



Smart Meter-Driven Distribution Grid Visibility and Control

Final Project Report

T-67

Power Systems Engineering Research Center

*Empowering Minds to Engineer
the Future Electric Energy System*



Smart Meter-Driven Distribution Grid Visibility and Control

Final Project Report

Project Team

Daniel Molzahn, Project Leader
Georgia Institute of Technology

Line Roald
University of Wisconsin–Madison

Daniel Fuhrmann
Michigan Technological University

Graduate Students

Samuel Talkington
Richard Asiamah
Georgia Institute of Technology

Jiaqi Chen
Aditya Rangarajan
University of Wisconsin-Madison

Pedro Aquino de Alcantara
Michigan Technological University

Rahul Gupta, Postdoctoral
Washington State University

PSERC Publication 25-07
December 2025

For information about this project, contact:

Daniel Molzahn
Georgia Institute of Technology
School of Electrical and Computer Engineering
E176 Van Leer Building
777 Atlantic Drive NW
Atlanta, GA 30313, USA
Phone: (404) 894-1876
Email: molzahn@gatech.edu

Power Systems Engineering Research Center

The Power Systems Engineering Research Center (PSERC) is a multi-university Center conducting research on challenges facing the electric power industry and educating the next generation of power engineers. More information about PSERC can be found at the Center's website: <http://www.pserc.org>.

For additional information, contact:

Power Systems Engineering Research Center
Arizona State University
527 Engineering Research Center
Tempe, Arizona 85287-5706
Phone: 480-965-1643
Fax: 480-727-2052

Notice Concerning Copyright Material

PSERC members are given permission to copy without fee all or part of this publication for internal use if appropriate attribution is given to this document as the source material. This report is available for downloading from the PSERC website.

© 2025 Georgia Institute of Technology. All rights reserved.

Acknowledgments

This is the final report for the Power Systems Engineering Research Center (PSERC) research project T-67, Smart Meter-Driven Distribution Grid Visibility and Control.

The research team members wish to express our appreciation for the support provided by PSERC's industry advisors who have followed the project and provided valuable feedback throughout:

- David Rieken (Hubbell, Inc.)
- Kurt LaFrance (Consumers Energy)
- Venkat Banunarayanan (NRECA)
- Bo Gong (Salt River Project)
- Dongbo Zhao, Shakil Hossan, Arun Sukumaran Nair (Eaton)
- Benjamin Kroposki (NREL)
- Jouni Peppanen (EPRI)
- John J. O'Donnell, Christopher Grzybowski (DTE Energy)
- Mohit Singh (ComEd)

Executive Summary

Project Overview

This report presents the outcomes of PSERC Project T-67, Smart Meter-Driven Distribution Grid Visibility and Control. The project addresses an emerging challenge in modern power systems: improving real-time visibility and control in distribution grids with increasing penetration of distributed energy resources (DERs). The project develops new computational methods for achieving visibility and control while relying on existing smart meter infrastructure constrained by communication bandwidth.

While smart meters are ubiquitous and capable of recording accurate and frequent measurements of voltage, current, and power, their potential for real-time monitoring and control has been under-utilized. The main barrier is communication bandwidth: utilities can only retrieve a limited number of high-frequency data streams, far fewer than the total number of meters deployed. We consider a setting where the smart meter communication capabilities can be dynamically allocated such that the bandwidth is limited in aggregate across a fleet of smart meters. As Fig. 1 illustrates conceptually, this is analogous to shining flashlights around a dark room: more data from certain smart meters can be obtained at the cost of transmitting less data from other meters. Accordingly, this project develops new theoretical frameworks, optimization methods, and machine learning techniques to maximize the value of limited smart meter data streams, thereby transforming smart meters from billing devices into critical sensing infrastructure for future distribution grids.

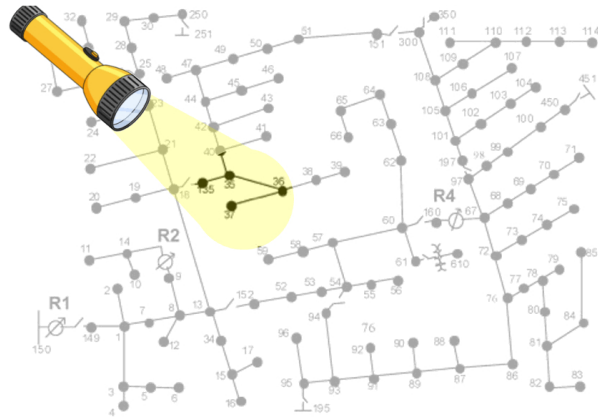


Figure 1: Conceptual illustration of dynamically using limited-bandwidth communication across a fleet of smart meters.

In this context, the project focuses on five particular topics: (1) online strategies for identifying worst-case violations of voltage magnitude limits in distribution networks, (2) machine learning methods for classifying the control laws of behind-the-meter solar inverters, (3) forecast-aided state estimation, (4) online identification of distribution network topologies using quantized measurements, and (5) information-theoretic approaches for managing bandwidth-limited smart meter data.

Online Strategies for Identifying Worst-Case Voltage Magnitude Violations

We developed two approaches for online sensor sampling to identify worst-case violations of voltage magnitude limits under bandwidth constraints. The first method builds on a stochastic lin-

earized power flow (LinDistFlow) model and casts the problem as a *spectral multi-armed bandit*. The algorithm iteratively selects subsets of smart meters in real time, guided by the graphical structure of distribution networks and sub-Gaussian bounds on voltage fluctuations. The second method is an iterative optimization algorithm that, at each step, identifies which buses may have the most significant voltage excursions given the measurements taken so far and known bounds on the power injections at each bus. The method updates the set of measurements to include voltages at unmeasured buses with the largest range of possible voltage magnitudes.

Simulations on benchmark feeders demonstrated that both measurement selection methods consistently outperform baseline algorithms, such as conventional linear bandit sampling. This work shows that limited communication capacity need not prevent effective monitoring of potential voltage limit violations. By exploiting the inherent network structure, utilities can adaptively reallocate bandwidth to measure critical buses, improving situational awareness without additional infrastructure investments.

Machine Learning for Classifying Control Laws of Behind-the-Meter Solar Inverters

Utility companies may not know the reactive power and voltage magnitude control laws of behind-the-meter solar inverters, but these control laws can have significant impacts on the behavior of distribution systems. Using a synthetic dataset of smart meter measurements reflecting different inverter control strategies (e.g., fixed power factor and volt-var controls), this work tested two methods for classifying inverter behavior: a nearest-centroid classifier and a neural network.

Both methods successfully identified inverter control laws, but the neural network substantially outperformed the simpler centroid approach, achieving higher accuracy and greater robustness to measurement noise. We also evaluated how measurement errors and data availability affect classification performance, identifying trade-offs between data quality and classifier reliability. The ability to infer control behaviors from smart meter data provides a low-cost means of improving grid models, forecasting, and operational decision-making.

Forecast-Aided State Estimation

State estimation is a key application for real-time visibility of power systems. Smart meter measurements can contribute to improving the accuracy of distribution system state estimation provided bandwidth limitations can be appropriately managed. Accordingly, we developed an algorithm for distribution system state estimation under bandwidth limitations using a hybrid approach that incorporates forecasting models. At unmeasured buses, a vector autoregressive (VAR) model with recursive least squares updates was used to forecast nodal power injections, providing priors for a state estimator. Real-time smart meter data were then incorporated to refine the estimates under bandwidth-limited conditions.

Case studies on unbalanced distribution feeders showed that forecast-aided distribution system state estimation achieved high accuracy even when only 10–20% of buses provided high-frequency measurements. The approach also adapted well under conditions with photovoltaic integration and vari-

able load demands. These results demonstrate that forecasting can compensate for measurement scarcity, enabling more accurate state estimation without deploying additional sensors beyond existing bandwidth-limited smart meters.

Online Identification of Distribution Network Topologies Using Quantized Measurements

In settings where the distribution network characteristics are unknown or rapidly changing (e.g., due to post-storm network restoration), it is valuable to infer the network topology and line parameters directly from measured voltages and power injections. We investigated the problem of learning radial distribution network topology and parameters from smart meter data subject to quantization (i.e., finite-resolution measurements). Building on the stochastic LinDistFlow power flow approximation, our method derives statistical error bounds for topology recovery when measurements are communicated at low precision due to bandwidth limitations.

Theoretical analysis established error bounds on topology estimation accuracy, quantifying how the number of samples and quantization levels affect recovery performance. Numerical experiments on test feeders confirmed that accurate topology learning is achievable even with coarse quantization, provided that sufficient time-series data are collected. By showing that reliable topology inference is possible from bandwidth-limited smart meter data, this research provides a pathway for enhancing grid visibility and data quality in regions where records are incomplete or outdated.

Information-Theoretic Approaches for Managing Bandwidth-Limited Smart Meter Data

Methods for managing limited communication capacity among a large number of sensors may be categorized into two groups: 1) sensor sampling or measurement selection, and 2) bit rate allocation. The online strategies for identifying worst-case voltage magnitude violations described above are a good example of the first approach. In contrast, we have also considered methods which continually acquire data from all the sensors in the network, but which allocate varying bit rates to each of the sensors based on the value of the data in a particular application such as state estimation. Well-known tools from information theory, including rate distortion theory and quantization, play a key role.

In order to determine the bit rate for a continuous random variable subject to quantization, knowledge of the prior probability distribution of that random variable is required. Hence, a probabilistic description of the underlying system from which the measurements are derived comes into play. We use a state variable description of the system, and a Kalman filter which gives the evolution of the state probability distribution functions based on the measurements. Our approach differs from conventional Kalman filtering in that we assume that the measurements are ideal, and that the measurement “error” is the result of the quantization which leads to finite bit rate communications.

Project Publications

- [1] S. Talkington, R. Gupta, R. Asiamah, P. Buason, and D.K. Molzahn, “Strategic Electric Distribution Network Sensing via Spectral Bandits,” *63rd IEEE Conference on Decision and Control (CDC)*, December 16-19, 2024.
- [2] J. Chen and L.A. Roald, “Real-time Assessment of Distribution Grid Security through Adaptive Smart Meter Measurements,” *63rd IEEE Conference on Decision and Control (CDC)*, December 16-19, 2024.
- [3] R. Asiamah, S. Talkington, M. Boateng, M. Vanin, F. Geth, and D.K. Molzahn, “Classifying Reactive Power Control Laws of Behind-the-Meter Solar Photovoltaic Inverters,” *6th Kansas Power and Energy Conference (KPEC)*, April 24-25, 2025.
- [4] A. Rangarajan, R.K. Gupta, D.K. Molzahn, and L.A. Roald, “Forecast-Aided State Estimation in Unbalanced Distribution Networks Using Smart Meter Data under Limited Communication Bandwidth,” *Sustainable Energy, Grids and Networks*, vol. 43, no. 101863, September 2025. Presented at the *12th IREP Bulk Power system Dynamics and Control Symposium*, June 22-27, 2025.
- [5] S. Talkington, A. Rangarajan, P.A. de Alcântara, L.A. Roald, D.K. Molzahn, and D.R. Fuhrmann, “Error Bounds for Radial Network Topology Learning from Quantized Measurements”, submitted, arXiv:2508.05620.
- [6] P. Aquino de Alcantara, F. Costa, and D. Fuhrmann, “One-Dimensional Kalman Filter with Measurement Noise Replaced by Quantization”, *59th Annual Conference on Information Science and Systems (CISS)*, March 2025.

Student Theses

- [1] Samuel Talkington, “Approximate and Differentiable Optimal Power Flow Models,” *PhD Dissertation, Georgia Institute of Technology, School of Electrical and Computer Engineering*, Spring 2026 (expected).
- [2] Richard Asiamah, “Development and Applications of Synthetic Electric Grid Models for Underrepresented Regions,” *PhD Dissertation, Georgia Institute of Technology, School of Electrical and Computer Engineering*, Fall 2026 (expected).
- [3] Jiaqi Chen, “Data-Driven Approaches for Distribution Grid Analysis,” *PhD Dissertation, University of Wisconsin–Madison, Department of Electrical and Computer Engineering*, July 2025.
- [4] Aditya Rangarajan, “Management of Renewable-rich Power Systems with Partial System Information,” *PhD Dissertation, University of Wisconsin–Madison, Department of Electrical and Computer Engineering*, Spring 2026 (expected).
- [5] Pedro Aquino de Alcantara, “Information-Theoretic Methods in Smart Meter Communication,” *PhD Dissertation, Michigan Technological University, Department of Electrical and Computer Engineering*, Spring 2027 (expected).

Table of Contents

1	Introduction	1
1.1	Motivation	1
1.2	Key Contributions	1
1.3	Chapter Summaries	2
1.4	Closing Remarks	3
2	Strategic Electric Distribution Network Sensing via Spectral Bandits	4
2.1	Introduction.....	4
2.1.1	Related work.....	5
2.1.2	Proposed Framework and Contributions	6
2.1.3	Notation	6
2.2	Problem setting	6
2.2.1	Power flow model for distribution systems.....	7
2.2.2	Review of the LinDistFlow approximation	7
2.2.3	Action set and reward formulation	8
2.3	Stochastic LinDistFlow model	9
2.3.1	Concentration of nodal voltage perturbations	10
2.3.2	Condition for positive semi-definite Laplacian	10
2.3.3	Confidence ellipsoid, sampling policy, and regret	11
2.4	Sensor Sampling Algorithm	12
2.4.1	Optimism principle	12
2.4.2	Spectral UCB	12
2.5	Numerical Experiments	13
2.6	Discussion and Conclusion	15
2.7	Proofs of Lemmas	16
2.7.1	Proof of Lemma 1	16
2.7.2	Proof of Lemma 2	16
2.7.3	Proof of Lemma 3 (sub-Gaussianity of nodal voltages)	17

2.7.4	Proof of Theorem 1	18
2.7.5	Proof sketch for Theorem 2	19
3	Real-time Assessment of Distribution Grid Security through Adaptive Smart Meter Measurements	21
3.1	Introduction.....	21
3.1.1	Literature Review.....	22
3.1.2	Contributions	23
3.2	Problem Description	24
3.2.1	Optimization Problem for Extreme Achievable Voltages	24
3.2.2	Criteria for Success.....	26
	Successful Certification for Secure Operations	26
	Successful Identification of Highest and Lowest Voltages	26
3.2.3	Algorithm for Adapting the Measurement Set	27
	Metrics to assess the value of a measured node	27
	Avoiding recursive loops	28
	Iterative Algorithm to Adjust the Measurement Set	28
3.3	Numerical Test	29
3.3.1	Simulation Setup	29
3.3.2	Test Result	30
3.4	Conclusion.....	34
4	Classifying Reactive Power Control Laws of Behind-the-Meter Solar Photovoltaic Inverters	36
4.1	Introduction.....	36
4.1.1	Related work.....	37
4.1.2	Contributions	37
4.2	Methodology	38
4.2.1	Nearest Centroid Classifier Method	38
4.2.2	Neural Network Based Method	39
4.3	Numerical Results.....	40
4.3.1	Setup and Dataset.....	41

4.3.2	Nearest Centroid Classifier Results	41
4.3.3	Neural Network-Based Classification Results	42
4.3.4	Comparing Methods	43
4.4	Conclusion	47
5	Forecast-Aided State Estimation in Unbalanced Distribution Networks Using Smart Me- ter Data under Limited Communication Bandwidth	48
	Abbreviations	48
5.1	Introduction	49
5.1.1	Related Work	49
5.1.2	Contributions	50
5.2	Proposed Approach	52
5.2.1	Notation	52
5.2.2	Overview	52
5.3	Three-phase LinDistFlow	53
5.4	Forecasting	54
5.4.1	Vector Autoregressive Model (VAR)	55
5.4.2	Estimating the Coefficient and Covariance Matrices	55
	Yule-Walker	56
	Recursive Least Squares (RLS)	56
	RLS with Partial Measurements	57
5.4.3	Forecasting the System States	58
5.5	State Estimation	59
5.6	Numerical Simulations	60
5.6.1	Setup	61
	Test System	61
	Load and Renewable Data	61
	Measurements	61
	Scenarios	61
5.6.2	Forecasting Performance	62

5.6.3	State Estimation Performance	64
5.6.4	Impact of the Number of Measurements	68
5.7	Conclusions.....	68
6	Error Bounds for Radial Network Topology Learning from Quantized Measurements	70
6.1	Introduction.....	70
6.2	Problem formulation	71
6.2.1	Communication model	71
6.2.2	Statistical tools.....	72
6.2.3	Network and power flow model	72
	Linear Coupled Power Flow (LCPF) model	73
	Lifting to the complete graph	73
6.3	Main result.....	74
6.3.1	Topology learning problem	74
6.3.2	Bounding the error of the parameter estimate.....	74
6.3.3	Numerical results	75
6.4	Conclusion.....	76
7	Kalman Filter-based State Estimator with Measurement Quantization.....	77
7.1	Introduction.....	77
7.1.1	Literature Review.....	78
7.1.2	Problem Statement and Contributions	79
7.2	Methodology	80
7.2.1	One-Dimensional System	80
	Conventional One-Dimensional Kalman Filter	80
	Proposed Quantized One-Dimensional Kalman Filter.....	82
	Gaussian Approximation.....	82
	2-bin 1-bit Quantizer	84
	Proposed 3-bin 1-bit Quantizer	85
7.2.2	Multi-Dimensional System	88
	Multi-Dimensional Kalman Filter	88

	Multi-Dimensional Quantized Kalman Filter	90
	Rate distortion of Independent Gaussian Sources	90
	Reverse Water-Filling and Quantized Kalman Filter	92
	Calculating Maximum Individual Distortion (δ) For a Fixed Total Rate (R)....	92
	Practical Implementation.....	93
7.3	Results	95
7.3.1	One-Dimensional System Results	95
7.3.2	Multi-Dimensional System Results.....	97
7.4	Conclusion.....	102

List of Figures

Figure 1 Conceptual illustration of dynamically using limited-bandwidth communication across a fleet of smart meters.	ii
Figure 2.1 Illustration of the problem: Given an aggregate bandwidth limit across nodes, adaptively design sampling a policy $\mathcal{S} \subseteq \mathcal{N}$ to expose violations of voltage magnitude limits.	8
Figure 2.2 Performance with unity power factor at all nodes. The plots show regret relative to the LinDistFlow and AC optimal sampling vs. time with spectral (left) and ℓ_2 (right) regularization.	14
Figure 2.3 Performance with random, non-unity power factors. The plots show regret relative to the LinDistFlow and AC optimal sampling vs. time with spectral (left) and ℓ_2 (right) regularization.	15
Figure 3.1 Flowchart of the proposed algorithm.	30
Figure 3.2 Value of load factor on scenario 1.	32
Figure 3.3 Value of load factor on scenario 5.	34
Figure 4.1 Sample mean curve for the different control laws.	40
Figure 4.2 The architecture of the fully connected neural network model for classifying a solar inverter's control law.	42
Figure 4.3 Impact of the number of input time periods on the neural network's classification accuracy.	44
Figure 4.4 Receiver operating characteristic curve for the neural network classifier.	46
Figure 4.5 Variation of classification accuracy vs. smart meter measurement error.	46
Figure 5.1 Overview of the proposed state estimation framework, partial real time measurements	51
Figure 5.2 Correlation in the active power injections in (a) the base case and (b) the low-correlation scenario.	60

Figure 5.3 Forecast accuracy of the VAR(1) model when trained using (a) 1000 time steps and (b) 200 time steps. In both plots, the dashed black line represents the true nodal power injections. Both methods perform comparably when the VAR model is trained using 1000 time steps of data. However, with 200 samples, the YW method fails to accurately predict the power injections, while the RLS method is more robust. 63

Figure 5.4 Performance of the proposed state estimation method in different scenarios. Figs. 5.4a–5.4c show the substation active power injections under different scenarios. Figs. 5.4d–5.4i show the performance of the proposed method with 10% (top) and 20% (bottom) measurements, in the base case (left, BC), with PV integration (middle, PV), and in the low-correlation scenario (right, LC). As an illustrative example, we plot the voltage at one particular three-phase node but emphasize that the performance is consistent across all the nodes in the network as shown by Tables 5.2–5.4. 65

Figure 5.5 Impact of the number of measured nodes on state estimation performance. 67

Figure 6.1 Conceptual illustration of the proposed problem. Left: Distribution network smart meter measurements have realistic quantization noise corresponding to *bin width* $\Delta > 0$. Right: Estimate the topology *and* parameters of the distribution network (green, solid lines). 71

Figure 6.2 Relative error vs. number of samples for repeated runs of the topology learning program (6.3) on the Baran & Wu (left) and CIGRE low-voltage (right) test cases. The scatter markers show the experimental error obtained for one run of (6.3), where each color corresponds to a bin width Δ . The solid lines show the error prescribed by (6.12). 75

Figure 7.1 Studied sensed network architecture in a distribution network with smart-meters. 80

Figure 7.2 Propagation of the state distribution with the conventional Kalman Filter approach. 81

Figure 7.3 Propagation of the state distribution with noisy measurements replaced by quantization. 82

Figure 7.4 Steps of the propagation of state distribution with quantization and Gaussian approximation.	83
Figure 7.5 2-bin 1-bit quantization.	84
Figure 7.6 Proposed 3-bin 1-bit quantization.	85
Figure 7.7 Propagation of the states distributions in the conventional Kalman filter approach.	89
Figure 7.8 Propagation of the state distributions in the quantized Kalman filter approach. ...	90
Figure 7.9 Reverse water filling for independent Gaussian sources.	91
Figure 7.10 Quantized Kalman filter and reverse water-filling.	93
Figure 7.11 Practical implementation example.	95
Figure 7.12 True state and quantized state estimator for each quantization method.	96
Figure 7.13 Portion of all cases random walk simulations with Quantizer 1, and prior/posterior variances in steady-state.	98
Figure 7.14 Theoretical and achieved rate-distortion curves of each quantizer and all study cases.	

List of Tables

Table 3.1 Result of Scenario 1	31
Table 3.2 Measurement Update Process	33
Table 3.3 Status of Different Scenarios	34
Table 4.1 Prediction accuracy of Nearest Centroid classifier	43
Table 4.2 Confusion matrices for the two classification models. Diagonal entries represent correct predictions as a percentage of the total predictions.	45
Table 5.1 Forecasting accuracy with varying training dataset sizes and measurement sets ...	63
Table 5.2 Voltage estimation error in the base case	64
Table 5.3 Voltage estimation error with PV integration	64
Table 5.4 Voltage estimation error under the low correlation	66
Table 7.1 Simulation Results.	96
Table 7.2 Results for $R = 3$ bits.	99
Table 7.3 Results for $R = 12$ bits.	100

1. Introduction

This document is the final report for the PSERC project T-67: Smart Meter-Driven Distribution Grid Visibility and Control.

1.1 Motivation

The proliferation of distributed energy resources such as rooftop solar photovoltaic generators, electric vehicles, and flexible loads is transforming power distribution systems from passive to highly active networks. This transition creates both opportunities and challenges for utilities. On the one hand, distributed energy resources offer new capabilities for local generation and grid services; on the other hand, their variability and lack of direct utility control increase uncertainty and risk.

To address these challenges, utilities require improved real-time visibility and control of distribution grids. Smart meters, already widely deployed for billing and customer services, are a promising source of system data. They provide frequent, accurate measurements of voltage, current, and power. However, their potential for real-time monitoring has been limited by communication constraints: utilities cannot continuously retrieve data from every meter. Instead, only a fraction of meters can provide high-frequency data at a given time. This creates a need for new algorithms that allocate bandwidth dynamically and extract maximum value from limited measurements.

This project, PSERC T-67: Smart Meter-Driven Distribution Grid Visibility and Control, develops computational methods to leverage the capabilities of smart meters. The research leverages optimization, machine learning, and information theory to design algorithms that operate effectively under bandwidth limitations. By doing so, the project contributes to enabling secure, reliable, and cost-effective operation of modern power grids without requiring costly new infrastructure.

1.2 Key Contributions

The project makes several contributions to distribution system monitoring and control:

1. **Online sensing strategies for voltage security** that adaptively allocate smart meter bandwidth to identify worst-case violations of voltage magnitude limits.
2. **Machine learning methods for inverter characterization** that classify behind-the-meter solar inverter control laws using only smart meter data.
3. **Forecast-aided state estimation** methods that combine predictive models with bandwidth-limited measurements to improve distribution system state estimation.

4. **Topology identification under quantization** that develops statistical methods for recovering distribution grid topology using quantized smart meter data.
5. **Information-theoretic analysis** of smart meter communication, providing fundamental limits and guidelines for managing constrained measurement resources.

Together, these advances provide a unified framework for extending the role of smart meters from billing tools to critical infrastructure for real-time grid operation.

1.3 Chapter Summaries

Chapters 2 and 3: Online Strategies for Identifying Worst-Case Voltage Magnitude Violations.

These chapters develop two approaches to detect potential violations of voltage limits under communication constraints. The first method in Chapter 2 formulates sensor selection as a spectral multi-armed bandit problem, using stochastic LinDistFlow models and spectral upper confidence bounds to adaptively choose which meters should provide measurements at each time period. The second method in Chapter 3 is an iterative optimization algorithm that dynamically selects the measurement set by targeting buses with the largest range of potential voltage magnitude violations. Simulations on benchmark feeders show both methods outperform baseline sampling strategies, demonstrating that limited communication capacity can still support effective monitoring when guided by intelligent algorithms.

Chapter 4: Machine Learning for Classifying Inverter Control Laws.

Utilities often lack information about the reactive power control laws implemented by customer-owned photovoltaic inverters, yet these settings strongly influence power grid performance. This chapter presents classification algorithms that infer inverter control behavior from smart meter data. Using synthetic datasets, the study compares nearest-centroid and neural network models. Results show that neural networks deliver higher accuracy and robustness to noise, enabling utilities to improve distribution models and operational forecasts without direct access to inverter settings.

Chapter 5: Forecast-Aided Distribution System State Estimation.

State estimation is key to system monitoring, but traditional approaches for transmission networks rely on dense sensing that is rarely feasible in practice for distribution networks. This chapter integrates forecasting with state estimation to compensate for sparse measurements. A vector autoregressive (VAR) model with recursive updates provides forecasts of nodal power injections, which are then refined using available smart meter data. Case studies on unbalanced feeders demonstrate that forecast-aided state estimation achieves high accuracy even when only 10–20% of buses report in real time, highlighting the value of predictive models in bandwidth-limited environments.

Chapter 6: Online Identification of Distribution Network Topologies Using Quantized Measurements.

Accurate feeder topology information is essential but often incomplete or outdated. This chapter addresses the challenge of learning network topology from smart meter data that are subject to quantization error due to communication limits. Building on the LinDistFlow power flow approximation, this chapter derives statistical error bounds for topology recovery and validates the approach on test feeders. Results show that reliable topology inference is possible with coarse data, provided sufficient time-series samples are available.

Chapter 7: Information-Theoretic Approaches to Communication-Constrained Smart Meter Data.

Methods for managing limited communication capacity among a large number of sensors may be categorized into two groups: 1) sensor sampling or measurement selection, and 2) bit rate allocation. The online strategies for identifying worst-case voltage magnitude violations described in Chapter 2 above are a good example of the first approach. In contrast, we consider methods which continually acquire data from all the sensors in the network, but which allocate varying bit rates to each of the sensors based on the value of the data in a particular application such as state estimation. Well-known tools from information theory, including rate distortion theory and quantization, play a key role.

In order to determine the bit rate for a continuous random variable subject to quantization, knowledge of the prior probability distribution of that random variable is required. Hence, a probabilistic description of the underlying system from which the measurements are derived comes into play. We use a state variable description of the system, and a Kalman filter which gives the evolution of the state probability distribution functions based on the measurements. Our approach differs from conventional Kalman filtering in that we assume that the measurements are ideal, and that the measurement “error” is the result of the quantization which leads to finite bit rate communications. We first consider a one-dimensional version of the problem, in which the resulting communication system is highly reminiscent of delta modulation. A vector version of the system can also be developed, which relies on an assumption that the sensors at the network edge employ sidelink communication for joint processing prior to communication with a central processor.

1.4 Closing Remarks

The research presented in this report demonstrates that even with severe communication limitations, smart meters can play a transformative role in distribution grid monitoring and control. By integrating ideas from optimization, machine learning, forecasting, and information theory, the project advances both the theoretical foundations and the practical tools needed for future distribution grids. These contributions lay the groundwork for utilities to leverage existing infrastructure more fully, improving the reliability and efficiency of power system operations in an era of growing distributed energy resource adoption.

2. Strategic Electric Distribution Network Sensing via Spectral Bandits

Despite their wide-scale deployment and ability to make accurate high-frequency voltage measurements, communication network limitations have largely precluded the use of smart meters for real-time monitoring purposes in electric distribution systems. Although smart meter communication networks have limited bandwidth available per meter, they also have the ability to dedicate higher bandwidth to varying subsets of meters. Using this capability to enable real-time monitoring from smart meters, this chapter proposes an online bandwidth-constrained sensor sampling algorithm that takes advantage of the graphical structure inherent in the power flow equations. The key idea is to use a spectral bandit framework where the estimated parameters are the graph Fourier transform coefficients of the nodal voltages. The structure provided by this framework promotes a sampling policy that strategically accounts for electrical distance. Maxima of sub-Gaussian random variables model the policy rewards, which relaxes distributional assumptions common in prior work. The scheme is implemented on a synthetic electrical network to dynamically identify meters exposing violations of voltage magnitude limits, illustrating the effectiveness of the proposed method.

This chapter is based on the following paper:

S. Talkington, R. Gupta, R. Asiamah, P. Buason, and D.K. Molzahn, “Strategic Electric Distribution Network Sensing via Spectral Bandits,” *63rd IEEE Conference on Decision and Control (CDC)*, December 16-19, 2024.

2.1 Introduction

Sensing infrastructure plays a critical role in the control and operation of electricity networks and is crucial for applications such as state estimation, system identification, and data-driven control schemes [1–4]. For transmission networks, observability is maintained by using phasor measurement units and micro-phasor measurement units spread across multiple substations. This allows for continuous monitoring of the entire system. The same level of real-time observability is not available for distribution networks, however. Smart meters with advanced metering infrastructure (AMI) could help address this problem.

The number of smart meters installed throughout the United States continues to increase, reaching 107 million units in 2021 [5], accounting for more than 50% of all electricity meters installed throughout the country. By regularly measuring and recording nodal voltages, currents, and power injections [6], smart meters have the capability to increase visibility into distribution networks. The frequency and consistency of the recorded data make them good candidates for real-time control and operation applications in distribution networks. Smart meters have the added benefit of being widely available in existing systems, eliminating the need for new infrastructure investments to

increase visibility.

Smart meters send measurements to a central server that performs various calculations, often for customer billing purposes. Smart meters communicate these measurements through channels with limited bandwidth connections to the central server [7]. Although the bandwidth of the communication network is sufficient to send power consumption data at 15-minute to hourly intervals for billing purposes, bandwidth limitations pose a significant challenge for real-time monitoring purposes, especially when the limited bandwidth is shared by many smart meters. However, new smart meters can be dynamically queried so that different subsets of the meters can report data at high frequencies. This motivates the development of new algorithms to take advantage of this capability for real-time monitoring purposes.

This chapter focuses on the task of dynamically identifying smart meters to query with the goal of revealing violations of voltage magnitude limits so that system operators can undertake corrective actions. With rapidly growing deployments of distributed energy resources that contribute to substantial variations in power injections, system operators would benefit from algorithms for real-time *online* identification of voltage violations. This requires dynamically selecting varying sets of smart meters to identify voltage violations.

2.1.1 Related work

Smart meter data has been used for real-time monitoring in electric distribution networks [8, 9], and many other applications in load forecasting, demand response, and consumer characterization applications [10]. Moreover, there is a rapidly growing number of works that develop sensor selection methods for various monitoring tasks in infrastructure network settings. The work of [11] developed a game-theoretic approach to network monitoring, while [12] developed a change-point method based on bandit algorithms—a highly similar line of research to ours. Distinct from the bandit formulation, [13] developed a method to detect change points. More broadly, adaptive sampling techniques have seen recent innovation in continuous environments [14]. Recent work developed an Upper Confidence Bound (UCB)-type algorithm where the rewards are extreme quantities derived from a Gaussian process reward function [15].

Additional authors have developed a wealth of methods for sensor placement in generic network settings. A rich literature from the authors of [11, 16, 17] has developed analytical zero-sum game approaches to time-invariant network inspection. Additional online efforts have explored reinforcement learning algorithms [18].

In the setting of electric distribution networks, sensor selection strategies for network topology identification [19] have been developed with identifiability guarantees. The work in [20] developed submodular algorithms for sensing applications in water distribution networks. Nevertheless, online algorithms of this kind have yet to be utilized for strategic detection of network constraint violations.

2.1.2 Proposed Framework and Contributions

This work proposes an online resource-constrained sensor sampling algorithm that exploits the graphical structure inherent in the power flow equations. The key idea is to use a spectral bandit framework, where the parameters we estimate are the graph Fourier transform coefficients of the nodal voltages. The structure provided by this framework promotes a sampling policy that accounts for electrical distances.

This chapter is part of an extensive and growing literature on multi-armed bandit algorithms [21, 22]. We are particularly close to works on context-varying linear bandits [22], the large body of work on bandit algorithms leveraging graph spectral structure [23–26], and combinatorial bandits [27, 28]. Recent work in [26] introduced a graph-based upper confidence bound algorithm similar to ours.

Our work uses the language of bandits, which is a particular sequential game formulation. This enables us to recursively solve the sensor sampling problem to detect violations of constraints on nodal states in a streaming fashion. To the best of our knowledge, this is the first such work to do this.

In summary, the contributions of this research are:

1. A stochastic LinDistFlow model, developed in Section 2.3, which is agnostic to the probability distributions of the nodal power injections.
2. A sufficient condition for nodal voltage magnitudes to be sub-Gaussian random variables; consequently, this yields concentration inequalities for their maximal fluctuations. In particular, these inequalities explicitly depend on the network model topology and parameters.
3. A new spectral bandit algorithm, developed in Section 2.4, which strategically samples a small subset of nodes to find maximal voltage magnitude fluctuations.

2.1.3 Notation

For a matrix $A \in \mathbb{C}^{n \times d}$, we denote its transpose as $A^\top$. The p norm of a vector $x \in \mathbb{R}^n$ is denoted as $\|x\|_p$. The norm of a vector $x \in \mathbb{R}^n$ with respect to matrix $A \in \mathbb{R}^{n \times n}$ is denoted as $\|x\|_A := \sqrt{x^\top A x}$. The imaginary unit is $j := \sqrt{-1}$, and $\bar{x}$ is the conjugate of any complex number $x \in \mathbb{C}$. Expectation and probability are denoted as $\mathbb{E}[\cdot]$ and $\Pr(\cdot)$, respectively. We write $a \lesssim b$ if $a \leq Cb$ for some universal constant C . The symbol $\mathbf{1}$ is a vector of all ones, and $\text{diag}(\cdot)$ is a diagonal matrix with entries given by the argument.

2.2 Problem setting

In this chapter, we focus on analyzing an undirected graph $\mathcal{G}$ that represents a distribution network in power systems, with nodes denoted as $\mathcal{N} := \{1, \dots, n\}$ and lines represented by $\mathcal{E} \subseteq \mathcal{N} \times \mathcal{N}$.

The structure of the graph is akin to that of a power distribution network, although the framework we develop could be extended to other network types. Let $A \in \{-1, 0, 1\}^{m \times n}$ be the edge-to-node incidence matrix of the network, and let $w := [w_{ij}]_{(i,j) \in \mathcal{E}}$ be a vector of *edge weights* ordered corresponding to the rows of A ; they both may model self-edge weights.

2.2.1 Power flow model for distribution systems

We briefly review the graphical model for a single-phase distribution network. The network admittance matrix $Y := A^\top \text{diag}(w)A \in \mathbb{C}^{n \times n}$ is a weighted graph Laplacian matrix that encodes Kirchhoff's current and voltage laws. Let complex nodal voltages be $u \in \mathbb{C}^n$ and nodal power injections be $s \in \mathbb{C}^n$; the two are related by the *injection model* of the power flow equations, which take the form $s = \text{diag}(u)\overline{Y}u$ with real (active) and imaginary (reactive) components given as $p := \text{Re}\{s\}$ and $q := \text{Im}\{s\}$, respectively.

2.2.2 Review of the LinDistFlow approximation

We consider a linear power flow approximation known as LinDistFlow to represent the nodal voltage magnitudes in radial distribution networks based on Laplacian matrices; we refer the reader to [29, Sec. 2] for additional exposition. Due to the tree structure of distribution networks, the incidence matrix with the first column removed, $A \in \{-1, 0, 1\}^{n \times n}$, is *square*. Using $r, x \in \mathbb{R}^m$ to denote the line resistances and reactances, we can approximate the nodal voltage magnitudes $v \in \mathbb{R}^n$ as $v := v^\bullet \mathbf{1} + Rp + Xq$, where $v^\bullet$ denotes the nominal slack voltage, $R := G^{-1}$ and $X := B^{-1}$, where $G, B \in \mathbb{R}^{n \times n}$ are $n \times n$ Laplacian matrices of the form $G := A^\top \text{diag}(r)^{-1}A$ and $B := A^\top \text{diag}(x)^{-1}A$.

Assumption 1. *The ratio of reactive to active injections, $\kappa = q_i/p_i$, is the same for all nodes $i \in \mathcal{N}$ in the network.*

Physically, Assumption 1 is equivalent to the statement that *all power factors* are the same at every node in the network. Importantly, while Assumption 1 aids in the forthcoming theoretical analyses, our numerical results in Section 2.5 empirically indicate that it may be unnecessary in practice. The algorithm appears to perform well in our experiments, even if Assumption 1 is not satisfied.

Lemma 1. *Given Assumption 1, there exists a Laplacian matrix $L = (R + X\mathbf{I}\kappa)^{-1}$ and orthonormal eigenbasis W such that $L := W\Lambda W^\top$, $\Lambda := \text{diag}(\lambda_1, \dots, \lambda_n)$, and the LinDistFlow model is then $v = v^\bullet \mathbf{1} + L^\dagger p$ for any $p \in \mathbb{R}^n$, where $\{\cdot\}^\dagger$ denotes the Moore-Penrose pseudoinverse, and $v^\bullet$ is a nominal voltage, e.g., 1 per unit (pu).*

We prove the Lemma in Section 2.7.1. The Laplacian eigenbasis W allows us to construct the *graph Fourier transform* of the nodal injections and voltages, which we write as $\rho := W^\top p$ and $\psi := W^\top v$, respectively. Substituting these transformations into the LinDistFlow model yields the following representation of the voltages

$$v - v^\bullet \mathbf{1} = L^\dagger p = W\Lambda^{-1}W^\top p = W\psi \quad (2.1)$$

in the *graph Fourier basis* W . In the following, the graph Fourier transform coefficients ψ are the parameters we wish to estimate.

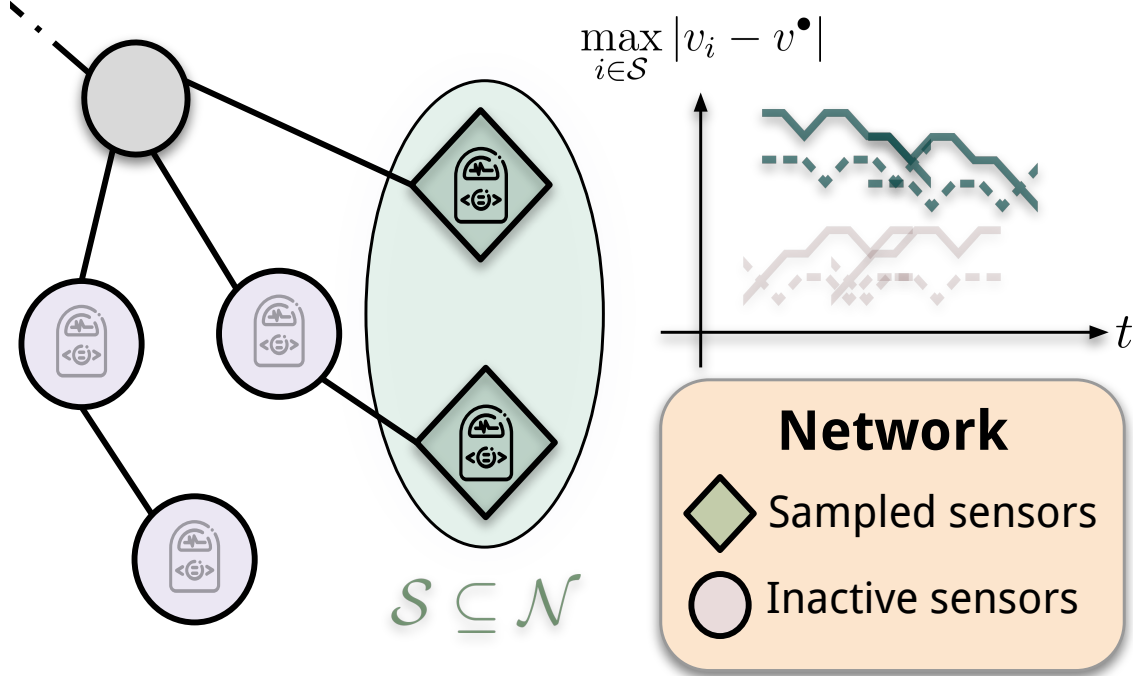


Figure 2.1: Illustration of the problem: Given an aggregate bandwidth limit across nodes, adaptively design sampling a policy $\mathcal{S} \subseteq \mathcal{N}$ to expose violations of voltage magnitude limits.

2.2.3 Action set and reward formulation

The proposed sensor sampling problem can be expressed naturally in the language of multi-armed bandits. At each measurement interval, or *round* t , the learner selects a subset of no more than b nodes to sample to identify possible voltage violations in the network. Thus, the set of all available sampling strategies is defined via the action set

$$\mathcal{A} = \left\{ \mathcal{S} \in 2^{\mathcal{N}} : |\mathcal{S}| \leq b \right\}, \quad (2.2)$$

of which there are $|\mathcal{A}| = \binom{n}{b}$ possible strategies. The action set is a subset of the n -dimensional hypercube: $\mathcal{A} \subset \{0, 1\}^n$. This is a classic example of a *combinatorial bandit* problem; see [27, 28], and [21, Ch. 30] for detailed discussions.

Often, engineers wish to use sensors in distribution systems to determine how far away nodal voltages are from their nominal values. Thus, an appropriate model for the effectiveness of a sampling strategy $\mathcal{S} \in \mathcal{A}$ is the *maximum deviation* from the nominal value observed at any node $i \in \mathcal{S}$. Concretely, after selecting sensors $\mathcal{S} \in \mathcal{A}$ to *query*, the learner then observes a *reward* $f : \mathcal{A} \rightarrow \mathbb{R}$ in the form

$$f(\mathcal{S}) = \max_{i \in \mathcal{S}} |v_i - v_i^*| = \max_{i \in \mathcal{S}} |\langle w_i \rangle \psi|, \quad (2.3)$$

where v_i and v_i^* are the sampled and nominal voltage at node $i \in \mathcal{S} \subseteq \mathcal{N}$, respectively. The voltages v_i are random variables distributed as $v_i \sim \mathcal{D}_i$, where $\mathcal{D}_i$ is an arbitrary distribution for

all nodes $i \in \mathcal{N}$. Hereafter, the nominal voltage $v_i^\bullet$ will take the form of the expected value of that nodal quantity, conditioned on past observations, $v_i^\bullet = \mathbb{E}[v_i | \mathcal{F}_{t-1}]$, where $\mathcal{F}_{t-1}$ is a filtration, which is a collection of all past input-output pairs (w_i, v_i) up to time t .

Note that the reward (2.3) is higher for sampling strategies that expose *extreme deviations* of the nodal quantities from their nominal values. In addition, it has additional appealing properties, described in Lemma 2.

Lemma 2. *The reward function $f : \mathcal{A} \rightarrow \mathbb{R}$ is monotone submodular in $\mathcal{A}$ and 1-Lipschitz (see the proof in Section 2.7.2).*

At each *round* or time step, $t = 1, \dots, m$, the power injections $p_1, \dots, p_t \in \mathbb{R}^n$ are a stochastic process. The learner then selects subsets of sensors $\mathcal{S}_1, \dots, \mathcal{S}_t \subseteq \mathcal{N}$ to *query*. The *regret* of the learner over m time steps is

$$R_m := \mathbb{E} \left[\max_{\mathcal{S} \in \mathcal{A}} \sum_{t=1}^m f(\mathcal{S}) - f(\mathcal{S}_t) \right], \quad (2.4)$$

which the learner wishes to minimize; we, in turn, wish to bound it. The regret (2.4) models the cumulative differences between the maximal average perturbation around a nominal point and the maximal perturbations observed by our chosen sampling policy.

2.3 Stochastic LinDistFlow model

This section proposes a stochastic, distribution-agnostic LinDistFlow model. Given a mild assumption on power consumption in distribution grids, we use the structure provided by LinDistFlow to probabilistically model the extrema of voltage magnitude fluctuations about a nominal point. Many of our results are based on sub-Gaussian concentration. We recall the definition of these random variables below.

Definition 1. *A random variable $x \in \mathbb{R}$ is sub-Gaussian with parameter σ if $\mathbb{E} \left[e^{\lambda(x - \mathbb{E}[x])} \right] \leq e^{\lambda^2 \sigma^2 / 2}$ for all $\lambda \in \mathbb{R}$.*

We first make a general statement on the fluctuation of random nodal voltages in the LinDistFlow model around their nominal values. Critically, this statement holds for a broad family of possible active power injection distributions—that is, any distribution with bounded support. Specific knowledge of power injection distributions is unnecessary.

