DEPARTMENT OF APPLIED MECHANICS AND
BIOMEDICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY MADRAS
CHENNAI – 600036

Advanced Estimation Strategies for Sparsely Observed Cyber-Physical Systems

A Thesis

Submitted by

RAHUL MADBHAVI

For the award of the degree

Of

DOCTOR OF PHILOSOPHY

July 2023

DEPARTMENT OF APPLIED MECHANICS AND
BIOMEDICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY MADRAS
CHENNAI – 600036

Advanced Estimation Strategies for Sparsely Observed Cyber-Physical Systems

A Thesis

Submitted by

RAHUL MADBHAVI

For the award of the degree

Of

DOCTOR OF PHILOSOPHY

July 2023

This thesis is dedicated to my parents and grandparents for their support, motivation, and encouragement during my education.

THESIS CERTIFICATE

This is to undertake that the Thesis titled **ADVANCED ESTIMATION STRATEGIES FOR SPARSELY OBSERVED CYBER-PHYSICAL SYSTEMS**, submitted by me to the Indian Institute of Technology Madras, for the award of **Doctor of Philosophy**, is a bona fide record of the research work done by me under the supervision of **Dr. Babji Srinivasan**. The contents of this Thesis, in full or in parts, have not been submitted to any other Institute or University for the award of any degree or diploma.

Chennai 600036 **Rahul Madbhavi**

Date: July 2023

Dr. Babji Srinivasan
Research advisor
Associate Professor
Department of Applied Mechanics
IIT Madras

LIST OF PUBLICATIONS

I. REFEREED JOURNALS BASED ON THESIS

a. **R. Madbhavi**, B. Natarajan and B. Srinivasan, "Enhanced Tensor Completion Based Approaches for State Estimation in Distribution Systems," in IEEE Transactions on Industrial Informatics, vol. 17, no. 9, pp. 5938-5947, Sept. 2021, doi: 10.1109/TII.2020.3035449.

b. **R. Madbhavi**, B. Natarajan and B. Srinivasan, "Graph Neural Network-Based Distribution System State Estimators," in IEEE Transactions on Industrial Informatics, doi: 10.1109/TII.2023.3248082.

c. **R. Madbhavi**, B. Natarajan, and B. Srinivasan, "Meter placement approaches for matrix completion-based distribution system state estimator," in Electric Power Systems Research, vol. 213, p. 108687, 2022, doi: 10.1016/j.epsr.2022.108687.

d. **R. Madbhavi** and B. Srinivasan, "Data-driven dictionaries to enhance the performance of compressive sensing-based state estimators," in International Journal of Advances in Engineering Sciences and Applied Mathematics, vol. 14, no. 3, pp. 94–107, Dec. 2022, doi: 10.1007/s12572-022-00324-1.

II. PRESENTATIONS IN CONFERENCES

a. **R. Madbhavi**, A. Joshi, S. Munikoti, L. Das, P. K. Mohapatra and B. Srinivasan, "Data Driven Leak Localization in Water Distribution Networks," in 2019 Indian Process Systems Engineering Conference, IIT Madras, Chennai, Tamil Nadu, India **(Poster Session)**.

III. PUBLICATIONS IN CONFERENCE PROCEEDINGS

a. **R. Madbhavi**, H. S. Karimi, B. Natarajan and B. Srinivasan, "Tensor Completion based State Estimation in Distribution Systems," in 2020 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT), Washington, DC, USA, 2020, pp. 1-5, doi: 10.1109/ISGT45199.2020.9087747.

b. **R. Madbhavi**, B. Natarajan and B. Srinivasan, "Distribution System Topology Identification using Graph Neural Networks," in 2023 IEEE PES GT&D Conference and Exposition, Istanbul, Turkey, May 2023.

c. **R. Madbhavi**, A. Joshi, S. Munikoti, L. Das, P. K. Mohapatra and B. Srinivasan, "Sensor Placement for Leak Localization in Water Distribution

Networks using Machine Learning," in 2020 IEEE International Conference on Computing, Power and Communication Technologies (GUCON), Greater Noida, India, 2020, pp. 95-100, doi: 10.1109/GUCON48875.2020.9231148.

d. **R. Madbhavi** and B. Srinivasan, "Enhancing Performance of Compressive Sensing-based State Estimators using Dictionary Learning," in 2022 IEEE International Conference on Power Systems Technology (POWERCON), Kuala Lumpur, Malaysia, 2022, pp. 1-6, doi: 10.1109/POWERCON53406.2022.9930028.

IV. OTHER WORKS

a. R. Srinivasan, B. Srinivasan, **R. Madbhavi**, S. Sriram, T. Valuthottiyil Shajahan, "Device and method for multi-user eye-tracking," India, 421055, Feb. 21, 2022.

b. T. Valuthottiyil Shajahan, **R. Madbhavi**, M. A. Shahab, B. Srinivasan, R. Srinivasan, "Dhrushti-AI: A multi-screen multi-user eye-tracking system to understand the cognitive behavior of humans in process industries," in 2023 ESCAPE Conference, Athens, Greece, 2023.

ACKNOWLEDGEMENTS

Many people have helped me during my time as a Ph.D. research scholar. I want to take this opportunity to thank them all.

Firstly, I am sincerely grateful to my research advisor Dr. Babji Srinivasan for his support and guidance. The works presented in this dissertation would not have been possible without open discussions and the freedom to pursue ideas. Dr. Babji has always encouraged me to think independently and has assisted me in every way he could throughout my Ph.D. study. His willingness to be available at any time of the day for discussions motivated me and gave me timely feedback and guidance. I truly value my time as his student.

Secondly, I thank Dr. Balasubramaniam Natarajan from Kansas State University for the learnings, guidance, and opportunities. Dr. Bala has always encouraged me to think independently and guided me in my research throughout my Ph.D. study. I also thank Dr. Bala for the opportunity to work with him as a visiting scholar. It helped me develop as a research scholar and taught me to maintain a good work-life balance.

My special appreciation is towards my doctoral committee members: Dr. Ramakrishnan S, Dr. Nirav Pravinbhai Bhatt, Dr. Shaikh Faruque Ali, and Dr. Satyanarayanan Seshadri, for their suggestions and insights. I also thank the committee members for helping me at various stages during my transfer to IIT Madras. I also thank my previous doctoral committee members from IIT Gandhinagar, Dr. Ravi Hegde, and Dr. Mithun Radhakrishna, for their input during my initial years as a research scholar.

I have also enjoyed working with a helpful and supportive group of friends over the years. I want to thank them all. Firstly, I thank my friends from IIT Gandhinagar, Amit, Ayush, Jaivik, Laya, Parag, Punit, Sai, Shraddha, and Umair. The environment in the lab made me feel at ease in my formative years as a Ph.D. scholar. Secondly, I thank my friends

from IIT Gandhinagar, Aabila, Aatif, Ahtesham, Deepesh, Dinesh, Mahindra, Nasre, Pravesh, and Shubhankar for their support and numerous discussions over the years. I thank my friends from Kansas State University, Alaleh and Hazhar. Thank you both for your help throughout my time as a visiting scholar at the K-State campus. I also thank my friends at IIT Madras, Anish, Anushree, Darpan, Haider, Jagadeeshwari, Karthikeyan A S, Karthikeyan, Mayank, Pragathi, Saravanan C, Saravanan M, Sriram, Syam, and Thasnimol. I have enjoyed the time spent with all of you. Special thanks to my kind, caring, understanding, and supportive friend, Prakalya. Finally, a thank you to all of my friends and fellow researchers I missed mentioning. It has been a long journey, and your presence has made the experience memorable.

I'm grateful to my parents, Gopal S Madbhavi and Anjali, for their love, support, and encouragement. Special thanks to my grandparents, Sheshappa Madbhavi, Shantabai Madbhavi, Monappa Pattar, and Sulochana Pattar, and great-grandparents Malleshappa Madbhavi and Ambamma Madbhavi. Thank you for teaching me the value of education and encouraging my career path.

Some parts of this work are partly supported by the Science and Engineering Research Board, Department of Science and Technology, Government of India, under award number MTR/2019/000328.

Some parts of this work are based upon work supported by the Department of Energy, Office of Energy Efficiency and Renewable Energy (EERE), Solar Energy Technologies Office, under Award Number DE-EE0008767.

ABSTRACT

Modernizing traditional systems (such as power, water, and gas distribution systems) with sensing, computation, control, and communication infrastructure has led to the development of smart systems. This modernization has been motivated by the need for improved reliability, resiliency, operational efficiency, and safety. However, various challenges exist in the control and operation of these systems. Consider power distribution systems delivering electrical power to consumers. Firstly, unlike transmission systems, power distribution systems operate under low-observable conditions due to the lack of measurements. Secondly, these systems suffer from other issues, such as unbalanced operation and communication channel congestions due to rising deployments of granular-level metering equipment. Finally, the increased penetration of newer technologies on the distribution side, such as PV generation, distributed energy generation, and electric vehicles, have resulted in bi-directional power flow. This new prosumer paradigm has increased the risk of grid instabilities and dangerous operating conditions. To ensure the safe operation of distribution systems, grid operators require knowledge of the system states obtained using state estimators to take corrective actions. However, the low-observability conditions result in the poor performance of current state-of-the-art techniques. Therefore, in this dissertation, we develop novel state estimators that utilize spatiotemporal correlations, network topology, and power flow constraints to improve state estimation performance under low-observability conditions. We also tackle problems such as topology identification, meter placement, and error analysis to aid the developed state estimators.

Firstly, we formulate the state estimation problem as a low-rank tensor completion problem with power flow equations as constraints. The formulation introduces feature scaling and weighted tensor nuclear norm to improve the sensitivity of the estimates to all physical quantities in the tensor. We demonstrate the superiority of this approach over existing

sparsity-aware techniques. Secondly, we address the challenges of reducing computational time and eliminating the need for complete knowledge of network parameters. We achieve this objective by converting the tensor completion-based state estimators into their graph neural network (GNN) equivalents. A comprehensive evaluation of over twenty techniques is performed to demonstrate the effectiveness of the developed GNN-based state estimators. These estimators perform comparably to the tensor completion-based approaches while requiring several orders of magnitude lower computational time. Further, these estimators learn the interdependencies in the data without prior knowledge of the network parameters.

The model-based approaches are then optimized for performance. We derive error bounds for state estimates of the model-based state estimators and introduce a technique to optimize hyperparameters. We then exploit the interdependencies among the data to develop meter placement strategies for determining metering locations. These techniques maximize the information available to the developed state estimators and thus aim to reduce state estimation errors. Next, we address the data channel congestion issues associated with increasing metering deployments by developing data-driven dictionaries. We then utilize these dictionaries with compressive sensing-based frameworks. We demonstrate the need to use dictionaries tailored to the data to obtain superior data compression and reconstruction performance over deterministic dictionaries. Finally, we study problems in other cyber-physical systems, such as leak detection in water distribution networks, and develop meter placement strategies.

In summary, this dissertation offers novel state estimators that have superior performance over current state-of-the-art state estimators and provides solutions for their associated problems, such as computational time reduction, topology identification, error bounds for state estimates, meter placement, and data channel decongestion. This thesis can also be a foundation for future works to create distributed implementations capable of operating on large real-world distribution systems within computational time constraints. Utilities

may also not have complete knowledge of all network parameters. Therefore, updating the developed approaches to incorporate partial knowledge of the network parameters would improve the real-world applicability of the developed techniques. Finally, these updated formulations can be used to estimate the unknown network parameters.

KEYWORDS Cyber-physical systems; smart grids; distribution systems; state estimation; topology identification; error analysis; meter placement;

CONTENTS

LIST OF TABLES

LIST OF FIGURES

NOTATIONS

Θ	Distribution system model parameters.
Ω	Set of indices of known entries in the measurement matrix
m, n	Number of buses and branches, respectively.
p	Number of phases.
N_T	Number of time instants in the state tensor.
$\boldsymbol{P_t^{in}}$	Active power injections vector at time t.
$\boldsymbol{Q_t^{in}}$	Reactive power injections vector at time t.
$\boldsymbol{v_{ref}}$	Reference bus voltage vector.
$\boldsymbol{V_t}$	Bus voltage phasor vector at time t.
$\boldsymbol{M_t}$	State measurement matrix at time t.
$\boldsymbol{X}$	Estimate of the measurement matrix
$\boldsymbol{\mathcal{T}}$	State measurement tensor.
$\boldsymbol{\Lambda_t}$	Availability matrix at time t.
$\boldsymbol{\Lambda}$	Availability tensor.
$\mathbf{G}$	Graph representation of the distribution system.
$\mathbf{V}$	Set of all buses in the distribution system.
$\mathbf{E}$	Set of all branches in the distribution system.
$\otimes$	Kronecker product of matrices.
$\odot$	Hadamard product of matrices.
$\partial f(.)$	Subgradient of a function
$diag(.)$	Diagonal matrix using the elements in the provided vector
$rank(.)$	Rank of a matrix
$trace(.)$	Trace of a matrix
$\mathcal{P}_\Omega(.)$	Orthogonal projection of a matrix onto the span of matrices vanishing outside of Ω
$\mathcal{D}_\tau(.)$	Soft-thresholding at level τ applied on the singular values of a matrix
$\lvert.\rvert$	Cardinality of a set
$\lVert.\rVert_*$	Nuclear norm
$\lVert.\rVert_1$	L1 norm
$\lVert.\rVert_F$	Frobenius norm
$\lVert.\rVert_{spectral}$	Spectral norm
$\langle .\,,. \rangle$	Inner product

CHAPTER 1

INTRODUCTION

1.1 OVERVIEW AND MOTIVATION

The push toward grid modernization has resulted in the introduction of newer technologies and metering equipment into traditional distribution systems. This modernization is motivated by the need for improved reliability, resiliency, operational efficiency, and safety. Governments worldwide have shown significant interest in grid modernization owing to their benefits. In recent years, there has been a significant push towards adopting renewable generation. These developments together with newer technologies such as electric vehicles have resulted in new dynamics between the grid components, see Chen *et al.* (2016). For example, modern grids allow bi-directional energy transfer and thus results in grid stability and security issues, see Powalko *et al.* (2012). These developments have necessitated the need for better situational awareness to control these systems.

Distribution systems have generally been low-observable systems due to limited measurement equipments. Most state estimation techniques rely on the system to be observable or to be made observable through pseudo-measurements. While the observability issues can be resolved by installing more grid sensors, it is not feasible due to monetary and resource constraints. On the other hand pseudo-measurement generation depend on historical data and may result in large estimation errors. Therefore, there is a need to develop state estimation techniques that can provide accurate state estimate even under low-observability conditions, see Primadianto and Lu (2017) and Dehghanpour *et al.* (2019). The desired characteristics of distribution system state estimators include low state estimation errors, low parameter sensitivity, robustness to noise, and robustness to bad data. Model-free approaches that do not explicitly use the knowledge of the underlying system model have shorter computation times than

model-based approaches. However, it was shown by Madbhavi *et al.* (2021) that they fail to produce accurate state estimates. Therefore, model-based approaches can be created to exploit the system constraints and provide superior state estimates at the cost of higher computation times. Additionally, system constraints also result in specific combinations of measurements to provide better state estimation performance. Therefore, meter placement strategies can be developed by analyzing these dependencies. Finally, there is no existing work on error bounds for model-based distribution system state estimators, such as matrix completion based estimators. Therefore, it is desired to develop error bounds to quantify the state estimation errors. These bounds can enable new techniques for existing problems, such as meter placement.

While sparsity-aware techniques provide estimates with low state estimation errors, they have large computational times. Several works in the literature aim to replace computationally heavy algorithms with deep learning models that learn from the data. Therefore, the computation time challenges of model-based approaches can be overcome by developing deep learning methods which learn the interactions between the network nodes and shift the computation burden to their training phase. Among such models are deep neural networks (DNNs) that utilize multiple hidden layers between the input and output layers to perform classification and regression tasks. Several works have been proposed that utilize DNNs for distribution system state estimation such as Mestav *et al.* (2019); Manitsas *et al.* (2012). However, these models are prone to over-fitting. Furthermore, DNNs cannot incorporate known relationships in graph structures, such as node or edge information. Therefore, new techniques to improve state estimation performance and reduce computation time are desired that can also incorporate known dependencies into the deep learning approaches.

Alternatively, granular level data such as measurements obtained from smart meters can be utilized for state estimation. However, acquiring these measurements have their own challenges. For example, existing communication networks cannot handle the

data deluge due to smart meters. Currently, smart meter measurements are transmitted in 15 minutes, 60 minutes or greater time intervals. Therefore, techniques (such as data compression) need to be developed that can maximize the throughput of existing communication networks. Additionally, state estimators can also be developed that operate directly on these compressed measurements. One such example are compressive sensing-based states estimators. These estimators rely on the assumption that real-world signals are sparse in some domains. It has been shown in the literature that deterministic transformation matrices such as DCT, Haar, Hankel, Toeplitz can be effectively utilized to compress smart meter measurements. However, these transformation matrices do not necessarily generate the sparsest solutions. Data-driven dictionaries can be developed that are tailored to the data and provided superior reconstruction performance and low reconstruction errors. Therefore, this work develops data-driven dictionaries to enhance the performance of existing compressive sensing-based state estimators.

1.2 CURRENT CHALLENGES

There are several research challenges in distribution systems. For example:

- Existing model-based approaches, such as matrix completion-based state estimators, cannot utilize measurements across time to improve state estimation, this is particularly of interest in cases with sensors operating at different sampling rates. The incapability of current matrix completion-based approaches in utilizing measurements across time can be overcome by developing tensor completion-based approaches.

- Matrix completion-based distribution system state estimators that utilize power flow equations as constraints are currently not optimized from a meter placement perspective. Existing works do not address the problem of placement of sensors that optimize their performance. The introduction of power flow equations as constraints results in certain combinations of sensors to provide better state

estimation performance than others. Therefore, it is necessary to study this problem of meter placement to optimize the performance of these estimators under the given resource constraints (number/type of sensors).

- Current state estimation techniques for distribution systems assume that the topology of the network is known. However, this is not always true. Therefore, topology identification is an important precursor to state estimation. Current topology identification techniques have several drawbacks such as the need for many sensors and repeated application of state estimators to identify the topology with least deviation from observed results. These drawbacks result in financial burden on utilities to install the required number of sensors and large latencies due to the need to evaluate all possible topologies. However, advances in fields such as graph neural networks have enabled models to be built that can learn system interdependencies through data. Therefore, the problem of topology identification could be solved elegantly using graph neural networks (GNNs) while overcoming existing drawbacks.

- Current compressive sensing-based distribution system state estimators utilize deterministic transformation matrices. However, the reconstruction performance can be enhanced by learning the transformation matrices tailored for the data. This can by achieved by utilizing sensor data to learn the transformation matrices through techniques such as dictionary learning.

Given these challenges, this dissertation aims to address the following research objectives:

- Developing state estimators for distribution system state estimation (DSSE) that can provide superior performance over conventional state estimators under low observability conditions.

- Develop error bounds for the model-based distribution system state estimators.

- Developing topology identification approaches to enable topology-based model selection in model-based distribution system state estimators.

- Developing optimal sensor placement strategies which maximize performance of the proposed state estimators.

- Reducing congestion of communication channels by developing more efficient transformation matrices.

1.3 CONTRIBUTIONS

The research objectives highlighted in the previous section have been addressed through works whose contributions are summarized below.

1. For the first time, tensor completion-based state estimators are developed that overcome the limitations of existing state-of-the-art estimators. These estimators provide superior voltage angle and magnitude estimation performance even under low data availability conditions. Feature scaling and pseudo measurement generation are introduced which provide greater sensitivity to all physical quantities in the tensor and improved temporal tracking performance. It is also observed that some unfoldings of the tensors are more informative than others. Therefore, attaching higher weights to more informative unfoldings resulted in superior state estimation performance. Further, graph neural network equivalents to model-based estimators are developed to reduce computational time. Interestingly, these estimators are able to learn the interdependencies among the measurements without any prior knowledge of the network parameters.

   More details related to contributions in tensor completion-based approaches developed in Chapter 3 and GNN-based approaches developed in Chapter 4 can be accessed through the following published articles.

- **R. Madbhavi**, B. Natarajan and B. Srinivasan, "Enhanced Tensor Completion Based Approaches for State Estimation in Distribution Systems," in IEEE Transactions on Industrial Informatics, vol. 17, no. 9, pp. 5938-5947, Sept. 2021, doi: 10.1109/TII.2020.3035449.

- **R. Madbhavi**, H. S. Karimi, B. Natarajan and B. Srinivasan, "Tensor Completion based State Estimation in Distribution Systems," in 2020 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT), Washington, DC, USA, 2020, pp. 1-5, doi: 10.1109/ISGT45199.2020.9087747.

- **R. Madbhavi**, B. Natarajan and B. Srinivasan, "Graph Neural Network-Based Distribution System State Estimators," in IEEE Transactions on Industrial Informatics, doi: 10.1109/TII.2023.3248082.

2. To aid the developed state estimators in model selection and topology detection, a graph neural network based topology identification approach is developed. This approach converts the topology identification problem into a link prediction problem at the switch locations. The developed technique provides comparable topology identification accuracies with significantly lower computational times.

   Additional information on the topology identification techniques presented in Chapter 5 can be accessed using the following paper:

   - **R. Madbhavi**, B. Natarajan and B. Srinivasan, "Distribution System Topology Identification using Graph Neural Networks," in IEEE PES GT&D Conference and Exposition, Istanbul, Turkey, May 2023.

3. Under low data availability conditions, the choice of metering locations plays a critical role in reducing state estimation errors. The power flow constraints used in the model-based state estimators can be exploited to get the interdependencies among the measurements. Therefore, these constraints are utilized to place sensors that minimize the voltage residual. The proposed approaches have better scalability and overcome the limitations of multiple evaluations as required by other techniques.

More details on meter placement strategies presented in Chapter 7 are available in the following published article:

- **R. Madbhavi**, B. Natarajan, and B. Srinivasan, "Meter placement approaches for matrix completion-based distribution system state estimator," in Electric Power Systems Research, vol. 213, p. 108687, 2022, doi: 10.1016/j.epsr.2022.108687.

4. The rising smart meter deployments results in a data deluge, that cannot be handled by existing communication infrastructure. Existing compressive sensing-based approaches utilize deterministic dictionaries. However, by tailoring the dictionaries to the data, superior reconstruction and state estimation performance can be obtained using compressive sensing-based frameworks. The developed data driven dictionaries outperform conventional deterministic dictionaries used in power systems such as Haar and Hankel.

   For further reading on the advantages of the data-driven dictionaries developed in Chapter 8, the following published articles can be referred:

   - **R. Madbhavi** and B. Srinivasan, "Data-driven dictionaries to enhance the performance of compressive sensing-based state estimators," in International Journal of Advances in Engineering Sciences and Applied Mathematics, vol. 14, no. 3, pp. 94–107, Dec. 2022, doi: 10.1007/s12572-022-00324-1.

   - **R. Madbhavi** and B. Srinivasan, "Enhancing Performance of Compressive Sensing-based State Estimators using Dictionary Learning," in 2022 IEEE International Conference on Power Systems Technology (POWERCON), Kuala Lumpur, Malaysia, 2022, pp. 1-6, doi: 10.1109/POWERCON53406.2022.9930028.

5. Non-revenue water is a major loss for water distribution systems. Therefore, leak localization strategies are developed using a machine learning framework. Classifiers such as support vector machines, random forests are utilized along with heuristic-based feature selection strategies. The developed techniques work with only pressure measurements at the locations selected as features by feature selection strategies. More details on the meter placement approaches for leak

localization in water distribution networks presented in Chapter 9 can be accessed through the following published paper.

- **R. Madbhavi**, A. Joshi, S. Munikoti, L. Das, P. K. Mohapatra and B. Srinivasan, "Sensor Placement for Leak Localization in Water Distribution Networks using Machine Learning," in 2020 IEEE International Conference on Computing, Power and Communication Technologies (GUCON), Greater Noida, India, 2020, pp. 95-100, doi: 10.1109/GUCON48875.2020.9231148.

1.4 ORGANIZATION OF THIS DISSERTATION

The remainder of this dissertation is organized as follows. Chapter 2 provides a background into techniques used in this study. In chapter 3, spatial, temporal, and physics-based relations are exploited to develop tensor completion-based distribution system state estimators. In chapter 4, graph neural network-equivalents of state estimators developed in chapter 3 are developed. Then, chapter 5 presents the topology identification technique that can aid state estimators developed in prior chapters for model selection and topology detection. In chapter 6, error bounds are developed for model-based state estimators. Then, in chapter 7, meter placement strategies are developed that minimize state estimation errors with model-based state estimators. The data congestion problems due to increased metering equipment is then addressed in chapter 8. Associated problems in other cyber-physical systems, such as leak detection in water distribution systems are explored in chapter 9. Finally, chapter 10 concludes this dissertation with conclusions and future directions.

CHAPTER 2

BACKGROUND

This chapter summarizes the literature on various topics discussed in this work such as state estimation, meter placement, matrix and tensor completion, compressive sensing, and graph neural networks.

2.1 DATA IMPUTATION

Data reconstruction from partially sampled signals is a critical requirement in several applications. In general, sampled data might contain missing entries and incorrect values. However, data analysis on such data requires data-cleansing to remove incorrect values and imputation techniques to recover the missing entries. This need has resulted in the development of several techniques which can be categorized as matrix and tensor completion approaches. Traditionally, matrix and tensor completion are performed through factor models. The factor models learnt using the known measurements are used to estimate the unknown measurements. However, these models require repeated computation of the latent factors. Therefore, other matrix and tensor completion approaches were developed that formulate the problem as optimization problems that enforce constraints on their structure (for example, rank minimization).

Consider a vector $x \in \mathbb{R}^n$. Now, suppose that only a few entries of x are known. Candes *et al.* (2006) proved that the recovery is possible under certain conditions on signal sparsity and even exact under certain additional conditions . Now, consider a table containing data arranged in the form of m rows and n columns, for example, an image from a camera. This table (matrix) might contain missing values or corrupted values that need to be corrected. This problem of imputing the missing entries of a matrix is called matrix completion. Furthermore, some applications have multidimensional data

with dimensionality greater than two, for example multi-channel images and videos. The algorithms imputing missing entries in such data structures are called tensor completion approaches. Some examples of applications of matrix and tensor completion are as follows:

- **Guessing user responses**: Consider the case of partially filled survey forms. The responses of all candidates can be combined to form a matrix such that each row contains the responses of one candidate. Thus, one might be interested in predicting all the candidates' missing answers.

- **Recommender systems**: In this class of problems, user preferences are assumed to be based on a small set of factors. The knowledge of such factors could be exploited to develop recommender systems. The Netflix problem is one of the popular example of such use cases. Here, given user responses for ratings of a few movies, the task is to suggest titles for the user which would suit his/her preferences.

- **Video and image processing**: Video and images can contain several artefacts. The quality of these images can be improved by first creating a mask identifying regions of the images containing such artefacts and then imputing these pixel values using matrix and tensor completion algorithms such as in Liu *et al.* (2013).

- **Background model initialization**: Motion recognition from a stationary camera video can be achieved by subtracting the background model from the video. A fixed background model cannot be utilized due to various factors in the observed volume of space, such as lighting or environmental changes. Therefore, the background model has to be dynamically created. In this regard, matrix and tensor completion have been utilized to generate these background models, see Sobral and hadi Zahzah (2017).

- **State estimation**: Due to growing complexities in modern cyber-physical systems (CPS), there has been a surge in sensor deployments to monitor these systems. State measurement matrices and tensors can be formed by using the measurements of these sensors. Communication infrastructure limitations limit the number of such sensors that can be deployed. Thus, these matrices and tensors contain several missing entries. However, the control of such complex systems requires accurate state estimates. Therefore, matrix and tensor completion techniques have been utilized to reconstruct these matrices and thus provide estimates for the unmeasured quantities, as developed by Madbhavi *et al.* (2021, 2020).

While the above examples present specific use cases of matrix and tensor completion, these algorithms can be applied across domains by converting the problem into a task of imputing missing matrix or tensor entries.

Consider a matrix $\mathbf{M} \in \mathbb{R}^{m \times n}$ in which only a fraction of the total entries are known. Let Ω be the set of all indices (i, j) of known elements of the matrix $\mathbf{M}$. Now the aim of low-rank matrix completion is to find a low-rank matrix $\mathbf{X} \in \mathbb{R}^{m \times n}$ that satisfies the observed entries in $\mathbf{M}$. This can be written mathematically as follows.

$$\begin{aligned} \underset{\mathbf{X}}{\text{argmin}} \quad & rank(\mathbf{X}) \\ \text{S.T.} \quad & \mathcal{P}_\Omega(\mathbf{X}) = \mathcal{P}_\Omega(\mathbf{M}) \end{aligned} \tag{2.1}$$

Here, $\mathcal{P}_\Omega(.)$ is the orthogonal projection of a given matrix onto the set of all matrices vanishing outside of Ω as described below.

$$\mathcal{P}_\Omega(\mathbf{X}) = [\mathcal{P}_\Omega(\mathbf{X})]_{ij} = \begin{cases} \mathbf{X}_{ij}, & (i, j) \in \Omega \\ 0, & \text{otherwise} \end{cases} \tag{2.2}$$

However, the $rank(.)$ function is non-convex, with no algorithms to solve it in practical time limits. It has been proven that the nuclear norm can be used as a convex relaxation of

the $rank(.)$ function, see Candès and Tao (2010). Therefore, low-rank matrix completion formulation can be re-formulated as shown below.

$$\begin{aligned} \underset{\mathbf{X}}{\text{argmin}} \quad & \|\mathbf{X}\|_* \\ \text{S.T.} \quad & \mathcal{P}_\Omega(\mathbf{X}) = \mathcal{P}_\Omega(\mathbf{M}) \end{aligned} \tag{2.3}$$

Here $\|.\|_*$ is the nuclear norm of a matrix and is equal to the sum of singular values of the matrix, such that $\sigma_i(\mathbf{X})$ is the i^{th} largest singular value of $\mathbf{X}$ and k is the rank of the matrix.

$$\|\mathbf{X}\|_* = \sum_{i}^{k} \sigma_i(\mathbf{X}) \tag{2.4}$$

This formulation (2.3) can further be relaxed as follows.

$$\underset{\mathbf{X}}{\text{argmin}} \quad \lambda\|\mathbf{X}\|_* + \frac{1}{2}\|\mathcal{P}_\Omega(\mathbf{X}) - \mathcal{P}_\Omega(\mathbf{M})\|_F^2 \tag{2.5}$$

Now consider the shrinkage operation $\mathcal{D}_\tau(.)$ on a matrix $\mathbf{A}$ with singular value decomposition $\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$. For a given soft-thresholding parameter τ, the shrinkage operation is defined as follows.

$$\begin{aligned} \mathcal{D}_\tau(\mathbf{A}) &= \mathbf{U}\mathcal{D}_\tau(\mathbf{\Sigma})\mathbf{V}^T \\ \mathcal{D}_\tau(\mathbf{\Sigma}) &= diag(\{max(\sigma_i(\mathbf{A}) - \tau, 0)\}) \end{aligned} \tag{2.6}$$

Cai *et al.* (2010) proved that $\mathcal{D}_\lambda(.)$ is the proximity operator of $\lambda\|\mathbf{X}\|_*$ and proposed an iterative algorithm based on singular value thresholding (SVT) to solve matrix completion formulation in (2.5). This algorithm is presented in equation (2.7). Where, $\mathbf{Y}^k \in \mathbb{R}^{m\times n}$ is an intermediate matrix with $\mathbf{Y}^0 = 0 \in \mathbb{R}^{m\times n}$ and $\delta_k > 0$.

$$\begin{aligned} \mathbf{X}^k &= \mathcal{D}_{\lambda\delta_{k-1}}(\mathbf{Y}^{k-1}) \\ \mathbf{Y}^k &= \mathbf{X}^k + \delta_k \mathcal{P}_\Omega(\mathbf{M} - \mathbf{X}^k) \end{aligned} \tag{2.7}$$

Another class of approaches for matrix completion involves learning the factorization of the matrices. In these approaches the matrix factorization obtained using the known entries are utilized to impute the missing entries. For example, Wang *et al.* (2014) develop the orthogonal rank-one matrix pursuit (OR1MP) and economic OR1MP (EOR1MP) by extending the idea of orthogonal matching pursuit of vectors to matrices. These algorithms aim to identify a set of rank-one matrices to form an overcomplete basis that satisfy the observed entries. The Grassmannian rank-one update sub-space estimation (GROUSE) developed by Balzano and Wright (2013) identifies subspaces in $\mathbb{R}^n$ using partially observed vectors. A significant drawback of this class of techniques is the repeated computation of factor matrices. Matrix factorization is a computational intensive step and could result in scalability issues with increase in matrix size.

Low-rank modeling using convex optimization has also been utilized in matrix and tensor completion approaches to impute missing measurements. Lin *et al.* (2010) develops an augmented Lagrange multipliers approach to solve a convex optimization problem that minimizes a combination of nuclear and l_1 norms. Video foreground and background separation algorithms: MAMR (motion-assisted matrix restoration) and RMAMR (robust MAMR) proposed by Ye *et al.* (2015) separate foreground and background from video clips by modeling the background as low-rank matrices and the foreground by sparse matrices.

Cai *et al.* (2010) developed a singular value thresholding algorithm (SVT) for the matrix completion problem. Thresholding the singular values to remove the lowest singular values allows for the removal of noise components from the data. Nie *et al.* (2012) developed a generalization of the norm-based approaches by converting the problem into a Schatten-p norm minimization problem.

Tensor completion allows for the imputation of partially filled higher dimensional arrays. Trivial approaches for tensor completion are to convert the tensors into matrices through a process called unfolding. An n-dimensional tensor can be unfolded in n-ways allowing n interpretations of the same data structure. The first tensor completion algorithms were extended from matrix completion algorithms by utilizing these unfoldings of the tensor. Liu *et al.* (2013) develops three low-rank tensor completion algorithms: simple low-rank tensor completion (SiLRTC), fast low-rank tensor completion (FaLRTC), and high accuracy low-rank tensor completion (HaLRTC). The SiLRTC formulation is as follows:

$$\min_{\mathcal{M}_i,\mathcal{X}} : \quad \sum_{i=1}^{n} \alpha_i \left\|\mathcal{M}_i\right\|_* + \frac{\beta_i}{2} \left\|\mathcal{M}_i - \mathcal{X}_{(i)}\right\|_F^2 \tag{2.8}$$

$$S.T. \quad \mathcal{X}_\Omega = \mathcal{T}_\Omega \tag{2.9}$$

The algorithm introduces additional matrices $\mathcal{M}_1$, $\mathcal{M}_2$,... $\mathcal{M}_n$ to convert the tensor completion problem into a combined matrix completion problem of the unfoldings of the tensor. As mentioned earlier in equations 2.5 and 2.6, an iterative solution exists for this formulation. HaLRTC improves upon SiLRTC by using replacing matrices $\mathcal{M}_1$, $\mathcal{M}_2$,... $\mathcal{M}_n$ with their tensor equivalents.

Other variants utilize CP factorization such as the one proposed by Zhao *et al.* (2015) and tucker factorization. CP factorization can be thought of as a generalization of SVD to tensors, wherein the tensor is decomposed into "rank-1" tensors created by outer product of columns of the n factor matrices.

2.2 COMPRESSIVE SENSING

Compressive sensing (CS) utilizes the fundamentals of sparsity and allows the recovery of finite dimensional signals from a set of linear, nonadaptive measurements. The theory

of compressive sensing as described in Donoho (2006) states that: *it is possible with high probability to reconstruct data from a small number of random projections of the original data, provided that the original data is sparse by itself or in a transformed domain*. Compressive sensing has been used in several applications such as data compression Joshi *et al.* (2019), medical imaging as in Lustig *et al.* (2008), and state estimation applications such as Alam *et al.* (2014); Karimi and Natarajan (2017); Dahale and Natarajan (2021).

A signal $x \in \mathbb{R}^n$ can be compressed by taking projections using a random matrix $\mathbf{\Phi} \in \mathbb{R}^{m \times n}$ to obtain the compressed signal $h \in \mathbb{R}^m$. The signal x is assumed to be sparse under the transformation $\mathbf{\Psi}$. The problem of recovering the compressed signal can then be viewed as the solution of the optimization problem,

$$\begin{aligned} z^* \quad = \quad & \text{argmin}_z \, \|z\|_1 \\ s.t. \quad & h = \mathbf{\Phi}\mathbf{\Psi} z \end{aligned}$$

The original signal can be recovered using $w^* = \mathbf{\Psi} z^*$. The choice of transformation basis has a significant impact on the reconstruction accuracy. Traditionally, deterministic transformation matrices such as Haar, Hankel, Hadamard, Toeplitz have been utilized for compressive sensing. The generation of these deterministic dictionaries is described in Appendix B.

The compressibility of the signals over space and time can be exploited by using 2-D compressive sensing. Here, the original load data $W \in \mathbb{R}^{N_{space} \times N_{time}}$ having sparse representations in transformation matrices $\mathbf{\Psi}_{\boldsymbol{space}}$ and $\mathbf{\Psi}_{\boldsymbol{time}}$, is used to generated compressed measurements $\boldsymbol{H} = \mathbf{\Phi}_{\boldsymbol{space}} \boldsymbol{W} {\mathbf{\Phi}_{\boldsymbol{time}}}^T$, The reconstruction of these compressed measurements can then be formulated as follows,

$$\begin{aligned} z^* \quad = \quad & \text{argmin}_z \, \|z\|_1 \\ s.t. \quad & vec(\boldsymbol{H}) = (\mathbf{\Phi}_{\boldsymbol{space}} \otimes \mathbf{\Phi}_{\boldsymbol{time}})(\mathbf{\Psi}_{\boldsymbol{space}} \otimes \mathbf{\Psi}_{\boldsymbol{time}}) z \end{aligned}$$

The original data can be recovered using $vec(\boldsymbol{W}^*) = (\boldsymbol{\Psi}_{\boldsymbol{space}} \otimes \boldsymbol{\Psi}_{\boldsymbol{time}})z^*$.

Compressive sensing (CS) exploits the sparsity in the data to generate sparse representations of signals in some transformation domains, see Candès and Romberg (2007). Compressive sensing has been used in medical imaging such as Graff and Sidky (2015) (x-ray tomography and MRI), hyperspectral imaging in Arce *et al.* (2014), data compression and many other applications. Current works, such as Joshi *et al.* (2019) utilize CS for data compression have studied the effectiveness of deterministic matrices such as Haar, Hadamard, Hankel and Toeplitz matrices as transformation bases for compressive sensing. This work also develops measures for sparsity and thus enables dynamic compression. A combined data compression and state estimation technique is proposed in Alam *et al.* (2014) to estimate the distribution system states from compressed measurements. However, these articles utilize deterministic dictionaries, which may not generate the sparsest solutions. In this regard, dictionary learning proposed by Mairal *et al.* (2009) can be utilized to obtain concise representations and improve reconstruction accuracy.

2.3 DICTIONARY LEARNING

Sparse coding involves decomposition of input vectors into sparse linear combinations of basis elements. This field has evolved from the need to exploit the inherent sparsity in real-world data in certain transformation domains for various applications such as signal processing and machine learning. Conventional transformation bases (example: Hadamard), utilize orthonormal basis vectors. However, this constraint of orthonormality between basis vectors may not produce the sparsest representation of the given data. Therefore, recent techniques such as dictionary learning overcome this limitation by relaxing this constraint and allowing more atoms than the dimension of the vector. Since these vectors do not form an independent vector set, the vectors are referred to as dictionary atoms.

Consider the data arranged in the form of a matrix $\boldsymbol{X} = [\boldsymbol{x_1}, \boldsymbol{x_2}, ..., \boldsymbol{x_N}]$, where $\boldsymbol{x_i} \in \mathbb{R}^n$ such that $\boldsymbol{X}$ represents data from N time windows of length n each. The objective of dictionary learning is to find a dictionary $\boldsymbol{D} = [\boldsymbol{d_1}, \boldsymbol{d_2}, ..., \boldsymbol{d_k}] \in \mathbb{R}^{n \times k}$, $k > n$ with atoms $\boldsymbol{d_i} \in \mathbb{R}^n$ such that $\boldsymbol{D}$ provides sparse representations for all $\boldsymbol{x_i}$.

The conventional dictionary learning problems such as Olshausen and Field (1997); Aharon *et al.* (2006) utilize the set of training signals $\boldsymbol{X}$ and optimizes cost functions of the form,

$$f_n(\boldsymbol{D}) = \frac{1}{n}\Sigma_{i=1}^{n} l(x_i, D) \tag{2.10}$$

This problem was formulated in Mairal *et al.* (2009) as shown below,

$$\begin{aligned} (\boldsymbol{D}, \boldsymbol{\alpha}) = & \underset{\boldsymbol{D} \in \boldsymbol{C}, \boldsymbol{\alpha} \in \mathbb{R}^{k \times N}}{\operatorname{argmin}} \frac{1}{N} \sum_{i=1}^{N} \left(\frac{1}{2} \|\boldsymbol{x_i} - \boldsymbol{D}\boldsymbol{\alpha_i}\|_2^2 + \lambda \|\boldsymbol{\alpha_i}\|_1 \right) \\ & s.t. \quad \boldsymbol{C} \in \{\boldsymbol{D} \in \mathbb{R}^{n \times k} : \|\boldsymbol{d_i}\|_2 \leq 1, \forall i = 1, 2, ..., k\}, \lambda > 0 \end{aligned} \tag{2.11}$$

The constraints ensure that the dictionary entries do not become arbitrarily large, resulting in small values for elements in $\boldsymbol{\alpha_i}$.

2.4 DISTRIBUTION SYSTEM STATE ESTIMATION

Distribution system state estimation (DSSE) is the process of estimating system states using the available grid measurements. Conventionally, weighted least squares (WLS) was utilized to obtain the state estimates. However, WLS requires pseudo-measurements to handle the low-observability conditions. In recent years, sparsity-aware state estimators have become popular due to their ability to exploit the sparsity in data to provide superior performance over conventional techniques. Furthermore, existing model-free approaches

require more measurements to provide accurate state estimates. This drawback can be overcome by introducing network constraints in the state estimation formulation. A brief review on current sparsity-aware state estimators is as follows:

Load data has been shown to be sparse under certain transformation such as Haar, Hankel, Hadamard, see Joshi *et al.* (2019); Alam *et al.* (2014). Moreover, due to the inherent correlation structure, the real and reactive power injections are sparse under transformations $\mathbf{\Psi_P}$ and $\mathbf{\Psi_Q}$, which together generate the block diagonal transformation matrix $\mathbf{\Psi}$. Therefore, the temporal and spatial compression capabilities were utilized to generate two voltage estimation methods, namely the indirect approach and the direct approach by Alam *et al.* (2014).

The indirect approach, first reconstructs the power injections and then utilizes Newton Raphson to obtain the state estimates. Here, $F(x)$ is a system of non-linear equations that maps the state vector x with the power injections vector y. The compressed loads are reconstructed to obtain the active and reactive power injections. A flat start is assumed for the states of the system. The algorithm then iteratively updates the state vector by minimizing the mismatch with the reconstructed voltages using the Jacobian matrix.

The indirect approach requires the reconstruction of all power injections individually before it can perform state estimation. This can be avoided by combining the reconstruction and state estimation procedures and this results in the direct approach. Unlike the indirect method, the direct method utilizes the mismatch of the compressed power injections to perform state estimation. This step eliminates the need for individual reconstructions of the power injections at all buses.

Liu *et al.* (2019*a*) proposed a low-rank matrix completion-based formulation for distribution system state estimation (DSSE) using linearized power flow equations as constraints. Consider a three-phase distribution system with m buses and n branches. The measurements from this system can be used to create state measurement matrices.

