf{x}$ and measurements $\mathbf{z}$ can be linearized around an operating point as in [81] resulting in,

$$\mathbf{x} = \mathbf{B}\mathbf{z} + \mathbf{w} \quad (3.2)$$

$$|\mathbf{x}| = \mathbf{C}\mathbf{z} + |\mathbf{w}|, \quad (3.3)$$

where $\mathbf{z} = ((\mathbf{P})^T, (\mathbf{Q})^T)^T$ represents the measurement vector of active and reactive power injections, $|\mathbf{x}|$ represents the magnitudes at all the buses.

$$\mathbf{B} = \left(\mathbf{Y}_{\text{LL}}^{-1} \text{diag}(\bar{\mathbf{w}})^{-1}, -j \mathbf{Y}_{\text{LL}}^{-1} \text{diag}(\bar{\mathbf{w}})^{-1} \right) \text{ and,}$$

$$\mathbf{C} = \left(|\text{diag}(\bar{\mathbf{w}})|^{-1} \Re(\text{diag}(\bar{\mathbf{w}}) \mathbf{M}) \right),$$

$$\mathbf{w} = -\mathbf{Y}_{\text{LL}}^{-1} \mathbf{Y}_{\text{L0}} \mathbf{v}_0 \text{ is the zero-load voltage}$$

where $\mathbf{v}_0$ denotes the complex vectors collecting the three-phase nodal voltage at the slack bus. Here, $\mathbf{Y}_{\text{LL}} \in \mathbb{C}^{|\mathcal{P}| \times |\mathcal{P}|}$ and $\mathbf{Y}_{\text{L0}} \in \mathbb{C}^{|\mathcal{P}| \times 3}$ are the submatrices of the three-phase admittance matrix,

$$\mathbf{Y} = \begin{bmatrix} \mathbf{Y}_{00} & \mathbf{Y}_{0\text{L}} \\ \mathbf{Y}_{\text{L0}} & \mathbf{Y}_{\text{LL}} \end{bmatrix} \in \mathbb{C}^{(|\mathcal{P}|+3) \times (|\mathcal{P}|+3)} \quad (3.4)$$

Next, we will discuss the weighted least squares (WLS) approach conventionally used for state estimation in the distribution grid.

3.1.1 WLS method

In a traditional DSSE approach, the state vector is estimated from the measurement equation in (3.1) using the weighted least squares method. In particular, the DSSE problem is solved by finding

$$\hat{\mathbf{x}} = \min_{\mathbf{x}} [\mathbf{z} - h(\mathbf{x})]^\top \mathbf{W}^{-1} [\mathbf{z} - h(\mathbf{x})] \quad (3.5)$$

where $\hat{\mathbf{x}}$ is the estimated state vector, $\top$ is the matrix transposition operation, and $\mathbf{W}$ denotes the weight matrix that represents the user's confidence in the measured data. The weighting matrix $\mathbf{W}$ is commonly taken as diagonal elements related to background noise covariance as $\mathbf{W} = \text{diag}\{\sigma_1^{-2}, \dots, \sigma_m^{-2}\}$, where σ_j^2 represents the variance of the measurement error corresponding to the j^{th} element of $\mathbf{z}$. This choice of the weight matrix is based on two assumptions: 1) the error vector (ϵ) has a Gaussian distribution with zero mean, and 2) the measurement errors of different elements of the measurement vector are statistically independent. Under these assumptions, the WLS problem is transformed to the maximum likelihood problem.

The solution for $\hat{\mathbf{x}}$ is obtained in an iterative fashion by linearizing (3.1) around the available estimate (at iteration k) and applying the Gauss-Newton algorithm to improve the estimate, using the following equations:

$$\mathbf{G}(j)\Delta\mathbf{x}(j) = \mathbf{H}^\top(j)\mathbf{W}^{-1}[\mathbf{z} - h(\mathbf{x}(j))] \quad (3.6)$$

$$\hat{\mathbf{x}}(j+1) = \hat{\mathbf{x}}(j) + \Delta\mathbf{x}(j) \quad (3.7)$$

where $\mathbf{G}(j) = \mathbf{H}^\top(j)\mathbf{W}^{-1}\mathbf{H}(j)$ is the gain matrix at iteration j . The Jacobian matrix $\mathbf{H}(j)$ needed at each iteration is the first-order derivative of $h(\mathbf{x})$, with respect to $\mathbf{x}$, evaluated at $\hat{\mathbf{x}}(j)$, i.e., $\mathbf{H}(j) = \{\frac{\partial h(\mathbf{x})}{\partial \mathbf{x}}\}_{\mathbf{x}=\hat{\mathbf{x}}}$. The iterative procedure is terminated when the norm of the residual falls below a predefined value, i.e., for some $\delta > 0$, $\|\mathbf{z} - h(\hat{\mathbf{x}}(j))\|^2 \leq \delta$.

According to the update equation in (3.6), an invertible Jacobian matrix is required to implement WLS. However, this condition cannot be satisfied due to insufficient measure-

ments at the grid edge. In other words, the low observability in distribution grids hinders the applicability of the conventional WLS method for DSSE. To address the low observability issue in DSSE, the correlations among states and measurements are exploited. Specifically, a spatial correlation exists among DERs based on their geographical proximity [4]. Moreover, power injection/consumption patterns of nodes may exhibit temporal correlation. The presence of spatial and/or temporal correlation results in the sparsity of states/measurements in a linear transformation basis. Thus, very few measurements from network buses are required for state estimation. The sparsity characteristics can be exploited by the following DSSE methods.

3.1.2 Compressive Sensing

1-D Compressive Sensing

This technique exploits the sparsity of the signal of interest in a linear transformation basis. It is used to efficiently reconstruct a signal by finding sparse solutions to an under-determined linear system. Let $\mathbf{x} \in \mathbb{R}^n$ be the original state, compressible in a linear transformation basis such that,

$$\mathbf{x} = \Psi \mathbf{a} \quad (3.8)$$

where $\mathbf{a}$ has at most $k \ll n$ significant coefficients i.e., $\mathbf{x}$ is k -sparse in sparsifying basis Ψ . If the sensing mechanism is such that,

$$\mathbf{y} = \Phi \mathbf{x}; \mathbf{y} \in \mathbb{R}^m, \Phi \in \mathbb{R}^{m \times n}, \quad (3.9)$$

where Φ is a random measurement/projection matrix (e.g., matrix elements distributed as i.i.d. Gaussian random variable with mean 0 and variance $1/m$ or Bernoulli random variables), then the original state $\mathbf{x}$ can be reconstructed by solving the following l_1 minimization problem

$$\begin{aligned}
\hat{\mathbf{a}} &= \min_{\mathbf{s}} \|\mathbf{s}\|_1 \\
&\text{subject to } \mathbf{y} = \Phi\Psi\mathbf{s} \\
&\text{(3.2), (3.3)}
\end{aligned} \tag{3.10}$$

Here, (3.2) represents the linearized power-flow constraints. This formulation can be modified to include noisy measurements as shown in [82]. The reconstructed state is

$$\hat{\mathbf{x}} = \Psi\hat{\mathbf{a}} \tag{3.11}$$

The result of the optimization problem in (5.2.1) provides an exact reconstruction with overwhelming probability [82] if there exists a $\delta \in (0, 1)$ such that,

$$(1 - \delta)\|\mathbf{s}\|_2^2 \leq \|\Phi\Psi\mathbf{s}\|_2^2 \leq (1 + \delta)\|\mathbf{s}\|_2^2, \tag{3.12}$$

holds for all K -sparse signal $\mathbf{s}$. This is called the Restricted Isometry property (RIP) of order K . The ratio M/N is termed as Compressed Measurement Ratio (CMR). More details regarding 1-D CS-based DSSE can be found in [4].

2-D Compressive Sensing

2-D CS aims at reconstructing the signal with insufficient data using the underlying sparsity across both space and time. Let $\mathbf{X} \in \mathbb{R}^{N_{space} \times N_{time}}$ be the spatio-temporal state over N_{space} nodes and for N_{time} number of observations. It has been shown that $\mathbf{X}$ is sparse in sparsifying basis $\Psi_{N_{space}}$ and $\Psi_{N_{time}}$ such that [4],

$$\mathbf{X} = \Psi_{N_{space}} \mathbf{A} \Psi_{N_{time}}^T \tag{3.13}$$

The spatio-temporal compressed sensing of $\mathbf{X}$ corresponds to,

$$\mathbf{Y} = \Phi_{space} \mathbf{X} \Phi_{time}^T \tag{3.14}$$

where, $\Phi_{space} \in \mathbb{R}^{m_{space} \times N_{space}}$ and $\Phi_{time} \in \mathbb{R}^{m_{time} \times N_{time}}$ with entries that are i.i.d. Gaussian random variables with zero mean and respective variance of $1/m_{space}^2$ and $1/m_{time}^2$ with $m_{space} \ll N_{space}$ and $m_{time} \ll N_{time}$. The spatio-temporal data is recovered by solving the following l_1 minimization problem

$$\begin{aligned} \hat{\mathbf{a}} = \min_{\mathbf{s}} \quad & \|\mathbf{s}\|_1 \\ \text{subject to} \quad & \text{vec}(\mathbf{Y}) = (\Phi_{space} \otimes \Phi_{time})(\Psi_{N_{space}} \otimes \Psi_{N_{time}})\mathbf{s} \end{aligned} \quad (3.15)$$

(3.2), (3.3)

Here, $\text{vec}(\cdot)$ represents the vectorized version of a matrix and $\otimes$ represents the kronecker product. The reconstructed state is therefore,

$$\text{vec}(\hat{\mathbf{X}}) = (\Psi_{N_{space}} \otimes \Psi_{N_{time}})\hat{\mathbf{a}} \quad (3.16)$$

More details on 2-D CS based DSSE can be found in [4].

3.1.3 Matrix completion

To utilize the matrix completion approach for DSSE, a structured matrix $\mathbf{X}$ whose rows correspond to measurement locations and columns correspond to measurement types (e.g. power or voltage) is constructed for a given time. For example, each row represents a bus and each column represents a measurement associated with the bus [9]. Therefore, for every bus $b \in \mathcal{P}$, the corresponding row in the matrix $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2}$ contains:

$$[\Re(v_b), \Im(v_b), |v_b|, \Re(s_b), \Im(s_b)] \quad (3.17)$$

where $n_1 = |\mathcal{P}|$ and $n_2 = 5$ quantities per row. Since the distribution grid is unobservable, $\mathbf{X}$ will be an incomplete matrix. The smoothness in the spatial variation of the physical power system quantities (voltage, power, etc) translates into the low-rank property for this matrix $\mathbf{X}$. This low-rank property can be exploited to impute the unknown entries of $\mathbf{X}$.

Specifically, matrix completion aims to determine the unknown elements in the matrix by minimizing the rank of the matrix. The convex relaxation for the rank of a matrix is the nuclear norm. The sampling operator $P_\Omega(\mathbf{X})$ represents the observation matrix as,

$$[P_\Omega(\mathbf{X})]_{ij} = \begin{cases} \mathbf{X}_{ij}, & (ij) \in \Omega \\ 0, & \text{otherwise} \end{cases} \quad (3.18)$$

Therefore, the optimization formulation (3.19) recovers the complete low rank matrix, as

$$\begin{aligned} \hat{\mathbf{L}} = \min_{\mathbf{L} \in \mathbb{R}^{n_1 \times n_2}} \quad & \|\mathbf{L}\|_* \\ \text{subject to} \quad & \|P_\Omega(\mathbf{X}) - P_\Omega(\mathbf{L})\|_2^2 \leq \epsilon \end{aligned} \quad (3.19)$$

(3.2), (3.3)

where, (3.2) and (3.3) captures the linearized power-flow constraint relating voltage states $\mathbf{x}$ to power measurements $\mathbf{z}$. More information on matrix completion-based DSSE can be found in [9].

3.1.4 Tensor completion

Similar to matrix completion, the starting point for tensor completion approach is the construction of a tensor corresponding to system states and measurements. In essence, the matrix corresponding to the state/measurement at one time instant can be extended to a sequence of matrices each corresponding to one time instant. The resulting tensor $\mathcal{T}$ will be of dimension $\mathbb{R}^{\mathcal{P} \times 5 \times N_t}$ with $|\mathcal{P}|$ buses, N_t time instants and 5 physical variables as stated in (3.17). $\mathcal{T}$ is a 3-way tensor indexed by two spatial variables and one temporal variable.

The goal of tensor completion is to fill in the missing entries of $\mathcal{T}$ by exploiting the sparsity in the data. Tensor completion utilizes tensor trace norm minimization formulations with linearized power-flow constraints. The tensor trace norm can also be expressed as the convex combination of trace norms of all matrices obtained by unfolding the tensor along all its modes [83]. Additional matrices $\mathcal{M}_1, \mathcal{M}_2, \dots, \mathcal{M}_n$ are introduced to eliminate the

interdependency in the elements across the different unfoldings. In [10], a simple low rank tensor completion (SiLRTC) approach is considered that uses block coordinate descent to obtain the n matrices $\mathcal{M}_i$ and tensor $\mathcal{X}$. The suffix (i) denotes the unfolding operation applied on the tensor along the mode i . The optimal M_i can be obtained by applying a shrinkage operator $D_{\mathcal{T}}(\mathcal{X}_{(i)})$, where $D_{\mathcal{T}}(\mathcal{X}_{(i)}) = U\Sigma_{\mathcal{T}}V^T$ and $\Sigma_{\mathcal{T}} = \text{diag}(\max(\sigma_j - \tau, 0))$. $\mathcal{X}$ is obtained by solving the optimization problem for certain positive values of β_i s,

$$\begin{aligned} \min_{\mathcal{X}} \quad & \sum_{i=1}^n \frac{\beta_i}{2} \|\mathcal{M}_i - \mathcal{X}_{(i)}\|_F^2 \\ \text{subject to} \quad & \mathcal{X}_{\Omega} = \mathcal{T}_{\Omega} \end{aligned} \tag{3.20}$$

(3.2), (3.3)

The $\mathcal{M}_i$'s and $\mathcal{X}$ are alternatively updated until $\mathcal{X}$ converges. More details on tensor completion based DSSE can be found in [10].

Chapter 4

Robust and Efficient Sparsity-based state estimators

All the sparsity-based approaches discussed in chapter 3 do not take into account the presence of outlier measurements and network parameter uncertainty. The presence of bad measurement data can degrade state estimation accuracy [84]. Furthermore, all the DSSE methods discussed in chapter 3 assumes accurate knowledge of the power network. The network model parameters include series resistances, reactances, tap values, and susceptance. Parameter errors may create serious bias in the state estimate solutions that tend to last for a long time [85]. Furthermore, efficient formulations of these DSSE approaches are not exploited. Therefore, we aim to pursue this direction for the first time. To this end, we first develop robust formulations of the DSSE approaches to deal with the presence of - 1) bad data and 2) network parameter uncertainty. We then develop joint formulations of the sparsity-based estimators for enhancing the DSSE accuracy at highly unobservable conditions. To validate the efficacy of the proposed approaches, we perform extensive simulations on IEEE 37 bus test system.

4.1 Robust DSSE approaches

4.1.1 Proposed Approach: Robustness to bad data

“Bad data” refers to the data measurements that significantly deviate from the underlying actual behavior. They occur due to instrument failures, impulsive communication noises, infrequent instrument calibration, or jammed sensors. In the smart grid context, bad data can also refer to some form of malicious data injections [86], [87], [88], [89]. The sparsity of bad data reflects the fact that cyber-attacks, impulsive noise, or faulty sensors are infrequent as compared to the total number of measurements. In practice, prior research efforts [90],[91] have also affirmed that at most 10% of total data ends up being bad. This section discusses the robust formulations of sparsity-based approaches in the presence of bad data.

Robust CS

Considering bad data and measurement noise, the sensing matrix $\mathbf{y}$ in (3.9) can be defined as,

$$\mathbf{y} = \Phi \mathbf{x} + \mathbf{o} + \mathbf{e} \quad (4.1)$$

where, $\mathbf{o} \in \mathbb{R}^m$ and $\mathbf{e} \in \mathbb{R}^>$ represent outlier noise and measurement noise respectively. $\mathbf{o}$ is an unknown vector with its entry $\mathbf{o}(i)$ being non-zero only if $\mathbf{y}(i)$ is a bad datum. The joint reconstruction of both $\mathbf{x}$ and $\mathbf{o}$ essentially reveals the state and identifies the faulty measurements. It is reasonable to assume that the presence of bad data in the measurement dataset is sparse. Hence, by exploiting the sparsity of $\mathbf{o}$, we can recover both the states and bad data. The robust optimization formulation corresponds to,

$$\begin{aligned} \min_{\mathbf{a}, \mathbf{o}} \quad & \|\mathbf{a}\|_1 + \lambda_1 \|\mathbf{o}\|_1 + \lambda_2 \|\mathbf{y} - \Phi \Psi \mathbf{a} - \mathbf{o}\|_2 \\ \text{subject to} \quad & (3.2), (3.3) \end{aligned} \quad (4.2)$$

where λ_1 and λ_2 are the tuning parameters. The reconstructed signal is $\hat{\mathbf{x}} = \Psi \hat{\mathbf{a}}$. A similar robust formulation is obtained for 2-D CS by incorporating the l_1 minimization of bad data

vector in (3.15).

Robust Matrix Completion

As mentioned earlier, the partially observed matrix $\mathbf{X}$ serves as the starting point for matrix completion. However, entry-wise measurement noise and outlier noise could exist. This in turn may degrade the performance of matrix completion-based state estimation. A robust formulation is proposed to withstand these errors. The key idea behind the robust formulation is to recover the low-rank matrix as well as the outlier matrix from the observed matrix $\mathbf{X}$ at different levels of system observability. The partially observed matrix $\mathbf{X}$ is,

$$\mathbf{X} = \mathbf{L} + \mathbf{S} + \mathbf{Z} \quad (4.3)$$

where $\mathbf{L}$ – low rank matrix, $\mathbf{S}$ – sparse bad data matrix and $\mathbf{Z}$ – measurement noise matrix

Accordingly, the optimization problem (3.19) is reformulated to incorporate bad data and measurement noise as,

$$\begin{aligned} \min_{\mathbf{L}, \mathbf{S}} \quad & \|\mathbf{L}\|_* + \lambda_1 \|\text{vec}(\mathbf{S})\|_1 + \lambda_2 \|\mathbf{X} - \mathbf{L} - \mathbf{S}\|_F \\ \text{subject to} \quad & (3.2), (3.3) \end{aligned} \quad (4.4)$$

where λ_1 and λ_2 are the tuning parameters. By solving (4.4), the low rank matrix ($\mathbf{L}$) and corrupted matrix ($\mathbf{S}$) is jointly recovered [92].

Robust Tensor Completion

In the tensor completion formulation (3.20), we need to remove outliers and recover the low-rank tensor based on the global structure of the tensor [93]. Therefore, similar to the matrix completion case, the robust tensor completion formulation of tensor $\mathcal{L}$ corresponds

to,

$$\begin{aligned}
(\hat{\mathcal{L}}, \hat{\mathcal{S}}) = \min_{\mathcal{L}, \mathcal{S}} \quad & \sum_{i=1}^N \|\mathcal{L}_{(i)}\|_* + \lambda \|\text{vec}(\mathcal{S})\|_1 \\
\text{subject to} \quad & \mathcal{X} = \mathcal{L} + \mathcal{S} \\
& (3.2), (3.3)
\end{aligned} \tag{4.5}$$

where λ is the tuning parameter. $\mathcal{L}, \mathcal{S}$ are low rank and bad data tensor respectively.

4.1.2 Proposed Approach: Robustness to network parameter uncertainty

In all the DSSE methods discussed so far, the utilities assume the perfect knowledge of the network parameters stored in the database. However, in reality, parameters may be incorrectly documented or unknown. Furthermore, it is reasonable to assume that such errors in network parameters do not exist universally. That is, only a subset of these values may be erroneous i.e., these parameter errors occur sparsely [85]. Consider a general norm minimization problem with objective $\|\mathbf{Ax} - \mathbf{b}\|$ and variable $\mathbf{x}$ along with uncertainty for a set of possible values of $\mathbf{A} \in \mathcal{A}$. The worst-case robust approximation formulation [94] can be applied here. i.e., the goal is to minimize the worst-case error e_{wc} ,

$$\begin{aligned}
& \underset{\mathbf{x}}{\text{minimize}} \quad e_{wc}(\mathbf{x}) \\
& \text{where} \quad e_{wc}(\mathbf{x}) = \sup \{ \|\mathbf{Ax} - \mathbf{b}\| \mid \mathbf{A} \in \mathcal{A} \}.
\end{aligned} \tag{4.6}$$

(4.6) can be cast as a LP with variables $\mathbf{x}$ and t when $\mathcal{A}$ is the singleton $\mathcal{A} = \{\mathbf{A}\}$,

$$\begin{aligned}
& \underset{\mathbf{x}, t}{\text{minimize}} \quad t \\
& \text{subject to} \quad -t\mathbf{1} \preceq \mathbf{Ax} - \mathbf{b} \preceq t\mathbf{1}
\end{aligned} \tag{4.7}$$

For DSSE, uncertainties in the network parameter values result in uncertainties in submatrices of admittance matrix $\mathbf{Y}_{\mathbf{LL}} \in \mathbb{C}^{|\mathcal{P}| \times |\mathcal{P}|}$. The robust formulation of the power-flow

constraints (3.2) with voltage $\tilde{\mathbf{x}}$ as the optimization variable corresponds to,

$$\begin{aligned}
& \underset{\tilde{\mathbf{x}}, t}{\text{minimize}} && t \\
& \text{subject to} && -t\mathbf{1} \preceq \Re\{\mathbf{Y}_{\mathbf{LL}}\mathbf{x} - (\mathbf{M}_{\mathbf{YL}}\mathbf{z} - \mathbf{Y}_{\mathbf{L0}}\mathbf{v}_0)\} \preceq t\mathbf{1} \\
& && -t\mathbf{1} \preceq \Im\{\mathbf{Y}_{\mathbf{LL}}\mathbf{x} - (\mathbf{M}_{\mathbf{YL}}\mathbf{z} - \mathbf{Y}_{\mathbf{L0}}\mathbf{v}_0)\} \preceq t\mathbf{1}
\end{aligned} \tag{4.8}$$

where

$$\mathbf{M}_{\mathbf{YL}} := \left(\text{diag}(\bar{\mathbf{w}})^{-1}, -j\text{diag}(\bar{\mathbf{w}})^{-1} \right)$$

Thus, by minimizing the maximum residual error of the powerflow constraint, we can get robust solutions for $\tilde{\mathbf{x}}$. The complete robust formulation of CS with network parameter uncertainty corresponds to,

$$\begin{aligned}
& \underset{\mathbf{s}, t}{\text{minimize}} && \|\mathbf{s}\|_1 + \lambda t \\
& \text{subject to} && \mathbf{y} = \Phi\Psi\mathbf{s} \\
& && -t\mathbf{1} \preceq \Re\{\mathbf{Y}_{\mathbf{LL}}(\Psi\mathbf{s}) - (\mathbf{M}_{\mathbf{YL}}\mathbf{z} - \mathbf{Y}_{\mathbf{L0}}\mathbf{v}_0)\} \preceq t\mathbf{1} \\
& && -t\mathbf{1} \preceq \Im\{\mathbf{Y}_{\mathbf{LL}}(\Psi\mathbf{s}) - (\mathbf{M}_{\mathbf{YL}}\mathbf{z} - \mathbf{Y}_{\mathbf{L0}}\mathbf{v}_0)\} \preceq t\mathbf{1}
\end{aligned} \tag{4.9}$$

where λ is the tuning parameter. A similar robust formulation can be derived for matrix completion and tensor completion approaches. Due to space constraints, those formulations have been omitted.

4.1.3 Simulation Results

In this section, the DSSE approaches introduced earlier are evaluated and compared on IEEE 37 test systems. This test system is a standard three-phase unbalanced distribution grids [95]. IEEE 37 test system is a highly unbalanced delta-connected system with an operating voltage of 4.8 kV and spot loads. The CVX solver [96] is utilized to solve the involved optimization formulations. For quantifying the performance of sparsity-based DSSE methods, MAPE (Mean Absolute Percentage Error) and MIAE (Mean Integrated Absolute Error) metrics are

used for voltage magnitude and voltage angle, respectively. They are defined as,

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{x_i - \hat{x}_i}{x_i} \right| \quad (4.10)$$

$$MIAE = \sum_{i=1}^N \frac{|\theta_i - \hat{\theta}_i|}{N} \quad (4.11)$$

where x_i and θ_i represent the true magnitude and angle at bus i respectively. $\hat{x}_i$ and $\hat{\theta}_i$ are the estimated magnitude and angle, respectively. Different case studies are presented for evaluating the performance of these methods.

Case I - Robustness to bad data

The robustness of DSSE methods is evaluated on the IEEE 37 bus test system with power and voltage measurement data. The percentage of bad data is increased from 0% to 10% of the total measurements with a fixed CMR of 75%. For each percentage, 50 cases are generated with different randomly generated bad data. We assume that the measurement noise distribution has a mean zero and standard deviation of 1% of the actual parameter value, and the standard deviation of bad measurements is 100% of the actual value. The performance of WLS, robust 1-D CS, and matrix completion is illustrated in the presence of bad data and measurement noise in Fig. 4.1 and Fig. 4.2. It can be inferred that WLS is more sensitive to bad data than the sparse-aware approaches. For instance, with the presence of 10% of bad data, the magnitude performance of the matrix completion approach is 94% better compared to the WLS method.

Fig 4.3 and Fig 4.4 illustrate the performance of robust versions of tensor completion and 2-D CS-based DSSE. These methods deliver comparable performance in terms of voltage magnitude estimation. However, 2-D CS-based method has improved accuracy in voltage angle estimation, compared to tensor completion. This is because tensor completion focuses more on voltage magnitude error reduction relative to voltage angle accuracy due to the different scales of those constituent tensor elements.

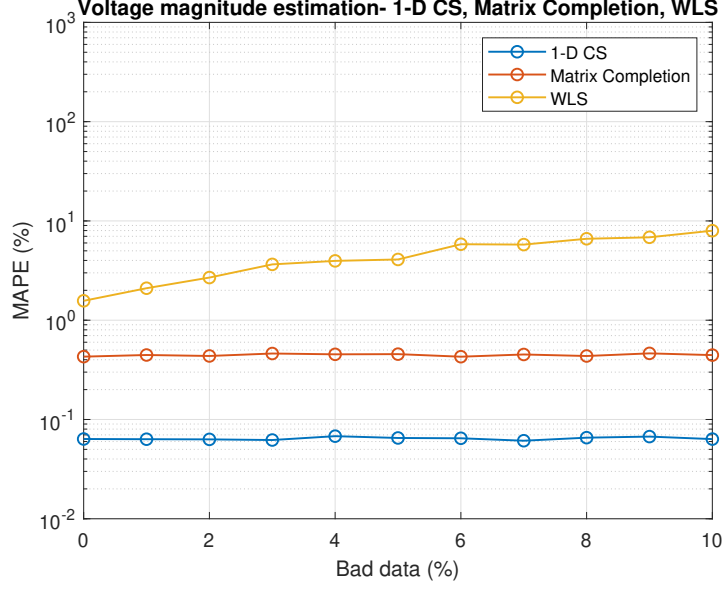


Figure 4.1: *Case I: Voltage magnitude performance of spatial approaches for different bad data*

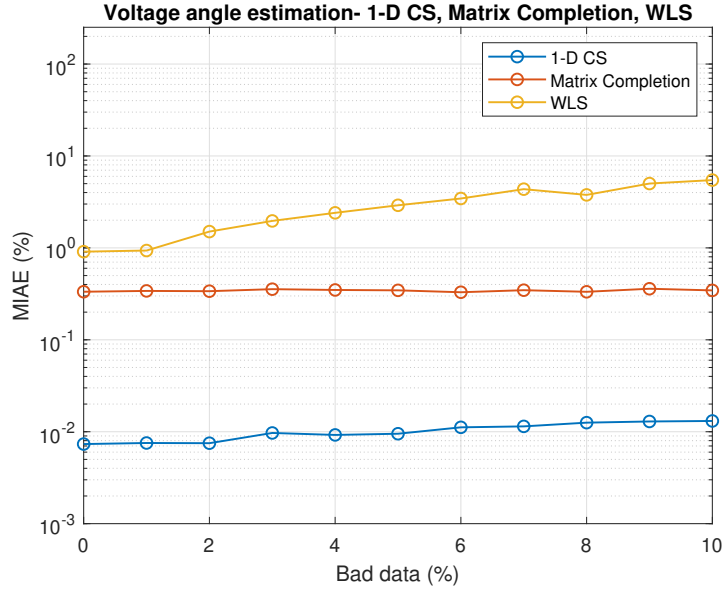


Figure 4.2: *Case I: Voltage angle performance of spatial approaches for different bad data*

Case II - Robustness to network parameter uncertainty

The performance of the robust formulation in (4.9) is compared against WLS for different CMRs in Fig. 4.5 and Fig. 4.6. The measurement set includes real power injections, reactive power injections, and voltage phasors. The states to be estimated are voltage magnitude and

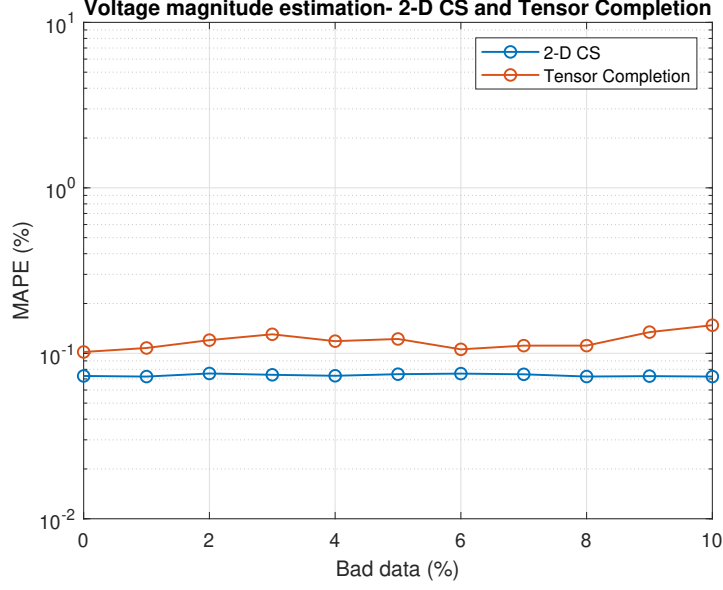


Figure 4.3: *Case I: Voltage magnitude performance of spatiotemporal approaches for different bad data*

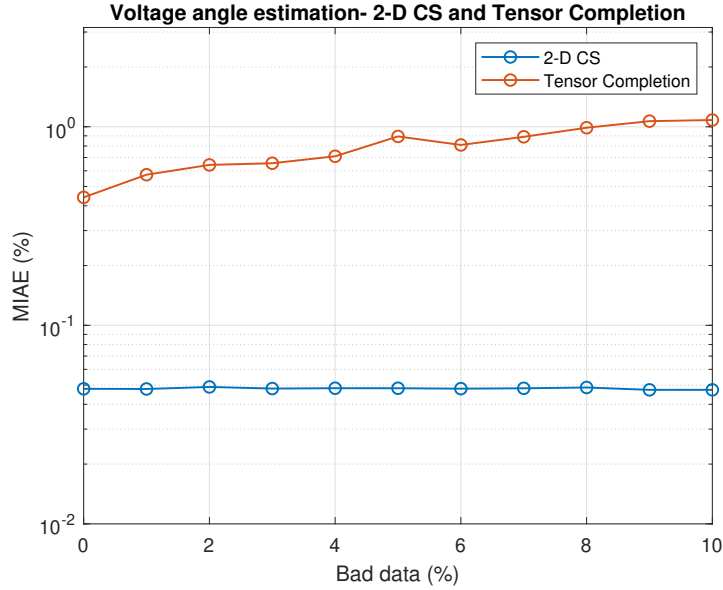


Figure 4.4: *Case I: Voltage angle performance of spatiotemporal approaches for different bad data*

voltage angle. Since implementing the WLS approach requires a full-rank jacobian matrix, i.e., the number of measurements needs to be equal to the number of states, this approach is not feasible with less than 66% measurements. Therefore, pseudo-measurements based

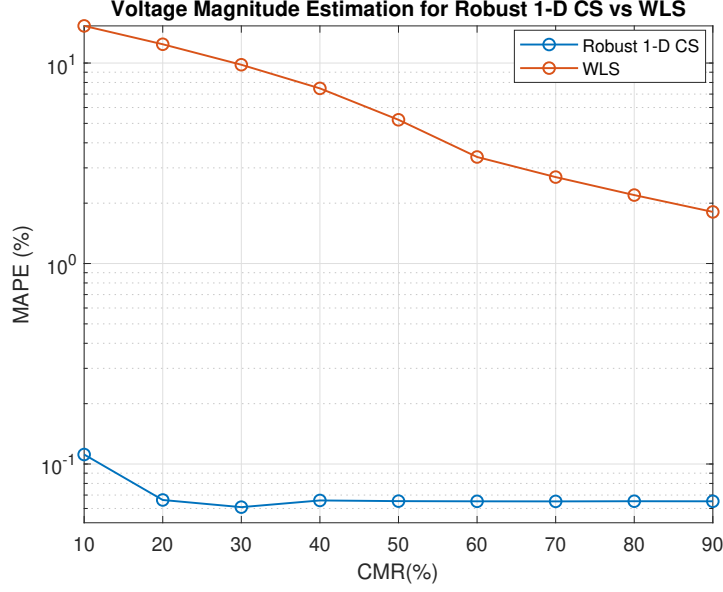


Figure 4.5: *Case II: Voltage magnitude performance of spatial approaches in presence of network parameter uncertainty*

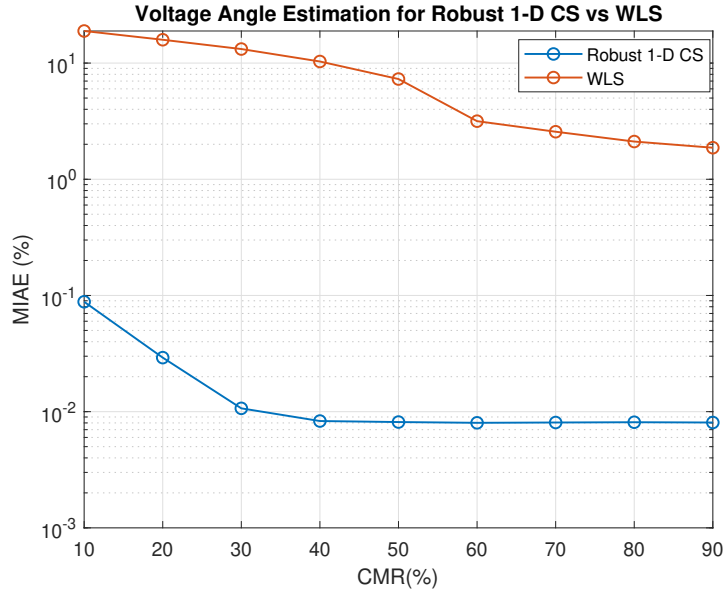


Figure 4.6: *Case II: Voltage angle performance of spatial approaches in presence of network parameter uncertainty*

on historical data of the network are added to make the system observable for CMR less than 66%. To simulate network parameter uncertainties, the resistance in the branch 2-3 for phase A is intentionally made erroneous by 10 p.u. Also, the reactance in the branch 12-13

for phase B is modified from 0.29 p.u to 4.65 p.u. The measurements are generated with this perturbed model whereas the model known to the estimator remains unchanged. It can be inferred that the states estimated using the WLS approach are not robust to parameter uncertainty. However, the 1-D CS-based approach reconstructs the states with high fidelity even at low CMR values. Also, the WLS estimates are poor as the pseudo-measurements are inaccurate compared to the real-time measurements.

4.2 Efficient DSSE approaches

In order to accurately estimate the states with high probability using the sparsity-based approaches, the requirement of a minimum number of measurements must be satisfied. In compressive sensing, the reconstruction of length N states using M measurements where $M \ll N$ is possible by exploiting the sparsity of states in a transformed basis. K -sparse states are recovered accurately using $M \geq cK \log(N/K)$ i.i.d gaussian measurements [97]. In a matrix completion approach, the minimum number of measurements (m) required to recover the matrix of size $n_1 \times n_2$ with high probability is $m \geq Cn^{1.2}r \log n$ where $n = \max(n_1, n_2)$ and r is the rank of the matrix [98]. These requirements restrict the application of sparsity-based approaches for highly unobservable distribution systems. Furthermore, in the matrix completion approach, the minimization of the nuclear norm requires solving a semi-definite program which becomes computationally inefficient for large matrices. An efficient alternating minimization algorithm is proposed in [99] that reformulates the matrix completion problem with time-series data.

This chapter proposes two unique approaches to estimating the system states when the availability of spatio-temporal measurements at the local control center is very limited. We consider a commonly occurring practical scenario where the sensors at specific spatial locations send data to the local center at a particular sampling rate. The first proposed approach estimates the states by performing matrix completion across space and compressive sensing across time using an alternating minimization approach. This formulation estimates the states in a single shot by exploiting the low-rank property as well as temporal sparsity in

the measurements while incorporating the power-flow constraints. In the second approach, compressive sensing and matrix completion is performed in two stages. In the first stage, the compressed measurements from a single sensor are recovered by exploiting the sparsity of the states on a linear transformation basis. In the second stage, matrix completion across the network is performed at individual time instants. In the following subsections, we describe these two approaches, namely the joint MC-CS approach and CS-MC approach.

4.2.1 Proposed Joint MC-CS Approach

The framework of the joint MC-CS approach is shown in Fig.4.7. Consider a power distribution grid with $|\mathcal{P}|$ three-phase non-slack buses. Sensors are deployed throughout the grid but due to communication and other constraints, only a fraction of data is aggregated from these sensors and used for DSSE. Assume the sensors located at a subset of buses send data to the local control center at time $t = 1, 2, \dots, T$ at a predefined sampling rate. Let $\mathbf{M}^t$ denote the measurement matrix at time t whose structure for every bus $b \in \mathcal{P}$ is given as,

$$[\Re(v_i), \Im(v_i), |v_i|, \Re(s_i), \Im(s_i)]^\top \quad (4.12)$$

A block matrix $\mathbf{M}$ is constructed as,

$$\mathbf{M} = [\mathbf{M}^1; \mathbf{M}^2; \dots, \mathbf{M}^T] \in \mathbb{R}^{m \times n} \quad (4.13)$$

Here, $m = 5T$ and $n = |\mathcal{P}|$. The linearized powerflow constraints from (3.2) and (3.3) at time $t = 1, \dots, T$ can be written as $\mathbf{y} \approx \mathbf{A}\mathbf{p} + \mathbf{b}$.

where, $\mathbf{y} = [\Re(\mathbf{v}^1), \Im(\mathbf{v}^1), |\mathbf{v}^1|, \dots, \Re(\mathbf{v}^T), \Im(\mathbf{v}^T), |\mathbf{v}^T|]^\top$,

$$\mathbf{A} = \begin{bmatrix} \mathbf{B}_1 & \mathbf{B}_2 & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{B}_3 & \mathbf{B}_4 & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{C}_1 & \mathbf{C}_2 & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{B}_1 & \mathbf{B}_2 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{B}_3 & \mathbf{B}_4 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{C}_1 & \mathbf{C}_2 \end{bmatrix},$$

$$\mathbf{p} = [\Re(\mathbf{s}^1)^\top \ \Im(\mathbf{s}^1)^\top \dots \ \Re(\mathbf{s}^T)^\top \ \Im(\mathbf{s}^T)^\top]^\top$$

and $\mathbf{b} = [\Re(\mathbf{w})^\top, \Im(\mathbf{w})^\top, |\mathbf{w}|^\top \dots \Re(\mathbf{w})^\top, \Im(\mathbf{w})^\top, |\mathbf{w}|^\top]^\top$

where, the term $[\mathbf{B}_1 \ \mathbf{B}_2] = \Re(\mathbf{B})$, $[\mathbf{B}_3 \ \mathbf{B}_4] = \Im(\mathbf{B})$ and $[\mathbf{C}_1 \ \mathbf{C}_2] = \mathbf{C}$.

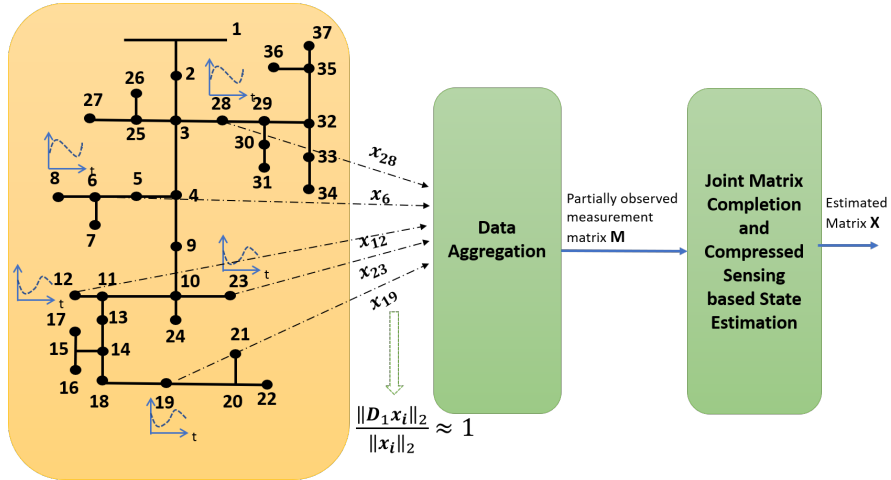


Figure 4.7: *Framework of the Joint MC-CS approach*

It is important to note that the rows/columns of the matrix $\mathbf{M}$ are observed to exhibit low-rank feature and discrete cosine transform (DCT) compactness properties as discussed next.

Low rank property

The columns of the matrix $\mathbf{M}$ are dependent on each other as there exists a spatial correlation between different locations in a power grid. Furthermore, the physics of power-flow relates to the different measurements. Hence, $\mathbf{M}$ possess a low-rank property which can be evaluated by calculating the SVD of the matrix as,

$$\mathbf{M} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T \quad (4.14)$$

where the matrix $\mathbf{U} \in \mathbb{R}^{m \times m}$, $\mathbf{V} \in \mathbb{R}^{n \times n}$ and $\mathbf{\Sigma} \in \mathbb{R}^{m \times n}$ containing $p = \min(m, n)$ singular values arranged in descending order ($\sigma_1 > \sigma_2 > \dots > \sigma_p$). For our experiments involving practical data from IEEE 37 bus test system, it can be inferred that the largest 5 singular values occupy about 99.9% of the energy confirming the low-rank property of the measurement matrix. This property has also been confirmed by other prior efforts on matrix completion [9].

