UNIVERSITY OF CALIFORNIA
RIVERSIDE

Estimation of the Topology, Parameters, and Distributed Energy Resources in Power Distribution Systems

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Electrical Engineering

by

Wenyu Wang

June 2021

Dissertation Committee:

Dr. Nanpeng Yu, Chairperson
Dr. Yingbo Hua
Dr. Weixin Yao

The Dissertation of Wenyu Wang is approved:

Committee Chairperson

University of California, Riverside

Acknowledgments

First and foremost, I would like to express my sincere gratitude to my Ph.D. advisor, Dr. Nanpeng Yu, for his excellent supervision, continuous support, and kind patience. I am extremely grateful for his guidance with profound knowledge, insightful advice, and hearty encouragement, which make my research explorations into wonderful adventures. His expertise, dedication, and persistence in scientific research and education have inspired me to push myself to a higher level in work and life.

I greatly appreciate my committee members, Dr. Yingbo Hua and Dr. Weixin Yao for their service and help. They have provided invaluable suggestions and comments, which help a lot to shape my research work.

I am thankful to all the collaborators in academia and industry for providing valuable data, consultation, and comments. I would also like to thank all my labmates and friends for numerous discussions, sincere advice, and mutual support.

I am grateful to all the professors and educators who have instructed or taught me during and before my study in UCR. Without them I would not be making progress in my study.

Finally and most importantly, I would like to express my deep gratitude to my family, especially my parents, Fugui Wang and Huimin Sun. They not only have raised me up with all the best they have, but also respect my own decision and encourage me to chase my dreams. I would have not been able to complete this work without their love, understanding, and unconditional support.

The content of this dissertation is a reprint of the materials that appeared in the following publications:

- W. Wang, N. Yu, B. Foggo, J. Davis, and J. Li, "Phase identification in electric power distribution systems by clustering of smart meter data," in Machine Learning and Applications 2016, 15th IEEE International Conference on. IEEE, Dec. 2016, pp. 259–265. (Chapter 2.3)

- W. Wang and N. Yu, "Advanced metering infrastructure data driven phase identification in smart grid," in The Second International Conference on Green Communications, Computing and Technologies, Sep. 2017, pp. 16–23. (Chapter 2.4)

- W. Wang and N. Yu, "Maximum marginal likelihood estimation of phase connections in power distribution systems," IEEE Transactions on Power Systems, vol. 35, no. 5, pp. 3906–3917, 2020. (Chapter 3)

- W. Wang and N. Yu, "Parameter estimation in three-phase power distribution networks using smart meter data," in 2020 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS). IEEE, Aug. 2020, pp. 1–6. (Chapter 4)

- W. Wang and N. Yu, "Estimate three-phase distribution line parameters with physics-informed graphical learning method," arXiv preprint arXiv:2102.09023 [cs.LG], Feb. 2021.[Online]. Available: https://arxiv.org/pdf/2102.09023v1.pdf (Chapter 5)

- W. Wang, N. Yu, and R. Johnson, "A model for commercial adoption of photovoltaic systems in California," Journal of Renewable and Sustainable Energy, vol. 9, no. 2, p. 025904, 2017. (Chapter 6)

- W. Wang, N. Yu, J. Shi, and N. Navarro, "Diversity factor prediction for distribution feeders with interpretable machine learning algorithms," in 2020 IEEE Power & Energy Society General Meeting (PESGM). IEEE, Aug. 2020, pp. 1–5. (Chapter 7)

To my parents for all the support and love.

To my grandparents, in loving memory.

ABSTRACT OF THE DISSERTATION

Estimation of the Topology, Parameters, and Distributed Energy Resources in Power Distribution Systems

by

Wenyu Wang

Doctor of Philosophy, Graduate Program in Electrical Engineering
University of California, Riverside, June 2021
Dr. Nanpeng Yu, Chairperson

Distributed energy resources (DERs), such as distributed solar photovoltaic (PV) systems, electric vehicles, distributed energy storage, and demand response, are being deployed in the power distribution system at an unprecedented pace. Though DERs bring environmental and technological benefits, challenging technical problems arise as well. Distribution networks must be actively managed and planned/upgraded accordingly to accommodate DERs and coordinate their operations. All of these depend on the solving of the technical problems of accurate phase identification, network parameter estimation, DER adoption prediction, and long-term load forecasting in the distribution system. In this dissertation, I use machine learning and data analytic techniques to address these challenging problems, which are critical to the adoption of DERs.

To address the problem of phase identification, we study multiple methodology approaches. Two unsupervised learning algorithms are developed based on smart meter data and supervisory control and data acquisition (SCADA) data. The first algorithm leverages linear dimension reduction and centroid-based clustering. The second algorithm further improves the phase identification accuracy by nonlinear dimension reduction and density-based clustering. In the third approach, a

maximum marginal likelihood estimation approach based on physics-informed model is proposed, which is physically interpretable and more accurate.

To address the problem of three-phase network parameter estimation, we develop a maximum likelihood estimation approach based on a physics-informed model to estimate the serial impedance of three-phase lines. A more advanced method based on graphical learning model is then developed to provide more accurate parameter estimation.

To address the problem of DER adoption prediction, we study the adoption of distributed commercial solar PV systems by developing a generalized Bass diffusion model. This model is capable of forecasting solar PV adoptions and quantifies the impact of solar PV costs and government incentive programs on the adoption.

To address the problem of long-term load forecasting, we develop comprehensive models based on supervised learning to forecast the diversity factor (DF) of distribution feeders at high accuracy. We also quantifies the importance of different influential factors and analyze how they affect DF.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

Our power distribution system is experiencing a tremendous transformation in the last decade. Driven by stricter environmental regulations, technological advances, business model innovations, declining system costs, and the government's supporting policies and incentives, there is a rapid expansion of distributed energy resources (DERs) in the distribution system. DERs are power generation resources or controllable loads connected to a local distribution system. Examples of DERs include distributed solar photovoltaic (PV) systems, electric vehicles, distributed energy storage, demand response, etc. The DERs are expected to have faster growth than the centralized generation. According to a report by Guidehouse [1], the annual new DER capacity additions are expected to grow from 200 GW in 2020 to 500 GW in 2030; in comparison, the new centralized generation capacity additions are expected to grow from 220 GW in 2020 to 270 GW in 2030.

DER technology brings both environmental and technological benefits. First, by using more renewable energy resources, DER technology produces less greenhouse gas and pollution to

the environment, reducing global warming and health risks. Second, DER technologies improve the distribution system's reliability, power supply quality, and energy efficiency. Third, by conserving and providing more flexible energy resources, DERs reduce the overcrowding on the grid and reduces the pressure of needing more generation and transmission capacity.

However, challenging technical problems arise with the rapid growth of DERs, due to their impacts on the distribution system in terms of network operation and planning/upgrading. First, to monitor and coordinate the operations of DERs, utility operators must rely on accurate three-phase distribution network models, including network phase connections and parameters, which are usually unavailable [2, 3, 4, 5, 6, 7, 8, 9, 10]. Second, to accommodate DERs, the distribution networks must be planned/upgraded according to the growth of DERs and customer load. Thus, it is very critical to solve the technical problems of accurate phase identification, network parameter estimation, DER adoption prediction, and long-term load forecasting in the distribution system.

The widespread installation of smart meters and sensors in the distribution system provides the hardware basis to use data-driven techniques in solving the research problems introduced by DERs. Smart meters can record users' data such as power consumption, voltage level, current, and power factor at a high resolution and they also enables two-way communication with the central system. The distribution networks world-wide are experiencing a fast expansion of advanced metering infrastructure (AMI). The smart meter penetration level in North America is expected to reach 81% in 2024 [11]. In European countries such as Italy, Sweden, Finland, and the Netherlands, smart meter penetration levels reached 80% by 2019 and are still increasing [12].

New research opportunities to study the problems in distribution system are created by advances in data mining and machine learning techniques. Data mining and machine learning show

strong potential in discovering hidden data patterns and relationships, modeling complicated systems, and decision making, which inspires many researchers. For example, decision tree and classification techniques are used for fault detection in microgrids [13]; deep reinforcement learning techniques are used for volt-var control [14, 15, 16] and network reconfiguration [17, 18]. In this dissertation, techniques in data mining and machine learning are leveraged to study the research problems introduced by DERs.

1.2 Technical Challenges and Research Opportunities

The rapid growth of DERs bring technical challenges and new research opportunities at the same time. These challenges and research opportunities originate from the DERs' impacts on distribution network operation and planning/upgrading. In this section, the two kinds of impacts will be explained first, then the details of the challenging technical problems will be elaborated.

The first kind of impacts is on the operation and management of the distribution system. DERs introduce bidirectional power flow to the distribution network, while the traditional distribution network has one-way power flow from the centralized power source to the electricity users. In addition, DERs introduce high variability and dynamics to the distribution system. For example, the varied load/generation patterns of distributed solar PV systems, wind turbine, and electric vehicles are either highly dynamic or intermittent [19, 20]. To coordinate the operations of a large number of heterogeneous DERs, advanced distribution system control applications such as Volt-VAR control, network reconfiguration, and three-phase optimal power flow need to be implemented. All of these applications rely on accurate three-phase distribution network models, of which the phase connections and network parameters are critical information.

The second kind of impacts is on the planning and upgrading of distribution networks. On one hand, the fast expansion of DERs impacts the voltage, power quality, and protection coordination of distribution networks. To mitigate these impacts, system planners must upgrade or adjust the distribution system according to the deployment of DERs. Thus, system planners must have accurate predictions of the future DER adoptions. On the other hand, with the increasing penetration of DERs, forecasting the long-term load of distribution networks becomes more complicated. The network planners must rely on accurate forecasts of the long-term load to determine the network design and equipment ratings of new distribution feeders.

1.2.1 Identification of Phase Connections in Power Distribution Systems

The electric power distribution system is the final portion of the power delivery infrastructure, which carries electricity from highly interconnected, high-voltage transmission systems to end-use customers. The power distribution system serves electricity in three phases: A, B, and C. The electricity users such as family houses are connected to one, two, or three phases of the wires of the distribution system. Different from the transmission system, the distribution system is usually unbalanced, thus the phase connection is very important in the modeling of distribution systems. However, utility companies typically do not have accurate phase connectivity information; moreover, the phase connections may change over time during distribution system maintenance.

Traditionally, to determine phase connections, electric utilities need to send field crews to measure phase angles with special equipment such as phase meters [21]. Such practice is not widely applicable because it is very labor-intensive, time-consuming, and expensive. With the widespread availability of smart meter and sensor data, data-driven methods become a promising approach for efficient and inexpensive phase identification.

1.2.2 Network Parameter Estimation in Power Distribution Systems

Similar to phase connections, the network parameter such as the wire impedance is another important part of distribution network modeling. Utility companies typically do not have accurate information of network parameters because their geographic information system (GIS) may contain errors due to unreliable documentation during the system modifications and upgrades.

The network parameter estimation in distribution systems is more challenging than in transmission systems. Although many studies have examined the network parameter estimation in transmission systems [22, 23, 24, 25], they cannot be easily applied to the distribution system. This is because, different from the transmission system, distribution lines are rarely transposed. Untransposed lines will lead to unequal diagonal and off-diagonal terms in the phase impedance matrix. Thus, instead of single-phase models, three-phase line segment models need to be developed. Specifically, the elements of a 3×3 phase impedance matrix need to be estimated for each three-phase distribution line segment. The readily available smart meters open up opportunities to estimate distribution network parameters from data-driven approaches.

1.2.3 Predicting the Growth of DER Adoptions

Predicting the growth of DER adoptions is very important to the planning and upgrading of both distribution and transmission systems. Failing to incorporate accurate DER forecasts in the power system planning can lead to wide-ranging consequences [26]. First, inaccurate forecast of DER growth may lead to a mismatch between generation resources and the demand. In the case of over-forecasting, there may not be adequate bulk generation resources due to the predicted DER contribution, which will lead to a less reliable and resilient system. In the case of under-forecasting,

unnecessary bulk generation resources and infrastructure could be added, which will result in over-building. Second, DERs may impact the voltage, power quality, and protection coordination of distribution networks. For these reasons, planners need accurate forecasts of DER adoptions in order to mitigate the aforementioned impacts accordingly. The mitigation measures include upgrading the equipment, changing settings of existing devices, and applying new coordination and management tools.

Predicting the growth of future DER adoptions is challenging because it is influenced by multiple factors such as installation cost, government incentive programs, DER types, DER capacity, etc. In the study of product adoptions, adoption growth models such as the Bass diffusion model [27] and the generalize Bass model [28] have been widely used and thus have the potential to enable the prediction of DER adoptions.

1.2.4 Long-Term Load Forecast of Distribution Feeders

Determining the maximum diversified demand is one of the most important parts of long-term load forecast of distribution feeders. The maximum diversified demand is the maximum of the sum of demands of a group of electricity customers over a particular period. It is a critical factor to consider when utilities develop plans to build new distribution systems. Both network topology design and equipment ratings depend on the maximum diversified demand. Underestimating the maximum diversified demand will cause reliability and safety issues. If the peak load exceeds the circuit rating, then equipment such as transformers and cables will be overloaded, which results in shortened lifespan and premature failure. Overestimating the maximum diversified demand often leads to installation of oversized distribution system equipment and under-utilization of system assets.

The maximum diversified demand is usually estimated by using the maximum noncoincident demand and the diversity factor (DF). The DF is defined as the ratio of maximum noncoincident demand to the maximum diversified demand [29]. Here, the maximum noncoincident demand is the sum of each individual customer's maximum demand.

Estimating DF is challenging, while the maximum noncoincident demand is straightforward to estimate. In current practice, DF is often estimated based on a simple relationship with the number of customers and this relationship is often derived by load surveys. However, DF is influenced by many other factors, such as customer demographics and climate conditions. The rapid growth of DERs should also be considered when estimating DF. It will be very handy to system planners if a DF prediction model accounts for a variety of influential factors and can explain how different factors affect the DF. Thanks to the technical advances in machine learning, new opportunities arise to build such forecasting models.

1.3 Contributions

This dissertation studies the four targeted topics mentioned in Section 1.2. The main contributions are elaborated as follows.

1.3.1 Phase Identification by Unsupervised Learning

Two works on unsupervised machine learning algorithms are developed to identify the phase connectivity of customers based on smart meter data and supervisory control and data acquisition (SCADA) data. In the first work [30], an innovative constrained k-means clustering algorithm of smart meter data is proposed to solve the phase identification problem. First, unique features

are extracted from the voltage time series of smart meters instead of directly using the voltage time series data. Then customer phase constraints are defined based on the known information about line configurations in the network connectivity model. At last, a constrained k-means clustering algorithm is applied to accurately identify the phase connection of each customer.

In light of the existing literature, the unique contributions of this work are as follows. First, the proposed algorithm utilizes the known information of line configurations in the network connectivity model to avoid mislabeling of the customers on the same secondary feeder, which can occur in the existing methods. Second, the proposed algorithm is computationally efficient compared with the 0-1 integer linear programming method and the correlation-based methods. Third, the proposed algorithm is highly accurate in distribution circuits where the majority of loads are connected to two-phase laterals. Fourth, the proposed algorithm can still determine the phase connections of metered customers when the distribution circuit has some unmetered customers.

In the second work [31], an AMI data driven machine learning algorithm is developed to solve the phase identification problem. First, key features are extracted from the voltage time series by a nonlinear dimension reduction technique. A constraint-driven hybrid clustering (CHC) algorithm is then designed to group smart meters/structures into various clusters. Finally, the phase connection of each cluster is identified by field validations on the phase connections of very few smart meters. In comprehensive case studies on 5 distribution circuits, the proposed data driven machine learning algorithm yields high accuracies. In addition, this work reveals that more granular voltage readings will lead to even more accurate phase identification results.

Compared to the existing data-driven phase identification algorithms, the proposed method has the following advantages. First, the proposed algorithm does not require prior knowledge about

the number of phase connections in the distribution system, which is required by most of the existing AMI data driven methods. Second, the proposed algorithm works well with distribution feeders that have both phase-to-neutral and phase-to-phase connections, while most of the existing techniques works only with one of the two kinds of connections. Third, the accuracy of the proposed phase identification algorithm is insensitive to the level of unbalance in a distribution feeder.

1.3.2 Phase Identification by the Physics-Informed Model Approach

To further improve the phase identification accuracy and provide a theoretical foundation for the problem, a physically inspired machine learning method is developed for phase identification [32]. First, a physical model linking phase connections to the smart meter measurements is developed by linearizing the three-phase power flow manifold. Then the phase identification task is formulated as a maximum likelihood estimation (MLE) problem and the correct phase connection is proved to yield the highest log likelihood value. The MLE problem is difficult to solve due to the nonlinearity and nonconvexity nature. Thus, the MLE problem is reformulated as a maximum marginal likelihood estimation (MMLE) problem and it is proved that the correct phase connection also yields the highest marginal log likelihood value. Finally, an efficient solution algorithm is developed for the MMLE problem by dividing it into sub-problems, which can be solved by least squares integer programming.

Compared to the existing data-driven phase identification algorithms, this approach has the following advantages. First, the physically interpretable MMLE formulation brings a solid theoretical foundation to the phase identification problem. Second, the proposed algorithm works for not only radial distribution feeders, but also heavily meshed networks. Third, the proposed algorithm is more accurate for complex circuits with both single-phase and two-phase connections and

a lower level of unbalance, while existing data-driven methods have a lot of problems in such circuits. Fourth, the proposed algorithm is robust with respect to inaccurate feeder models, incomplete measurements, and bad measurements.

1.3.3 Parameter Estimation by the Physics-Informed Model Approach

A data-driven algorithm is proposed to estimate the serial conductance and serial susceptance of the π equivalent model for three-phase distribution lines by using the readily available smart meter measurements of voltage magnitude, real power consumption, and reactive power consumption [33]. First, a physical model linking smart meter measurements and the three-phase serial conductance and susceptance is built by linearizing the three-phase power flow manifold. Then the three-phase parameter estimation problem is formulated as a maximum likelihood estimation (MLE) problem and it is proved that the correct network parameters yield the highest likelihood value. At last, the stochastic gradient descent (SGD) algorithm with early stopping is adopted to solve the MLE problem.

Compared to the existing parameter estimation methods, the proposed algorithm has two advantages. First, the proposed approach takes unequal self and mutual serial conductance and susceptance into consideration, which is specifically designed to estimate parameters of three-phase distribution networks. Second, the proposed approach can be easily applied in real-world distribution circuits because it only uses readily available smart meter data.

1.3.4 Parameter Estimation by the Graphical Learning Model Approach

A physics-informed graphical learning algorithm is developed to estimate the 3×3 series resistance and reactance matrices of three-phase distribution line model using readily available smart

meter measurements [34]. The proposed method is inspired by the emerging graph neural network (GNN), which is designed for estimation problems in networked systems. Three-phase power flow-based physical transition functions are designed to replace the ones based on deep neural networks in the GNN. The gradient of the voltage magnitude loss function with respect to the line segments' resistance and reactance parameters is then derived with an iterative method. Finally, the stochastic gradient descent (SGD) approach is used to estimate distribution network parameters by minimizing the error between the physics-based graph learning model and the smart meter measurements. Prior estimates and bounds of network parameters are also leveraged to improve the estimation accuracy. To improve computation efficiency, partitions can be introduced so that parameter estimations are executed in parallel in sub-networks.

The main technical contributions of this work are as follows. First, a physics-informed graphical learning method is developed to estimate line parameters of three-phase distribution networks. Second, the proposed algorithm can be easily applied to real-world distribution circuits because it only uses readily available smart meter data. Third, the proposed approach is more accurate on test feeders than the state-of-the-art benchmark because it preserves the nonlinearity of three-phase power flows in the graphical learning framework.

1.3.5 Modeling of Commercial Adoption of Photovoltaic Systems

The solar PV system is one of the fast expanding DER types. A model for commercial solar PV adoption is developed with explanatory variables such as government incentive programs and solar PV system installation costs to quantify their impacts [35]. The adoption model is built on top of the generalized Bass diffusion framework and is applied to forecast commercial solar PV adoption in Southern California. The nonlinear least squares approach is used to estimate the model

parameters and their asymptotic standard errors. The proposed model yields a lower root-mean-square error (RMSE) than the basic Bass diffusion model.

Compared with other related work, this work makes the following unique contributions. First, this work develops a model for commercial PV adoption, which quantifies the impact of solar PV costs and government incentive programs on the adoption. Second, the proposed model is also capable of forecasting the eventual commercial PV adoption rate and quantifying the delayed effect of explanatory variables on adoption. It is found that the eventual adoption rate of solar PV system is higher for large commercial customers. Third, the model is applied to fit the empirical commercial PV adoption data in Southern California. The empirical results show that large commercial customers are more susceptible to the influence of PV costs and government incentives than small commercial customers. Fourth, by changing the cost and incentive rates fed into the model, adoption curves can be forecasted under different cost and policy conditions. This can be a useful tool for the government to evaluate its renewable energy technology incentive policies.

1.3.6 Diversity Factor Prediction for Distribution Feeders

A set of comprehensive models based on supervised machine learning algorithms are developed to predict the DF of distribution feeders, accounting for a variety of influential factors, such as customer type, weather, demographics, and socioeconomic conditions [36]. The machine learning algorithms not only yield high prediction accuracy on real-world distribution feeders but also provide useful insights on how input features influence DF. The key factors that affect the DF are identified by using the interpretation method called SHapley Additive exPlanations (SHAP) [37].

1.4 Thesis Organization

The remainder of this dissertation is organized as follows. Chapter 2 presents the work of phase identification by unsupervised learning, including the algorithm using linear dimension reduction with centroid-based clustering [30] and the algorithm using nonlinear dimension reduction with density-based clustering [31]. Chapter 3 presents the work of phase identification by the physics-informed model approach [32]. In Chapter 4, a physics-informed model approach is proposed for parameter estimation[33]. In Chapter 5, the parameter estimation problem is solved by a graphical learning model approach [34]. Chapter 6 presents a model for commercial adoption of solar photovoltaic systems [35]. In Chapter 7, the diversity factor of distribution feeders is predicted by using a set of comprehensive supervised machine learning models [36]. Chapter 8 states the conclusion.

Chapter 2

Phase Identification by Unsupervised Learning

2.1 Introduction

Driven by stricter environmental regulations, technological advances, and business model innovations, distributed energy resources (DERs) are being deployed in the electric power distribution systems at an unprecedented pace. According to a technical report from Navigant Research [38], the annual installed capacity across the global DER market is expected to grow from 136.4 GW in 2015 to 530.7 GW in 2024.

To fully exploit the benefits of the DERs, the distribution network must be actively managed. To operate the distribution system in an efficient and reliable manner, the distribution system operators typically rely on a set of tools and applications including three-phase power flow, distribution system state estimation, three-phase optimal power flow, distribution system restoration and

distribution network reconfiguration. All of these applications require an accurate distribution network and phase connectivity model. Although the network connectivity model is mostly accurate, phasing errors are common [39]. Therefore, an accurate phase identification method is in critical need.

Electric utility companies typically do not have accurate phase connectivity information. Moreover, the phase connectivity of the distribution network changes over time when new customers are connected to the system. With more DERs connected to the power distribution systems, correct phase connectivity data become increasingly important to efficient and reliable operations of power distribution systems. In this chapter, we develop two unsupervised machine learning algorithms to identify the phase connectivity of customers based on smart meter data and supervisory control and data acquisition (SCADA) data. The first algorithm leverages the technique of linear dimension reduction with centroid-based clustering; the second algorithm leverages the technique of nonlinear dimension reduction with density-based clustering.

The rest of this chapter is organized as follows. Section 2.2 introduces the background and defines the problem of phase identification. Section 2.3 presents the phase identification algorithm by linear dimension reduction with centroid-based clustering. Section 2.4 presents the phase identification algorithm by nonlinear dimension reduction with density-based clustering.

2.2 Background and Problem Definition of Phase Identification

To understand the phase identification problem, we first briefly introduce the electric power distribution system. The electric power distribution system is the final portion of the power delivery infrastructure that carries electricity from highly interconnected, high-voltage transmission

systems to end-use customers. An illustration of a simple electric distribution system is depicted in Figure 2.1. The starting point of the distribution system is the distribution substation. In the distribution substation, a step-down transformer lowers the transmission-level voltage (35 to 230 kV) to a medium-level voltage (4 to 35 kV) in the primary distribution circuits [40]. The electric power then flows through the primary feeders and laterals ($L1$-$L5$) to distribution transformers ($T1$-$T8$), which further step down the voltage to low-voltage secondary circuits. The secondary circuits serve end-use customers and operate at 120/240 V three-wire, 120/208 V three-phase, or 277/480 V three-phase. Laterals can be single-phase ($L2$), two-phase, also called "V" phase ($L3$, $L4$), or three-phase ($L1$, $L5$).

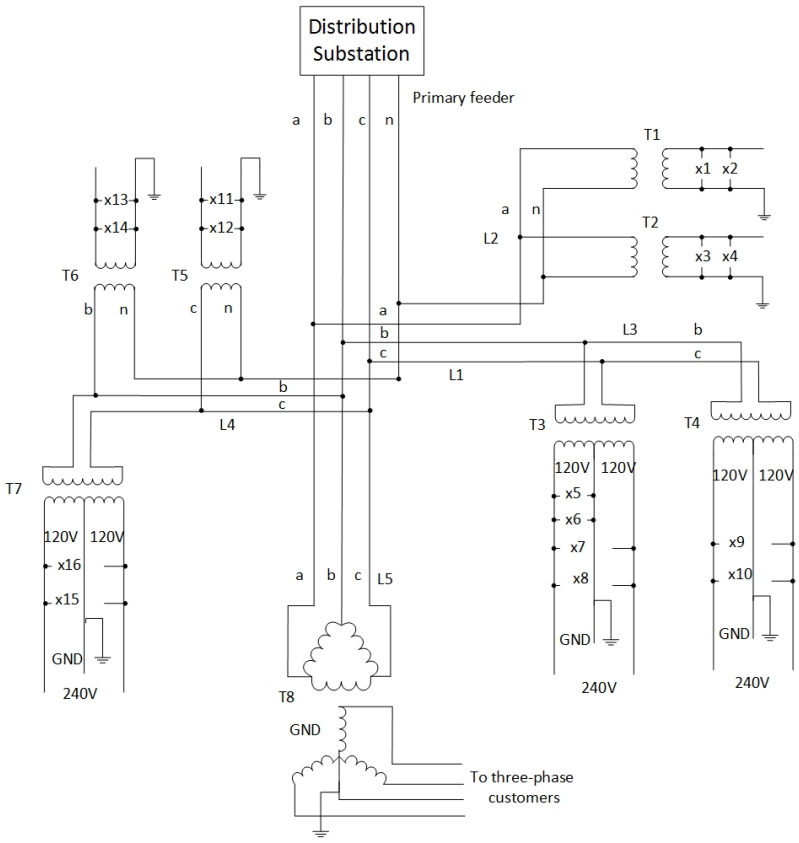


Figure 2.1: Illustration of a distribution system. Labels a, b, and c represent the three phases. L stands for a lateral, T stands for a transformer, and x denotes a customer.

The majority of the electric power is supplied by three-phase generators. In balanced conditions, the electric power circuits are 3-phase circuits and the three voltage phasors, V_{an}, V_{bn}, and V_{cn}, differ only in their angles, with 120-degree differences between any pair. Residential customers can be served by either a 120/240 V three-wire secondary through a center-tapped transformer (e.g., $T3$, $T4$, $T7$) or a 120 V single-phase secondary through a single-phase transformer (e.g., $T1$, $T2$, $T5$, $T6$). Commercial customers are typically served by a 208 V or 480 V three-phase four-wire secondary through a three-phase transformer (e.g., $T8$).

The phase identification problem is defined as identifying the phase connectivity of each customer and structure in the power distribution network.

2.3 Phase Identification by Linear Dimension Reduction With Centroid-Based Clustering

2.3.1 Literature Review and Contributions of Our Work

Very few studies on phase identification have been carried out. The existing methods for solving the phase identification problem can be separated into two general approaches. In the first approach, only smart meter data and SCADA information are assumed to be available [39, 41, 42, 43]. In the second approach, special equipments such as micro-synchrophasors [44], signal generators and discriminators [45] need to be installed to accurately identify the phase of distribution system customers and/or structures.

In the first approach, 0-1 integer linear programming and correlation-based methods are proposed to solve the phase identification problem. The phase identification problem is formulated

as a 0-1 integer linear programming problem where the phase connection of smart meters are treated as binary variables. Tabu search [41] and branch & bound search [42] are used to solve the integer optimization problem. There are two drawbacks associated with the 0-1 integer programming method. The first drawback is its computational complexity. A typical distribution feeder serves 1000 to 3000 customers on average. Therefore, the 0-1 integer programming problem for phase identification has thousands of binary decision variables, which requires daunting computational time. The second drawback is its low tolerance for erroneous and missing measurements. The existing methods only work when there are no unmetered loads or erroneous load measurements.

In correlation-based methods [39, 43], correlation coefficients or R^2 (coefficient of determination) are calculated between the voltage profile of individual smart meters and the voltage profile of the substation on each phase. These correlation coefficients or R^2 are assumed to have the highest value when the customer's phase is correctly labeled. Although correlation-based methods have been shown to be effective in identifying single-phase customers, it is not clear if the method can be successfully applied in the distribution circuits where the majority of the loads are connected to two-phase laterals. In addition, the algorithm may incorrectly label customers on the same single-phase secondary differently.

In the second approach, micro-synchrophasors, signal generators and discriminators are leveraged to accurately identify the phase of each customer. In [44], micro-synchrophasors are deployed at the target bus for phase identification. Micro-synchrophasors can measure voltage phase angles in addition to voltage magnitude. The main idea behind the method is that the correct customer phase label should yield the highest voltage magnitude and phase correlation with the corresponding phase at the substation. The advantage of the micro-synchrophasor approach is that

the method is applicable to all types of customer phase connections. In [45], a signal generator is deployed at the distribution substation and signal discriminators are deployed at the target customer sites to accurately identify the phases of smart meters. The disadvantage of the methods in the second approach is the expensive capital and maintenance costs for the additional equipments.

In this work, an innovative constrained k-means clustering algorithm of smart meter data is proposed to solve the phase identification problem. Instead of directly using the voltage time series data, we propose to first extract unique features from the voltage time series of smart meters. Then we define customer phase constraints by exploiting the known information about line configurations in the network connectivity model. At last, a constrained k-means clustering algorithm is applied to accurately identify the phase connection of each customer.

In light of the existing literature, the unique contributions of this work are as follows:

1. The proposed phase identification algorithm utilizes the known information about line configurations in the network connectivity model to avoid mislabeling of the customers on the same secondary feeder which can occur in the existing methods.

2. The proposed phase identification algorithm is computationally efficient compared with the 0-1 integer linear programming method and the correlation-based methods.

3. The proposed phase identification algorithm can identify phase connections with high accuracy in distribution circuits where the majority of loads are connected to two-phase laterals.

4. The proposed phase identification algorithm can still determine the phase connections of metered customers when the distribution circuit has some unmetered customers.

2.3.2 Technical Methods

The framework of our proposed phase identification algorithm by clustering smart meter data is illustrated in Figure 2.2. In the first step, voltage measurements are collected from smart meters and the SCADA system. In the second step, we normalize the customer voltage time series by their standard deviations and apply principal component analysis (PCA) on the normalized time series to extract the top q components. In the third step, we define the constraints in the clustering process by inspecting the network connectivity data. The k-means constrained clustering method is then applied to partition customers into clusters. At last, we identify the phase of each cluster by solving a minimization problem. The rest of this section is divided into three parts. First, we briefly review the methods in clustering of time series data. Second, the k-means constrained clustering algorithm for smart meter data is presented. Third, the algorithm for identifying the phase of each cluster is introduced.

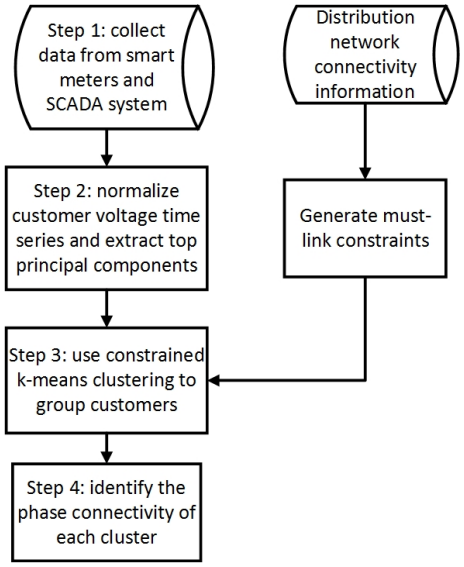


Figure 2.2: Diagram of the phase identification procedure.

2.3.2.1 Brief Review of Clustering Time Series Data

The goal of clustering is to identify the structure in an unlabeled dataset by objectively organizing data into homogeneous groups such that the objects in the same group are more similar to each other than those in different groups [46]. Various algorithms have been developed to cluster time series data. One of the widely used clustering algorithms is k-means, in which the objects are divided into k clusters so that the within-cluster sum of squares is minimized. Though typically it is not practical to find the minimal sum of squares among all possible partitions, many algorithms have been proposed to find local optimal solutions [47].

Almost all clustering algorithms require a similarity or distance function. There are many different types of distance functions. We only consider two of them here. The first one is Euclidean distance. If a_i and a_j are two p-dimensional time series, then their Euclidean distance is defined by

$$d_E = \sqrt{\sum_{k=1}^{p}(a_{ik} - a_{jk})^2} \tag{2.1}$$

Another type of distance function is related to Pearson's correlation coefficient. For two p-dimensional time series a_i and a_j, their Pearson's correlation factor is defined by

$$cc = \frac{\sum_{k=1}^{p}(a_{ik} - \mu_i)(a_{jk} - \mu_j)}{s_i s_j} \tag{2.2}$$

where μ_i and μ_j are the mean values of a_i and a_j, and $s_i = \sqrt{\sum_{k=1}^{p}(a_{ik} - \mu_i)^2}$ [46]. Then the distance between a_i and a_j can be defined based on cc as $d_1 = 1 - cc$ or $d_2 = (\frac{1-cc}{1+cc})^\beta$, $(\beta > 0)$ [48].

Smart meter time series data are high-dimensional. It is not desirable to work with high-dimensional noisy raw data in practice [46]. Therefore, we adopt a feature-based clustering method for the phase identification problem. Drawing features from data often requires expert knowledge

of the data, but in the phase identification problem, little knowledge is known on what features are important. PCA is a useful tool to reduce the data dimension and extract key features hidden in the time series data. PCA transforms a dataset into a new set of uncorrelated variables called principal components (PCs). PCs are ordered such that the first component retains the most of the variation in the original variables, the second component retains the second most of the variation, and so on [49]. In this work, PCA is used to select the most important features of the voltage time series data by picking the first q components. Euclidean distance in the chosen principal components' space will be used as the distance metric in the subsequent clustering process.

2.3.2.2 Clustering of Smart Meter Data with Constraints

The intuition behind identifying phase connectivity through clustering of voltage time series data is that the distribution system is typically operated in an unbalanced manner. The unbalanced impedances and electric loads on three phases lead to unbalanced line currents and voltages [29]. This implies that the trajectory of voltage time series of customers with the same phase connectivity will have more similar behavior than those with different phase connectivity. Instead of working directly with the raw voltage data, a feature-based clustering approach is used with features extracted from the voltage time series by PCA. Preprocessing including normalization and centering of the raw voltage data is conducted before applying PCA. We will show in the case study in Section 2.3.3 that a small number of features can yield very accurate clustering results.

The goal of clustering the voltage data from smart meters is to identify distinct groups of customers such that all customers in the same group have the same phase connectivity. Using the distribution feeder shown in Figure 2.1 as an example, customers $x7$, $x8$, $x9$, $x10$, $x15$, and $x16$ are all connected to phase BC through a three-wire system (120/240 V) and they should be

clustered into the same group. Similarly, consumers $x1$, $x2$, $x3$, and $x4$ should also be in one cluster because they are all connected to phase A and have the same voltage level (120 V). Before applying the clustering algorithm, we first separate customers based on their service voltage levels (120 V, 120/240 V, 208 V, 277 V, 480 V). These voltage levels can be easily identified by inspecting the voltage magnitude data from smart meters. The algorithm proposed in this work aims at clustering customers of the same voltage level. For example, meters of 120/240 V three-wire service have 6 possible phase connections: AB, BC, CA, AN, BN, and CN; meters of 120 V single phase service have 3 possible phase connections: AN, BN, and CN.

Various studies have been carried out to improve clustering/learning performances by utilizing constraints from background knowledge [50, 51, 52, 53]. In reference [50], two kinds of hard constraints are introduced: *must-link* constraints and *cannot-link* constraints. Must-link constraints specify that two data points have to be in the same cluster; cannot-link constraints specify that two data points cannot be in the same cluster. The must-link or cannot-link constraints for the phase identification problem can be formed based on the network connectivity information, which is typically available for power distribution systems. The network connectivity information includes line segment configurations and the connectivity between customers, distribution transformers, laterals, and primary feeders. If two customers are connected to the same secondary laterals and have the same voltage level, then they must have the same phase connectivity and should be linked together in the clustering process. For example, in Figure 2.1, customers $x7$, $x8$, $x9$, and $x10$ are all connected to the same lateral $L3$, and receive power through a three-wire (120/240 V) configuration. Therefore, these customers should be grouped into the same cluster. On the other hand, customers $x7$ and $x15$ should not be linked to each other because they are connected to different laterals.

A scheme is introduced in [51] for constrained k-means clustering. It is similar to the standard k-means clustering algorithm except that in the constrained clustering algorithm, each data point is assigned to the closest cluster such that it does not violate the constraints. The phase identification problem has must-link constraints where certain data points must be in the same cluster. We first put customers on the same laterals into a subset. Then an augmented k-means clustering algorithm is performed to the subsets themselves to obtain the full partition. Let $D = D_1 \bigcup D_2 \bigcup ... \bigcup D_n$ be the whole dataset, and $D_1, ..., D_n$ are the subsets in which every data point is linked together by the constraints. If a data point is not linked to any other data point, then it forms a subset in D itself. The constrained k-means clustering algorithm for phase identification is described in Algorithm 1, which is a modification of the scheme in [51]. As mentioned in Section 2.3.2.1, it is difficult to find the optimal result(s) by k-means clustering. To get a relatively good clustering result in our approach, the clustering algorithm is performed multiple times with different sets of random initial cluster centers. The clustering result with the smallest sum of squared distances is selected in the end.

Algorithm 1 Constrained k-means clustering algorithm

1: **procedure** CON-K-MEANS($D = D_1 \bigcup D_2 \bigcup ... \bigcup D_n$)
2: Choose data points randomly from D as the initial cluster centers $C_1, ..., C_k$.
3: Calculate each subset D_i's distance to each cluster. The distance is defined as the sum of squared distances of all the data points in D_i with the cluster center.
4: Assign each subset to the cluster that has the minimum summed distance.
5: For each cluster C_i, update its center by averaging all the data points that have been assigned to it.
6: Iterate between (3) and (5) until convergence.
7: **return** $\{C_1, ..., C_k\}$.
8: **end procedure**

2.3.2.3 Identify the Phase Connectivity of Each Cluster

Once the customers are clustered as described in Section 2.3.2.2, the next and last step is to identify the phase of each cluster. Since the customers in the same cluster should have the same phase connection, we can identify the phase of each cluster by picking a small number of customers from that cluster and identify their phase. This is a huge workload reduction compared with performing phase identification on every single customer. One may identify the phase of these few customers by micro-synchrophasors, signal generators and discriminators as in [44, 45].

However, to further reduce the computational workload, and to save the expense of equipments used in [44, 45], we can identify the phase of each cluster by a one-to-one matching between the set of clusters and the set of possible phase connections. The one-to-one matching can be found by solving the following minimization problem. Suppose there are k clusters to be identified with centers $C_1, ..., C_k$, and there are k substation voltage time series on the k possible phases. The k substation voltage series are centered and normalized by their standard deviations, and then projected onto the chosen principal components' space used for clustering. Let $V_1, ..., V_k$ be the coordinates of the k voltage series in the chosen principal components' space, and let $f : \{C_1, ..., C_k\} \to \{V_1, ..., V_k\}$ be an unknown bijection between the cluster set and the substation voltage set. The solution of the minimization in (2.3) is the one-to-one matching for phase identification. The phase of each cluster's paired voltage data is the cluster's identified phase.

$$\underset{\forall \text{ bijection } f:\{C_1,...,C_k\}\to\{V_1,...,V_k\}}{\arg\min} \sum_{i=1}^{k} d_E(C_i, f(C_i))^2 \tag{2.3}$$

Here $d_E(C_i, f(C_i))$ is the Euclidean distance between C_i and $f(C_i)$. The minimization can be solved by exhaustive search, because there are only $k!$ possible bijections, where k is small (e.g., $k = 3$ at 120/240 V level).