However, these matrices may be incomplete due to a lack of measurements. These missing measurements can be imputed using a low-rank matrix completion based approach. In this approach, the measurements from the distribution system are shaped into state measurement matrices $\boldsymbol{M} \in \mathbb{R}^{3m\times 5}$ such that the voltage phasor components, voltage magnitude and power injection phasor components ($Re\{\boldsymbol{V_i}\}$, $Im\{\boldsymbol{V_i}\}$, $|\boldsymbol{V_i}|$, P_i^{in} and Q_i^{in}) at the three phases of the i^{th} bus are stored in the $(3i-2)$, $(3i-1)$ and $(3i)^{th}$ rows of the matrix respectively. The availability of data in $\boldsymbol{M}$ is characterized as a binary matrix $\boldsymbol{\Lambda}$ which indicates the presence or absence (1 or 0) of corresponding measurements in $\boldsymbol{M}$. The missing measurements are imputed by formulating the state estimation problem as a nuclear norm minimization problem (2.12). This formulation aims to find a low-rank matrix $\boldsymbol{X}$ such that the imputed measurements are constrained to satisfy the linearized power flow equations while allowing for minor deviations to account for measurement noise. This approach is an adaptation of the matrix completion approach in Liu *et al.* (2019*a*) and incorporates the learnings on improving the state estimation accuracy of nuclear-norm minimization approaches in Madbhavi *et al.* (2021).

$$\begin{aligned} \min_{\boldsymbol{X}} \quad & \|\boldsymbol{X}.diag(\boldsymbol{s})\|_* + \lambda\|((\boldsymbol{X}-\boldsymbol{M}) \bigodot \boldsymbol{\Lambda}).diag(\boldsymbol{f})\|_F^2 \\ \text{s.t.} \quad & \|\boldsymbol{v} - \boldsymbol{D}\boldsymbol{x} - \boldsymbol{w}\|_\infty \leq \gamma \\ & \| \, |\boldsymbol{v}| - \boldsymbol{K}\boldsymbol{x} - |\boldsymbol{w}| \, \|_\infty \leq \delta \end{aligned} \tag{2.12}$$

The constraints in (2.12) are the linearized approximations of the bus voltage phasors and voltage magnitudes. A feature scaling vector $\boldsymbol{s} \in \mathbb{R}^5$ is used to improve the sensitivity of the nuclear norm to all the physical quantities in the low-rank matrix $\boldsymbol{X}$ as described in Madbhavi *et al.* (2021). $\boldsymbol{f} \in \mathbb{R}^5$ is a scaling vector used to weight the errors in the estimation of different physical quantities. $\bigodot$ is the Hadamard product or the element-wise product. $\|.\|_*$, $\|.\|_F$ and $\|.\|_\infty$ are the trace, Frobenius and infinity norms, respectively. $\boldsymbol{v} \in \mathbb{C}^{3(m-1)\times 1}$ is the column vector of bus voltage phasors, and $\boldsymbol{x} \in \mathbb{R}^{6(m-1)\times 1}$ is the

concatenated column vector of the active and reactive power injections of the $(m-1)$ PQ buses. A tunable weight λ is used to increase or decrease the penalty applied due to deviations of entries in $\boldsymbol{X}$ from measured quantities in $\boldsymbol{M}$. The weight λ can be tuned using a grid search approach. State measurement matrices with randomized sensor locations and a range of data availabilities (10%, 20%, 30%, ... 90%) can be used as input data. The mean absolute error of estimated voltage angles can be used as the performance metric to evaluate the state estimator performance over a range of values to find the best λ. The allowed deviation of the voltage phasors and magnitudes are defined as γ and δ, respectively. The coefficient matrices $\boldsymbol{D} \in \mathbb{C}^{3(m-1)\times 6(m-1)}$, $\boldsymbol{K} \in \mathbb{R}^{3(m-1)\times 6(m-1)}$ and the zero-load voltage vector $\boldsymbol{w} \in \mathbb{C}^{3(m-1)}$ are calculated using the following equations as derived in Bernstein and Dall'Anese (2017).

$$\begin{aligned} \boldsymbol{D} &= \left[\boldsymbol{Y_{LL}^{-1}}.diag(\overline{\boldsymbol{w}})^{-1}, -j\boldsymbol{Y_{LL}^{-1}}.diag(\overline{\boldsymbol{w}})^{-1}\right] \\ \boldsymbol{K} &= |diag(\boldsymbol{w})|\,.Re\{diag(\boldsymbol{w})^{-1}.\boldsymbol{D}\} \\ \boldsymbol{w} &= -\boldsymbol{Y_{LL}^{-1}}\boldsymbol{Y_{L0}}\boldsymbol{v_0} \end{aligned} \tag{2.13}$$

where, $\boldsymbol{Y_{LL}} \in \mathbb{C}^{3(m-1)\times 3(m-1)}$ and $\boldsymbol{Y_{L0}} \in \mathbb{C}^{3(m-1)\times 3}$ are sub-matrices of the three-phase admittance matrix $\boldsymbol{Y}$, while $\boldsymbol{v_0}$ is the three-phase voltage column vector of the slack bus.

$$\boldsymbol{Y} = \begin{bmatrix} \boldsymbol{Y_{00}} & \boldsymbol{Y_{0L}} \\ \boldsymbol{Y_{L0}} & \boldsymbol{Y_{LL}} \end{bmatrix} \in \mathbb{C}^{3m\times 3m} \tag{2.14}$$

Graph neural networks (GNNs) have also been used in various power systems applications such as power flow calculations; data generation; fault detection, isolation, and diagnosis; and time series predictions such as Liao *et al.* (2022). However, applications of GNN in distribution system state estimation are still largely unexplored, with very few works in literature, see Kundacina *et al.* (2022*b*); Zamzam and Sidiropoulos (2019); Wu *et al.* (2022*a*); Kundacina *et al.* (2022*a*). GNN-based state estimators are proposed in

Kundacina *et al.* (2022*b*,*a*), which model individual bus measurements as nodes and their interactions through edges. Zamzam and Sidiropoulos (2019) utilizes graph pruning techniques to reduce the number of neural network connections, thereby preventing over-fitting. Wu *et al.* (2022*a*) incorporate edge weights using network parameters and multi-sampling rate measurements.

Given the limitations of existing works, it is desired to utilize spatio-temporal and physics-based dependencies to develop new sparsity-aware state estimators. Additionally, the computation time limitations can be addressed by developing graph neural network equivalents.

2.5 DISTRIBUTION SYSTEM: METER PLACEMENT

Sensor placement in distribution systems is motivated by various applications such as outage detection in Lotfifard (2017), observability enhancement in Babu and Bhattacharyya (2016), and state estimation in Liu *et al.* (2014), to name a few. However, meter placement is often challenging due to the conflicting nature of sensor placement objectives. Several sensor placement approaches have been proposed to optimally place sensors to achieve one or more of the following objectives: namely, minimizing the number of sensors installed, minimizing the cost of sensors in Liu *et al.* (2014); Chauhan and Sodhi (2020), optimizing performance under a resource constraint in Raposo *et al.* (2017), minimizing state estimation errors in Ghasemi Damavandi *et al.* (2015), and improving observability.

State estimator performance can be enhanced by improving the observability of the distribution systems. The observability of a distribution system depends on the number, location, and types of sensors along with the network's topology. The system's observability can be improved by increasing the number of installed sensors or utilizing pseudo-measurements. However, resource constraints limit the number of available sensors. Thus, optimal sensor placement is essential to maximize state estimation

accuracy under the given constraints. Several approaches have been proposed to meet the observability requirements of distribution systems, refer Liu *et al.* (2014); Chauhan and Sodhi (2020); Raposo *et al.* (2017); Ding and Emesih (2007); Alvarez *et al.* (2015); Rezaeian Koochi and Hemmatpour (2020); Peng *et al.* (2006). A numerical observability analysis approach based on the Jacobian matrix is proposed in Ding and Emesih (2007) to identify unobservable branches. The observable islands are then merged by evaluating the candidate branch measurements and power injections. A μPMU (micro phasor measurement unit) and smart meter placement strategy is proposed in Teimourzadeh *et al.* (2019), which formulates the sensor placement problem as a mixed-integer linear programming problem. This approach aims to maintain observability at steady state and during contingencies. The distribution-level phasor measurement unit (D-PMU) placement has been formulated as a mixed-integer programming problem in Chauhan and Sodhi (2020) to minimize cost and maximize system observability. A network connectivity and unreachability index (URI) is proposed in Alvarez *et al.* (2015) to quantify the difficulty in observing the nodes of the system. Thus, sensors are placed on node locations with high URI values. A PMU placement approach is proposed in Rezaeian Koochi and Hemmatpour (2020) to achieve full system observability. This approach formulates the meter placement problem as a binary-integer program using the connectivity matrix to identify the minimum number of PMUs and locations while covering the maximum number of line outages.

Heuristic and metaheuristic approaches have also been used for sensor placement Peng *et al.* (2006); Vigliassi *et al.* (2019); De Oliveira Rocha *et al.* (2013). A heuristic optimization approach is proposed in Peng *et al.* (2006) that utilizes Tabu search to place PMUs to obtain full system observability. A hybrid meter placement approach (PMUs and SCADA metering system) is proposed in Vigliassi *et al.* (2019) that utilizes a multi-objective evolutionary algorithm based on subpopulation tables and triangular factorization of the Jacobian matrix to obtain reliable meter placement. A genetic algorithm is utilized in Liu *et al.* (2012) to obtain the optimum number, type, and

locations of sensors to ensure given state estimation accuracy. However, recent works on sparsity-based methods have shown that similar state estimation results can be achieved using sensor sets with sizes less than the bounds predicted by observability-based methods. Therefore, there is a need for meter placement approaches that provide small sensor sets capable of providing low state estimation errors with sparsity-based state estimators.

State estimation is key to obtaining situational awareness and taking necessary control actions in distribution systems. Several meter placement approaches have been proposed to improve state estimation accuracy, refer Raposo *et al.* (2017); Ghasemi Damavandi *et al.* (2015); Yao *et al.* (2019); Chen *et al.* (2021); Ferreira *et al.* (2020); Xygkis *et al.* (2018); Ali *et al.* (2016). A mixed-integer semidefinite programming-based approach was proposed in Yao *et al.* (2019) to place a limited number of sensors to improve state estimation accuracy. An optimum PMU placement approach was proposed in Chen *et al.* (2021) to extend existing metering systems with PMUs to improve state estimation accuracy. This approach utilizes the gain matrix to place PMUs to improve state estimation accuracies of estimators such as maximum likelihood and multiple-segment state estimators. The sensor placement problem for distribution systems with the number of sensors as a constraint has also been solved using binary particle swarm optimization Raposo *et al.* (2017). A mixed linear integer programming approach was proposed in Ferreira *et al.* (2020) to optimize sensor placement to reduce the standard error in state estimates at all buses under load uncertainty. Principles of Fisher information are used to place sensors in Xygkis *et al.* (2018) by minimizing errors in the estimated state vector. A submodular saturation algorithm was proposed in Ghasemi Damavandi *et al.* (2015) to optimize worst-case estimation accuracy among all system configurations. This method was shown to outperform greedy algorithms and genetic algorithms. Information-theoretic approaches are increasingly being used in sensor placement and have several advantages over heuristic and topological observability-based approaches. These approaches maximize the value of information given by the sensors to the state estimators for any given sensor set size. These meter placement approaches also

highlight the need to develop meter placement strategies specific to the state estimator to minimize state estimation errors.

Recently, sparsity-aware state estimators that utilize compressive sensing, matrix completion by Liu *et al.* (2019*a*), and tensor completion based approaches by Madbhavi *et al.* (2020, 2021) for estimating system states in low observability conditions have been developed. However, current works evaluate the state estimation performance by assuming that a fraction of the total measurements are known and do not delve into the meter placement problem. The choice of sensors significantly affects the state estimation performance, particularly under low observability conditions. Therefore, meter placement strategies are essential to optimize the performance of state estimators.

2.6 GRAPH NEURAL NETWORKS

Recently, graph neural networks (GNN) have gained popularity as they can handle rich relational information among elements. GNNs enable the modeling of dependencies among nodes in a graph and use them in various applications that can be converted to node, edge, or graph-level prediction/classification tasks. The vanilla GNN model has a few limitations, such as its inability to use different parameters in different layers and failure to account for edge features, heterogeneous nodes, and edge types. Several variants of GNNs have been developed to overcome these limitations. Graph convolution networks (GCN) were developed to extend the idea of CNNs to graphs. Multiple relations among nodes were handled through relational graph convolutional network. Other variants include graph attention networks, graph recurrent networks, graph residual networks, and changes to aid in training and optimizing GNNs. These developments enable the use of graph neural networks on problems that can be converted to node/edge/graph-level prediction/classification problems of GNNs.

2.6.1 Graph Convolutional Networks

Graph convolutional networks (GCNs) are a type of GNNs that generalize convolutions to the graph domain. The GCN model proposed by Kipf and Welling (2017) can be written as follows:

$$\boldsymbol{H^{(l+1)}} = \sigma\left(\tilde{\boldsymbol{D}}^{-0.5}\tilde{\boldsymbol{A}}\tilde{\boldsymbol{D}}^{-0.5}\boldsymbol{H^{(l)}}\boldsymbol{W^{(l)}}\right)$$

Here, $\boldsymbol{H^{(l)}}$ represents the feature matrix of the l^{th} layer of the model. Non-linearity is introduced through the activation function $\sigma(.)$. $\boldsymbol{W^{(l)}}$ is the weight matrix of the l^{th} layer. The structural information of the network is introduced through the degree matrix ($\boldsymbol{D}$) and the adjacency matrix ($\boldsymbol{A}$). A re-normalization trick was developed in Kipf and Welling (2017) to alleviate the problems of exploding or vanishing gradients in the GCN model. This trick involves adding a self-connection for each node in the graph and reconstructing the degree and adjacency matrices. The resulting matrices are as follows:

$$\tilde{\boldsymbol{A}} = \boldsymbol{A} + \boldsymbol{I}, \quad [\tilde{\boldsymbol{D}}_{ii}] = \Sigma_j \tilde{\boldsymbol{A}}_{\boldsymbol{ij}}$$

The GCN model accepts an input matrix $\boldsymbol{H^{(0)}} \in \mathbb{R}^{N \times F_{in}}$, where N is the number of nodes and F_{in} is the number of features in the input layer. Multiple GCN layers can be stacked together to obtain the desired output feature-length F_{out}. The output of the final layer can then be used to perform node-level prediction/classification.

2.7 LEAK DETECTION IN WATER DISTRIBUTION NETWORKS

A broad class of techniques for leak detection involve characterizing the normal operation of a network, followed by detection of an anomaly in the operation with measurements obtained from sensors installed throughout the network. In this class of methods, a set of methods make predictions of flow/pressure ahead of time and compare them against the measurements from the network to identify discrepancies caused by leaks, see Mounce *et al.* (2009). This is achieved with methods such as non-linear Kalman filter, Cumulative Sum (CUSUM), Hotelling T^2 by Jung and Lansey (2014), predictive Kalman filter by Karray *et al.* (2016), evolutionary polynomial regression by Laucelli

et al. (2016), weighted least squares with expectation maximization by Ye and Fenner (2013). These methods however rely heavily on the predictions and hence assume that the hydraulic model used for making the predictions represents the real system perfectly. However, several of the model parameters that determine the performance of these techniques cannot be determined accurately, thus limiting the reliability of the techniques. Approaches involving pattern recognition based on simulations of transient flow of water in the network have also been explored in the literature, see Wang *et al.* (2018). However, most of these methods employ computationally expensive simulations and require a large number of sensors to be installed, limiting their applicability in real time applications, see Romano *et al.* (2012). Several other approaches such as graph partitioning followed by mass balance in Rajeswaran *et al.* (2018) and linear programming methods in Berglund *et al.* (2017) have been used to detect and localize leaks in the network. However, these approaches are limited, either by the ability/extent of graph partitioning schemes or by the inherent assumption of a linear relationship between multiple leaks and the observed quantities.

The problem of leak localization is increasingly being viewed in the literature (from a machine learning perspective) as a classification problem, refer Chan *et al.* (2018). In this approach one makes use of a machine learning tool to predict the leaky scenario in a network to which a given set of measurements (are most likely to) correspond. An ensemble convolutional neural network-support vector machine model was used to extract features from data and obtain the probability of presence of a leak, followed by a graph theoretical approach to localize the leak in Kang *et al.* (2018). A pattern matching and ANN based approach for leak classification was proposed in Mounce *et al.* (2014). This method matches new profiles against a library of predefined patterns to find similarly shaped profiles and alert the user if a match with any anomaly in the library exceeds a predefined threshold. An unsupervised clustering based technique to localize leaks in which the data are clustered based on similarities of the vectors after detection and removal of outliers in Wu *et al.* (2016). Zhang et al. divided a network into regions

with k-means clustering and used a support vector machine (SVM) to localize leaks to one of the cluster regions such as in Zhang *et al.* (2016). Soldevila *et al.* (2016) used a k-nearest neighbor based classifier to localize leaks in a network . Although several techniques have been proposed in the literature, they typically begin by assuming a specific number of sensors and identify the location of sensors in the network that results in the best localization accuracy. As a result, the problem of leak localization becomes an optimal sensor placement problem with the aim of maximizing the performance of a classification algorithm. However, such an approach restricts itself by looking at a fixed number of sensors and does not allow the user to strike a trade-off between the number of sensors to be placed and the accuracy achieved.

Subsequent chapters aim to address the limitations of existing state estimators, meter placement techniques, and data compression and reconstruction by addressing the research problems outlined in Section 1.2.

CHAPTER 3

MODEL-BASED AND DATA-DRIVEN DSSE

3.1 INTRODUCTION

Grid modernization has resulted in the adoption of newer technologies such as renewable generation, electric vehicles, and metering equipment. The complex interactions among the nodes such as bi-directional energy transfer have introduced new dynamics in the power grids as highlighted by Chen *et al.* (2016). These dynamics have resulted in new voltage and security issues, see Powalko *et al.* (2012). Therefore, the operation of these networks requires better situational awareness.

Unlike transmission systems, low system observability has always been an issue in distribution systems. These issues can be resolved by increasing the number of sensors/ metering equipment in the network. However, existing communication networks cannot handle the significant load force on these networks by the new metering equipment, see Gupta and Kumar (2000). Furthermore, the costs associated with the installation of new metering equipment and additional communication infrastructure may be infeasible. Alternatively, state estimation techniques can be utilized to utilize measurements from the grid to obtain the system states. Thus, developing state estimation methods for distribution systems with limited measurements has become an active area of research, refer Primadianto and Lu (2017) and Dehghanpour *et al.* (2019). Recently, sparsity-aware state estimators, such as Liu *et al.* (2019*a*), have improved state estimation performance. However, they still have poor voltage angle estimation and tracking performance. Therefore, in this chapter we propose tensor completion-based approaches which utilize spatial and temporal correlations, and physics-based dependencies to improve state estimation performance even under very low data availability conditions.

3.2 PROPOSED APPROACHES

Tensors are n-dimensional generalizations of vectors and matrices. Tensors are used in several fields to represent data such as images, videos, medical imaging data and social network data. However, in real-world applications, the tensors contain missing entries due to the unavailability of data. The process of imputing the missing or unobserved elements of partial tensors is referred to as tensor completion. In this chapter, we propose tensor completion based approaches for state estimation in distribution systems.

The state measurement matrices proposed by Liu *et al.* (2019*a*) are extended to form state measurement tensors by stacking the state measurement matrices along the third dimension. Consider a p-phase distribution system consisting of m buses, n branches and measurements from N_T time instants. Then, the state measurement matrix M at any given time instant is a matrix of dimension $np^2 \times 12$. The columns in M are $Re\{V_i\}$, $Im\{V_i\}$, $|V_i|$, P_i^{in}, Q_i^{in}, $Re\{V_j\}$, $Im\{V_j\}$, $|V_j|$, P_j^{in}, Q_j^{in}, $Re\{I_{ij}\}$ and $Im\{I_{ij}\}$. Here, i and j are the buses connected by branch (i, j). V_i and V_j are the bus voltage phasors of buses i and j. P_i^{in} and Q_i^{in} are the active and reactive power injections at bus i, while, P_j^{in} and Q_j^{in} are active and reactive power injections at bus j. I_{ij} is the branch current in branch (i, j). This process of stacking state measurement matrices results in the generation of a state measurement tensor of dimension $np^2 \times 12 \times N_T$. The state measurement tensors created corresponding to a distribution system may be partial tensors with missing measurements. Since the control of these distribution systems requires complete situational awareness, it is required to estimate the missing measurements to obtain complete situational awareness of the system. Thus, there is a need for tensor completion approaches for state measurement tensors. The following sub-sections describe the proposed tensor completion based state estimation approaches.

Consider a partial state measurement tensor $\mathcal{T} \in \mathbb{R}^{np^2 \times 12 \times N_T}$ created by using the available measurements from N_T time instants. A binary-valued tensor called the data availability operator Λ is created such that the elements of Λ corresponding to available

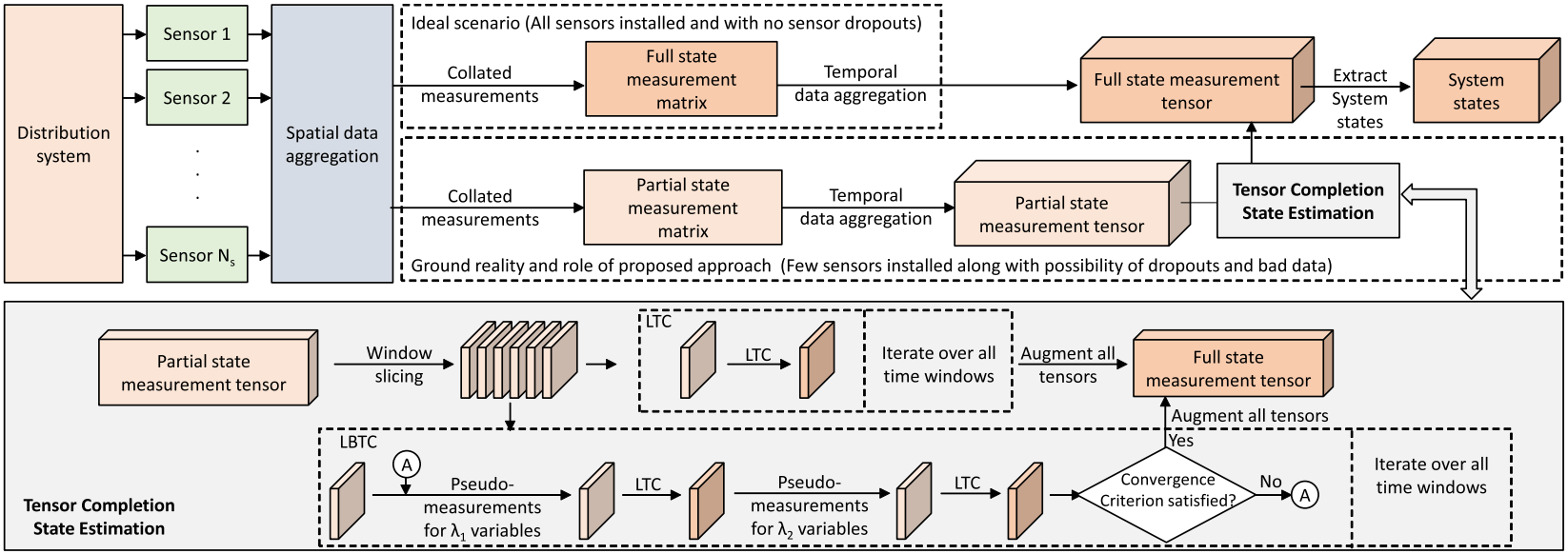


Figure 3.1: Role of proposed approaches in distribution system state estimation.

measurements in $\mathcal{T}$ are one and are zero otherwise. Fraction of Available Data (FAD) is the ratio of the available number of measurements to the total number of measurements and quantifies the availability of data in the tensor. Also, the temporal correlation of the system states is significant up to a few adjacent time frames. Hence, the tensor completion methods are applied on windows of the state measurement tensor. Considering time windows of length N_t (where $N_t \leq N_T$), tensors with dimension $np^2 \times 12 \times N_t$ are formed. Two enhanced tensor completion methods are proposed in the following subsections to estimate the missing measurements in the state measurement tensors.

3.2.1 Low-Rank Tensor Completion

In this classic approach, the state estimation problem in distribution systems is formulated as a tensor-completion problem in which the tensor trace norm of the state measurement tensor is minimized with linearized power flow equations as constraints. Low-rank tensor completion finds its roots in the traditional rank minimization of matrices. However, this optimization problem is non-convex due to the function $rank(X)$. The trace norm of the matrix ($||.||_*$) is the tightest convex envelope for the rank of matrices, see Goldfarb and Qin (2014). Liu *et al.* (2013) extends the trace norm of matrices to tensors by defining the tensor trace norm as the convex combination of trace norms of all the unfoldings of the tensor. However, the trace norms of the different unfoldings of the tensor $\mathcal{X}$ are not independent. This dependency is eliminated from the trace norm minimization by defining three independent tensors ($\mathcal{M}_1$, $\mathcal{M}_2$ and $\mathcal{M}_3$) corresponding to the three unfoldings of the tensor. The problem of minimizing the tensor trace norm can then be written as in equation (3.15). The first constraint ensures consistency of the observed measurements in the state measurement tensor with the corresponding elements in the imputed tensor. The second constraint allows the entries of $\mathcal{M}_1$, $\mathcal{M}_2$ and $\mathcal{M}_3$ to be independent of each other.

$$
\begin{aligned}
\min_{\mathcal{X},\mathcal{M}_1,\mathcal{M}_2,\mathcal{M}_3} : \quad & \sum_{i=1}^{3} \alpha_i \left\| (\mathcal{M}_i)_{(i)} \right\|_* \\
\text{S.T.} \quad & \mathcal{X}_\Omega = \mathcal{T}_\Omega \\
& \|\mathcal{X} - \mathcal{M}_i\|_F^2 \leq d_i, \quad i = 1, 2, 3
\end{aligned}
\tag{3.15}
$$

Due to the different orders of magnitude of the columns in the state measurement tensor, some columns influence the value of the tensor norm more than others. Hence, there is a need to perform feature scaling to make the tensor norm used in these algorithms sensitive to variations in all physical quantities in the state measurement tensor. The vector $s \in \mathbb{R}^{12}$, is a vector consisting of scaling factors for the 12 columns in the state measurement tensor. The scaling matrix ($\mathcal{S}$) for the first-mode unfolding of the state tensor is then generated by taking the Kronecker product ($\otimes$) of the identity matrix of dimension $N_t \times N_t$ with the diagonal matrix created from s. Here, N_t is the number of time frames of the tensor that are considered in the tensor completion process (current time window). The feature scaled version of the tensor $\mathcal{X}$ is given by $(fold_1(\mathcal{X}_{(1)}\mathcal{S}))$. The subscript "(1)" in $\mathcal{X}_{(1)}$ denotes that the tensor unfolding operation is applied on $\mathcal{X}$ across its first mode, while $fold_i()$ is the tensor unfolding operator along the i^{th} mode. Also, $\langle A, B \rangle = vec(A).vec(B)$. In other words, the <> operation on two tensors is defined to be the sum of element-wise products of the two tensors.

Lin *et al.* (2010) solve the optimization problem thus formulated using the Alternating Direction Method of Multipliers (ADMM). ADMM is chosen as it can efficiently solve optimization problems with non-smooth terms. The augmented Lagrangian used by Madbhavi *et al.* (2020) is modified to implement feature scaling, as shown in equation (3.16).

$$
\begin{aligned}
&\mathcal{L}_p(\mathcal{X}, \mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3, \mathcal{Y}_1, \mathcal{Y}_2, \mathcal{Y}_3) \\
&= \sum_{i=1}^{3} (\alpha_i \left\| (\mathcal{M}_i)_{(i)} \right\|_* + \left\langle (fold_1(\mathcal{X}_{(1)}S)) - \mathcal{M}_i, \mathcal{Y}_i \right\rangle \\
&\quad + \frac{\rho_i}{2} \|\mathcal{M}_i - (fold_1(\mathcal{X}_{(1)}S))\|_F^2)
\end{aligned} \tag{3.16}
$$

Using ADMM, the problem of solving (3.15) using (3.16) is done by converting them into sub-problems (3.17), (3.18) to obtain $\mathcal{M}_i$, $\mathcal{X}$ and the dual update (3.19). The linearized power flow equations, as described in Bernstein and Dall'Anese (2017) are added, as constraints in the optimization of $\mathcal{X}$. The full state measurement tensor is thus obtained by solving (3.17), (3.18) and (3.19) iteratively until convergence.

1. Sub-problem 1: Obtaining $\mathcal{M}_1$, $\mathcal{M}_2$ and $\mathcal{M}_3$.

$$
\begin{aligned}
&\{\mathcal{M}_1^{k+1}, \mathcal{M}_2^{k+1}, \mathcal{M}_3^{k+1}\} = \\
&\underset{\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3}{\text{argmin}} : \mathcal{L}_p(\mathcal{X}^k, \mathcal{M}_1^k, \mathcal{M}_2^k, \mathcal{M}_3^k, \mathcal{Y}_1^k, \mathcal{Y}_2^k, \mathcal{Y}_3^k)
\end{aligned} \tag{3.17}
$$

2. Sub-problem 2: Obtaining $\mathcal{X}$.

$$
\begin{aligned}
&\mathcal{X}^{k+1} = \\
&\underset{\mathcal{X}}{\text{argmin}} : \mathcal{L}_p(\mathcal{X}^k, \mathcal{M}_1^{k+1}, \mathcal{M}_2^{k+1}, \mathcal{M}_3^{k+1}, \mathcal{Y}_1^k, \mathcal{Y}_2^k, \mathcal{Y}_3^k) \\
&\qquad \text{S.T.} \quad \left| I_{ij,t} - (V_{i,t} - V_{j,t}) Y_{ij} \right| \leq \varepsilon_{ij,t} \\
&\qquad\qquad \|v_t - Dx_t - w\|_\infty \leq \gamma \\
&\qquad\qquad \| |v_t| - Kx_t - |w| \|_\infty \leq \alpha \\
&\qquad\qquad \forall (i,j) \in \Omega, \forall t \in \{1, 2, ..., N_t\}
\end{aligned} \tag{3.18}
$$

3. Sub-problem 3: Dual update.

$$
\mathcal{Y}_i^{k+1} = \mathcal{Y}_i^k - \rho_i \left(\mathcal{M}_i^{k+1} - \mathcal{X}^{k+1} \right) \tag{3.19}
$$

In (3.18), the first constraint is Ohm's law, while the second and third constraints are the linearized approximations of the voltage phasor and magnitudes. $I_{ij,t}$ is the instantaneous branch current at time t in the branch (i, j). For each branch (i, j) in Ω (set of all branches), the voltages of the two constituent buses i and j at time t are $V_{i,t}$ and $V_{j,t}$. Y_{ij}

is the admittance of the branch (i, j). The bus voltage phasor vector at time t is denoted by v_t. $D \in \mathbb{C}^{p(m-1)\times 2p(m-1)}$, $K \in \mathbb{R}^{p(m-1)\times 2p(m-1)}$ and $w \in \mathbb{C}^{p(m-1)}$ are obtained from the linearization of bus voltage phasors and magnitudes, see Bernstein and Dall'Anese (2017).

3.2.2 Low-Rank Block Tensor Completion

It was observed by Madbhavi *et al.* (2020) that higher availability of data significantly improves the voltage estimation. Thus, we improve the bus voltage estimation at low data availabilities by introducing pseudo-measurements and block tensor completion into a model-based tensor completion framework. In this unique approach, the dependent and independent quantities (columns) in the tensors are estimated alternatively while using pseudo-measurements for the unknown tensor quantities. The 12 columns in the state measurement tensors are split into two sets λ_1 and λ_2. The set λ_1 contains the columns of the tensor containing the components of power injections and the branch currents, while the set λ_2 contains columns corresponding to the voltage measurements. The method involves alternatively estimating physical quantities in one set of columns (λ_1 or λ_2) while considering that quantities in the other columns set as known. This alternating estimation of tensor columns helps in overcoming the problem of conflicting updates while simultaneously updating multiple variables to satisfy the constraints. Also, the pseudo-measurements increase the data availability in the state measurement tensors at every stage in the tensor completion process. Four new data availability operators are generated to simplify the update expressions in the algorithm. Λ_{10} and Λ_{11} are created by setting the elements in columns λ_1 of Λ to zero and one, respectively. Similarly, Λ_{20} and Λ_{21} are created using Λ and λ_2.

The tensor created by applying zero-order hold on the partial state measurement tensor ($\mathcal{T}$) is used as the initial estimate for the estimated state measurement tensor ($\widehat{\mathcal{T}}$). Two new tensors $\widehat{\mathcal{T}_1}$ and $\widehat{\mathcal{T}_2}$ hold the intermediate results of the two stages in the iterations. These tensors are initialized to be equal to $\widehat{\mathcal{T}}$.

Algorithm 1 Low-rank Block Tensor Completion (LBTC)

Input: $\mathcal{X}, \Lambda, \mathcal{S}$
Output: $\widehat{\mathcal{T}}$

$\lambda_1 \leftarrow \{4, 5, 9, 10, 11, 12\}$, $\lambda_2 \leftarrow \{1, 2, 3, 6, 7, 8\}$.
$\Lambda_{10} \leftarrow \Lambda$; $\Lambda_{11} \leftarrow \Lambda$; $\Lambda_{10}(:, \lambda_1, :) \leftarrow 0$, $\Lambda_{11}(:, \lambda_1, :) \leftarrow 1$
$\Lambda_{20} \leftarrow \Lambda$; $\Lambda_{21} \leftarrow \Lambda$; $\Lambda_{20}(:, \lambda_2, :) \leftarrow 0$, $\Lambda_{21}(:, \lambda_2, :) \leftarrow 1$
$\widehat{\mathcal{T}} \leftarrow zeroOrderHold(\mathcal{X}, \Lambda)$, $\widehat{\mathcal{T}_1} \leftarrow \widehat{\mathcal{T}}$, $\widehat{\mathcal{T}_2} \leftarrow \widehat{\mathcal{T}}$
repeat
    $\widehat{\mathcal{T}}_{prev} \leftarrow \widehat{\mathcal{T}}$
    $\mathcal{R}_1 \leftarrow \widehat{\mathcal{T}_1} \bigodot \Lambda_{10} + \widehat{\mathcal{T}_2} \bigodot (\Lambda_{11} - \Lambda_{10})$ *// Stage 1*
    $\widehat{\mathcal{T}_1} \leftarrow LTC(\mathcal{R}_1, \Lambda_{11}, \mathcal{S})$
    $\mathcal{R}_2 \leftarrow \widehat{\mathcal{T}_2} \bigodot \Lambda_{20} + \widehat{\mathcal{T}_1} \bigodot (\Lambda_{21} - \Lambda_{20})$ *// Stage 2*
    $\widehat{\mathcal{T}_2} \leftarrow LTC(\mathcal{R}_2, \Lambda_{21}, \mathcal{S})$, $\widehat{\mathcal{T}} \leftarrow \widehat{\mathcal{T}_2}$
until $\|\widehat{\mathcal{T}} - \widehat{\mathcal{T}}_{prev}\|_{max} < \Delta$ *// Convergence criterion*

The reduced tensor for stage 1 ($\mathcal{R}_1$) is generated by combining the entries corresponding to observed voltages in $\widehat{\mathcal{T}_1}$, with all the power injections and branch currents in $\widehat{\mathcal{T}_2}$. The new estimates for the voltage components are obtained by applying LTC on $\mathcal{R}_1$. $\widehat{\mathcal{T}_1}$ is updated with the imputed tensor obtained in this stage. The reduced tensor for stage 2 ($\mathcal{R}_2$) is generated by combining all voltages in $\widehat{\mathcal{T}_1}$, with entries corresponding to observed measurements of power injections and branch currents in $\widehat{\mathcal{T}_2}$. The new estimates for the power injections and currents are obtained by applying LTC on $\mathcal{R}_2$. $\widehat{\mathcal{T}_2}$ is updated with the imputed tensor obtained in this stage.

These stages are repeated until the imputed tensor $\widehat{\mathcal{T}}$ (equal to $\widehat{\mathcal{T}_2}$) converges. In this process, the bus voltages, power injections and branch currents are estimated at higher FADs than the LTC. Hence, LBTC is expected to provide better estimates of the voltage magnitude and angle than LTC. Algorithm 1 and Fig. 3.1 summarize the LBTC approach.

3.3 RESULTS

The proposed methods are validated on three test networks, namely, the IEEE 33 and 37 node distribution systems, and a 100-node test network, refer Appendix A. In the following subsections, the impacts of feature scaling and weighted tensor norm on voltage

estimation is studied. The methods are evaluated for their performance at different FADs. Finally, the results of studies on the robustness of proposed methods to sensor noise and bad data are presented [1].

3.3.1 IEEE 33-node Test Feeder

The proposed methods are evaluated on state measurement tensors of IEEE 33-node distribution system generated by using the simulation results from MATPOWER 7.0 developed by Zimmerman *et al.* (2011), an open-source power flow analysis tool. State measurement tensors were created as outlined in Madbhavi *et al.* (2020). State measurement tensors with 1% sensor noise and FADs ranging from 30% to 90% are used to study the performance of the proposed methods at different percentages of data availability. The robustness of the proposed methods to sensor noise is evaluated by applying the methods on state measurement tensors created with 5% of true measurement as the standard deviation. Finally, state measurement tensors with 2%, 4%, 6%, 8% and 10% bad measurements at 70% FAD are used to study the robustness of the proposed methods to bad data.

Effect of Feature Scaling

One of the drawbacks of existing matrix and tensor completion based state estimation is the poor voltage angle estimation accuracy as highlighted in Madbhavi *et al.* (2020). In these methods, the voltage angle is determined from the state measurement tensors by extracting the columns containing the real and imaginary components of the voltage phasors and thereby obtaining the phase angle. However, due to the different orders of magnitudes of the columns of the state measurement tensor, some columns influence

[1] Results in this section have been published in:

- **R. Madbhavi**, B. Natarajan, and B. Srinivasan "Enhanced Tensor Completion Based Approaches for State Estimation in Distribution Systems," IEEE Trans. Ind. Informatics, vol. 17, no. 9, pp. 5938–5947, Sep. 2021, doi: 10.1109/TII.2020.3035449.
- **R. Madbhavi**, H. S. Karimi, B. Natarajan and B. Srinivasan, "Tensor completion based state estimation in distribution systems," 2020 IEEE Power Energy Soc. Innov. Smart Grid Technol. Conf. ISGT 2020, Feb. 2020, doi: 10.1109/ISGT45199.2020.9087747.

Table 3.1: Performance of tensor completion methods in bus voltage magnitude estimation by using feature scaling and weighted tensor norm on the IEEE 33-node distribution system.

Parameter			Voltage Magnitude Estimation MAPE						
FAD			30%	40%	50%	60%	70%	80%	90%
Category	Method	Case							
Model Free	HaLRTC	Base	5.30	1.84	1.00	0.86	0.76	0.75	0.75
		FS	1.76	1.01	0.82	0.75	0.73	0.74	0.75
		WTN	4.00	1.46	0.95	0.83	0.76	0.76	0.76
		FS+WTN	1.38	0.86	0.76	0.74	0.74	0.75	0.75
Model Based	RMCSE	Base	0.44	0.35	0.30	0.26	0.22	0.18	0.14
		FS	0.28	0.22	0.19	0.16	0.13	0.11	0.11
	LTC	Base	0.38	0.30	0.26	0.22	0.17	0.14	0.10
		FS	0.22	0.16	0.12	0.10	0.09	0.09	0.10
		WTN	0.32	0.25	0.21	0.18	0.14	0.12	0.10
		FS+WTN	0.21	0.15	0.11	0.09	0.08	0.08	0.10
	LBTC	FS+WTN	0.13	0.12	0.11	0.11	0.11	0.11	0.11

the magnitude of the tensor norm more than others. Hence, there is a need to perform feature scaling to make the tensor norm used in these algorithms sensitive to variations in all physical quantities in the state measurement tensor.

From Tables 3.1 and 3.2, it is observed that model-based approaches performed significantly better than their model-free counterparts across all FADs. It is also observed that in all the methods, the scaled tensor implementations (FS) perform significantly better than the unscaled implementations (Base). It was observed that all bus voltage angle magnitudes are less than 1 degree. Therefore, the columns containing the real components of the voltage phasors in the state measurement tensors have a higher order of magnitude than the columns containing the imaginary components of the voltage phasors. Feature scaling provides a significant improvement in voltage angle estimation by making the matrix nuclear norm and tensor trace norms more sensitive to the columns containing the imaginary components of voltage phasor. A significant improvement in voltage angle estimation is observed from the results in Tables 3.1 and

Table 3.2: Performance of tensor completion methods in bus voltage angle estimation by using feature scaling and weighted tensor norm on the IEEE 33-node distribution system.

Parameter			Voltage Angle Estimation MAPE						
FAD			30%	40%	50%	60%	70%	80%	90%
Category	Method	Case							
Model Free	HaLRTC	Base	108.62	66.62	41.94	30.30	19.41	11.44	5.07
		FS	5.94	3.10	1.83	1.45	1.16	1.01	0.87
		WTN	82.54	34.27	15.45	11.85	7.93	5.26	2.68
		FS+WTN	6.16	2.24	1.37	1.14	0.97	0.87	0.81
Model Based	RMCSE	Base	15.39	13.28	11.42	9.70	7.97	6.42	5.15
		FS	4.02	2.53	1.87	1.51	1.39	1.38	1.42
	LTC	Base	15.45	12.67	10.72	9.76	8.81	8.57	8.52
		FS	3.08	2.14	1.72	1.55	1.46	1.52	1.62
		WTN	11.38	9.26	7.84	7.14	6.48	6.34	6.34
		FS+WTN	2.43	1.84	1.72	1.63	1.53	1.45	1.40
	LBTC	FS+WTN	2.10	1.59	1.24	1.11	1.07	1.14	1.26

Table 3.3: Voltage magnitude estimation performance of tensor completion methods on the IEEE-33 node system at different percentages of sensor noise.

Parameter			Voltage Magnitude Estimation MAPE						
FAD			30%	40%	50%	60%	70%	80%	90%
Category	Method	Noise							
Conventional	WLS	1%	-	-	-	2.44	1.57	0.87	0.50
Model Free	ZOH	1%	0.76	0.77	0.77	0.77	0.77	0.78	0.77
		5%	3.88	3.87	3.91	3.87	3.88	3.87	3.85
	HaLRTC	1%	1.38	0.86	0.76	0.74	0.74	0.75	0.75
		5%	7.86	4.76	4.02	3.70	3.67	3.66	3.72
Model Based	RMCSE	1%	0.28	0.22	0.19	0.16	0.13	0.11	0.11
		5%	0.70	0.56	0.47	0.40	0.33	0.25	0.19
	LTC	1%	0.21	0.15	0.11	0.09	0.08	0.08	0.10
		5%	0.49	0.37	0.30	0.23	0.18	0.15	0.15
	LBTC	1%	0.13	0.12	0.11	0.11	0.11	0.11	0.11
		5%	0.21	0.20	0.19	0.18	0.17	0.14	0.15

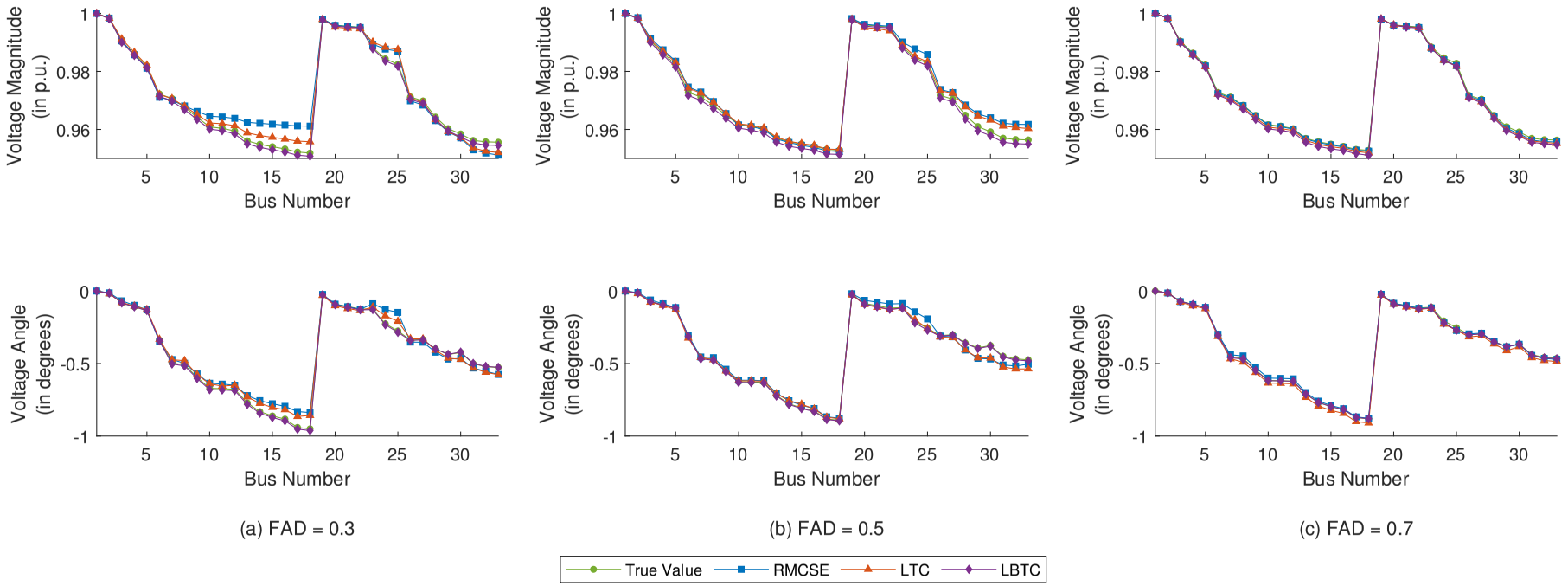


Figure 3.2: Bus voltage estimation on the IEEE-33 node distribution system using the proposed methods at different FADs.

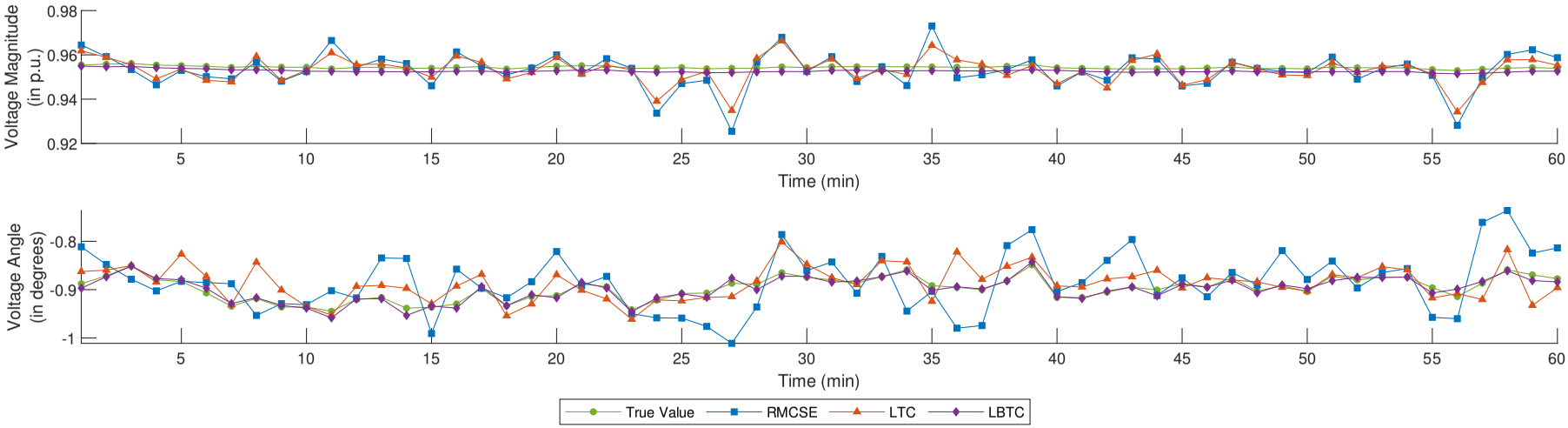


Figure 3.3: Bus voltage estimation across time for bus 18 of IEEE-33 node distribution system at 70% FAD using the proposed methods.