DCT compactness analysis

In a distribution network, the loads are observed to be slowly changing over time. The temporal data in the matrix $\mathbf{M}$ represented by $\mathbf{x}_i$ is observed to exhibit sparsity in a linear transformation basis [4]. Discrete cosine transform (DCT) enables to the representation of the data in fewer coefficients. The DCT matrix $\mathbf{D} = \{d(k, n)\}$ of dimension $T \times T$ as defined in [100] can be split as,

$$\mathbf{D} = \begin{bmatrix} \mathbf{D}_1 \\ \mathbf{D}_2 \end{bmatrix}$$

where $\mathbf{D}_1$ consists of first ' j ' rows of $\mathbf{D}$ and $\mathbf{D}_2$ consists of last ' $T - j$ ' rows. To exhibit temporal sparsity for the time-series data $\mathbf{x}_i$, only a few DCT coefficients will capture most of the energy i.e.,

$$\frac{\|\mathbf{D}_1 \mathbf{x}_i\|_2}{\|\mathbf{x}_i\|_2} \approx 1, \quad \frac{\|\mathbf{D}_2 \mathbf{x}_i\|_2}{\|\mathbf{x}_i\|_2} \approx 0$$

This property can be observed from the practical time-series data from the IEEE 37 bus test system where 1-2 DCT coefficients occupy 99% of the energy, thus proving the temporal data in matrix $\mathbf{M}$ is compact. However, due to limited system observability, only limited entries of the matrix $\mathbf{M}$ are observed. In order to recover the complete matrix, the low-rank feature property, DCT compactness, and the linearized power-flow constraints are exploited in an integrated optimization formulation corresponding to,

$$\begin{aligned}
\min_{\mathbf{X}} \quad & \|\mathbf{X}\|_* + \lambda_1 \|P_\Omega(\mathbf{X}) - P_\Omega(\mathbf{M})\|_F^2 + \nu \|\mathbf{y} - (\mathbf{A}\mathbf{p} + \mathbf{b})\|_2^2 + \lambda_2 \|\mathbf{s}\|_2 \\
\text{s.t.} \quad & \mathbf{y} = [a_1^\top \mathbf{X} \ a_2^\top \mathbf{X} \ \dots \ a_{3T-2}^\top \mathbf{X} \ a_{3T-1}^\top \mathbf{X} \ a_{3T}^\top \mathbf{X}]^\top, \\
& \mathbf{p} = [c_1^\top \mathbf{X} \ c_2^\top \mathbf{X} \ \dots \ c_{2T-1}^\top \mathbf{X} \ c_{2T}^\top \mathbf{X}]^\top, \\
& \mathbf{s} = \begin{bmatrix} \mathbf{D}_2 \ (e_4^\top \text{reshape}(\mathbf{X}(:, 1), [5, T]))^\top \\ \mathbf{D}_2 \ (e_5^\top \text{reshape}(\mathbf{X}(:, 1), [5, T]))^\top \\ \mathbf{D}_2 \ (e_1^\top \text{reshape}(\mathbf{X}(:, 1), [5, T]))^\top \\ \vdots \\ \mathbf{D}_2 \ (e_1^\top \text{reshape}(\mathbf{X}(:, n), [5, T]))^\top \end{bmatrix}
\end{aligned} \tag{4.15}$$

where, $m = 5T$, $n = |\mathcal{P}|$, $\mathbf{X} \in \mathbb{R}^{m \times n}$, $\mathbf{y} \in \mathbb{R}^{\frac{3}{5}mn}$, $\mathbf{p} \in \mathbb{R}^{\frac{2}{5}mn}$

Here, $a_{3(t-1)+i} = e_{5(t-1)+i}$ and $c_{2(t-1)+i} = e_{5(t-1)+3+i}$ are the standard basis vectors in $\mathbb{R}^m$, e_1 , e_4 and e_5 are the standard basis vectors in $\mathbb{R}^5$. The parameters $\lambda_1 \geq 0$, $\nu \geq 0$, $\lambda_2 \geq 0$ are the tuning parameters.

The matrix $\mathbf{X}$ can be factorized into two matrices $\mathbf{U}$ and $\mathbf{V}$. The nuclear norm of $\mathbf{X}$ can be expressed by the Frobenius norm of matrix $\mathbf{U}$ and $\mathbf{V}$ given as,

$$\begin{aligned}
\|\mathbf{X}\|_* &= \min_{\mathbf{U}, \mathbf{V}} \|\mathbf{U}\|_F^2 + \|\mathbf{V}\|_F^2 \\
&\text{subject to } \mathbf{X} = \mathbf{UV}
\end{aligned} \tag{4.16}$$

Substituting (4.16) in (4.15), we obtain the following optimization problem,

$$\begin{aligned}
& \min_{\mathbf{U}, \mathbf{V}} \|\mathbf{U}\|_F^2 + \|\mathbf{V}\|_F^2 + \lambda_1 \|P_\Omega(\mathbf{UV}) - P_\Omega(\mathbf{M})\|_F^2 + \\
& \quad \nu \|f_1(\mathbf{UV}) - (\mathbf{A}f_2(\mathbf{UV}) + \mathbf{b})\|_2^2 + \lambda_2 \|f_3(\mathbf{UV})\|_2 \\
& \text{s.t. } f_1(\mathbf{UV}) = [a_1^\top \mathbf{X} \ a_2^\top \mathbf{X} \ \dots \ a_{3T-2}^\top \mathbf{X} \ a_{3T-1}^\top \mathbf{X} \ a_{3T}^\top \mathbf{X}]^\top, \\
& \quad f_2(\mathbf{UV}) = [c_1^\top \mathbf{X} \ c_2^\top \mathbf{X} \ \dots \ c_{2T-1}^\top \mathbf{X} \ c_{2T}^\top \mathbf{X}]^\top, \\
& \quad f_3(\mathbf{UV}) = \begin{bmatrix} \mathbf{D}_2 \ (e_4^\top \text{reshape}(\mathbf{X}(:, 1), [5, T]))^\top \\ \mathbf{D}_2 \ (e_5^\top \text{reshape}(\mathbf{X}(:, 1), [5, T]))^\top \\ \mathbf{D}_2 \ (e_1^\top \text{reshape}(\mathbf{X}(:, 1), [5, T]))^\top \\ \vdots \\ \mathbf{D}_2 \ (e_1^\top \text{reshape}(\mathbf{X}(:, n), [5, T]))^\top \end{bmatrix} \\
& \quad \mathbf{X} = \mathbf{UV}
\end{aligned} \tag{4.17}$$

The matrix completion formulation in (4.17) is a non-convex problem. But using alternating minimization algorithm [99], the problem becomes convex when either $\mathbf{U}$ or $\mathbf{V}$ is fixed. This algorithm updates the variables $\mathbf{U}$ and $\mathbf{V}$ at each iteration k in an alternating fashion while fixing the other factor. The update rules are given by the following update equations as,

$$\begin{aligned}
& \underset{\mathbf{U}^{(k)}}{\text{minimize}} \quad \|\mathbf{U}\|_F^2 + \lambda_1 \|P_\Omega(\mathbf{UV}^{(k-1)}) - P_\Omega(\mathbf{M})\|_F^2 + \nu \\
& \quad \|f_1(\mathbf{UV}^{(k-1)}) - (\mathbf{A}f_2(\mathbf{UV}^{(k-1)}) + \mathbf{b})\|_2^2 \\
& \quad + \lambda_2 \|f_3(\mathbf{UV}^{(k-1)})\|_2
\end{aligned} \tag{4.18}$$

$$\begin{aligned}
& \underset{\mathbf{V}^{(k)}}{\text{minimize}} \quad \|\mathbf{V}\|_F^2 + \lambda_1 \|P_\Omega(\mathbf{U}^k \mathbf{V}) - P_\Omega(\mathbf{M})\|_F^2 + \nu \\
& \quad \|f_1(\mathbf{U}^k \mathbf{V}) - (\mathbf{A}f_2(\mathbf{U}^k \mathbf{V}) + \mathbf{b})\|_2^2 \\
& \quad + \lambda_2 \|f_3(\mathbf{U}^k \mathbf{V})\|_2
\end{aligned} \tag{4.19}$$

The alternating minimization approach for the proposed approach is given in Algorithm 1.

4.2.2 Proposed CS-MC Approach

The joint MC-CS approach proposed in section 4.2.1 directly incorporates the raw information from each of the sensors at a particular bus for state estimation. However, due to the network bandwidth limitation, it may not be practical to collect all the temporal measurements. Furthermore, as stated earlier, solving (4.13) involving large matrices can be computationally inefficient. Therefore, we propose a state estimation approach that alleviates these drawbacks.

The framework of the CS-MC approach is shown in Fig.4.8. Assume the measurements collected from each buses is $\Phi \mathbf{x}$ where $\Phi \mathbf{x} \ll \mathbf{x}$. The recovery of the states is performed in two stages. Firstly, the incomplete temporal measurements are recovered by compressive sensing using (3.10). It should be noted that no power-flow constraints are used at this stage. Once the estimates of all the temporal measurements are obtained, the second stage consists of recovering the spatial states by classic matrix completion based state estimation using (3.19). The state estimation is performed separately at each time steps along-with the power-flow constraints in (3.2) and (3.3).

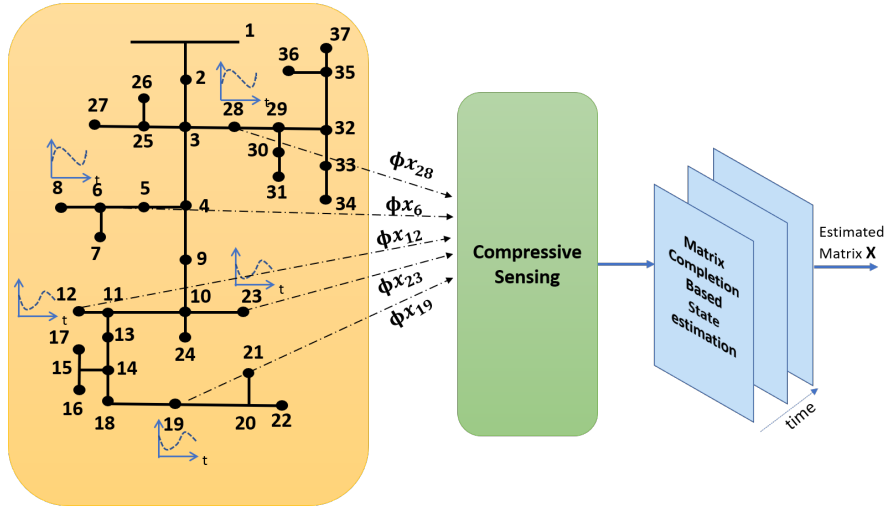


Figure 4.8: Framework of the CS-MC approach

Algorithm 1 Alternating minimization Algorithm for matrix completion across space and compressive sensing across time

Input: Measurement matrix $\mathbf{M}$, Set of known indices Ω , DCT matrix $\mathbf{D}$, system model $\mathbf{A}$, $\mathbf{b}$, number of iterations N .

initialization: Compute the SVD of the matrix $\mathbf{M} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$, set $\mathbf{V}^{(0)} = \mathbf{\Sigma}^{0.5}\mathbf{V}$

- 1: **for** $k = 1, \dots, N$ **do**
 - 2: Solve (4.18)
 - 3: Solve (4.19)
 - 4: **end for**
 - 5: **return** $\mathbf{X} = \mathbf{U}^{(N)}\mathbf{V}^{(N)}$
-

4.2.3 Computational Complexity

In this section, the computational complexity of the proposed approaches for a matrix $\mathbf{X} \in \mathbb{R}^{5T \times |\mathcal{P}|}$ is discussed. In the joint MC-CS approach, in order to solve the power-flow constraints, $\mathcal{O}(|\mathcal{P}|^3 T^3)$ computations are required at each iteration. In a classic MC approach, the matrix $\mathbf{X} \in \mathbb{R}^{5 \times |\mathcal{P}|}$ is estimated separately for each time T . Denoting m and n as the rows and columns of $\mathbf{X}$, the main computation is involved in calculation of the SVD of the matrix which is $\mathcal{O}(mn \cdot \min(m, n))$ along-with the power-flow constraints $\mathcal{O}(n^3)$ at each time. Hence, the overall worst case computational complexity is $\mathcal{O}(|\mathcal{P}|^3 T)$. The CS-MC approach performs CS for each temporal state and classic matrix completion in the next step. CS requires $\mathcal{O}(T \cdot k \cdot \min(T, k))$ computations where T denotes the time and k as constraints. The overall worst case complexity in both the steps is $\mathcal{O}(|\mathcal{P}|^3 T)$. This illustrates that the proposed joint MC-CS approach is computationally expensive as compared to the conventional matrix completion as well as the CS-MC approach.

4.2.4 Simulation Results

In this section, the efficacy of the proposed formulation is demonstrated on the IEEE 37 unbalanced three phase test system. We characterize the performance of power and voltage magnitude recovery using the mean absolute percentage error (MAPE) metric and the voltage angle recovery using mean integrated absolute error (MIAE) metric [101]. For simulation, we consider a matrix that includes 8 time steps and whose entries are randomly available

representing different fractions of available data (FAD). The compression of the temporal measurements in the CS-MC approach is indicated by CMR as defined earlier. Fig. 4.9 shows the recovery performance of power (active and reactive) using the two proposed approaches. Fig. 4.10 -4.11 shows the comparative performance of the proposed formulation with the classic MC in the recovery of voltage magnitude and voltage angle states respectively. It can be inferred that the proposed joint MC-CS approach as well as the CS-MC approach outperforms the classic matrix completion at all FADs. The performance is superior especially in the low observability region. This is due to the fact that this approach exploits both the spatial correlation as well as temporal correlation of the states. The joint MC-CS approach, although computationally expensive, is superior than CS-MC approach. This is due to the fact that raw measurements are used for estimating the states rather than the compressed measurements. The recovery of states in CS-MC approach also depends on the CMR of the temporal measurements. The error in the recovery performance increases as the CMR is decreased. Fig. 4.9-4.11 shows the performance of CS-MC at 40% and 80% CMR.

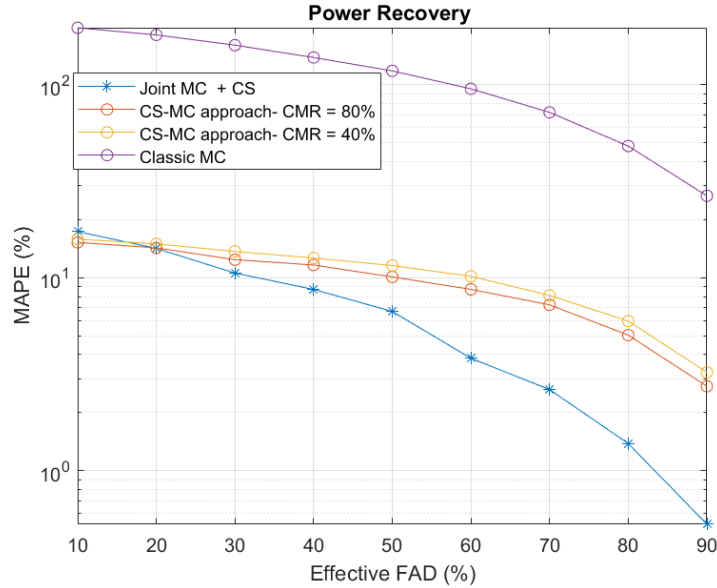


Figure 4.9: *Power recovery at different FADs*

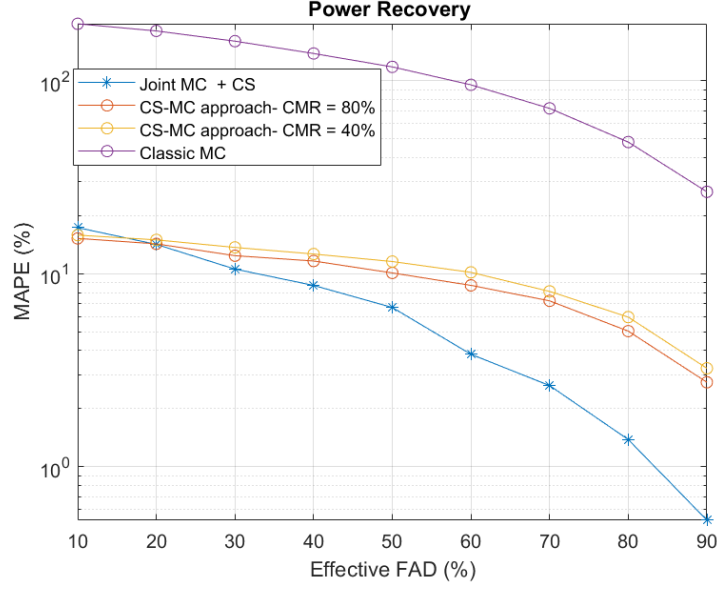


Figure 4.10: Voltage magnitude recovery at different FADs

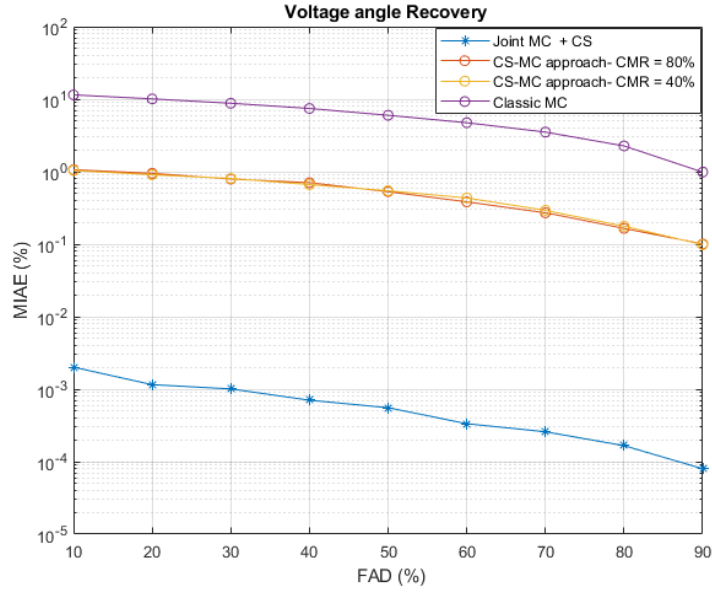


Figure 4.11: Voltage angle recovery at different FADs

4.3 Summary

In this chapter, our goal was to develop new robust formulation for these approaches to tackle the presence of bad data and network parameter uncertainties. Numerical studies show the superior performance of compressive sensing-based approaches, compared to matrix

completion and tensor completion-based methods. We also develop an efficient joint matrix completion and compressive sensing approach to perform DSSE. The proposed approach employs an alternating minimization approach to estimate an incomplete spatiotemporal matrix. The efficacy of the proposed approach was demonstrated using the IEEE 37 bus test system. The performance gains of these robust and efficient sparsity-based DSSE methods are especially pronounced in low observability conditions. Furthermore, the proposed robust optimization formulations of sparse-aware approaches are compared with conventional WLS-based estimation methods and shown to offer significant performance improvements.

Chapter 5

Handling multi time-scale measurements using Multi-task GP approach

As discussed in chapter 4, the lack of unobservability at the grid edge hinders the notion of real-time situational awareness. However, with the advent of the smart grid, there has been a plethora of efforts to extract and exploit information at the distribution feeder level. The increase in the deployment of smart meters and sensors have contributed to improved monitoring of the distribution system. The sources of information available from a smart distribution system include (1) Remote terminal unit (RTU) measurements aggregated by the SCADA system that typically includes bus voltage and head line currents. These measurements are sampled at rates ranging from a few seconds to minutes. (2) Smart meters concentrators or advanced metering infrastructure (AMI) data comprising of active and reactive power consumption details of customers. The sampling rate is usually around 15 min. The AMI sensors located on the secondary side of the distribution systems collect consumer data for billing purposes. The load composition of a primary feeder can be calculated according to the energy consumption of all the customers served by the feeder. The measurements at the primary feeder are averaged over 15 min intervals [102]. These aggregated measure-

ments serve as pseudo-measurements for SE purpose and are critical for increasing the data redundancy in the distribution system [23]. (3) Voltage phasor measurements collected by PMUs (phasor measurement units) typically sampled at 120 samples/sec. The information collected by these metering devices presents some challenges. Firstly, the measurements from heterogeneous sources are sampled at different rates and are rarely synchronized. The sources of information discussed above can be broadly classified as - (1) Fast rate measurements collected by PMUs or SCADA systems that are typically sampled every few seconds, and (2) Slow rate measurements at the primary feeder averaged over every 15 minutes or hours. Fig. 5.1 shows the sources of information available in a smart grid at different timescales [102]. There are two options for real-time DSSE when confronted with multi time-scale measurements [102], namely: (1) perform the DSSE at the resolution of the lowest measurement frequency, or (2) perform the DSSE at a resolution higher than the lowest measurement frequency using a method to reconcile the older and future measurements. For a reliable DSSE, the first option may be unsuitable. In order to pursue the second option, all the information sources should be properly combined as their inclusion may significantly affect the estimation of the system state. Secondly, the information aggregated from these sources can be intermittent and/or corrupted due to communication network impairments. Hence, one of the key challenges in distribution system automation is to properly aggregate and reconcile noisy, corrupted, heterogeneous, and incomplete time-series data. Secondly, understanding how to exploit these measurements for state estimation at any desired timescale will be critical in developing optimal control mechanisms.

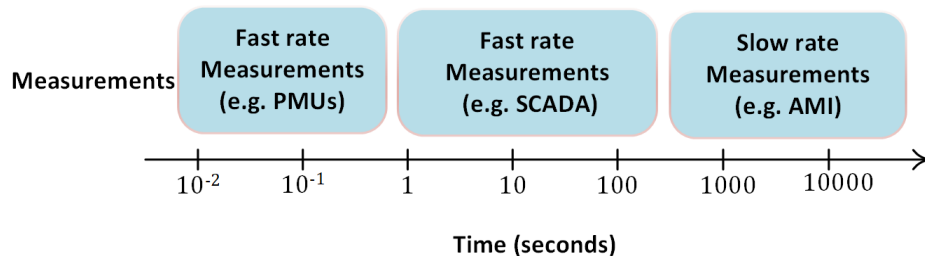


Figure 5.1: *Relative frequency of data acquisition from different sources in distribution systems*

5.1 System model

Assume we have S sensors at different spatial nodes in the distribution grid, each collecting data at a different rate. The measurements collected by these sensors include bus voltages and power injections at various nodes. The data-set $\mathcal{M}$ consists of $\{\mathbf{x}_s, \mathbf{y}_s\}_{s=1}^S$ where the s^{th} sensor data contains time $\mathbf{x}_s \in \mathbb{R}^{n_s \times 1}$ and its corresponding measurement values $\mathbf{y}_s \in \mathbb{R}^{n_s \times 1}$. The number of observations for the s^{th} sensor data is represented by n_s which are different as each sensor collects data at different temporal resolutions. The observations $\mathbf{y}_s$ consists of $\mathbf{y}_{AMI}, \mathbf{y}_{SCADA}$ or $\mathbf{y}_{PMU}$ measurements. Here, the observations $\mathbf{y}_{AMI}$ are the aggregated AMI data of the s^{th} sensor obtained at $\mathbf{x}_{AMI}$ times (15-min interval data over 24-hr period). $\mathbf{y}_{SCADA}$ and $\mathbf{y}_{PMU}$ are the voltage magnitude measurements of s^{th} sensor obtained from SCADA and PMU sensors, respectively. These measurements are obtained at time points $\mathbf{x}_{SCADA}$ (1-min interval) and $\mathbf{x}_{PMU}$ (0.1 s interval). The sampling rates of the sensors (AMI, SCADA, or PMUs) are different, hence the time instants $\mathbf{x}_s$ for each of them would also be different.

5.2 Proposed Bayesian Approach

Distribution systems have poor observability due to insufficient measurements. In order to deal with this challenge, the data collected by multiple sensors at the grid edge could be leveraged. Hence, before the DSSE operation, we reconcile the heterogeneous measurements at a single time-scale. In order to perform this, the proposed Bayesian framework is divided into two stages as illustrated in Fig.5.2. In stage 1, the temporal measurements at the buses are considered for imputation. In stage 2, these imputed timeseries are leveraged at a single snapshot of time along with their associated uncertainty for Bayesian Matrix completion based DSSE. These stages are discussed briefly in the following sections.

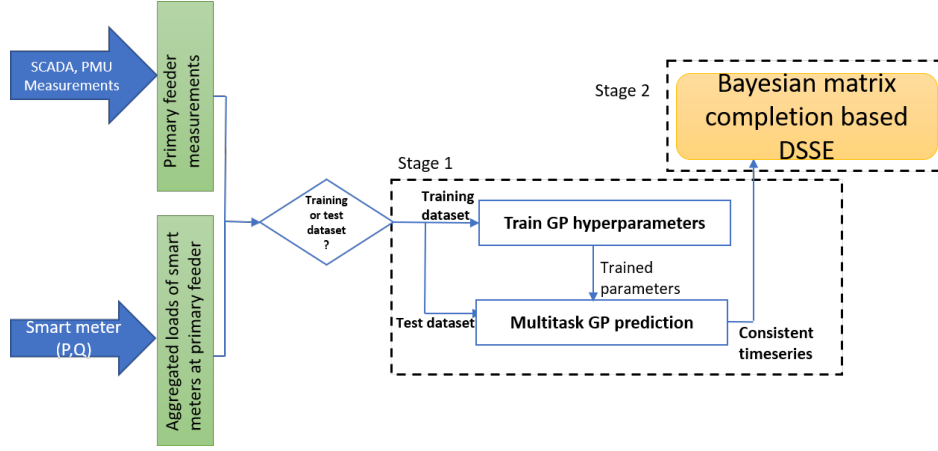


Figure 5.2: *Proposed Bayesian framework*

5.2.1 Stage 1 - Multi-task Gaussian Process

In this stage, a multi-task Gaussian process (GP) is used to impute data at any required timescale based on unevenly sampled multiple time-series data aggregated from different sensors in the distribution grid. A GP is a collection of random variables, any finite number of which has a joint normal distribution [103]. GP defines a prior over functions given as,

$$f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')) \quad (5.1)$$

where, $m(\mathbf{x})$ is the mean function and $k(\mathbf{x}, \mathbf{x}')$ is the covariance (kernel) function. The kernel function $k(\cdot, \cdot)$ dictates the correlation among data points in the modelled function. There are different kernel choices with one of the most popular being RBF (radial basis function) kernel corresponding to:

$$k(x, x') = \sigma_s^2 \exp\left(-\frac{(x - x')^2}{2l^2}\right) \quad (5.2)$$

where hyperparameters l and σ_s^2 are the length-scale and signal variance respectively. These hyperparameters dictate the smoothness of the function. The training of the GP prior refers to the estimation of the mean and kernel function hyper-parameters based on the dataset

$\mathcal{M}$.

In the process of imputation, we have target tasks consisting of time-series data $\mathcal{D} = \{\tilde{\mathbf{x}}_s, \tilde{\mathbf{y}}_s, \tilde{\mathbf{x}}_s^*, \tilde{\mathbf{y}}_s^*\}_{s=1}^S$ where $\tilde{\mathbf{x}}_s$ and $\tilde{\mathbf{y}}_s$ are observed data points with their corresponding values respectively. Multi-task GP predicts the target task $\{\tilde{\mathbf{x}}_s^*, \tilde{\mathbf{y}}_s^*\}$ including the missing measurements given the observed data and the prior parameters learned from the training dataset.

This prediction is accomplished across multiple time-series simultaneously by capturing the spatial and temporal correlation across them. In a distribution system, this is especially relevant because voltage/power at one node is typically correlated with voltage/power values at other connected nodes [23]. Temporal correlation is also a common characteristic of power and voltages. The GP prior for multi-task GP is defined as,

$$f(\cdot) = \mathcal{GP}(m_\phi(\cdot), k_\theta(\cdot, \cdot)), \quad (5.3)$$

where, $m_\phi(\cdot)$ is the mean function and $k_\theta(\cdot, \cdot)$ is the co-variance function parameterized by the sets of parameters ϕ and θ respectively. The function $m_\phi(\cdot)$ could be based on a deep neural network (e.g. multilayer perceptron (MLP)) [104]. This function models the global trend among the entire set of diverse time-series in a multi-task paradigm. The ϕ parameters associated with the mean function are globally shared across all the time-series. $k_\theta(\cdot, \cdot)$ is the kernel for the GP prior that captures the temporal correlation within each time-series. The θ parameters refer to the hyperparameters of the kernel function. The effective encoding of the prior knowledge into the GP model enables us to use only a few measurements for achieving a high imputation performance [105].

The GP prior function can be perceived as,

$$f_s(\tilde{\mathbf{x}}_1) = g(\tilde{\mathbf{x}}_1) + l_s(\tilde{\mathbf{x}}_1) \quad (5.4)$$

where $g(\cdot)$ models global trends among the diverse sensor signals and hence it is shared across all the sensors. We consider the function $g(\cdot)$ as a mean function in GP prior ($m_\phi(\cdot)$).

On the other hand, $l_s(\cdot)$ models the sensor-specific signal (for s^{th} sensor) using the kernel function ($K_\theta(\cdot, \cdot)$). The GP prior function f can be written as,

$$\begin{bmatrix} \mathbf{f}_1 \\ \vdots \\ \mathbf{f}_s \end{bmatrix} \sim \mathcal{GP} \left(\begin{bmatrix} m_\phi(\tilde{\mathbf{x}}_1) \\ \vdots \\ m_\phi(\tilde{\mathbf{x}}_s) \end{bmatrix}, \begin{bmatrix} K_\theta^{\tilde{\mathbf{x}}_1 \tilde{\mathbf{x}}_1}, & \dots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & K_\theta^{\tilde{\mathbf{x}}_s \tilde{\mathbf{x}}_s} \end{bmatrix} \right)$$

The GP prior learnable parameters $\psi = \{\phi, \theta\}$ as defined in (5.3) are optimized by maximizing the log marginal likelihood (LML) given as,

$$\psi^* = \underset{\psi}{\operatorname{argmax}} \sum_{s=1}^S \log \mathbf{p}(\mathbf{y}_s | \mathbf{x}_s, \psi) \quad (5.5)$$

where the log-likelihood is the sum of individual time-series data. An individual log-likelihood can be computed in closed form as [103],

$$\begin{aligned} \log \mathbf{p}(\mathbf{y}_s | \mathbf{x}_s, \psi) = & \frac{-1}{2} (\mathbf{y}_s - m_\phi(\mathbf{x}_s))^T (K_\theta^{xx} + \sigma^2 I)^{-1} - \\ & (\mathbf{y}_s - m_\phi(\mathbf{x}_s)) - \frac{1}{2} \log |K_\theta^{xx} + \sigma^2 I| - \frac{n_s}{2} \log(2\pi) \end{aligned} \quad (5.6)$$

Here, σ^2 represents the noise variance of the Gaussian process. The LML is computed using the stochastic gradient descent method [104]. After the GP prior parameters are optimized and effectively encoded, we predict the target tasks based on observed data. The posterior distribution can be calculated using the Gaussian identities.

Let $\mathbf{x}$ and $\mathbf{y}$ be jointly Gaussian random vectors i.e.,

$$\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} \mu_{\mathbf{x}} \\ \mu_{\mathbf{y}} \end{bmatrix}, \begin{bmatrix} A & C \\ C^\top & B \end{bmatrix} \right).$$

The conditional distribution of $\mathbf{x}$ given $\mathbf{y}$ is,

$$p(\mathbf{x} | \mathbf{y}) = \mathcal{N}(\mu_{\mathbf{x}} + CB^{-1}(\mathbf{y} - \mu_{\mathbf{y}}), A - CB^{-1}C^\top) \quad (5.7)$$

Using (5.7), the predictive distribution of $\tilde{\mathbf{y}}_s^*$ is,

$$\mathbf{p}(\tilde{\mathbf{y}}_s^* | \tilde{\mathbf{x}}_s^*, \tilde{\mathbf{y}}_s, \tilde{\mathbf{x}}_s) = \mathcal{N}(\boldsymbol{\mu}^*, \mathbf{C}^*) \quad (5.8)$$

with

$$\boldsymbol{\mu}^* = m_\phi(\tilde{\mathbf{x}}_s^*) + K_\theta^{*x} (K_\theta^{xx} + \sigma^2 I)^{-1} (\tilde{\mathbf{y}}_s - m_\phi(\tilde{\mathbf{x}}_s)) \quad (5.9)$$

$$\mathbf{C}^* = K_\theta^{**} - K_\theta^{*x} (K_\theta^{xx} + \sigma^2 I)^{-1} (K_\theta^{x*}) \quad (5.10)$$

where, $(K_\theta^{x*}) = (K_\theta^{*x})^\top$ and $(K_\theta^{x*})_{st} = k_\theta(\tilde{\mathbf{x}}_s, \tilde{\mathbf{x}}_t^*)$ and $(K_\theta^{xx})_{st} = k_\theta(\tilde{\mathbf{x}}_s, \tilde{\mathbf{x}}_t)$

The posterior mean and variance as defined in (5.9), (5.10) constitute the consistent time-series that are computed at those time instances where measurements are unobserved. Here, the consistent time series refers to reconciling all the multi time-scale measurements at a common time scale. The proposed approach is capable of imputing the multi time-scale measurements at any desired time-scale. In this work, as an example, the narrowest measurement time resolution is chosen when the measurement types are equal to or larger than two. A confidence or credible interval can be evaluated based on the predicted mean, variance, and desired confidence level.

5.2.2 Stage II - Bayesian matrix completion based DSSE

The consistent time-series obtained from GP is still limited to specific spatial locations in the network where measurements are aggregated. Matrix completion (MC) approaches aim to estimate the complete matrix from a limited set of observed entries. Specifically, matrix completion aims to find the complete matrix $\mathbf{X}$ from an incomplete and noisy observation matrix by suitable low-rank approximation. The low dimensionality in the matrix $\mathbf{X}$ results due to: (i) Spatial correlation between measurements at different locations; (ii) the correlation between different types of measurements via linearized power-flow equations [106].

Assume that the measurements at the slack bus are known. Thus, we use the measurements at the non-slack buses to construct the data matrix, whose structure is given in (3.17). The deterministic matrix completion formulation recovers the complete low-rank

matrix, whose optimization formulation is shown in (3.19). The formulation in (3.19) can be recast as a semidefinite program which is computationally inefficient for large matrices. A prior knowledge on the rank of the matrix is generally required for this formulation [107]. Furthermore, this formulation requires extensive parameter tuning as well as data and application-dependent supervision. In order to overcome these shortcomings and effectively incorporate the output from stage 1, a hierarchical sparse Bayesian matrix completion framework [107] is proposed. In the hierarchical sparse Bayesian matrix completion framework, the unknown low rank matrix is factorized into two matrices as

$$\mathbf{X} = \mathbf{A}\mathbf{B}^\top \quad (5.11)$$

where $\mathbf{A}$ is an $m \times r$ matrix, and $\mathbf{B}$ is an $n \times r$ matrix such that $\text{rank}(\mathbf{X}) = r$. The matrix $\mathbf{X}$ is the sum of the outer-products of the columns of $\mathbf{A}$ and $\mathbf{B}$ such that,

$$\mathbf{X} = \sum_{i=1}^k \mathbf{a}_i \mathbf{b}_i^\top \quad (5.12)$$

where, $k \geq r$, $\mathbf{a}_i$ and $\mathbf{b}_i$ denote the i^{th} column of matrix $\mathbf{A}$ and $\mathbf{B}$ respectively. The i^{th} row of matrix $\mathbf{A}$ and $\mathbf{B}$ is represented by $\mathbf{a}_i$ and $\mathbf{b}_i$ respectively. In order to obtain the low rank matrix, most of the columns in $\mathbf{A}$ and $\mathbf{B}$ should be set to zero. To achieve this condition, the columns of $\mathbf{A}$ and $\mathbf{B}$ are associated with Gaussian priors of precisions γ_i , that is

$$\mathbf{p}(\mathbf{A}|\boldsymbol{\gamma}) = \prod_{i=1}^k \mathcal{N}(\mathbf{a}_i|\mathbf{0}, \gamma_i^{-1}\mathbf{I}_m) \quad (5.13)$$

$$\mathbf{p}(\mathbf{B}|\boldsymbol{\gamma}) = \prod_{i=1}^k \mathcal{N}(\mathbf{b}_i|\mathbf{0}, \gamma_i^{-1}\mathbf{I}_n) \quad (5.14)$$

During inference, most of the γ_i 's take large values, thus forcing the columns of $\mathbf{A}$ and $\mathbf{B}$ to go to zero. The columns of $\mathbf{A}$ and $\mathbf{B}$ have the same sparsity profile enforced by the common precisions γ_i . These sparsity priors on the factorized matrix encourages low-rank solutions.

The precision γ_i are assumed to have a Gamma hyperprior given as,

$$\mathbf{p}(\gamma_i) = \text{Gamma}(c, \frac{1}{d}) \quad (5.15)$$

The parameters c and d are set to small values to obtain broad hyperpriors. The known entries in matrix $\mathbf{Z}$ as described in (3.17) can also be written as,

$$\mathbf{Z}_{ij} = \mathbf{X}_{ij} + \mathbf{N}_{ij}, \quad (i, j) \in \Omega \quad (5.16)$$

where, $\mathbf{X}_{ij}$ and $\mathbf{N}_{ij}$ refers to the row entry i and column entry j in the matrix $\mathbf{X}$ and $\mathbf{N}$, respectively. Using the model (5.16) and factorized matrices $\mathbf{A}$ and $\mathbf{B}$, the conditional distribution of the observations are obtained as,

$$\mathbf{p}(\mathbf{Z}|\mathbf{A}, \mathbf{B}) = \prod_{(i,j) \in \Omega} \mathcal{N}(\mathbf{Z}_{ij}|\mathbf{X}_{ij}, \beta_{ij}^{-1}) \quad (5.17)$$

where β is the noise precision. The noise precisions (β) and observations ($\mathbf{Z}$) are obtained from the stage 1 output (consistent time-series output from the multitask Gaussian process framework). The joint distribution is therefore given as,

$$\mathbf{p}(\mathbf{Z}, \mathbf{A}, \mathbf{B}, \gamma) = \mathbf{p}(\mathbf{Z}|\mathbf{A}, \mathbf{B})\mathbf{p}(\mathbf{A}|\gamma)\mathbf{p}(\mathbf{B}|\gamma)\mathbf{p}(\gamma) \quad (5.18)$$

The evaluation of posterior distributions of the latent variables $(\mathbf{A}, \mathbf{B}, \gamma)$ denoted as $\mathbf{h}$ given the observed matrix $\mathbf{Z}$ requires the computation of $\mathbf{p}(\mathbf{z})$. However, this computation is intractable as it involves marginalizing all latent variables. Mean field variational Bayes is a method that computes the posterior distribution approximations [108]. Factorized distributions partition the latent variables into disjoint groups such that,

$$\mathbf{q}(\mathbf{h}) = \prod q(\mathbf{h}_k) \quad (5.19)$$

For all the distributions of the form in (5.19), the posterior approximation of each latent

variable $\mathbf{h}_k$ is found using

$$\log \mathbf{q}(\mathbf{h}_k) = \langle \log \mathbf{p}(\mathbf{Z}, \mathbf{h}) \rangle_{\mathbf{h} \setminus \mathbf{h}_k} + \text{const} \quad (5.20)$$

The posterior distribution of $\mathbf{A}$ and $\mathbf{B}$ decompose as independent distributions of their rows. The approximate posterior distributions of the latent variables are updated as,

$$\mathbf{q}(\mathbf{a}_{i.}) = \mathcal{N}(\mathbf{a}_{i.} | \langle \mathbf{a}_{i.} \rangle, \Sigma_i^a) \quad (5.21)$$

where the mean and covariance is defined by calculating the mean iterated over all the columns of $\mathbf{a}_{i.}$ and accounting each measurement's precision β_{ij} ,

$$\langle \mathbf{a}_{i.} \rangle^\top = \langle \beta \rangle \Sigma_i^a \langle \mathbf{B}_i \rangle^\top \mathbf{z}_{i.}^\top \quad (5.22)$$

$$\Sigma_i^a = (\langle \beta \rangle \langle \mathbf{B}_i^\top \mathbf{B}_i \rangle + \Gamma)^{-1} \quad (5.23)$$

Here,

$$\langle \mathbf{B}_i^\top \mathbf{B}_i \rangle = \sum_{j: (i,j) \in \Omega} \left(\langle \mathbf{b}_{j.}^\top \rangle \langle \mathbf{b}_{j.} \rangle + \Sigma_j^b \right) \quad (5.24)$$

and

$$\Gamma = \text{diag}(\gamma)$$

Similarly, the posterior density of j^{th} row of $\mathbf{B}$ is found as,

$$\mathbf{q}(\mathbf{b}_{j.}) = \mathcal{N}(\mathbf{b}_{j.} | \langle \mathbf{b}_{j.} \rangle, \Sigma_j^b) \quad (5.25)$$

where the mean and covariance are defined as,

$$\langle \mathbf{b}_{j.} \rangle^\top = \langle \beta \rangle \Sigma_j^b \langle \mathbf{A}_j \rangle^\top \mathbf{z}_{.j}^\top \quad (5.26)$$

$$\Sigma_j^b = (\langle \beta \rangle \langle \mathbf{A}_j^\top \mathbf{A}_j \rangle + \Gamma)^{-1} \quad (5.27)$$

The posterior density of γ_i becomes a gamma distribution

$$\mathbf{q}(\gamma_i) \propto \gamma_i^{(c-1+\frac{m+n}{2})} \exp\left(-\gamma_i \frac{2d + \langle \mathbf{a}_i^\top \mathbf{a}_i \rangle + \langle \mathbf{b}_i^\top \mathbf{b}_i \rangle}{2}\right) \quad (5.28)$$

with mean,

$$\langle \gamma_i \rangle = \frac{2c + m + n}{2d + \langle \mathbf{a}_i^\top \mathbf{a}_i \rangle + \langle \mathbf{b}_i^\top \mathbf{b}_i \rangle} \quad (5.29)$$

The required expectations are given by

$$\langle \mathbf{a}_i^\top \mathbf{a}_i \rangle = \langle \mathbf{a}_i \rangle^\top \langle \mathbf{a}_i \rangle + \sum_j \left(\Sigma_i^a \right)_{ii}, \quad (5.30)$$

$$\langle \mathbf{b}_i^\top \mathbf{b}_i \rangle = \langle \mathbf{b}_i \rangle^\top \langle \mathbf{b}_i \rangle + \sum_j \left(\Sigma_i^b \right)_{ii} \quad (5.31)$$

It is important to remember that the output of stage 1 provides the uncertainty in imputations which can be used as an estimate of the noise variances for the imputed measurements. However, if a linear interpolation technique is used in stage 1, the noise precisions are not available. Hence, β will be treated as a stochastic quantity whose approximate posterior distribution is given as,

$$\langle \beta \rangle = \frac{(FAD) \times m \times n}{\langle \|\mathbf{Z} - P_\Omega(\mathbf{A}\mathbf{B}^\top)\|_F^2 \rangle} \quad (5.32)$$

The proposed Bayesian matrix completion based DSSE approach is summarized in Algorithm 2.

5.3 Computational complexity

The computational complexity of Stage 1 consisting of multitask GP framework is $\mathcal{O}(Sn_s^3)$ where n_s is the total number of data points in S^{th} sensor and S is the total number of sensors. This complexity is associated with the calculation of the predictive distributions of the multi-task GP approach using (5.9) and (5.10).