We emphasize that this work does not assume specific distributions for any random quantity. Instead, a *family* of possible distributions is considered. Concretely, for each sampling interval $t = 1, \dots, m$, let $p_t := p_\star + \tilde{p}_t \in \mathbb{R}^n$ be the nodal injections, where $p_\star$ is a nominal value, and $\tilde{p}_t$ is a random change in the nodal injections drawn from an unknown distribution. The only assumption is that the *range* of power consumption can be inferred, as we describe now.

Assumption 2. *Assume that the power consumption at each node $i \in \mathcal{N}$ is an independent vari-*

able whose range is a predictable process, and therefore we can infer bounds on the injections $\{(\underline{p}_t, \bar{p}_t)\}_{t=1}^m$ from historical data such that $\underline{p}_t \leq \tilde{p}_t \leq \bar{p}_t$ almost surely.

Hereafter, denote $\Delta_t := \bar{p}_t - \underline{p}_t > 0$.

2.3.1 Concentration of nodal voltage perturbations

Using the structure of the LinDistFlow model, Assumption 2 immediately gives rise to concentration bounds for the voltages. Moreover, Assumption 2 relaxes the Gaussianity assumptions common in prior work (e.g. [30, 31]), which we discuss in Section 2.7.3. The following results are valid even if $(\underline{p}_t, \bar{p}_t)$ are the largest physically plausible injections, i.e., grid constraint bounds. Access to tighter bounds forecast by the sensors only serves to make the results more precise.

Lemma 3. *If Assumption 2 holds, then for each $i \in \mathcal{N}$ we have that $v_i - v^\bullet$ is a sub-Gaussian random variable with parameter $\frac{1}{2}\Delta \|\Lambda^{-1}w_i\|_2$.*

Lemma 3, proven in Section 2.7.3, characterizes a family of possible distributions on the voltage magnitudes that arise from Assumption 2. Applying sub-Gaussian maxima concentration leads to our first primary result.

Theorem 1. *Let $\mathcal{S} \subseteq \mathcal{N}$ be a sampling of b nodes. Suppose that $\Delta_t := \Delta$ for all t , and suppose that LinDistFlow accurately represents the network model. If Assumption 2 holds, we have*

$$\mathbb{E} \left[\max_{i \in \mathcal{S}} |v_i - v_i^\bullet| \right] \lesssim \frac{1}{2} \Delta \max_{i \in \mathcal{S}} \|\Lambda^{-1}w_i\|_2^2 \sqrt{2 \log(b)}; \quad (2.5)$$

moreover, for all $\varepsilon > 0$

$$\Pr \left(\max_{i \in \mathcal{S}} |v_i - v_i^\bullet| > \varepsilon \right) \leq 2b \exp \left\{ \frac{-2\varepsilon^2}{\Delta^2 \max_{i \in \mathcal{S}} \|\Lambda^{-1}w_i\|_2^2} \right\}. \quad (2.6)$$

Theorem 1 is proven in Section 2.7.4. There is an intuitive physical interpretation for these results. The results provide probabilistic predictions of the worst-case fluctuation of nodal voltages that would be observed by a sensor sampling strategy $\mathcal{S} \subseteq \mathcal{N}$; these predictions can inform how sensors should be selected. The result is achieved by virtue of the graphical structure of the LinDistFlow model, coupled with the mild Assumption 2 on bounded load behavior.

2.3.2 Condition for positive semi-definite Laplacian

Regret guarantees for spectral bandits typically assume that the Laplacian L is positive semidefinite. This is not guaranteed in general for distribution networks—the power factor ratio κ or the reactances x may be negative, potentially rendering L indefinite [32]. However, we can construct a condition using the physical structure of distribution networks to ensure that L satisfies this requirement. This pursuit begins with an assumption that mirrors [33].

Assumption 3. Let $\kappa \in \mathbb{R}$ as in Assumption 1, and let $\mathcal{E}_+$ (resp. $\mathcal{E}_-$) be the subsets of all lines $(i, j) \in \mathcal{E}$ where $r_{ij} + \kappa x_{ij}$ is positive (resp. negative). Let L_+ be the Laplacian constructed from the subnetwork $(\mathcal{N}, \mathcal{E}_+)$. Assume that

$$|r_{ij} + \kappa x_{ij}| \geq e_{ij}^\top L_+^\dagger e_{ij} \quad \forall (i, j) \in \mathcal{E}_-, \quad (2.7)$$

where $e_{ij} := e_i - e_j$ is the difference between the i -th and j -th basis vectors and $e_{ij}^\top L_+^\dagger e_{ij}$ is the effective resistance between nodes $i, j \in \mathcal{N}$.

Intuitively, Assumption 3 requires the effective impedance between any two points in the network to be non-negative. The above condition leads to the following result.

Lemma 4. If and only if Assumption 3 holds,

$$L \succeq 0. \quad (2.8)$$

Proof. If the network lacks cycles, Assumption 3 is necessary and sufficient for the claim to hold; see [33, As. 3.1]. By radially, a single-phase distribution network lacks cycles; the claim is then a corollary of [33, Thm. 3.2]. $\square$

2.3.3 Confidence ellipsoid, sampling policy, and regret

We now provide a bound on the regret of the strategic sampling policy, when v_t is generated as noisy observations of the form $(v_t)_i - v_i^\bullet = \langle r_i + \kappa x_i \rangle p_\star + \eta_t = \langle w_i \rangle \psi_\star + (\eta_t)_i$, where $(\eta_t)_i = \langle r_i + \kappa x_i \rangle \tilde{p}_t$.

Theorem 2. Let $V_t = \Lambda + \sum_{s=1}^m w_s w_s^\top$, then there exists a $\delta \in (0, 1)$ and constants c_t for all $t = 1, \dots, m$ such that for any $w \in \mathbb{R}^n$, with probability $1 - \delta$:

$$|w^\top (\hat{\psi}_t - \psi_\star)| \leq c_t \|w\|_{V_t^{-1}}. \quad (2.9)$$

Moreover, let $b = 1$. If Assumptions 1, 2, and 3 hold, and $|w_i^\top \psi_\star| \leq 1$ for all $i \in \mathcal{N}$, the regret (2.4) of the sampler over m periods is bounded as

$$R_m \leq \tilde{\mathcal{O}}(d\sqrt{m}), \quad (2.10)$$

where d is the effective dimension, defined in [23] as

$$d := \max_{i \in \mathcal{N}} i \quad \text{s.t.} \quad (i-1)\lambda_i \leq \frac{m}{\log(1 + m/\lambda_1)}, \quad (2.11)$$

where λ_1 is the smallest eigenvalue of L .

The result in Theorem 2 can essentially be understood as stating that the regret of the proposed sampling algorithm grows sublinearly, with a scaling factor (2.11) that is at most the number of nodes, $d \leq n$.

Hereafter, for a sampling strategy $\mathcal{S} \subseteq \mathcal{N}$, we let $W_{\mathcal{S}} \in \mathbb{R}^{b \times n}$ denote the submatrix of the graph Fourier basis W with rows corresponding to $\mathcal{S}$. It is straightforward to combine [27, Thm. 1] by Lemma 2, while using the regularization framework of [23] to produce a $\tilde{\mathcal{O}}(d\sqrt{bm})$ bound. We leave the completion of this proof to future work.

2.4 Sensor Sampling Algorithm

The greedy algorithm in [34] approximately maximizes a normalized, monotone, submodular objective function $F : 2^{\mathcal{N}} \mapsto \mathbb{R}$ subject to cardinality constraints. In particular, the algorithm solves the program $\max_{\mathcal{S} \in 2^{\mathcal{N}}} F(\mathcal{S})$, subject to $|\mathcal{S}| \leq b$, within a tolerance of $(1 - \frac{1}{e})$ of the global solution.

This algorithm will be integrated into our bandit framework; see [27] for an explanation on how the integration of this step into the algorithm affects the regret bound.

2.4.1 Optimism principle

To construct our proposed algorithm, we first define an optimistic upper bound on the reward function $\tilde{f} : \mathcal{A} \rightarrow \mathbb{R}$. This reward leverages the confidence ellipsoid developed in Section 2.3.3; it takes the form

$$\tilde{f}(\mathcal{S}) = \max_{i \in \mathcal{S}} |w_i^\top \psi - v_i^\bullet| + c \|w_i\|_{V_t^{-1}}, \quad (2.12)$$

where c is the exploration parameter. In Algorithm 1, $W_{\mathcal{S}} \in \mathbb{R}^{b \times n}$ is the submatrix formed from the b rows of the basis W indexed by $\mathcal{S}$.

2.4.2 Spectral UCB

Online decision-making problems in which rewards are generated from an underlying graphical model can be solved with the spectral bandit framework [23, 25]. These algorithms recursively solve a regularized least squares program

$$\hat{\psi}_t = \operatorname{argmin}_{\psi \in \mathbb{R}^n} \sum_{s=1}^{t-1} (v_s - \langle w_s, \psi \rangle)^2 + \beta \|\psi\|_{\Lambda}^2, \quad (2.13)$$

where $w_s \in \mathbb{R}^n$ is the row of the graph Fourier basis W selected at time step s , which in our case, corresponds to having sampled the sensor at the node $i_s \in \mathcal{N}$. The regularization parameter $\beta > 0$ is chosen by the user; a particular choice of β that depends on the *effective dimension* of the Laplacian allows the program (2.13) to achieve superior regret in comparison to standard least squares [25, Thm. 8]. The regularizer $\|\cdot\|_{\Lambda}$ is the following norm defined in terms of the graph Laplacian:

$$\|\psi\|_{\Lambda} := \sqrt{\psi^\top \Lambda \psi} = \sqrt{p^\top \mathcal{W} \Lambda^{-1} \mathcal{W}^\top p} = \sqrt{p^\top \mathcal{L}^{-1} p}. \quad (2.14)$$

The regularizer (2.14) promotes spatially smooth voltage predictions. Specifically, this regularizer promotes solutions where predicted voltages are similar at well-connected nodes; see [23] for

Algorithm 1 EXTREMESPECTRALSAMPLER

Require: $\{n, m, b\}$: # nodes, pulls, and samples/pull,
 $\{\Lambda, W\}$: spectral basis of L ,
 $\{\lambda, \delta\}$: regularization and confidence parameters,
 $\{p, \bar{p}\}$: report bounds on nodal injections,

- 1: $\Delta \leftarrow \bar{p} - p$
- 2: $d \leftarrow (2.11)$
- 3: **for** $t = 1, \dots, T$ **do**
- 4: $\mathcal{S}_0 \leftarrow \emptyset$,
- 5: $i \leftarrow 1$
- 6: **while** $i \leq b$ **do**
- 7: $k_i \leftarrow \arg \max_{j \in \mathcal{N}} \tilde{f}(\mathcal{S}_{i-1} \cup \{j\}) - \tilde{f}(\mathcal{S}_{i-1})$
- 8: $\mathcal{S}_i \leftarrow \mathcal{S}_{i-1} \cup \{k_i\}$
- 9: $i \leftarrow i + 1$
- 10: **end while**
- 11: Observe noisy voltages $v_t \leftarrow v^\bullet + W_{\mathcal{S}} \psi + \eta_t$
- 12: Observe noisy reward $f_t \leftarrow \max_{i \in \mathcal{S}_t} |w_i^\top \psi_\star + \eta_t|$
- 13: $V_{t+1} \leftarrow V_t + W_{\mathcal{S}}^\top W_{\mathcal{S}}$
- 14: $\hat{\psi}_{t+1} \leftarrow V_{t+1}^{-1} \sum_{s=1}^t W_{\mathcal{S}}^\top (v_s - v^\bullet)$
- 15: **end for**

further details. The program (2.13) has a closed form solution at each time t :

$$\hat{\psi}_t = \left(\sum_{s=1}^{t-1} w_s w_s^\top + \beta \Lambda \right)^{-1} \left(\sum_{s=1}^{t-1} w_s v_s \right). \quad (2.15)$$

2.5 Numerical Experiments

Experimentally, we demonstrate the proposed method on the `case33bw` network [35]. A demonstration of the online sampling algorithm is shown in Fig. 2.2 for the simple case where all power factors are unity ($\alpha_i = 1$ for each node $i \in \mathcal{N}$). The results with the spectral regularizer (2.14) in the left subfigures are juxtaposed with the conventional LinUCB algorithm—which uses ℓ_2 -norm regularization—shown in the rightmost subfigures. As expected, a sublinear growth in regret is observed for both, and the growth rate is markedly improved for the SpectralUCB algorithm compared to the LinUCB algorithm.

In the next experiment, we do not assume unity power factors. Instead, the reactive power control parameters vary for each node. We generate these control parameters randomly as fixed power factors bounded between 0.90 and 1, that is, $0 \leq \kappa \leq 0.48$. Similarly, the signs of the reactive power injections are generated according to a Rademacher distribution, i.e., $\text{sgn}(\kappa) = \pm 1$ equiprobably. These parameters allow us to construct reactive power injections for any fixed active power injection, as we discuss in Section 2.7.1. Similarly to the finding in Fig. 2.2, the result shown in Fig. 2.3

Cumulative regret vs. time by sample bandwidth and regularization (unity)

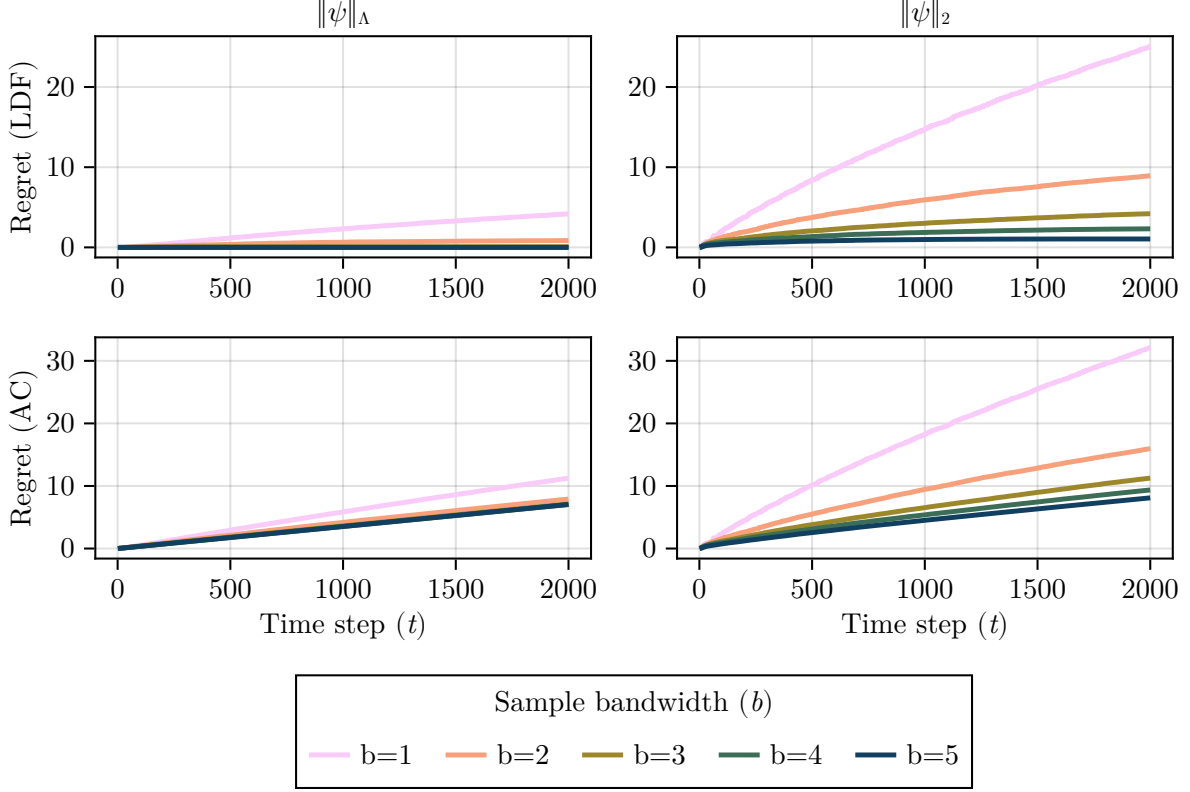


Figure 2.2: Performance with unity power factor at all nodes. The plots show regret relative to the LinDistFlow and AC optimal sampling vs. time with spectral (left) and ℓ_2 (right) regularization.

illustrates a significant reduction in cumulative regret using the SpectralUCB algorithm compared to the LinUCB algorithm.

Finally, we compute an “AC regret” quantity. Specifically, over the entire time horizon $t = 1, \dots, m$, we precompute voltage magnitudes corresponding to the true AC power flow solution. This differs from the previous experiments depicted in the top subfigures of Figs. 2.2 and 2.3, where the LinDistFlow approximation of the voltage magnitudes serves as the ground truth. Instead, we compute the regret (2.4) as $R_t = \mathbb{E} [\sum_{\tau=1}^t f(\mathcal{S}_{ac}) - f(\mathcal{S}_t)]$, where $\mathcal{S}_{ac}$ is the optimal clairvoyant sampling strategy given from the AC solutions. The AC regret results shown in the bottom subfigures of Fig. 2.2 and Fig. 2.3 similarly show a reduction of the growth rate in regret when using the SpectralUCB algorithm.

Cumulative regret vs. time by sample bandwidth and regularization (rand. non-unity)

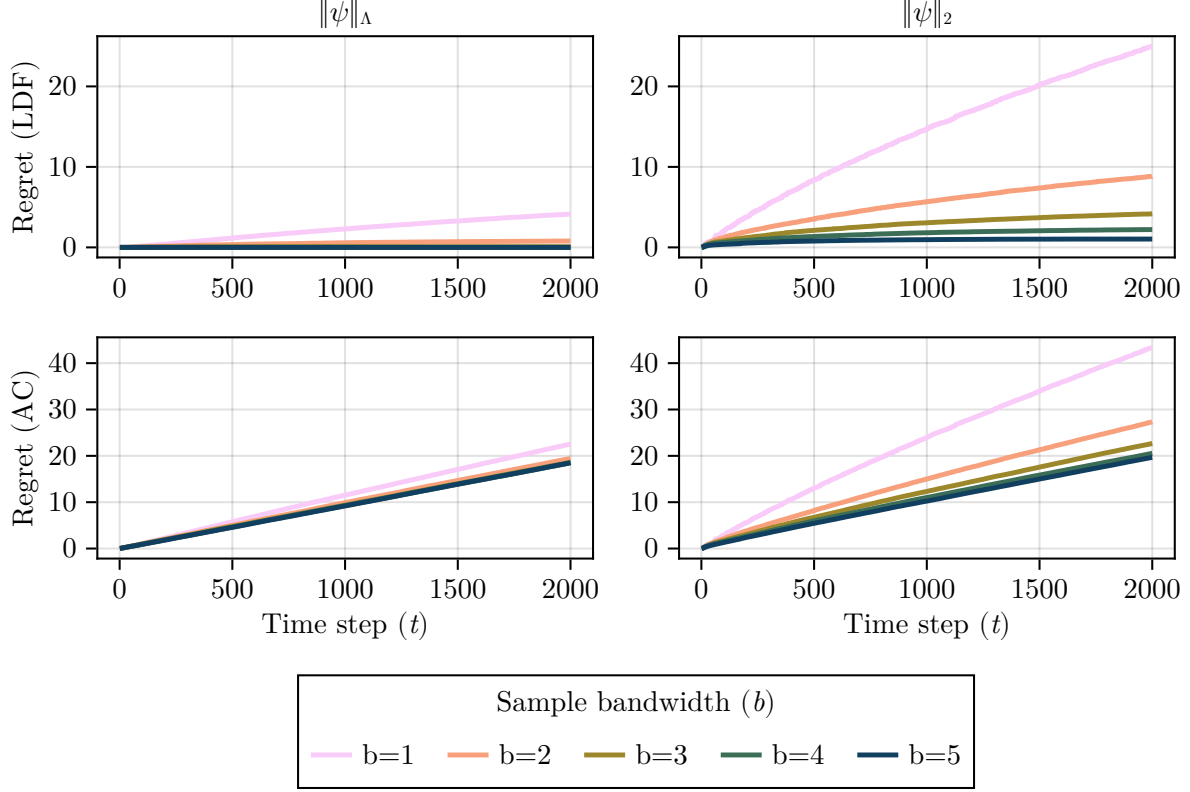


Figure 2.3: Performance with random, non-unity power factors. The plots show regret relative to the LinDistFlow and AC optimal sampling vs. time with spectral (left) and ℓ_2 (right) regularization.

2.6 Discussion and Conclusion

This work proposed a method to selectively sample streaming sensors in distribution networks according to a security criterion—the maximal voltage fluctuations around a nominal value. The key idea is to embed the graphical structure of distribution networks into a spectral bandit algorithm, promoting an electrically dispersed sampling policy.

In contrast with previous work that relies on Gaussian assumptions, we used a distribution-agnostic framework to develop a stochastic LinDistFlow model. This enabled us to develop a principled model for the perturbations of nodal voltages around an operating point without knowledge of the specific distribution driving network loads.

Although the greedy submodular maximization approach we used is well known for solving optimization problems over sets—including bandit problems—it has limitations. Chief among these limitations are the limited convergence guarantees and its inherent approximation error in the ob-

jective function. Future work by the authors will explore more modern extreme bandit frameworks, such as the inverse gap weighting approach [36].

Critically, the strongest theoretical assumption we made to simplify our analysis, Assumption 1, did not appear to significantly affect the numerical experiments in Section 2.5. Moreover, while we have presently used synthetic test data, ongoing work by the authors is developing a testbed to investigate the performance of the proposed algorithms on actual networks.

2.7 Proofs of Lemmas

2.7.1 Proof of Lemma 1

Proof. First, note that for any fixed active injection p and known reactive settings $\alpha \in (0, 1]^n$ and $\text{sgn}(q) \in (0, \pm 1)^n$, we can express the corresponding reactive injection $q \in \mathbb{R}^n$ as $q := Kp$, where K is a diagonal matrix with entries that take the form $K_{ii} = \text{sgn}(q_i)\alpha_i^{-1}\sqrt{1 - \alpha_i^2}$ for all $i \in \mathcal{N}$. Then, note that

$$v = v^\bullet \mathbf{1} + Rp + Xq = v^\bullet \mathbf{1} + (R + XK)p \quad (2.16a)$$

$$= v^\bullet \mathbf{1} + A^{-1} \text{diag}(r + \kappa Ix) A^{-\top} p. \quad (2.16b)$$

Now let $L^\dagger := R + XK$, where R, X are inverses of Laplacian matrices. Under Assumption 1, $K_{ii} := \kappa$ for all $i \in \mathcal{N}$, hence $L^\dagger$ is symmetric, as $L^{-\top} = (R^\top + K^\top X^\top) = (R + \kappa IX) = L^\dagger$. As $L^\dagger$ is real and symmetric, it has an orthonormal eigenbasis $W \in \mathbb{R}^{n \times n}$ such that $L^\dagger := W\Lambda^{-1}W^\top$, where $\Lambda^{-1} = \text{diag}(\lambda_1^{-1}, \dots, \lambda_n^{-1})$ is a diagonal matrix of the reciprocals of the Laplacian eigenvalues. $\square$

2.7.2 Proof of Lemma 2

Proof. Let $a : \mathcal{A} \rightarrow \{0, 1\}^n$ map a sampling strategy $\mathcal{S} \in \mathcal{A}$ to a binary vector whose entries take $a_i = 1$ if $i \in \mathcal{S}$ and 0 otherwise. Notice that we can equivalently write $f : \mathcal{A} \rightarrow \mathbb{R}$ for any $\mathcal{S} \in \mathcal{A}$ as

$$f(\mathcal{S}) = \left\| \text{diag}(a(\mathcal{S})) L^\dagger (y - y_i^\bullet) \right\|_\infty, \quad (2.17)$$

where $\|\cdot\|_\infty : \mathbb{R}^n \rightarrow \mathbb{R}_+$ is the ℓ_∞ norm. Now, recall that any norm $\|\cdot\| : \mathbb{R}^n \rightarrow \mathbb{R}_+$ is L -Lipschitz, where L is the smallest $L > 0$ that satisfies $\|x\| \leq L\|x\|_2$ for any $x \in \mathbb{R}^n$. By norm equivalence, we have $\|x\|_\infty \leq \|x\|_2$ for any $x \in \mathbb{R}^n$. We conclude that f is 1-Lipschitz.

Monotonicity clearly holds; for all $\mathcal{S}, \mathfrak{S} \in 2^{\mathcal{N}}$ such that $\mathcal{S} \subseteq \mathfrak{S}$, we have $f(\mathcal{S}) \leq f(\mathfrak{S})$. Furthermore, normalization also follows clearly as $f(\emptyset) = 0$.

Let the centered voltages be $\tilde{v}_i := v_i - \mathbb{E}[v_i]$ for all $i \in \mathcal{N}$. To establish submodularity, we want to show that for all $\mathcal{S} \subseteq \mathfrak{S} \in 2^{\mathcal{N}}$ that

$$f(\mathcal{S}) + f(\mathfrak{S}) \geq f(\mathcal{S} \cap \mathfrak{S}) + f(\mathcal{S} \cup \mathfrak{S}). \quad (2.18)$$

Using the mapping $a : \mathcal{A} \rightarrow \{0, 1\}^n$ as before, we obtain

$$\begin{aligned} f(\mathcal{S} \cap \mathfrak{S}) + f(\mathcal{S} \cup \mathfrak{S}) &= \|(a(\mathcal{S}) \circ a(\mathfrak{S})) \circ \tilde{v}\|_\infty \\ &\quad + \|(a(\mathcal{S}) + a(\mathfrak{S}) - a(\mathcal{S}) \circ a(\mathfrak{S})) \circ \tilde{v}\|_\infty. \end{aligned} \quad (2.19)$$

We complete the proof in cases. First, suppose that $\mathcal{S} \subseteq \mathfrak{S}$; this means $a(\mathcal{S}) \circ a(\mathfrak{S}) \leq a(\mathcal{S})$, where the inequality is elementwise. The inequality is legal because, for all $i \in \mathfrak{S} \setminus \mathcal{S}$, we have $a(\mathcal{S})_i a(\mathfrak{S})_i = 0$. Hence,

$$\begin{aligned} f(\mathcal{S} \cap \mathfrak{S}) + f(\mathcal{S} \cup \mathfrak{S}) &\leq \|a(\mathcal{S}) \circ \tilde{v}\|_\infty + \|a(\mathfrak{S}) \circ \tilde{v}\|_\infty \\ &:= f(\mathcal{S}) + f(\mathfrak{S}). \end{aligned} \quad (2.20)$$

For the second case, take $\mathfrak{S} \subseteq \mathcal{S}$, and note that we can equivalently have $a(\mathcal{S}) \circ a(\mathfrak{S}) \leq a(\mathfrak{S})$ and thereby produce the same bound (2.20). For the final case where $\mathfrak{S} \cap \mathcal{S} = \emptyset$, by applying the triangle inequality, we obtain

$$\begin{aligned} f(\mathcal{S} \cap \mathfrak{S}) + f(\mathcal{S} \cup \mathfrak{S}) &= \|(f(\mathfrak{S}) + f(\mathcal{S})) \tilde{v}\|_\infty \\ &\leq \|f(\mathfrak{S}) \circ \tilde{v}\|_\infty + \|f(\mathcal{S}) \circ \tilde{v}\|_\infty \\ &:= f(\mathcal{S}) + f(\mathfrak{S}). \end{aligned} \quad (2.21)$$

So, we conclude that (2.18) holds for all $\mathcal{S}, \mathfrak{S} \subseteq 2^{\mathcal{N}}$, and therefore, the reward f is submodular. $\square$

2.7.3 Proof of Lemma 3 (sub-Gaussianity of nodal voltages)

Proof. Assumption 2 implies that p is a vector of sub-Gaussian variables with parameter $\frac{1}{2}\Delta$ [37, Ex. 2.4]. For each node $i \in \mathcal{N}$ and any $s \in \mathbb{R}$, the moment-generating function of fluctuations in v_i is conditionally bounded as

$$\mathbb{E} \left[e^{s(v_i - \mathbb{E}[v_i])} \right] = \mathbb{E} \left[\prod_{j=1}^n e^{sW_{ij}\lambda_j^{-1}\rho_j} \right] \quad (2.22a)$$

$$\stackrel{(1)}{=} \prod_{j=1}^n \mathbb{E} \left[e^{sW_{ij}\lambda_j^{-1}\rho_j} \right] \quad (2.22b)$$

$$\stackrel{(2)}{\leq} \prod_{j=1}^n e^{\frac{1}{8}s^2 W_{ij}^2 \lambda_j^{-2} \Delta^2} \quad (2.22c)$$

$$= e^{\frac{1}{8}s^2 (\sum_{j=1}^n W_{ij}^2 \lambda_j^{-2} \Delta^2)}. \quad (2.22d)$$

In the above display, step (1) is by the assumption of independence of the injections, and step (2) is by sub-Gaussianity of ρ with parameter $\frac{1}{2}\Delta$. Thus, by Definition 1 we see that v_i is sub-Gaussian with parameter $\frac{1}{2}\Delta \|\Lambda^{-1}w_i\|_2$, where w_i is the i -th row of the eigenbasis of the Laplacian.

Next, we apply the Cramér-Chernoff bound [21, 37]. For any $s > 0$, by Markov's inequality we obtain

$$\begin{aligned} \Pr([v_i - \mathbb{E}_t[v_i]] > \varepsilon) &\leq \frac{\mathbb{E}\left[e^{s(v_i - \mathbb{E}_t[v_i])}\right]}{s\varepsilon} \\ &\stackrel{(2.22)}{\leq} e^{\frac{1}{8}s^2\Delta^2\|\Lambda^{-1}w_i\|_2^2 - s\varepsilon}. \end{aligned} \quad (2.23)$$

To make the upper bound (2.23) as small as possible, we minimize the exponent with respect to $s > 0$, which yields

$$\inf_{s>0} \frac{1}{8}s^2\Delta^2\|\Lambda^{-1}w_i\|_2^2 - s\varepsilon = \frac{-2\varepsilon^2}{\Delta^2\|\Lambda^{-1}w_i\|_2^2}. \quad (2.24)$$

By including both tails in the bound, for all $i \in \mathcal{N}$ we have

$$\Pr([v_i - \mathbb{E}[v_i]] > \varepsilon) \leq 2 \exp\left\{-\frac{2\varepsilon^2}{\Delta^2\|\Lambda^{-1}w_i\|_2^2}\right\}. \quad (2.25)$$

□

2.7.4 Proof of Theorem 1

We emphasize that the following proof does not rely on independence of the voltage magnitudes.

Proof. To simplify notation, define the centered nodal voltages $\tilde{v}_i := v_i - \mathbb{E}[v_i]$. Applying the union bound yields

$$\Pr([\max_{i \in \mathcal{J}} |\tilde{v}_i| > \varepsilon) = \Pr([\bigcup_{i \in \mathcal{J}} |\tilde{v}_i| > \varepsilon) \quad (2.26a)$$

$$\leq \sum_{i \in \mathcal{J}} \Pr([|\tilde{v}_i| > \varepsilon) \quad (2.26b)$$

$$\stackrel{(2.25)}{\leq} 2b \exp\left\{-\frac{2\varepsilon^2}{\Delta^2 \max_{i \in \mathcal{J}} \|\Lambda^{-1}w_i\|_2^2}\right\}. \quad (2.26c)$$

Therefore, in expectation, we have

$$\mathbb{E}\left[\max_{i \in \mathcal{J}} |\tilde{v}_i|\right] := \int_{\varepsilon=0}^{\infty} \Pr([\max_{i \in \mathcal{J}} |\tilde{v}_i| \geq \varepsilon) d\varepsilon \quad (2.27a)$$

$$\stackrel{(1)}{\leq} c + \int_{\varepsilon=c}^{\infty} \sum_{i \in \mathcal{J}} \Pr([|\tilde{v}_i| > \varepsilon) d\varepsilon \quad (2.27b)$$

$$\stackrel{(2)}{\leq} c + \int_{\varepsilon=c}^{\infty} 2b \exp\left\{-\frac{2\varepsilon^2}{\Delta^2 \max_{i \in \mathcal{J}} \|\Lambda^{-1}w_i\|_2^2}\right\} d\varepsilon, \quad (2.27c)$$

where step (1) follows from the union bound and the fact that the probability of any event is upper bounded by 1, combined with the fact that $\int_0^c \Pr([\cdot]) \max_{i \in \mathcal{S}} |\tilde{v}_i| \geq \varepsilon d\varepsilon \leq c$, and step (2) follows from the tail bound (2.26). Differentiating the upper bound with Leibniz's rule and solving for the minimizing c yields the desired result (2.5) with a factor of $\sqrt{\log(2b)}$, which can be sharpened to $\sqrt{\log(b)}$, as discussed in [38, 2.5.10]. $\square$

2.7.5 Proof sketch for Theorem 2

The following proof for the confidence ellipsoid is a straightforward update to [25, Lemma 20].

Proof. First, we show that there exists a $C > 0$ such that $\|\psi_\star\|_\Lambda \leq C$. Let $v_t := v^\bullet + W\psi_\star + \eta_t$, where $\eta_t := W\Lambda^{-1}W^\top p_t$ and $\psi_\star := \Lambda^{-1}W^\top p_\star$. Observe that

$$\begin{aligned} \|\psi^\star\|_\Lambda &= \sqrt{\tilde{v}^\top W \Lambda W^\top \tilde{v}} \\ &\stackrel{(1)}{=} \sqrt{\sum_{k=1}^n \lambda_k^{-1} |\langle p_\star \rangle w_k|^2} \\ &\stackrel{(2)}{\leq} \|p\|_2 \sqrt{\sum_{k=1}^n \lambda_k^{-1}} \\ &\stackrel{(3)}{\leq} \|\Delta\|_2 \sqrt{\text{Tr } R + XK} := C > 0, \end{aligned}$$

where step (1) is by definition, $v := W\Lambda^{-1}W^\top p$, step (2) is by Cauchy-Schwarz and the fact that $\|w_k\|_2 = 1$ for all k , step (3) is by the boundedness of p and the definition of $\text{Tr } \cdot$, and λ_k is the k -th eigenvalue of L .

Furthermore, note that $p_t \in [\underline{p}_t, \bar{p}_t]^n$ a.s., then by Lemma 3 we obtain that for all $i = 1, \dots, n$ the noise $(\eta_t)_i$ is sub-Gaussian with parameter at most

$$\begin{aligned} \sigma_i &\leq \frac{1}{2} \Delta \|\Lambda^{-1} w_i\|_2 \\ &\stackrel{(1)}{\leq} \frac{1}{2} \Delta \max_{w: \|w\|_2 \leq 1} \|\Lambda^{-1} w\|_2 \\ &\stackrel{(2)}{=} \frac{1}{2} \Delta \|\Lambda^{-1}\|_{\text{op}} \stackrel{(3)}{=} \frac{1}{2} \Delta \|R + XK\|_{\text{op}} := \sigma_\star, \end{aligned}$$

where step (1) is true because $\|w_i\|_2 \leq 1$ for all nodes i , step (2) is by definition of the operator norm, and step (3) is true because $R + XK \succeq 0$ by Assumption 1, hence, $\|R + XK\|_{\text{op}} := \sigma_{\max}(L^\dagger) = (\sigma_{\min}(L))^{-1} = \|\Lambda^{-1}\|_{\text{op}}$, where $\sigma_{\max}(\cdot)$ (resp. $\sigma_{\min}(\cdot)$) denotes the maximum (resp. minimum) singular value.

As $b = 1$, redefine $V_t := \Lambda + \sum_{s=1}^{t-1} w_s w_s^\top$ and let $\xi_t := \sum_{\tau=1}^{t-1} \eta_\tau w_\tau$, where the η_τ 's are independent sub-Gaussian random variables with parameter $\sigma_\star$. We have that for any node $i \in \mathcal{N}$, the voltage

prediction error at time t satisfies

$$|\langle w \rangle \hat{\psi} - \psi^*| = |\langle w \rangle - V_t^{-1} \Lambda^{-1} \psi^* + V_t^{-1} \xi_t| \quad (2.28a)$$

$$\stackrel{(1)}{\leq} \left| \left\langle V_t^{-\frac{1}{2}} w \right\rangle V_t^{-\frac{1}{2}} \Lambda \psi^* \right| + \left| \left\langle V_t^{-\frac{1}{2}} w \right\rangle V_t^{-\frac{1}{2}} \xi_t \right| \quad (2.28b)$$

$$\stackrel{(2)}{\leq} \|w\|_{V_t^{-1}} \left(\|\xi_t\|_{V_t^{-1}} + \|\Lambda^{-1} \psi^*\|_{V_t^{-1}} \right), \quad (2.28c)$$

where steps (1) and (2) are by the triangle inequality and Cauchy-Schwarz, respectively.

By Assumption 3 and Lemma 4, the initial gram matrix is positive definite, $V_0 := \lambda I + \Lambda \succ 0$. We need the following result, which appears in [39, Lemma 11], [25, Lemma 19], and [21, Lemma 19.4].

Lemma 5. *Let $\{w_t\}_{t=1}^m$ be a sequence of rows of W chosen for the sampling strategy. Then,*

$$\sum_{t=1}^m \min \left\{ 1, \|w_t\|_{V_t^{-1}}^2 \right\} \leq 2 \log \left(\frac{\det V_m}{\det V_0} \right). \quad (2.29)$$

We then apply [39, Lemma 9] under our Assumption 2; with probability at least $1 - \delta$,

$$\|\xi\|_{V_t^{-1}} \leq \frac{1}{2} \Delta^2 \|\Lambda^{-1} w_t\|_2^2 \log \left(\frac{\det V_t^{1/2}}{\delta \det \Lambda^{1/2}} \right). \quad (2.30)$$

By [39, Thm. 2], for any $w \in \mathbb{R}^n$ and for all $t \geq 1$, as $\|\psi^*\|_{\Lambda} \leq C$, then with probability at least $1 - \delta$ it holds that

$$|\langle w \rangle \hat{\psi}_t - \psi^*| \leq \|w\|_{V_t^{-1}} \underbrace{\left(\sigma_* \sqrt{2 \log \frac{\det V_t^{1/2}}{\delta \det \Lambda^{1/2}}} + C \right)}_{:= C_t}, \quad (2.31)$$

where σ_*, C are as defined in the bounds above. Then, the regret bound for $b = 1$ is found by directly invoking [23, Thm. 1] and Lemma 5. For $b \neq 1$, a $\tilde{\mathcal{O}}(d\sqrt{bm})$ bound can be produced by combining [27, Thm. 1] and [23, Thm. 1], then applying Lemma 2. $\square$

3. Real-time Assessment of Distribution Grid Security through Adaptive Smart Meter Measurements

The rapid expansion of distributed energy resources is heightening uncertainty and variability in distribution system operations, potentially leading to power quality challenges such as voltage magnitude violations and excessive voltage unbalance. Ensuring the dependable and secure operation of distribution grids requires system real-time assessment. However, constraints in sensing, measurement, and communication capabilities within distribution grids result in limited awareness of the system's state. To achieve better real-time estimates of distribution system security, we propose a real-time security assessment based on data from smart meters, which are already prevalent in most distribution grids. Assuming that it is possible to obtain a limited number of voltage magnitude measurements in real time, we design an iterative algorithm to adaptively identify a subset of smart meters whose real-time measurements allow us to certify that all voltage magnitudes remain within bounds. This algorithm iterates between (i) solving optimization problems to determine the worst possible voltage magnitudes, given a limited set of voltage magnitude measurements, and (ii) leveraging the solutions and sensitivity information from these problems to update the measurement set. Numerical tests on the IEEE 123 bus distribution feeder demonstrate that the proposed algorithm consistently identifies and tracks the nodes with the highest and lowest voltage magnitude, even as the load changes over time.

This chapter is based on the following paper:

J. Chen and L.A. Roald, "Real-time Assessment of Distribution Grid Security through Adaptive Smart Meter Measurements," *63rd IEEE Conference on Decision and Control (CDC)*, December 16-19, 2024.

3.1 Introduction

The stochastic fluctuations in power injections from distributed energy resources (DERs), alongside load variability, can induce constraint violations within distribution systems [40]. These violations may culminate in equipment malfunction, failure of electrical components, and, in extreme circumstances, power outages. Ensuring the dependable and secure operation of distribution grids requires system real-time situational awareness, as this enables distribution system operators (DSOs) to implement appropriate measures to uphold the network's security, reliability, and efficiency without violating operating constraints [41]. Consequently, there is a pressing need for the development and implementation of more efficient monitoring tools and strategies to ensure the secure operation of distribution networks. Unfortunately, distribution grids typically have very limited sensing, measurement, and communication capabilities, resulting in limited awareness for the system operator concerning the system state [42]. Consequently, methods that enable real-time security assessment of the power distribution grid with limited information about the system state, has become a focal

point of interest for distribution grid operators.

3.1.1 Literature Review

A fundamental requirement for monitoring distribution grids is the sufficiency and uninterrupted availability of measurement data. Considering the practical scarcity of such data in distribution networks, state estimation (SE) arises as a cost-effective alternative to overcome the lack of measurement data [43]. The efficacy of these strategies significantly relies on the volume and accuracy of information, particularly operational data [44]. The introduction of digital relays, microphaser measurement units (μ PMUs), Intelligent Electronic Devices (IEDs), automated feeder switches, voltage regulators, and smart inverters has provided an opportunity to enhance system observability [2]. However, in practice, only a few real-time measurement devices, such as PMUs, have been installed to monitor medium and low-voltage feeders [41], resulting in limited real-time measurements, primarily of current and voltage magnitudes. To address the challenge of limited real-time sensors, prior approaches to distribution system state estimation (DSSE) have incorporated load forecasts as pseudo-measurements [45, 46]. However, it has been demonstrated in [42] that the utilization of forecasted or pseudo-measurements in real-time can compromise estimation performance. Furthermore, due to the increased penetration of renewable energy resources, highly variable loads in the power grid, and the integration of electric vehicles (EVs) into the grid, conventional DSSE methods encounter challenges in effectively managing heightened uncertainty and monitoring dynamic changes and topology configurations [47, 48]. Additionally, conventional DSSE approaches, relying on parameters and models of distribution networks, are complex, time-consuming, and highly sensitive to initial conditions [49]. Addressing these challenges has prompted the emergence of intelligent, data-driven, and model-free state estimation algorithms, leveraging artificial intelligence (AI) and machine learning (ML) techniques [42, 50, 51]. These advancements aim to reduce computation time and enhance result accuracy [52].

The primary objective of DSSE methods is to attain full observability of the system state. However, managing significant volumes of data poses a considerable challenge. Moreover, conventional wisdom dictates the necessity of deploying PMUs across the distribution network to obtain accurate state estimates. However, the implementation of these devices is associated with considerable costs and may require extensive installations to deliver meaningful returns on investment [45]. In contrast, smart meters possess the capability to sense customers' phases, voltage, current, active and reactive powers, and power factors at frequent intervals. Moreover, smart meters are already widely deployed across most distribution networks, thereby eliminating the need for additional overhead in terms of installation and maintenance. Prior work has suggested that such data could be used as near-real-time measurements for inferring system state [53]. However, at present, limited communication bandwidth poses a challenge in acquiring real-time measurements from all smart meters. Consequently, only a selected subset of smart meter readings can be transmitted to the utility in real-time [41].

In order to alleviate the computational burden in DSSE and account for bandwidth considerations associated with smart meter data, our work, in contrast to DSSE, refrains from ensuring *full observ-*

ability of the distribution system. Instead, our focus is on enhancing distribution system real-time assessment by identifying the minimum and maximum voltage magnitudes in the feeder, which allows us to achieve the more limited goal of certifying *secure operations*. *Secure operation* refers to the situation where we can certify that there are no violations of the voltage constraints, despite significant uncertainty regarding the exact power consumption of most loads [54]. Specifically, we consider the scenario with specified ranges of load variability (i.e. unknown, but bounded real and reactive power consumption) and a limited subset of nodes where voltages are measured and reported in real-time by smart meters, and seek to verify whether there are any voltage violations in the feeder.

Since the choice of the nodes with real-time measurements impact what we know about the system, a specific set of measurement nodes may be unable to certify *secure operations*, while an alternative subset of measurements could potentially ascertain the absence of voltage violations. Consequently, a key contribution of our work is to devise an appropriate algorithm for selecting the measurement subset.

We note that [54] investigates the certification of secure operation using an optimization-based method given limited measurement and control capabilities, but does not propose a systematic method for locating the measurement data. In [55], the sensor placement problem is addressed by formulating a bilevel optimization problem, aiming to identify optimal locations for installing sensors capable of capturing all potential violations of voltage magnitude limits across different operating conditions. In contrast with [55], which chooses sensor locations and then keeps these locations constant, our work seeks to adaptively update the measurement locations (i.e., which smart meter we obtain data from) in real-time based on the current system conditions.

3.1.2 Contributions

In our work, we propose an approach to obtain real-time assessment of distribution grid security through adaptive smart meter measurements. We assume that we can access a subset of voltage magnitude measurements in real-time (with the number of measurements limited by bandwidth constraints in the smart meter communications infrastructure), that we can adaptively choose which measurements to obtain, and that the power consumption at each node is unknown, but bounded. Given these assumptions, we propose an algorithm that iterates between (i) solving optimization problems to certify whether our current measurement set includes the nodes with the highest and lowest voltage magnitudes in the feeder, and (ii) adapting the measurement set given information obtained from the optimization problem. In each iteration, the algorithm provides bounds on the voltage magnitude ranges in the feeder which can be used to assess distribution grid security.