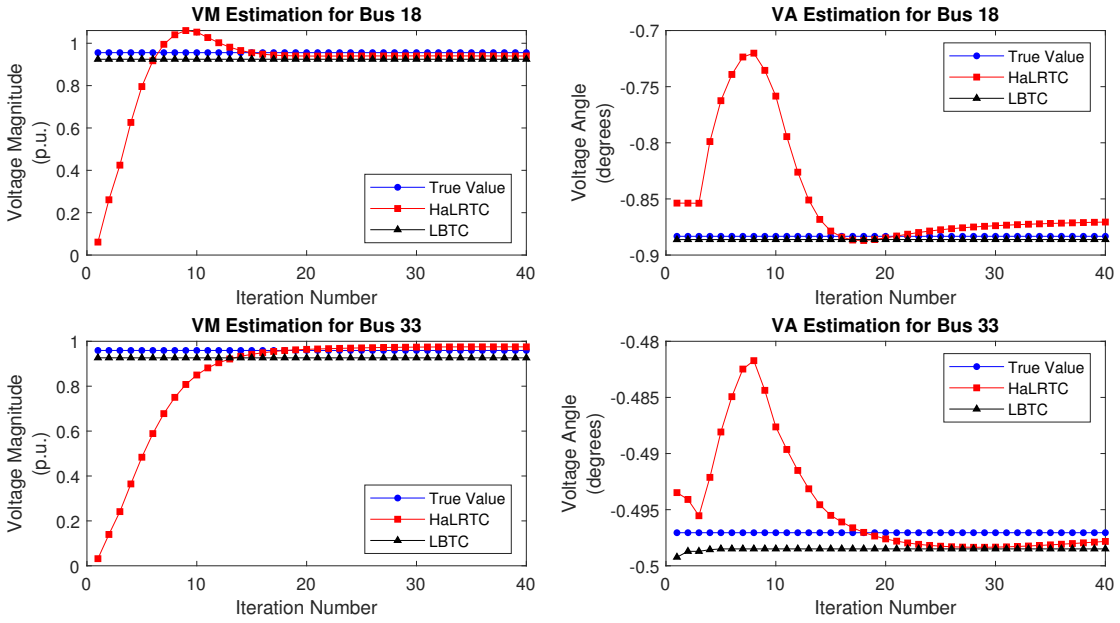


Figure 3.4: Convergence plots of LBTC and HaLRTC.

Table 3.4: Voltage angle estimation performance of tensor completion methods on the IEEE-33 node system at different percentages of sensor noise.

Parameter			Voltage Angle Estimation MAPE						
FAD			30%	40%	50%	60%	70%	80%	90%
Category	Method	Noise							
Conventional	WLS	1%	-	-	-	93.73	14.46	4.93	1.37
Model Free	ZOH	1%	2.44	2.03	1.82	1.52	1.30	1.10	0.94
		5%	4.76	4.45	4.29	4.20	4.14	4.02	3.93
	HaLRTC	1%	6.16	2.24	1.37	1.14	0.97	0.87	0.81
		5%	8.86	5.32	4.36	4.07	3.93	3.85	3.85
Model Based	RMCSE	1%	4.02	2.53	1.87	1.51	1.39	1.38	1.42
		5%	8.03	5.94	4.89	4.16	3.72	3.20	2.87
	LTC	1%	2.43	1.84	1.72	1.63	1.53	1.45	1.40
		5%	6.19	4.71	3.91	3.36	3.02	2.65	2.47
	LBTC	1%	2.10	1.59	1.24	1.11	1.07	1.14	1.26
		5%	3.49	3.18	2.91	2.78	2.72	2.53	2.44

Table 3.5: Voltage magnitude estimation performance of tensor completion methods on the IEEE-33 node system at different percentages of bad data.

Parameter		Voltage Magnitude Estimation MAPE					
Percentage of Bad Data		0%	2%	4%	6%	8%	10%
Category	Method						
Model Free	Zero-order hold	0.77	1.62	2.70	3.68	4.90	5.74
	HaLRTC	0.74	1.93	3.12	4.23	5.59	6.46
Model Based	RMCSE	0.13	0.31	0.48	0.60	0.72	0.76
	LTC	0.08	0.14	0.19	0.21	0.25	0.26
	LBTC	0.11	0.12	0.12	0.13	0.14	0.15

Table 3.6: Voltage estimation performance of tensor completion methods on the IEEE-33 node system at different percentages of bad data.

Parameter		Voltage Angle Estimation MAPE					
Percentage of Bad Data		0%	2%	4%	6%	8%	10%
Category	Method						
Model Free	Zero-order hold	1.30	539.1	942.6	1,017	1,471	2,141
	HaLRTC	0.97	414.1	706.8	851	1,113	1,506
Model Based	RMCSE	1.39	3.14	4.64	6.20	7.74	8.65
	LTC	1.53	2.68	3.86	5.03	6.44	7.30
	LBTC	1.07	2.68	3.90	4.91	6.16	6.90

3.2.

Impact of Selective Weighting of Modes

The tensor completion methods used in this study utilize tensor trace norm minimization for state estimation. The tensor trace norm is the convex combination of the trace norms of the unfoldings of the tensor. Due to the structure of the state measurement tensors, some unfoldings of the tensor provide more meaningful matrix structures than others. For example, the first mode unfolding results in the augmentation of frames of the state measurement tensor to form a single matrix. The second mode unfolding results in the augmentation of the transposed frames of the state measurement tensors. In contrast, the third mode unfolding generates a matrix in which the time series column vectors of each physical quantity in the state measurement tensor are augmented to form a single matrix.

The third mode unfolding generates a well-structured matrix. However, at low FADs, the probability of obtaining state measurement tensors in which some measurements are unavailable across the entire time window is higher. This data unavailability results in entire columns of time-series measurements of those quantities in the third mode unfolding to be zero. But, due to the nature of tensor unfolding, a missing column in the third mode unfolding does not result in entire rows or columns in the first and second mode unfoldings to be zero. Hence, a weighted tensor norm with higher weightage to the third mode unfolding due to its structure and lower weightage to the other unfoldings to account for missing entire measurements across time is introduced. Also, it can be observed that among the unfoldings across modes one and two, the first mode unfolding is more structured than the second mode unfolding. Each column of the first mode unfolding is a vector of the measurements of one of the 12 physical quantities at one time instant, whereas, no such significant structure arises from the second mode unfolding.

The importance of these unfoldings was confirmed by using the proposed methods with the tensor trace norm containing only one unfolding at a time. It was observed that

the state estimation benefited with the introduction of the first mode unfolding, while it deteriorated with the introduction of second mode unfolding. The weightage combination (10%, 0%, 90%) for the three unfoldings provided the best results. This combination provides a good trade-off between the performance gained by increasing the weightage of the first mode unfolding and the deterioration in performance caused by decreasing the weightage of the third mode unfolding. Tables 3.1 and 3.2 shows the results of different methods with and without weighted tensor norm (WTN and Base respectively). The methods with weighted tensor norm provide better angle estimation than their original counterparts.

Effect of utilizing both Feature Scaling and Weighted Tensor Norm

The proposed feature scaling and weighted tensor norms are aimed at solving two different issues, namely, improving the angle estimation by rescaling the columns and improving the estimation of all physical quantities in the tensor using the structures hidden in the different unfoldings of the tensors. Hence, both feature scaling and weighted tensor norm are applied together to evaluate the performance of the proposed methods.

The results of voltage estimation are shown in Tables 3.1 and 3.2 under the case name FS+WTN. It is observed that all methods benefit significantly from the combination of feature scaling and weighted tensor norms in the voltage angle estimation. The voltage magnitude estimation is slightly improved as compared to the tensor completion with feature scaling only.

The performance of the proposed methods are compared with the results of RMCSE at one time instant for different FADs is plotted in Fig. 3.2. It can be observed that at lower FADs, the performance of RMCSE and LTC deteriorate. However, it is observed that LBTC is affected the least even at very low FADs due to the improved data availability caused by using the data from prior time instants.

The proposed methods are also analysed for their performance in estimating the voltage

phasor across time. The voltage magnitude and angle estimation of bus 18 using the model-based methods is shown in Fig. 3.3. It is observed that LBTC performs significantly better than RMCSE and LTC in both voltage magnitude and angle estimation. Also, LBTC follows the true profile more precisely and accurately than the other methods.

Effect of Sensor Noise

The robustness of the proposed methods to sensor noise was studied by evaluating the methods at two different sensor noise levels, namely 1% and 5%. Zero-order hold, the simplest tensor completion approach with least computation cost is used as a minimum-benchmark that proposed methods must outperform. WLS is the conventionally used DSSE method. However, it is evident from Tables 3.3 and 3.4 that WLS is inferior to both the matrix and tensor completion based approaches utilized in this study. Therefore, we have omitted WLS and utilized the results of RMCSE as the benchmark for the other two systems.

The model-based methods handle noise better than the model-free methods as their formulations allow minor deviations from the available measured values. From Tables 3.3 and 3.4, it can be observed that the model-free approaches are more sensitive to higher magnitudes of noise in the data than the model-based approaches, resulting in higher voltage magnitude and angle estimation MAPEs. Therefore, the model-based implementations are more robust to sensor noise. LTC performs better than RMCSE as it utilizes the patterns in temporal evolution to fill missing values in one frame using the values from other frames. Also, the increased data availability due to the hybrid approach in LBTC enables it to perform better than the other model-based methods at all FADs.

Effect of Bad Data

The distribution system like other cyber-physical systems is often faced with the problem of bad data. This arises due to several reasons such as malfunctioning sensors, data corruption or malicious attacks. Hence, there is a need for state estimation methods to be

robust up to a reasonable percentage of bad data as per the design requirements. State measurement tensors were created having FAD of 70% and with 2%, 4%, 6%, 8% and 10% of measurements being bad measurements, respectively as outlined in Madbhavi *et al.* (2020). The performance of the proposed methods in the presence of bad data was studied by using the proposed methods on these state measurement tensors. The results of voltage magnitude and angle estimation performance of model-free and model-based methods in the presence of bad data are shown in Tables 3.5 and 3.6, respectively. It is observed that the model-based methods are more robust to bad data. The proposed methods (LTC and LBTC) perform significantly better than the model-free methods. They also show lower degradation in voltage estimation performance when compared with RMCSE.

Convergence of Proposed Methods

Under low data availability conditions, there arises several cases where entire row(s) or column(s) of the matrix are unknown. This is severely detrimental to the state estimation process and results in poor estimation accuracies in the model-free methods and classic matrix completion methods. This low data availability results in the problem of conflicting updates. At low data availability conditions, especially in the cases where all variables in a constraint equation are unknown, a lot of freedom arises for filling the missing measurements. This results in the oscillatory convergence of system variables. Fig. 3.4 shows this nature of the classic tensor completion approaches. In this figure, the convergence of the voltage magnitude and voltage angle of terminal buses 18 and 33 in a tensor with 30% FAD are shown. For the considered tensor, HaLRTC takes 40 iterations to converge to the final solution. However, LBTC converges in just 7 iterations (the algorithm exit criteria was removed to simulate the LBTC upto 40 iterations). Since LBTC estimates one set of measurements while using pseudo measurements for the complement set of measurements from the previous time frames, it results in an increase in data availability and more accurate estimation of system states. Also, the pseudo measurements generated in LBTC provide a better initial estimate than the flat start

considered in the classic approaches.

3.3.2 100 Node Test System

The performance of the proposed methods in larger networks and with real power usage data is studied by using tensors generated from a 100-node test network using publicly available energy usage data along the procedures outlined in Malekpour and Pahwa (2015). The IEEE 33-node distribution system is extended to form a 100-node test system by adding lateral feeders to the primary feeders. Load and PV generation data is extracted from actual customers as provided by eGauge website: egauge (2018*a*). The load data from these homes are aggregated and distributed to the 100 nodes such that the loads fulfil the nominal loads of the original IEEE 33-node test system. It was observed that the performance of the proposed methods on the 100-node test system was consistent with the results of the IEEE 33-node distribution system.

The performance of the proposed methods in voltage magnitude and angle estimation in the 100-node test system is presented in Tables 3.7 and 3.8, respectively. The proposed methods LTC and LBTC performed better than their model-free counterparts. The proposed methods also performed better than zero-order hold at FADs higher than 40%. Zero-order hold performs better than the model-based methods at low FADs (30% and 40%), primarily due to the lower level of noise present in the sensor data as compared to the assumed sensor noise in the IEEE 33-node system.

The proposed methods are used on the state measurement tensors of the 100-node test system to study their robustness to bad data. The results are presented in Tables 3.9 and 3.10. It is observed that model-based methods are more robust to bad data and show relatively lower deterioration in performance when compared with the model-free counterparts. Among the model-based methods, the proposed LBTC performed better then the other methods at all FADs.

Table 3.7: Voltage magnitude estimation performance of tensor completion methods on the 100-node system at different percentages of data availability.

Parameter		Voltage Magnitude Estimation MAPE						
FAD		30%	40%	50%	60%	70%	80%	90%
Category	Method							
Model Free	Zero-order hold	0.66	0.52	0.40	0.30	0.21	0.13	0.06
	HaLRTC	6.49	2.44	1.34	0.94	0.20	0.05	0.01
Model Based	RMCSE	0.79	0.64	0.49	0.33	0.20	0.11	0.02
	LTC	0.56	0.36	0.30	0.25	0.22	0.14	0.05
	LBTC	0.45	0.27	0.13	0.08	0.08	0.02	0.02

Table 3.8: Voltage angle estimation performance of tensor completion methods on the 100-node system at different percentages of data availability.

Parameter		Voltage Angle Estimation MAPE						
FAD		30%	40%	50%	60%	70%	80%	90%
Category	Method							
Model Free	Zero-order hold	3.33	2.57	2.01	1.49	1.05	0.65	0.32
	HaLRTC	28.79	8.24	3.10	1.87	0.76	0.68	0.25
Model Based	RMCSE	7.35	4.29	2.46	1.17	0.73	0.41	0.18
	LTC	7.64	2.79	1.60	1.04	0.56	0.18	0.05
	LBTC	6.20	2.81	1.26	1.07	0.53	0.18	0.07

Table 3.9: Voltage magnitude estimation performance of tensor completion methods on 100-node system at different percentages of bad data.

Parameter		Voltage Magnitude Estimation MAPE					
Percentage of Bad Data		0%	2%	4%	6%	8%	10%
Category	Method						
Model Free	Zero-order hold	0.21	0.99	1.98	3.00	4.25	5.35
	HaLRTC	0.20	1.83	3.29	4.64	6.10	7.18
Model Based	RMCSE	0.20	0.50	0.85	1.13	1.40	1.65
	LTC	0.22	0.45	0.71	0.90	1.07	1.26
	LBTC	0.08	0.12	0.12	0.13	0.14	0.15

Table 3.10: Voltage angle estimation performance of tensor completion methods on 100-node system at different percentages of bad data.

Parameter		Voltage Angle Estimation MAPE					
Percentage of Bad Data		0%	2%	4%	6%	8%	10%
Category	Method						
Model Free	Zero-order hold	1.05	2,525	3,645	5,012	7,825	8,673
	HaLRTC	0.76	1,783	2,704	3,943	4,973	5,835
Model Based	RMCSE	0.73	6.24	9.96	13.31	16.68	19.07
	LTC	0.56	6.23	9.32	11.67	13.70	15.44
	LBTC	0.53	2.68	3.90	4.91	6.16	6.90

Table 3.11: Voltage magnitude estimation performance of tensor completion methods on the IEEE-37 node system at different data availabilities.

Parameter		Voltage Magnitude Estimation MAPE						
FAD		30%	40%	50%	60%	70%	80%	90%
Category	Method							
Model Free	Zero-order hold	0.22	0.17	0.13	0.10	0.07	0.04	0.02
	HaLRTC	9.46	2.64	0.76	0.26	0.12	0.05	0.01
Model Based	RMCSE	0.14	0.12	0.09	0.09	0.08	0.06	0.06
	LTC	0.15	0.10	0.08	0.06	0.04	0.04	0.05
	LBTC	0.13	0.10	0.06	0.05	0.05	0.05	0.05

Table 3.12: Voltage angle estimation performance of tensor completion methods on the IEEE-37 node system at different data availabilities.

Parameter		Voltage Angle Estimation MAPE						
FAD		30%	40%	50%	60%	70%	80%	90%
Category	Method							
Model Free	Zero-order hold	3.21	2.49	1.92	1.43	1.01	0.65	0.31
	HaLRTC	3.50	1.05	0.46	0.21	0.10	0.04	0.01
Model Based	RMCSE	0.24	0.65	1.39	0.99	0.49	0.12	0.15
	LTC	0.28	0.25	0.21	0.16	0.13	0.11	0.11
	LBTC	0.16	0.14	0.13	0.12	0.12	0.11	0.11

Table 3.13: Voltage magnitude estimation performance of tensor completion methods on the IEEE-37 node system at different percentages of bad data.

Parameter		Voltage Magnitude Estimation MAPE					
Percentage of Bad Data		0%	2%	4%	6%	8%	10%
Category	Method						
Model Free	Zero-order hold	0.07	1.21	2.28	3.44	4.62	5.78
	HaLRTC	0.12	1.74	3.06	4.25	5.49	6.65
Model Based	RMCSE	0.08	0.75	0.93	1.02	1.44	1.97
	LTC	0.04	0.65	0.61	0.72	0.89	1.13
	LBTC	0.05	0.45	0.55	0.45	0.52	0.52

Table 3.14: Voltage angle estimation performance of tensor completion methods on the IEEE-37 node system at different percentages of bad data.

Parameter		Voltage Angle Estimation MAPE					
Percentage of Bad Data		0%	2%	4%	6%	8%	10%
Category	Method						
Model Free	Zero-order hold	1.01	20.08	40.76	60.69	84.60	105.38
	HaLRTC	0.10	14.95	30.22	45.50	61.46	78.14
Model Based	RMCSE	0.49	1.94	2.90	4.24	4.84	5.69
	LTC	0.13	1.69	2.47	3.20	3.80	3.75
	LBTC	0.12	1.57	2.53	2.42	2.66	2.96

3.3.3 IEEE 37-node Test Feeder

The performance of the proposed methods was also studied on tensors generated using the IEEE 37-node test network. This case study utilizes the load data from Malekpour and Pahwa (2015) and thus is equivalent to the bus voltage estimation of the primary nodes of the 559-node test system (an extension of the IEEE 37-node system) contained in the aforementioned work. It was observed that the performance of the proposed methods on the IEEE 37-node test system was consistent with the results of the IEEE 33-node distribution system.

The performance of the proposed methods in voltage magnitude and angle estimation in the IEEE 37-node test system is presented in Tables 3.11 and 3.12, respectively. The deterioration in state estimation performance in the presence of bad data was studied and the results are summarized in Tables 3.13 and 3.14. Similar to the observations made in the first two case studies, LTC and LBTC show the lowest deterioration in bus voltage estimation performance. The angle estimation accuracy of the proposed methods in the presence of bad data is far superior than the model-free methods.

3.3.4 Computational Time

The average computation time taken for completing tensors of the three systems is summarized in table 3.15. The implementations of these methods in MATLAB R2018b were run on a workstation with an Intel Core i7-9700K CPU and 64GB RAM. CVX developed by Grant and Boyd (2014) was used to solve the optimization problems. While the branch format of the state measurement tensor allows for a generic framework that incorporates both branch and bus measurements in the state estimation process, it results in larger tensors which increase the computation time. If only bus measurements are being considered, the state measurement tensor can be reduced to contain only 4 columns namely $Re\{V_i\}$, $Im\{V_i\}$, P_i^{in} and Q_i^{in}. There is a further reduction in the computation time since each measurement appears only once in the state measurement tensor, thereby eliminating the need for equality constraints that are required in the branch-format of

Table 3.15: Computation time of the model-based methods.

Test System	Parameters				Average computation time (s)		
	m	n	p	N_t	RMCSE	LTC	LBTC
IEEE 33	33	32	1	4	4.27	4.01	5.28
100-node	100	99	1	4	14.64	42.78	48.53
IEEE 37	37	36	3	4	140.44	183.24	186.66

state measurement tensors.

3.4 SUMMARY

In this chapter, new and enhanced tensor completion based approaches for state estimation in distribution systems have been proposed. The proposed methods were evaluated on the IEEE 33 and 37 node distribution systems and a 100-node test system. The need for feature scaling and its impact on state estimation methods in improving the voltage angle estimation by making the tensor trace norm sensitive to all tensor columns has been presented. The structural significance of different unfoldings of the tensor and the benefits of using weighted tensor norm in the optimization formulation has been studied. The state estimation performance has been further improved by increasing the data availability by dynamically generating pseudo-measurement and alternatively estimating the dependent and independent variables. Also, the proposed methods were observed to be more robust to sensor noise and bad data. The proposed inclusion of governing equations into the tensor completion process need not be restricted to state estimation in distribution systems and could potentially be extended to other networks with suitably designed state measurement tensors and constraints in the optimization. Such extensions would also benefit from the proposed feature scaling and weighted tensor norm implementation introduced in this work. However, the developed model-based approaches have high-computational complexities, and require long computational times. Therefore, in the next chapter, graph neural network equivalents are developed to overcome the computational time limitations of the model-based approaches.

CHAPTER 4

GNN-BASED DSSE

4.1 INTRODUCTION

Model-based approaches such as Liu *et al.* (2019*a*) and state estimators developed in the previous chapter Madbhavi *et al.* (2020, 2021) rely on solving optimization problems to obtain the system states. However, these approaches have large computational complexities rendering them unsuitable for online state estimation. Over the last decade, computationally heavy algorithms are being replaced by deep learning-based methods to reduce the computation time. Among such methods are graph neural networks (GNN).

GNNs have gained popularity as they can handle rich relational information among elements, see Liu and Zhou (2020). GNNs enable the modeling of dependencies among nodes in a graph and use them in various applications that can be converted to node, edge, or graph-level prediction/classification tasks. GNNs have been used in various applications, such as node classification, such as Kipf and Welling (2017), the resiliency of complex networks in Munikoti *et al.* (2022*b*), prediction uncertainty quantification in Munikoti *et al.* (2022*a*), and knowledge graphs in Schlichtkrull *et al.* (2018), to name a few.

GNNs have also been used in various power systems applications such as power flow calculations; data generation; fault detection, isolation, and diagnosis; and time series predictions such as Liao *et al.* (2022). However, applications of GNN in distribution system state estimation are still largely unexplored, with very few works in literature, see Kundacina *et al.* (2022*b*); Zamzam and Sidiropoulos (2019); Wu *et al.* (2022*a*); Kundacina *et al.* (2022*a*). GNN-based state estimators are proposed in Kundacina *et al.* (2022*b*,*a*), which model individual bus measurements as nodes and their interactions

through edges. Zamzam and Sidiropoulos (2019) utilizes graph pruning techniques to reduce the number of neural network connections, thereby preventing over-fitting. Wu *et al.* (2022*a*) incorporate edge weights using network parameters and multi-sampling rate measurements.

Current works require the knowledge of network parameters such as line impedances. However, detailed network models are typically unavailable, particularly in the secondary and tertiary branches of large distribution systems. Furthermore, unscaled features can result in large disparities in model weights, resulting in reduced sensitivity to some features over others and poor estimation performance. Finally, a cold start for unknown features in GNNs typically lowers their performance. These limitations motivate the need to develop new GNN-based state estimators.

The state estimation problem in distribution systems can be considered a node-level prediction problem of the GNN operating on the graph representation of the distribution system with state measurement matrices or tensors as input features. Therefore, GNNs are used in this chapter to learn the inherent relationships among the state measurement matrix and tensor elements.

4.2 BACKGROUND

The state measurement data structures and graph model of the distribution system are described in the following sub-sections.

4.2.1 Data Structure

Consider a three-phase distribution system with m buses, n branches, and measurements from N_T time instants. The measurements from this system at a time instant t can be

rearranged to form a state measurement matrix $\boldsymbol{M_t} \in \mathbb{R}^{3m\times 4}$ as described below,

$$\boldsymbol{M_t} = \begin{bmatrix} | & | & | & | \\ Re\{\boldsymbol{V_t}\} & Im\{\boldsymbol{V_t}\} & \boldsymbol{P_t^{in}} & \boldsymbol{Q_t^{in}} \\ | & | & | & | \end{bmatrix} - \begin{bmatrix} | & | & | & | \\ Re\{\boldsymbol{V_{ref}}\} & Im\{\boldsymbol{V_{ref}}\} & \boldsymbol{0}_{m\times 1} & \boldsymbol{0}_{m\times 1} \\ | & | & | & | \end{bmatrix} \tag{4.20}$$

$$\boldsymbol{V_{ref}} = \boldsymbol{1}_{m\times 1} \otimes \boldsymbol{v_{ref}} \tag{4.21}$$

$$\boldsymbol{v_{ref}} = \left[1, e^{-j(2\pi/3)}, e^{-j(4\pi/3)}\right]^T \tag{4.22}$$

Here, $\boldsymbol{V_t}$, $\boldsymbol{P_t^{in}}$, and $\boldsymbol{Q_t^{in}}$ are the complex bus voltage vector and the rectangular components of the complex power injection vector at time instant t. $Re\{\boldsymbol{X}\}$ and $Im\{\boldsymbol{X}\}$ are the real and imaginary components of a complex matrix $\boldsymbol{X}$, while $\otimes$ is the Kronecker product and $\boldsymbol{v_{ref}}$ is the reference bus voltage vector. The state measurement matrices $\boldsymbol{M_1}, \boldsymbol{M_2}, ..., \boldsymbol{M_{N_T}}$ can be stacked along the third dimension to form a state measurement tensor $\boldsymbol{\mathcal{T}} \in \mathbb{R}^{3m\times 4\times N_T}$. However, the state measurement tensor may be incomplete due to missing measurements. The available measurements in the N_T time instants can be characterized by binary availability matrices $\boldsymbol{\Lambda}_1, \boldsymbol{\Lambda}_2, ..., \boldsymbol{\Lambda}_{N_T}$, where each element in $\boldsymbol{\Lambda}_k$ represents the availability of the corresponding element in $\boldsymbol{M}_k$. The availability matrices can be stacked to form the availability tensor $\boldsymbol{\Lambda}$. Therefore, the state estimation problem of distribution systems is now a matrix completion problem of the incomplete state measurement matrices $\boldsymbol{M_1}, \boldsymbol{M_2}, ..., \boldsymbol{M_{N_T}}$. The problem can also be viewed as the tensor completion problem of the incomplete state measurement tensor $\boldsymbol{\mathcal{T}}$.

4.2.2 System Model

The distribution system can be represented using a graph $G_M = (V_M, E_M)$, such that V_M is the set of all buses and E_M is the set of all branches in the system. Fig. 4.1b illustrates the graph representation of a simple distribution system shown in Fig. 4.1a. For a 3-phase distribution system, each node can be replaced by three nodes representing the individual phases, and each edge (v_i, v_j) can be replaced by nine edges interconnecting the three nodes of v_i with the three nodes of v_j. Fig. 4.1c illustrates the conversion of the nodes and edges for 3-phase systems. The graph representations of the system

considering a window size of k-time instants can be formed by making k copies of the graph G_M and connecting the corresponding nodes of consecutive time-instant pairs $(t_j, t_{j+1}), \forall j \in 1, 2, ..., k-1$. The resulting graph $G_T = (V_T, E_T)$ is illustrated in Fig. 4.1d.

4.3 PROPOSED APPROACHES

Two GNN-based distribution system state estimators have been developed in this chapter. These estimators aim to provide comparable state estimation performance as their sparsity-aware counterparts RMCSE in Liu *et al.* (2019*a*), LRTC, and LBTC in Madbhavi *et al.* (2021) under low data availability conditions.

The first approach, Matrix GCN with feature scaling and pseudo-measurements or MGCN-FP, utilizes multiple GCN layers operating on graph representations of the distribution system with state measurement matrices as input features. Feature scaling is introduced to avoid large disparities between model weights and to ensure greater sensitivity to all features. Learnings from Madbhavi *et al.* (2021) are then utilized to introduce a pseudo-measurement generation phase to reduce state estimation errors. Finally, measurements across time are exploited by updating the graph structure of the network with additional copies of the network graph. This modification enables measurements across time instants to be utilized and results in the tensor GCN with feature scaling and pseudo-measurements approach (TGCN-FP). In practice, the TGCN-FP approach should be preferred in networks with mixed sampling rates since it allows utilizing measurements across time, thereby improving performance. Their formulations are described in the following sub-sections.

4.3.1 Matrix GCN with Feature Scaling and Pseudo-measurements (MGCN-FP)

The first state estimator extends graph convolution networks (GCN) to accommodate state measurement matrices as input features and additional modules that implement feature scaling and pseudo-measurement generation. The core of the MGCN-FP estimator

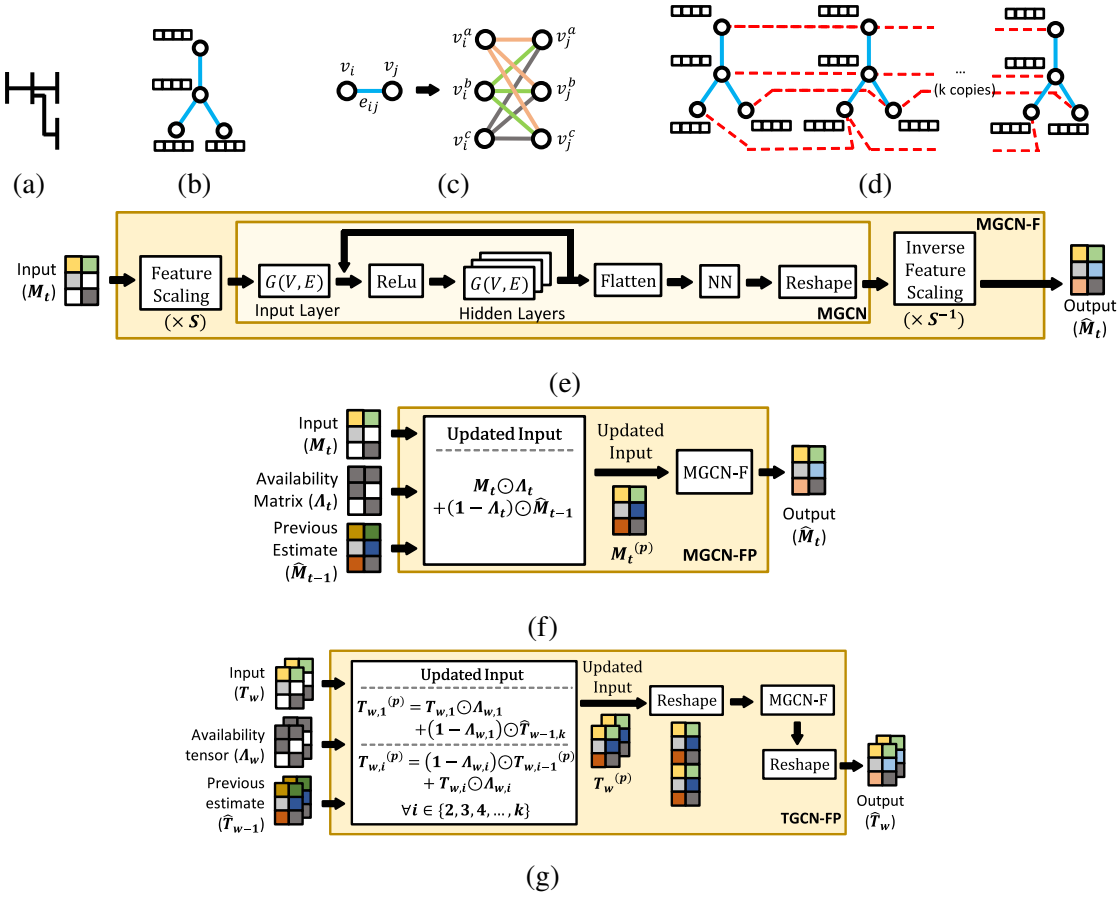


Figure 4.1: Proposed GNN-based approaches (a) Distribution system (DS) diagram (b) Graph representation of DS for all variants of MGCN (c) Conversion of nodes and edges for 3-phase DS (d) Graph representation of DS for all variants of TGCN (e) MGCN and MGCN-F (f) MGCN-FP (g) TGCN-FP.

comprises multiple GCN layers operating on G_M with the rectified linear unit (ReLu) as the activation function. The flattened outputs of the GCN layers are passed through a neural network to obtain the scaled imputed entries of the state measurement matrix, as illustrated in Fig. 4.1e. The feature scaling and pseudo-measurement generation modules are described in the following sub-sections.

Feature scaling

The initial feature vector for each node is obtained from its corresponding row in the scaled state measurement matrix $\boldsymbol{M_t S}$. Here, $\boldsymbol{S} = diag([s_1, s_2, s_3, s_4])$ and s_1, s_2, s_3, and s_4 are the scaling factors for the four columns of the state measurement matrix. Feature scaling prevents large disparities among weights corresponding to different physical quantities in the state measurement matrices and tensors. It also improves state estimation accuracy (similar to the benefits observed in Madbhavi *et al.* (2021)). Finally, the results from the GCN layers are rescaled using the inverse of the scaling performed in the input stage to obtain the imputed state measurement matrix. This module results in the matrix GCN with feature scaling (MGCN-F), as illustrated in Fig. 4.1e.

Pseudo-measurement generation

It was observed in Madbhavi *et al.* (2021) that state estimator performance could be enhanced by introducing pseudo-measurements from previous state estimates. Therefore, a pseudo-measurement generation module is introduced in MGCN-FP. Consider a time instant t, the corresponding state measurement is $\boldsymbol{M_t}$, and the availability matrix is $\boldsymbol{\Lambda_t}$. The estimated state measurement at the previous time instant is $\boldsymbol{\hat{M}_{t-1}}$. Similar to Madbhavi *et al.* (2021), pseudo-measurements are generated by replacing the unknown entries in $\boldsymbol{M_t}$ with their estimates from $\boldsymbol{\hat{M}_{t-1}}$. Therefore, the state measurement matrix at time t with pseudo-measurements can be created as follows,

$$\boldsymbol{M_t}^{(p)} = \boldsymbol{M_t} \bigodot \boldsymbol{\Lambda_t} + (\mathbf{1} - \boldsymbol{\Lambda_t}) \bigodot \boldsymbol{\hat{M}_{t-1}}$$

The updated state measurement matrix $\boldsymbol{M_t}^{(p)}$ is then used as input to MGCN-F. The approach is illustrated in Fig. 4.1f.

4.3.2 Tensor GCN with Feature Scaling and Pseudo-measurements (TGCN-FP)

Measurements across time instants are utilized in the second state estimator: tensor GCN with feature scaling and pseudo-measurements (TGCN-FP). In this approach, a window size of k-time instants is considered for state estimation. The corresponding graph representation of the network G_T is used instead of G_M. The input to this state estimator model is a tensor $\boldsymbol{T_w} \in \mathbb{R}^{3m \times 4 \times k}$ obtained by a moving non-intersecting window on the original state measurement tensor $\boldsymbol{\mathcal{T}}$. Pseudo-measurements for a state measurement tensor $\boldsymbol{T_w}$ are generated using the state estimate $\hat{\boldsymbol{T}}_{\boldsymbol{w-1}}$ of the previous time window. A cold start is assumed for the first time window, where unknown entries are assumed to be zero. The unknown measurements for the first temporal slice of $\boldsymbol{T_w}$ (i.e., $\boldsymbol{T_{w,1}}$) are initialized using the corresponding entries from their estimates in the last temporal slice of $\hat{\boldsymbol{T}}_{\boldsymbol{w-1}}$ (i.e., $\hat{\boldsymbol{T}}_{\boldsymbol{w-1,k}}$). Similarly, pseudo-measurements for unknown measurements in subsequent temporal slices of $\boldsymbol{T_w}$ (i.e., $\boldsymbol{T_{w,i}}$) are generated using measurements from their previous temporal slice with pseudo-measurements ($\boldsymbol{T_{w,i-1}}^{(p)}$). The pseudo-measurement generation steps can be written as follows,

$$\boldsymbol{T_{1,1}}^{(p)} = \boldsymbol{T_{w,1}} \bigodot \boldsymbol{\Lambda_{1,1}}$$

$$\boldsymbol{T_{w,1}}^{(p)} = \boldsymbol{T_{w,1}} \bigodot \boldsymbol{\Lambda_{w,1}} + (\mathbf{1} - \boldsymbol{\Lambda_{w,1}}) \bigodot \hat{\boldsymbol{T}}_{\boldsymbol{w-1,k}},$$
$$\forall w \geq 2$$

$$\boldsymbol{T_{w,i}}^{(p)} = \boldsymbol{T_{w,i}} \bigodot \boldsymbol{\Lambda_{w,i}} + (\mathbf{1} - \boldsymbol{\Lambda_{w,i}}) \bigodot \boldsymbol{T_{w,i-1}}^{(p)},$$
$$\forall i \in 2, 3, ..., k$$

Here, $\boldsymbol{T_{w,j}}$ refers to the j^{th} time frame in the w^{th} window of tensor $\boldsymbol{T}$. The state measurement matrices in the windowed-state measurement tensor are then stacked vertically to form the input matrix for the MGCN-F. The output from MGCN-F is reshaped into the original windowed tensor dimensions to obtain the estimated state

measurement tensor $\hat{\boldsymbol{T}}_w$.

4.4 RESULTS

The proposed approaches have been evaluated on the IEEE 33, 37-node distribution systems, and a 559-node distribution system, refer Appendix A. MGCN-FP and TGCN-FP have been compared with matrix completion-based methods: IALM, GROUSE, OR1MP, EOR1MP, MAMR, RMAMR, SVT, and SchattenPMC; and tensor completion-based methods SiLRTC, FaLRTC, HaLRTC, BCPFic, and tSVD. The proposed methods have also been compared with a neural tensor completion-based method CoSTCo, a physics-informed DNN approach PhyDNN, a GNN approach GNSE; and state estimators: RMCSE, LRTC, and LBTC. The algorithms compared in this study are summarized in Table. 4.1. Ablation studies were performed on the proposed methods to study the effectiveness of the proposed modules. Finally, an alternative approach to scale errors is studied. These approaches have been categorized into ten groups and their state estimation performance are tabulated in Tables 4.2, 4.3, 4.4, 4.5, 4.6, and 4.7 [1]

4.4.1 Data Generation

State measurement tensors ($\boldsymbol{\mathcal{T}}$) for the three distribution systems were created with 5% sensor noise, as described by Madbhavi *et al.* (2020). Availability tensors ($\boldsymbol{\Lambda}$) were created with various fractions of available data (FAD) of (30%, 50%, and 70%), (15%, 25%, and 35%), and (15%, 25%, and 35%) for the IEEE 33, 37-node distribution systems and the 559-node distribution system, respectively. Here, the fraction of available data (FAD) is the ratio of available measurements to the total number of measurements. Load data for generating the data sets are sourced from smart meter data of the eGuage website. Minute-wise load data for seven days (10080 samples) of 42 homes were collected.

[1] Results in this section have been published in:

- **R. Madbhavi**, B. Natarajan and B. Srinivasan, "Graph Neural Network-Based Distribution System State Estimators," in IEEE Transactions on Industrial Informatics, doi: 10.1109/TII.2023.3248082.

Combinations of load data are assigned to nodes of the systems to match their peaks with the nominal nodes. Load-flow analysis is performed to obtain bus voltages and generate state measurement matrices and tensors. Gaussian noise with a 5% standard deviation of the measurement magnitude is added to simulate noisy measurements. Half of the dataset was reserved as the test set. The remaining half was split in an 80:20 ratio as training and validation sets. All results are reported on the test set.

4.4.2 Model Generation

In this study, 3-layer GCN models were created and connected to a fully connected neural network. Training a GCN model involves forward passes through the layers and updating the weights in the backward propagation step. The models were trained using partial state measurement matrices as input features and full state measurement matrices as the expected output. Mean squared error was chosen as the loss function. The feature scaling phase incorporated into the three approaches (through MGCN-F) enables the use of mean-squared error as the loss function. Here, the pre-scaling step ensures that no feature dominates over the other. Hence, improving the sensitivity to all features. In their current form, the proposed models need to be retrained for each topology configuration.

4.4.3 Voltage Magnitude and Angle Estimation

It can be observed from Tables 4.2, 4.3, 4.4, 4.5, 4.6, and 4.7 that MC and TC algorithms have the highest voltage magnitude estimation errors. These approaches fail to provide good estimation under low data availability conditions, as discussed in Madbhavi *et al.* (2021). The CoSTCo approach allows non-linear relationships among tensor elements and has lower voltage magnitude estimation errors at all FADs than model-free methods. The MC-SE and TC-SE approaches utilize power flow equations as constraints in their low-rank matrix and tensor completion formulations and provide voltage magnitude estimation errors of less than 0.07%, 0.25%, and 0.20% for the IEEE 33, 37-node distribution systems and the 559-node distribution system at all FADs considered in this study. The TC-SE approaches provide the best state estimation performance.

The physics-informed DNN approach PhyGNN developed by Ostrometzky *et al.* (2019) has slightly inferior state estimation performance compared to the TC-SE approaches. Their comparable performance can be attributed to their usage of spatio-temporal measurements and power flow constraints. GNSE developed by Wu *et al.* (2022*b*) does not incorporate temporal measurements and therefore has poor performance compared to TC-SE, PhyDNN, and the proposed approaches.

Among the proposed approaches (group X), MGCN-FP has the highest state estimation errors. In contrast, TGCN-FP utilizes measurements across time to improve state estimation accuracy. Furthermore, TGCN-FP has comparable performance to the LRTC approach at low FADs. This superior performance is expected as it emulates the best features of the tensor completion-based approaches, such as utilizing measurements across time and feature scaling into the GNN framework. While PhyDNN performs better than the proposed approaches, it requires complete knowledge of network parameters which is generally unavailable in distribution systems.

4.4.4 Impact of Sub-modules on State Estimation Performance

Ablation studies were performed on the proposed approaches to study the effectiveness of the feature scaling and pseudo-measurement generation modules in improving the state estimation performance of the proposed methods. The results are tabulated in Tables 4.2, 4.3, 4.4, 4.5, 4.6, and 4.7 as groups VIII.a, VIII.b, and VIII.c. It can be observed from groups VIII.a, VIII.b, VIII.c, and X that removing either or both modules results in a deterioration in state estimation performance. Interestingly, it can be observed from groups VIII.a and VIII.b that feature scaling on its own may not significantly improve state estimation performance but performs better when paired with pseudo-measurement generation. This behavior is expected as feature scaling has the effect of scaling errors and reducing large disparities among weights. Feature scaling is more effective with the increased availability of measurements through pseudo-measurement generation. Finally, an alternate approach to scale errors is implemented through weighted mean squared

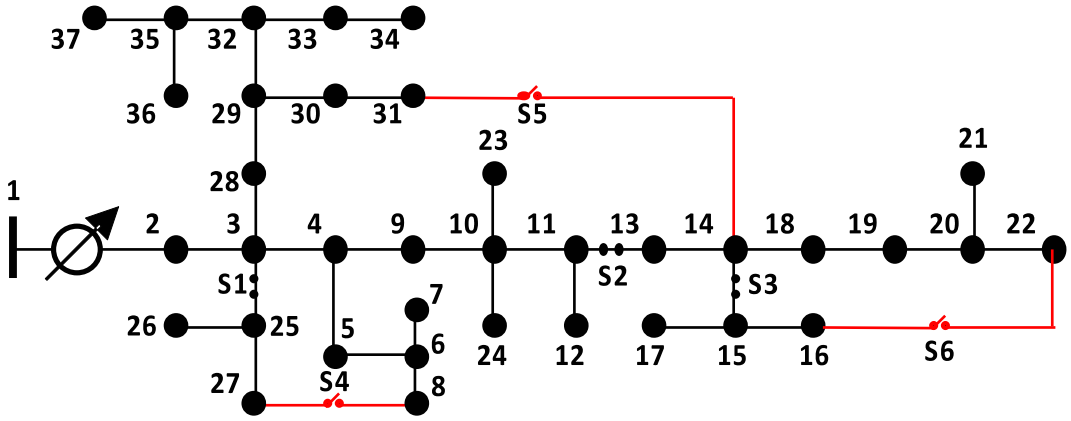


Figure 4.2: The primary distribution network in the 559-node system.

error (WMSE) instead of MSE. It can be observed from results from groups VIII.b and IX that while feature scaling scales errors, it also results in better performance than WMSE since it also avoids large disparities among the model weights.

4.4.5 Topology Identification

The primary distribution network in the 559-node distribution system is shown in Fig. 4.2. Three additional branches are added to generate different network topologies (marked in red). The switches $S1$, $S2$, ..., and $S6$ generate eight radial configurations of the network. The corresponding topology and switch configurations are provided in Table 4.8. The proposed approach MGCN-FP is used with the estimator pool technique to incorporate topology identification. In the estimator pool technique, a fraction of the measurements from the test system is provided to models built for each topology configuration. The topology corresponding to the model providing the best estimates of the remaining fraction of measurements is selected as the network's topology. This technique has been utilized in literature to incorporate state estimators in the topology identification process. MGCN-FP provided switch status classification accuracies of 92.1%, slightly higher than RMCSE (91.7%) and BCSE (90.5%). However, due to the computational time advantages of GNNs, MGCN-FP had an average computation time of 3.18s compared to 181.23s for RMCSE. Future work will incorporate a topology processor.

4.4.6 Computation Time

The implementations of these methods in MATLAB R2018b and python 3.8 were run on a laptop with an Intel Core i7-10750H CPU, NVIDIA Quadro P620 GPU, and 48GB RAM.

As expected and shown in prior works Madbhavi *et al.* (2021), tensor completion approaches require higher computational times than their matrix completion counterparts. Among the best state estimator categories, it can be observed that the DNN and GNN approaches require significantly lower computational times. It can be observed from Tables 4.2, 4.3, 4.4, 4.5, 4.6, and 4.7 that the TGCN-FP (best GNN method) provides a reduction in the computational time by a factor in the range of 100 to 1000 as compared to the LBTC approach (best model-based approach). The proposed approaches have also been compared with CoSTCo and GNSE, which are neural network tensor completion and GNN-based state estimator approaches. CoSTCo requires the model to be generated for each input tensor. Therefore, the training time for the CNN is an added computational time for this approach. However, the reusability of GNN-based approaches allows the model to be created once and utilized for new inputs. GNSE requires computational times of similar order of magnitude as the proposed approaches. It can also be observed that feature scaling and pseudo-measurement generation phases do not affect the computational time significantly.