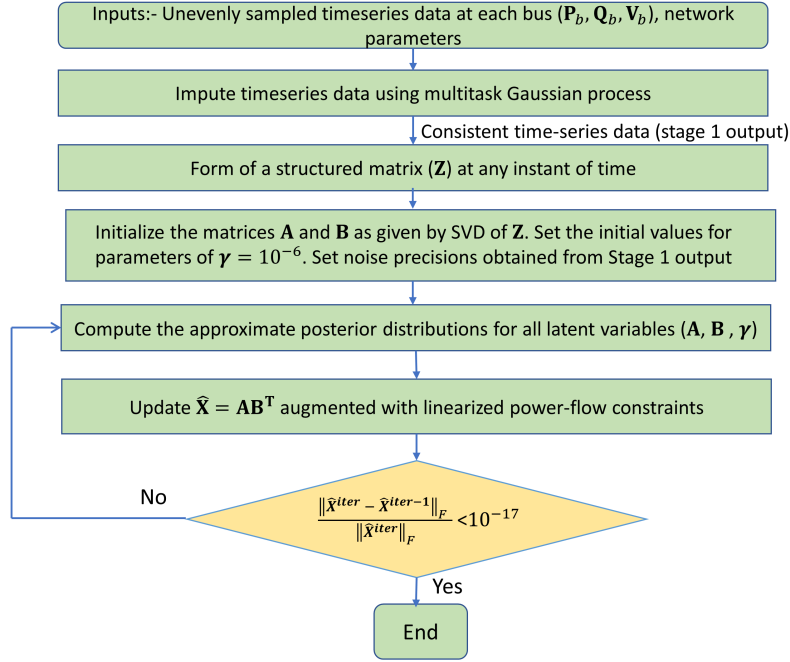


Figure 5.3: Flowchart of the proposed Bayesian framework in distribution systems

Algorithm 2 Bayesian matrix completion based DSSE

Input: Measurement matrix $\mathbf{Z}$, Set of known indices Ω , System model $(\mathbf{M}, \mathbf{w}, \mathbf{K})$, tolerance limit (η) , β, c, d

Initialization: Initialize the matrices $\mathbf{A}$ and $\mathbf{B}$ as $\mathbf{A} = \mathbf{U}\mathbf{D}^{\frac{1}{2}}$ and $\mathbf{B}^T = \mathbf{D}^{\frac{1}{2}}\mathbf{V}^T$ where $\mathbf{Z} = \mathbf{U}\mathbf{D}\mathbf{V}^T$. Initialize γ . Set $\hat{\mathbf{X}}^0 = \mathbf{Z}$.

- 1: **while** $\frac{\|\hat{\mathbf{X}}^{iter} - \hat{\mathbf{X}}^{iter-1}\|_F}{\|\hat{\mathbf{X}}^{iter-1}\|_F} < \eta$ **do**
 - 2: Calculate the mean and covariance matrix of $\mathbf{A}$ using (5.22) and (5.23).
 - 3: Calculate the mean and covariance matrix of $\mathbf{B}$ using (5.26) and (5.27).
 - 4: Calculate the posterior distribution of γ using (5.29) - (5.31).
 - 5: Set $\hat{\mathbf{X}}^{iter} = \mathbf{A}\mathbf{B}^T$ alongwith additional linearized powerflow constraints (3.2), (3.3).
 - 6: $iter \leftarrow iter + 1$
 - 7: **end while**
 - 8: **return** $\hat{\mathbf{X}}$.
-

The complexity in stage 2 includes the complexity associated with the matrix completion formulation. A deterministic matrix completion requires the computation of singular value decomposition (SVD) of the matrix at each iteration. This involves semidefinite programming that is solved by interior point methods. The computational complexity for calculating the SVD of the matrix in a deterministic method is $\mathcal{O}(mn \cdot \min(m, n))$ where m and n are the numbers of rows and columns respectively [109]. Furthermore, in order to solve the linearized powerflow equations, the computation is $\mathcal{O}(m^3)$. The overall complexity of the sparse Bayesian matrix completion is $\mathcal{O}(m \cdot \min((FAD)^3 n^3, k^3) + n \cdot \min((FAD)^3 m^3, k^3))$ [107]. However, this complexity is also dependent on the FAD and estimated rank at each iteration (k). In a sparse Bayesian matrix completion, the effective rank is reduced in the first few iterations and thus, the convergence is rapid. Hence, a Bayesian matrix completion is more computationally efficient than deterministic matrix completion.

5.4 Simulation results

The efficacy of the proposed Bayesian framework is verified on the three phase unbalanced IEEE 37 bus test system [110] as well as IEEE 123 test system [95]. Fig. 5.3 shows the structure of the proposed Bayesian framework consisting of stages 1 and 2. The Bayesian matrix completion based DSSE is represented by (5.21)-(5.31) alongwith the linearized powerflow constraints (3.2), (3.3). An aggregated load profile at the primary node is obtained by combining individual residential loads connected on the secondary side of the feeder. The 24-hr active power consumption profile at a load bus is synthetically generated [23], [111]. Load reactive power can be defined in proportion to the real load connected to the same bus with a power factor of 0.9 lagging. Other load profiles at the load buses of IEEE 37 and IEEE 123 bus test systems are generated by adding an error pattern by using a sinusoidal wave of random amplitude spanning the 24-hr period and random noise terms. With this data, the voltage profile at all nodes is obtained by running the load flow. The aggregated primary feeder measurements (active and reactive power injections) are averaged over 15 min intervals, SCADA measurements (voltage magnitudes) are sampled at 1 min intervals

and the PMU measurements are sampled at 0.1s intervals. These are all real measurements and not pseudo-measurements. The measurement errors are assumed to be 0 mean with a standard deviation equal to 1% of the actual values for both power and voltage measurements [9], [12]. The power injection measurements at a subset of load buses are monitored using the AMI sensors. The first stage consists of reconciling the slow-rate and fast-rate time series data. The time-series data is split into training and testing datasets.

5.4.1 IEEE 37 bus test system

The locations of the sensors used as training dataset are indicated in Fig.5.4. For our case study, we have considered data from 10 aggregated AMI sensors, 9 SCADA sensors, and 3 PMU in the IEEE 37 bus test system whose locations are given as,

1. 10 time-series data from the aggregated AMI sensors are from the following load buses:
 - (a) Phase A - Bus 2, 6, 7, 8.
 - (b) Phase B - Bus 2, 7, 17.
 - (c) Phase C - Bus 2, 5, 7.
2. The 9 time-series data of SCADA sensors are obtained at the three-phase buses 3, 4, and 9.
3. A PMU sensor is placed at bus 25, which measures the voltages at three phases at 0.1s intervals.

The AMI time-series data aggregated at a single bus is sampled at 15-min intervals. Thus, there are 96 measurements over the 24-hr duration for an individual aggregated AMI location. For 10 such aggregated AMI locations, the total measurements are 10×96 . This is the base-case scenario of AMI data. Similarly, the SCADA measurements at a single bus for one phase have 1440 data points obtained over a day. So, for 9 SCADA sensors, the total measurements are 9×1440 . This is the base-case scenario of SCADA data. At bus 25,

we have 3 PMU time-series measurements (as three-phase time-series voltages are obtained) where each time-series data is sampled at 0.1s intervals. It is to be noted that training of the GP prior function can be performed using alternate sets of time-series data, and this setup was used just for demonstration. This is the base-case scenario of PMU data. This base-case scenario of AMI, SCADA, and PMU data represents the training dataset. There are no measurements that were considered to be missing in this training dataset. It is important to note that even without missing measurements, the dataset represents an *unobservable condition* to perform DSSE.

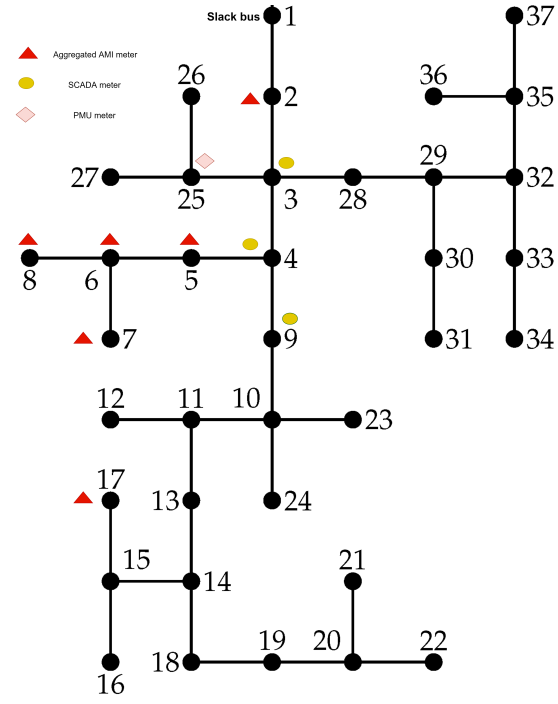


Figure 5.4: *Training dataset for IEEE 37 bus test system*

Using the training dataset, the prior GP function is learned. GP prior mean function consists of a deep neural network with 2 hidden layers and ReLU activation functions. The mean value at a time instant x corresponds to,

$$m_\phi(x) = \mathbf{W}_3\sigma(\mathbf{W}_2\sigma(\mathbf{W}_1(x) + b_1) + b_2) + b_3 \quad (5.33)$$

where, the $\mathbf{W}_i$'s are the weight matrices of the layers, the b_i 's are the biases, and $\sigma(\cdot)$ is the

ReLU activation function. The GP kernel function consists of a radial basis function. The parameters of GP prior are optimized by maximizing the log marginal likelihood. Training of parameters was performed for 100 epochs using stochastic gradient descent (Adam optimizer).

Table 5.1: *Root mean square error of imputed time-series data - 1% Gaussian noise (LI: Linear Interpolation approach, GP: multitask Gaussian process approach)*

System	Type	60% missing		40% missing		20% missing		10% missing	
		LI	GP	LI	GP	LI	GP	LI	GP
IEEE 37 system	Active P	16.8%	4%	5.8%	1.8%	1.82%	0.99%	1.5%	0.7%
	Reactive P	13.25%	2.6%	5.3%	1.4%	1.24%	1.07%	1.3%	0.8%
IEEE 123 system	Active P	22.5%	4.7%	11.39%	2.2%	2.3%	1.7%	1.47%	1.07%
	Reactive P	14.668%	1.798%	7.9%	1.67%	4.76%	1.32%	2.46%	1.27%

The imputation is performed for the test data after the GP prior parameters are optimized. The test data consists of slow-rate and fast-rate measurements with random missing measurements over the 24-hr period. The sensors and the data available in the test dataset are decided by the FAD. Fig.5.8 shows the meter locations for 20% FAD. We use the SCADA, AMI, and PMU measurements from these meters as test data. The observed measurements include the base case scenario (as discussed in the training dataset) minus the missing measurements at the bus locations shown in Fig.5.8. The missing measurements occur for every time-series data corresponding to each sensor. Consider a 60% temporal missing measurement scenario. Each SCADA sensor will have only 576 (1440×0.4) observed measurements randomly placed over the 24-hr duration. Similarly, every PMU sensor will have 345600 (864000×0.4) observed measurements over the 24-hr duration. Each AMI sensor in the test dataset will have 38 observed measurements (96×0.4) over the 24-hr duration.

We have demonstrated the imputation of the unobserved measurements for 1-min interval as well as 0.1s interval [112]. We compare the effective performance of imputation using GP based approach and the simple linear interpolation technique [23]. Both the fast and slow rate measurements are imputed at the same time frame.

It can be inferred from Table 5.1 that the performance gain of GP based approach is highly

pronounced when the percentage of missing measurements is high. The performance of GP based approach and the linear interpolation approach reconciled at the 1-min intervals for a single time-series of active and reactive power at 60% missing measurements is illustrated in Fig.5.5 and Fig.5.6 respectively. Fig.5.5 and Fig.5.6 illustrates the time-series data of aggregated active power at node 8 at phase A and aggregated reactive power injection profile at node 28 of phase C, respectively. The error at 60% missing temporal measurements in GP based approach is only 2.6% as compared to the 13.28% error with the linear interpolation approach. The imputations are also performed at 0.1s intervals. For a 12-hour duration, the corresponding error of GP based imputation at node 17 of phase B is 1.5% compared to the linear interpolation error of 3.5%. The uncertainty bounds associated with the imputed measurements provide a measure of confidence in the imputations. The confidence bounds on the imputed measurements are indicated by a 95% confidence interval. The uncertainty increases as we move away from the measurement. As shown in Fig.5.5, the sudden widening of the confidence region as we move from hour 12 to hour 16 is because of the lack of actual measurements during this time. The linear interpolation technique fails in providing such confidence bounds. As seen from the figures, the GP-based imputation provides accurate as well as smoother imputation as compared to the linear interpolation technique.

We have also simulated the proposed GP approach in the presence of 10% Gaussian noise for power measurements in the IEEE 37 bus test system. Here, the AMI measurements in both the training and test datasets consist of random 10% Gaussian noise. Fig.5.7 shows the performance of GP based approach against the linear interpolation for noisy AMI time-series data at node 2 of phase A where we have considered a 60% temporal missing measurement scenario. We have simulated the performance of GP and linear interpolation using the noisy time-series data for different missing measurements as tabulated in Table 5.3. It can be inferred that the GP-based approach provides smoother and more accurate imputation even when slow-rate measurements are corrupted by higher noise levels.

The predicted mean and the corresponding variances obtained from the consistent time-series are further used in stage 2 for performing state estimation. At any given instant of time, an incomplete observation matrix whose structure illustrated in (3.17) is what is

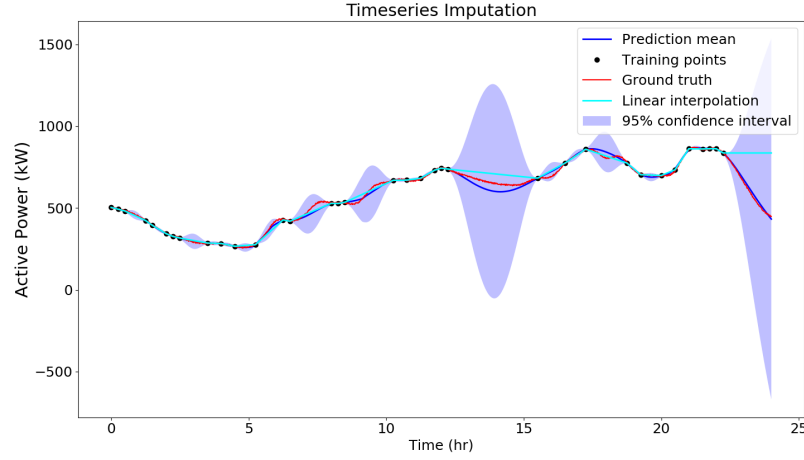


Figure 5.5: *Active power time-series imputation by multi-task GP framework*

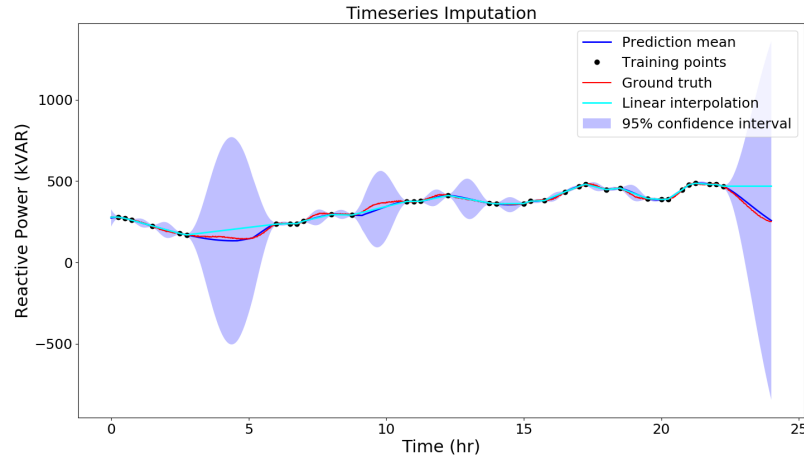


Figure 5.6: *Reactive power time-series imputation by multi-task GP framework*

available in a distribution system. A sparse Bayesian matrix completion approach discussed in section 5.2.2 is used in stage 2 for state estimation. The observation matrix are sampled randomly to reflect different FADs. While choosing the known elements in the matrix $\mathbf{Z}$, the zero injection measurements are also selected. Similar to the FAD, a global measurement redundancy defined in [113] quantifies the ratio of available measurements to the number of states. By performing the SVD of the observed matrix ($\mathbf{Z}$), an initial factorized matrices ($\mathbf{A}$ and $\mathbf{B}$) are obtained given as,

$$\mathbf{Z} = \mathbf{U}\mathbf{D}\mathbf{V}^\top \quad (5.34)$$

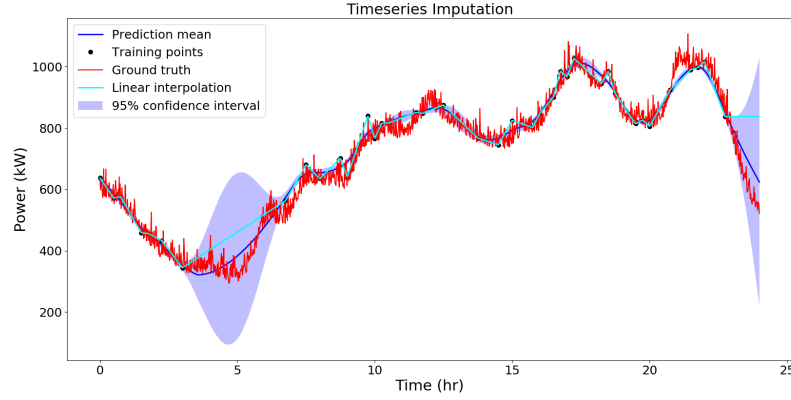


Figure 5.7: Active power time-series imputation by multi-task GP framework - 10% Gaussian Error

with $\mathbf{A} = \mathbf{U}\mathbf{D}^{\frac{1}{2}}$ and $\mathbf{B}^\top = \mathbf{D}^{\frac{1}{2}}\mathbf{V}^\top$. The initial parameters c and d associated with the precisions γ are assumed to be 10^{-6} . The approximate posterior distributions for each of the latent variables are updated using the mean field variational Bayes approach discussed in section 5.2.2. If the observation matrix is obtained from linearly interpolated time-series, the β precisions are also updated. These computations are performed until the convergence criteria $\frac{\|\hat{\mathbf{X}}^{iter} - \hat{\mathbf{X}}^{iter-1}\|_F}{\|\hat{\mathbf{X}}^{iter-1}\|_F} < 10^{-17}$ is satisfied.

In order to quantify the performance of the matrix completion based approaches, MAPE (Mean Absolute Percentage Error) and MIAE (Mean Integrated Absolute Error) metrics are used for voltage magnitude, power and voltage angle, respectively as defined in 4.10 and 4.11 respectively. We consider the IEEE 37 bus unbalanced test system where the FAD is changed from 10% to 90%. We randomly vary the locations and number of measurements for a given FAD.

1. The load buses for an IEEE 37 test system are,
 - (a) Phase A load buses - Bus 2, 6, 7, 8, 13, 18, 19, 24, 27, 30, 31.
 - (b) Phase B load buses - Bus 2, 7, 17, 21, 23, 24, 27, 30, 34, 36, 37.
 - (c) Phase C load buses - Bus 2, 5, 7, 9, 12, 14, 16, 22, 24, 26, 28, 36.

Column 1 and Column 2 of matrix $\mathbf{Z}$ consist of active and reactive power injection

measurements of all the phases of the non-slack buses. A FAD of 100% represents the availability of load measurements at each phase of the non-slack load buses.

2. Column 3 and Column 4 correspond to real and imaginary parts of voltages. For a FAD of 100%, all these measurements are available.
3. All the voltage magnitude measurements in Column 5 of matrix $\mathbf{Z}$ are available for a FAD of 100%.

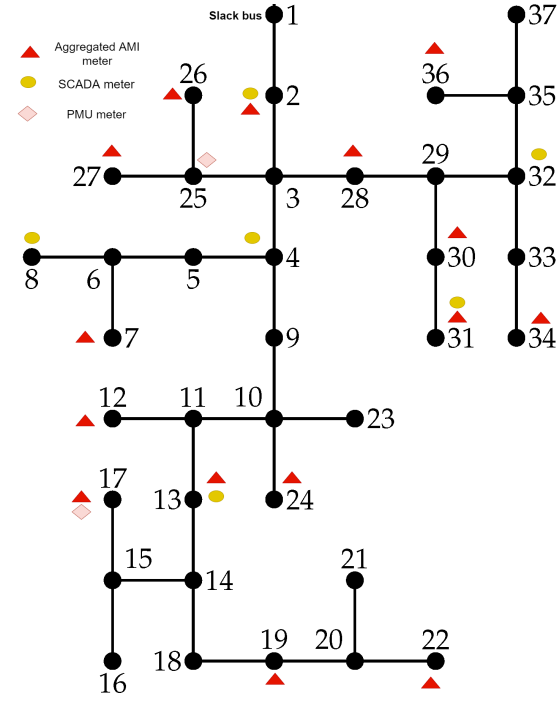
Hence, the maximum number of possible entries in columns 1 and 2 of $\mathbf{Z}$ is 68. The maximum possible entries of voltage magnitudes (Column 3) and voltage phasors (Columns 4 and 5) are 324. For a FAD of 20%, we have presented the position and type of measurements on the IEEE 37 bus test system in Fig. 5.8. Specifically, we have placed 6 SCADA sensors, and 2 PMU sensors to obtain 30 non-zero entries in columns 3-5. The rest of the measurements in columns 1-2 are obtained via the aggregated smart meter measurements. Thus, for FAD of 20%, we have a partially observed matrix $\mathbf{Z}$ with 78 total measurement entries. This case represents a highly unobservable distribution system. In this scenario, the states are voltage magnitude and voltage angles while the measurements are obtained from aggregated smart meters, PMU and SCADA. The number of states at non-slack buses for the IEEE 37 test system is 218. The total measurements are 78. Hence, the global measurement redundancy is $78/218 = 0.357$.

Fig. 5.9, 5.10 and 5.11 show the errors in the estimated spatial states obtained from multitask GP output corresponding to 60% temporal missing measurements and measurement noise level corresponding to 1% and 10%. We have considered measurements obtained from the consistent time-series data at 5th hour. Fig. 5.9 shows the comparative performance of power recovery using GP based and linearly interpolated time-series data followed by state estimation using Bayesian matrix completion for IEEE 37 test system. It can be inferred that estimates obtained from the GP outperform the linearly interpolated time-series at all FADs. For example, in the 1% noise scenario, error reductions up to 10% are achieved in the power recovery at 20% FAD. Fig. 5.10 and Fig. 5.11 show the superior performance of GP

Table 5.2: *Root mean square error of imputed time-series data - 10% Gaussian noise*

System	Type	60% missing		40% missing		20% missing		10% missing	
IEEE 37 test system		Linear Int	GP	Linear Int	GP	Linear Int	GP	Linear Int	GP
	Active P	15.7%	5%	9.09%	4.85%	5.12%	4.35%	4.25%	4.12%
	Reactive P	11.29%	6.9%	4.54%	3.27%	2.81%	2.79%	2.75%	2.66%

in the recovery of voltage magnitude as well as the angle at all the FADs, respectively. As expected, the performance of state estimation degrades with 10% noise in measurements, as seen from the results. However, the proposed Bayesian MC based DSSE approach ensures that the MAPE and MIAE of voltage estimation are less than 2%.

**Figure 5.8:** *Measurement configuration at 20% FAD for three phase unbalanced IEEE 37 bus test system.*

5.4.2 IEEE 123 bus test system

The proposed Bayesian approach is also simulated on three phase unbalanced IEEE 123 test system. In this system, the Bayesian approach has better performance than a simple linear

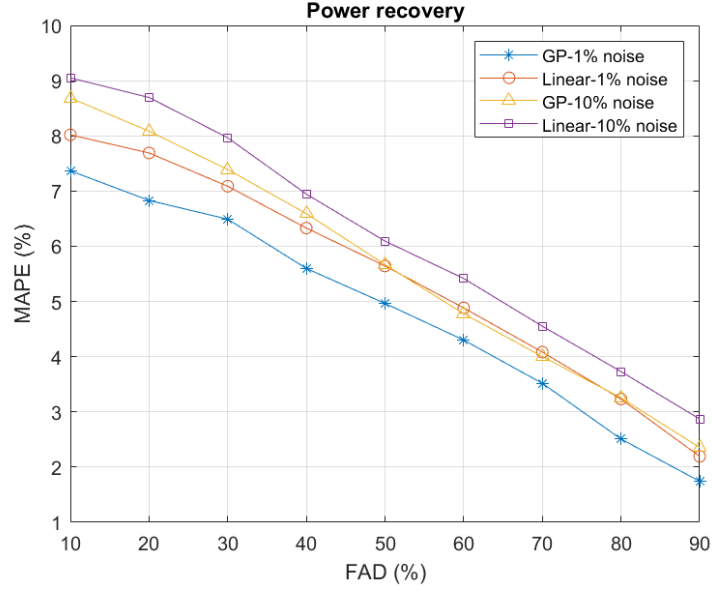


Figure 5.9: Power recovery performance at different FADs for IEEE 37 test system corresponding to 1% noise and 10% noise

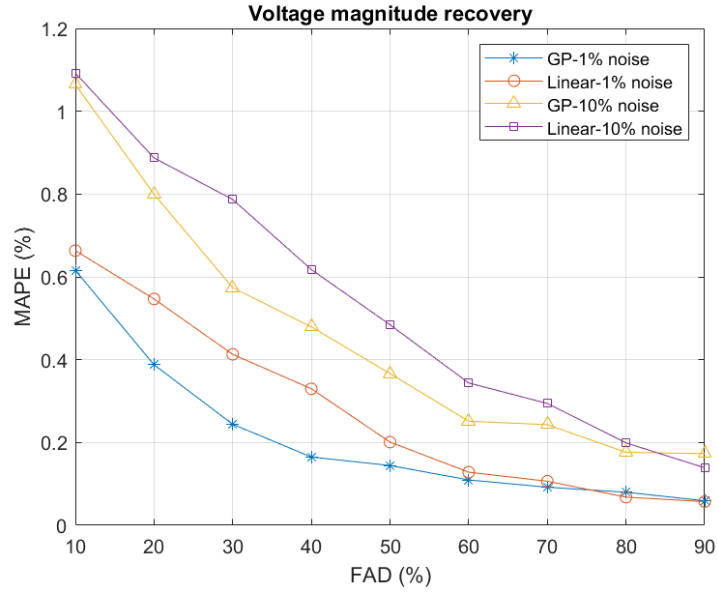


Figure 5.10: Voltage magnitude recovery performance at different FADs for IEEE 37 test system corresponding to 1% noise and 10% noise

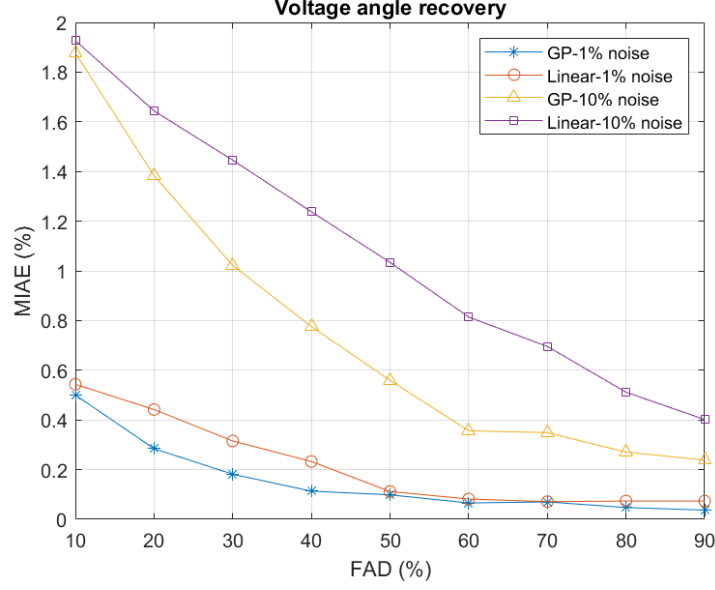


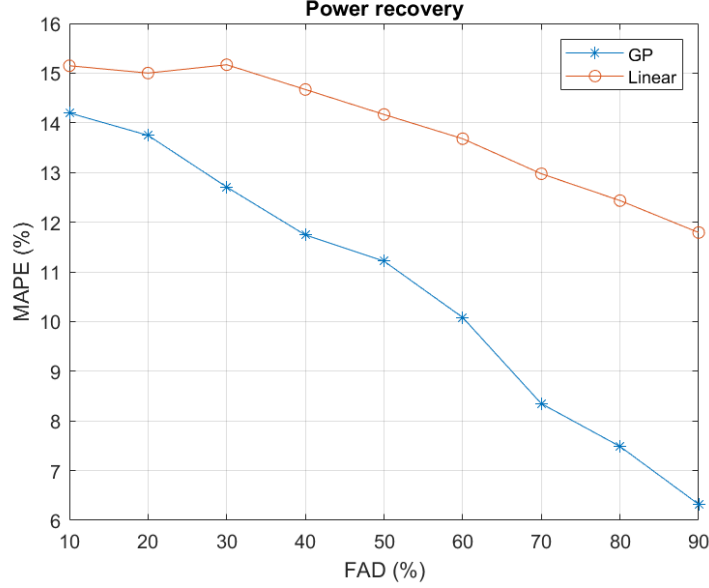
Figure 5.11: Voltage angle recovery performance at different FADs for IEEE 37 test system corresponding to 1% noise and 10% noise

interpolation based technique. The recovery of power using the two methods is shown in Fig. 5.12. Comparative performance of voltage magnitude and voltage angle recovery between the two approaches is illustrated in Fig. 5.13 and Fig. 5.14 respectively. The GP-based approach estimates the states with high accuracy as compared to the linear interpolation-based approach.

The mean and variance of computational times based on monte-carlo simulation and FAD of 75% for the Bayesian matrix completion and deterministic matrix completion approach is tabulated in Table.5.3. In accordance with the complexity analysis in section 5.3, the time complexity of Bayesian MC is reduced at lower FADs. These time calculations are carried out on Intel i9 core, 32 GB RAM with CVX Mosek package [114]. From Table.5.3, it is evident that the proposed Bayesian MC is an order faster than the deterministic MC. This is due to the fact that the computation of the SVD of the matrix at each iteration is eliminated.

Table 5.3: Average time requirement ($FAD = 75\%$)

Test system	Deterministic MC	Bayesian MC
IEEE 37 test system	Mean - 1.267s, Variance - 0.42	Mean - 0.02s, Variance - 3.037×10^{-5}
IEEE 123 test system	Mean - 8.1s, Variance - 1.32	Mean - 0.48s, Variance - 1.4733×10^{-4}

**Figure 5.12:** Power recovery performance at different FADs for IEEE 123 test system

5.5 Summary

This chapter proposes a Bayesian approach for effectively aggregating heterogeneous intermittent time-series data and using it to accurately estimate the distribution system states in low-observability conditions. The proposed approach leverages the spatio-temporal correlations in multi- timescale measurements. Superior performance in the recovery of power, as well as voltage states, are obtained in the proposed matrix completion approach. The probabilistic formulation allows us to compute the uncertainty in both data imputation and state estimation while being computationally efficient. In the presence of noisy measurements, the uncertainty associated with the imputation can be quantified. Knowledge of the variances obtained in the imputation stage provides different noise precision parameters (β),

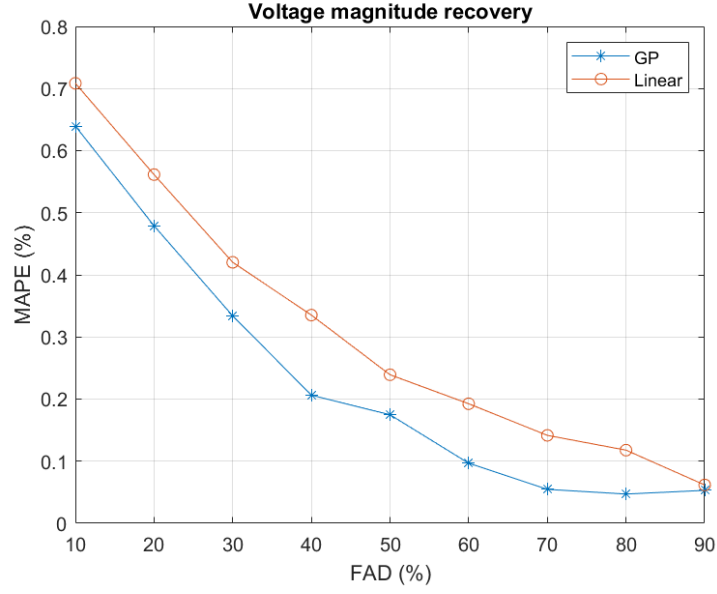


Figure 5.13: Voltage magnitude recovery performance at different FADs for IEEE 123 test system

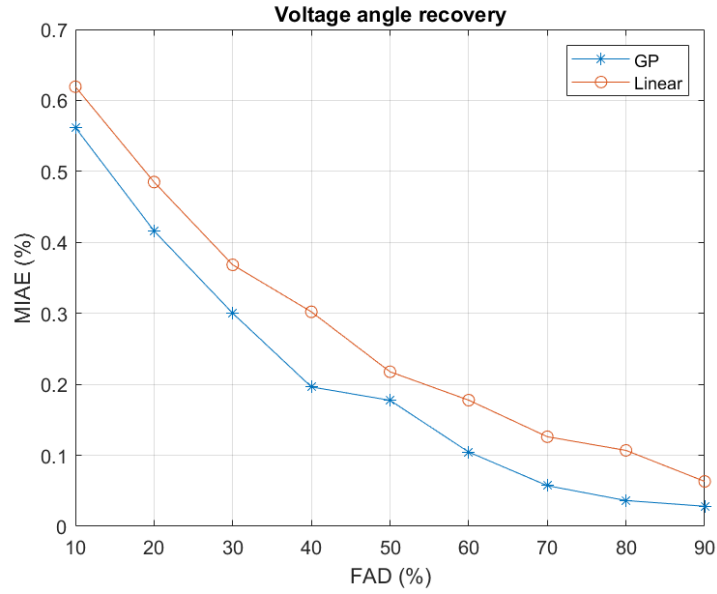


Figure 5.14: Voltage angle recovery performance at different FADs for IEEE 123 test system

which can then be used to guide the state estimation process. Test studies in IEEE 37, as well as IEEE 123 bus system, reveal that the power and voltage states are recovered with high fidelity.

Chapter 6

Streaming multi time-scale data using Recursive GP approach

The multi-task GP approach discussed in Chapter 5 performs imputations using all the measurements at once. That is, the multi-task GP approach involves batch processing and cannot be used to perform imputations in real-time as measurements arrive. Furthermore, this approach does not exploit the graphical structure of the grid. To address this limitations, this chapter proposes a recursive multi-task Gaussian process approach that sequentially aggregates multi-time scale measurements while incorporating the network connectivity information.

6.1 Background

Consider a distribution system which can be perceived as a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ where $\mathcal{V} \in \mathbb{R}^M$ are the nodes and $\mathcal{E}$ denotes the edges. The adjacency matrix $\mathbf{A}$ is defined as,

$$\mathbf{A}(i, j) = \begin{cases} 1, & (i, j) \in \mathcal{E} \\ 0, & \text{otherwise} \end{cases} \quad (6.1)$$

In the distribution system, sensors are placed at a subset of M nodes. These sensors measure power injections or voltages at different locations in the network at different sampling rates. Hence, the main goal is to first reconcile these multi time-series measurements at the desired time scale and estimate the states. To do so, we sequentially process the measurements using a recursive multi-task Gaussian process-aided state estimation approach. This approach exploits the spatial and temporal correlations of the sequential measurements located on the graph $\mathcal{G}$. We propose to leverage the graphical structure of the grid for better imputation of the unevenly sampled measurements. Here, we consider the topology of the primary distribution system. Also, the imputation of the multi time-scale measurements is performed at the primary feeder. Therefore, the proposed approach does not rely on the model information of the secondary side of the network. Based on the available network connectivity information, we propose two approaches, as shown in Fig.6.1. The inputs in all these approaches are the unevenly sampled time-series measurements, and the output is the coherent set of measurements along with their variance. Conventional full GP and multitask Recursive GP (RGP) process the unevenly sampled measurements without utilizing any graph structure. Recursive processing of heterogeneous measurements using graph information in real-time and batch mode is performed using RGP-G interpolation and RGP-G prediction, respectively. We will review the relevant concepts of graph signal processing before introducing the proposed approaches.

Background of Graph signal processing

The graph Laplacian matrix $\mathbf{L}$ for graph $\mathcal{G}$ is defined as $\mathbf{L} = \mathbf{D} - \mathbf{A}$ where $\mathbf{D}$ is the diagonal degree matrix whose i th diagonal element is given by the sum of the elements in the i th row of $\mathbf{A}$. The observations $\mathbf{y} = [y(1), \dots, y(M)] \in \mathbb{R}^M$ represents a signal on graph $\mathcal{G}$. The signal variation of $\mathbf{y}$ over graph $\mathcal{G}$ is measured as,

$$l(\mathbf{y}) = \sum_{(i,j) \in \mathcal{E}, i \neq j} A(i,j)((y(i) - y(j))^2 = \mathbf{y}^T \mathbf{L} \mathbf{y}$$

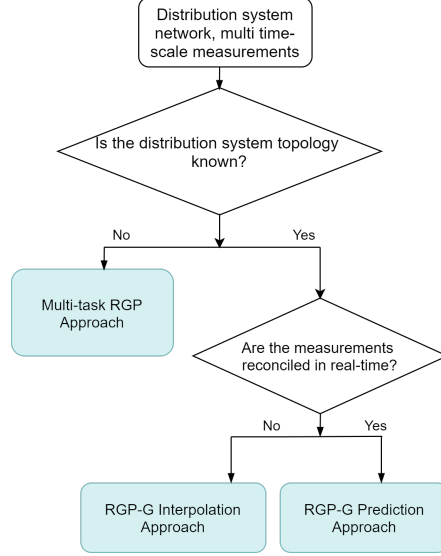


Figure 6.1: *Classification of the proposed approaches for processing streaming data*

The Laplacian quadratic form $\mathbf{y}^T \mathbf{L} \mathbf{y}$ denotes the smoothness of the $\mathbf{y}$. Suppose we want to recover the smooth signal $\mathbf{y}_d$ from noisy observation $\mathbf{y} = \mathbf{y}_d + \mathbf{w}$, over the graph $\mathcal{G}$. In order to recover this signal, an optimization problem is formulated as,

$$\mathbf{y}_d^* = \min_{\mathbf{y}_d \in \mathbb{R}^M} \|\mathbf{y} - \mathbf{y}_d\|_2^2 + \alpha \mathbf{y}_d^T \mathbf{L} \mathbf{y}_d \quad (6.2)$$

where $\alpha \geq 0$. The global solution is,

$$\mathbf{y}_d^* = (\mathbf{I}_M + \alpha \mathbf{L})^{-1} \mathbf{y} \quad (6.3)$$

Here, $\mathbf{I}_M$ is the identity matrix. The optimal solution $\mathbf{y}_d^*$ can be seen as the graph filtering of $\mathbf{y}$ using the graph filter $\mathbf{S} = (\mathbf{I}_M + \alpha \mathbf{L})^{-1}$ [115][116]. This graph filter will be used in the proposed RGP-G approach for inducing the graph structure of the distribution system. In the next section, we will review the conventional full GP approach.

6.1.1 Full GP Approach

Consider a distribution system with M buses and d types of sensor tasks. Here, the sensor tasks refer to different sensor measurements available, e.g., aggregated active and reactive power injections or voltage magnitudes at the primary feeder. We consider the availability of measurements for T time instances. Let the measurements corresponding to time instant x_t be $\mathbf{y}_t \in \mathbb{R}^{dM}$. The measurements $\mathbf{y}_t$ are obtained by concatenating measurements from different sensor locations i.e., $\mathbf{y}_t = [\mathbf{y}_t^1, \dots, \mathbf{y}_t^d]^\top$. Each $\mathbf{y}_t^d$ is drawn from a noisy process as,

$$\mathbf{y}_t^d(x_t) = \mathcal{N}(\mathbf{f}_t^d(x_t), \sigma_\epsilon^2 \mathbf{I}_M) \quad (6.4)$$

where, $\mathbf{f}_t^d \in \mathbb{R}^M$ and σ_ϵ^2 is the noise variance. The entries in $\mathbf{y}_t^d$ are zero at the locations where there are no sensor measurements.

The GP prior function $\mathbf{f}_t^d$ associated with d^{th} sensor task at time t has distribution given as,

$$\mathbf{f}_t^d = \mathcal{N}(\mathbf{0}, \mathcal{K}(x_t, x_t^\top) \mathbf{I}_M) \quad (6.5)$$

The function $\mathbf{f}_t = [\mathbf{f}_t^1, \dots, \mathbf{f}_t^d]^\top$ is a Gaussian prior with distribution,

$$\mathbf{f}_t(x_t) = \mathcal{N}(\mathbf{0}, \mathbf{I}_c \otimes \mathcal{K}(x_t, x_t^\top) \mathbf{I}_M) \quad (6.6)$$

where $\otimes$ denotes the Kronecker product, $\mathbf{I}_c \in \mathbb{R}^{d \times d}$ is an identity matrix between different sensor tasks in the distribution grid. For instance, a distribution grid may have active power (**P**), reactive power injections (**Q**), and voltage (**V**) measurements at the primary feeder. Thus, there are three sensors tasks, and $\mathbf{I}_c$ has a size of 3×3 . In some distribution systems, there are only **P** and **Q** measurements available. The voltage measurements are available only at the substation. Therefore, in this case, the size of $\mathbf{I}_c$ is 2×2 . The kernel matrix $\mathcal{K}$ represents the temporal covariance functions within this sensor task. There are different kernel choices, with one of the most popular being RBF (radial basis function) kernel [103]

corresponding to:

$$\mathcal{K}(x_1, x_2) = \sigma_s^2 \exp \frac{-(x_1 - x_2)^2}{2l^2} \quad (6.7)$$

where hyperparameters l and σ_s^2 are the length-scale and signal variance respectively. The lengthscale of the kernel function controls the smoothness of the GP function [103].

If all the measurements upto time T are represented as $\tilde{\mathbf{y}} = \text{vec}(\mathbf{y}_1, \dots, \mathbf{y}_T)$, the distribution of $\tilde{\mathbf{y}}$ using (6.4) and (6.6) is given as,

$$\tilde{\mathbf{y}} = \mathcal{N}(\mathbf{0}, ((\mathbf{I}_c \otimes \mathbf{I}) \otimes \mathbf{K}) + \sigma_\epsilon^2 \mathbf{I}) \quad (6.8)$$

where $\mathbf{K}$ is the kernel matrix defined for all time instances $t = 1, \dots, T$. Here, the i^{th} and j^{th} entry of $\mathbf{K}$ is given as $\mathbf{K}_{ij} = \mathcal{K}(x_i, x_j)$ and σ_ϵ^2 is the noise variance.