Compared to the load matching approach in [42], which assumes aggregated electricity consumption of all customers matches that of the substation, our proposed method is less sensitive to the presence of unmetered customers.

2.3.3 Case Study

In this section, the proposed phase identification method is validated through a case study of a distribution feeder in Southern California Edison's service territory. The results show that the constrained k-means clustering algorithm yields highly accurate phase connectivity on a typical distribution feeder.

2.3.3.1 Description of Datasets and Preprocessing of Data

The distribution feeder used for case study is a 12.47 kV network with a peak load of about 5.2 MW. The feeder serves about 1500 customers. The majority of the customers are residential customers.

The raw data collected to test the phase identification algorithm include: 1) hourly smart meter readings of voltages; 2) feeder line-to-line voltage readings of three phases from the SCADA system; 3) network connectivity of the distribution system. The number of a smart meter's readings varies by month. In months with 30 days, there are 720 readings (yielding measurement vectors of dimension 720), while months with 31 days have 744 reading hours. The SCADA system only records new feeder measurements when the difference between the new measurement and the previous measurement exceeds certain threshold. For example, the threshold setting for the line-to-line voltage is 0.02 kV. At last, to evaluate the accuracy of the proposed phase identification method, the correct phase connectivity of each meter is also gathered to serve as the ground truth.

Since the SCADA readings are recorded at nonuniform timestamps, linear interpolation is used to create a new set of voltages that have the same timestamps as the smart meter readings. All the readings are centered and normalized by their standard deviations. PCA and k-means clustering are performed on the readings of the same time period with the same timestamps. The timestamps are chosen such that most meters have a complete set of measurements. A smart meter is removed from the case study if it has missing readings at the chosen timestamps in the study period. In the testing distribution feeder, most of the customers are served by a three-wire system (120/240 V) based on the smart meter voltage levels, and all of them are connected to phases AB, BC, or CA. A few customers are served by three-phase laterals and there is no need to perform phase identification for these customers. Less than 1% of the customers are served by two-wire single-phase systems (120 V). Due to the small number of datasets, they are removed from the clustering process and their phase connectivity can be identified using methods introduced in [44, 45].

After preprocessing the test data, about 1500 customers/meters need to be clustered into 3 groups: phase AB, phase BC, and phase CA. PCA is conducted on the preprocessed time series data. Only the first two principal components are used to calculate Euclidean distances among customers. Based on the simulation results, including additional principal components does not further improve the performance of the phase identification results. The phase of each cluster is identified by finding the bijection described in Section 2.3.2.3. In this case, the bijection is between 3 clusters and the substation voltages of phases AB, BC, and CA.

Table 2.1: Clustering Result

Unconstrained Clustering Results of August 2015				
Cluster	Identified Phase	Number of Meters	Accuracy	Overall Accuracy
1	AB	674	92.58%	87.55%
2	BC	518	87.64%	
3	CA	246	73.58%	
Constrained Clustering Results of August 2015				
Cluster	Identified Phase	Number of Meters	Accuracy	Overall Accuracy
1	AB	636	98.27%	90.40%
2	BC	560	87.68%	
3	CA	242	76.03%	
Unconstrained Clustering Results of September 2015				
Cluster	Identified Phase	Number of Meters	Accuracy	Overall Accuracy
1	AB	678	93.36%	93.12
2	BC	547	93.60%	
3	CA	244	91.39%	
Constrained Clustering Results of September 2015				
Cluster	Identified Phase	Number of Meters	Accuracy	Overall Accuracy
1	AB	645	98.29%	97.28%
2	BC	559	97.67%	
3	CA	265	93.96%	
Unconstrained Clustering Results of October 2015				
Cluster	Identified Phase	Number of Meters	Accuracy	Overall Accuracy
1	AB	662	95.02%	93.09%
2	BC	531	93.60%	
3	CA	254	87.01%	
Constrained Clustering Results of October 2015				
Cluster	Identified Phase	Number of Meters	Accuracy	Overall Accuracy
1	AB	630	99.84%	97.86%
2	BC	550	98.36%	
3	CA	267	92.13%	

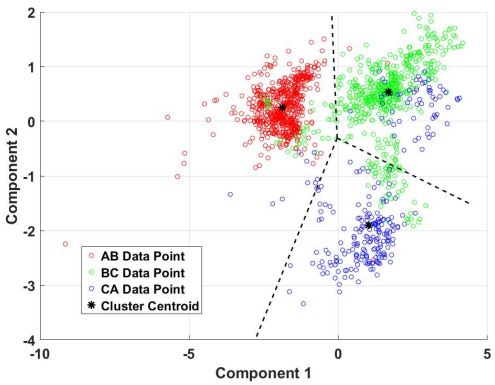


Figure 2.3: Principal components of August voltage time series data.

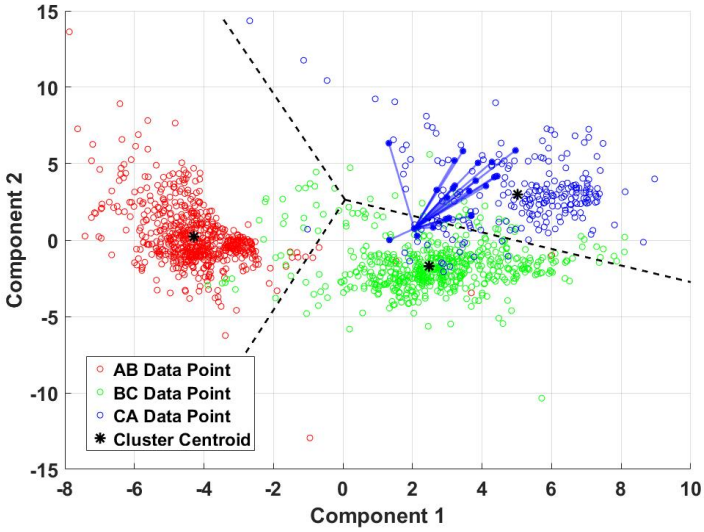


Figure 2.4: Principal components of October voltage time series data.

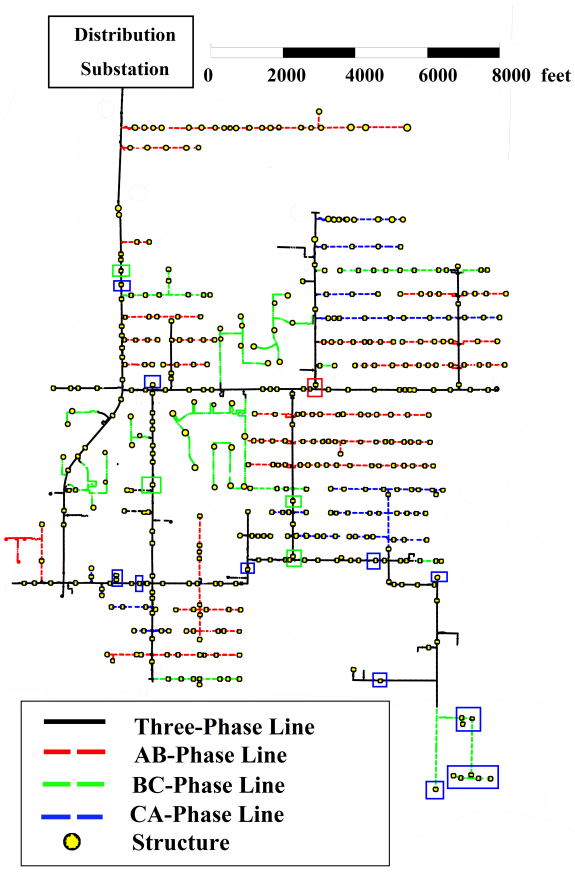


Figure 2.5: Phase identification results.

2.3.3.2 Clustering Results

Three months of SCADA, smart meter, and network connectivity data are collected from August 1, 2015 to October 31, 2015. 1438 smart meters' data are available in August. According to the ground truth, 629 of them are connected to phase AB laterals, 557 of them are connected to phase BC laterals, and 252 of them are connected to phase CA laterals. In September, 1469 smart meters' data are available. According to the ground truth, 638 of them are connected to phase AB laterals, 571 of them are connected to phase BC laterals, and 260 of them are connected to phase CA laterals. In October, 1447 smart meters' data are available. According to the ground truth, 633 of them are connected to phase AB laterals, 562 of them are connected to phase BC laterals, and 252 of them are connected to phase CA laterals.

The clustering and phase identification results are shown in Table 2.1, which can be interpreted as follows. The clustering and phase identification algorithms group the smart meters into three clusters. The phase identified for each cluster is listed in the identified phase column. If a meter is assigned to a cluster whose identified phase is the same as the meter's actual phase, then it is assigned to the correct cluster. The accuracy column shows the percentage of correct assignments in each cluster and the overall accuracy column shows the overall accuracy of the algorithm.

Table 2.1 shows that the phase identification algorithm of both unconstrained and constrained clustering achieve at least 90% overall accuracy in September and October. In addition, in all months, the constrained clustering algorithm yields a higher accuracy than the unconstrained k-means clustering algorithm. The constrained clustering outperforms the unconstrained clustering by letting must-link constraints pull a linked meter back to the correct cluster when it is near the boundary of two clusters.

Figure 2.3 and 2.4 show the distributions of two months' voltage data points in the space of the first two principal components. Dashed lines are the boundaries of Voronoi cells associated with cluster centers derived from the constrained clustering algorithm. Figure 2.4 also shows an example of how the constrained clustering algorithm improves the accuracy. In Figure 2.4, a set of blue data points grouped by must-link constraints are connected by solid lines. Although this set of data points are separated by a boundary, they are closer to the CA cluster as a whole. Therefore, they are assigned to the CA cluster, which is the correct phase. Without these must-link constraints, some of the data points will be assigned to the BC cluster, which is incorrect. Figure 2.3 and 2.4 show that data points of different phases are separated in the space of the first two principal components. However, there are more data points of phase BC and CA overlapped in Figure 2.3 than Figure 2.4. As a result, the overall accuracy of phase BC and CA are lower when using data from August, compared with October.

Figure 2.5 shows the clustering results on the distribution circuit map based on the smart meter data of October 2015. In Figure 2.5, each line is colored according to its actual phase. Each structure (e.g., transformer) is represented by a small dot. The three-phase black lines are primary feeder lines. Structures can be connected to primary feeder lines through a three-wire (120/240 V) system, so they can be connected to phases of AB, BC, and CA. A colored rectangle is overlayed on top of a structure if it is assigned to a wrong cluster. The color of the rectangular shows the identified phase of the cluster. Note that the number of structures is smaller than the number of smart meters/customers as a distribution transformer typically serves several customers.

In summary, the results above show that the constrained k-means clustering algorithm groups the meters by phase at high accuracy, and the identification method correctly identifies the

phase of each cluster, in a circuit where the majority of customers are connected to two-phase laterals.

The proposed algorithm is computationally more efficient than the integer linear programming method. The running time of the proposed algorithm is the sum of the running time of the PCA step and the k-means clustering step. The running time of the PCA is $O(p^2m + p^3)$ [54], and the running time of Lloyd's algorithm for k-means clustering is given by as $O(mkqi)$. Here p is the number of dimensions of the raw time series data, m is the number of data points (i.e., the number of meters), k is the number of clusters, q is the number of principal components used in clustering, and i is the number of algorithm iterations. In the case study, the typical value for i is less than 10. Therefore, the total running time of the proposed algorithm increases linearly with m. On the other hand, the running time of branch and bound search, which solves the integer linear programming problem, is not bounded by a polynomial function of m [55].

2.3.4 Conclusion

An innovative distribution system phase identification algorithm using constrained k-means clustering of smart meter data is proposed in this work. The proposed algorithm leverages the network connectivity information to avoid mislabeling customers on the same secondary feeder. Utilizing only the smart meter and SCADA information, the proposed algorithm is not only computationally efficient but also yields high accuracy. A real-world distribution feeder is used to validate the algorithm. Case study results show that the constrained k-means clustering algorithm outperforms the unconstrained algorithm. The overall accuracy of the proposed algorithm is at least 90%.

Table 2.1 shows that this algorithm performs better during some months than others. Future research is needed to determine over which time periods the phase identification algorithm

performs best. In addition, it is desirable to develop algorithms that not only perform phase identification but also estimate the confidence level of clustering result for each individual meter.

2.4 Phase Identification by Nonlinear Dimension Reduction With Density-Based Clustering

2.4.1 Literature Review and Contributions of Our Work

Currently, most electric utilities conduct phase identification using special phase meters [56][21]. Typically, two phase meters/units are used. One unit is located at the substation to serve as the reference. The other is called the field unit and is located at the smart meter/structure of interest in the distribution feeder. The working mechanism of these special phase meters is very similar to that of the phasor measurement units except that the phase meters are mobile. With GPS time, the phase angle difference between the reference point and the field structure can be accurately measured, which then determines the phase connectivity of the field structure. Although phase meters provide highly accurate phase identification results, this solution is very time consuming and labor intensive, which make it unsuitable for large-scale deployment.

The existing data-driven algorithms use electric load and voltage magnitude measurements from the AMI to identify the phase connections of the smart meters and structures in the distribution network. These algorithms include supply and consumption balancing [41][42], linear regressions and correlation analysis [39][57], and constrained k-means clustering algorithm (CK-Means) [30]. However, the existing data-driven algorithms have three drawbacks. First, all these methods assume that the number of phase connections are known in advance. Second, these meth-

ods can not provide accurate phase identification when there is a mix of both phase-to-neutral and phase-to-phase connected smart meters and structures. Third, the existing methods are quite sensitive to the level of unbalance in a distribution feeder. The proposed AMI data driven phase identification algorithm addresses these drawbacks by leveraging a nonlinear dimensionality reduction technique that can extract hidden features from voltage time series and using the CHC algorithm to dynamically create smart meter clusters with arbitrary shapes. Field validation results show that the proposed algorithm outperforms the existing methods in all of the 5 distribution feeders.

In this work, an AMI data driven machine learning algorithm is developed to solve the phase identification problem. This algorithm leverages voltage magnitude data recorded by the AMI to identify the phase connection of each smart meter and structure. A nonlinear dimensionality reduction technique is first used to extract key features from the voltage time series. A constraint-driven hybrid clustering (CHC) algorithm is then developed to separate smart meters/structures into various clusters. Finally, the phase connection of each cluster can be identified by performing field validations on the phase of very few smart meters. Comprehensive case studies are conducted on 5 distribution circuits, which went through detailed field validations. The AMI data driven machine learning algorithm has yielded high accuracies on all circuits. In addition, this work discovers that more granular voltage readings will lead to even more accurate phase identification results.

Compared to the existing data-driven phase identification algorithms, the proposed method has the following advantages:

1. The proposed algorithm does not require prior knowledge about the number of phase connections in the distribution system. Most of the existing AMI data driven methods need the number of phase connections as an input parameter.

2. The proposed algorithm works well with distribution feeders that have both phase-to-neutral and phase-to-phase connections. Most of the existing techniques are only capable of identifying the phase connections in distribution feeders with only phase-to-neutral connections or phase-to-phase connections.

3. The accuracy of the proposed phase identification algorithm is not very sensitive to the level of unbalance in a distribution feeder.

2.4.2 Drawbacks of the Existing Data-driven Phase Identification Methods

Three main drawbacks of the existing phase identification methods are studied in detail below. As the CK-Means method is the most promising algorithm among the existing data-driven phase identification methods, it will be used as an example in the performance evaluation. A comprehensive study is conducted on 5 distribution feeders and 18 data sets to analyze the impact of unbalance level and the mix of phase connection types on the phase identification accuracy for the CK-Means method.

The general descriptions of the 5 distribution feeders and 18 data sets are shown in Table 2.2. The feeder and smart meter data is provided by the Pacific Gas & Electric Company and Southern California Edison. The number of customers, feeder voltage level, proportion of the major phase connection types, and feeder peak load are listed in the second column of the table. A distribution feeder can have 3 possible phase-to-neutral connections, AN, BN, and CN, and/or 3 possible phase-to-phase connections, AB, BC, and CA, where A, B, C, and N denote the three phases' wires and the neutral wire. 2 months of smart meters' voltage data with 5-minute granularity is gathered from feeder 1, 2, and 3. 6 months of smart meters' voltage data with hourly granularity is

Table 2.2: Descriptions of the Distribution Feeders

Feeder	Number of Customers, Feeder Voltage, and Peak Load	Month	Data Set
1	3200 customers (99.8% phase-to-neutral), 12.47 kV, 4.4 MW.	Nov 2016	s_1
		Dec 2016	s_2
2	4800 customers (98.8% phase-to-neutral), 12.47 kV, 8.3 MW.	Nov 2016	s_3
		Dec 2016	s_4
3	4000 customers (97% phase-to-neutral), 12.47 kV, 6.4 MW.	Nov 2016	s_5
		Dec 2016	s_6
4	1500 customers (100% phase-to-phase), 12.47 kV, 5.2 MW.	Aug 2015	s_7
		Sep 2015	s_8
		Oct 2015	s_9
		Nov 2015	s_{10}
		Dec 2015	s_{11}
		Jan 2016	s_{12}
5	2400 customers (84% phase-to-phase), 12.47 kV, 8.5 MW.	Aug 2015	s_{13}
		Sep 2015	s_{14}
		Oct 2015	s_{15}
		Nov 2015	s_{16}
		Dec 2015	s_{17}
		Jan 2016	s_{18}

gathered from feeder 4 and 5.

In feeder 1, 2, and 3, some meters have missing voltage readings at different time intervals, making up 9%, 21%, and 18% of the total customer population respectively. The missing readings are filled in using the k-nearest neighbor (k-NN) imputation method. A meter's missing readings are imputed using the average values of the five nearest neighbor meters' corresponding readings. The distance between meters are measured by the Euclidean distance of the voltage time series of the corresponding meters.

In order to make the results comparable, the hourly average voltage magnitudes are calculated for feeder 1, 2, and 3. The hourly average voltage magnitudes are used as inputs in this section. Each of the 18 data sets includes one month of voltage magnitude data from a feeder. The

drawbacks of the existing data-driven phase identification algorithms are explored in the next three subsections.

2.4.2.1 Number of Phase Connections

In order to solve the phase identification problem, the supply and consumption balancing approach [41][42] requires the number of phase connections in the distribution feeder as an input. In fact, the problem formulation in [41][42] only allows the identification of phase-to-neutral connections where the number of phase connections is 3. In the linear regression and correlation analysis [39][57], the number of phase connections in the feeder is also a mandatory input. In fact, both linear regression and correlation analysis work well when there are only three phase-to-neutral connections. The k-means clustering algorithm is used in the CK-Means method [30], where the number of phase connections/clusters needs to be known as prior knowledge. When applying the CK-Means method to identify the phase connections of the 5 distribution feeders, the number of clusters is set to be 3 for feeders 1 to 4, given that over 97% of the smart meters in these feeders only have 3 connection types. The number of clusters is set to be 6 for feeder 5.

2.4.2.2 Impact of Unbalance Level on the Phase Identification Accuracy

This subsection evaluates the impact of the distribution feeder's unbalance level on the phase identification accuracy of the CK-Means algorithm. The CK-Means algorithm works as follows: The voltage magnitude measurements are first standardized. Linear features are then extracted by using principal component analysis (PCA) and the top d components are selected. To provide a fair comparison with the proposed phase identification algorithm in Section 2.3.3, the number of principal components is set to 30. Next, the data points in the low-dimensional space are clustered

by using a constrained k-means clustering algorithm. Must-link constraints are derived from the distribution feeders' connectivity information, which is typically available from the Geographical Information System (GIS). The must-link constraints state that if some smart meters are connected to the same lateral or transformer, then they must be linked together and grouped into the same cluster. To identify the phase of each cluster, field validations are performed on a must-link group of at least 20 smart meters that has the least mean squared distance to the cluster center.

The CK-Means algorithm is applied on the 18 voltage time series from the 5 distribution feeders. The phase identification accuracy is calculated based on independent field validations conducted by the electric utility companies. To measure the level of unbalance of a distribution feeder, define $u(t)$ as the level of unbalance of a feeder at time interval t:

$$u(t) = \frac{|I_A(t) - I_m(t)| + |I_B(t) - I_m(t)| + |I_C(t) - I_m(t)|}{3I_m(t)} \tag{2.4}$$

where $I_m(t) = \frac{1}{3}(I_A(t) + I_B(t) + I_C(t))$ is the mean of the distribution substation line currents of the three phases. $u(t)$ can be interpreted as the ratio of the average three-phase current deviation to the mean. The average level of unbalance for $u(t)$ over a month is calculated for each data set.

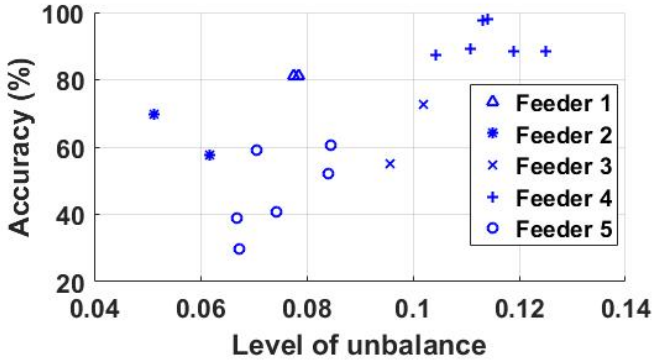


Figure 2.6: The phase identification accuracy of the CK-Means method under different levels of unbalance.

Figure 2.6 plots the phase identification accuracy against the level of unbalance. It shows that the CK-Means algorithm is very accurate for the highly unbalanced data sets. As the level of unbalance decreases, the phase identification accuracy drops quickly. This result is very intuitive. Imagine there is a perfectly balanced distribution feeder whose three phase wires have the same load distribution all the time. In this case, the level of unbalance should be zero. Therefore, it is impossible to distinguish the phase connections of the smart meters on the three phases with unsynchronized voltage magnitude measurements.

2.4.2.3 A Mix of Phase-to-Neutral and Phase-to-Phase Connections

In general, the existing data-driven phase identification algorithms do not perform well for the distribution feeders with a mix of phase-to-neutral and phase-to-phase connections. For example, Figure 2.6 shows that the phase identification accuracy is the lowest for feeder 5. This is because feeder 5 not only has a lower degree of unbalance, but also has all 6 possible phase connections types, AN, BN, CN, AB, BC, and CA. In this case, the default phase identification accuracy is only 16.7% instead of 33.3% for the distribution feeders with only three possible phase connections.

2.4.3 Technical Methods

The overall framework of the proposed phase identification algorithm is illustrated in Figure 2.7. The phase identification methodology involves three stages. In stage 1, voltage magnitude measurements are collected from the smart meters. Each smart meter's readings are centered and normalized by their standard deviation. Key features are then extracted from the preprocessed volt-

age time series with a nonlinear dimensionality reduction method. In stage 2, the CHC algorithm is leveraged to cluster the low-dimensional data points generated in stage 1. In stage 3, the phase connection of each cluster is identified by performing field validations on a very small number of smart meters. The three stages are explained in detail below.

2.4.3.1 Stage 1: Feature Extraction from Voltage Time Series

It is undesirable to directly work with raw voltage readings, which are high-dimensional and noisy. Therefore, in the first stage, dimensionality reduction techniques will be applied to extract key features from the raw voltage time series. The extracted features will then be fed into the CHC algorithm in stage 2.

Dimensionality reduction techniques can be divided into two categories, linear dimensionality reduction methods and nonlinear ones. Linear dimensionality reduction techniques, such as PCA, are restricted to learning only linear manifolds. However, high-dimensional data typically lies on or near a low-dimensional, nonlinear manifold [58]. Furthermore, it is very difficult for linear

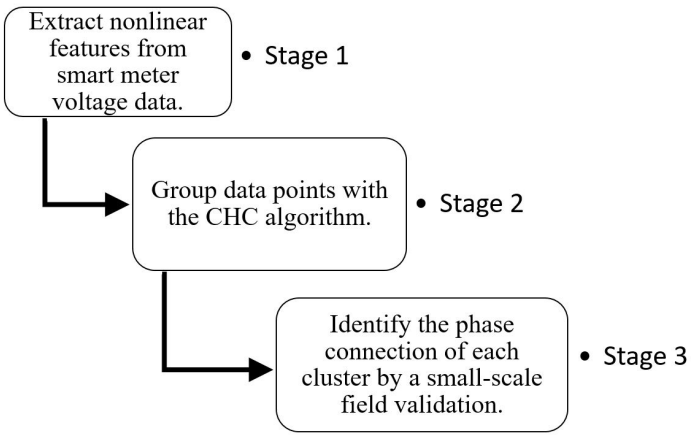


Figure 2.7: The overall framework of the proposed phase identification algorithm.

mappings to keep the low-dimensional representations of very similar points close together. This explains the lower accuracy of the phase identification algorithm using linear features for less unbalanced feeders. To address this problem, we turn to nonlinear dimensionality reduction methods. Many nonlinear dimensionality reduction techniques have been proposed, e.g., Sammon mapping [59], curvilinear components analysis (CCA) [60], Isomap [61], and t-distributed stochastic neighbor embedding (t-SNE) [58]. This work adopts t-SNE, because it has been shown to work well with a wide range of data sets and captures both local and global data structures. t-SNE improves upon SNE [62] by 1) simplifying the gradient calculation with a symmetrized version of the SNE cost function and 2) adopting a Student's t-distribution rather than a Gaussian distribution to compute the similarity between two points in the low-dimensional space [58].

The basic idea of t-SNE is to convert the high-dimensional Euclidean distances between data points into joint probabilities and represent the data points in a low-dimensional space, so that similar joint probabilities are preserved. Suppose we need to map a high-dimensional data set $X = \{x_1, x_2, ..., x_n\}$ to a low-dimensional data set $Y = \{y_1, y_2, ..., y_n\}$. Define p_{ji} as a joint probability of X. p_{ji} is a symmetric approximation of the conditional probability that x_i would pick x_j as its neighbor. The neighbors are picked in proportion to their probability density under a Gaussian distribution centered at x_i with a variance σ_i. p_{ji} is calculated as $p_{ji} = p_{ij} = (p_{j|i} + p_{i|j})/2n$, where $p_{j|i}$ is calculated as:

$$p_{j|i} = \frac{\exp(-\|x_i - x_j\|^2/2\sigma_i^2)}{\sum_{l \neq i} \exp(-\|x_i - x_l\|^2/2\sigma_i^2)} \tag{2.5}$$

In the same way, define q_{ji} as a joint probability in Y, but under a Student's t-distribution with one degree of freedom. Then q_{ji} can be calculated as:

$$q_{ji} = q_{ij} = \frac{(1 + \|y_i - y_j\|^2)^{-1}}{\sum_{l \neq m}(1 + \|y_l - y_m\|^2)^{-1}} \tag{2.6}$$

Then given X, the mapping Y is found by minimizing the Kullback-Leibler divergence between joint probability distribution P, in the high-dimensional space, and the joint probability distribution Q, in the low-dimensional space:

$$C = D_{\mathrm{KL}}(P\|Q) = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}} \tag{2.7}$$

The t-SNE algorithm requires three input parameters: 1) the output dimension d_{out} (typically selected to be either 2 or 3); 2) the initial dimension d_{in}, which is the dimension that the original data set is reduced to by PCA before performing t-SNE; 3) perplexity p, which is a measure of effective number of neighbors and controls σ_i. Since the objective function (2.7) is minimized using a gradient descent optimization that is initiated randomly, each run of t-SNE produces a slightly different mapping result. In practice, it is recommended to run t-SNE multiple times and select the result with the lowest cost function value in (2.7). More details of the t-SNE algorithm can be found in [58].

2.4.3.2 Stage 2: Group Data Points with the CHC Algorithm

After the preprocessed voltage time series are mapped to a 2-dimensional or 3-dimensional feature space through t-SNE, they need to be grouped into clusters. Three features of the phase identification problem need to be considered when designing the clustering algorithm. First, many utility companies do not know the number of phase connections for each of their distribution feeders. Second, the customers with the same phase connection in the low-dimensional feature space do not necessarily form a convex-shape cluster, which is very common in t-SNE applications [58][63][64]. Third, valuable distribution network connectivity information that defines the mapping between smart meters and laterals/transformers should be incorporated into the clustering algorithm.

In order to leverage the features of the phase identification problem, the CHC algorithm is developed and applied to solve the smart meter clustering problem. The proposed CHC framework synergistically combines the merits of an unsupervised density-based clustering algorithm and a supervised classification algorithm. This work selects the density-based spatial clustering of applications with noise (DBSCAN) [65] as the unsupervised clustering algorithm in the CHC framework, because it naturally incorporates the first two features of the phase identification problem. Unlike centroid-based or medoid-based methods, DBSCAN does not need the number of clusters as an input parameter. In addition, DBSCAN is capable of discovering clusters with arbitrary shapes.

DBSCAN separates data points into different clusters and noise/outliers. The noise/outliers do not belong to any cluster. However, in the phase identification application, all smart meters must have a particular phase connection. To mitigate this drawback, k-nearest neighbor (k-NN) classification is adopted as the supervised machine learning algorithm in the CHC framework to assign these outliers and points in the low-density region into one of the existing output clusters from DBSCAN. At last, the must-link constraints defined by the feeder connectivity model will be considered in reassigning smart meters connected to the same lateral/transformer to the same cluster.

2.4.3.2.1 Review of DBSCAN A brief review of DBSCAN is provided here. DBSCAN is one prominent example of density-based clustering approach with high computational efficiency. The good efficiency of DBSCAN is crucial for deploying phase identification algorithms in electric utilities with thousands of distribution feeders. The DBSCAN algorithm defines clusters and outliers based on four key concepts: ϵ neighborhood of a point, directly density-reachable, density-reachable, and density-connected. The algorithm requires two parameters: ϵ, the radius of neighborhood, and *MinPts*, the minimum number of data points in an ϵ neighborhood. The ϵ neigh-

Algorithm 2 The CHC algorithm

1: Run the DBSCAN algorithm on a preprocessed data set D with n data points with parameters ϵ and *MinPts*.
2: Define a threshold coefficient $\alpha \in (0, 1)$. Given the output of step 1, keep the data points from the clusters of size greater than or equal to αn as the training data set. Suppose there are c clusters kept. All the data points outside these clusters are "un-clustered" data points.
3: Assign all un-clustered data points to one of the c clusters with the k-NN algorithm.
4: With must-link constraints, the data set D can be divided into N groups $D_1, ..., D_N$. If a data point has no links to others, it forms a group itself. In each group D_i, the data points may have been assigned to different clusters. To enforce the constraints, assign all data points in group D_i to the cluster that contains the largest number of data points in D_i.
5: Return the final clustering result.

borhood of a point p is defined as the set of points in the data set with a distance to p less than ϵ. A point p is a core point if it has at least *MinPts* neighbors within the radius ϵ. These neighbors are directly density-reachable from p. A point q is density-reachable from p if there is a path $p, p_1, p_2, ..., p_m, q$ such that each point is directly reachable from the previous point. Two points are considered density-connected if they have a distance of less than ϵ. These four definitions allow us to define the transitive hull of density-connected points, forming density-based clusters. The points on the border of the clusters are called border points. Any point(s) not reachable from a core point is counted as an outlier or noise.

2.4.3.2.2 The CHC Algorithm The framework of the algorithm is shown in Algorithm 1. It requires four input parameters, α, k, ϵ, and *MinPts*. α is a threshold used to filter out very small clusters. k is the parameter in the k-NN algorithm representing the number of nearest neighbors.

The CHC algorithm has 5 steps. Step 1 runs the DBSCAN algorithm on features extracted by the t-SNE algorithm. Depending on the distribution of data points in the low-dimensional feature space, the DBSCAN output may include large clusters, small clusters, and noise/outliers. Step 2 filters out the points in the small clusters and noise/outliers and only keeps the large clusters as the

training data set for the next step. Step 3 classifies the points from small clusters and noise/outliers with k-NN algorithm using the training data points from the large clusters. Step 4 enforces the must-link constraints by assigning all smart meters connected to the same lateral/transformer to the same cluster. The final clustering results will be returned in step 5.

Note that researchers have proposed alternative approaches, such as C-DBSCAN [66] to integrate constraints into density-based clustering algorithms. In the C-DBSCAN algorithm, the data points from different clusters involved in a must-link constraint are simply forced to merge together. However, when the preprocessed voltage time series are mapped to the low-dimensional space, we often encounter cases where a very small number of meters connected to one phase are spread over two clusters representing two phases. To address this issue, in step 4 of the proposed CHC algorithm, we only reassign all the data points connected by a must-link constraint to the same cluster without affecting the grouping of other data points.

2.4.3.3 Stage 3: Phase Identification for Clustered Customers

The final stage identifies the phase connection of the clusters determined in stage 2. This can be accomplished by performing field validations on a small number of samples of smart meters with phase measurement tools [56][21]. The cost associated with the field validation is minimal as the number of customers that require phase measurement is quite small. To achieve the highest accuracy, the small sample of customers should be chosen as close to the clusters' centers as possible. Depending on the availability of must-link constraints, two sampling strategies can be implemented:

1. If there are no must-link constraints, then in each cluster choose m smart meters that are closest to the cluster center. Field validations can then be performed on these m smart meters.

The most frequent phase connection of these m meters is selected as the phase connection of all the customers in the cluster.

2. If must-link constraints are available, then in each cluster choose the group D_g that contains at least w customers and has the least mean squared distance to the cluster center. Field validations will be performed on any of the smart meters in group D_g. The phase connection of the group is selected as the phase connection of all the customers in the cluster.

2.4.4 Numerical Study

2.4.4.1 Experimental Design

Two types of experiments are designed below to 1) examine the performance of the proposed phase identification algorithm and 2) explore the impact of smart meter data granularity on the phase identification accuracy.

The first set of experiments compare the performance of the constrained k-means clustering algorithm with linear dimensionality reduction [30] and the CHC algorithm with nonlinear dimensionality reduction proposed in this work. The constrained k-means clustering algorithm with linear dimensionality reduction is referred to as "CK-Means" method. Both methods are evaluated over 18 hourly voltage time series gathered from 5 distribution feeders as described in Table 2.2.

The second set of experiments evaluate the impact of smart meter sampling frequency on the accuracy of the proposed phase identification algorithm. The experiments are conducted over 6 voltage time series gathered from 3 distribution feeders. The smart meters on distribution feeder 1-3 were configured to record voltage magnitudes every 5 minutes. The average voltage magnitudes with hourly, 15-minute, and 5-minute granularity are used as inputs.

2.4.4.2 Parameter Selection

A few parameters need to be set up in order to run the proposed phase identification algorithm. In the feature extraction stage, three parameters from the t-SNE algorithm need to be selected. The dimensionality of the PCA output and t-SNE input d_{in} is set to be 30. The perplexity p is set to be 100. Note that these two parameters can be tuned by running the optimization several times on a data set and picking the parameters that yield the best map [58]. The dimensionality of the t-SNE output d_{out} is typically set to be 2 or 3. For better visualization, we set d_{out} to 2. In fact, the case study results with $d_{\text{out}} = 2$ and $d_{\text{out}} = 3$ are very similar.

In the proposed CHC algorithm, three key parameters *MinPts*, ϵ, and α need to be tuned first. The typical ranges for the three parameters are 8 to 20 for *MinPts*, 1 to 3 for ϵ, and 0.005 to 0.01 for α. When tuning these parameters, the aim is to see the data points in the t-SNE space being clustered appropriately. For example, assume we select some initial settings for *MinPts*, ϵ, and α, and get the clustering results as shown in Figure 2.9. Intuitively, cluster 11 and 15 should be two separate clusters. If the initial parameter setting merges these two clusters, then the parameters need to be tuned so that they are separated in the clustering results. In this particular case, we should decrease ϵ and/or increase *MinPts* to separate cluster 11 and 15. Note that ϵ is the radius of neighborhood and *MinPts* is the threshold for determining if a point p is a core point or a border point in a cluster. The parameter α controls the number of output clusters. If the value of α is too large, then the phase identification accuracy will be lower. However, if the value of α is too small, then a large number of meters need to be field validated, which increases implementation costs. k, the parameter of the k-NN, can be selected to be equal to *MinPts*. At last, in the field validation, choose the must-link group with at least $w = 20$ customers.

2.4.4.3 Performance of the Proposed Phase Identification Algorithm

The phase identification accuracies of the CK-Means method and the proposed phase identification algorithm are calculated based on field validation results. For the proposed algorithm, 30 runs of t-SNE are conducted. 10 runs with the lowest cost function values are kept. The average accuracy over the 10 runs are reported in Table 2.3 and Figure 2.8.

As shown in the table, the proposed phase identification algorithm significantly outperforms the CK-Means method with all the data sets in terms of accuracy. On average, the proposed phase identification algorithm improves the identification accuracy by 19.81% over the CK-Means algorithm. Figure 2.8 shows that the improvement in phase identification accuracy varies by the

Table 2.3: Phase Identification Accuracies

Feeder	Data Set	Level of Unbalance	CK-Means Accuracy (%)	Proposed Algorithm Accuracy (%)
1	s_1	0.0785	81.21	93.06
	s_2	0.0776	81.18	93.62
2	s_3	0.0514	69.67	87.55
	s_4	0.0617	57.51	87.79
3	s_5	0.0956	54.91	83.94
	s_6	0.1019	72.78	82.83
4	s_7	0.1109	89.29	98.60
	s_8	0.1141	97.82	98.94
	s_9	0.1131	97.79	99.63
	s_{10}	0.1190	88.42	99.66
	s_{11}	0.1043	87.49	99.88
	s_{12}	0.1250	88.34	99.65
5	s_{13}	0.0673	29.80	73.18
	s_{14}	0.0668	38.80	73.32
	s_{15}	0.0705	59.07	67.01
	s_{16}	0.0742	40.56	88.19
	s_{17}	0.0846	60.49	87.11
	s_{18}	0.0842	52.02	89.84

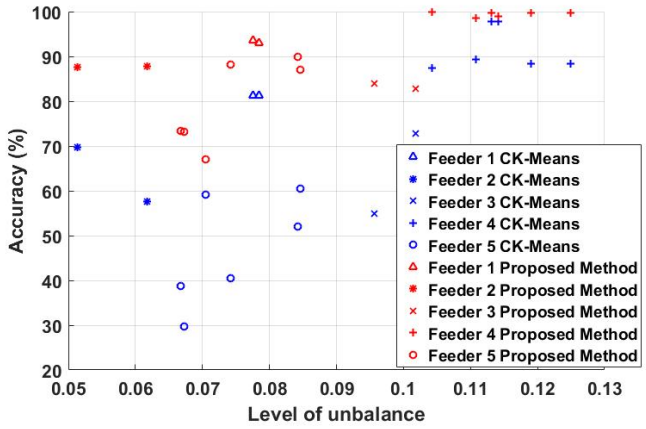


Figure 2.8: The phase identification accuracy with CK-Means and proposed algorithm.

unbalance level of the distribution circuit. The improvement is more significant for periods when the distribution feeder is less unbalanced.

The combinations of phase connections in the 5 testing feeders include 3 phase-to-neutral connections, 3 phase-to-phase connections, and a mix of all 6 possible connections. The accuracy of the proposed phase identification algorithm is very high under most cases. s_{13}, s_{14}, and s_{15} have relatively lower accuracy, because they have lower levels of unbalance and they have all 6 possible connections, which is more difficult to identify than other feeders. When the level of unbalance is higher, the accuracy is greatly improved in s_{16}, s_{17}, and s_{18}, whose accuracies are very decent for a feeder with all the 6 possible phase connections. Figure 2.9 illustrates the clustering result of data set s_{18} in the 2-dimensional t-SNE map, using the proposed phase identification algorithm. In the figure, each dot represents a smart meter. Figure 2.10 depicts the actual phase connection of each smart meter. By comparing Figure 2.9 and Figure 2.10, it is shown that the proposed phase identification algorithm not only groups phase-to-phase meters accurately, but also groups

phase-to-neutral meters with a high accuracy. Cluster 2, 11, 12, 13, and 15 each represents one of the phase-to-neutral connections AN, BN, and CN, as indicated by the arrows in Figure 2.9 and Figure 2.10.