4.4.7 Conformity of State Estimates with Power Flow Constraints

Consider the power flow model $v = F(\Theta, P, Q)$ of a p-phase distribution system with parameters Θ. This model utilizes the power flow constraints to compute the bus voltage vector v using the active and reactive power injections, P and Q, respectively. In model-free approaches, several solutions for the imputed matrix/tensor may exist which satisfy the known measurements. However, these solutions may violate power flow constraints. Model-based approaches incorporate the power constraints in their formulation and provide significantly lower voltage magnitude and angle estimation errors compared to

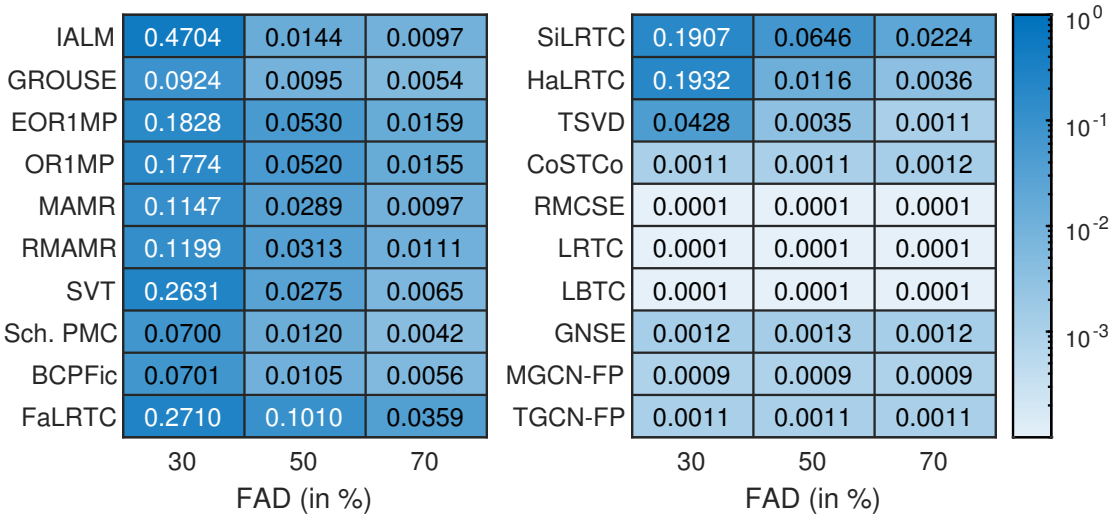


Figure 4.3: Conformance of state estimation results (AVPE) of IEEE 33-node distribution system with power flow equations.

model-free approaches Madbhavi *et al.* (2021). Therefore, conformance with power flow equations is the desired quality of distribution system state estimators. To verify the conformance of the state estimates with the power flow constraints, we define an average voltage phasor error (AVPE) metric as follows,

$$AVPE = \frac{1}{mpN_t} \sum_{t=1}^{N_t} \sum_{b=1}^{mp} |\boldsymbol{e}_{t,b}|$$

$$\boldsymbol{e}_t = \hat{\boldsymbol{V}}_t - F(\Theta, \hat{\boldsymbol{P}}_t^{in}, \hat{\boldsymbol{Q}}_t^{in})$$

Here, $\hat{\boldsymbol{V}}_t$, $\hat{\boldsymbol{P}}_t^{in}$, and $\hat{\boldsymbol{Q}}_t^{in}$ are the estimated bus voltage phasor vector, active power injection vector, and reactive power injection vector at time t. $\boldsymbol{e}_t$ is the vector of deviations of bus voltage phasors estimated using the state estimator and the bus voltage phasors estimated using the power flow model $F(.)$. $\hat{\boldsymbol{e}}_{t,b}$ is the corresponding error of bus b. Finally, AVPE is calculated using the estimates over the entire test dataset.

An AVPE of 0 implies that the state estimates are exactly as predicted by the power flow model. Therefore, the algorithms must generate state estimates with AVPE close to zero. Fig. 4.3 shows the AVPE values for all the methods analyzed in the previous sub-sections. The state estimation results of the three networks show that MC and TC approaches have

inferior performance at low FADs. Since low-observability conditions are considered in the study, several solutions that satisfy the known measurements may exist for model-free approaches. These solutions may not necessarily conform to the power flow equations. In contrast, model-based approaches force conformance with power flow equations through their constraints. Therefore MC and TC approaches have higher AVPE than model-based approaches. GNN-based approaches have very low AVPEs, indicating that these models can learn the inter-dependencies among the state measurement tensor elements and conform to the power flow equations. These results also demonstrate the power of GNNs, as the proposed approaches learned the interdependencies without prior knowledge of network parameters. PhyGNN is excluded from these results as it does not compute power injections.

4.5 SUMMARY

Recent works on distribution system state estimation have shown significant gains in state estimation performance using matrix and tensor completion-based approaches. However, these methods are computationally demanding and do not consider non-linear interactions among measurements. These limitations can be overcome by developing deep learning-based alternatives. For example, neural network models that utilize power flow constraints have been developed for distribution system state estimation. However, these approaches require complete knowledge of the network parameters, which is typically unavailable. Thus, new models that deliver performance on par with model-based approaches without relying on network constraints or necessitating complete knowledge of the network parameters are desired. Therefore, GNN-based approaches are proposed in this work. These approaches incorporate node and edge connectivity information by operating on the graph representation of the distribution systems. The proposed approaches utilize state measurement matrices and tensors as input features to graph neural network layers with additional feature scaling and pseudo-measurement generation phases to improve state estimation performance. The proposed approaches are evaluated on the IEEE

33, 37-node distribution systems and an unbalanced three-phase 559-node distribution system to show their applicability across single/three-phase systems and different sizes of distribution systems. These approaches provide comparable results to sparsity-aware state estimators at data availabilities as low as 15% on the 559-node distribution system without requiring knowledge of the network parameters. The computational times of the proposed approaches are several orders of magnitude lower than the model-based approaches and are comparable to existing deep learning approaches. In the next chapter, we develop topology identification techniques to aid in model selection and topology detection in the previously developed state estimators.

Table 4.1: Table of algorithms used in this chapter.

Category	Group	Algorithm	Reference
MC	I	IALM	Lin *et al.* (2010)
		GROUSE	Balzano and Wright (2013)
		EORIMP ORIMP	Wang *et al.* (2014)
		MAMR RMAMR	Ye *et al.* (2015)
		SVT	Cai *et al.* (2010)
		SchattenPMC	Nie *et al.* (2012)
TC	II	BCPFic	Zhao *et al.* (2015)
		FaLRTC SiLRTC HaLRTC	Liu *et al.* (2013)
		tSVD	Zhang *et al.* (2014)
NN-TC	III	CoSTCo	Liu *et al.* (2019*b*)
MC-SE	IV	RMCSE	Liu *et al.* (2019*a*)
TC-SE	V	LRTC LBTC	Madbhavi *et al.* (2021)
DNN-SE	VI	PhyDNN	Ostrometzky *et al.* (2019)
GNN-SE	VII	GNSE	Wu *et al.* (2022*b*)
Proposed GNN-SE (Ablation Study)	VIII.a	MGCN TGCN	Madbhavi *et al.* (2023)
	VIII.b	MGCN-F TGCN-F	
	VIII.c	MGCN-P TGCN-P	
Proposed GNN-SE (Using WMSE)	IX	MGCN-WP TGCN-WP	
Proposed GNN-SE	X	MGCN-FP TGCN-FP	

Table 4.2: State estimation results of IEEE 33-node distribution system at 5% sensor noise and different fractions of data availabilities.

Evaluation Metric							VM MAPE (in %)			VA MAE (in degrees)			Avg. time per tensor frame (in ms)
FAD							30	50	70	30	50	70	
Category	Group	Algorithm	Modules		Loss Fn./ Constraints								
			F	P	W	M							
MC	I	IALM	–	–	–	–	45.45	1.45	0.98	7.986	0.310	0.031	5.88
		GROUSE	–	–	–	–	6.22	0.77	0.55	5.905	0.226	0.011	10.12
		EORIMP	–	–	–	–	18.06	5.29	1.58	0.292	0.023	0.014	0.07
		ORIMP	–	–	–	–	17.52	5.18	1.54	0.228	0.027	0.014	0.05
		MAMR	–	–	–	–	11.48	2.92	0.99	0.033	0.012	0.010	0.32
		RMAMR	–	–	–	–	11.95	3.12	1.09	0.033	0.012	0.010	0.31
		SVT	–	–	–	–	25.63	2.69	0.63	0.036	0.014	0.012	4.37
		SchattenPMC	–	–	–	–	6.70	1.06	0.30	0.034	0.015	0.015	24.13
TC	II	BCPFic	–	–	–	–	6.71	1.04	0.56	0.363	0.059	0.011	89.34
		FaLRTC	–	–	–	–	26.24	9.69	3.38	0.038	0.049	0.016	12.94
		SiLRTC	–	–	–	–	18.43	6.16	2.07	0.035	0.016	0.015	33.88
		HaLRTC	–	–	–	–	18.79	1.03	0.25	0.060	0.017	0.015	4.57
		tSVD	–	–	–	–	4.12	0.25	0.01	0.063	0.018	0.016	96.39

F - feature scaling, P - pseudo-measurements, W - weighted errors, M - system model, VM - voltage magnitude, VA - voltage angle.

Table 4.3: State estimation results of IEEE 33-node distribution system at 5% sensor noise and different fractions of data availabilities.

Evaluation Metric							VM MAPE (in %)			VA MAE (in degrees)			Avg. time per tensor frame (in ms)
FAD							30	50	70	30	50	70	
Category	Group	Algorithm	Modules		Loss Fn./ Constraints								
			F	P	W	M							
NN-TC	III	CoSTCo	✓[1]	×	×	×	0.05	0.02	0.02	0.015	0.012	0.015	309.93
MC-SE	IV	RMCSE	✓[1]	×	×	✓	0.07	0.03	0.02	0.023	0.012	0.008	372.97
TC-SE	V	LRTC	✓	×	✓	✓	0.03	0.02	0.01	0.013	0.010	0.007	861.43
		LBTC	✓	✓	✓	✓	0.04	0.02	0.02	0.004	0.002	0.001	1404.51
DNN-SE	VI	PhyDNN	–	–	×	✓	0.03	0.03	0.03	0.019	0.018	0.018	6.25
GNN-SE	VII	GNSE	–	–	×	×	0.17	0.17	0.15	0.018	0.015	0.012	6.52
Proposed GNN-SE (Ablation Study)	VIII.a	MGCN	×	×	×	×	0.17	0.17	0.16	0.015	0.014	0.014	4.55
		TGCN	×	×	×	×	0.16	0.15	0.13	0.013	0.014	0.014	7.71
	VIII.b	MGCN-F	✓	×	×	×	0.18	0.17	0.14	0.012	0.011	0.009	4.82
		TGCN-F	✓	×	×	×	0.16	0.15	0.13	0.013	0.013	0.014	7.74
	VIII.c	MGCN-P	×	✓	×	×	0.12	0.12	0.12	0.018	0.018	0.017	4.85
		TGCN-P	×	✓	×	×	0.08	0.08	0.08	0.020	0.020	0.020	7.77
Proposed GNN-SE (Using WMSE)	IX	MGCN-WP	×	✓	✓	×	0.11	0.11	0.11	0.016	0.016	0.016	4.81
		TGCN-WP	×	✓	✓	×	0.07	0.07	0.06	0.021	0.021	0.021	7.71
Proposed GNN-SE	X	MGCN-FP	✓	✓	×	×	0.05	0.04	0.04	0.014	0.013	0.013	4.77
		TGCN-FP	✓	✓	×	×	0.04	0.03	0.03	0.012	0.011	0.011	7.80

F - feature scaling, P - pseudo-measurements, W - weighted errors, M - system model, VM - voltage magnitude, VA - voltage angle.
[1]The original papers do not utilize feature scaling. However, it was observed that these algorithms benefit from feature scaling.

Table 4.4: State estimation results of IEEE 37-node distribution system at 5% sensor noise and different fractions of data availabilities.

Evaluation Metric							VM MAPE (in %)			VA MAE (in degrees)			Avg. time per tensor frame (in ms)
FAD							15	25	35	15	25	35	
Category	Group	Algorithm	Modules		Loss Fn./ Constraints								
			F	P	W	M							
MC	I	IALM	–	–	–	–	3.95	2.91	2.06	0.625	0.467	0.338	9.75
		GROUSE	–	–	–	–	1.58	0.49	0.26	0.377	0.121	0.068	21.90
		EORIMP	–	–	–	–	2.96	1.61	0.98	0.472	0.263	0.166	0.14
		ORIMP	–	–	–	–	2.96	1.62	0.99	0.472	0.264	0.167	0.10
		MAMR	–	–	–	–	1.93	1.07	0.81	0.393	0.262	0.214	1.60
		RMAMR	–	–	–	–	1.93	1.07	0.81	0.392	0.262	0.213	1.50
		SVT	–	–	–	–	3.95	3.13	2.55	0.625	0.501	0.413	8.50
		SchattenPMC	–	–	–	–	1.45	0.49	0.25	0.238	0.094	0.058	81.28
TC	II	BCPFic	–	–	–	–	2.04	1.12	1.68	0.461	0.221	0.389	501.05
		FaLRTC	–	–	–	–	1.70	0.69	0.39	0.275	0.122	0.077	103.72
		SiLRTC	–	–	–	–	1.59	0.58	0.31	0.258	0.105	0.065	44.92
		HaLRTC	–	–	–	–	3.72	2.70	2.02	0.589	0.435	0.330	61.10
		tSVD	–	–	–	–	1.56	0.66	0.41	0.258	0.120	0.081	414.23

F - feature scaling, P - pseudo-measurements, W - weighted errors, M - system model, VM - voltage magnitude, VA - voltage angle.

Table 4.5: State estimation results of IEEE 37-node distribution system at 5% sensor noise and different fractions of data availabilities.

Evaluation Metric							VM MAPE (in %)			VA MAE (in degrees)			Avg. time per tensor frame (in ms)
FAD							15	25	35	15	25	35	
Category	Group	Algorithm	Modules		Loss Fn./ Constraints								
			F	P	W	M							
NN-TC	III	CoSTCo	✓[1]	×	×	×	0.23	0.12	0.08	0.074	0.043	0.037	5426.00
MC-SE	IV	RMCSE	✓[1]	×	×	✓	0.25	0.09	0.05	0.054	0.027	0.020	765.97
TC-SE	V	LRTC	✓	×	✓	✓	0.18	0.05	0.03	0.056	0.028	0.021	11123.62
		LBTC	✓	✓	✓	✓	0.14	0.05	0.03	0.029	0.010	0.006	25488.42
DNN-SE	VI	PhyDNN	–	–	×	✓	0.16	0.15	0.14	0.058	0.051	0.050	12.67
GNN-SE	VII	GNSE	–	–	×	×	0.97	0.89	0.60	0.193	0.145	0.113	27.34
Proposed GNN-SE (Ablation Study)	VIII.a	MGCN	×	×	×	×	1.02	1.00	0.96	0.162	0.158	0.151	18.02
		TGCN	×	×	×	×	1.12	1.00	0.92	0.177	0.158	0.141	38.03
	VIII.b	MGCN-F	✓	×	×	×	0.81	0.73	0.62	0.135	0.122	0.102	17.98
		TGCN-F	✓	×	×	×	1.18	1.14	1.07	0.194	0.186	0.172	37.78
	VIII.c	MGCN-P	×	✓	×	×	0.80	0.79	0.79	0.128	0.127	0.126	18.71
		TGCN-P	×	✓	×	×	0.62	0.57	0.55	0.095	0.088	0.085	38.41
Proposed GNN-SE (Using WMSE)	IX	MGCN-WP	×	✓	✓	×	0.71	0.71	0.70	0.117	0.115	0.114	18.80
		TGCN-WP	×	✓	✓	×	0.62	0.57	0.55	0.095	0.088	0.085	38.10
Proposed GNN-SE	X	MGCN-FP	✓	✓	×	×	0.43	0.38	0.35	0.083	0.073	0.069	18.17
		TGCN-FP	✓	✓	×	×	0.19	0.16	0.14	0.051	0.049	0.047	38.47

F - feature scaling, P - pseudo-measurements, W - weighted errors, M - system model, VM - voltage magnitude, VA - voltage angle.
[1]The original papers do not utilize feature scaling. However, it was observed that these algorithms benefit from feature scaling.

Table 4.6: State estimation results on primary and secondary nodes of 559-node distribution system at 5% sensor noise and different fractions of data availabilities.

Evaluation Metric							VM MAPE (in %)			VA MAE (in degrees)			Avg. time per tensor frame (in s)
FAD							15	25	35	15	25	35	
Category	Group	Algorithm	Modules		Loss Fn./ Constraints								
			F	P	W	M							
MC	I	IALM	–	–	–	–	4.27	3.33	2.01	0.669	0.528	0.330	0.014
		GROUSE	–	–	–	–	1.68	0.68	0.70	0.354	0.159	0.191	0.070
		EORIMP	–	–	–	–	2.86	1.46	0.79	0.454	0.240	0.136	0.010
		ORIMP	–	–	–	–	2.86	1.46	0.77	0.455	0.239	0.134	0.011
		MAMR	–	–	–	–	2.06	1.11	0.70	0.392	0.248	0.175	0.003
		RMAMR	–	–	–	–	2.06	1.11	0.70	0.392	0.247	0.174	0.003
		SVT	–	–	–	–	4.41	3.77	3.06	0.691	0.594	0.490	0.015
		SchattenPMC	–	–	–	–	1.67	0.69	0.30	0.271	0.124	0.066	0.256
TC	II	BCPFic	–	–	–	–	5.73	4.91	3.70	1.877	1.994	1.705	1.669
		FaLRTC	–	–	–	–	1.77	0.76	0.34	0.285	0.132	0.071	0.025
		SiLRTC	–	–	–	–	1.76	0.71	0.31	0.285	0.126	0.067	0.073
		HaLRTC	–	–	–	–	4.39	3.73	3.01	0.687	0.588	0.482	0.090
		tSVD	–	–	–	–	1.88	0.94	0.53	0.304	0.163	0.100	5.652

F - feature scaling, P - pseudo-measurements, W - weighted errors, M - system model, VM - voltage magnitude, VA - voltage angle.

Table 4.7: State estimation results on primary and secondary nodes of 559-node distribution system at 5% sensor noise and different fractions of data availabilities.

Evaluation Metric							VM MAPE (in %)			VA MAE (in degrees)			Avg. time per tensor frame (in s)
FAD							15	25	35	15	25	35	
Category	Group	Algorithm	Modules		Loss Fn./ Constraints								
			F	P	W	M							
NN-TC	III	CoSTCo	✓[1]	×	×	×	0.25	0.25	0.21	0.095	0.094	0.099	24.420
MC-SE	IV	RMCSE	✓[1]	×	×	✓	0.20	0.13	0.11	0.043	0.031	0.025	13.610
TC-SE	V	LRTC	✓	×	✓	✓	0.16	0.11	0.09	0.040	0.030	0.025	173.300
		LBTC	✓	✓	✓	✓	0.12	0.05	0.03	0.032	0.020	0.016	328.100
DNN-SE	VI	PhyDNN	–	–	×	✓	0.15	0.15	0.13	0.050	0.049	0.046	0.066
GNN-SE	VII	GNSE	–	–	×	×	0.91	0.90	0.67	0.182	0.129	0.112	0.374
Proposed GNN-SE (Ablation Study)	VIII.a	MGCN	×	×	×	×	1.21	1.17	1.16	0.189	0.182	0.181	0.318
		TGCN	×	×	×	×	1.18	1.10	1.01	0.188	0.173	0.158	0.426
	VIII.b	MGCN-F	✓	×	×	×	0.73	0.49	0.50	0.131	0.098	0.097	0.313
		TGCN-F	✓	×	×	×	1.31	1.28	1.24	0.213	0.208	0.202	0.424
	VIII.c	MGCN-P	×	✓	×	×	1.12	1.07	1.05	0.179	0.172	0.168	0.348
		TGCN-P	×	✓	×	×	0.70	0.63	0.60	0.107	0.095	0.090	0.439
Proposed GNN-SE (Using WMSE)	IX	MGCN-WP	×	✓	✓	×	1.07	1.02	1.00	0.183	0.175	0.173	0.333
		TGCN-WP	×	✓	✓	×	0.66	0.59	0.56	0.122	0.112	0.109	0.445
Proposed GNN-SE	X	MGCN-FP	✓	✓	×	×	0.30	0.25	0.23	0.097	0.091	0.091	0.339
		TGCN-FP	✓	✓	×	×	0.27	0.22	0.20	0.050	0.045	0.042	0.422

F - feature scaling, P - pseudo-measurements, W - weighted errors, M - system model, VM - voltage magnitude, VA - voltage angle.
[1]The original papers do not utilize feature scaling. However, it was observed that these algorithms benefit from feature scaling.

Table 4.8: Switch configurations for the 559-node distribution system.

Switches	Topology Index							
	1	2	3	4	5	6	7	8
S1	✓	✓	✓	✓	✕	✕	✕	✕
S2	✓	✓	✕	✕	✓	✓	✕	✕
S3	✓	✕	✓	✕	✓	✕	✓	✕
S4	✕	✕	✕	✕	✓	✓	✓	✓
S5	✕	✕	✓	✓	✕	✕	✓	✓
S6	✕	✓	✕	✓	✕	✓	✕	✓

(On: ✓, Off: ✕)

CHAPTER 5

MODEL AND NETWORK CONFIGURATION SELECTION: TOPOLOGY IDENTIFICATION

5.1 INTRODUCTION

Distribution system applications such as state estimation, fault detection, and control require accurate knowledge of networks' topology. However, due to insufficient measurements and monitoring, utilities fail to capture the status of all switches in the network. Several methods proposed for state estimation and control cannot be implemented without the knowledge of network topology and parameters, such as Madbhavi *et al.* (2020, 2021). Therefore, topology identification techniques have received significant research interest, for example: Singh *et al.* (2010); Karimi and Natarajan (2021); Baran *et al.* (2009). Therefore, a topology identification approach is proposed and presented in this chapter to assist the state estimation techniques developed in the previous chapters. This technique aids the developed state estimators in model selection and network configuration identification.

5.2 PROPOSED APPROACH

The state measurement matrix structure, graph model of the distribution system, and the graph convolutional network (GCN) model for topology identification are described in the following sub-sections.

5.2.1 Data Structure

Consider a three-phase distribution system with m buses, n branches, and measurements from N_T time instants. The measurements from this system at a time instant t can be rearranged to form a state measurement matrix $\boldsymbol{M_t} \in \mathbb{R}^{m\times 12}$ as described below,

$$\boldsymbol{M_t} = \begin{bmatrix} \boldsymbol{M_t^{(a)}} & \boldsymbol{M_t^{(b)}} & \boldsymbol{M_t^{(c)}} \end{bmatrix}$$

Where, $\boldsymbol{M_t^{(a)}}, \boldsymbol{M_t^{(b)}}, \boldsymbol{M_t^{(c)}} \in \mathbb{R}^{m \times 4}$ are the state measurement matrices for the individual phases of the three-phase distribution system. Thus, for a phase k, the state measurement matrix is defined as follows,

$$\begin{aligned}
\boldsymbol{M_t^{(k)}} &= [Re\{\boldsymbol{V_t^{(k)}}\}, Im\{\boldsymbol{V_t^{(k)}}\}, \boldsymbol{P_t^{in(k)}}, \boldsymbol{Q_t^{in(k)}}] - \\
&\quad [Re\{\boldsymbol{V_{ref}^{(k)}}\}, Im\{\boldsymbol{V_{ref}^{(k)}}\}, \boldsymbol{[0]_{m\times 1}}, \boldsymbol{[0]_{m\times 1}}] \\
\boldsymbol{V_{ref}^{(a)}} &= \boldsymbol{[1]_{m\times 1}} \\
\boldsymbol{V_{ref}^{(b)}} &= \boldsymbol{[1]_{m\times 1}} \times e^{-j(2\pi/3)} \\
\boldsymbol{V_{ref}^{(c)}} &= \boldsymbol{[1]_{m\times 1}} \times e^{-j(4\pi/3)}
\end{aligned}$$

Here, $\boldsymbol{V_t^{(k)}}$, $\boldsymbol{P_t^{in(k)}}$, and $\boldsymbol{Q_t^{in(k)}}$ are the complex bus voltage vector and the rectangular components of the complex power injection vector at time instant t for the phase k. $\boldsymbol{[1]_{r\times c}}$ and $\boldsymbol{[0]_{r\times c}}$ are matrices of size $r \times c$ having only ones and zeros, respectively.

5.2.2 System Model

The distribution system can be represented using a graph $G = (V, E)$, such that V is the set of all buses and E is the set of all branches in the system. Fig. 5.1b illustrates the graph representation of a simple distribution system shown in Fig. 5.1a. The graph obtained by retaining only the edges containing switches is denoted by the graph $G_T = (V, E_T)$ as shown in Fig. 5.1c. The graph obtained by disconnecting all edges containing switches is denoted by the graph $G_I = (V, E_I)$ as shown in Fig. 5.1d.

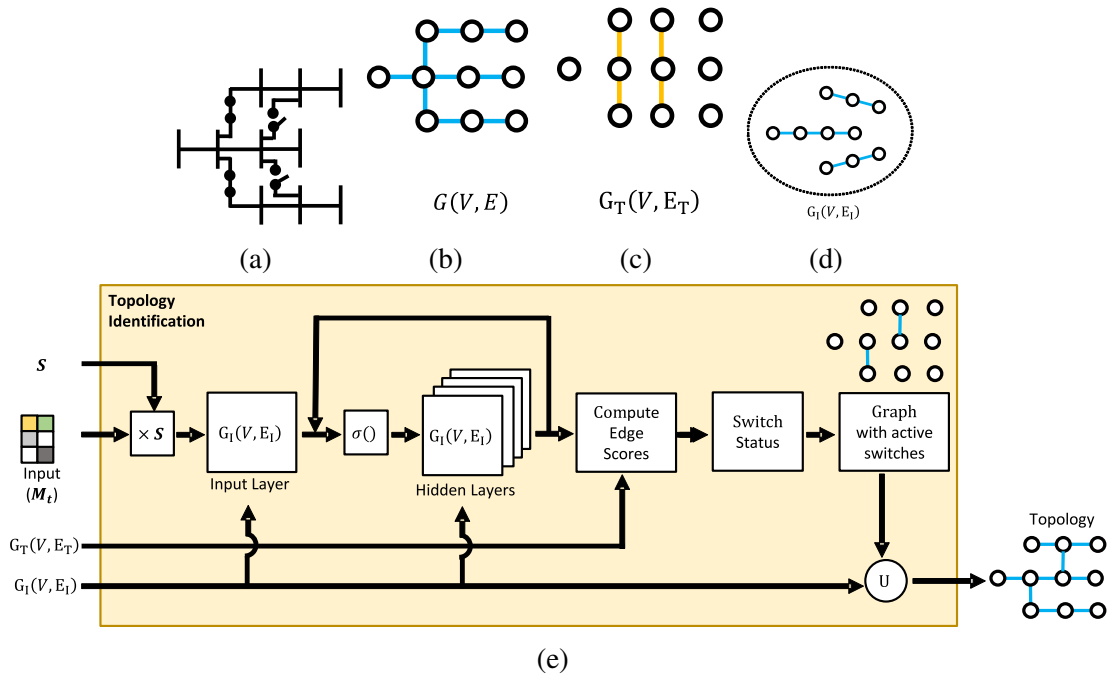


Figure 5.1: Topology identification using graph neural networks (a) Distribution system diagram (b) Graph representation of distribution system (c) Graph of distribution system with only edges containing switches (d) Graph of distribution system after removing edges containing switches (e) Overview of topology identification.

5.2.3 GCN Model

The GCN-based topology identification approach comprises multiple GCN layers operating on G_I with the rectified linear unit (ReLu) as the activation function. The initial feature vector for each node is obtained from its corresponding row in the scaled state measurement matrix $\boldsymbol{M_t S}$. Here, $\boldsymbol{S} = diag([s_1, s_2, s_3, s_4, ..., s_{12}])$ and s_1, s_2, s_3, ... s_{12} are the scaling factors for the columns of the state measurement matrix. Feature scaling prevents large disparities among GCN weights corresponding to different physical quantities in the state measurement matrices and tensors. The final node embeddings are used to compute branch scores. The links in each set of switch combinations with the highest scores are predicted to be 'on'. For example, consider the primary distribution system in the 559-node network as shown in Fig. 5.2. Among the switch pairs $(S1, S4)$, $(S2, S5)$, $(S3, S6)$, only one switch from each pair can be on if the radial structure is to be maintained. Thus, links corresponding to switches with the highest edge score in each switch pair are predicted to be connected. The approach is illustrated in Fig. 5.1e.

5.3 CASE STUDIES

The proposed approach has been evaluated on a three-phase unbalanced 559-node distribution system, see Appendix A. The GCN-based approach has been compared with the estimator pool technique utilizing the matrix completion-based method state estimator RMCSE by Liu *et al.* (2019*a*) (robust matrix completion state estimator). The proposed approach has also been compared with BCSE, from Baran *et al.* (2009).

The primary distribution network in the 559-node distribution system is shown in Fig. 5.2. Three additional branches are added to generate different network topologies (marked in red). The switches $S1$, $S2$, ..., and $S6$ generate eight radial network configurations. Table 5.1 provides the corresponding topologies. Smart meter data was sourced from the eGauge website. The load data was assigned to the nodes and scaled so that the

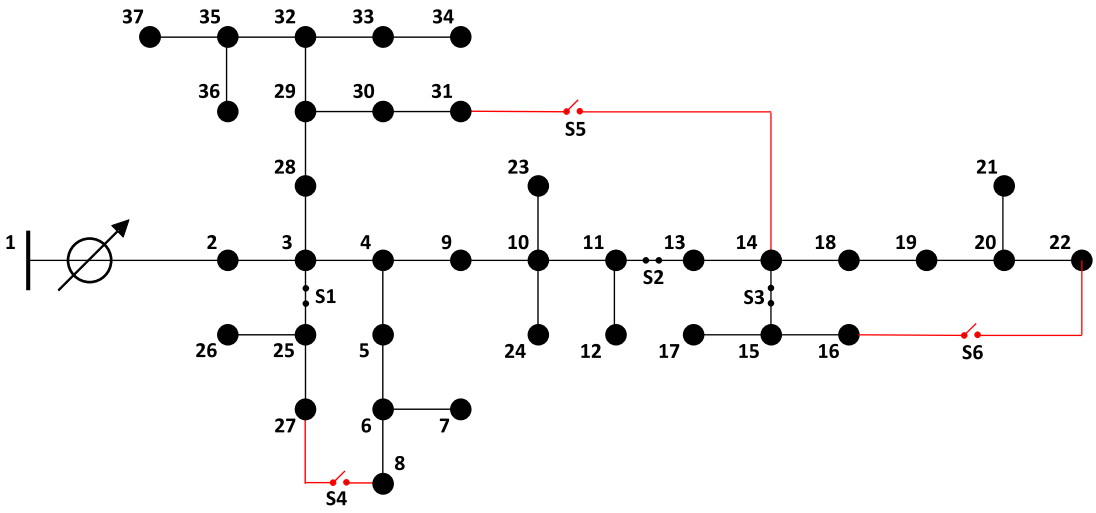


Figure 5.2: The primary distribution network in the 559-node system.

bus's peak loads were within the nominal loads. Power flow analysis was performed for various network topologies to create the state measurement tensors ($\mathcal{T}$). Gaussian noise with a 5% standard deviation of the measurement magnitude is added to simulate noisy measurements. Further, tensors were created with various fractions of available data (FAD) of 15%, 25%, and 35%. Here, the fraction of available data (FAD) is the ratio of available measurements to the total number of measurements.

5.3.1 Model Generation

State measurement matrices and tensors for training the GNN model were created using load profiles different from those of the test sets on which the performance of all algorithms has been evaluated. 3-layer GCN models were created and connected to an output layer. The output layer consists of six nodes each corresponding to one of the switches $S1$, $S2$, ..., and $S6$. Training a GCN model involves forward passes through the layers and updating the weights in the backward propagation step. The models were trained using partial state measurement matrices as input features and the switch statuses as the expected output. Cross-entropy loss was chosen as the loss function.

Table 5.1: Switch configurations for different topologies of the 559-node distribution system.

Topology Index	Switches (On: ✓) S1	S2	S3	S4	S5	S6
1	✓	✓	✓			
2	✓	✓				✓
3	✓		✓		✓	
4	✓				✓	✓
5		✓	✓	✓		
6		✓		✓		✓
7			✓	✓	✓	
8				✓	✓	✓

5.3.2 Topology Identification

The proposed approach provided 92.4% accuracy at 15% FAD. The confusion matrix for classification is provided in Fig. 5.3. This accuracy is comparable and slightly higher than the estimator pool approach, which provides 91.7% accuracy, and the BCSE, which provides 90.5% accuracy. However, at higher FADs of 25% and 35%, all approaches provided near 100% accuracy, with only a few misclassifications. BCSE provided slightly lower accuracy than the other methods because it is sensitive to the threshold, which is used as the stopping criterion. Lowering the stopping criteria improves the accuracy at the cost of more iterations.

Finally, the approaches are compared for their computation times in Table 5.2. It can be observed that the proposed GCN-based approach requires computational time several orders of magnitude lower than the estimator-based methods. Estimator pool techniques utilizing sparsity-aware state estimators such as RMCSE require large computational times since they have large computational complexities and are required to evaluate all possible topologies. The computational time can be improved using the BCSE approach, which iteratively switches branches from the on/off states to get the residuals within the predefined threshold. Therefore, the computation time depends on the approach's

Table 5.2: Computational time of various topology identification strategies on the 559-node distribution system.

Method	Computation Time (s)
Estimator pool with RMCSE	181.23
BCSE	22.41 - 170.97
Proposed GCN	0.48

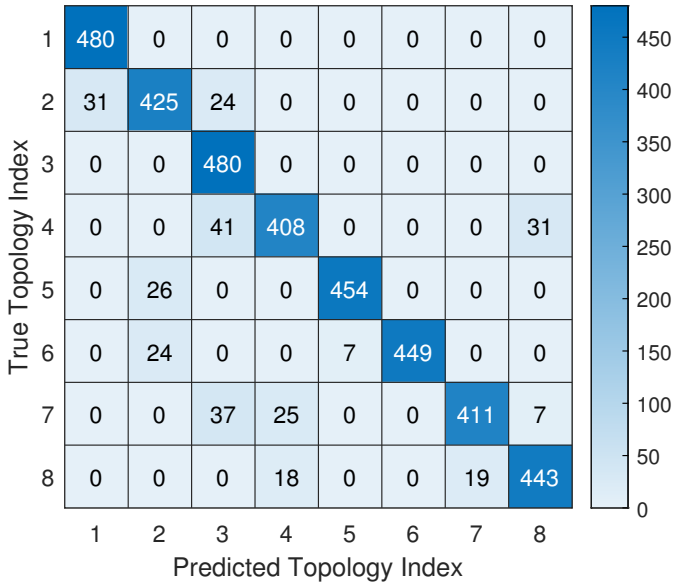


Figure 5.3: Topology identification confusion matrix for the proposed approach with 15% FAD.

number of iterations. However, it is still several orders of magnitude higher than the proposed GCN-based approach.

5.4 SUMMARY

Topology identification is essential for several distribution system analyses. Current works on topology identification in distribution systems suffer from large computational complexity, rendering them unsuitable for real-time applications. Therefore, a topology identification approach using graph neural networks has been developed in this chapter. The proposed link prediction-based topology identification approach avoids evaluating

multiple network topologies using state estimators as required by conventional approaches. The proposed approach was evaluated on a three-phase unbalanced 559-node distribution system and provided comparable performance to existing approaches with significantly lower computational times. In the next chapter, we develop error bounds for model-based estimators.

CHAPTER 6

OPTIMIZING DSSE PERFORMANCE: ERROR ANALYSIS

6.1 INTRODUCTION

The state estimators developed in the previous chapters have shown significant improvement in state estimation accuracy compared to prior state-of-the-art techniques. However, no studies have been performed that quantify the error bounds of such sparsity-aware estimators. Further, the choice of scaling factors have a significant impact on the performance of these estimators. Therefore, in this chapter, we develop error bounds for a matrix completion-based distribution system state estimator. The proposed approach can be extended to incorporate errors due to parameter uncertainty. Additionally, we develop an algorithm to solve the optimization problem and implement an approach to automatically update scaling factors.

6.2 PROPOSED APPROACHES

6.2.1 Error Bounds

Consider a p-phase distribution system with one slack bus, N_b PQ buses and N_{br} branches. The rectangular components of bus voltage phasors, magnitude of bus voltage phasors, and rectangular components of power injections at any time instant t can be rearranged to form a state measurement matrix ($\boldsymbol{M_t} \in \mathbb{R}^{pN_b \times 5}$). Let the data availability be characterized by a binary matrix $\boldsymbol{\Lambda_t}$. The linearized power flow equations derived in Bernstein and Dall'Anese (2017) have been used as constraints in matrix and tensor completion-based distribution system state estimators in Liu *et al.* (2019*a*); Madbhavi *et al.* (2020, 2021). Consider the matrix completion-based state estimator developed

in Liu *et al.* (2019*a*). It can be rewritten as follows,

$$\begin{aligned} \hat{\boldsymbol{X}} = \min_{\boldsymbol{X}} : \quad & \|\boldsymbol{X}\|_* + \alpha \left\|(\boldsymbol{M_t} - \boldsymbol{X}) \bigodot \boldsymbol{\Lambda_t}\right\|_F^2 \\ \text{S.T.} \quad & \left\|\boldsymbol{v_r} - \boldsymbol{\mathcal{D}_r}\left[\boldsymbol{p}^T \boldsymbol{q}^T\right]^T - \boldsymbol{w_r}\right\|_\infty \leq \gamma_r \\ & \left\|\boldsymbol{v_i} - \boldsymbol{\mathcal{D}_i}\left[\boldsymbol{p}^T \boldsymbol{q}^T\right]^T - \boldsymbol{w_i}\right\|_\infty \leq \gamma_i \\ & \left\|\boldsymbol{v_m} - \boldsymbol{\mathcal{K}}\left[\boldsymbol{p}^T \boldsymbol{q}^T\right]^T - \boldsymbol{w_m}\right\|_\infty \leq \gamma_m \end{aligned} \tag{6.23}$$

where, $\bigodot$ is the Hadamard product. To simplify notations, the columns of $\boldsymbol{X}$ are referred as $\boldsymbol{v_r}$, $\boldsymbol{v_i}$, $\boldsymbol{v_m}$, $\boldsymbol{p}$, and $\boldsymbol{q}$, respectively. γ_r, γ_i, and γ_m are the allowed deviations of the real components, imaginary components, and magnitudes of the voltage phasors from the linearized model, respectively. The coefficient matrices ($\boldsymbol{\mathcal{D}}$ and $\boldsymbol{\mathcal{K}}$) and the constant phasor vector $\boldsymbol{w}$, can be obtained from Bernstein and Dall'Anese (2017), Further, to simplify notations, consider,

$$\begin{aligned} \boldsymbol{\mathcal{D}_r} = Re\left\{\boldsymbol{\mathcal{D}}\right\}, \boldsymbol{\mathcal{D}_i} = Im\left\{\boldsymbol{\mathcal{D}}\right\} \\ \boldsymbol{w_r} = Re\left\{\boldsymbol{w}\right\}, \boldsymbol{w_i} = Im\left\{\boldsymbol{w}\right\}, \boldsymbol{w_m} = \left|\boldsymbol{w}\right|. \end{aligned} \tag{6.24}$$

Now, consider the true measurements for the real part of bus voltage vector ($\boldsymbol{v_{r_0}}$), imaginary part of bus voltage vector ($\boldsymbol{v_{i_0}}$), magnitude of bus voltage vector ($\boldsymbol{v_{m_0}}$), and the rectangular components of power injections at the buses ($\boldsymbol{p_0}$ and $\boldsymbol{q_0}$). These measurements can be introduced in the linearized power flow equations derived in Bernstein and Dall'Anese (2017) as follows,

$$\begin{aligned} \boldsymbol{v_{r_0}} &= \boldsymbol{\mathcal{D}_r}\left[\boldsymbol{p_0}^T \boldsymbol{q_0}^T\right]^T + \boldsymbol{w_r} + \boldsymbol{\eta_r} \\ \boldsymbol{v_{i_0}} &= \boldsymbol{\mathcal{D}_i}\left[\boldsymbol{p_0}^T \boldsymbol{q_0}^T\right]^T + \boldsymbol{w_i} + \boldsymbol{\eta_i} \\ \boldsymbol{v_{m_0}} &= \boldsymbol{\mathcal{K}}\left[\boldsymbol{p_0}^T \boldsymbol{q_0}^T\right]^T + \boldsymbol{w_m} + \boldsymbol{\eta_m} \end{aligned} \tag{6.25}$$

where, $\boldsymbol{\eta_r}$, $\boldsymbol{\eta_i}$, and $\boldsymbol{\eta_m}$ are the modeling errors introduced by linearizing the power flow equations. It was observed in Bernstein and Dall'Anese (2017), that the errors in approximating the real components, imaginary components and magnitudes of the bus voltage vectors by their linearized forms is quite low (less than 0.5%). Therefore, the modeling errors are dropped from these equations[1].

[1] Future studies will attempt to incorporate modeling errors.