The main goal of the Gaussian process-based imputation process is to infer the unknown test values $\mathbf{y}_*$ corresponding to the time $\mathbf{x}_*$ given the measurements $\tilde{\mathbf{y}}$ at time $\mathbf{x}$ and the modeled GP prior function $\mathbf{f}(\cdot)$. The measurements $\tilde{\mathbf{y}}$ and the test values $\mathbf{y}_*$ are jointly Gaussian whose distribution is given as,

$$\begin{pmatrix} \tilde{\mathbf{y}} \\ \mathbf{y}_* \end{pmatrix} \sim \mathcal{N} \left(\mathbf{0}, \begin{pmatrix} \mathbf{A} & \mathbf{D} \\ \mathbf{D}^\top & \mathbf{F} \end{pmatrix} \right) \quad (6.9)$$

where, the matrices $\mathbf{A}$, $\mathbf{D}$ and $\mathbf{F}$ corresponds to,

$$\mathbf{A} = (\mathbf{I}_c \otimes \mathbf{I}) \otimes \mathbf{K} + \sigma_\epsilon^2 \mathbf{I}, \quad (6.10)$$

$$\mathbf{D} = (\mathbf{I}_c \otimes \mathbf{I}) \otimes \mathbf{K}_* + \sigma_\epsilon^2 \mathbf{I}, \quad (6.11)$$

$$\mathbf{F} = (\mathbf{I}_c \otimes \mathbf{I}) \otimes \mathbf{K}_{**} + \sigma_\epsilon^2 \mathbf{I}. \quad (6.12)$$

Here, $\mathbf{K}_* = \mathcal{K}(\mathbf{x}, \mathbf{x}_*)$, $\mathbf{K}_{**} = \mathcal{K}(\mathbf{x}_*, \mathbf{x}_*)$ and $\mathbf{x} = [x_1, \dots, x_T]^\top$.

The conditional distribution of the test values $\mathbf{y}_*$ given $\tilde{\mathbf{y}}$ is a Gaussian distribution [103]

with mean and covariance,

$$\mathbf{m}^* = \mathbf{D}^\top \mathbf{A}^{-1} \tilde{\mathbf{y}} \quad (6.13)$$

and

$$\mathbf{C}^* = \mathbf{F} - \mathbf{D}^\top \mathbf{A}^{-1} \mathbf{D} \quad (6.14)$$

Algorithm 3 summarizes the full-GP approach. (6.13) and (6.14) involves inverting the matrix $\mathbf{A}$ for all the $\mathbf{x}$ time instances which is computationally expensive. The full GP approach suffers from the following drawbacks:

- The GP prior function and the corresponding measurements $\tilde{\mathbf{y}}$ as defined in (6.8) is a simple multi-task Gaussian process with an independent kernel function among the different measurements obtained at M nodes.
- This approach performs training using all the measurements in the batch, and thus the training is performed off-line in a batch mode.
- The computational complexity is $\mathcal{O}((TdM)^3)$, where T is the size of $\mathbf{x}$, d is the total number of sensor tasks, and M are the nodes. The inversion of the matrix $\mathbf{A}$ is the key contributor to this complexity.

To overcome these challenges, we propose a recursive GP approach that sequentially processes the measurements corresponding to each x_t by using the knowledge of graphical structure of the distribution grid. In the next section, we will formulate the RGP-G approach when the network connectivity information is known. Then, we will develop the RGP-G method when the graph information is unknown.

6.2 Proposed Recursive GP approach

This section presents the formulation for recursively imputing the multi time-scale measurements with and without topology information.

Algorithm 3 Full GP Approach

Input: Aggregated Active and Reactive power injection measurements at load bus $\mathbf{y}_t$ corresponding to time $\mathbf{x}_t$, $\tilde{\mathbf{y}} = \text{vec}(\mathbf{y}_1, \dots, \mathbf{y}_T)$, Kernel choice and hyper-parameters associated to kernel function

- 1: Calculate the kernel matrix $\mathbf{K}$ that exploits the temporal correlation using any kernel function (e.g., RBF kernel).
 - 2: Calculate the matrix $\mathbf{I}_c \in \mathbb{R}^{d \times d}$.
 - 3: Calculate the matrix $\mathbf{A}$, $\mathbf{D}$ and $\mathbf{F}$ by means of (6.10), (6.11) and (6.12) respectively.
 - 4: Perform imputation at time x_* by means of the mean $\mathbf{m}^*$ and covariance matrix $\mathbf{C}^*$ using (6.13) and (6.14).
 - 5: **return** $\mathbf{m}^*$, $\mathbf{C}^*$.
-

6.2.1 RGP-G Approach

One of the challenges in the full GP approach is the need to use the complete vector $\mathbf{x} \in \mathbb{R}^T$. To overcome this challenge, we aim to use the basis vectors $\mathbf{x} = [x_1, x_2, \dots, x_n]^\top$ where $n \ll T$. We perform all the calculations on the basis vectors $\mathbf{x} \in \mathbb{R}^n$ which are fixed in number and locations. The function $\mathbf{f} = \mathbf{f}(\mathbf{x})$ is the GP function corresponding to the basis vectors $\mathbf{x}$. When the network connectivity information is known, we can construct the graph filter matrix $\mathbf{S}$ as defined in (6.3). In order to exploit the spatial correlation and induce the graph information, the observations in the GP function (4) and (5) are modified as,

$$\mathbf{y}_t^d(x_t) = \mathcal{N}(\mathbf{S}\mathbf{f}_t^d(x_t), \sigma_\epsilon^2 \mathbf{I}_M) \quad (6.15)$$

where, $\mathbf{f}_t^d$ is defined in (6.5). Using (6.5) and (6.15), we obtain the distribution of $\mathbf{y}_t^d$ as,

$$\mathbf{y}_t^d = \mathcal{N}(\mathbf{0}, \mathbf{S}\mathcal{K}(x, x^\top)\mathbf{S}^\top + \sigma_\epsilon^2 \mathbf{I}_M) \quad (6.16)$$

The distribution of $\mathbf{y}_t$ obtained by concatenating $\mathbf{y}_t^d$ is given as,

$$\mathbf{y}_t(x_t) = \mathcal{N}(\mathbf{0}, (\mathbf{K}_c \otimes \mathbf{S}^2)\mathcal{K}(x_t, x'_t) + \sigma_\epsilon^2 \mathbf{I}) \quad (6.17)$$

Here, the identity matrix $\mathbf{I}_c$ given in (6.6) is replaced by the kernel matrix $\mathbf{K}_c$ that represents correlation among different sensor tasks.

The main aim of this section is to recursively update the mean and covariance of the multi-task prior function $\mathbf{f}$ as the measurements arrive at time $t = 1, \dots, T$ by incorporating the graph structure of the grid. We assume that the hyperparameters of the kernel function are known apriori using historical-based data. The proposed RGP-G approach has the flexibility of performing both interpolation and prediction described as,

1. RGP-G Interpolation- This approach operates over a set time frame (24 hours as an example). The GP function is updated recursively at those time instances where the measurements are obtained. Once the GP function in that batch is updated, we perform imputation at the finest time resolution. Here, the finest time resolution refers to the narrowest time resolution between the different measurement sources.
2. RGP-G Prediction- This approach reconciles the multi-time scale measurements in real-time. The imputation at the desired time resolution is performed by predicting the GP function until the subsequent measurements is observed. Here, the prediction is performed at the finest time resolution.

RGP-G Interpolation

We assume that the GP prior function $\mathbf{f}$ at time $t = 0$ has an initial distribution,

$$p_0(\mathbf{f}) = \mathcal{N}(\mathbf{f}; \boldsymbol{\mu}_{g,0}^{\mathbf{f}}, \mathbf{C}_{g,0}^{\mathbf{f}}) \quad (6.18)$$

with mean $\boldsymbol{\mu}_{g,0}^{\mathbf{f}} = \mathbf{0}$ and covariance $\mathbf{C}_{g,0}^{\mathbf{f}}$ defined as,

$$\mathbf{C}_{g,0}^{\mathbf{f}} = (\mathbf{K}_c \otimes \mathbf{S}^2) \otimes \mathbf{K}. \quad (6.19)$$

Here, $\mathbf{K} = \mathcal{K}(\mathbf{x}, \mathbf{x})$. The initial covariance of the GP prior function exploits the spatial correlation using graph structure and the temporal correlation between the time instances. The measurements $\mathbf{y}_t \in \mathbb{R}^{dM}$ arrive sequentially at time $t = 1, \dots, T$. The goal is to calculate

the posterior distribution

$$p(\mathbf{f}|\mathbf{y}_{1:t}) = \mathcal{N}(\mathbf{f}; \boldsymbol{\mu}_{g,t}^{\mathbf{f}}, \mathbf{C}_{g,t}^{\mathbf{f}}) \quad (6.20)$$

at time t , where $\mathbf{y}_{1:t} = (\mathbf{y}_1, \dots, \mathbf{y}_t)$, by combining the new measurements $\mathbf{y}_t$ with the distribution,

$$p_{t-1}(\mathbf{f}|\mathbf{y}_{1:t-1}) = \mathcal{N}(\mathbf{f}; \boldsymbol{\mu}_{g,t-1}^{\mathbf{f}}, \mathbf{C}_{g,t-1}^{\mathbf{f}}) \quad (6.21)$$

The desired posterior distribution is expanded according to [117],

$$p(\mathbf{f}|\mathbf{y}_{1:t}) = \int \underbrace{c_t \cdot p(\mathbf{y}_t|\mathbf{f}, \mathbf{f}_t)}_{p(\mathbf{f}, \mathbf{f}_t|\mathbf{y}_{1:t}) \text{ update}} \cdot \underbrace{\overbrace{p(\mathbf{f}_t|\mathbf{f}) \cdot p(\mathbf{f}|\mathbf{y}_{1:t-1})}^{p(\mathbf{f}, \mathbf{f}_t|\mathbf{y}_{1:t-1}) \text{ inference}}}_{p(\mathbf{f}, \mathbf{f}_t|\mathbf{y}_{1:t}) \text{ update}} d\mathbf{f}_t$$

where, $\mathbf{f}_t$ is the GP function at time t and c_t is the normalization constant.

Calculation of the posterior is performed in two steps:

a) Inference: In this step, we infer the joint prior $p(\mathbf{f}, \mathbf{f}_t|\mathbf{y}_{1:t-1})$ using the measurements received upto time $t - 1$. Here, the matrices $\mathbf{A}_g, \mathbf{D}_g, \mathbf{F}_g$ calculated at time x_t be defined as,

$$\mathbf{A}_g = (\mathbf{K}_c \otimes \mathbf{S}^2) \otimes \mathbf{K} + \sigma_\epsilon^2 \mathbf{I}, \quad (6.22)$$

$$\mathbf{D}_g = (\mathbf{K}_c \otimes \mathbf{S}^2) \otimes \mathbf{K}_t + \sigma_\epsilon^2 \mathbf{I}, \quad (6.23)$$

$$\mathbf{F}_g = (\mathbf{K}_c \otimes \mathbf{S}^2) \otimes \mathcal{K}_{tt} + \sigma_\epsilon^2 \mathbf{I} \quad (6.24)$$

Here, the subscript g refers to notations related to recursive GP with graphs approach. The goal is to calculate the joint prior $p(\mathbf{f}, \mathbf{f}_t|\mathbf{y}_{1:t-1})$ using the information from the prior $p(\mathbf{f}|\mathbf{y}_{1:t-1})$. This can be achieved using the chain rule as,

$$p(\mathbf{f}, \mathbf{f}_t|\mathbf{y}_{1:t-1}) = p(\mathbf{f}_t|\mathbf{f}) \cdot p(\mathbf{f}|\mathbf{y}_{1:t-1}) \quad (6.25)$$

$$= \mathcal{N}(\mathbf{f}_t; \boldsymbol{\mu}_{g,t}^{\mathbf{p}}, \mathbf{B}_g) \cdot \mathcal{N}(\mathbf{f}; \boldsymbol{\mu}_{g,t-1}^{\mathbf{f}}, \mathbf{C}_{g,t-1}^{\mathbf{f}}) \quad (6.26)$$

The first term $p(\mathbf{f}_t|\mathbf{f})$ follows from the assumption that $\mathbf{f}_t$ is conditionally independent of the

past measurements $\mathbf{y}_{1:t-1}$ given $\mathbf{f}$. As any finite representation of a GP is Gaussian, the joint prior is also Gaussian. Hence, the conditional distribution $p(\mathbf{f}_t|\mathbf{f})$ is Gaussian and calculated by Gaussian identities given as,

$$\boldsymbol{\mu}_{g,t}^{\mathbf{p}} = \mathbf{J}_{g,t} \boldsymbol{\mu}_{g,t-1}^{\mathbf{f}} \quad (6.27)$$

$$\mathbf{B}_g = \mathbf{F}_g - \mathbf{J}_{g,t} \mathbf{D}_g \quad (6.28)$$

$$\mathbf{J}_{g,t} = \mathbf{D}_g^{\top} \mathbf{A}_g^{-1} \quad (6.29)$$

Using Gaussian identities and Woodbury formula, the solution to (6.26) is a joint Gaussian $p(\mathbf{f}, \mathbf{f}_t | \mathbf{y}_{1:t-1}) = \mathcal{N}(\mathbf{q}, \mathbf{Q})$ with mean and covariance defined as,

$$\mathbf{q} = \begin{bmatrix} \boldsymbol{\mu}_{g,t}^{\mathbf{f}} \\ \boldsymbol{\mu}_{g,t}^{\mathbf{p}} \end{bmatrix} \quad (6.30)$$

and,

$$\mathbf{Q} = \begin{bmatrix} \mathbf{C}_{g,t}^{\mathbf{f}} & \mathbf{C}_{g,t-1}^{\mathbf{f}} \mathbf{J}_{g,t}^{\top} \\ \mathbf{J}_{g,t} \mathbf{C}_{g,t-1}^{\mathbf{f}} & \mathbf{C}_{g,t}^{\mathbf{p}} \end{bmatrix} \quad (6.31)$$

where,

$$\mathbf{C}_{g,t}^{\mathbf{p}} = \mathbf{B}_g + \mathbf{J}_{g,t} \mathbf{C}_{g,t-1}^{\mathbf{f}} \mathbf{J}_{g,t}^{\top} \quad (6.32)$$

b) Update: This step updates the joint prior $\mathbf{f}$ with new measurements $\mathbf{y}_t$ arriving at time t . The function $\mathbf{f}_t$ is updated by Kalman filter update step which yields,

$$p(\mathbf{f}_t | \mathbf{y}_{1:t}) = \mathcal{N}(\mathbf{f}_t; \boldsymbol{\mu}_{g,t}^{\mathbf{e}}, \mathbf{C}_{g,t}^{\mathbf{e}}) \quad (6.33)$$

where,

$$\boldsymbol{\mu}_{g,t}^{\mathbf{e}} = \boldsymbol{\mu}_{g,t}^{\mathbf{p}} + \mathbf{G}_t (\mathbf{y}_t - \boldsymbol{\mu}_{g,t}^{\mathbf{p}}), \quad (6.34)$$

$$\mathbf{C}_{g,t}^{\mathbf{e}} = \mathbf{C}_{g,t}^{\mathbf{p}} - \mathbf{G}_t (\mathbf{C}_{g,t}^{\mathbf{p}}), \quad (6.35)$$

Here, $\boldsymbol{\mu}_{g,t}^{\mathbf{p}}$ and $\mathbf{C}_{g,t}^{\mathbf{p}}$ are obtained from (6.27) and (6.32), respectively. Here, $\mathbf{G}_t = \mathbf{C}_{g,t}^{\mathbf{p}} (\mathbf{C}_{g,t}^{\mathbf{p}} +$

$\sigma_\epsilon^2 \mathbf{I})^{-1}$ is the Kalman gain. The update is performed at time t where measurements are available. For instance, the aggregated smart meter at the load buses provides measurements at intervals $t = 1, 16, 31, \dots, T$. At $t = 16$, assume that few of the measurements $\mathbf{y}_t$ are missing due to communication bottleneck. Therefore, at those locations, we do not update but use the predicted measurements $\boldsymbol{\mu}_{g,t}^p$. The posterior function $\mathbf{f}$ has Gaussian distribution $\mathcal{N}(\mathbf{f}; \boldsymbol{\mu}_{g,t}^f, \mathbf{C}_{g,t}^f)$ which is defined as,

$$\boldsymbol{\mu}_{g,t}^f = \boldsymbol{\mu}_{g,t-1}^f + \tilde{\mathbf{G}}_t \cdot (\mathbf{y}_t - \boldsymbol{\mu}_{g,t}^p), \quad (6.36)$$

$$\mathbf{C}_{g,t}^f = \mathbf{C}_{g,t-1}^f - \tilde{\mathbf{G}}_t \mathbf{J}_{g,t} \mathbf{C}_{g,t-1}^f, \quad (6.37)$$

$$\tilde{\mathbf{G}}_t = \mathbf{C}_{g,t-1}^f \mathbf{J}_{g,t}^\top (\mathbf{C}_{g,t}^p + \sigma_\epsilon^2 \mathbf{I})^{-1}. \quad (6.38)$$

The function $\mathbf{f}$ is sequentially updated with the observations $\mathbf{y}_t$ until $t = T$. Once updated, the imputation of $\mathbf{y}_*$ at time x_* is performed using the following steps,

$$\mathbf{m}_g^* = \mathbf{D}_g^\top \mathbf{A}_g^{-1} \boldsymbol{\mu}_{g,T}^f \quad (6.39)$$

and

$$\mathbf{C}_g^* = \mathbf{B}_g + \mathbf{J}_{g,t^*} (\mathbf{C}_{g,T}^f) \mathbf{J}_{g,t^*}^\top \quad (6.40)$$

where the matrices $\mathbf{D}_g$, $\mathbf{A}_g$ and $\mathbf{B}_g$ are evaluated for time x_* . The complete RGP-G Interpolation approach is illustrated in Algorithm 4.

RGP-G Prediction Approach

The RGP-G interpolation approach operates over a set time frame. However, it is critical to reconcile the measurements as and when they arrive. The RGP-G prediction approach achieves the reconciliation in real-time by performing the step ahead prediction of the GP function using the knowledge of the past measurements. These predictions are the imputed values at the narrowest time resolution.

The complete algorithm of the proposed sequential prediction over graphs is summarized

Algorithm 4 RGP-G Interpolation Approach

Input: Basis vector $\mathbf{x}$, Distribution system graph laplacian $\mathbf{L} \in \mathbb{R}^{M \times M}$, total time instants T , α , $\mathbf{S}$, $\mathbf{K}$, $\mathbf{K}_c$, $\mathbf{y}_t$

- 1: Initialization: $\boldsymbol{\mu}_{g,0}^{\mathbf{f}} = \mathbf{0}$ and $\mathbf{C}_{g,0}^{\mathbf{f}} = (\mathbf{K}_c \otimes \mathbf{S}^2) \otimes \mathbf{K}$.
- 2: **for** $t = 1, \dots, T$ **do**
- 3: Calculate the gain matrix $\mathbf{J}_{g,t}$ according to (6.29).
- 4: Calculate mean $\boldsymbol{\mu}_{g,t}^{\mathbf{p}}$ using (6.27) and covariance matrix $\mathbf{C}_{g,t}^{\mathbf{p}}$ using (6.32).
- 5: Calculate the gain matrix $\tilde{\mathbf{G}}_t$ according to (6.38).
- 6: Set $\mathbf{y}_t = \boldsymbol{\mu}_{g,t}^{\mathbf{p}}$ at locations where $\mathbf{y}_t$ is missing. Calculate mean $\boldsymbol{\mu}_{g,t}^{\mathbf{f}}$ by means of (6.36) and covariance matrix $\mathbf{C}_{g,t}^{\mathbf{f}}$ by means of (6.37).
- 7: **end for**
- 8: Perform imputation at time x_* by means of mean $\mathbf{m}_g^*$ (6.39) and covariance matrix $\mathbf{C}_g^*$ (6.40).
- 9: **return** $\mathbf{m}_g^*$, $\mathbf{C}_g^*$.

in Algorithm 5. Fig.6.2 shows the proposed RGP-G prediction approach. We demonstrate this approach by illustrating an example. At time $t = 0$, we initialize the GP function as given in (6.18). At time $t = 1$, we receive measurements $\mathbf{y}_1$. The GP function is updated using these measurements $\mathbf{y}_1$ by means of (6.36) and (6.37). The updation of the mean and covariance of the GP function $\mathbf{f}$ are denoted by $\boldsymbol{\mu}_{g,t}^{\mathbf{f}}$ and $\mathbf{C}_{g,t}^{\mathbf{f}}$. If any measurements at time $t = 1$ are missing, they are predicted using (6.27) and (6.32). From time $t = 2$ onwards, we perform step-ahead prediction of the GP function using (6.41) and (6.42) until the subsequent measurements are observed. We perform prediction for all $x(t_*)$ satisfying $x(t_*) > x(t)$ using the knowledge of function $\mathbf{f}$ updated at the previous time instant. The predicted mean $\mathbf{m}_g^*$ and their associated variances $\mathbf{C}_g^*$ are given as,

$$\mathbf{m}_g^* = \mathbf{D}_g^{\top} \mathbf{A}_g^{-1} \boldsymbol{\mu}_{g,t}^{\mathbf{f}} \quad (6.41)$$

and

$$\mathbf{C}_g^* = \mathbf{B}_g + \mathbf{J}_{g,t*} (\mathbf{C}_{g,t}^{\mathbf{f}}) \mathbf{J}_{g,t*}^{\top} \quad (6.42)$$

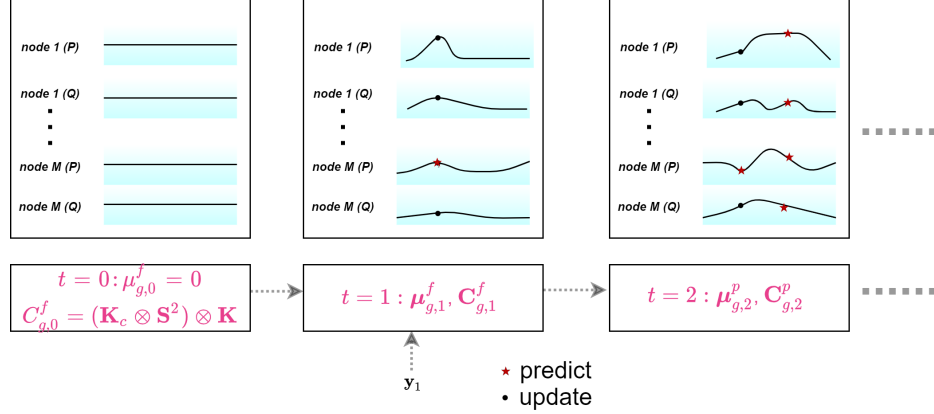


Figure 6.2: RGP-G prediction approach

Algorithm 5 RGP-G prediction Approach

Input: Basis vector $\mathbf{x}$, Distribution graph $\mathbf{L} \in \mathbb{R}^{M \times M}$, total time instants T , α , $\mathbf{y}_t$, $\mathbf{S}$, $\mathbf{K}$, $\mathbf{K}_c$

- 1: Set $\boldsymbol{\mu}_{g,0}^f = \mathbf{0}$ and $\mathbf{C}_{g,0}^f = (\mathbf{K}_c \otimes \mathbf{S}^2) \otimes \mathbf{K}$.
 - 2: Calculate the gain matrix $\mathbf{J}_{g,t}$ according to (6.29).
 - 3: Calculate mean $\boldsymbol{\mu}_{g,t}^p$ using (6.27) and covariance matrix $\mathbf{C}_{g,t}^p$ using (6.32).
 - 4: Calculate the gain matrix $\hat{\mathbf{G}}$ according to (6.38).
 - 5: Set $\mathbf{y}_t = \boldsymbol{\mu}_{g,t}^p$ at locations where $\mathbf{y}_t$ is missing. Calculate mean $\boldsymbol{\mu}_{g,t}^f$ by means of (6.36) and covariance matrix $\mathbf{C}_t^f$ by means of (6.37).
 - 6: Predict for all x_* which satisfy $x(t) < x(t_*)$ with mean $\mathbf{m}_g^*$ and variance $\mathbf{C}_g^*$ using (6.41) and (6.42) respectively.
 - 7: **return** $\mathbf{m}_g^*$, $\mathbf{C}_g^*$.
-

6.2.2 RGP Approach

When the network topology is unknown, the graph structure information by means of the graph filter matrix is not possible. In this case, we aim to use the recursive multi-task GP without graphs approach (RGP). Here, we initialize the prior function $\mathbf{f} = \mathbf{f}(\mathbf{X})$ at time $t = 0$ as,

$$p_0(\mathbf{f}) = \mathcal{N}(\mathbf{f}; \boldsymbol{\mu}_0^f, \mathbf{C}_0^f) \quad (6.43)$$

with mean $\boldsymbol{\mu}_0^f = \mathbf{0}$ and covariance $\mathbf{C}_0^f$ defined as,

$$\mathbf{C}_0^f = (\mathbf{K}_c \otimes \mathbf{I}) \otimes \mathbf{K} + \sigma_\epsilon^2 \mathbf{I} \quad (6.44)$$

We define the matrices $\mathbf{A}$, $\mathbf{D}$ and $\mathbf{F}$ as,

$$\mathbf{A} = (\mathbf{K}_c \otimes \mathbf{I}) \otimes \mathbf{K} + \sigma_\epsilon^2 \mathbf{I}, \quad (6.45)$$

$$\mathbf{D} = (\mathbf{K}_c \otimes \mathbf{I}) \otimes \mathbf{K}_t + \sigma_\epsilon^2 \mathbf{I}, \quad (6.46)$$

$$\mathbf{F} = (\mathbf{K}_c \otimes \mathbf{I}) \otimes \mathcal{K}_{tt} + \sigma_\epsilon^2 \mathbf{I} \quad (6.47)$$

Instead of the matrices (6.22), (6.23) and (6.24) defined for RGP-G approach, we use the matrices defined in (6.45), (6.46) and (6.47). The sequential inference and update step remains the same for both approaches. In this approach, we can perform both the interpolation and prediction similar to the RGP-G approach. In case of RGP Interpolation approach, the predicted mean and covariance is denoted by $\mathbf{m}^*$ and $\mathbf{C}^*$ given as,

$$\mathbf{m}^* = \mathbf{D}^\top \mathbf{A}^{-1} \boldsymbol{\mu}_T^f \quad (6.48)$$

$$\mathbf{C}^* = \mathbf{B} + \mathbf{J}_{t^*} (\mathbf{C}_T^f) \mathbf{J}_{t^*}^\top \quad (6.49)$$

The RGP approach used for interpolation is summarized in Algorithm 6.

Algorithm 6 Multi-task RGP Approach

Input: Basis vector $\mathbf{x}$, total time instants T , α , sequential measurements $\mathbf{y}_t \in \mathbb{R}^{d^M}$, $\mathbf{K}$, $\mathbf{K}_c$

- 1: Initialization: $\boldsymbol{\mu}_0^f = \mathbf{0}$ and $\mathbf{C}_0^f = (\mathbf{K}_c \otimes \mathbf{I}) \otimes \mathbf{K}$.
 - 2: **for** $t = 1, \dots, T$ **do**
 - 3: Calculate the gain matrix $\mathbf{J}_t$ according to (6.29) using the matrix $\mathbf{A}$ and $\mathbf{D}$ from (6.45) and (6.46) respectively .
 - 4: Calculate mean $\boldsymbol{\mu}_t^p$ using (6.27) and covariance matrix $\mathbf{C}_t^p$ using (6.32).
 - 5: Calculate the gain matrix $\tilde{\mathbf{G}}_t$ according to (6.38).
 - 6: Set $\mathbf{y}_t = \boldsymbol{\mu}_t^p$ at locations where $\mathbf{y}_t$ is missing. Calculate mean $\boldsymbol{\mu}_t^f$ by means of (6.36) and covariance matrix $\mathbf{C}_t^f$ by means of (6.37).
 - 7: **end for**
 - 8: Perform imputation at time x_* by means of mean $\mathbf{m}^*$ (6.48) and covariance matrix $\mathbf{C}^*$ (6.49) .
 - 9: **return** $\mathbf{m}^*$, $\mathbf{C}^*$.
-

6.3 Computational Complexity

The computational complexity associated with Algorithms 4, 5, and 6 for n basis vectors and dM number of observations at step t is $\mathcal{O}(dMn^2)$. This complexity is driven by the gain matrix calculation in (6.29). Use of recursive GP significantly reduces the computational complexity as compared to the full GP approach.

We next show the use of graph information in RGP-G approach reduces the uncertainty of the posterior distribution when the measurements are recursively processed at time $t = 1, \dots, T$.

Theorem 1. *The variance of the estimator of $\mathbf{f}$ using the RGP-G (Algorithm 4 and 5) of the distribution $p(\mathbf{f})$ is less than the variance of the estimator of $\mathbf{f}$ using RGP at time $t = 0$ i.e.,*

$$\text{tr}(\mathbf{C}_0^{\mathbf{f}}) > \text{tr}(\mathbf{C}_{g,0}^{\mathbf{f}})$$

where $\mathbf{C}_0^{\mathbf{f}}$ is defined in (6.44) and $\mathbf{C}_{g,0}^{\mathbf{f}}$ is defined in (6.19).

Proof. In order to prove this theorem, we need to show that the trace of $\Delta\mathbf{C}_0 = \mathbf{C}_0^{\mathbf{f}} - \mathbf{C}_{g,0}^{\mathbf{f}}$ is nonnegative. The initial covariance matrix $\mathbf{C}_{g,0}^{\mathbf{f}}$ at time $t = 0$ is defined with a graph filter $\mathbf{S}$ given in (6.3). The Laplacian matrix $\mathbf{L}$ used in this graph filter has an eigen-decomposition corresponding to,

$$\mathbf{L} = \mathbf{V}\mathbf{U}_{\mathbf{G}}\mathbf{V}^{\top} \quad (6.50)$$

where, $\mathbf{U}_{\mathbf{G}} = \text{diag}(U(1), U(2), \dots, U(M))$ and $\mathbf{V}$ denote the diagonal eigenvalue matrix and the associated eigenvectors respectively. Every eigenvalue of the laplacian matrix $\mathbf{L}$ is non-negative [118]. We need to prove,

$$\text{tr}(\mathbf{C}_0^{\mathbf{f}} - \mathbf{C}_{g,0}^{\mathbf{f}}) \geq 0 \quad (6.51)$$

$$\text{tr}((\mathbf{K}_c \otimes \mathbf{I}) \otimes \mathbf{K} - (\mathbf{K}_c \otimes \mathbf{S}^2) \otimes \mathbf{K}) \geq 0,$$

The trace of the Kronecker product of three matrices is the product of the traces of the

matrices. Hence, we get,

$$\text{tr}(\mathbf{K}_c)\text{tr}(\mathbf{I})\text{tr}(\mathbf{K}) - (\text{tr}(\mathbf{K}_c)\text{tr}(\mathbf{S}^2)\text{tr}(\mathbf{K})) \geq 0,$$

The kernel matrix $\mathbf{K}$ and $\mathbf{K}_c$ is positive semidefinite by construction. Hence, we need to prove,

$$\text{tr}(\mathbf{I} - \mathbf{S}^2) > 0,$$

$$\text{tr}(\mathbf{I} - \mathbf{V}(\mathbf{I} + \alpha\mathbf{U}_G)^{-2}\mathbf{V}^\top) \geq 0,$$

$$\sum_{i=1}^M \left(1 - \frac{1}{(1 + \alpha U(i))^2} \right) \geq 0.$$

Let $U(s)$ denote the smallest non-zero eigenvalue of $\mathbf{L}$, we then need to prove

$$(1 + \alpha U(s))^{-2} \leq 1,$$

$$(1 + \alpha U(s)) \geq 1$$

$$\alpha U(s) \geq 0.$$

For $\alpha > 0$, the smallest non-zero eigenvalue of $\mathbf{L}$ i.e., $U(s) > 0$. Hence, $\text{tr}(\mathbf{C}_0^{\mathbf{f}} - \mathbf{C}_{g,0}^{\mathbf{f}}) > 0$.

■

Theorem 2. The posterior covariance matrix of the estimator of function $\mathbf{f}$ for RGP-G approach at time t i.e., $\mathbf{C}_g^*$ in (6.40) is smaller than the posterior covariance matrix for the RGP based estimator of $\mathbf{f}$ given as $\mathbf{C}^*$ in (6.49) evaluated without graph information $\forall t = 1, \dots, T$.

Proof: We need to prove, $\text{tr}(\mathbf{C}_g^*) < \text{tr}(\mathbf{C}^*)$ for all $t = 1, \dots, T$, where

$$\mathbf{C}^* = \mathbf{B} + \mathbf{J}_{t*}(\mathbf{C}_T^{\mathbf{f}})\mathbf{J}_{t*}^\top$$

and $\mathbf{C}_g^*$ is defined in (6.40). From Theorem 1, we have proved that $\Delta\mathbf{C}_0 = \text{tr}(\mathbf{C}_0^{\mathbf{f}} - \mathbf{C}_{g,0}^{\mathbf{f}}) > 0$.

The matrix $\mathbf{B}$ for RGP approach can be defined as,

$$\mathbf{B} = \mathbf{F} - \mathbf{D}^\top \mathbf{A}^{-1} \mathbf{D}$$

Similarly, we have $\mathbf{B}_g$ for RGP-G approach as defined in (6.28). Thus, we have,

$$\begin{aligned} \text{tr}(\mathbf{B} - \mathbf{B}_g) = \text{tr}((\mathbf{F} - \mathbf{F}_g) - (\mathbf{D} - \mathbf{D}_g)^\top (\mathbf{C}_0^{\mathbf{f}} - \mathbf{C}_{g,0}^{\mathbf{f}})^{-1} \\ (\mathbf{D} - \mathbf{D}_g)) \end{aligned} \quad (6.52)$$

As $\text{tr}(\mathbf{C}_0^{\mathbf{f}} - \mathbf{C}_{g,0}^{\mathbf{f}}) > 0$ and the Schur complement of a positive definite matrix is also positive definite. Therefore, we have

$$\text{tr}(\mathbf{B} - \mathbf{B}_g) > 0$$

Similarly,

$$\begin{aligned} \text{tr}(\mathbf{C}_t^{\mathbf{p}} - \mathbf{C}_{g,t}^{\mathbf{p}}) = \text{tr}((\mathbf{B} - \mathbf{B}_g) + (\mathbf{J}_t - \mathbf{J}_{g,t})(\mathbf{C}_{g,t-1}^{\mathbf{f}} - \mathbf{C}_{t-1}^{\mathbf{f}}) \\ \times (\mathbf{J}_t - \mathbf{J}_{g,t})^\top) \end{aligned} \quad (6.53)$$

At $t = 1$ we have $\mathbf{C}_0^{\mathbf{f}} - \mathbf{C}_{g,0}^{\mathbf{f}} > 0$ (proved in theorem 1) and $\text{tr}(\mathbf{B} - \mathbf{B}_g) > 0$. Hence,

$$\text{tr}(\mathbf{C}_1^{\mathbf{p}} - \mathbf{C}_{g,1}^{\mathbf{p}}) > 0$$

Similarly, $\text{tr}(\mathbf{C}_t^{\mathbf{f}} - \mathbf{C}_{g,t}^{\mathbf{f}}) > 0 \ \forall t = 1, \dots, T$. The posterior distribution of the imputations

$$\text{tr}(\mathbf{C}^* - \mathbf{C}_g^*) > 0. \blacksquare$$

6.4 Simulation results

The efficacy of the proposed approach is verified on the three-phase unbalanced IEEE 37 [110], IEEE 123 [95], and 11,000 bus test systems. An aggregated 24-hr load profile at the

primary nodes consists of a mixture of load profiles, i.e., industrial and commercial load profiles obtained from [111], and residential loads obtained from [119]. Reactive power profiles are obtained by assuming a power factor of 0.9 lagging. Other profiles at different nodes were obtained by adding a random noise term and a sinusoidal wave of random amplitude spanning the 24-hr period. By utilizing this data, the voltage profile at all nodes is obtained by running load flow. The aggregated smart meter data are averaged over 15-min intervals while the voltage magnitude measurements are sampled at a 1-min interval. Thus, we have considered two sensor types for the case study. We have assumed RBF kernel for all the GP-based approaches. The imputation is performed for the aggregated smart meter data at a 1-min interval. We compare the performance of the proposed RGP-G Interpolation against the linear interpolation approach [23], RGP (Algorithm 6), and full GP (Algorithm 3) approach. The RGP-G prediction approach is compared with [120]. Algorithms 4, 5, 6 are initialized using their respective mean and co-variance functions associated with the GP function at time $t = 0$. Here, the hyper-parameters associated with the GP function can be obtained by either training the proposed approaches using historical data or using cross-validation techniques. The hyper-parameters involved in the proposed approach are $\theta = [l, \sigma_s^2, \sigma_\epsilon^2]$, where $l, \sigma_s^2, \sigma_\epsilon^2$, are defined in (6.7). We have used the grid search method guided by a five-fold cross-validation technique to obtain the hyper-parameters for our problem. In the cross-fold validation technique, one fold of the measurement set is retained as a validation set and the other folds as a training set. Each time a different set is chosen as the validation set, and this procedure is repeated five times. We select a finite set of reasonable hyper-parameter values to perform a grid search. The performance of each combination is evaluated through cross-validation on the training set. This approach evaluates the MAPE for each possible combination of hyperparameter values and chooses the set that minimizes the error on the validation set. More details on the grid-search-based cross-fold validation technique for Gaussian process hyperparameter tuning can be found in [103]. Another approach is to consider the historical data for hyperparameter tuning. The historical measurements of multi time-scale measurements can be used to obtain the hyper-parameters by maximizing the log marginal likelihood of the historical time-series data. The log-likelihood

can be computed in closed form as given in [103]. It is important to note that the proposed approach does not require any extra training set for imputation. The parameter α for the RGP-G approach is set to 0.05.

6.4.1 IEEE 37 bus test system

There are three cases by which we illustrate the performance of the multi-task RGP-G approach.

1. Case 1: In this case, we consider the measurement noise as mean 0 with standard deviation equal to 1% of the actual values. Fig. 6.3 shows the performance of the RGP-G interpolation approach at 0% missing measurement case for an IEEE 37 bus test system. Here, the time-series is the active power injection at node 11 of phase A. The RGP-G interpolation approach recursively updates the GP function in the 24-hr batch and later performs imputation at 1-min interval. The 95% confidence interval indicates the uncertainty bounds associated with the imputed measurements. The ideal case, i.e., 0% missing measurements dataset, has no missing measurements, but the dataset consists of a subset of the total measurements, representing an unobservable condition. For instance, let us assume that the aggregated AMI measurements in this dataset are available at 15-min intervals. Therefore, if we consider a 24-hr duration to perform imputations every minute, we have only 96 measurements per AMI sensor out of the total 1440-time instances.
2. Case 2: In this case, we perform the reconciliation and state estimation using the noisy time-series measurements corrupted by 0 mean and 10% standard deviation for a total of four hours duration. In this case, we fix the number of sensors and their locations corresponding to a particular FAD. The meters are placed randomly in the network. Fig.6.4 shows the meter deployment for IEEE 37 bus test system for 50% FAD. Thus, there are no time-series measurements at the nodes where the sensors are absent. We then perform imputation using these incomplete measurements. Fig.6.5 shows the comparison of the RGP-G approach with linear interpolation. It can be seen that the

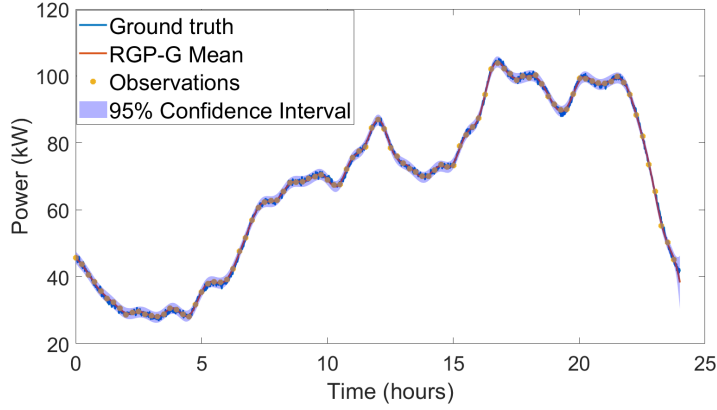


Figure 6.3: *RGP-G Interpolation approach of active power injection timeseries at node 11, Phase A*

Table 6.1: *Case 2: MAPE of active and reactive power imputed time-series data (IEEE 37 bus test system)*

Scenario	Proposed RGP-G I	RGP I	Full GP	Linear I
0% missing	1.4% (P)	1.8% (P)	8.5% (P)	3.18% (P)
	2.3% (Q)	2.6% (Q)	5.3% (Q)	6.5% (Q)
10% missing	3.23% (P)	4.2% (P)	8.1% (P)	3.37% (P)
	3.4% (Q)	5.35% (Q)	8.7% (Q)	7.72% (Q)
20% missing	3.02% (P)	4.2% (P)	7.8% (P)	3.6% (P)
	3.07% (Q)	9.89% (Q)	8.9% (Q)	8.1% (Q)

former approach provides smoother imputation than the latter one. We compare the performance of all five approaches using the mean absolute percentage error (MAPE) metric. Table 6.1 tabulate their performances for IEEE 37 bus test systems. It can be inferred that the performance of the RGP-G interpolation is superior to the other four approaches. In the linear interpolation approach, each time-series data is imputed individually without exploiting any spatio-temporal property of the data. In contrast, the full GP approach (Algorithm 3) exploits temporal relationships for imputation. The GP function update is performed using all the measurements in batch mode at once, which is computationally expensive.

The proposed recursive GP approaches assumes that the measurement data is corrupted by Gaussian noise as seen from (6.15). Also, the optimization formulation in

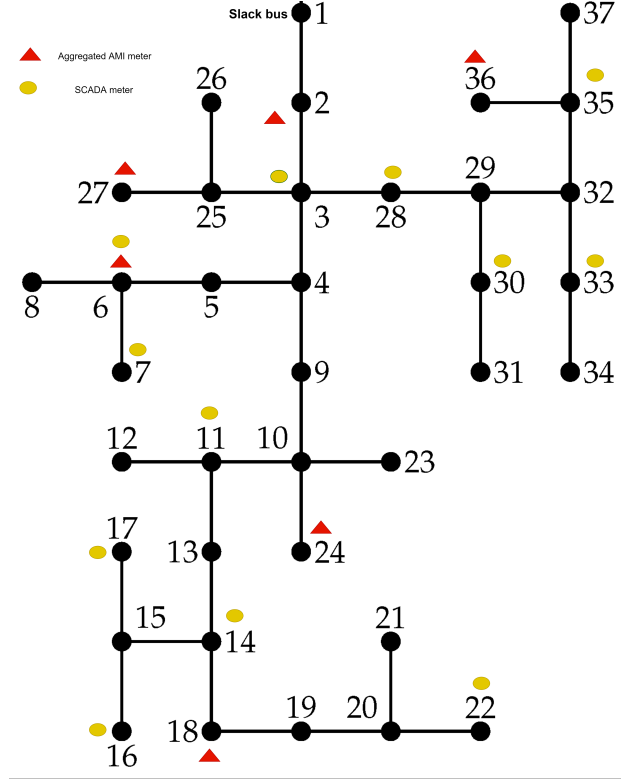


Figure 6.4: Meter deployment corresponding to 50% FAD for IEEE 37 bus test system

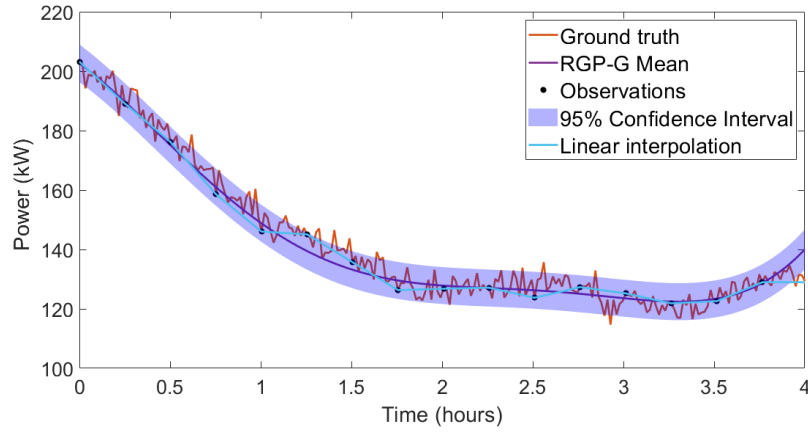


Figure 6.5: Comparison of RGP-G interpolation and linear interpolation approach of an active power injection time-series at Node 2 of Phase A

(6.2) assumes Gaussian noise. These are common assumptions used in many prior efforts [115, 116, 103]. The proposed approach in its current form can be applied to non-Gaussian noise but will not be optimal. Table 6.2 shows the performance of the RGP-G approach with Laplacian noise with 0 mean with standard deviation equal to

5%, 10% of the actual power values. As seen from Table 6.2, it can be inferred that the performance of the recursive GP approach under non-Gaussian noise scenarios is not optimal. Alternately, under a non-Gaussian measurement noise scenario, a warped GP approach [121] can be used. In this approach, the observations are transformed into a latent space such that the transformed data has Gaussian noise and will be better modeled by the GP. Developing a multi-task recursive GP framework using the warped GP will be pursued as part of our future work.