As a comparison, Figure 2.11 shows the distribution of smart meters from data set s_{18} in the 2-dimensional PCA map. The data points are not well separated according to phase connection. From Figure 2.11 and Figure 2.10, it is clear that the nonlinear dimensionality reduction technique, t-SNE, does a much better job in extracting hidden features from the voltage time series during a less unbalanced period for the feeders.

As shown in Figure 2.9, the clusters are in different sizes and shapes. Some of the clusters are non-convex. The proposed CHC algorithm has a great advantage in identifying clusters with such data point distributions. Figure 2.9 also shows how the must-link constraints could improve the phase identification accuracy. In the top right cluster 8, a few data points are linked together.

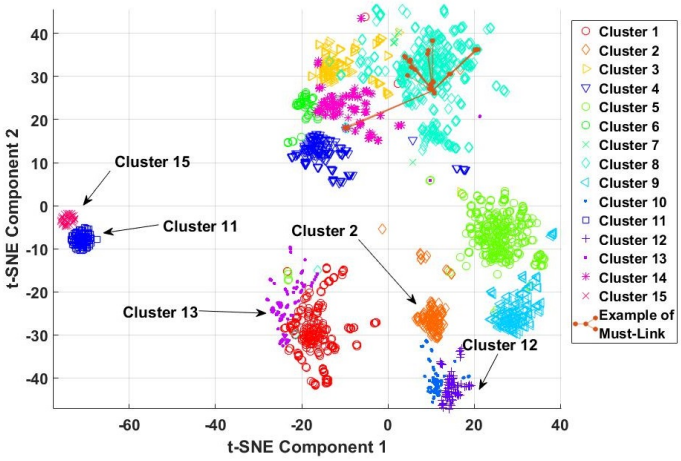


Figure 2.9: The clustering result in the 2-dimensional t-SNE map on data set s_{18}.

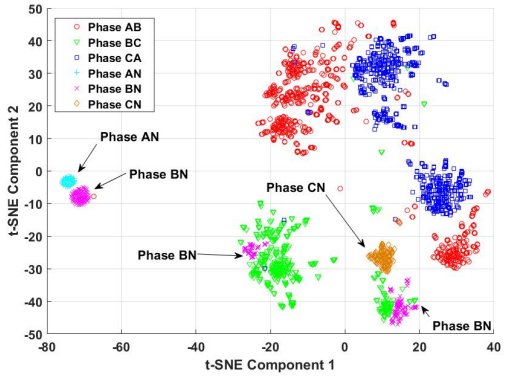


Figure 2.10: Field validated phase connections of data set s_{18} in the 2-dimensional t-SNE map.

Although a small number of the data points are located in cluster 14, they will eventually be assigned to cluster 8 due to the must-link constraint. From Figure 2.10, these data points should belong to cluster 8, which is connected to phase CA instead of phase AB.

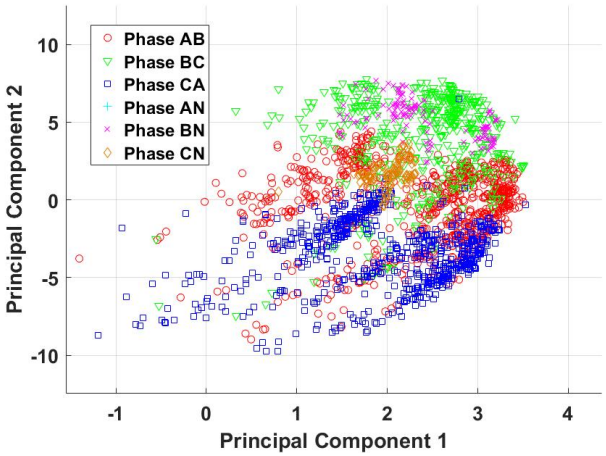


Figure 2.11: Field validated phase connections of data set s_{18} in the 2-dimensional PCA map.

Table 2.4: Impact of Sampling Frequency on the Phase Identification Accuracy

Feeder	Data Set	Granularity of Meter Readings		
		1 hour	15-minute	5-minute
1	s_1	93.06%	93.93%	93.88%
	s_2	93.62%	94.32%	94.40%
2	s_3	87.55%	88.86%	92.03%
	s_4	87.79%	90.47%	89.93%
3	s_5	83.94%	90.02%	91.56%
	s_6	82.83%	84.51%	87.16%

2.4.4.4 Impact of the Smart Meter Sampling Frequency on the Phase Identification Accuracy

The phase identification accuracies of the proposed algorithm under 3 different meter reading granularity levels are calculated and summarized in Table 2.4. It shows that as the granularity of meter readings increases from hourly to every 15 minutes and then 5 minutes, the phase identification accuracy increases. The average increase in the phase identification accuracy over the 3 distribution circuits is 3.36% when the meter reading granularity increases from hourly to 5 minutes. More granular voltage readings allow extractions of features/patterns that may not be present in coarse data sets. However, it should be noted that there are additional costs associated with gathering more granular smart meter data. Note that the phase identification accuracy decreases slightly for data set s_1 and s_4 when the sampling frequency increases from 15-minute to 5-minute. This is partly due to the randomness of the t-SNE mapping.

2.4.5 Conclusion

This work develops an AMI data driven phase identification algorithm that addresses the drawbacks of the existing solutions. Compared to the existing solutions, the proposed algorithm has three main advantages. First, the proposed algorithm does not require prior knowledge about the

number of phase connections in the distribution system. Second, the proposed algorithm works well with distribution feeders that have both phase-to-neutral and phase-to-phase connections. Third, the accuracy of the proposed phase identification algorithm is not very sensitive to the level of unbalance in a distribution feeder. Comprehensive field testing results on 5 distribution feeders show that the proposed algorithm significantly outperforms the existing methods. In addition, we discover that more granular voltage time series leads to higher phase identification accuracy.

In the proposed CHC algorithm, a few parameters need to be tuned manually. To implement the proposed AMI data driven phase identification algorithm on thousands of distribution feeders, future work can be done to develop an automatic parameter tuning algorithm.

Chapter 3

Phase Identification by the Physics-Informed Model Approach

3.1 Introduction

With declining costs, distributed energy resources (DERs) such as energy storage systems, distributed generation, and electric vehicles are rapidly penetrating power distribution systems world-wide. To coordinate the operations of a large number of heterogeneous DERs, advanced distribution system control applications such as Volt-VAR control, network reconfiguration, and three-phase optimal power flow are necessary. The successful implementation of these applications requires accurate information about the phase connectivity of power distribution systems. However, the phase connectivity information in electric utilities is usually missing or highly unreliable.

In this chapter, by linearizing the three-phase power flow manifold, we develop a physics-informed model, which links the phase connections to the smart meter measurements. The phase

identification problem is first formulated as a maximum likelihood estimation problem and then reformulated as a maximum marginal likelihood estimation problem. We prove that the correct phase connection achieves the highest log likelihood values for both problems. An efficient solution method is proposed by decomposing the original problem into subproblems with a binary least-squares formulation. The numerical tests on a comprehensive set of distribution circuits show that our proposed method yields very high accuracy on both radial and meshed distribution circuits with a combination of single-phase, two-phase, and three-phase loads. The proposed algorithm is robust with respect to inaccurate feeder models, incomplete measurements, and bad measurements. It also outperforms the existing methods on complex circuits.

The rest of this chapter is organized as follows. Section 3.2 reviews the literature and summarizes our work's contribution. Section 3.3 formulates the phase identification problem. Section 3.4 presents the linearized three-phase power flow model. Section 3.5 derives the physics-informed model that links the phase connections to the smart meter measurements. Section 3.6 formulates the phase identification problem as an MLE and MMLE problem and presents an efficient solution algorithm. Section 3.7 performs a comprehensive numerical test to evaluate the performance of the proposed MMLE-based phase identification method. Section 3.8 states the conclusion.

3.2 Literature Review and Contributions of Our Work

Traditionally, electric utilities send field crews to measure phase angles and determine phase connections with special equipment such as phase meters [21]. Although such practices provide very accurate phase connections information, they are very labor-intensive, time-consuming, and expensive. The time synchronized measurements from micro-phasor measurement units (μPMUs)

can also provide highly accurate estimations of phase connections [44, 67]. However, a system-wide installation is cost prohibitive. State estimation can also be used to verify phase connection information [68]. However, this method only applies to circuits with mostly accurate phase connections and the area of incorrect phase connections needs to be known. In order to develop more cost effective phase identification algorithms, researchers have turned to data-driven methods, which use measurements from the advanced metering infrastructure (AMI). The existing data-driven approaches can be categorized into three approaches: energy supply and consumption matching, correlation-based analysis, and clustering-based analysis.

The energy supply and consumption matching approach is based on the principle of conservation of energy. With complete coverage of load measurements, the aggregate power consumption of downstream loads in each phase plus losses is equal to the corresponding phase's power flow measured at the upstream point. In this approach, Ref. [41] formulates the problem as integer programming and solves it using tabu search. Ref. [69] uses relaxed integer programming and improves the phase identification accuracy by actively managing the power injections of DERs. In [70], principal component analysis (PCA) and its graph-theoretic interpretation are used to infer phase connections. However, algorithms in this approach cannot identify phase connections in the presence of delta-connected two-phase loads.

In the correlation-based analysis approach, correlation analysis is performed using smart meters' and the substation's measurements or the three-phase primary line's measurements. Each smart meter is assigned to a phase, which has the highest correlation coefficient with it. In this approach, Ref. [39, 57] use voltage magnitude profiles for the correlation analysis. In [71], salient features are extracted from load profiles for the correlation analysis. Although the correlation-based

analysis has achieved good performance on radial circuits with only single-phase loads, it does not work well for a meshed circuit, which has all seven possible phase connections of single-phase, two-phase, and three-phase loads.

In the clustering-based approach, smart meters are grouped based on the mutual similarity of their voltage magnitude profiles. It is assumed that each resulting cluster represents a single phase connection. Ref. [30, 31] project the voltage magnitude profiles onto low-dimension spaces and leverage constrained clustering algorithms to identify both single-phase and two-phase connections. Ref. [72] designs an algorithm by combining clustering and the minimum spanning tree method to identify phase connections. However, it has been shown that the performance of the clustering-based approach deteriorates as the feeder becomes more balanced [31].

To further improve the phase identification accuracy and provide a theoretical foundation for the problem, we develop a physically inspired method for phase identification. By linearizing the three-phase power flow manifold, we first develop a physical model, which links phase connections to the smart meter measurements. We then formulate the phase identification task as a maximum likelihood estimation (MLE) problem and prove that the correct phase connection yields the highest log likelihood value. The nonlinearity and nonconvexity nature of the MLE problem makes it difficult to solve. Thus, we reformulate the MLE problem as a maximum marginal likelihood estimation (MMLE) problem and prove that the correct phase connection also yields the highest marginal log likelihood value. Finally, an efficient solution algorithm is developed for the MMLE problem by dividing it into sub-problems, which can be solved by least squares integer programming.

Compared to the existing data-driven phase identification algorithms, our approach has the following advantages: first, the physically interpretable MMLE formulation brings a solid theoreti-

cal foundation to the phase identification problem; second, our proposed algorithm not only works for radial distribution feeders, but also heavily meshed networks; third, our proposed algorithm achieves higher accuracy for complex circuits with both single-phase and two-phase connections and a lower level of unbalance, which create a lot of problems to existing data-driven methods; fourth, our proposed algorithm is robust with respect to inaccurate feeder models, incomplete measurements, and bad measurements.

3.3 Problem Formulation

3.3.1 Problem Setup

We intend to identify the type of phase connection for all loads on a distribution feeder. The distribution feeder's three-phase primary line contains $N + 1$ nodes, indexed as node 0 to N, in which node 0 is the source/substation. A load can connect to a three-phase node directly, or indirectly through a single-phase or two-phase branch (e.g., the dashed lines and dash-dot lines in Fig. 3.1). Note that nodes and loads are two different concepts. In the technical derivation, all variables are in per unit or radian angles unless otherwise specified.

3.3.2 Assumptions

Note that the assumptions described below are only used to prove that the correct phase connection yields the highest log likelihood value of the MLE/MMLE problem formulated in this chapter. Some of these assumptions may not hold in the real world. However, the numerical study will show that our proposed algorithm works well even when some of these assumptions no longer hold. In these cases, we can no longer guarantee that the algorithm will result in 100% accuracy.

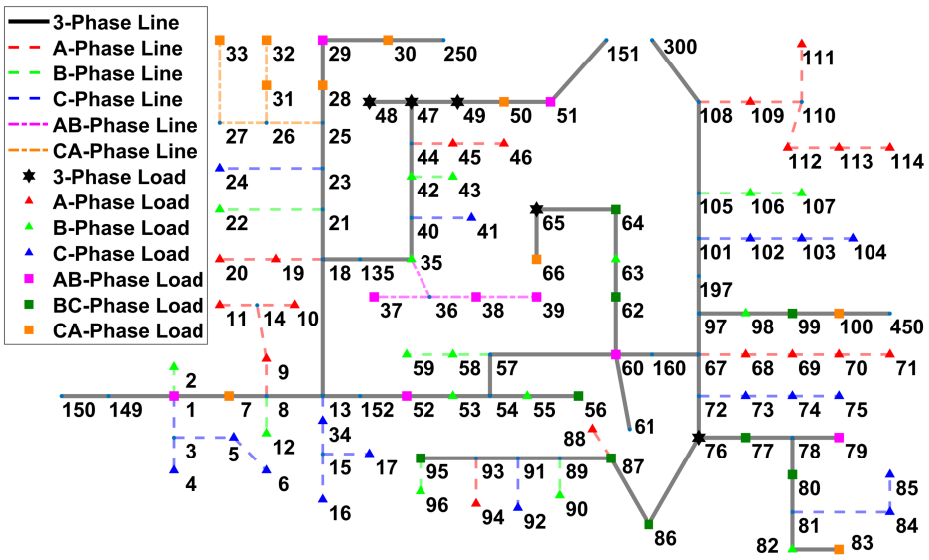


Figure 3.1: Schematic of a modified IEEE 123-node test feeder.

3.3.2.1 Data and Model Availability

First, the information about whether the load is single-phase, two-phase, or three-phase is assumed to be available. Usually, this information can be deduced by examining the distribution transformer configuration and customer billing information. Second, for a single-phase load on phase i, we know its power injection (both real and reactive power) and voltage magnitude of phase i. Third, for a two-phase delta-connected load between phase i and j, we know its power injection and voltage magnitude across phase i and j. Fourth, for a three-phase load, we know its total power injection and the voltage magnitude of one of the phases, which needs to be identified. Fifth, for the source node, we know the voltage measurement. Sixth, the connectivity model and the parameters of the primary feeder are known. Finally, we assume that the distribution feeder is not severely unbalanced. The task of phase identification is to determine which phase(s) each single-phase or two-phase load connects to and which phase's voltage magnitude the three-phase smart meter mea-

sures. Note that our proposed algorithm does not assume a 100% smart meter penetration rate. The numerical study will show that our algorithm is robust with respect to incomplete measurements.

3.3.2.2 Statistical Assumptions

First, it is assumed that the incremental changes in measured real, reactive power, and voltage magnitudes across one time interval are independent over time. Second, it is assumed that the noise terms which represent the model errors and the measurement errors are i.i.d. Gaussian. Note that the noise terms will be derived later in Section 3.6. Third, it is assumed that theses noise terms are independent of the incremental changes in smart meter measurements. Note that these statistical assumptions will be verified in the numerical study section.

3.4 Linearized Three-Phase Power Flow Model

The very first step of our phase identification framework is to build a three-phase power flow model for the primary feeder. To do so, we need a procedure that we call *reduction*, and the resulting network is called a *reduced network*. The reduction is simply converting any loaded single-phase or two-phase branch into an equivalent load so that the reduced network contains only three-phase lines. The details of the reduction procedure is explained in Appendix A.1. In the rest of the chapter, we use M to denote the number of loads in the reduced network and *load* refers to the equivalent load in the reduced network.

Based on the reduced primary feeder, by following [73], we can derive the linearized three-phase power flow model shown in (3.1), in which the variables organized by phase. In the linearized model, shunt admittance is ignored because its value is very small. Later on in the numerical

study, it will be verified that ignoring shunt admittance does not affect the phase identification accuracy.

$$A \begin{bmatrix} \boldsymbol{v} - \overline{\boldsymbol{v}} \\ \boldsymbol{\theta} - \overline{\boldsymbol{\theta}} \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{v} - \overline{\boldsymbol{v}} \\ \boldsymbol{\theta} - \overline{\boldsymbol{\theta}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{p} \\ \boldsymbol{q} \end{bmatrix} \tag{3.1}$$

Here A_{11}, A_{12}, A_{21}, and A_{22} are $3(N+1) \times 3(N+1)$ matrices. $\boldsymbol{v}$, $\boldsymbol{\theta}$, $\boldsymbol{p}$, and $\boldsymbol{q}$ are the nodes' voltage magnitude, voltage angle, and real and reactive power of three phases. $\overline{\boldsymbol{v}} = \mathbf{1}_{3(N+1)}$ and $\overline{\boldsymbol{\theta}} = [0 \times \mathbf{1}_{N+1}^T, -\frac{2\pi}{3} \times \mathbf{1}_{N+1}^T, \frac{2\pi}{3} \times \mathbf{1}_{N+1}^T]^T$ are the flat feasible solution for the underlying nonlinear power flow model. Let $\alpha = e^{-j\frac{2\pi}{3}}$, define $\Phi \triangleq \text{diag}(I_{(N+1)},\ \alpha I_{(N+1)},\ \alpha^2 I_{(N+1)})$ and define

$$Y \triangleq \begin{bmatrix} Y^{aa} & Y^{ab} & Y^{ac} \\ Y^{ba} & Y^{bb} & Y^{bc} \\ Y^{ca} & Y^{cb} & Y^{cc} \end{bmatrix} \tag{3.2}$$

where Y^{ij} is the $(N+1) \times (N+1)$ nodal admittance matrix between phase i and j. Then A_{11}, A_{12}, A_{21}, and A_{22} can be calculated as $A_{11} = -A_{22} = Re(\Phi^{-1} Y \Phi)$ and $A_{12} = A_{21} = -Im(\Phi^{-1} Y \Phi)$.

It has been shown in [74] that for a connected three-phase network, $\text{rank}(Y) = 3N$. Thus, $\text{rank}(A)$ is at most $6N$. For subsequent derivations, we need to transform A into a nonsingular form. Following Appendix A.2, the transformed power flow model becomes

$$\check{A} \begin{bmatrix} \check{\boldsymbol{v}} \\ \check{\boldsymbol{\theta}} \end{bmatrix} = \begin{bmatrix} \check{A}_{11} & \check{A}_{12} \\ \check{A}_{21} & \check{A}_{22} \end{bmatrix} \begin{bmatrix} \check{\boldsymbol{v}} \\ \check{\boldsymbol{\theta}} \end{bmatrix} = \begin{bmatrix} \check{\boldsymbol{p}} \\ \check{\boldsymbol{q}} \end{bmatrix} \tag{3.3}$$

where $\check{A}_{mn}$ is a $3N \times 3N$ matrix obtained by removing the rows and columns corresponding to the substation node in A_{mn}. We denote the difference of voltage magnitudes and voltage angles between the non-substation nodes and the substation nodes as $\check{\boldsymbol{v}}$, $\check{\boldsymbol{\theta}}$. We denote the non-substation nodes' real and reactive power as $\check{\boldsymbol{p}}$ and $\check{\boldsymbol{q}}$.

In theory, $\check{A}$ is not guaranteed to be invertible. However, for the majority of real-world distribution feeders, $\text{rank}(\check{A}) = 6N$. It will be shown in the numerical study section that for all IEEE distribution test feeders, $\check{A}$ has a full rank.

Solving for $\check{\boldsymbol{v}}$ with $\check{\boldsymbol{p}}$ and $\check{\boldsymbol{q}}$ from (3.3), we have

$$\check{\boldsymbol{v}} = (\check{A}_{11} - \check{A}_{12}\check{A}_{22}^{-1}\check{A}_{21})^{-1}\check{\boldsymbol{p}} - (\check{A}_{11} - \check{A}_{12}\check{A}_{22}^{-1}\check{A}_{21})^{-1}\check{A}_{12}\check{A}_{22}^{-1}\check{\boldsymbol{q}} \tag{3.4}$$

or in condensed form as

$$\check{\boldsymbol{v}} = K\check{\boldsymbol{p}} - L\check{\boldsymbol{q}} \tag{3.5}$$

It can be shown that $(\check{A}_{11} - \check{A}_{12}\check{A}_{22}^{-1}\check{A}_{21})$ is invertible if $\check{A}$ is invertible. Similarly, we can link $\check{\boldsymbol{\theta}}$ with $\check{\boldsymbol{p}}$ and $\check{\boldsymbol{q}}$ as

$$\check{\boldsymbol{\theta}} = (\check{A}_{12} - \check{A}_{11}\check{A}_{21}^{-1}\check{A}_{22})^{-1}\check{\boldsymbol{p}} - (\check{A}_{12} - \check{A}_{11}\check{A}_{21}^{-1}\check{A}_{22})^{-1}\check{A}_{11}\check{A}_{21}^{-1}\check{\boldsymbol{q}} \tag{3.6}$$

or in condensed form as

$$\check{\boldsymbol{\theta}} = \mathcal{K}\check{\boldsymbol{p}} - \mathcal{L}\check{\boldsymbol{q}} \tag{3.7}$$

3.5 Physics-Informed Model for Phase Identification

In this section, we develop a mathematical model that relates the phase connections of loads to voltage magnitude and power injection measurements. Section 3.5.1 explains how to express smart meter measurements in terms of nodal voltages and power injections of the three-phase power flow model. Section 3.5.2 derives the phase connection model, which relates phase connections to network measurements.

3.5.1 Link Smart Meter Measurements with the Nodal Voltages and Power Injections

The linearized three-phase power flow models (3.5) and (3.7) are derived in terms of nodal voltages and power injections $\check{\boldsymbol{v}}$, $\check{\boldsymbol{\theta}}$, $\check{\boldsymbol{p}}$, and $\check{\boldsymbol{q}}$, which are often not directly measured by

smart meters. Thus, we need to embed the smart meter measurements into these two equations. This is straightforward for single-phase and three-phase loads. For a single-phase load m on node n, its voltage measurement $\hat{v}_m$ is equal to one of the three phase-to-neutral voltage magnitudes v_n^i $(i = a, b, c)$, which is related to $\check{v}_n^i$ in (3.3) via $\check{v}_n^i \triangleq v_n^i - v_0^i$, where v_0^i is the source voltage magnitude in phase i. Similarly, a single-phase load's power injection measurement $\hat{p}_m + j\hat{q}_m$ corresponds to the power injection of one of the three phases $\check{p}_n^i + j\check{q}_n^i$ at node n. For a three-phase load m at node n, the single-phase voltage measurement $\hat{v}_m$ is equal to one of the three nodal voltage magnitudes v_n^i $(i = a, b, c)$. We can assume that the three-phase power injections $\hat{p}_m + j\hat{q}_m$ is distributed relatively evenly to three phases at node n. For a delta-connected two-phase load, we need the following derivations to link its measurements to the three-phase power flow model.

3.5.1.1 Link Power Injection Measurements with Power Flow Model

Without loss of generality, we use a phase AB load as an example. Suppose the two-phase power injection measurement is $S_{ab} = P_{ab} + jQ_{ab} = S_a + S_b = (P_a + jQ_a) + (P_b + jQ_b)$. Here, S_a and S_b are the power injections at the phase A and phase B ports. We can estimate S_a and S_b based on S_{ab} as follows: (see the proof in Appendix A.3)

$$S_a \approx \left(\frac{1}{2}P_{ab} + \frac{\sqrt{3}}{6}Q_{ab}\right) + j\left(\frac{1}{2}Q_{ab} - \frac{\sqrt{3}}{6}P_{ab}\right) \tag{3.8}$$

$$S_b \approx \left(\frac{1}{2}P_{ab} - \frac{\sqrt{3}}{6}Q_{ab}\right) + j\left(\frac{1}{2}Q_{ab} + \frac{\sqrt{3}}{6}P_{ab}\right) \tag{3.9}$$

3.5.1.2 Link Voltage Magnitude Measurements with Power Flow Model

Here we need to establish a relationship between the phase-to-phase voltage magnitude measurements and the nodal phase-to-neural voltage magnitudes in (3.5) and (3.7). For a load m across phase ij ($ij \in \{ab, bc, ca\}$) at node n, the relationship can be written as: (see the proof in Appendix A.4)

$$\hat{v}_m - v_0^{ij} \approx \frac{\sqrt{3}}{2}(v_n^i - v_0^i) + \frac{\sqrt{3}}{2}(v_n^j - v_0^j) + \frac{1}{2}(\theta_n^i - \theta_0^i) - \frac{1}{2}(\theta_n^j - \theta_0^j) \tag{3.10}$$

where $\hat{v}_m$ is load m's voltage magnitude measurement. v_0^{ij} is the voltage magnitude across phase ij at the substation. v_n^i and v_0^i are the voltage magnitudes of phase i at node n and the substation. θ_n^i and θ_0^i are the voltage angles of phase i at node n and the substation. Note that in above derivations, voltages are in per unit and angles are in radian.

3.5.2 Modeling Phase Connections in Three-phase Power Flow

3.5.2.1 Decision Variables for Phase Connections

We use three decision variables, x_m^1, x_m^2, and x_m^3 to denote the phase connection for each load m. $x_m^i = 0$ or 1, and $\sum_i x_m^i = 1,\ \forall\ m$. If load m is single-phase, then x_m^1, x_m^2, and x_m^3 represent AN, BN, and CN connections. If m is two-phase, then x_m^1, x_m^2, and x_m^3 represent AB, BC, and CA connections. If m is three-phase, and the measured voltage is between one phase and the neutral, then x_m^1, x_m^2, and x_m^3 represent which of the phases AN, BN, and CN is measured. As stated in the assumptions, we know whether a load is single-phase, two-phase, or three-phase from the distribution transformer configuration and customer billing information. The phase connection decision variables form an $M \times 3M$ matrix X defined as $X \triangleq \text{diag}([x_1^1\ x_1^2\ x_1^3], ..., [x_M^1\ x_M^2\ x_M^3])$.

3.5.2.2 Additional Definitions

Several matrices and variables are defined here to build the model for phase connections.

Define matrices W_1 and W_2 as

$$W_1 \triangleq \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}, \; W_2 \triangleq \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ -1 & 0 & 1 \end{bmatrix} \tag{3.11}$$

Let I_n denote an identity matrix of size n, $\mathbb{0}_{k\times l}$ denote a $k \times l$ all-0 matrix, and $\mathbb{1}_{k\times l}$ denote a $k \times l$ all-1 matrix. Define U^1 and U^2 as $3M \times 3N$ matrices of 3×3 blocks. Define $\hat{U}^1$ and $\hat{U}^2$ as $3N \times 3M$ matrices of 3×3 blocks. Define U^1_{mn} and U^2_{mn} as the mn-th block of U^1 and U^2. Define $\hat{U}^1_{nm}$ and $\hat{U}^2_{nm}$ as the nm-th block of $\hat{U}^1$ and $\hat{U}^2$. If load m is not connected to node n, then U^1_{mn}, U^2_{mn}, $\hat{U}^1_{nm}$, and $\hat{U}^2_{nm}$ are equal to $\mathbb{0}_{3\times 3}$. If load m is connected to node n, then U^1_{mn}, U^2_{mn}, $\hat{U}^1_{nm}$, and $\hat{U}^2_{nm}$ are defined based on load m's phase connection type, as shown in Table 3.1.

Table 3.1: Values of 3×3 Blocks by Phase Connection Type if Load m is Connected to Node n

Load m's Phase Connection Type	U^1_{mn}	U^2_{mn}	$\hat{U}^1_{nm}$	$\hat{U}^2_{nm}$
single-phase	I_3	$\mathbb{0}_{3\times 3}$	I_3	$\mathbb{0}_{3\times 3}$
two-phase	$\frac{\sqrt{3}}{2}W_1$	$\frac{1}{2}W_2$	$\frac{1}{2}W_1^T$	$\frac{\sqrt{3}}{6}W_2^T$
three-phase	I_3	$\mathbb{0}_{3\times 3}$	$\frac{1}{3}\mathbb{1}_{3\times 3}$	$\mathbb{0}_{3\times 3}$

Define $\hat{\boldsymbol{v}}^{\text{ref}} \triangleq [\hat{\boldsymbol{v}}_1^{\text{ref}}, \ldots, \hat{\boldsymbol{v}}_M^{\text{ref}}]^T$, where $\hat{\boldsymbol{v}}_m^{\text{ref}} = [v_0^a, v_0^b, v_0^c]$ if load m is single-phase or three-phase; $\hat{\boldsymbol{v}}_m^{\text{ref}} = [v_0^{ab}, v_0^{bc}, v_0^{ca}]$ if load m is two-phase. Here, v_0^i denotes the substation's voltage magnitude of phase i, and v_0^{ij} denotes the substation's voltage magnitude across phase ij.

3.5.2.3 Phase Connection Model

Now we can build the model, which links phase connections with the smart meter measurements. Let $\hat{\boldsymbol{v}}$, $\hat{\boldsymbol{p}}$, and $\hat{\boldsymbol{q}}$ be $M \times 1$ vectors of measured voltage magnitudes, real power, and

reactive power of the M loads. From (3.8) - (3.10), Section 3.5.2.1, and 3.5.2.2, we have:

$$\check{\boldsymbol{p}} \approx \hat{U}^1 X^T \hat{\boldsymbol{p}} + \hat{U}^2 X^T \hat{\boldsymbol{q}} \tag{3.12}$$

$$\check{\boldsymbol{q}} \approx -\hat{U}^2 X^T \hat{\boldsymbol{p}} + \hat{U}^1 X^T \hat{\boldsymbol{q}} \tag{3.13}$$

$$\hat{\boldsymbol{v}} \approx X \hat{\boldsymbol{v}}^{\text{ref}} + X U^1 \check{\boldsymbol{v}} + X U^2 \check{\boldsymbol{\theta}} \tag{3.14}$$

With a slight abuse of notations, the entries of $\check{\boldsymbol{p}}$, $\check{\boldsymbol{q}}$, $\check{\boldsymbol{v}}$, and $\check{\boldsymbol{\theta}}$ are organized by node in (3.12)-(3.14) (instead of by phase as in (3.5) and (3.7)). Equations (3.12) and (3.13) map the measured power injection of each load to the corresponding nodal power injections in the linearized power flow model. Take load m connected to node n as an example and suppose $x_m^1 = 1$. If load m is single-phase, then its power injection is mapped to phase A at node n. If load m is two-phase, then its power injection is distributed to phase A and B at node n according to (3.8) and (3.9). If load m is three-phase, then its power injection is evenly distributed to all three phases of node n.

Equation (3.14) links the voltage measurement $\hat{\boldsymbol{v}}$ with $\check{\boldsymbol{v}}$ and $\check{\boldsymbol{\theta}}$, i.e., the nodal line-to-neutral voltage magnitude and angle difference with the substation in the linearized power flow model. Take load m connected to node n as an example and suppose $x_m^1 = 1$. If load m is single-phase or three-phase, then (3.14) can be reduced to $\hat{v}_m = v_0^a + (v_n^a - v_0^a)$, where v_n^a is node n's voltage magnitude in phase A. If load m is two-phase, then (3.14) is equivalent to (3.10).

Substituting (3.5), (3.7), (3.12) and (3.13) into (3.14) yields

$$\hat{\boldsymbol{v}} \approx X \hat{\boldsymbol{v}}^{\text{ref}} + X \hat{K} X^T \hat{\boldsymbol{p}} + X \hat{L} X^T \hat{\boldsymbol{q}} \tag{3.15}$$

where $\hat{K} \triangleq [(U^1K+U^2\mathcal{K})\hat{U}^1+(U^1L+U^2\mathcal{L})\hat{U}^2]$ and $\hat{L} \triangleq [(U^1K+U^2\mathcal{K})\hat{U}^2-(U^1L+U^2\mathcal{L})\hat{U}^1]$. Here, with a slight abuse of notations, K, L, $\mathcal{K}$, and $\mathcal{L}$'s entries are organized by node (instead of by phase as in (3.5) and (3.7)). Thus, (3.15) provides the physical model, which relates power injection measurements and phase connections to voltage magnitude measurements.

To remove trends and seasonality in time series data, we define the difference of the voltage measurement and its lagged variable as $\tilde{\boldsymbol{v}}(t)$, with $\tilde{\boldsymbol{v}}(t) \triangleq \hat{\boldsymbol{v}}(t)-\hat{\boldsymbol{v}}(t-1)$. $\tilde{\boldsymbol{v}}^{\text{ref}}(t)$, $\tilde{\boldsymbol{p}}(t)$, and $\tilde{\boldsymbol{q}}(t)$ are defined in a similar way. Thus, we have the time difference version of the physical model:

$$\tilde{\boldsymbol{v}}(t) = X\tilde{\boldsymbol{v}}^{\text{ref}}(t) + X\hat{K}X^T\tilde{\boldsymbol{p}}(t) + X\hat{L}X^T\tilde{\boldsymbol{q}}(t) + \boldsymbol{n}(t) \tag{3.16}$$

where $\boldsymbol{n}(t)$ is the "noise term" representing the error of the linearized power flow model, the measurement error, and all the other sources of noise not considered. In (3.16), $\tilde{\boldsymbol{v}}(t)$, $\tilde{\boldsymbol{p}}(t)$, $\tilde{\boldsymbol{q}}(t)$, and $\tilde{\boldsymbol{v}}^{\text{ref}}(t)$ can be calculated from the smart meter and substation measurements. $\hat{K}$ and $\hat{L}$ can be derived from the feeder model. Thus, the task of phase identification is to estimate the phase decision variables in X.

3.6 Maximum Marginal Likelihood Estimation of Phase Connections

In this section, we first formulate phase identification as an MLE problem and then as an MMLE problem. Next, we prove that the correct phase connection is a global optimizer of the MMLE problem. Lastly, we develop a computationally efficient algorithm to solve the MMLE problem.

3.6.1 MLE Problem Formulation

Let $\boldsymbol{x} \triangleq [x_1^1, x_1^2, x_1^3, ..., x_M^1, x_M^2, x_M^3]^T$ be the phase connection decision variable vector. Define $\tilde{\boldsymbol{v}}(t, \boldsymbol{x})$ as the theoretical differenced voltage measurement $\tilde{\boldsymbol{v}}(t)$ with phase connection $\boldsymbol{x}$:

$$\tilde{\boldsymbol{v}}(t, \boldsymbol{x}) \triangleq X\tilde{\boldsymbol{v}}^{\text{ref}}(t) + X\hat{K}X^T\tilde{\boldsymbol{p}}(t) + X\hat{L}X^T\tilde{\boldsymbol{q}}(t) \tag{3.17}$$

Then $\tilde{\boldsymbol{v}}(t) = \tilde{\boldsymbol{v}}(t, \boldsymbol{x}) + \boldsymbol{n}(t)$, where $\boldsymbol{x}$ is the phase connection decision variable vector that we need to estimate. As stated in Section 3.3.2, we assume that the noise $\boldsymbol{n}(t)$ is independent of $\tilde{\boldsymbol{v}}^{\text{ref}}(t)$, $\tilde{\boldsymbol{p}}(t)$, and $\tilde{\boldsymbol{q}}(t)$ and is i.i.d. Gaussian $\boldsymbol{n}(t) \sim \mathcal{N}(\mathbb{0}_{M\times 1}, \Sigma_n)$, where Σ_n is an unknown underlying covariance matrix. Given these conditions, $\boldsymbol{n}(t)$ is also independent of $\tilde{\boldsymbol{v}}(t, \boldsymbol{x})$. Thus, the likelihood of observing $\{\tilde{\boldsymbol{v}}(t)\}_{t=1}^T$ given $\{\tilde{\boldsymbol{v}}^{\text{ref}}(t)\}_{t=1}^T$, $\{\tilde{\boldsymbol{p}}(t)\}_{t=1}^T$, and $\{\tilde{\boldsymbol{q}}(t)\}_{t=1}^T$ is a function of $\boldsymbol{x}$:

$$\begin{aligned} &Prob(\{\tilde{\boldsymbol{v}}(t)\}_{t=1}^T | \{\tilde{\boldsymbol{v}}^{\text{ref}}(t)\}_{t=1}^T, \{\tilde{\boldsymbol{p}}(t)\}_{t=1}^T, \{\tilde{\boldsymbol{q}}(t)\}_{t=1}^T; \boldsymbol{x}) = \\ &\frac{|\Sigma_n|^{-\frac{T}{2}}}{(2\pi)^{\frac{MT}{2}}} \times \exp\Big\{-\frac{1}{2}\sum_{t=1}^{T}[\tilde{\boldsymbol{v}}(t) - \tilde{\boldsymbol{v}}(t, \boldsymbol{x})]^T\Sigma_n^{-1}[\tilde{\boldsymbol{v}}(t) - \tilde{\boldsymbol{v}}(t, \boldsymbol{x})]\Big\} \end{aligned} \tag{3.18}$$

Taking the negative logarithm of (3.18), removing the constant term, and scaling by $\frac{2}{T}$, we get

$$f(\boldsymbol{x}) \triangleq \frac{1}{T}\sum_{t=1}^{T}[\tilde{\boldsymbol{v}}(t) - \tilde{\boldsymbol{v}}(t, \boldsymbol{x})]^T\Sigma_n^{-1}[\tilde{\boldsymbol{v}}(t) - \tilde{\boldsymbol{v}}(t, \boldsymbol{x})] \tag{3.19}$$

It will be shown in Lemma 1 that the correct phase connection $\boldsymbol{x}^*$ maximizes the likelihood function (3.18) and minimizes $f(x)$ under two mild assumptions.

Lemma 1. *Let $\boldsymbol{x}^*$ be the correct phase connection. If the following two conditions are satisfied, then as $T \to \infty$, $\boldsymbol{x}^*$ is a global optimizer to minimize $f(\boldsymbol{x})$.*

1. *$\boldsymbol{n}(t_k)$ is i.i.d. and independent of $\tilde{\boldsymbol{v}}^{ref}(t_l)$, $\tilde{\boldsymbol{p}}(t_l)$, and $\tilde{\boldsymbol{q}}(t_l)$, for $\forall t_k, t_l \in Z^+$.*

2. *$\tilde{\boldsymbol{v}}^{ref}(t_k)$, $\tilde{\boldsymbol{p}}(t_k)$, and $\tilde{\boldsymbol{q}}(t_k)$ are independent of $\tilde{\boldsymbol{v}}^{ref}(t_l)$, $\tilde{\boldsymbol{p}}(t_l)$, and $\tilde{\boldsymbol{q}}(t_l)$, for $\forall t_k, t_l \in Z^+,\ t_k \neq t_l$*

The proof of Lemma 1 can be found in Appendix A.5. By substituting (3.17) into (3.19), we can see that directly minimizing $f(\boldsymbol{x})$ is very difficult due to its nonlinearity and nonconvexity. Furthermore, the actual value of Σ_n is unknown. To address this technical challenge, in Section 3.6.2, we will convert the phase identification problem into an MMLE problem and prove that the correct phase connection is also a global optimizer of the MMLE problem.