The contributions are summarized as follows:

- 1) We formulate optimization problems to bound the voltage magnitudes in the network by adapting the maximization/minimization framework proposed in [54] to a setting with a limited set of voltage magnitude measurements. The goal of this step is to determine the extreme achievable voltage

magnitudes at each node within the distribution system, given the current set of measurements.

2) Using the solutions and sensitivities obtained from the optimization problems, we propose an algorithm to adjust the measurement set to improve (i.e. narrow) our estimates of the extreme achievable voltage magnitudes.

3) The effectiveness of the proposed algorithm is validated through numerical demonstrations in the provided case study. The results demonstrate that the algorithm can effectively identify the nodes in the system where the highest/lowest voltages occur and certify that there is indeed no other nodes with worse voltage magnitudes.

The remainder of this chapter is organized as follows. Section 3.2 describes the formulation of the secure operation certification problem and provides the approach to update the measurement set to enhance the real-time assessment. In Section 3.3, numerical simulations are conducted to analyze the effectiveness of the proposed method, and conclusions are drawn in Section 3.4.

3.2 Problem Description

In this section, we first adapt the optimization model proposed in [54] to a setting with limited voltage measurements. Then we propose an algorithm to adjust the measurement set to enhance the real-time assessment.

3.2.1 Optimization Problem for Extreme Achievable Voltages

In distribution grids, the power consumption varies due to variability in generation from DERs and load patterns. Here, we assume that the exact generation or load is unknown, but bounded to lie within a range of net active and reactive power injections $\mathbf{p} \in [\underline{\mathbf{p}}, \bar{\mathbf{p}}]$, $\mathbf{q} \in [\underline{\mathbf{q}}, \bar{\mathbf{q}}]$. Here $\mathbf{p}, \mathbf{q}$ represent vectors of real and reactive power injections, while $\underline{\mathbf{p}}, \bar{\mathbf{p}}, \underline{\mathbf{q}},$ and $\bar{\mathbf{q}}$ are vectors of their respective upper and lower bounds.

Due to the bandwidth limitations in the smart meter communications infrastructure, the number of nodes where we have access to the real-time voltage magnitude is limited. We use the set $\mathcal{V}$ to represent the subset of nodes from which we obtain voltage magnitude data in real-time. The number of such nodes $|\mathcal{V}|$ is limited due to the bandwidth, but we can choose which nodes is contained in $\mathcal{V}$.

Our objective is to determine whether there exist a load realization $\mathbf{p}, \mathbf{q}$ in the specified ranges that causes voltage violations at any nodes, given the measured voltage magnitudes at nodes in $\mathcal{V}$. In contrast to DSSE, which estimates the voltage magnitude for each node, our (more limited) objective is to determine the *extreme achievable voltage magnitudes*. Once we have bounds on the extreme achievable voltages, we can compare these values with the upper and lower engineering

limits to ascertain whether there is a potential violation of voltage constraints¹.

The secure operation certification problem is formulated as the following optimization problem, aiming to find the extreme achievable voltage magnitude at all nodes subject to constraints representing the operation of a three-phase distribution system:

$$(P1) \quad \underline{u}_n^\phi = \min_{\mathbf{p}, \mathbf{q}, \mathbf{u}} u_n^\phi \quad \text{or} \quad \bar{u}_n^\phi = \max_{\mathbf{p}, \mathbf{q}, \mathbf{u}} u_n^\phi \quad (3.1a)$$

$$\text{s.t.} \quad (\forall n \in \mathcal{N})$$

$$\underline{p}_k^\phi \leq p_k^\phi \leq \bar{p}_k^\phi, \quad \forall k \in \mathcal{N} \setminus ref, \quad (3.1b)$$

$$\underline{q}_k^\phi \leq q_k^\phi \leq \bar{q}_k^\phi, \quad \forall k \in \mathcal{N} \setminus ref, \quad (3.1c)$$

$$u_i^\phi = u_i^{\text{measure}, \phi}, \quad \forall i \in \mathcal{V}, \quad (\lambda_i^\phi), \quad (3.1d)$$

$$u_{\text{root}}^\phi = u_{\text{ref}}^\phi, \quad (3.1e)$$

$$p_{ij}^\phi = \sum_{k \in N_d(j)} p_{jk}^\phi + p_j^\phi, \quad (3.1f)$$

$$q_{ij}^\phi = \sum_{k \in N_d(j)} q_{jk}^\phi + q_j^\phi, \quad (3.1g)$$

$$u_i^\phi - u_j^\phi = 2 \cdot (r_{ij}^\phi \cdot p_{ij}^\phi + x_{ij}^\phi \cdot q_{ij}^\phi). \quad (3.1h)$$

The objective function (3.1a) minimizes or maximizes the square of voltage magnitude achievable at a particular node $n \in \mathcal{N}$, where $\underline{u}_n^\phi$ and $\bar{u}_n^\phi$ are the minimum and maximum achievable squared voltage magnitude at phase ϕ at node n , u_n^ϕ is the square of voltage magnitude at phase ϕ at node n , $\mathcal{N}$ is the set of all the nodes in the distribution system, and $\mathbf{p}, \mathbf{q}$ and $\mathbf{u}$ are the vectors of three-phase active power injection, reactive power injection and square of voltage magnitude.

Constraints (3.1b) and (3.1c) model the range of variability in the net power injections, where p_k^ϕ and q_k^ϕ are the active power injection, reactive power injection at phase ϕ at all nodes except for the reference node.

Constraint (3.1d) enforces that the square of voltage magnitudes equals the measured value if the smart meter at the respective node is selected to report its data in real-time. Here, $\mathcal{V}$ denotes the subset of nodes where smart meters are chosen to provide real-time measured voltage magnitude to the operator.

Constraints (3.1e) to (3.1h) represent three-phase power flow constraints. The power flow with a

¹ We note that while our focus is primarily on voltage magnitude constraints, which are typically of greatest concern in distribution grid operations [56], a similar methodology could be applied to investigate other factors such as current flow or voltage unbalance.

radial structure can be described using the branch power flow model. We adopt a linear distribution power flow (LinDistFlow) model [57] to construct a linear optimization model. Here, u_{root}^ϕ and u_{ref}^ϕ represent the squared root node voltage magnitudes and the squared reference voltage magnitudes at phase ϕ , while p_{ij}^ϕ and q_{ij}^ϕ are the branch active power flow and reactive power flow on line ij at phase ϕ , respectively. $N_d(j)$ is the set of nodes connected by the lateral branching out from node j .

3.2.2 Criteria for Success

Based on the problem above, we define two criteria for success of our problem.

Successful Certification for Secure Operations

Our first criteria for success is whether we can verify the absence of voltage violations given the range of load variability (3.1b) and (3.1c) and the set of measurement nodes $\mathcal{V}$. We use $u^{\min}$, $u^{\max}$ to represent the lower and upper voltage engineering limits corresponding to secure operating ranges that adhere to power quality and safety requirements. We propose the following criterion to certify the *secure operation*,

$$\textbf{Criterion 1: } u_i^{\phi, \min} \leq \underline{u}_i^\phi \text{ and } \bar{u}_i^\phi \leq u_i^{\phi, \max}, \forall i \in \mathcal{N}. \quad (3.2)$$

If the minimum and maximum achievable voltage magnitudes ($\underline{u}_i^\phi$ and $\bar{u}_i^\phi$) obtained from (P1) fall within the specified engineering limits $u^{\min}$, $u^{\max}$ for all nodes, the power flow solutions associated with all potential power injections within the specified ranges (3.1b) and (3.1c) will adhere to the engineering limits on voltage magnitudes, provided we have real-time measured voltage magnitude data of the subset nodes $\mathcal{V}$. In this case, we can guarantee that the system is operating securely.

Successful Identification of Highest and Lowest Voltages

Our second criteria for success is whether we can verify that the measurement set $\mathcal{V}$ contains the nodes with the highest and lowest voltage magnitudes in the feeder. We consider the problems (P1) with the most extreme achievable voltage magnitudes, i.e. the node with the lowest achievable voltage $\underline{u}_i^\phi$, which we will refer to as $\underline{u}^{\phi, \text{lowest}}$, and the node with the largest achievable voltage magnitude $\bar{u}_i^\phi$, which we will refer to as $\bar{u}^{\phi, \text{largest}}$. We also identify the highest and lowest measured voltage magnitudes $\underline{u}^{\text{measure}}$, $\bar{u}^{\text{measure}}$ among our measured nodes $\mathcal{V}$. We then calculate the differences $\underline{\Delta u}$, $\overline{\Delta u}$ between the highest (and lowest) achievable voltage magnitudes, as identified by the optimization problems, and the highest (and lowest) measured voltage magnitudes,

$$\underline{\Delta u} = \underline{u}^{\text{measure}} - \underline{u}^{\phi, \text{lowest}}, \quad (3.3)$$

$$\overline{\Delta u} = \bar{u}^{\phi, \text{largest}} - \bar{u}^{\text{measure}}. \quad (3.4)$$

If the difference $\underline{\Delta u}$, $\overline{\Delta u}$ is equal or less than a predefined small enough value ε , we can certify that we have found the nodes in the network with the highest and lowest voltage magnitudes (within the

tolerance ϵ). We also know that it is not possible to identify a measurement set $\mathcal{V}$ that would lead to tighter bounds on the minimum and maximum achievable voltage magnitudes $\underline{u}^{\phi, \text{lowest}}$, $\bar{u}^{\phi, \text{largest}}$, as those bounds are already equal to actual measured voltages.

3.2.3 Algorithm for Adapting the Measurement Set

The extreme achievable voltages $\underline{u}_i^{\phi}$ and $\bar{u}_i^{\phi}$ depend on the considered range of load variability (3.1b) and (3.1c) and the subset of nodes $\mathcal{V}$ where we can access real-time voltage magnitudes. While we consider the range of load variability to be an input to our problem, we can choose the set of nodes $\mathcal{V}$ which provides the *narrowest range of extreme achievable voltages*. Specifically, we aim to choose $\mathcal{V}$ in a way that maximizes the lowest achievable voltages $\underline{u}_i^{\phi}$ and minimizes the largest extreme achievable voltage $\bar{u}_i^{\phi}$ to let $\underline{u}_i^{\phi}$, $\bar{u}_i^{\phi}$ become as close as possible to the actual highest and lowest voltage in the system, and ideally make them equal (as outlined in our second criterion for success above).

The challenge in updating the measurement set arises from the fact that we do not know the new voltage measurement until we have measured it. Assuming that our sampling procedure allows us to update the measurement set $\mathcal{V}$ and obtain a new measurement before the load changes significantly (e.g. within a few seconds or a minute), we propose to update $\mathcal{V}$ for the next time step based on the solutions and sensitivity information obtained from our previous solution of (P1).

Our goal is to change our measurement set by “move in” (i.e. adding a measurement to) the node that has the most impact on the values of the lowest and largest extreme achievable voltages ($\underline{u}^{\phi, \text{lowest}}$, $\bar{u}^{\phi, \text{largest}}$) of the problem (P1) and “remove” (i.e. stop taking a measurement from) the node with the least impact. Since the dual variable of a constraint represents the impact of this constraint to the objective function, we use the dual variable λ_i^{ϕ} of the constraint (3.1d) to analyze the impact of the measured nodes on the values of extreme achievable voltages ($\underline{u}_i^{\phi}$ and $\bar{u}_i^{\phi}$).

Metrics to assess the value of a measured node

We define two metrics to help us assess the value (i.e. the information provided by) different nodes in our measurement set. For each measured node, we define:

Metric I: We consider the problem (P1) with has the lowest $\underline{u}_i^{\phi}$ and largest $\bar{u}_i^{\phi}$ (i.e. the problems that provide $\underline{u}^{\phi, \text{lowest}}$ and $\bar{u}^{\phi, \text{largest}}$). We take the absolute value of the dual variable of constraint (1d) λ_i^{ϕ} for each measured node $i \in \mathcal{V}$ as our Metric I.

$$\underline{\lambda}^I = \left\{ |\lambda_{i,k}^{\phi}| \mid \forall i \in \mathcal{V}, \phi \in \{a, b, c\} \right\}, \quad (3.5)$$

$$\bar{\lambda}^I = \left\{ |\lambda_{i,\ell}^{\phi}| \mid \forall i \in \mathcal{V}, \phi \in \{a, b, c\} \right\}, \quad (3.6)$$

where subscript k refers to the problem (P1) for the node with the lowest achievable voltage and ℓ refers to the problem (P1) for the note with the highest achievable voltage.

Metric II: We consider the average impact of the measured nodes on the extreme achievable voltage magnitude across all problems (P1). However, since most nodes only provide information regarding other, nearby nodes (e.g. nodes on the same branch of the feeder), we introduce a special way of computing the average that only considers nodes for which the impact (as measured by the value of the dual variable associated with constraint (1d)) exceeds a given threshold $\lambda^{\text{threshold}}$. This ensures that we not only choose nodes that are located in the longest feeder branch. We calculate the mean of the absolute value of the dual variable of these scenarios and use it as Metric II as follows,

$$\lambda^{\text{II}} = \left\{ \text{mean}|\lambda_{i,k}^{\phi}| \text{ s.t. } |\lambda_{i,k}^{\phi}| \geq \lambda^{\text{threshold}}, \forall i \in \mathcal{V}, \forall k \in \mathcal{N}, \phi \in \{a, b, c\} \right\} \quad (3.7)$$

Avoiding recursive loops

One important challenge of our approach is to prevent recursive loops, where nodes are switched in a cyclic manner (e.g., switching from node A to node B in iteration 1 and then reverting from node B to node A in iteration 2, and so forth). To address this, we implement a *loop check* after we choose a new measurement set $\hat{\mathcal{V}}$. We define *the existence of a loop* if the new measurement set is the same as one of the past M measurement sets $\mathcal{V}$, where M is a predefined value. If a loop is identified, we select another new measurement set $\hat{\mathcal{V}}$ to avoid the loop, which will be illustrated in details in the following algorithm.

Iterative Algorithm to Adjust the Measurement Set

At each time step, we use the following algorithm to adjust the measurement set. Assuming that we exchange one node at a time, the goal is to find a new subset $\hat{\mathcal{V}}$ to achieve a larger $\underline{u}^{\phi, \text{lowest}}$ or a lower $\bar{u}^{\phi, \text{largest}}$ to narrow the the extreme achievable range for a better real-time assessment.

Step 1: In the first step, we solve the secure operation certification problem (P1) for all the nodes and phases. Check if **Criterion 1** is satisfied; if yes, we can guarantee that the system is operating securely.

Step 2: In second step, we check if we need to update the measurement set by checking if $\Delta \underline{u} \leq \varepsilon, \Delta \bar{u} \leq \varepsilon$. If these conditions are satisfied, we have already identified the nodes with the highest lowest voltage, and it is not possible to further improve the set of measurements to achieve a tighter range. In this case, we keep the measurement set the same and move back to Step 1 once we receive new measurements. Moreover, if **Criterion 1** is not satisfied at this time, we can confirm that the system is operating insecurely.

If the differences $\Delta \underline{u}, \Delta \bar{u}$ are larger than ε , it may be possible to achieve a better bounds on $\underline{u}^{\phi, \text{lowest}}$ and $\bar{u}^{\phi, \text{largest}}$, by updating the measurement set. In this case, we continue to step 3.

Step 3 (Identify “Move in” node): Because we want to know if the node with the lowest or largest extreme achievable voltage value $\underline{u}^{\phi, \text{lowest}}, \bar{u}^{\phi, \text{largest}}$ really has such a low or large voltage value, we select this node to be added to the new measurement set $\hat{\mathcal{V}}$ to obtain an exact real-time measurement on this node.

Step 4 (Identify “Remove” node): To identify the node we want to remove, we use Metric I ($\underline{\lambda}^I, \bar{\lambda}^I$) and Metric II (λ^{II}) to choose the node that should be removed. Assuming that number of measurements per phase is fixed, we remove a node from the same phase where we want to add a node (i.e. the phase with the lowest or largest extreme achievable voltage value $\underline{u}^{\phi, \text{lowest}}, \bar{u}^{\phi, \text{largest}}$). Therefore, we compare Metric I and II for nodes in that phase. We follow a hierarchical approach to decide which node to remove.

(i) First, we consider the measured nodes that have the least impact on lowest $\underline{u}^{\phi, \text{lowest}}$ or $\bar{u}_i^{\phi, \text{largest}}$ by looking at Metric I. We compare $\underline{\lambda}^I$ and $\bar{\lambda}^I$ of each measured node and remove the node with the smallest value from $\mathcal{V}$.

(ii) If multiple nodes have the same smallest value for Metric I, we remove the node that has the least average impact. Specifically, we choose the node with the smallest value for Metric II λ^{II} among the nodes that have the the same smallest Metric I, and remove this node from $\mathcal{V}$.

Step 5 (Loop check): Next, we do a *loop check*. If *the loop exists*, according to the criterion defined above, we exclude the previously identified node as a possible “Remove” node and go back to Step 4. In this case, we turn to remove the node with the next smallest Metric value.

If no *loop exists*, we have identified a new measurement set $\hat{\mathcal{V}}$ and return to Step 1 once we obtain the new measurements.

The steps of the algorithm are summarized in the flowchart in Fig. 3.1.

3.3 Numerical Test

3.3.1 Simulation Setup

The proposed approach is tested on a modified unbalanced three-phase 123-bus distribution system [58]. We simulate the load change by multiplying the given load consumption with a factor generated based on the real load data of the oak-park substation located in a suburb Portland, OR. The information of branch parameters and load profiles is available online [59]. Considering the the bandwidth limitations in the smart meter communications infrastructure, we assume for each phase, the number of measurements K is fixed and we can switch one measurement in one phase at one time step. The power flow calculated with the LinDistFlow method is used as the measured value for the measured nodes. We check the loop in past 20 time steps ($M = 20$). The predefined value ε is set to be $1e^{-4}$ and the threshold $\lambda^{\text{threshold}}$ is set to be 0.3. We test the approach with different initial measurement sets, different fixed measurement numbers K , and different load profiles. All simulations were implemented using MATLAB on a personal laptop with an Apple M1 Pro processor and 32 GB of RAM.

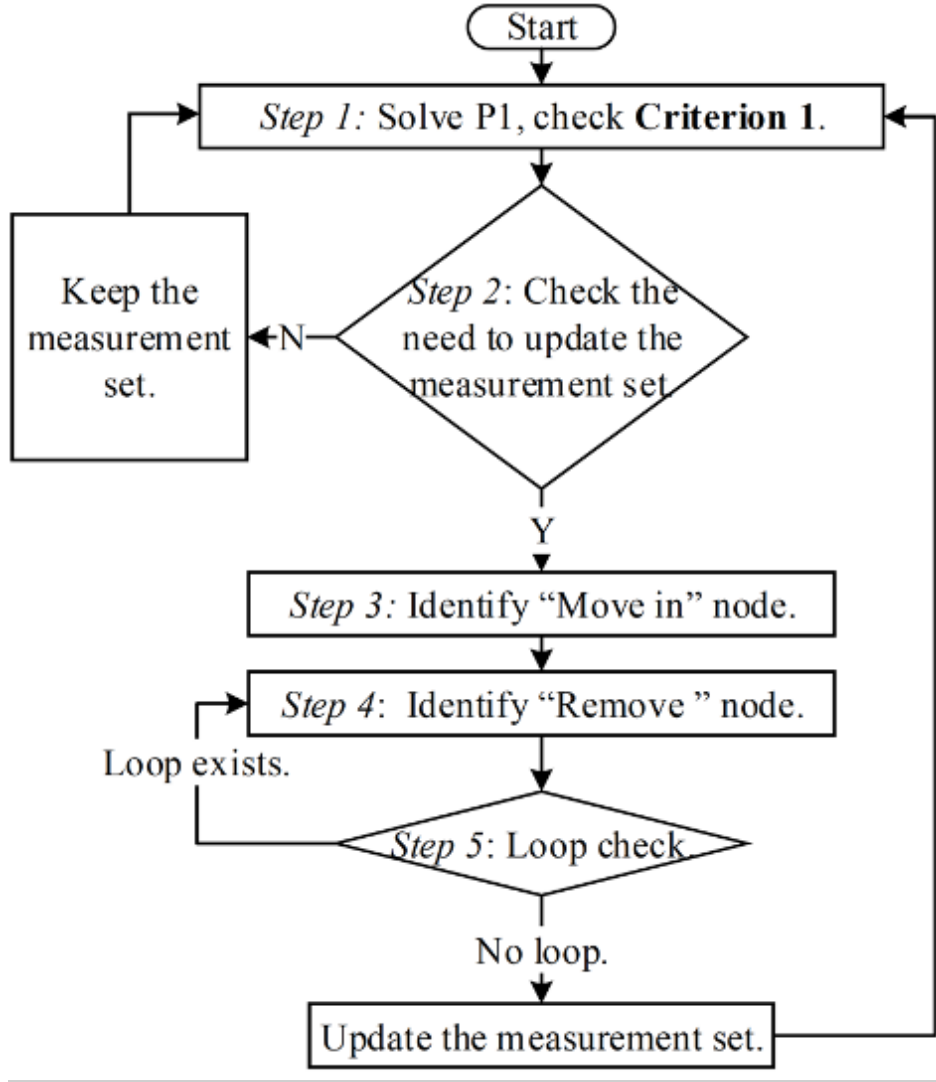


Figure 3.1: Flowchart of the proposed algorithm.

3.3.2 Test Result

First, we test the approach in scenario 1 with the initial measurement set on nodes $\{22, 23, 24, 26, 27, 28, 29, 30, 31, 32\}$ at each phase with the fixed measurement number $K = 10$. The real-time assessment process started at 00 : 00 AM of one day and the time step to update the measurement set is 1 min. Because in scenario 1 there are no distributed generations, we focus the lower bound of the voltage magnitude. Due to the space limitation, we only illustrate the result of the first 30 time steps in Table 3.1. The corresponding factor that multiplies with the given load is illustrated in Fig. 3.2.

In Table 3.1, status 0 represents the difference Δu between the lowest achievable voltage value $u^{\phi, \text{lowest}}$ and the lowest measured voltage value u^{measure} is larger than the predefined value ε , which indicates the need to update the measurement. Status 1 represents the node with the lowest achiev-

Table 3.1: Result of Scenario 1

Time	Status	Lowest Measured Voltage			Lowest Achievable Voltage			Lowest Voltage			Voltage Difference	
		Node	ϕ	u	Node	ϕ	u	Node	ϕ	u	Δu	Δu^{real}
0	0	30,31	a	0.9140	115	a	0.7413	115	a	0.8605	1.73e-1	1.19e-1
1	0	115	a	0.8601	86	c	0.8044	115	a	0.8601	5.57e-2	5.57e-2
2	0	115	a	0.8604	97	b	0.8135	115	a	0.8604	4.69e-2	4.69e-2
3	0	115	a	0.8610	95	a	0.8290	115	a	0.8610	3.20e-2	3.20e-2
4	0	115	a	0.8604	67	c	0.8420	115	a	0.8604	1.84e-2	1.84e-2
5	0	115	a	0.8601	72	a	0.8480	115	a	0.8601	1.21e-2	1.21e-2
6	0	115	a	0.8612	83	a	0.8578	115	a	0.8612	3.41e-3	3.41e-3
7	0	115	a	0.8621	112	a	0.8595	115	a	0.8621	2.62e-3	2.62e-3
8	0	115	a	0.8626	72	a	0.8524	115	a	0.8626	1.02e-2	1.02e-2
9	0	112	a	0.8662	115	a	0.8556	115	a	0.8634	1.06e-2	7.85e-3
10	2	115	a	0.8634	114	a	0.8634	115	a	0.8634	0	0
11	2	115	a	0.8631	114	a	0.8631	115	a	0.8631	0	0
12	0	115	a	0.8639	105	c	0.8626	115	a	0.8639	1.32e-3	1.32e-3
13	2	115	a	0.8639	114	a	0.8639	115	a	0.8639	0	0
14	1	115	a	0.8650	115	a	0.8650	115	a	0.8650	0	0
15	1	115	a	0.8650	115	a	0.8650	115	a	0.8650	0	0
16	1	115	a	0.8655	115	a	0.8655	115	a	0.8655	0	0
17	1	115	a	0.8646	115	a	0.8646	115	a	0.8646	0	0
18	2	115	a	0.8646	114	a	0.8646	115	a	0.8646	0	0
19	1	115	a	0.8651	115	a	0.8651	115	a	0.8651	0	0
20	1	115	a	0.8653	115	a	0.8653	115	a	0.8653	0	0
21	2	115	a	0.8659	114	a	0.8659	115	a	0.8659	0	0
22	2	115	a	0.8663	114	a	0.8663	115	a	0.8663	0	0
23	2	115	a	0.8659	114	a	0.8659	115	a	0.8659	0	0
24	1	115	a	0.8659	115	a	0.8659	115	a	0.8659	0	0
25	1	115	a	0.8653	115	a	0.8653	115	a	0.8653	0	0
26	2	115	a	0.8660	114	a	0.8660	115	a	0.8660	0	0
27	2	115	a	0.8661	114	a	0.8661	115	a	0.8661	0	0
28	2	115	a	0.8665	114	a	0.8665	115	a	0.8665	0	0
29	1	115	a	0.8661	115	a	0.8661	115	a	0.8661	0	0

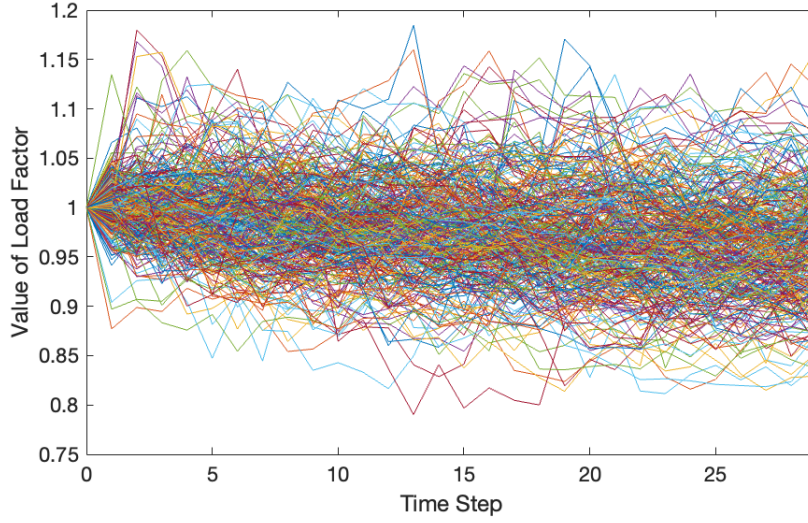


Figure 3.2: Value of load factor on scenario 1.

able voltage value $\underline{u}^{\phi, \text{lowest}}$ is the same as the node with the lowest measured voltage value $\underline{u}^{\text{measure}}$. Status 2 represents the difference $\Delta \underline{u}$ is equal or less than the value ε but the node with the lowest achievable voltage value $\underline{u}^{\phi, \text{lowest}}$ is not the same as the node with the lowest measured voltage value $\underline{u}^{\text{measure}}$. In status 1 and 2, we don't need to update the measurement set. The last two columns illustrate the difference of the square of voltage magnitude, where $\Delta \underline{u}$ represents the difference between the lowest achievable voltage value $\underline{u}^{\phi, \text{lowest}}$ and the lowest measured voltage value $\underline{u}^{\text{measure}}$, $\Delta \underline{u}^{\text{real}}$ represents the difference between the lowest achievable voltage value $\underline{u}^{\phi, \text{lowest}}$ and the system's lowest voltage value $\underline{u}^{\text{real, lowest}}$.

We can observe that on status 1 and 2, when the lowest achievable voltage $\underline{u}^{\phi, \text{lowest}}$ is the same as the lowest measurement voltage value $\underline{u}^{\text{measure}}$ within the tolerance ε ($\Delta \underline{u} \leq \varepsilon$), the lowest achievable voltage $\underline{u}^{\phi, \text{lowest}}$ is also always the same as the system's real lowest voltage $\underline{u}^{\text{real, lowest}}$ within the tolerance ε ($\Delta \underline{u}^{\text{real}} \leq \varepsilon$). This is because the lowest achievable voltage $\underline{u}^{\phi, \text{lowest}}$ is always less than all the system real operating voltage, and if the lowest achievable voltage $\underline{u}^{\phi, \text{lowest}}$ reaches the lowest measured voltage value, which is an actual voltage value, (status 1 or 2), it reaches the lowest real voltage value. Therefore, under status 1 and 2, the system's highest and lowest voltage value and node where it occurs are the same as those of the highest and lowest measured voltage. In this case, we obtain a *good system real-time assessment*.

It can be observed that with the initial measurement set, it takes 11 time steps to find a *good measurement set* where the lowest achievable voltage $\underline{u}^{\phi, \text{lowest}}$ is almost the same as the lowest real voltage value $\underline{u}^{\text{real, lowest}}$ (status 2 at time 10), which indicates a *good system real-time assessment*. The value of the lowest achievable voltage $\underline{u}^{\phi, \text{lowest}}$ shows an overall upward trend during time 0 to time 10. After that, since the load does not change significantly, we don't update the measurement set except at time 12. We observe after we get the *good system real-time assessment* at time 9, for most following time steps, we can always get a *good system real-time assessment* (status 1 or 2).

Moreover, we observe that although at time 12 we don't have a *good system real-time assessment*, we can find a new *good measurement set* in just one iteration. It is because the initial measurement set at time 12 is better than that at time 0.

To further illustrates the measurement update process, we use Table 3.2 to show the update process from time 12 to time 14. Due to space limitation, we only show the information of the phase where the lowest achievable voltage value $\underline{u}^{\phi, \text{lowest}}$ occurs. It can be observed that at time 12, we solve the problem (P1) with the measurement set of phase c on nodes $\{22, 8, 24, 26, 27, 67, 29, 30, 31, 32\}$, the lowest achievable voltage value $\underline{u}^{\phi, \text{lowest}}$ occurs on node 105 on phase c . The voltage difference Δu is larger than the value ε . Therefore, we move node 105 on phase c into the new measurement set $\hat{\mathcal{V}}$. Next, we compare Metric I on phase c and find the smallest value of Metric I is 0 and occurs on nodes $\{29, 30, 31\}$. Therefore, we compare Metric II of nodes $\{29, 30, 31\}$ and find node 29 has the smallest Metric II value. So we remove node 29 on phase c and check that there no *loop exists*. The updated new measurement set $\hat{\mathcal{V}}$ is nodes $\{22, 8, 24, 26, 27, 67, 29, 30, 31, 32\}$ on phase c . At time 13 and time 14, we take *Step 1* and *Step 2* and find the voltage difference Δu is 0, which is less than the value ε . Therefore, we don't update the measurement set. The difference between the scenarios at time 13 and time 14 is that at time 14, the node with lowest achievable voltage value $\underline{u}^{\phi, \text{lowest}}$ (node 115) is already in the measurement set (status 1).

Table 3.2: Measurement Update Process

Time 12 Phase c				Time 13 Phase a				Time 14 Phase a			
Start Set	Metric I	Metric II	New Set	Start Set	Metric I	Metric II	New Set	Start Set	Metric I	Metric II	New Set
22	0.07740	3.02490	22	22	1.48e-16	5.18140	22	22	0	5.18044	22
86	0.78066	0.81971	86	72	2.66e-17	0.49237	72	72	0	0.49237	72
24	0.04222	2.98695	24	24	2.35e-16	7.13266	24	24	0	7.13392	24
26	0.00304	6.19873	26	95	8.33e-17	0.96107	95	95	0	0.96107	95
27	0.00468	1.04175	27	112	0	0.63394	112	112	0	0.63394	112
67	0.16017	0.56466	67	28	2.78e-17	1.28261	28	28	0	1.28280	28
29	0	0	105	29	6.61e-17	3.09891	29	29	0	3.09936	29
30	0	4.95827	30	83	0	0.87210	83	83	0	0.87210	83
31	0	3.23078	31	31	1.39e-17	1	31	31	0	1	31
32	0.00164242	2.33333	32	115	1	0.48570	115	115	1	0.48570	115
Node of $\underline{u}^{\phi, \text{lowest}}$	Phase of $\underline{u}^{\phi, \text{lowest}}$	$\underline{u}^{\phi, \text{lowest}}$		Node of $\underline{u}^{\phi, \text{lowest}}$	Phase of $\underline{u}^{\phi, \text{lowest}}$	$\underline{u}^{\phi, \text{lowest}}$		Node of $\underline{u}^{\phi, \text{lowest}}$	Phase of $\underline{u}^{\phi, \text{lowest}}$	$\underline{u}^{\phi, \text{lowest}}$	
105	c	0.86256		114	a	0.86394		115	a	0.86498	

Next, we analyze the impact of the size of measurement set, the initial measurement set and the load variation. Table 3.3 illustrate the status of different scenarios. The settings of scenario 2, 3 and 4 are the same as scenario 1 except for the initial measurement set, where the initial measurement sets are nodes $\{7, 12, 31, 40, 52, 72, 84, 97, 105, 115\}$, $\{22, 23, 24, 26, 27, 28, 29\}$, $\{19, 20, 21, 22, 23, 24, 26, 27, 28, 29, 30, 31, 32, 33, 34\}$ on each phase for scenario 2, 3 and 4, respectively. The difference between scenario 5 and scenario 1 is that scenario 5 has more significant load fluctuations. The corresponding factor that multiplies with the given load of scenario 5 is illustrated in Fig. 3.3. It can be observed that with the same number of the measurement, scenario 2 finds the *good system real-time assessment* (status 1 or 2) more quickly with 3 time steps. It is

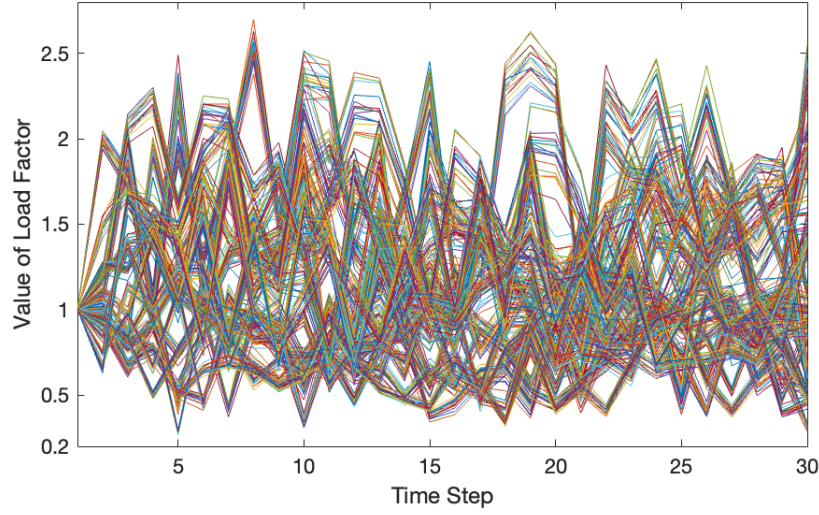


Figure 3.3: Value of load factor on scenario 5.

because its initial measurement set is distributed more evenly so it can get more information of the system to build a better optimization problem (P1). Comparing scenario 1 with scenario 3 and 4, which have 10, 7 and 15 measurements on each phase respectively, we observe that as the number of measurements increases, we can reach status 1 or 2 more quickly. In scenario 5, we observe that although we reach status 1 and find a *good system real-time assessment* at time 9, at time 11 to 14 and time 23 to 24 we need to update the measurement set and cannot get a *good system real-time assessment*. This is because in scenario 5, there are significant changes in the load, and the measurement set obtained from the current time step may not be able to improve the lowest achievable voltage value $\underline{u}^{\phi, \text{lowest}}$ at the next time step. Nevertheless, we observe that during time steps 11 to 14 and time steps 23 to 24, it uses fewer time steps to reach status 1 or 2. This is because at these times, the initial measurement set is better, helping us formulate a better problem (P1).

Table 3.3: Status of Different Scenarios

Time	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
Scenario 1	0	0	0	0	0	0	0	0	0	0	2	2	0	2	1	1	1	1	2	1	1	2	2	2	1	1	2	2	2	1
Scenario 2	0	0	1	1	1	1	1	2	1	1	1	2	2	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Scenario 3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Scenario 4	0	0	1	1	1	1	1	2	1	1	1	2	2	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Scenario 5	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	1	1	1	2	2	1	1	2	0	0	1	1	2	1	2

3.4 Conclusion

In this chapter, we have proposed a real-time assessment approach of distribution power security through adaptive smart meter measurements. Considering the bandwidth limitations in the smart meter communications infrastructure, we formulate an optimization problem to determine the ex-

treme achievable voltage magnitude across all nodes for a setting with limited measurements. We identify criteria under which we can certify that the system is operating securely, or where we can identify that we have found the nodes with the highest and/or lowest voltage magnitudes in the system. Further, we propose an algorithm to adjust the set of smart meters we obtain measurements from based on solutions from our optimization problems. This algorithm is designed to help narrow the range of extreme achievable voltages, and ideally identify the true range of voltages in the system. Numerical analyses were tested using the modified three-phase unbalanced IEEE 123-bus distribution system. The results illustrate that the proposed algorithm can find better bounds on the extreme achievable voltage magnitude and enhance the real-time assessment, even when load fluctuations are present. The algorithm also successfully identifies the node with the most extreme voltage and obtain voltage magnitudes at this node. Moreover, test results illustrate with more measurements and better initial measurement set, it takes fewer time steps to achieve a good real-time assessment. If the load fluctuation is significant, it needs to update the measurement set more frequently to improve the real-time assessment. In future work, we want to investigate how to determine an optimal number of the measurement and how to deal with large load fluctuations.

4. Classifying Reactive Power Control Laws of Behind-the-Meter Solar Photovoltaic Inverters

Various types of reactive power control laws are heterogeneously used for inverter-based resources (IBRs) in distribution grids. As a result, grid operators may not be aware of *which type* of control law is used by a particular IBR. Different control laws imply different voltage support behaviors, which need to be known for the development of accurate computational models for power system analyses. To help mitigate this challenge, this paper develops two *classification* algorithms to identify which type of control law governs the reactive power output of a behind-the-meter solar photovoltaic inverter when the specific control law selected by the IBR owner is unknown. In particular, the two algorithms require only aggregated smart meter measurements to identify candidate reactive power control laws (constant power factor or volt-var control) for distribution network inverters. We present a case study to assess their classification accuracy and evaluate the algorithms' performance in the context of noisy measurements. Our neural network-based classifier is shown to have a higher classification accuracy and performs better under varying levels of noise.

This chapter is based on the following paper:

R. Asiamah, S. Talkington, M. Boateng, M. Vanin, F. Geth, and D.K. Molzahn, "Classifying Reactive Power Control Laws of Behind-the-Meter Solar Photovoltaic Inverters," *6th Kansas Power and Energy Conference (KPEC)*, April 24-25, 2025.

4.1 Introduction

Reactive power control of inverter-based assets is a crucial feature for advanced distribution management systems. Thus, there is significant research interest in the design [60, 61] and analysis [62, 63] of inverter control characteristics, as well as their stability assessment [64]. Historically, solar photovoltaic (PV) inverters were often operated in constant power factor mode (lagging to mitigate voltage rise). However, more recently, piece-wise linear volt-var control with a dead-band has gained popularity, as it is superior at mitigating extreme voltage swings while avoiding unnecessary reactive flows at normal voltages. In practice, standards such as IEEE 1547-2018 [65] provide references for utilities to shape their interconnection requirements (see, e.g., the Hawaiian Rule 14H [66]). In Australia and New Zealand, the AS/NZS 4777.2 standard [67] specifies the requirements for inverters in low-voltage networks.

However, it is common for network operators to be unaware of the control laws selected for individual PV inverters; these are usually set by inverter control designers, installers, and field engineers and are not reported or logged by the distribution utility. The decentralized nature of these activities and responsibilities, in addition to the loss of system information, may ultimately lead to violations of grid connection standards (e.g., voltage requirements). This concern was evidenced in an anal-

ysis by the Australian Energy Market Operator, which showed that only a minority of the many examined inverters comply with the AS/NZS 4777.2 requirements [68].

Poor standard compliance levels can have significant implications, such as increased PV curtailment and reduced network stability limits [68]. Furthermore, the lack of knowledge of inverter control capabilities makes the results of optimal power flow [69] (or any “digital twin” functionality that aims to capture actual network behavior) untrustworthy, often precisely when it matters most, e.g., near the statutory limit of overvoltage. Incorrect assumptions on control settings could impact digital twins similarly to erroneous network models [70].

While it is unclear from public information how often IBR settings are changed in practice, the IEEE 1547-2018 standard mandates that IBRs have the ability to have their control settings “remotely programmed” by their “managing entities” [65]. As this feature becomes deployed in practice, grid operators will need to be aware of evolving control conditions across a portfolio of interconnected IBRs.

4.1.1 Related work

While extensive research has been performed on the data-driven estimation of network properties (impedances, connectivity, etc.), e.g., [71, 72], there is limited work on the identification of inverter control settings. Although increased telemetry in distribution networks facilitates system identification, PV systems are usually installed behind the meter (BTM) and utilities do not always have access to separate PV generation measurements, but only to net grid injection/demand. While this depends on the location and jurisdiction, if only net power measurements are available, disaggregation of the PV injection from the net load is necessary [73]. Furthermore, as these are privately owned assets, operators may lack knowledge of other PV system properties, e.g., their rating, tilt, and azimuth. Estimation of these properties can be beneficial for operators, e.g., to improve generation forecasts [74].

To the authors’ knowledge (and as per the review in [70]), the only prior research paper on inverter control settings estimation is reference [75]. However, [75] assumes prior knowledge of the control law (i.e., constant power factor vs. volt-var), whereas even the latter is often unknown in practice.

4.1.2 Contributions

This paper fills this gap in the literature by developing methods to *classify* the type of reactive power control law that governs the behavior of a BTM inverter-based resource (IBR), particularly focusing on solar photovoltaic inverters. The candidate laws we consider are volt-var (vv) and fixed power factor (pf), the most common reactive power control modes. The data input to these classification methods are 15-minute-granularity advanced metering infrastructure (AMI) data, the most common data available to utilities [76].

We develop two methods: an intuitive method using the nearest centroid classifier and another method using a fully connected neural network (FCNN). Our numerical experiments with varying

noise levels demonstrate both methods' performance. The results in this paper indicate that the neural network-based method achieves significantly higher accuracy.

The remainder of the paper is organized as follows. First, in Section 4.2, we propose the two methods for classifying the control laws governing solar inverters based on smart meter data. In Section 6.3, we present and analyze our results. In Section 6.4, we conclude the paper and provide future research directions.

4.2 Methodology

This section introduces two methods for addressing the reactive power control law classification problem. Relying on data-driven techniques, both methods require the availability of synthetic timeseries smart meter data associated with nodes where the governing control laws are correctly labeled as either volt-var, fixed power factor, or no solar inverter present. As we will discuss further in Section 4.3.1, our numerical studies use data from a timeseries simulation of the 1379-node Ckt. 5 model in OpenDSS [77] with 701 solar PV inverters that have a mix of volt-var and fixed power factor control laws.

4.2.1 Nearest Centroid Classifier Method

At a conceptual level, our first proposed method is a clustering-style heuristic that classifies each recorded timeseries power injection profile according to its closeness to the mean profiles ("centroid") for each control law obtained from labeled timeseries data, after appropriate data normalization. This method relies on the assumption that nodes where the net injection is always negative (consistent net power consumption) do not have a solar generator present. While relatively simplistic, this method serves as a baseline for assessing the possible performance improvements that may be achieved via more sophisticated methods, such as our second method based on neural networks discussed later in this section.

The intuition for this clustering heuristic is based on the fact that the measurements obtained for the active and reactive power at a node are significantly impacted by the PV inverter's control law. More specifically, the reactive power output from a PV inverter is determined by the active power when set in fixed power factor control whereas the nodal voltage determines the reactive power output when volt-var control is employed. The relationships between the different quantities are given by the control laws' characteristic curves as discussed in [78, 79].

With this knowledge, we develop simulated, empirical sample estimates of the sample mean profiles for the active and reactive power injections associated with different control laws to estimate expected behaviors. Define $\mathcal{N} = \{1, \dots, N\}$ as the set of nodes for an N -bus system and $\mathcal{T} = \{1, \dots, T\}$ the set of T time periods. Let $P_{n,t}^*$ and $Q_{n,t}^*$ denote the measured values for active and reactive power injections at node $n \in \mathcal{N}$ and time $t \in \mathcal{T}$. To compute sample mean profiles that are meaningful across nodes with differing demands and solar PV inverter sizes, we first normalize the timeseries profiles of active and reactive power injection data, P^* and Q^* , by their maximum

values to obtain normalized power injection profiles $P_{n,t}$ and $Q_{n,t}$:

$$P_{n,t} = \frac{P_{n,t}^*}{\max_{\tau \in \mathcal{T}} P_{n,\tau}^*}, \quad Q_{n,t} = \frac{Q_{n,t}^*}{\max_{\tau \in \mathcal{T}} Q_{n,\tau}^*}, \quad n \in \mathcal{N}, t \in \mathcal{T}.$$

Define the sets of nodes with volt-var and constant power factor control laws and no PV inverter present as $\mathcal{N}_{vv}$, $\mathcal{N}_{pf}$, and $\mathcal{N}_{no}$, respectively. From the normalized data P and Q , we obtain the empirical average of the normalized active and reactive power injections, $\hat{P}$ and $\hat{Q}$, across nodes with a particular control law at each time $t \in \mathcal{T}$:

$$\begin{aligned} \hat{P}_{vv,t} &= \frac{1}{|\mathcal{N}_{vv}|} \sum_{n \in \mathcal{N}_{vv}} P_{n,t}, & \hat{Q}_{vv,t} &= \frac{1}{|\mathcal{N}_{vv}|} \sum_{n \in \mathcal{N}_{vv}} Q_{n,t}, \\ \hat{P}_{pf,t} &= \frac{1}{|\mathcal{N}_{pf}|} \sum_{n \in \mathcal{N}_{pf}} P_{n,t}, & \hat{Q}_{pf,t} &= \frac{1}{|\mathcal{N}_{pf}|} \sum_{n \in \mathcal{N}_{pf}} Q_{n,t}, \\ \hat{P}_{no,t} &= \frac{1}{|\mathcal{N}_{no}|} \sum_{n \in \mathcal{N}_{no}} P_{n,t}, & \hat{Q}_{no,t} &= \frac{1}{|\mathcal{N}_{no}|} \sum_{n \in \mathcal{N}_{no}} Q_{n,t}, \end{aligned}$$

where $|\cdot|$ denotes the number of elements in a set. We thus have three sample mean profiles for active power, $\hat{P}_{pf}$, $\hat{P}_{vv}$, and $\hat{P}_{no}$, and three sets for reactive power, $\hat{Q}_{pf}$, $\hat{Q}_{vv}$, and $\hat{Q}_{no}$. A timeseries simulation of the Ckt. 5 model using OpenDSS [77] yields the sample mean curves shown in Fig. 4.1.