Table 6.2: *MAPE of proposed RGP-G approach for Laplacian and Gaussian noise with standard deviation set as percentage of actual power values*

Standard deviation of noise	5%	10%
Laplacian noise	4.14%	12.47%
Gaussian noise	2.6%	3.07%

The consistent time-series measurements are further used to estimate the states using the matrix completion-based DSSE approach discussed in section 3.19. While performing the matrix completion for a particular FAD, the corresponding entries in the measurements matrix are zero, with no sensor measurements. Table 6.3 shows the absolute errors and relative error reductions for RGP-G and linear interpolation methods. It can be deduced that RGP-G based technique significantly reduces error at all FADs. For example, the error in estimating reactive power using GP-based imputed time-series is reduced by 46% at 90% FAD compared to the linearly interpolated time-series. The reduction in voltage state estimation error is more modest mainly due to the robustness of matrix completion based DSSE that includes the topology information.

3. Case 3:

In this case, we compare the RGP-G prediction aided matrix completion with [120]. The method in [120] uses a data collation method to reconcile heterogeneous measurements and a Kalman filter method to perform DSSE. The data collation consists of an exponential moving average method to extrapolate the slow-rate measurements. Table 6.4 shows the prediction errors for different percentages of missing temporal data. Our proposed approach has several advantages over [120]. Firstly, the DSSE in [120]

Table 6.3: Absolute errors and relative error reductions (%) compared to the actual measurements (LI : Linear Interpolation approach, %: percentage reductions)

Scenario	FAD = 50%			FAD = 70%			FAD = 90%		
Measurements	LI	RGP-G	%	LI	RGP-G	%	LI	RGP-G	%
Active power (kW)	8.3	7.9	5.1%	8.35	7.5	11.33%	2.8	2.4	16.67%
Reactive power (kVAR)	3.87	3.46	11.8%	3.8	3.1	22.5%	1.57	1.07	46.7%
Voltage magnitude (p.u)	0.85	0.84	1.19%	0.85	0.83	2.41%	0.22	0.18	22.22%

Table 6.4: Case 3: MAPE of imputed time-series data of active and reactive power (IEEE 37 bus test system)

Scenario	Proposed RGP-G Prediction	Data Collation method [120]
0% missing	2.26% (P) 2.87% (Q)	6.44% (P) 6.75% (Q)
10% missing	3.5% (P) 5.65% (Q)	7.75% (P) 7.16% (Q)
20% missing	4.8% (P) 7.3% (Q)	8.13% (P) 8.86% (Q)

requires measurement redundancy (ratio of the number of measurements to the total states) higher than two. Thus, unlike our proposed approach, the method in [120] fails in low-observable conditions. Secondly, it can be observed that the imputation error in [120] is higher than the proposed RGP-G prediction approach as seen from Table 6.4.

6.4.2 IEEE 123 bus test system

In this case, we perform the reconciliation and state estimation using the noisy time-series measurements corrupted by 0 mean and 10% standard deviation for a total of four hours duration. In this case, we fix the number of sensors and their locations corresponding to a

Table 6.5: *Case 2: MAPE of active and reactive power imputed time-series data (IEEE 123 bus test system)*

Scenario	Proposed RGP-G I	RGP I	Full GP	Linear I
0% missing	4.11% (P) 1.7% (Q)	4.38% (P) 1.72% (Q)	8.9% (P) 5.68% (Q)	6.67% (P) 2.5% (Q)
10% missing	4.45% (P) 1.8% (Q)	5.1% (P) 2.5% (Q)	15.6% (P) 2.7% (Q)	6.75% (P) 10.3% (Q)
20% missing	4.6% (P) 2.18% (Q)	5.41% (P) 2.9% (Q)	15.8% (P) 10.48% (Q)	17.87% (P) 12.28% (Q)

particular FAD. Thus, there are no time-series measurements at the nodes where the sensors are absent. We then perform imputation using these incomplete measurements. Table 6.5 illustrates the performance of the proposed RGP-G and RGP Interpolation approach against the linear interpolation and full GP approach. Similar to the results as seen for IEEE 37 bus test systems on page 92, it can be inferred that the Recursive GP with graphs (RGPG-I) provides accurate imputation than the other counterparts.

6.4.3 11,000 bus test system

As discussed in section 6.3, the computational complexity associated with Algorithms 4, 5, 6 for M node system with d different sensor data streams and n time instants will be of the order of $\mathcal{O}(dMn^2)$. Hence, while the approach can be used for $M = 8500$, the complexity grows as M increases. To address the scalability issue, a distributed implementation of the proposed multi-task recursive GP approach is possible. We perform the distributed implementation on the 11,000-node feeder proposed in [122]. The 11,000-node test the feeder is constructed by connecting an IEEE 8,500-node test feeder and an EPRI Ckt7 test feeder at the substation. To perform the distributed recursive GP approach, we partition the 11,000-node network into four areas. We assume that smart meter measurements are available at 30% of nodes in each area. The load profiles assigned to each node consist of industrial, residential, and commercial load profiles and are scaled according to their base loads provided in [122]. Reactive power profiles are obtained by assuming a power factor

randomly varying between 0.9 and 0.95 lagging. The smart meter measurements are averaged over 15-minute intervals with measurement noise as mean 0 and standard deviation equal to 1% of the actual power values. We perform the multi-task RGP Interpolation approach in each area for 4 hours. The MAPE for each area is tabulated in Table 6.6.

Table 6.6: *MAPE of active and reactive power imputed time-series data (11,000 node feeder)*

Area	MAPE (%)
Area 1	1.66%
Area 2	1.5%
Area 3	1.59%
Area 4	2.26%

6.5 Summary

This chapter proposes a recursive Gaussian process with graphs for effectively aggregating heterogeneous intermittent time-series data and using it to estimate the distribution system states in low observability conditions. The proposed approach leverages the graphical structure of the network for accurately imputing the multi time-scale measurements. It has the flexibility to perform imputations in batch mode or real-time mode. Superior imputation performance of the active and reactive power time-series measurements are obtained with the proposed approach. Further, state estimation in IEEE 37 and the IEEE 123 bus system reveals that the power and voltage states are recovered with high fidelity.

Chapter 7

Handling multi time-scale data using neural network based approaches

The Gaussian process based approaches for integrating multi time-scale measurements discussed in Chapter 5 and 6 is computationally expensive as it requires an inversion operation of the kernel matrix to impute or predict the slow-rate measurements. Therefore, in this chapter, we propose a novel neural-network based approach that leverages spatio-temporal dependencies in time-series data without involving matrix inverse operations. The proposed approach uses neural ordinary differential equations (ODE) [33] that are ideal for imputing and predicting time-series measurements collected at non-uniform intervals.

7.1 Background: Neural ODE

The measurements obtained from multiple grid sensors are obtained at different sampling rates. The multivariate time-series sensor data with D variables and of N length can be written as,

$$X_t = x_1, x_2, \dots, x_T^D \in \mathbb{R}^{N \times D} \quad (7.1)$$

This data may contain missing values due to the sensor sampling rate or communication impairments. A mask $M \in \mathbb{R}^{N \times D}$ identifies the missing measurements. The entries in M

i.e., m_t^d is set to 1 if the corresponding measurement x_t^d is observed; else, they are set to 0. The goal is to reconcile the unevenly sampled measurements at the finest time resolution. Learning a generative model for the multivariate time-series will help accomplish this goal.

Generative models using a deep neural network is typically built on the concept of a fixed number of *layers*. In the forward pass, each network consists of a stack of L transformations, where L is the depth of the model. In order to update these models, a backpropagation algorithm is run through the same L layers via chain rule. This process necessitates that we store the intermediate values of the layers. Thus, training standard deep neural networks are computationally challenging as the memory requirement for storing the intermediate quantity increases as the model depth is increased. Furthermore, limited number of transformations are performed due to the fixed number of layers.

Neural ODE is a recent novel framework that is effective for modeling irregularly-sampled time series commonly encountered in various real-world application, including smart grid and medical data. It combines deep neural network principles with ordinary differential equations, and thus are more effective than conventional time series models. Particularly, in this work, Neural ODE is used for learning generative models for multivariate time series data from the distribution grid. Neural ODE offers a continuous time transformation of variables from input state to final predictions unlike standard deep neural network which only performs a limited number of transformations depending on the number of layers. The transformed values (or intermediate values) are obtained via ODE solvers by providing initial state and dynamics as inputs. The dynamics of the transformation function is determined by a neural network as shown below:

$$\frac{dz_t}{dt} = f(z_t, \theta) \quad (7.2)$$

where, f is a neural network parameterized by θ that defines the ODE dynamics. z_t is the hidden state of the Neural ODE. Thus, starting from an initial point $z(t_0)$, the transformed

state at any time t_i is given by integrating an ODE forward in time, given as,

$$z_i = z_0 + \int_{t_0}^{t_i} \frac{dz_t}{dt} dt \quad (7.3)$$

$$z_i = \text{ODESolve}(f, z_0, t_0, t_i, \theta)$$

(7.3) can be solved numerically using any ODE Solver (e.g., Euler’s method). In order to train the parameters of the ODE function f , an adjoint sensitivity approach is proposed in [33]. This approach computes the derivatives of the loss function with respect to the model parameters θ by solving a second augmented ODE backwards in time. Some of the advantages of using Neural ODE solvers over other conventional approaches are: (1) *Memory efficiency*: The adjoint sensitivity approach allows us to train the model with constant memory cost independent of the layers in the ODE function f ; (2) *Adaptive computation*: In deep neural networks, the number of layers are fixed and therefore, they generally have a fixed amount of function evaluations. However, the number of layers in a neural ODE is the number of steps an adaptive ODE solver decides to take. This means that the neural ODE can effectively adapt the number of layers on the fly for different datasets and take adaptive steps wherever necessary to determine the solution with desired accuracy; (3) *Effective formulation*: In addition to above two advantages, the continuously defined dynamics can naturally incorporate data which arrives at arbitrary times. Therefore, we propose to use the Neural ODE for reconciling the unevenly sampled distribution grid data.

Based on the foundations of NeuODEs, ODE-RNN was developed. ODE-RNN is autoregressive in nature unlike the generative form of NeuODES [34]. It models the irregularly sampled time series by applying ODE and RNN interchangeably through a time series. The states evolve continuously using the ODE model, while they are updated using the RNN cell at the instances where the measurements are available. The function f at time t is described in a neural ODE with initial hidden state h_t :

$$\dot{\mathbf{h}} = f(\mathbf{h}_t, \theta) \quad (7.4)$$

$$o_t = c(\mathbf{h}_t, \theta_c) \quad (7.5)$$

where, $\mathbf{h} \in \mathbb{R}^m$ is the hidden state of the data, $\dot{\mathbf{h}} = \frac{d\mathbf{h}}{dt}$ is the derivative of the hidden state, $o \in \mathbb{R}^d$ is the output of the ODE-RNN. $f : \mathbb{R}^m \rightarrow \mathbb{R}^m$ and $c : \mathbb{R}^m \rightarrow \mathbb{R}^d$ are the neural ODE operator and output function parameterized by neural network parameters θ and θ_c , respectively. The hidden states are modeled using a Neural ODE, where they obey the solution of an ODE as,

$$\mathbf{h}'_i = ODEsolve(\mathbf{h}_{i-1}, (t_{i-1}, t_i), f) \quad (7.6)$$

Lets represent the RNN cell function by $v(\cdot)$ with parameters θ_v . At each observation x_i , the hidden states are updated by an RNN as,

$$\mathbf{h}_i = v(\mathbf{h}'_i, x_i, \theta_v). \quad (7.7)$$

We now discuss the approaches based on the principles of neural differential equations.

7.2 Proposed Generative model - Latent Neural ODE

The basic neural ODEs evaluate the hidden state values at any desired time instants. However, such models are hard to interpret, especially for multi-time scale power measurements, due to the combined dynamics of power system and the ODE solver. Therefore, we propose to use Latent ODE (LODE) approach for reconciling the multi-time scale power measurements. The LODE approach is a continuous-time generative process for integrating multi-time scale measurements. This approach uses Neural ODEs, and variational autoencoder within a single framework [34]. LODE has two key advantages over neural ODE: First, it explicitly decouples the dynamics of the power system, the likelihood of observations, and the recognition model so that each component can be analyzed separately. Secondly, the posterior distribution over an initial latent state provides a measure of uncertainty which further increases the reliability of our predictions. The proposed framework (LODE) has

three different modules, namely an encoder, a decoder, and the ODE solver. The architecture of LODE for smart distribution grid is illustrated in Fig. 7.1. Each module of LODE

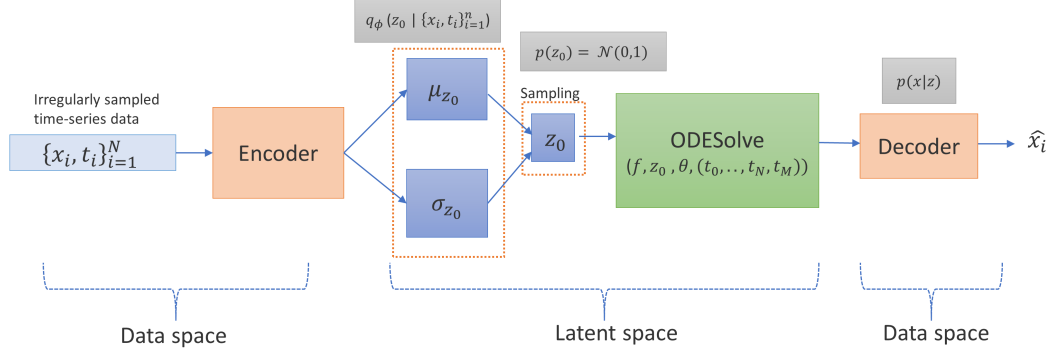


Figure 7.1: *Architecture of proposed Latent ODE approach for smart distribution grid*

is described in the forthcoming subsection.

7.2.1 Encoder: Recognition network

This module encodes the input measurements from data space and transforms it into a latent space. Encoding is typically carried out via a recurrent neural network since it is effective in capturing long term dependencies of time series data. Encoder takes $\{x_i, t_i\}_{i=1}^N$ as an input, where x_i represents the observations and t_i represents the corresponding observation times. The data is processed backward in time from time t_N to t_0 . An approximate posterior over the initial state $q_\phi(z_0 | \{x_i, t_i\}_{i=1}^N)$ is computed from the last hidden layer of the encoder network. In our approach, the mean and standard deviation of the approximate posterior $q_\phi(z_0 | \{x_i, t_i\}_{i=1}^N)$ are the function of the final hidden state of an encoder network, characterized by,

$$q_\phi(z_0 | \{x_i, t_i\}_{i=1}^N) = \mathcal{N}(\mu_{z_0}, \sigma_{z_0}) \quad (7.8)$$

where, $\mu_{z_0}, \sigma_{z_0} = g(RNN_\phi(\{x_i, t_i\}_{i=1}^N))$. Here, the function g represents a neural network layer, translating the final hidden state of the encoder into the mean and variance of z_0 .

7.2.2 ODE Solver

Once an approximate posterior distribution $q(z_0|\{x_i, t_i\}_{i=1}^N)$ is obtained from the encoder, an initial latent state (z_0) for the ODE solver is sampled from the corresponding distribution. The initial latent state serves as an input to the ODE solver together with a ODE dynamic function f . Then, a ODE solver is used to obtain latent space observations for all the given times,

$$z_1, z_2, \dots, z_N = \text{ODESolve}(f, z_0, \theta, \{t_0, t_1, \dots, t_N\}) \quad (7.9)$$

Thus, given observation times $t_0, t_1, \dots, t_N$ and an initial state z_0 , an ODE solver produces $z_1, \dots, z_N$, which describes the latent state at each observation. The ODE solver is capable of imputing historical time points as well as forecasting future values by providing appropriate latent states at desired time instants. As the function f is time-invariant, a unique latent trajectory can be defined given the initial latent state.

7.2.3 Decoder

the decoder transforms the latent trajectory defined at various time instants back into the data space using the neural network. Standard multi layer perceptron can be used in decoder network since we only need to map two function spaces.

7.2.4 Training of LODE

The training of the proposed LODE framework is similar to that of a variational autoencoder, and it is end-to-end. The encoder-decoder model is trained by maximizing the Evidence Lower Bound (ELBO) given as,

$$\begin{aligned} ELBO(\phi, \theta) = & \mathbb{E}_{q_\phi(z_0|\{x_i, t_i\}_{i=0}^N)} [\log(p_\theta(x_0, \dots, x_N))] \\ & - KL(q_\phi(z_0|\{x_i, t_i\}_{i=0}^N) || p(z_0)) \end{aligned} \quad (7.10)$$

The first term in the ELBO represents the log probability of the decoder estimates. The second term denotes the KL divergence or degree of “dissimilarity” between the two distributions $q_\phi(z_0|\{x_i, t_i\}_{i=0}^N)$ and $p(z_0)$. Here, the prior over latent states $p(z_0)$ is chosen as $\mathcal{N}(0, 1)$. The overall pipeline of the proposed Latent ODE approach summarized in Algorithm 7.

Algorithm 7 Latent ODE Approach

Input: Datapoints $\{x_i\}_{i=1}^N$ and the corresponding times $\{t_i\}_{i=1}^N$

- 1: $z_0 = RNN(\{x_i\}_{i=1}^N)$
 - 2: $\mu_{z_0}, \sigma_{z_0} = g(z_0)$
 - 3: $z_0 = N(\mu_{z_0}, \sigma_{z_0}) = q_\phi(z_0|\{x_i, t_i\}_{i=0}^N)$
 - 4: $z_1, z_2, \dots, z_N = ODEsolve(f, \theta, z_0, (t_0, \dots, t_N))$
 - 5: $\hat{x}_i = OutputNN\{z_i\}$
 - 6: **return** $\hat{x}_i$
-

7.3 Proposed Autoregressive model - SDE-RNN

The ODE-RNN approach proposed in [34] and discussed in 7.1 provides imputations but fails to quantify the uncertainty associated with it. Therefore, we offer a novel SDE-RNN framework, which, in addition to the predictions, is capable of quantifying the uncertainties in a principled manner. In this section, we describe the key elements of the proposed SDE-RNN approach. Our approach involves the modification of ODE-RNN to include a stochastic differential equation (SDE-RNN) to capture the evolution of the hidden states (as opposed to ODE). Stochastic differential equation that governs the dynamics of the hidden states corresponds to,

$$d\mathbf{h}_t = f(\mathbf{h}_t, t)dt + g(\mathbf{h}_t, t)d\mathbf{B}_t \quad (7.11)$$

where, f and g are the drift and diffusion functions, respectively. The transformations f and g are carried through neural networks. $\mathbf{B}_t$ represents the Brownian motion with the distribution $d\mathbf{B}_t \sim \mathcal{N}(0, \mathbf{Q}\Delta t)$ where $\mathbf{Q}$ is the diffusion matrix. $g(\mathbf{h}_t, t)$ denotes the variance of the Brownian motion and represents the epistemic uncertainty in the hidden state dynamics. Both the drift and diffusion functions are nonlinear, thanks to the nonlinear transformation of neural networks f and g . Therefore, the SDEs in (7.11) is a nonlinear

SDE. The hidden states in the SDE-RNN model are the solution of the SDE in (7.11), i.e.,

$$\mathbf{h}'_i = \text{SDEsolve}(\mathbf{h}_{i-1}, (t_{i-1}, t_i), f, g) \quad (7.12)$$

Here, SDEsolve represents the SDE Solver. It is important to note that states are updated based on observations x_i at time t_i using the RNN model. Using this SDE-RNN approach, we can quantify the uncertainty in the hidden and output states, as discussed in the following subsection.

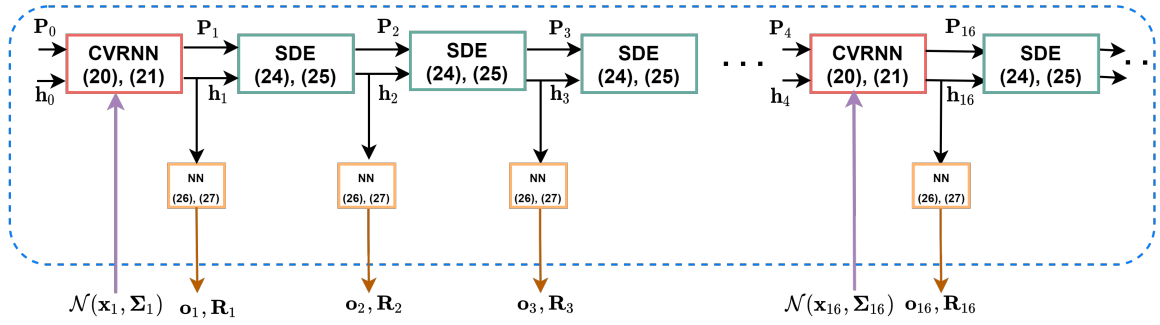


Figure 7.2: *Propagation of uncertainty. x_i : Input at irregular intervals*

7.3.1 Uncertainty quantification

We propose a novel uncertainty propagation approach in the SDE-RNN model. The proposed method quantifies the uncertainty in a holistic manner accounting for both aleatoric and epistemic, without relying on prior specification of model parameters and complicated Bayesian inference. The SDE-RNN approach propagates the uncertainty from the noisy input observations to the final outputs using the CVRNN model and the SDE model. Here, we refer to the CVRNN model as a modified form of RNN, that propagates the mean and uncertainty arising from the previous hidden states and the inputs to the next hidden state. Fig. 7.2 and Algorithm 8 illustrate the overall uncertainty propagation in the SDE-RNN framework. We can initialize the mean and covariance matrix of the hidden states as zero and then update them continuously. The CVRNN model updates the hidden states at the

time when an observation is available along with its associated uncertainty. This approach accounts for the uncertainty in both the input observation and the previous hidden states. SDE model captures the epistemic uncertainty of the SDE-RNN approach. In the following subsections, we elaborate the uncertainty propagation in both the CVRNN model and SDE model.

1. CVRNN Model

The RNN cell in Eq. (7.7) updates the hidden states at time instants where measurements are available. The dynamics of a general RNN cell are modeled by,

$$\mathbf{h}_i = v(\mathbf{h}_{i-1}, x_i, \theta_v), \quad (7.13)$$

$$o_i = c(\mathbf{h}_i, \theta_c). \quad (7.14)$$

Assume the input x_i is corrupted with additive Gaussian noise given as

$$\tilde{x}_i = x_i + w_i, \quad (7.15)$$

where $w_i = \mathcal{N}(0, \Sigma_i)$ with Σ_i representing the covariance matrix of the measurement. These noisy observations are fed to the RNN model, where the hidden states and the outputs become a random variable corresponding to,

$$\tilde{\mathbf{h}}_i = v(\tilde{\mathbf{h}}_{i-1}, \tilde{x}_i, \theta_v), \quad (7.16)$$

$$\tilde{o}_i = c(\tilde{\mathbf{h}}_i, \theta_c). \quad (7.17)$$

Let $\hat{\mathbf{h}}_i = \mathbb{E}[\tilde{\mathbf{h}}_i]$ and $\hat{P}_i$ denotes the estimate of the covariance matrix $P_i = \mathbb{E}\{(\tilde{\mathbf{h}}_i - \hat{\mathbf{h}}_i)(\tilde{\mathbf{h}}_i - \hat{\mathbf{h}}_i)^\top\}$. Based on (7.16), the linearization around $\hat{\mathbf{h}}_{i-1}$ implies that,

$$\begin{aligned}
\tilde{\mathbf{h}}_i &= v(\tilde{\mathbf{h}}_{i-1}, \tilde{x}_i) \\
&= v(\hat{\mathbf{h}}_{i-1} + \delta h_{i-1}, \hat{x}_i + w_i) \\
&= v(\hat{\mathbf{h}}_{i-1}, \hat{x}_i) + \Delta_h v \delta \mathbf{h}_{i-1} + \Delta_x v w_i + o(\delta \mathbf{h}_{i-1}^2, w_i^2)
\end{aligned} \tag{7.18}$$

where, $\Delta_h v$ and $\Delta_x v$ are calculated at operating points $(\hat{\mathbf{h}}_{i-1}, \hat{x}_i)$. According to the transformation of uncertainty [123] given in Theorem 1, we can evaluate the expected value and covariance of $\tilde{\mathbf{h}}_i$.

Theorem 3. *Transformation of uncertainty: Consider a recurrent neural network represented by $v(\cdot)$ with neural network parameters θ_v and hidden states given by $\tilde{\mathbf{h}}_i = v(\tilde{\mathbf{h}}_{i-1}, \tilde{x}_i, \theta_v)$. The input $\tilde{x}_i$ is corrupted with noise $w_i = \mathcal{N}(0, \Sigma_i)$. The estimation of the expected value and covariance matrix of the hidden state $\tilde{\mathbf{h}}_i$ at time i can be recursively calculated as,*

$$\hat{\mathbf{h}}_i = v(\hat{\mathbf{h}}_{i-1}, x_i) \tag{7.19}$$

$$\hat{\mathbf{P}}_i = (\Delta_h v) \hat{\mathbf{P}}_{i-1} (\Delta_h v)^\top + (\Delta_x v) \Sigma_i (\Delta_x v)^\top \tag{7.20}$$

where, the expected value of the previous hidden state $\mathbb{E}[\tilde{\mathbf{h}}_{i-1}] = \hat{\mathbf{h}}_{i-1}$ and its associated covariance matrix is $\hat{\mathbf{P}}_{i-1}$ are given.

Thus, the mean and covariance can be computed in a CVRNN framework with noisy input measurement data at time i using (7.19) and (7.20). We discuss the SDE model in the following subsection.

2. SDE Model

A neural SDE model aims to capture the epistemic uncertainty with Brownian motion and propagate it to the next time step. It also aims to capture the uncertainties arising from the previously obtained hidden states and their co-variances. A neural SDE can be expressed using the dynamical equation in Eq. (7.11). As this SDE is nonlinear,

its statistics can be computed by adopting certain approximations. Linearized approximation of SDEs [124] is a technique which computes the statistics by linearizing the drift and diffusion function around a certain point. Using the Taylor series, the drift function $f(\mathbf{h}, t)$ around the mean $\mathbf{m}$ is linearized as,

$$f(\mathbf{h}, t) \approx f(\mathbf{m}, t) + \mathbf{F}_h(\mathbf{h}, t)(\mathbf{h} - \mathbf{m}) \quad (7.21)$$

The diffusion function is linearized as,

$$g(\mathbf{h}, t) \approx g(\mathbf{m}, t) + \mathbf{G}_h(\mathbf{h}, t)(\mathbf{h} - \mathbf{m}) \quad (7.22)$$

A linearized approximation to estimate the mean and covariance of an SDE (7.11) can be obtained by integrating the following differential equations from the initial conditions $\mathbf{m}(t_0) = E[\mathbf{h}(t_0)]$ and $\mathbf{P}(t_0) = \text{Cov}[\mathbf{h}(t_0)]$ to the target time t :

$$\frac{d\mathbf{m}}{dt} = f(\mathbf{m}, t), \quad (7.23)$$

$$\frac{d\mathbf{P}}{dt} = \mathbf{P}\mathbf{F}_h^\top(\mathbf{m}, t) + \mathbf{F}_h(\mathbf{m}, t)\mathbf{P} + \mathbf{G}_h(\mathbf{m}, t)\mathbf{Q}\mathbf{G}_h^\top(\mathbf{m}, t) \quad (7.24)$$

Thus, (7.23) and (7.24) captures the evolution of both the mean and covariance of the hidden states. After the hidden states are updated, they are transformed by another neural network to get final predictions (see Fig.7.2). The estimation of the mean and the covariance of the output is obtained according to the transformation of uncertainty (Theorem 1) as,

$$\hat{o}_i = c(\hat{\mathbf{h}}_i) \quad (7.25)$$

$$\hat{R}_i = (\Delta_h c) \hat{\mathbf{P}}_i (\Delta_h c)^\top \quad (7.26)$$

Algorithm 8 illustrates the working mechanism of the proposed SDE-RNN approach. As discussed earlier, the hidden states $\mathbf{h}_i$ are updated at times i where measurements

are present. The availability of the measurements is indicated by the vector mask $\in \{0, 1\}$. The imputed measurements along-with the associated variances are indicated by $\{o_i\}_{i=0}^N$ and $\{R_i\}_{i=0}^N$ respectively.

Algorithm 8 SDE-RNN Approach

Input: Datapoints $\{x_i\}_{i=1}^N$ and the corresponding times $\{t_i\}_{i=1}^N$, drift function f , diffusion function g

Initialization: $\mathbf{h}_0, \mathbf{P}_0$

```

1: for  $i = 1, \dots, N$  do
2:    $\mathbf{h}'_i, \mathbf{P}'_i = SDESolve(\mathbf{h}_{i-1}, \mathbf{P}_{i-1}, (t_{i-1}, t_i), f, g)$ 
3:    $\mathbf{h}_i, \mathbf{P}_i = CVRNN(\mathbf{h}'_i, \mathbf{P}'_i, x_i)$ 
4:    $\mathbf{h}_i = \text{mask} \times \mathbf{h}_i + (1-\text{mask}) \times \mathbf{h}'_i$ 
5:    $\mathbf{P}_i = \text{mask} \times \mathbf{P}_i + (1-\text{mask}) \times \mathbf{P}'_i$ 
6:    $\{o_i, R_i\}_{i=1}^N = c(\mathbf{h}_i, \mathbf{P}_i)$ 
7: end for
8: return  $\{o_i\}_{i=0}^N, \{R_i\}_{i=0}^N$ 

```

Uncertainty propagation for GRU

The CVRNN model discussed in the previous section can be any form of the recurrent neural network; for example, it could be either RNN, Long Short Term Memory (LSTM), or Gated Recurrent Unit (GRU) model whose hidden states are given as,

$$h_t = v(h_{t-1}, x_t, \theta_v), \quad (7.27)$$

In this section, we will derive the uncertainty propagation for a GRU model by using the transformation of uncertainty as discussed in Theorem 1. To begin with, we need to compute the $\Delta_h v$ and $\Delta_x v$. In a GRU model, the hidden states are given as,

$$h_t = z_t \circ h_{t-1} + (1 - z_t) \circ h'_t \quad (7.28)$$

$$h'_t = \tanh(W_{in}x_t + b_{in} + r_t(W_{hn}h_{t-1} + b_{hn})) \quad (7.29)$$

$$z_t = \sigma(W_{iz}x_t + b_{iz} + W_{hz}h_{t-1} + b_{hz}) \quad (7.30)$$

$$r_t = \sigma(W_{ir}x_t + b_{ir} + W_{hr}h_{t-1} + b_{hr}) \quad (7.31)$$

Here, x_t , z_t and r_t are the inputs, update and reset gates of the GRU model, respectively. $\tanh$ and σ are the Tanh and Sigmoid activation functions, respectively. $\circ$ represents the Hadamard product.

The gradient of the hidden states (7.28) at time t computed at operating points $(\hat{h}_{t-1}, x_t)$ are given as,

$$\Delta_h v = \frac{\partial z_t}{\partial h_{t-1}} h_{t-1} + z_t \frac{\partial h_{t-1}}{\partial h_{t-1}} + \frac{\partial(1-z_t)}{\partial h_{t-1}} h'_t + \frac{\partial h'_t}{\partial h_{t-1}} (1-z_t) \quad (7.32)$$

The derivative of the sigmoid function is denoted as $\sigma'(x) = \sigma(x)(1-\sigma(x))$. Each gradient in the (7.32) is given as,

$$\frac{\partial z_t}{\partial h_{t-1}} = \sigma'(W_{iz}x_t + b_{iz} + W_{hz}h_{t-1} + b_{hz}) \circ W_{hz} \quad (7.33)$$

$$\frac{\partial r_t}{\partial h_{t-1}} = \text{diag}(r_t(1-r_t)) \circ W_{hr} \quad (7.34)$$

$$\frac{\partial h'_t}{\partial h_{t-1}} = \text{diag}(1-h_t'^2) \circ \left\{ \frac{\partial r_t}{\partial h_{t-1}} (W_{hn}h_{t-1} + b_{hn}) + W_{hn}r_t \right\} \quad (7.35)$$

Similarly, $\Delta_x v$ is given as,

$$\Delta_x v = \frac{\partial z_t}{\partial x_t} h_{t-1} + \frac{\partial(1-z_t)}{\partial x_t} h'_t + \frac{\partial h'_t}{\partial x_t} (1-z_t) \quad (7.36)$$

Each gradient in (7.36) is given as,

$$\frac{\partial z_t}{\partial x_t} = \text{diag}(z_t(1-z_t)) \circ W_{iz} \quad (7.37)$$

$$\frac{\partial r_t}{\partial x_t} = \text{diag}(r_t(1-r_t)) \circ W_{ir} \quad (7.38)$$

$$\frac{\partial h'_t}{\partial x_t} = \text{diag}((1-h_t'^2) \circ (W_{in} + \frac{\partial r_t}{\partial x_t} (W_{hn}h_{t-1} + b_{hn}))) \quad (7.39)$$

Using Eq. (7.32) and (7.36), the covariance matrix of the hidden states P_t at time t is given by (7.20).

7.4 Simulation results

This section evaluates the efficacy of the proposed framework for the imputation and predictions tasks related to smart distribution grid. Experiments are carried out in standard IEEE 37 bus system.

7.4.1 Data processing

We consider measurements from smart meter and SCADA sensors. The smart meter measurements consist of 24-hr active and reactive power injection time-series data aggregated at the primary nodes. This 24-hr load profile consists of a mixture of load profiles, i.e., industrial/commercial load profiles [111], and residential loads [119]. Reactive power profiles are obtained by assuming a power factor of 0.9 lagging. The SCADA measurements are obtained by executing load flows on the test network. The SCADA measurements consists of the voltage magnitude measurements at a subset of node locations. The aggregated smart meter data are averaged over 15-min intervals while the SCADA measurements are sampled at a 1-min interval. A Gaussian noise with 0 mean and standard deviation equal to 10% of the actual power values is added in the smart meter data to mimic real-world patterns. The smart meter and SCADA measurements constitute our training dataset.

Once the distribution grid's measurements are obtained, we represent the dataset as a list of records. Each record represents the information about the time-series data with the format given as, $record = [measurement\ type, values, times, mask]$. Here, time-series data at each node of the IEEE 37 bus network represents one *record*. The *measurement type* denotes the sensor type, i.e., P, Q , or V . $Values \in \mathbb{R}^{N \times 1}$ represents the sensor measurements with $times \in \mathbb{R}^N$ as the corresponding time instants. $Mask \in \mathbb{R}^{N \times 1}$ represents the availability of the corresponding measurements. The dataset is further normalized between $[0,1]$ interval.

We take the union of all time points across different nodes in the dataset that are irregularly sampled. This is needed to perform batching during training.

We now discuss the model specifications and simulation results for LODE and SDE-RNN approach.

7.4.2 LODE approach

Model Specifications

The encoder is a gated recurrent unit (GRU) [125]. We consider 40-dimensional hidden states of encoder with tanh activation functions. The ODE function is a feedforward neural network with three layers and 100 units on each layer. The ODE solver is a fifth-order ‘dopri5’ solver. The decoder consists of a feedforward neural network with a single layer. Here, we consider an adaptive learning rate with an initial value of the learning rate set to 0.01. We consider batch size as 10, and report loss as mean squared error (MSE) and negative ELBO. The model is trained using stochastic gradient descent through Adam optimizer for 200 iterations. All the experiments are conducted on the system with Intel i9 core processor, 32 GB RAM, 8 GB GPU. Python is explicitly used for coding the entire framework with the support of PyTorch’s TorchdiffEq [33].

The results of our experiments using LODE approach are discussed in the following subsections.

Imputation

In this task, the smart meter measurements are interpolated at a 1-min interval. The training is performed using the observed 15-minute interval data. In order to perform interpolation, the encoder runs backward in time to compute the approximate posterior distribution at the initial time t_0 . Fig. 7.3 demonstrates the imputation performance using the proposed LODE approach. Table. 7.1 shows the comparison of LODE approach with linear interpolation approach [23] and recursive GP with graphs (RGP-G) approach [126]. As seen from Table. 7.1, the proposed approach is accurate with 0.2% MSE error in the test data. The trajectory

Table 7.1: *MSE performance of the proposed LODE, linear interpolation and recursive GP approach*

Approach	MSE (%)
Latent ODE	0.2%
Linear Interpolation	2%
Recursive Gaussian process	0.7%

of MSE and negative ELBO on the test data are illustrated in Fig. 7.4 and Fig. 7.5, respectively. The convergence of the losses demonstrate the effective training of the model.

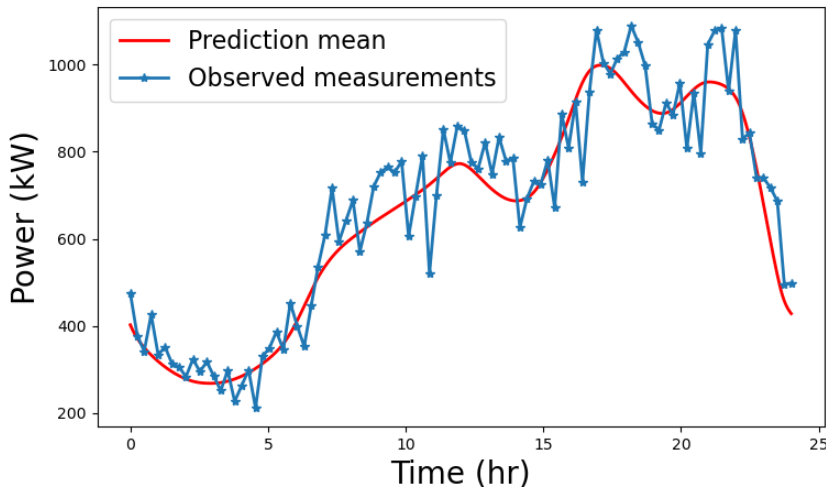


Figure 7.3: *Imputation at node 1 using Latent ODE approach*

Prediction

In this task, we split the time series into two halves, t_0 to $t_{N/2}$ and $t_{N/2}$ to t_N . The model is trained by conditioning the observations in the first 12 hrs of the time-series data and reconstructing the other half, i.e., training loss is considered on the second half. Once the model is trained, it can perform predictions for any desired time horizon. As seen from Fig. 7.6, the model only observes the first 12 hours of measurement data (blue in color) and extrapolates the next 12 hours (red in color). The predictions on the testing data are accurate with an MSE of 0.72%.

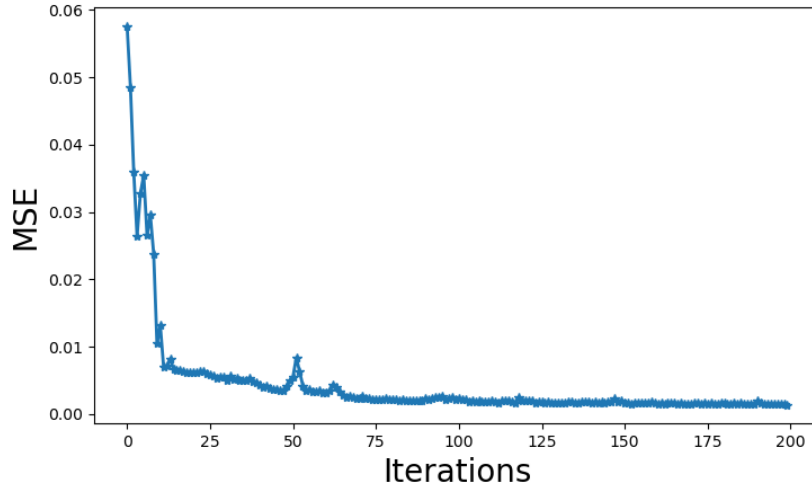


Figure 7.4: *MSE on the test dataset using Latent ODE approach*

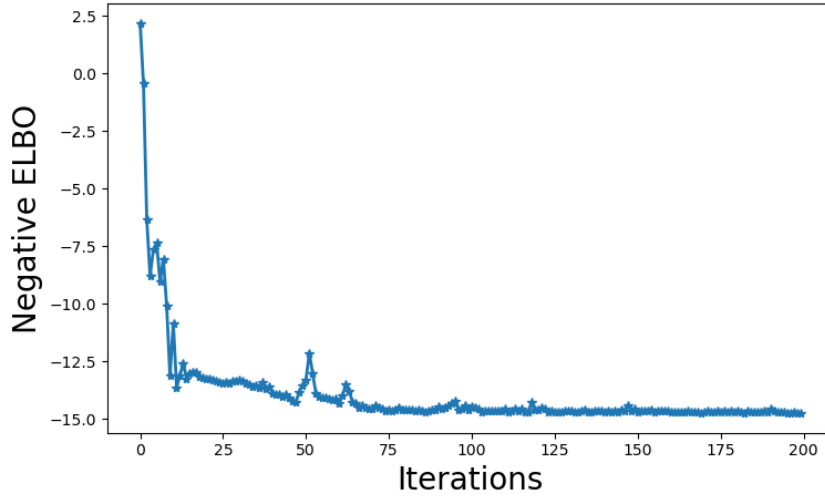


Figure 7.5: *Negative ELBO on the test dataset using Latent ODE approach*

7.4.3 SDE-RNN approach

Model Specifications

In the case of SDE-RNN approach, we use the GRU cell with hidden size 5 to encode the observations. The drift function of the SDE is a feedforward neural network with 1 layer and 100 units. The diffusion function is a feedforward neural network with 1 layer, 100 units, and a sigmoid activation function. We consider Ito SDE with diagonal noise of Brownian

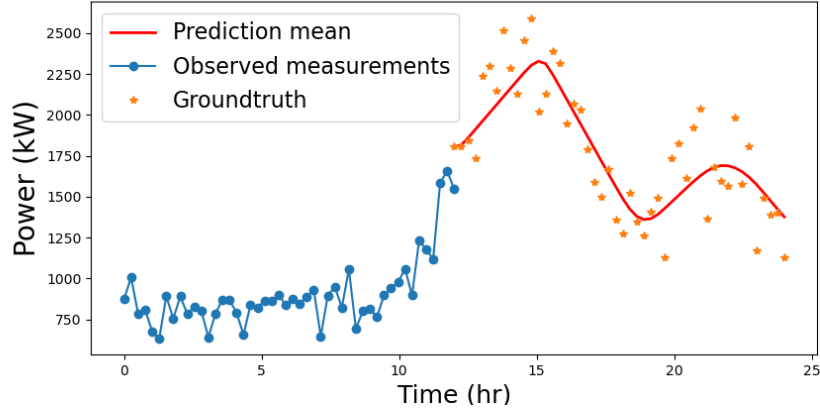


Figure 7.6: *Predictions at node 12 using Latent ODE approach*

motion. Python is used for coding the entire framework with the support of PyTorch’s torchsde package [35]. We consider the output neural network as a 1-layer feedforward network with an input size of 5 and an output size of 1. We set the learning rate to 0.01, batch size as 10 and report the loss as mean squared error (MSE). We compare the efficacy of the proposed SDE-RNN with the classic GRU approach. The classic GRU approach consists of a GRU model with 3 input features (observations, times, and mask) and 5 output features. The predictions are obtained by transforming the output of the GRU cell via another feedforward neural network with 2 layers. The first layer is a neural network with output features 100 followed by the Tanh activation function. We then employ a dropout rate of 0.3 after this activation function followed by another feedforward neural network of input features 100 and output feature 1.