3.6.2 MMLE Problem Formulation

Let $\tilde{v}_m(t)$ be the mth entry of $\tilde{\boldsymbol{v}}(t)$, $\tilde{v}_m(t,\boldsymbol{x})$ be the mth entry of $\tilde{\boldsymbol{v}}(t,\boldsymbol{x})$, and $n_m(t)$ be the mth entry of $\boldsymbol{n}(t)$. The marginal likelihood of observing $\{\tilde{v}_m(t)\}_{t=1}^T$ given $\{\tilde{\boldsymbol{v}}^{\text{ref}}(t)\}_{t=1}^T$, $\{\tilde{\boldsymbol{p}}(t)\}_{t=1}^T$, and $\{\tilde{\boldsymbol{q}}(t)\}_{t=1}^T$ is a function of $\boldsymbol{x}$:

$$\begin{aligned} &Prob(\{\tilde{v}_m(t)\}_{t=1}^T|\{\tilde{\boldsymbol{v}}^{\text{ref}}(t)\}_{t=1}^T,\{\tilde{\boldsymbol{p}}(t)\}_{t=1}^T,\{\tilde{\boldsymbol{q}}(t)\}_{t=1}^T;\boldsymbol{x}) \\ &= \frac{\Sigma_n(m,m)^{-\frac{T}{2}}}{(2\pi)^{\frac{T}{2}}}\exp\Big\{-\frac{1}{2}\sum_{t=1}^{T}\frac{[\tilde{v}_m(t)-\tilde{v}_m(t,\boldsymbol{x})]^2}{\Sigma_n(m,m)}\Big\} \end{aligned} \tag{3.20}$$

where $\Sigma_n(m,m)$ is the mth diagonal entry of Σ_n. Taking the negative logarithm of (3.20), removing the constant term, and scaling by $\frac{2\Sigma_n(m,m)}{T}$, we have

$$f_m(\boldsymbol{x}) \triangleq \frac{1}{T}\sum_{t=1}^{T}[\tilde{v}_m(t)-\tilde{v}_m(t,\boldsymbol{x})]^2 \tag{3.21}$$

Lemma 2. *Let $\boldsymbol{x}^*$ be the correct phase connection. If the two conditions in Lemma 1 hold, then $\boldsymbol{x}^*$ is a global optimizer to minimize $f_m(\boldsymbol{x})$ as $T\to\infty$. In addition, any $\boldsymbol{x}$ is a global optimizer of $f_m(\boldsymbol{x})$ if it satisfies all the following conditions:*

1. $x_m^i = x^{*i}_m, \forall i$;
2. $x_k^i = x^{*i}_k, \forall i$, $k\neq m$ *and load* k *is not three-phase.*

The proof of Lemma 2 can be found in Appendix A.6.

3.6.3 Solution Method for the MMLE Problem

Directly minimizing $f_m(\boldsymbol{x})$ from (3.21) is still a difficult task. Thus, we further simplify the optimization problem by first solving three subproblems $min f_{m,i}(\boldsymbol{x}_{-m}), i \in \{1,2,3\}$. $f_{m,i}(\boldsymbol{x}_{-m})$ are defined as

$$\begin{gathered} f_{m,i}(\boldsymbol{x}_{-m}) \triangleq f_m(\boldsymbol{x}) \\ \text{subject to } x_m^i = 1 \text{ and } x_m^j = 0 \text{ for } j \neq i \end{gathered} \tag{3.22}$$

where $\boldsymbol{x}_{-m}$ is a $(3M-3)\times 1$ vector containing every element in $\boldsymbol{x}$ except x_m^1, x_m^2, and x_m^3. Since $x_m^i = 0$ or 1, and $\sum_i x_m^i = 1$, then from (3.22) we have:

$$\min_{\boldsymbol{x}} f_m(\boldsymbol{x}) = \min_{i=1,2,3} \min_{\boldsymbol{x}_{-m}} f_{m,i}(\boldsymbol{x}_{-m}) \tag{3.23}$$

To solve the sub-problems, we first define $\tilde{v}_{m,i}(t, \boldsymbol{x}_{-m})$ as

$$\begin{gathered} \tilde{v}_{m,i}(t, \boldsymbol{x}_{-m}) \triangleq \tilde{v}_m(t, \boldsymbol{x}) \\ \text{subject to } x_m^i = 1 \text{ and } x_m^j = 0 \text{ for } j \neq i \end{gathered} \tag{3.24}$$

Substituting (3.17) into (3.24), we have

$$\begin{gathered} \tilde{v}_{m,i}(t, \boldsymbol{x}_{-m}) = \tilde{v}_{m,i}^{\text{ref}}(t) + \hat{K}_{m,i} X^T \tilde{\boldsymbol{p}}(t) + \hat{L}_{m,i} X^T \tilde{\boldsymbol{q}}(t) \\ \text{subject to } x_m^i = 1 \text{ and } x_m^j = 0 \text{ for } j \neq i \end{gathered} \tag{3.25}$$

where $\tilde{v}_{m,i}^{\text{ref}}(t)$ is the entry of $\tilde{\boldsymbol{v}}^{\text{ref}}(t)$ corresponding to x_m^i, $\hat{K}_{m,i}$ and $\hat{L}_{m,i}$ are the row vectors of $\hat{K}$ and $\hat{L}$ corresponding to x_m^i.

Define an $M \times 3M$ matrix $\mathfrak{D}$ as:

$$\mathfrak{D} \triangleq \text{diag}(\underbrace{[1\ 1\ 1], \ldots, [1\ 1\ 1]}_{\text{repeat } M \text{times}}) \tag{3.26}$$

Then matrix X can be expressed by decision vector $\boldsymbol{x}$ as $X = \mathfrak{D}\ \text{diag}(\boldsymbol{x})$. Thus, we can simplify the second term on the right-hand-side (RHS) of (3.25) as

$$\begin{aligned}\hat{K}_{m,i}X^T\tilde{\boldsymbol{p}}(t) &= \hat{K}_{m,i}\ \text{diag}(\boldsymbol{x})\ \mathfrak{D}^T\ \tilde{\boldsymbol{p}}(t)\\ =&\boldsymbol{x}^T\ \text{diag}(\hat{K}_{m,i})\ \mathfrak{D}^T\ \tilde{\boldsymbol{p}}(t) = \boldsymbol{x}^T\ \boldsymbol{\zeta}_{m,i}(t) = \boldsymbol{\zeta}_{m,i}^T(t)\ \boldsymbol{x}\end{aligned} \tag{3.27}$$

where $\boldsymbol{\zeta}_{m,i}(t) \triangleq \text{diag}(\hat{K}_{m,i})\ \mathfrak{D}^T\ \tilde{\boldsymbol{p}}(t)$. Similarly, simplify the third term on the RHS of (3.25) as

$$\hat{L}_{m,i}X^T\tilde{\boldsymbol{q}}(t) = \boldsymbol{\xi}_{m,i}^T(t)\ \boldsymbol{x} \tag{3.28}$$

where $\boldsymbol{\xi}_{m,i}(t) \triangleq \text{diag}(\hat{L}_{m,i})\ \mathfrak{D}^T\ \tilde{\boldsymbol{q}}(t)$.

Substituting (3.27) and (3.28) into equation (3.25), we have

$$\begin{aligned}&\tilde{v}_m(t) - \tilde{v}_{m,i}(t, \boldsymbol{x}_{-m})\\ =&\tilde{v}_m(t) - \tilde{v}_{m,i}^{\text{ref}}(t) - \boldsymbol{\zeta}_{m,i}^T(t)\boldsymbol{x} - \boldsymbol{\xi}_{m,i}^T(t)\boldsymbol{x}\\ =&\tilde{v}_m(t) - \tilde{v}_{m,i}^{\text{ref}}(t) - \boldsymbol{\psi}_{m,i}^T(t)\boldsymbol{x}\\ =&\tilde{v}_m(t) - \tilde{v}_{m,i}^{\text{ref}}(t) - [\boldsymbol{\varphi}_{m,i}^T(t)\boldsymbol{x}_{-m} + \eta_{m,i}(t)]\\ =&v_{m,i}^{\text{tot}}(t) - \boldsymbol{\varphi}_{m,i}^T(t)\boldsymbol{x}_{-m}\end{aligned} \tag{3.29}$$

Where $\boldsymbol{\psi}_{m,i}(t) \triangleq \boldsymbol{\zeta}_{m,i}(t) + \boldsymbol{\xi}_{m,i}(t)$. $\boldsymbol{\varphi}_{m,i}(t)$ is a vector containing all the elements in $\boldsymbol{\psi}_{m,i}(t)$ except the three elements corresponding to x_m^1, x_m^2, and x_m^3. $\eta_{m,i}(t)$ is the element in $\boldsymbol{\psi}_{m,i}(t)$ corresponding to x_m^i. In the last line of (3.29), $v_{m,i}^{\text{tot}}(t)$ is defined as $v_{m,i}^{\text{tot}}(t) \triangleq \tilde{v}_m(t) - \tilde{v}_{m,i}^{\text{ref}}(t) - \eta_{m,i}(t)$.

Note that our proposed phase identification method still works even if there is a topology change in the primary feeder. If such topology change occurs at time t_c, then we can simply update $v_{m,i}^{\text{tot}}(t)$ and $\boldsymbol{\varphi}_{m,i}(t)$ in (3.29) according to the new primary feeder topology.

With (3.29), the function $f_{m,i}(\boldsymbol{x}_{-m})$ can be transformed into

$$f_{m,i}(\boldsymbol{x}_{-m}) = \frac{1}{T}\sum_{t=1}^{T}[v_{m,i}^{\text{tot}}(t) - \boldsymbol{\varphi}_{m,i}^{T}(t)\boldsymbol{x}_{-m}]^2 \tag{3.30}$$

Now each MMLE sub-problem in (3.23) can be formulated as

$$\begin{aligned} \text{find} \qquad & \boldsymbol{x}_{-m,i}^{\dagger} = \arg\min_{\boldsymbol{x}_{-m}} \ f_{m,i}(\boldsymbol{x}_{-m}) \\ \text{subject to} \qquad & x_k^j = 0 \text{ or } 1 \quad \forall j \text{ and } k \neq m \\ & \sum_j x_k^j = 1 \quad \forall k \neq m. \end{aligned} \tag{3.31}$$

This is a binary least-square problem. To solve it efficiently, we can further relax the problem by replacing the binary constraint by its convex hull. Now the problem is equivalent to convex quadratic programming, which can be solved in polynomial time [75]. The continuous solution of $\boldsymbol{x}_{-m}$ in the convex hull can then be rounded to binary values as follows: for each load $k \neq m$, round x_k^j to 1 if it is the largest among x_k^1, x_k^2, and x_k^3, and round the other two variables to 0.

3.6.4 Phase Identification Algorithm

Our proposed MMLE-based phase identification algorithm is summarized in Algorithm 3 and explained as follows. From step 1 to 6, we solve M MMLE problems, each of which contains three binary least-square sub-problems. Step 3 solves the sub-problems of MMLE based on (3.31). Based on (3.23), step 5 solves the mth MMLE problem by finding which of the three $\boldsymbol{x}_{-m,i}^{\dagger}$ $(i = 1, 2, 3)$ minimizes $f_m(\boldsymbol{x})$. The chosen $\boldsymbol{x}_{-m,i}^{\dagger}$, combined with the corresponding $x_m^i = 1$ and $x_m^j = 0$ $(j \neq i)$, forms the $3M \times 1$ solution $\boldsymbol{x}^{\dagger}{}_m$ of the mth MMLE problem. The M sets of $\boldsymbol{x}^{\dagger}{}_m$ may not be all correct due to the limited number of measurements and measurement noise. Thus, in step 7, we design two approaches to integrate M sets of $\boldsymbol{x}^{\dagger}{}_m$ into two phase identification solutions:

1. *Target-only Approach.* The phase connection of each load m is the corresponding connection shown in the mth solution ${\boldsymbol{x}^\dagger}_m$.

2. *Voting Approach.* For a single-phase or two-phase load m, the phase connection is the corresponding phase connection that receives the most votes in the M sets of ${\boldsymbol{x}^\dagger}_m$. For a three-phase load m, the phase connection is still determined by the target-only approach.

In step 8, we calculate $\sum_{m=1}^{M} f_m(\boldsymbol{x})$ based on the phase identification solution of both the target-only and the voting approaches. The final phase identification solution is the one that has the lower sum of square error.

Algorithm 3 Phase Identification Algorithm

Input: $\tilde{\boldsymbol{v}}(t)$, $\tilde{\boldsymbol{v}}^{\text{ref}}(t)$, $\tilde{\boldsymbol{p}}(t)$, $\tilde{\boldsymbol{q}}(t)$, $\hat{K}$, and $\hat{L}$, $t = 1, ..., T$.
Output: Estimated phase connections for the M loads.

1: **for** $m = 1$ to M **do**
2: **for** $i = 1$ to 3 **do**
3: Use the input to calculate $v_{m,i}^{\text{tot}}(t)$ and $\boldsymbol{\varphi}_{m,i}^{T}(t)$ and find the solution $\boldsymbol{x}_{-m,i}^{\dagger}$ to the sub-problem in (3.31).
4: **end for**
5: Use $\boldsymbol{x}_{-m,i}^{\dagger}$, $i \in \{1, 2, 3\}$ to find the $\boldsymbol{x}$ that minimizes $f_m(\boldsymbol{x})$ in (3.21). Record the solution as ${\boldsymbol{x}^\dagger}_m$.
6: **end for**
7: Generate two phase identification results based on M sets of ${\boldsymbol{x}^\dagger}_m$ using two approaches: the target-only approach and the voting approach.
8: Calculate $\sum_{m=1}^{M} f_m(\boldsymbol{x})$ based on both the target-only and the voting approach. Select the solution with the lower sum of square error.

3.7 Numerical Study

3.7.1 Setup for Numerical Tests

The performance of our proposed MMLE-based algorithm is evaluated using the IEEE 37-bus, 123-bus, and 342-bus test circuits. Fig. 3.1 illustrates the schematic of the 123-bus circuit. The results show that the proposed algorithm works well for distribution networks with either tree structured feeders (37-bus and 123-bus) or heavily meshed primary feeders (342-bus). The 342-bus feeder represents meshed distribution systems that are often used in urban centers, which have a high load density and require very high reliability. In North America, about 80 cities currently operate such distribution systems [76]. To make the task more difficult, we modify the test feeders to include all possible phase connection types (single-phase, two-phase, and three-phase). The number of loads by phase connection type is summarized in Table 3.2.

Table 3.2: Number of Loads Per Phase in the IEEE Test Circuits and Level of Unbalance

Feeder	Phase Connection								Level of
	A	B	C	AB	BC	CA	ABC	Total	Unbalance
37-bus	5	5	6	3	2	2	2	25	0.027
123-bus	18	17	17	9	9	10	5	85	0.0164
342-bus	30	38	31	35	31	33	10	208	0.0097

The hourly average real power consumption measurements from smart meters of a distribution feeder managed by FortisBC are used in test feeders. The length of the real power consumption time series is 2160, which represents 90 days of hourly smart meter measurements. The reactive power time series are generated by randomly sampling power factors from a uniform distribution $\mathcal{U}(0.9, 1)$ to represent lagging loads. The peak loads for the three IEEE test circuits are 2.4 MW, 4 MW, and 43 MW. The power flows of the test circuits are simulated using OpenDSS. All

smart meter measurements contain noise that follows zero-mean Gaussian distributions with three-sigma deviation matching 0.1% to 0.2% of the nominal values. The 0.1 and 0.2 accuracy class smart meters established in ANSI C12.20-2015 are typical in real-world implementations. To make the phase identification task even more challenging, we assume that older generations of smart meters are adopted. That is to say, after adding measurement noise, the voltage measurements are rounded to the nearest 1 V for primary line loads and 0.1 V for secondary loads. The real and reactive power measurements are rounded to the nearest 0.1 kW or 0.1 kVAr. The relaxed optimization problems in equation (3.31) are solved using CPLEX on a DELL workstation with 3.3 GHz Intel Xeon CPU and 16 GB of RAM.

In the simulation, the power consumption time series are allocated relatively evenly to each phase so that the test feeders are close to balance. Following [31], the level of unbalance of a feeder at time interval t can be measured as

$$u(t) = \frac{|I_A(t) - I_m(t)| + |I_B(t) - I_m(t)| + |I_C(t) - I_m(t)|}{3I_m(t)} \tag{3.32}$$

where $I_m(t) = \frac{1}{3}(I_A(t) + I_B(t) + I_C(t))$ is the mean of the distribution substation line current magnitudes of the three phases at time interval t. $u(t)$ represents the power deviation of each phase from the average value at time interval t. We use the 90-day average value of $u(t)$ in this simulation to measure the level of unbalance of a feeder. The level of unbalance of the three feeders are shown in Table 3.2.

Before presenting the main numerical results, we first verify the Gaussianity assumption for the noise term $\boldsymbol{n}(t)$ in equation (3.16). The Kolmogorov-Smirnov test is used to verify the Gaussianity assumption. With a significance level of 5%, the noise terms for all loads pass the test except 9 loads at 0.1% meter accuracy level and 1 load at 0.2% meter accuracy level in the 342-bus

circuit. By checking the normalized auto-correlations of $n(t)$, we found the noise to be uncorrelated over time. For Gaussian random variables, this indicates independence over time.

3.7.2 Performance of the Proposed Phase Identification Method

The phase identification accuracy of our proposed MMLE-based algorithm is shown in Table 3.3, which covers three IEEE test feeders, two meter accuracy classes (0.1% and 0.2%), and three time windows (30 days, 60 days, 90 days). With 90 days of hourly meter measurements and both accuracy class meters, the proposed algorithm achieved 100% accuracy for all three IEEE distribution test circuits. The proposed algorithm works well not only for radial feeders (37-bus, 123-bus), but also the meshed circuit (342-bus). As shown in the table, the accuracy of the MMLE-based phase identification algorithm increases as the smart meter measurement error decreases. When additional smart meter data becomes available, the phase identification accuracy of the proposed algorithm also increases as expected. The average computation time of the algorithm with 90 days of data is only around 1.3 seconds, 6.5 seconds, and 256 seconds for the three circuits, respectively.

Table 3.3: Accuracy of the Proposed Phase Identification Method in Feeders Close to Balance

Feeder	Meter Class	30 Days	60 Days	90 Days
37-bus	0.1%	100%	100%	100%
	0.2%	92%	100%	100%
123-bus	0.1%	96.47%	100%	100%
	0.2%	63.53%	96.47%	100%
342-bus	0.1%	96.63%	100%	100%
	0.2%	72.60%	99.52%	100%

We also test our proposed method with higher unbalance levels by adjusting the load levels in each phase of the test feeders. The result shows that the method is very accurate even if the

feeder is severely unbalanced. Table 3.4 shows the phase identification accuracy using 0.1% meter class in three unbalance levels: the original close-to-balance level, 0.05, and 0.1. The feeders are considered to be moderately unbalanced when the unbalance level is 0.05. The feeders are deemed as severely unbalanced, when the unbalance level is 0.1. As shown in Table 3.4, the phase identification accuracy of our proposed algorithm gradually decreases as the unbalance level increases for the 123-bus feeder. This is because our proposed method is derived from an approximate model of distribution feeders that are close to balance. Thus, the approximation error will increase when the feeder becomes more unbalanced. However, at 0.1 unbalance level, our proposed method still attains high accuracy with sufficient smart meter data.

Table 3.4: Accuracy of the Proposed Phase Identification Method With Different Unbalance Levels (Meter Accuracy Class 0.1%)

Feeder	Level of Unbalance	30 Days	60 Days	90 Days
37-bus	0.027	100%	100%	100%
	0.05	100%	100%	100%
	0.1	100%	100%	100%
123-bus	0.0164	96.47%	100%	100%
	0.05	95.29%	100%	100%
	0.1	85.88%	100%	100%
342-bus	0.0097	96.63%	100%	100%
	0.05	96.63%	99.52%	100%
	0.1	96.63%	99.52%	100%

We verify the robustness of our proposed phase identification algorithm against bad data. We assume that a meter with bad data has erroneous voltage and power measurements in 10% of the hours. The erroneous voltage measurements are assumed to have a uniform distribution within $\pm 20\%$ of the true values. The erroneous real and reactive power measurements are assumed to follow uniform distributions within $\pm 100\%$ of the true values. We test our method when 1%, 5%, and 10% of the smart meters have bad data on the 123-bus and 342-bus feeders. The 37-bus feeder

only has 25 smart meters. Thus, we increase percentage of meters with bad data to 4%, 8%, and 12%. Table 3.5 shows the phase identification accuracy of our proposed method using 0.1% meter class when the feeders are close to balance. The average accuracy of 30 test cases are reported in the table. In each test case, we randomly select the meters with bad data. As shown in Table 3.5, the phase identification accuracy of our proposed algorithm gradually decreases as more meters are compromised with bad data. However, even with 10% bad meters, our algorithm can still achieve over 93% accuracy on the most complex circuit with 90 days of smart meter data.

Table 3.5: Average Accuracy of the Proposed Phase Identification Method With Bad Data (Meter Accuracy Class 0.1%)

Feeder	% of Meters with Bad Data	30 Days	60 Days	90 Days
37-bus	0%	100%	100%	100%
	4%	98.53%	98.53%	100%
	8%	96.53%	97.20%	99.07%
	12%	91.47%	95.73%	97.87%
123-bus	0%	96.47%	100%	100%
	1%	94.94%	99.80%	99.92%
	5%	90.51%	99.33%	99.76%
	10%	88.51%	98.43%	99.61%
342-bus	0%	96.63%	100%	100%
	1%	95.06%	98.93%	99.17%
	5%	91.54%	96.47%	97.15%
	10%	86.71%	93.35%	93.83%

3.7.3 Comparison With Existing Methods

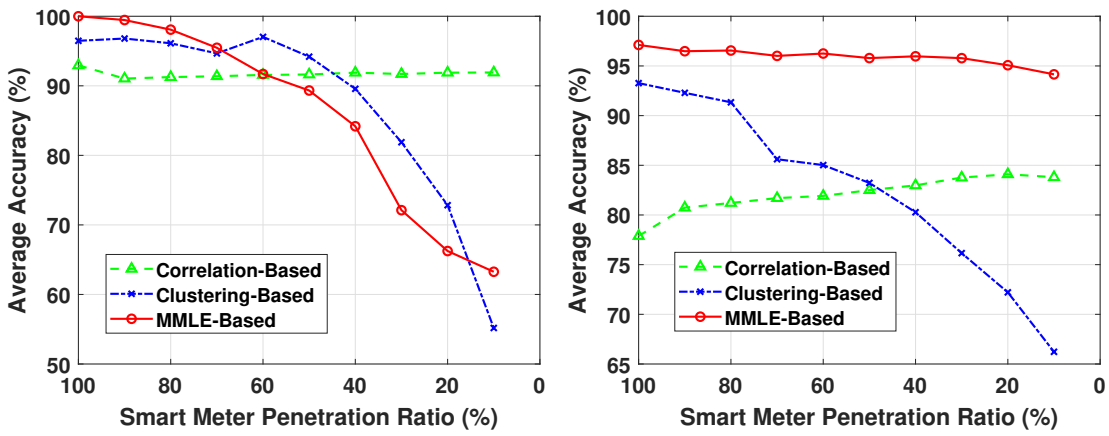
The phase identification accuracy of our proposed MMLE-based method is compared with two state-of-the-art methods: the correlation-based approach [71] and the clustering-based approach [31]. We also evaluate the robustness of the phase identification algorithms with respect to inaccurate feeder models, incomplete measurements, and bad data.

The 123-bus and 342-bus test feeders with 90 days of 0.1% accuracy class smart meter measurements are used for the comparison. To introduce incomplete smart meter measurements, we gradually decrease the penetration ratio of smart meters from 100% to 10% with a 10% step. To create inaccurate feeder models, we introduce noisy network parameters and inaccurate topology information. Specifically, we add zero-mean Gaussian noise with three-sigma deviation matching 30% of the nominal values to the actual line admittance of the 123-bus and 342-bus feeders. Eight secondary branches are assumed to be missing in the topology model of the 342-bus feeder.

Note that the correlation-based method [71] was originally designed to handle single-phase loads only. Thus, we extend it to accommodate two-phase loads. To make it a fair comparison, we assume that the information of whether a particular load is one-phase, two-phase, or three-phase is known to all algorithms. Inaccurate feeder models and incomplete measurements do not affect the correlation-based and clustering-based algorithms directly. This is because these two methods do not rely on the primary feeder model. Similarly, the MMLE-based method simply constructs a formulation with a smaller decision vector $\boldsymbol{x}$ when dealing with incomplete meter measurements.

The average phase identification accuracies of the proposed algorithm and two benchmark algorithms with different smart meter penetration ratios and inaccurate feeder models are shown in Fig. 3.2. When the smart meter penetration rate is not 100%, we randomly select the location of smart meters around 50 times and calculate the average accuracies.

As shown in Fig.3.2, our proposed MMLE-based algorithm achieves around 97% accuracy on the 342-bus feeder at the 100% smart meter penetration rate. This is lower than the 100% accuracy reported in Table 3.3 due to an inaccurate primary feeder model. Our proposed algorithm yields higher accuracy for the 123-bus radial feeder when the smart meter penetration rate is at 70%



(a) 123-bus (b) 342-bus

Figure 3.2: Average phase identification accuracy of three methods with inaccurate feeder models (0.1% meter class, 90 days' data).

or higher. Note that the smart meter penetration level is already higher than 70% in hundreds of thousands of distribution circuits and keeps increasing around the world. In the U.S., more than 40 electric companies have fully deployed smart meters [77] by the end of 2016. The smart meter penetration level in North America is expected to reach 81% in 2024 [11]. In European countries such as Italy, Sweden, Finland, and the Netherlands, smart meter penetration levels have reached 80% and are expected to pass 95% in 2020 [12]. As the penetration level of smart meters continues to increase around the world, the comparative advantage of our proposed algorithm will become more pronounced.

For the more complex 342-bus feeder, which is heavily meshed, our proposed algorithm outperforms both existing algorithms across all smart meter penetration levels. Our proposed algorithm is more robust with respect to incomplete measurements on the heavily meshed 342-bus feeder than on the radial 123-bus feeder. To explain this phenomenon, we examine the sensitivity of $\tilde{v}_m(t, \boldsymbol{x})$, the smart meter voltage measurement for load m, with respect to the phase connection

decision vector $\boldsymbol{x}$. It turns out that in the 342-bus feeder, load m's voltage measurement is more sensitive to its own phase connection decision variables and less sensitive to the phase connection decision variables of other loads.

Finally, we compare the performance of all three phase identification methods when there is bad data in the smart meter measurements. Table 3.6 shows the phase identification accuracy under different ratios of meters with bad data, using 90 days of 0.1% accuracy class meter measurements. The result is based on 100% smart meter penetration level and accurate feeder parameters. The average accuracies over 30 test cases are reported. In each test case, we randomly select the meters with bad data. As shown in Table 3.6, in the presence of bad data, our proposed phase identification algorithm always yields higher accuracy than the two benchmark algorithms when the smart meter penetration level is 100%.

Table 3.6: Average Accuracy of Three Phase Identification Methods With Bad Data (Meter Accuracy Class 0.1%, 90 Days' Data)

Feeder	% of Meters with Bad Data	Correlation-Based Approach [71]	Clustering-Based Approach [31]	MMLE-Based Algorithm
37-bus	0%	92%	100%	100%
	4%	91.87%	98.40%	100%
	8%	92%	97.60%	99.07%
	12%	91.73%	96.40%	97.87%
123-bus	0%	92.94%	96.47%	100%
	1%	88.35%	96.43%	99.92%
	5%	89.14%	94.24%	99.76%
	10%	89.33%	92.51%	99.61%
342-bus	0%	77.88%	93.27%	100%
	1%	80.51%	51.76%	99.17%
	5%	80.54%	49.55%	97.15%
	10%	79.25%	48.51%	93.83%

3.8 Conclusion

In this chapter, we develop a physics-informed data-driven algorithm for the phase identification in power distribution systems. The phase identification problem is first formulated as an MLE and MMLE problem based on the three-phase power flow manifold. We prove that the correct phase connection is a global optimum for both the MLE and the MMLE problems. A computationally efficient algorithm is developed to solve the MMLE problem, which involves synthesizing the solutions from the sub-problems via the voting and the target-only approaches. The sub-problems are further transformed into an equivalent binary least square form and solved efficiently by relaxing the binary constraints. Comprehensive simulation results with real-world smart meter data and IEEE distribution test circuits show that our proposed phase identification algorithm yields high accuracy and outperforms existing methods. The proposed algorithm is also fairly robust with respect to inaccurate feeder models, incomplete measurements, and bad measurements.

3.9 List of Symbols in Chapter 3

I_n	$n \times n$ identity matrix.
$Im(\cdot)$	Imaginary part of a complex variable.
M	Number of loads in a circuit.
N	Number of three-phase non-substation nodes in the distribution network.
$Re(\cdot)$	Real part of a complex variable.
Y^{ij}	Bus admittance matrix between phase i and j.

$\text{diag}(\cdot)$	$\text{diag}(\boldsymbol{x})$ of a vector $\boldsymbol{x}$ is a diagonal matrix with $\boldsymbol{x}$ on the main diagonal. $\text{diag}(X_1, ..., X_n)$ is a block diagonal matrix with diagonal matrices of $X_1, ..., X_n$.
$\boldsymbol{v}$, $\boldsymbol{\theta}$, $\boldsymbol{p}$, $\boldsymbol{q}$	Vector of voltage magnitudes, voltage angles, real power injections, and reactive power injections of 3 phases of the nodes.
$\check{\boldsymbol{v}}$, $\check{\boldsymbol{\theta}}$, $\check{\boldsymbol{p}}$, $\check{\boldsymbol{q}}$	Non-substation nodes' voltage magnitude and angle difference with the substation, and their real and reactive power.
$\hat{\boldsymbol{v}}$, $\hat{\boldsymbol{p}}$, $\hat{\boldsymbol{q}}$	Vectors of load measurements of voltage magnitudes, real power, and reactive power injections.
$\tilde{\boldsymbol{v}}$, $\tilde{\boldsymbol{p}}$, $\tilde{\boldsymbol{q}}$	Time differenced load measurements of voltage magnitudes, real, and reactive power injections.
$\overline{\boldsymbol{v}}$, $\overline{\boldsymbol{\theta}}$	Flat voltage solution of three-phase power flow.
$\hat{\boldsymbol{v}}^{\text{ref}}$	Vector of reference voltage for loads.
x_m^i	Decision variable of load m's phase connection.
$\boldsymbol{x}$	Vector of decision variables x_m^i.
$\boldsymbol{x}^*$	True value of the decision variable vector $\boldsymbol{x}$.
α	Rotation operator, $\alpha = e^{-j\frac{2\pi}{3}}$.
$\mathbf{1}_n$	An all-1 vector of size n.
$(\cdot)^i$	A variable in phase i.
$(\cdot)^{ij}$	A variable between phase i and j.
$(\cdot)_n$	A variable at node or load n.
$(\cdot)_{-n}$	A variable excluding node or load n.
$\cdot(t)$	The value of a variable at time t.

Chapter 4

Parameter Estimation by the Physics-Informed Model Approach

4.1 Introduction

Accurate modeling of three-phase power distribution systems is gaining importance with the the rapid increasing penetration of distributed energy resources (DERs). To monitor and coordinate the operations of DERs in distribution networks, distribution system operators need key applications such as three-phase power flow, distribution system state estimation, three-phase optimal power flow, and distribution network reconfiguration. All of these applications rely on accurate models of three-phase distribution networks, which include the network topology and parameters. However, the distribution network parameters and topology in the geographic information system (GIS) may contain errors due to unreliable documentation during the system modifications and upgrades.

Though the topic of topology estimation for distribution networks has been studied extensively [30, 31, 78, 67, 79], the estimation of distribution network parameters such as line impedance has not received sufficient attention. The task of parameter estimation in power distribution networks is more challenging than that in transmission networks because the distribution lines are almost always not transposed. Untransposed lines will lead to unequal diagonal and off-diagonal terms in the phase impedance matrix. Thus, instead of single-phase models, three-phase line segment models need to be developed. Specifically, the elements of a 3×3 phase impedance matrix need to be estimated for each three-phase distribution line segment.

In this chapter, we propose a parameter estimation approach, which considers three-phase series impedance and only leverages readily available smart meter measurements. We first build a physical model based on the linearized three-phase power flow manifold, which links the network parameters with the smart meter measurements. The parameter estimation problem is then formulated as a maximum likelihood estimation (MLE) problem. We prove that the correct network parameters yield the highest likelihood value. A stochastic gradient descent (SGD) method with early stopping is then adopted to solve the MLE problem. Comprehensive numerical tests show that the proposed algorithm improves the accuracy of the network parameters.

The rest of this chapter is organized as follows. Section 4.2 reviews the literature and summarizes our work's contribution. Section 4.3 formulates the problem of network parameter estimation. Section 4.4 presents the physics-informed model linking network parameters with smart meter measurements. Section 4.5 formulates the parameter estimation problem as an MLE problem and proposes an SGD algorithm to solve it. Section 4.6 shows the results of numerical study. Section 4.7 states the conclusion.

4.2 Literature Review and Contributions of Our Work

Many technical methods have been developed to estimate transmission network parameters. However, very few of them can be applied to the three-phase distribution networks with readily available sensor data. The existing parameter estimation literature can be roughly classified into three groups based on the sensor data used. The sensor data that were used for parameter estimation include supervisory control and data acquisition (SCADA) system information, phasor measurement unit (PMU) data, and smart meter data.

The first group of literature [22, 23] uses SCADA data such as power and current injections to estimate network parameters of the transmission system with a single-phase model. Most of these works perform joint state and parameter estimation by residual sensitivity analysis, state vector augmentation, and Kalman filter.

The second group of literature uses time synchronized measurements such as voltage and current phasors to estimate single-phase line models in transmission systems and three-phase line models in distribution networks [80, 81, 82, 83, 84]. Although these algorithms can achieve highly accurate network parameter estimates, they require widespread installation of PMUs, which are extremely costly.

The third group of literature uses readily available smart meter data to estimate network parameters of distribution systems [39, 85, 86]. Linear regression models are fitted based on voltage magnitude and complex power consumption measurements to estimate line parameters of single-phase secondary feeders [39, 86]. By solving power flow equations with voltage magnitude and complex power measurements, the parameters of a single-phase distribution line model can be estimated [85].

In this chapter, we propose a data-driven algorithm to estimate the serial conductance and serial susceptance of the π equivalent model for three-phase distribution lines by using the readily available smart meter measurements of voltage magnitude, real power consumption, and reactive power consumption. The serial conductance and susceptance are the real and imaginary part of the inverse of a line's phase impedance matrix. By linearizing the three-phase power flow manifold, we first build a physical model, which links smart meter measurements and the three-phase serial conductance and susceptance. We then formulate the three-phase parameter estimation problem as a maximum likelihood estimation (MLE) problem and prove that the correct network parameters yield the highest likelihood value. At last, we adopt the stochastic gradient descent (SGD) algorithm with early stopping to solve the MLE problem.

Compared to the existing parameter estimation methods, our proposed algorithm has two advantages. First, our proposed approach is specifically designed to estimate network parameters of three-phase distribution networks, which takes unequal self and mutual serial conductance and susceptance into consideration. Second, our proposed approach only uses readily available smart meter data and can be easily applied in real-world distribution circuits.

4.3 Problem Formulation

4.3.1 Problem Setup

We aim to estimate the serial conductance and susceptance (i.e., the real and imaginary part of the inverse of the phase impedance matrix) of three-phase primary lines of a distribution feeder's network. The network contains $\mathfrak{L}$ lines/edges and $N+1$ nodes, indexed as node 0 to N, in which node 0 is the substation. In the distribution feeder, there are M loads connected to the

primary lines through the non-substation nodes. The loads in the distribution feeder can be single-phase, two-phase, or three-phase.

4.3.2 Assumptions

4.3.2.1 Availability of Measurement Data and Network Model

First, for a single-phase load on phase i, we know its power injection (both real and reactive power) and voltage magnitude of phase i. Second, for a two-phase delta-connected load between phase i and j, we know its power injection and voltage magnitude across phase i and j. Third, for a three-phase load, we know its total power injection and the voltage magnitude of a known phase i. Fourth, for the source node, we know the voltage measurement. Fifth, it is assumed that each load's phase connection is known. Sixth, the topology of the primary line network is known. Seventh, we assume that the GIS has rough estimates of the network parameters, which are inaccurate but not far from the correct values. Finally, we assume that the distribution feeder is not severely unbalanced. Assumptions one to four are based on the typical measurement configurations of smart meters and SCADA. Assumptions five to seven are based on the available information in GIS. The last assumption holds for distribution feeders under normal operations. The task of network parameter estimation is to estimate the 3×3 serial conductance and susceptance matrices of the three-phase primary line segments.

4.3.2.2 Statistical Assumptions

First, we assume that the incremental changes in measured real, reactive power, and voltage magnitudes across different time intervals are independent. Second, we assume that the noise

term which represents the model errors and the measurement errors is i.i.d. Gaussian. Note that the noise term will be derived later in Section 4.4.1. Third, we assume that the noise term is independent of the incremental changes in smart meter measurements. These statistical assumptions have been verified in [32].

4.4 Physics-Informed Model for Network Parameter Estimation

4.4.1 Linearized Power Flow Model of Distribution Feeders

In order to build the model of network parameter estimation, we first introduce a linearized three-phase power flow model [32] as shown in (4.1). This linearized model links three parts of a distribution system: the first difference of smart meter measurement time series ($\tilde{\boldsymbol{v}}(t)$, $\tilde{\boldsymbol{v}}^{\text{ref}}(t)$, $\tilde{\boldsymbol{p}}(t)$, and $\tilde{\boldsymbol{q}}(t)$), the load phase connection X, and the primary feeder's topology and parameters (U^1, U^2, $\hat{U}^1$, $\hat{U}^2$, P, and $\check{A}$). Next, we will explain these three parts in detail. For the detailed derivation of the linearized three-phase power flow model, please refer to [32]. Note that $\boldsymbol{n}(t)$ is the noise term, which represents the model errors and the measurement errors and is assumed to be i.i.d. Gaussian.

$$\tilde{\boldsymbol{v}}(t) = X\tilde{\boldsymbol{v}}^{\text{ref}}(t) + X\begin{bmatrix} U^1 & U^2 \end{bmatrix} P\check{A}^{-1}P^T \begin{bmatrix} \hat{U}^1 & \hat{U}^2 \\ -\hat{U}^2 & \hat{U}^1 \end{bmatrix} \cdot \begin{bmatrix} X^T & \\ & X^T \end{bmatrix} \begin{bmatrix} \tilde{\boldsymbol{p}}(t) \\ \tilde{\boldsymbol{q}}(t) \end{bmatrix} + \boldsymbol{n}(t) \tag{4.1}$$

4.4.1.1 The Smart Meter Measurements

The measurements are modeled as follows. Let $\hat{\boldsymbol{v}}(t)$, $\hat{\boldsymbol{p}}(t)$, and $\hat{\boldsymbol{q}}(t)$ denote $M \times 1$ vectors of the measurement of voltage magnitude, real power, and reactive power of the M loads at time t. Let v_0^i denote the substation's voltage magnitude of phase i, and let v_0^{ij} denote the substation's

voltage magnitude across phase ij. Define a $3M \times 1$ vector $\hat{\boldsymbol{v}}^{\text{ref}}(t) \triangleq [\hat{\boldsymbol{v}}_1^{\text{ref}}(t), \ldots, \hat{\boldsymbol{v}}_M^{\text{ref}}(t)]^T$, where $\hat{\boldsymbol{v}}_m^{\text{ref}}(t) = [v_0^a(t), v_0^b(t), v_0^c(t)]$ if load m is single-phase or three-phase; $\hat{\boldsymbol{v}}_m^{\text{ref}} = [v_0^{ab}(t), v_0^{bc}(t), v_0^{ca}(t)]$ if load m is two-phase. $\tilde{\boldsymbol{v}}(t) \triangleq \hat{\boldsymbol{v}}(t) - \hat{\boldsymbol{v}}(t-1)$. $\tilde{\boldsymbol{p}}(t)$, $\tilde{\boldsymbol{q}}(t)$, and $\tilde{\boldsymbol{v}}^{\text{ref}}(t)$ are defined in a similar way as $\tilde{\boldsymbol{v}}(t)$.

4.4.1.2 The Load Phase Connection

The $M \times 3M$ block diagonal matrix $X \triangleq \text{diag}([x_1^1\ x_1^2\ x_1^3], ..., [x_M^1\ x_M^2\ x_M^3])$ represents the loads' phase connections. $x_m^i = 0$ or 1, and $\sum_i x_m^i = 1,\ \forall\ m$. If load m is single-phase, then x_m^1, x_m^2, and x_m^3 represent AN, BN, and CN connections. If m is two-phase, then x_m^1, x_m^2, and x_m^3 represent AB, BC, and CA connections. If m is three-phase and one of AN, BN, and CN voltages is measured, then x_m^1, x_m^2, and x_m^3 represent which phase is measured.