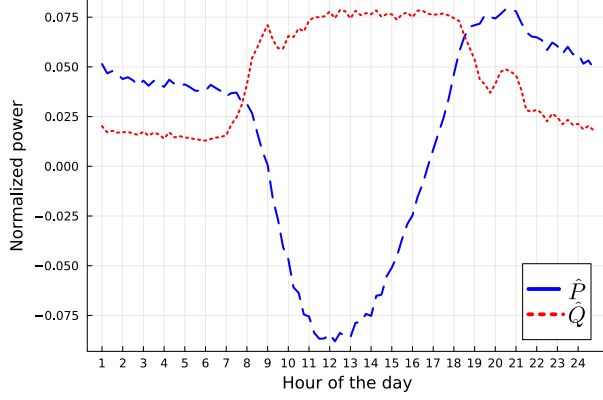
After establishing the mean profile for each of our three cases, we proceed with the classification. Given a timeseries of measured points $P_{n,t}^*$ and $Q_{n,t}^*$, $t \in \mathcal{T}$, from a smart meter at node n , the control law for the associated node is found by calculating the Euclidean norm between the normalized points and the sample mean profiles for the three different cases (vv, pf, and no). The input measurements are sampled over 24 hours at 15-minute granularity, although we note that any sufficiently long period could work. The *nearest centroid* for the solar PV inverter at node $n \in \mathcal{N}$ is obtained by determining which sample mean profile ($\hat{Q}_{vv}$ or $\hat{Q}_{pf}$) has smallest squared Euclidean distance in the reactive power injections across time:

$$\hat{h}_n = \operatorname{argmin}_{h \in \{vv, pf\}} \sum_{t \in \mathcal{T}} (Q_{n,t} - \hat{Q}_{h,t})^2. \quad (4.1)$$

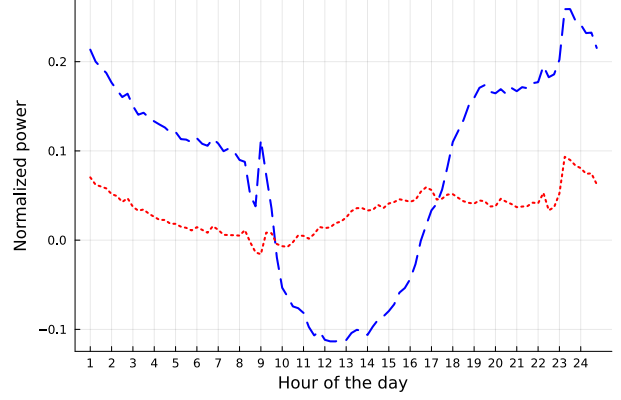
Only computing the distances for Q with respect to volt-var and power factor control is required in (4.1) because we perform a heuristic rejection step for nodes that never achieve a net positive active power injection, i.e., we assume that nodes where the net active power injections are always negative do not have solar PV inverters. This leads to the method shown in Algorithm 2.

4.2.2 Neural Network Based Method

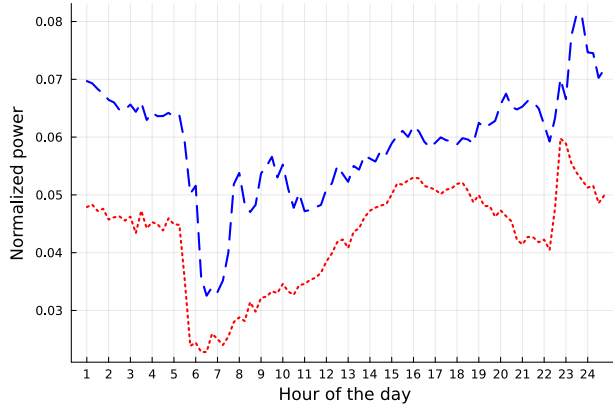
The varying nature of smart meter measurements makes neural-network-based techniques viable candidates for classifying inverter control laws. In a supervised learning approach, we train a fully connected neural network (FCNN) that takes the timeseries of active and reactive power injections



(a) Sample mean profile for volt-var control



(b) Sample mean profile for power factor control



(c) Sample mean profile for no control (no inverter present)

Figure 4.1: Sample mean curve for the different control laws.

and voltage magnitudes at node $n \in \mathcal{N}$ as inputs to estimate a classification for the control law at this node.

The size of the input layer is an $n \times 3$ matrix, where n is the length of the time series measurement fed into the FCNN. The output is a 3×1 vector representing the three possible classes. The output is modeled as a one-hot classifier, which produces 1 for the predicted/classified control law and 0 for the other outputs, as shown in Fig. 4.2. We design the FCNN to have two hidden layers, each with a `ReLU` activation function, while the output has the `softmax` activation function [80]. The number of neurons in the hidden layers is scaled by the size of the input matrix.

4.3 Numerical Results

In this section, we investigate how well the methods described in Section 4.2 perform the classification task and analyze the results obtained.

Algorithm 2 Nearest Centroid Classifier

Require: P^*, Q^* **Ensure:** Classifications for each node

```
1: Normalize:  $P^* \rightarrow P, Q^* \rightarrow Q$ 
2: for all nodes  $n \in \mathcal{N}$  do
3:   if  $\min_{t \in \mathcal{T}} (P_{n,t}) \leq 0$  then
4:     if  $\sum_{t \in \mathcal{T}} (Q_{n,t} - \hat{Q}_{\text{pf},t})^2 \leq \sum_{t \in \mathcal{T}} (Q_{n,t} - \hat{Q}_{\text{vv},t})^2$  then
5:       Classify node  $n$  as power factor controlled
6:     else
7:       Classify node  $n$  as volt-var controlled
8:     end if
9:   else
10:    Classify node  $n$  as no control (no inverter present)
11:   end if
12: end for
```

4.3.1 Setup and Dataset

To evaluate the performance of the methods, we utilized synthetic data obtained from the Ckt. 5 model of OpenDSS [77]. This dataset corresponds to a distribution network of 1379 nodes, 701 ($\approx 50\%$) of which have installed solar PV generators. The dataset includes solar panels with both power factor and volt-var control laws. The generated synthetic data consists of the P , Q , and V measurement values for all nodes and solar panels stemming from power flow calculations at 15-minute intervals. The control laws are correctly labeled for each solar panel, enabling them to be used for all experiments. The labels are distributed as follows: 32.70% nodes with volt-var control, 18.13% nodes with power factor control, and 49.17% nodes with no PV.

The main metric for assessing performance in our results section is prediction accuracy, which is defined as the percentage of nodes for which the control law is correctly classified:

$$\text{accuracy} = \frac{1}{|\mathcal{N}|} \sum_{n \in \mathcal{N}} \mathbb{1}\{\hat{y}_n = y_n\} \times 100 \quad [\%], \quad (4.2)$$

where $\hat{y}_i$ is the predicted label, y_i is the true label, and $\mathbb{1}\{\hat{y}_i = y_i\}$ is the indicator function (1 if $\hat{y}_i = y_i$ and 0 otherwise).

4.3.2 Nearest Centroid Classifier Results

We classify the measurements into three control laws using the algorithm detailed in Algorithm 2. Out of the 1379 nodes in our dataset, we generate our sample mean profiles using 1100 nodes ($\approx 80\%$) and test on the remaining 279 nodes. We refer to these as our training and testing datasets, respectively. For the training dataset, we utilize all the data points to generate the sample mean daily profiles for the active and reactive power. The testing is done by applying Algorithm 2 to

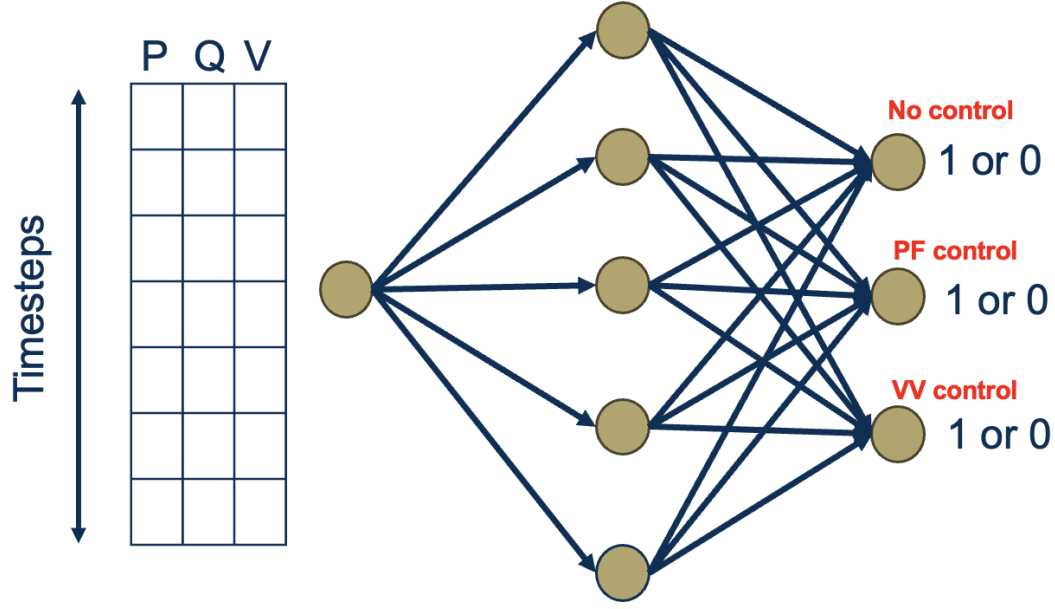


Figure 4.2: The architecture of the fully connected neural network model for classifying a solar inverter's control law.

day-long samples for each node in the test dataset, i.e., 96 samples each day.

Due to the seasonal differences in the load profile and changes in solar radiation, we present the results on a month-by-month basis vs. a year-long basis to observe the difference. We do this to investigate the effect of seasonally dependent curves versus implementing a year-long analysis (seasonal effects ignored). The results are given in Table 4.1.

The results show that using a year-long estimate performs better on average, despite some months having higher prediction accuracies than the year-long dataset.

4.3.3 Neural Network-Based Classification Results

The flexibility of neural networks allows for variation in the structure of the inputs. This subsection focuses on experiments to find the FCNN structure in terms of the timeseries length that gives the most accurate results.

Smart meter measurements come in abundance for different electrical quantities. This raises the question of how many data points are “enough” to predict the control laws correctly. To answer this, we trained different FCNNs that receive varying numbers of inputs for differing amounts of timeseries data and applied these neural networks to classify the testing dataset. This allows us to assess how the number of time periods affects the performance of the neural network classifier. The fully connected neural network was built and trained using TensorFlow [81] with the AdaMax Optimizer. For uniformity, the batch size was set to 128, with a dropout of 0.25 and a maximum

Table 4.1: Prediction accuracy of Nearest Centroid classifier

Month	Prediction Accuracy (%)
January	86.18
February	86.37
March	86.68
April	85.17
May	82.99
June	69.73
July	66.73
August	66.83
September	70.91
October	84.62
November	83.49
December	81.50
Monthly Average	79.27
Yearly	83.22

number of epochs capped at 50 for all experiments. Similar to Section 4.3.2, the first 1100 nodes are used to train neural networks, while the rest are reserved as testing datasets.

Fig. 4.3 shows that the accuracy of the neural network-based classification increases with the length of the input time series, which is expected. The steeper increases in accuracy occur between eight and twenty-four time periods, and limited gains are registered after forty-eight time periods (i.e., 12 hours). Using the entire day yields a maximum accuracy of 92.75%. Analyzing the effects of seasonal variation on our neural network classifier will be investigated in our future work.

4.3.4 Comparing Methods

Finally, we present a detailed comparison of the two methods to assess their performance. We analyze the methods' ability to accurately classify the control law used in a particular inverter, given the set of smart meter measurements under the same conditions. We add noise to the measurement to emulate real-world conditions and assess how such measurement errors affect the two methods.

As is conventional in the literature, measurement errors are assumed to be independent and normally distributed, with a standard deviation equal to one-third of the maximum error [82]. The value of the measured quantity with error is usually represented as

$$X \sim \mathcal{N}\left(X^\dagger, \frac{0.005 \times X^\dagger}{3}\right), \quad (4.3)$$

where $X^\dagger$ is the actual value of the quantity and the maximum error is 0.5% of the true value

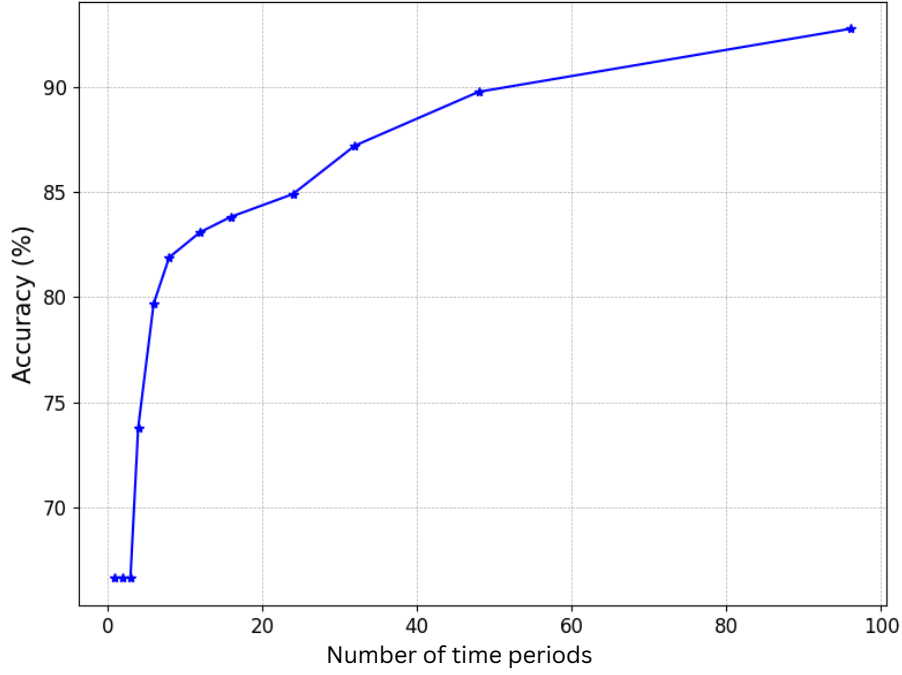


Figure 4.3: Impact of the number of input time periods on the neural network's classification accuracy.

$(0.005 \times X^\dagger)$. Using this formulation for randomized error, we compare the two methods and their ability to make accurate predictions.

The accuracy of the centroid classifier is 82.79%. In contrast, the neural network classifier has a classification accuracy of 92.52%, i.e., a 10% accuracy advantage.

For further analysis, we investigate the estimates in more detail. Tables 4.2(a) and 4.2(b) are two confusion matrices showing the percentage of true positives and false negatives in the predictions for the different methods.

For the centroid-based classifier results in Table 4.2a, we observe a perfect prediction, with no false positives, for nodes without PV inverters. This classifier tends to make more false positive predictions of the volt-var control law, indicating a bias towards this prediction. On the other hand, in addition to being more accurate, the neural network classifier results in Table 4.2b show a relatively more uniform distribution of false predictions for the different control laws. Unlike the centroid classifier, the neural network method had no clear propensity towards any of the control laws.

Another metric for studying the performance of these methods is their receiver operating characteristic curve, which describes the classifier's performance under varying threshold levels [83]. This is only relevant to the neural network classifier. Fig. 4.4 shows the receiver operating characteris-

		Predicted scheme		
		NO	PF	VV
Actual scheme	NO	51.97%	0.00%	0.00%
	PF	0.77%	3.12%	12.60%
	VV	1.20%	2.22%	28.13%

(a) Centroid-based classifier

		Predicted scheme		
		NO	PF	VV
Actual scheme	NO	49.08%	0.02%	0.08%
	PF	0.21%	12.31%	5.71%
	VV	0.42%	4.85%	27.33%

(b) Neural network classifier

Table 4.2: Confusion matrices for the two classification models. Diagonal entries represent correct predictions as a percentage of the total predictions.

tics curve. The area under the curve (AUC) values are all near one, indicating good classification performance.

Finally, we illustrate the robustness of the different methods to varying levels of measurement noise. In (4.3), we model the maximum error possible as 0.5% of the actual value. To assess the performance of these methods in the presence of varying noise, Fig. 4.5 shows the classifier accuracy with maximum errors ranging from 0% to 5%.

The figure shows that the neural network consistently outperforms the centroid classifier. Interestingly, while the accuracies of both methods decrease with increasing measurement noise, the rate of decline of the centroid classifier is steeper, particularly in considerably noisy scenarios ($\geq 3\%$).

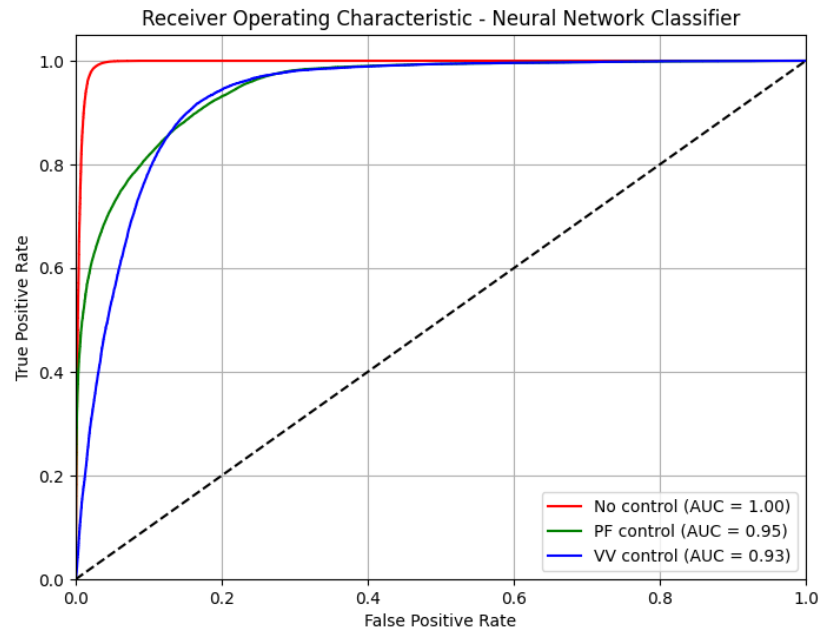


Figure 4.4: Receiver operating characteristic curve for the neural network classifier.

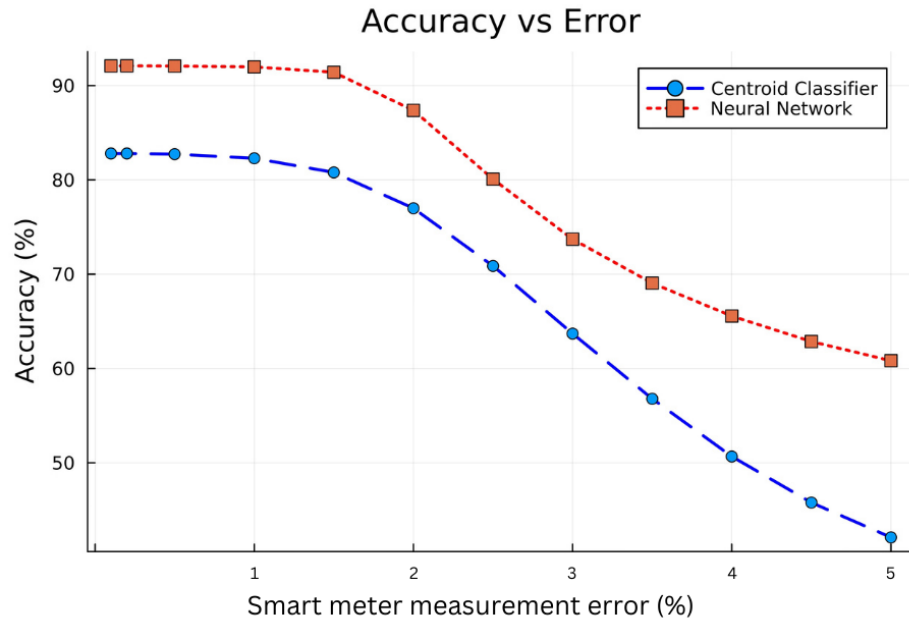


Figure 4.5: Variation of classification accuracy vs. smart meter measurement error.

4.4 Conclusion

As distribution utilities face increasing installations of inverter-based resources, identifying their control laws, which are usually unknown, is essential for network safety and efficient management.

This paper presents two data-driven methods to classify reactive power control laws of solar PV inverters in distribution networks from smart meter measurements. The first method is based on a nearest centroid approach and the second on a neural network. The results show that the methods can accurately classify nodes as having an inverter with constant power factor control, an inverter with volt-var control, or the absence of an inverter. We compare the two methods, showing their performance under varying levels of measurement noise, and investigating their biases towards any of the control laws. Our neural network-based method performs well even with increasing levels of measurement noise and shows no biases towards any of the control laws.

A primary limitation of the work conducted to date is that we have only considered volt-var and fixed power factor control laws. Several other control paradigms, such as volt-watt control and frequency droop control, could be considered in future work.

Furthermore, as this is, to our knowledge, the first paper to address the problem of inverter control law identification, we deem it of interest to develop and test additional methods. A simple potential improvement, for instance, could combine the two methods: using the nearest centroid classifier to detect the nodes with no PV with high accuracy, and then applying a neural network classifier to the remaining datasets.

Acknowledgment

The authors thank Joseph Azzolini for his early assistance with the OpenDSS simulation pipeline and Jaimie Swartz for her insightful comments on an early version of this chapter.

5. Forecast-Aided State Estimation in Unbalanced Distribution Networks Using Smart Meter Data under Limited Communication Bandwidth

With the growing integration of stochastic renewable generation and adaptable resources in electrical distribution systems, distribution utilities are increasingly eager to improve the visibility of their networks using distribution system state estimation (DSSE). However, scarcity of measurements and limited communication bandwidth challenges the ability of the distribution utilities to estimate distribution system states. This chapter presents a forecast-aided real-time state estimation method for distribution networks, using forecasts for nodes lacking direct measurements. While other recent studies have also used forecast-aided state estimation methods, existing approaches require large amounts of historical data to train the forecasting model or depend on phasor measurements, both of which are not easily accessible to distribution utilities. In contrast, we introduce a joint forecasting and state estimation methodology. Our forecasts are generated by a Vector-Autoregressive (VAR) model, which is recursively trained as new measurements are acquired, and thus does not rely on a full set of historical data or phasor measurements. These forecasts, together with the available measurements, are subsequently used for DSSE. We validate the effectiveness of our approach on the IEEE 123-bus benchmark network, taking into account various correlation assumptions and differing quantities of accessible measurements.

This chapter is based on the following paper:

A. Rangarajan, R.K. Gupta, D.K. Molzahn, and L.A. Roald, “Forecast-Aided State Estimation in Unbalanced Distribution Networks Using Smart Meter Data under Limited Communication Bandwidth,” *Sustainable Energy, Grids and Networks*, vol. 43, no. 101863, September 2025. Presented at the *12th IREP Bulk Power system Dynamics and Control Symposium*, June 22-27, 2025.

Abbreviations

DER	Distributed Energy Resources
EV	Electric vehicle
PV	Photovoltaic
DSO	Distribution System Operators
DSSE	Distribution System State Estimation
VAR	Vector Autoregression
YW	Yule-Walker
RLS	Recursive least squares
RMSE	Root mean square error
MAE	Mean absolute error

5.1 Introduction

Distribution networks are constantly changing with the rapid integration of DERs, such as EV chargers and privately owned solar PV installations. This evolution has, on one hand, boosted flexible resources within distribution grids. On the other hand, it has also led to challenges with maintaining adequate voltages and avoiding overloads due to the significant magnitude and stochastic nature of solar PV generation and EV demand. These factors have increased the volatility of power injections within power distribution systems, causing line flow congestion, transformer overloads, voltage imbalance, and overvoltages. As DSOs are required to ensure quality of supply while adhering to the physical limits of the network [84–86], they are increasingly focused on enhancing the visibility of their networks through DSSE. Additionally, the growing number of flexible resources in the distribution system brings increased risks, such as cyber attacks [87]. DSSE provides DSOs with the real-time situational awareness needed to react appropriately when threats arise.

Normally, state estimation involves an overdetermined system of equations where the number of measurements is higher than the number of states to be estimated. However, implementing state estimation in power distribution systems is challenging as the number of available measurements is usually limited. In addition, distribution grids also have many branches and nodes, each with associated states, resulting in an under-determined system of equations. The extensive installation of the smart meters in recent years could fill the gap of measurements as illustrated in [88–90], therefore enhancing the observability of DSSE. However, the key challenge is the communication of the smart meter data to a central server where DSSE is hosted. This issue arises from the limited investment in the communication infrastructure in the distribution systems. Currently, smart meters are typically pinged either monthly for billing purposes or daily [7, 91], with a maximum time resolution of 15 minutes. This limitation is primarily due to the high cost of communication, as noted in [92, 93], with measurements recorded at intervals ranging from 15 minutes to an hour. Therefore, even if there are enough measurements, limitations in the communication bandwidth make it challenging to transmit these measurements at a high sampling rate to the DSO. Thus, the lack of measurements and limited communication bandwidth motivates the development of DSSE methods relying on sparse measurements.

5.1.1 Related Work

Several methods in the literature consider solving the state estimation problem in contexts with limited measurements. Such methods can broadly be categorized into three different approaches. The first approach relies on algorithms for *compressed sensing* and *matrix completion*, where the problem of low observability is addressed by leveraging sparsity properties of the DSSE problem while it remains underdetermined. For example, [94] proposed compressing the measurements based on correlation information among the nodal voltages. The work in [95] assumed that the substation injected current is much higher than the load current, resulting in a sparse problem. References [96] and [97] propose approaches based on matrix completion, which is generally used to estimate missing values in low-rank matrices. The key drawback of these approaches is that they are highly sensitive to measurement noise and require significant computational effort.

The second approach relies on *pseudo-measurements* to address the problem of sparse measurements [98]. Pseudo-measurements specify estimated power consumption or generation based on historical observations. For example, [98] proposes using standard profiles of different consumers based on clustering. Reference [99] uses artificial neural networks to create pseudo-measurements. The key drawback of these methods is that they depend on fixed heuristics that are determined based on past observations and are not updated frequently. Heuristics based on historical data may no longer be valid due to the stochastic nature of injections in distribution systems with high levels of solar PV generation and EV demand and may be outdated particularly fast in areas where the number of solar PV installations and EVs is rapidly changing. Furthermore, the construction of heuristics requires access to historical data, which is not necessarily easy to obtain for technical and privacy reasons.

The third approach, which is the focus of our work, relies on a two-stage framework referred to as *forecast-aided* DSSE. In the first stage, a forecasting model is used for predicting measurement values for the non-observed nodes. In the second stage, a DSSE is solved considering the measurements from the observed nodes and the predicted values of the non-observed nodes, making the DSSE problem overdetermined. References [100–104] propose different approaches to achieve forecast-aided DSSE. Reference [100] uses load forecasts in the state estimation process; however, their motivation was to improve the estimation performance, rather than to tackle the challenge of low observability. References [102] and [103] consider using the forecasting model in a low-observable setting and showed that forecasting schemes can make the DSSE observable. However, the proposed scheme assumes measurements from phasor measurement units (PMUs), which are not usually available at the distribution level. Further, this scheme has a long offline training phase that assumes access to a large set of historical data from all nodes, which is not necessarily realistic. The method in [103] also utilizes a separate forecasting model for each node and ignores the spatial correlation with injections from the other nodes. Other drawbacks of the method in [103] is that they use a Gaussian process offline to train the forecast model and that the forecast model is not updated as new measurements arrive. Such a scheme may fail when distribution injections have substantial real-time variations. The works in [101] and [104] also propose forecast-aided DSSE, but rely on a large number of measurements from PMUs and remote terminal units, which are unavailable in typical distribution systems.

In summary, existing forecast-aided DSSE methods assume the availability of phasor measurements, a large number of historical measurements to train the forecast model, and/or do not account for the spatial correlation between different nodal measurements. We tackle these limitations of existing work as described in the next section.

5.1.2 Contributions

We address the above-mentioned drawbacks of the forecast-aided DSSE by (i) developing a new recursive forecasting model that relies only on existing smart meter measurement and communication infrastructure, and (ii) integrating this forecasting model with the state estimation step. We demonstrate the efficacy of the proposed method on a case study. We describe each contribution in

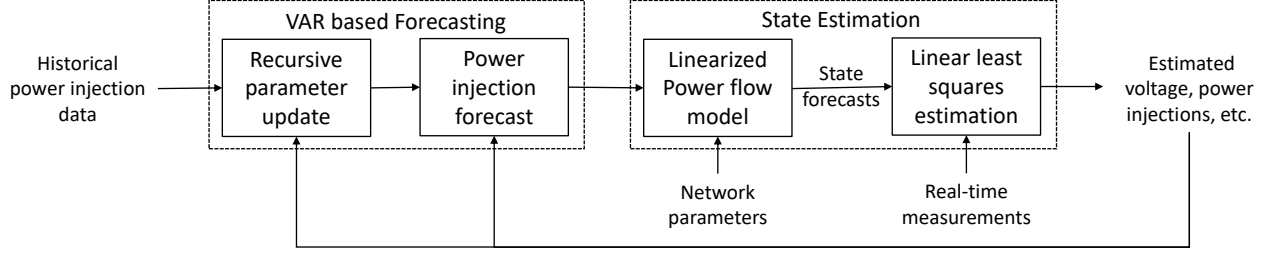


Figure 5.1: Overview of the proposed state estimation framework, partial real time measurements

more detail below.

- **Forecasting model:** We propose a forecasting model that relies only on smart meters measurements that are already present in the grid, thus allowing DSSE without the installation of PMUs, as required by the methods proposed in [102–104]. To comply with existing bandwidth limitations in the communication with smart meters, our models only rely on accessing a few measurements at a time and are updated recursively. We achieve this using a recursive least-square approach, which is computationally inexpensive to train and can be updated in each timestep based on measurements from a small number of locations in the network. As a result, we do not rely on a large amount of historical data to train the forecasting model, as assumed by the previous work in [102–104].
- **State Estimation:** We utilize the above-mentioned forecasts along with smart meter measurements from a limited number of nodes to solve a linear DSSE. The DSSE is formulated as a linear least squares estimator which is based on the linearized Lin3DistFlow approximation of the power flow equations [105–107]. Unlike previous research where the smart meter locations are fixed, the proposed DSSE is capable of reliably estimating the network state with a limited set of smart meter measurements whose locations can be changed dynamically.
- **Validation:** We validate the proposed scheme on the unbalanced IEEE 123-bus benchmark network and show that the proposed forecast-aided approach successfully achieves good estimation performance while directly measuring only 10% of the total states at any time.

The paper is organized as follows. Section 5.2 gives a schematic overview of the proposed approach. Section 5.3 introduces the linearized power flow model, which is later used for state estimation. Section 5.4 presents the forecasting schemes. Section 5.5 describes the state estimation scheme. Section 5.6 presents the numerical setup and results. Finally, Section 5.7 concludes the work.

5.2 Proposed Approach

In this section, we describe the notation used in the paper and provide an overview of the proposed approach.

5.2.1 Notation

We consider a generic unbalanced power distribution network represented by the graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$, where $\mathcal{N}$ is the set of nodes and $\mathcal{E}$ is the set of all distribution lines. $\mathbf{A}$ is the incidence matrix of the graph $\mathcal{G}$ with the column corresponding to the slack bus removed. The nodes in the network are indexed from $1, 2, \dots, N$ where node 1 represents the substation. The three phases are denoted by the superscripts a, b , and c . The lines are indexed from $1, \dots, E$, where $E = N - 1$, as distribution networks are typically operated in a radial manner.

Throughout the paper, we use **boldface** to represent vectors and matrices, while normal fonts are used to represent scalars and elements of vectors/matrices. The vectors $\mathbf{v}$, $\boldsymbol{\theta}$, $\mathbf{P}$, and $\mathbf{Q} \in \mathbb{R}^{3N}$ represent the squared voltage magnitudes, phase angles, and active and reactive power injections, respectively. The vectors $\mathbf{p}$ and $\mathbf{q} \in \mathbb{R}^{3E}$ denote active and reactive power flows along lines. We use single subscripts i and j to represent quantities associated with the corresponding node in the network, e.g., $\mathbf{v}_i$ and $\mathbf{P}_i$ for node i . Double subscripts ij are used to represent quantities associated with the distribution line connecting nodes i and j , e.g., $\mathbf{p}_{ij}$, $\mathbf{q}_{ij}$. Let $\mathbf{I}_n$, $\mathbf{0}_n$, and $\mathbf{1}_n$ be the $n \times n$ identity matrix, the length- n zero vector, and the length- n vector of ones, respectively. The operators $\text{Tr}(\cdot)$ and $\text{vec}(\cdot)$ represent the trace and vectorization of matrices, respectively. The operators $\text{diag}(\cdot)$ and $\text{blkdiag}(\cdot)$ are used to represent diagonal and block-diagonal matrices, respectively. Let $^\top$ and $^{-\top}$ denote the transpose and inverse transpose of a matrix, and $\|\cdot\|$ represents the ℓ_2 -norm of a vector. Finally, $\mathbb{E}(\cdot)$ and $\mathfrak{N}(\cdot)$ represent the expectation operator and normally distributed random variables, respectively.

5.2.2 Overview

Each node in the network is assumed to be equipped with a smart meter capable of measuring voltage magnitudes $\mathbf{v}$ as well as active and reactive power injections $\mathbf{P}$ and $\mathbf{Q}$. However, due to limited communication bandwidth, the DSO can only access measurements from a subset of smart meters, denoted by $\mathcal{M} \subset \mathcal{N}$, at any given time. Given this limitation, the DSO's objective is to estimate the network state, i.e., voltage magnitudes $\mathbf{v}$, phase angles $\boldsymbol{\delta}$, and power injections $\mathbf{P}$ and $\mathbf{Q}$ for all nodes, using the available measurements. This communication bottleneck challenges the DSO's ability to achieve an accurate distribution system state estimate. As only a few measurements are available at any given time, the state estimation problem is underdetermined. Therefore, the key difficulty lies in developing a DSSE approach that can accurately estimate the network state when only a limited number of smart meters are sampled due to bandwidth constraints.

In this chapter, we tackle the challenge of limited bandwidth constraints by developing a forecast-aided DSSE that is capable of reliably estimating the network state using a small number of dynamically sampled smart meter measurements, i.e., the set of sampled smart meters changes at

each time instant. The proposed scheme consists of two stages, as shown in Fig. 5.1. The first stage implements a recursive forecasting scheme based on vector auto-regressive (VAR) model, which is trained using power injection measurements from a time-varying subset of smart meters. The trained VAR model is used to predict power injections at all nodes in the network for the next time step using data from the previous time step. Subsequently, the second stage performs linear least-squares estimation to provide both the voltage magnitudes and phase angles and refine the power injection forecasts. This second step uses both voltage magnitude and power injection measurements from the observed nodes. To account for load variability during the day, the measurements are used to recursively refine the forecast model to keep it updated with the new load information. We model the grid using a linearized approximation of the power flow equations known as the Lin3DistFlow model [105–107], resulting in a linear DSSE. While we use the Lin3DistFlow approximation for state estimation, we evaluate our proposed method by comparing the estimated values to the true states obtained by solving a non-linear AC power flow. The prediction error of the forecasting model and the noise in smart meter measurements are modeled as additive white Gaussian noise with zero mean.

In the following sections, we present the key components of the proposed framework which are the power flow model, the forecasting scheme, and the state estimation formulation.

5.3 Three-phase LinDistFlow

In this section, we describe the linearized three-phase DistFlow approximation (Lin3DistFlow) [105–107] used in our model. To simplify the non-linear DistFlow equations, the Lin3DistFlow approximation assumes that the system is lossless, the voltage angle differences across lines are small, and the voltage magnitudes are close to unity. As a consequence of these assumptions, the voltage phasors are assumed to be almost balanced. Under these assumptions, the voltages at adjacent nodes i and j are related as:

$$\Delta \mathbf{v}_{ij} = \mathbf{v}_i - \mathbf{v}_j = \mathbf{H}_{ij}^p \mathbf{p}_{ij} - \mathbf{H}_{ij}^q \mathbf{q}_{ij}, \quad (5.1a)$$

$$\Delta \delta_{ij} = \delta_i - \delta_j = -\mathbf{H}_{ij}^q \mathbf{q}_{ij} - \mathbf{H}_{ij}^p \mathbf{p}_{ij}, \quad (5.1b)$$

where $\Delta \mathbf{v}_{ij} \in \mathbb{R}^3$ represents the difference of squared voltage magnitudes $\mathbf{v}_i, \mathbf{v}_j \in \mathbb{R}^3$ at nodes i and j . Similarly, $\Delta \delta_{ij} \in \mathbb{R}^3$ represents the difference in phase angles $\delta_i, \delta_j \in \mathbb{R}^3$ at nodes i and j . The quantities $\mathbf{p}_{ij}$ and $\mathbf{q}_{ij}$ are the active and the reactive power flows on the distribution line connecting nodes i and j . Finally, the matrices $\mathbf{H}_{ij}^p, \mathbf{H}_{ij}^q \in \mathbb{R}^{3 \times 3}$ are

$$\mathbf{H}_{ij}^p = \begin{bmatrix} -2r_{ij}^{aa} & r_{ij}^{ab} - \sqrt{3}x_{ij}^{ab} & r_{ij}^{ac} + \sqrt{3}x_{ij}^{ac} \\ r_{ij}^{ba} + \sqrt{3}x_{ij}^{ba} & -2r_{ij}^{bb} & r_{ij}^{bc} - \sqrt{3}x_{ij}^{bc} \\ r_{ij}^{ca} - \sqrt{3}x_{ij}^{ca} & r_{ij}^{cb} - \sqrt{3}x_{ij}^{cb} & -2r_{ij}^{cc} \end{bmatrix}, \quad (5.2a)$$

$$\mathbf{H}_{ij}^q = \begin{bmatrix} -2x_{ij}^{aa} & x_{ij}^{ab} + \sqrt{3}r_{ij}^{ab} & x_{ij}^{ac} - \sqrt{3}r_{ij}^{ac} \\ x_{ij}^{ba} - \sqrt{3}r_{ij}^{ba} & -2x_{ij}^{bb} & x_{ij}^{bc} + \sqrt{3}r_{ij}^{bc} \\ x_{ij}^{ca} + \sqrt{3}r_{ij}^{ca} & x_{ij}^{cb} + \sqrt{3}r_{ij}^{cb} & -2x_{ij}^{cc} \end{bmatrix}. \quad (5.2b)$$

Using (5.1a), we can express the voltage difference $\Delta \mathbf{v} \in \mathbb{R}^{3E}$ across all lines as

$$\begin{aligned}\Delta \mathbf{v} &= \text{vec}(\Delta \mathbf{v}_{ij}), \\ &= \begin{bmatrix} \text{blkdiag}(\mathbf{H}_{ij}^p) & \text{blkdiag}(\mathbf{H}_{ij}^q) \end{bmatrix} \begin{bmatrix} \text{vec}(\mathbf{p}_{ij}) \\ \text{vec}(\mathbf{q}_{ij}) \end{bmatrix}, \\ &= \begin{bmatrix} \text{blkdiag}(\mathbf{H}_{ij}^p) & \text{blkdiag}(\mathbf{H}_{ij}^q) \end{bmatrix} \mathbf{S}_f,\end{aligned}\tag{5.3}$$

where $\mathbf{S}_f = [\text{vec}(\mathbf{p}_{ij})^\top \text{vec}(\mathbf{q}_{ij})^\top]^\top \in \mathbb{R}^{6E}$ is the vector of active and reactive branch flows.

For radial networks, we can express the squared voltage magnitudes $\mathbf{v}$ and the nodal power injections $\mathbf{S}$ as

$$\mathbf{v} = \mathbf{v}_1 \mathbf{1} + \mathbf{A}^{-1} \Delta \mathbf{v},\tag{5.4}$$

$$\mathbf{S} = \mathbf{A}^\top \mathbf{S}_f,\tag{5.5}$$

where $\mathbf{v} = [\mathbf{v}_2^\top, \dots, \mathbf{v}_N^\top]^\top \in \mathbb{R}^{3N}$, $\mathbf{S} = [\mathbf{P}_2^\top, \dots, \mathbf{P}_N^\top, \mathbf{Q}_2^\top, \dots, \mathbf{Q}_N^\top]^\top \in \mathbb{R}^{6N}$, $\mathbf{v}_1$ is the voltage at the substation, and $\mathbf{P}_i$ and $\mathbf{Q}_i$ are the active and reactive power injections at node i , respectively. Combining (5.3), (5.4), and (5.5) results in

$$\mathbf{v} = \mathbf{v}_1 \mathbf{1} + [2\mathbf{R} \quad 2\mathbf{X}] \mathbf{S} = \mathbf{v}_1 \mathbf{1} + \mathbf{M}_v \mathbf{S},\tag{5.6}$$

where $\mathbf{R} = \mathbf{A}^{-1} \text{blkdiag}(\mathbf{H}_{ij}^p) \mathbf{A}^{-\top}$, $\mathbf{X} = -\mathbf{A}^{-1} \text{blkdiag}(\mathbf{H}_{ij}^q) \mathbf{A}^{-\top}$, and $\mathbf{M}_v = [2\mathbf{R} \quad 2\mathbf{X}]$. Similarly, we can derive the following expression for the voltage phase angles δ :

$$\delta = \delta_1 \mathbf{1} + [\mathbf{X} \quad -\mathbf{R}] \mathbf{S} = \delta_1 \mathbf{1} + \mathbf{M}_\delta \mathbf{S},\tag{5.7}$$

where $\delta = [\delta_2, \dots, \delta_N]^\top \in \mathbb{R}^{3N}$, δ_1 is the voltage angle at the substation, and $\mathbf{M}_\delta = [\mathbf{X} \quad -\mathbf{R}]$.

5.4 Forecasting

We next describe the forecasting model used to predict the power injections at every node. The power injection forecasts, in conjunction with the Lin3DistFlow model, are used to forecast the state variables, which include the squared voltage magnitude ($\mathbf{v}$) and the voltage angles (δ), and to perform state estimation, as shown in Fig. 5.1.

In the literature, there are two broad classes of methods used for forecasting load and renewable generation: (a) Machine learning (ML) based methods and (b) statistical time series models like autoregressive processes (AR), autoregressive moving average (ARMA) processes, etc. Recent studies highlight the application of ML-based models to forecast loads and renewable generation [108–111]. While ML-based methods can accurately capture complicated non-linear relationships between loads, these methods often require large datasets for effective training, which makes them computationally expensive. Consequently, ML-based methods are less effective at adapting to short-term variations in power injections.

On the other hand, statistical methods such as ARMA processes are simpler, require less data, and are computationally more efficient. Several studies have highlighted the utility of different statistical methods for short term load and renewable generation forecasts [101, 112–115]. Due to the low computational effort involved in training these statistical models, they can adapt quickly to changes in the network, making them suitable for short-term forecasting. Hence, AR-type models are a better choice in the context of forecast-aided DSSE. In addition to the short-term temporal correlations, power injections also exhibit spatial correlations influenced by geographical and behavioral factors [116]. To capture both spatial and temporal correlations effectively, we selected a vector autoregressive (VAR) model.