The estimated bus voltage vector components and magnitudes can be written as follows,

$$\begin{aligned}
\hat{\boldsymbol{v}}_{\boldsymbol{r}} &= \mathcal{D}_{\boldsymbol{r}} \left[\hat{\boldsymbol{p}}^T \hat{\boldsymbol{q}}^T\right]^T + \boldsymbol{w}_{\boldsymbol{r}} + \boldsymbol{\delta}_{\boldsymbol{r}} \\
\hat{\boldsymbol{v}}_{\boldsymbol{i}} &= \mathcal{D}_{\boldsymbol{i}} \left[\hat{\boldsymbol{p}}^T \hat{\boldsymbol{q}}^T\right]^T + \boldsymbol{w}_{\boldsymbol{i}} + \boldsymbol{\delta}_{\boldsymbol{i}} \\
\hat{\boldsymbol{v}}_{\boldsymbol{m}} &= \mathcal{K} \left[\hat{\boldsymbol{p}}^T \hat{\boldsymbol{q}}^T\right]^T + \boldsymbol{w}_{\boldsymbol{m}} + \boldsymbol{\delta}_{\boldsymbol{m}} \\
|\boldsymbol{\delta}_{\boldsymbol{r}}|_\infty \leq \gamma_r, &\quad |\boldsymbol{\delta}_{\boldsymbol{i}}|_\infty \leq \gamma_i, \quad |\boldsymbol{\delta}_{\boldsymbol{m}}|_\infty \leq \gamma_m
\end{aligned} \tag{6.26}$$

Let $\mathcal{A} = \left[\mathcal{D}_{\boldsymbol{r}}{}^T \quad \mathcal{D}_{\boldsymbol{i}}{}^T \quad \mathcal{K}^T\right]^T$ and Ω_r, Ω_i, Ω_m. Ω_p, and Ω_q be the buses whose $\boldsymbol{v}_{\boldsymbol{r}}$, $\boldsymbol{v}_{\boldsymbol{i}}$, $\boldsymbol{v}_{\boldsymbol{m}}$. $\boldsymbol{p}$, and $\boldsymbol{q}$ are known, respectively. Also, Ω_z^c denotes the complement of the set Ω_z. Additionally, the notation $\mathcal{A}_{\boldsymbol{x},\boldsymbol{y}}$ denotes the sub-matrix of $\mathcal{A}$ comprising of rows x and columns y. Then, the estimation error can be written as,

$$\begin{bmatrix} (\boldsymbol{e}_{\boldsymbol{r}})_{\Omega_r} \\ (\boldsymbol{e}_{\boldsymbol{r}})_{\Omega_r^c} \\ (\boldsymbol{e}_{\boldsymbol{i}})_{\Omega_i} \\ (\boldsymbol{e}_{\boldsymbol{i}})_{\Omega_i^c} \\ (\boldsymbol{e}_{\boldsymbol{m}})_{\Omega_m} \\ (\boldsymbol{e}_{\boldsymbol{m}})_{\Omega_m^c} \end{bmatrix} = \begin{bmatrix} \mathcal{A}_{\Omega_r,\Omega_p} & \mathcal{A}_{\Omega_r,\Omega_q} \\ \mathcal{A}_{\Omega_r^c,\Omega_p} & \mathcal{A}_{\Omega_r^c,\Omega_q} \\ \mathcal{A}_{\Omega_i,\Omega_p} & \mathcal{A}_{\Omega_i,\Omega_q} \\ \mathcal{A}_{\Omega_i^c,\Omega_p} & \mathcal{A}_{\Omega_i^c,\Omega_q} \\ \mathcal{A}_{\Omega_m,\Omega_p} & \mathcal{A}_{\Omega_m,\Omega_q} \\ \mathcal{A}_{\Omega_m^c,\Omega_p} & \mathcal{A}_{\Omega_m^c,\Omega_q} \end{bmatrix} \begin{bmatrix} (\boldsymbol{e}_{\boldsymbol{p}})_{\Omega_p} \\ (\boldsymbol{e}_{\boldsymbol{q}})_{\Omega_q} \end{bmatrix} + \begin{bmatrix} \mathcal{A}_{\Omega_r,\Omega_p^c} & \mathcal{A}_{\Omega_r,\Omega_q^c} \\ \mathcal{A}_{\Omega_r^c,\Omega_p^c} & \mathcal{A}_{\Omega_r^c,\Omega_q^c} \\ \mathcal{A}_{\Omega_i,\Omega_p^c} & \mathcal{A}_{\Omega_i,\Omega_q^c} \\ \mathcal{A}_{\Omega_i^c,\Omega_p^c} & \mathcal{A}_{\Omega_i^c,\Omega_q^c} \\ \mathcal{A}_{\Omega_m,\Omega_p^c} & \mathcal{A}_{\Omega_m,\Omega_q^c} \\ \mathcal{A}_{\Omega_m^c,\Omega_p^c} & \mathcal{A}_{\Omega_m^c,\Omega_q^c} \end{bmatrix} \begin{bmatrix} (\boldsymbol{e}_{\boldsymbol{p}})_{\Omega_p^c} \\ (\boldsymbol{e}_{\boldsymbol{q}})_{\Omega_q^c} \end{bmatrix}$$

$$+ \begin{bmatrix} (\boldsymbol{\delta}_{\boldsymbol{r}})_{\Omega_r} \\ (\boldsymbol{\delta}_{\boldsymbol{r}})_{\Omega_r^c} \\ (\boldsymbol{\delta}_{\boldsymbol{i}})_{\Omega_i} \\ (\boldsymbol{\delta}_{\boldsymbol{i}})_{\Omega_i^c} \\ (\boldsymbol{\delta}_{\boldsymbol{m}})_{\Omega_m} \\ (\boldsymbol{\delta}_{\boldsymbol{m}})_{\Omega_m^c} \end{bmatrix} \tag{6.27}$$

A larger magnitude of the penalty weight α in (6.23) results in the estimated values of known quantities to be close to the measured values. Furthermore, these estimates can

be constrained to be within the 2-norm ball of radius equal to the corresponding sensor accuracy (ϵ_r, ϵ_i, ϵ_m, ϵ_p, or ϵ_q). Thus, the generalized expression for estimation errors of the known quantities in (6.27) can be written as follows,

$$\left\|(\boldsymbol{e_x})_{\boldsymbol{\Omega_x}}\right\|_F^2 \leq 4\left|\Omega_x\right|\epsilon_x^2 \tag{6.28}$$

Using Frobenius norm expansion properties on $(\boldsymbol{e_r})_{\boldsymbol{\Omega_r}}$ $(\boldsymbol{e_i})_{\boldsymbol{\Omega_i}}$, and $(\boldsymbol{e_m})_{\boldsymbol{\Omega_m}}$ in (6.27), the estimation errors of unknown power injections can be bounded as follows,

$$\begin{aligned}
E_1 &= \frac{4\left\|\mathcal{A}_{\boldsymbol{\Omega_r},\boldsymbol{\Omega_p}\cup\boldsymbol{\Omega_q}}\right\|_F^2\left(\left|\Omega_p\right|\epsilon_p^2+\left|\Omega_q\right|\epsilon_q^2\right)+\left|\Omega_r\right|\left(4\epsilon_r^2+\gamma_r^2\right)}{\left\|\mathcal{A}_{\boldsymbol{\Omega_r},\boldsymbol{\Omega_p^c}\cup\boldsymbol{\Omega_q^c}}\right\|_F^2} \\
E_2 &= \frac{4\left\|\mathcal{A}_{\boldsymbol{\Omega_i},\boldsymbol{\Omega_p}\cup\boldsymbol{\Omega_q}}\right\|_F^2\left(\left|\Omega_p\right|\epsilon_p^2+\left|\Omega_q\right|\epsilon_q^2\right)+\left|\Omega_i\right|\left(4\epsilon_i^2+\gamma_i^2\right)}{\left\|\mathcal{A}_{\boldsymbol{\Omega_i},\boldsymbol{\Omega_p^c}\cup\boldsymbol{\Omega_q^c}}\right\|_F^2} \\
E_3 &= \frac{4\left\|\mathcal{A}_{\boldsymbol{\Omega_m},\boldsymbol{\Omega_p}\cup\boldsymbol{\Omega_q}}\right\|_F^2\left(\left|\Omega_p\right|\epsilon_p^2+\left|\Omega_q\right|\epsilon_q^2\right)+\left|\Omega_m\right|\left(4\epsilon_m^2+\gamma_m^2\right)}{\left\|\mathcal{A}_{\boldsymbol{\Omega_m},\boldsymbol{\Omega_p^c}\cup\boldsymbol{\Omega_q^c}}\right\|_F^2} \\
\left\|(\boldsymbol{e_p})_{\boldsymbol{\Omega_p^c}}\right\|_F^2 &+ \left\|(\boldsymbol{e_q})_{\boldsymbol{\Omega_q^c}}\right\|_F^2 \leq \max\left(E_1, E_2, E_3\right)
\end{aligned} \tag{6.29}$$

The generalized expansions for squared Frobenius norms of $(\boldsymbol{e_r})_{\boldsymbol{\Omega_r^c}}$, $(\boldsymbol{e_i})_{\boldsymbol{\Omega_i^c}}$, and $(\boldsymbol{e_m})_{\boldsymbol{\Omega_m^c}}$ in (6.27) can be written as follows,

$$\begin{aligned}
\left\|(\boldsymbol{e_x})_{\boldsymbol{\Omega_x^c}}\right\|_F^2 \quad \leq \quad & 4\left|\Omega_p\right|\epsilon_p^2\left\|\mathcal{A}_{\boldsymbol{\Omega_x^c},\boldsymbol{\Omega_p}}\right\|_F^2 \\
& +4\left|\Omega_q\right|\epsilon_q^2\left\|\mathcal{A}_{\boldsymbol{\Omega_x^c},\boldsymbol{\Omega_q}}\right\|_F^2 \\
& +\left\|\mathcal{A}_{\boldsymbol{\Omega_x^c},\boldsymbol{\Omega_p^c}\cup\boldsymbol{\Omega_q^c}}\right\|_F^2\left\|(\boldsymbol{e_p})_{\boldsymbol{\Omega_p^c}}\right\|_F^2 \\
& +\left\|\mathcal{A}_{\boldsymbol{\Omega_x^c},\boldsymbol{\Omega_p^c}\cup\boldsymbol{\Omega_q^c}}\right\|_F^2\left\|(\boldsymbol{e_q})_{\boldsymbol{\Omega_q^c}}\right\|_F^2
\end{aligned} \tag{6.30}$$

Futhermore, parameter uncertainty can be accounted in this formulation by re-writing the above steps with $\mathcal{A}+\boldsymbol{\Delta}\mathcal{A}$ instead of $\mathcal{A}$. The first order approximation of the inverse can be used for small perturbations. The upper bound for the estimation errors of $\boldsymbol{v_r}$, $\boldsymbol{v_i}$, and $\boldsymbol{v_m}$ can be obtained by substituting (6.28), (6.29), and (6.30) in the squared Frobenius

norm expansion of (6.27).

6.2.2 Auto-Scaling Factor Update

Consider the regularized form of the estimator defined in 6.23,

$$\boldsymbol{g} : \underset{\boldsymbol{X}}{\operatorname{argmin}} \quad \|\boldsymbol{X}\|_* + \lambda_1 \left\|\mathcal{A}_\Lambda.\mathrm{vec}\left(\left(\boldsymbol{M_t} - \boldsymbol{X}\right)\boldsymbol{S}\right)\right\|_F^2 + \lambda_2 \left\|\mathcal{A}_M.\mathrm{vec}\left(\boldsymbol{X}\right)\right\|_F^2 \tag{6.31}$$

where, $\Omega = \left\{(i,j)\,|\boldsymbol{\Lambda}_{ij} = 1\right\}$. The selection operator and model matrices can be written as transformation matrices. $\mathcal{A}_\Lambda$ is a matrix with number of rows equal to known elements in $\boldsymbol{M_t}$, and number of columns equal to the number of elements in $\boldsymbol{M_t}$. Each row in $\mathcal{A}_\Lambda$, indicates the presence of a measurement in $\boldsymbol{M_t}$ as a binary row vector. $\mathcal{A}_M$ is the model transformation matrix.

$$\mathcal{A}_\Lambda = \left[a_{ij}\right]_{|\Omega|\times 5mp} \tag{6.32}$$

$$\mathcal{A}_M = \begin{bmatrix} \boldsymbol{I}_{mp} & \boldsymbol{0}_{mp} & \boldsymbol{0}_{mp} & -\mathcal{D}_{\boldsymbol{r}} \\ \boldsymbol{0}_{mp} & \boldsymbol{I}_{mp} & \boldsymbol{0}_{mp} & -\mathcal{D}_{\boldsymbol{i}} \\ \boldsymbol{0}_{mp} & \boldsymbol{0}_{mp} & \boldsymbol{I}_{mp} & -\mathcal{K} \end{bmatrix} \tag{6.33}$$

The scaling factors $s_1^{(0)}$, $s_2^{(0)}$, $s_3^{(0)}$, $s_4^{(0)}$, and $s_5^{(0)}$ which form the initial scaling matrix $\boldsymbol{S}^{(0)} = \mathrm{diag}(s_1^{(0)}, s_2^{(0)}, s_3^{(0)}, s_4^{(0)}, s_5^{(0)})$ are assumed to be equal to 1. This starts the scaling with equal weights to all columns in the residuals.

The estimate of the state measurement matrix is initialized to the observed measurements ($\hat{\boldsymbol{X}}_0 = \boldsymbol{M_t}$). To increase the rate of convergence, a momentum term $\boldsymbol{P}_0 = \boldsymbol{0}_{mp\times 5}$ is introduced. the number of known measurements of each physical quantity can be written as follows,

$$n_j = \sum_{j=1}^{N_b} \mathbf{\Lambda}_{i,j} \tag{6.34}$$

The direction of descent can be obtained as $-\partial \boldsymbol{X}$, where,

$$\partial \boldsymbol{X}^{(k)} = \boldsymbol{U}\boldsymbol{V}^T \bigodot \mathbf{\Lambda} + \left[\lambda_1 \nabla_{\mathrm{vec}(X)} f_1(\hat{\boldsymbol{X}}^{(k-1)})\right]_{mp\times 5} + \left[\lambda_2 \nabla_{\mathrm{vec}(X)} f_2(\hat{\boldsymbol{X}}^{(k-1)}\right]_{mp\times 5} \tag{6.35}$$

The rate of convergence can be improved using the momentum update,

$$\boldsymbol{P}^{(k)} = \phi \boldsymbol{P}^{(k-1)} - \kappa \frac{\|\boldsymbol{X}^{(k-1)}\|_F}{\|\partial \boldsymbol{X}^{(k)}\|_F} \partial \boldsymbol{X}^{(k)} \tag{6.36}$$

The new estimate can then be written as,

$$\boldsymbol{X}^{(k)} = \boldsymbol{X}^{(k-1)} + \boldsymbol{P}^{(k)} \tag{6.37}$$

The update rule for the scaling factors is based on the assumption that each weighted column has equal contribution to the error. This enables equal sensitivity to all columns in the state measurement matrix residuals. We define a thresholding function to limit the rate of update of scaling factors.

$$f_{thres}(x, s_r^{(max)}) = \begin{cases} s_{min}, & x < s_{min} \\ s_{max}, & x > s_{max} \\ x, & otherwise \end{cases} \tag{6.38}$$

The total sum of squared errors in known quantities in each column are determined,

$$e_j = \sum_{(i,j)\in\Omega} \left(\hat{\boldsymbol{X}}^{(k)}_{(i,j)} - \boldsymbol{M}_{(i,j)} \right)^2 \tag{6.39}$$

The scaling weights are then updated using,

$$s_j^{(k)} = s_j^{(k-1)} \times f_{thres} \left(\sqrt{\frac{e_{ref}}{e_j}}, s_r^{(max)} \right) \tag{6.40}$$

These steps are iterated until convergence.

6.3 CASE STUDIES

The IEEE 33-node distribution system and a 559 node distribution system were used as the test systems, see Appendix A. State measurement matrices were generated as described in Madbhavi *et al.* (2021). It can be observed from Fig. 6.1 that the estimation errors obtained by using the state estimator (6.23) are within the predicted error bounds. The figure also shows that the estimation error in the presence of parameter uncertainty (10% absolute value) is within the predicted error bounds. Future work will incorporate modelling errors and attempt to tighten the error bounds.

The results of the state estimates using the proposed auto-scaling approach are shown in Figure 6.2. The true measurements are plotted in blue, while the estimated measurements are plotted in red. It can be observed that the auto-scaling results in good estimation performance with the matrix completion-based state estimator.

6.4 SUMMARY

Quantifying error bounds and optimizing state estimator performance are essential for real world applications. Therefore, for the first time, the error bounds for a model-based matrix completion-based distribution system state estimator was developed. The powerflow constraints were used to quantify the maximum errors in the unknown measurements.

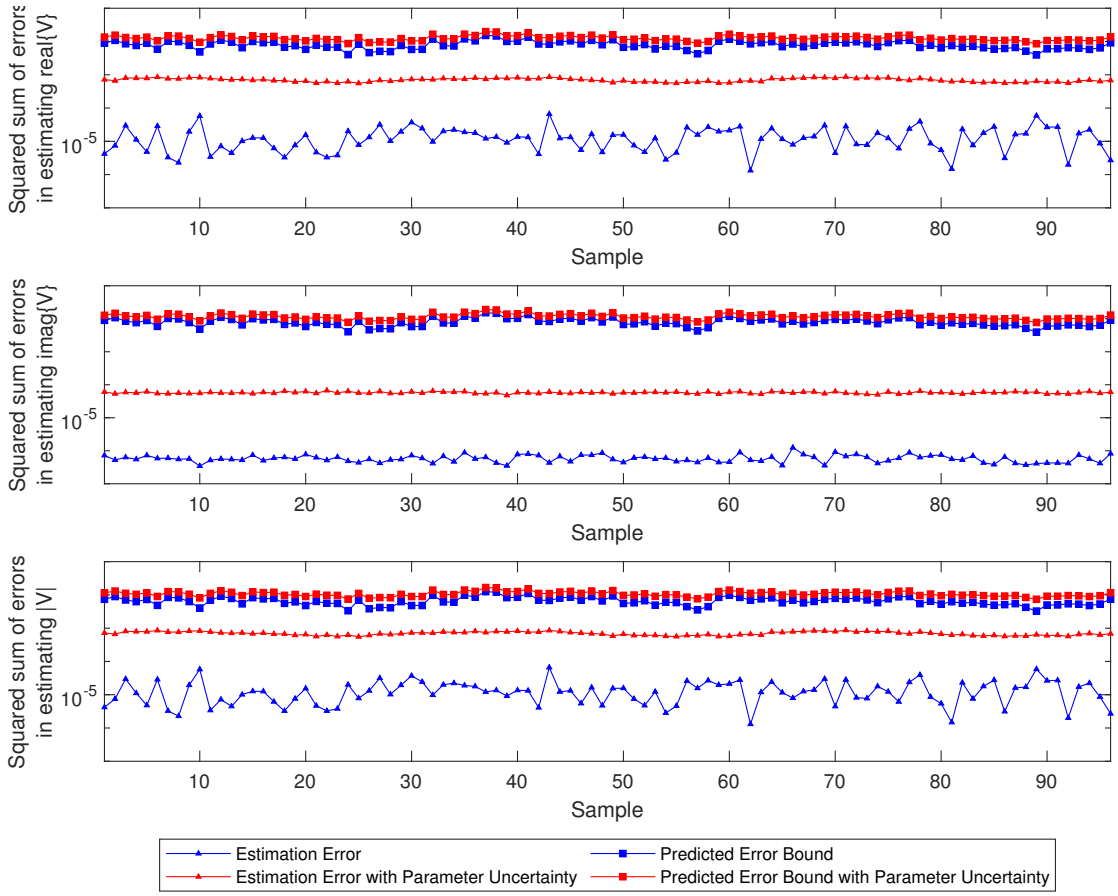
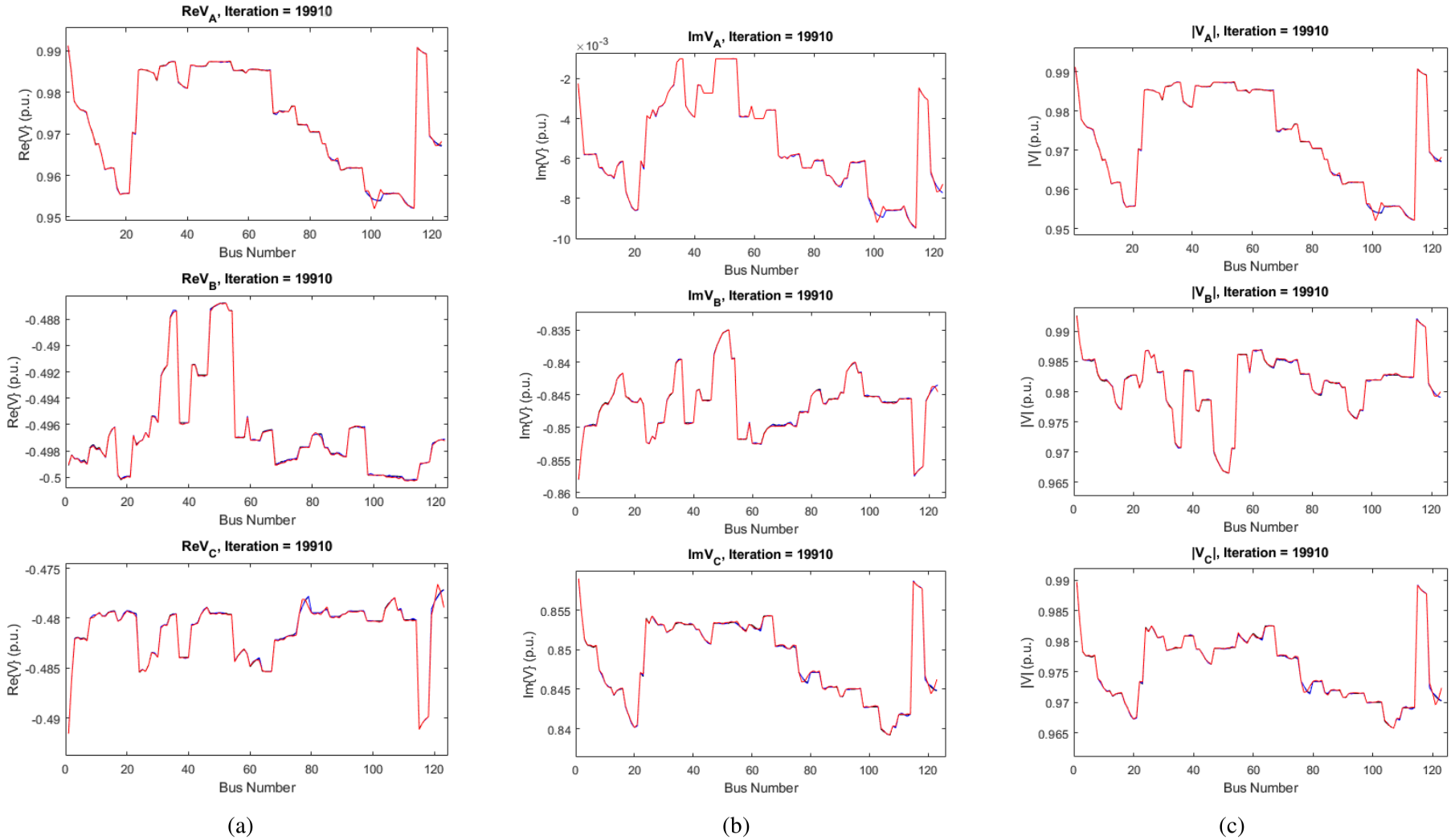


Figure 6.1: Comparison of state estimator error with predicted error bounds on the IEEE 33-node distribution system.



(a) (b) (c)

Figure 6.2: Estimation of voltage phasor on the primary and secondary nodes of 559-node distribution system while using auto-scaling (a) Real part of voltage phasor (b) Imaginary part of voltage phasor (c) Magnitude of voltage phasor.

The developed bounds also take into account parameter uncertainty. Case studies on the IEEE 33-node distribution system indicate that the estimation errors are within the estimated error bounds. Future works will work on tightening the developed bounds and exploring their applicability in other distribution system applications such as meter placement. Also, the choice of hyper-parameters is critical in obtaining better state estimation performance. Moreover, the choice of these parameters typically requires a grid search which can be computationally expensive. Further, these scaling factors could vary over time based on the values in the state measurement tensor. Therefore, an automatic scaling factor update mechanism was introduced to update the scaling factors. In the next chapter, we develop strategies for meter placement, to reduce state estimation errors.

CHAPTER 7

OPTIMIZING DSSE PERFORMANCE: METER PLACEMENT

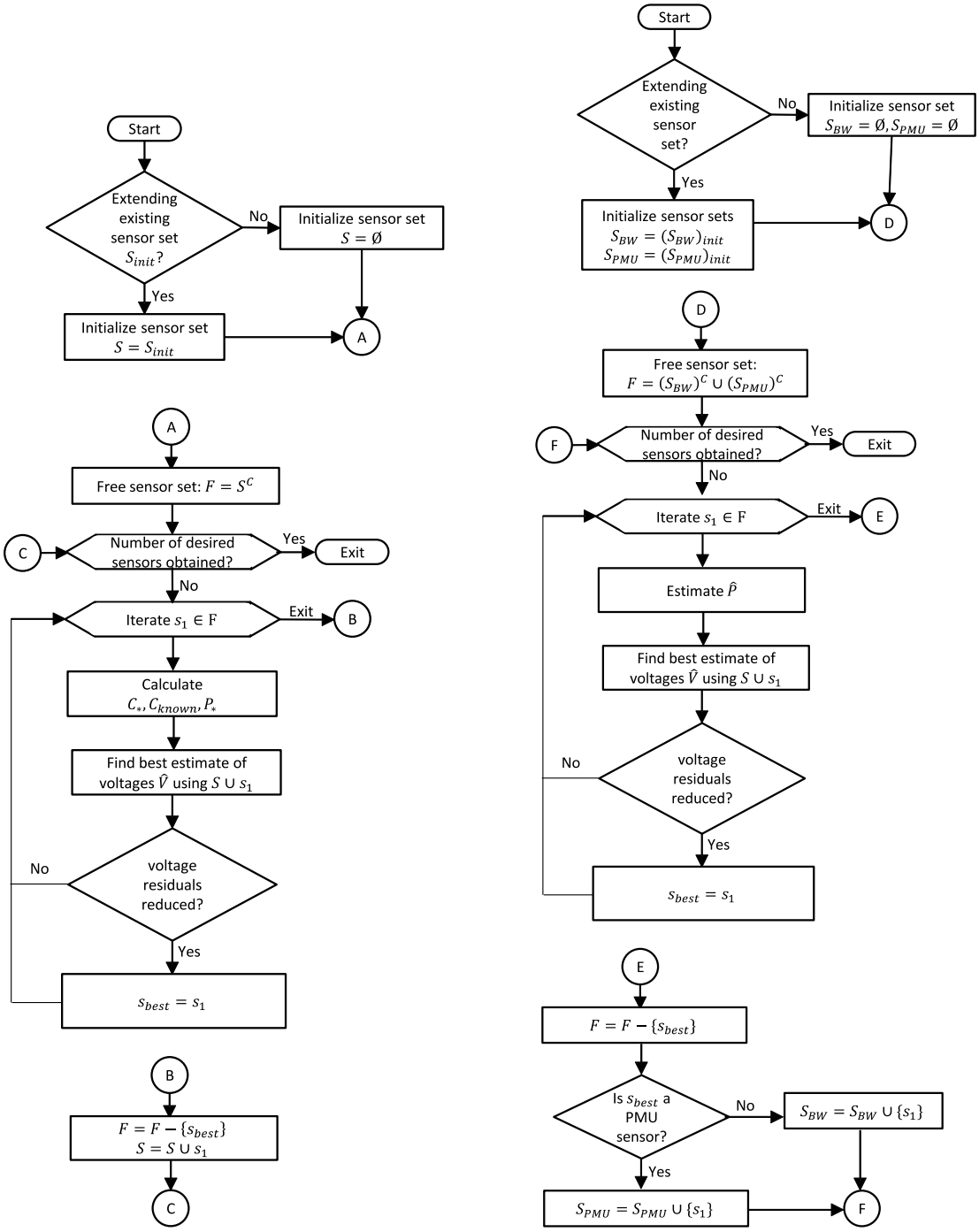
7.1 INTRODUCTION

Advanced metering infrastructure (AMI) includes integrated systems of smart meters, communication networks, and data management systems. Recent mandates by governments across the world have resulted in large scale deployments of such metering equipment. Current electric smart meter deployments in the US and Europe already exceed 100 million as per the U.S. Energy Information Administration (2021) and 200 million units as per Berg Insight (2021*b*), respectively. These deployments are expected to grow at rates exceeding 6% compounded annual growth rate (CAGR). Research studies also project smart meter deployments of over 570 million units in China, India, Japan, and South Korea between 2021 and 2025, see Berg Insight (2021*a*). Therefore, an exponential increase in smart meter deployments is expected in this decade.

State estimation is one of the use case of these grid sensors. Several sparsity-aware state estimators have been developed to utilize the measurements from these grid sensors, see Dahale *et al.* (2020). Model-based approaches incorporate constraints into traditional matrix and tensor completion-based approaches. These constraints result in certain combinations of measurements to result in the better state estimation result over others. Therefore, minimizing the state estimation errors from these estimators requires meter placement approaches.

7.2 PROPOSED APPROACHES

The parameters of a distribution system completely define its response to perturbations. Therefore, an approach is proposed that utilizes the linearized power flow equations



(a) Bus selection approach (b) Element selection approach

Figure 7.1: Flowcharts of the proposed meter placement approaches.

to place the sensors. Consider the linear approximation of the voltage phasor and magnitudes as defined in the following equations by Bernstein and Dall'Anese (2017).

$$\begin{aligned} \boldsymbol{v} - \boldsymbol{D}\boldsymbol{x} - \boldsymbol{w} &= 0 \\ |\boldsymbol{v}| - \boldsymbol{K}\boldsymbol{x} - |\boldsymbol{w}| &= 0 \end{aligned} \tag{7.41}$$

In the above equations, the model is linearized about the zero power injection condition. Therefore, $\boldsymbol{w} \in \mathbb{C}^{N_p(m-1)\times 1}$ (N_p is the number of phases) contains the offsets at the point of linearization, and coefficient matrices $\boldsymbol{D}$ and $\boldsymbol{K}$ quantify the impact of power injection at all nodes on the bus voltage phasor and magnitude vectors. The column pictures of equations (7.41) are essential for the sensor placement problem. Here, the bus voltage phasor vector $\boldsymbol{v}$ is the sum of the offset vector $\boldsymbol{w}$ and the differential changes in $\boldsymbol{v}$ due to the individual power injections, and likewise for $|\boldsymbol{v}|$. The equations in (7.41) can be combined to form (7.42).

$$\begin{bmatrix} Re\{\boldsymbol{v}\} \\ Im\{\boldsymbol{v}\} \\ |\boldsymbol{v}| \end{bmatrix} - \boldsymbol{C}\boldsymbol{x} - \begin{bmatrix} Re\{\boldsymbol{w}\} \\ Im\{\boldsymbol{w}\} \\ |\boldsymbol{w}| \end{bmatrix} = 0, \quad \boldsymbol{C} = \begin{bmatrix} Re\{\boldsymbol{D}\} \\ Im\{\boldsymbol{D}\} \\ \boldsymbol{K} \end{bmatrix} \tag{7.42}$$

In the proposed approaches, a set of reference cases similar to dynamic programming are chosen to evaluate each candidate location's effectiveness. The measurements from sensors at N_t time instants are used as reference cases to create a state measurement tensor $\boldsymbol{\mathcal{T}} \in \mathbb{R}^{3m\times 5\times N_t}$. A consolidated voltage matrix $\boldsymbol{\mathcal{V}}$ containing voltage measurements and a power injection measurement matrix $\boldsymbol{\mathcal{P}}$ are generated by concatenating the sub-matrices of the lateral slices of the tensor $\boldsymbol{\mathcal{T}}$.

$$\boldsymbol{\mathcal{V}} = \begin{bmatrix} \boldsymbol{\mathcal{T}}[N_p + 1 : m \times N_p, 1, :] \\ \boldsymbol{\mathcal{T}}[N_p + 1 : m \times N_p, 2, :] \\ \boldsymbol{\mathcal{T}}[N_p + 1 : m \times N_p, 3, :] \end{bmatrix} \tag{7.43}$$

$$\boldsymbol{\mathcal{P}} = \begin{bmatrix} \boldsymbol{\mathcal{T}}[N_p + 1 : m \times N_p, 4, :] \\ \boldsymbol{\mathcal{T}}[N_p + 1 : m \times N_p, 5, :] \end{bmatrix} \tag{7.44}$$

Therefore, the consolidated voltage matrix during zero power injection conditions is given by (7.45). While, the impact of power injections in altering the bus voltages can be quantified by (7.46).

$$\boldsymbol{\mathcal{V}_0} = \left[Re\{\boldsymbol{w}\}^T Im\{\boldsymbol{w}\}^T |\boldsymbol{w}|^T \right]^T \times ones(1, N_t) \tag{7.45}$$

$$\Delta\boldsymbol{\mathcal{V}} = \boldsymbol{Cx} = \boldsymbol{\mathcal{V}} - \boldsymbol{\mathcal{V}_0} \tag{7.46}$$

Using the above matrix definitions and equations, the sensor placement approach has been presented in the following subsections in two forms: a bus selection problem and an element selection problem (elements corresponding to measurements obtained from a sensor).

7.2.1 Bus selection

In this type of sensor placement, it is assumed that all measurements ($Re\{\boldsymbol{V_i}\}$, $Im\{\boldsymbol{V_i}\}$, $|\boldsymbol{V_i}|$, P_i^{in} and Q_i^{in}) of the selected sensor set are known. Bus 1 is assumed to be the reference bus. This approach aims to identify a set of buses whose measurements minimize the voltage magnitude and angle estimation errors. Prior efforts on matrix and tensor completion based state estimators have achieved voltage magnitude estimation as low as 0.3% even in the presence of 1% sensor noise at 30% data availability

in Madbhavi *et al.* (2021). Moreover, it was reported that voltage angle estimation accuracy deteriorated much faster than voltage magnitude estimation accuracy at low data availabilities. Therefore, it is essential to prioritize minimizing the voltage angle estimation error. Consider the function $g(.)$ which operates on the sensor set S' and provides the set of column indices in $\boldsymbol{\mathcal{V}}$ corresponding to these sensor locations.

$$g(S') = \bigcup_{\forall s \in S'} \bigcup_{\forall j \in 1,2,...,N_p} \{N_p(s-2)+j, Np(m-1)+N_p(s-2)+j\} \quad (7.47)$$

Now, the candidate sensor set S is initialized as a null set, while the free sensor set is initialized as $F = \{2, 3, 4, ..., m\}$. The algorithm then proceeds with the first sensor s_1 in the free sensor set F. The current sensor set of column indices is $g(S \cup s_1)$. Thus, the best estimate of $\boldsymbol{\mathcal{V}}$ using the current sensor set to minimize the voltage residual is,

$$\hat{\boldsymbol{\mathcal{V}}} = \begin{cases} \boldsymbol{\mathcal{V}_0} + \boldsymbol{C_*}(\boldsymbol{C_*}^T\boldsymbol{C_*})^{-1}\boldsymbol{C_*}^T\Delta\boldsymbol{\mathcal{V}}, & S = \emptyset \\ \boldsymbol{\mathcal{V}_0} + \boldsymbol{C_{known}}\boldsymbol{\mathcal{P}_*} + \boldsymbol{C_*}(\boldsymbol{C_*}^T\boldsymbol{C_*})^{-1}\boldsymbol{C_*}^T(\Delta\boldsymbol{\mathcal{V}} - \boldsymbol{C_{known}}\boldsymbol{\mathcal{P}_*}), & S \neq \emptyset \end{cases} \quad (7.48)$$

where, (7.49)

$$\boldsymbol{C_*} = \boldsymbol{C}[:, g(\{s_1\})] \quad (7.50)$$

$$\boldsymbol{C_{known}} = \boldsymbol{C}[:, g(S)] \quad (7.51)$$

$$\boldsymbol{\mathcal{P}_*} = \boldsymbol{\mathcal{P}}[g(S), :] \quad (7.52)$$

The first term $\boldsymbol{\mathcal{V}_0}$ in $\hat{\boldsymbol{\mathcal{V}}}$ is the offset term. The $\boldsymbol{C_{known}}\boldsymbol{\mathcal{P}_*}$ term in $\hat{\boldsymbol{\mathcal{V}}}$ quantifies the contribution of measurements from sensors in S towards $\Delta\boldsymbol{\mathcal{V}}$. While the last term accounts for the contribution of sensor s_1 in estimating the residual from $\Delta\boldsymbol{\mathcal{V}}$ after removing the contribution of sensors in S.

The mean absolute error (MAE) in angle estimation is then calculated using the estimated phasor components in $\hat{\boldsymbol{\mathcal{V}}}$ with respect to the true voltage phasor components in $\boldsymbol{\mathcal{V}}$. This step is then repeated with all the remaining sensors in the free sensor set F. The sensor

s_{best} which results in the lowest voltage angle MAE is added to the candidate set S while it is removed from the free sensor set F. This process is iterated until the desired number of buses are selected in the candidate sensor set S. The flowchart of the bus selection approach is depicted in Fig. 7.1a.

7.2.2 Element selection

In practice, all measurements from a bus may not be available since the placed sensors may report only some quantities. Moreover, it may be required to extend the existing heterogeneous set of sensors to improve state estimation accuracy. Thus, in this approach, two sensors are considered: distribution phasor measurement units (D-PMU) and SCADA sensors (e.g., Bellwether meters). These meters provide ($Re\{\boldsymbol{V_i}\}$, $Im\{\boldsymbol{V_i}\}$, $|\boldsymbol{V_i}|$) and ($|\boldsymbol{V_i}|$, P_i^{in} and Q_i^{in}) measurements, respectively. The algorithm for placing bellwether meters to improve state estimation accuracy is explained below.

Firstly, two functions are defined to generate matrix indices of known measurements in the state measurement tensors. Given a set of D-PMUs at buses S'_{PMU}, the set of row indices corresponding to these locations is given by $h_1(.)$. Similarly, given a set of bellwether meters at buses S'_{BW}, the set of column indices corresponding to these locations is given by $h_2(.)$. Here, $phase(S)$ is 1, 2 and 3 respectively for phases A, B and C.

$$h_1(S'_{PMU}) = \bigcup_{\forall s \in S'_{PMU}} \bigcup_{\forall j \in 1,2,...,N_p} \{N_p(s-2)+j, Np(m-1)+N_p(s-2)+j, 2Np(m-1)+N_p(s-2)+j\} \quad (7.53)$$

$$h_2(S'_{BW}) = \bigcup_{\forall s \in S'_{BW}} \{N_p(busNumber(s)-2)+phase(s), N_p(busNumber(s)-2)+phase(s)+N_p(m-1)\} \quad (7.54)$$

The sets S_{PMU} and S_{BW} contain the existing D-PMU and bellwether meters. The candidate sensor set S is initialized as the null set, while F is the free sensor set containing all allowed locations for bellwether meters. The algorithm proceeds with the first sensor s_1 in the free sensor set F. The current set of sensor row indices is $h_1(S_{PMU})$, while the set of sensor column indices is $h_2(S_{BW} \cup \{s_1\})$. Thus, the best estimate of the power injections to minimize the voltage residuals is given by,

$$\hat{\boldsymbol{P}} = \underset{\boldsymbol{P}}{\operatorname{argmin}} ||\Delta\boldsymbol{\mathcal{V}} - \boldsymbol{C}[:, h_2(S_{BW})]\boldsymbol{P}[h_2(S_{BW}), :] - \boldsymbol{C}[:, h_2(\{s_1\})]\boldsymbol{P}[h_2(\{s_1\}), :]||_F^2$$
$$s.t.\ |\boldsymbol{C}[h_1(S_{PMU}), :]\boldsymbol{P} - \Delta\boldsymbol{\mathcal{V}}[h_1(S_{PMU}), :]|_\infty <= \delta$$

If S_{BW} or S_{PMU} is a null set, then the corresponding term or constraint respectively are removed. Thus, the best estimate of $\boldsymbol{\mathcal{V}}$ using the current sensor set is,

$$\hat{\boldsymbol{\mathcal{V}}} = \boldsymbol{\mathcal{V}_0} + \boldsymbol{C}\hat{\boldsymbol{P}} + \boldsymbol{C}[:, h_2(S_{BW})]\boldsymbol{P}[h_2(S_{BW}), :]$$

The mean absolute error (MAE) in angle estimation is then calculated using the estimated phasor components in $\hat{\boldsymbol{\mathcal{V}}}$ with respect to the true voltage phasor components in $\boldsymbol{\mathcal{V}}$. This step is then repeated with all the remaining sensors in the free sensor set F. The sensor s_{best} which results in the lowest voltage angle MAE is added to the candidate set S_{BW}, while it is removed from the free sensor set F. This process is iterated until the desired number of buses are selected in the candidate sensor set S_{BW}. While the above process has been shown with bellwether meter placements, combinations of D-PMU and bellwether meters can be placed by using the free sensor set containing both remaining D-PMUs and bellwether meters. The flowchart of the element selection approach is depicted in Fig. 7.1b.

These approaches are collectively referred to as the LinPF approaches in the subsequent

sections. A comparative study on the performance of the proposed and existing approaches is provided in the next section.

7.3 CASE STUDIES

The sensor placement approaches are evaluated on the IEEE 33-node distribution system and a three-phase unbalanced 559-node test system (modelled on the IEEE 37-node distribution system), see Appendix A. The state estimation performance of the proposed approach and existing approaches such as generalized linear integer programming, dynamic programming and active matrix completion approaches is evaluated in the following subsections. The impact of PV penetration on state estimation has been studied. The mean absolute percentage error (MAPE) and mean absolute error (MAE) metrics have been utilized to quantify the voltage magnitude and angle estimation errors, respectively [1].

7.3.1 IEEE 33-node distribution system

The state measurement matrices for the IEEE 33-node distribution system were generated using the results of load flow analysis. Table 7.1 summarizes the results of matrix completion on state measurement matrices of IEEE 33-node distribution system using different meter placement strategies. The sensor placement strategies have been categorized into three types. Type I corresponds to sensor placement where all measurements ($Re\{\boldsymbol{V_i}\}$, $Im\{\boldsymbol{V_i}\}$, P_i^{in}, Q_i^{in}) are known. Type II corresponds to bellwether meter placement where only ($|V_i|$, P_i^{in}, Q_i^{in}) are known. Finally, type III placement extends existing D-PMU sensor sets (Ghadikolaee *et al.* (2020); Ghasemi Damavandi *et al.* (2015); Liu *et al.* (2012)) with bellwether meters to improve state estimation performance. Also, 1% sensor noise was considered while creating the state

[1] Results in this section have been published in:

- **R. Madbhavi**, B. Natarajan and B. Srinivasan, "Meter Placement Approaches for Matrix Completion-based Distribution System State Estimator", Energy Power Systems Research, vol. 213, Dec. 2022, doi: 10.1016/j.epsr.2022.108687.

measurement matrices. The choice of noise level is consistent with case studies in existing works on matrix and tensor completion-based state estimators: Liu *et al.* (2019*a*); Madbhavi *et al.* (2020, 2021); Dahale *et al.* (2020). It can be observed from Table 7.1 that all approaches provide good voltage magnitude estimation performance.

The generalized linear integer programming (GLIP) approach is a topological observability based approach. This approach determines the smallest sensor set such that every node in the network is either a candidate node or a node adjacent to a candidate node. This approach has been used in applications such as optimal PMU placement in Gou (2008). GLIP proposes 11 sensors $(2, 5, 8, 11, 14, 17, 21, 24, 27, 29, 32)$ for obtaining full topological observability. With all the 11 sensors placed, the matrix completion state estimator provided voltage magnitude MAPE of 0.2% and voltage angle MAE of 0.003 degrees. Unlike the other approaches, GLIP cannot suggest sensor locations for a given sensor set size. Therefore, it has not been used in subsequent sensor placement evaluations.

The second approach utilizes dynamic programming to obtain the sensor locations. Dynamic programming is an iterative algorithm that progressively adds sensors to the selected sensor set to maximize/minimize a chosen overall quality metric (depending on its nature) in Muscas *et al.* (2009). The reference cases used in this study were created by introducing unit demand to the components of power injections and calculating the system states. Voltage angle MAE is used as the quality metric to evaluate the performance of different sensor placements in state estimation. Dynamic programming provides the best voltage angle MAPEs across different set sizes. Dynamic programming achieves this by exhaustively evaluating all sensor locations using (7.55) to iteratively add the next best sensor to the candidate sensor set.

Table 7.1: State estimation performance of matrix completion based state estimator using different meter placement approaches on the IEEE 33-node distribution system.

Estimation Performance Metric				VM MAPE (%)				VA MAE (degrees)				Time
Number of Sensors Placed				1	2	3	4	1	2	3	4	(s)
Sensor Placement												
Type	Measurements	Method	Locations (i)									
I	$Re\{\boldsymbol{V_i}\}$, $Im\{\boldsymbol{V_i}\}$, P_i^{in}, Q_i^{in}	Dynamic Prog.	14, 8, 22, 25	0.50	0.37	0.39	0.28	0.022	0.016	0.015	0.011	466.1
		QBC	14, 18, 17, 33	0.41	0.39	0.36	0.33	0.024	0.041	0.029	0.023	159.0
		QBC Stability	14, 33, 22, 13	0.45	0.32	0.35	0.29	0.024	0.019	0.017	0.015	352.5
		LinPF	8, 24, 18, 32	0.30	0.22	0.23	0.19	0.083	0.083	0.026	0.022	0.1
II	$\lvert\boldsymbol{V_i}\rvert$, P_i^{in}, Q_i^{in}	Dynamic Prog.	9, 28, 23, 13	0.68	0.82	0.59	0.53	0.027	0.027	0.029	0.024	345.2
		QBC	23, 18, 17, 16	0.62	0.49	0.44	0.42	0.268	0.260	0.262	0.263	182.4
		QBC Stability	14, 18, 17, 33	0.36	0.37	0.35	0.33	0.426	0.426	0.428	0.427	335.3
		LinPF	8, 24, 18, 32	0.32	0.32	0.22	0.16	0.131	0.150	0.141	0.170	16.5
III.a	(PMUs 14, 15, 30)Ghadikolaee *et al.* (2020) + ($\lvert\boldsymbol{V_i}\rvert$, P_i^{in}, Q_i^{in})	Dynamic Prog.	30, 22, 21, 20	0.12	0.21	0.26	0.29	0.020	0.015	0.013	0.012	585.3
		QBC	18, 33, 17, 32	0.33	0.36	0.34	0.24	0.037	0.042	0.028	0.048	32.0
		QBC Stability	18, 17, 33, 16	0.26	0.21	0.22	0.21	0.033	0.030	0.030	0.029	67.8
		LinPF	8, 17, 32, 24	0.28	0.21	0.19	0.16	0.027	0.028	0.032	0.024	72.0
III.b	(PMUs 17, 18, 33)Ghasemi Damavandi *et al.* (2015) + ($\lvert\boldsymbol{V_i}\rvert$, P_i^{in}, Q_i^{in})	Dynamic Prog.	30, 15, 29, 13	0.14	0.11	0.11	0.10	0.018	0.018	0.018	0.018	490.6
		QBC	18, 17, 16, 33	0.34	0.32	0.32	0.41	0.039	0.044	0.043	0.027	33.0
		QBC Stability	18, 17, 33, 16	0.25	0.20	0.24	0.23	0.025	0.022	0.021	0.020	70.9
		LinPF	9, 31, 16, 24	0.35	0.29	0.25	0.14	0.025	0.033	0.033	0.036	71.7
III.c	(PMUs 13, 18, 33)Liu *et al.* (2012) + ($\lvert\boldsymbol{V_i}\rvert$, P_i^{in}, Q_i^{in})	Dynamic Prog.	25, 24, 13, 17	0.18	0.18	0.17	0.17	0.014	0.010	0.010	0.010	553.4
		QBC	18, 17, 16, 33	0.32	0.30	0.30	0.38	0.041	0.045	0.045	0.028	32.7
		QBC Stability	18, 17, 33, 16	0.25	0.20	0.24	0.23	0.019	0.021	0.020	0.020	60.5
		LinPF	8, 17, 32, 24	0.26	0.20	0.19	0.16	0.018	0.019	0.026	0.017	70.9

$$\begin{aligned} \min_{\boldsymbol{X}} \quad & \|\boldsymbol{X}.diag(\boldsymbol{s})\|_* + \lambda\|((\boldsymbol{X}-\boldsymbol{M}) \bigodot \boldsymbol{\Lambda}).diag(\boldsymbol{f})\|_F^2 \\ \text{s.t.} \quad & \|\boldsymbol{v}-\boldsymbol{D}\boldsymbol{x}-\boldsymbol{w}\|_\infty \leq \gamma \\ & \|\,|\boldsymbol{v}|-\boldsymbol{K}\boldsymbol{x}-|\boldsymbol{w}|\,\|_\infty \leq \delta \end{aligned} \tag{7.55}$$

The meter placement problem in distribution systems can be viewed as a matrix sampling problem of the state measurement matrices. Therefore, the meter placement problem can also be solved using active-matrix completion. In active matrix completion, a measure of prediction uncertainty is utilized to compute the uncertainty in predicting matrix entries. The entries with the highest uncertainties are then queried to minimize matrix completion errors. Among these techniques are query by committee (QBC) and query by committee stability , see Chakraborty *et al.* (2013). The query by committee framework utilizes a committee of estimators to identify matrix indices that exhibit high prediction variance. These element indices thus form the selected senor set for the meter placement problem. On the other hand, query by committee stability utilizes a single estimator but generates multiple estimators using different estimator parameters. The selected sensor set is thus formed in the same way as in QBC. Three variants of the matrix completion state estimator ((7.55) and its block completion counterparts) were used in the Query by committee (QBC) approach. In comparison, eight variants of the state estimator (7.55) were utilized in the Query by committee stability approach. However, the approaches could not perform better than dynamic programming. Moreover, these approaches require multiple evaluations of each state measurement matrix, thereby increasing the computation time.

Finally, the LinPF approach provides the best voltage magnitude estimation. The reduction in voltage residuals with the increasing size of the sensor set is depicted in Fig. 7.2. The decrease in the magnitude of voltage residuals with increasing sensor set size can be seen in Fig. 7.2 (a), (b), and (c). Fig. 7.2 (d) shows the voltage residuals in a non-optimal location for the third sensor location (bus 18 vs. bus 12). However, voltage

angle estimation using LinPF is lacking when compared with dynamic programming. It should also be noted that because of the small size of the IEEE 33 node network, the results vary significantly even by small changes in sensor locations. Therefore, a 559-node distribution system is utilized in the following sub-section.