4.4.1.3 The Primary Feeder's Topology and Parameters

The primary feeder's topology and parameters are modeled as follows. Let $\alpha = e^{-j\frac{2\pi}{3}}$ and let $I_{(N+1)}$ be an identity matrix of size $N+1$. Let's first define matrix Y as follows:

$$Y \triangleq \begin{bmatrix} Y^{aa} & Y^{ab} & Y^{ac} \\ Y^{ba} & Y^{bb} & Y^{bc} \\ Y^{ca} & Y^{cb} & Y^{cc} \end{bmatrix} \tag{4.2}$$

where Y^{ij} is the $(N+1) \times (N+1)$ nodal admittance matrix between phase i and j. Define a block diagonal matrix $\Phi \triangleq \text{diag}(I_{(N+1)},\ \alpha I_{(N+1)},\ \alpha^2 I_{(N+1)})$ and define matrix A as

$$A \triangleq \begin{bmatrix} Re(\Phi^{-1}Y\Phi) & -Im(\Phi^{-1}Y\Phi) \\ -Im(\Phi^{-1}Y\Phi) & -Re(\Phi^{-1}Y\Phi) \end{bmatrix} \tag{4.3}$$

By removing the rows and columns corresponding to the substation from A, we obtain a $6N \times 6N$ matrix $\breve{A}$. P is a known constant $6N \times 6N$ permutation matrix that regroups the rows and columns of $\breve{A}$ by nodes instead of by phases.

U^1 and U^2 are $3M \times 3N$ matrices. $\hat{U}^1$ and $\hat{U}^2$ are $3N \times 3M$ matrices. These four matrices represent which of the N non-substation nodes each load is connected to and how many phases each load is connected to. Please refer to [32] for the details on the calculation of these matrices. The elements of these matrices are determined once each load's location and the number of phases are given. In this work, these four matrices are treated as constant matrices.

4.4.2 Explicit Model of Distribution Line Parameters in Linearized Power Flow Model

The distribution line parameters are implicitly considered in $\breve{A}^{-1}$ of the linearized power flow model (4.1) derived in Section 4.4.1. In this subsection, we explicitly model the distribution line parameters in the linearized power flow model.

A three-phase line segment l's serial conductance and susceptance can be represented by two symmetric matrices, the serial conductance matrix $[g]_l$ and the serial susceptance matrix $[b]_l$:

$$[g]_l \triangleq \begin{bmatrix} g_l^{aa} & g_l^{ab} & g_l^{ac} \\ g_l^{ba} & g_l^{bb} & g_l^{bc} \\ g_l^{ca} & g_l^{cb} & g_l^{cc} \end{bmatrix}, \; [b]_l \triangleq \begin{bmatrix} b_l^{aa} & b_l^{ab} & b_l^{ac} \\ b_l^{ba} & b_l^{bb} & b_l^{bc} \\ b_l^{ca} & b_l^{cb} & b_l^{cc} \end{bmatrix} \tag{4.4}$$

Since both matrices are symmetric, only 12 independent parameters need to be derived for each line segment. Define Λ as the set of all line parameters, i.e., $\Lambda \triangleq \{g_l^{ij}, b_l^{ij} \,|\, l \in \{1,\ldots,\mathfrak{L}\},\, ij \in \{aa,ab,ac,bb,bc,cc\}\}$. Define $\boldsymbol{g}^{ij}$ and $\boldsymbol{b}^{ij}$ as $\boldsymbol{g}^{ij} \triangleq [g_1^{ij}, \ldots, g_{\mathfrak{L}}^{ij}]$ and $\boldsymbol{b}^{ij} \triangleq [b_1^{ij}, \ldots, b_{\mathfrak{L}}^{ij}]$. Then, the serial conductances can be grouped in a $3\mathfrak{L} \times 3\mathfrak{L}$ matrix as in (4.5). Λ_b can be defined in a similar

way for serial susceptances. Next we define four $3\mathfrak{L} \times 3\mathfrak{L}$ matrices as in (4.6). Define two rotation matrices $R(\Phi^{-1})$ and $R(\Phi)$ as in (4.7).

$$\Lambda_g \triangleq \begin{bmatrix} \text{diag}(\boldsymbol{g}^{aa}) & \text{diag}(\boldsymbol{g}^{ab}) & \text{diag}(\boldsymbol{g}^{ac}) \\ \text{diag}(\boldsymbol{g}^{ab}) & \text{diag}(\boldsymbol{g}^{bb}) & \text{diag}(\boldsymbol{g}^{bc}) \\ \text{diag}(\boldsymbol{g}^{ac}) & \text{diag}(\boldsymbol{g}^{bc}) & \text{diag}(\boldsymbol{g}^{cc}) \end{bmatrix} \tag{4.5}$$

$$\begin{aligned} \sin(\Phi^{-1}) &\triangleq \text{diag}\big(\sin(0)\cdot I_{\mathfrak{L}}, \sin(\frac{2\pi}{3})\cdot I_{\mathfrak{L}}, \sin(-\frac{2\pi}{3})\cdot I_{\mathfrak{L}}\big) \\ \cos(\Phi^{-1}) &\triangleq \text{diag}\big(\cos(0)\cdot I_{\mathfrak{L}}, \cos(\frac{2\pi}{3})\cdot I_{\mathfrak{L}}, \cos(-\frac{2\pi}{3})\cdot I_{\mathfrak{L}}\big) \\ \sin(\Phi) &\triangleq \text{diag}\big(\sin(0)\cdot I_{\mathfrak{L}}, \sin(-\frac{2\pi}{3})\cdot I_{\mathfrak{L}}, \sin(\frac{2\pi}{3})\cdot I_{\mathfrak{L}}\big) \\ \cos(\Phi) &\triangleq \text{diag}\big(\cos(0)\cdot I_{\mathfrak{L}}, \cos(-\frac{2\pi}{3})\cdot I_{\mathfrak{L}}, \cos(\frac{2\pi}{3})\cdot I_{\mathfrak{L}}\big) \end{aligned} \tag{4.6}$$

$$\begin{aligned} R(\Phi^{-1}) &\triangleq \begin{bmatrix} \cos(\Phi^{-1}) & \sin(\Phi^{-1}) \\ -\sin(\Phi^{-1}) & \cos(\Phi^{-1}) \end{bmatrix} \\ R(\Phi) &\triangleq \begin{bmatrix} \cos(\Phi) & \sin(\Phi) \\ -\sin(\Phi) & \cos(\Phi) \end{bmatrix} \end{aligned} \tag{4.7}$$

Let $\mathcal{A}$ denote the $(N+1) \times \mathfrak{L}$ incidence matrix representing the topology of the primary feeder. If line segment l connects node i and j ($i < j$), then $\mathcal{A}_{il} = 1$, $\mathcal{A}_{jl} = -1$, and $\mathcal{A}_{kl} = 0, \forall k \neq i, j$. By removing the row corresponding to the substation, we obtain a $N \times \mathfrak{L}$ matrix $\check{\mathcal{A}}$. Define $\check{\mathcal{A}}_{6N}$ as $\check{\mathcal{A}}_{6N} \triangleq \text{diag}(\check{\mathcal{A}}, \check{\mathcal{A}}, \check{\mathcal{A}}, \check{\mathcal{A}}, \check{\mathcal{A}}, \check{\mathcal{A}})$ and define Λ_y as:

$$\Lambda_y \triangleq \begin{bmatrix} \Lambda_g & -\Lambda_b \\ -\Lambda_b & -\Lambda_g \end{bmatrix} \tag{4.8}$$

Then, it can be shown that

$$\check{A} = \check{A}(\Lambda) = \check{\mathcal{A}}_{6N} R(\Phi^{-1}) \Lambda_y R(\Phi)^T \check{\mathcal{A}}_{6N}^T \tag{4.9}$$

By plugging (4.9) into (4.1), we can obtain an explicit model of network parameters in the linearized power flow model.

4.5 Maximum Likelihood Estimation of Distribution Network Parameters

In this section, we first show how to formulate the network parameter estimation problem using maximum likelihood estimation (MLE). Then, we derive the gradient of the negative log likelihood function with respect to network parameters. Lastly, we develop an SGD-based algorithm with early stopping to solve the MLE problem.

4.5.1 MLE Problem Formulation

Define $\tilde{\boldsymbol{v}}(t,\Lambda)$ as the theoretical value of $\tilde{\boldsymbol{v}}(t)$, i.e., the first difference of voltage time series with network parameters Λ as in (4.10). Then, $\tilde{\boldsymbol{v}}(t) = \tilde{\boldsymbol{v}}(t,\Lambda) + \boldsymbol{n}(t)$, in which Λ is the set of network parameters to estimate.

$$\begin{aligned}\tilde{\boldsymbol{v}}(t,\Lambda) \triangleq & X\tilde{\boldsymbol{v}}^{\text{ref}}(t) + X\begin{bmatrix} U^1 & U^2 \end{bmatrix} P\check{A}(\Lambda)^{-1}P^T \begin{bmatrix} \hat{U}^1 & \hat{U}^2 \\ -\hat{U}^2 & \hat{U}^1 \end{bmatrix} \\ & \cdot \begin{bmatrix} X^T & \\ & X^T \end{bmatrix} \begin{bmatrix} \tilde{\boldsymbol{p}}(t) \\ \tilde{\boldsymbol{q}}(t) \end{bmatrix}\end{aligned} \tag{4.10}$$

As stated in Section 4.3.2.2, we assume that the noise $\boldsymbol{n}(t)$ is independent of $\tilde{\boldsymbol{v}}^{\text{ref}}(t)$, $\tilde{\boldsymbol{p}}(t)$, and $\tilde{\boldsymbol{q}}(t)$ and is i.i.d. Gaussian $\boldsymbol{n}(t) \sim \mathcal{N}(\mathbb{0}_{M\times 1}, \Sigma_n)$, in which Σ_n is an unknown underlying covariance matrix. Given these conditions, $\boldsymbol{n}(t)$ is also independent of $\tilde{\boldsymbol{v}}(t,\Lambda)$. Thus, the likelihood

of observing $\{\tilde{\boldsymbol{v}}(t)\}_{t=1}^{T}$ given X, $\{\tilde{\boldsymbol{v}}^{\text{ref}}(t)\}_{t=1}^{T}$, $\{\tilde{\boldsymbol{p}}(t)\}_{t=1}^{T}$, and $\{\tilde{\boldsymbol{q}}(t)\}_{t=1}^{T}$ is a function of Λ show in (4.11):

$$\begin{aligned} &Prob(\{\tilde{\boldsymbol{v}}(t)\}_{t=1}^{T}|X,\{\tilde{\boldsymbol{v}}^{\text{ref}}(t)\}_{t=1}^{T},\{\tilde{\boldsymbol{p}}(t)\}_{t=1}^{T},\{\tilde{\boldsymbol{q}}(t)\}_{t=1}^{T};\Lambda) = \\ &\frac{|\Sigma_n|^{-\frac{T}{2}}}{(2\pi)^{\frac{MT}{2}}}\times\exp\Big\{-\frac{1}{2}\sum_{t=1}^{T}[\tilde{\boldsymbol{v}}(t)-\tilde{\boldsymbol{v}}(t,\Lambda)]^T\Sigma_n^{-1}[\tilde{\boldsymbol{v}}(t)-\tilde{\boldsymbol{v}}(t,\Lambda)]\Big\} \end{aligned} \tag{4.11}$$

Taking the negative logarithm of (4.11), removing the constant term, and scaling by $\frac{2}{T}$, we get

$$f(\Lambda) \triangleq \frac{1}{T}\sum_{t=1}^{T}[\tilde{\boldsymbol{v}}(t)-\tilde{\boldsymbol{v}}(t,\Lambda)]^T\Sigma_n^{-1}[\tilde{\boldsymbol{v}}(t)-\tilde{\boldsymbol{v}}(t,\Lambda)] \tag{4.12}$$

It will be shown in Lemma 3 that the correct network parameters Λ maximize the likelihood function (4.11) and minimizes $f(\Lambda)$ under two mild assumptions.

Lemma 3. *Let Λ^* be the correct network parameters. If the following two conditions are satisfied, then as $T\to\infty$, Λ^* is a global minimizer of $f(\Lambda)$.*

1. *$\boldsymbol{n}(t_k)$ is i.i.d. and independent of $\tilde{\boldsymbol{v}}^{ref}(t_l)$, $\tilde{\boldsymbol{p}}(t_l)$, and $\tilde{\boldsymbol{q}}(t_l)$, for $\forall t_k, t_l \in Z^+$.*
2. *$\tilde{\boldsymbol{v}}^{ref}(t_k)$, $\tilde{\boldsymbol{p}}(t_k)$, and $\tilde{\boldsymbol{q}}(t_k)$ are independent of $\tilde{\boldsymbol{v}}^{ref}(t_l)$, $\tilde{\boldsymbol{p}}(t_l)$, and $\tilde{\boldsymbol{q}}(t_l)$, for $\forall t_k, t_l \in Z^+,\ t_k \neq t_l$*

For the proof of Lemma 3, please refer to Appendix E of Ref. [32]. The only difference is that in this work, the decision variable is the network parameter Λ, while in Ref. [32], the decision variable is the phase connection $\boldsymbol{x}$. In real-world applications, Σ_n is unknown, so we can use I_M instead. With I_M, Lemma 3 still holds and the proof is similar.

By substituting (4.10) into (4.12), we can see that directly minimizing $f(\Lambda)$ is very difficult because it is nonconvex and highly nonlinear. Thus, we adopt SGD to solve the problem.

4.5.2 Derive Gradient of the Negative Log-likelihood Function

In this subsection, we derive the gradient of $f(\Lambda)$, which will be used to find the Λ that minimizes $f(\Lambda)$. To derive the gradient in matrix form, we define the following terms:

$$\begin{aligned} &\boldsymbol{y}(t) \triangleq \tilde{\boldsymbol{v}}(t) - X\tilde{\boldsymbol{v}}^{\text{ref}}(t),\ \boldsymbol{z}(t) \triangleq \begin{bmatrix} \tilde{\boldsymbol{p}}(t) \\ \tilde{\boldsymbol{q}}(t) \end{bmatrix} \\ &\mathcal{C} \triangleq X \begin{bmatrix} U^1 & U^2 \end{bmatrix} P,\ \mathcal{D} \triangleq P^T \begin{bmatrix} \hat{U}^1 & \hat{U}^2 \\ -\hat{U}^2 & \hat{U}^1 \end{bmatrix} \begin{bmatrix} X^T & \\ & X^T \end{bmatrix} \end{aligned} \tag{4.13}$$

Then $\tilde{\boldsymbol{v}}(t) - \tilde{\boldsymbol{v}}(t,\Lambda) = \boldsymbol{y}(t) - \mathcal{C}\check{A}(\Lambda)^{-1}\mathcal{D}\boldsymbol{z}(t)$. Using the chain rule, for $\forall\lambda \in \Lambda$, we have

$$\frac{\partial f(\Lambda)}{\partial \lambda} = \text{Tr}\left(\left[\frac{\partial f(\Lambda)}{\partial(\mathcal{C}\check{A}(\Lambda)^{-1}\mathcal{D})} \right]^T \times \frac{\partial(\mathcal{C}\check{A}(\Lambda)^{-1}\mathcal{D})}{\partial \lambda} \right) \tag{4.14}$$

where

$$\frac{\partial f(\Lambda)}{\partial(\mathcal{C}\check{A}(\Lambda)^{-1}\mathcal{D})} = -\frac{2}{T}\Sigma_n^{-1} \cdot \sum_{t=1}^{T} \big(\boldsymbol{y}(t) - \mathcal{C}\check{A}(\Lambda)^{-1}\mathcal{D}\boldsymbol{z}(t)\big)\boldsymbol{z}(t)^T \tag{4.15}$$

Calculating $\partial(\mathcal{C}\check{A}(\Lambda)^{-1}\mathcal{D})/\partial\lambda$ is equivalent to calculating every element's derivative $\partial[\mathcal{C}\check{A}(\Lambda)^{-1}\mathcal{D}]_{i,j}/\partial\lambda$, in which $[\mathcal{C}\check{A}(\Lambda)^{-1}\mathcal{D}]_{i,j}$ is the ijth element of $(\mathcal{C}\check{A}(\Lambda)^{-1}\mathcal{D})$, $i = 1,\ldots,M$ and $j = 1,\ldots,2M$. Using the chain rule, we have

$$\frac{\partial[(\mathcal{C}\check{A}(\Lambda)^{-1}\mathcal{D})]_{i,j}}{\partial\lambda} = \text{Tr}\left(\left[\frac{\partial[\mathcal{C}\check{A}(\Lambda)^{-1}\mathcal{D}]_{i,j}}{\partial\check{A}(\Lambda)} \right]^T \times \frac{\partial\check{A}(\Lambda)}{\partial\lambda} \right) \tag{4.16}$$

Define $E_{m\times n}^{(i,j)}$ as an $m \times n$ matrix, in which the ij-th element is 1 and the rest of elements are all 0. Using the rules of matrix derivatives [87], we have

$$\frac{\partial[\mathcal{C}\check{A}(\Lambda)^{-1}\mathcal{D}]_{i,j}}{\partial\check{A}(\Lambda)} = -\check{A}(\Lambda)^{-T}\mathcal{C}^T E_{M\times 2M}^{(i,j)} \mathcal{D}^T \check{A}(\Lambda)^{-T} \tag{4.17}$$

Let $[\check{A}(\Lambda)]_{i,j}$ be the ijth element of $\check{A}(\Lambda)$. Using the chain rule, we get

$$\frac{\partial[\check{A}(\Lambda)]_{i,j}}{\partial\lambda} = \text{Tr}\left(\left[\frac{\partial[\check{A}(\Lambda)]_{i,j}}{\partial\Lambda_y} \right]^T \times \frac{\partial\Lambda_y}{\partial\lambda} \right) \tag{4.18}$$

Using the rules of matrix derivatives [87], we have

$$\frac{\partial[\check{A}(\Lambda)]_{i,j}}{\partial\Lambda_y} = R(\Phi^{-1})^T \check{\mathcal{A}}_{6N}^T E_{6N\times 6N}^{(i,j)} \check{\mathcal{A}}_{6N} R(\Phi) \tag{4.19}$$

From (4.8), we have

$$\frac{\partial\Lambda_y}{\partial\lambda} = \begin{bmatrix} \partial\Lambda_g/\partial\lambda & -\partial\Lambda_b/\partial\lambda \\ -\partial\Lambda_b/\partial\lambda & -\partial\Lambda_g/\partial\lambda \end{bmatrix} \tag{4.20}$$

The calculation of $\partial\Lambda_g/\partial\lambda$ is straightforward. By (4.5), we have

$$\frac{\partial\Lambda_g}{\partial\lambda} = \begin{cases} \mathbb{0}_{3\mathfrak{L}\times 3\mathfrak{L}} & \text{if } \lambda \notin \Lambda_g \\ E_{3\mathfrak{L}\times 3\mathfrak{L}}^{(i,i)} & \text{if } \lambda \text{ is the } ii\text{-th diagonal element in } \Lambda_g \\ E_{3\mathfrak{L}\times 3\mathfrak{L}}^{(i,j)} + E_{3\mathfrak{L}\times 3\mathfrak{L}}^{(j,i)} & \text{if } \lambda \text{ is the } ij\text{-th and } ji\text{-th} \\ & \text{off-diagonal elements in } \Lambda_g \end{cases} \tag{4.21}$$

$\partial\Lambda_b/\partial\lambda$ can be calculated in a similar way. Based on the derivations above, we can calculate the gradient $\nabla f(\Lambda)$ for any given Λ by calculating $\partial f(\Lambda)/\partial\lambda$ for all $\lambda \in \Lambda$ as follows. First, calculate $\partial[\check{A}(\Lambda)]_{i,j}/\partial\lambda$ for $\forall i, j$ using (4.18), (4.19), (4.20), and (4.21). Next, calculate $\partial[(\mathcal{C}\check{A}(\Lambda)^{-1}\mathcal{D})]_{i,j}/\partial\lambda$ for $\forall i, j$ using (4.16) and (4.17). Lastly, calculate $\partial f(\Lambda)/\partial\lambda$ using (4.14) and (4.15).

4.5.3 The SGD Algorithm

We design an SGD-based algorithm with early stopping to minimize $f(\Lambda)$ and estimate the network parameters Λ. As shown in Algorithm 4, in step 1, the parameters Λ are initialized with their original values in the GIS. The initial values for the parameters are assumed to be not far from the correct values. In steps 2 to 17, we iteratively update Λ using SGD, in which we update Λ by descending $f(\Lambda)$'s gradient over a small group of samples (i.e., a batch) of size n_{batch}. We use

patience $n_{patience}$ to decide when to stop the iterative process. That is to say, the algorithm will be stopped if $f(\Lambda)$ over all T samples is not improved in $n_{patience}$ epochs (an epoch goes through all T samples in batches). Steps 4 to 9 show the procedure of updating Λ over each batch of samples, in which we use the backtracking line search of parameters a_{step}, α, and β to determine the step size in each move. In step 18, the parameters Λ with the lowest $f(\Lambda)$ is selected as the output.

Algorithm 4 Network Parameter Estimation Algorithm

Input: First difference of smart meter measurements $\tilde{\boldsymbol{v}}(t)$, $\tilde{\boldsymbol{v}}^{\text{ref}}(t)$, $\tilde{\boldsymbol{p}}(t)$, and $\tilde{\boldsymbol{q}}(t)$, $t = 1, ..., T$; feeder constant matrices $\mathcal{C}$, $\mathcal{D}$, $R(\Phi)$, $R(\Phi^{-1})$; hyperparameters n_{batch}, $n_{patience}$, a_{step}, α, and β; an initial estimate of Λ for the $\mathfrak{L}$ primary line segments.

Output: Updated estimates of Λ.

1: Let $\Lambda_{best} = \Lambda$ and $f_{best} = f(\Lambda)$, in which $f(\Lambda)$ is calculated over all T measurements. $n_{epoch} = 0$.
2: **while** $n_{epoch} < n_{patience}$ **do**
3: Randomly divide the T measurements into batches of size n_{batch}.
4: **for** each batch **do**
5: Calculate $\nabla f(\Lambda)$ over the batch following Section 4.5.2. The descent direction is $\Delta\Lambda = -\nabla f(\Lambda)$. $s = a_{step}$.
6: **while** $f(\Lambda + s\Delta\Lambda) > f(\Lambda) + \alpha s \nabla f(\Lambda)^T \Delta\Lambda$ **do**
7: $s = \beta s$
8: **end while**
9: $\Lambda = \Lambda + s\Delta\Lambda$
10: **end for**
11: Calculate $f(\Lambda)$ over all T measurements.
12: **if** $f(\Lambda) < f_{best}$ **then**
13: $f_{best} = f(\Lambda)$, $\Lambda_{best} = \Lambda$, $n_{epoch} = 0$.
14: **else**
15: $n_{epoch} = n_{epoch} + 1$
16: **end if**
17: **end while**
18: Output the Λ_{best}.

4.6 Numerical Study

4.6.1 Setup for Numerical Tests

We evaluate the performance of our proposed parameter estimation algorithm with the modified IEEE 13-bus test feeder, which is shown in Figure 4.1. We modify the standard 13-bus test feeder by introducing loads with all 7 types of phase connections, AN, BN. CN, AB, BC, CA, and ABC. The test circuits' primary feeder contains 6 line segments and 7 nodes, which serve 10 loads.

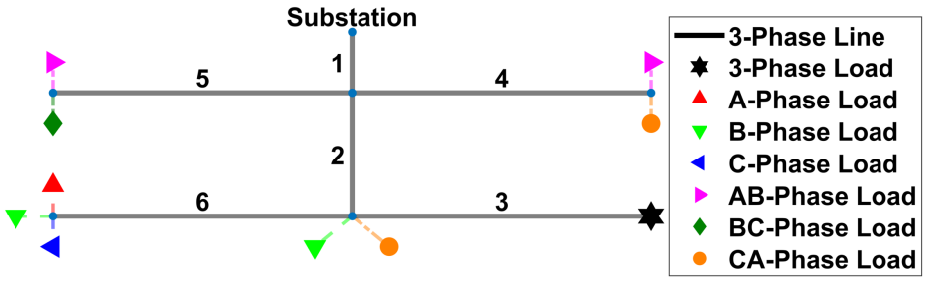


Figure 4.1: Schematic of the modified IEEE 13-bus test feeder.

We aggregated the hourly average real power consumption data from the smart meters of a distribution feeder managed by an electric utility in North America, as the hourly loads on the test feeder. The length of the real power consumption time series is 2160, which represents 90 days of measurements. The reactive power time series of the lagging loads are calculated with power factors randomly sampled from a uniform distribution $\mathcal{U}(0.9, 1)$ (a typical range for distribution network loads). The peak load of the feeder is 3 MW. The power flow results are generated using OpenDSS. All smart meter measurements contain noise, which follows a zero-mean Gaussian distribution with three-sigma deviation matching 0.1% to 0.2% of the nominal values. The 0.1 and 0.2 accuracy

class smart meters established in ANSI C12.20-2015 represent the typical noise levels in real-world systems. To make the parameter estimation task more challenging, we assume the smart meters have limited precision. That is to say, after adding measurement noise, the voltage measurements are rounded to the nearest 1V. The real and reactive power measurements are rounded to the nearest 0.1 kW and 0.1 kVar. We assume that the initial estimates for network parameters Λ are randomly sampled from a uniform distribution within $\pm 50\%$ of the correct values, which are very inaccurate starting values.

We set hyperparameters of the SGD algorithm as $n_{batch} = 10$, $n_{patience} = 10$, $a_{step} = 1e8$, $\alpha = 0.3$, and $\beta = 0.5$. These values are set empirically so that the algorithm updates $f(\Lambda)$ adequately and stops when it saturates. The SGD algorithm is implemented using MATLAB on a DELL workstation with 3.3 GHz Intel Xeon CPU and 16 GB RAM.

4.6.2 Performance of Parameter Estimation Algorithm

We demonstrate the effectiveness of our proposed network parameter estimation algorithm with two meter accuracy classes (0.1% and 0.2%) and two time windows (30 days and 90 days). We use the mean absolute deviation ratio (MADR) to measure the parameter estimation error. The MADR between the estimated Λ and the correct value Λ^* is defined in (4.22).

$$\text{MADR} \triangleq \frac{\sum_{i=1}^{12\mathfrak{L}} |\lambda_i - \lambda_i^*|}{\sum_{i=1}^{12\mathfrak{L}} |\lambda_i^*|} \times 100\% \tag{4.22}$$

The percentage of MADR improvements resulting from applying our proposed algorithm is reported in Table 4.1. In other words, we report $(\text{MADR}_{initial} - \text{MADR}_{final})/\text{MADR}_{initial} \times 100\%$, where $\text{MADR}_{initial}$ and MADR_{final} represent the MADR of the initial and the final network parameter estimates. The maximum possible MADR improvement is 100% with perfect estimation,

i.e., $\text{MADR}_{final} = 0\%$. As shown in the table, our proposed algorithm significantly reduces the network parameter estimation error. The improvement is more pronounced with longer periods of more accurate smart meter data.

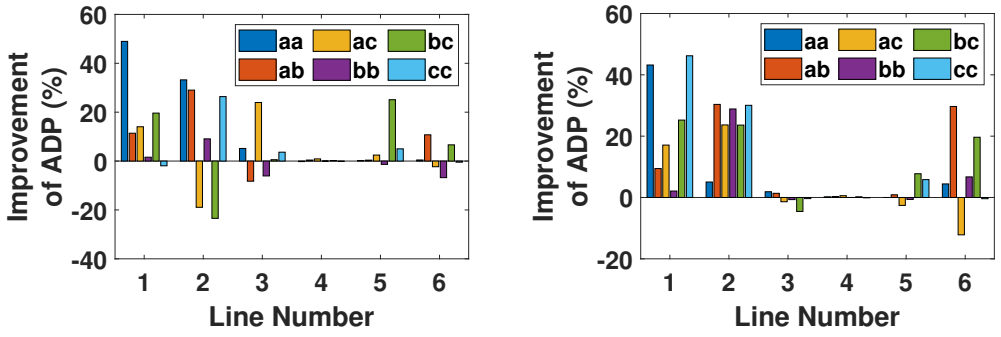
Table 4.1: Improvement in MADR of the Network Parameter Estimates of the Proposed Algorithm

Meter Class	Number of Days	MADR Improvement
0.1%	30	12.53%
	90	13.54%
0.2%	30	8.76%
	90	11.64%

To quantify the estimation error of each network parameter, we define the absolute deviation percentage (ADP) of a parameter λ_i as $|\lambda_i - \lambda_i^*|/|\lambda_i^*| \times 100\%$. Figure 4.2 shows the improvement in ADP due to the proposed algorithm, i.e., $\text{ADP}_{initial} - \text{ADP}_{final}$. As shown in the figure, our proposed algorithm reduces ADP for most of the network parameters. The improvement is more significant for line segments 1 and 2, which are the "backbones" of the feeder. Some parameters' estimation deteriorates with negative improvement, indicating that the algorithm may converge to a local minimum.

4.6.3 Performance with Different Smart Meter Penetration Levels

We also evaluate the performance of our proposed method with different smart meter penetration levels. In the 13-bus test feeder, there are 10 and 45 possible meter placement combinations with 90% and 80% smart meter penetration levels. The reduction in MADR are calculated for each case and the average reduction in MADR due to our proposed algorithm are reported in Table 4.2. As shown in the table, the improvement in MADR decreases when the penetration level of the smart



(a) Serial Conductance

(b) Serial Susceptance

Figure 4.2: Improvement in ADP of all network parameter estimates due to the proposed algorithm (0.1% meter class, 90 day data).

meters decreases. When the smart meter penetration level drops to around 80%, our proposed algorithm is no longer effective. This is because the linearized power flow model becomes inaccurate when we have incomplete smart meter measurements. Note that this limitation of our proposed algorithm will be less concerning as the penetration level of smart meters continues to increase.

Table 4.2: Impact of Smart Meter Penetration Level on the Performance of the Proposed Algorithm

Meter Class	100% Penetration	90% Penetration	80% Penetration
0.1%	13.54%	6.24%	-1.41%
0.2%	11.64%	2.23%	-7.70%

4.7 Conclusion

In this chapter, we develop a data-driven parameter estimation algorithm for three-phase power distribution networks. Our proposed algorithm uses only the readily available smart meter data to estimate the three-phase serial conductance and susceptance of the primary line segments. The network parameter estimation problem is first formulated as an MLE problem based on the

linearized three-phase power flow. It can be proven that the correct network parameters yield the highest likelihood value. We design an SGD-based algorithm with early stopping to solve the MLE problem. The comprehensive numerical study results show that our proposed algorithm is capable of improving the accuracy of the parameter estimates.

Future works can be done to develop an algorithm to jointly estimate network parameters and phase connections of distribution feeders.

4.8 List of Symbols in Chapter 4

$E_{m\times n}^{(i,j)}$	An $m \times n$ matrix, in which the ij-th element is 1 and the rest of elements are all 0.
I_n	$n \times n$ identity matrix.
$Im(\cdot)$	Imaginary part of a complex variable.
$Re(\cdot)$	Real part of a complex variable.
diag$(\cdot)$	diag$(\boldsymbol{x})$ of a vector $\boldsymbol{x}$ is a diagonal matrix with $\boldsymbol{x}$ on the main diagonal. diag$(X_1, ..., X_n)$ is a block diagonal matrix with diagonal matrices of $X_1, ..., X_n$.
$\mathbb{0}_{m\times n}$	An $m \times n$ all-0 matrix.

Chapter 5

Parameter Estimation by the Graphical Learning Model Approach

5.1 Introduction

Accurate modeling of three-phase power distribution systems is crucial to accommodating the increasing penetration of distributed energy resources (DERs). To monitor and coordinate the operations of DERs, several key applications such as three-phase power flow, state estimation, optimal power flow, and network reconfiguration are needed. All of these depend on accurate three-phase distribution network models, which include the network topology and parameters [30]. However, the distribution network topology and parameters in the geographic information system (GIS) often contain errors because the model documentation usually becomes unreliable during the system modifications and upgrades [79]. Thus, the network topology and parameters need to be accurately estimated.

Although topology estimation for distribution networks has been studied extensively [67, 32], the estimation of distribution network parameters such as line impedances still needs further development. It is more challenging to estimate parameters of power distribution networks than that of transmission networks. This is because the distribution lines are rarely transposed, which lead to unequal diagonal and off-diagonal elements in the impedance matrix. Thus, three-phase line models need to be developed instead of single-phase equivalent models. Specifically, the elements of the 3×3 phase impedance matrix need to be estimated for each three-phase line segment.

In this chapter, we develop a physics-informed graphical learning algorithm to estimate network parameters of three-phase power distribution systems. Our proposed algorithm uses only readily available smart meter data to estimate the three-phase series resistance and reactance of the primary distribution line segments. We first develop a parametric physics-based model to replace the black-box deep neural networks in the conventional graphical neural network (GNN). Then we derive the gradient of the loss function with respect to the network parameters and use stochastic gradient descent (SGD) to estimate the physical parameters. Prior knowledge of network parameters is also considered to further improve the accuracy of estimation. Comprehensive numerical study results show that our proposed algorithm yields high accuracy and outperforms existing methods.

The rest of this chapter is organized as follows. Section 5.2 reviews the literature and summarizes our work's contribution. Section 5.3 formulates the problem of network parameter estimation. Section 5.4 presents the overall framework of the proposed method and briefly introduces the GNN. Section 5.5 provides the technical methods for construction and parameter estimation based on the physics-informed graphical model. Section 5.6 evaluates the performance of the proposed algorithm with a comprehensive numerical study. Section 5.7 states the conclusion.

5.2 Literature Review and Contributions of Our work

Many methods have been proposed to estimate transmission network parameters. However, very few of them can be applied to the three-phase distribution networks using readily available sensor data. The existing parameter estimation literature can be roughly classified into three groups based on the type of sensor data used.

In the first group of literature, supervisory control and data acquisition (SCADA) system data such as power and current injections are used to estimate transmission network parameters of a single-phase model. Most of the algorithms in this group perform joint state and parameter estimation by residual sensitivity analysis and state vector augmentation [23]. Parameter errors are detected by identification indices [24, 88], enhanced normalized Lagrange multipliers [25], and projection statistics [89]. Adaptive data selection [90] is used to improve estimation accuracy.

In the second group of literature, phasor measurement unit (PMU) data such as voltage and current phasors are used to estimate line parameters of transmission and distribution systems [91, 92, 93, 94, 95, 96, 84]. Although these methods achieve highly accurate parameter estimates, they require costly and widespread installation of PMUs. Linear least squares is used to estimate transmission line parameters [91]. Parallel Kalman filter for a bilinear model is used to estimate both states and line parameters of the transmission system [92]. With single-phase transmission line models, nonlinear least squares is used to estimate line parameters and calibrate remote meters [93]. Traveling waves are used to estimate parameters of series compensated lines [94]. An augmented state estimation method is developed to estimate three-phase transmission line parameters [95]. Maximum likelihood estimation (MLE) is used to estimate single-phase distribution line parameters [96]. Lasso is adopted to estimate three-phase admittance matrix in distribution systems [84].

In the third group of literature, smart meter data such as voltage magnitude and complex power consumption are used to estimate distribution line parameters [97, 98, 99, 100, 33]. Particle swarm [97] and linear regression [86, 99] are used to estimate single-phase line parameters. Linear approximation of voltage drop [98] is used to estimate the parameters of single-phase and balanced three-phase distribution lines. Multiple linear regression model is used to estimate three-phase line impedance in [100], but it does not work with delta-connected smart meters with phase-to-phase measurement. In [33], three-phase line parameters are estimated through MLE based on a linearized physical model.

The existing methods for parameter estimation either assume a single-phase equivalent distribution network model or require widespread installation of micro-PMUs, which are cost prohibitive. To fill the knowledge gap, this work develops a physics-informed graphical learning algorithm to estimate the 3×3 series resistance and reactance matrices of three-phase distribution line model using readily available smart meter measurements. Our proposed method is inspired by the emerging graph neural network (GNN), which is designed for estimation problems in networked systems. We develop three-phase power flow-based physical transition functions to replace the ones based on deep neural networks in the GNN. We then derive the gradient of the voltage magnitude loss function with respect to the line segments' resistance and reactance parameters with an iterative method. Finally, the estimates of distribution network parameters can be updated with the stochastic gradient descent (SGD) approach to minimize the error between the physics-based graph learning model and the smart meter measurements. Prior estimates and bounds of network parameters are also leveraged to improve the estimation accuracy. To improve computation efficiency, partitions can be introduced so that parameter estimations are executed in parallel in sub-networks.

The main technical contributions of this work are:

- A physics-informed graphical learning method is developed to estimate line parameters of three-phase distribution networks.
- Our proposed algorithm only uses readily available smart meter data and can be easily applied to real-world distribution circuits.
- By preserving the nonlinearity of three-phase power flows in the graphical learning framework, our proposed approach yields more accurate parameter estimates on test feeders than the state-of-the-art benchmark.

5.3 Problem Formulation

5.3.1 Problem Setup

The objective of this work is to estimate the series resistance and reactance in the 3×3 phase impedance matrix of three-phase primary lines of a distribution feeder. The impedance matrix of a line l can be written as, $Z_l = R_l + jX_l$, where

$$R_l \triangleq \begin{bmatrix} r_l^{aa} & r_l^{ab} & r_l^{ac} \\ r_l^{ab} & r_l^{bb} & r_l^{bc} \\ r_l^{ac} & r_l^{bc} & r_l^{cc} \end{bmatrix}, \quad X_l \triangleq \begin{bmatrix} x_l^{aa} & x_l^{ab} & x_l^{ac} \\ x_l^{ab} & x_l^{bb} & x_l^{bc} \\ x_l^{ac} & x_l^{bc} & x_l^{cc} \end{bmatrix}. \tag{5.1}$$

Since Z_l is symmetric, for each line segment there are 6 resistance and 6 reactance parameters. The network contains $\mathfrak{L}$ lines and $N+1$ nodes, indexed as node 0 to N. Node 0 is the source node (e.g., a substation). In total, there are $12\mathfrak{L}$ parameters to estimate. M loads are connected to the primary lines through the non-source nodes. The loads can be single-phase, two-phase, or three-phase.

5.3.2 Assumptions

The assumptions of measurement data and the network model are summarized below. First, for a single-phase load on phase i, the smart meter records real and reactive power injections and voltage magnitude of phase i. Second, for a two-phase delta-connected load between phase i and j, the smart meter records the power injection and voltage magnitude across phase i and j. Third, for a three-phase load, the smart meter records total power injection and voltage magnitude of a known phase i. Fourth, SCADA system records the voltage measurements at the source node. Fifth, it is assumed that the phase connections of all loads are known. Sixth, the topology of the primary three-phase feeder is known. Seventh, we assume that the GIS contains rough estimates of the network parameters. The first four assumptions one to four are based on the typical measurement configurations of smart meters and SCADA. Assumptions five to seven are based on the available information in GIS.

5.4 Overall Framework and Review of GNN

5.4.1 Overall Framework of the Proposed Method

The overall framework of the proposed graphical learning method for distribution line parameter estimation is illustrated in Fig. 5.1. As shown in the figure, a physics-informed graphical learning engine is constructed based on nonlinear power flow. The inputs to the graphical learning engine include power injection measurements from smart meters, distribution network topology, and distribution line parameters. In the graphical learning engine, each node corresponds to a physical bus in the distribution network. The nodal states, i.e., the three-phase complex voltage are iteratively updated by a set of transition functions. The graphical learning engine's outputs are the estimated

smart meter voltage magnitudes, which are used to calculate the graphical learning engine's loss function. The gradient of the line parameters is computed from the loss function and subsequently used to update the line parameters using stochastic gradient descent. The technical details of the proposed method will be explained in Section 5.5.