5.4.1 Vector Autoregressive Model (VAR)

We next provide a detailed description of the vector autoregressive (VAR) model for predicting nodal power injections and the associated training process. Consider the predicted nodal power injections at time t , $\hat{\mathbf{S}}_t$, described by an order- p VAR(p) model:

$$\hat{\mathbf{S}}_t = \sum_{i=1}^p \mathbf{B}_i \mathbf{S}_{t-i} + \boldsymbol{\varepsilon}_t, \quad (5.8)$$

where $\mathbf{S} = [\mathbf{P}, \mathbf{Q}]^\top \in \mathbb{R}^{6N}$ is the vector of active and reactive power injections, $\mathbf{B}_p, \dots, \mathbf{B}_1 \in \mathbb{R}^{6N \times 6N}$ are the VAR model's coefficient matrices, and $\boldsymbol{\varepsilon}_t \sim \mathcal{N}(0, \boldsymbol{\Sigma}_s)$ is a zero-mean Gaussian noise with a covariance of $\boldsymbol{\Sigma}_s \in \mathbb{R}^{6N \times 6N}$. The temporal correlation in the power injections $\mathbf{S}$ is captured by the diagonal terms of $\mathbf{B}_i$ and the covariance matrix $\boldsymbol{\Sigma}_s$, while the spatial correlations are captured by the off-diagonal terms of $\mathbf{B}_i$ and $\boldsymbol{\Sigma}_s$. While potentially providing improved forecasting accuracy, there are certain drawbacks to using higher-order models. *First*, higher-order models have many more parameters, which increases the amount of data and computational effort required for training. *Second*, higher-order models can result in overfitting, which can reduce the accuracy improvements outside of the training setting. Thus, we select a VAR(1) model, which has been shown to be adequately accurate for many short-term forecasting applications [114, 117]:

$$\hat{\mathbf{S}}_t = \mathbf{B} \mathbf{S}_{t-1} + \boldsymbol{\varepsilon}_t. \quad (5.9)$$

The VAR(1) model, described by (5.9), is used to predict values for the power injections at time step t ($\hat{\mathbf{S}}_t$) based on the power injections at the previous time step ($\mathbf{S}_{t-1}$). While the forecasting model is simple, we face two hurdles: (i) we have to determine appropriate values for the matrix $\mathbf{B}$ and (ii) we do not have access to all the measurements $\mathbf{S}_{t-1}$ from the previous time step. In the following subsection, we will first describe two methods to determine $\mathbf{B}$ while assuming that we could access all measurements simultaneously. Then we will describe our proposed approach for relaxing this requirement.

5.4.2 Estimating the Coefficient and Covariance Matrices

This subsection describes three methods for estimating the coefficient matrix $\mathbf{B}$ and the covariance matrix $\boldsymbol{\Sigma}_s$ for the prediction error $\boldsymbol{\varepsilon}_t$ in (5.9).

Yule-Walker

The Yule-Walker (YW) method is a standard method for estimating the parameters of autoregressive models [118]. Multiplying both sides of (5.9) on the right with $\mathbf{S}_{t-1}^\top$ and taking the expected value yields

$$\mathbf{\Pi}(1) = \mathbf{B}\mathbf{\Pi}(0), \quad (5.10)$$

where $\mathbf{\Pi}(0)$ and $\mathbf{\Pi}(1)$ represent the auto-covariances of $\mathbf{S}$ at lags of 0 and 1 time periods, respectively. These can be estimated using past observations of $\mathbf{S}$ as follows:

$$\hat{\mathbf{\Pi}}(j) = \frac{1}{M} \sum_{i=1}^M \mathbf{S}_i \cdot \mathbf{S}_{i+j}^\top, \quad j \in \{0, 1\}, \quad (5.11)$$

where M is the number of past observations available for training the VAR(1) model. The coefficient matrix $\mathbf{B}$ is estimated by solving the system of equations (5.10). Finally, the forecasting error covariance is computed as

$$\hat{\Sigma}_s = \hat{\mathbf{\Pi}}(0) - \hat{\mathbf{B}}\hat{\mathbf{\Pi}}(1), \quad (5.12)$$

where $\hat{\Sigma}_s$ is the estimated forecasting error covariance and $\hat{\mathbf{B}}$ is the coefficient matrix estimated by solving (5.10) [118].

We highlight that the accuracy of the YW method improves with the number of available observations, M . Additionally, as the network size grows, the number of parameters in the VAR model increases, necessitating a larger dataset to accurately train the model [116]. Furthermore, the YW method assumes that power injection measurements from all nodes are available at every time instant. However, in reality, this assumption is often unrealistic due to the limited availability of sensors and constraints on communication bandwidth. To address this, we use a recursive least-squares algorithm to estimate the parameters of the VAR(1) model.

Recursive Least Squares (RLS)

Consider the VAR(1) model described in (5.9). Rather than utilizing all historical data simultaneously, we can iteratively update our estimate of the matrix $\mathbf{B}$ at each time step, as power injection measurements are collected, using the following iterative update rule:

$$\hat{\mathbf{B}}_t = \hat{\mathbf{B}}_{t-1} + (\mathbf{S}_t - \hat{\mathbf{B}}_{t-1}\mathbf{S}_{t-1})\mathbf{K}_t^\top, \quad (5.13)$$

where $\hat{\mathbf{B}}_t$ is our estimate of the matrix $\mathbf{B}$ at time t . The optimal gain matrix $\mathbf{K}_t \in \mathbb{R}^{6N}$ is found by solving the following optimization problem:

$$\begin{aligned} \min_{\mathbf{K}_t} \sum_{i=1}^k \lambda^{t-i} \|\mathbf{S}_t - \hat{\mathbf{B}}_{t-1}\mathbf{S}_{t-1}\|^2 &\iff \\ \min_{\mathbf{K}_t} \text{Tr} ((\bar{\mathbf{S}}_t - \hat{\mathbf{B}}_{t-1}\bar{\mathbf{S}}_{t-1})\Lambda(\bar{\mathbf{S}}_t - \hat{\mathbf{B}}_{t-1}\bar{\mathbf{S}}_{t-1})^\top), &\end{aligned} \quad (5.14)$$

where the scalar $\lambda \in (0, 1]$ is a forgetting factor, $\mathbf{\Lambda} = \text{diag}(\lambda^{t-1}, \dots, \lambda^{t-i+1}, 1)$, $\bar{\mathbf{S}}_t = [\mathbf{S}_2, \dots, \mathbf{S}_t]$, and $\bar{\mathbf{S}}_{t-1} = [\mathbf{S}_1, \dots, \mathbf{S}_{t-1}]$. The optimization problem (5.14) has a closed form solution that is expressed as

$$\mathbf{K}_t = (\bar{\mathbf{S}}_t \mathbf{\Lambda} \bar{\mathbf{S}}_{t-1}^\top)(\bar{\mathbf{S}}_{t-1} \mathbf{\Lambda} \bar{\mathbf{S}}_{t-1}^\top)^{-1}. \quad (5.15)$$

Using the Sherman-Morrison matrix inversion formula, we can write (5.15) using the following recursion:

$$\mathbf{K}_t = \mathbf{C}_{t-1} \mathbf{S}_t (\lambda + \mathbf{S}_t^\top \mathbf{C}_{t-1} \mathbf{S}_t)^{-1}, \quad (5.16a)$$

$$\mathbf{C}_t = \frac{1}{\lambda} (\mathbf{I}_{6N} - \mathbf{K}_t \mathbf{S}_t^\top) \mathbf{C}_{t-1}, \quad (5.16b)$$

where $\mathbf{C} = (\bar{\mathbf{S}}_{t-1} \mathbf{\Lambda} \bar{\mathbf{S}}_{t-1}^\top)^{-1}$ is the covariance of the estimated matrix $\hat{\mathbf{B}}$. The covariance Σ_s of the prediction error ε is estimated as

$$\hat{\Sigma}_s = \frac{1}{t-1} \sum_{i=1}^t (\tilde{\mathbf{S}}_i - \hat{\mathbf{B}}_i \tilde{\mathbf{S}}_{i-1})(\tilde{\mathbf{S}}_i - \hat{\mathbf{B}}_i \tilde{\mathbf{S}}_{i-1})^\top, \quad (5.17)$$

where $\hat{\Sigma}_s$ is the estimated covariance of the prediction error ε_t . Therefore, $\mathbf{S}_t \sim \mathcal{N}(\boldsymbol{\mu}_s, \hat{\Sigma}_s)$, where $\boldsymbol{\mu}_s = \mathbf{S}_{t-1}$. Selecting an appropriate forgetting factor λ can yield convergence of the RLS algorithm with significantly fewer training data samples. Moreover, RLS provides computational advantages over the YW method. The YW method requires solving the system of equations in (5.10), which can be computationally prohibitive for large systems, and requires storing extensive historical data for accurate training. In contrast, RLS updates parameters iteratively, eliminating the need for large-scale data storage and reducing computational costs.

RLS with Partial Measurements

As formulated above, the RLS method assumes that power injection measurements from all the nodes are available at each time step to perform the iterative update. To address this, we next describe an approach that only accesses measurements from subset of smart meters at each time step. Let $\tilde{\mathbf{S}}_t$ represent the training data sample used to perform the iterative update of the estimated matrix $\hat{\mathbf{B}}$, described by (5.13) and (5.16). To perform this iterative update, we use the power injection measurements at time t for the observed nodes and the previously measured values for the unobserved nodes, i.e.,

$$\tilde{S}_t^m = S_t^m, \quad \forall m \in \mathcal{M}_t, \quad (5.18)$$

$$\tilde{S}_t^m = \tilde{S}_{t-1}^m, \quad \forall m \in \mathcal{N} \setminus \mathcal{M}_t, \quad (5.19)$$

where $\mathcal{M}_t$ is the set of observed nodes at time t and $\tilde{S}_t^m$ and S_t^m represent the entries of the vectors $\tilde{\mathbf{S}}_t$ and $\mathbf{S}_t$ corresponding to node m . Here, we assume that the nodal power injections are in a quasi-steady-state, meaning they change slowly over time. While this assumption may not strictly hold in the presence of DERs, we will empirically demonstrate that the method remains effective under

realistic conditions. The subset of observed nodes changes at each time step, ensuring full network coverage in a finite number of steps. The proposed method does not depend on how the set of observed nodes is chosen at every time step and optimally selecting the nodes to be observed for obtaining the measurements is part of our future work. However, it is important to ensure that the measurements are obtained from all the nodes in the network in the shortest time possible.

5.4.3 Forecasting the System States

Using (5.6) and (5.7), the system states $\mathbf{y} = [\mathbf{v}^\top, \boldsymbol{\theta}^\top, \mathbf{P}^\top, \mathbf{Q}^\top]^\top \in \mathbb{R}^{12N}$, consisting of the squared voltage magnitudes, voltage angles, and power injections, are expressed as

$$\mathbf{y} = \mathbf{y}_1 + \mathbf{M}\mathbf{S}, \quad (5.20)$$

where the matrix $\mathbf{M} = [\mathbf{M}_v^\top \quad \mathbf{M}_\delta^\top \quad \mathbf{I}_{6N}]^\top \in \mathbb{R}^{12N \times 6N}$ and $\mathbf{y}_1 = [\mathbf{v}_1^\top \quad \boldsymbol{\delta}_1^\top \quad \mathbf{0}_{6N}^\top]^\top \in \mathbb{R}^{12N}$. Using (5.20), we express the observed states $\mathbf{y}_o$ as

$$\mathbf{y}_o = \mathbf{Z}_o \mathbf{y} = \mathbf{y}_{1o} + \mathbf{M}_o \mathbf{S}, \quad (5.21)$$

where $\mathbf{Z}_o$ is a permutation matrix composed of the rows of the identity matrix $\mathbf{I}_{12N}$ corresponding to the observed states such that $\mathbf{y}_{1o} = \mathbf{Z}_o \mathbf{y}_1$ and $\mathbf{M}_o = \mathbf{Z}_o \mathbf{M}$. Similarly, we express the unobserved states $\mathbf{y}_u$ as

$$\mathbf{y}_u = \mathbf{Z}_u \mathbf{y} = \mathbf{y}_{1u} + \mathbf{M}_u \mathbf{S}, \quad (5.22)$$

where $\mathbf{Z}_u$ is a permutation matrix composed of the rows of the identity matrix $\mathbf{I}_{12N}$ corresponding to the unobserved states such that $\mathbf{y}_{1u} = \mathbf{Z}_u \mathbf{y}_1$ and $\mathbf{M}_u = \mathbf{Z}_u \mathbf{M}$. Since we do not actually measure the unobserved states, we do not model any measurement noise in the unobserved states.

Combining (5.21), (5.22), and the forecasted power injections $\hat{\mathbf{S}}$ from (5.9), we obtain forecasts of the state variables:

$$\hat{\mathbf{y}}_o = \mathbf{y}_{1o} + \mathbf{M}_o \hat{\mathbf{S}}, \quad (5.23a)$$

$$\hat{\mathbf{y}}_u = \mathbf{y}_{1u} + \mathbf{M}_u \hat{\mathbf{S}}, \quad (5.23b)$$

where $\hat{\mathbf{S}} \sim \mathcal{N}(\boldsymbol{\mu}_s, \hat{\boldsymbol{\Sigma}}_s)$ are the forecasts of the nodal power injections and $\hat{\mathbf{y}}_o$ and $\hat{\mathbf{y}}_u$ are the forecasts of the state at the observed and the unobserved nodes, respectively. Since we assume that the power injection forecasts $\hat{\mathbf{S}}$ are normally distributed and, as described in (5.23), the states and the power injections are linearly related, the forecasted states are also normally distributed, i.e., $\hat{\mathbf{y}}_o \sim \mathcal{N}(\boldsymbol{\mu}_o, \hat{\boldsymbol{\Sigma}}_o)$ and $\hat{\mathbf{y}}_u \sim \mathcal{N}(\boldsymbol{\mu}_u, \hat{\boldsymbol{\Sigma}}_u)$. Their respective means are

$$\boldsymbol{\mu}_o = \mathbb{E}(\hat{\mathbf{y}}_o) = \mathbf{M}_o \boldsymbol{\mu}_s, \quad (5.24a)$$

$$\boldsymbol{\mu}_u = \mathbb{E}(\hat{\mathbf{y}}_u) = \mathbf{M}_u \boldsymbol{\mu}_s. \quad (5.24b)$$

The covariances of $\hat{\mathbf{y}}_o$ and $\hat{\mathbf{y}}_u$, $\hat{\Sigma}_o$ and $\hat{\Sigma}_u$, and the cross-covariance, $\hat{\Sigma}_{uo}$, between $\hat{\mathbf{y}}_o$ and $\hat{\mathbf{y}}_u$ are computed as

$$\hat{\Sigma}_o = \mathbb{E}((\hat{\mathbf{y}}_o - \boldsymbol{\mu}_o)(\hat{\mathbf{y}}_o - \boldsymbol{\mu}_o)^\top) = \mathbf{M}_o \hat{\Sigma}_s \mathbf{M}_o^\top, \quad (5.25a)$$

$$\hat{\Sigma}_u = \mathbb{E}((\hat{\mathbf{y}}_u - \boldsymbol{\mu}_u)(\hat{\mathbf{y}}_u - \boldsymbol{\mu}_u)^\top) = \mathbf{M}_u \hat{\Sigma}_s \mathbf{M}_u^\top, \quad (5.25b)$$

$$\hat{\Sigma}_{uo} = \mathbb{E}((\hat{\mathbf{y}}_u - \boldsymbol{\mu}_u)(\hat{\mathbf{y}}_o - \boldsymbol{\mu}_o)^\top) = \mathbf{M}_u \hat{\Sigma}_s \mathbf{M}_o^\top. \quad (5.25c)$$

The estimated means and the covariances of the forecasts for the observed and unobserved states are used in conjunction with actual measurements to estimate the network states. We model the measurement noise in the observed states as additive Gaussian noise, i.e.,

$$\mathbf{y}_o^m = \mathbf{y}_o + \boldsymbol{\eta}, \quad (5.26)$$

where $\boldsymbol{\eta} \sim \mathcal{N}(0, \boldsymbol{\Omega}_o)$ is the measurement noise with a covariance of $\boldsymbol{\Omega}_o \in \mathbb{R}^{M \times M}$ and $\mathbf{y}_o^m \sim \mathcal{N}(\mathbf{y}_o, \boldsymbol{\Omega}_o)$ is the measured value of the observed states $\mathbf{y}_o$. Using (5.26) and forecasts of the observed states, $\hat{\mathbf{y}}_o$, we forecast the measurements

$$\hat{\mathbf{y}}_o^m = \hat{\mathbf{y}}_o + \boldsymbol{\eta}, \quad (5.27)$$

where $\hat{\mathbf{y}}_o^m \sim \mathcal{N}(\boldsymbol{\mu}_o, \hat{\Sigma}_o + \boldsymbol{\Omega}_o)$ is the forecast of the measurements of the observed states. We assume that the measurement noise $\boldsymbol{\eta}$ and the forecast error $\boldsymbol{\varepsilon}$ are uncorrelated, i.e., $\mathbb{E}(\boldsymbol{\varepsilon} \boldsymbol{\eta}^\top) = \mathbb{E}(\boldsymbol{\eta} \boldsymbol{\varepsilon}^\top) = 0$. However, in practice, this assumption may not hold as the VAR model used for forecasting is trained using power injection measurements, meaning that the accuracy of these measurements directly influences the accuracy of the forecasts. Consequently, any errors in the measurements could propagate into the forecasted states, leading to potential correlation between $\boldsymbol{\eta}$ and $\boldsymbol{\varepsilon}$. However, our numerical assessment shows that, despite the potential correlation in prediction errors and measurements noise, the proposed method performs well over the time periods considered. Furthermore, this motivates the periodic retraining of the VAR model to avoid the accumulation of errors caused due to this assumption. Since the measurement noise and the prediction errors are assumed to be uncorrelated, the covariance between the forecasts of the unobserved states $\hat{\mathbf{y}}_u$ and the measurements $\hat{\mathbf{y}}_o^m$ is also given by (5.25c).

5.5 State Estimation

Building upon the models described in the prior two sections, we next describe the state estimator based on a linear least-squares approach. We use the forecasts of nodal power injections to forecast the system states using the Lin3DistFlow approximation. Then, treating these state forecasts as priors, we incorporate limited smart meter measurements to refine our state estimate by solving a least-squares optimization problem. With the means $\boldsymbol{\mu}$ and covariances $\hat{\Sigma}$ of the state forecasts $\hat{\mathbf{y}}$ estimated in Section 5.4.3, we construct a linear estimator for the unobserved states $\mathbf{y}_u$ of the form $\tilde{\mathbf{y}}_u = \mathbf{L} \mathbf{y}_o^m + \boldsymbol{\rho}$, where $\boldsymbol{\rho}$ is a normally distributed random variable representing the estimation error, by minimizing the squared norm of the estimation residual:

$$\min \mathbb{E}(\|\tilde{\mathbf{y}}_u - \mathbf{y}_u\|^2), \quad (5.28)$$

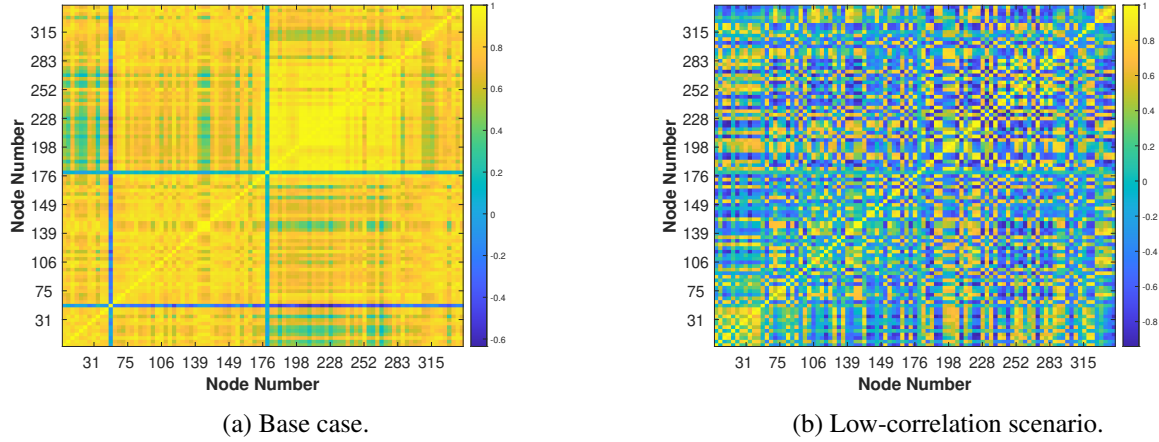


Figure 5.2: Correlation in the active power injections in (a) the base case and (b) the low-correlation scenario.

where $\tilde{\mathbf{y}}_u$ is the vector of estimated states. The optimization problem described in (5.28) has a closed-form solution [119]:

$$\begin{aligned}\tilde{\mathbf{y}}_u &= \mathbb{E}(\mathbf{y}_u | \mathbf{y}_o^m), \\ &= \boldsymbol{\mu}_u + \hat{\boldsymbol{\Sigma}}_{uo} (\hat{\boldsymbol{\Sigma}}_o + \boldsymbol{\Omega}_o)^{-1} (\mathbf{y}_o^m - \boldsymbol{\mu}_o).\end{aligned}\quad (5.29)$$

The proposed estimator uses the statistical information of states computed in Section 5.4.3, which only depend on the forecasted power injections, the network topology, and the line parameters. The estimator given by (5.29) is linear with respect to the measurements and, thus, is computationally efficient and easy to implement. Moreover, this estimator does not require full network observability as the lack of measurements are compensated for by the spatial correlation in power injections estimated by the VAR(1) model. For accurate state estimation, it is important to choose the observed nodes such that $(\hat{\boldsymbol{\Sigma}}_o + \boldsymbol{\Omega}_o)$ is invertible, which will happen as long as the observed nodes are uniformly spread across the system.

While our simulations will show that our proposed method performs well under many different scenarios, a possible drawback of this method as currently implemented is that it might not be robust to outliers, as we have not accounted for other sources of measurement error. Existing machinery for detecting and mitigating outliers (see, e.g., [120]) could be applied in this context, which is a direction for future work.

5.6 Numerical Simulations

In this section, we describe the numerical simulations performed under various scenarios to validate the proposed method. First, we describe the test feeder utilized and the modifications applied to it. Next, we provide details on the load and renewable generation data used, along with the different simulation scenarios. Finally, we present a detailed analysis of the results.

5.6.1 Setup

Test System

The performance of the proposed DSSE method is tested on the IEEE 123-bus distribution test feeder [121], using the network data available at [122]. The feeder includes capacitor banks, voltage regulators, and switches for topology changes. To simplify the modeling requirements, we neglect the capacitor banks, replace all the voltage regulators by short transmission lines of equivalent impedance, and remove all the switches in the network. The nodes connected to the ends of normally closed switches are merged to maintain connectivity, while normally open switches are left open. Extensions of the proposed approach to consider more detailed network models are possible without substantive changes to our proposed methodology.

Load and Renewable Data

We use load profiles based on actual measurements from a substation in Oak Park, Portland, OR [123]. These measurements include a full day of load data sampled at 2 kHz, which is subsequently down-sampled to a resolution of 1 minute. For data on renewable generation, we used actual active power measurements from a rooftop PV system from the EPFL Distributed Electrical Systems Laboratory [124], with data available at [125]. In our simulations, we assume that PV plants with a capacity of 50% of the peak load are installed at every node with non-zero power injection. Using the load and renewable generation data, we obtain the true values of the network states by solving the non-linear AC power flow (ACPF), which are used to evaluate the accuracy of our proposed scheme and to obtain noisy measurements as described in the following sections.

Measurements

We model the measurement noise as unbiased additive Gaussian noise according to the standard defined by the Instrument Transformer (IT) class [126, 127]. For simulations, we consider IT 1.0 which translates to the maximum measurement noise of 1% relative to the true values. Additionally, we assume that only a small set of measurements, as low as 10%, are accessible in real-time at each time step. The measurement set can be updated at each time step to capture measurements from all the nodes in the network. To select the set of observed nodes, we begin by constructing a random permutation $\mathcal{N}_p$ by shuffling the set of nodes $\mathcal{N}$. The set $\mathcal{N}_p$ is then partitioned into subsets of equal size. At each time step, one subset is selected as the measurement set.

Scenarios

To validate the performance of the proposed method under different operating conditions, we consider the following scenarios:

- a) *Base case (BC)*: Each node in the feeder is randomly assigned one of the six available load profiles, which are described in Section 5.6.1. The load profiles are first normalized to have a peak value of one. The load data for each node are then generated by multiplying the node's assigned load profile by the nominal load value specified in the network data. We note that

this results in the load data having a high degree of spatial correlation, as shown in Fig. 5.2a, as may occur if the distribution feeder is spread over a small geographic area.

- b) *Higher renewable penetration (PV)*: To verify the effectiveness of the proposed method in networks with high renewable penetration, we simulate a scenario where each node is equipped with a PV plant with a capacity of up to 50% of its nominal load. In this scenario, we use the same load data as the base case scenario.
- c) *Low correlation (LC)*: Finally, we simulate a scenario with reduced correlation in the load data, as illustrated in Fig. 5.2b, to assess whether the results from the base case are overly influenced by the high correlation in the base case scenario. Since we do not have access to actual load data from many different customers, we generate a synthetic dataset with low correlation by shuffling each load profile by a random amount in time. This captures the underlying patterns in the typical customer behaviors inherent to the measured load data from [123] while modeling customers with offset daily schedules.

For each of the scenarios, we use the root mean square error (RMSE) and the mean absolute error (MAE) to evaluate the performance of the proposed method:

$$RMSE = \sqrt{\frac{1}{N_s} \sum_{i=1}^{N_s} \|\tilde{\mathbf{v}}_t - \mathbf{v}_t\|_2^2}, \quad (5.30)$$

$$MAE = \frac{1}{N_s} \sum_{t=1}^{N_s} \|\tilde{\mathbf{v}}_t - \mathbf{v}_t\|_1, \quad (5.31)$$

where N_s refers to the number of samples and $\tilde{\mathbf{v}}_t$ and $\mathbf{v}_t$ are vectors of the estimated and true voltage magnitudes.

5.6.2 Forecasting Performance

In this section, we evaluate the performance of the VAR(1) model, which, as described by (5.9), predicts the power injections $\mathbf{S}_t$ at the current time step using power injection measurements $\mathbf{S}_{t-1}$ from the previous time step. The VAR(1) model is trained using the Yule-Walker (YW) and the Recursive Least Squares (RLS) algorithms explained in Section 5.4.2, with the YW algorithm providing a baseline. To compare the effectiveness of the YW and RLS algorithms under varying amounts of available data, we consider two training datasets that include either 1000 or 200 time steps. For the YW algorithm, the training dataset is gathered over time and then used simultaneously to estimate the parameters of the VAR(1) model. Conversely, for the RLS algorithm, the dataset size represents the number of time steps required for iteratively training the VAR(1) model. At each time step, the RLS algorithm updates the parameters of the VAR(1) model using measurements only from the current time step. In addition to varying the size of the training dataset in terms of the number of time steps, we also assess how varying the percentage of sampled nodes impacts the RLS algorithm's performance. To obtain a baseline, we first train the VAR(1) model using the YW algorithm with power injection measurements from all nodes. We then train the VAR(1)

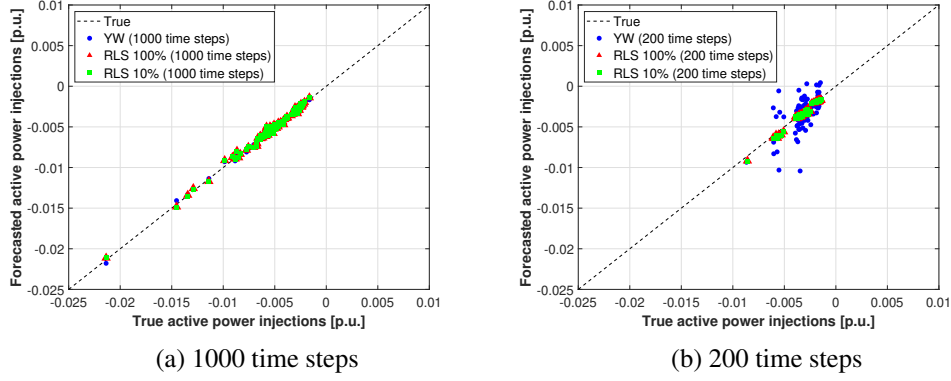


Figure 5.3: Forecast accuracy of the VAR(1) model when trained using (a) 1000 time steps and (b) 200 time steps. In both plots, the dashed black line represents the true nodal power injections. Both methods perform comparably when the VAR model is trained using 1000 time steps of data. However, with 200 samples, the YW method fails to accurately predict the power injections, while the RLS method is more robust.

Table 5.1: Forecasting accuracy with varying training dataset sizes and measurement sets

# Training samples	Method	MAE [p.u.]	RMSE [p.u.]
1000	YW	8.9120×10^{-5}	1.3291×10^{-4}
	RLS 100%	8.2374×10^{-5}	1.2946×10^{-4}
	RLS 10%	9.6501×10^{-5}	1.4401×10^{-4}
200	YW	1.3965×10^{-3}	2.1458×10^{-3}
	RLS 100%	6.7451×10^{-4}	4.4499×10^{-4}
	RLS 10%	7.0130×10^{-4}	4.4706×10^{-4}

model with the RLS algorithm assuming that differing subsets of nodes can be observed at every time step. Specifically, we consider two cases where the operator can access measurements from 100% and 10% of nodes in the network. At every time step, the set of observed nodes is varied according to the procedure described in Section 5.6.1. We evaluate the performance of both the methods via predicting the power injections at the current time step $\mathbf{S}_t$ by assuming that we have perfect knowledge of the power injections at the previous time step $\mathbf{S}_{t-1}$.

Fig. 5.3 and Table 5.1 summarize the forecasting performance of each method with different dataset sizes and percentages of sampled measurements. We first analyze the scenario where data from 1000 time steps are used to train the VAR(1) model. Here, we observe that all the considered methods have comparable prediction accuracy, as shown by Fig. 5.3a. Importantly, the RLS method accurately trains the VAR model even when only 10% of the nodal power injections are accessible at each time step due to bandwidth constraints. The forecasting accuracy of both methods deteriorates when we have access to less training data. We can observe this in Figs. 5.3b when compared to the plots in Figs. 5.3a. In the case of the RLS method with access to 200 samples of data, the RMSE increases by almost a factor three and the MAE increases by almost an order of magnitude,

Table 5.2: Voltage estimation error in the base case

% Measurements	RMSE [p.u.]	MAE [p.u.]
Fixed Measurement set		
10	5.4991×10^{-4}	4.2027×10^{-4}
20	4.9729×10^{-4}	3.8687×10^{-4}
Dynamic Measurement set		
10	5.6587×10^{-4}	4.3026×10^{-4}
20	5.2314×10^{-4}	4.0294×10^{-4}

Table 5.3: Voltage estimation error with PV integration

% Measurements	RMSE [p.u.]	MAE [p.u.]
Fixed Measurement set		
10	5.1374×10^{-4}	3.8790×10^{-4}
20	4.6347×10^{-4}	3.5473×10^{-4}
Dynamic Measurement set		
10	5.2448×10^{-4}	3.9446×10^{-4}
20	4.8866×10^{-4}	3.7037×10^{-4}

as shown in Table 5.1. Despite the degraded performance, the RLS method is, in general, still capable of capturing the trends in power injections with an acceptable accuracy. Additionally, the performance of the RLS method is insensitive to the percentage of measurements accessed in each training iteration. The performance with only 10% of the measurements is only slightly worse than when we have access to all the measurements. On the other hand, with a training dataset of 200 samples, the YW method fails to accurately estimate the parameters of the VAR model, resulting in poor forecasting performance, as shown by Fig. 5.3b. The YW method involves solving (5.10), where Π_0 is computed as a sum of M rank-1 matrices, with M denoting the number of training samples. Therefore, the YW method is only feasible with a sufficiently large M so that Π_0 is not ill-conditioned.

From Fig. 5.3, it is evident that the forecasting accuracy of the RLS method is more sensitive to the size of the training dataset than to the number of sampled measurements. Thus, during the offline training phase, it is more important to train the VAR model for a longer duration, resulting in a more accurate estimate of its parameters $\mathbf{B}$ and Σ_s .

5.6.3 State Estimation Performance

Next, we compare the performance of the proposed state estimation method under the scenarios described in Section 5.6.1, with the results summarized in Fig. 5.4 and Tables 5.2–5.3. In all three scenarios, we train the VAR(1) model using the RLS method with 200 data samples. As shown

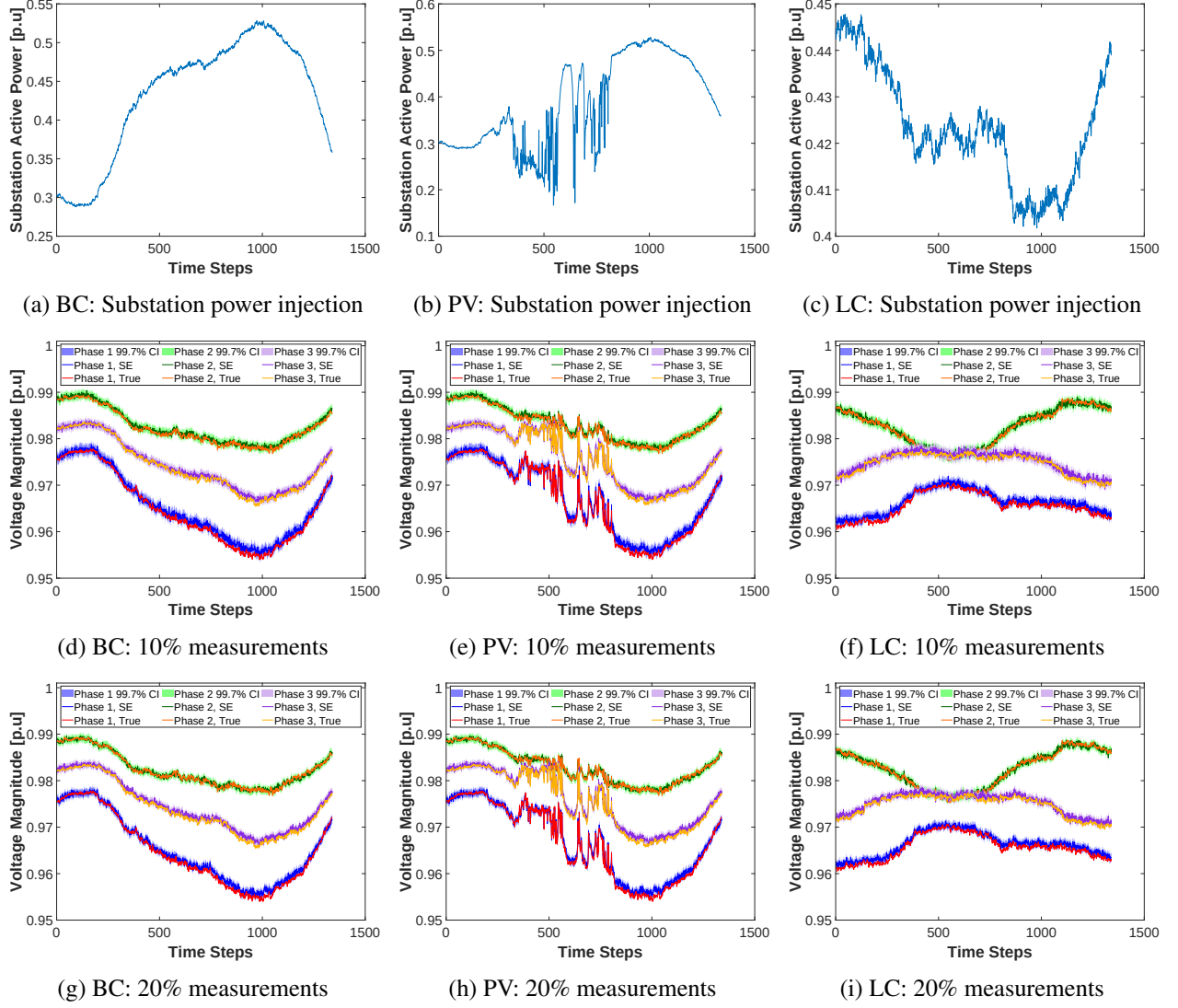


Figure 5.4: Performance of the proposed state estimation method in different scenarios. Figs. 5.4a–5.4c show the substation active power injections under different scenarios. Figs. 5.4d–5.4i show the performance of the proposed method with 10% (top) and 20% (bottom) measurements, in the base case (left, BC), with PV integration (middle, PV), and in the low-correlation scenario (right, LC). As an illustrative example, we plot the voltage at one particular three-phase node but emphasize that the performance is consistent across all the nodes in the network as shown by Tables 5.2–5.4.

in Section 5.6.2, despite the lower accuracy, the VAR(1) model trained with 200 data samples still captures the general trends in the power injections. As we will show in this section, the state estimation step is able to account for this increased error to recover accurate state estimates. While the VAR(1) model can be updated regularly to adapt to changes in load and renewable generation, for simplicity, we train the VAR model at the beginning of the simulation and keep it fixed thereafter.

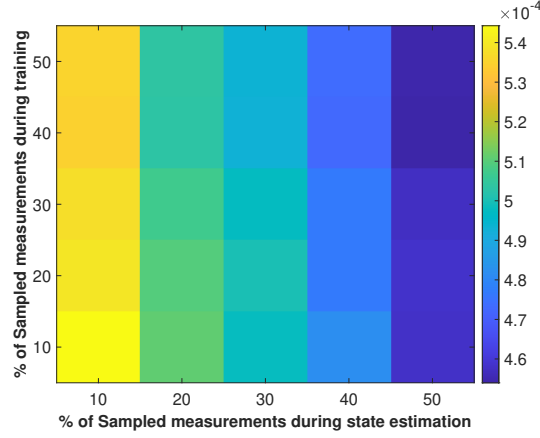
Table 5.4: Voltage estimation error under the low correlation

% Measurements	RMSE [p.u.]	MAE [p.u.]
Fixed Measurement set		
10	5.7959×10^{-4}	4.4772×10^{-4}
20	5.3669×10^{-4}	4.3154×10^{-4}
Dynamic Measurement set		
10	6.3749×10^{-4}	4.8921×10^{-4}
20	5.4284×10^{-4}	4.2703×10^{-4}

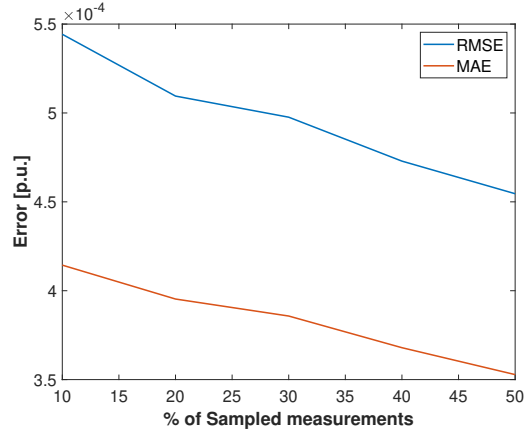
To evaluate the performance of the proposed method, voltage magnitudes estimated using the proposed method are compared against the true nodal voltages obtained by solving an AC power flow using the backward-forward sweep algorithm. To assess the performance of the proposed method under different communication infrastructures, we simulate both fixed and dynamically varying measurement sets, with sizes corresponding to 10% and 20% of all nodes in the network. The fixed measurement set represents a case where only a few nodes in the network are equipped with smart meters, while the dynamically varying measurement set represents a case where every node in the network is equipped with a smart meter but measurements can only be accessed from a few nodes each time step due to limited bandwidth. In the case of a fixed measurement set, we randomly choose one of the subsets constructed using the procedure described in Section 5.6.1 at the beginning of the simulation. The measurement set is kept fixed during the entire simulation. In the case of a dynamically varying measurement set, a different subset of measurements is selected at each time step such that all nodes in the network are sampled within a finite number of time steps. State estimation is initialized using the final sample of the training data. The estimated power injections from the previous time step are used as inputs to the VAR model for predicting the power injections at the current time step.

We first analyze the performance of the proposed method in the base case scenario. In this scenario, we assume there is no renewable generation in the network, ensuring that there are no sudden changes in the power injections, as shown in Fig. 5.4a. We also observe that the loads are highly correlated. Under these conditions, the proposed method is expected to work well with few measured nodes. Thus, the base case scenario serves as a baseline, allowing us to compare the performance of our proposed method under other scenarios. Figs. 5.4d and 5.4g show the true and the estimated voltages of all phases of a particular node, computed using a fixed measurement set with 20% and 10% of all nodes, respectively, while the estimation errors with different measurement configurations are tabulated in Table 5.2. Here, we can observe that our proposed method performs consistently across different measurement configurations. Notably, even with only 10% of the nodes sampled, the state estimation error remains low.

To assess whether the observed performance in the base case is related to the lack of renewable generation, we next analyze our proposed method in a scenario with high renewable penetration,



(a) Variation in the RMSE of estimated voltage magnitude as we vary the percentage of nodes sampled during VAR training (vertical axis) and during state estimation (horizontal axis).



(b) RMSE and MAE of the estimated voltage magnitude with an increasing percentage of sampled measurements for both VAR model training and state estimation.

Figure 5.5: Impact of the number of measured nodes on state estimation performance.

the results of which are summarized in Figs. 5.4e and 5.4h and in Table 5.3. The selected PV profiles exhibit significant fluctuations in active power injections, as shown by the substation active power injection in Fig. 5.4b. These sudden variations represent the uncertainty in PV generation due to factors such as cloud cover and partial shading. Despite the rapid changes in voltage magnitudes induced by the fluctuations in power injections, the proposed method accurately tracks the voltage, as shown by Figs. 5.4e and 5.4h. This indicates that method performs well even when our assumption of quasi-steady-state power injections (see Section 5.4.2) does not hold in the presence of significant quantities of renewable generation. Additionally, from Tables 5.2 and 5.3, we observe that our proposed method performs similarly in both the high renewable generation and the base case scenarios despite the sudden changes in active power injections.

Finally, to assess the impacts of load demand correlations, we simulate the case with low correlation

and compare to the base case. Figs. 5.4f–5.4i and Table 5.4 illustrate that the method remains effective for low-correlation scenarios. The proposed method performs comparably to the base case across the different measurement configurations, with some slight deterioration since the lack of measurements is partially compensated by the estimated correlation in nodal power injections. In the extreme case when the power injections are uncorrelated, the forecasted power injections do not provide additional information, necessitating a larger measurement set.

The results presented above illustrate that the proposed method performs consistently across all scenarios. However, we can observe from Tables 5.2–5.4 that, for all three scenarios, the performance of the proposed method with a dynamic measurement set is slightly worse than when compared to fixed measurement set due to the random sampling of observed nodes at each time step. Since the network consists of both three- and single-phase nodes, we note that random sampling does not ensure a consistent number of measurements at each time step. Thus, choosing a measurement set more strategically is likely to improve our method’s performance. This is a topic for our future work.

5.6.4 Impact of the Number of Measurements

Finally, we provide a more detailed assessment on how the number of sampled measurements impacts the accuracy of the state estimate. To accomplish this, we vary the number of measurements sampled during VAR parameter estimation and state estimation. For each combination of sampled measurements across these two stages, we simulate the base case and the low-correlation scenarios to assess the sensitivity of the proposed method to the number of measurements under different conditions. To ensure that the comparison is fair across all cases, we fix the forgetting factor λ at 1. We use 200 data samples to train the VAR model and use a fixed measurement set for the state estimation.

As expected, the performance of the proposed state estimation framework improves when we increase the number of measurements sampled at different stages, as shown in Figs. 5.5a–5.5b. From Fig. 5.5a, we observe that the performance of the state estimation is more sensitive to the number of measurements sampled at the state estimation step than to the number of measurements sampled during VAR training. While only the results for the base case scenario are shown, we observe similar trends for the low-correlation scenario. The results presented above suggest that careful selection of measurements during the state estimation stage is more critical.

5.7 Conclusions

This chapter has presented a forecast-aided state estimation scheme for distribution systems that addresses the problem of scarcity of measurements caused by limited communication bandwidth. This scheme consists of two stages. The first stage trains a forecasting model using power injection measurements from a subset of smart meters that are randomly sampled over a period of time. The second stage then uses the trained forecasting model along with the measurements from a few observable nodes for the state estimation. The proposed method improves on the recent literature by (i) introducing a recursive approach for the forecasting which does not require a large amount of

historical measurements, (ii) developing a state estimation approach without any phasor measurements, and (iii) accurately estimating the system state with a limited set of dynamically sampled smart meter measurements.

We validated the proposed scheme on the IEEE 123-bus benchmark network. The results show that the proposed scheme can accurately estimate the system state with measurements from only 10% of nodes, with estimation error reducing as the number of observed nodes increases. We also analyzed the performance of the proposed scheme using three scenarios with different correlation properties in power injections. The proposed method performs consistently across scenarios with varying measurement configurations and thus remains robust under a range of scenarios.

Future work will improve upon the key assumptions used in this work as follows.

- a) Our model assumes that every node has a smart meter (with few sampled simultaneously), which may not hold in distribution grids with limited smart meter deployment. However, we also demonstrated that our methods achieve accurate estimates with a constant set of measurements, representing scenarios with limited measurement units. We plan to further investigate the impacts that reduced smart meter coverage has on estimation accuracy.
- b) We used a linearized power flow approximation, Lin3DistFlow, and compared the state estimation results against the true AC power flow solutions. In the future, we will extend the state estimation model to consider the AC three-phase non-linear power flow model.
- c) We assumed that all the prediction errors and measurement noise are additive white Gaussian noise with zero mean, which may not reflect the real noise characteristics. Future work could relax this assumption by considering additional noise models.
- d) We also assumed that the prediction errors and measurement noise are uncorrelated, which may not always be true. Our results show that, for the range of time periods we considered, the proposed method remains accurate despite possible correlations. To mitigate the potential issues that may arise if this assumption does not hold, the VAR model could be retrained periodically.

Furthermore, our future research will focus on developing an optimization framework for dynamically sampling smart meter data, with the aim of reducing the state estimation error while satisfying communication constraints. We will also develop algorithms for considering adjustable precision from smart meters and determining different bit rates for various smart meters to minimize the overall state estimation error, considering the impact of quantization error and communication bandwidth limits.

6. Error Bounds for Radial Network Topology Learning from Quantized Measurements

We probabilistically bound the error of a solution to a radial network topology learning problem where both connectivity and line parameters are estimated. In our model, data errors are introduced by the precision of the sensors, i.e., quantization. This produces a nonlinear measurement model that embeds the *operation of the sensor communication network* into the learning problem, expanding beyond the additive noise models typically seen in power system estimation algorithms. We show that the error of a learned radial network topology is proportional to the quantization bin width and grows sublinearly in the number of nodes, provided that the number of samples per node is logarithmic in the number of nodes.