The uncertainty estimates provided by the SDE-RNN is compared with that of Monte-Carlo dropout approach in classic RNN using the Expected normalized calibration error (ENCE) metric [127]. Calibration error is typically used to evaluate uncertainty estimates since they are not associated with any groundtruth values. In a well calibrated model, average confidence score should match with the average model accuracy [128]. Thus, in a classification setting, calibration implies that whenever a forecaster assigns a probability of p to an event, that event should occur about $p\%$ of the time. On the other hand, in a regression setting, calibration signifies that the prediction y_t should fall in a $c\%$ (e.g., 90%)

confidence interval approximately $c\%$ (e.g., 90%) of the time. The expected normalized calibration error (ENCE) is based on the similar regression setting that for each value of uncertainty, measured through the standard deviation σ , the expected mistake (measured in MSE) matches the predicted error σ^2 , i.e.,

$$E_{x,y}[(\mu(x) - y)^2 | \sigma(x)^2 = \sigma^2] = \sigma^2 \quad (7.40)$$

ENCE metric is evaluated using the binning approach, where we assume that the number of bins N divides by the number of time points T . We divide the indices of the examples to N bins, $\{B_j\}_{j=1}^N$, such that $B_j = \{(j-1)\frac{T}{N} + 1, \dots, j\frac{T}{N}\}$. Each bin represents an interval in the standard deviation axis: $[\min_{t \in B_j} \{\sigma_t\}, \max_{t \in B_j} \{\sigma_t\}]$. Expected Normalized Calibration Error (*ENCE*) is calculated as,

$$ENCE = \sqrt{\frac{1}{N} \sum_{j=1}^N \frac{|mVar(j) - RMSE(j)|}{mVar(j)}} \quad (7.41)$$

where, the root of the mean variance ($mVar(j)$) at bin j is,

$$mVar(j) = \sqrt{\frac{1}{|B_j|} \sum_{t \in B_j} \sigma_t^2}, \quad (7.42)$$

and root mean squared error ($RMSE(j)$) at bin j is,

$$RMSE(j) = \sqrt{\frac{1}{|B_j|} \sum_{t \in B_j} (y_t - \hat{y}_t)^2}. \quad (7.43)$$

It is expected that the $mVar$ will equal the $RMSE$ for each bin, i.e., the plot of $RMSE$ as a function of $mVar$ should be an identity function. However, in reality, the models are not well calibrated and the plot is not an identity function. Thus, a model is said to provide better uncertainty estimates if they have a lower ENCE. We compare the uncertainty estimation of our SDE-RNN approach with the classic RNN in the following subsection.

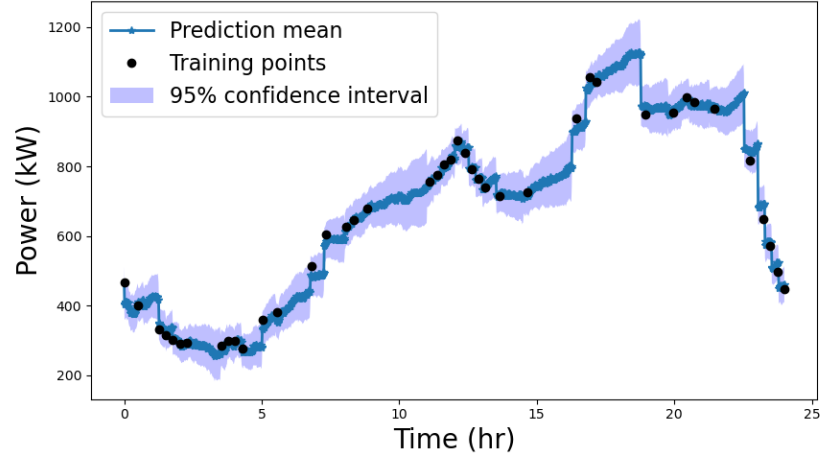


Figure 7.7: *Uncertainty quantification in SDE-RNN approach- node 1*

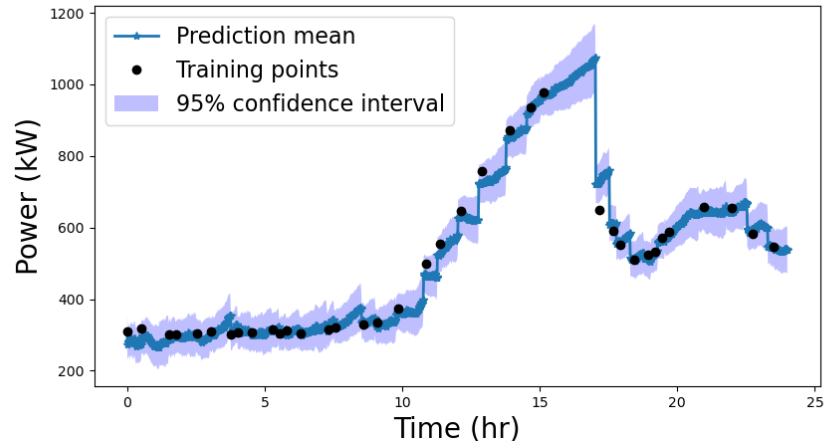


Figure 7.8: *Uncertainty quantification in SDE-RNN approach- node 12*

Table 7.2: *Imputation results for different percentages of missing data in test dataset*

Metric	MSE			ENCE		
Missing data (%)	40%	60%	80%	40%	60%	80%
SDE-RNN	0.0005	0.0008	0.0013	57.45	49.11	54.35
Classic-GRU	0.0506	0.0978	0.1527	534.05	419.06	457.92

Imputation:

In the first experiment, the smart meter measurements are imputed at a 1-min interval using the proposed SDE-RNN approach. The training and test dataset contains the available AMI (15-minute) and SCADA (1-minute interval) measurements. In addition, in the test dataset, we introduce missing data as a percentage of the total number of time instants (minutes) in the 24-hr interval (i.e., 1440 time instants). The prediction and uncertainty estimates provided by the SDE-RNN model are shown in Fig. 7.7 and Fig. 7.8 for nodes 1 and 12, respectively. Here, 5% of total interpolated points (i.e., 1440 data points) are observed. It can be observed that the epistemic uncertainty increases as we move away from the observed values. For instance in Fig. 7.7, the sparse region of 10th-12th hr depicts less confidence compared to data-rich regions of 5th-9th hr. The aleatoric uncertainty due to the sensor noise is well captured by the SDE-RNN model, reflected by the uncertainty estimates at the observations. The uncertainty estimates at the observation times is nonzero and present due to the noisy input data, which is the aleatoric uncertainty part.

In the second experiment, we compare the uncertainty estimates provided by the SDE-RNN with the classic GRU + MC dropouts approach. We assume the number of bins is 5. Table 7.2 shows the performance of both these approaches for different percentages of missing data in the test dataset. We provide the test dataset’s mean squared error (MSE) and ENCE. It is evident from the Table that the SDE-RNN approach offers better accuracy and uncertainty estimates than the classic GRU model for all the levels of missingness in the test dataset.

7.5 Summary

This chapter proposes two neural ODE-based approaches, namely an autoregressive SDE-RNN model and a Latent ODE approach for integrating heterogeneous measurements in a smart distribution grid. The proposed Latent ODE approach uses neural ODEs for learning generative models, which can provide accurate imputations and predictions at any desired time instant. The Neural SDE-RNN approach is capable of quantifying both the aleatoric and epistemic uncertainty in the measurement data. Specifically, we derive analytical expressions for quantifying and propagating the epistemic and aleatoric uncertainty across time instants.

Chapter 8

Phase and outage awareness in distribution network

State estimation approaches discussed in chapter 4 require sufficient measurements and system model information to estimate the system states. The system model includes the system topology, network parameters, and phase connectivity information. The network connectivity model is mostly accurate, but phasing errors are quite common [129]. Phase changes due to restoration, reconfiguration, and maintenance activities frequently happen in the distribution networks and such changes are not always tracked continuously. Utilities have limited or unreliable information to identify the connected loads' phases (A, B, or C) [53]. Real-time awareness of the operational topology is also necessary for performing state estimation, demand response management, and distributed resource monitoring. However, line outages change the network topology and lead to failures if not identified quickly. Line outages are caused due to fault or any topology attacks, which causes the protective devices to isolate the faulty areas quickly. Due to the operation of protective devices in response to faults, the loads downstream of the protective devices will also experience an outage. Thus, the network will be divided into different islands, some energized and others that might not be energized. Hence, outage detection is an important task to return stability to the distribution network. However, the presence of limited real-time measurements in the distribution

grid affects the real-time monitoring capabilities in distribution systems. The sheer size of low-voltage distribution grids has made the cost of meter placement prohibitive. Hence, one of the critical challenges is the optimal selection of the number of meters, type, and their locations in the grid, for outage identifiability. Furthermore, any outage detection strategy should detect outages in DN as quickly as possible.

Therefore, in this chapter, we develop frameworks to enhance the situational awareness of the low-observable distribution system. To this end, we first develop a novel phase identification approach that accurately identifies the phase labels of the multiphase unbalanced distribution system. In the second approach, we develop an outage detector and meter placement approach for detecting outages with high probability. Both these approaches are discussed in the following sections.

8.1 Phase Identification approach

A distribution system can be perceived as a graph with a tree structure consisting of substations, feeders, laterals, and customers (see Fig. 8.1). The paths from the substation to each consumer are unique. Typically, the feeders are three-phase, whereas the lateral connections are either single-phase, two-phase, or three-phase. Transformers connected to the end of these laterals step down the voltages for energizing the customers. The phase labels of these laterals are incorrectly documented or unknown to the utility. Therefore, all the customers connected to a lateral may have unknown phase information. In order to estimate the unknown phase labels, the proposed approach learns a new graph structure using the aggregated voltage magnitude measurements of the customers connected to each lateral. The estimated graph structure provides the phase connectivity information by solving an optimization problem that clusters the measurement data into k clusters, where each cluster represents a phase in the distribution network.

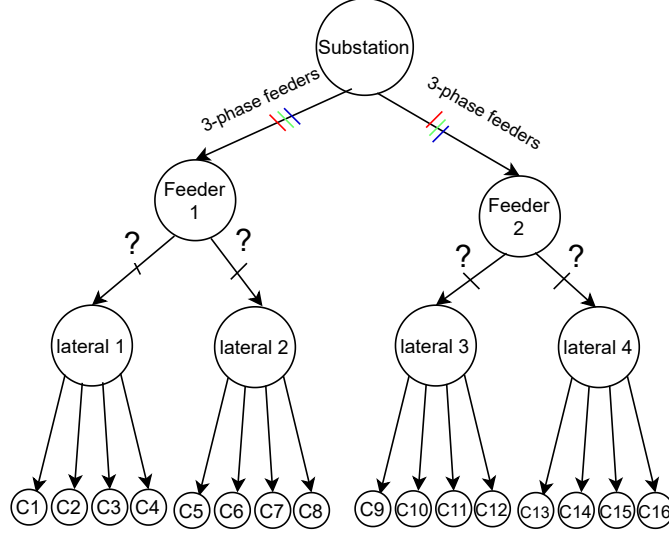


Figure 8.1: Tree representation of network topology with question mark ? representing the unknown phase labels

8.1.1 System model

A distribution system as shown in Fig.8.1 can be perceived as a graph where nodes are assigned to each of the bus phases. This graph is denoted by $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{A})$, where $\mathcal{V} \in \mathbb{R}^M$ denotes the vertices, $\mathcal{E}$ are the edges and $\mathbf{A} \in \mathbb{R}^{M \times M}$ is the adjacency matrix. Here, M represents set of phases at all buses in the graph $\mathcal{G}$. The adjacency matrix for an undirected graph is a symmetric matrix. The problem of phase identification can be mapped to the problem of discovering the adjacency matrix $\mathbf{A}$.

Consider a matrix of voltage magnitude measurements $\mathbf{V} \in \mathbb{R}^{M \times N}$, where N represents the total time over which the samples are collected. The entries in the rows of the matrix $\mathbf{V}$ are non-zero where sensors are placed in the grid. The matrix $\mathbf{V} = [\mathbf{V}_1^{p_1}, \dots, \mathbf{V}_K^{p_K}]^\top$ where $\mathbf{V}_i^{p_i}$ is the nodal voltages on bus i with phases p_i . For instance, a three-phase bus will be represented by $\mathbf{V}_i^{p_i} = [\mathbf{V}_i^a, \mathbf{V}_i^b, \mathbf{V}_i^c]$. Here, $\mathbf{V}_i^\phi$ is the line-to-ground voltage magnitude time-series on phase ϕ and $\top$ is the transpose operator. Here, the voltage measurements refer to the phase-to-neutral voltages [130].

The aim is to determine the ϕ labels with respect to the reference of substation phases. Let $\mathbf{s}_i$ and $\mathbf{t}_j$ specify the two terminals between bus i and bus j , $\mathbf{s}_i \in \mathbb{R}^{p_i}$ is the source

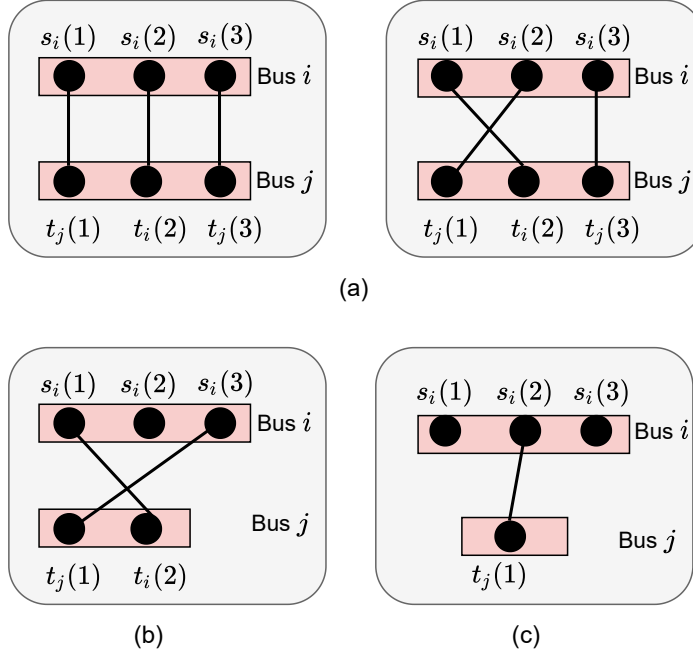


Figure 8.2: Different possible configurations for a two-bus connection. (a) three phase end (b) two-phase end (c) one-phase end

terminal and $\mathbf{t}_j \in \mathbb{R}^{p_j}$ is the target terminal. For instance in Fig.8.2 , $\mathbf{s}_i = [s_i(1), s_i(2), s_i(3)]$ and $\mathbf{t}_j = [t_j(1), t_j(2), t_j(3)]$. The vectors $\mathbf{s}$ and $\mathbf{t}$ are the concatenated vectors of $\mathbf{s}_i$ and $\mathbf{t}_j$, $\forall i, j \in M$, respectively. The dimension of $\mathbf{s}$ and $\mathbf{t}$ will correspond to the phases p_i and p_j on the buses i and j , respectively. There are different configurations in which terminals of bus i and bus j are connected depending on whether they are three-phase, two-phase, or single-phase connections. Two combinations for a three-phase end lateral are shown in Fig. 8.2(a).

8.1.2 Proposed phase identification approach

In this work, we exploit a key property that was validated in [131].

Theorem 4. *In a multi-phase distribution grid, if two terminal buses of a branch are connected on the same phase, their phase voltage correlation is the largest.*

Proof: See [131] for the complete proof.

Higher voltage correlation is equivalent to smooth signals over the graph structure [132]. This smoothness property can be quantified as,

$$l(\mathbf{v}) = \sum_{(i,j)} \mathbf{A}(i,j) \|\mathbf{v}_i - \mathbf{v}_j\|^2 = \text{tr}(\mathbf{V}^\top \mathbf{L} \mathbf{V}) \quad (8.1)$$

where $\mathbf{A}$ is the adjacency matrix and $\mathbf{A}(i,j)$ denotes the connection between the nodes i and j . The derivation of (8.1) can be found in [133]. The Laplacian matrix $\mathbf{L}$ is a positive semidefinite matrix defined as $\mathbf{L} = \mathbf{D} - \mathbf{A}$, where $\mathbf{D}$ is the diagonal degree matrix ($\mathbf{D} = \text{diag}(\mathbf{A}\mathbf{1})$). Here, $\mathbf{1} \in \mathbb{R}^M$ is a vector of all ones. The Laplacian quadratic form $\mathbf{V}^\top \mathbf{L} \mathbf{V}$ measures the smoothness or voltage signal variance over the graph. Two nodes i and j are said to be connected if they have small distance $\|\mathbf{v}_i - \mathbf{v}_j\|$ or the term $\text{tr}(\mathbf{V}^\top \mathbf{L} \mathbf{V})$ is small. The quantity $l(\mathbf{v})$ is small when $\mathbf{v}$ takes similar values across all connected nodes. Therefore, the main aim of the phase-identification approach is to find the adjacency matrix $\mathbf{A}$ that minimizes the term $l(\mathbf{v})$ along with additional constraints. The graph is thus obtained by solving the general optimization problem:

$$\min_{\mathbf{A}} \text{tr}(\mathbf{V}^\top \mathbf{L} \mathbf{V}) + \eta s(\mathbf{A}), \text{ subject to } \textit{constraints} \quad (8.2)$$

where $\eta > 0$ is a regularization parameter and $\mathbf{A}$ is symmetric matrix. The Laplacian $\mathbf{L} = (\text{diag}(\mathbf{A}\mathbf{1}) - \mathbf{A})$. Here, $s(\mathbf{A})$ is a regularizer. The distribution system graph is not a fully connected graph and is usually sparse. Hence, $s(\mathbf{A})$ term includes the sparsity term $\|\text{vec}(\mathbf{A})\|_1$. Furthermore, there are few phase label entries ($\mathbf{A}_{\text{known}} \in \mathbb{R}^{M \times M}$) which are known to the utility. Hence, we impose constraints on the entries corresponding to the known phase labels while estimating the adjacency matrix. Let $\Omega \subseteq \{1, \dots, M\} \times \{1, \dots, M\}$ describe the known entries in $\mathbf{A}_{\text{known}}$. The operator P_Ω is defined as,

$$[P_\Omega(\mathbf{A}_{\text{known}})]_{mn} = \begin{cases} [\mathbf{A}_{\text{known}}]_{mn}, & \text{if } (m,n) \in \Omega \\ 0, & \text{otherwise} \end{cases} \quad (8.3)$$

The challenge is to form k clusters using the estimated adjacency matrix, such that each cluster denotes a phase label. Thus, we need to estimate a graph $\mathcal{G}$ with $\mathbf{A} \in \mathbb{R}^{M \times M}$ which has k connected components (for e.g., $k = 3$ to form 3 clusters representing 3 different phases). The entries of $\mathbf{A}$ are binary variables. In order to formulate an objective based on this strategy, we start from the following theorem.

Theorem 5. *The multiplicity k of the zero eigenvalue of the Laplacian matrix $\mathbf{L}$ is equal to the number of connected components in the graph associated with $\mathbf{A}$.*

Proof: The proof of this classic result can be found in [134]. According to Theorem 5, if the $\text{rank}(\mathbf{L}) = M - k$, the graph is an ideal graph based on which the voltage time-series data points are clustered into k clusters. For instance, to form 3 clusters, the number of zero eigenvalues in the Laplacian matrix should be three or equivalently, $\text{rank}(\mathbf{L}) = M - 3$. Let $\sigma_i(\mathbf{L})$ denote the i -th smallest eigenvalue of $\mathbf{L}$. As $\mathbf{L}$ is positive semi-definite, $\sigma_i(\mathbf{L}) \geq 0$. Hence, minimizing the ' k ' smallest eigenvalue of $\mathbf{L}$ will form a graph with ' k ' clusters.

By adding the sparsity term $\|\text{vec}(\mathbf{A})\|_1$, the information of the known phase labels, and a term that minimizes the k smallest eigenvalues of the Laplacian matrix, the optimization problem in (8.2) can be modified as,

$$\begin{aligned} \min_{\mathbf{A}} \quad & \text{tr}(\mathbf{V}^\top \mathbf{L} \mathbf{V}) + \eta \|\text{vec}(\mathbf{A})\|_1 + \lambda_1 \|P_\Omega(\mathbf{A} - \mathbf{A}_{\text{known}})\|_F^2 \\ & + \beta \sum_{i=1}^k \sigma_i(\mathbf{L}) \\ \text{subject to } & \mathbf{A} \in \mathcal{A} \end{aligned} \quad (8.4)$$

where, $\mathcal{A}$ represents the set of valid adjacency matrices given as, $\mathbf{A}_{ij} = \mathbf{A}_{ji}$, $\mathbf{A}_{ij} \geq 0$.

However, (8.4) is non-convex due to the last term in the objective function. However, according to Ky Fan's Theorem [135], we can write it's equivalent convex form as,

$$\sum_{i=1}^k \sigma_i(\mathbf{L}) = \min_{\mathbf{S} \in \mathbb{R}^{M \times k}, \mathbf{S}^\top \mathbf{S} = \mathbf{I}} \text{tr}(\mathbf{S}^\top \mathbf{L} \mathbf{S}) \quad (8.5)$$

Therefore, the problem in (8.4) is equivalent to the following problem:

$$\begin{aligned}
& \min_{\mathbf{A}, \mathbf{S}} \text{tr}(\mathbf{V}^\top \mathbf{L} \mathbf{V}) + \eta \|\text{vec}(\mathbf{A})\|_1 + \beta \text{tr}(\mathbf{S}^\top \mathbf{L} \mathbf{S}) \\
& \quad + \lambda_1 \|P_\Omega(\mathbf{A} - \mathbf{A}_{\text{known}})\|_F^2 \\
& \text{subject to } \mathbf{S}^\top \mathbf{S} = \mathbf{I} \\
& \quad \mathbf{A} \in \mathcal{A}
\end{aligned} \tag{8.6}$$

In addition, we impose must-link constraints and cannot-link constraints in the optimization problem. The must-link constraints specify that there should be only one connection among the phases between any two connected buses. The topological information of the network is either assumed to be known or estimated using [136]. Mathematically, we can formulate the must-link constraints between two connected buses i and j as,

$$\begin{aligned}
0.5 & \leq \sum_{m=1}^{p_j} \mathbf{A}(\mathbf{s}_i(l), \mathbf{t}_j(m)) \leq 1, \forall i, \forall j, i \sim j \\
0.5 & \leq \sum_{l=1}^{p_i} \mathbf{A}(\mathbf{s}_i(l), \mathbf{t}_j(m)) \leq 1, \forall i, \forall j, i \sim j
\end{aligned} \tag{8.7}$$

Here, $i \sim j$ implies that bus i is a neighbor of bus j . Eq.(8.7) must be satisfied for at least one l or one m . The cannot-link constraints specify that the different phases in the same bus cannot be connected together i.e., the corresponding elements in the adjacency matrix should be set to zero. They can be formulated as,

$$\mathbf{A}(\mathbf{s}_i(l), \mathbf{s}_i(m)) \leq 0, \mathbf{A}(\mathbf{t}_i(l), \mathbf{t}_i(m)) \leq 0, \tag{8.8}$$

where, $m, l = 1, \dots, p_i, i = 1, \dots, K$.

The must-link constraints (8.7) and cannot-link constraints (8.8) are added in the optimization problem (8.6) to yield,

$$\begin{aligned}
& \min_{\mathbf{A}, \mathbf{S}} \text{tr}(\mathbf{V}^\top \mathbf{L} \mathbf{V}) + \eta \|\text{vec}(\mathbf{A})\|_1 + \beta \text{tr}(\mathbf{S}^\top \mathbf{L} \mathbf{S}) \\
& + \lambda_1 \|P_\Omega(\mathbf{A} - \mathbf{A}_{\text{known}})\|_F^2 \\
& \text{subject to (8.7), (8.8)} \\
& \mathbf{A} \in \mathcal{A} \\
& \mathbf{S}^\top \mathbf{S} = \mathbf{I}
\end{aligned} \tag{8.9}$$

The optimization formulation in (8.9) is a non-convex problem due to the variables $\mathbf{A}$ and $\mathbf{S}$ in the objective function. But using an alternating minimization approach where either $\mathbf{A}$ or $\mathbf{S}$ is fixed, convex forms can be obtained. Specifically, the alternating minimization algorithm updates the variables $\mathbf{A}$ and $\mathbf{S}$ at each iteration z in an alternating fashion while fixing the other variable. As the iterations increase, the optimization problem (8.10) and (8.11) using the alternating minimization algorithm converge to an optimal solution. The convergence of the alternating minimization algorithm to an optimal solution is proved in [137]. The number of iterations “iter” is chosen to be sufficiently large such that the alternating minimization algorithm (Algorithm 9) will converge. Alternatively, we can use criteria to check that the difference between the variables obtained at the previous and current iteration is less than a specified threshold. If the requirements are satisfied, the optimization problem is solved. The update rules are given by the following update equations as,

$$\begin{aligned}
\mathbf{A}^{(z)} = \min_{\mathbf{A}^{(z)}} & \text{tr}(\mathbf{V}^\top \mathbf{L} \mathbf{V}) + \eta \|\text{vec}(\mathbf{A})\|_1 + \beta \text{tr}(\mathbf{S}^{(z-1)\top} \mathbf{L} \mathbf{S}^{(z-1)}) \\
& + \lambda_1 \|P_\Omega(\mathbf{A} - \mathbf{A}_{\text{known}})\|_F^2 \\
& \text{subject to (8.7), (8.8)}
\end{aligned} \tag{8.10}$$

$$\mathbf{A} \in \mathcal{A}$$

$$\begin{aligned}
\mathbf{S}^{(z)} = \min_{\mathbf{S}^{(z)}} & \text{tr}(\mathbf{S}^\top \mathbf{L}^{(z)} \mathbf{S}) \\
& \text{subject to } \mathbf{S}^\top \mathbf{S} = \mathbf{I}
\end{aligned} \tag{8.11}$$

The estimated matrix $\mathbf{S}$ in (8.11) is formed by the k eigenvectors of $\mathbf{L}$ corresponding to the

k smallest eigenvalues. Algorithm 9 summarizes the proposed phase identification approach.

Algorithm 9 Alternating minimization Algorithm for phase identification in distribution system

Input: Voltage measurements $\mathbf{V}$, Set of known bus label indices Ω , $\mathbf{s}$, $\mathbf{t}$, number of iterations $iter$, $\mathbf{A}_{\text{known}}$

initialization: Initialize $\mathbf{S}^{(0)} = \text{eigenvector}(\text{diag}(\mathbf{A}_{\text{known}}\mathbf{1}) - \mathbf{A}_{\text{known}})$

- 1: **for** $z = 1, \dots, iter$ **do**
- 2: Solve (8.10)
- 3: Update $\mathbf{S}$, which is formed by the k eigenvectors of $\mathbf{L}^{(z)} = \mathbf{D}^{(z)} - \mathbf{A}^{(z)}$ corresponding to k smallest eigenvalues. $\mathbf{D}^{(z)} = \text{diag}(\mathbf{A}^{(z)}\mathbf{1})$.
- 4: **end for**
- 5: **return** $\mathbf{A}$

8.1.3 Joint Phase Identification and state estimation

In this section, we aim to estimate the system states and the phase labels in a single framework. Estimation of spatio-temporal states in an unobservable distribution system is achieved by tensor completion-based DSSE proposed in [138]. In this approach, a measurement tensor $\mathcal{T} \in \mathbb{R}^{M \times 5 \times N}$ is constructed by creating an array of state measurement matrices of the system. The matrices obtained at a single time instant t is $\mathbf{T} \in \mathbb{R}^{M \times 5}$ and structured as,

$$\mathbf{T} = [\mathbf{P}_t, \mathbf{Q}_t, \Re(\mathbf{v}_t), \Im(\mathbf{v}_t), |\mathbf{v}_t|], \quad (8.12)$$

where, $\mathbf{P}_t$ and $\mathbf{Q}_t$ represent the active power and reactive power injections at time t respectively. The term $\Re(\mathbf{v}_t)$ and $\Im(\mathbf{v}_t)$ represent the real and imaginary parts of voltage phasors at each phase of non-slack buses respectively. The measurement tensor $\mathcal{T}$ corresponding to system states and measurements are partially filled due to various issues like unobservability, sensor dropouts, and measurements sampled at different rates. The buses with no power injection measurements will have entries zero in the measurement tensor. Let the three unfoldings of the tensor $\mathcal{T}$ be denoted by $\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3$. Tensor completion fills the missing elements in the measurement tensor by utilizing the sparsity in data with linearized powerflow constraints[81] at every time instant.

The linearized powerflow constraints require knowledge of the network admittance matrix. The utility has the knowledge of the admittance matrix ($\mathbf{Y}_{LL}^s$) which is stored in the distribution management database. However, this system model information may be incorrectly documented if the phase labels are incorrect. The bus admittance matrix is updated by using the estimated adjacency matrix from the phase identification method (Algorithm 9). When the nominal topology of the system is known, the estimated admittance matrix ($\mathbf{Y}_{LL}^{(i,j)} \in \mathbb{R}^{3 \times 3}$) between two connected buses i and j is obtained by suitable permutation of the stored admittance matrix ($\mathbf{Y}_{LL}^{s(i,j)}$) as,

$$\mathbf{Y}_{LL}^{(i,j)} = \mathbf{Y}_{LL}^{s(i,j)} \pi \quad (8.13)$$

Here, the permutation matrix (π) is a matrix of 0 and 1. This permutation matrix obtained from Algorithm 9 reflects the phase label information between bus i and bus j . Using the updated admittance matrix information, the system model parameters $\mathbf{D}$ and $\mathbf{K}$ as defined in [81] are given as,

$$\mathbf{D} = \left(\mathbf{Y}_{LL}^{-1} \text{diag}(\bar{\mathbf{w}})^{-1}, -j \mathbf{Y}_{LL}^{-1} \text{diag}(\bar{\mathbf{w}})^{-1} \right), \quad (8.14)$$

$$\mathbf{K} = \left(|\text{diag}(\bar{\mathbf{w}})|^{-1} \Re(\text{diag}(\bar{\mathbf{w}}) \mathbf{D}) \right), \quad (8.15)$$

$$\text{and, } \mathbf{w} = -\mathbf{Y}_{LL}^{-1} \mathbf{Y}_{L0} v_0 \text{ is the zero-load voltage} \quad (8.16)$$

The tensor-completion optimization formulation [138] is given as,

$$\begin{aligned} \hat{\mathcal{X}} = \min_{\mathcal{X}} \sum_{i=1}^3 \frac{\beta_i}{2} \|\mathcal{M}_i - \mathcal{X}_{(i)}\|_F^2 + w_2 \gamma + w_3 \alpha \\ \text{s.t. } \|\mathbf{v}_t - \mathbf{D} \mathbf{x}_t - \mathbf{w}\|_{\infty} \leq \gamma \\ \||\mathbf{v}_t| - \mathbf{K} \mathbf{x}_t - |\mathbf{w}|\|_{\infty} \leq \alpha \\ t \in 1, 2, \dots, N \end{aligned} \quad (8.17)$$

where $\mathbf{v}_t$ is the bus voltage phasor vector at time t . $\mathbf{x}_t$ is the vector of active and reactive power injections of all buses at time t . The suffix (i) denotes the unfolding operation applied

on the tensor along the mode i . More details about the tensor completion based DSSE can be found in [101], [138].

The phase identification and state estimation are achieved jointly as shown in Algorithm 10. In this algorithm, the input measurements are limited spatio-temporal voltage and power measurements. At each iteration z , the following tasks are performed - First, the adjacency matrix representing the phase label connectivity is estimated. Secondly, the system model is updated using the knowledge of the estimated adjacency matrix. Thirdly, tensor completion-based DSSE using the updated model provides a complete knowledge of the system states, which further aids in performing phase label identification.

The proposed approach as discussed in Algorithm 10 assumes batch-wise measurements. However, in a real control room, the measurements arrive sequentially at different times. In addition, the different types of measurements (voltage and power) may come at different rates as they may be measured by various sensors (for example, AMI, PMU, or SCADA). Hence, to implement our approach in an actual control room, we propose two modifications to our existing approach:

1. In a practical setting, we can use the measurements obtained for a specific duration of time and stack them as tensors. The measurement tensor will have zero entries for those instances where they are not available. Once the measurement tensor is formed, we can perform the joint phase identification and tensor completion as discussed in Algorithm 10.
2. In a control room setting, the multi-timescale measurements can be either reconciled in real-time or batch-wise using the approaches proposed in our prior work [139], [126]. The measurements can be first imputed or predicted at a common timescale before using it for joint phase identification and tensor completion approach (Algorithm 10).

8.1.4 Simulation results

In this section, the proposed approach is validated on two test systems, namely a three-phase unbalanced IEEE 37 bus and IEEE 123 bus test system. The substation is selected as the slack bus, whose phase labels are known. We consider a 1-day and 2-day load profile at the

Algorithm 10 Alternating minimization Algorithm for phase identification and tensor completion in distribution system

Input: Limited voltage measurements $\mathbf{V}$, Limited Power injection measurements $\mathbf{X}$, Set of known bus label indices Ω , $\mathbf{s}$, $\mathbf{t}$, number of iterations $iter$, $\mathbf{A}_{\text{known}}$, $\mathbf{Y}_{LL}^s$

Initialization: Initialize $\mathbf{S}^{(0)} = \mathbf{eigenvector}(\text{diag}(\mathbf{A}_{\text{known}}\mathbf{1}) - \mathbf{A}_{\text{known}})$

- 1: **for** $z = 1, \dots, iter$ **do**
- 2: Solve (8.10).
- 3: Update $\mathbf{S}$, which is formed by the k eigenvectors of $\mathbf{L}^{(z)} = \mathbf{D}^{(z)} - \mathbf{A}^{(z)}$ corresponding to k smallest eigenvalues. $\mathbf{D}^{(z)} = \text{diag}(\mathbf{A}^{(z)}\mathbf{1})$.
- 4: Perform DSSE (8.17) and estimate the tensor $\hat{\mathcal{X}}$.
- 5: **end for**
- 6: **return** $\mathbf{A}$, $\hat{\mathcal{X}}$

load bus for IEEE 37 bus and IEEE 123 bus respectively. This profile contains hourly smart meter readings from Pacific Gas and Electric residential and commercial customers. Since IEEE 37 bus is a primary distribution grid, the real power at each bus is an aggregation of customers which are randomly considered in the range 75-100. We have considered a power factor of 0.95 and calculated the corresponding reactive power for all the loads. It is important to note that our proposed phase identification approach is general and can be applied to other cases of reactive power loads as well. The reactive and active power measurements are corrupted with measurement noise of mean 0 and standard deviation of 10% of the actual power values. In order to generate the voltage time-series ground-truth data on an hourly basis, power-flow analysis is run for one day. Errors are introduced in the voltage magnitude measurements with mean 0 and standard deviation of 1% of the actual voltage values. The loads attached to each phase are unequal. Hence, the system is unbalanced.

IEEE 37 bus test system

We have compared the proposed approach with the spectral clustering approach as proposed in [58] for IEEE 37 bus system under different scenarios as,

- Limited temporal measurement data
- Unknown phase labels

- Incorrect phase labels
- Limited spatial measurement data

1) Limited temporal measurement data - In this scenario, we introduce random missing measurements in the voltage measurement matrix. The voltage entries in the matrix $\mathbf{V}$ are set to 0 where the measurements are missing. We vary the percentage of available data from 10% to 90% and plot the results as shown in Fig.8.3. Here, the percentage of known phase labels is set to 10%. It can be inferred that as the data availability increases, the proposed approach accurately identifies the phase labels. At 90% of the available measurements, the accuracy of the proposed approach is 90%. Across all fractions of available data, the proposed approach consistently outperforms the spectral clustering approach in [58].

2) Unknown phase labels - In this case, the percentage of correctly known phase labels is varied from 0% to 90%. In order to simulate this scenario, the entries in the sampling operator matrix are 1 for a given percentage of the known phase labels. It can be seen from Fig.8.4 that as the known phase labels are increased, the accuracy of phase detection increases. At just 50% of the known phase labels, the accuracy achieved is higher than 90%. Here, the available voltage measurements is set to 40% of the matrix $\mathbf{V}$. When the phase label information is entirely unknown (0% in the x-axis of Fig.8.4), the proposed approach provides an accuracy of 40%. We have not compared this case with the spectral clustering approach as it does not incorporate any information about the known phase labels in its formulation.

3) Incorrect phase labels - In this scenario, we randomly switch the phases of the buses. In some distribution grids, up to 10% of the phase labels are incorrect or unknown. Several buses are randomly chosen to simulate the incorrect phase labels and switch their phase a voltage measurements to phase b voltage or phase c voltage data. We have considered the availability of known labels and the voltage measurements data at 50%. Fig.8.5 shows the performance of the proposed approach for different percentages of the bus with incorrect phase labels. It can be inferred that the proposed approach is insensitive to incorrect bus phase labels and achieves high fidelity with an accuracy higher than 80%. Its performance

is superior relative to the spectral clustering approach proposed in [58].

4) Limited spatial measurements - In this scenario, the voltage measurements are available at only a subset of nodes, representing an unobservable distribution system. As the availability of measurements increases from 10% to 90% with the percentage of known phase labels set to 60%, the accuracy of correctly estimating the phase labels increases as shown in Fig.8.6. The accuracy of the spectral clustering approach is less than the proposed approach for all the fraction of available data.

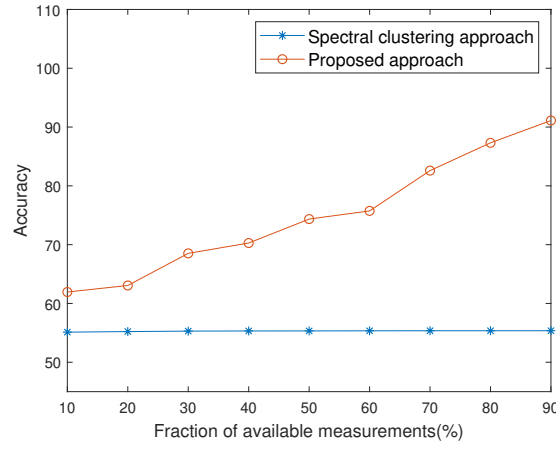


Figure 8.3: *Case I: Performance of phase identification approaches in the presence of missing measurements - IEEE 37 bus test system*

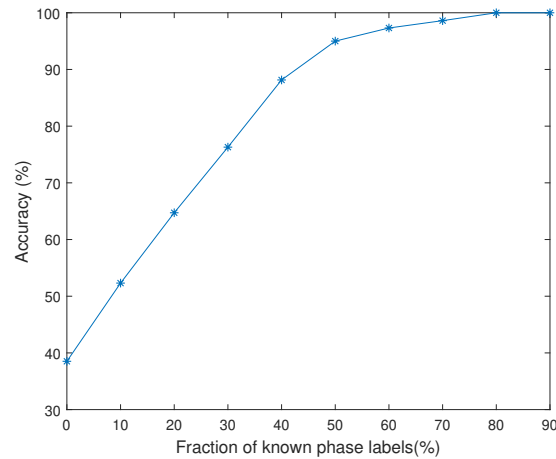


Figure 8.4: *Case II: Performance of the proposed phase identification approach in the presence of unknown phase labels- IEEE 37 bus test system*

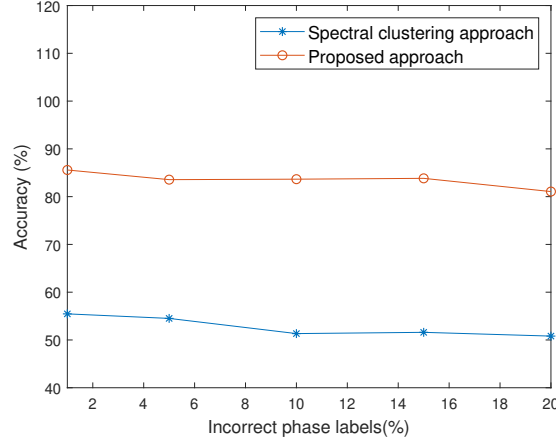


Figure 8.5: *Case III: Performance of phase identification approaches in the presence of incorrect phase labels*

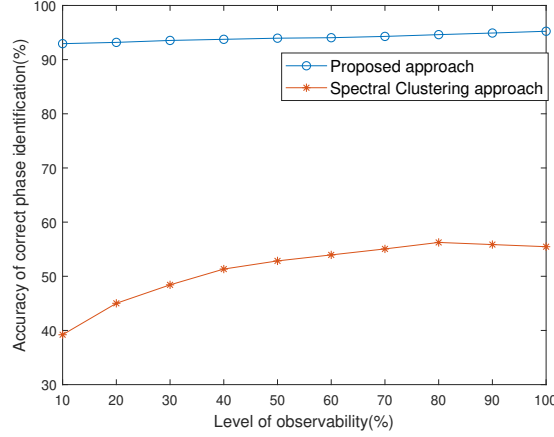


Figure 8.6: *Case IV: Performance of phase identification approaches in unobservable conditions- IEEE 37 bus test system*

C. Performance of State Estimation

The proposed approach can jointly estimate the phase labels and DSSE states as discussed in section 8.1.3. In this case, we infer the phase labels and system states jointly using Algorithm 10. We compare the performance of Algorithm 10 with respect to the tensor completion approach with incorrect model information. The incorrect model is obtained by randomly perturbing the admittance matrix entries to different phases. The proposed approach updates the correct model information by finding the correct phase label information. The estimation error of voltage magnitude and voltage angle for different fractions

of available measurements is shown in Fig. 8.7 and Fig. 8.8 respectively. It can be seen from 8.7 and Fig. 8.8 that the voltage states recovery using the proposed approach has low estimation errors at all FADs as compared to the recovery with incorrect model information. For example, error reductions up to 60% are achieved in voltage magnitude recovery at 10% FAD.