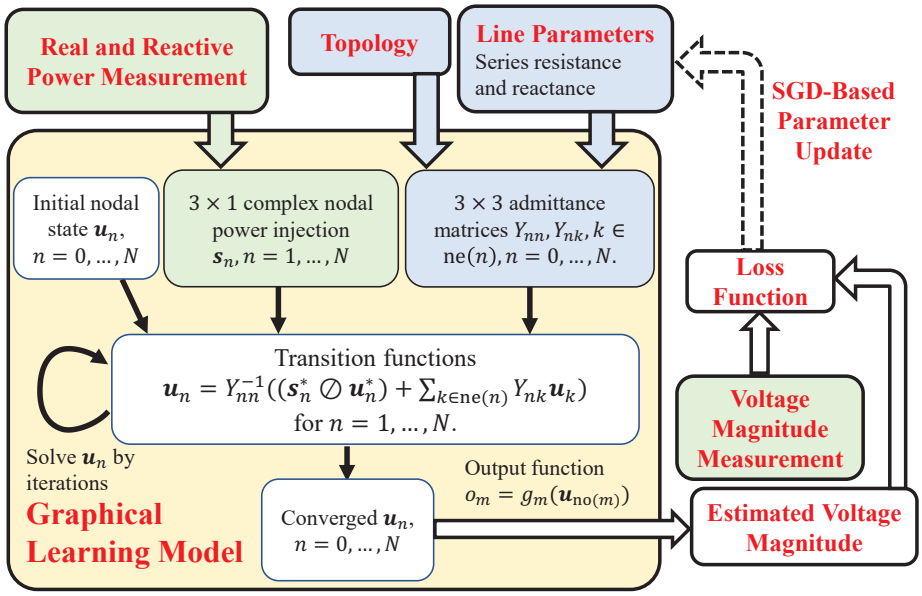


Figure 5.1: Framework of the method. The bold boxes with red titles represent higher-level elements. Green boxes represent smart meter data, and blue boxes represent distribution network information.

5.4.2 A Brief Overview of the GNN

A GNN is a neural network model, which uses a graph's topological relationships between nodes to incorporate the underlying graph-structured information in data [101]. GNNs have been successfully applied in many different domains, such as social networks, image processing, and chemistry [102]. Our proposed physics-informed graphical learning model is developed by embedding physics of power distribution networks into the standard GNN.

The GNN is comprised of nodes connected by edges. The nodes represent objects or concepts, and the edges represent the relationships between nodes. Two vectors are attached to a node n: the state vector $\boldsymbol{x}_n$ and the feature vector $\boldsymbol{l}_n$. A feature vector $\boldsymbol{l}_{(m,n)}$ is attached to edge (m, n). The state $\boldsymbol{x}_n$, which embeds information from its neighborhood with arbitrary depth, is naturally defined by the features of itself and the neighboring nodes and edges through a local parametric transition function $f_{\boldsymbol{w},n}$. A local output o_n of node n, representing a local decision, is produced through a parametric output function $g_{\boldsymbol{w},n}$. The local transition and output functions are defined as follows:

$$\begin{aligned} \boldsymbol{x}_n &= f_{\boldsymbol{w},n}(\boldsymbol{l}_n, \boldsymbol{l}_{\text{co}(n)}, \boldsymbol{x}_{\text{ne}(n)}, \boldsymbol{l}_{\text{ne}(n)}) \\ o_n &= g_{\boldsymbol{w},n}(\boldsymbol{x}_n, \boldsymbol{l}_n) \end{aligned} \tag{5.2}$$

Here, $\boldsymbol{l}_{\text{co}(n)}$, $\boldsymbol{x}_{\text{ne}(n)}$, and $\boldsymbol{l}_{\text{ne}(n)}$ are the features of edges connected to node n, the states of node n's neighbor nodes, and the features of node n's neighbor nodes. $\boldsymbol{w}$ is the set of parameters defining the transition and output functions. An example of a node and its neighbor area in a GNN is depicted in Fig. 5.2. The local transition function for node 1 is $\boldsymbol{x}_1 = f_{\boldsymbol{w},1}(\boldsymbol{l}_1, \boldsymbol{l}_{(1,2)}, \boldsymbol{l}_{(1,3)}, \boldsymbol{l}_{(1,4)}, \boldsymbol{x}_2, \boldsymbol{x}_3, \boldsymbol{x}_4, \boldsymbol{l}_2, \boldsymbol{l}_3, \boldsymbol{l}_4)$. The implementation of the transition and output functions are flexible. They can be modeled as linear or nonlinear functions (e.g., neural networks). Let $[\boldsymbol{x}]$, $[o]$, $[\boldsymbol{l}]$, and $[\boldsymbol{l}_N]$ represent the vectors

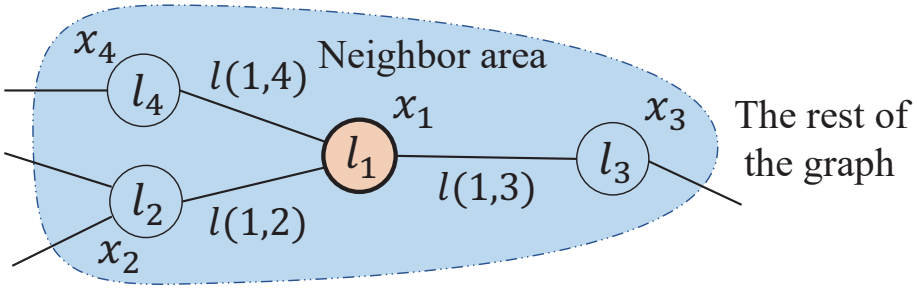


Figure 5.2: An illustration of a node and its neighbor area in a GNN.

constructed by stacking all the states, all the outputs, all the features, and all the node features, respectively. Then (5.2) can be represented in a compact form:

$$\begin{aligned} [\boldsymbol{x}] &= F_{\boldsymbol{w}}([\boldsymbol{x}], [\boldsymbol{l}]) \\ [o] &= G_{\boldsymbol{w}}([\boldsymbol{x}], [\boldsymbol{l}_N]) \end{aligned} \tag{5.3}$$

Here, $F_{\boldsymbol{w}}$ and $G_{\boldsymbol{w}}$ are the global transition function and global output function, which stacks all nodes' $f_{\boldsymbol{w},n}$ and $g_{\boldsymbol{w},n}$, respectively.

With the sufficient condition provided by the Banach fixed point theorem [103], one can find a unique solution of the state $[\boldsymbol{x}]$ for (5.3) using the classic iterative scheme:

$$[\boldsymbol{x}]^{\tau+1} = F_{\boldsymbol{w}}([\boldsymbol{x}]^{\tau}, [\boldsymbol{l}]) \tag{5.4}$$

Here, $[\boldsymbol{x}]^{\tau}$ is the τ-th iteration of $[\boldsymbol{x}]$. The dynamic system of (5.4) converges exponentially fast to the solution of system (5.3) for any initial value $[\boldsymbol{x}]^0$.

The parameters $\boldsymbol{w}$ of a GNN's global transition and output functions $F_{\boldsymbol{w}}$ and $G_{\boldsymbol{w}}$ are updated and learned such that the output $[o]$ approximate the target values, i.e., minimizing a quadratic loss function:

$$loss = \sum_{m=1}^{M} (o_m - \check{o}_m)^2 \tag{5.5}$$

Here, M is the number of elements (number of measurements) in $[o]$, and o_m and $\check{o}_m$ are the m-th output and target value. The learning algorithm is based on a gradient-descent strategy. Since the iterative scheme in (5.4) is equivalent to a recurrent neural network, the gradient is calculated in a more efficient approach based on the Almeida-Pineda algorithm. Additional technical details of the GNN can be found in [101, 104, 105].

5.5 Technical Methods

This section is organized as follows. Section 5.5.1 describes the construction of transition function $F_{\boldsymbol{w}}$. Section 5.5.2 describes the formulation of the output function $G_{\boldsymbol{w}}$ and the loss function. Section 5.5.3 derives the gradient of the loss function. The use of prior knowledge of line parameters is described in Section 5.5.4. Section 5.5.5 presents the parameter estimation algorithm. The network partition method, which improves the scalability of the algorithm is described in Section 5.5.6.

Our proposed physics-informed graphical learning model is different from the GNN [101]. In the GNN, $F_{\boldsymbol{w}}$ and $G_{\boldsymbol{w}}$ are often represented by neural networks whose weights are being learned. However, in our proposed framework, $F_{\boldsymbol{w}}$ and $G_{\boldsymbol{w}}$ are built based on the physical model of the power distribution system. The parameters to be estimated are the line resistance and reactance.

5.5.1 Construction of the Transition Function

The transition function is constructed based on the nonlinear power flow model of the distribution system. Let $\boldsymbol{s}_n \triangleq [s_n^a, s_n^b, s_n^c]^T$ be a 3×1 vector of nodal three-phase complex power injection of node n. $s_n^i \triangleq p_n^i + jq_n^i, i = a, b, c$, where p_n^i and q_n^i are node n's real and reactive power injection of phase i. $\boldsymbol{s}_n$ can be derived from smart meters' power consumption data and phase connections as described in Section III-A of [32]. Similarly, we define three-phase complex nodal voltage as $\boldsymbol{u}_n \triangleq [u_n^a, u_n^b, u_n^c]^T$, $u_n^i \triangleq \alpha_n^i + j\beta_n^i, i = a, b, c$. Let $Y_{nk} = Z_{nk}^{-1}$ be the 3×3 admittance matrix of the line between node n and k, which can be calculated by using the topology and line parameters of the distribution network. Ignoring the negligible shunt, the three-phase power flow equation of node n can be written as in (5.6). Here, $Y_{nn} = \sum_{k\in \text{ne}(n)} Y_{nk}$, $\odot$ is the element-

wise multiplication, ne(n) is the set of n's neighbor nodes, and $(\cdot)^*$ represents complex conjugate. An equivalent form of (5.6) is (5.7).

$$\boldsymbol{s}_n = \boldsymbol{u}_n \odot \left(Y_{nn}^* \boldsymbol{u}_n^* - \sum_{k \in \text{ne}(n)} Y_{nk}^* \boldsymbol{u}_k^* \right) \tag{5.6}$$

$$\boldsymbol{u}_n = Y_{nn}^{-1} \left((\boldsymbol{s}_n^* \oslash \boldsymbol{u}_n^*) + \sum_{k \in ne(n)} Y_{nk} \boldsymbol{u}_k \right) \tag{5.7}$$

Here, $\oslash$ represents element-wise division.

Next we convert (5.7) from a complex equation to a real-valued equation. For a matrix A, we define

$$\langle A \rangle \triangleq \begin{bmatrix} Re(A) & -Im(A) \\ Im(A) & Re(A) \end{bmatrix} \tag{5.8}$$

Here, $Re(A)$ and $Im(A)$ are the real and imaginary part of A. Then, (5.7) can be rewritten as the local transition function:

$$\begin{bmatrix} Re(\boldsymbol{u}_n) \\ Im(\boldsymbol{u}_n) \end{bmatrix} = \langle Z_{nn} \rangle \left(\begin{bmatrix} Re(\boldsymbol{s}_n^* \oslash \boldsymbol{u}_n^*) \\ Im(\boldsymbol{s}_n^* \oslash \boldsymbol{u}_n^*) \end{bmatrix} + \sum_{k \in \text{ne}(n)} \langle Y_{nk} \rangle \begin{bmatrix} Re(\boldsymbol{u}_k) \\ Im(\boldsymbol{u}_k) \end{bmatrix} \right) \tag{5.9}$$

Here $Z_{nn} \triangleq Y_{nn}^{-1}$. We define 6×1 state vector $\boldsymbol{x}_n$ and feature vector $\boldsymbol{l}_n$ of node n as

$$\boldsymbol{x}_n \triangleq \begin{bmatrix} Re(\boldsymbol{u}_n) \\ Im(\boldsymbol{u}_n) \end{bmatrix}, \ \boldsymbol{l}_n \triangleq \begin{bmatrix} Re(\boldsymbol{s}_n) \\ Im(\boldsymbol{s}_n) \end{bmatrix} \tag{5.10}$$

Now, we can convert the local transition function (5.9) into the standard form and the global compact form:

$$\begin{aligned} \boldsymbol{x}_n &= f_{\boldsymbol{w},n}(\boldsymbol{x}_n, \boldsymbol{l}_n, \boldsymbol{x}_{\text{ne}(n)}) \text{ (local form of node } n) \\ [\boldsymbol{x}] &= F_{\boldsymbol{w}}([\boldsymbol{x}], [\boldsymbol{l}]) \text{ (global compact form)} \end{aligned} \tag{5.11}$$

For each node in a distribution system, we can derive a local transition function and stack them to obtain the global form of $F_{\boldsymbol{w}}$ as in (5.11). Note that $[\boldsymbol{l}]$ only contains all the nodes' features and does not contain any edge features. The model's parameter $\boldsymbol{w}$ is the set of all lines' three-phase resistance and reactance, which is embedded in $\langle Z_{nn} \rangle$ and $\langle Y_{nk} \rangle$ of (5.9).

Given line parameter $\boldsymbol{w}$, we can calculate the theoretical node state values of each time instance t by iteratively applying the transition function (5.11). This iteration procedure is formulated as a function called FORWARD shown in Algorithm 5. In the algorithm, step 1 initializes all nodes' states. In step 2, the global transition function is constructed. Step 3–6 estimate the nodes' states iteratively, while $\boldsymbol{x}_0(t)$ is fixed to its initial value because it is the measurement at the reference node. The iteration continues until convergence, which is controlled by a small ratio $\epsilon_{\text{forward}}$.

Algorithm 5 FORWARD($\boldsymbol{w}$, t)

Input: Current line parameter $\boldsymbol{w}$ and the time instance t.

Output: Theoretical $[\boldsymbol{x}(t)]$ of the distribution system with line parameter $\boldsymbol{w}$.

1: Initialize the source nodes' state $\boldsymbol{x}_0(t)$ with the known measurement at the source node. Initialize the other nodes' state $\boldsymbol{x}_n(t)$ as defined in (5.10) with balanced flat node voltage, i.e. $\boldsymbol{u}_n(t) = [1, e^{-j\frac{2\pi}{3}}, e^{j\frac{2\pi}{3}}]^T$, $(n = 1, ..., N)$.

2: Construct the initial $[\boldsymbol{x}(t)]^0$ by stacking all the initial $\boldsymbol{x}_n(t)$, $(n = 0, ..., N)$. Construct function $F_{\boldsymbol{w}}$ with $\boldsymbol{w}$.

3: **repeat**

4: $\quad [\boldsymbol{x}(t)]^{\tau+1} = F_{\boldsymbol{w}}([\boldsymbol{x}(t)]^{\tau}, [\boldsymbol{l}(t)])$ and fix $\boldsymbol{x}_0(t)$ to its initial value.

5: $\quad \tau = \tau + 1$

6: **until** $\|[\boldsymbol{x}(t)]^{\tau} - [\boldsymbol{x}(t)]^{\tau-1}\|^2 < \epsilon_{\text{forward}} \cdot \|[\boldsymbol{x}(t)]^{\tau-1}\|^2$

7: **return** $[\boldsymbol{x}(t)] = [\boldsymbol{x}(t)]^{\tau}$.

5.5.2 Construction of the Output and Loss Function

The output of our proposed graphical learning model is the estimated smart meters' voltage measurements. For smart meter m, the estimated output o_m is in the form of:

$$\begin{aligned} o_m &= g_m(\boldsymbol{x}_{\text{no}(m)}) \text{ (local form of meter } m) \\ [o] &= G([\boldsymbol{x}]) \text{ (global compact form)} \end{aligned} \tag{5.12}$$

Here, $\boldsymbol{x}_{\text{no}(m)}$ is the state of the node, which the smart meter m is connected to. Suppose we have a solution of the state $[\boldsymbol{x}(t)] = \text{FORWARD}(\boldsymbol{w}, t)$, then $[o(t)] = G([\boldsymbol{x}(t)])$. Though $\boldsymbol{x}_{\text{no}(m)}$ has 6 elements from 3 phases, a smart meter only measures one single-phase or one phase-phase voltage magnitude. Based on the assumptions in Section 5.3.2, if $k = \text{no}(m)$, then g_m is defined as follows:

$$g_m(\boldsymbol{x}_k) = \begin{cases} \sqrt{(\alpha_k^i)^2 + (\beta_k^i)^2} & \text{if meter } m \text{ is single-phase or} \\ & \text{three-phase, measuring phase } i \\ \sqrt{(\alpha_k^i - \alpha_k^j)^2 + (\beta_k^i - \beta_k^j)^2} & \text{if meter } m \text{ is} \\ & \text{two-phase, measuring phase } ij \end{cases} \tag{5.13}$$

Note that in the line parameter estimation formulation, g_m does not depend on the parameter vector $\boldsymbol{w}$. Thus, it is not a parametric function.

Next we derive the loss function. To remove trends, instead of directly using the voltage output $[o]$, we use the first difference of the output time series. The estimated first difference of output time series for meter m is:

$$\tilde{o}_m(t) \triangleq o_m(t) - o_m(t-1) \tag{5.14}$$

The loss of first difference voltages at time t is:

$$e_{\boldsymbol{w}}(t) = \frac{1}{M}\sum_{m=1}^{M}\left(\tilde{v}_m(t) - \tilde{o}_m(t)\right)^2 \tag{5.15}$$

Here, M is the number of meters, $\tilde{v}_m(t) = v_m(t) - v_m(t-1)$ is the first difference of actual voltage magnitude measured by meter m. In the graphical learning model, we need to calculate the loss function over both the whole data set (i.e., all first difference instances) and mini-batch data (i.e., a smaller set of first difference instances). Thus, we define the gross loss function over a batch of data with time index set $\mathfrak{T}$ as:

$$e_{\boldsymbol{w}}(\mathfrak{T}) \triangleq \frac{1}{|\mathfrak{T}|}\sum_{t\in\mathfrak{T}} e_{\boldsymbol{w}}(t) \tag{5.16}$$

Here, $|\mathfrak{T}|$ is the size of $\mathfrak{T}$. Suppose we have measurement data over $t = 0, ..., T$, and define $\mathfrak{T}_{\text{full}} \triangleq \{t|t = 1, ..., T\}$ as the full batch for first difference time series. Then the gross error of the model over all first difference instances is $e_{\boldsymbol{w}}(\mathfrak{T}_{\text{full}})$.

5.5.3 Gradient of the Loss Function With Respect to the Line Parameters

We design a new algorithm to calculate the gradient of the loss function (5.16) of first difference voltage time series with respect to the line parameters $\boldsymbol{w}$. The gradient calculation formula in the GNN cannot be directly applied because it is derived for the data of a particular time instance, and not for time series. To derive the gradient of the loss function (5.16), we define an equivalent graphical learning model, with new state and feature vectors as follows:

$$\hat{\boldsymbol{x}}_n(t) \triangleq \begin{bmatrix} \boldsymbol{x}_n(t-1) \\ \\ \boldsymbol{x}_n(t) \end{bmatrix},\ \hat{\boldsymbol{l}}_n(t) \triangleq \begin{bmatrix} \boldsymbol{l}_n(t-1) \\ \\ \boldsymbol{l}_n(t) \end{bmatrix} \tag{5.17}$$

The corresponding equivalent transition function is:

$$\begin{aligned}\hat{\boldsymbol{x}}_n(t) &= \hat{f}_{\boldsymbol{w},n}(\hat{\boldsymbol{x}}(t)_n, \hat{\boldsymbol{l}}(t)_n, \hat{\boldsymbol{x}}(t)_{\text{ne}(n)}) \\ &\triangleq \begin{bmatrix} f_{\boldsymbol{w},n}(\boldsymbol{x}_n(t-1), \boldsymbol{l}_n(t-1), \boldsymbol{x}_{\text{ne}(n)}(t-1)) \\ f_{\boldsymbol{w},n}(\boldsymbol{x}_n(t), \boldsymbol{l}_n(t), \boldsymbol{x}_{\text{ne}(n)}(t)) \end{bmatrix}\end{aligned} \tag{5.18}$$

The compact form of (5.18) is:

$$[\![\hat{\boldsymbol{x}}(t)]\!] = \hat{F}_{\boldsymbol{w}}([\![\hat{\boldsymbol{x}}(t)]\!], [\hat{\boldsymbol{l}}(t)]) \triangleq \begin{bmatrix} F_{\boldsymbol{w}}([\boldsymbol{x}(t-1)], [\boldsymbol{l}(t-1)]) \\ F_{\boldsymbol{w}}([\boldsymbol{x}(t)], [\boldsymbol{l}(t)]) \end{bmatrix} \tag{5.19}$$

Here,

$$[\hat{\boldsymbol{x}}(t)] \triangleq \begin{bmatrix} [\boldsymbol{x}(t-1)] \\ [\boldsymbol{x}(t)] \end{bmatrix}, [\hat{\boldsymbol{l}}(t)] \triangleq \begin{bmatrix} [\boldsymbol{l}(t-1)] \\ [\boldsymbol{l}(t)] \end{bmatrix} \tag{5.20}$$

The output function of first difference voltage time series for meter m is:

$$\tilde{o}_m([\![t]\!]) = \hat{g}_m(\hat{\boldsymbol{x}}_{\text{no}(m)}([\![t]\!])) \triangleq g_m(\boldsymbol{x}_{\text{no}(m)}([\![t]\!])) - g_m(\boldsymbol{x}_{\text{no}(m)}([\![t-1]\!])) \tag{5.21}$$

The compact form of (5.21) is:

$$[\tilde{o}(t)] = \hat{G}([\hat{\boldsymbol{x}}(t)]) \triangleq G([\boldsymbol{x}(t)]) - G([\boldsymbol{x}(t-1)]) \tag{5.22}$$

Using the equivalent graphical learning model defined in (5.17)-(5.22), we can calculate the gradient of $e_{\boldsymbol{w}}(\mathfrak{T})$ over any batch of data $\mathfrak{T}$ with respect to $\boldsymbol{w}$ using an efficient function BACKWARD shown in Algorithm 6. The iterative FORWARD function can be represented as a recurrent neural network. Thus, $e_{\boldsymbol{w}}(\mathfrak{T})$'s gradient is difficult to calculate in the conventional way. To evaluate the gradient more efficiently, we design Algorithm 6 following the same backpropagation principle in [101] based on the Almeida-Pineda algorithm [104, 105]. Algorithm 6 calculates the gradient by using an intermediate variable $\boldsymbol{z}(t)$ through iterative applications of steps 5–10. The theoretical

details of designing such algorithms can be found in [101, 104, 105]. In Algorithm 6, the lengthy derivations of $\hat{A}(t)$, $\hat{\boldsymbol{b}}(t)$, and $\frac{\partial \hat{F}_{\boldsymbol{w}}([\hat{\boldsymbol{x}}(t)],[\hat{\boldsymbol{l}}(t)])}{\partial \boldsymbol{w}}$ are omitted. Please refer to the detailed derivations in Appendix B.1, B.2, and B.3, respectively. $\epsilon_{\text{backward}}$ is a small ratio controlling the convergence threshold and $\tilde{\mathfrak{T}}$ is the backward shift batch index defined as:

$$\tilde{\mathfrak{T}} \triangleq \{t-1 | t \in \mathfrak{T}\} \tag{5.23}$$

Algorithm 6 BACKWARD($\boldsymbol{w}$, $\mathfrak{T}$)

Input: Current line parameter $\boldsymbol{w}$ and the first difference instance batch index $\mathfrak{T}$.
Output: Gradient $\frac{\partial e_{\boldsymbol{w}}(\mathfrak{T})}{\partial \boldsymbol{w}}$.

```
[x(t)]=FORWARD(w, t), t ∈ 𝔗 ∪ 𝔗̃.
Construct [x̂(t)] as (5.20), t ∈ 𝔗.
Calculate [õ(t)] = Ĝ([x̂(t)]), Â(t) = ∂F̂_w([x̂(t)],[l̂(t)]) / ∂[x̂(t)], b̂(t) = ∂e_w(t)/∂[õ(t)] · ∂Ĝ([x̂(t)])/∂[x̂(t)], for t ∈ 𝔗.
for t ∈ 𝔗 do
    Initialize z(t)^0 = 0_{1×12N}, τ = 0.
    repeat
        z(t)^{τ+1} = z(t)^τ · Â(t) + b̂(t)
        τ = τ + 1
    until ‖z(t)^τ − z(t)^{τ−1}‖² < ε_backward · ‖z(t)^{τ−1}‖²
    ∂e_w(t)/∂w = z(t)^τ · ∂F̂_w([x̂(t)],[l̂(t)])/∂w, for t ∈ 𝔗.
end for
∂e_w(𝔗)/∂w = (1/|𝔗|) Σ_{t∈𝔗} ∂e_w(t)/∂w
return ∂e_w(𝔗)/∂w
```

5.5.4 Utilization of Prior Distribution of Line Parameters Through MAP and Constraints

Electric utilities often have reasonable estimates of distribution systems' line impedance in GIS, which serve as key statistics for the prior distributions of the line parameters. This subsection describes how to use these information to improve estimates of line parameters using maximum a posteriori probability (MAP) and parameter constraints.

5.5.4.1 Use of Prior Line Parameter Distribution in MAP Estimate

The posterior distribution of the line parameters is:

$$P(\boldsymbol{w} \mid [\tilde{v}(t)]_{t=1}^{T}) = \frac{P([\tilde{v}(t)]_{t=1}^{T} \mid \boldsymbol{w})P(\boldsymbol{w})}{P([\tilde{v}(t)]_{t=1}^{T})} \tag{5.24}$$

Here $[\tilde{v}(t)]$ represents a stack of $\tilde{v}_m(t)$, $(m = 1, ..., M)$, and $[\tilde{v}(t)]_{t=1}^{T}$ represents $[\tilde{v}(t)]$ of $t = 1, ..., M$, i.e., the observed first difference voltage time series over the entire time period. Maximizing (5.24) is equivalent to the minimization in (5.25):

$$\min_{\boldsymbol{w}} -\log P([\tilde{v}(t)]_{t=1}^{T} \mid \boldsymbol{w}) - \log P(\boldsymbol{w}) \tag{5.25}$$

We assume $\tilde{v}_m(t) \sim N(\tilde{o}_m(t), \sigma_{v_m}^2)$ and are independent across smart meters $m = 1, ..., M$ and time steps $t = 1, ..., T$. We also assume a Gaussian prior of the line parameters $w_i \sim N(\mu_i, \sigma_{w_i}^2)$, $i = 1, ..., |\boldsymbol{w}|$. $\tilde{o}_m(t)$ is the output of the graphical learning model with parameter $\boldsymbol{w}$, i.e., the theoretical $\tilde{v}_m(t)$ with parameter $\boldsymbol{w}$. For simplification, we further assume $\sigma_{v_m} \approx \sigma_v$, $\forall m$, so that (5.25) can be approximated by:

$$\min_{\boldsymbol{w}} \sum_{t=1}^{T} \sum_{m=1}^{M} \frac{(\tilde{v}_m(t) - \tilde{o}_m(t))^2}{\sigma_v^2} + \sum_{i=1}^{|\boldsymbol{w}|} \frac{(w_i - \mu_i)^2}{\sigma_{w_i}^2} \tag{5.26}$$

By scaling (5.26), we have:

$$\begin{aligned} &\min_{\boldsymbol{w}} \frac{1}{TM} \sum_{t=1}^{T} \sum_{m=1}^{M} (\tilde{v}_m(t) - \tilde{o}_m(t))^2 + \frac{\sigma_v^2}{TM} \sum_{i=1}^{|\boldsymbol{w}|} \frac{(w_i - \mu_i)^2}{\sigma_{w_i}^2} \\ &= \min_{\boldsymbol{w}} e_{\boldsymbol{w}}(\mathfrak{T}_{\text{full}}) + R(\boldsymbol{w}) \end{aligned} \tag{5.27}$$

where $R(\boldsymbol{w}) \triangleq \frac{\sigma_v^2}{TM} \sum_{i=1}^{|\boldsymbol{w}|} \frac{(w_i - \mu_i)^2}{\sigma_{w_i}^2}$. The prior distribution of line parameters specifies, μ_i and $\sigma_{w_i}^2$. The only unknown term in $R(\boldsymbol{w})$ is σ_v^2, which needs to be estimated. With the Gaussian assumption $\tilde{v}_m(t) \sim N(\tilde{o}_m(t), \sigma_v^2)$, σ_v^2 can be estimated from data samples by:

$$\sigma_v^2 \approx \frac{1}{M(T-1)} \sum_{t=1}^{T} \sum_{m=1}^{M} (\tilde{v}_m(t) - \tilde{o}_m(t))^2 = \frac{T}{T-1} e_{\boldsymbol{w}}(\mathfrak{T}_{\text{full}}) \tag{5.28}$$

The approximation in (5.28) holds when $\boldsymbol{w}$ is close to the true parameter value. The MAP estimation of line parameters consists of two steps. First, we estimate $\boldsymbol{w}$ by minimizing $e_{\boldsymbol{w}}(\mathfrak{T}_{\text{full}})$ without prior knowledge and calculate σ_v^2 with (5.28). Second, we obtain the MAP estimate with (5.27).

Since we work with both the entire dataset and mini-batches, we define the loss function over a data batch $\mathfrak{T}$ as:

$$J_{\boldsymbol{w}}(\mathfrak{T}) = e_{\boldsymbol{w}}(\mathfrak{T}) + \gamma R(\boldsymbol{w}) \tag{5.29}$$

where γ is the regularization factor that controls the weight of prior. (5.24)-(5.27) corresponds to MAP with $\gamma = 1$. Note that $R(\boldsymbol{w})$ does not depend on $|\mathfrak{T}|$, because $R(\boldsymbol{w})$ is defined on the full batch size $T = |\mathfrak{T}_{\text{full}}|$. This definition ensures that when $\mathfrak{T}_{\text{full}}$ is split into mini-batches, the average $J_{\boldsymbol{w}}(\mathfrak{T})$ over all mini-batches equals $J_{\boldsymbol{w}}(\mathfrak{T}_{\text{full}})$.

The gradient of $R(\boldsymbol{w})$ can be calculated as follows:

$$\frac{\partial R(\boldsymbol{w})}{\partial w_i} = \frac{2\sigma_v^2(w_i - \mu_i)}{TM\sigma_{w_i}^2}, \quad i = 1, ..., |\boldsymbol{w}| \tag{5.30}$$

5.5.4.2 Constraints on Line Parameter Estimates

We can also add constraints to the line parameter estimates if we know their upper and lower limits. Assume that we know $w_{\text{min},i} \leq w_i \leq w_{\text{max},i}$, $i, = 1, ..., |\boldsymbol{w}|$. Then we can apply projected gradient descent to ensure that the learned parameters from the SGD-based estimation procedure stays within the allowable range. Here we denote the projection as $\boldsymbol{w}_{\text{proj}} = \text{CONS}(\boldsymbol{w}, \boldsymbol{w}_{\text{min}}, \boldsymbol{w}_{\text{max}})$, in which $w_{\text{proj},i} = \min(w_{\text{max},i}, \max(w_i, w_{\text{min},i}))$ for $i, = 1, ..., |\boldsymbol{w}|$.

5.5.5 SGD-Based Line Parameter Estimation Algorithm

Our proposed SGD-based line parameters estimation method is summarized in Algorithm 7. In step 1, the parameter set $\boldsymbol{w}_{\text{iter}}$ is initialized with its original value in the GIS. The initial values for the parameters are assumed to be not far from the correct ones. In steps 2 to 20, we iteratively update $\boldsymbol{w}_{\text{iter}}$ by descending $J_{\boldsymbol{w}_{\text{iter}}}(\mathfrak{T}_{\text{batch}})$'s gradient over a small group of samples (i.e., a mini-batch) of size n_{batch}. We use patience n_{patience} to decide when to stop the iterative update process. That is to say, the algorithm will be stopped if J_{best} is not improved in n_{patience} epochs (an epoch goes through all T samples in mini-batches). Steps 5 to 15 show the procedure of updating $\boldsymbol{w}_{\text{iter}}$ over each mini-batch, in which we use the backtracking line search of parameters s_{initial}, α, and β to determine the step size in each move. In step 21, the parameters $\boldsymbol{w}_{\text{best}}$, which has the lowest loss value $J_{\boldsymbol{w}_{\text{best}}}(\mathfrak{T}_{\text{full}})$ is selected as the output. The use of prior distribution of the distribution line parameters is controlled by μ_i, $\sigma_{w_i}^2$, $i = 1, ..., |\boldsymbol{w}|$, γ, $\boldsymbol{w}_{\text{min}}$, and $\boldsymbol{w}_{\text{max}}$.

5.5.6 Distributed Parameter Estimation With Network Partition

For large-scale networks, the FORWARD function takes a larger number iterations to converge and is thus more time consuming. To solve this problem, we propose a network partitioning method to enable parallel computing over smaller sub-networks. The proposed approach works as follows. First, we identify a few edges of the network, which partition the network into sub-networks with similar sizes. Second, for each selected edge, one end of it is used as a quasi-source. The quasi-source's three-phase power injection, voltage magnitude of each phase, and the voltage angle difference between phases are measured. Now, each sub-network contains at least one quasi-source node or substation. Third, each sub-network is treated as an independent network

Algorithm 7 SGD-Based Line Parameter Estimation

Input: First difference of smart meter voltage magnitude $[\tilde{v}(t)]$ and three-phase nodal power injection $[\tilde{\boldsymbol{l}}(t)]$, $t \in \mathfrak{T}_{\text{full}}$; prior distribution information μ_i, $\sigma_{w_i}^2$ of line parameters, $i = 1, ..., |\boldsymbol{w}|$, regularization factor γ, parameter constraints $\boldsymbol{w}_{\min}$, $\boldsymbol{w}_{\max}$; hyperparameters n_{batch}, n_{patience}, s_{initial}, α, β and ϵ_{stop}; an initial estimate $\boldsymbol{w}_{\text{initial}}$ of $\boldsymbol{w}$ for the $12\mathfrak{L}$ line parameters.

Output: Updated estimate of $\boldsymbol{w}$.

Initialize $\boldsymbol{w}_{\text{iter}} = \boldsymbol{w}_{\text{best}} = \boldsymbol{w}_{\text{initial}}$ and $J_{\text{best}} = J_{\boldsymbol{w}_{\text{best}}}(\mathfrak{T}_{\text{full}})$ as (5.29). $n_{\text{epoch}} = 0$. $J_{\text{history}}(n_{\text{epoch}}) = J_{\text{best}}$.
repeat
  $n_{\text{epoch}} = n_{\text{epoch}} + 1$
  Randomly split $\mathfrak{T}_{\text{full}}$ into mini-batches of size n_{batch}.
  for each mini-batch $\mathfrak{T}_{\text{batch}}$ **do**
    Calculate $\frac{\partial R(\boldsymbol{w}_{\text{iter}})}{\partial \boldsymbol{w}_{\text{iter}}}$ as (5.30).
    $\nabla J_{\boldsymbol{w}_{\text{iter}}} = \text{BACKWARD}(\boldsymbol{w}_{\text{iter}}, \mathfrak{T}_{\text{batch}}) + \gamma \frac{\partial R(\boldsymbol{w}_{\text{iter}})}{\partial \boldsymbol{w}_{\text{iter}}}$
    Set $s = s_{\text{initial}}$ and $\Delta \boldsymbol{w} = -\nabla J_{\boldsymbol{w}_{\text{iter}}}$.
    $\boldsymbol{w}_{\text{temp}} = \text{CONS}(\boldsymbol{w}_{\text{iter}} + s\Delta \boldsymbol{w}, \boldsymbol{w}_{\min}, \boldsymbol{w}_{\max})$
    while $J_{\boldsymbol{w}_{\text{temp}}}(\mathfrak{T}_{\text{batch}}) > J_{\boldsymbol{w}_{\text{iter}}}(\mathfrak{T}_{\text{batch}}) + \alpha s \nabla J_{\boldsymbol{w}_{\text{iter}}}^T \Delta \boldsymbol{w}$ **do**
      $s = \beta s$
      $\boldsymbol{w}_{\text{temp}} = \text{CONS}(\boldsymbol{w}_{\text{iter}} + s\Delta \boldsymbol{w}, \boldsymbol{w}_{\min}, \boldsymbol{w}_{\max})$
    end while
    $\boldsymbol{w}_{\text{iter}} = \boldsymbol{w}_{\text{temp}}$
  end for
  if $J_{\boldsymbol{w}_{\text{iter}}}(\mathfrak{T}_{\text{full}}) < J_{\text{best}}$ **then**
    $J_{\text{best}} = J_{\boldsymbol{w}_{\text{iter}}}(\mathfrak{T}_{\text{full}})$, $\boldsymbol{w}_{\text{best}} = \boldsymbol{w}_{\text{iter}}$.
  end if
  $J_{\text{history}}(n_{\text{epoch}}) = J_{\text{best}}$
until $1 - \frac{J_{\text{history}}(n_{\text{epoch}})}{J_{\text{history}}(n_{\text{epoch}} - n_{\text{patience}})} < \epsilon_{stop}$
return $\boldsymbol{w}_{\text{best}}$.

and one quasi-source node or substation is selected as the source node; the other quasi-source nodes or substations in this sub-network are treated as ordinary nodes with three additional single-phase pseudo-loads in phase A, B, and C respectively, whose voltage and power injections are measured. Fourth, we execute Algorithm 7 for all sub-network in parallel.

We can take the IEEE 37-bus test feeder shown in Fig. 5.3 as an example of the network partition method. The feeder is partitioned into three sub-networks with similar size by edge 702-703 and 708-733. Node 702 and 708 are used as quasi-sources. Sub-network 1's source node is

799, and node 702 has 3 additional pseudo loads. Sub-network 2's source node is 702, and node 708 has 3 additional pseudo loads. Sub-network 3's source node is 708. Since sub-network 3 has no other quasi-source nodes, it does not contain any pseudo loads.

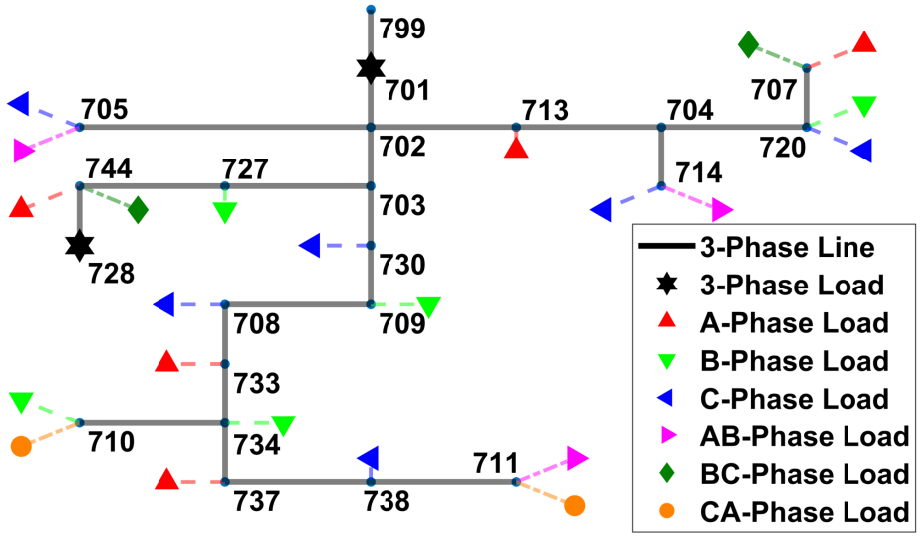


Figure 5.3: Schematic of the modified IEEE 37-bus test feeder.

5.6 Numerical Study

5.6.1 Setup for Numerical Tests

We evaluate the performance of our proposed graphical learning-based parameter estimation algorithm and a few state-of-the-art algorithms on the modified IEEE 13-bus and 37-bus test feeders. We modify these two test feeders by introducing loads with all 7 types of phase connections, AN, BN, CN, AB, BC, CA, and ABC. The basic information of the two modified IEEE test feeders are shown in Table 5.1. The modified 37-bus test feeder is shown in Fig. 5.3 and the

modified 13-bus feeder is described in [33].