This chapter is based on the following paper:

S. Talkington, A. Rangarajan, P.A. de Alcântara, L.A. Roald, **D.K. Molzahn**, and D.R. Fuhrmann, “Error Bounds for Radial Network Topology Learning from Quantized Measurements”, submitted, arXiv:2508.05620.

6.1 Introduction

Efficiently allocating *limited smart meter bandwidth* is an emerging challenge at the intersection of power and communication engineering [7, 128]. As smart meters proliferate, power engineers will increasingly rely on analog-to-digital *quantization* methods to improve computational and communication efficiency by mapping continuous measurements to discrete intervals [129]; see Fig. 6.1 (left). While coarse quantization accelerates computations [130], it introduces *nonlinear*, typically *non-Gaussian*, measurement noise [131, 132], contrasting with the additive Gaussian assumptions common in power system estimation.

To help address these challenges, we present a new *sample complexity* analysis method for distribution topology learning, a well-studied inference task in power engineering; see [29] for a comprehensive review. Sample complexity refers to the *number of measurements needed* to ensure that the *error of an estimated parameter is bounded* by a chosen tolerance [133]. In particular, this work provides such topology learning error bounds under the effects of practically relevant communication non-idealities.

To the knowledge of the authors, these contributions are the first of their kind and expand upon recent deterministic studies of measurement requirements for topology learning [134]. Concretely, we address the following question:

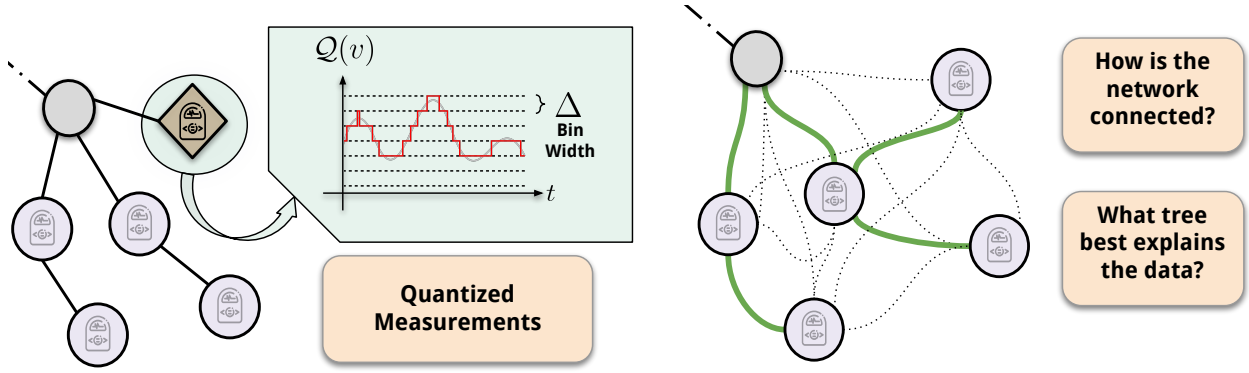


Figure 6.1: Conceptual illustration of the proposed problem. Left: Distribution network smart meter measurements have realistic quantization noise corresponding to *bin width* $\Delta > 0$. Right: Estimate the topology *and* parameters of the distribution network (green, solid lines).

Consider an n -node distribution network with unknown topology (connectivity and parameters). Suppose that we collect measurements at every node with a uniform quantization bin width $\Delta > 0$. How many samples per node $s = m/n$ are needed to recover the topology of the network from the $m = sn$ quantized measurements, up to a desired error tolerance?

This research question is illustrated in Fig. 6.1 (right), outlined in Section 6.2. We give a precise answer in Section 6.3 by providing a bound on the error of a topology estimate given a prescribed number of measurements.

6.2 Problem formulation

6.2.1 Communication model

We wish to recover an unknown vector of line parameters $\mathbf{w}_\star \in \mathbb{R}^d$ from $m \geq d$ *quantized measurements* of the form

$$p_i \triangleq \mathcal{Q}(\langle \mathbf{a}_i, \mathbf{w}_\star \rangle), \quad i = 1, \dots, m, \quad (6.1)$$

where $\mathcal{Q} : \mathbb{R} \rightarrow \mathbb{R}$ is a nonlinear *quantization function* and $(p_i, \mathbf{a}_i)$, $i = 1, \dots, m$ are measurements collected from smart meters (both spatially *and* temporally). Vectors $\{\mathbf{a}_i\}$ are rows of a particular sensing matrix $\mathbf{A} \in \mathbb{R}^{m \times d}$ and contain pairwise voltage differences measured co-temporally with active power injection p_i . We will define matrix $\mathbf{A}$ explicitly in Section 6.3.1 as a consequence of linearizing the power flow equations. In the same way, we will argue that the sparsity pattern of $\mathbf{w}_\star$ represents the network topology (cf. [29, Sec. IV-B]).

Concretely, we focus on the setting in which coarsely quantized measurements (6.1) are generated from a uniformly dithered quantization function (cf. [129, 132]) with bin width $\Delta > 0$. These

measurements take the form

$$p_i = \Delta \cdot \left(\left\lfloor \frac{\langle \mathbf{a}_i \rangle \mathbf{w}_\star + \tau_i}{\Delta} \right\rfloor + \frac{1}{2} \right) \quad i = 1, \dots, m, \quad (6.2)$$

where $\tau_i \sim \text{Uniform}(-\frac{\Delta}{2}, \frac{\Delta}{2})$, and $\mathbb{E}[p_i] = \langle \mathbf{a}_i \rangle \mathbf{w}_\star$. The *dither* τ_i is a purposely applied random noise component that is generated by a sensor and added to an input signal prior to quantization. This technique is rather well-established and commonly used by sensors in practice, and by statisticians in theory; see [129, 135].

6.2.2 Statistical tools

Suppose that $\mathbf{w}_\star$ lies in a convex constraint set $\mathcal{K} \subseteq \mathbb{R}^d$ that encodes some sort of *known structure* of the parameter $\mathbf{w}_\star$, e.g., knowledge of where certain lines are located (but not their switching status), or whether the network is operated radially. We can solve for a constrained estimate $\hat{\mathbf{w}} \in \mathbb{R}^d$ that obeys this structure via the program (6.3), known as the *generalized LASSO*:

$$\hat{\mathbf{w}} = \underset{\mathbf{w} \in \mathcal{K}}{\operatorname{argmin}} \frac{1}{2m} \sum_{i=1}^m (\langle \mathbf{a}_i \rangle \mathbf{w} - p_i)^2. \quad (6.3)$$

The quality of the estimate produced by the program (6.3) can be quantified with a statistical tool known as the *Gaussian width*. We introduce this tool briefly and refer the reader to [136], [38, Ch. 7.5] and [131, 132] for further technical information. Formally, the Gaussian width $\omega(\mathcal{T})$ of a set $\mathcal{T} \subseteq \mathbb{R}^d$ is defined as

$$\omega(\mathcal{T}) \triangleq \mathbb{E} \left[\sup_{\mathbf{u} \in \mathcal{T}} \langle \mathbf{u} \rangle \mathbf{g} \right], \quad (6.4)$$

where $\mathbf{g} \sim \mathcal{N}(0, \mathbf{I})$ is a vector of i.i.d. standard Gaussians. The Gaussian width ω is a useful tool for predicting the behavior of structured convex problems with random input data [136]. It represents the “size” of a standard Gaussian process over $\mathcal{T}$, which we can use as a stochastic *comparison* for our structured problem. Two classic examples of structured constraint sets with well-behaved Gaussian widths are sparse and low-rank constraint sets [136].

The squared Gaussian width, $\omega^2(\mathcal{T})$, provides a measure of the effective dimension of the set $\mathcal{T}$; informally, this represents the minimum number of orthogonal directions required to capture the variation or structure within $\mathcal{T}$. We will also need another object for our analysis—the *tangent cone* of a set $\mathcal{K} \subset \mathbb{R}^d$ at $\mathbf{x} \in \mathbb{R}^d$, which is defined as

$$\mathcal{D}(\mathcal{K}, \mathbf{x}) \triangleq \{ \lambda \mathbf{u} : \lambda \geq 0, \mathbf{u} \in \mathcal{K} - \mathbf{x} \}, \quad (6.5)$$

where $\mathcal{K} - \mathbf{x}$ is the set $\mathcal{K}$ translated by $\mathbf{x}$. The tangent cone (6.5) generalizes the idea of the tangent space to a nonlinear surface that may have non-differentiable points.

6.2.3 Network and power flow model

The abstract quantities in (6.1) and (6.2) can be applied to the Linear Coupled Power Flow (LCPF) model [29, 137]. As the focus of this work is on sample complexity, we make the following simplifying assumption on the reactive power.

Assumption 4 (Fixed power factor). Let $\mathbf{p} \in \mathbb{R}^n$ be active power injections. Assume that reactive power injections satisfy $\mathbf{q} = \kappa \mathbf{p}$, where $\kappa \in \mathbb{R}$ is a known constant.

Assumption 4 is equivalent to defining a fixed power factor $\phi \in (0, 1]$ such that $\kappa = \pm \phi^{-1} \sqrt{1 - \phi^2}$, where $\text{sgn}(\kappa) = 1$ if the injections are inductive and $\text{sgn}(\kappa) = -1$ if the injections are capacitive. For more information on the consequences of this assumption, see [138, 139].

Linear Coupled Power Flow (LCPF) model

For a radial network with $n + 1$ nodes, let $\mathbf{C} \in \{-1, 0, 1\}^{n \times n}$ be the invertible, reduced, branch-to-node incidence matrix of a tree graph with the column corresponding to the slack node removed. Let $\{(v_t, \mathbf{p}_t)\}_{t=1}^m$ be a sequence of nodal voltage magnitude and active power measurements. Under Assumption 4, the voltage magnitudes satisfy [29, 137]:

$$\mathbf{v}_t - \mathbf{1} = (\mathbf{R} + \mathbf{X} \cdot \kappa) \mathbf{p}_t, \quad t = 1, \dots, m \quad (6.6)$$

where $\mathbf{R} + \mathbf{X} \kappa \triangleq \mathbf{Z}$ is the (reduced) *equivalent impedance matrix*. We can write $\mathbf{Z}$ as the inverse of a *real-valued* equivalent admittance matrix $\mathbf{Y} \in \mathbb{R}^{n \times n}$, where

$$\mathbf{Z} = \mathbf{C}^{-1} \text{diag}(\mathbf{r} + \kappa \mathbf{x}) \mathbf{C}^{-\top} \triangleq \mathbf{Y}^{-1}, \quad (6.7)$$

where $\mathbf{z} \triangleq \mathbf{r} + \kappa \mathbf{x} \in \mathbb{R}^n$ are the line impedances (scaled). The matrix $\mathbf{C}^{-1} \in \{-1, 0\}^{n \times n}$ is a lower triangular matrix where the non-zero entries of each column j represent the descendants of node j in the tree. In particular, over the indices $i, j = 1, \dots, n$, we have $(\mathbf{C}^{-1})_{ij} = -1$ if $i = j$ or i is a descendant of j , and $(\mathbf{C}^{-1})_{ij} = 0$ otherwise.

Lifting to the complete graph

Echoing [134], we can view the problem of recovering $\mathbf{w}_*$ as *sparsifying* a complete undirected graph K_{n+1} with $n + 1$ nodes and $\binom{n+1}{2}$ lines. Accordingly, let $\tilde{\mathbf{Y}}(\cdot)$ be the *admittance matrix operator* that sends line parameters $\mathbf{w} \in \mathbb{R}^{\binom{n+1}{2}}$ to a corresponding (non-reduced) $(n + 1) \times (n + 1)$ admittance matrix:

$$\tilde{\mathbf{Y}}(\mathbf{w}) = \tilde{\mathbf{C}}^\top \text{diag}(\mathbf{w}) \tilde{\mathbf{C}} = \sum_{i=1}^n \sum_{j=i+1}^n \mathbf{E}_{ij} w_{ij}. \quad (6.8)$$

Here, $\mathbf{E}_{ij} \triangleq (\mathbf{i} - \mathbf{j})(\mathbf{i} - \mathbf{j})^\top$ is an *elementary Laplacian matrix*, where $\mathbf{i}$ is the i -th standard basis vector in $\mathbb{R}^{n+1}$. The matrix $\tilde{\mathbf{C}} \in \{-1, 0, 1\}^{\binom{n+1}{2} \times n+1}$ is the incidence matrix that corresponds to a complete graph. We can now represent the set of all radial networks as the set of all line parameter vectors $\mathbf{w}$ that correspond to a connected radial network with $n + 1$ nodes. This set takes the form

$$\mathcal{R} = \left\{ \mathbf{w} \in \mathbb{R}^{\binom{n+1}{2}} : \|\mathbf{w}\|_0 = n, \lambda_2(\tilde{\mathbf{Y}}(\mathbf{w})) > 0 \right\}. \quad (6.9)$$

The sparsity condition $\|\mathbf{w}\|_0 = n$ ensures n lines, and the condition on the second smallest eigenvalue $\lambda_2(\tilde{\mathbf{Y}}(\mathbf{w})) > 0$ means the graph is connected, i.e., $\mathbf{x}^\top \tilde{\mathbf{Y}}(\mathbf{w}) \mathbf{x} > 0$ for all $\mathbf{x} \perp \mathbf{1}$. The constraint

set (6.9) is intractable; however, in Section 6.3.1 we provide a relaxation that performs well both experimentally and theoretically. Hereafter, we implicitly remove the slack node column of $\tilde{\mathbf{C}}$ in our calculations, and refer to the network as having n nodes.

6.3 Main result

6.3.1 Topology learning problem

Let $\mathbf{P}, \mathbf{V} \in \mathbb{R}^{n \times s}$ be data matrices whose columns are s samples of active power and voltage magnitude measurement vectors, respectively, across all n nodes. We seek to recover an n -sparse vector $\mathbf{w}_\star$ of line parameters, where $w_{\star,i} = 1/z_{\star,i}$ if $z_{\star,i} \neq 0$, and 0 otherwise. This problem can be written as a sparse recovery problem over the *complete graph* by defining the measurement system

$$\mathbf{p} = \mathcal{Q}(\text{vec}(\mathbf{P})) = \mathcal{Q}(\mathbf{A}\mathbf{w}_\star), \quad \mathbf{p} \in \mathbb{R}^{sn}. \quad (6.10)$$

The $sn \times \binom{n+1}{2}$ sensing matrix is $\mathbf{A} \triangleq \mathbf{V}^\top \tilde{\mathbf{C}}^\top \odot \tilde{\mathbf{C}}^\top$, where $\odot$ is the Khatri-Rao matrix product.

6.3.2 Bounding the error of the parameter estimate

Taking advantage of the sparsity, we can characterize the Gaussian width of the tree set (6.9).

Lemma 6. *Define the set $\mathcal{T}_{\mathcal{R}, \mathbf{w}_\star} \triangleq \mathcal{D}(\mathcal{R}, \mathbf{w}_\star) \cap \mathbb{S}^{d-1}$ as the intersection of the tangent cone (6.5) of constraint set $\mathcal{R}$ with the unit shell $\mathbb{S}^{d-1} \triangleq \{\mathbf{u} \in \mathbb{R}^d : \|\mathbf{u}\|_2 = 1\}$, $d = \binom{n+1}{2}$. The squared Gaussian width of the set of radial networks (6.9) satisfies $\omega^2(\mathcal{T}_{\mathcal{R}, \mathbf{w}_\star}) \leq 2n \log\left(\frac{n+1}{2}\right) + \frac{3}{2}n$.*

Proof. The ground truth $\mathbf{w}_\star$ corresponds to a spanning tree of the complete graph with $n+1$ nodes. Since this tree must have n edges, vector $\mathbf{w}_\star$ must be n -sparse, i.e., $\|\mathbf{w}_\star\|_0 = n$. Dropping connectivity, define the relaxation $\mathcal{K} \supseteq \mathcal{R}$ as

$$\mathcal{K} \triangleq \left\{ \mathbf{w} \in \mathbb{R}^{\binom{n+1}{2}} : \|\mathbf{w}\|_1 \leq \|\mathbf{w}_\star\|_1 \right\}, \quad (6.11)$$

i.e., the ℓ_1 -norm ball of radius $\|\mathbf{w}_\star\|_1$. It is well known (cf. [132, 136]) that, as $\mathbf{w}_\star$ is n -sparse, we can conclude that $\omega^2(\mathcal{T}_{\mathcal{R}, \mathbf{w}_\star}) \leq \omega^2(\mathcal{K}, \mathbf{w}_\star) \leq 2n \log\left(\frac{n+1}{2}\right) + \frac{3}{2}n$. $\square$

Using Lemma 6, the error of the topology learned by (6.3) with constraint set (6.11) and bin width $\Delta > 0$ is bounded and can be quantified.

Theorem 3. *Suppose that $s = \frac{m}{n}$ samples with bin width $\Delta > 0$ are collected throughout an n -node distribution network, and suppose that the $\{\mathbf{a}_i\}$ are any i.i.d. sub-Gaussian random vectors. Then, there exist constants $C, c_1, c_2 > 0$ such that the error of the topology estimate from solving (6.3), with probability at least 0.99 (i.e., 99%) is bounded as*

$$\|\hat{\mathbf{w}} - \mathbf{w}_\star\|_2 \leq C\Delta \cdot \sqrt{\frac{2 \log\left(\frac{n+1}{2}\right) + 3/2}{s}}, \quad (6.12)$$

provided the number of samples $s \geq \mathcal{O}(\log n)$.

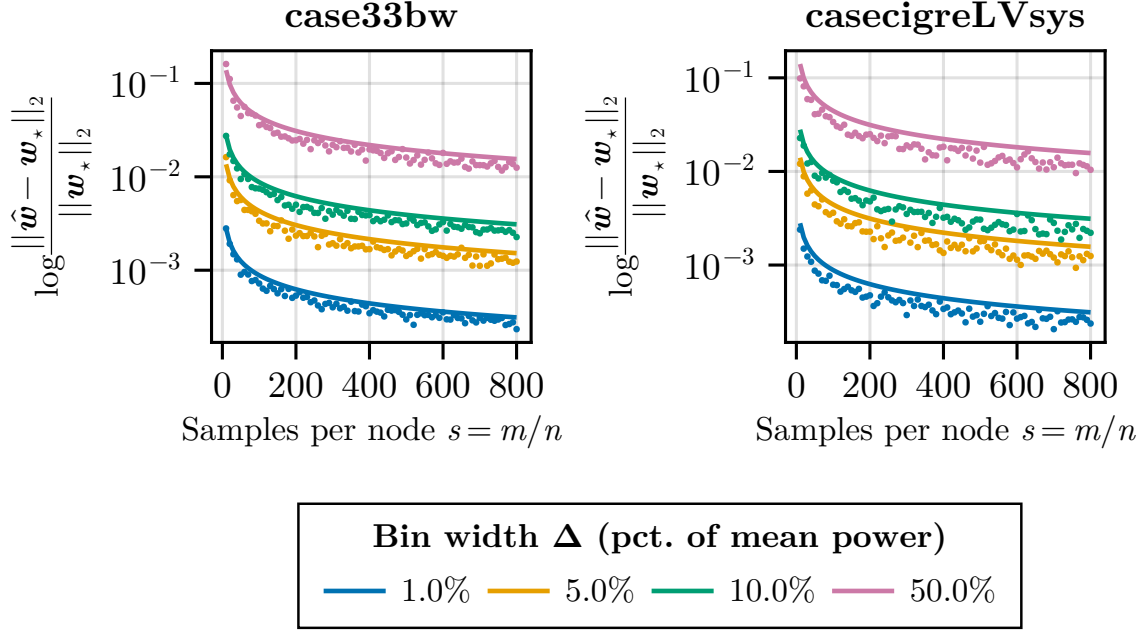


Figure 6.2: Relative error vs. number of samples for repeated runs of the topology learning program (6.3) on the Baran & Wu (left) and CIGRE low-voltage (right) test cases. The scatter markers show the experimental error obtained for one run of (6.3), where each color corresponds to a bin width Δ . The solid lines show the error prescribed by (6.12).

Proof. Combining Lemma 6 with (6.13) and applying [132, Thm. III.1], there exists a constant $C > 0$ such that the minimizer $\hat{w}$ of the program (6.3) satisfies, with probability at least 0.99,

$$\|\hat{w} - w_*\|_2 \leq C\Delta \cdot \frac{\omega(\mathcal{T}_{\mathcal{K}}, w_*)}{\sqrt{m}}, \quad (6.13a)$$

$$\leq C\Delta \cdot \sqrt{\frac{2n \log(\frac{n+1}{2}) + \frac{3}{2}n}{sn}}, \quad (6.13b)$$

provided the total number of measurements $m = sn \geq c_1 \omega^2(\mathcal{T}_{\mathcal{K}}, w_*) + c_2$ for some absolute constants $c_1, c_2 > 0$. Due to Lemma 6, if the number of samples per node $s \geq c_1(2 \log(\frac{n+1}{2}) + 3/2) + c_2 = \mathcal{O}(\log n)$ we satisfy this measurement requirement. $\square$

6.3.3 Numerical results

Fig. 6.2 compares the error bound (6.12) with the parameter error obtained by numerically solving many instances of the program (6.3). The errors are plotted against the number of samples per node, varied over 100 discrete uniformly spaced points $10 \leq s \leq 800$. For each instance of the

program (6.3), a random matrix $\mathbf{V} \in \mathbb{R}^{n \times s}$ of independent Gaussian voltages is generated, where the mean of each column is the AC power flow solution of the feeder, and the variance of each column entry is 10% thereof. Active power measurements are then generated as in (6.10). The quantization bin width Δ is varied as a percentage of the sample mean of the absolute active power injections $\frac{1}{m} \|\mathbf{A}\mathbf{w}_*\|_1$. The solid curves in Fig. 6.2 depict the error bounds given by (6.12), for each bin width Δ .

Following [132], the constant C is not specified by the theorem. Therefore, we numerically select C for each feeder by computing the maximum ratio of empirical errors and (6.12) over all s, Δ . We found $C \approx 13$ for `case33bw` and $C \approx 10$ for `casecigreLV`, which we use for the solid curves in Fig. 6.2. The scaling $\mathcal{O}(\Delta\sqrt{\log n/s})$ predicted by (6.12) matches the experimental errors across all Δ .

6.4 Conclusion

We provided a framework to *predict the error of a learned distribution network topology—before solving any optimization problem*. By exploiting the radial structure, we achieved an error *sublinear* in the number of nodes, up to constant factors determined by the precision of the sensor and the underlying probability distributions. Formally, we showed that the topology *and* line parameters of a radial network can be recovered up to a relative error of $\mathcal{O}(\Delta\sqrt{\log n/s})$, given a collection of $\mathcal{O}(\log n)$ samples from every node with quantization bin width $\Delta > 0$. Thus, before collecting data, one can size quantization and sampling budgets to hit a target topology error. Note that (6.13) is determined by the bin width Δ . Deriving bounds involving communication bit rates would require assumptions on the variance of the underlying measurement being quantized.

7. Kalman Filter-based State Estimator with Measurement Quantization

The development of new sensing technologies has spurred many studies in related fields, including network communication and information processing. The diversity of fields ranges from environmental monitoring to industrial applications, including smart grids, that rely on state estimators for dynamic monitoring and control. Kalman Filters have been widely used for state estimation of stochastic processes. Communication between the sensors and the processor becomes an issue in applications where the sensors are widely dispersed and physically separated from a central processing facility. Therefore, low communication bandwidth adds more complexity to state estimator models, as state quantization should be considered. This chapter studies a Kalman Filter with a quantization process replacing noisy measurements. The methodology considers a centralized Kalman Filtering method in which the local central processor communicates with a remote central processor through a low-bandwidth communication channel. Initially, the analysis is carried out in a one-dimensional system with a 1-bit quantization process in which 2-bin and 3-bin 1-bit quantizers are considered. State distributions are calculated using discrete convolutions on a fine grid, and with Gaussian approximations. For the 3-bin 1-bit quantizer, two variations of the Gaussian approximation are developed. MATLAB simulations for the five resulting methods indicate that discrete convolutions are not worth the high computational cost, and that the 3-bin 1-bit quantizer eliminates the granular noise of the conventional 2-bin 1-bit quantizer. The problem is extended to a vector version in a multi-dimensional system with several states. The method focuses on optimally allocating bit-rates to each sensor given a maximum channel capacity between the local and the remote data processors, based on the Gaussian rate-distortion curve for independent sources. Assuming a constant state transition matrix, the results show the methodology may allocate the bit-rates by calculating the steady-state maximum individual distortion based on a constant innovation covariance matrix of the states. In a practical implementation, the difference between the achieved rate-distortion curve and the theoretical one depends not only on the type of quantizer, but also on the assumed state transition and innovation covariance matrices.

This chapter is based in part on the following paper:

P. A. de Alcantara, F. Bezerra Costa and D. R. Fuhrmann, "One-Dimensional Kalman Filter with Measurement Noise Replaced by Quantization", *259th Annual Conference on Information Sciences and Systems (CISS)*, Baltimore, MD, USA, 2025.

7.1 Introduction

The growing development of new sensing technologies in conjunction with the need for robust data processing capabilities has encouraged research in the most diverse areas. The large amount of available data leads to new studies in network communication and information processing, enabling the optimal leverage of important information [140]. The diversity of fields ranges from

environmental monitoring to industrial applications, including smart grids, as well as human health-care products. In all of them, optimally selecting key information from the sensors is dependent not only on the specific nature of the application but also on how the sensors are disposed in the communication network architecture. Networks with widely dispersed sensors, for instance, may require communication channels with high capacity, adding more complexity to systems with centralized data processing. Additionally, most systems have a dynamic and stochastic nature, which spurs even more the seek for optimal sensing and information processing.

7.1.1 Literature Review

Dynamic stochastic systems require robust state estimators to operate safely and reliably. The Kalman filter (KF), first published in [141], is a well-established state estimation method for dynamic systems used in several applications [142]. Although it was originally developed for linear systems to solve the problem of state estimation in continuous-time [143], the Unscented (UKF) and Extended (EKF) Kalman filters are the most popular examples of further improvements to deal with non-linear systems [144]. Additionally, the KF literature may be divided into three main classes, each taking advantage of different levels of cooperation between sensors [145]. The centralized KF methods work with a central processing node to which all other nodes report, with no inter-node communication. In these cases, the central node performs Kalman filtering to calculate the network estimates. In decentralized methods, all sensors calculate network estimates locally, leveraging full or partial node cooperation. Algorithms like [146] are designed to guarantee that the local state estimates are identical to the results obtained from centralized methods, with the advantage of avoiding a single point of failure, and, consequently, increasing the system reliability. However, fully decentralized methods may result in inconsistent and/or suboptimal estimates, as the algorithms are not designed to incorporate information from other sensors collaboratively. Therefore, distributed methods like [145, 147, 148] are designed to achieve consensual results by leveraging the local estimates of different sensors and actively collaborating with neighboring sensors.

Despite the advantages of decentralized and distributed methods, centralized KF is still the best option in many scenarios. Processing data in one place simplifies mathematical calculations, reducing computational effort and implementation complexity, as well as taking full advantage of correlations when modeling the states as a single state vector [149]. Also, centralized methods are optimal in finding global minimum mean square error under Gaussian assumptions [150]. However, one of the main challenges of having a central processor is dealing with the limitation of the communication bandwidth. If communication channels have limited capacity, then the sensor data must be quantized or otherwise reduced, even if the sensor measurements themselves are perfect, imposing more complexity on the state estimation problem, and hence, several researchers have addressed Kalman filter-based state estimation with quantization effects.

In [142], the proposed quantized Kalman filter method based on the quantization of innovations is compared to an approach based on quantized observations. The authors provide a study of the tradeoff between energy consumption and estimation accuracy to show that the method reduces

information loss. Reference [143] analyzes the state estimation of networked linear systems with quantization effects in both noise-free and unreliable channel situations. They show the mean-squared error (MSE) lower bounds of quantized state estimation. Further, the UKF is mixed with optimal quantization by [151] to propose a novel approach to finding the best approximation of a random variable (RV) density. The effect of quantization is considered in [144] to tackle the state estimation problem in distribution networks using improved UKF. Another application of quantized Kalman filtering is presented in [152], where a two-level quantizer is used to predict vehicle states in traffic. Reference [153] designs a method for stochastic nonlinear systems with quantization effects using the EKF, including the impact of Round-Robin communication protocols. The results validate the method through simulations. In the Quantized Gossip-based Interactive Kalman Filter (QGIKF) proposed by [154], the authors show that the estimation error variance converges weakly in distribution in a wireless sensor network where there exists inter-sensor communication.

Previous works address the quantization effects by modeling them as additive measurement noise, a technique that works well in low-noise/high-bit rate applications. However, these algorithms require high computational effort, especially for nonlinear systems. In contrast, [155] compares a Quantized Kalman Filter based on quantized observations and shows that it has mean square error (MSE) and computational complexity similar to those estimators based on non-quantized observations.

7.1.2 Problem Statement and Contributions

This chapter proposes an approach to quantization effects in a dynamic system state estimation problem, in which we directly model the quantization of the measurements. The context is well depicted in Figure 7.1, using a small portion of a distribution network with smart-meters as an example. The dispersed smart-meters sensors send measurement information to a local central processor with high-bandwidth communication, but a low bandwidth communication link connects the local and the remote central processors.

The first contribution of this chapter consists of an approach to quantization effects in a simple one-dimensional dynamic system state estimation problem, in which we directly model the quantization of the measurements [140]. The applications are low-bit-rate systems, where the communication bandwidth is on the order of 1 bit per measurement sample. For the intuitively simple 2-bin 1-bit quantizer, the proposed method reduces to ordinary delta modulation of the measurement [156]. We also introduce a novel 3-bin 1-bit quantizer, which slightly reduces the reconstruction mean square error (MSE) and eliminates granular noise or “dither”. Furthermore, we consider both the exact calculation of the state probability density function (pdf) on a fine discrete grid and a Gaussian approximation, which significantly reduces computational effort with little performance degradation.

The second contribution expands the system to a multi-dimensional dynamic model, where all states are individually quantized. The method proposes the use of rate-distortion functions for independent Gaussian sources to allocate bits according to their variances, resulting in a reverse

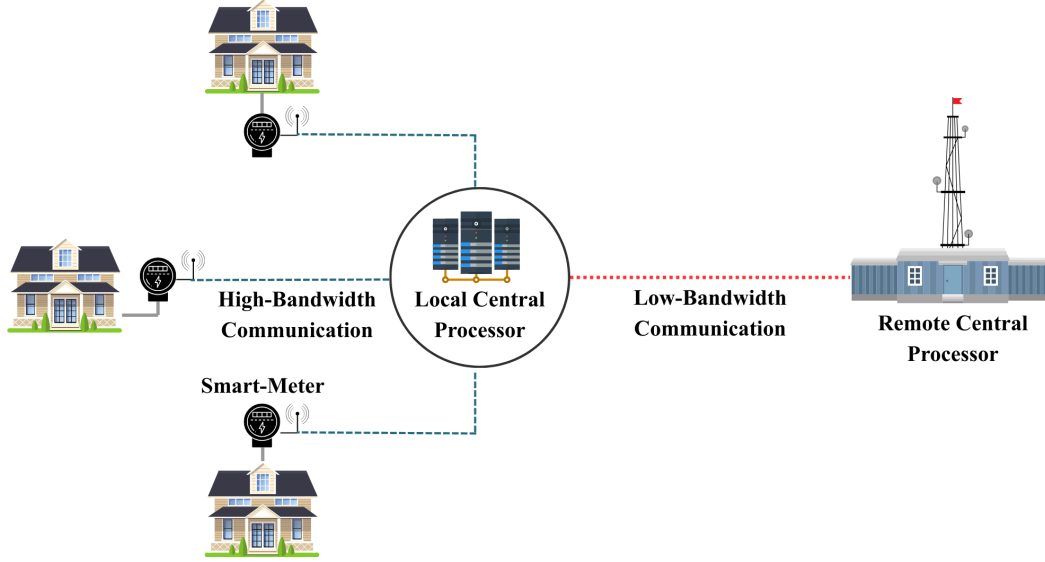


Figure 7.1: Studied sensed network architecture in a distribution network with smart-meters.

water-filling approach. The objective is to keep the individual quantization distortion of each state at the same level by making setting the quantizers bin-widths equal. The bin-widths and reconstruction points are calculated by approximating the distribution of the bins as uniform and Gaussian approximation, or merging them. The results show that those quantizers with Gaussian approximated reconstructed points present better performance at low bit rates. The achieved rate-distortion curves differ from the theoretical lower bound by the expected amount specified by Gish and Pierce [157], validating the method.

7.2 Methodology

7.2.1 One-Dimensional System

Conventional One-Dimensional Kalman Filter

This paper comprises applications where the sensors are widely dispersed and separate from the command center. The command center receives the measurements and tracks the distribution of the states. The analysis and propositions in this section consider the scalar/one-dimensional case of the Kalman filter-based state estimator.

The classical approach for this case regards the state model and the measurement dynamic respectively given by:

$$x(k) = x(k-1) + u(k), \quad (7.1)$$

$$y(k) = x(k) + w(k), \quad (7.2)$$

where $x(k)$ and $y(k)$ are the states and measurements at time instant k , respectively; $u(k)$ and $w(k)$

are state prediction and measurement noises, respectively.

The estimator leverages measurement information to track the distribution of the states by updating it every time step. Consider the random variables X and Y to represent the states and measurements, respectively, and the random variables U and W to represent state prediction and measurement noises, respectively. Before receiving a measurement, the estimator has prior knowledge of the distribution of X , denoted by $p_1(x)$. After receiving measurement information, the distribution of X is updated into a posterior distribution here denoted by $p_2(x)$. The state distribution propagation is depicted in Fig. 7.2.

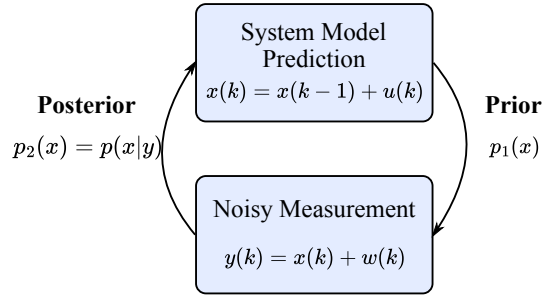


Figure 7.2: Propagation of the state distribution with the conventional Kalman Filter approach.

The distribution propagation process consists of two steps: system model prediction and noisy measurement step according to equations (7.1) and (7.2), respectively. Every time instant, the system model prediction step uses the old $p_2(x)$ (i.e., the posterior distribution at time instant $(k-1)$) to generate the new $p_1(x)$ as the convolution of $p_1(x)$ and the pdf of $u(k)$. The measurement step obtains the new $p_2(x)$ by calculating the conditional distribution of X given the measurement information Y , that is, $p_2(x) = p(x|y)$.

If the noises are normally distributed with zero mean and given variance as $U \sim \mathcal{N}(0, \sigma_u^2)$ and $W \sim \mathcal{N}(0, \sigma_w^2)$, the prior and posterior distributions are Gaussian, where $p_1(x) \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $p_2(x) = p(X|Y) \sim \mathcal{N}(\mu_2, \sigma_2^2)$. This way it is possible to calculate $p_1(x)$ and $p_2(x)$ in steady-state.

As the mean of U is zero ($\mu_u = 0$), then the prior mean μ_1 equals the old posterior mean μ_2 , i.e., the mean of $p_2(x)$ at the previous time instant, as follows:

$$\mu_1(k) = \mu_2(k-1). \quad (7.3)$$

Similarly, the prior variance σ_1^2 is the sum of old posterior variance and the variance of state prediction noise σ_u^2 , as follows:

$$\sigma_1^2(k) = \sigma_2^2(k-1) + \sigma_u^2(k). \quad (7.4)$$

As the measurement $Y = X + W$, then $Y \sim \mathcal{N}(\mu_1, \sigma_1^2 + \sigma_w^2)$. Also, as $p_2(x) = p(x|y)$, then the new

μ_2 is the conditional mean of X given $Y = y$. In steady-state, μ_2 is given by:

$$\mu_2(k) = \left(\frac{\sigma_w^2(k)}{\sigma_1^2(k) + \sigma_w^2(k)} \right) \mu_1(k) + \left(\frac{\sigma_1^2(k)}{\sigma_1^2(k) + \sigma_w^2(k)} \right) y(k). \quad (7.5)$$

Similarly, the variance of $p_2(x)$ in steady-state is given by:

$$\sigma_2^2(k) = \sigma_1^2(k) \left(\frac{\sigma_w^2(k)}{\sigma_1^2(k) + \sigma_w^2(k)} \right). \quad (7.6)$$

Proposed Quantized One-Dimensional Kalman Filter

We aim to analyze the quantization effect on state distribution tracking. We use a similar idea to the conventional approach but focus on quantization. The idea is to replace the noisy measurement with a model given by a quantization function Q as follows:

$$y(k) = Q(x(k)) \quad (7.7)$$

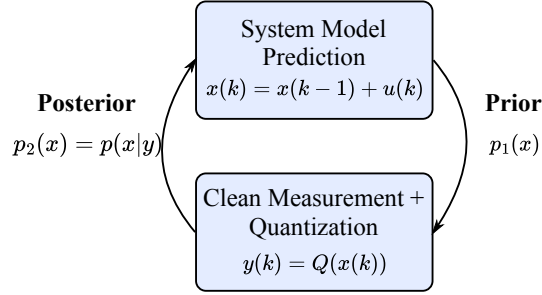


Figure 7.3: Propagation of the state distribution with noisy measurements replaced by quantization.

The measurements prior to quantization are considered noise-free as the focus is to evaluate the effect of the quantization on the state distribution propagation. Then, the approach for the distribution propagation is depicted in Fig. 7.3.

Consider $p_1(x)$ has a given support set $\mathcal{X}_1$. The scalar quantizer of the measurements splits $\mathcal{X}_1$ into R quantization regions $\mathcal{R}_i$ with each respective reconstruction points $a_i, \forall i \in \{1, 2, \dots, R\}$. The posterior distribution is then the conditional distribution given that $X \in \mathcal{R}_i$ ($p_2(x) = p(x|X \in \mathcal{R}_i)$).

In steady-state, using the same idea as in the conventional Kalman filter approach, μ_1 and σ_1^2 are given by equations (7.3) and (7.4), respectively; μ_2 and σ_2^2 are the mean and variance, respectively, of the conditional distribution of X , conditional on the region R_i in which it lies.

Gaussian Approximation

The new $p_1(x)$ is given by the convolution of the old $p_2(x)$ with the distribution of U . But if the prediction noise is normally distributed, the results will show that it is reasonable to approximate $p_2(x)$ as Gaussian, so $p_1(x)$ is consequently Gaussian.

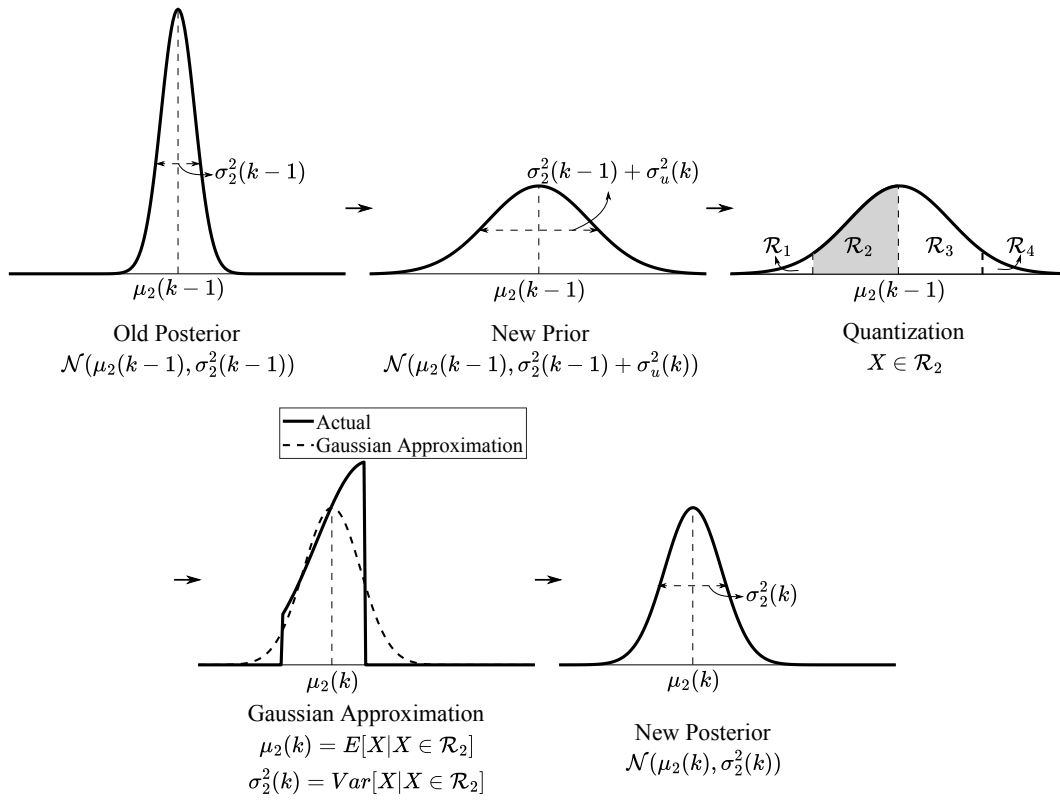


Figure 7.4: Steps of the propagation of state distribution with quantization and Gaussian approximation.

To help perception, the whole tracking process is presented in Fig. 7.4. Given initial knowledge for the old posterior as Gaussian distributed, the new prior is obtained by adding the prediction noise variance. In Fig. 7.4, the quantization step is illustrated considering four regions and the state value lies hypothetically in the region $\mathcal{R}_2$. The new posterior is then approximated as the Gaussian with mean and variance equal to the mean and variance of the actual conditional distribution. Finally, the new posterior is updated as the old posterior in the next instant, and the process starts again.

This work compares the results of simulations using the Gaussian approximations and simulations where the real distributions are calculated every time instant, i.e., $p_2(x)$ is the actual conditional distribution and $p_1(x)$ is the convolution of the old $p_2(x)$ with the distribution of U , computed on a fine grid.

This work comprises low-bit rate systems where the communication bandwidth is 1 bit per measurement sample. The following sections delve into 1-bit quantization using the quantized Kalman filter approach.

2-bin 1-bit Quantizer

In the most obvious 1-bit quantization, the prior state distribution is split into two quantization regions (bins) with equal probability [158, 159]. The quantizer then represents the measurement depending on the region where the state $x(k)$ lies. The new posterior mean is quantized to a_1 or a_2 , according to Fig. 7.5.

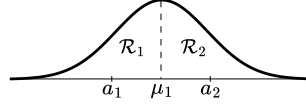


Figure 7.5: 2-bin 1-bit quantization.

If $p_1(x)$ is Gaussian, a_1 and a_2 can be expressed in terms of the mean and variance of a half-Gaussian distribution derived as a conditional distribution of the standard Gaussian. Then, if $X \sim \mathcal{N}(0, 1)$, the mean and variance of a half-Gaussian are respectively given by:

$$\mu_n = \sqrt{\frac{2}{\pi}}, \quad (7.8)$$

$$\sigma_n^2 = \left(1 - \frac{2}{\pi}\right). \quad (7.9)$$

Further, it is possible to calculate the prior variance of the state in steady-state as follows:

$$\sigma_1^2(k) = \sigma_1^2(k) \left(1 - \frac{2}{\pi}\right) + \sigma_u^2 = \frac{\sigma_u^2 \pi}{2} \quad (7.10)$$

Then, a_1 and a_2 are calculated as follows:

$$a_{1,2}(k) = \mu_1(k) \pm \mu_n \sqrt{\frac{\sigma_u^2 \pi}{2}} = \mu_1(k) \pm \sqrt{\frac{2}{\pi}} \sqrt{\frac{\sigma_u^2 \pi}{2}} \quad (7.11)$$

$$a_1(k) = \mu_1(k) - \sigma_u \quad (7.12)$$

$$a_2(k) = \mu_1(k) + \sigma_u. \quad (7.13)$$

Using the Gaussian approximations, the posterior distribution has mean and variance calculated, respectively, as follows:

$$\mu_2(k) = \begin{cases} a_1(k), & \text{if } x(k) < \mu_1(k) \\ a_2(k), & \text{if } x(k) > \mu_1(k) \end{cases} \quad (7.14)$$

$$\sigma_2^2(k) = \sigma_n^2 \sigma_1^2(k) = \left(1 - \frac{2}{\pi}\right) \frac{\sigma_u^2 \pi}{2}. \quad (7.15)$$

In this case, $\alpha = \sigma_n^2 = \left(1 - \frac{2}{\pi}\right) \frac{\sigma_u^2 \pi}{2}$.

Proposed 3-bin 1-bit Quantizer

In the conventional 1-bit quantizer with two bins, the estimator constantly changes the posterior mean value regardless of how close the measurement is to the previous value. This leads to granular noise, where the estimator output oscillates around a slowly changing state value. Especially if the state distribution is Gaussian, the state value is more likely to lie in a central region.

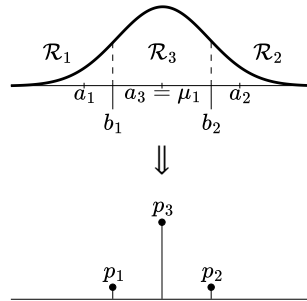


Figure 7.6: Proposed 3-bin 1-bit quantization.