The bellwether meter placement by different placement strategies is then evaluated in the absence of all other sensors. These results are tabulated as Type II in Table 7.1. Dynamic programming still provides the best voltage angle estimation. However, the LinPF approach performs significantly better than QBC and QBC Stability.

D-PMU placements from the literature were used to evaluate the performance of the sensor placement approaches in extending existing sensor sets. Dynamic programming continues to outperform all other placement methods. However, it is important to note that this approach has very high computational complexity. The computational complexity of the matrix completion based state estimator used in this study is $O(m^3 N_t N_i)$, where N_i is the number of iterations of the algorithm. Therefore, the computational complexity of the dynamic programming approach is $O(m^4 k N_t N_i)$ where k is the number of sensors to be placed. However, QBC and QBC stability have computational complexities $O(m^4 N_e N_t N_i)$ where N_e is the number of estimators in the committee. Since $N_e < k$ for most systems and $N_e << k$ for large distribution systems (such as the 559-node network), QBC and QBC-Stability are significantly faster than the dynamic programming approach. The computational complexity of the bus selection approach of LinPF is $O(m^3 k.max(k, N_t))$.

The algorithms were implemented in MATLAB R2020b on a laptop with an Intel Core i7-10750H processor and 16GB RAM. The computation times for sensor placement are tabulated in Table 7.1. It can be observed that dynamic programming is the slowest among the evaluated methods, as predicted by the computational complexity analysis. The LinPF approach is the fastest among the assessed methods, since it does not evaluate

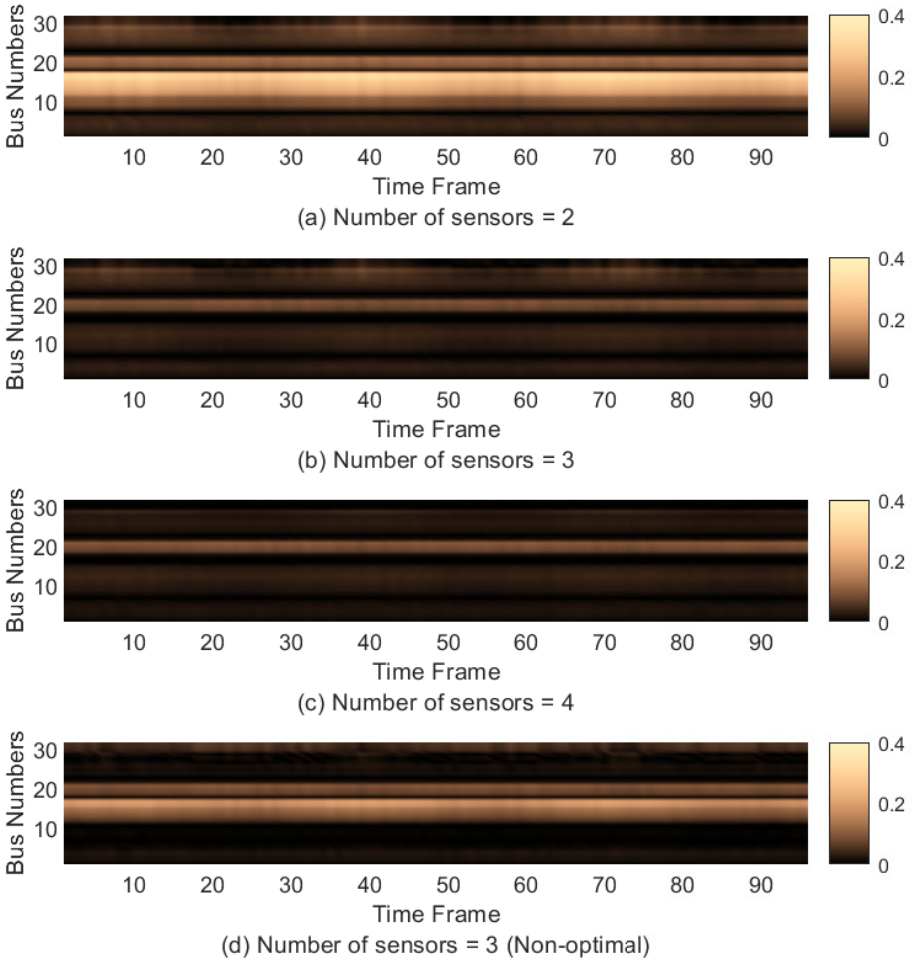


Figure 7.2: Reduction in voltage residuals with increasing sensor set size.

all candidate sensor placements using the matrix completion estimator. Moreover, type I sensor placement for linPF approach is significantly faster than other methods, particularly for smaller networks because the best estimate of $\hat{\boldsymbol{V}}$ has a closed-form solution. However, its computational time increases with network size due to the matrix inverse operation in its solution.

7.3.2 559-node test system

The sensor placement approach is also evaluated on a 3-phase unbalanced 559-node test system (modelled on the IEEE-37 node distribution system), see Appendix A. The following subsections investigate the performance of the sensor placement approaches on state estimation performance. The sensor placement strategies have been categorized into three types, similar to the case study on the IEEE 33-node distribution system. The state estimation results have been tabulated in Table 7.2, while the sensor locations are provided in Table 7.3.

The sensor placement types I and II have been evaluated with PV generation at all nodes and the results are summarized as I.b and II.b in Table 7.2. In practice, only some buses have PV generation; therefore, 50% of the nodes were selected at random as nodes with PV generation. The state estimation performance of these datasets been tabulated as I.c and II.c in Table 7.2.

The results of state estimation using datasets with PV data (Types I.b, I.c, II.b and II.c in Table 7.2) indicate that introduction of PV generation does not negatively impact the state estimation performance. Also, the results highlight the need for utilizing both D-PMUs and bellwether meters. When only one type of sensor is considered in the sensor placement process, some columns of the state measurements are entirely unknown and severely deteriorates the state estimation performance, as seen in sensor placement type II. Another aspect of utilizing active learning approaches for completing state measurement tensors is the need for better approaches for combining the results from

Table 7.2: State estimation performance of matrix completion based state estimator using different meter placement approaches on the 559-node distribution system.

Estimation Performance Metric			VM MAPE (%)				VA MAE (degrees)				Time
Percentage of total sensor locations with sensors			2.5	5	7.5	10	2.5	5	7.5	10	
Sensor Placement											
Type	Measurements	Method									
I.a	$Re\{\boldsymbol{V_i}\}$, $Im\{\boldsymbol{V_i}\}$, P_i^{in}, Q_i^{in}	Dynamic Prog.	0.008	0.005	0.004	0.003	0.008	0.004	0.002	0.003	11d 19h 32m 55s
		QBC	0.013	0.012	0.009	0.004	0.127	0.121	0.118	0.055	17m 16s
		QBC Stability	0.012	0.006	0.006	0.004	0.013	0.012	0.018	0.007	1h 24m 49s
		LinPF	0.007	0.004	0.003	0.003	0.005	0.001	0.001	0.001	43m 52s
I.b	$Re\{\boldsymbol{V_i}\}$, $Im\{\boldsymbol{V_i}\}$, P_i^{in}, Q_i^{in}	Dynamic Prog.	0.007	0.004	0.003	0.003	0.008	0.004	0.002	0.002	Same placement as I.a
		QBC	0.012	0.011	0.008	0.004	0.104	0.100	0.099	0.047	
		QBC Stability	0.011	0.006	0.006	0.004	0.072	0.022	0.007	0.006	
		LinPF	0.006	0.003	0.003	0.003	0.004	0.001	0.001	0.002	
I.c	$Re\{\boldsymbol{V_i}\}$, $Im\{\boldsymbol{V_i}\}$, P_i^{in}, Q_i^{in}	Dynamic Prog.	0.007	0.005	0.003	0.003	0.008	0.004	0.003	0.002	Same placement as I.a
		QBC	0.012	0.012	0.008	0.004	0.115	0.111	0.109	0.051	
		QBC Stability	0.011	0.006	0.006	0.004	0.019	0.019	0.019	0.006	
		LinPF	0.006	0.004	0.003	0.003	0.005	0.001	0.001	0.001	
II.a	$\lvert\boldsymbol{V_i}\rvert$, P_i^{in}, Q_i^{in}	Dynamic Prog.	0.252	0.246	0.252	0.242	0.156	0.150	0.137	0.134	9d 16h 00m 34s
		QBC	0.491	0.481	0.462	0.444	0.183	0.182	0.179	0.172	14m 22s
		QBC Stability	0.523	0.462	0.460	0.211	0.192	0.173	0.170	0.154	1h 10m 37s
		LinPF	0.054	0.025	0.021	0.022	0.172	0.169	0.167	0.165	1h 19m 02s
II.b	$\lvert\boldsymbol{V_i}\rvert$, P_i^{in}, Q_i^{in}	Dynamic Prog.	0.238	0.232	0.238	0.229	0.134	0.128	0.116	0.112	Same placement as II.a
		QBC	0.459	0.450	0.433	0.416	0.165	0.164	0.160	0.153	
		QBC Stability	0.491	0.433	0.431	0.198	0.175	0.155	0.151	0.124	
		LinPF	0.050	0.023	0.020	0.021	0.140	0.138	0.136	0.134	
II.c	$\lvert\boldsymbol{V_i}\rvert$, P_i^{in}, Q_i^{in}	Dynamic Prog.	0.244	0.239	0.244	0.235	0.143	0.138	0.134	0.131	Same placement as II.a
		QBC	0.474	0.465	0.447	0.429	0.173	0.172	0.169	0.161	
		QBC Stability	0.506	0.446	0.444	0.204	0.183	0.163	0.159	0.138	
		LinPF	0.052	0.024	0.021	0.022	0.156	0.153	0.152	0.149	

Table 7.3: Meter placement using different meter placement approaches for sensor types I and II on the 559-node distribution system.

Type	Method	Meter Locations (i)
I	Dynamic Prog.	265, 89, 365, 3, 19, 38, 338, 8, 516, 559, 330, 119, 69, 520, 72, 373, 168, 558, 253, 524, 459, 164, 309, 369, 479, 443, 286, 31, 430, 379, 556, 11, 177, 465, 527, 493, 294, 99, 533, 55, 395, 328, 39, 391, 133, 142, 233, 519, 538, 539, 257, 546, 237, 166, 495, 521
	QBC	11, 491, 507, 492, 486, 503, 474, 449, 395, 556, 352, 127, 498, 508, 315, 550, 311, 452, 433, 555, 542, 422, 552, 420, 410, 545, 537, 546, 540, 538, 302, 196, 353, 430, 460, 354, 415, 428, 426, 435, 431, 424, 417, 406, 533, 150, 28, 116, 6, 281, 252, 73, 516, 258, 132, 31
	QBC Stability	11, 452, 311, 546, 454, 361, 474, 431, 433, 137, 48, 410, 486, 301, 258, 204, 58, 69, 255, 550, 70, 84, 492, 127, 536, 537, 303, 538, 352, 337, 338, 411, 339, 450, 353, 472, 487, 268, 276, 88, 49, 60, 86, 114, 187, 205, 188, 270, 207, 256, 528, 547, 139, 429, 548, 526
	LinPF	11, 99, 19, 2, 15, 24, 407, 236, 533, 20, 296, 3, 43, 74, 202, 348, 274, 479, 508, 370, 195, 64, 141, 40, 102, 165, 394, 76, 442, 334, 6, 5, 298, 18, 8, 363, 464, 322, 190, 541, 551, 511, 521, 206, 39, 552, 42, 227, 36, 201, 497, 148, 270, 473, 234, 436
II	Dynamic Prog.	559C, 511C, 122B, 253A, 43A, 113B, 105B, 132B, 330C, 339A, 131B, 353A, 355A, 126B, 350A, 349A, 194C, 181C, 192C, 342A, 348A, 558C, 196C, 529C, 528C, 527C, 556C, 524C, 522C, 343A, 197C, 344A, 198C, 523C, 340A, 554C, 553C, 240A, 246A, 552C, 536C, 546C, 548C, 514A
	QBC	355A, 559C, 558C, 556C, 553C, 544C, 463C, 537C, 328C, 324C, 365C, 353A, 331C, 329C, 348A, 468C, 293C, 288C, 282C, 289C, 287C, 343A, 448C, 283C, 284C, 485A, 423A, 417A, 497A, 438A, 502A, 325C, 326C, 295C, 292C, 543C, 542C, 447C, 445C, 279C, 278C, 429A, 258C, 253C
	QBC Stability	355A, 354A, 43A, 353A, 352A, 348A, 359C, 360C, 362C, 233A, 234A, 235A, 251C, 361C, 448C, 232A, 445C, 447C, 446C, 455C, 453C, 228A, 457C, 252C, 253C, 254C, 469C, 264C, 271C, 256C, 257C, 263C, 266C, 269C, 265C, 268C, 270C, 259C, 456C, 258C, 458C, 400B, 475C, 474C
	LinPF	355A, 511C, 511B, 295C, 547B, 130B, 516A, 156C, 323C, 483A, 536A, 244A, 426A, 58A, 107C, 113B, 337A, 311B, 536C, 183C, 210B, 97C, 66B, 541B, 270C, 137B, 43A, 516B, 521C, 79C, 279C, 538A, 234A, 542B, 556C, 436A, 511A, 191C, 543B, 105B, 544B, 521B, 516C, 121B

the committee state estimators. Unlike other applications, simple averaging across all committee estimator results on state measurement matrices does not yield good results. This effect is amplified in the presence of significant variance in the results of different committee state estimators. Finally, the LinPF approach outperforms both QBC and QBC stability in all types of sensor placements and number of sensors considered in this case study.

Active learning methods QBC and QBC Stability rely on a sizeable pool of estimators to quantify the uncertainties in state estimates. In most cases, the size of the estimator pool is limited to the number of available state estimators (QBC) or the number of variants that can be developed considering the model parameters (QBC Stability). Therefore, it may not be possible to generate a sufficiently large pool of state estimators for QBC and QBC stability to outperform the proposed approach. In the IEEE 33-node distribution system, the proposed approach has computation time comparable to QBC and QBC Stability. QBC stability always outperformed QBC in this study because QBC stability had a larger pool of estimators than QBC. As a result, it can be seen from the 559-node distribution system that QBC stability performs better than QBC in most cases. However, LinPF still outperforms both QBC and QBC stability. Improving the performance of active learning methods would require more state estimators to be added to the pool of estimators. Therefore, there exists a trade-off between the quality of sensor placements and the computation time, which depends on the number of constituent estimators.

Dynamic programming requires a set of reference cases to iteratively place sensors by evaluating their performance with the state estimator. Therefore, the dynamic programming approach requires a sizeable sample of reference cases to cover a sufficiently large number of load configurations. However, this increases the computation time as each of these cases has to be evaluated using the state estimator. However, the LinPF approach works directly with the state estimator's constraints. Therefore it does not require an evaluation of the state estimator to place the sensors. Thus, the proposed

approach has significantly lower computational time than dynamic programming.

7.4 SUMMARY

This chapter proposes a meter placement approach that leverages the linearized power flow equations to place sensors for improving the state estimation performance of a matrix completion-based state estimator. Sparsity-aware state estimators do not address the sensor placements that minimize their estimation errors. Therefore, this work bridges this gap by iteratively placing sensors with the highest contribution towards minimizing the voltage residual in the selected reference cases. The proposed approach avoids the limitations of existing approaches, such as scalability issues of dynamic programming and the need for multiple matrix completion evaluations of QBC and QBC stability. The proposed approaches eliminate the need for repeated application of the matrix completion-based state estimator, which contributes most to computation time in the meter placement problem. These approaches solve optimization problems that reduce the voltage residual in the selected reference cases. The first approach addresses the meter placement problem as a bus selection problem that results in the highest voltage residual reduction. This simplification enables a closed-form solution and reduces computation time. On the other hand, the element selection approach deals with the problem of placing a heterogeneous set of sensors that minimize the voltage residual. These approaches also provide a mechanism to extend existing sensor sets to improve state estimation performance by including them in the initial sensor set. The proposed approaches have been evaluated on the 33-node distribution system and a 559-node unbalanced 3-phase test system. These approaches provide superior state estimation accuracy for any given sensor set size over the active learning methods. The limited variations that can be introduced into the matrix completion-based state estimators result in the poor performance of the active learning approaches. These approaches also outperform the dynamic programming approach in the bus selection problem, while they provide better voltage magnitude estimation in the problem of placing a heterogeneous set of sensors.

In the next chapter, we generate data-driven dictionaries to improve compressibility and to reduce reconstruction and state estimation errors.

CHAPTER 8

OPTIMIZING DSSE PERFORMANCE: DATA-DRIVEN DICTIONARIES

8.1 INTRODUCTION

Smart grids utilize sensing and communication infrastructure to enable two-way communication exchange between utilities and consumers. These advanced metering infrastructure (AMI) allow utilities to obtain the situational awareness necessary to monitor, control and operate these networks. Newer technologies such as distributed energy resources, PV generation and electric vehicles have introduced complex interactions among grid elements. These interactions can result in grid instabilities and thus necessitate better situational awareness. Moreover, technologies such as automatic fault detection, automated reconfiguration, VAR control and outage management require close monitoring of the network through sensing equipment, see El-hawary (2014). Therefore, state estimation (SE) becomes essential for many of these technological changes. State estimation plays a crucial role in the control and planning of smart grids, such as real-time network models for the grid, optimizing power flows, and bad data detection , in Huang *et al.* (2012). Distribution system state estimation also facilitates the automation of control operations and network model management functions necessary for the smooth operation of smart grids , see Hayes and Prodanovic (2014). The increase in grid sensors due to the implementations of these systems, results in an excessive stress on the communication networks, thereby limiting the use cases of these sensors. For example, this limitation has resulted in several utilities utilizing valuable metering equipment such as smart meters only for billing purposes. Therefore, there is a need for efficient information transfer to maximize the throughput of existing communication networks.

Several approaches can be used to make the information transfer more efficient such as sensing data only from representative nodes and exploiting the sparsity in the signal to compress the sensed measurements. Among the latter are techniques such as compressive sensing. Compressive sensing exploits the inherent property of many natural signals to have concise representations in some transformed domains. This sparseness allows these signals to be recovered with fewer samples than required by the Nyquist-Shannon sampling theorem. In this regard, compressive sensing generates sparse solutions by mapping the problem to an under-determined system of linear equations and then finds the sparsest solution. Moreover, the choice of transformation basis plays a critical role in compressive sensing to obtain sparse representations. Therefore, this chapter presents a framework for generating and utilizing dictionary learning to learn the transformation basis, resulting in sparse representations of smart meter data. The motivation behind using dictionary learning is to generate data-driven dictionaries that provide better reconstruction accuracy than deterministic transformation bases. These dictionaries could then be utilized in existing compressive sensing-based state estimation techniques to obtain more accurate estimates of the system states and facilitate subsequent control actions.

8.2 PROPOSED APPROACH

The proposed framework involves dictionary generation, data compression, aggregation and state estimation. The dictionary generation, data compression, and aggregation steps are described below. Existing 1-D and 2-D compressive sensing-based state estimators developed by Alam *et al.* (2014) are utilized as the state estimators in this study.

8.2.1 Dictionary Generation

The first step involves computing the data-driven dictionaries for temporal reconstruction. A subset of smart meters are chosen as candidates for obtaining load data to create the data-driven dictionaries. Dictionary learning is then utilized to learn the transformation

matrix, which allows sparse representation of this load data.

The load data from the smart meters is obtained in the form $\boldsymbol{L} \in \mathbb{R}^{N_s \times N_t}$, where N_s is the number of sample smart meters and N_t is the number of time instants. The window length considered for temporal compression is N_w. Now, the load data matrix is truncated to remove the last incomplete window and thus results in the updated data matrix $\boldsymbol{L_{(truncated)}} \in \mathbb{R}^{N_s \times kN_w}$, where k is the number of time windows in the data matrix. The load data is then reshaped to obtain the training data for the dictionary learning.

$$\boldsymbol{X} = \left[reshape(vec(\boldsymbol{L_{(truncated)}}^T)) \right]_{N_w \times kN_s} \tag{8.56}$$

The training data is then used to find the data-driven dictionary by solving the optimization problem (8.57). Once the dictionaries are generated, the compressed measurements sent from the smart meters can be reconstructed using the steps in the following sub-sections.

$$\begin{aligned} (\boldsymbol{D}, \boldsymbol{\alpha}) = & \underset{\boldsymbol{D} \in \boldsymbol{C}, \boldsymbol{\alpha} \in \mathbb{R}^{k \times N}}{\operatorname{argmin}} \frac{1}{N} \sum_{i=1}^{N} \left(\frac{1}{2} \|\boldsymbol{x_i} - \boldsymbol{D}\boldsymbol{\alpha_i}\|_2^2 + \lambda \|\boldsymbol{\alpha_i}\|_1 \right) \\ s.t. \quad & \boldsymbol{C} \in \{\boldsymbol{D} \in \mathbb{R}^{n \times k} : \|\boldsymbol{d_i}\|_2 \leq 1, \forall i = 1, 2, ..., k\}, \lambda > 0 \end{aligned} \tag{8.57}$$

8.2.2 Data Compression and Aggregation

The second step involves compressing the data at the consumer homes and transmitting the compressed measurements to the data aggregators. Data imputation is done as a pre-processing step before compression to fill the missing data. These measurements are then compressed using a random projection matrix $\Phi_{time} \in \mathbb{R}^{m \times n}$, where $m < n$. Thus, for every measurement window $w \in \mathbb{R}^n$, a compressed measurement vector $h = \Phi_{time} w$ is generated and transmitted to the data aggregator. The data aggregator creates the aggregated compressed measurements for each bus using the compressed measurements

from the smart meters connected to them. This step helps in reducing the computational complexity involved in individually reconstructing all compressed measurements from smart meters. A pictorial representation of the steps described above is shown in figure 8.1.

If spatio-temporal compressive sensing is utilized, the data aggregators further compress the measurements from the buses $(h_i, i = 1, 2, ..., q)$ into a matrix $\boldsymbol{H}$ using a random projection matrix $\Phi_{space} \in \mathbb{R}^{p \times q}$, where $p < q$ and q is the number of buses. The compressed measurement matrix is then computed as follows,

$$\boldsymbol{H} = \Phi_{space} \left[h_1, h_2, ..., h_q\right]^T \tag{8.58}$$

8.3 RESULTS

The effectiveness of different transformation matrices and data-driven dictionaries on the recovery performance of loads and state estimation accuracy is studied in this section. The reconstruction performance is quantified using the integrated normalized absolute error (INAE) metric (8.59).

$$INAE = \frac{\sum_{k=1}^{N} \left| \hat{x_k} - x_k \right|}{\sum_{k=1}^{N} \left| x_k \right|} \times 100 \tag{8.59}$$

The IEEE 33-node distribution system and a 100-node distribution system are chosen as the test systems. The descriptions of the test systems are available in Appendix A. Loads from the egauge data set egauge (2018*b*) are distributed such that the peak aggregate load is equal to the nominal load. The minimization problems in this article are solved using CVX (Grant and Boyd (2014, 2008)). Voltage magnitude and angle estimation errors are quantified using the mean absolute percentage error (MAPE) and mean absolute error

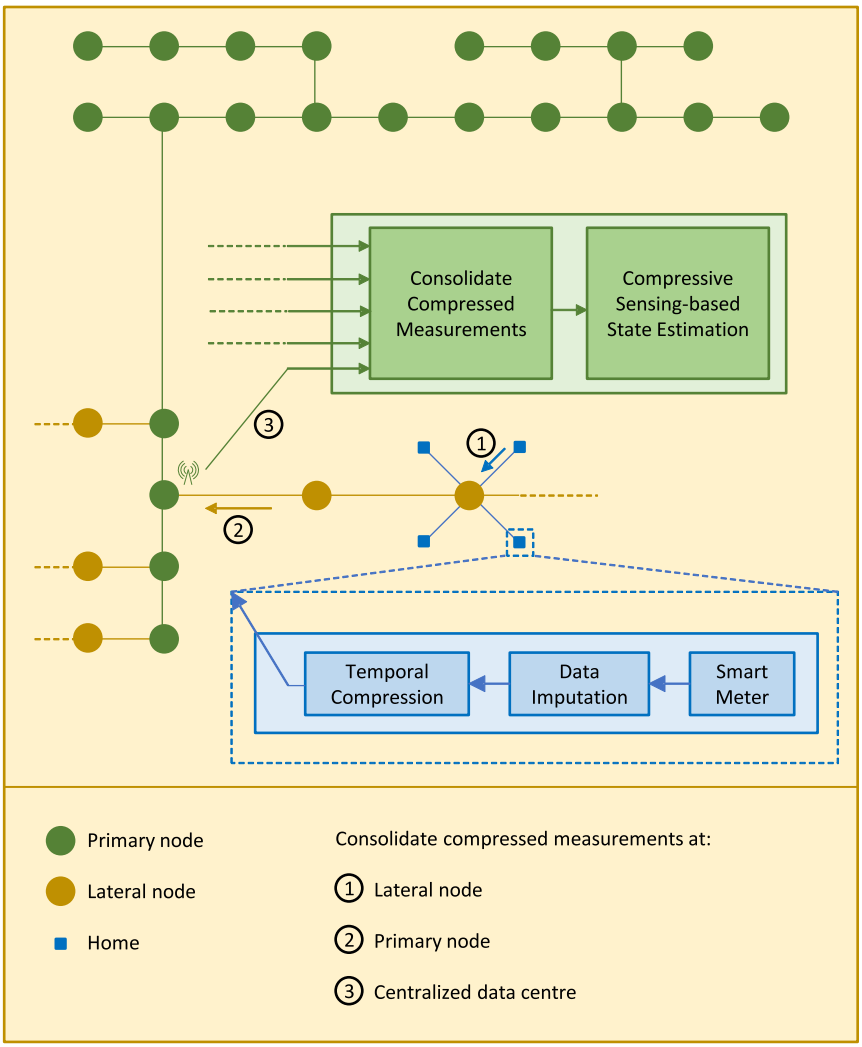


Figure 8.1: Data compression, aggregation and state estimation.

(MAE), respectively [1].

$$MAPE = \frac{1}{N}\sum_{k=1}^{N}\left|\frac{\hat{x_k}-x_k}{x_k}\right| \times 100 \tag{8.60}$$

$$MAE = \frac{1}{N}\sum_{k=1}^{N}\left|\hat{x_k} - x_k\right| \tag{8.61}$$

8.3.1 Data Compression and Reconstruction

The reconstruction performance of different transformation matrices at various compression levels is evaluated in this section. The data was divided into windows of 128 and 256-time instants. The smart meter data was then compressed to compression factors 0.125, 0.25, 0.375, 0.5, 0.625, 0.75 and 0.875. The reconstruction performance using the different transformation matrices on the Egauge data set from egauge (2018*b*) is tabulated in Table 8.1. Similarly, reconstruction performance of different transformation matrices on the AMPds by Makonin (2016), iAWE by Batra *et al.* (2013), PecanStreet by Parson *et al.* (2015), RED by Kolter and Johnson (2011), REFIT by Murray *et al.* (2017) and UK-DALE by Kelly and Knottenbelt (2015) are provided in tables 8.2, 8.3, 8.4, 8.5, 8.6, and 8.7, respectively. It can be observed from these tables that data-driven dictionaries generated through dictionary learning provides better reconstruction performance than all other transformation matrices. Hankel and Toeplitz are the next best choice as transformation matrices for compressing smart meter load data.

[1] Results in this section have been published in:

- **R. Madbhavi** and B. Srinivasan, "Data-driven dictionaries to enhance the performance of compressive sensing-based state estimators", International Journal of Advances in Engineering Sciences and Applied Mathematics, Dec. 2022, doi:10.1007/s12572-022-00324-1.
- **R. Madbhavi** and B. Srinivasan, "Enhancing Performance of Compressive Sensing-based State Estimators using Dictionary Learning," 2022 IEEE International Conference on Power Systems Technology (POWERCON), 2022, pp. 1-6, doi: 10.1109/POWERCON53406.2022.9930028.

Table 8.1: Comparison of reconstruction INAEs for different transformation matrices on the Egauge dataset.

Transformation Matrix	Window Size	Compression Factor (CF)						
		0.125	0.250	0.375	0.500	0.625	0.750	0.875
DCT	256	40.15	22.07	15.46	11.58	8.32	6.57	3.93
Haar	256	26.79	17.14	11.21	10.26	6.87	4.00	1.20
Hadamard	256	27.96	21.64	14.71	12.60	10.36	7.45	4.80
Toeplitz	256	21.69	10.83	7.22	4.74	3.42	1.82	0.52
Hankel	256	21.66	10.74	7.22	4.73	3.41	1.83	0.53
	128	21.68	9.67	7.89	5.15	3.29	2.69	1.09
Hankel-Haar	256	25.57	15.54	11.27	9.79	6.77	3.78	1.29
Hankel-Hadamard	256	27.57	19.02	14.88	12.12	9.92	7.83	4.78
Hankel-DCT	256	29.97	17.94	12.86	10.02	7.45	5.37	3.17
Haar-DCT	256	25.07	14.18	9.71	7.66	5.44	3.05	1.46
Haar-Hadamard	256	27.21	18.12	11.34	8.93	6.62	3.89	2.12
Hadamard-DCT	256	26.98	18.38	13.57	9.87	7.87	5.51	3.57
Data-driven	256	2.58	2.24	2.13	2.00	1.79	1.53	1.29
Dictionary	128	9.66	5.31	4.61	4.22	3.77	3.48	3.03

Table 8.2: Comparison of reconstruction INAEs for different transformation matrices on the AMPds dataset with window size 256.

Transformation Matrix	Compression Factor (CF)						
	0.125	0.250	0.375	0.500	0.625	0.750	0.875
DCT	77.138	42.916	32.543	26.029	19.497	14.651	9.946
Haar	66.673	40.134	24.433	10.628	5.560	2.754	0.972
Hadamard	72.249	46.687	36.538	28.034	22.743	17.097	11.496
Hankel	40.709	28.784	16.974	8.904	3.900	2.112	0.860
Toeplitz	40.885	28.790	16.911	8.951	3.901	2.115	0.860
Hankel-Haar	64.975	38.939	23.979	10.603	5.905	2.922	1.054
Hankel-Hadamard	68.636	45.310	35.840	28.132	22.935	16.977	11.488
Hankel-DCT	69.073	41.675	32.103	24.978	19.023	14.662	9.634
Haar-DCT	70.593	37.582	25.217	15.227	8.291	5.006	2.601
Haar-Hadamard	72.757	39.241	27.189	15.728	7.606	3.965	1.711
Hadamard-DCT	70.432	41.690	32.553	25.378	19.841	15.104	9.770
Data-driven Dictionary	22.764	18.471	14.826	5.931	3.289	2.935	1.567

Table 8.3: Comparison of reconstruction INAEs for different transformation matrices on the iAWE dataset with window size 256.

Transformation Matrix	Compression Factor (CF)						
	0.125	0.250	0.375	0.500	0.625	0.750	0.875
DCT	42.298	13.764	11.044	8.254	5.976	4.543	3.511
Haar	6.706	1.873	0.811	0.509	0.336	0.147	0.038
Hadamard	11.412	6.611	5.131	3.940	3.304	2.553	1.721
Hankel	4.886	2.729	0.903	0.495	0.303	0.162	0.053
Toeplitz	4.827	2.706	0.907	0.499	0.305	0.158	0.051
Hankel-Haar	8.117	2.257	1.141	0.606	0.409	0.199	0.073
Hankel-Hadamard	10.269	5.736	5.033	3.829	3.272	2.506	1.724
Hankel-DCT	17.319	6.428	4.912	3.568	2.924	2.203	1.619
Haar-DCT	9.474	2.391	1.172	0.671	0.451	0.288	0.151
Haar-Hadamard	9.444	2.500	1.097	0.588	0.418	0.229	0.100
Hadamard-DCT	10.614	5.816	4.333	3.383	2.941	2.179	1.727
Data-driven Dictionary	3.523	1.653	0.262	0.202	0.134	0.109	0.092

Table 8.4: Comparison of reconstruction INAEs for different transformation matrices on the PecanStreet dataset with window size 256.

Transformation Matrix	Compression Factor (CF)						
	0.125	0.250	0.375	0.500	0.625	0.750	0.875
DCT	47.023	22.116	17.444	12.355	9.026	6.893	4.786
Haar	27.242	13.388	5.984	2.302	0.945	0.517	0.240
Hadamard	34.018	20.617	15.957	11.953	9.071	6.851	4.370
Hankel	15.061	8.519	5.064	1.936	0.837	0.403	0.142
Toeplitz	15.044	8.530	5.058	1.957	0.832	0.402	0.143
Hankel-Haar	28.067	12.433	6.030	2.808	1.252	0.699	0.305
Hankel-Hadamard	33.234	19.310	15.368	11.505	9.049	6.718	4.309
Hankel-DCT	31.778	18.671	14.564	10.265	7.628	5.798	3.907
Haar-DCT	30.237	14.252	7.697	3.754	2.013	1.189	0.574
Haar-Hadamard	33.689	15.272	7.728	3.558	1.547	0.872	0.377
Hadamard-DCT	33.336	18.847	14.177	10.382	7.669	5.900	3.924
Data-driven Dictionary	12.106	8.103	5.051	1.698	1.117	0.800	0.266

Table 8.5: Comparison of reconstruction INAEs for different transformation matrices on the REDD dataset with window size 256.

Transformation Matrix	Compression Factor (CF)						
	0.125	0.250	0.375	0.500	0.625	0.750	0.875
DCT	41.715	12.727	10.313	7.659	4.893	4.074	3.124
Haar	1.571	0.309	0.255	0.222	0.189	0.148	0.104
Hadamard	1.897	1.380	1.207	0.912	0.711	0.582	0.417
Hankel	1.168	0.389	0.304	0.258	0.202	0.158	0.113
Toeplitz	1.158	0.389	0.308	0.264	0.203	0.161	0.113
Hankel-Haar	2.161	0.317	0.266	0.222	0.191	0.146	0.105
Hankel-Hadamard	2.393	1.221	1.167	0.774	0.716	0.577	0.395
Hankel-DCT	11.858	3.231	1.869	1.207	0.877	0.651	0.419
Haar-DCT	2.292	0.557	0.319	0.223	0.180	0.132	0.094
Haar-Hadamard	1.859	0.316	0.260	0.210	0.179	0.137	0.097
Hadamard-DCT	2.159	1.254	1.021	0.732	0.615	0.532	0.378
Data-driven Dictionary	1.188	0.299	0.154	0.155	0.166	0.158	0.190

Table 8.6: Comparison of reconstruction INAEs for different transformation matrices on the REFIT dataset with window size 256.

Transformation Matrix	Compression Factor (CF)						
	0.125	0.250	0.375	0.500	0.625	0.750	0.875
DCT	49.384	23.052	17.123	13.525	9.613	7.464	5.383
Haar	29.954	11.855	6.246	3.470	2.207	1.061	0.420
Hadamard	34.434	21.890	17.035	13.470	10.177	7.496	5.594
Hankel	17.528	10.229	5.258	2.876	1.573	0.793	0.349
Toeplitz	17.607	10.258	5.254	2.881	1.573	0.787	0.345
Hankel-Haar	27.695	12.283	6.662	3.626	2.395	1.005	0.488
Hankel-Hadamard	30.752	20.874	17.050	13.697	10.271	7.434	5.588
Hankel-DCT	33.890	19.383	13.985	11.330	8.384	6.201	4.146
Haar-DCT	28.313	13.678	8.338	4.393	2.683	1.462	0.690
Haar-Hadamard	34.584	13.078	8.560	4.562	2.855	1.793	0.696
Hadamard-DCT	30.751	19.102	14.189	11.671	8.598	6.292	4.518
Data-driven Dictionary	13.308	9.599	4.809	2.807	1.347	0.465	0.323

Table 8.7: Comparison of reconstruction INAEs for different transformation matrices on the UK-DALE dataset with window size 256.

Transformation Matrix	Compression Factor (CF)						
	0.125	0.250	0.375	0.500	0.625	0.750	0.875
DCT	47.431	20.485	17.729	13.958	10.939	8.005	5.728
Haar	27.066	7.155	4.722	3.400	2.504	2.183	1.033
Hadamard	32.822	20.391	16.940	11.679	8.506	7.253	4.749
Hankel	16.055	10.704	5.372	4.066	3.395	2.712	1.747
Toeplitz	16.090	10.778	5.378	4.057	3.391	2.720	1.753
Hankel-Haar	25.557	8.076	5.809	3.886	2.741	2.056	1.119
Hankel-Hadamard	29.208	18.983	16.382	11.456	8.694	7.034	4.717
Hankel-DCT	32.127	17.791	14.088	11.519	8.809	6.529	4.725
Haar-DCT	28.175	9.734	7.050	4.311	3.059	2.241	1.421
Haar-Hadamard	29.303	8.482	5.156	4.050	2.543	2.367	1.380
Hadamard-DCT	26.732	17.542	13.517	10.672	8.531	6.649	4.497
Data-driven Dictionary	7.348	5.711	4.113	3.713	3.080	2.038	1.358

Deterministic Dictionaries

The deterministic dictionaries used in this study can be further classified as complete and over-complete dictionaries. The complete dictionaries considered comprise of DCT, Haar, Hadamard, Toeplitz and Hankel matrices. Over-complete dictionaries are generated by combining these complete dictionaries.

Among the complete dictionaries, it can be observed from Table 8.1 that the Hankel and Toeplitz matrices provide the best reconstruction performance. Hankel matrices can be generated such that any step or ramp signal can be generated by linear combinations of only two of their columns. Since such changes are expected in load data, the Hankel and Toeplitz matrices perform better than other deterministic dictionaries.

Over-complete dictionaries contain more atoms than the length of the atoms themselves. This redundancy helps to provide sparse representations of the signal. Therefore, combinations of the complete dictionaries were used to evaluate their performance. However, none of the combinations provided any improvement over the Hankel matrix.

Table 8.8: Voltage magnitude estimation MAPE (%) using compressive sensing on the IEEE 33-node distribution system.

Transformation Matrix	Spatial Compression Factor / Temporal Compression Factor	1.00	0.95	0.90	0.85	0.80
Data-driven Dictionary	0.500	0.004	0.011	0.063	0.089	0.144
	0.375	0.004	0.011	0.065	0.094	0.145
	0.250	0.282	0.283	0.292	0.295	0.323
Haar	0.500	0.055	0.057	0.062	0.122	0.172
	0.375	0.085	0.091	0.090	0.157	0.208
	0.250	1.906	1.940	2.137	2.246	2.388

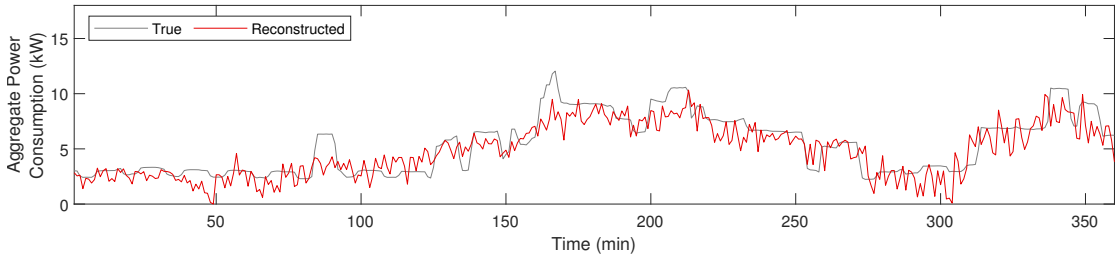
Table 8.9: Voltage angle estimation MAE (degrees) using compressive sensing on the IEEE 33-node distribution system.

Transformation Matrix	Spatial Compression Factor / Temporal Compression Factor	1.00	0.95	0.90	0.85	0.80
Data-driven Dictionary	0.500	0.002	0.009	0.007	0.016	0.081
	0.375	0.006	0.009	0.008	0.018	0.087
	0.250	0.022	0.024	0.025	0.029	0.093
Haar	0.500	0.005	0.009	0.011	0.017	0.088
	0.375	0.007	0.015	0.012	0.018	0.090
	0.250	0.153	0.152	0.170	0.171	0.175

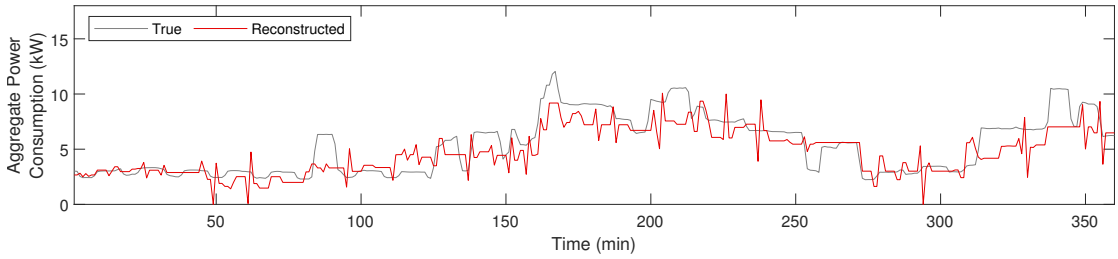
The Hankel-Toeplitz combination is not chosen in this study since they are interchangeable using exchange matrices and do not provide new atoms to the dictionary.

Data-driven Dictionaries

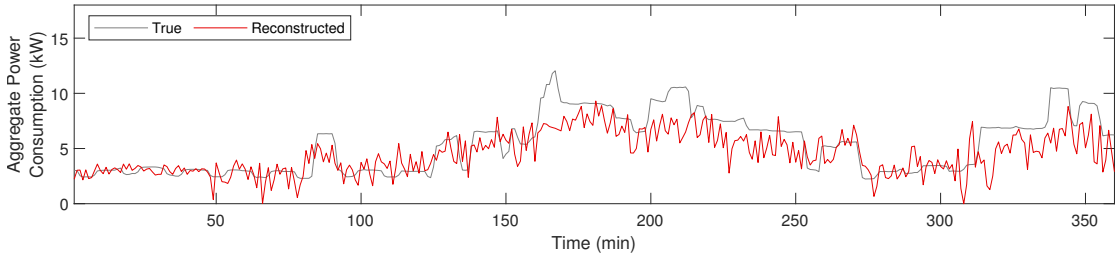
Data-driven dictionaries were generated using dictionary learning as described in Section 8.2.1. These dictionaries provided better reconstruction than the deterministic dictionaries. These dictionaries were more accurate in reconstructing the load data at lower compression factors than the Hankel and Toeplitz matrices (Table 8.1). It can be observed from Fig. 8.2 that reconstruction with data-driven dictionaries results in the least signal reconstruction errors.



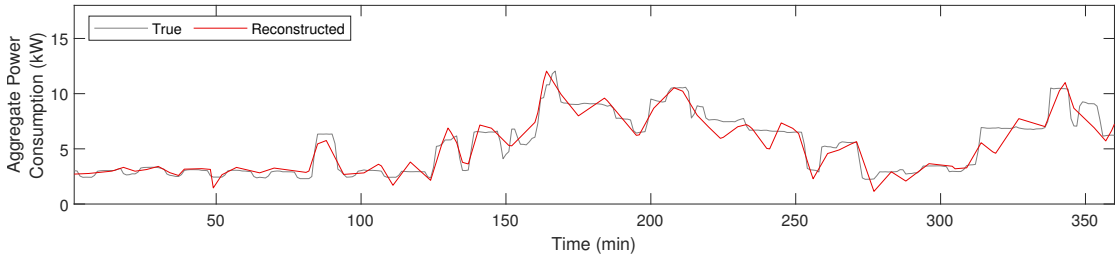
(a) DCT



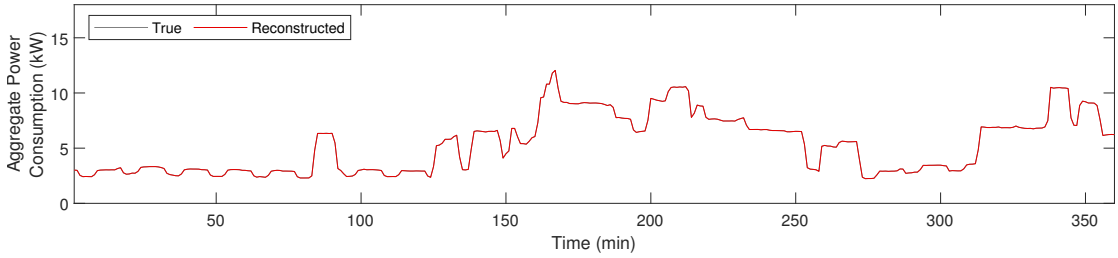
(b) Haar



(c) Hadamard



(d) Hankel



(e) Data-driven dictionary

Figure 8.2: Comparison of load data reconstruction using different transformation matrices (compression ratio = 4).

Table 8.10: Voltage magnitude estimation MAPE (%) using compressive sensing on the 100-node distribution system.

Transformation Matrix	Spatial Compression Factor / Temporal Compression Factor	1.00	0.95	0.90	0.85	0.80
Data-driven Dictionary	0.500	0.005	0.014	0.061	0.091	0.151
	0.375	0.006	0.020	0.080	0.103	0.165
	0.250	0.312	0.343	0.359	0.371	0.405
Haar	0.500	0.080	0.091	0.0117	0.145	0.192
	0.375	0.105	0.121	0.143	0.162	0.219
	0.250	2.473	2.540	2.597	2.866	2.992

Table 8.11: Voltage angle estimation MAE (degrees) using compressive sensing on the 100-node distribution system.