Table 5.1: The Basic Information of the IEEE Test Feeders

Feeder	No. of Loads	No. of Edges	Peak Loads	Level of Unbalance
13-bus	10	6	3 MW	0.0376
37-bus	25	21	2.4 MW	0.0270

The hourly real power consumptions on the test feeders are calculated based on the real power consumption time series from the smart meters of a real-world distribution feeder in North America. The length of the real power consumption time series is 2160, which corresponds to 90 days of measurements. The reactive power time series are calculated by assuming a lagging power factor, which follows a uniform distribution $\mathcal{U}(0.9, 1)$. The peak loads of the 13-bus and 37-bus test feeders are 3MW and 2.4MW respectively. The nodal voltages are calculated by power flow analysis using OpenDSS. To simulate the smart meter measurement noise, we use a zero-mean Gaussian distribution with three standard deviation matching 0.1% to 0.2% of the nominal values. The 0.1 and 0.2 accuracy class smart meters established in ANSI C12.20-2015 represent the typical noise levels in real-world advanced metering infrastructure. We assume that the initial estimates for the distribution line parameters, $\boldsymbol{w}_{\text{initial}}$, are randomly sampled from a uniform distribution within $\pm 50\%$ of the correct values.

When generating simulated time series data, the power consumptions are allocated relatively evenly to each phase so that the test feeders are close to balance. Following [31], the level of unbalance of a feeder at time interval t can be measured as

$$u(t) = \frac{|I_A(t) - I_m(t)| + |I_B(t) - I_m(t)| + |I_C(t) - I_m(t)|}{3I_m(t)} \tag{5.31}$$

where $I_m(t) = \frac{1}{3}(I_A(t) + I_B(t) + I_C(t))$ is the mean of the distribution substation line current

magnitudes of the three phases at time interval t. We use the 90-day average of $u(t)$ to measure the level of unbalance of the test feeders, which are shown in Table 5.1.

The hyperparameters for SGD of the proposed graphical learning model is set up as follows. $n_{\text{batch}} = 10$, $n_{\text{patience}} = 10$, $s_{\text{initial}} = 1000$, $\alpha = 0.3$, $\beta = 0.5$, and $\epsilon_{stop} = 0.01$. The $\epsilon_{\text{forward}}$ in the FORWARD function and $\epsilon_{\text{backward}}$ in the BACKWARD function are set to be $1e-20$. These values are set empirically so that the algorithm updates $J_{\boldsymbol{w}_{\text{iter}}}(\mathfrak{T})$ adequately and stops when it saturates.

The setup corresponding to the prior distribution component of the proposed algorithm is set up as follows. $\boldsymbol{w}_{\text{initial}}$ is assumed to be within $\pm 50\%$ of the correct values. Thus, the lower and upper bounds of the parameter w_i are selected to be $\frac{w_{\text{initial},i}}{1+50\%} = \frac{2}{3} w_{\text{initial},i}$ and $\frac{w_{\text{initial},i}}{1-50\%} = 2 w_{\text{initial},i}$, where $w_{\text{initial},i}$ is the ith element in $\boldsymbol{w}_{\text{initial}}$. For the MAP estimation of each parameter w_i, we set $\mu_i = w_{\text{initial},i}$ and $\sigma_{w_i} = w_{\text{initial},i} \times 50\% \times \frac{1}{3}$, which represents a Gaussian distribution centered at $w_{\text{initial},i}$ and its three standard deviation matching $\pm 50\%$ of $w_{\text{initial},i}$. Though this Gaussian assumption is different from the actual uniform distribution of $\boldsymbol{w}_{\text{initial}}$, simulation results show the MAP is still effective.

The proposed graphical learning model uses SGD to update line parameter estimates. To reliably evaluate the performance of the proposed model, we execute the algorithm multiple times with different random seeds and calculate the average performance. The numerical tests are implemented using MATLAB on a DELL workstation with two 3.0 GHz Intel Xeon 8-core CPUs and 192 GB RAM.

5.6.2 Performance Measurement

We use the mean absolute deviation ratio (MADR) to measure the estimation error of distribution line parameters. The MADR between the estimated $\boldsymbol{w}$ and the correct value $\boldsymbol{w}^{\dagger}$ is

defined as in (5.32). The performance of a distribution line parameter estimation algorithm is evaluated by the percentage of MADR improvement, which is defined as in (5.33). Here $\text{MADR}_{initial}$ and MADR_{final} represent the MADR of the initial and the final line parameter estimates. The maximum possible MADR improvement is 100%, which corresponds to a perfect estimation (i.e., $\text{MADR}_{final} = 0\%$).

$$\text{MADR} \triangleq \sum_{i=1}^{12\mathfrak{L}} |w_i - w_i^{\dagger}| \div \sum_{i=1}^{12\mathfrak{L}} |w_i^{\dagger}| \times 100\% \tag{5.32}$$

$$\text{MADR improvement} \triangleq \frac{\text{MADR}_{initial} - \text{MADR}_{final}}{\text{MADR}_{initial}} \times 100\% \tag{5.33}$$

5.6.3 Performance Comparison of the Proposed Graphical Learning Method and State-of-the-Art Algorithms

The performance of our proposed graphical learning algorithm (GL) with MAP and parameter constraints (abbreviated as CON) is compared with the state-of-the-art algorithm, linearized power flow model based maximum likelihood estimation (LMLE) [33]. In addition, we perform an ablation study to evaluate the relative importance of the MAP and parameter constraints modules in our proposed graphical learning model. These methods are tested with three smart meter accuracy class: noiseless (0%), 0.1%, and 0.2%. Due to the randomness of the SGD component of the proposed and comparison algorithms, the combination of each algorithm and smart meter class are tested 20 times with different random seeds. The average MADR improvement of the proposed and comparison algorithms are reported in Table 5.2.

From Table 5.2, we can see that the MADR improvement of the GL algorithm is significantly higher than that of LMLE in both test feeders under all meter classes. The increase in

Table 5.2: Average MADR Improvement of Parameter Estimation Methods

Feeder	Meter Class	LMLE	GL	GL+ CON	GL+ MAP	GL+ CON&MAP
13-bus	0%	59.9%	69.2%	74.7%	69.5%	75.0%
	0.1%	59.3%	68.0%	70.0%	70.3%	73.4%
	0.2%	56.4%	64.7%	66.4%	68.0%	70.2%
37-bus	0%	35.8%	40.5%	41.7%	40.5%	41.7%
	0.1%	17.0%	22.0%	25.2%	25.4%	25.6%
	0.2%	-10.9%	10.7%	18.7%	20.3%	20.9%

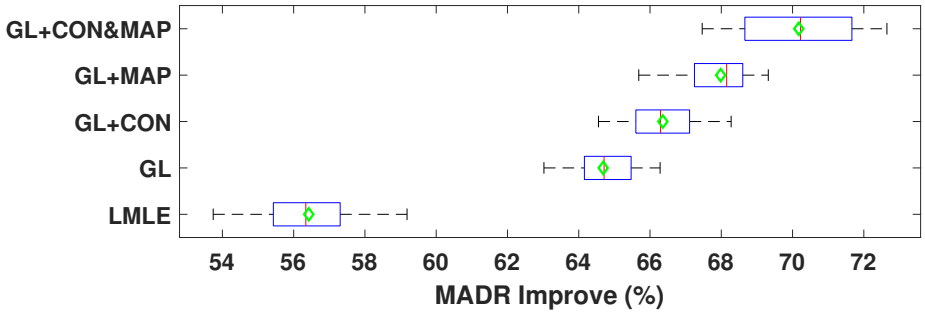


Figure 5.4: Box plot of 20 random tests for each different algorithms in the 13-bus test feeder, 0.2% noise level.

MADR improvement ranges from 8.3% to 9.3% in the 13-bus feeder and 4.7% to 21.6% in the 37-bus feeder. The estimation accuracy of both GL and LMLE increases as the meter noise level decreases. In the 37-bus feeder under 0.2% meter class, the LMLE has negative MADR improvement, which means the LMLE fails to obtain a more accurate parameter estimation from the initial parameters. On the other hand, the GL algorithm still obtains a more accurate parameter estimation under the same condition. These results show that by preserving the nonlinearity of three-phase power flows, the GL algorithm is significantly more accurate than the LMLE.

In addition to the advantage of GL algorithm, Table 5.2 shows the benefit of CON and MAP. Compared with GL algorithm, using only CON has a higher MADR improvement by 1.7% to 5.5% in the 13-bus feeder, and 1.2% to 8% in the 37-bus feeder. Compared with GL algorithm,

using only MAP has a higher MADR improvement by 0.3% to 3.3% in the 13-bus feeder, and 0% to 9.6% in the 37-bus feeder. The GL algorithm using both CON and MAP has the highest MADR improvement, which is higher than LMLE by 13.8% to 15.1% in the 13-bus feeder, and 5.9% to 31.8% in the 37-bus feeder. The box plot of Fig. 5.4 compares the accuracy of different algorithms in the 13-bus test feeder, 0.2% meter class. These results show that both MAP and CON are effective in utilizing the prior distribution of line parameters to further improve the parameter estimation accuracy.

The estimation accuracy of all algorithms increases as the meter noise level decreases, with one exception. In Table 5.2, we note that under meter class 0%, the MAP's improvement over the GL algorithm is not as significant as 1% and 2% meter classes. This is because under the noiseless 0% meter class, the σ_v^2 for MAP is much smaller than 1% and 2% meter classes. The smaller σ_v^2 put less weight on $R(\boldsymbol{w})$ in (5.27) and thus the MAP is less effective under the 0% meter class.

In Table 5.2, we also note that the overall accuracy of 13-bus feeder is higher than 37-bus. This is because the 37-bus feeder has lower meter number to line number ratio and longer average node-to-node distances (in terms of number of line segments). The material of line segments, configurations, and load profiles are also different between the two feeders.

5.6.4 Performance on Unbalanced Distribution Feeders

We test our proposed method with higher unbalance levels by adjusting the load levels in each phase of the test feeders. The result shows that our proposed method is very accurate even if the feeder is severely unbalanced. Table 5.3 shows the average MADR improvement of different parameter estimation methods when the feeder's unbalance level is 0.1, which is deemed

as severely unbalanced. From Table 5.3, We can draw similar conclusions as in Table 5.2. The GL algorithm and its combination with CON and MAP significantly outperform LMLE. Compared with the LMLE, the GL algorithm has a higher MADR improvement by 8.4% to 8.7% in the 13-bus feeder, and 4.5% to 19.7% in the 37-bus feeder. The GL+CON&MAP has the most accurate estimation result. Its MADR improvement is higher than LMLE by 14.7% to 16.1% in the 13-bus feeder, and 6.6% to 29.5% in the 37-bus feeder.

Table 5.3: Average MADR Improvement of Parameter Estimation Methods in Highly Unbalanced Feeders(Unbalance Level=0.1)

Feeder	Meter Class	LMLE	GL	GL+ CON	GL+ MAP	GL+ CON&MAP
13-bus	0%	58.2%	66.6%	73.1%	67.7%	73.8%
	0.1%	57.6%	66.3%	68.8%	69.6%	72.3%
	0.2%	54.2%	62.9%	65.2%	67.4%	70.3%
37-bus	0%	35.0%	40.4%	41.6%	40.3%	41.6%
	0.1%	17.5%	22.0%	25.3%	25.5%	25.7%
	0.2%	-8.4%	11.3%	19.2%	20.7%	21.1%

5.7 Conclusion

In this chapter, we develop a physics-informed graphical learning algorithm to estimate line parameters of three-phase power distribution networks. Our proposed algorithm is broadly applicable as it uses only readily available smart meter data to estimate the three-phase series resistance and reactance of the primary line segments. We leverage the domain knowledge of power distribution systems by replacing the deep neural network-based transition functions in the graph neural network with three-phase power flow-based physical transition functions. A rigorous derivation of the gradient of the loss function for first difference voltage time series with respect to line parameters is provided. The network parameters are estimated through iterative application of stochastic gradi-

ent descent. The prior distribution of the line parameters is also considered to further improve the accuracy of the proposed parameter estimation algorithm. Comprehensive numerical study results on IEEE test feeders show that our proposed algorithm significantly outperforms the state-of-the-art algorithm. The relative advantage of the proposed algorithm becomes more pronounced when smart meter measurement noise level is higher.

5.8 List of Symbols in Chapter 5

$E_{m\times n}^{(i,j)}$	An $m \times n$ matrix, in which the ij-th element is 1 and the rest of elements are all 0.
$Im(\cdot)$	Imaginary part of a complex variable.
M	Number of loads in a circuit.
N, $\mathfrak{L}$	Number of non-substation nodes and lines in the three-phase primary distribution network.
$Re(\cdot)$	Real part of a complex variable.
$\mathfrak{T}$	A batch of time difference instances.
co(n), ne(n), no(m)	Edge set connected to node n, node neighbor set of node n, node that meter m connects to.
p_n^i, q_n^i, α_n^i, β_n^i	Real and imaginary part of power injection, real and imaginary part of node n's voltage at phase i.
$\boldsymbol{u}$, $\boldsymbol{s}$	3×1 Vector of a node's three-phase complex voltage and complex power injection.
v	Voltage magnitude scalar of a smart meter.

$\tilde{v}$, $\tilde{o}$	Time difference of v and o.
$\boldsymbol{w}$	The set of parameters of the GNN/distribution system.
$\boldsymbol{x}$, $\boldsymbol{l}$, o	State vector, feature vector, and output scalar of an element in the GNN.
$[\boldsymbol{x}]$, $[\boldsymbol{l}]$, $[o]$	A vector that stacks all the states, features, and output.
$\hat{\boldsymbol{x}}_n(t)$, $\hat{\boldsymbol{l}}_n(t)$	12×1 vectors that stack $\boldsymbol{x}_n(t-1)$ and $\boldsymbol{x}_n(t)$, $\boldsymbol{l}_n(t-1)$ and $\boldsymbol{l}_n(t)$, respectively.
$[\hat{\boldsymbol{x}}(t)]$, $[\hat{\boldsymbol{l}}(t)]$	Vectors of stacking $[\boldsymbol{x}(t-1)]$ and $[\boldsymbol{x}(t)]$, $[\boldsymbol{l}(t-1)]$ and $[\boldsymbol{l}(t)]$, respectively.
$(\cdot)^i$	A variable in phase i.
$(\cdot)^{ij}$	A variable between phase i and j.
$(\cdot)_n$	A variable of the n-th node, load, or element.
$\cdot(t)$	The value of a variable at time t.
$\mathbb{0}_{m\times n}$	All-0 matrix of size $m \times n$.
$\odot$, $\oslash$	Element-wise multiplication and division.

Chapter 6

Modeling of Commercial Adoption of Photovoltaic Systems

6.1 Introduction

Solar energy, including both the solar thermal and photovoltaic, grew rapidly in the U.S. from 64 trillion Btu in 2001 to 427 trillion Btu in 2014 [106]. The U.S. solar energy's share of total new electricity generation capacity increased from 10% in 2012 to 32% in 2014 [107]. In particular, the small-scale distributed solar PV system, which accounts for 33% of the total solar generation has grown significantly in the United States over the past several years [108]. As a vital component of the U.S. renewable energy portfolio, continuing adoption of solar PV systems is key to both reducing greenhouse gas emissions and building a clean energy workforce. For example, the solar industry workforce in the United States grew more than 50% in the past four years and now employs more than 140,000 workers [109].

To stimulate the growth of the solar PV market, the federal and state governments have been supporting research and strengthening U.S. solar manufacturing capabilities to drive down the installation cost of solar PV [109]. In addition, the U.S. important a significant portion of solar equipment from Asia [110, 111, 112]. The boom in global solar module production also led to a precipitous decline in solar PV module prices. In California, the median installed cost for systems of 10-100 kW-dc in size dropped 56% from $10.7/W-dc in 2001 to $4.7/W-dc in 2013 [113]. In addition, the federal and state governments developed many incentive programs to directly promote the adoption of solar PV systems. At the federal level, the Investment Tax Credit (ITC) was implemented in 2006, which provides 30% tax credit for solar systems on residential and commercial properties [114]. At the state level, California has been leading the way by implementing an array of incentive programs including the California Solar Initiative (CSI), the New Solar Homes Partnership (NSHP), the Self-Generation Incentive Program (SGIP), etc [115, 116, 117].

The drop in solar PV cost and direct government incentives have contributed to the rapid growth in the penetration of small-scale distributed solar PV systems in the energy market. However, there is a lack of rigorous analyses that quantify the impact of government incentives and solar PV costs on adoption. Such analyses will provide critical and useful feedback to government agencies to improve the design of future renewable energy incentive programs, and serve as the basis of forecasting the adoption.

In this chapter, a model for commercial solar PV adoption is developed with explanatory variables such as government incentive programs and solar PV system installation costs. The adoption model is built on top of the Generalized Bass diffusion framework. The model is applied to forecast commercial solar PV adoption in Southern California. Asymptotic standard errors of the

parameter estimates are calculated to verify the significance of the explanatory variables. Empirical results show that decreasing solar PV installation costs and government incentive programs are the main forces that drove the growth of commercial solar PV adoption. In the case of Southern California, we also discover that government incentive programs and PV installation costs have a much higher impact on large commercial customers than small commercial customers. Our Generalized Bass diffusion model of commercial solar PV adoption yields a lower root-mean-square error (RMSE) than the basic Bass Diffusion model. In addition, the commercial solar PV adoption model predicted that the eventual adoption rate of solar PV system is higher for large commercial customers.

The rest of the chapter is organized as follows. Section 6.2 reviews the literature and summarizes our work's contribution. Section 6.3 reviews the Bass Model and Generalized Bass Model. Section 6.4 presents the model for commercial PV adoption. In Section 6.5, the solar PV adoption model is fitted and validated using historical adoption data. Section 6.6 quantifies the impact of federal and state incentives on California commercial PV adoption. Section 6.7 presents the conclusions and policy implications of the work.

6.2 Literature Review and Contributions of Our Work

Several diffusion models have been developed to describe the adoption of new products and technologies. The Bass Model (BM) [27] is a simple but effective model which described the empirical adoption of a wide range of products and services. The BM was extended to the Generalized Bass Model (GBM)[28], by including decision variables such as price, marketing effort, etc. In the BM [27],the model parameters were estimated by ordinary least squares (OLS). A maximum

likelihood estimation (MLE) approach was proposed by [118], and it fits better to the observations than OLS and allows for one-step ahead forecasts. A nonlinear least squares approach was proposed by [119], and it corrects the underestimates of standard errors of the estimated parameters in the MLE approach.

Many researchers have studied the mechanics of residential solar PV adoption using various approaches. [120] and [121] analyzed multiple factors that influence residential customers' decisions on solar PV adoption. These factors include government incentives and solar PV costs. However, these analyses were conducted in the form of surveys and interviews. [122] used BM without any decision variables to model residential solar PV adoption. The model parameter estimates were not as robust as the MLE and nonlinear least squares approach. GBM was used by [123] to model adoption patterns of PV systems in many countries. In the GBM, institutional measures, policies, and government interventions were modeled as perturbations in the form of exponential shocks and rectangular shocks. The model proposed was helpful in explaining the impact of short-term interventions, such as a moment of opinion change due to social incidents. However, it did not clearly explain the effects of long-term interventions such as solar PV incentive programs and the dropping installation cost of PV systems on adoption.

Since its emergence in 2007, the third-party ownership (TPO) model has earned a significant share in the solar PV market [124]. The rapidly growing third-party PV ownership has prompted a few researchers to study its impact on residential solar PV adoptions. [125] found that the introduction of third-party PV ownership enticed a new demographic to adopt residential PV systems which increased the total demand for PV systems. The economics of buy or lease a residential PV system is studied in [126]. It is shown that the choice of contract type and payment

structure have implications for the total cost to residential solar PV customers over the lifetime of the contract [127].

To the best of our knowledge, this work is the first to study adoption of distributed commercial PV systems. The adoption of commercial PV systems include both purchasing/owning the solar PV system and deployment of solar PV system through a TPO contract.[128]. Note that the drivers and barriers to distributed commercial PV adoption are different from that of residential adoption. For example, a higher percentage of buildings are leased in the commercial sector than the residential sector [129, 130]. Therefore, the commercial PV adoption faces bigger challenges in the form of incentive splitting between the building owner and multiple-tenants. In addition, many limited liability companies (LLC) that own commercial properties have low credit ratings which make solar financing more difficult. Finally, though the total number of commercial customers are smaller than residential customers, their average solar PV size is much larger. Figure 6.1 shows the number of solar PV installations, and Figure 6.2 shows their average solar PV size each year in Southern California. From 2001 to 2014, the number of commercial installations were only about 3% of the total installations of all types, but the commercial solar PV capacity was about 40% of the total installed PV capacity.

Compared with other related work, this work makes the following unique contributions:

1. This work develops a GBM for commercial PV adoption, which quantifies the impact of solar PV costs and government incentive programs on the adoption.

2. The GBM for commercial PV adoption is also capable of forecasting the eventual commercial PV adoption rate and quantifying the delayed effect of explanatory variables on adoption.

3. The model is applied to fit the empirical commercial PV adoption data in Southern

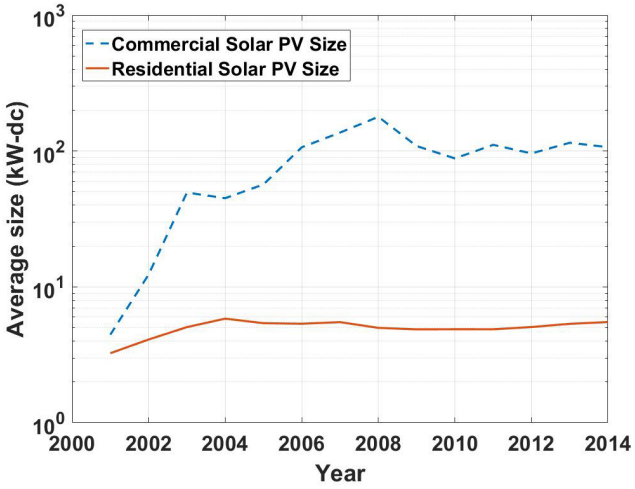


Figure 6.1: Average installed PV size by customer type in Southern California

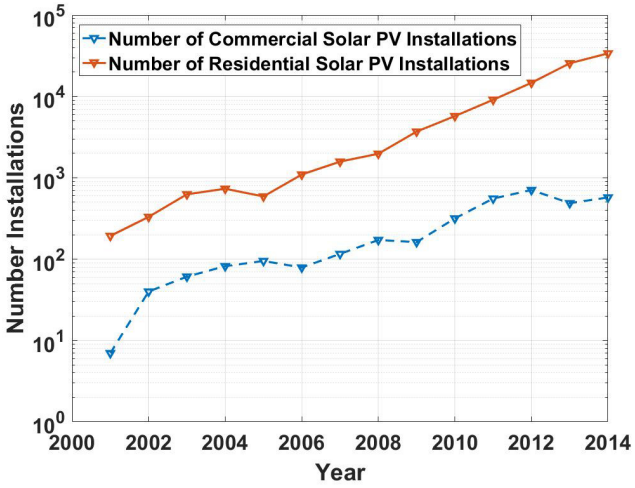


Figure 6.2: Number of solar PV installations by customer type in Southern California

California. Nonlinear least squares is applied to estimate the model parameters and their asymptotic standard errors. The empirical results show that large commercial customers are more susceptible to the influence of PV costs and government incentives than small commercial customers.

4. By changing the cost and incentive rates fed into the model, adoption curves can be forecasted under different cost and policy conditions. This can be a useful tool for the government to evaluate its renewable energy technology incentive policies.

6.3 Overview of Bass Model, Generalized Bass Model and Parameter Estimation Approaches

6.3.1 Bass Model and Generalized Model

The Bass diffusion model is a well-established model of innovation and technology adoption in the market. It can be described by the following formulation [27]:

$$\frac{f(t)}{1-F(t)} = p + qF(t) \tag{6.1}$$

$F(t)$ is the cumulative adoption function. $F(t) \to 1$ as $t \to \infty$. $f(t) = \frac{dF(t)}{dt}$ is the adoption rate. The left hand side of the function describes the conditional adoption rate at time t, and it is controlled by two factors: p and q. p is the innovation factor, describing innovative adopters who are willing to adopt the product themselves, and q is the imitation factor, describing the adopters who follow other adopters' use of the product. Both p and q are positive.

The solution of equation (6.1) is as following [27]:

$$F(t) = \frac{1-e^{-(p+q)t}}{1+\frac{q}{p}e^{-(p+q)t}} \tag{6.2}$$

$$f(t) = \frac{\frac{(p+q)^2}{p} e^{-(p+q)t}}{(1 + \frac{q}{p} e^{-(p+q)t})^2} \tag{6.3}$$

To include marketing effort factors in the diffusion model, Frank M. Bass introduced the generalized Bass model [28]. The GBM is described as follows:

$$\frac{f(t)}{1 - F(t)} = [p + qF(t)]x(t) \tag{6.4}$$

where $x(t)$ is called "current marketing effort", reflecting the influence of market factors on the adoption rate at time t. Define the cumulative marketing effort $X(t) = \int_0^t x(\tau)d\tau$ and let $X(0) = 0$. The solution of (6.4) is given by:

$$F(t) = \frac{1 - e^{-X(t)(p+q)}}{1 + \frac{q}{p} e^{-X(t)(p+q)}} \tag{6.5}$$

$$f(t) = x(t) \frac{\frac{(p+q)^2}{p} e^{-X(t)(p+q)}}{(1 + \frac{q}{p} e^{-X(t)(p+q)})^2} \tag{6.6}$$

$x(t)$ is a function of one or more decision variables. For example, product price was chosen as a decision variable by [28]. Under the argument of diminishing returns, $x(t)$ can be defined as below:

$$x(t) = 1 + [(dPr(t)/dt)/Pr(t)]\beta_0 \tag{6.7}$$

where $Pr(t)$ is the price at time t, and β_0 is a weight coefficient expected to be negative. Let $\Phi_0(t) = \ln \frac{Pr(t)}{Pr(0)}$, then

$$X(t) = t + \beta_0 \Phi_0(t) \tag{6.8}$$

6.3.2 Parameter Estimation Methods

In this subsection, we describe how to perform parameter estimations for diffusion models. Let M denote the total customer population. The eventual cumulative solar adoption denoted by m is only a portion of M. Define the eventual adoption rate c ($0 \leq c \leq 1$) such that $m = Mc$. To estimate the parameters, we use the observation of a series of historical adoption s_i, ($i = 1, 2, 3, ..., N$), which is the number of solar PV systems installed in time interval i. For example, let s_i be the number of adopted solar PV systems in month i and let $t_i = i$. Then from (6.1), the following is derived:

$$mf(t_i) = mp + (q - p)(mF(t_i)) - (q/m)(mF(t_i))^2 \tag{6.9}$$

In the work of [27, 122], $mf(t_i)$ was replaced by s_i , and $mF(t_i)$ was replaced by $\sum_{j=1}^{i} s_j$. Then OLS estimation can be applied to (6.9) to estimate parameters p, q, and m. The OLS approach has two drawbacks. First, the replacement is not precise because it uses the aggregated adoption of time intervals to replace instantaneous adoption rates. The diffusion model is time continuous, but it is estimated using discrete time series data. Second, disturbances such as noise and parameter misspecifications were not appropriately modeled. [118] showed that there is bias in such methods.

To overcome these drawbacks, nonlinear least squares estimation was proposed by [119], which also yields valid estimation of the standard errors of the estimators. The disturbance is modeled as follows:

$$s_i = m[F(t_i) - F(t_{i-1})] + u_i \tag{6.10}$$

where $F(t_i)$ can be in forms of either (6.2) or (6.5). u_i is the net disturbance of sampling errors, the impact of excluded factors, misspecified parameters, etc. Then the parameters in $x(t)$, p, q, and m

can be estimated by the nonlinear least squares (NLS) approach as in (6.11). In this work the NLS approach is used to estimate the model parameters.

$$\min \sum_{i=1}^{N} \{m[F(t_i) - F(t_{i-1})] - s_i\}^2 \tag{6.11}$$

6.4 Technical Methods

6.4.1 Choice of Decision Variables

To apply the GBM, appropriate decision variables need to be chosen first. As mentioned in Section 6.1, two explanatory variables have significant influence on PV adoption: installed PV system costs and government incentives. The installed PV system costs include the PV module price and non-module costs such as inverters, mounting hardware, labor and permitting fees, and installer profit. Government incentives include both federal level and state level incentive programs. They can be in forms of tax credits like the ITC, rebates like the CSI, etc.

Government incentives have different impacts on different groups of commercial customers. Based on PV system sizes, the adopters can be divided into 3 groups: 0-10 kW-dc, 10-100 kW-dc, and 100-1000 kW-dc. Figure 6.3 depicts the percentage of commercial customers in each group who applied and received California solar PV installation incentives on an annual basis. From 2007 to 2014, 88.27% of commercial customers in the 100-1000 kW group received state level incentives, whereas 63.22% of customers in the 10-100 kW group received incentives; only 34.20% of customers in the 0-10 kW group received incentives. Though no information is available on the application rate of the ITC, it can be inferred that the ITC may also have a higher application rate in the 100-1000 kW range than the other two ranges. A decision variable should describe a factor that impacts most of the PV adopters. If there is a large portion of adopters who do not apply for incen-

tive programs, then the incentive may not be a good decision variable. In this case, the incentive can be a suitable decision variable for the 100-1000 kW group, probably a suitable decision variable for the 10-100 kW-dc group, and not a good decision variable for the 0-10 kW-dc group. Therefore, in the 0-10 kW-dc group, we use only the cost as the decision variable, while in the 10-100 kW-dc group and the 100-1000 kW-dc group we use both incentives and the cost. We can verify if the incentive is a suitable decision variable by checking the significance level of parameter estimates.

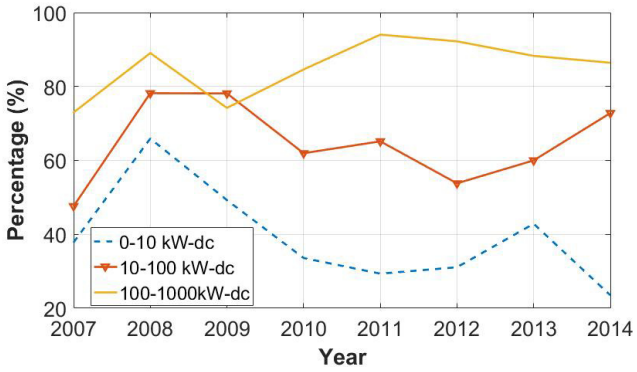


Figure 6.3: Percentage of commercial PV installations that benefited from the CSI program in 3 size ranges.

One may propose that savings in electricity bills are also important factors. This is true, because customers with solar PV systems consume less electricity from utility companies. This factor is also related to government policy. For example, the Net Energy Metering program(NEM) [131] is supported by the government in California, and the program allows customers with solar PV to pay their utility bills based on the net energy consumption from the grid. If the solar PV generates more energy than a customer's consumption, the energy can be sold to utility companies at the retail price. However, electricity bill savings are not considered in our model, because savings are difficult to estimate for a large group of potential commercial customers. The estimation of

savings is difficult for two reasons. First, the tariff rates of utility companies are hard to forecast. There are many different types of tariffs for commercial customers, and they change frequently. To make the issue more complicated, the growth of solar PV adoption can affect the tariff in turn [132]. The second reason is that it is difficult to forecast each customer's electricity usage pattern in the long term. The electricity usage patterns of commercial customers depend on a variety of factors such as their locations, business, and building types.

Overall, in the solar PV adoption model, the installation cost is one general decision variable for all customer groups, and the incentive is another decision variable for the customer groups that have high incentive adoption rate.

6.4.2 Commercial Solar PV Adoption Model with Costs and Incentive Rates as Decision Variables

In this subsection, we address the question of how to map incentives and costs to $x(t)$ in the GBM. When the installation cost is the only decision variable, $x(t)$ and $X(t)$ can be defined as (6.7) and (6.8). When both the incentive and the cost are decision variables, they are mapped to $x(t)$ in the following way. Government incentive programs have varying rates, which can be transformed into monetary savings. Then the monetary saving is combined with the solar PV cost. Let $Pr(t)$ be the solar PV cost at time t, and $INC(t)$ be the money saved from incentive programs at time t. Both $Pr(t)$ and $INC(t)$ have the unit of \$/W-dc. Then we define the net expense $E(t)$ of installing solar PV as follows:

$$E(t) = Pr(t) - INC(t) \tag{6.12}$$

Similar to (6.7), with the law of diminishing returns, $x(t)$ is derived as follows:

$$x(t) = 1 + [(dE(t)/dt)/E(t)]\beta_0 \tag{6.13}$$

where β_0 is a weight coefficient expected to be negative. Let $\Phi_0(t) = \ln \frac{E(t)}{E(0)}$, then

$$X(t) = t + \beta_0 \Phi_0(t) \tag{6.14}$$

From (6.5), the cumulative solar PV adoption function at time t is:

$$F(t) = \frac{1 - e^{-[t+\beta_0\Phi_0(t-d)](p+q)}}{1 + \frac{q}{p}e^{-[t+\beta_0\Phi_0(t-d)](p+q)}} \tag{6.15}$$

In (6.15) a variable d is introduced to represent the effect of time delay. d represents the time lag between the time when new incentive program or pricing information is made available and the time when a PV system is installed using the new information. d includes the decision making time for commercial customers, the time taken to apply for incentive programs and acquire necessary permits, time for installation, etc. When fitting the GBM, M is known, so c can be estimated directly. The NLS parameter estimation problem (6.11) can be reformulated as:

$$\min \sum_{i=1}^{N} \{Mc[F(t_i) - F(t_{i-1})] - s_i\}^2 \tag{6.16}$$

M and s_i are exogenous variables. p, q, β_0, d, and c are model parameters that need to be estimated by solving NLS problem. For simplicity in estimation, define $b = p + q$ and $a = q/p$. Then the parameters to be estimated are c, d, a, b, and β_0.

6.5 Case Study

In this section, a case study is conducted to validate the usefulness of the proposed GBM in modeling the commercial solar PV adoption. The historical PV adoption data, from 2001 to

2014 in the service territory of Southern California Edison [133], is used in the empirical study. For California solar PV adopters, the major incentive programs considered include the CSI at the state level and the ITC at the federal level. The impact of these two incentive programs and the PV cost is analyzed in this section. The GBM parameter fitting results will be presented and compared with the BM.

6.5.1 Description of Datasets

The raw datasets include four parts: 1) commercial customers' aggregated electrical energy usage; 2) solar PV adopters' information; 3) historical installed PV cost; and 4) historical information about incentive program applications. The details of the datasets are as follows:

1) There are about 676000 commercial electric customers in SCE's service territory. These commercial customers are divided into various groups based on building/business types derived from the North American Industry Classification system. In each building type, customers are further divided into several subcategories based on their annual electricity usage. The average annual usage for customers in each subcategory is recorded.

2) Information on 3000 commercial solar PV adopters was gathered, including their adoption dates, PV system sizes, annual electricity usages, and building types from 2001 to 2014. The 3000 commercial solar PV adoptions include both purchasing/owning the solar PV and deployment of solar PV system through TPO contracts.

3) The median installed PV cost in California from 1998 to 2013 [113] of three PV size ranges: 0-10 kW-dc, 10-100 kW-dc, and 100-1000 kW-dc.

4) Commercial solar PV incentive application data includes incentive application dates, proposed solar PV system sizes, and rebates from the CSI program [134].

6.5.2 Preprocess of the Datasets

6.5.2.1 Forecast the Commercial Solar PV Size of Potential Adopters

As mentioned in Section 6.4.1, incentive programs have different impacts on different size groups. Three Generalized Bass Models will be developed for customers with an estimated PV system size of 0-10 kW-dc, 10-100 kW-dc, and 100-1000 kW-dc respectively. The entire commercial customer population needs to be divided based on their forecasted PV system size, so that we can obtain M for each size group. We assume that customers of the same building type have similar preference in choosing the size of a solar PV system. Intuitively, customers who use more electricity are likely to adopt larger PV systems. However, it is difficult to find a deterministic relationship between the PV size and the annual electricity usage. For example, in Figure 6.4, there is a large variation in the PV system size for customers who have very similar annual electricity consumptions.

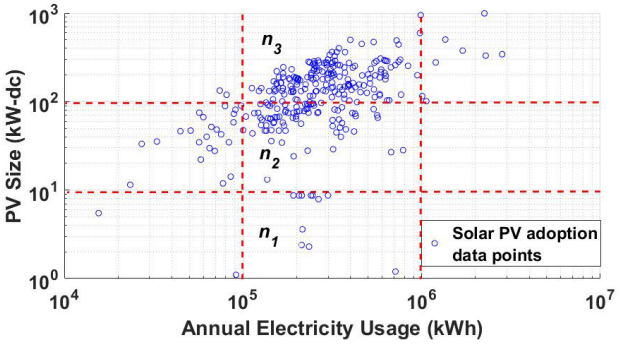


Figure 6.4: Adoption records of 297 customers in the group "education-primary," divided into different PV size groups and annual electricity usage groups.

Nevertheless, given a potential customer's annual electricity usage and building type, we can still estimate the probability that the customer will install a solar PV system in a certain size

range. For example, in Figure 6.4, for customers with an annual usage of 10^5-10^6 kWh, there are n_1, n_2, and n_3 adoption records in the three size groups. Then we can estimate that for any "education-primary" customer with an annual usage of 10^5-10^6 kWh, the probabilities to adopt a solar PV system of size 0-10 kW-dc, 10-100 kW-dc, and 100-1000 kW-dc are $n_1/(n_1+n_2+n_3)$, $n_2/(n_1+n_2+n_3)$, and $n_3/(n_1+n_2+n_3)$ correspondingly. The procedure to estimate these probabilities is carried out for each building type, and the detailed process is as follows:

Step 1: Divide the adoption records into four groups by their annual electricity usage: $<10^4$ kWh, 10^4-10^5 kWh, 10^5-10^6 kWh, and $\geq 10^6$ kWh. In each electricity usage group, further divide the records into three groups by the solar PV sizes: 0-10 kW-dc, 10-100 kW-dc, and 100-1000 kW-dc.

Step 2: In each electricity usage group, calculate probabilities that a customer will adopt a solar PV system in the three size groups.

Step 3: In each size group of each electricity usage group, calculate the average solar PV size. This is used to forecast the average solar PV size of potential adopters. If a building type has too few adoption records, we merge it with other similar building types.

In each building type, by multiplying the probability by the customer population in dataset 1 of the same electricity usage range, we can get the number of potential adopters in each size group. Assuming that customers' electricity usage increases by 2% annually, we can repeatedly estimate M each year, and we replace the constant customer population M in (6.16) with $M(t)$. Figure 6.5 shows the estimated total customer population over time for each PV system size range. $M(t)$ changes slightly over time, since we assume customers' electricity usage increases gradually on an annual basis.

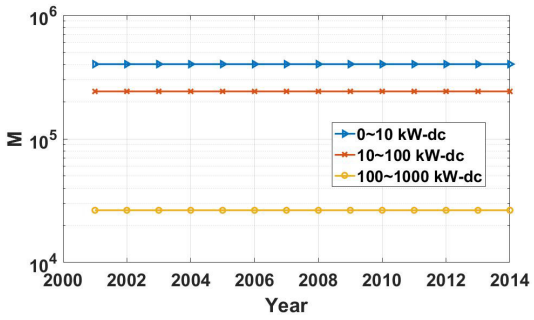


Figure 6.5: M of different size groups.

6.5.2.2 Calculation of Incentive Savings

The incentives received by commercial customers from the CSI program are estimated by taking the average of each year's cash rebates of the entire state of California, based on part 4) of the data set. This model is a simplification of the actual CSI incentive program, which has a 10-tier structure with available incentive funds decreasing over time. The federal level ITC has been in effect since 2006. According to the latest ITC amendment, the incentive credit for a solar PV system will be 30% of expenditures till 2019. This incentive credit rate is scheduled to decrease from 30% in 2019 to 10% in 2022 and beyond [135].

6.5.3 Estimation of Model Parameters and Standard Errors

Model parameter estimation is conducted by using 168 data points of monthly solar PV system adoption from January 2001 to December 2014. The RMSE of the estimated adoption is calculated for the same time range. The NLS problem is solved by the Nelder-Mead simplex algorithm [136]. The monthly installed PV system cost and incentive are calculated by linear interpolation, using the annual installed PV cost and incentives data.

Before model parameters are estimated, the monthly adoption data series is smoothed using moving average method with a window of 3, 5, and 7 months. The smoothing process mitigates the spikes in the historical monthly adoption data. As mentioned in Section 6.4.1, GBM with both incentives and costs is used for the 10-100 kW-dc and the 100-1000 kW-dc groups, and GBM with only costs is used for the 0-10 kW-dc group. However, for the size groups 0-10 kW-dc and 10-100 kW-dc, the asymptotic standard errors are larger than the parameter estimates of β_0. This means β_0 is insignificant in these two cases. Therefore instead of GBM, BM is used in the 0-10 kW-dc group, and incentives are removed from the GBM for the 10-100 kW-dc group.