Alternatively, this paper proposes to split the prior state distribution into three bins, as shown in Fig. 7.6. In this situation, the estimator does not need to change the value of the estimate if the change is not considerable. It yields bins with different probabilities, where b_1 and b_2 are the region bounds, and a_1 , a_3 and a_2 are the reconstruction points. Given that the central bin has a high probability p_3 , and the left and right tails have the same probability $p_1 = p_2 = \frac{1-p_3}{2}$, it is possible to calculate the discrete distribution of the quantizer such that the entropy is 1 bit. This means that

for sufficiently large sequences, it is possible to encode the measurement with an entropy rate of 1 bit/sample.

$$\begin{aligned}
H &= - \sum_{i=1}^3 p_i \log_2 p_i = 1 \\
&= -(1-p_3) \log_2 \left(\frac{1-p_3}{2} \right) - p_3 \log_2 p_3 = 1.
\end{aligned} \tag{7.16}$$

Conventional numerical methods can be used to show that $p_3 \approx 0.7730$ and consequently $p_1 = p_2 \approx 0.1135$. These values are used to calculate the boundaries b_1 and b_2 using the state pdf and cumulative density function (cdf). Given a state cdf $F(x)$, b_1 and b_2 are respectively calculated as follows:

$$b_1 = F^{-1}(p_1) \tag{7.17}$$

$$b_2 = F^{-1}(p_1 + p_3). \tag{7.18}$$

In the 2-bin 1-bit quantizer, the boundary that separated the regions was given by the prior distribution mean itself, so there was no need for further calculation. However, in the proposed approach, the boundaries need to be calculated along with the reconstruction points. Also, in the conventional approach, the Gaussian approximations were performed by calculating the prior variance in steady-state, given that the posterior distribution always has a half-Gaussian shape. However, in the proposed approach, this is not possible since the posterior distribution has a different shape depending on the bin the state belongs to. It means that the variance changes constantly. This makes room to address Gaussian approximations in two ways: using a changing variance or an average steady-state variance.

Changing Variance Gaussian Approximation: consider the posterior variances calculated at each time instant. First, if the state prior distribution $p_1(x)$ is Gaussian, b_1 and b_2 can be expressed in terms of a standard normal Gaussian. If $X \sim \mathcal{N}(0, 1)$, then $b_1 = -b_2$. The area below the curve in the interval $[0, b_2]$ is equal to $\frac{p_3}{2}$, then the boundaries related to the standard normal b_{1n} and b_{2n} are calculated as:

$$b_{1n} = -\Phi^{-1} \left(\frac{p_3 + 1}{2} \right) \approx -1.2079 \tag{7.19a}$$

$$b_{2n} = \Phi^{-1} \left(\frac{p_3 + 1}{2} \right) \approx 1.2079 \tag{7.19b}$$

where Φ is the standard Gaussian CDF.

Further, the posterior mean and variance can also be expressed in terms of the standard Gaussian. They will vary according to the bin to which the measurement belongs. They are calculated using

well-known conditional mean and variance equations in the Gaussian distribution literature. If the measurement belongs to the tail bins ($\mathcal{R}_1$ or $\mathcal{R}_2$), the conditional standard normal variance and mean are respectively:

$$\mu_{nT} = E[X|X > 1.2079] = 1.6940 \quad (7.20a)$$

$$\sigma_{nT}^2 = \text{Var}[X|X > 1.2079] = 0.1764 \quad (7.20b)$$

If the measurement belongs to the central bin $\mathcal{R}_3$, the conditional standard normal central bin variance is given by:

$$\sigma_{nC}^2 = \text{Var}[X|-1.2079 < X < 1.2079] = 0.3988 \quad (7.21)$$

There is no need to calculate the conditional standard normal mean of the central bin since the posterior mean will receive the same value as the prior mean ($\mu_2(k) = \mu_1(k)$).

This approach updates the prior variance every instant as in (7.4). Then, the boundaries are calculated every instant as:

$$b_1(k) = \mu_1(k) - 1.2079\sqrt{\sigma_1^2(k) + \sigma_u^2}, \quad (7.22a)$$

$$b_2(k) = \mu_1(k) + 1.2079\sqrt{\sigma_1^2(k) + \sigma_u^2}. \quad (7.22b)$$

The reconstruction points are finally calculated as:

$$a_1(k) = \mu_1(k) - 1.6940\sqrt{\sigma_1^2(k) + \sigma_u^2}, \quad (7.23a)$$

$$a_3(k) = \mu_1(k), \quad (7.23b)$$

$$a_2(k) = \mu_1(k) + 1.6940\sqrt{\sigma_1^2(k) + \sigma_u^2}. \quad (7.23c)$$

Finally, the posterior variance calculated using the Gaussian approximations has mean and variance given by:

$$\mu_2(k) = \begin{cases} a_1(k), & \text{if } x(k) < b_1(k) \\ a_3(k), & \text{if } b_1(k) < x(k) < b_2(k) \\ a_2(k), & \text{if } x(k) > b_2(k), \end{cases} \quad (7.24)$$

$$\sigma_2^2(k) = \begin{cases} 0.1764(\sigma_1^2(k) + \sigma_u^2), & \text{if } x(k) < b_1(k) \\ 0.3988(\sigma_1^2(k) + \sigma_u^2), & \text{if } b_1(k) < x(k) < b_2(k) \\ 0.1764(\sigma_1^2(k) + \sigma_u^2), & \text{if } x(k) > b_2(k). \end{cases} \quad (7.25)$$

Average Variance Gaussian Approximation: in this approach, the posterior variances are calculated using a steady-state variance based on average bin variances. Instead of having a conditional standard normal variance for each bin, an average value between them is used to calculate a steady-state

variance. It is calculated according to the respective bin probabilities as follows:

$$\sigma_{avg}^2 = \sigma_{nT}^2(1 - p_3) + \sigma_{nC}^2 p_3 \quad (7.26)$$

$$= 0.1764 \times 0.2270 + 0.3988 \times 0.7730 \approx 0.3483. \quad (7.27)$$

The posterior distribution has an approximate variance in steady-state σ_{avg}^2 . Then, it is possible to calculate the variance of the prior distribution in steady-state as:

$$\sigma_1^2(k) = \sigma_{avg}^2 \sigma_1^2(k) + \sigma_u^2 = \frac{\sigma_u^2}{1 - \sigma_{avg}^2} \approx 1.5344 \sigma_u^2. \quad (7.28)$$

The boundaries are then calculated:

$$b_1(k) = \mu_1(k) - 1.2079 \sqrt{1.5344 \sigma_u^2} \quad (7.29a)$$

$$b_2(k) = \mu_1(k) + 1.2079 \sqrt{1.5344 \sigma_u^2}. \quad (7.29b)$$

The reconstruction points are given by:

$$a_1(k) = \mu_1(k) - 1.6940 \sqrt{1.5344 \sigma_u^2} \quad (7.30a)$$

$$a_3(k) = \mu_1(k) \quad (7.30b)$$

$$a_2(k) = \mu_1(k) + 1.6940 \sqrt{1.5344 \sigma_u^2}. \quad (7.30c)$$

The posterior mean is chosen as in (7.24) using reconstruction points as in (7.30), and variances are as follows:

$$\sigma_2^2(k) = \begin{cases} 0.1764(1.5344 \sigma_u^2), & \text{if } x(k) < b_1(k). \\ 0.3988(1.5344 \sigma_u^2), & \text{if } b_1(k) < x(k) < b_2(k) \\ 0.1764(1.5344 \sigma_u^2), & \text{if } x(k) > b_2(k). \end{cases} \quad (7.31)$$

7.2.2 Multi-Dimensional System

Multi-Dimensional Kalman Filter

The context of this work comprises applications where the sensors are widely dispersed and located far from the central command center. The low bit-rate information from the sensors is used to estimate the network state through a Kalman filter state estimator. This section considers the vector case of the Kalman filter-based state estimator. It is assumed that the sensors can communicate with each other, allowing the states to be modeled as a state vector.

In a conventional Kalman Filter-based state estimator approach, the distribution of the states is tracked in two steps: the prediction step, and the update step. The first predicts the state distribution according to a state transition model (equation (7.32)), and the second updates the knowledge level of the states distribution after receiving measurement information from the measurement model (equation (7.33)). Consider the states and measurement models, respectively:

$$\bar{\mathbf{x}}(\mathbf{k}) = \mathbf{A}\bar{\mathbf{x}}(\mathbf{k} - 1) + \bar{\mathbf{u}}(\mathbf{k}) \quad (7.32)$$

$$\bar{\mathbf{y}}(\mathbf{k}) = \mathbf{H}\bar{\mathbf{x}}(\mathbf{k}) + \bar{\mathbf{w}}(\mathbf{k}) \quad (7.33)$$

where $\bar{\mathbf{x}}$ is the state vector; $\mathbf{A}$ is the state transition matrix; $\bar{\mathbf{u}}$ is the innovation vector; $\bar{\mathbf{y}}$ is the measurement vector; $\mathbf{H}$ is the observation matrix; $\bar{\mathbf{w}}$ is the measurement noise vector; and $\mathbf{k}$ is the time instant. For convenience, the prior and posterior states distributions are named $\mathbf{p}_1(\bar{\mathbf{x}})$ and $\mathbf{p}_2(\bar{\mathbf{x}})$, respectively, where $\mathbf{p}_2(\bar{\mathbf{x}}) = \mathbf{p}(\bar{\mathbf{x}}|\bar{\mathbf{y}})$ is the conditional distribution, given the measurement information. Figure 7.7 depicts the propagation of the states distribution.

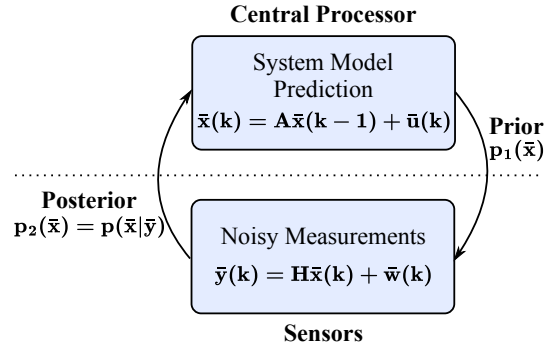


Figure 7.7: Propagation of the states distributions in the conventional Kalman filter approach.

Consider that $\bar{\mathbf{x}} \sim \mathcal{N}(\bar{\boldsymbol{\mu}}_{\mathbf{x}}, \boldsymbol{\Sigma}_{\mathbf{x}})$, $\bar{\mathbf{u}} \sim \mathcal{N}(0, \boldsymbol{\Sigma}_{\mathbf{u}})$, $\bar{\mathbf{w}} \sim \mathcal{N}(0, \boldsymbol{\Sigma}_{\mathbf{w}})$ are normally distributed vectors, where $\bar{\boldsymbol{\mu}}_{\mathbf{x}}$ is the mean vector of the states, and $\boldsymbol{\Sigma}_{\mathbf{x}}$, $\boldsymbol{\Sigma}_{\mathbf{u}}$, $\boldsymbol{\Sigma}_{\mathbf{w}}$ are the states, innovation and measurement noise covariance matrices, respectively. At each step, we have:

1. Prediction:

$$\bar{\boldsymbol{\mu}}_{\mathbf{x}1}(k) = \mathbf{A}\bar{\boldsymbol{\mu}}_{\mathbf{x}1}(k-1) \quad (7.34)$$

$$\boldsymbol{\Sigma}_{\mathbf{x}1}(k) = \mathbf{A}\boldsymbol{\Sigma}_{\mathbf{x}1}(k-1)\mathbf{A}^T + \boldsymbol{\Sigma}_{\mathbf{u}} \quad (7.35)$$

Then, $\mathbf{p}_1(\bar{\mathbf{x}}) \sim \mathcal{N}(\bar{\boldsymbol{\mu}}_{\mathbf{x}1}(k), \boldsymbol{\Sigma}_{\mathbf{x}1}(k))$.

2. Update:

$$\bar{\boldsymbol{\mu}}_{\mathbf{x}2}(k) = [\mathbf{I} - \mathbf{K}(k)\mathbf{H}]\bar{\boldsymbol{\mu}}_{\mathbf{x}1}(k) + \mathbf{K}(k)\bar{\mathbf{y}}(k) \quad (7.36)$$

$$\boldsymbol{\Sigma}_{\mathbf{x}2}(k) = [\mathbf{I} - \mathbf{K}(k)\mathbf{H}]\boldsymbol{\Sigma}_{\mathbf{x}1}(k) \quad (7.37)$$

where $\mathbf{K}(k) = \boldsymbol{\Sigma}_{\mathbf{x}1}(k)\mathbf{H}^T[\mathbf{H}\boldsymbol{\Sigma}_{\mathbf{x}1}(k)\mathbf{H}^T + \boldsymbol{\Sigma}_{\mathbf{w}}]^{-1}$ is the optimal Kalman gain. Then $\mathbf{p}_2(\bar{\mathbf{x}}) \sim \mathcal{N}(\bar{\boldsymbol{\mu}}_{\mathbf{x}2}(k), \boldsymbol{\Sigma}_{\mathbf{x}1}(k))$.

Multi-Dimensional Quantized Kalman Filter

The multi-dimensional quantized Kalman filter approach follows the same idea as the scalar version in [140]. The goal is to analyze the impact of quantization in a Kalman Filter state estimation, considering that the measurements are a noise-free representation of the predicted states through a quantization process represented by a function $\mathbf{Q}(\bar{\mathbf{x}})$. Now, the measurement model is replaced by (7.38):

$$\bar{\mathbf{y}}(\mathbf{k}) = \mathbf{Q}(\bar{\mathbf{x}}(\mathbf{k})) \quad (7.38)$$

Note that for simplicity, $\mathbf{H} = \mathbf{I}$. The state propagation for this approach is displayed in Figure 7.8.

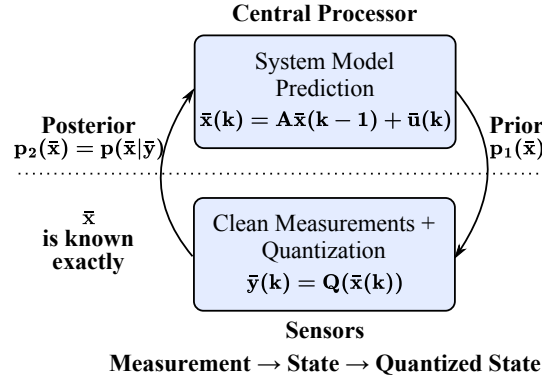


Figure 7.8: Propagation of the state distributions in the quantized Kalman filter approach.

The goal is to communicate a finite-bit representation of $\bar{\mathbf{x}}$ ($\bar{\mathbf{x}}_q$). The sum of the number of bits allocated to each state is the rate R . The distortion D associated with this representation is given by:

$$D = \sum_{i=1}^m |x_i - x_{qi}|^2. \quad (7.39)$$

The greater R , the lower the distortion D . The relation of D as a function of R is called the rate-distortion function $R(D)$.

Rate distortion of Independent Gaussian Sources

Assuming that the components of $\bar{\mathbf{x}}$ are i.i.d. random variables $x_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ for $i = 1, 2, \dots, n$, where n is the number of states, $\Sigma_{\mathbf{x}}$ is a diagonal matrix whose diagonal entries are the individual variances σ_i^2 . The rate-distortion function for independent Gaussian sources is given by [160]:

$$R(D) = \sum_{i=1}^m \frac{1}{2} \log_2 \left(\frac{\sigma_i^2}{D_i} \right) \quad (7.40)$$

where $D = \sum_{i=1}^m D_i$ is given by:

$$D_i = \begin{cases} \delta, & \delta < \sigma_i^2 \\ \sigma_i^2, & \delta \geq \sigma_i^2 \end{cases} \quad (7.41)$$

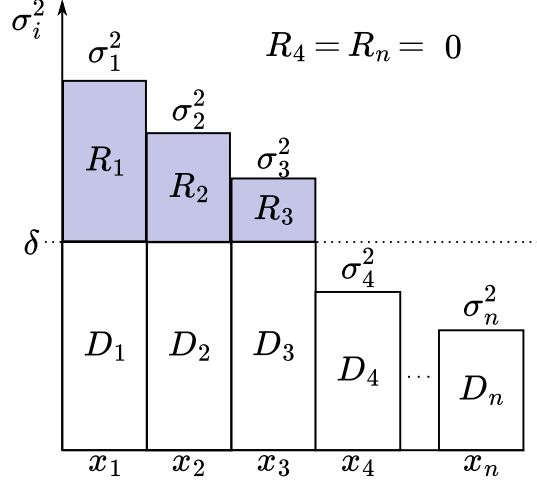


Figure 7.9: Reverse water filling for independent Gaussian sources.

where δ is the maximum individual distortion for per state, set to be the same for all states. This process results in reverse water filling, where the higher the variance, the more bits are allocated to each individual state, as shown in Figure 7.9. Those states with a variance lower than δ will be assigned 0 bits.

Even if the components of $\bar{\mathbf{x}}$ are not independent, $\Sigma_{\mathbf{x}}$ is a symmetric positive definite matrix and, therefore, diagonalizable. A change of basis can be used for this covariance matrix. Let $E[\bar{\mathbf{x}}\bar{\mathbf{x}}^T] = \Sigma_{\mathbf{x}} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T$ be an eigendecomposition. Set $\bar{\mathbf{z}} = \mathbf{Q}^T\bar{\mathbf{x}}$. The covariance of $\bar{\mathbf{z}}$ is:

$$\Sigma_{\mathbf{z}} = E[\bar{\mathbf{z}}\bar{\mathbf{z}}^T] = \mathbf{Q}^T \Sigma_{\mathbf{x}} \mathbf{Q} \quad (7.42)$$

$$= \mathbf{Q}^T \mathbf{Q} \mathbf{\Lambda} \mathbf{Q}^T \mathbf{Q} \quad (7.43)$$

$$= \mathbf{\Lambda} \quad (7.44)$$

According to [160], $\bar{\mathbf{x}}$ and $\bar{\mathbf{z}}$ have the same rate-distortion function. The reverse water filling is then performed using the eigenvalues of $\Sigma_{\mathbf{x}}$. It means that the operation is performed on the basis of the eigenvectors of $\Sigma_{\mathbf{x}}$.

On top of that, it is possible to predetermine the basis in which the reverse water-filling is performed. From equation (7.35), and considering a fixed state transition matrix $\mathbf{A}$, the prior state covariance matrix in steady-state is:

$$\Sigma_{\mathbf{x}_1} = \delta \mathbf{A} \mathbf{A}^T + \Sigma_{\mathbf{u}} \quad (7.45)$$

Four possible cases are described below.

- **Case 1:** when $\mathbf{A} = c\mathbf{I}$ and $\Sigma_{\mathbf{u}} = \sigma_{\mathbf{u}}^2\mathbf{I}$, where c is a known constant. Then:

$$\Sigma_{\mathbf{x}_1} = (c\delta + \sigma_{\mathbf{u}}^2)\mathbf{I} \quad (7.46)$$

In this case, $\Sigma_{\mathbf{x}_1}$ is diagonal, so the reverse water-filling can be performed on any basis.

- **Case 2:** when $\mathbf{A} = c\mathbf{I}$ and $\Sigma_{\mathbf{u}} \neq \sigma_{\mathbf{u}}^2\mathbf{I}$. Then:

$$\Sigma_{\mathbf{x}_1} = c\delta\mathbf{I} + \Sigma_{\mathbf{u}}. \quad (7.47)$$

In this case, $\Sigma_{\mathbf{x}_1}$ and $\Sigma_{\mathbf{u}}$ share the same eigenvectors. Then, the basis for the reverse water-filling is the eigenvectors of $\Sigma_{\mathbf{u}}$.

- **Case 3:** when $\mathbf{A} \neq c\mathbf{I}$ and $\Sigma_{\mathbf{u}} = \sigma_{\mathbf{u}}^2\mathbf{I}$. Then:

$$\Sigma_{\mathbf{x}_1} = \delta\mathbf{A}\mathbf{A}^T + \sigma_{\mathbf{u}}^2\mathbf{I} \quad (7.48)$$

In this case, $\Sigma_{\mathbf{x}_1}$ and $\mathbf{A}\mathbf{A}^T$ share the same eigenvectors. Then, the basis for the reverse water-filling is the eigenvectors of $\mathbf{A}\mathbf{A}^T$.

- **Case 4:** when $\mathbf{A} \neq c\mathbf{I}$ and $\Sigma_{\mathbf{u}} \neq \sigma_{\mathbf{u}}^2\mathbf{I}$. In this case, according to equation (7.45) the basis for the reverse water-filling is given by the eigenvectors of $(\delta\mathbf{A}\mathbf{A}^T + \Sigma_{\mathbf{u}})$.

Reverse Water-Filling and Quantized Kalman Filter

It turns out that the update step computes the posterior distribution by individually quantizing each state. It is important to note that the precise way of individually quantizing the states would be calculating real posterior pdfs by convolving each one with their respective noise pdf. However, in [140] is demonstrated that approximating the individual posterior distributions as Gaussian is a fair approach since the prediction innovations are normally distributed.

After the prediction step given by equations (7.34) and (7.35), the posterior mean and covariance are:

$$\bar{\mu}_{\mathbf{x}_2}(k) = \bar{\mu}_{\mathbf{x}_1}(k) + \mathbf{Q}(\bar{\mathbf{y}}(k) - \bar{\mu}_{\mathbf{x}_1}(k)) \quad (7.49)$$

$$\Sigma_{\mathbf{x}_2} = \text{diag}(\Delta) \quad (7.50)$$

where Δ is the vector of the computed distortions from the Reverse water-filling process according to equation (7.41), so $\Delta = [D_1 \ D_2 \ \dots \ D_n]$.

In steady-state, all variances will be increased above the threshold δ in the prediction step according to $\Sigma_{\mathbf{u}}$, and reduced to the same level δ in the update step. This process is well depicted in Figure 7.10.

Calculating Maximum Individual Distortion (δ) For a Fixed Total Rate (R)

The reverse water-filling process defines the minimum individual rate necessary to maintain a constant individual distortion level δ . However, the goal is to calculate what would be δ for a total rate $R = \sum_{i=1}^m R_i$, and how to allocate the bits for each state individually. In steady state $\Sigma_{\mathbf{x}}$ vary from $\Sigma_{\mathbf{x}_1}$ to $\Sigma_{\mathbf{x}_2}$ and back. Then:

$$(\Sigma_{\mathbf{x}_1} = \delta\mathbf{A}\mathbf{A}^T + \Sigma_{\mathbf{u}}) \leftrightarrow (\Sigma_{\mathbf{x}_2} = \delta\mathbf{I}) \quad (7.51)$$

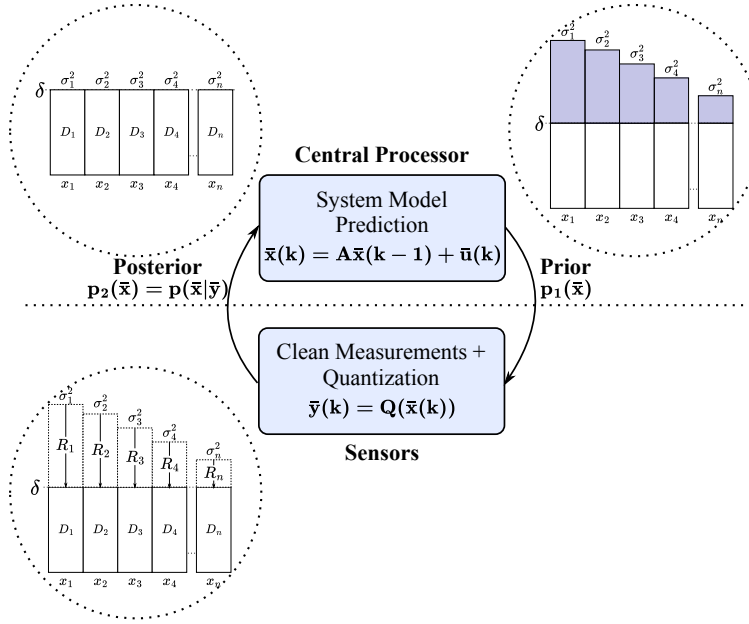


Figure 7.10: Quantized Kalman filter and reverse water-filling.

Let λ be the vector of eigenvalues of the matrix $(\delta \mathbf{A} \mathbf{A}^T + \Sigma_{\mathbf{u}})$. For a fixed R , equation (7.52) may be solved for δ through conventional numerical methods:

$$R = \sum_{i=1}^m \frac{1}{2} \log_2 \left(\frac{\lambda_i}{\delta} \right) \quad (7.52)$$

Practical Implementation

As a practical implementation we employ scalar quantization of each of the states independently. It is understood that there is a cost in bit rate; getting closer to the rate distortion bound requires vector quantization which we have not yet considered. For scalar quantizers, uniform quantization also simplifies the implementation and also achieves excellent performance [159]. In the context of this document, the uniform quantization process requires calculating the bin widths and the reconstruction points based on the maximum allowed distortion δ .

In a simple implementation, the same state individual distortion can be achieved by making the quantizers' bin widths (W_i) equal ($W_1 = W_2 = \dots = W_n = W$). The following subsections introduce 2 possible bin width distribution approximations explored in this document. Two ways of calculating W and the reconstruction points are explored.

- **Uniform Distribution Approximation**

Using the distortion measure from (7.39), the individual distortion is equivalent to the poste-

rior variance, i.e., the conditional variance:

$$D_i = \delta = \text{var}(x_i|y_i), \forall i = 1, 2, \dots, n. \quad (7.53)$$

The posterior state distribution is approximated as a uniform distribution, and consequently, the individual distortions are set to a uniform variance. Then, W is given by:

$$D_i = \delta = \frac{W^2}{12} \rightarrow W = \sqrt{12\delta} \quad (7.54)$$

After calculating W , the quantization edges $b_i, \forall i \in \{1, 2, \dots, n, n+1\}$ are determined. The reconstruction points $a_i, \forall i \in \{1, 2, \dots, n\}$ are calculated as the middle point of the bins:

$$a_i = \frac{b_{i+1} - b_i}{2} \quad (7.55)$$

- **Gaussian Approximation** The second way approximates the posterior distribution as Gaussian, using the same concept depicted in Fig. 7.4. It calculates W using the actual conditional variance for Gaussian random variables. Conservatively, let $j = \arg \max_{j \in \{1, \dots, n\}} \Sigma_{\mathbf{x}, jj}$ be the index of the state with larger variance. Then, knowing that the j^{th} element of $\bar{\mathbf{x}}$ is $x_j \sim \mathcal{N}(\mu_j, \sigma_j^2)$ in steady-state, then:

$$\delta = \text{var}[x_j|y_j] = \text{var}\left[x_j \mid -\frac{W}{2} \leq x_j \leq \frac{W}{2}\right] \quad (7.56)$$

$$\delta = E\left[x_j^2 \mid -\frac{W}{2} \leq x_j \leq \frac{W}{2}\right] - \left(E\left[x_j \mid -\frac{W}{2} \leq x_j \leq \frac{W}{2}\right]\right)^2 \quad (7.57)$$

where,

$$E\left[x_j \mid -\frac{W}{2} \leq x_j \leq \frac{W}{2}\right] = \frac{\int_{-\frac{W}{2}}^{\frac{W}{2}} x_j f(x_j) dx_j}{F(\frac{W}{2}) - F(-\frac{W}{2})} \quad (7.58)$$

$$E\left[x_j^2 \mid -\frac{W}{2} \leq x_j \leq \frac{W}{2}\right] = \frac{\int_{-\frac{W}{2}}^{\frac{W}{2}} x_j^2 f(x_j) dx_j}{F(\frac{W}{2}) - F(-\frac{W}{2})} \quad (7.59)$$

Note that f is the pdf, and F is the cdf of x_j . W may be found solving equations (7.56) and (7.57) through conventional numerical methods. The reconstruction points are calculated as the conditional mean conditional on the bin interval:

$$a_i = E[x|b_i \leq x \leq b_{i+1}] \quad (7.60)$$

When assigning all quantizers the same bin width W , it turns out that those states with higher variances result in quantizers with higher entropy rates (H), and, consequently, more bits are needed to keep constant distortion. Then, it is easy to see from the example of Figure 7.11 that:

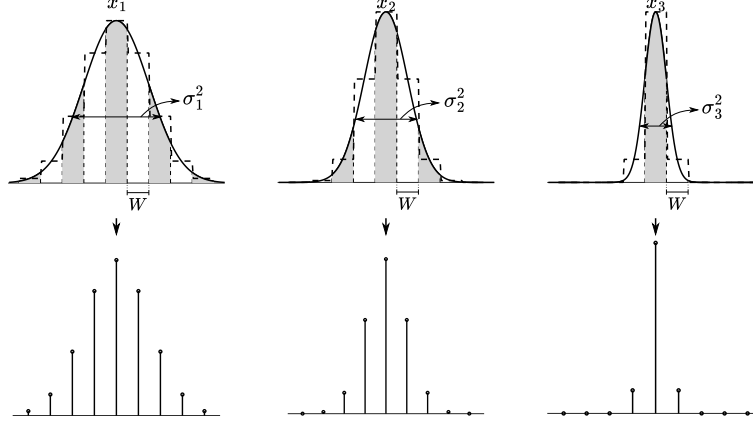


Figure 7.11: Practical implementation example.

$$\sigma_1^2 > \sigma_2^2 > \sigma_3^2 \rightarrow H_1 > H_2 > H_3 \rightarrow R_1 > R_2 > R_3 \quad (7.61)$$

It is possible to combine the two approximations to facilitate implementation and keep the accuracy within a desired range. For instance, one may calculate the bin widths using the uniform approximation, but determine the reconstruction points using the Gaussian approximation.

7.3 Results

7.3.1 One-Dimensional System Results

Using (7.1), a one-dimensional random walk of 10^6 points was implemented in MATLAB. Each of the quantized Kalman Filter-based state estimators was performed to track the states. This Section compares the results of the 2-bin and the proposed 3-bin 1-bit quantization methods. It also compares the results obtained using the Gaussian approximations with those obtained by calculating the convolution between the posterior distribution and the state prediction noise distribution. The compared quantized 1-bit estimators are listed below:

- 2-bin quantizer with pdf calculations
- 2-bin quantizer with Gaussian approximations
- 3-bin quantizer with pdf calculations
- 3-bin quantizer with changing variance Gaussian approximations
- 3-bin quantizer with average variance Gaussian Approximations

The script initiates the prior distribution as standard Gaussian, so $p_1(x) \sim \mathcal{N}(0, 1)$ at $k = 0$. Also, the state prediction noise is Gaussian $U \sim \mathcal{N}(0, 3)$.

The methods are evaluated by calculating the MSE between the actual and the estimated state sequence. The estimated sequence comprises the posterior means $\mu_2(k)$ calculated at each instant. The plot of the random walk simulations and the state estimates are depicted in Fig. 7.12 for each quantization method. In the plots, only 350 points from the 10^6 are displayed to help visualization. Additionally, Table 7.1 lists the respective calculated MSE and calculated steady-state entropy.

Table 7.1: Simulation Results.

Quantizer	MSE	Steady-state Entropy
2-bin with pdf calculations	1.7678	1.0000 bits
2-bin with Gaussian approximations	1.7627	1.0000 bits
3-bin with pdf calculations	1.6314	1.0010 bits
3-bin with changing variance Gaussian approximations	1.6016	1.0085 bits
3-bin with average variance Gaussian approximations	1.5978	1.0068 bits

The proposed 3-bin quantizer results improve the estimator MSE, while the steady-state entropy is practically kept at 1 bit. Along with MSE improvement, the plots show that the proposed 1-bit quantization with three bins shows a reduced granular noise around the estimation compared to the conventional 2-bin approach, where the granular noise is very present. Furthermore, the MSE results show that calculating the prior distributions by discrete convolution does not improve the MSE; in fact, it is slightly worse. This indicates that if the noise of state prediction is Gaussian distributed, the posterior state distributions can be approximated as Gaussian without significant performance loss. Also, calculating convolutions in each time frame may require high computational performance. It is aggravated once the calculations need to be performed at both the command center and the sensors.

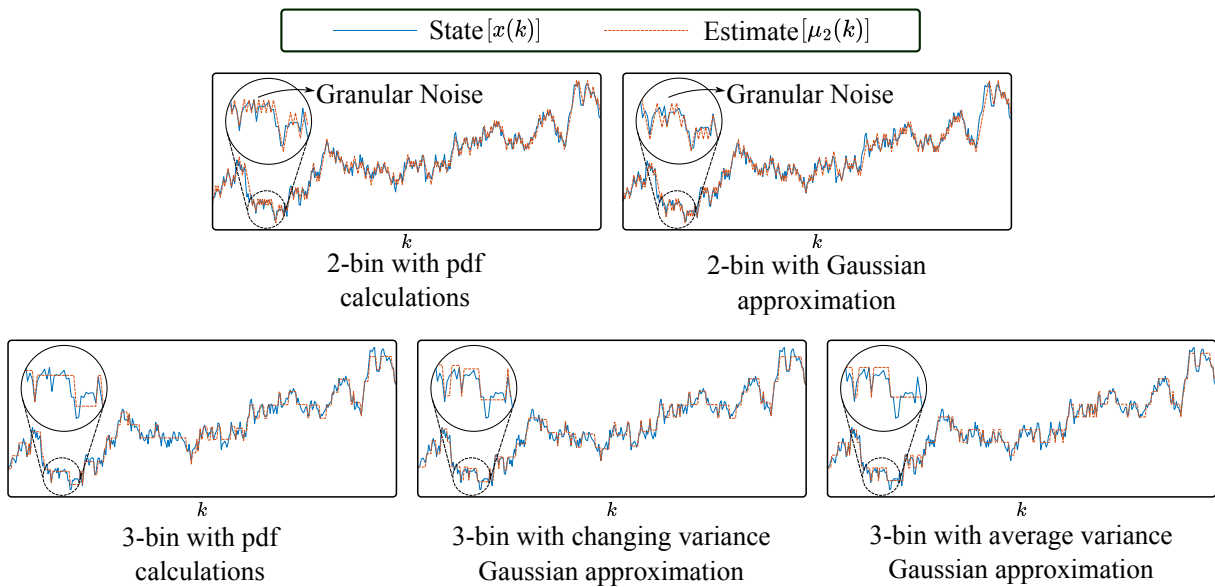


Figure 7.12: True state and quantized state estimator for each quantization method.

Note that if the state prediction noise is non-Gaussian, the prior distribution may not be well approximated by a Gaussian. Hence, the calculations for the bin edges may change, and there may be some performance degradation. Further examination of the performance under a variety of conditions remains an area for further study.

7.3.2 Multi-Dimensional System Results

For the multi-dimensional method, vector random walks of 10^4 points were simulated in MATLAB for a system with three states, i.e., $\bar{\mathbf{x}} = [x_1 \ x_2 \ x_3]^T$. For each case described in 7.2.2, simulations were performed for different values of total bit-rate R , and also for different quantizers, depending on how the bin width and the reconstruction points are calculated, according to the uniform quantizers described in 7.2.2. Four quantizers are explored:

- **Quantizer 1:** bin width and reconstruction points calculated through uniform distribution approximation.
- **Quantizer 2:** bin width calculated through uniform approximation, and reconstruction points calculated through Gaussian approximation.
- **Quantizer 3:** bin width calculated through Gaussian approximation, and reconstruction points calculated through uniform approximations.
- **Quantizer 4:** bin width and reconstruction points calculated through Gaussian approximation.

The parameters used in the simulations of each case are described below.

- Σ_u for cases 2 and 4:

$$\Sigma_u = \begin{bmatrix} 0.505 & 0.099 & 0.23 \\ 0.099 & 0.531 & 0.083 \\ 0.23 & 0.083 & 0.764 \end{bmatrix} \quad (7.62)$$

- A matrix for cases 3 and 4:

$$\mathbf{A} = \begin{bmatrix} 0.5 & 0.2 & 0.1 \\ 0.1 & 0.4 & 0.2 \\ 0.2 & 0.1 & 0.3 \end{bmatrix} \quad (7.63)$$

- $c = 1$, for cases 1 and 2.
- $\sigma_u^2 = 1$, for cases 1 and 3.

The script keeps track of the prior and posterior variances of each state. Using the reverse water-

filling method, the posterior variances are updated until they reach the calculated maximum allowed distortion δ . It is also important to note that different cases have different δ values, as they are associated with distinct parameters.

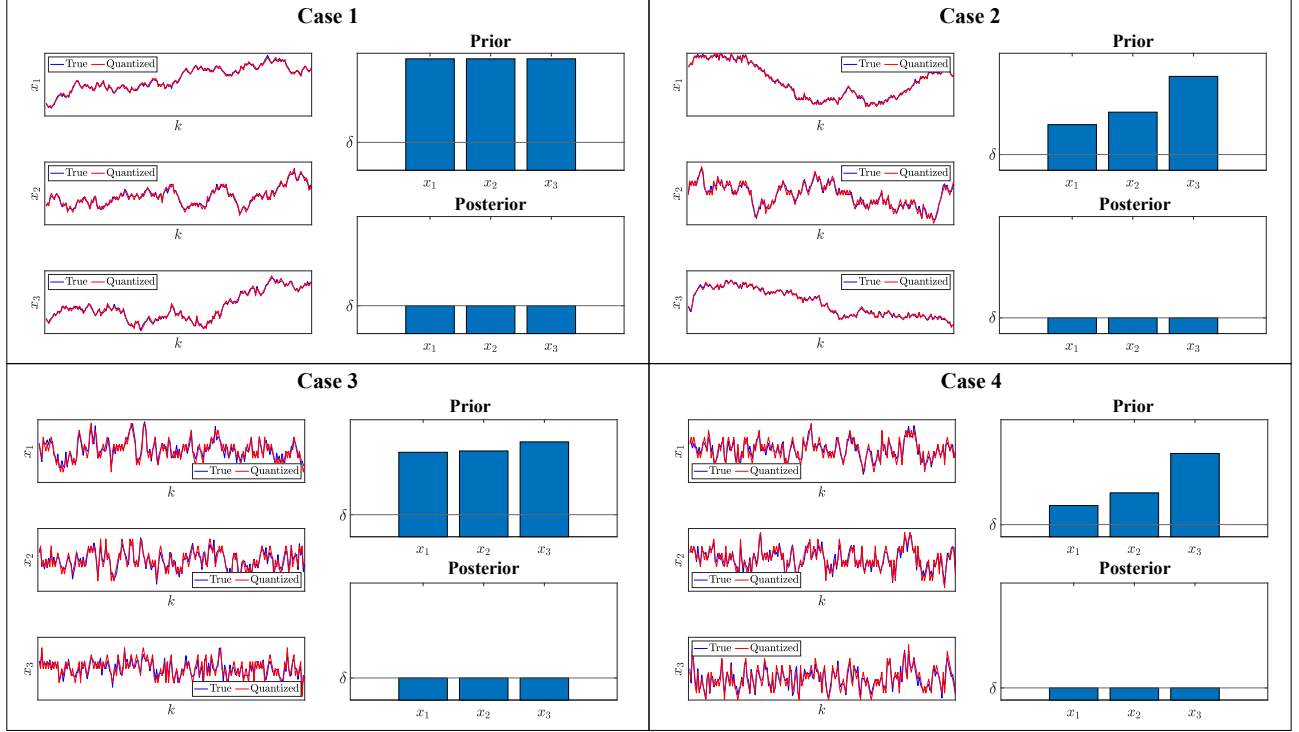


Figure 7.13: Portion of all cases random walk simulations with Quantizer 1, and prior/posterior variances in steady-state.

The plots in Figure 7.13 show, as an example, the true and quantized random walks for each state in all cases, as well as the steady-state prior and posterior variances, when using Quantizer 1 and the total bit-rate is $R = 3$ bits. Note that for case 1, in the prediction step, the prior variances are increased by the same amount, since the $\mathbf{A} = \mathbf{I}$ and $\Sigma_{\mathbf{u}}$ is diagonal. In the other cases, either $\mathbf{A} \neq \mathbf{I}$ or $\Sigma_{\mathbf{u}}$ is not diagonal, and therefore the prior variances are increased by different amounts, and consequently, quantized with different rates.

More details are listed in tables 7.2 and 7.3. Tables 7.2a, 7.2b, 7.2c, and 7.2d show the results of cases 1, 2, 3, and 4, respectively, when the total bit is $R = 3$ bits. They display the calculated δ , the achieved rates, calculated as the practical entropies of the sequences, the theoretical rates from the reverse water-filling process, and the mean squared error (MSE) between the true and quantized sequences for each state. All those are repeated using the different quantizers described previously.

For case 1, the theoretical bit rates are evenly distributed, with 1 bit per state, while the results for the other cases have different rates per state. The achieved rates are slightly above the theoretical values due to the approximations in the bin distributions. The MSEs are close to the calculated value of δ , but different quantizers show different performances. Quantizers 1 and 3 yield worse MSE

Table 7.2: Results for $R = 3$ bits.

(a) Case 1 ($\delta = 0.3333$)

Quantizer	State	Achieved Rate (bits)	Theoretical Rate (bits)	MSE
1	x_1	1.4014	1.0000	0.3318
	x_2	1.4324	1.0000	0.3387
	x_3	1.4155	1.0000	0.3351
2	x_1	1.3842	1.0000	0.2701
	x_2	1.3707	1.0000	0.2679
	x_3	1.3896	1.0000	0.2663
3	x_1	1.3645	1.0000	0.3779
	x_2	1.3714	1.0000	0.3724
	x_3	1.3632	1.0000	0.3691
4	x_1	1.3329	1.0000	0.2941
	x_2	1.3412	1.0000	0.2950
	x_3	1.3128	1.0000	0.2861

(b) Case 2 ($\delta = 0.1881$)

Quantizer	State	Achieved Rate (bits)	Theoretical Rate (bits)	MSE
1	x_1	1.3734	0.7681	0.1918
	x_2	1.3647	0.9428	0.1883
	x_3	1.5549	1.2890	0.1872
2	x_1	1.3323	0.7681	0.1452
	x_2	1.3454	0.9428	0.1474
	x_3	1.5432	1.2890	0.1555
3	x_1	1.3258	0.7681	0.2048
	x_2	1.3538	0.9428	0.2052
	x_3	1.5096	1.2890	0.2053
4	x_1	1.2947	0.7681	0.1614
	x_2	1.3137	0.9428	0.1575
	x_3	1.4728	1.2890	0.1688

(c) Case 3 ($\delta = 0.2639$)

Quantizer	State	Achieved Rate (bits)	Theoretical Rate (bits)	MSE
1	x_1	1.6544	0.9684	0.2636
	x_2	1.7067	0.9795	0.2624
	x_3	1.7452	1.0521	0.2638
2	x_1	1.6081	0.9684	0.2248
	x_2	1.6175	0.9795	0.2243
	x_3	1.6362	1.0521	0.2268
3	x_1	1.6301	0.9684	0.2981
	x_2	1.6673	0.9795	0.2966
	x_3	1.6745	1.0521	0.2973
4	x_1	1.5310	0.9684	0.2531
	x_2	1.5697	0.9795	0.2521
	x_3	1.6022	1.0521	0.2532

(d) Case 4 ($\delta = 0.1450$)

Quantizer	State	Achieved Rate (bits)	Theoretical Rate (bits)	MSE
1	x_1	1.6157	0.6815	0.1444
	x_2	1.6971	0.9282	0.1434
	x_3	1.9120	1.3903	0.1448
2	x_1	1.5426	0.6815	0.1212
	x_2	1.6095	0.9282	0.1261
	x_3	1.8435	1.3903	0.1299
3	x_1	1.5739	0.6815	0.1558
	x_2	1.6565	0.9282	0.1592
	x_3	1.8574	1.3903	0.1558
4	x_1	1.4958	0.6815	0.1300
	x_2	1.5446	0.9282	0.1350
	x_3	1.7692	1.3903	0.1418

results compared to quantizers 2 and 4, indicating that the accurate calculation of reconstruction points is crucial. Additionally, it was noted that the difference between the theoretical and achieved bit rates is higher in cases 3 and 4. Especially in case 4, the difference is very close to 1 bit, suggesting that the system is very sensitive to the state transition matrix $\mathbf{A}$ and the innovation covariance matrix Σ_u .

Table 7.3: Results for $R = 12$ bits.