Transformation Matrix	Spatial Compression Factor / Temporal Compression Factor	1.00	0.95	0.90	0.85	0.80
Data-driven Dictionary	0.500	0.003	0.008	0.011	0.021	0.100
	0.375	0.006	0.010	0.010	0.023	0.106
	0.250	0.025	0.035	0.043	0.060	0.124
Haar	0.500	0.008	0.011	0.019	0.032	0.108
	0.375	0.008	0.018	0.017	0.032	0.143
	0.250	0.192	0.216	0.223	0.248	0.259

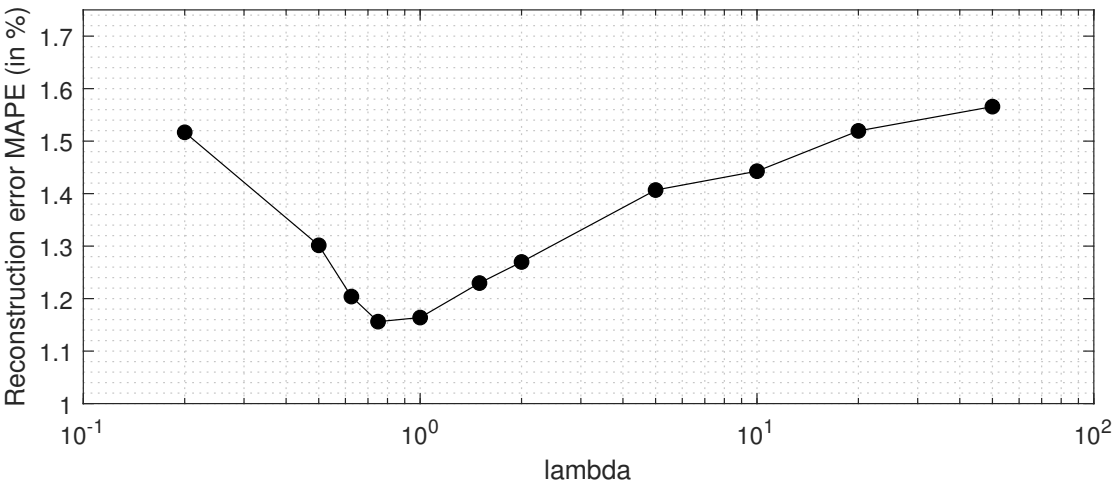


Figure 8.3: Impact of lambda on reconstruction error of bus loads.

Impact of Lambda on Reconstruction Error

The parameter λ in the dictionary learning formulation provides a trade-off between the squared sum of reconstruction errors and the l_1 norm of atom weights. Therefore, the parameter λ plays a critical role in obtaining sparse solutions while still having low reconstruction errors.

The impact of this parameter λ was studied by generating dictionaries with different values of λ. It can be observed from Fig. 8.3 that at lower values of $\lambda(< 0.5)$, the reconstruction error keeps increasing. Under these conditions, the learnt dictionary was forced to prioritize minimizing the sum squared of errors. However, these dictionaries could not learn atoms that generalize for data outside the training sample. On the other hand, at higher values of $\lambda(> 2)$, the reconstruction error again increases. In this case, the learnt dictionary is forced to prioritize generating sparse representations on the training data at the cost of reconstruction accuracy on test sets.

Therefore, the choice of λ is critical for obtaining low reconstruction errors. However, the optimum choice of λ depends on the data and needs to be obtained by sampling the system for measurements.

8.3.2 State Estimation

The direct and indirect state estimation approaches (Section **??**) are evaluated using data-driven dictionaries as the transformation matrix. The voltage magnitude and angle estimation results for the IEEE 33-node distribution system are tabulated in tables 8.8 and 8.9. Similarly, the results for the 100-node distribution system are tabulated in tables 8.10 and 8.11, respectively. Results with spatial compression factor 1 correspond to the indirect state estimation approach, while the other columns show the performance of the direct state estimation approach. It is observed that minimal spatial compression is possible as compared to temporal compression. The state estimation accuracy reduces significantly below the spatial compression factor of 0.85. These results are compared with the results of the Haar matrix utilized in Alam *et al.* (2014). The results of the 100-node distribution system show that the method is scalable to larger systems.

Figure 8.4 shows the state estimation result at different spatial and temporal compression factors at a single time instant. It is observed that the voltage angle estimation is very poor at the spatial compression factor of 0.8.

8.4 SUMMARY

Compressive sensing is an effective tool to compress smart meter data and thus reduce the load on communication networks. The compressibility of the signals and reconstruction errors so obtained are dictated by the effectiveness of the transformation under which the signal is assumed to be sparse. Previous works utilize deterministic matrices such as Haar, Hankel, Hadamard, DCT, and Toeplitz as transformation bases. However, these matrices do not generate the sparsest representations of smart meter data. On the other hand, data-driven dictionaries computed from smart meter data provide better reconstruction than complete and over-complete dictionaries even while using lower compression factors, thereby reducing the network load. These dictionaries also provide better state estimation performance with the compressive sensing based state estimator

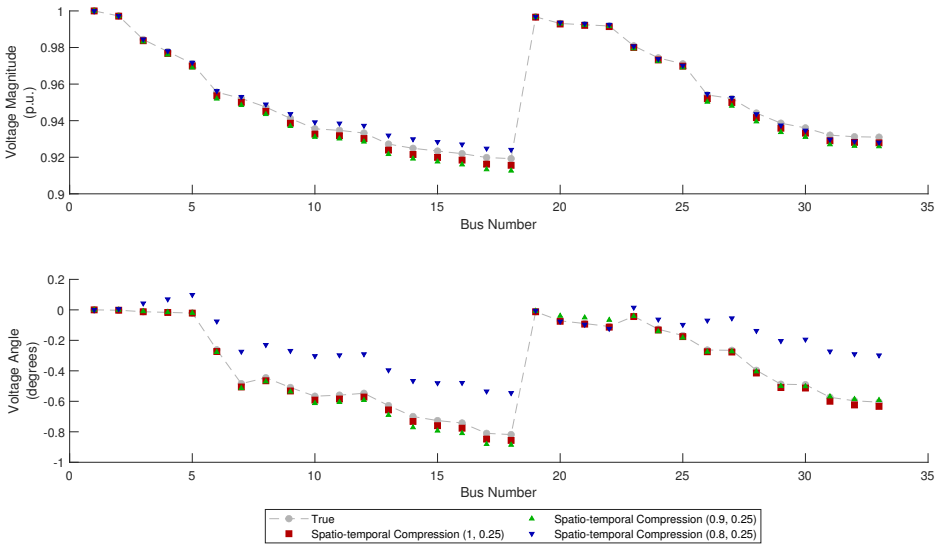


Figure 8.4: Comparison of state estimation results at different compression levels. The spatial and temporal compression factors are provided in brackets.

than deterministic transformation matrices. Therefore, data-driven dictionaries should be preferred over traditional deterministic transformation bases to minimize signal reconstruction errors and improve the performance of estimators utilizing compressed measurements. In the next chapter, we explore problems in other CPS such as leak detection in water distribution networks.

CHAPTER 9

EXPLORING OTHER CPS: LEAK DETECTION IN WATER DISTRIBUTION NETWORKS

9.1 INTRODUCTION

Water distribution networks constitute an important infrastructure of cities that manage a highly valuable and indispensable natural resource. Water distributed through these networks is consumed for household, agricultural, commercial as well as industrial purposes. Water distribution networks (WDNs) are tasked with providing water to consumers at desirable levels of pressure while accommodating variations in demand. In addition, WDNs are also required to meet standard requirements with respect to the quality of water delivered to consumers. Although WDNs are designed to ensure timely delivery of healthy and safe water, they are prone to disruptions such as leaks and clogging that can lower the quantity and/or quality of water being delivered, see Yazdani and Jeffrey (2012). Such a disruption can be due to harsh environmental conditions imposing say, excessive pressure on the pipes, deterioration of physical components over time, see Berardi *et al.* (2008), or adversarial attacks. Among several such scenarios, *leaks* are a common disruption and constitute a significant threat to the operation of WDNs.

A leak in the distribution network is characterized by breakage of the physical structure (nodes and/or pipes) carrying water to the consumers, resulting in a significant amount of water being lost to the environment. The amount of water lost by the network due to a leak is proportional to the capacity of the leaky node and/or pipeline, and also a function of the location of the leak. For example, a leak in a location that is close to pumps or reservoirs can result in excessive loss of water owing to continuous pumping through the region or proximity to the source. Loss of water due to leaks can be as high

as 20% to 30% of the total amount of water supplied by the network, see Chan *et al.* (2018). In addition to loss, leaks can also lead to the water flowing through the pipes to be contaminated due to exposure to the surrounding environment, in Collins and Boxall (2012). Furthermore, multiple leaks can also occur simultaneously at different locations in a distributed network, further accentuating the impact on operation of WDNs. Several of these factors act as confounding agents that make it a highly challenging task to detect and localize leaks in a water distribution network. Detection and localization of leaks has received significant research attention resulting in a host of techniques proposed in the literature for addressing the problem, see Chan *et al.* (2018).

9.2 PROPOSED APPROACH

The problem of leak localization is first formulated as a classification problem in the framework of machine learning. This allows identification of the location of a given number of sensors that maximises the accuracy of localization. Feature selection is then incorporated into the problem formulation to adjust the number of sensors from a small value to an upper threshold (a fraction of the number of nodes in the network). This approach is applied on four standard water distribution network case studies, in which the accuracy of localization of leak is calculated for different number of sensors placed in the network. This exercise is performed with an exhaustive set of classification and feature selection techniques and the performance of different methods are compared. Two feature selection techniques are proposed that perform at least as good as the methods available in the literature.

9.2.1 Problem Formulation

Consider a water distribution network $\mathcal{N}$ that consists of N_p pipelines and N_n nodes (or junctions) where two or more pipelines meet. Assuming the magnitude of leak at the i^{th} node in the network to range from $L_i^{(Min)}$ to $L_i^{(Max)}$ and can take M_i different values. For the case of a single leak at a node in the network as discussed above, this results

in a total of $\sum_{i=1}^{N_n} M_i$ possible leaky scenarios for the network. Additionally, assuming that it is possible to install N ($\leq N_n$) sensors that collect measurements at the nodes and communicate to a central location (eg. SCADA center) for further processing. Let $x \in \mathbb{R}^N$ denote the vector of measurements at steady state from the N sensors installed in $\mathcal{N}$. This data vector, can at any time, correspond to the case of normal operation (no leak in the network), or a leak at the i^{th} node (with $i \in [1, N_n]$). In a supervised machine learning framework, x is said to belong to one of N_n classes, which is typically represented in a classification formulation as a target $y \in [1, N_n]$. A classification model Ξ then learns the relationship between x and y, represented as:

$$\Xi(x, \theta): \qquad y = f(x; \theta) \tag{9.62}$$

where $f(\cdot)$ represents structure of the model (support vector machine, neural network, etc.) and θ represents the parameters used in the model to construct the relationship between x and y. The task of machine learning then involves using a set of data vectors and classes (called the training data) to adjust the parameters θ to minimize the classification error. After the parameters have been learnt from the data, the model can then be used to predict the class of any new input data vector (called testing data).

It can be observed from Equation 9.62 that the performance of the model Ξ depends on the structure of Ξ, the parameters θ and the information content in x at the time of training that can be used to predict the corresponding y. It is then straightforward that for a given structure of Ξ and a network $\mathcal{N}$ with N_n nodes, a model that employs measurements from all nodes in the network ($N = N_n$) is likely to perform better than one that employs measurements from only a few nodes ($N < N_n$). This suggests that in order to reliably detect and localize leaks in a network, it is necessary to install sensors at all nodes, which is not an economically viable solution to the problem. However, this challenge can be addressed by acknowledging the fact that there exist some nodes in the network that, owing to their location in the network, carry significantly more information relevant for solving the classification problem than some others (Any additional sensors installed

in the network will then result in only a limited improvement in performance of the classifier, or be redundant, making the classification exercise robust to noisy or missing measurements). This in turn suggests that it is possible to strategically place sensors at only a few nodes in the network that provide a significant amount of information for localization of the leak. Identifying the most informative nodes in the network for installing sensors then provides a pragmatic solution to the leak localization problem. Furthermore, in doing so, one is provided with the flexibility of striking a trade-off between the accuracy of localization and the number of sensors required to be installed. This problem of identifying sensors that are more important to solve another problem is referred to as feature selection/extraction, which in the present case, involves identifying the components of x that can be used to solve the classification problem while delivering acceptable levels of performance.

Formulating the problem of leak localization as one of feature selection followed by classification therefore allows one to simultaneously,

1. Identify a suitable number of sensors to be installed.
2. Identify the locations of nodes where the sensors are required to be installed.
3. Perform classification to localize the leak.

In addition, the accuracy of leak localization can also be traded-off against the number of sensors to be installed in the network. The different classification techniques and the associated feature selection methods are discussed in the following sub-sections.

9.2.2 Correlation-based Feature Selection

Correlation-based feature selection techniques exploit the underlying correlation structure of the data of different classes. This approach involves a set of *selected features* $\mathcal{S}$ which is iteratively populated with features that have the most *novel information* for predicting the output class. In this approach, the correlation of all features that are not members of

$\mathcal{S}$ ($x^{(\mathcal{S}')}$) with all features that are members of $\mathcal{S}$ ($x^{(\mathcal{S})}$) at every step are calculated as:

$$C\left(x_i^{(\mathcal{S}')}, x_j^{(\mathcal{S})}\right) = \frac{\sum_{k=1}^{K}\left(x_{i,k}^{(\mathcal{S}')} - \bar{x}_i^{(\mathcal{S}')}\right)\left(x_{j,k}^{(\mathcal{S})} - \bar{x}_j^{(\mathcal{S})}\right)}{\sqrt{\sum_{k=1}^{K}\left(x_{i,k}^{(\mathcal{S}')} - \bar{x}_i^{(\mathcal{S}')}\right)^2}\sqrt{\sum_{k=1}^{K}\left(x_{j,k}^{(\mathcal{S})} - \bar{x}_j^{(\mathcal{S})}\right)^2}} \tag{9.63}$$

where K represents the total number of observations for all classes; $\bar{x}_i^{(\mathcal{S}')}$ and $\bar{x}_j^{(\mathcal{S})}$ are the means of the K observations corresponding to features $x_i \in \mathcal{S}'$ and $x_j \in \mathcal{S}$ respectively. This is followed by scoring of the features that are not members of $\mathcal{S}$ and including those feature(s) with the highest score(s) to populate $\mathcal{S}$.

The average correlation of every $x_i^{(\mathcal{S}')}$:

$$C_{LC}(x_i^{(\mathcal{S}')}) = \frac{1}{|\mathcal{S}|}\sum_{j=1}^{|\mathcal{S}|} C\left(x_i^{(\mathcal{S}')}, x_j^{(\mathcal{S})}\right) \tag{9.64}$$

$|\mathcal{S}|$ representing the number of elements in $\mathcal{S}$ then represents the amount of *novel information* carried by the features. This average correlation is used as a score to rank the features for inclusion in $\mathcal{S}$, and a predefined number (m) of features with the highest score are added to $\mathcal{S}$. This process is repeated until the desired number of features is reached, or the desired performance attained.

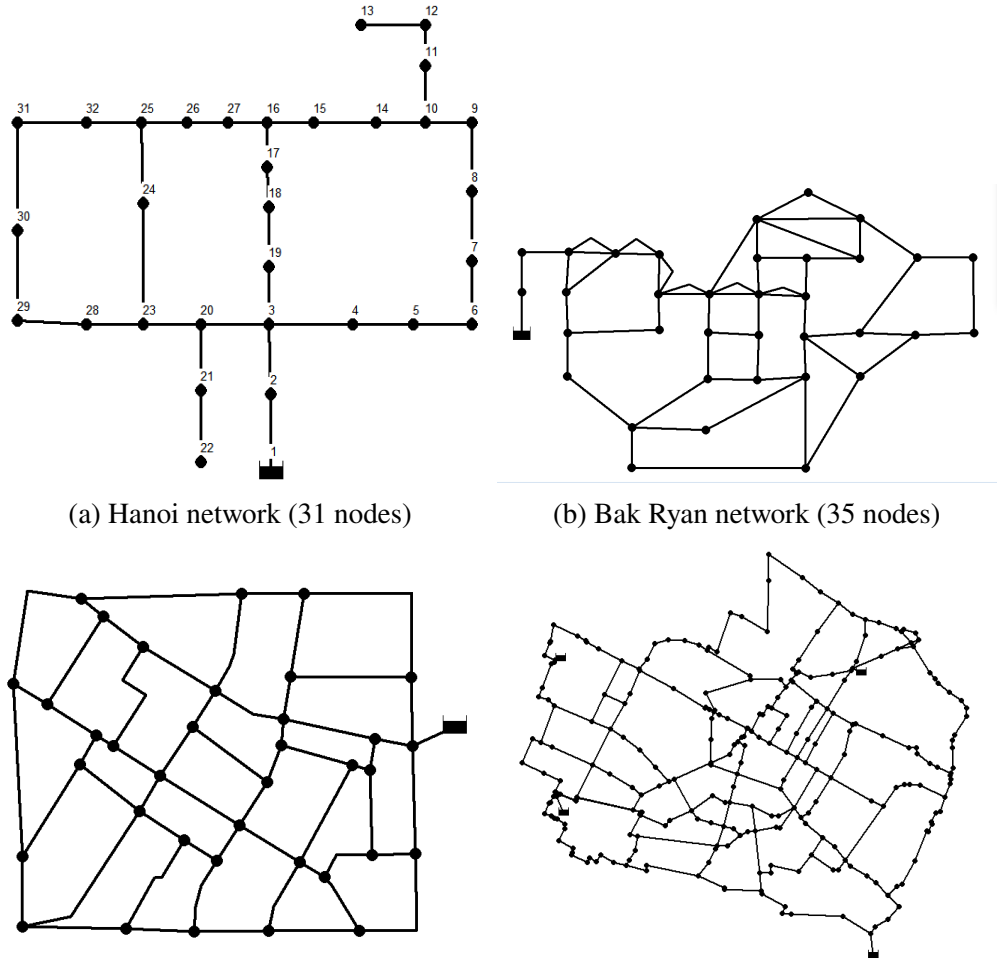
It must be noted that in the above approach, all variables that are not members of $\mathcal{S}$ are assigned equal importance, and the ones with the least average correlation with existing features are chosen for inclusion in the model. However, it does not account for the variability in the individual features, which can be different for different features. Therefore, a *Weighted Least Correlation* (WLC) approach can be utilized. WLC weighs the correlation of each variable (feature) differently so as to account for its variance and calculates the weighted score of $x_i^{(\mathcal{S}')}$ as:

$$C_{WLC}(x_i^{(\mathcal{S}')}) = \frac{W_i}{|\mathcal{S}|}\sum_{j=1}^{|\mathcal{S}|} C\left(x_i^{(\mathcal{S}')}, x_j^{(\mathcal{S})}\right) \tag{9.65}$$

Here, W_i represents the weight assigned to $x_i^{(S)}$. This allows assigning more importance to features that have more variability, i.e., that exhibit more sensitivity to the different classes and hence that can be used to classify the data better. This is achieved in this work by using the standard deviation of measured pressures across the different leak scenarios as the W_i for each sensor. Correlation based feature selection approaches are used along with SVM and kNN classifiers for identifying nodes for sensor placement and localizing leaks in the network.

9.3 DATA GENERATION AND MODEL TRAINING

In this chapter, four water distribution networks: Hanoi, Bak Ryan, Fossolo and Modena networks are considered. Their layouts are shown in Fig. 9.1. The number of nodes and pipes in the networks are listed in Table 9.1. The networks were simulated with the files obtained from Wang *et al.* (2015) and generated the data for training and evaluation of different classifiers according to the procedure depicted in Fig. 9.2. Each network was first simulated with EPANET under normal operating conditions and the steady state pressure at all nodes were recorded. This was followed by introducing leaks in the network by increasing the demand at the leaky node. The range of magnitude of leak introduced in the four networks are listed in Table 9.1. Measurements of steady state pressure collected at all nodes in the network for all the above cases then resulted in a data matrix $\mathcal{D} \in \mathbb{R}^{P \times N_n}$ where $P = \sum_i M_i + 1$ as shown in Fig. 9.2(a). In order to account for noise in measurements due sensor characteristics, Gaussian random noise is added to the measurements at each node, for each magnitude of leak at each node in the network. This was performed 10 times resulting in a matrix $\mathcal{X} \in \mathbb{R}^{10P \times N_n}$, referred to as the predictor matrix that contains 10 different instances of noisy measurements from all nodes in the network for all the leaky cases and under normal operating conditions. In order to perform the classification exercise, it is necessary to provide the target vector containing the class to which each example (row vector) in the predictor matrix belongs. This resulted in a target vector $\mathcal{Y} \in \mathbb{R}^{10P}$. This process is depicted in Fig. 9.2(b). The



(a) Hanoi network (31 nodes)

(b) Bak Ryan network (35 nodes)

(c) Fossolo network (36 nodes)

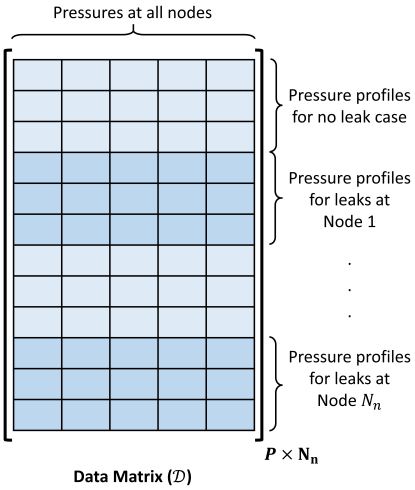
(d) Modena network (268 nodes)

Figure 9.1: Layout of four water distribution networks.

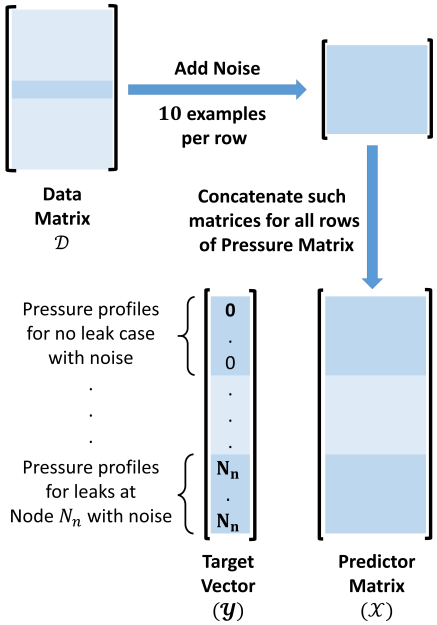
Table 9.1: Network and EPANET simulation parameters for the four WDNs.

Network	**Network Parameters**			**EPANET Simulation Parameters**		
	Nodes	**Pipes**	**Total Demand (LPS)**	**Minimum Leak (LPS)**	**Maximum Leak (LPS)**	**Step Size (LPS)**
Hanoi (HAN)	31	34	5538.9	5.0	28.0	1.0
BakRyan (BAK)	35	58	1146	1.0	6.0	1.0
Fossolo (FOS)	36	58	33.9	0.5	1.0	0.1
Modena (MOD)	268	317	406.9	0.5	3.0	0.5

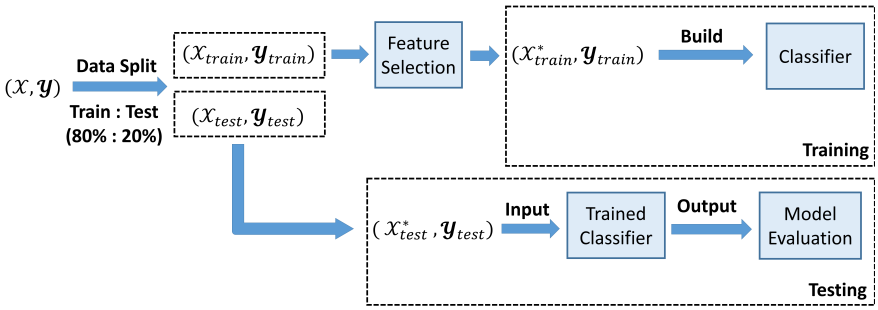
predictor matrix $\mathcal{X}$ and the corresponding target vector $\mathcal{Y}$ were then divided into a training dataset $(\mathcal{X}_{Train}, \mathcal{Y}_{Train})$ and testing set $(\mathcal{X}_{Test}, \mathcal{Y}_{Test})$ by random splitting and selection with the former containing 80% examples as shown in Fig. 9.2(c).



(a) Generation of data matrix $\mathcal{D}$ from pressure data obtained with EPANET.



(b) Generation of predictor matrix and target labels from data matrix.



(c) Generation of training and testing data for classifier training and evaluation.

Figure 9.2: Data generation and model training procedure.

Having generated the training and testing datasets, the training datasets are utilized to learn the parameters of different classifiers, coupled with different feature selection techniques. The training of classifiers was performed with categorical cross-entropy as the objective function. The training as well as existing feature selection techniques were implemented with the scikit-learn by Pedregosa *et al.* (2011) package in Python, while the proposed correlation based feature selection was performed with custom functions in Python. This was followed by testing the performance of a classifier for different number of sensors identified with different feature selection techniques for all classification

models.

9.4 RESULTS

In this chapter, four water distribution networks have been considered. The results of classification tasks can be analyzed using confusion matrices and metrics derived therefrom. In addition to the traditional metrics associated with machine learning, a topographical measure of the performance of classification models referred to as the Average Topological Distance (ATD) developed by Soldevila *et al.* (2016) is considered. While accuracy, precision, recall, and F1-score together quantify the ability of a classifier to identify the exact class, ATD allows one to quantify the classification error in terms of the distance between the predicted and true location of leaks. Specifically, ATD is defined as the mean value of the minimum *topological distance* between the node which has a leak and the node at which the classifier predicts the leak to have occurred. In the above definition, the topological distance between any two nodes in a network can be obtained as the length (number of nodes) of the shortest path between the two nodes. Mathematically, ATD can be represented as:

$$ATD = \frac{\sum_{i,j} \Gamma_{i,j} \Delta_{i,j}}{\sum_{i,j} \Gamma_{i,j}} \tag{9.66}$$

where Δ represents the topological distance matrix and $\Gamma_{i,j}$ represents the number of data points in class C_i classified as class C_j.

The results and a comparative analysis of classification with different combinations of feature selection methods for the four networks shown in Fig. 9.1. In each case, localisation of leak in the networks is performed with SVM, kNN and RF as classifiers with Recursive Feature Elimination (RFE) and proposed feature selection techniques such as: $RFE-SVM$, $RFE-RF$, LC, and WLC. A comparison of accuracy and ATD obtained with each combination of classifier (CL) and feature selection (FS) techniques

is also presented. The results are reported for number of sensors ranging from 2 to $\sim$ 25% of the number of nodes in a network, the latter being an empirically observed threshold beyond which addition of sensors does not improve the accuracy by a significant amount [1].

9.4.1 Hanoi

The first network considered in this work is a relatively small network with only 31 nodes and one tank, the Hanoi network, refer Fujiwara and Khang (1990). Leak localisation with the approaches mentioned above was performed with the number of sensors, $N \in [2, 8]$ for this network. Fig. 9.3 shows the accuracy obtained with all the $CL-FS$ combinations for the Hanoi network, with each plot comparing the performance of different methods of feature selection for a single classifier. It can be observed from Fig. 9.3 that the performance of all $CL - FS$ combinations (except $kNN - RFE - SVM$) improves dramatically from $N = 2$ to $N = 4, 5$ after which the increase in performance per sensor installed in the network reduces. In addition, it can also be observed that the results obtained with LC and WLC are identical for this network with SVM as well as kNN as the classifiers. However, as shown later, this phenomenon is observed only for the Hanoi network, and the results of LC and WLC are different for the other networks. SVM is observed to perform poor compared to all other FS techniques. Although RF is, in general expected to perform better due to the ensemble of classifiers built into it compared to a single classifier in SVM, the performance of kNN is also observed to be higher than that of SVM for almost all values of N.

The ATD of different techniques shown in Table. 9.4 also suggest that the errors made by the techniques are low, with a worst case value of $\sim$ 0.7 node. In order to compare

[1] Results in this section have been published in:

- **R. Madbhavi**, A. Joshi, S. Munikoti, L. Das, P. K. Mohapatra and B. Srinivasan, "Sensor placement for leak localization in water distribution networks using machine learning," 2020 IEEE Int. Conf. Comput. Power Commun. Technol. GUCON 2020, pp. 95–100, Oct. 2020, doi: 10.1109/GUCON48875.2020.9231148.

Table 9.2: ATD obtained with approach proposed by Soldevila *et al.* (2016).

Feature Selection	Sensor Location	Classifier	ATD
Exhaustive Search	14, 28		1.15
Wrapper	14, 28	Bayesian	1.27
Hybrid Feature Selection	14, 28		1.18

Table 9.3: ATD obtained with feature selection and classification scheme proposed in this chapter.

Feature Selection	Sensor Location	Classifier	ATD
RFE-SVM	13, 30	SVM	0.67
		kNN	0.27
LC	13, 22	SVM	0.38
		kNN	0.31
WLC	13, 22	SVM	0.38
		kNN	0.31
RFE-RF	13, 22	RF	0.33
		kNN	0.24

the results with the literature, localisation of leaks with sensors placed according to the scheme proposed in Soldevila *et al.* (2016) is performed and compared the performance the techniques used in this chapter. Specifically, Soldevila *et al.* (2016) assumed $N = 2$ sensors to be placed in the network and with different feature selection techniques, identified that the two sensors should be placed at nodes 14 and 28. However, the techniques used in this chapter identified different nodes as being important for localising leaks in the network. Table 9.2 and Table 9.3 list the ATD reported in Soldevila *et al.* (2016) and those obtained with the techniques proposed in this chapter for $N = 2$ sensors placed in the network respectively. It can be observed from Table 9.2 and Table 9.3 that while the technique in the literature has a best-case ATD of 1.15, the ATD of the techniques used in this chapter are much smaller and lie in the range $[0.24, 0.67]$, delivering a worst-case localisation error that is approximately half ($\sim$ 58%) of the best-case error reported in the literature.

Table 9.4: ATD obtained for the four WDNs using all the feature selection methods.

Net.	Clf.	FS	Number of Sensors						
			2	**3**	**4**	**5**	**6**	**7**	**8**
HAN	SVM	RFE-SVM	0.6672	0.1290	0.1611	0.1115	0.1137	0.0046	0.0023
		LC	0.3809	0.1084	0.0137	0.0076	0.0053	0.0046	0.0061
		WLC	0.3809	0.1084	0.0137	0.0069	0.0053	0.0046	0.0061
	kNN	RFE-SVM	0.2649	0.1053	0.1122	0.1076	0.1168	0.0031	0.0015
		RFE-RF	0.2353	0.0626	0.0023	0.0023	0.0008	0.0015	0.0008
		LC	0.3809	0.1000	0.0038	0.0061	0.0053	0.0069	0.0053
		WLC	0.3089	0.1000	0.0038	0.0046	0.0053	0.0069	0.0053
	RF	RFE-RF	0.3344	0.0771	0.0130	0.0076	0.0053	0.0031	0.0038
BAK	SVM	RFE-SVM	1.4595	1.3000	1.0714	1.0929	0.8310	0.8571	0.7310
		LC	1.7071	1.3310	0.9595	0.7095	0.5571	0.5714	0.5000
		WLC	1.7071	1.3310	0.9095	0.7190	0.5381	0.5571	0.4357
	kNN	RFE-SVM	1.3310	1.0262	0.7857	0.7024	0.6619	0.6310	0.5429
		RFE-RF	1.2795	1.1731	0.7198	0.6169	0.6159	0.5855	0.4530
		LC	1.5571	1.0357	0.8810	0.5190	0.5214	0.4429	0.4286
		WLC	1.5571	1.0357	0.6238	0.5262	0.5214	0.4429	0.4095
	RF	RFE-RF	1.3654	1.0072	0.5995	0.5492	0.5769	0.5670	0.4293
FOS	SVM	RFE-SVM	0.1690	0.0718	0.0093	0.0000	0.0000	0.0000	0.0000
		LC	0.4259	0.2060	0.0255	0.0139	0.0139	0.0324	0.0069
		WLC	0.4259	0.2060	0.0255	0.0139	0.0069	0.0046	0.0139
	kNN	RFE-SVM	0.0347	0.0208	0.0185	0.0000	0.0000	0.0000	0.0000
		RFE-RF	0.1412	0.0635	0.0070	0.0023	0.0000	0.0000	0.0000
		LC	0.1481	0.0486	0.0000	0.0000	0.0000	0.0000	0.0000
		WLC	0.1481	0.0486	0.0000	0.0000	0.0000	0.0000	0.0000
	RF	RFE-RF	0.1247	0.0610	0.0213	0.0212	0.0094	0.0047	0.0024
MOD	SVM	RFE-SVM	5.8557	3.0992	0.9807	0.8128	0.7749	0.7261	0.6356
		LC	5.6549	2.6297	2.3243	2.2366	2.0830	1.2525	0.8206
		WLC	5.6549	2.6297	2.3243	2.2366	1.2845	1.2491	1.2242
	kNN	RFE-SVM	4.3986	2.3927	0.5299	0.4658	0.4437	0.4226	0.3535
		RFE-RF	5.5656	1.2235	0.6222	0.5445	0.4086	0.3506	0.2759
		LC	4.2404	1.9400	1.7273	1.6346	1.4966	0.8489	0.5320
		WLC	4.2404	1.9400	1.7273	1.6346	0.9797	0.8489	0.8560
	RF	RFE-RF	5.4313	1.3225	0.6679	0.6469	0.4575	0.3585	0.3496

9.4.2 BakRyan

The second network considered is the BakRyan network which has 35 nodes, 58 pipes and one reservoir, refer Lee and Lee (2001). The same feature selection methods along with classifiers as used in the Hanoi network were used on the BakRyan network to localize the leaks. Fig. 9.3 presents the accuracy as plots. It can be observed from Fig. 9.3 that the accuracy of classification for this network is always small compared to the Hanoi network. In addition, RFE with SVM as the classifier, $RFE-SVM$ performs better than other FS techniques, while with kNN as the classifier, the WLC approach outperformed the other techniques. As opposed to the observations made in the Hanoi network however, the performance of the LC and WLC approaches are different in this network, suggesting that they identify different sensor locations for the same localization problem. In addition, it can also be observed that the performance of the WLC approach is better than the LC as well as RFE methods, when used with SVM and kNN as classifiers. This suggests that WLC is a more informative and strategically superior manner of identifying the locations of sensors to be place in the network. Similar observations can also be made regarding the ATD of different $CL-FS$ combinations. kNN performs better than SVM and RF for all the cases studied for this network.

9.4.3 Fossolo

The third network considered in this chapter is the Fossolo network, refer Bragalli and D'Ambrosio (2012), which is a highly interconnected network (Fig. 9.1). The Fossolo network comprises 36 nodes, 58 pipes and 1 reservoir. The performance of the feature selection methods and classifiers in leak classification on the Fossolo network are presented in Fig. 9.4. Similar to the BakRyan network, the classification accuracy increases consistently with the increase in the number of sensors placed (features selected). However, the accuracy of classification is comparable with that obtained for the Hanoi network. The kNN is observed to perform exceptionally well in classifying leaks using $RFE-SVM$, LC and WLC as the FS techniques, with accuracies between $99.5\%-100\%$ for $N>4$ for this network. These results are further supported by the

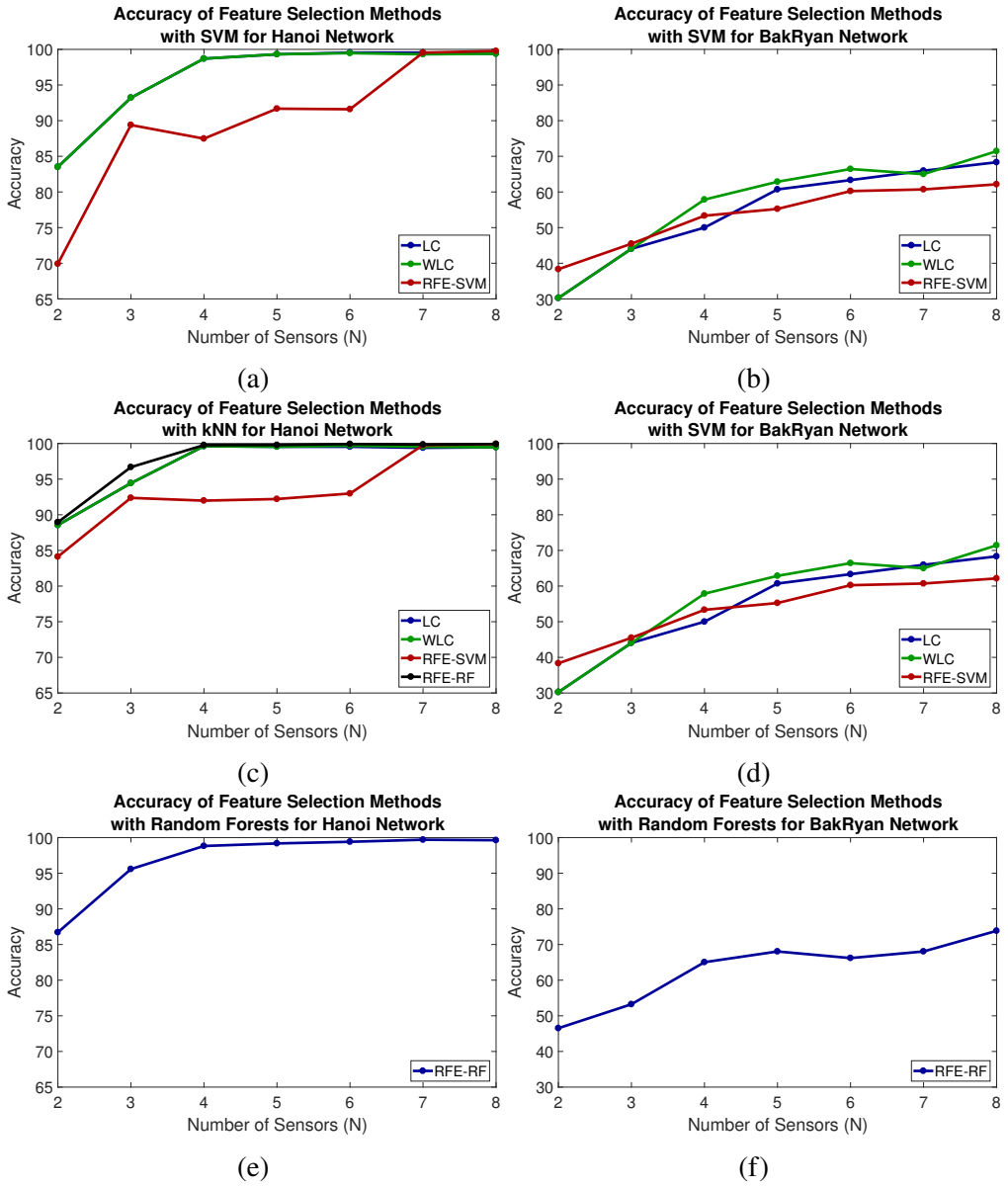


Figure 9.3: Comparison of the performance of different leak localisation techniques for Hanoi and BakRyan networks. The plots on the left show the accuracy of localisation obtained with a classifier for the Hanoi WDN while the plots on the right show the results for BakRyan network.

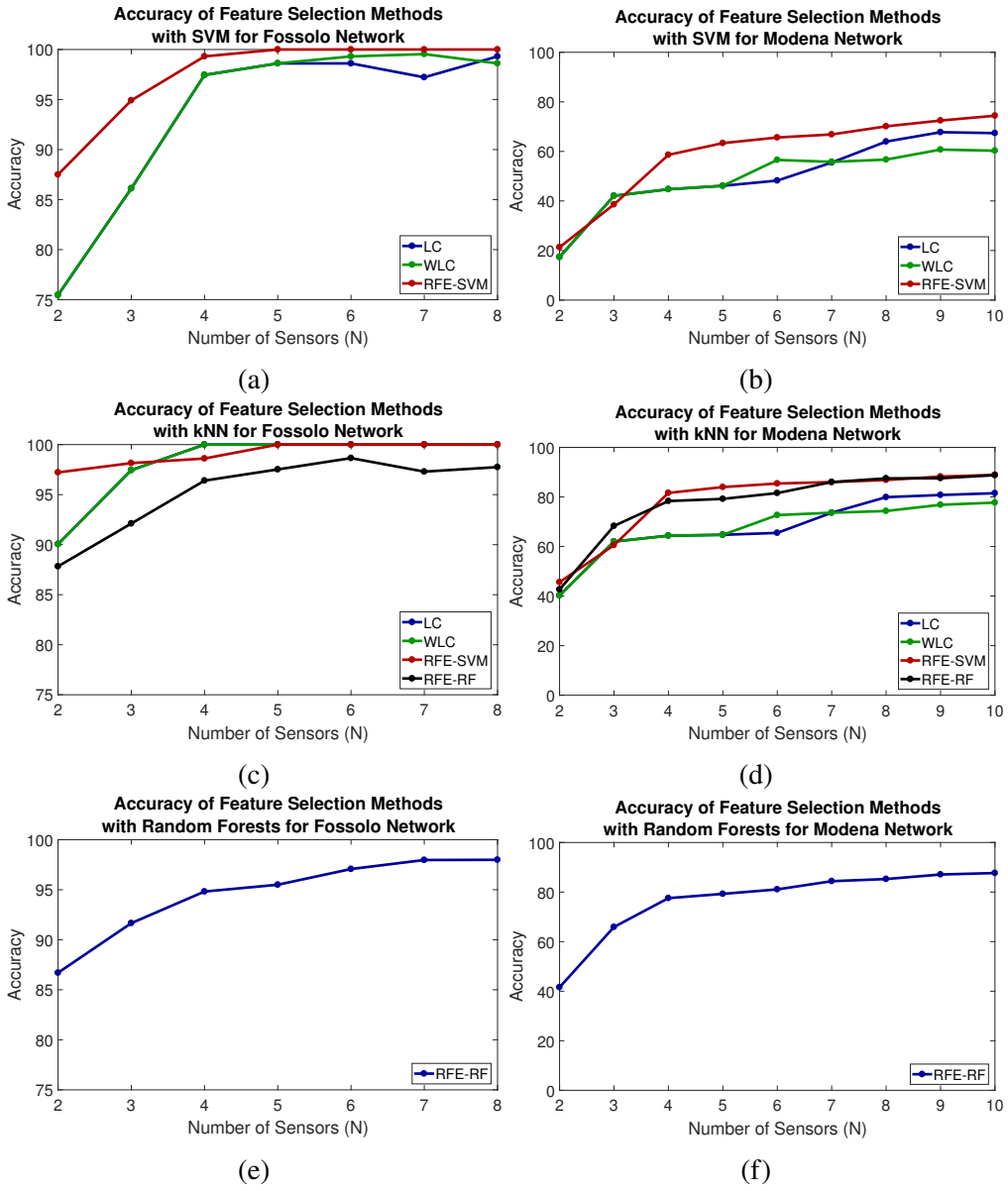


Figure 9.4: Comparison of the performance of different leak localisation techniques for Fossolo and Modena WDNs. The plots on the left show the accuracy of localisation obtained with a classifier for the Fossolo Network, while the plots on the right show the plots for the Modena network.

values of ATD obtained with for the different combinations. Specifically, the ATD for all feature selection methods for $N \geq 4$ was observed to be in the range $[0, 0.0325]$. Also, high F1-scores (> 0.97) was obtained when more than 4 sensors are considered in all the FS methods.

The three networks considered here are small networks with different configurations of pipelines for approximately the same number of nodes. In order to study the applicability of the proposed approach to larger networks, a similar study is performed with a much larger network, the results of which are described in the following section.

9.4.4 Modena

The final network considered in this work is the Modena network, refer Bragalli and D'Ambrosio (2012) that consists of 268 nodes, 317 pipes and 4 reservoirs, representing a significant increase in size compared to the other networks considered in this work. Fig. 9.4 shows the accuracies of the different $FS - CL$ combinations studied for this network. It can be observed from Fig. 9.4 that there is a significant increase in the performance of the methods as the number of sensors is increased till $n = 4$, after which the percentage increase in performance per sensor added to the network decreases. The $RFE - SVM$ provides good performance with SVM as well as kNN as the classifier (74.4% and 88.87% for $N = 10$), while RF provides accuracy of 88.78%.

For the Modena network, there is a significant increase in classification accuracy as the number of nodes increase to 4 nodes and later the increment slows down. This behaviour is consistent for all combinations of classifier and feature selection methods. For SVM as classifier, RFE-SVM has the highest average classification accuracy and lowest average ATD score corresponding to different number of nodes used. For KNN as classifier, RFE-SVM outperforms other feature selection algorithms followed by REF-RF. Overall, KNN as classifier outperforms SVM as classifier for different feature selection algorithms.

Table 9.5: Comparison of the accuracy of leak localisation technique - classifier combinations for the Modena network.

FS	Classifier	Number of Sensors (N)						
		15	**25**	**35**	**45**	**55**	**65**	**75**
RFE-SVM	SVM	79.01	83.15	84.61	86.97	88.62	88.00	87.62
	kNN	91.23	94.15	94.12	95.52	95.86	96.42	96.42
RFE-RF	RF	90.64	92.56	93.80	94.17	95.01	95.63	95.57
	kNN	91.23	93.03	94.11	95.23	95.57	96.19	96.20
LC	SVM	73.29	82.25	83.99	86.26	91.20	91.11	90.11
	kNN	84.48	89.71	91.88	93.03	94.15	94.53	95.02
WLC	SVM	73.41	79.79	82.74	83.83	84.89	86.26	89.89
	kNN	82.84	89.99	91.17	91.57	93.72	94.31	95.18

These $FS - Classifier$ methods were also used to analyse the network for higher values of N, where $N \in \{15, 25, 35, 45, 55, 65, 75\}$. For the Modena network placing 75 sensors is approximately placing sensors at 25% of the total number of nodes. These results are presented in table 9.5. It is observed that unlike the accuracy plots in Fig. 9.4, the accuracies of the $RFE - SVM$, LC and WLC methods become nearly equal for higher values of N. Also, $N = 75$ which is approximately 25% of the sensors provides high classification accuracies in the range $95\% - 96.5\%$ when the FS methods are used with kNN.

9.5 DISCUSSIONS

The results shown in the previous section demonstrate the potential of the proposed approach for detection and localization of leaks in water distribution networks. The networks considered in this chapter exhibit markedly different characteristics. For example, while the first three networks are fairly small networks with comparable number of nodes with different configurations, the final network is a large network with the number of nodes that is an order of magnitude larger than the others. Furthermore, the small networks, i.e., Hanoi, BakRyan and Fossolo networks are unique in the way their nodes are interconnected by pipes. Specifically, while the Hanoi network has chains of nodes with a degree of 2 with only six of the 31 nodes having degree greater than 2, the

BakRyan network has very few nodes with a degree of 2, and has multiple edges and small loops. The Fossolo network, on the other hand has a high average degree of 3.22 with nearly all nodes being degree 3. These configurations impart the networks with different abilities, that manifest as different patterns in the data collected from different nodes under normal and leaky conditions. For example, the Fossolo network exhibits a large number of loops in the network, which allows the network to compensate for leaks in one node by supplying the extra demand from adjacent nodes. This results in a change in the observed pressure at the adjacent nodes and this effect propagates through the network. Therefore, even small leaks have perceivable effects on the pressure at other nodes of the network. As a result, this network has an overall tendency to be amenable towards detection and localization of leaks. The BakRyan network, on the other hand has parallel lines as a result of which, the leaks might remain undetected at several locations.