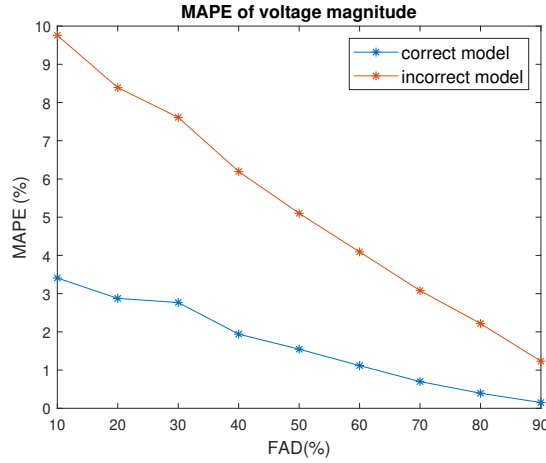


Figure 8.7: Case V: Voltage magnitude recovery for IEEE 37 bus system

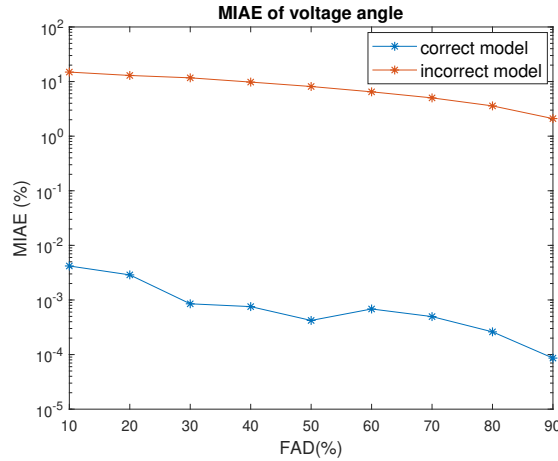


Figure 8.8: Case V: Voltage angle recovery for IEEE 37 bus system

IEEE 123 bus test system

We simulate a scenario for the IEEE 123 bus test system where the voltage measurements are available at only a subset of nodes. We vary the percentage of availability of measurements

from 10% to 90% with the percentage of known phase labels set to 40%. The accuracy of correctly estimating the phase labels is shown in Fig.8.9. As seen, the accuracy increases as the level of observability increases. We also simulate varying the percentage of phase labels known to the utility. As seen from Fig. 8.10, the accuracy improves as more phase label information is available.

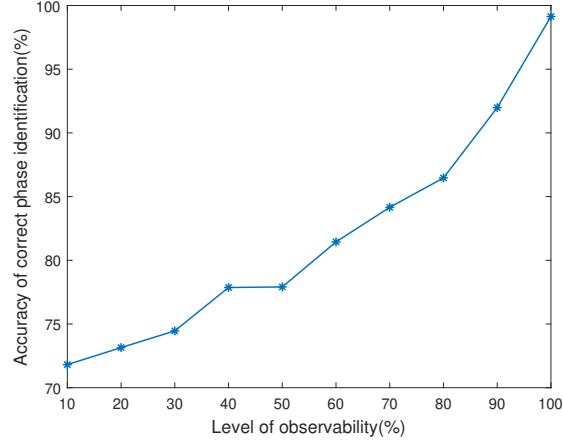


Figure 8.9: *Performance of the proposed phase identification approach under unobservable conditions - IEEE 123 bus test system*

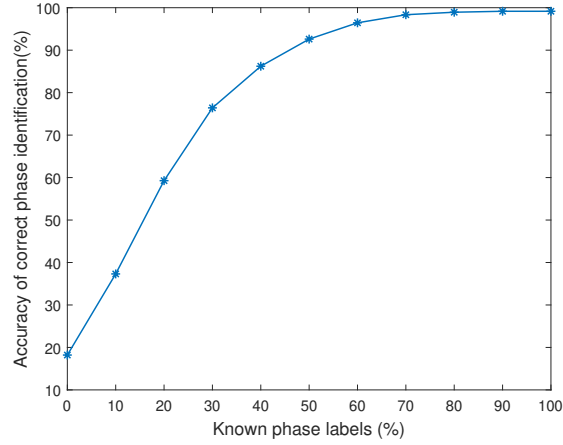


Figure 8.10: *Performance of the proposed phase identification approach in the presence of unknown phase labels - IEEE 123 bus test system*

8.2 Sensor placement and outage detection in reconfigurable distribution grid

8.2.1 System model

Consider a distribution network with V re-configurable topologies. Such a network can be perceived as a graph with M nodes and edges in each topology denoted by $\mathcal{E}_v$, $v = 1, \dots, V$. Hence, the total possible edges are $\mathcal{E} = \bigcup_{v=1}^V \mathcal{E}_v$. Let the zero injection buses in the network be Z . The sensors placed in the distribution system can be briefly classified as node and line sensors. A node sensor is installed at a distribution system node, which measures the voltage and real power flow on every edge from that node. The number of powerflow measurement channels in the node sensors may be limited or unlimited [77]. Examples of node sensors are microPMUs, whose typical sampling rate is 512 samples/cycle at 60 Hz. Thus, there are 30,720 samples/sec measurements provided by microPMUs [140]. A line sensor installed on edge (i, j) provides power-flow measurements on the edge (i, j) and voltage magnitude measurement at node j . An example of a line sensor is Sentient MM3 [141], whose waveform capturing rate is 130 samples/cycle for a frequency of 60 Hz. Thus, it is capable of providing 7800 samples of measurements per sec. We plan to leverage the high-resolution measurements provided by these sensors for outage detection. The load and measurement models for a radial distribution system are discussed in the forthcoming subsections.

Load Model

Each node $m \in M$ in the distribution network has a consumption load $x(m)$. The forecast of each load is $\hat{x}(m)$ with error $\epsilon(m) = x(m) - \hat{x}(m)$. We assume that errors are mutually independent random variables that follow $\epsilon(m) \sim N(0, \sigma^2(m))$. Given load forecasts, the true load is unknown and modeled as a random variable distributed as $x(m) \sim N(\hat{x}(m), \sigma^2(m))$. In the vector case,

$$\mathbf{x} \sim N(\hat{\mathbf{x}}, \Sigma) \quad (8.18)$$

Any load forecasting process can estimate the load forecast error covariance Σ [76]. The proposed meter placement and outage detection strategy are agnostic to the specific load forecast approach used. Only the mean and covariance of the forecasts have a bearing on the proposed approach.

Measurement Model

Sensors placed at a subset of nodes provide measurements, including branch power flow, voltage magnitude, and voltage angle. The measured power flow on edge e_i provides information about all the active downstream loads. The measured flow depends on the network topology, outage situation, and actual loads. The vector of all power flow measurements is $\mathbf{y} \in \mathbb{R}^p$. Here, p denotes the total powerflow measurements available from the node and line sensors. The measured power consumption of the i th sensor measurement under a particular outage scenario and topology is,

$$y_i(H) = \sum_{m \in M_i} x_i(m) \quad (8.19)$$

where M_i is the set of nodes to be summed over for a particular outage hypothesis H and network topology. The set represents the loads that remain to be served via that line under a particular outage scenario. Thus, y_i represents the sum of the power in nodes M_i . The nodes present in M_i represent the set of all child nodes of the node i . In general, the observed flows $\mathbf{y} \in \mathbb{R}^p$ can be expressed as,

$$\mathbf{y} = \Gamma_H \mathbf{x} \quad (8.20)$$

where $\Gamma_H \in \{0, 1\}^{p \times M}$ depends on the outage scenario for particular network topology. Correct estimation of Γ_H provides the exact information about the location of the outage in the network. The following section discusses estimating Γ_H in a radial distribution system.

8.2.2 Proposed sensor placement approach

As discussed earlier, the key challenge in outage detection is the estimation of Γ_H . To correctly estimate Γ_H , the proposed approach involves two tasks: outage identifiability and outage detection. The first task deploys meters at strategic locations such that any outage can be uniquely identified for a reconfigurable distribution network. The second task leverages the measurements from the deployed meters to detect the exact location of the outages in DN. Therefore, our approach is divided into two tasks, meter placement, and outage detection. We propose an optimization approach for placing meters at different locations in the network. Outage detection uses these measurements for detecting outages. However, the proposed outage detection is general and can be performed independently with any other sensor placement approach as well.

In this section, we map the meter placement problem to a convex optimization formulation to ensure the outage identifiability at a minimal cost. Unlike the approach used in [77], we develop a convex formulation that takes into account the possibility of different possible switch configurations of the network. Let s_i be a binary variable that determines if the node sensor is placed at node $i \in M$, i.e., $s_i = 1$ if the meter is placed at node i otherwise $s_i = 0$. Similarly, let the edge sensor placed at edge (i, j) be represented by the binary variable $t_{(i,j)}$. Let a_i and $b_{(i,j)}$ be the cost of placing node sensor and line sensors on the vertex i and edge (i, j) respectively. $\mathbf{a}$ and $\mathbf{b}$ represent the vectors of a_i and $b_{(i,j)}$. The goal of the meter placement problem is to minimize the cost of installing node sensors ($\mathbf{s}$) and line sensors ($\mathbf{t}$) for outage identifiability for different switch configurations. The degree of node k in configuration v is $d_{v,k}$. $C_{v,k}$ represents the child nodes of k th node for the v th configuration. The optimization problem proposed in [77] can be extended for re-configurable distribution systems in (8.21). The objective function and constraints associated with the optimization problem (8.21) are detailed below:

Objective function (8.21a):

The objective is to minimize the cost of installing the node sensors $\mathbf{s}$ and line sensors $\mathbf{t}$ in the network.

Constraint (8.21b):

This constraint indicates that the sensor locations are binary variables.

Constraint (8.21c):

Constraint (8.21c) ensures that the outages of child edges of root node for each configuration are identifiable. The root node that provides power to all the loads downstream in the network may differ for each configuration. The degree of the root node for configuration v is $d_{v,1}$. Placing a sensor at the root node or its child edges is essential to know if they are in an outage or not. Constraint (8.21c) requires a combination of node sensor and line sensor to monitor all $d_{v,1} - 1$ child edges of the root node.

$$\min_{\mathbf{s}, \mathbf{t}} \quad \sum_{i \in V} a_i s_i + \sum_{(i,j) \in \mathcal{E}} b_{(i,j)} t_{(i,j)} \quad (8.21a)$$

$$\text{subject to} \quad \mathbf{s}, \mathbf{t} \in \{0, 1\} \quad (8.21b)$$

$$d_{v,1} s_1 + \sum_{j \in C_{v,1}} s_j + \sum_{(1,j) \in \mathcal{E}_v} t_{1,j} \geq d_{v,1} - 1, \quad (8.21c)$$

$$d_{m,k} s_k + \sum_{j \in C_{v,k}} s_j + \sum_{(k,j) \in \mathcal{E}_v} t_{k,j} \geq d_{v,k} - 2 \quad \forall m \in V, \forall k \in M \setminus \{1\}, d_{v,k} \geq 3 \quad (8.21d)$$

$$s_j + t_{(i,j)} \leq 1 \quad \forall (i,j) \in \mathcal{E}_v, \quad \forall v \in V \quad (8.21e)$$

$$s_k + t(p_k, k) = 1 \quad \forall k \in Z \quad (8.21f)$$

Constraint (8.21d):

This constraint ensures that the outages of child edges of non-root nodes are identifiable for each configuration. Here, the sensors are placed at the non-root nodes whose degree is greater than or equal to 3. A non-root node $k \in M$ for a configuration $v \in V$ and degree $d_{v,k} \geq 3$ has $d_{v,k} - 1$ child edges and one parent edge. According to constraint (P1d), at least $d_{v,k} - 2$ child edges need to be monitored. The outage of the un-monitored child edge can be identified using the powerflow measurements of the $d_{v,k} - 2$ child edges and the powerflow on the parent edge.

Constraint (8.21e):

In addition to constraint (8.21e), this constraint (P1e) ensures that we do not have both a line sensor on line (i, j) and a node sensor at node j as this is redundant.

Constraint (8.21f):

This constraint ensures the outage identifiability of edges of zero injection nodes. Additional sensors are required if zero injection nodes are present in the network as described in constraint (P1f). The zero injection nodes are identified as those with no loads connected. Constraint (P1f) is independent of the load levels at non-zero nodes.

The optimization formulation of (8.21) is a binary integer problem that is not solvable in polynomial time. Therefore, we transform (8.21) into a convex problem, where the binary

variables $\mathbf{s}$ and $\mathbf{t}$ are relaxed to continuous variables. The convex formulation is given as,

$$\min_{\mathbf{s}, \mathbf{t}} \quad \mathbf{a}^\top \mathbf{s} + \mathbf{b}^\top \mathbf{t} \quad (8.22a)$$

$$\text{subject to} \quad A_v \begin{bmatrix} \mathbf{s} \\ \mathbf{t} \end{bmatrix} \leq c_v, \forall v \in V, \quad (8.22b)$$

$$G_v \begin{bmatrix} \mathbf{s} \\ \mathbf{t} \end{bmatrix} \leq \mathbf{1}, \forall v \in V, \quad (8.22c)$$

$$D \begin{bmatrix} \mathbf{s} \\ \mathbf{t} \end{bmatrix} = \mathbf{1}, \quad (8.22d)$$

$$0 \leq \mathbf{s} \leq 1, \quad (8.22e)$$

$$0 \leq \mathbf{t} \leq 1 \quad (8.22f)$$

Here A_v, c_v concatenates the root-node constraints (8.21c) and non-root constraints (8.21d). G_v forms the constraints (8.21e), whereas D forms the equality constraints (8.21f). c_v for the v th configuration is a vector whose elements are defined as,

$$c_v = \begin{bmatrix} -(d_{v,1} - 1) \\ -(d_{v,k} - 2) \end{bmatrix} \quad (8.23)$$

The convex formulation resembles the relaxed linear programming form of (8.21). As the problem is relaxed, the objective value $f(8.22)$ is a lower bound to the objective value $f(8.21)$. The integrality gap is denoted by $\frac{f(8.22)}{f(8.21)}$. However, the solution obtained in (8.22) is continuous and not integral; hence we round them to binary values. The variables s and t are rounded to 1 if their value is greater than 0.5 and 0 if they are less than 0.5. Such a rounding technique increases the objective function value. However, rounding the solution $f(\text{ALG})$ should be within a factor $\alpha > 1$ to the optimum cost as, $f(\text{ALG}) \leq \alpha f(8.22) \leq \alpha f(8.21)$. Hence, the rounded solution is an α -approximation of the optimal solution [142].

As seen from the optimization problem (8.22), the constraints matrix (8.22b) is defined

separately for each configuration. The constraint (8.22b) is modified by merging the elements for all the configurations given as,

$$[A_1; A_2; \dots A_V] \begin{bmatrix} \mathbf{s} \\ \mathbf{t} \end{bmatrix} \leq [c_1; c_2; \dots, c_V] \quad (8.24)$$

However, several constraints are redundant in accumulating the constraints matrix (8.22b). The elements in the matrices $A_2, A_3, \dots, A_V$ are repetitive if the degree of the nodes with respect to the reference configuration 1 is unchanged. This leads to many constraints, increasing the dimension of the problem. In order to solve this issue, the constraint (8.24) can be reduced by removing the redundant elements. To perform this, the nodes whose degree is changed with respect to the reference configuration 1 are identified. These nodes are extracted, and a vector R_k is formed. These nodes in R_k are involved in network reconfiguration in any subsequent $(k + 1)th$ configuration. Using the vector R_k , a reduced size constraints matrix is derived, represented as,

$$[A_1; A_{r2}; \dots A_{rV}] \begin{bmatrix} \mathbf{s} \\ \mathbf{t} \end{bmatrix} \leq [c_1; c_{r2}; \dots c_{rV}] \quad (8.25)$$

Here, $A_{r2}, \dots, A_{rV}$ are the reduced rows of the matrix $A_2, A_3, \dots, A_V$ respectively. Using (8.25), we develop a full dimension constraint matrix for the reference configuration, and a reduced size matrix for the rest of the configurations [143].

The solution to the convex optimization problem (8.22) using the reduced size constraints (8.25) provides the location of meters in a distribution grid. These meter locations ensure outage identification for any outages in a re-configurable distribution system. The forthcoming subsection proposes a novel outage detection algorithm that leverages the powerflow measurements from the deployed meters.

8.2.3 Proposed outage detector

The goal of outage detection is to identify the set of edges that are isolated from the network. Given the vector of load forecasts $\hat{\mathbf{x}}$, forecast error covariance $\mathbf{\Sigma}$ and the real-time load flow measurements $\mathbf{y}$ along a set of edges acquired for N snapshots of time, the aim is to determine the correct number and location of the edges in an outage. This is equivalent to determining the matrix Γ_H as defined in (8.20) for each outage scenario H . As discussed in section 8.2.1, the matrix Γ_H is $\{0, 1\}^{p \times M}$. Using (8.18), we can write (8.20) as,

$$\mathbf{y} = \Gamma_H(\hat{\mathbf{x}} + \mathbf{e}) = \Gamma_H\hat{\mathbf{x}} + \mathbf{e} \quad (8.26)$$

where, $\mathbf{e} = N(0, \mathbf{Q})$, $\mathbf{Q} = \Gamma_H\mathbf{\Sigma}\Gamma_H^\top$ is an unknown covariance matrix.

Let $\{\mathbf{y}_1, \mathbf{x}_1, \dots, \mathbf{y}_N, \mathbf{x}_N\}$, $N \geq p + M$ denote the samples of available measurements and loads. The N samples of $\mathbf{x}_t$ are obtained from the multivariate distribution $N(\hat{\mathbf{x}}, \mathbf{\Sigma})$. The likelihood function is given as,

$$L = (2\pi)^{-\frac{1}{2}Np} |\mathbf{Q}^*|^{-\frac{1}{2}N} \exp\left(-\frac{1}{2} \sum_{t=1}^N (\mathbf{y}_t - \Gamma_H^* \mathbf{x}_t)^\top \mathbf{Q}^{*-1} (\mathbf{y}_t - \Gamma_H^* \mathbf{x}_t)\right) \quad (8.27)$$

The maximum likelihood estimation involves determining the Γ_H^* and $\mathbf{Q}^*$ given $\mathbf{y}_1, \mathbf{x}_1, \dots, \mathbf{y}_N, \mathbf{x}_N$ that maximize the likelihood function. The maximum likelihood estimates of Γ_H^* and $\mathbf{Q}^*$ are obtained using Theorem 6.

Theorem 6. *If $\mathbf{y}_t$ is a measurement from $N(\Gamma_H\hat{\mathbf{x}}, \mathbf{Q})$, $t = 1, \dots, N$, with $(\mathbf{x}_1, \dots, \mathbf{x}_N)$ are the samples with rank M from the distribution $N(\hat{\mathbf{x}}, \mathbf{\Sigma})$, the maximum likelihood estimator of Γ_H is given as, $\Gamma_H^* = \mathbf{C}\mathbf{A}^{-1}$, where $\mathbf{C} = \frac{1}{N} \sum_{t=1}^N \mathbf{y}_t \mathbf{x}_t^\top$ and $\mathbf{A} = \frac{1}{N} \sum_{t=1}^N \mathbf{x}_t \mathbf{x}_t^\top$. The unknown covariance matrix $\mathbf{Q}^*$ is given as, $\mathbf{Q}^* = \frac{1}{N} \sum_{t=1}^N (\mathbf{y}_t - \Gamma_H^* \mathbf{x}_t)(\mathbf{y}_t - \Gamma_H^* \mathbf{x}_t)^\top$.*

Proof: The proof of this classic theorem can be found in [144].

In order to perform outage detection, we leverage the power flow measurements provided by the meters using convex optimization formulation (8.22) at multiple snapshots of time. The powerflow measurements are obtained in real-time, capturing the network's actual op-

erating state. Furthermore, load forecasts and associated variances are required at each load bus of the DN. Using the load forecasts and their variances, we obtain random samples from the multivariate distribution as given in (8.18). Using the powerflow measurements $\mathbf{y}$ as well as the load samples $\mathbf{x}$, the matrix Γ_H^* , as defined in Theorem 6, is estimated. As the elements of the true Γ_H matrix are binary i.e., $\{0, 1\}$, we first normalize the elements of the estimated matrix Γ_H^* in the $(0, 1)$ interval. After normalization, we compare each matrix element value with a threshold of 0.5. A threshold of 0.5 was chosen to get unbiased solutions. The element value is set to 1 if its normalized value is greater than the threshold, else it is set to 0. Algorithm 11 illustrates the proposed outage detection algorithm.

Algorithm 11 Outage detection Algorithm

Input: Meter placement sensors locations, N samples of powerflow measurements $\mathbf{y}_t \in \mathbb{R}^p$, $\hat{\mathbf{x}} \in \mathbb{R}^M$, $\Sigma \in \mathbb{R}^{M \times M}$

- 1: Obtain N samples of $\mathbf{x}_t$ from the distribution $N(\hat{\mathbf{x}}, \Sigma)$.
- 2: Obtain samples covariance between sensor measurements and load, as $\mathbf{C} = \frac{1}{N} \sum_{t=1}^N \mathbf{y}_t \mathbf{x}_t^\top$.
- 3: Obtain samples covariance between load as $\mathbf{A} = \frac{1}{N} \sum_{t=1}^N \mathbf{x}_t \mathbf{x}_t^\top$.
- 4: Obtain $\Gamma_H^* = \mathbf{C} \mathbf{A}^{-1}$
- 5: Normalize $\Gamma_H^* = \frac{\Gamma_H^* - \min(\text{vec}(\Gamma_H^*))}{\max(\text{vec}(\Gamma_H^*)) - \min(\text{vec}(\Gamma_H^*))}$.
- 6: If the element value of the matrix $\Gamma_H^* \geq 0.5$, the element value is set to 1 else it is set to 0.
- 7: **return** Γ_H^*

8.2.4 Simulation Results

We simulate the proposed approach in IEEE 33 bus and IEEE 119 bus test systems. Both these test systems considered are reconfigurable distribution systems. To demonstrate the capability of the proposed approach for a secondary distribution network, we have developed a modified IEEE 33 bus system by adding secondary circuits. We discuss the results for these three distribution systems in the following subsections. All the simulations are performed in a MATLAB environment.

IEEE 33 bus distribution system

IEEE 33 bus system is a standard 12.66 kV radial DN with 33 tie switches and five sectional switches. We have considered two configurations as discussed in [145]. Using the convex formulation of the meter placement approach as given in (8.22), we obtain the number, type, and locations of meters in an IEEE 33 bus DN. We simulate different case scenarios to evaluate the efficacy of the proposed approach for the IEEE 33 bus system.

1. *Effect of sensor costs:* In this scenario, we vary the costs of the placements of different meters. Typically, the node sensors are 2-5 times more costly than line sensors [77]. Table 8.1 shows the IEEE 33 bus system's meter placement solution for different line and node sensor costs. It can be inferred that the node sensors are reduced as their relative cost to line sensors increases. In all three cases shown in Table 8.1, it is found that the objective value using the convex formulation is bounded by a factor of $\alpha = 1.6$ to the integer programming solution. Fig.8.11 and 8.12 show the meter location for two configurations of a 33-bus distribution system. Here, we assume the cost of node sensors is five times that of the cost of line sensors (i.e., $\mathbf{a} = 5\mathbf{b}$). As seen from Figs 8.11-8.12, it can be inferred that all line sensors are placed in a distribution system as they are less expensive than node sensors.

Table 8.1: *Meter placement solution for IEEE 33 bus reconfigurable distribution system*

Scenario	$ V_p $	$ E_p $	meter locations
Case 1 $\mathbf{a} = 2\mathbf{b}$	5	1	t: 1-2 s: 2,3,6,12,21
Case 2 $\mathbf{a} = 3\mathbf{b}$	2	4	t: 1-2, 6-7, 12-13, 21-22 s: 2,3
Case 3 $\mathbf{a} = 5\mathbf{b}$	0	6	t: 1-2, 2-3, 3-4, 6-7, 12-13, 21-22

2. *Effect of load forecast error:* After the sensors are placed in the distribution system, we perform outage detection using the approach proposed in section 8.2.3. We leverage the power flow measurements from the node and line sensors in the distribution system. The load covariance matrix is chosen to be $\Sigma = \text{diag}(k\hat{\mathbf{x}})$, where k is a constant [76]. The typical values of k observed during load forecast are 0.1-0.3. We obtain sampled measurements from the multivariate normal distribution as given in (8.26).

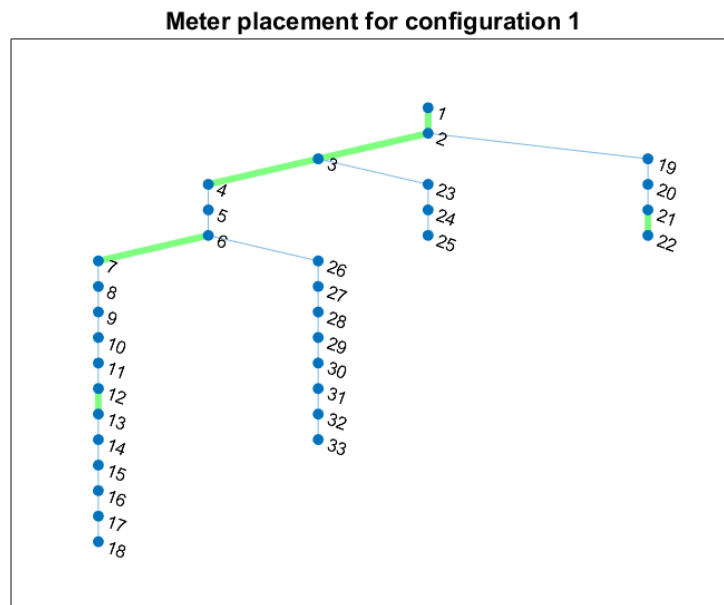


Figure 8.11: *Meter placement for IEEE 33 bus system - Configuration 1*

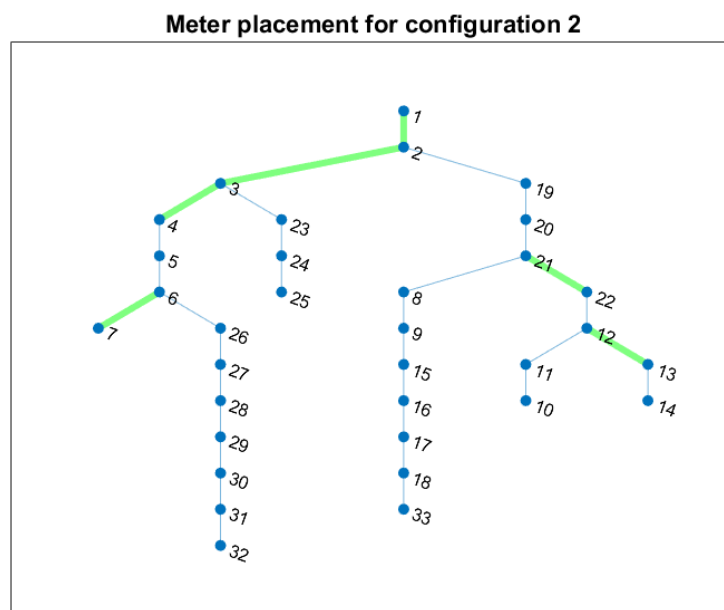


Figure 8.12: *Meter placement for IEEE 33 bus system - Configuration 2*

We randomly introduce multiple edge outage scenarios and obtain the corresponding noisy power flow measurements. Here, the power flow measurements also consider the dis-

tribution network's line losses. Using the powerflow measurements and load forecasts, we perform outage detection using Algorithm 11. Fig. 8.13 illustrates the probability of correctly estimating the matrix for different values of load covariance factor k . It can be inferred that, as k is increased, the probability of detection is reduced. The decrease in the accuracy is especially pronounced for fewer measurement samples ($N < 200$). However, as more measurement samples are available, the probability of detection is 1 for any value of load covariance factor k . Fig. 8.14 shows the results for two optimal network configurations of IEEE 33 bus systems. The proposed approach can estimate the correct Γ_H matrix for all the configurations with probability one as the number of samples is increased.

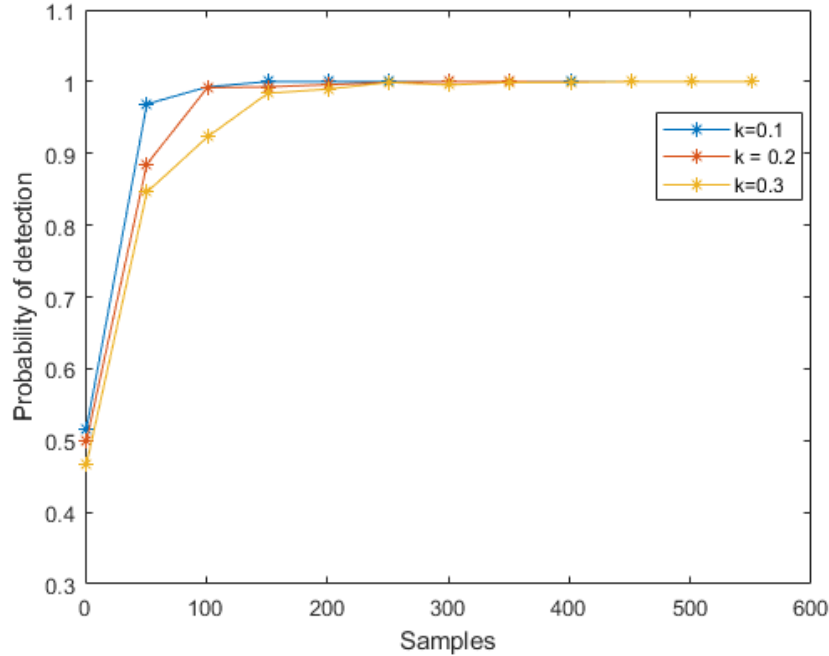


Figure 8.13: *Probability of outage detection for different load covariances of 33 bus test system*

The proposed meter placement approach for re-configurable DN is compared with [77] for outage detection. Ref. [77] considers a single configuration and there are four line sensors placed at edge (1, 2), (2, 3), (3, 23), (6, 7). Using these meter locations, we perform outage detection as discussed in section 8.2.3. The probability of correctly detecting the outages is tabulated in Table.8.2. Using the method in [77], the probability of detecting outages in the

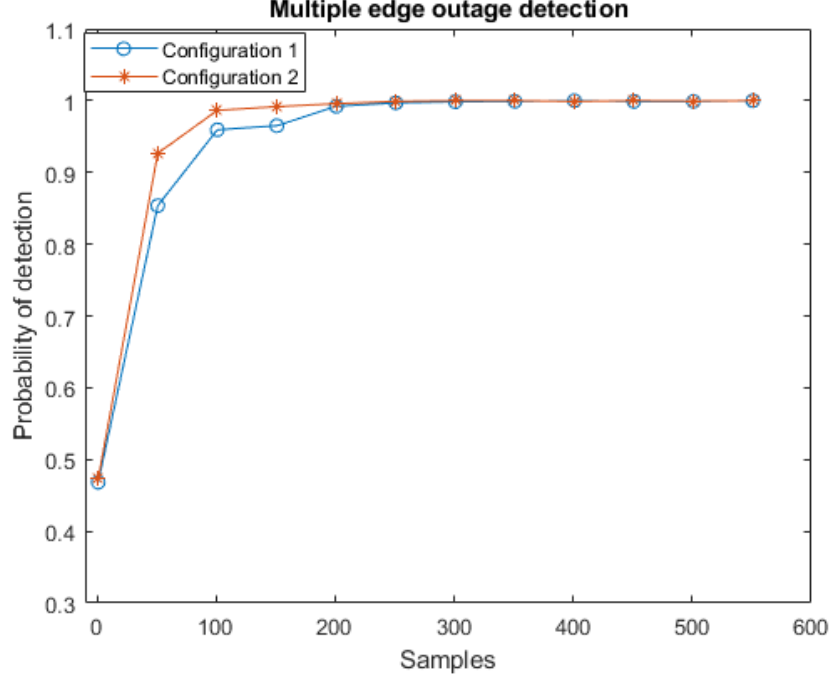


Figure 8.14: Probability of outage detection for different configurations of 33 bus test system

Table 8.2: Outage detection using different meter placement approaches

Approach	Probability of detection
Proposed meter placement for reconfigurable DN	0.99
Meter placement considering a single configuration [77]	0.84

entire network is lower than our proposed approach because the meter placement approach [77] is designed only for a single configuration.

3. Comparison of the proposed approach with ML detector [76]:

We compare the proposed outage detector approach with the maximum likelihood outage detector proposed in [76]. Note that the sensor placement is the same for both approaches with costs $\mathbf{a} = 5\mathbf{b}$. Here, we simulate single-edge outages. The ML single-edge outage detector computes the likelihood for each of the hypotheses in $H \in \mathcal{H}$ and chooses the maximum. That is,

$$\hat{H} = \max_{H \in \mathcal{H}} p(\mathbf{y}|\mathbf{x}, H) \quad (8.28)$$

where $\mathbf{x} \sim N(\hat{\mathbf{x}}, \Sigma)$. We generate 100 Monte Carlo simulations. We randomly generate

Table 8.3: *Comparison of the proposed approach with state-of-the-art approaches*

Approach	Probability of detection (with measurement noise)	Probability of detection (without measurement noise)
Proposed outage detector	0.9912	0.9995
ML detector	0.70	0.90

a single edge outage and corresponding noisy sensor measurements at each iteration. We consider a particular hypothesis scenario H and compute the corresponding Γ_H . We then compare the likelihood of the measurements $\mathbf{y}$ given the load forecast, covariance matrix, and Γ_H . We compute the probability of successful detection of the hypothesis. Table 8.3 shows the performance of the naive ML detector with the proposed outage detector for single-edge outages in the DN. We have considered two scenarios where we consider the sensor measurements with and without noise. The measurement errors are assumed to be 0 mean with standard deviation equal to 1% of the actual values. For the case with no measurement noise and only load forecast covariance Σ , the probability of successfully detecting single-edge outages is 0.99 for the proposed approach compared to 0.9 using the naive ML detector. With measurement noise, the probability detection of the ML detector is reduced to 0.7. The performance of the ML detector is poor due to the reason that many of the hypothesis maps to the same observed flows, making the detector output non-unique [76]. Furthermore, the ML detector approach does not consider the effect of measurement noise statistics.

We have considered only single-edge outages to compare the two approaches. In reality, however, multiple edge outage scenario takes place in DN. Therefore, as the outage size K increases, enumerating the entire ML detector requires $\sum_{k=1}^K \binom{|E|}{k}$ evaluations. Evaluating a large set of hypotheses is computationally expensive, making the ML approach impractical.

IEEE 119 bus distribution system

IEEE 119 bus system is an 11 kV distribution system with 118 sectionalizing switches, and 15 tie switches [146]. We have considered two re-configurable topologies described in [146]. Simulations for IEEE 119 bus system are performed with similar case scenarios as

seen for the IEEE 33 bus system in section 8.2.4. In the first case, the meter placement varies with different placement costs. Meter locations for various costs of node and line sensors are tabulated in Table 8.4. In the second case, we perform outage detection using Algorithm 11, and meter placement cost $\mathbf{a} = 5\mathbf{b}$. Here, we have introduced an outage in lines 6-7 and 40-41. Fig.8.15 shows the probability of correctly estimating the Γ_H matrix for load covariance factor $k = 0.3$. As seen from the results, both configurations achieve a maximum detection probability of 0.995 with 1000 measurement samples. We have tested our proposed approach for different possible outage scenarios. It can be inferred that for all the possible combinations, the proposed approach is able to estimate the matrix correctly. We can accurately estimate Γ_H within a few seconds. For instance, it takes 12 ms for line sensors to provide 1000 samples of measurements, while for node sensors, 1000 samples can be captured in 3.25 ms. Hence, our proposed approach can detect the outage scenario quickly and correctly.

Table 8.4: *Meter placement solution for IEEE 119 bus re-configurable distribution system*

Scenario	$ V_p $	$ E_p $	meter locations
Case 1 $\mathbf{a} = 2\mathbf{b}$	15	0	s: 1,2,4,8, 11, 25, 29, 30, 64,65, 73 79, 91, 100, 110
Case 2 $\mathbf{a} = 3\mathbf{b}$	10	5	t: 8-9, 11-12, 25-26, 79-80, 110-111 s: 1, 2, 4, 29, 30, 64, 65, 73, 91, 100
Case 3 $\mathbf{a} = 5\mathbf{b}$	2	15	t: 1-2, 4-5, 8-9, 11-12, 25-26, 29-30 29-38, 30-31, 64-65, 65-66, 73-74, 79-80, 91-92, 100-101, 110-111 s: 1, 2

Modified IEEE 33 bus system

Most of the outages occur in the lateral feeders of the distribution network. These laterals feeders are connected to the step-down service transformers, which feed the power to the consumers. We modify the IEEE 33 bus system to check the performance of the outage detection in lateral feeders. The network is modified by adding lateral feeders to primary nodes of the original IEEE 33 bus system. These lateral feeders are ten and have line configuration 725. The admittance matrices for configuration type 725 are available in [95].

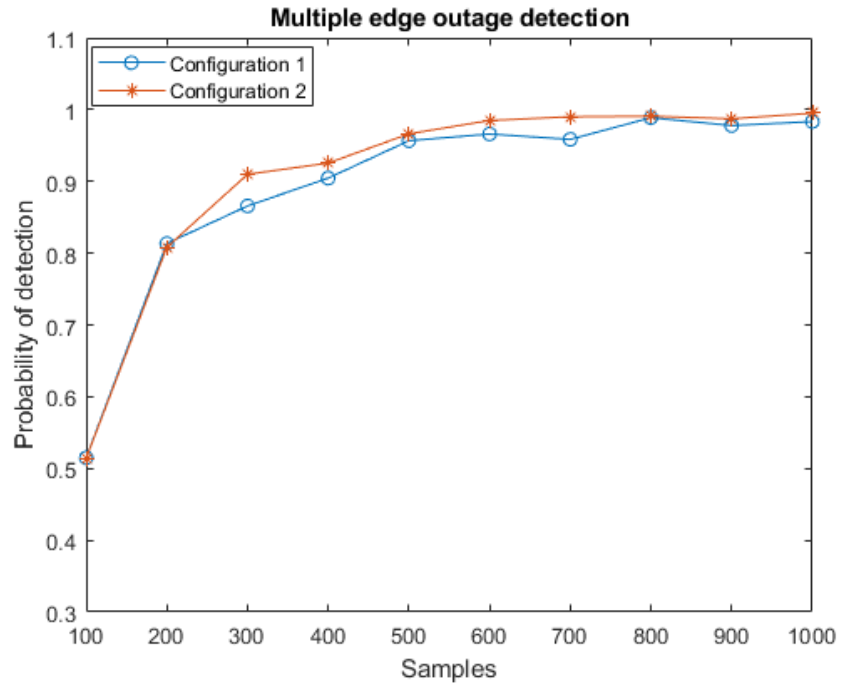


Figure 8.15: *Probability of outage detection for different configurations of IEEE 119 bus test system*

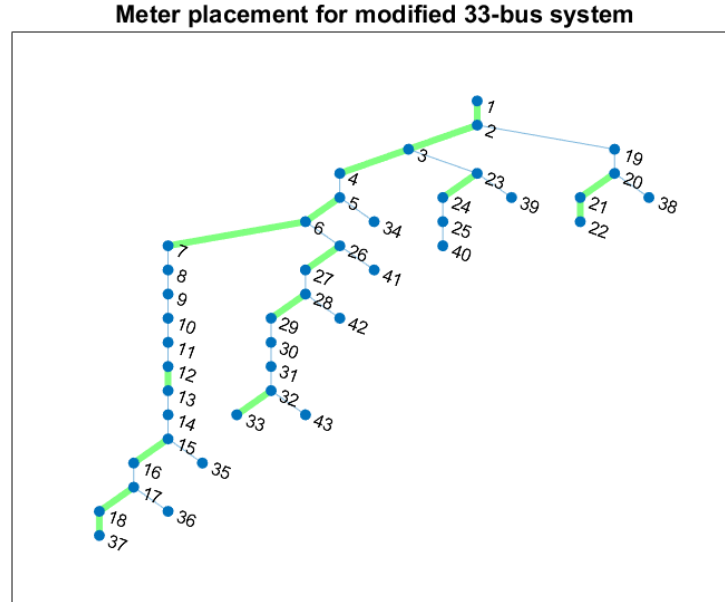


Figure 8.16: *Meter placement for modified IEEE 33 bus system*

Loads connected at these lateral feeders represent the aggregated consumer loads connected to their service transformers. Fig.8.16 shows the meter locations in the modified IEEE 33 bus system for the case when node sensors are five times more costly than edge sensors ($\mathbf{a} = 5\mathbf{b}$). After placing meters, we randomly generate outages at different edges and corresponding noisy sensor measurements for 100 iterations. We consider total sensor measurements as 500 and load forecast covariance matrix is chosen as $\Sigma = \text{diag}(0.3\hat{\mathbf{x}})$, where $\hat{\mathbf{x}}$ represents the load forecasts. The probability of detecting the outages was observed to be 0.997. Thus, the proposed outage detection is also effective for detecting outages in feeder laterals.

8.3 Summary

This chapter presents a novel optimization formulation for inferring the phase connectivity of the entire distribution network. The approach is robust to incorrect phase label errors with minimum measurement data requirements. Simulation results for IEEE 37 bus system reveal the accurate performance of phase label identification and state estimation with limited measurement data. We also propose a novel outage detection approach for a re-configurable distribution network. In this regard, we propose a convex formulation of the meter placement approach for outage identifiability. We leverage the measurements from the deployed sensors and formulate a maximum likelihood outage detection approach. Simulation results for IEEE 33 and IEEE 119 bus systems reveal that outages can be identified accurately and efficiently within a few seconds.

Chapter 9

Conclusion and future work

9.1 Conclusions

This chapter provides a summary of the dissertation’s contributions and outlines potential avenues for future research. The dissertation focuses on improving situational awareness in distribution grids and addresses fundamental questions. The proposed methods share a common thread of utilizing sparsity, reducing complexity, and enhancing efficacy. To achieve this objective, various data-driven and optimization formulations are developed, incorporating the physics of the power grid. These approaches are useful for distribution grid utility companies to improve the visibility of the grid. The accomplishments of this dissertation can be summarized as follows:

Chapter 4 focuses on developing optimization formulations for robust and efficient state estimation to tackle the uncertainties in the system model and measurement data. Specifically, it develops a robust sparsity-based DSSE and incorporates the sparsity of both the bad data and measurements in a single optimization formulation. The optimization formulation achieves robustness to network parameter uncertainty using a worst-case robust approximation. Additionally, this chapter develops a novel joint sparsity-based DSSE framework effective in highly unobservable conditions. Results in the standard test networks demonstrate that the error performance of the proposed approach relative to the conventional

methods offers significant improvements.

Chapter 5 develops a framework to integrate the multi time-scale measurements with sparsity based state estimators in a distribution network. This chapter focuses on imputing the multiple unevenly sampled time-series data using Multi-task Gaussian process approach. Once the consistent time-series along-with the confidence bounds in the imputations are obtained, they are fed into a Bayesian matrix completion method augmented with linearized powerflow constraints. Simulation results in the unbalanced IEEE test systems demonstrate the approach effectively imputes the time-series data in the presence of missing measurements. However, this approach is computationally expensive as it involves kernel matrix inversion operation considering all measurements at once. To address this issue, Chapter 6 develop a recursive Gaussian process based approach which handles streaming data and requires kernel matrix inversion only at the initial time step. In this approach, the network connectivity information can also be accounted for aggregating the multi time-scale measurements. Simulation results on different IEEE test systems reflects the capability of the proposed approach in aggregating measurements batch-wise or real-time.