(a) Case 1 ($\delta = 0.0039$)

Quantizer	State	Achieved Rate (bits)	Theoretical Rate (bits)	MSE
1	x_1	4.2439	4.0000	0.0039
	x_2	4.2682	4.0000	0.0039
	x_3	4.2588	4.0000	0.0039
2	x_1	4.2548	4.0000	0.0038
	x_2	4.2729	4.0000	0.0038
	x_3	4.2506	4.0000	0.0039
3	x_1	4.2688	4.0000	0.0039
	x_2	4.2510	4.0000	0.0039
	x_3	4.2703	4.0000	0.0039
4	x_1	4.2494	4.0000	0.0039
	x_2	4.2600	4.0000	0.0040
	x_3	4.2521	4.0000	0.0039

(b) Case 2 ($\delta = 0.0022$)

Quantizer	State	Achieved Rate (bits)	Theoretical Rate (bits)	MSE
1	x_1	4.1882	3.6861	0.0022
	x_2	4.2208	3.9369	0.0022
	x_3	4.4833	4.3771	0.0022
2	x_1	4.1950	3.6861	0.0021
	x_2	4.2212	3.9369	0.0021
	x_3	4.4790	4.3771	0.0022
3	x_1	4.1954	3.6861	0.0022
	x_2	4.2168	3.9369	0.0022
	x_3	4.4853	4.3771	0.0022
4	x_1	4.1998	3.6861	0.0022
	x_2	4.2237	3.9369	0.0021
	x_3	4.4716	4.3771	0.0022

(c) Case 3 ($\delta = 0.0039$)

Quantizer	State	Achieved Rate (bits)	Theoretical Rate (bits)	MSE
1	x_1	4.4639	3.9995	0.0039
	x_2	4.5248	3.9997	0.0039
	x_3	4.5690	4.0008	0.0039
2	x_1	4.4822	3.9995	0.0039
	x_2	4.5218	3.9997	0.0039
	x_3	4.5738	4.0008	0.0039
3	x_1	4.5017	3.9995	0.0039
	x_2	4.5371	3.9997	0.0039
	x_3	4.5810	4.0008	0.0039
4	x_1	4.4798	3.9995	0.0039
	x_2	4.5282	3.9997	0.0039
	x_3	4.5997	4.0008	0.0039

(d) Case 4 ($\delta = 0.0022$)

Quantizer	State	Achieved Rate (bits)	Theoretical Rate (bits)	MSE
1	x_1	4.3720	3.6847	0.0022
	x_2	4.4807	3.9367	0.0022
	x_3	4.7746	4.3786	0.0022
2	x_1	4.3571	3.6847	0.0022
	x_2	4.5061	3.9367	0.0022
	x_3	4.7766	4.3786	0.0022
3	x_1	4.3869	3.6847	0.0021
	x_2	4.4893	3.9367	0.0022
	x_3	4.7805	4.3786	0.0022
4	x_1	4.3731	3.6847	0.0021
	x_2	4.4911	3.9367	0.0022
	x_3	4.7558	4.3786	0.0022

To illustrate a high-bit rate scenario, Tables 7.3a, 7.3b, 7.3c and 7.3d show the same analysis for when $R = 12$ bits. Again, in case 1, the theoretical rates are evenly distributed among the states, with 4 bits per state, while in the other cases, they are distributed unevenly. Now, the MSE are very consistent with the calculated δ , for all quantizers. The achieved rates are also very consistent among the quantizers, even though they are still above the theoretical bounds. However, it is also possible to notice that when compared to the low bit rate scenario, the difference between the

theoretical and achieved rates is lower. This suggests that the performance of the quantizers is more severely affected in low-bit-rate scenarios.

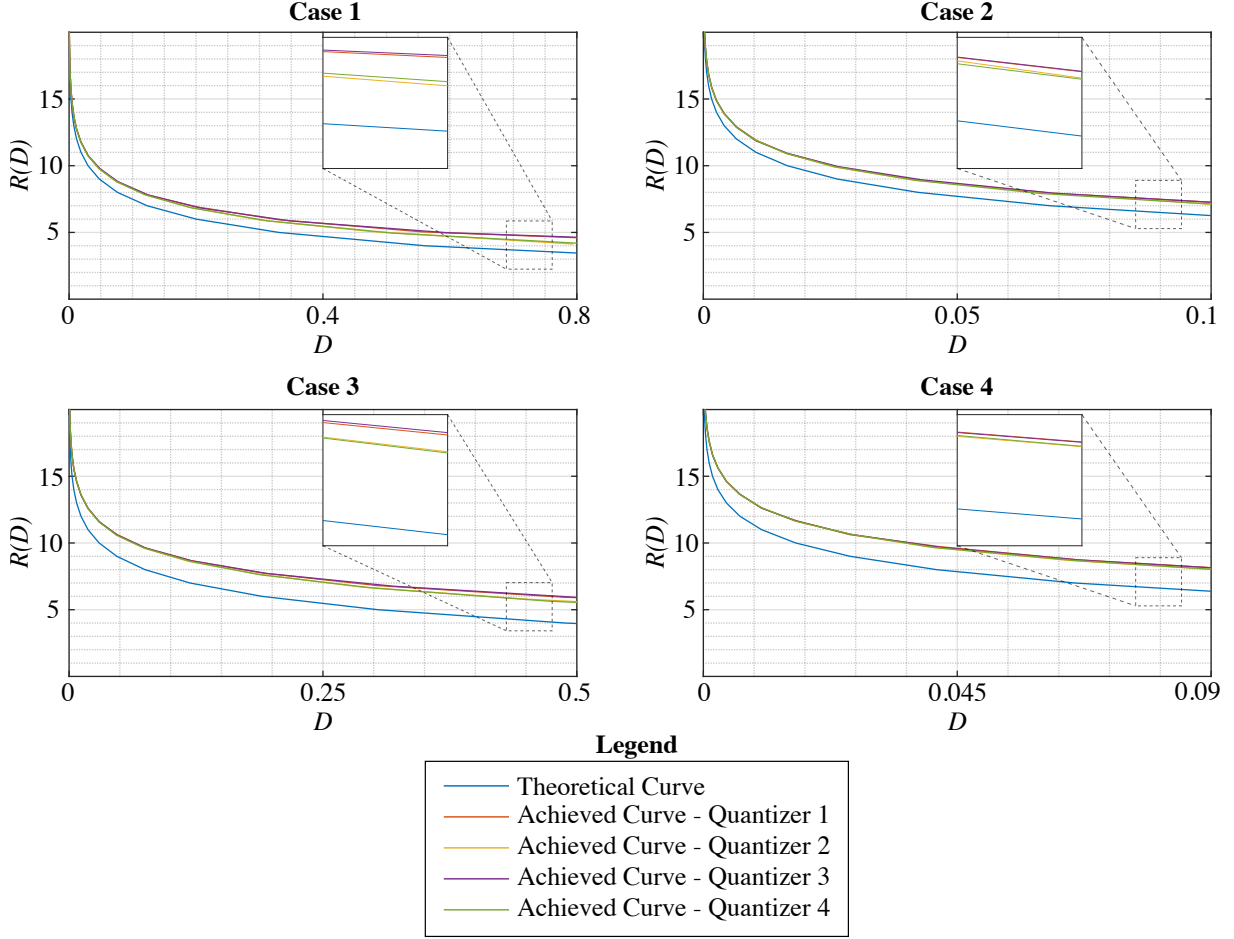


Figure 7.14: Theoretical and achieved rate-distortion curves of each quantizer and all study cases.

For high bit rates, Gish and Pierce [157] proved that the cost in bits associated with using a high-bit-rate scalar quantization on Gaussian sources is given by [161, 162]:

$$R_c = \frac{1}{2} \log_2 \left(\frac{\pi e}{6} \right) \approx 0.255 \text{ bits} \quad (7.64)$$

The simulations were performed for $R = 3$ up to $R = 33$ bits, and the rate-distortion curves are plotted in Figure 7.14. The plots show that the results are off from the Gish-Pierce theory by a factor of 2 per state, especially in case 4.

7.4 Conclusion

The one-dimensional Kalman Filter-based state estimation with quantization proposed and approach that replaces the measurement noise with the quantization process. It also proposes that approximating the posterior state distribution as Gaussian is possible with no significant information loss, given that state prediction noise is normally distributed. The analyzed system model considers state estimation with 1-bit quantization to demonstrate applications with communication bandwidth constraints.

The work also proposes a novel 1-bit quantization with three bins as opposed to the 2-bin conventional approach. It shows that for sufficiently large sequences, it is possible to encode the state by splitting the distribution into three bins and still have an entropy rate of 1 bit since the central bin is highly probable. The results indicate an improvement in MSE from the conventional to the proposed method and a significant reduction in granular noise.

The results also compare simulations where the actual prior distributions were calculated by discrete convolution of the quantized posterior distribution with the state prediction noise distribution with those where Gaussian approximations were considered. They indicate that calculating the convolutions is not worth the computational effort, most likely due to errors introduced in the discrete convolution itself. The Gaussian approximation approaches performed fairly well with little or no performance degradation.

The work was expanded to a multi-dimensional system, where the same idea of replacing measurement noise with quantization noise stands. The method proposes to quantize each element of the state vector independently, according to the rate-distortion function of independent Gaussian sources. The process uses the reverse-water approach to distribute a fixed total bit-rate among the states according to their individual variances. Even if the states are not independent the method proposes to change the operation basis in order to make independent scalar quantization possible, and because of that, four possible cases are enabled. The first case considers the state transition matrix and the innovation covariance matrices to be multiples of the identity, allowing the calculations to be performed on any basis. The second case changes the innovation covariance matrix to any symmetric positive-definite matrix, and keep the state transition matrix as a multiple of the identity. In this case, the basis is given by the eigenvectors of the innovation covariance matrix. The third case assumes the state transition matrix to be any matrix different from identity, and the innovation covariance matrix to be a multiple of the identity. In this case, the operation basis is given by the eigenvectors of the product of the transition matrix and its transpose. The last case considers both the state transition and the innovation covariance matrices to be different from the identity. Now, the basis is given by the eigenvectors of the product of the transition matrix with its transpose plus the innovation covariance matrix.

The goal is to build a practical implementation where the individual distortion of the states are reduced to the same level in steady-state. This was achieved by making the bin widths of all quantizers equal. For all cases, simulations were performed considering four different scalar uniform

quantizers, determined by how the bin widths and the reconstruction points are calculated, whether using a uniform distribution, Gaussian distribution approximations, or a mix of them. The results show better results for those quantizers in which the reconstruction points are calculated through Gaussian approximations at low bit-rates. In all cases, the mean-squared-error are reduced to the theoretical maximum allowed distortion. In addition, the achieved rate-distortion curves differ from the theoretical one by the expected amount estipulated by the Gish and Pierce constant.

References

- [1] I. Akingeneye, J. Wu, and J. Yang, “Optimum PMU Placement for Power System State Estimation,” in *IEEE Power and Energy Society General Meeting*, 2017.
- [2] T. Baldwin, L. Mili, M. Boisen, and R. Adapa, “Power System Observability with Minimal Phasor Measurement Placement,” *IEEE Transactions on Power Systems*, vol. 8, no. 2, pp. 707–715, 1993.
- [3] S. Chakrabarti, E. Kyriakides, and D. G. Eliades, “Placement of Synchronized Measurements for Power System Observability,” *IEEE Transactions on Power Delivery*, vol. 24, no. 1, pp. 12–19, 2009.
- [4] W. Yuill, A. Edwards, S. Chowdhury, and S. P. Chowdhury, “Optimal PMU Placement: A Comprehensive Literature Review,” in *IEEE Power and Energy Society General Meeting*, 2011.
- [5] J. L. Adam Cooper, Mike Shusterr, “Electric company smart meter deployments: Foundation for a smart grid (2021 update),” Institute for Electric Innovation, The Edison Foundation, Tech. Rep., 2021.
- [6] Aclara, “SGM1300 Smart Energy Meter Family,” Hubbell, Tech. Rep. M-0917SGM13900uk, 2017.
- [7] C. Huang, C.-C. Sun, N. Duan, Y. Jiang, C. Applegate, P. D. Barnes, and E. Stewart, “Smart meter pinging and reading through AMI two-way communication networks to monitor grid edge devices and DERs,” *IEEE Transactions on Smart Grid*, vol. 13, no. 5, pp. 4144–4153, 2022.
- [8] M. Baran and A. Kelley, “State estimation for real-time monitoring of distribution systems,” *IEEE Trans. on Power Syst.*, vol. 9, no. 3, pp. 1601–1609, 1994.
- [9] A. Angioni, J. Shang, F. Ponci, and A. Monti, “Real-Time Monitoring of Distribution System Based on State Estimation,” *IEEE Transactions on Instrumentation and Measurement*, vol. 65, no. 10, pp. 2234–2243, 2016.
- [10] Y. Wang, Q. Chen, T. Hong, and C. Kang, “Review of Smart Meter Data Analytics: Applications, Methodologies, and Challenges,” *IEEE Transactions on Smart Grid*, vol. 10, no. 3, pp. 3125–3148, 2019.
- [11] J. Milosevic, M. Dahan, S. Amin, and H. Sandberg, “A network monitoring game with heterogeneous component criticality levels,” in *58th IEEE Conference on Decision and Control (CDC)*. IEEE, 2019, pp. 4379–4384.
- [12] W. Zhang and Y. Mei, “Bandit Change-Point Detection for Real-Time Monitoring High-Dimensional Data Under Sampling Control,” *Technometrics*, vol. 65, no. 1, pp. 33–43, 2023.
- [13] X. Xian, A. Semenov, Y. Hu, A. Wang, and Y. Jin, “Adaptive Sampling and Quick Anomaly Detection in Large Networks,” *IEEE Transactions on Automation Science and Engineering*, vol. 20, no. 4, pp. 2253–2267, Oct. 2023.

- [14] J. Grant, A. Boukouvalas, R.-R. Griffiths, D. Leslie, S. Vakili, and E. M. De Cote, “Adaptive Sensor Placement for Continuous Spaces,” in *36th International Conference on Machine Learning (ICML)*, vol. 97. PMLR, Jun 2019, pp. 2385–2393.
- [15] Y. Yang, A. Blanchard, T. Sapsis, and P. Perdikaris, “Output-Weighted Sampling for Multi-Armed Bandits with Extreme Payoffs,” *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 478, no. 2260, p. 20210781, Apr. 2022.
- [16] B. Bahamondes and M. Dahan, “Network Inspection from Locations with Imperfect Detection Capabilities,” in *American Control Conference (ACC)*, 2022, pp. 613–620.
- [17] J. Milosevic, M. Dahan, S. Amin, and H. Sandberg, “Strategic monitoring of networked systems with heterogeneous security levels,” *IEEE Transactions on Control of Network Systems*, vol. 11, no. 3, pp. 1165–1176, 2023.
- [18] O. Abramenko and A. Jung, “Graph signal sampling via reinforcement learning,” in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2019, pp. 3077–3081.
- [19] G. Cavraro, A. Bernstein, V. Kekatos, and Y. Zhang, “Real-Time Identifiability of Power Distribution Network Topologies With Limited Monitoring,” *IEEE Control Systems Letters*, no. 2, pp. 325–330, 2020.
- [20] A. Krause, J. Leskovec, C. Guestrin, J. VanBriesen, and C. Faloutsos, “Efficient Sensor Placement Optimization for Securing Large Water Distribution Networks,” *Journal of Water Resources Planning and Management*, vol. 134, no. 6, pp. 516–526, 2008.
- [21] T. Lattimore and C. Szepesvari, *Bandit Algorithms*. Cambridge University Press, Jul. 2020.
- [22] C. Zeng, Q. Wang, S. Mokhtari, and T. Li, “Online Context-Aware Recommendation with Time Varying Multi-Armed Bandit,” in *22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Aug. 2016, pp. 2025–2034.
- [23] M. Valko, R. Munos, B. Kveton, and T. Kocák, “Spectral Bandits for Smooth Graph Functions,” in *Proc. 31st ICML*, E. P. Xing and T. Jebara, Eds., vol. 32, no. 2. PMLR, 22–24 Jun 2014, pp. 46–54.
- [24] T. Kocák, M. Valko, R. Munos, and S. Agrawal, “Spectral Thompson Sampling,” *AAAI Conference on Artificial Intelligence*, vol. 28, no. 1, Jun. 2014, number: 1.
- [25] T. Kocák, R. Munos, B. Kveton, S. Agrawal, and M. Valko, “Spectral Bandits,” *Journal of Machine Learning Research*, vol. 21, no. 1, Jan. 2020.
- [26] P. Thaker, M. Malu, N. Rao, and G. Dasarathy, “Maximizing and Satisficing in Multi-armed Bandits with Graph Information,” in *Advances in Neural Information Processing Systems*, vol. 35. Curran Associates, Inc., 2022, pp. 2019–2032.
- [27] Y. Yue and C. Guestrin, “Linear Submodular Bandits and their Application to Diversified Retrieval,” in *Advances in Neural Information Processing Systems*, J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, and K. Q. Weinberger, Eds., vol. 24. Curran Associates, Inc., 2011.
- [28] N. Cesa-Bianchi and G. Lugosi, “Combinatorial Bandits,” *Journal of Computer and System Sciences*, vol. 78, no. 5, pp. 1404–1422, 2012.
- [29] D. Deka *et al.*, “Learning Distribution Grid Topologies: A Tutorial,” *IEEE Trans. Smart Grid*, vol. 15, no. 1, pp. 999–1013, 2024. [Online]. Available: <https://ieeexplore.ieee.org/document/10113752/>

- [30] Q. Li, R. Negi, and M. D. Ilić, “Phasor Measurement Units Placement for Power System State Estimation: A Greedy Approach,” in *IEEE Power and Energy Society General Meeting*. IEEE, 2011.
- [31] V. Kekatos, G. B. Giannakis, and B. Wollenberg, “Optimal Placement of Phasor Measurement Units via Convex Relaxation,” *IEEE Transactions on Power Systems*, vol. 27, no. 3, pp. 1521–1530, 2012.
- [32] D. Zelazo and M. Bürger, “On the Definiteness of the Weighted Laplacian and its Connection to Effective Resistance,” in *53rd IEEE Conference on Decision and Control*, 2014, pp. 2895–2900.
- [33] Y. Chen, S. Z. Khong, and T. T. Georgiou, “On the Definiteness of Graph Laplacians with Negative Weights: Geometrical and Passivity-Based Approaches,” in *American Control Conference (ACC)*, 2016, pp. 2488–2493.
- [34] G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher, “An Analysis of Approximations for Maximizing Submodular Set Functions—I,” *Mathematical Programming*, vol. 14, pp. 265–294, 1978.
- [35] M. Baran and F. Wu, “Network Reconfiguration in Distribution Systems for Loss Reduction and Load Balancing,” *IEEE Transactions on Power Delivery*, vol. 4, no. 2, pp. 1401–1407, Apr. 1989.
- [36] R. Sen, A. Rakhlin, L. Ying, R. Kidambi, D. Foster, D. N. Hill, and I. S. Dhillon, “Top-k eXtreme Contextual Bandits with Arm Hierarchy,” in *38th International Conference on Machine Learning (ICML)*, ser. Proceedings of Machine Learning Research, M. Meila and T. Zhang, Eds., vol. 139. PMLR, 18–24 Jul 2021, pp. 9422–9433.
- [37] M. J. Wainwright, *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*, ser. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2019.
- [38] R. Vershynin, *High-Dimensional Probability: An Introduction with Applications in Data Science*, ser. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2025.
- [39] Y. Abbasi-yadkori, D. Pál, and C. Szepesvári, “Improved Algorithms for Linear Stochastic Bandits,” in *Advances in Neural Information Processing Systems*, vol. 24. Curran Associates, Inc., 2011.
- [40] J. A. Ghijsselen and A. P. Van den Bossche, “Exact voltage unbalance assessment without phase measurements,” *IEEE Transactions on Power Systems*, vol. 20, no. 1, pp. 519–520, 2005.
- [41] M. Zeraati, M. Shabanzadeh, M. R. Sheibani, and F. Jabari, “Meter placement algorithms to enhance distribution systems state estimation: Review, challenges and future research directions,” *IET Renewable Power Generation*, vol. 16, no. 15, pp. 3422–3444, 2022.
- [42] K. R. Mestav, J. Luengo-Rozas, and L. Tong, “Bayesian state estimation for unobservable distribution systems via deep learning,” *IEEE Transactions on Power Systems*, vol. 34, no. 6, pp. 4910–4920, 2019.
- [43] S. De la Torre, A. Conejo, and J. Contreras, “Simulating oligopolistic pool-based electricity markets: A multiperiod approach,” *IEEE Transactions on Power Systems*, vol. 18, no. 4, pp. 1547–1555, 2003.

- [44] G. Celli, P. A. Pegoraro, F. Pilo, G. Pisano, and S. Sulis, "Dms cyber-physical simulation for assessing the impact of state estimation and communication media in smart grid operation," *IEEE Transactions on Power Systems*, vol. 29, no. 5, pp. 2436–2446, 2014.
- [45] V. Zamani and M. E. Baran, "Meter placement for conservation voltage reduction in distribution systems," *IEEE Transactions on Power Systems*, vol. 33, no. 2, pp. 2109–2116, 2017.
- [46] A. S. Zamzam and N. D. Sidiropoulos, "Physics-aware neural networks for distribution system state estimation," *IEEE Transactions on Power Systems*, vol. 35, no. 6, pp. 4347–4356, 2020.
- [47] Y. Liao, Y. Weng, M. Wu, and R. Rajagopal, "Distribution grid topology reconstruction: An information theoretic approach," in *2015 North American Power Symposium (NAPS)*, 2015.
- [48] J.-H. Menke, N. Bornhorst, and M. Braun, "Distribution system monitoring for smart power grids with distributed generation using artificial neural networks," *International Journal of Electrical Power & Energy Systems*, vol. 113, pp. 472–480, 2019.
- [49] A. S. Zamzam, X. Fu, and N. D. Sidiropoulos, "Data-driven learning-based optimization for distribution system state estimation," *IEEE Transactions on Power Systems*, vol. 34, no. 6, pp. 4796–4805, 2019.
- [50] B. Zargar, A. Angioni, F. Ponci, and A. Monti, "Multiarea parallel data-driven three-phase distribution system state estimation using synchrophasor measurements," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 9, pp. 6186–6202, 2020.
- [51] A. Akrami, S. Asif, and H. Mohsenian-Rad, "Sparse tracking state estimation for low-observable power distribution systems using D-PMUs," *IEEE Transactions on Power Systems*, vol. 37, no. 1, pp. 551–564, 2021.
- [52] S. Radhoush, M. Bahramipanah, H. Nehrir, and Z. Shahoei, "A review on state estimation techniques in active distribution networks: Existing practices and their challenges," *Sustainability*, vol. 14, no. 5, p. 2520, 2022.
- [53] A. Alimardani, F. Therrien, D. Atanackovic, J. Jatskevich, and E. Vaahedi, "Distribution system state estimation based on nonsynchronized smart meters," *IEEE Transactions on Smart Grid*, vol. 6, no. 6, pp. 2919–2928, 2015.
- [54] D. K. Molzahn and L. A. Roald, "Grid-aware versus grid-agnostic distribution system control: A method for certifying engineering constraint satisfaction," in *52nd Hawaii International Conference on System Sciences (HICSS)*, January 2019.
- [55] P. Buason, S. Misra, S. Talkington, and D. K. Molzahn, "A Data-Driven Method for Locating Sensors and Selecting Alarm Thresholds to Identify Violations of Voltage Limits in Distribution Systems," *Electric Power Systems Research*, vol. 232, no. 110387, July 2024.
- [56] S. C. Ross, G. Vuylsteke, and J. L. Mathieu, "Effects of load-based frequency regulation on distribution network operation," *IEEE Transactions on Power Systems*, vol. 34, no. 2, pp. 1569–1578, 2019.
- [57] M. Baran and F. F. Wu, "Optimal sizing of capacitors placed on a radial distribution system," *IEEE Transactions on power Delivery*, vol. 4, no. 1, pp. 735–743, 1989.
- [58] W. Zheng, W. Wu, B. Zhang, and C. Lin, "Distributed optimal residential demand response considering operational constraints of unbalanced distribution networks," *IET generation, transmission & distribution*, vol. 12, no. 9, pp. 1970–1979, 2018.

- [59] “Supplemental file for real-time assessment of distribution power quality through adaptive smart meter measurements,” [ONLINE], available (2024/3/22) at https://docs.google.com/document/d/1LBtGVJJa3eQ_BRX51cdBsOK65T2OW9A7/edit?usp=sharing&oid=113500991377257474573&rtpof=true&sd=true, March 2024.
- [60] K. Baker, A. Bernstein, C. Zhao, and E. Dall’Anese, “Network-Cognizant Design of decentralized Volt/VAR controllers,” in *IEEE Power Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, Apr. 2017.
- [61] M. Zholbaryssov and A. D. Domínguez-García, “Safe Data-Driven Secondary Control of Distributed Energy Resources,” *IEEE Transactions on Power Systems*, vol. 36, no. 6, pp. 5933–5943, Nov. 2021.
- [62] X. Zhou, M. Farivar, Z. Liu, L. Chen, and S. H. Low, “Reverse and Forward Engineering of Local Voltage Control in Distribution Networks,” *IEEE Transactions on Automatic Control*, vol. 66, no. 3, pp. 1116–1128, Mar. 2021.
- [63] J. Swartz and A. Von Meier, “The Impact of Three-phase Impedances on the Stability of DER Systems,” in *North American Power Symposium (NAPS)*, 2021.
- [64] J. Swartz, F. Celi, F. Pasqualetti, and A. Von Meier, “Parameter Conditions to Prevent Voltage Oscillations Caused by LTC-Inverter Hunting on Power Distribution Grids,” in *European Control Conference (ECC)*, 2022, pp. 1118–1125.
- [65] “IEEE Standard for Interconnection and Interoperability of Distributed Energy Resources with Associated Electric Power Systems Interfaces,” *IEEE Std 1547-2018 (Revision of IEEE Std 1547-2003)*, pp. 1–138, Apr. 2018.
- [66] Hawaiian Electric Company, Inc, “Rule No. 14,” *Service connections and facilities on customer’s premises, Hawaiian Electric*, 2021. [Online]. Available: https://www.hawaiianelectric.com/documents/billing_and_payment/rates/hawaiian_electric_rules/14.pdf
- [67] “Grid Connection of Energy Systems via Inverters, Part 2: Inverter Requirements,” *AS/NZS 4777.2:2020*, 2020. [Online]. Available: <https://www.standards.org.au/standards-catalogue/standard-details?designation=as-nzs-4777-2-2020>
- [68] Australian Energy Market Operator (AEMO), “Compliance of Distributed Energy Resources with Technical Settings. Compliance to AS/NZS4777.2; A Technical Report for Australia,” pp. 1–75, April 2023. [Online]. Available: <https://aemo.com.au/-/media/files/initiatives/der/2023/compliance-of-der-with-technical-settings.pdf?la=en>
- [69] A. Inaolaji, A. Savasci, and S. Paudyal, “Distribution Grid Optimal Power Flow in Unbalanced Multiphase Networks With Volt-VAR and Volt-Watt Droop Settings of Smart Inverters,” *IEEE Trans. Ind. Appl.*, vol. 58, no. 5, pp. 5832–5843, 2022.
- [70] F. Geth, M. Vanin, W. Van Westering, T. Milford, and A. Pandey, “Making Distribution State Estimation Practical: Challenges and Opportunities,” *arXiv:2311.07021*, 2023.
- [71] V. C. Cunha et al., “Automated Determination of Topology and Line Parameters in Low Voltage Systems Using Smart Meters Measurements,” *IEEE Trans. Smart Grid*, vol. 11, no. 6, pp. 5028–5038, 2020.
- [72] M. Vanin, F. Geth, R. D’hulst, and D. Van Hertem, “Combined Unbalanced Distribution System State and Line Impedance Matrix Estimation,” *Int. J. Elect. Power Energy Syst.*, vol. 151, p. 109155, 2023.

- [73] C. C. Liu, H. Chen, J. Shi, and L. Chen, "Self-Supervised Learning Method for Consumer-level Behind-the-meter PV Estimation," *Applied Energy*, vol. 326, p. 119961, 2022.
- [74] K. Mason, M. J. Reno, L. Blakely, S. Vejdani, and S. Grijalva, "A Deep Neural Network Approach for Behind-the-meter Residential PV Size, Tilt and Azimuth Estimation," *Solar Energy*, vol. 196, pp. 260–269, 2020.
- [75] S. Talkington, S. Grijalva, M. J. Reno, and J. A. Azzolini, "Solar PV Inverter Reactive Power Disaggregation and Control Setting Estimation," *IEEE Transactions on Power Systems*, vol. 37, no. 6, pp. 4773–4784, 2022.
- [76] J. Peppanen, M. Hernandez, J. Deboever, M. Rylander, and M. J. Reno, "Distribution Load Modeling – Survey of the Industry State, Current Practices and Future Needs," in *52nd North American Power Symposium (NAPS)*, Nov. 2021.
- [77] R. C. Dugan and T. E. McDermott, "An Open-Source Platform for Collaborating on Smart Grid Research," in *IEEE Power and Energy Society General Meeting*, 2011.
- [78] M. Jafari, T. O. Olowu, and A. I. Sarwat, "Optimal Smart Inverters Volt-VAR Curve Selection with a Multi-Objective Volt-VAR Optimization using Evolutionary Algorithm Approach," *IEEE Trans. Smart Grid*, 2025.
- [79] R. Heidari and F. Geth, "Improved Algebraic Inverter Modeling for Four-Wire Power Flow Optimization," *Electric Power Systems Research*, vol. 234, p. 110825, 2024, presented at *23rd Power Systems Computation Conference (PSCC)*.
- [80] J. S. Bridle, "Probabilistic Interpretation of Feedforward Classification Network Outputs, with Relationships to Statistical Pattern Recognition," in *Neurocomputing: Algorithms, Architectures and Applications*. Springer, 1990, pp. 227–236.
- [81] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, "TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems," *arXiv:1603.04467*, 2016.
- [82] M. Deakin, M. Vanin, Z. Fan, and D. Van Hertem, "Smart energy network digital twins: Findings from a UK-based demonstrator project," *Int. J. Elect. Power Energy Syst.*, vol. 162, p. 110302, 2024.
- [83] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, 2006.
- [84] Power Quality Application Guide, "Voltage disturbances," *Standard EN*, vol. 50160, 2004.
- [85] CIGRE Task Force C6.04.02, "Benchmark systems for network integration of renewable and distributed energy resources," CIGRE International Council on Large Electric Systems, Tech. Rep., July 2009.
- [86] IEEE Std 1159-2009, "IEEE recommended practice for monitoring electric power quality," Tech. Rep., June 2009.
- [87] J. Zhao, G. Zhang, M. La Scala, Z. Y. Dong, C. Chen, and J. Wang, "Short-term state forecasting-aided method for detection of smart grid general false data injection attacks," *IEEE Transactions on Smart Grid*, vol. 8, no. 4, pp. 1580–1590, 2015.

- [88] FERC, “Assessment of Demand Response and Advanced Metering,” December 2024. [Online]. Available: <https://www.ferc.gov/power-sales-and-markets/demand-response/reports-demand-response-and-advanced-metering>
- [89] U.S. Dept. Energy, Washington, DC, USA, “Voice of experience: Leveraging AMI networks and data,” March 2019. [Online]. Available: https://www.energy.gov/sites/default/files/2024-02/01-03-2019_doe-voe-leveraging-ami-networks-and-data-report_508.0.pdf
- [90] —, “Advanced metering infrastructure and customer systems,” September 2016. [Online]. Available: https://www.energy.gov/sites/prod/files/2016/12/f34/AMI%20Summary%20Report_09-26-16.pdf
- [91] N. Duan, C. Huang, C.-C. Sun, and L. Min, “Smart meters enabling voltage monitoring and control: The last-mile voltage stability issue,” *IEEE Transactions on Industrial Informatics*, vol. 18, no. 1, pp. 677–687, 2022.
- [92] Y. Wang, Q. Chen, T. Hong, and C. Kang, “Review of smart meter data analytics: Applications, methodologies, and challenges,” *IEEE Transactions on Smart Grid*, vol. 10, no. 3, pp. 3125–3148, 2019.
- [93] S. S. S. R. Depuru, L. Wang, V. Devabhaktuni, and N. Gudi, “Smart meters for power grid—Challenges, issues, advantages and status,” in *IEEE/PES Power Systems Conference & Exposition*, 2011.
- [94] S. S. Alam, B. Natarajan, and A. Pahwa, “Distribution grid state estimation from compressed measurements,” *IEEE Transactions on Smart Grid*, vol. 5, no. 4, pp. 1631–1642, 2014.
- [95] M. Majidi, M. Etezadi-Amoli, and H. Livani, “Distribution system state estimation using compressive sensing,” *International Journal of Electrical Power & Energy Systems*, vol. 88, pp. 175–186, 2017.
- [96] P. L. Donti, Y. Liu, A. J. Schmitt, A. Bernstein, R. Yang, and Y. Zhang, “Matrix completion for low-observability voltage estimation,” *IEEE Transactions on Smart Grid*, vol. 11, no. 3, pp. 2520–2530, 2019.
- [97] R. Madbhavi, B. Natarajan, and B. Srinivasan, “Enhanced tensor completion based approaches for state estimation in distribution systems,” *IEEE Transactions on Industrial Informatics*, vol. 17, no. 9, pp. 5938–5947, 2020.
- [98] A. Mutanen, M. Ruska, S. Repo, and P. Jarventausta, “Customer classification and load profiling method for distribution systems,” *IEEE Transactions on Power Delivery*, vol. 26, no. 3, pp. 1755–1763, 2011.
- [99] E. Manitsas, R. Singh, B. C. Pal, and G. Strbac, “Distribution system state estimation using an artificial neural network approach for pseudo measurement modeling,” *IEEE Transactions on Power Systems*, vol. 27, no. 4, pp. 1888–1896, 2012.
- [100] E. A. Blood, B. H. Krogh, and M. D. Ilic, “Electric power system static state estimation through Kalman filtering and load forecasting,” in *IEEE Power and Energy Society General Meeting*, 2008.
- [101] J. Zhao, G. Zhang, Z. Y. Dong, and M. La Scala, “Robust forecasting aided power system state estimation considering state correlations,” *IEEE Transactions on Smart Grid*, vol. 9, no. 4, pp. 2658–2666, 2016.

- [102] R. Dobbe, D. Arnold, S. Liu, D. Callaway, and C. Tomlin, “Real-time distribution grid state estimation with limited sensors and load forecasting,” in *7th ACM/IEEE International Conference on Cyber-Physical Systems (ICCPS)*, 2016.
- [103] R. Dobbe, W. van Westering, S. Liu, D. Arnold, D. Callaway, and C. Tomlin, “Linear single- and three-phase voltage forecasting and Bayesian state estimation with limited sensing,” *IEEE Transactions on Power Systems*, vol. 35, no. 3, pp. 1674–1683, 2019.
- [104] R. Dutta, S. J. Geetha, S. Chakrabarti, and A. Sharma, “An l-1 regularized forecasting-aided state estimator for active distribution networks,” *IEEE Transactions on Smart Grid*, vol. 13, no. 1, pp. 191–201, 2021.
- [105] L. Gan and S. H. Low, “Convex relaxations and linear approximation for optimal power flow in multiphase radial networks,” in *18th Power Systems Computation Conference (PSCC)*, 2014.
- [106] D. B. Arnold, M. Sankur, R. Dobbe, K. Brady, D. S. Callaway, and A. Von Meier, “Optimal dispatch of reactive power for voltage regulation and balancing in unbalanced distribution systems,” in *IEEE Power and Energy Society General Meeting (PESGM)*, 2016.
- [107] M. D. Sankur, R. Dobbe, A. v. Meier, E. M. Stewart, and D. B. Arnold, “Optimal voltage phasor regulation for switching actions in unbalanced distribution systems,” in *IEEE Power & Energy Society General Meeting (PESGM)*, 2020.
- [108] A. A. Mamun, M. Sohel, N. Mohammad, M. S. Haque Sunny, D. R. Dipta, and E. Hos-sain, “A comprehensive review of the load forecasting techniques using single and hybrid predictive models,” *IEEE Access*, vol. 8, pp. 134 911–134 939, 2020.
- [109] J. Gorka and L. Roald, “Classification models for forecasting and real-time identification of solar curtailment in the California grid,” in *15th ACM International Conference on Future and Sustainable Energy Systems (e-Energy)*, 2024, pp. 158–169.
- [110] H. Shi, M. Xu, and R. Li, “Deep learning for household load forecasting—A novel pooling deep RNN,” *IEEE Transactions on Smart Grid*, vol. 9, no. 5, pp. 5271–5280, 2018.
- [111] D. Markovics and M. J. Mayer, “Comparison of machine learning methods for photovoltaic power forecasting based on numerical weather prediction,” *Renewable and Sustainable Energy Reviews*, vol. 161, p. 112364, 2022.
- [112] M. He, V. Vittal, and J. Zhang, “A sparsified vector autoregressive model for short-term wind farm power forecasting,” in *IEEE Power & Energy Society General Meeting*, 2015.
- [113] D. C. Hill, D. McMillan, K. R. W. Bell, and D. Infield, “Application of auto-regressive models to U.K. wind speed data for power system impact studies,” *IEEE Transactions on Sustainable Energy*, vol. 3, no. 1, pp. 134–141, 2012.
- [114] Y. Chakhchoukh, V. Vittal, and G. T. Heydt, “PMU based state estimation by integrating correlation,” *IEEE Transactions on Power Systems*, vol. 29, no. 2, pp. 617–626, 2014.
- [115] S. Córdova, H. Rudnick, Á. Lorca, and V. Martínez, “An efficient forecasting-optimization scheme for the intraday unit commitment process under significant wind and solar power,” *IEEE Transactions on Sustainable Energy*, vol. 9, no. 4, pp. 1899–1909, 2018.
- [116] J. Shi, Y. Liu, and N. Yu, “Spatio-temporal modeling of electric loads,” in *North American Power Symposium (NAPS)*, 2017.

- [117] M. Hassanzadeh, C. Y. Evrenosoğlu, and L. Mili, “A short-term nodal voltage phasor forecasting method using temporal and spatial correlation,” *IEEE Transactions on Power Systems*, vol. 31, no. 5, pp. 3881–3890, 2016.
- [118] P. J. Brockwell and R. A. Davis, *Modeling and Forecasting with ARMA Processes*. Cham: Springer International Publishing, 2016.
- [119] J. Walrand, *Tracking—A*. Cham: Springer International Publishing, 2021, pp. 163–192.
- [120] A. Abur and A. Gómez-Expósito, *Power System State Estimation: Theory and Implementation*. Marcel Dekker, 2004.
- [121] W. Kersting, “Radial distribution test feeders,” in *IEEE Power Engineering Society Winter Meeting*, vol. 2, 2001, pp. 908–912.
- [122] IEEE PES AMPS DSAS Test Feeder Working Group. [Online]. Available: <https://cmte.ieee.org/pes-testfeeders/resources/>
- [123] Y. Gu, “Meteorological and customer composition factors in power load forecasting,” Masters Thesis, University of Wisconsin–Madison, 2023.
- [124] R. Gupta, F. Sossan, and M. Paolone, “Grid-aware distributed model predictive control of heterogeneous resources in a distribution network: Theory and experimental validation,” *IEEE Transactions on Energy Conversion*, vol. 36, no. 2, pp. 1392–1402, 2021.
- [125] EPFL’s Distributed Electrical Systems Laboratory, “DESL-Photovoltaic-timeseries.” [Online]. Available: <https://github.com/DESL-EPFL/DESL-Photovoltaic-timeseries>
- [126] IEC, *Instrument transformers - Part 3: Additional requirements for inductive voltage transformers*, IEC Std 61869-3:2011, 2017.
- [127] —, *Instrument transformers - Part 2: Additional requirements for current transformers*, IEC Std 61869-2:2012, 2012.
- [128] A. Rangarajan, R. K. Gupta, D. K. Molzahn, and L. A. Roald, “Forecast-Aided State Estimation in Unbalanced Distribution Networks Using Smart Meter Data under Limited Communication Bandwidth,” *Sustainable Energy, Grids and Networks*, vol. 43, no. 101863, September 2025, Presented at the *12th IREP Bulk Power system Dynamics and Control Symposium*.
- [129] R. Gray and D. Neuhoff, “Quantization,” *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2325–2383, 1998.
- [130] A. Gholami, S. Kim, Z. Dong, Z. Yao, M. W. Mahoney, and K. Keutzer, “A Survey of Quantization Methods for Efficient Neural Network Inference,” in *Low-Power Computer Vision*. Chapman and Hall/CRC, 2022, pp. 291–326.
- [131] Y. Plan and R. Vershynin, “The Generalized LASSO with Non-Linear Observations,” *IEEE Trans. Inf. Theory*, vol. 62, no. 3, pp. 1528–1537, 2016.
- [132] C. Thrampoulidis and A. S. Rawat, “The Generalized LASSO for Sub-Gaussian Measurements With Dithered Quantization,” *IEEE Trans. Inf. Theory*, vol. 66, no. 4, pp. 2487–2500, 2020.
- [133] V. Vapnik, *The Nature of Statistical Learning Theory*. Springer Science & Business Media, 2013.
- [134] N. Rin, I. Shames, I. R. Petersen, and E. L. Ratnam, “Electric Grid Topology and Admittance Estimation: Quantifying Phasor-based Measurement Requirements,” to appear in the *American Control Conference (ACC)*, 2025. [Online]. Available: <https://arxiv.org/abs/2410.17553>

- [135] R. Gray and T. Stockham, “Dithered Quantizers,” *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 805–812, 1993.
- [136] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky, “The Convex Geometry of Linear Inverse Problems,” *Foundations of Computational Mathematics*, vol. 12, no. 6, pp. 805–849, 2012.
- [137] S. Bolognani and F. Dörfler, “Fast Power System Analysis via Implicit Linearization of the Power Flow Manifold,” in *53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 2015, pp. 402–409.
- [138] O. Stanojev, L. Werner, S. Low, and G. Hug, “Tractable Identification of Electric Distribution Networks,” in *62nd IEEE Conference on Decision and Control (CDC)*, 2023, pp. 549–555.
- [139] S. Bolognani and K. Moffat, “Chapter Six – Grid Topology Identification via Distributed Statistical Hypothesis Testing,” in *Big Data Application in Power Systems*, 2nd ed. Elsevier, 2024, pp. 109–134.
- [140] P. A. de Alcantara, F. Bezerra Costa, and D. R. Fuhrmann, “One-dimensional Kalman filter with measurement noise replaced by quantization,” in *59th Annual Conference on Information Sciences and Systems (CISS)*, 2025.
- [141] R. E. Kalman, “A new approach to linear filtering and prediction problems,” *J. Basic Eng.*, vol. 82, no. 1, pp. 35–45, 1960.
- [142] S. Sun, J. Lin, L. Xie, and W. Xiao, “Quantized Kalman Filtering,” in *22nd IEEE International Symposium on Intelligent Control*, 2007.
- [143] Z. Shen, J. Chen, and W. Su, “State Estimation for Linear Systems with Communication Channels,” in *32nd Chinese Control Conference*, 2013.
- [144] F. Yao, G. Li, X. Bai, and C. Hao, “Unscented Kalman Filter for Distribution Networks with Quantization Effects and False Data Injection Attacks,” in *2023 International Conference on Power System Technology (PowerCon)*, 2023.
- [145] M. D. Howard and Z. Qu, “An optimal kalman-consensus filter for distributed implementation over a dynamic communication network,” *IEEE Access*, vol. 9, pp. 66 696–66 706, 2021.
- [146] B. Rao and H. Durrant-Whyte, “Fully decentralised algorithm for multisensor Kalman filtering,” *IEE Proceedings D (Control Theory and Applications)*, vol. 138, pp. 413–420, 1991.
- [147] L. Sheng, W. Huai, Y. Niu, and M. Gao, “Consensus-based unscented kalman filtering over sensor networks with communication protocols,” *International Journal of Robust and Non-linear Control*, vol. 31, no. 13, pp. 6349–6368, 2021.
- [148] Q. Zhao, W. Zhang, T. Zhang, J. Chen, and H. Ye, “Robust consensus-based secondary control for microgrids considering communication delay and noises: A kalman filter-based approach,” *IEEE Transactions on Smart Grid*, vol. 16, no. 4, pp. 2791–2802, 2025.
- [149] M. S. Grewal and A. P. Andrews, *Kalman Filtering: Theory and Practice Using MATLAB*. John Wiley & Sons, 2015.
- [150] D. Simon, *Optimal state estimation: Kalman, H infinity, and nonlinear approaches*. John Wiley & Sons, 2006.
- [151] V. Pravong, J.-P. Condomines, G. O. Lundin, and S. Puechmorel, “Unscented Kalman Filter using Optimal Quantization,” in *27th International Conference on Information Fusion (FUSION)*, 2024.

- [152] L. P. Qian, A. Feng, N. Yu, W. Xu, and Y. Wu, “Vehicular Networking-Enabled Vehicle State Prediction via Two-Level Quantized Adaptive Kalman Filtering,” *IEEE Internet of Things Journal*, vol. 7, no. 8, 2020.
- [153] S. Liu, Z. Wang, J. Hu, and G. Wei, “Protocol-based extended Kalman filtering with quantization effects: The Round-Robin case,” *International Journal of Robust and Nonlinear Control*, vol. 30, no. 18, 2020.
- [154] D. Li, S. Kar, F. E. Alsaadi, A. M. Dobaie, and S. Cui, “Distributed kalman filtering with quantized sensing state,” *IEEE Transactions on Signal Processing*, vol. 63, no. 19, pp. 5180–5193, 2015.
- [155] E. J. Msechu, A. Ribeiro, S. I. Roumeliotis, and G. B. Giannakis, “Distributed Kalman filtering based on quantized innovations,” in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2008, pp. 3293–3296.
- [156] H. R. Schindler, “Delta Modulation,” *IEEE Spectrum*, vol. 7, no. 10, 1970.
- [157] H. Gish and J. Pierce, “Asymptotically efficient quantizing,” *IEEE Transactions on Information Theory*, vol. 14, no. 5, pp. 676–683, 1968.
- [158] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley & Sons, 2006.
- [159] R. Gallager, *Principles of Digital Communication*. Cambridge University Press Cambridge, UK, 2008, vol. 1.
- [160] T. M. Cover, *Elements of Information Theory*. John Wiley & Sons, 2006.
- [161] T. Koch and G. Vazquez-Vilar, “A general rate-distortion converse bound for entropy-constrained scalar quantization,” in *IEEE International Symposium on Information Theory (ISIT)*, 2016, pp. 735–739.
- [162] M. F. Pradier, P. M. Olmos, and F. Perez-Cruz, “Entropy-constrained scalar quantization with a lossy-compressed bit,” *Entropy*, vol. 18, no. 12, p. 449, 2016.