The proposed framework consisting of entirely data driven approaches with techniques of statistical machine learning, the variabilities between different networks as manifested in the data are automatically incorporated into the localization technique. The results further highlight this feature by being able to identify a small number of sensor locations that deliver a high accuracy of localization. For example, the localization accuracy for Hanoi and Fossolo networks were observed to be $\sim 99\%$ with an $ATD < 1$ with sensors placed at only 8 nodes ($\sim 25\%$ of the total number of nodes in the network) throughout the geographically distributed network. This signifies that the proposed technique is able to localize the leaks, on an average to within one node in the network with only 8 sensors. The performance of BakRyan network was however poor compared to the other networks, predominantly because of the presence of parallel lines that can mask the effect of leaks in the data. Furthermore, the accuracy of localization for the large network (with a total of 268 nodes) with only 10 sensors ($\sim 4\%$ of the total number of nodes) was close to 88%, which increased to 96% with 75 sensors ($\sim 25\%$ of the total number of nodes).

It can further be observed from the results that the kNN, in general delivers a better

performance than SVM as a classifier, irrespective of the feature selection technique used. A probable reason of this can be the high resemblance in pressure profile of nodes upon introduction of leaks at different locations, which matches with the ideology of kNN. The proposed feature selection algorithms, LC and WLC perform better for Hanoi and BakRyan networks and comparable for Fossolo and Modena networks.

9.6 SUMMARY

Leaks in water distribution networks contribute significantly to the non-revenue water. Leak localization in WDNs is therefore crucial for reducing the losses incurred by the water utilities in meeting the demands of its consumers. In this chapter, the leak localization problem is addressed with a statistical machine learning approach. A framework is proposed for leak detection and localization along with two feature selection methods. Different classifiers, namely, SVM, kNN and random forests classifiers have been used to classify the leaks. kNN was found to perform better than the other classifiers in all the performance metrics utilized.

The proposed feature selection methods were applied on four networks having different configurations to check the validity and generality of the approach. The methods performed better than the existing methods in literature. The different configurations of the networks and hence, the patterns observed in the data impart a large degree of diversity to the study and thus greater reliability to the applicability of the framework proposed in the chapter. The networks were chosen such that a variety of combinations could be studied which include: same number of nodes but having different connectivity between the nodes; with and without multiple edges in the network; low vs. high interconnectivity; with and without long chains of nodes; different total demands; and small vs. large network in terms of number of nodes. From this study, it was observed that the structure of the networks in terms of connectivity plays a significant role in the leak localizability. Multiple edges, while adding redundancy to the network, reduce the

localizability of the leaks. Whereas, introducing more interconnectivity in a network increases the leak localizability.

A comparison of the overall proposed framework of $FS-CL$ approach based on statistical machine learning with the literature also demonstrated the superiority of the proposed framework. An analysis of misclassifications was made to identify methods to improve the framework. One such observation was made in the leak localization in the presence of long chains of nodes in a network. In such networks, it was observed that adjacent nodes in a long chain of nodes are highly correlated in their pressure measurements, which resulted in high misclassifications when feature selection methods such as $RFE-SVM$ were utilized which do not capture these network dynamics. In such cases the proposed method was observed to outperform $RFE-SVM$. Future extensions of this work is to extend the framework to localize multiple simultaneous leaks, to introduce a multiple-objective feature selection approach to minimize misclassifications and maximize classification accuracy and to include networks with active elements such as pumps.

CHAPTER 10

CONCLUSIONS AND FUTURE WORK

This chapter summarizes the contributions of this dissertation and possible future research directions.

10.1 CONCLUSIONS

This dissertation addresses the problem of obtaining reliable state estimates under low-observability conditions, a condition typical in power distribution systems. It also addresses other research problems in distribution systems, such as topology identification, meter placement, and error analysis, to aid the developed state estimators.

- Chapter 3 develops model-based distribution system state estimators that operate even under low-observability conditions, which is typical of distribution systems. The state estimation problem was formulated as low-rank tensor completion-based approaches with power flow equations as constraints. The developed techniques utilize structures within the data (i.e., spatial, temporal, and physics-based dependencies) to provide state estimates with low state estimation errors even under low-observability conditions. Various enhancements were introduced to reduce state estimation errors. Firstly, the need for feature scaling is highlighted. It is shown that feature scaling reduces state estimation errors by improving the sensitivity of the nuclear norm to all features and weighting the errors for each physical quantity. Next, it was also identified that certain unfoldings of the tensor are more informative than others. Therefore, assigning higher weights to more informative structures reduced state estimation errors. Finally, the power flow equations introduced in the state estimator as constraints reduced the solution space and ensured that the solutions conform to the power flow equations.

- Chapter 4 tackled the problem of high computational time requirements of model-based approaches. Graph neural network equivalents to model-based approaches that transfer the computational burden to the training rather than the evaluation phase were developed. Feature scaling module was introduced to improve the sensitivity of each physical quantity towards state estimate errors and to weigh the errors in the estimation of each physical quantity. A pseudo-measurement generation module was introduced to improve data availability at each stage in the state estimation process. Ablation studies confirmed the effectiveness of these modules. The developed GNN-based state estimators provided state estimates with state estimation errors comparable to the model-based approaches while using several orders of magnitude lower computational times. Interstingly, the GNN-based approaches could conform to the power flow equations by learning the interdependencies without any prior knowledge of the network parameters.

- Chapter 5 addresses the problem of model selection in model-based approaches and topology detection in model-free approaches by developing GNN-based topology identification techniques. The developed technique uses link prediction to identify switch statuses, thus, the network topology. This approach avoids the conventional estimator pool techniques, which evaluate all possible topology configurations to determine the topology which best fits the observed data.

- Chapter 6, for the first time, quantifies the errors of model-based estimators. The power flow equations used as constraints in the state estimators are used to develop error bounds for the state estimates. Further, it was identified that hyperparameter selection is a critical step in minimizing state estimation errors. Therefore, an algorithm was developed that alternatively updates the state estimates and error scaling factors to minimize state estimation errors.

- Chapter 7 exploits the interdependencies among different entries in the state

measurement matrices to pose the meter placement problem as an element selection problem in the state measurement matrices. The developed approaches utilize a heuristic to add sensors that minimize the voltage residual in the selected reference cases. These approaches also enable extending the existing set of metering equipment with additional sensors to reduce state estimation errors.

- Chapter 8 improves the compressibility of smart meter data by tailoring transformation matrices to the data using dictionary learning. The developed dictionaries result in greater data compressibility without loss in reconstruction performance. The data-driven dictionaries also provide superior state estimation performance when used as the transformation basis in compressive sensing-based state estimators. These dictionaries outperformed existing deterministic dictionaries such as Haar and Hankel, which are typically used with compressive sensing-based distribution system state estimators.

- Chapter 9 explores other cyber-physical systems, namely water distribution systems, for related problems. It was identified that non-revenue water is a key cause of concern for utilities. Therefore, leak localization strategies were developed that use pressure measurements at various points in the network to identify leak locations. Various classifiers, such as support vector machines, k-nearest neighbors, and random forests, were utilized to identify the leak locations. The meter placement problem was converted into a feature selection problem, and a least correlation-based heuristic approach was developed to identify sets of sensors that maximize the information available to these classifiers.

10.2 FUTURE WORK

The state estimators and techniques developed for solving their associated problems open up the possibilities for future work highlighted below.

- Chapter 3 assumes that all the network parameters of the system are known. However, utilities may have insufficient or unreliable network parameters. Therefore, updating the constraints to only account for the known parameters can provide a trade-off between model-free and model-based approaches and enable the use of partial knowledge of network parameters.

- The GNN-based state estimators developed in Chapter 4 assume the network topology is known. However, future work may combine it with the topology identification technique developed in Chapter 5 to create a joint topology and state estimator.

- Given the size of some real-world distribution systems, centralized state estimation, as developed in Chapters 3 and 4, suffer from implementation issues due to the size of the networks. Therefore, distributed implementations of these state estimators would result in greater acceptance by utilities.

- Error bounds developed in Chapter 6 are loose bounds due to the expansions of the Frobenius norm terms. Future work will be aimed at tightening the error bounds. These bounds could potentially have other applications, such as optimal meter placement.

- The data-driven dictionaries developed in Chapter 8 are assumed to remain unchanged over time. However, over long durations, newer devices and consumer load patterns could reduce the effectiveness of data-driven dictionaries. Therefore, a mechanism to sample some metering locations and update the data-driven dictionaries over time could help reduce reconstruction and state estimation errors.

APPENDIX A

POWER DISTRIBUTION SYSTEMS

A.1 IEEE 33-NODE DISTRIBUTION SYSTEM

The IEEE 33-node distribution system is a hypothetical test system operating at 12.66kV with 32 PQ buses. The network diagram is shown in Fig. A.1. The network parameters for this system are available in Baran and Wu (1989).

A.2 100-NODE TEST SYSTEM

The IEEE 33-node distribution system is extended to form a 100-node distribution system to study the applicability for larger systems. The network diagram for this system is provided in Fig. A.2. This system was created by adding lateral feeders and nodes to the IEEE 33-node distribution system. The nominal load at each primary node is distributed among the corresponding lateral nodes connected to that node. The line resistance and reactance of the lateral feeders are assumed to be 0.08 p.u. and 0.04 p.u. resistances, respectively.

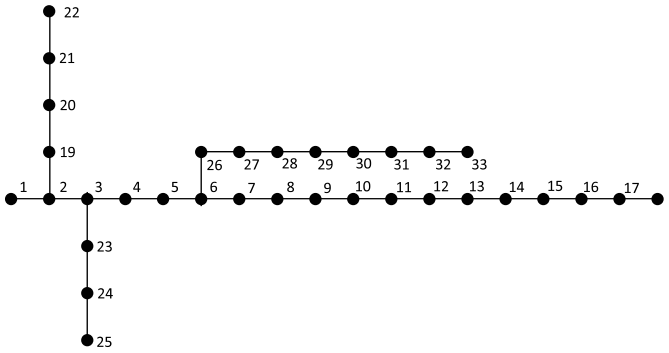


Figure A.1: IEEE 33-node distribution system.

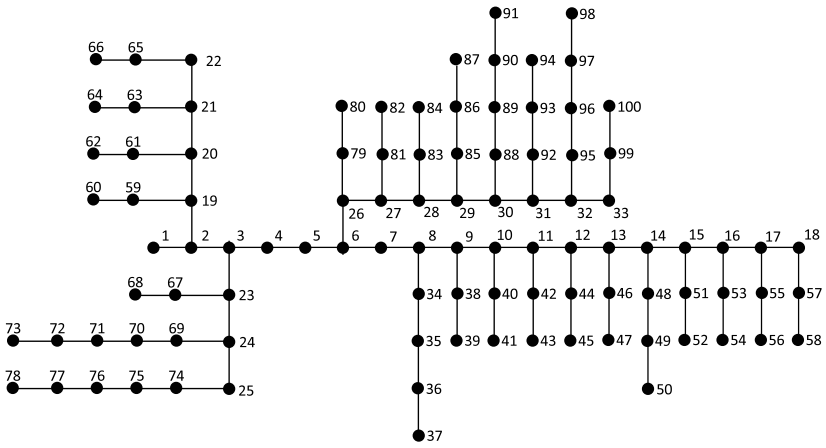


Figure A.2: 100-node distribution system.

A.3 IEEE 37-NODE DISTRIBUTION SYSTEM

The IEEE 37-node distribution system is an actual feeder in California, with a 4.8 kV operating voltage. The network details are available in IEEE (2017). The network diagram is shown in Fig. A.3. Additional switches used to evaluate topology identification techniques are marked in red.

A.4 559-NODE DISTRIBUTION SYSTEM

Malekpour and Pahwa (2015) developed a 559-node radial three-phase distribution system. This system extends the IEEE 37-node distribution system with secondary and tertiary branches. This system has been used to evaluate the applicability of proposed approaches on large networks.

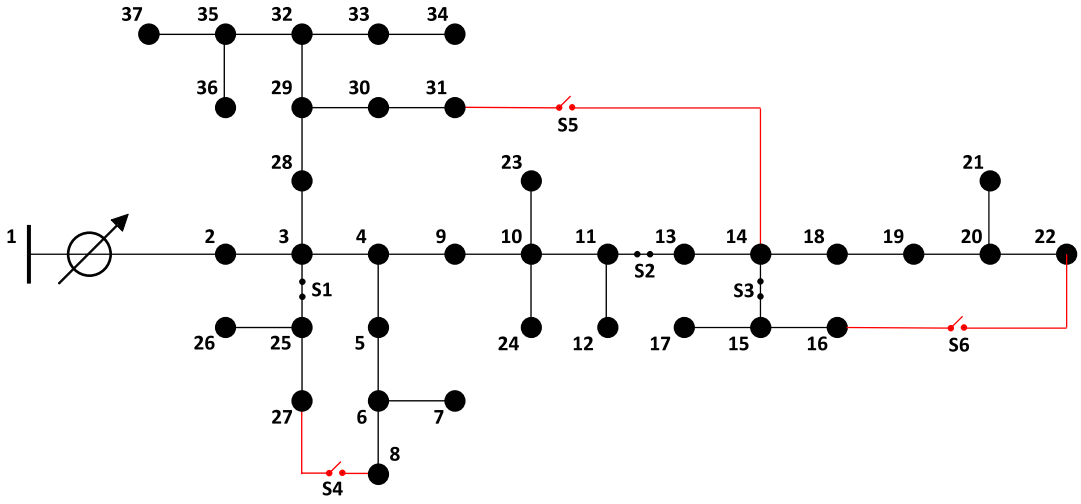


Figure A.3: IEEE 37-node distribution system.

APPENDIX B

COMPRESSIVE SENSING: TRANSFORMATION MATRICES

Transformation matrices are dictionaries used in compressive sensing to generate sparse representations of signals. These matrices can be categorized as deterministic and non-deterministic dictionaries. Deterministic dictionaries are generated by sampling transformation functions at specific points. Whereas, non-deterministic dictionaries are generated based on features from sample data.

B.1 COMPLETE DICTIONARIES

B.1.1 DCT Matrix

The DCT matrix is a series of data points expressed as cosine functions oscillating at different frequencies. It is a widely used transformation in signal processing and data compression.

B.1.2 Haar Matrix

The Haar transformation matrix enables the expression of local changes in the signal. It is a widely used transformation for signal and image compression. The Haar transformation matrix is derived from the Haar function.

B.1.3 Hadamard Matrix

The Hadamard matrix is a square matrix generated using only the entries +1 and -1 such that their rows are mutually orthogonal. The orthonormal variant of the Hadamard matrix is defined below

$$H_1 = [1] \tag{B.67}$$

$$H_n = \frac{1}{\sqrt{2}} \begin{bmatrix} H_{n-1} & H_{n-1} \\ H_{n-1} & -H_{n-1} \end{bmatrix} \tag{B.68}$$

B.1.4 Hankel Matrix

The Hankel matrix is defined to be a square matrix in which all entries in each skew-diagonal are equal. The elements of each skew-diagonal can be selected such that combinations of columns of the Hankel matrix can be used to generate step and ramp waveforms. Since power consumption data typically has step and ramp changes due to switching operations of connected devices, the Hankel matrix is expected to perform well as a sparsifying basis. The Hankel matrix generated from a vector $a \in \mathbb{R}^{(2n-1)}$ is shown below.

$$Hankel(a) = \begin{bmatrix} a_1 & a_2 & \dots & a_{n-1} & a_n \\ a_2 & a_3 & \dots & a_n & a_{n+1} \\ \dots & \dots & \dots & \dots & \dots \\ a_{n-1} & a_n & \dots & a_{2n-3} & a_{2n-2} \\ a_n & a_{n+1} & \dots & a_{2n-2} & a_{2n-1} \end{bmatrix} \tag{B.69}$$

B.1.5 Toeplitz Matrix

The Toeplitz matrix is a square matrix such that the elements within each diagonal are equal. These matrices can be generated from Hankel matrix using exchange matrices and thus are also expected to perform well as a sparsifying basis for power consumption data. The Toeplitz matrix generated from a vector $a \in \mathbb{R}^{(2n-1)}$ is shown below.

$$Toeplitz(a) = \begin{bmatrix} a_n & a_{n+1} & \dots & a_{(2n-2)} & a_{2n-1} \\ a_{n-1} & a_n & \dots & a_{2n-3} & a_{(2n-2)} \\ \dots & \dots & \dots & \dots & \dots \\ a_2 & a_3 & \dots & a_n & a_{n+1} \\ a_1 & a_2 & \dots & a_{n-1} & a_n \end{bmatrix} \tag{B.70}$$

B.2 OVER-COMPLETE DICTIONARIES

Over-complete dictionaries have more atoms than the dimension of the vector space. This redundancy in transformation matrix columns aims to create sparser representations of the signal than the complete dictionaries. Deterministic over-complete dictionaries can be generated by using combinations of deterministic complete dictionaries.

BIBLIOGRAPHY

1. **Aharon, M.**, **M. Elad**, and **A. Bruckstein** (2006). K-svd: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing*, **54**(11), 4311–4322.

2. **Alam, S. M.**, **B. Natarajan**, and **A. Pahwa** (2014). Distribution grid state estimation from compressed measurements. *IEEE Transactions on Smart Grid*, **5**(4), 1631–1642.

3. **Ali, S.**, **K. Wu**, **K. Weston**, and **D. Marinakis** (2016). A machine learning approach to meter placement for power quality estimation in smart grid. *IEEE Transactions on Smart Grid*, **7**(3), 1552–1561.

4. **Alvarez, M. J.**, **F. S. Sellschopp**, and **E. Vazquez** (2015). A PMUs placement methodology based on inverse of connectivity and critical measurements. *International Journal of Electrical Power and Energy Systems*, **68**, 336–344. ISSN 0142-0615.

5. **Arce, G. R.**, **D. J. Brady**, **L. Carin**, **H. Arguello**, and **D. S. Kittle** (2014). Compressive coded aperture spectral imaging: An introduction. *IEEE Signal Processing Magazine*, **31**(1), 105–115.

6. **Babu, R.** and **B. Bhattacharyya** (2016). Optimal allocation of phasor measurement unit for full observability of the connected power network. *International Journal of Electrical Power and Energy Systems*, **79**, 89–97. ISSN 0142-0615.

7. **Balzano, L.** and **S. J. Wright** (2013). On GROUSE and incremental SVD. *2013 5th IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing, CAMSAP 2013*, 1–4. URL `https://ieeexplore.ieee.org/document/6713992`.

8. **Baran, M. E.**, **J. Jung**, and **T. E. McDermott**, Topology error identification using branch current state estimation for distribution systems. *In 2009 Transmission Distribution Conference Exposition: Asia and Pacific*. 2009.

9. **Baran, M. E.** and **F. F. Wu** (1989). Network reconfiguration in distribution systems for loss reduction and load balancing. *IEEE Transactions on Power Delivery*, **4**, 1401–1407. ISSN 19374208.

10. **Batra, N.**, **M. Gulati**, **A. Singh**, and **M. B. Srivastava**, It's different: Insights into home energy consumption in india. *In Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings*, BuildSys'13. Association for Computing Machinery, New York, NY, USA, 2013. ISBN 9781450324311. URL `https://doi.org/10.1145/2528282.2528293`.

11. **Berardi, L.**, **O. Giustolisi**, **Z. Kapelan**, and **D. Savic** (2008). Development of pipe deterioration models for water distribution systems using epr. *Journal of Hydroinformatics*,

10(2), 113–126.

12. **Berg Insight** (2021*a*). More than 570 million smart electricity meters to be deployed in asian markets until 2025. URL https://www.berginsight.com/more-than-570-million-smart-electricity-meters-to-be-

13. **Berg Insight** (2021*b*). Smart metering in europe. URL https://www.berginsight.com/smart-metering-in-europe.

14. **Berglund, A.**, **V. S. Areti**, and **G. Mahinthakumar** (2017). Successive linear approximation methods for leak detection in water distribution systems. *Journal of Water Resources Planning and Management*, **143**(8), 04017042.

15. **Bernstein, A.** and **E. Dall'Anese**, Linear power-flow models in multiphase distribution networks. *In 2017 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*. 2017.

16. **Bragalli, C.** and **C. D'Ambrosio** (2012). On the optimal design of water distribution networks: a practical minlp approach. *Optimization and Engineering*, **13**(2), 219–246.

17. **Cai, J.-F.**, **E. J. Candès**, and **Z. Shen** (2010). A Singular Value Thresholding Algorithm for Matrix Completion. *http://dx.doi.org/10.1137/080738970*, **20**(4), 1956–1982.

18. **Candès, E.** and **J. Romberg** (2007). Sparsity and incoherence in compressive sampling. *Inverse Problems*, **23**(3), 969–985. URL https://doi.org/10.1088/0266-5611/23/3/008.

19. **Candes, E.**, **J. Romberg**, and **T. Tao** (2006). Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, **52**(2), 489–509.

20. **Candès, E. J.** and **T. Tao** (2010). The power of convex relaxation: Near-optimal matrix completion. *IEEE Transactions on Information Theory*, **56**(5), 2053–2080.

21. **Chakraborty, S.**, **J. Zhou**, **V. Balasubramanian**, **S. Panchanathan**, **I. Davidson**, and **J. Ye**, Active matrix completion. *In 2013 IEEE 13th International Conference on Data Mining*. 2013.

22. **Chan, T.**, **C. S. Chin**, and **X. Zhong** (2018). Review of current technologies and proposed intelligent methodologies for water distributed network leakage detection. *IEEE Access*, **6**, 78846–78867.

23. **Chauhan, K.** and **R. Sodhi** (2020). Placement of distribution-level phasor measurements for topological observability and monitoring of active distribution networks. *IEEE Transactions on Instrumentation and Measurement*, **69**(6), 3451–3460.

24. **Chen, Q.**, **D. Kaleshi**, **Z. Fan**, and **S. Armour** (2016). Impact of smart metering data

aggregation on distribution system state estimation. *IEEE Transactions on Industrial Informatics*, **12**(4), 1426–1437.

25. **Chen, T.**, **Y. Cao**, **X. Chen**, **L. Sun**, **J. Zhang**, and **G. A. Amaratunga** (2021). Optimal pmu placement approach for power systems considering non-gaussian measurement noise statistics. *International Journal of Electrical Power and Energy Systems*, **126**, 106577. ISSN 0142-0615. URL https://www.sciencedirect.com/science/article/pii/S0142061520318469.

26. **Collins, R.** and **J. Boxall** (2012). Influence of ground conditions on intrusion flows through apertures in distribution pipes. *Journal of Hydraulic Engineering*, **139**(10), 1052–1061.

27. **Dahale, S.**, **H. S. Karimi**, **K. Lai**, and **B. Natarajan** (2020). Sparsity based approaches for distribution grid state estimation - a comparative study. *IEEE Access*, **8**, 198317–198327.

28. **Dahale, S.** and **B. Natarajan**, Joint matrix completion and compressed sensing for state estimation in low-observable distribution system. *In 2021 IEEE PES Innovative Smart Grid Technologies Conference - Latin America (ISGT Latin America)*. 2021.

29. **De Oliveira Rocha, H. R.**, **J. C. S. De Souza**, and **M. B. Do Coutto Filho** (2013). Planning high quality metering systems for state estimation through a constructive heuristic. *International Journal of Electrical Power and Energy Systems*, **52**(1), 34–41. ISSN 0142-0615.

30. **Dehghanpour, K.**, **Z. Wang**, **J. Wang**, **Y. Yuan**, and **F. Bu** (2019). A survey on state estimation techniques and challenges in smart distribution systems. *IEEE Transactions on Smart Grid*, **10**(2), 2312–2322.

31. **Ding, Q.** and **V. A. Emesih** (2007). A simple factorization-based observability analysis and meter placement method. *International Journal of Electrical Power and Energy Systems*, **29**(10), 731–737. ISSN 0142-0615.

32. **Donoho, D. L.** (2006). Compressed sensing. *IEEE Transactions on Information Theory*, **52**, 1289–1306. ISSN 00189448.

33. **egauge** (2018*a*). Egauge center. URL http://egauge360.egaug.es/.

34. **egauge** (2018*b*). Monitor electricity on every circuit with precision and accuracy. URL https://www.egauge.net/.

35. **El-hawary, M. E.** (2014). The smart grid—state-of-the-art and future trends. *Electric Power Components and Systems*, **42**(3-4), 239–250. URL https://doi.org/10.1080/15325008.2013.868558.

36. **Ferreira, D. M.**, **P. M. Carvalho**, and **L. A. Ferreira** (2020). Optimal meter placement in low observability distribution networks with der.

Electric Power Systems Research, **189**, 106707. ISSN 0378-7796. URL https://www.sciencedirect.com/science/article/pii/S0378779620305101.

37. **Fujiwara, O.** and **D. B. Khang** (1990). A two-phase decomposition method for optimal design of looped water distribution networks. *Water Resources Research*, **26**(4), 539–549. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/WR026i004p00539.

38. **Ghadikolaee, E. T.**, **A. Kazemi**, and **H. A. Shayanfar** (2020). Novel multi-objective phasor measurement unit placement for improved parallel state estimation in distribution network. *Applied Energy*, **279**, 115814. ISSN 0306-2619. URL https://www.sciencedirect.com/science/article/pii/S0306261920312940.

39. **Ghasemi Damavandi, M.**, **V. Krishnamurthy**, and **J. R. Martí** (2015). Robust meter placement for state estimation in active distribution systems. *IEEE Transactions on Smart Grid*, **6**(4), 1972–1982.

40. **Goldfarb, D.** and **Z. T. Qin** (2014). Robust low-rank tensor recovery: Models and algorithms. *SIAM Journal on Matrix Analysis and Applications*, **35**(1), 225–253. ISSN 1095-7162.

41. **Gou, B.** (2008). Generalized integer linear programming formulation for optimal pmu placement. *IEEE Transactions on Power Systems*, **23**(3), 1099–1104.

42. **Graff, C. G.** and **E. Y. Sidky** (2015). Compressive sensing in medical imaging. *Applied optics*, **54**(8), C23–C44. ISSN 1539-4522. URL https://doi.org/10.1364/AO.54.000C23.

43. **Grant, M.** and **S. Boyd**, Graph implementations for nonsmooth convex programs. *In* **V. Blondel**, **S. Boyd**, and **H. Kimura** (eds.), *Recent Advances in Learning and Control*, Lecture Notes in Control and Information Sciences. Springer-Verlag Limited, 2008, 95–110.

44. **Grant, M.** and **S. Boyd** (2014). CVX: Matlab software for disciplined convex programming, version 2.1. http://cvxr.com/cvx.

45. **Gupta, P.** and **P. R. Kumar** (2000). The capacity of wireless networks. *IEEE Transactions on Information Theory*, **46**(2), 388–404.

46. **Hayes, B.** and **M. Prodanovic**, State estimation techniques for electric power distribution systems. *In 2014 European Modelling Symposium*. 2014.

47. **Huang, Y.-F.**, **S. Werner**, **J. Huang**, **N. Kashyap**, and **V. Gupta** (2012). State estimation in electric power grids: Meeting new challenges presented by the requirements of the future grid. *IEEE Signal Processing Magazine*, **29**(5), 33–43.

48. **IEEE** (2017). IEEE PES Test Feeder. URL

https://cmte.ieee.org/pes-testfeeders/.

49. **Joshi, A.**, **L. Das**, **B. Natarajan**, and **B. Srinivasan** (2019). A framework for efficient information aggregation in smart grid. *IEEE Transactions on Industrial Informatics*, **15**(4), 2233–2243.

50. **Jung, D.** and **K. Lansey** (2014). Water distribution system burst detection using a nonlinear kalman filter. *Journal of Water Resources Planning and Management*, **141**(5), 04014070.

51. **Kang, J.**, **Y.-J. Park**, **J. Lee**, **S.-H. Wang**, and **D.-S. Eom** (2018). Novel leakage detection by ensemble cnn-svm and graph-based localization in water distribution systems. *IEEE Transactions on Industrial Electronics*, **65**(5), 4279–4289.

52. **Karimi, H. S.** and **B. Natarajan**, Compressive sensing based state estimation for three phase unbalanced distribution grid. *In GLOBECOM 2017 - 2017 IEEE Global Communications Conference*. 2017.

53. **Karimi, H. S.** and **B. Natarajan** (2021). Joint Topology Identification and State Estimation in Unobservable Distribution Grids. *IEEE Transactions on Smart Grid*.

54. **Karray, F.**, **A. Garcia-Ortiz**, **M. W. Jmal**, **A. M. Obeid**, and **M. Abid** (2016). Earnpipe: A testbed for smart water pipeline monitoring using wireless sensor network. *Procedia Computer Science*, **96**, 285–294.

55. **Kelly, J.** and **W. Knottenbelt** (2015). The uk-dale dataset, domestic appliance-level electricity demand and whole-house demand from five uk homes. *Scientific Data*, **2**(1), 150007. ISSN 2052-4463. URL https://doi.org/10.1038/sdata.2015.7.

56. **Kipf, T. N.** and **M. Welling**, Semi-supervised classification with graph convolutional networks. *In International Conference on Learning Representations (ICLR)*. 2017.

57. **Kolter, J. Z.** and **M. J. Johnson**, Redd: A public data set for energy disaggregation research. *In SustKDD workshop on Data Mining Applications in Sustainability*. 2011.

58. **Kundacina, O.**, **M. Cosovic**, **D. Miskovic**, and **D. Vukobratovic** (2022*a*). Distributed nonlinear state estimation in electric power systems using graph neural networks. URL https://arxiv.org/abs/2207.11465.

59. **Kundacina, O.**, **M. Cosovic**, and **D. Vukobratovic** (2022*b*). State estimation in electric power systems leveraging graph neural networks. URL https://arxiv.org/abs/2201.04056.

60. **Laucelli, D.**, **M. Romano**, **D. Savić**, and **O. Giustolisi** (2016). Detecting anomalies in water distribution networks using epr modelling paradigm. *Journal of Hydroinformatics*, **18**(3), 409–427.

61. **Lee, S.-C.** and **S.-I. Lee** (2001). Genetic algorithms for optimal augmentation of water distribution networks. *Journal of Korea Water Resources Association*, **34**.

62. **Liao, W.**, **B. Bak-Jensen**, **J. R. Pillai**, **Y. Wang**, and **Y. Wang** (2022). A review of graph neural networks and their applications in power systems. *Journal of Modern Power Systems and Clean Energy*, **10**(2), 345–360.

63. **Lin, Z.**, **M. Chen**, and **Y. Ma** (2010). The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. *arxiv*.

64. **Liu, B.**, **H. Wu**, **Y. Zhang**, **R. Yang**, and **A. Bernstein** (2019*a*). Robust Matrix Completion State Estimation in Distribution Systems. *IEEE Power and Energy Society General Meeting*. ISSN 19449933.

65. **Liu, H.**, **Y. Li**, **M. Tsang**, and **Y. Liu** (2019*b*). CoSTCo: A Neural Tensor Completion Model for Sparse Tensors. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.

66. **Liu, J.**, **P. Musialski**, **P. Wonka**, and **J. Ye** (2013). Tensor completion for estimating missing values in visual data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **35**(1), 208–220. ISSN 01628828.

67. **Liu, J.**, **F. Ponci**, **A. Monti**, **C. Muscas**, **P. A. Pegoraro**, and **S. Sulis** (2014). Optimal meter placement for robust measurement systems in active distribution grids. *IEEE Transactions on Instrumentation and Measurement*, **63**(5), 1096–1105.

68. **Liu, J.**, **J. Tang**, **F. Ponci**, **A. Monti**, **C. Muscas**, and **P. A. Pegoraro** (2012). Trade-offs in pmu deployment for state estimation in active distribution grids. *IEEE Transactions on Smart Grid*, **3**(2), 915–924.

69. **Liu, Z.** and **J. Zhou** (2020). Introduction to Graph Neural Networks. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, **14**(2), 1–127. ISSN 19394616.

70. **Lotfifard, S.** (2017). Sparse sensing platform for line-outage identification in multiarea power systems. *IEEE Transactions on Industrial Informatics*, **13**(3), 947–955.

71. **Lustig, M.**, **D. L. Donoho**, **J. M. Santos**, and **J. M. Pauly** (2008). Compressed sensing mri: A look at how cs can improve on current imaging techniques. *IEEE Signal Processing Magazine*, **25**, 72–82. ISSN 10535888.

72. **Madbhavi, R.**, **H. S. Karimi**, **B. Natarajan**, and **B. Srinivasan** (2020). Tensor completion based state estimation in distribution systems. *2020 IEEE Power and Energy Society Innovative Smart Grid Technologies Conference, ISGT 2020*.

73. **Madbhavi, R.**, **B. Natarajan**, and **B. Srinivasan** (2021). Enhanced Tensor Completion Based Approaches for State Estimation in Distribution Systems. *IEEE Transactions on Industrial Informatics*, **17**(9), 5938–5947. ISSN 19410050.

74. **Madbhavi, R.**, **B. Natarajan**, and **B. Srinivasan** (2023). Graph neural network-based distribution system state estimators. *IEEE Transactions on Industrial Informatics*, 1–10.

75. **Mairal, J.**, **F. Bach**, **J. Ponce**, and **G. Sapiro**, Online dictionary learning for sparse coding. *In Proceedings of the 26th Annual International Conference on Machine Learning*, ICML '09. Association for Computing Machinery, New York, NY, USA, 2009. ISBN 9781605585161. URL https://doi.org/10.1145/1553374.1553463.

76. **Makonin, S.** (2016). AMPds2: The Almanac of Minutely Power dataset (Version 2). URL https://doi.org/10.7910/DVN/FIE0S4.

77. **Malekpour, A. R.** and **A. Pahwa**, Radial test feeder including primary and secondary distribution network. *In 2015 North American Power Symposium (NAPS)*. 2015.

78. **Manitsas, E.**, **R. Singh**, **B. C. Pal**, and **G. Strbac** (2012). Distribution system state estimation using an artificial neural network approach for pseudo measurement modeling. *IEEE Transactions on Power Systems*, **27**(4), 1888–1896.

79. **Mestav, K. R.**, **J. Luengo-Rozas**, and **L. Tong** (2019). Bayesian state estimation for unobservable distribution systems via deep learning. *IEEE Transactions on Power Systems*, **34**(6), 4910–4920.

80. **Mounce, S.**, **J. Boxall**, and **J. Machell** (2009). Development and verification of an online artificial intelligence system for detection of bursts and other abnormal flows. *Journal of Water Resources Planning and Management*, **136**(3), 309–318.

81. **Mounce, S.**, **R. Mounce**, **T. Jackson**, **J. Austin**, and **J. Boxall** (2014). Pattern matching and associative artificial neural networks for water distribution system time series data analysis. *Journal of Hydroinformatics*, **16**(3), 617–632.

82. **Munikoti, S.**, **D. Agarwal**, **L. Das**, and **B. Natarajan** (2022*a*). A General Framework for quantifying Aleatoric and Epistemic uncertainty in Graph Neural Networks. *arxiv*. URL https://arxiv.org/abs/2205.09968v1.

83. **Munikoti, S.**, **L. Das**, and **B. Natarajan** (2022*b*). Scalable graph neural network-based framework for identifying critical nodes and links in complex networks. *Neurocomputing*, **468**, 211–221. ISSN 0925-2312.

84. **Murray, D.**, **L. Stankovic**, and **V. Stankovic** (2017). An electrical load measurements dataset of united kingdom households from a two-year longitudinal study. *Scientific Data*, **4**(1), 160122. ISSN 2052-4463. URL https://doi.org/10.1038/sdata.2016.122.

85. **Muscas, C.**, **F. Pilo**, **G. Pisano**, and **S. Sulis** (2009). Optimal allocation of multichannel measurement devices for distribution state estimation. *IEEE Transactions on Instrumentation and Measurement*, **58**(6), 1929–1937.

86. **Nie, F.**, **H. Huang**, and **C. Ding**, Low-rank matrix recovery via efficient

schatten p-norm minimization. *In Proceedings of the National Conference on Artificial Intelligence*, volume 1. 2012. ISBN 9781577355687. URL https://dl.acm.org/doi/10.5555/2900728.2900822.

87. **Olshausen, B. A.** and **D. J. Field** (1997). Sparse coding with an overcomplete basis set: A strategy employed by v1? *Vision Research*, **37**(23), 3311–3325. ISSN 0042-6989. URL https://www.sciencedirect.com/science/article/pii/S0042698997001697.

88. **Ostrometzky, J.**, **K. Berestizshevsky**, **A. Bernstein**, and **G. Zussman** (2019). Physics-informed deep neural network method for limited observability state estimation.

89. **Parson, O.**, **G. Fisher**, **A. Hersey**, **N. Batra**, **J. Kelly**, **A. Singh**, **W. Knottenbelt**, and **A. Rogers**, Dataport and nilmtk: A building data set designed for non-intrusive load monitoring. *In 2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. 2015.

90. **Pedregosa, F.**, **G. Varoquaux**, **A. Gramfort**, **V. Michel**, **B. Thirion**, **O. Grisel**, **M. Blondel**, **P. Prettenhofer**, **R. Weiss**, **V. Dubourg**, **J. Vanderplas**, **A. Passos**, **D. Cournapeau**, **M. Brucher**, **M. Perrot**, and **E. Duchesnay** (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, **12**, 2825–2830.

91. **Peng, J.**, **Y. Sun**, and **H. F. Wang** (2006). Optimal PMU placement for full network observability using Tabu search algorithm. *International Journal of Electrical Power and Energy Systems*, **28**(4), 223–231. ISSN 0142-0615.

92. **Powalko, M.**, **A. G. Orths**, **H. Abildgaard**, **P. B. Eriksen**, **K. Rudion**, **I. I. Golub**, and **N. I. Voropai**, System observability indices for optimal placement of pmu measurements. *In 2012 IEEE Power and Energy Society General Meeting*. 2012.

93. **Primadianto, A.** and **C. Lu** (2017). A review on distribution system state estimation. *IEEE Transactions on Power Systems*, **32**(5), 3875–3883.

94. **Rajeswaran, A.**, **S. Narasimhan**, and **S. Narasimhan** (2018). A graph partitioning algorithm for leak detection in water distribution networks. *Computers & Chemical Engineering*, **108**, 11–23.

95. **Raposo, A. A.**, **A. B. Rodrigues**, and **M. da Guia da Silva** (2017). Optimal meter placement algorithm for state estimation in power distribution networks. *Electric Power Systems Research*, **147**, 22 – 30. ISSN 0378-7796. URL http://www.sciencedirect.com/science/article/pii/S037877961730072X.

96. **Rezaeian Koochi, M. H.** and **M. H. Hemmatpour** (2020). A general PMU placement approach considering both topology and system aspects of contingencies. *International Journal of Electrical Power and Energy Systems*, **118**, 105774. ISSN 0142-0615.

97. **Romano, M.**, **Z. Kapelan**, and **D. A. Savić** (2012). Automated detection of pipe bursts and other events in water distribution systems. *Journal of Water Resources Planning and*

Management, **140**(4), 457–467.

98. **Schlichtkrull, M.**, **T. N. Kipf**, **P. Bloem**, **R. van den Berg**, **I. Titov**, and **M. Welling**, Modeling relational data with graph convolutional networks. *In The Semantic Web*. Springer International Publishing, 2018. ISBN 978-3-319-93417-4.

99. **Singh, R.**, **E. Manitsas**, **B. C. Pal**, and **G. Strbac** (2010). A recursive bayesian approach for identification of network configuration changes in distribution system state estimation. *IEEE Transactions on Power Systems*, **25**(3), 1329–1336.

100. **Sobral, A.** and **E. hadi Zahzah** (2017). Matrix and tensor completion algorithms for background model initialization: A comparative evaluation. *Pattern Recognition Letters*, **96**, 22–33. ISSN 0167-8655.

101. **Soldevila, A.**, **J. Blesa**, **S. Tornil-Sin**, **E. Duviella**, **R. M. Fernandez-Canti**, and **V. Puig** (2016). Leak localization in water distribution networks using a mixed model-based/data-driven approach. *Control Engineering Practice*, **55**, 162–173.

102. **Teimourzadeh, S.**, **F. Aminifar**, and **M. Shahidehpour** (2019). Contingency-constrained optimal placement of micro-pmus and smart meters in microgrids. *IEEE Transactions on Smart Grid*, **10**(2), 1889–1897.

103. **U.S. Energy Information Administration** (2021). How many smart meters are installed in the united states, and who has them? URL `https://www.eia.gov/tools/faqs/faq.php?id=108&t=3`.

104. **Vigliassi, M. P.**, **J. A. Massignan**, **A. C. B. Delbem**, and **J. B. A. London** (2019). Multi-objective evolutionary algorithm in tables for placement of SCADA and PMU considering the concept of Pareto Frontier. *International Journal of Electrical Power and Energy Systems*, **106**, 373–382. ISSN 0142-0615.

105. **Wang, Q.**, **M. Guidolin**, **D. Savic**, and **Z. Kapelan** (2015). Two-objective design of benchmark problems of a water distribution system via moeas: Towards the best-known approximation of the true pareto front. *Journal of Water Resources Planning and Management*, **141**(3), 04014060.

106. **Wang, X.**, **D. P. Palomar**, **L. Zhao**, **M. S. Ghidaoui**, and **R. D. Murch** (2018). Spectral-based methods for pipeline leakage localization. *Journal of Hydraulic Engineering*, **145**(3), 04018089.

107. **Wang, Z.**, **M. J. Lai**, **Z. Lu**, **W. Fan**, **H. Davulcu**, and **J. Ye** (2014). Orthogonal Rank-One Matrix Pursuit for Low Rank Matrix Completion. *SIAM Journal on Scientific Computing*, **37**(1), A488–A514. ISSN 10957200. URL `https://arxiv.org/abs/1404.1377v2`.

108. **Wu, Y.**, **S. Liu**, **X. Wu**, **Y. Liu**, and **Y. Guan** (2016). Burst detection in district metering areas using a data driven clustering algorithm. *Water research*, **100**, 28–37.

109. **Wu, Z.**, **Q. Wang**, and **X. Liu**, State estimation for power system based on graph neural network. *In 2022 IEEE 5th International Electrical and Energy Conference (CIEEC)*. 2022*a*.

110. **Wu, Z.**, **Q. Wang**, and **X. Liu**, State estimation for power system based on graph neural network. *In 2022 IEEE 5th International Electrical and Energy Conference (CIEEC)*. 2022*b*.

111. **Xygkis, T. C.**, **G. N. Korres**, and **N. M. Manousakis** (2018). Fisher information-based meter placement in distribution grids via the d-optimal experimental design. *IEEE Transactions on Smart Grid*, **9**(2), 1452–1461.

112. **Yao, Y.**, **X. Liu**, and **Z. Li** (2019). Robust measurement placement for distribution system state estimation. *IEEE Transactions on Sustainable Energy*, **10**(1), 364–374.

113. **Yazdani, A.** and **P. Jeffrey** (2012). Water distribution system vulnerability analysis using weighted and directed network models. *Water Resources Research*, **48**(6).

114. **Ye, G.** and **R. A. Fenner** (2013). Weighted least squares with expectation-maximization algorithm for burst detection in uk water distribution systems. *Journal of Water Resources Planning and Management*, **140**(4), 417–424.

115. **Ye, X.**, **J. Yang**, **X. Sun**, **K. Li**, **C. Hou**, and **Y. Wang** (2015). Foreground-Background Separation from Video Clips via Motion-Assisted Matrix Restoration. *IEEE Transactions on Circuits and Systems for Video Technology*, **25**(11), 1721–1734. ISSN 10518215.

116. **Zamzam, A. S.** and **N. D. Sidiropoulos** (2019). Physics-aware neural networks for distribution system state estimation. URL `https://arxiv.org/abs/1903.09669`.

117. **Zhang, Q.**, **Z. Y. Wu**, **M. Zhao**, **J. Qi**, **Y. Huang**, and **H. Zhao** (2016). Leakage zone identification in large-scale water distribution systems using multiclass support vector machines. *Journal of Water Resources Planning and Management*, **142**(11), 04016042.

118. **Zhang, Z.**, **G. Ely**, **S. Aeron**, **N. Hao**, and **M. Kilmer** (2014). Novel methods for multilinear data completion and de-noising based on tensor-SVD. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 3842–3849. ISSN 10636919.

119. **Zhao, Q.**, **L. Zhang**, and **A. Cichocki** (2015). Bayesian CP factorization of incomplete tensors with automatic rank determination. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **37**(9), 1751–1763. ISSN 01628828.

120. **Zimmerman, R. D.**, **C. E. Murillo-Sánchez**, and **R. J. Thomas** (2011). Matpower: Steady-state operations, planning, and analysis tools for power systems research and education. *IEEE Transactions on Power Systems*, **26**(1), 12–19.

CURRICULUM VITAE

NAME Rahul Madbhavi

DATE OF BIRTH 27 August 1994

EDUCATION QUALIFICATIONS

2016 **Bachelor of Technology**

Institution	National Institute of Technology Karnataka
Specialization	Mechanical Engineering

Doctor of Philosophy

Institution	Indian Institute of Technology Madras
Specialization	Biomedical Engineering
Registration Date	24 July 2018

DOCTORAL COMMITTEE

Chairperson	Dr. Ramakrishnan S Professor Dept. of Applied Mechanics and Biomedical Engineering
Guide(s)	Dr. Babji Srinivasan Associate Professor Dept. of Applied Mechanics and Biomedical Engineering
Member(s)	Dr. Nirav Pravinbhai Bhatt Assistant Professor Dept. of Biotechnology Engineering
	Dr. Shaikh Faruque Ali Professor Dept. of Applied Mechanics and Biomedical Engineering
	Dr. Satyanarayanan Seshadri Associate Professor Dept. of Applied Mechanics and Biomedical Engineering