To further improve the scalability and accuracy, Chapter 7 develops a novel latent neural ODE (LODE) approach and a neural SDE-RNN approach. These leverages spatio-temporal dependencies in time-series data without involving matrix inverse operations. The proposed LODE approach is flexible in performing both imputations and predictions while being computationally efficient. The SDE-RNN approach allows us to quantify both aleatoric and epistemic uncertainty in the imputed measurements. Analytical results capturing these sources of uncertainty have been derived.

Finally, Chapter 8 enhances the situational awareness of the grid by leveraging the spatio-temporal correlations in the measurement data to perform two tasks, namely phase identification and outage detection. The phase identification approach is accurate in identifying the phase labels under varying availability of data, phase labels errors, and unknown phase connectivity information. The outage detection approach detects multi-edge outages in a re-configurable distribution grid using maximum likelihood principle and without enumerating over any hypothesis set. We conduct various types of simulations and indicate that

the outages can be identified with a high probability using the proposed approach within a few seconds. We compare the proposed outage detection with the existing state-of-the-art approaches, and it is observed that the approach performs well even at high percentages of measurement errors, load forecast errors, and any reconfigurable topology.

9.2 Future directions

This section presents possible future directions in the areas related to distribution system challenges and effective methodologies for addressing them. The following are the potential extensions to the work presented in this dissertation:

- Chapter 4 provides robust formulations of DSSE to deal with bad data and network parameter uncertainty. However, detecting the exact location of the measurements or the lines where bad data or parameter errors exist is also critical. Developing a framework for the localization of these errors can be pursued as a part of future research efforts.
- Chapter 4 discusses a joint matrix completion and compressed sensing framework for performing state estimation in a distribution grid. Implementation of this approach is challenging for large-scale distribution networks. Developing a distributed joint MC-CS framework for DSSE can be pursued in the future.
- Chapters 5, 6, and 7 discuss various approaches to reconcile the heterogeneous measurements in the distribution grid. The increase in the penetration of distributed generation has enhanced the need for distribution grid monitoring. Distributed generation needs to be monitored at different time scales, ranging from day-ahead forecasting to real-time data periodically submitted to the DMS. This information increases the measurement redundancy in the system. As a possible extension of this work, we can investigate aggregating this forecasted information for GP imputation.
- The multi-task GP approach proposed in Chapter 5 is sensitive to outliers in the mea-

surement data. Hence, we can develop a robust Gaussian process framework against outliers as a part of our future work.

- The Gaussian process-based approaches developed in Chapter 5 and 6 assumes that measurement data is distributed as multivariate Gaussian. Developing a multi-task recursive GP approach for non-Gaussian noise can be pursued as a part of our future work.
- The proposed GP approach’s performance in Chapter 6 depends on the GP function’s hyper-parameter values. Hence, future work will involve recursively learning the hyperparameters as the GP functions are updated.
- The visibility of utilities on the low-voltage side of the grid is limited, but various parties have installed sensors in the low-voltage distribution grid. These sensors may provide measurements that differ in accuracy from those used by utility companies. Therefore, it is crucial to establish methods for incorporating and verifying these measurements to enhance grid visibility.
- Chapter 7 presents the neural SDE-RNN approach for handling the irregularly sampled time series data. It will be interesting to analyze the epistemic uncertainty obtained using the SDE model with the conventional uncertainty quantification for neural network models, like Monte-Carlo dropouts, Deep Ensemble, etc. This could also be explored as a potential future task.
- Chapter 8 proposes a joint state estimation and phase identification framework. However, this approach is challenging to implement in a large real-world distribution system. To overcome this limitation, a distributed implementation of the proposed method can be possible by dividing the large-scale network into multiple areas. Implementation of a distributed method for state estimation and phase identification will be pursued as a part of our future work.
- It will also be interesting to explore a joint phase identification and state estimation

approach in Chapter 8 that works with multi-timescale measurements as a part of our future research work.

- Formulating an outage detection and meter placement approach in Chapter 8 considering a mesh topology and the presence of distributed generation can be pursued as a part of our future research efforts.

Bibliography

- [1] Federal Energy Regulatory Commission et al. Assessment of demand response and advanced metering staff report, 2022.
- [2] Efthymios Manitsas, Ravindra Singh, Bikash C Pal, and Goran Strbac. Distribution system state estimation using an artificial neural network approach for pseudo measurement modeling. *IEEE Transactions on Power Systems*, 27(4):1888–1896, 2012.
- [3] Kevin A Clements. The impact of pseudo-measurements on state estimator accuracy. In *2011 IEEE Power and Energy Society General Meeting*, pages 1–4. IEEE, 2011.
- [4] SM Shafiul Alam, Balasubramaniam Natarajan, and Anil Pahwa. Distribution grid state estimation from compressed measurements. *IEEE Transactions on Smart Grid*, 5(4):1631–1642, 2014.
- [5] Hazhar Sufi Karimi and Balasubramaniam Natarajan. Compressive sensing based state estimation for three phase unbalanced distribution grid. In *GLOBECOM 2017-2017 IEEE Global Communications Conference*, pages 1–6. IEEE, 2017.
- [6] Amit Joshi, Laya Das, Balasubramaniam Natarajan, and Babji Srinivasan. A framework for efficient information aggregation in smart grid. *IEEE Transactions on Industrial Informatics*, 15(4):2233–2243, 2018.
- [7] Hazhar Sufi Karimi and Balasubramaniam Natarajan. Recursive dynamic compressive sensing in smart distribution systems. In *2020 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, pages 1–5. IEEE, 2020.
- [8] Mohammad Babakmehr, Marcelo G Simões, Michael B Wakin, and Farnaz Harirchi. Compressive sensing-based topology identification for smart grids. *IEEE Transactions on Industrial Informatics*, 12(2):532–543, 2016.

- [9] Priya L Donti, Yajing Liu, Andreas J Schmitt, Andrey Bernstein, Rui Yang, and Yingchen Zhang. Matrix completion for low-observability voltage estimation. *IEEE Transactions on Smart Grid*, 2019.
- [10] Rahul Madbhavi, Hazhar Sufi Karimi, Balasubramaniam Natarajan, and Babji Srinivasan. Tensor completion based state estimation in distribution systems. *11th Conference on Innovative Smart Grid Technologies, Washington DC, Feb*, 2020.
- [11] Xiao Peng Li, Lei Huang, Hing Cheung So, and Bo Zhao. A survey on matrix completion: perspective of signal processing. *arXiv preprint arXiv:1901.10885*, 2019.
- [12] Bo Liu, Hongyu Wu, Yingchen Zhang, Rui Yang, and Andrey Bernstein. Robust matrix completion state estimation in distribution systems. *arXiv preprint arXiv:1902.02009*, 2019.
- [13] Mang Liao, Di Shi, Zhe Yu, Wendong Zhu, Zhiwei Wang, and Yingmeng Xiang. Estimate the lost phasor measurement unit data using alternating direction multipliers method. In *2018 IEEE/PES Transmission and Distribution Conference and Exposition (T&D)*, pages 1–9. IEEE, 2018.
- [14] VSG Thadikemalla, Ishan Srivastava, SS Bhat, and AS Gandhi. Data loss mitigation mechanism using compressive sensing for smart grids. In *2020 IEEE International Conference on Power Electronics, Smart Grid and Renewable Energy (PESGRE2020)*, pages 1–6. IEEE, 2020.
- [15] Ahmed S Zamzam, Yajing Liu, and Andrey Bernstein. Model-free state estimation using low-rank canonical polyadic decomposition. *arXiv preprint arXiv:2004.05741*, 2020.
- [16] Alcir Monticelli. Electric power system state estimation. *Proceedings of the IEEE*, 88(2):262–282, 2000.
- [17] Murat Göl and Ali Abur. Lav based robust state estimation for systems measured by pmus. *IEEE Transactions on Smart Grid*, 5(4):1808–1814, 2014.

- [18] Mehmet K Celik and Ali Abur. A robust wlvav state estimator using transformations. *IEEE Transactions on Power Systems*, 7(1):106–113, 1992.
- [19] Kaveh Dehghanpour, Zhaoyu Wang, Jianhui Wang, Yuxuan Yuan, and Fankun Bu. A survey on state estimation techniques and challenges in smart distribution systems. *IEEE Transactions on Smart Grid*, 10(2):2312–2322, 2018.
- [20] Ali Abur and Antonio Gomez Exposito. *Power system state estimation: theory and implementation*. CRC press, 2004.
- [21] W-HE Liu, Felix F Wu, and S-M Lun. Estimation of parameter errors from measurement residuals in state estimation (power systems). *IEEE Transactions on power systems*, 7(1):81–89, 1992.
- [22] W-HE Liu and Swee-Lian Lim. Parameter error identification and estimation in power system state estimation. *IEEE Transactions on power systems*, 10(1):200–209, 1995.
- [23] Antonio Gómez-Expósito, Catalina Gómez-Quiles, and Izudin Džafić. State estimation in two time scales for smart distribution systems. *IEEE Transactions on Smart Grid*, 6(1):421–430, 2014.
- [24] Arash Alimardani, Francis Therrien, Djordje Atanackovic, Juri Jatskevich, and Ebrahim Vaahedi. Distribution system state estimation based on nonsynchronized smart meters. *IEEE Transactions on Smart Grid*, 6(6):2919–2928, 2015.
- [25] Aleksandar M Stanković, Vanja Švenda, Andrija T Sarić, and Mark K Transtrum. Hybrid power system state estimation with irregular sampling. In *2017 IEEE Power & Energy Society General Meeting*, pages 1–5. IEEE, 2017.
- [26] Jianzhong Wu, Yan He, and Nick Jenkins. A robust state estimator for medium voltage distribution networks. *IEEE Transactions on Power Systems*, 28(2):1008–1016, 2012.
- [27] Kursat Rasim Mestav, Jaime Luengo-Rozas, and Lang Tong. Bayesian state estimation

- for unobservable distribution systems via deep learning. *IEEE Transactions on Power Systems*, 34(6):4910–4920, 2019.
- [28] Boris A Alcaide-Moreno, Claudio R Fuerte-Esquivel, Mevludin Glavic, and Thierry Van Cutsem. Electric power network state tracking from multirate measurements. *IEEE Transactions on Instrumentation and Measurement*, 67(1):33–44, 2017.
- [29] Côme Carquex, Catherine Rosenberg, and Kankar Bhattacharya. State estimation in power distribution systems based on ensemble kalman filtering. *IEEE Transactions on Power Systems*, 33(6):6600–6610, 2018.
- [30] Jianhan Song, Emiliano Dall’Anese, Andrea Simonetto, and Hao Zhu. Dynamic distribution state estimation using synchrophasor data. *IEEE Transactions on Smart Grid*, 11(1):821–831, 2019.
- [31] Zachary C Lipton, David Kale, and Randall Wetzel. Directly modeling missing data in sequences with rnns: Improved classification of clinical time series. In *Machine learning for healthcare conference*, pages 253–270. PMLR, 2016.
- [32] Zhengping Che, Sanjay Purushotham, Kyunghyun Cho, David Sontag, and Yan Liu. Recurrent neural networks for multivariate time series with missing values. *Scientific reports*, 8(1):6085, 2018.
- [33] Ricky TQ Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. *Advances in neural information processing systems*, 31, 2018.
- [34] Yulia Rubanova, Ricky TQ Chen, and David K Duvenaud. Latent ordinary differential equations for irregularly-sampled time series. *Advances in neural information processing systems*, 32, 2019.
- [35] Xuechen Li, Ting-Kam Leonard Wong, Ricky TQ Chen, and David K Duvenaud. Scalable gradients and variational inference for stochastic differential equations. In *Symposium on Advances in Approximate Bayesian Inference*, pages 1–28. PMLR, 2020.

- [36] Junteng Jia and Austin R Benson. Neural jump stochastic differential equations. *Advances in Neural Information Processing Systems*, 32, 2019.
- [37] Pashupati Hegde, Markus Heinonen, Harri Lähdesmäki, and Samuel Kaski. Deep learning with differential gaussian process flows. *arXiv preprint arXiv:1810.04066*, 2018.
- [38] Calypso Herrera, Florian Krach, and Josef Teichmann. Neural jump ordinary differential equations: Consistent continuous-time prediction and filtering. *arXiv preprint arXiv:2006.04727*, 2020.
- [39] John Denker and Yann LeCun. Transforming neural-net output levels to probability distributions. *Advances in neural information processing systems*, 3, 1990.
- [40] Charles Blundell, Julien Cornebise, Koray Kavukcuoglu, and Daan Wierstra. Weight uncertainty in neural network. In *International conference on machine learning*, pages 1613–1622. PMLR, 2015.
- [41] Hippolyt Ritter, Aleksandar Botev, and David Barber. A scalable laplace approximation for neural networks. In *6th International Conference on Learning Representations, ICLR 2018-Conference Track Proceedings*, volume 6. International Conference on Representation Learning, 2018.
- [42] Sai Munikoti, Deepesh Agarwal, Laya Das, and Balasubramaniam Natarajan. A general framework for quantifying aleatoric and epistemic uncertainty in graph neural networks. *Neurocomputing*, 521:1–10, 2023.
- [43] Chunyuan Li, Changyou Chen, David Carlson, and Lawrence Carin. Preconditioned stochastic gradient langevin dynamics for deep neural networks. In *Thirtieth AAAI Conference on Artificial Intelligence*, 2016.
- [44] Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *international conference on machine learning*, pages 1050–1059. PMLR, 2016.

- [45] Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. *Advances in neural information processing systems*, 30, 2017.
- [46] Ling kai Kong, Jimeng Sun, and Chao Zhang. Sde-net: Equipping deep neural networks with uncertainty estimates. *arXiv preprint arXiv:2008.10546*, 2020.
- [47] Kenneth J Caird. Meter phase identification, March 27 2012. US Patent 8,143,879.
- [48] Shankar V Achanta. Automatic phase identification for electric power delivery lines, September 15 2020. US Patent 10,775,448.
- [49] Vijay Arya, D Seetharam, Shivkumar Kalyanaraman, Kejitan Dontas, C Pavlovski, Steve Hoy, and Jayant R Kalagnanam. Phase identification in smart grids. In *2011 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, pages 25–30. IEEE, 2011.
- [50] Satya Jayadev Pappu, Nirav Bhatt, Ramkrishna Pasumarthi, and Aravind Rajeswaran. Identifying topology of low voltage distribution networks based on smart meter data. *IEEE Transactions on Smart Grid*, 9(5):5113–5122, 2017.
- [51] Murat Dilek, Robert P Broadwater, and Richard Sequin. Phase prediction in distribution systems. In *2002 IEEE Power Engineering Society Winter Meeting. Conference Proceedings (Cat. No. 02CH37309)*, volume 2, pages 985–990. IEEE, 2002.
- [52] Minghao Xu, Ran Li, and Furong Li. Phase identification with incomplete data. *IEEE Transactions on Smart Grid*, 9(4):2777–2785, 2016.
- [53] Miles HF Wen, Reza Arghandeh, Alexandra von Meier, Kameshwar Poolla, and Victor OK Li. Phase identification in distribution networks with micro-synchrophasors. In *2015 IEEE Power & Energy Society General Meeting*, pages 1–5. IEEE, 2015.
- [54] Brandon Foggo and Nanpeng Yu. A comprehensive evaluation of supervised machine

- learning for the phase identification problem. *International Journal of Computer and Systems Engineering*, 12(6):419–427, 2018.
- [55] Houman Pezeshki and Peter J Wolfs. Consumer phase identification in a three phase unbalanced lv distribution network. In *2012 3rd IEEE PES Innovative Smart Grid Technologies Europe (ISGT Europe)*, pages 1–7. IEEE, 2012.
 - [56] Frédéric Olivier, Antonio Sutera, Pierre Geurts, Raphaël Fonteneau, and Damien Ernst. Phase identification of smart meters by clustering voltage measurements. In *2018 Power Systems Computation Conference (PSCC)*, pages 1–8. IEEE, 2018.
 - [57] Logan Blakely, Matthew J Reno, and Wu-chi Feng. Spectral clustering for customer phase identification using ami voltage timeseries. In *2019 IEEE Power and Energy Conference at Illinois (PECI)*, pages 1–7. IEEE, 2019.
 - [58] Logan Blakely and Matthew J Reno. Phase identification using co-association matrix ensemble clustering. *IET Smart Grid*, 3(4):490–499, 2020.
 - [59] Hooman Firoozi and Habib Rajabi Mashhadi. Non-technical loss detection in limited-data low-voltage distribution feeders. *International Journal of Electrical Power & Energy Systems*, 135:107523, 2022.
 - [60] Marco ARS Cruz, Helder RO Rocha, Marcia HM Paiva, Marcelo EV Segatto, Eglantine Camby, and Gilles Caporossi. An algorithm for cost optimization of pmu and communication infrastructure in wams. *International Journal of Electrical Power & Energy Systems*, 106:96–104, 2019.
 - [61] Saeed Lotfifard. Sparse sensing platform for line-outage identification in multiarea power systems. *IEEE Transactions on Industrial Informatics*, 13(3):947–955, 2017.
 - [62] Rohit Babu and Biplab Bhattacharyya. Optimal allocation of phasor measurement unit for full observability of the connected power network. *International Journal of Electrical Power & Energy Systems*, 79:89–97, 2016.

- [63] Gagandeep Singh Dua, Barjeev Tyagi, and Vishal Kumar. Milp based deployment of micro-pmu in reconfigurable active distribution network. In *2019 North American Power Symposium (NAPS)*, pages 1–6, 2019. doi: 10.1109/NAPS46351.2019.9000347.
- [64] Junqi Liu, Ferdinanda Ponci, Antonello Monti, Carlo Muscas, Paolo Attilio Pegoraro, and Sara Sulis. Optimal meter placement for robust measurement systems in active distribution grids. *IEEE Transactions on Instrumentation and Measurement*, 63(5): 1096–1105, 2014.
- [65] Antonio A.M. Raposo, Anselmo B. Rodrigues, and Maria da Guia da Silva. Robust meter placement for state estimation considering distribution network reconfiguration for annual energy loss reduction. *Electric Power Systems Research*, 182:106233, 2020. ISSN 0378-7796. doi: <https://doi.org/10.1016/j.epsr.2020.106233>. URL <https://www.sciencedirect.com/science/article/pii/S0378779620300407>.
- [66] Kapil Chauhan and Ranjana Sodhi. Placement of distribution-level phasor measurements for topological observability and monitoring of active distribution networks. *IEEE Transactions on Instrumentation and Measurement*, 69(6):3451–3460, 2019.
- [67] Antonio AM Raposo, Anselmo B Rodrigues, and Maria da Guia da Silva. Optimal meter placement algorithm for state estimation in power distribution networks. *Electric power systems research*, 147:22–30, 2017.
- [68] Mohammad Ghasemi Damavandi, Vikram Krishnamurthy, and José R. Martí. Robust meter placement for state estimation in active distribution systems. *IEEE Transactions on Smart Grid*, 6(4):1972–1982, 2015. doi: 10.1109/TSG.2015.2394361.
- [69] Rahul Madbhavi, Balasubramaniam Natarajan, and Babji Srinivasan. Meter placement approaches for matrix completion-based distribution system state estimator. *Electric Power Systems Research*, 213:108687, 2022.
- [70] RC Leou and CN Lu. Improving feeder voltage calculation results with telemetered data. *IEEE transactions on power delivery*, 11(4):1914–1920, 1996.

- [71] Yu Xiang, Paulo F Ribeiro, and Joseph FG Cobben. Optimization of state-estimator-based operation framework including measurement placement for medium voltage distribution grid. *IEEE Transactions on Smart Grid*, 5(6):2929–2937, 2014.
- [72] Themistoklis C Xygkis, George N Korres, and Nikolaos M Manousakis. Fisher information-based meter placement in distribution grids via the d-optimal experimental design. *IEEE Transactions on Smart Grid*, 9(2):1452–1461, 2016.
- [73] Zhi Wu, Xiao Du, Wei Gu, Yafei Liu, Ping Ling, Jinsong Liu, and Chen Fang. Optimal pmu placement considering load loss and relaying in distribution networks. *IEEE Access*, 6:33645–33653, 2018.
- [74] Abdul Aziz G Mabaning, Jordan Rel C Orillaza, and Alexandra von Meier. Optimal pmu placement for distribution networks. In *2017 IEEE Innovative Smart Grid Technologies-Asia (ISGT-Asia)*, pages 1–6. IEEE, 2017.
- [75] Yue Zhao, Raffi Sevlian, Ram Rajagopal, Andrea Goldsmith, and H Vincent Poor. Outage detection in power distribution networks with optimally-deployed power flow sensors. In *2013 IEEE Power & Energy Society General Meeting*, pages 1–5. IEEE, 2013.
- [76] Raffi Avo Sevlian, Yue Zhao, Ram Rajagopal, Andrea Goldsmith, and H Vincent Poor. Outage detection using load and line flow measurements in power distribution systems. *IEEE Transactions on Power Systems*, 33(2):2053–2069, 2017.
- [77] Ananth Narayan Samudrala, M Hadi Amini, Soumya Kar, and Rick S Blum. Sensor placement for outage identifiability in power distribution networks. *IEEE Transactions on Smart Grid*, 11(3):1996–2013, 2019.
- [78] Hany A Abdelsalam, Almoataz Y Abdelaziz, and V Mukherjee. Optimal pmu placement in a distribution network considering network reconfiguration. In *2014 International Conference on Circuits, Power and Computing Technologies [ICCPCT-2014]*, pages 191–196. IEEE, 2014.

- [79] Van-khoi Tran, He-sheng Zhang, and Van-nghia Nguyen. Optimal pmu placement in multi-configuration power distribution networks. In *Proceedings of the 2016 4th International Conference on Electrical & Electronics Engineering and Computer Science (ICEEECS 2016)*, pages 508–514, 2016.
- [80] Mohamed A. Mohamed, Ameena Saad Al-Sumaiti, Mohamed Krid, Emad Mahrous Awwad, and Abdollah Kavousi-Fard. A reliability-oriented fuzzy stochastic framework in automated distribution grids to allocate μ -pmus. *IEEE Access*, 7:33393–33404, 2019. doi: 10.1109/ACCESS.2019.2902465.
- [81] Andrey Bernstein and Emiliano Dall’Anese. Linear power-flow models in multiphase distribution networks. In *2017 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*, pages 1–6. IEEE, 2017.
- [82] Emmanuel J Candès and Michael B Wakin. An introduction to compressive sampling. *IEEE signal processing magazine*, 25(2):21–30, 2008.
- [83] Ji Liu, Przemyslaw Musialski, Peter Wonka, and Jieping Ye. Tensor completion for estimating missing values in visual data. *IEEE transactions on pattern analysis and machine intelligence*, 35(1):208–220, 2012.
- [84] Vassilis Kekatos and Georgios B Giannakis. Distributed robust power system state estimation. *IEEE Transactions on Power Systems*, 28(2):1617–1626, 2012.
- [85] Yuzhang Lin and Ali Abur. Robust state estimation against measurement and network parameter errors. *IEEE Transactions on power systems*, 33(5):4751–4759, 2018.
- [86] Biyun Chen, Hongbin Li, and Bin Zhou. Real-time identification of false data injection attacks: A novel dynamic-static parallel state estimation based mechanism. *IEEE Access*, 7:95812–95824, 2019.
- [87] Mohammad Irshaad Oozeer and Simon Haykin. Cognitive risk control for mitigating cyber-attack in smart grid. *IEEE Access*, 7:125806–125826, 2019.

- [88] Dou An, Qingyu Yang, Wenmao Liu, and Yang Zhang. Defending against data integrity attacks in smart grid: A deep reinforcement learning-based approach. *IEEE Access*, 7:110835–110845, 2019.
- [89] Murilo EC Bento, Daniel Dotta, Roman Kuiava, and Rodrigo A Ramos. A procedure to design fault-tolerant wide-area damping controllers. *IEEE Access*, 6:23383–23405, 2018.
- [90] Dongliang Duan, Liuqing Yang, and Louis L Scharf. Phasor state estimation from pmu measurements with bad data. In *2011 4th IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, pages 121–124. IEEE, 2011.
- [91] Weiyu Xu, Meng Wang, Jian-Feng Cai, and Ao Tang. Sparse error correction from nonlinear measurements with applications in bad data detection for power networks. *IEEE Transactions on Signal Processing*, 61(24):6175–6187, 2013.
- [92] Zihan Zhou, Xiaodong Li, John Wright, Emmanuel Candes, and Yi Ma. Stable principal component pursuit. In *2010 IEEE international symposium on information theory*, pages 1518–1522. IEEE, 2010.
- [93] Donald Goldfarb and Zhiwei Qin. Robust low-rank tensor recovery: Models and algorithms. *SIAM Journal on Matrix Analysis and Applications*, 35(1):225–253, 2014.
- [94] Stephen Boyd, Stephen P Boyd, and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.
- [95] KP Schneider, BA Mather, BC Pal, C-W Ten, GJ Shirek, Hao Zhu, JC Fuller, José Luiz Rezende Pereira, Luis Fernando Ochoa, Leandro Ramos de Araujo, et al. Analytic considerations and design basis for the ieee distribution test feeders. *IEEE Transactions on power systems*, 33(3):3181–3188, 2017.
- [96] Michael Grant and Stephen Boyd. Cvx: Matlab software for disciplined convex programming, version 2.1, 2014.

- [97] Richard G Baraniuk. Compressive sensing [lecture notes]. *IEEE signal processing magazine*, 24(4):118–121, 2007.
- [98] Emmanuel J Candès and Benjamin Recht. Exact matrix completion via convex optimization. *Foundations of Computational mathematics*, 9(6):717, 2009.
- [99] Yajing Liu, April Sagan, Andrey Bernstein, Rui Yang, Xinyang Zhou, and Yingchen Zhang. Matrix completion using alternating minimization for distribution system state estimation. In *2020 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*, pages 1–6. IEEE, 2020.
- [100] Anil K Jain. *Fundamentals of digital image processing*. Prentice-Hall, Inc., 1989.
- [101] S. Dahale, H. S. Karimi, K. Lai, and B. Natarajan. Sparsity based approaches for distribution grid state estimation - a comparative study. *IEEE Access*, 8:198317–198327, 2020. doi: 10.1109/ACCESS.2020.3035378.
- [102] New York State Energy Research and Development Authority (NYSERDA). Fundamental research challenges for distribution state estimation to enable high-performing grids. *NYSERDA Report Number 18-37. Prepared by Smarter Grid Solutions, New York, NY.*, 2018.
- [103] Carl Edward Rasmussen. Gaussian processes in machine learning. In *Summer School on Machine Learning*, pages 63–71. Springer, 2003.
- [104] Vincent Fortuin, Heiko Strathmann, and Gunnar Rätsch. Meta-learning mean functions for gaussian processes. *arXiv preprint arXiv:1901.08098*, 2019.
- [105] Ingyo Chung, Saehoon Kim, Juho Lee, Kwang Joon Kim, Sung Ju Hwang, and Eunho Yang. Deep mixed effect model using gaussian processes: A personalized and reliable prediction for healthcare. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 3649–3657, 2020.

- [106] April Sagan, Yajing Liu, and Andrey Bernstein. Decentralized low-rank state estimation for power distribution systems. *IEEE Transactions on Smart Grid*, 2021.
- [107] S Derin Babacan, Martin Luessi, Rafael Molina, and Aggelos K Katsaggelos. Sparse bayesian methods for low-rank matrix estimation. *IEEE Transactions on Signal Processing*, 60(8):3964–3977, 2012.
- [108] Christopher M Bishop. *Pattern recognition and machine learning*. springer, 2006.
- [109] Thierry Bouwmans and El Hadi Zahzah. Robust pca via principal component pursuit: A review for a comparative evaluation in video surveillance. *Computer Vision and Image Understanding*, 122:22–34, 2014.
- [110] Ahmad Reza Malekpour and Anil Pahwa. Radial test feeder including primary and secondary distribution network. In *2015 North American Power Symposium (NAPS)*, pages 1–9. IEEE, 2015.
- [111] Cristina Carmona-Delgado, Esther Romero-Ramos, and Jesús Riquelme-Santos. Fast and reliable distribution load and state estimator. *Electric power systems research*, 101:110–124, 2013.
- [112] Ieee standard for synchrophasor measurements for power systems. *IEEE Std C37.118.1-2011 (Revision of IEEE Std C37.118-2005)*, pages 1–61, 2011. doi: 10.1109/IEEESTD.2011.6111219.
- [113] Arturo S Bretas, Aquiles Rossoni, Rodrigo D Trevizan, and Newton G Bretas. Distribution networks nontechnical power loss estimation: A hybrid data-driven physics model-based framework. *Electric Power Systems Research*, 186:106397, 2020.
- [114] Michael Grant and Stephen Boyd. CVX: Matlab software for disciplined convex programming, version 2.1. <http://cvxr.com/cvx>, March 2014.
- [115] Arun Venkitaraman, Saikat Chatterjee, and Peter Handel. Gaussian processes over

- graphs. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5640–5644. IEEE, 2020.
- [116] David I Shuman, Sunil K Narang, Pascal Frossard, Antonio Ortega, and Pierre Vandergheynst. The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. *IEEE signal processing magazine*, 30(3):83–98, 2013.
 - [117] Marco F Huber. Recursive gaussian process: On-line regression and learning. *Pattern Recognition Letters*, 45:85–91, 2014.
 - [118] Mark Newman. *Networks*. Oxford university press, 2018.
 - [119] Ali Al-Wakeel, Jianzhong Wu, and Nick Jenkins. State estimation of medium voltage distribution networks using smart meter measurements. *Applied energy*, 184:207–218, 2016.
 - [120] Hadis Karimipour and Venkata Dinavahi. Extended kalman filter-based parallel dynamic state estimation. *IEEE transactions on smart grid*, 6(3):1539–1549, 2015.
 - [121] Edward Snelson, Zoubin Ghahramani, and Carl Rasmussen. Warped gaussian processes. *Advances in neural information processing systems*, 16, 2003.
 - [122] Xinyang Zhou, Zhiyuan Liu, Changhong Zhao, and Lijun Chen. Accelerated voltage regulation in multi-phase distribution networks based on hierarchical distributed algorithm. *IEEE Transactions on Power Systems*, 35(3):2047–2058, 2020. doi: 10.1109/TPWRS.2019.2948978.
 - [123] Arash Amini, Guangyi Liu, and Nader Motee. Robust learning of recurrent neural networks in presence of exogenous noise. In *2021 60th IEEE Conference on Decision and Control (CDC)*, pages 783–788. IEEE, 2021.
 - [124] Simo Särkkä and Arno Solin. *Applied stochastic differential equations*, volume 10. Cambridge University Press, 2019.

- [125] Kyunghyun Cho, Bart Van Merriënboer, Dzmitry Bahdanau, and Yoshua Bengio. On the properties of neural machine translation: Encoder-decoder approaches. *arXiv preprint arXiv:1409.1259*, 2014.
- [126] Shweta Dahale and Balasubramaniam Natarajan. Recursive gaussian process over graphs for integrating multi-timescale measurements in low-observable distribution systems. *IEEE Transactions on Power Systems*, pages 1–12, 2022. doi: 10.1109/TPWRS.2022.3204415.
- [127] Dan Levi, Liran Gispán, Niv Giladi, and Ethan Fetaya. Evaluating and calibrating uncertainty prediction in regression tasks. *Sensors*, 22(15):5540, 2022.
- [128] Chuan Guo, Geoff Pleiss, Yu Sun, and Kilian Q Weinberger. On calibration of modern neural networks. In *International conference on machine learning*, pages 1321–1330. PMLR, 2017.
- [129] Tom A Short. Advanced metering for phase identification, transformer identification, and secondary modeling. *IEEE Transactions on Smart Grid*, 4(2):651–658, 2012.
- [130] Jose M Cano, Pablo Arboleya, Mahmoud Rashad Ahmed, Md Rejwanur R Mojumdar, and Gonzalo A Orcajo. Improving distribution system state estimation with synthetic measurements. *International Journal of Electrical Power & Energy Systems*, 129:106751, 2021.
- [131] Yizheng Liao, Yang Weng, Guangyi Liu, Zhongyang Zhao, Chin-Woo Tan, and Ram Rajagopal. Unbalanced multi-phase distribution grid topology estimation and bus phase identification. *IET Smart Grid*, 2(4):557–570, 2019.
- [132] Xiaowen Dong, Dorina Thanou, Pascal Frossard, and Pierre Vandergheynst. Learning laplacian matrix in smooth graph signal representations. *IEEE Transactions on Signal Processing*, 64(23):6160–6173, 2016.
- [133] Madeleine Udell. Introduction to spectral graph theory, 2011.

- [134] Ulrike Von Luxburg. A tutorial on spectral clustering. *Statistics and computing*, 17(4):395–416, 2007.
- [135] Feiping Nie, Xiaoqian Wang, Michael Jordan, and Heng Huang. The constrained laplacian rank algorithm for graph-based clustering. In *Proceedings of the AAAI conference on artificial intelligence*, volume 30, 2016.
- [136] Hazhar Sufi Karimi and Balasubramaniam Natarajan. Joint topology identification and state estimation in unobservable distribution grids. *IEEE Transactions on Smart Grid*, 2021.
- [137] Urs Niesen, Devavrat Shah, and Gregory Wornell. Adaptive alternating minimization algorithms. In *2007 IEEE International Symposium on Information Theory*, pages 1641–1645. IEEE, 2007.
- [138] Rahul Madbhavi, Hazhar Sufi Karimi, Balasubramaniam Natarajan, and Babji Srinivasan. Tensor completion based state estimation in distribution systems. In *2020 IEEE Power Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, pages 1–5, 2020. doi: 10.1109/ISGT45199.2020.9087747.
- [139] Shweta Dahale and Bala Natarajan. Bayesian framework for multi-timescale state estimation in low-observable distribution systems. *IEEE Transactions on Power Systems*, 2022.
- [140] Artis Riepnieks and Harold Kirkham. An introduction to goodness of fit for pmu parameter estimation. *IEEE Transactions on Power Delivery*, 32(5):2238–2245, 2016.
- [141] Sentient Energy, Inc mm3 overhead line sensor. <https://www.sentientenergy.com/products/mm3-line-sensor/>.
- [142] V.V. Vazirani. *Approximation Algorithms*. Springer Berlin Heidelberg, 2013. ISBN 9783662045657. URL <https://books.google.com/books?id=bJmqCAAQBAJ>.

- [143] Gagandeep Singh Dua, Barjeev Tyagi, and Vishal Kumar. Deploying micro-pmus with channel limit in reconfigurable distribution systems. *IEEE Systems Journal*, 16(1):832–843, 2021.
- [144] Theodore W Anderson. An introduction to multivariate statistical analysis. Technical report, 1958.
- [145] Ke-yan Liu, Wanxing Sheng, Yongmei Liu, and Xiaoli Meng. A network reconfiguration method considering data uncertainties in smart distribution networks. *Energies*, 10(5):618, 2017.
- [146] Dong Zhang, Zhengcai Fu, and Liuchun Zhang. An improved ts algorithm for loss-minimum reconfiguration in large-scale distribution systems. *Electric power systems research*, 77(5-6):685–694, 2007.

Appendix A

Reuse permissions from publishers



Home ? Live Chat Sign in Create Account



Bayesian Framework for Multi-Timescale State Estimation in Low-Observable Distribution Systems
Author: Shweta Dahale
Publication: IEEE Transactions on Power Systems
Publisher: IEEE
Date: November 2022
Copyright © 2022, IEEE

Thesis / Dissertation Reuse

The IEEE does not require individuals working on a thesis to obtain a formal reuse license, however, you may print out this statement to be used as a permission grant:

Requirements to be followed when using any portion (e.g., figure, graph, table, or textual material) of an IEEE copyrighted paper in a thesis:

- 1) In the case of textual material (e.g., using short quotes or referring to the work within these papers) users must give full credit to the original source (author, paper, publication) followed by the IEEE copyright line © 2011 IEEE.
- 2) In the case of illustrations or tabular material, we require that the copyright line © [Year of original publication] IEEE appear prominently with each reprinted figure and/or table.
- 3) If a substantial portion of the original paper is to be used, and if you are not the senior author, also obtain the senior author's approval.

Requirements to be followed when using an entire IEEE copyrighted paper in a thesis:

- 1) The following IEEE copyright/ credit notice should be placed prominently in the references: © [year of original publication] IEEE. Reprinted, with permission, from [author names, paper title, IEEE publication title, and month/year of publication].
- 2) Only the accepted version of an IEEE copyrighted paper can be used when posting the paper or your thesis on-line.
- 3) In placing the thesis on the author's university website, please display the following message in a prominent place on the website: In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of [university/educational entity's name goes here]'s products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink.

If applicable, University Microfilms and/or ProQuest Library, or the Archives of Canada may supply single copies of the dissertation.

BACK CLOSE WINDOW



Recursive Gaussian Process over graphs for Integrating Multi-timescale Measurements in Low-Observable Distribution Systems

Author: Shweta Dahale
Publication: IEEE Transactions on Power Systems
Publisher: IEEE
Date: 2022

Copyright © 2022, IEEE

Thesis / Dissertation Reuse

The IEEE does not require individuals working on a thesis to obtain a formal reuse license, however, you may print out this statement to be used as a permission grant:

Requirements to be followed when using any portion (e.g., figure, graph, table, or textual material) of an IEEE copyrighted paper in a thesis:

- 1) In the case of textual material (e.g., using short quotes or referring to the work within these papers) users must give full credit to the original source (author, paper, publication) followed by the IEEE copyright line © 2011 IEEE.
- 2) In the case of illustrations or tabular material, we require that the copyright line © [Year of original publication] IEEE appear prominently with each reprinted figure and/or table.
- 3) If a substantial portion of the original paper is to be used, and if you are not the senior author, also obtain the senior author's approval.

Requirements to be followed when using an entire IEEE copyrighted paper in a thesis:

- 1) The following IEEE copyright/ credit notice should be placed prominently in the references: © [year of original publication] IEEE. Reprinted, with permission, from [author names, paper title, IEEE publication title, and month/year of publication]
- 2) Only the accepted version of an IEEE copyrighted paper can be used when posting the paper or your thesis on-line.
- 3) In placing the thesis on the author's university website, please display the following message in a prominent place on the website: In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of [university/educational entity's name goes here]'s products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink.

If applicable, University Microfilms and/or ProQuest Library, or the Archives of Canada may supply single copies of the dissertation.

BACK

CLOSE WINDOW



Multi Time-scale Imputation aided State Estimation in Distribution System

Conference Proceedings: 2021 IEEE Power & Energy Society General Meeting (PESGM)
Author: Shweta Dahale
Publisher: IEEE
Date: 26 July 2021

Copyright © 2021, IEEE

Thesis / Dissertation Reuse

The IEEE does not require individuals working on a thesis to obtain a formal reuse license, however, you may print out this statement to be used as a permission grant:

Requirements to be followed when using any portion (e.g., figure, graph, table, or textual material) of an IEEE copyrighted paper in a thesis:

- 1) In the case of textual material (e.g., using short quotes or referring to the work within these papers) users must give full credit to the original source (author, paper, publication) followed by the IEEE copyright line © 2011 IEEE.
- 2) In the case of illustrations or tabular material, we require that the copyright line © [Year of original publication] IEEE appear prominently with each reprinted figure and/or table.
- 3) If a substantial portion of the original paper is to be used, and if you are not the senior author, also obtain the senior author's approval.

Requirements to be followed when using an entire IEEE copyrighted paper in a thesis:

- 1) The following IEEE copyright/ credit notice should be placed prominently in the references: © [year of original publication] IEEE. Reprinted, with permission, from [author names, paper title, IEEE publication title, and month/year of publication]
- 2) Only the accepted version of an IEEE copyrighted paper can be used when posting the paper or your thesis on-line.
- 3) In placing the thesis on the author's university website, please display the following message in a prominent place on the website: In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of [university/educational entity's name goes here]'s products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink.

If applicable, University Microfilms and/or ProQuest Library, or the Archives of Canada may supply single copies of the dissertation.

BACK

CLOSE WINDOW



Latent Neural ODE for Integrating Multi-timescale measurements in Smart Distribution Grids

Conference Proceedings: 2023 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)

Author: Shweta Dahale

Publisher: IEEE

Date: 16 January 2023

Copyright © 2023, IEEE

Thesis / Dissertation Reuse

The IEEE does not require individuals working on a thesis to obtain a formal reuse license, however, you may print out this statement to be used as a permission grant:

Requirements to be followed when using any portion (e.g., figure, graph, table, or textual material) of an IEEE copyrighted paper in a thesis:

- 1) In the case of textual material (e.g., using short quotes or referring to the work within these papers) users must give full credit to the original source (author, paper, publication) followed by the IEEE copyright line © 2011 IEEE.
- 2) In the case of illustrations or tabular material, we require that the copyright line © [Year of original publication] IEEE appear prominently with each reprinted figure and/or table.
- 3) If a substantial portion of the original paper is to be used, and if you are not the senior author, also obtain the senior author's approval.

Requirements to be followed when using an entire IEEE copyrighted paper in a thesis:

- 1) The following IEEE copyright/ credit notice should be placed prominently in the references: © [year of original publication] IEEE. Reprinted, with permission, from [author names, paper title, IEEE publication title, and month/year of publication]
- 2) Only the accepted version of an IEEE copyrighted paper can be used when posting the paper or your thesis on-line.
- 3) In placing the thesis on the author's university website, please display the following message in a prominent place on the website: In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of [university/educational entity's name goes here]'s products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink.

If applicable, University Microfilms and/or ProQuest Library, or the Archives of Canada may supply single copies of the dissertation.

BACK

CLOSE WINDOW



Joint Matrix Completion and Compressed Sensing for State Estimation in Low-observable Distribution System

Conference Proceedings: 2021 IEEE PES Innovative Smart Grid Technologies Conference - Latin America (ISGT Latin America)

Author: Shweta Dahale

Publisher: IEEE

Date: 15 September 2021

Copyright © 2021, IEEE

Thesis / Dissertation Reuse

The IEEE does not require individuals working on a thesis to obtain a formal reuse license, however, you may print out this statement to be used as a permission grant:

Requirements to be followed when using any portion (e.g., figure, graph, table, or textual material) of an IEEE copyrighted paper in a thesis:

- 1) In the case of textual material (e.g., using short quotes or referring to the work within these papers) users must give full credit to the original source (author, paper, publication) followed by the IEEE copyright line © 2011 IEEE.
- 2) In the case of illustrations or tabular material, we require that the copyright line © [Year of original publication] IEEE appear prominently with each reprinted figure and/or table.
- 3) If a substantial portion of the original paper is to be used, and if you are not the senior author, also obtain the senior author's approval.

Requirements to be followed when using an entire IEEE copyrighted paper in a thesis:

- 1) The following IEEE copyright/ credit notice should be placed prominently in the references: © [year of original publication] IEEE. Reprinted, with permission, from [author names, paper title, IEEE publication title, and month/year of publication]
- 2) Only the accepted version of an IEEE copyrighted paper can be used when posting the paper or your thesis on-line.
- 3) In placing the thesis on the author's university website, please display the following message in a prominent place on the website: In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of [university/educational entity's name goes here]'s products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink.

If applicable, University Microfilms and/or ProQuest Library, or the Archives of Canada may supply single copies of the dissertation.

BACK

CLOSE WINDOW