To validate the significance of estimated parameters, asymptotic standard errors are calculated as follows [137]. Assume N sample points are used in the estimation. Let $\theta = [c\ d\ a\ b\ \beta_0]'$ be the vector of parameters. Let K be the degree of θ. Based on (6.16), let $h(t_i, \theta) = Mc[F(t_i) - F(t_{i-1})]$. The variance of disturbance u_i is estimated as follows:

$$\hat{\sigma}^2 = \frac{1}{N-K}\sum_{i=1}^{N}[S_i - h(t_i, \hat{\theta})] \tag{6.17}$$

where $\hat{\theta}$ is the estimated parameters. Then

$$X^{0\prime}X^0 = \sum_{i=1}^{N}(\frac{\partial h(t_i, \theta_0)}{\partial \theta_0})(\frac{\partial h(t_i, \theta_0)}{\partial \theta_0'}) \tag{6.18}$$

The estimated covariance matrix of θ is given by:

$$Est.Asy.Cov[\theta] = \hat{\sigma}^2(X^{0\prime}X^0)^{-1} \tag{6.19}$$

The standard errors of p and q, as functions of a and b, can be estimated as follows [138]. Let $g(\theta)$ be a function of the parameter vector θ, the standard error of g can be estimated by:

$$SE(g) = (\frac{\partial g(\hat{\theta})}{\partial \theta})^T Est.Asy.Cov[\theta](\frac{\partial g(\hat{\theta})}{\partial \theta}) \tag{6.20}$$

By following the steps above, the estimated parameters and their asymptotic standard errors are calculated and shown in Tables 6.1, 6.2, and 6.3. The fitted PV adoption curves with a smoothing window of 3 months are shown in Figures 6.6, 6.7, and 6.8.

Table 6.1: Estimated parameter values and their standard errors (SE), using GBM with the incentives and cost, 100-1000 kW-dc

window		d	c	a	b	p	q	β_0	RMSE	BM RMSE
3	value	17	6.49E-02	2.10E+02	2.89E-02	1.37E-04	2.88E-02	-1.44E+01	2.52	2.89
	SE	5.78E-01	9.25E-03	7.01E+01	2.85E-03	3.38E-05	2.88E-03	3.22E+00	-	-
5	value	17	4.97E-02	7.01E+02	3.56E-02	5.07E-05	3.56E-02	-2.37E+01	1.86	2.43
	SE	2.85E-01	3.16E-03	2.13E+02	2.00E-03	1.28E-05	2.01E-03	2.86E+00	-	-
7	value	17	4.58E-02	1.21E+03	3.96E-02	3.26E-05	3.96E-02	-2.23E+01	1.60	2.16
	SE	2.67E-01	2.19E-03	3.42E+02	1.80E-03	7.88E-06	1.81E-03	2.49E+00	-	-

Table 6.2: Estimated parameter values and their standard errors (SE), using GBM with cost, 10-100 kW-dc

window		d	c	a	b	p	q	β_0	RMSE	BM RMSE
3	value	9	8.28E-03	1.13E+02	2.45E-02	2.15E-04	2.43E-02	-7.15E+01	3.60	3.81
	SE	5.66E-01	5.34E-04	4.32E+01	3.53E-03	5.32E-05	3.57E-03	1.97E+01	-	-
5	value	11	7.49E-03	2.56E+02	3.16E-02	1.23E-04	3.15E-02	-5.16E+01	3.09	3.32
	SE	5.99E-01	3.50E-04	1.02E+02	3.74E-03	3.49E-05	3.77E-03	1.47E+01	-	-
7	value	11	8.26E-03	1.80E+02	3.03E-02	1.64E-04	3.01E-02	-3.82E+01	2.75	2.93
	SE	6.54E-01	4.04E-04	5.78E+01	3.02E-03	3.77E-05	3.05E-03	1.06E+01	-	-

Table 6.3: Estimated parameter values and their standard errors (SE), using BM, 0-10 kW-dc

window		c	a	b	p	q	RMSE
3	value	3.28E-03	1.57E+03	5.31E-02	3.37E-05	5.30E-02	3.50
	SE	1.67E-04	6.94E+02	3.40E-03	1.27E-05	3.41E-03	-
5	value	3.25E-03	1.67E+03	5.36E-02	3.20E-05	5.36E-02	2.87
	SE	1.34E-04	6.08E+02	2.80E-03	9.98E-06	2.81E-03	-
7	value	3.24E-03	1.72E+03	5.39E-02	3.13E-05	5.39E-02	2.42
	SE	1.12E-04	5.29E+02	2.37E-03	8.27E-06	2.38E-03	-

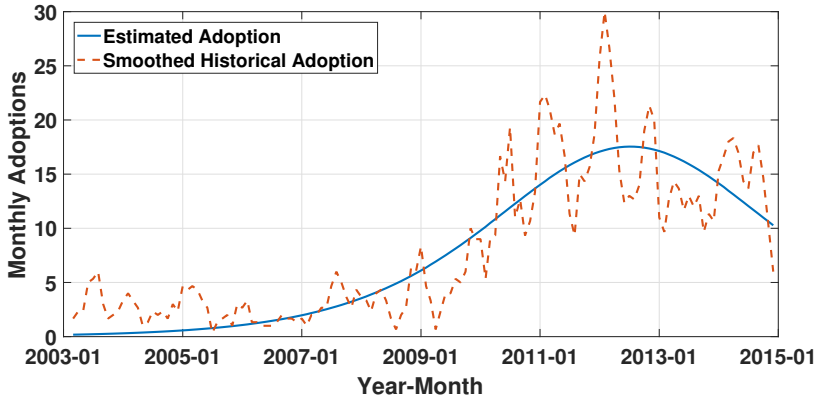


Figure 6.6: Adoption curve fitting of 0-10 kW-dc

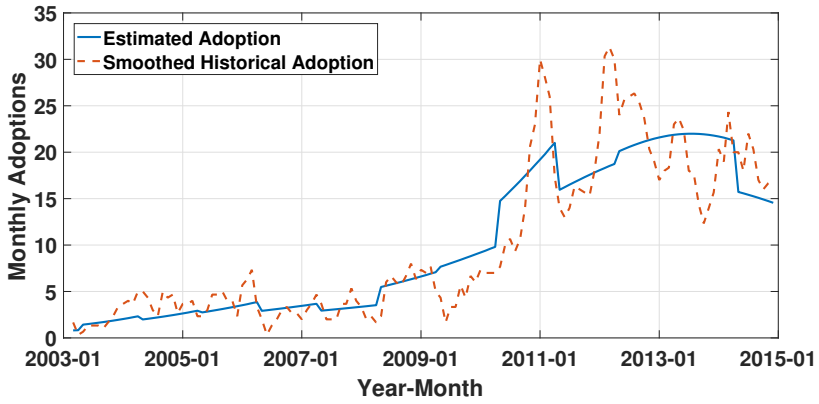


Figure 6.7: Adoption curve fitting of 10-100 kW-dc

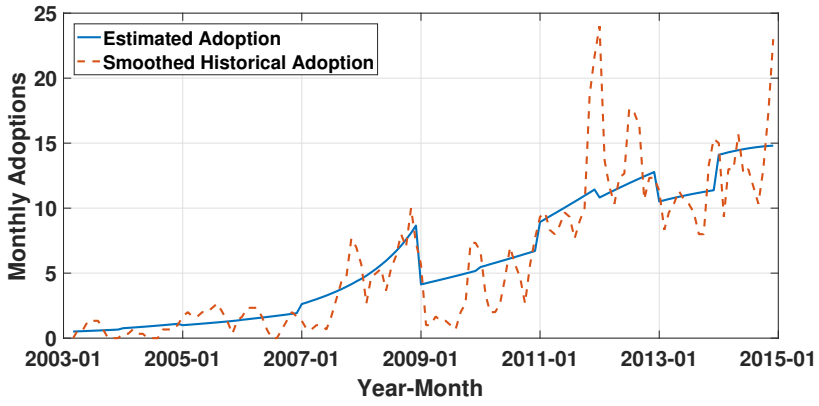


Figure 6.8: Adoption curve fitting of 100-1000 kW-dc

6.5.4 Interpretations of Model Fitting Results

6.5.4.1 Validity of the Parameter Estimation and RMSE

As shown in Tables 6.1, 6.2, and 6.3, the estimated parameters are stable across different smoothing windows, i.e., the estimated parameters under different smoothing windows are very similar. The estimated standard errors are much smaller than the parameter values. The explanatory variables in all three models are considered to be significant. In the 100-1000 kW-dc and the 10-100 kW-dc groups, the GBM have lower RMSE than the BM.

6.5.4.2 Interpretations of Parameter Estimates and Effective Decision Variables

As discussed in [139], the innovation factor p represents the contribution to the new adoptions that do not depend on the number of previous adoptions. These adoptions are due to some influence outside the social system of the customers, and let's call it the external influence. On the other hand, the imitation factor q represents the contribution to the new adoptions that are due to the prior adoptions. The probability of these adoptions are proportional to the prior market penetration level. The new adoptions due to q can be interpreted as an effect of word-of-mouth from customers who are satisfied with the product. In the case of solar PV adoption, the imitation factor q represents the effect of solar PV information spread by prior adopters. From Tables 6.1, 6.2, and 6.3, we can observe that the values of p in the 100-1000 kW-dc and the 10-100 kW-dc groups are larger than that in the 0-10 kW-dc group, while the values of q in the 100-1000 kW-dc and the 10-100 kW-dc groups are smaller than that in the 0-10 kW-dc group. This shows that larger-sized groups are more likely to be affected by external influence, compared with smaller-sized groups. The external influence includes advertisement and direct sales pitch from solar PV developers. This phenomena is called

the small commercial solar gap [140]. In addition, small commercial solar projects often have varying contract terms, power purchasers without credit ratings, and site-specific project requirements. The customer procurement and transaction costs associated with smaller commercial projects are nearly the same as those for larger deals. These difficulties have often led developers to focus their attention on larger commercial projects.

As mentioned in Section 6.4.1 and Section 6.5.3, the final effective decision variables for the three size groups are different. For the 100-1000 kW-dc group, both the incentive and the cost have strong impacts on the solar PV adoption; for the 10-100 kW-dc group, the impact of the incentive is much weaker and the incentive is insignificant; for the 0-10 kW-dc group, even the cost is no longer a significant decision variable. These observations further corroborate the small commercial solar gap phenomena.

6.5.4.3 Forecasting Eventual Adoption Rate and Delay Effect

Note that the 100-1000 kW-dc group has the largest c and the 0-10 kW-dc group has the smallest c. This shows that the eventual solar PV system adoption rate is higher for customer groups that plan to install larger solar PV systems. This observation can be explained intuitively as follows. With higher electricity costs, larger commercial customers are more motivated to install the solar PV system if it makes economic sense. In addition, with more building roof space, larger commercial customers typically have better solar PV mounting conditions than smaller customers. Finally, larger customers are more likely to be in better financial condition, which provides them easier access to solar PV system financing options.

Note that the eventual commercial solar PV adoption rate estimated in our model is much lower than the residential and/or commercial solar PV adoption rate estimated in other literatures

[141, 142, 143, 144, 122]. In the work of [142] and [122], the eventual adoption rate is simply assumed to be 30% for residential customers. In the work of [143], instead of eventual adoption rate, total roof area suitable for solar PV installation is estimated. It is concluded that 60%-65% of the roof area of commercial and industrial buildings, and 22%-27% of the roof area of residential buildings are suitable for solar PV installation. In the work of [144], the eventual adoption rate is estimated to be 51% for residential customers and 52% for commercial customers. The results of both [143] and [144] are based on analysis of the building roof space data. However, not all buildings suitable for solar PV installations will eventually adopt solar PV. The aforementioned references ignored the fact that the commercial customers may not be the owner of the building. This leads to the incentive splitting problem, where building owners pay for the solar PV system, but cannot easily recover savings from reduced electricity use that accrue to the tenants. Furthermore, if commercial building occupants only signed short-term leases, they will not have enough time to recoup the installation costs of solar PV systems [145].

Another observation from the model fitting result is that the effect of time delay d is larger in the 100-1000 kW-dc group than the 10-100 kW-dc group. As larger solar PV projects require substantial capital and may have a significant impact on the distribution network, deciding to adopt the solar PV system, securing the electrical permit and installing the system all require longer lead time.

6.6 Quantitative Evaluation and Forecast of the Impact of Incentives and Solar PV Costs on the Adoption

The solar PV adoption model is a useful tool for policy evaluations. Once the model parameters are estimated, the impact of government incentives and declining cost of solar PV systems on the adoption can be quantified and forecasted. The adoption forecasting model can provide useful feedback to government policy makers in developing future renewable energy policies. In this section, we first quantify the impact of federal and state solar incentive programs on the adoption. The assumptions regarding incentive programs and solar PV system costs are as follows. The incentive for the CSI has been in effect since 2007, and is assumed to be zero beyond 2015. The ITC program has been in effect since 2006, and its investment credit rate is assumed to decrease from 30% in 2019 to 10% in 2022 and beyond. It is assumed that in the next five years, the solar PV system cost will be declining at the same historical rate.

By setting the CSI and/or the ITC to zero, we simulated what would have happened to the commercial solar PV adoption without the incentive programs. Let's treat the cumulative solar PV adoption capacity without both incentive programs as the benchmark. Figure 6.9 shows the percentage increase in cumulative PV adoption capacity due to one or both of the incentive programs.

Figure 6.9 demonstrates that in 2008 the provisioning of the CSI program increased the cumulative solar PV adoption by 14.9% compared to the benchmark case, and the ITC program increased the cumulative PV adoption by 16.6%. By implementing both incentive programs, the cumulative solar PV adoption increased by 43.55% compared the benchmark case.

Based on the forecasting results, the ITC program has a greater influence than the CSI program on the commercial solar PV adoption in California. This is because the ITC provides more

rebates than the CSI. In 2007, the CSI's average rebate was about 27.5% of the installation cost, but the rate kept dropping, and was only 7.5% in 2014; on the other hand, the ITC has kept a tax credit rate of 30% of the installation cost since 2006.

Similarly, we can also quantify and forecast the impact of solar PV systems' cost on adoption. As mentioned in Section 6.1, the federal and state governments have been supporting research and strengthening solar manufacturing capabilities to drive down the cost. Let's treat the cumulative solar PV adoption capacity where the cost of solar PV system always stayed at the level of year 2001 as the benchmark. Figure 6.10 illustrates the impact of reduction in solar PV system cost on the cumulative PV system adoption. Figure 6.10 shows that in 2012, when the solar PV cost declined at the historical rate, the cumulative PV adoption capacity is 25.1% higher than the benchmark case; and if the cost had declined half as fast as the historical rate, the cumulative adoption is 9.9% higher than the benchmark case where the cost had stayed at the same level as year 2001.

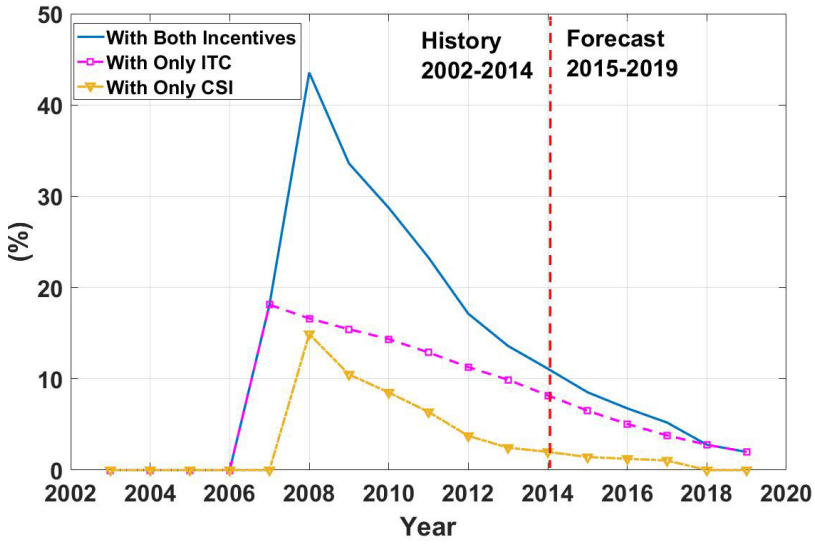


Figure 6.9: Percentage increase of cumulative installed solar PV capacity under different incentive scenarios, 100-1000 kW-dc.

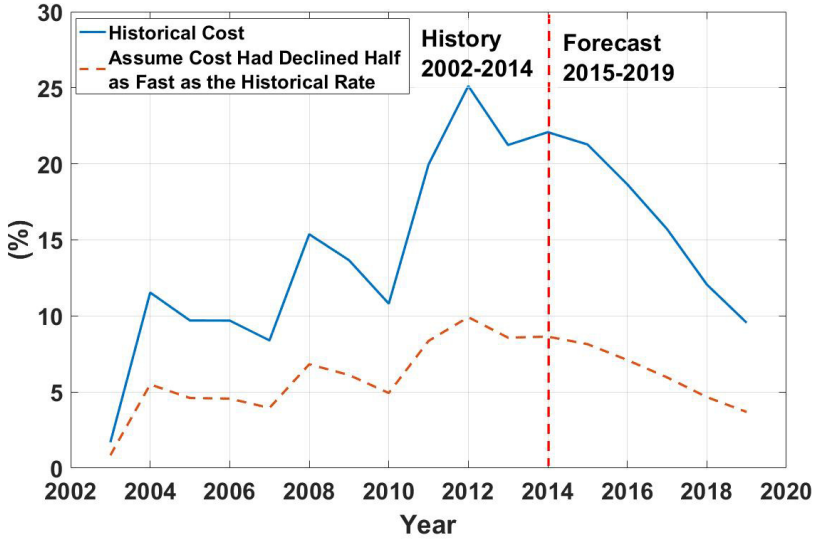


Figure 6.10: Percentage increase of cumulative installed solar PV capacity under different cost assumptions, 100-1000 kW-dc.

6.7 Conclusion

In this chapter, a commercial solar PV adoption forecasting model based on the Generalized Bass Model is developed in this chapter. This model not only provides robust parameter estimates, but also yields lower estimation error than the Bass Model. In addition, the model is capable of not only forecasting the eventual adoption rate of commercial solar PV systems, but also estimating the time delay of impacts on adoption. The proposed commercial solar PV adoption model is a very useful tool to quantify and forecast the impact of government policies and solar PV system costs on the commercial solar PV adoption. The simulation results show that both direct government solar incentive programs and declining solar PV system costs had a significant impact on the adoption of commercial solar PV systems in Southern California. Moreover, we introduced a method to forecast the size of the solar PV system that a customer is likely to adopt. This forecast can provide a valuable guidance to the power distribution system planners regarding optimal grid expansion and upgrade plans.

Based on the historical adoption data, the forecasted eventual adoption rate of commercial solar PV systems is much lower than that of residential systems. With a higher percentage of leased buildings in the commercial sector, the incentive splitting problem is more pronounced and has inhibited widespread adoption of commercial solar PV systems. Therefore, traditional government solar system incentive programs need to be complemented by policies that directly promote the adoption of solar PV system for non-owner-occupied commercial buildings. Future work can be done to explore and evaluate the impact of third-party ownership on the adoption of commercial solar PV system.

Chapter 7

Diversity Factor Prediction for Distribution Feeders

7.1 Introduction

The maximum diversified demand, i.e., the maximum of the sum of demands of a group of electricity customers over a particular period, is one of the most important factors to consider when utilities develop plans to build new distribution systems. The maximum diversified demand is very important to the design of both network topology and the ratings of equipment. Underestimating the maximum diversified demand will cause reliability and safety issues. If the peak load exceeds the circuit rating, then equipment such as transformers and cables will be overloaded, which results in shortened lifespan and premature failure. Overestimating the maximum diversified demand often leads to installation of oversized distribution system equipment and under-utilization of system assets.

The maximum diversified demand is usually estimated by using the maximum noncoincident demand and the diversity factor (DF), which is defined as follows [29]:

$$Diversity\ factor = \frac{Maximum\ noncoincident\ demand}{Maximum\ diversified\ demand}. \tag{7.1}$$

Here, the maximum noncoincident demand is the sum of each individual customer's maximum demand. Obviously, DF is greater than or equal to 1. A higher DF means that customers have more diversified usage patterns and their individual maximum loads have less coincidence in time. In general, as the number of customers increases, DF first increases and then gradually levels off.

The maximum noncoincident demand is straightforward to estimate because an individual customer's maximum demand is the customer's electric service rating, which can be obtained by survey. Thus, the key problem is how to estimate DF.

In practice, DF is often estimated based on a simple relationship. Engineers estimate DF by referring to a DF table, in which the DF value varies with the number of customers. The DF table is often derived by utilities through load surveys from a few groups of customers in the distribution system [29]. In the load survey, the maximum demand of each individual customer and their maximum diversified demand are recorded. However, DF is influenced by many other factors, such as customer demographics and climate conditions. Thus, DF tables, which ignore these factors, have limited accuracy. Furthermore, engineers cannot interpret or explain how various factors affect the DF.

In this chapter, by leveraging supervised machine learning algorithms, we build comprehensive DF prediction models that take a variety of factors into account. These models show high prediction accuracy and interpretabilty when applied to real-world distribution feeders. Using the interpretation method called SHapley Additive exPlanations, we quantify the importance of differ-

ent features in determining DFs. Finally, we offer more insights into how various factors affect DFs.

The rest of the chapter is organized as follows. Section 7.2 reviews the literature and summarizes our work's contribution. Section 7.3 explains the machine learning methodologies used to develop and interpret the DF prediction model. Section 7.4 summarizes the real-world distribution feeders and influential factors used to construct the dataset for the DF prediction model. Section 7.5 shows the DF prediction performance and provides interpretation for the model. Section 7.6 states the conclusions.

7.2 Literature Review and Contributions of Our Work

Researchers have studied different aspects of DF and demand diversity. However, very few research efforts have focused on developing comprehensive and interpretable prediction models for DF which account for various input features [146]. Early research [147] models DF as a function of the number of customers. Different DF functions are derived based on time of the year, day of the week, and whether electric-heating is used. Ref. [148] studies the distribution of DF and shows that DF follows gamma distributions rather than Gaussian distributions. Ref. [149] studies a metric called after-diversity maximum demand of n customers (ADMD^n), which is closely related to DF and demand diversity. This work shows that ADMD^n is affected by customers' household occupancy and wealth levels. In [150], a variable truncated R-vine copulas method is used to estimate the maximum diversified demand of customers of different household occupancy and wealth levels.

In this work, we develop comprehensive models based on supervised machine learning algorithms to predict the DF of distribution feeders, accounting for a variety of influential factors,

such as customer type, weather, demographics, and socioeconomic conditions. The machine learning algorithms not only yield high prediction accuracy on real-world distribution feeders but also provide useful insights on how input features influence DF. Using the interpretation method called SHapley Additive exPlanations (SHAP) [37], we identify the key factors that affect the DF.

7.3 Machine Learning Methodologies for DF Prediction Models

We adopt supervised machine learning algorithms to build the DF prediction model, which maps the input features to the output (i.e., DF of a feeder). In supervised machine learning, a model learns its mapping from a training dataset, which are samples of correct input-output pairs. Mean square error (MSE) is used to measure the model prediction performance. The details of DF prediction model development are provided in Section 7.3.1. To interpret the prediction model, we use a method called SHAP [37] to identify the most important input features that influence the DF prediction. The details of SHAP are explained in Section 7.3.2

7.3.1 Supervised Machine Learning Algorithms

To estimate DF of distribution feeders, we adopt 3 types of supervised machine learning algorithms: feed forward neural network (FNN), gradient boosted trees (GBT), and random forest. We choose these 3 algorithms, because they are widely used in the machine learning field and achieve great results in various problems. We further improve FNN by adding dropout layer(s) and introducing network pruning. Thus, in total, we deploy 6 algorithms: FNN, FNN+dropout, FNN+pruning, FNN+dropout+pruning, GBT, and random forest. The overall framework for building and evaluating DF prediction models consists of 3 steps. First, we preprocess the dataset and

split it into training, validation, and test datasets. Second, for each of the 6 models, we train the model and tune the model's hyperparameters. Third, we evaluate the performance of the 6 prediction models using the test dataset. Due to the underlying randomness in the training and model initialization processes, we train each model 10 times and report the average model prediction errors. The technical details related to the supervised machine learning algorithms are presented below.

7.3.1.1 FNN

Our base FNN consists of three components: an input layer of 45 nodes, three hidden layers of 200 nodes, and an output layer of 1 node. Each node has directed connections to the nodes of the subsequent layer and each connection has a corresponding weight. In the input layer, each node corresponds to an input variable. In the hidden layer, each node takes in the weighted sum of nodes from the previous layer (plus a bias term) and produces an output value by the ReLU activation function. The output layer is a linear function of the nodes in the last hidden layer. When training FNN and its variants, we use early stopping with patience=200 epochs.

7.3.1.2 Network Pruning

Pruning removes unnecessary branches to improve the performance of FNN. We adopt an innovative pruning method called lottery ticket [151, 152]. This pruning method comprises the following steps: a) randomly initialize a neural network with weights w_i; b) train the neural network, reaching the trained weights w_f; c) prune $p\%$ of the weights that have the smallest w_f in magnitude, i.e., set the pruned weights to 0; d) reset the unpruned weights in w_f to their initial values in w_i (i.e., winning tickets) and retrain the network while keeping the pruned weights to 0. It is believed that pruning produces a sparse neural network with less connections, which can reduce

overfitting. In addition, the winning tickets may discover a good initialization point that already lies in the randomly initialized network.

7.3.1.3 GBT

GBT is an ensemble learning method, which consists of a series of decision trees. The summed/aggregated prediction of the decision trees are used as the output. The GBT is trained by adding one tree a time while keeping the existing trees unchanged. Each new tree is trained using a gradient descent procedure so that the loss of the ensemble model is reduced. To avoid overfitting in the training process, we use early stopping technique with patience=200 to decide when to stop adding trees.

7.3.1.4 Random Forest

Random forest is another widely used ensemble learning method. It outputs the average prediction of multiple decision trees, which are fitted to various subsets of the dataset. Different from GBT, which trains a new tree based on the existing ones, random forest trains trees that are almost independent.

7.3.1.5 Data Preprocessing and Split

Every numerical input feature is standarized, i.e., centered and normalized by its standard deviation. This standarization shifts and rescales feature variables to similar ranges and thus improves convergence in the training process. Every categorical feature variable is represented by one-hot encoding. In our problem, only the climate zone featue is a categorical variable. For input features that are linearly dependent, we remove one of them. For example, the ratios of population

in different age ranges sum up to 1. Thus, we remove one of the ratios. Such features, called redundant features, are highly correlated with other features, so they do not provide relevant information. It is a common practice to remove them in machine learning.

64% of the samples in the dataset are used to train the prediction models. 16% of the samples are used as the validation dataset for hyperparameter tuning and early stopping. The remaining 20% of the samples are used as the test set to evaluate the models' prediction performance.

7.3.1.6 Hyperparameter Tuning

Hyperparameters are the settings and parameters that control the configuration and influence the performance of machine learning algorithms. Following the common pratice, we use the validation dataset to tune the hyperparameters. Under different hyperparameter settings, each model is trained 10 times using the training dataset and then evaluated on the validation set. For each of the 6 prediction models, the hyperparameter setting with the lowest average validation MSE is selected.

The possible hyperparameter settings for all 6 models are listed in Table 7.1. Every combination of the hyperparameter settings is examined when tuning the hyperparameters. For the model FNN+Dropout, all setting combinations between FNN and dropout are examined. For the FNN+Pruning model and FNN+Dropout+Pruning, we fix the hyperparameter settings already tuned for FNN and FNN+Dropout, and only tune the network pruning rate $p\%$.

7.3.2 The SHAP Method for Model Interpretation

It is important to understand how DF of a distribution feeder is influenced by different features. In this work, we use the SHAP [37] framework to interpret the DF prediction models. The

Table 7.1: Summary of Hyperparameters and Their Settings

Model	Hyperparameters and Their Possible Configurations
FNN	Batch size = [5,10,50,100]; optimizer = [Stochastic Gradient Descent, RMSprop, Adagrad, Adadelta, Adam, Adamax, Nadam].
Dropout	Input layer dropout ratio = [0.05, 0.1, 0.15, 0.2, 0.3]; hidden layer dropout ratio = [0.05, 0.1, 0.15, 0.2, 0.3, 0.4, 0.5].
Pruning	Pruning rate $p\%$ = [25%, 50%, 75%, 85%, 95%, 98%, 99%].
GBT	Max. tree depth= [2,4,6,8]; learning rate = [0.01, 0.02, 0.05, 0.1]; subsample ratio of training instances = [0.3, 0.4, ..., 0.8]; subsample ratio of features for each split = [0.1, 0.2, ..., 1].
Random Forest	Number of trees= [5,10,50,100,500,1000,5000,10000,100000]; max. number of features to consider for the best split = [m, $\log_2(m)$, $\sqrt{m}$, $m/3$] (m: the total number of features).

SHAP framework has a solid theoretical foundation in cooperative game theory. It calculates each input feature's contribution to the model's output so that the influence on the output can be fairly distributed to the input features. The SHAP framework is model-agnostic, meaning that it does not require the knowledge of the model structure. Thus, SHAP works well with all types of prediction models.

The inner workings of SHAP can be explained as follows. Suppose we have a prediction model $y = f(\boldsymbol{x})$, where $\boldsymbol{x} = [x_1, ..., x_m]$ is the input feature vector and y is the model output. All the samples of $\boldsymbol{x}$ form a set $\boldsymbol{X}$. For any sample $\boldsymbol{x}^{(i)} = [x_1^{(i)}, ..., x_m^{(i)}] \in \boldsymbol{X}$, SHAP calculates a vector $\boldsymbol{\phi}^{(i)} = [\phi_1^{(i)}, ..., \phi_m^{(i)}]$ representing the contribution of each input feature in $\boldsymbol{x^{(i)}}$, such that $\sum_{j=1}^{m} \phi_j^{(i)} = f(\boldsymbol{x}^{(i)}) - E_{\boldsymbol{x}\in\boldsymbol{X}}(f(\boldsymbol{x}))$. Here, $E_{\boldsymbol{x}\in\boldsymbol{X}}(f(\boldsymbol{x}))$ is the expectation of $f(\boldsymbol{x})$. We call $\phi_j^{(i)}$ the SHAP value of input feature j for sample i. For more details of SHAP, please refer to [37].

Note that SHAP is a local method, which explains a model prediction based on each individual sample input. Thus, in this study, the same input feature has different SHAP values for

different distribution feeders. By examining these SHAP values, we can discover which features have a significant contribution to the prediction output and how a input feature' contribution varies among different feeders.

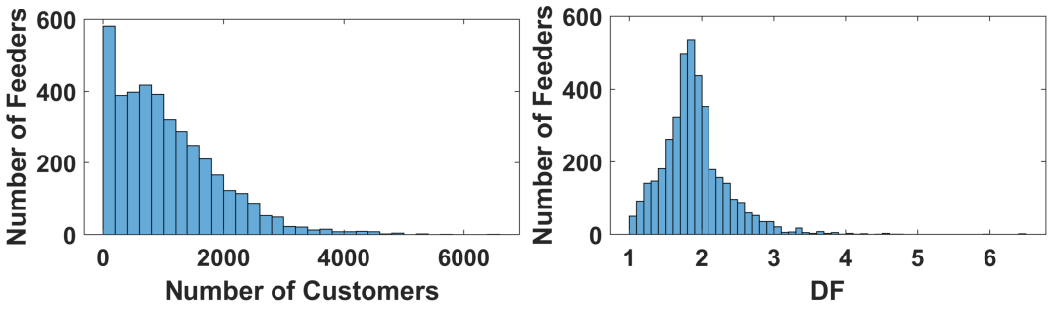
In this work, to interpret a prediction model, we calculate the SHAP value of all input features for every feeder in the dataset. Since each model is trained 10 times, the average SHAP value is reported as the final result.

7.4 Descriptions of Real-World Distribution Feeders and Input Features for DF Prediction

In this section, we first describe the distribution feeders used in the case study and summarize the statistics for their DFs. Then, we describe the input features used to predict DFs.

The case study covers 3,952 distribution feeders managed by Southern California Edison. In total, these feeders serve over 4,000,000 customers. The histogram of number of customers for the feeders is shown in Fig. 7.4. Using one year of hourly kWh readings of customers in 2015, we calculate the DFs of all distribution feeders according to equation (7.1). The histogram of DF is shown in Fig. 7.4.

We collect various types of input features to predict DFs of distribution feeders. The input features can be categorized into three classes: feeder characteristics, customer demographic and socioeconomic conditions, and environmental factors. The input features are summarized in Table 7.2. The sources of these input features are provided below.



(a) Histogram of # of customers.

(b) Histogram of DF.

Figure 7.1: Overview of feeders and DFs in the dataset.

7.4.1 Feeder Characteristics

Input features in this class represent the properties of the distribution feeder, which include number of customers, customer type, and the size and penetration rate of solar PV systems. These information is provided by Southern California Edison.

7.4.2 Demographic and Socioeconomic Conditions

Input features in this category are collected from the National Historical Geographic Information System (NHGIS) [153]. NHGIS organizes customers' data by census block groups (CBGs) instead of feeders. Thus, we derive the input feature values of each distribution feeder by matching the feeder's service area to the geographic locations of CBGs.

7.4.3 Environmental Factors

The California Energy Commission provides the climate zone information for each zip code [154]. By mapping the distribution feeders' locations to zip codes, we can obtain the climate zone information of each feeder. The weather data is collected from the National Centers for En-

vironmental Information [155], which organizes weather data by weather stations. By mapping the feeder locations to weather stations, we can obtain the weather data for each feeder. The elevation of distribution feeders are collected from U.S. Geological Survey by queries using feeder locations.

Table 7.2: Summary of Input Features

Class	Feature Type	Feature Description
Feeder Characteristics	No. of Customers	Number of customers in each feeder.
	Customer Type	Ratio of residential customers, ratio of commercial customers.
	Solar PV	Ratio of customers with solar PV and average solar PV size of commercial and residential customers, respectively.
Demographic and Socioeconomic Conditions	Age	Average age, ratio of population in 4 groups: child age ($\leq$5 years), school age (6$\sim$17 years), work age (18$\sim$61 years), retired age ($\geq$ 62 years).
	Education	Ratio of population in 4 educational levels: lower than college, less than 4 years' college, bachelor's degree, higher than bachelor's degree.
	Average Room No.	Average No. of rooms of a housing unit.
	Annual Income	Average household income, ratio of population in 3 income levels: $\leq$\$34,999, \$35,000$\sim$\$149,999, $\geq$\$150,000.
	Population	Population of each feeder's CBG.
	Occupancy Ratio	Occupancy ratio of housing units.
	Child Family Ratio	Ratio of families with children.
	Employment	Ratio of population in 4 conditions: employed, unemployed, army, not in labor.
Environmental Factors	Climate Zone	Building climate zone of California.
	Weather	Annual avg. of daily max. and min. temperature; annual highest, lowest, and avg. temperature; No. of days with max. temperature$\geq$90, $\geq$70, and $\leq$32; No. of days with min. temperature$\leq$32 and $\leq$0; heating degree days; cooling degree days.
	Elevation	Elevation of feeders' service area.

7.5 DF Prediction Performance and Interpretation of the Machine Learning Model

In this section, we first present the DF prediction performance of different machine learning models. Then, we quantify the features' importance in determining feeder DF. Lastly, we analyze how different features affect the DF prediction and provide more insights into how DF is determined. The case study is conducted in Python on an Oracle-Sun workstation with 2.3 GHz Intel Xeon CPUs and 128 GB of RAM.

7.5.1 Prediction Performance of Machine Learning Models

The MSEs of 6 machine learning models on the test dataset are shown in Fig. 7.2. Each model is trained 10 times with the tuned hyperparameter setting and the MSEs are represented by the box plot. The red bar represents the median value, and the green diamond marker represents the mean value. The variance of DF in the test dataset is 0.22811. The MSE of the benchmark linear regression model is 0.13445. As shown in the figure, all 6 supervised machine learning models yield more accurate DF prediction results than the linear regression model. Among all tested models, random forest has the lowest average MSE. The figure also shows that pruning improves the accuracy of FNN and FNN+dropout models. FNN+dropout+pruning and GBT have a similar level of performance. All 6 machine learning models take less than 1 second to predict the DFs of the 3,952 feeders. Since DF prediction is often conducted as part of the distribution system planing process, the model training can be done off-line.

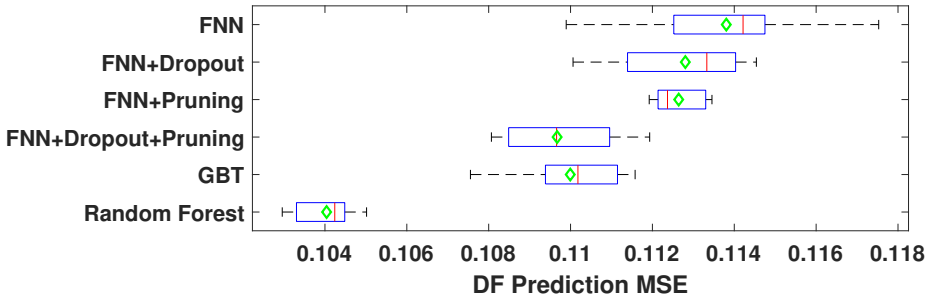


Figure 7.2: Box plot of the DF prediction MSEs of 6 models.

7.5.2 Importance of Different Feature Types

We calculate the SHAP values of all input features and samples for the random forest model, which yields the best prediction results. We then derive the feature type importance as follows. First, for each feeder, we sum up the SHAP values of input features by feature type. Then, for each feature type, we calculate the average absolute value of the SHAP values over all feeders, which quantifies the importance of each feature type in determining DFs. Fig. 7.3 shows the importance of all feature types, ranked from the highest to the lowest. The most influential feature types are customer type, weather, solar PV, climate zone, and number of customers.

7.5.3 Impacts of Input Features on DF Prediction

For the random forest model, we select a few features with high importance and analyze their effects on DF prediction. To do so, we plot the SHAP values of a feature vs. the values of the feature for all distribution feeders in Fig 7.4. In the subfigures, each circle represents a feeder, and we can see how a feature's contribution to DF (i.e., the SHAP value) changes when the feature's value changes. To demonstrate the interactions between features, we color the circles by the ratio

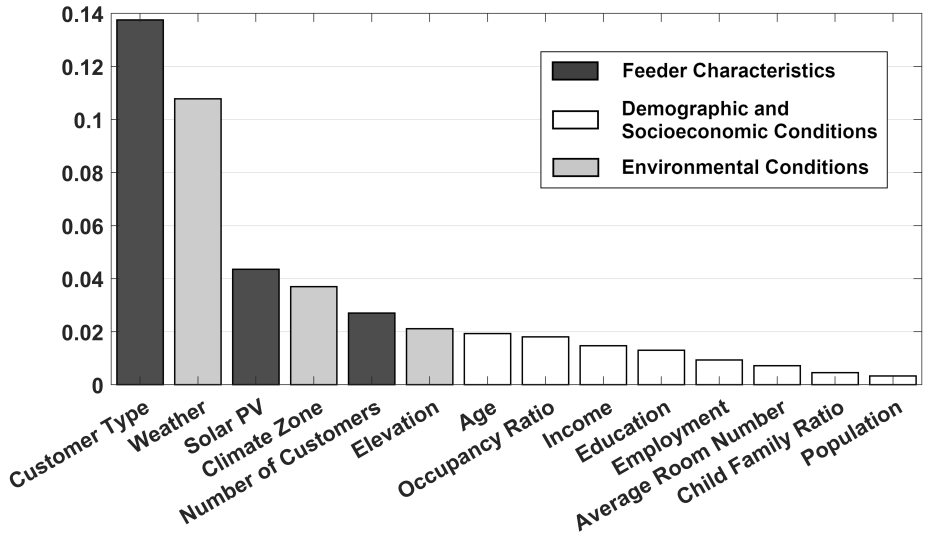
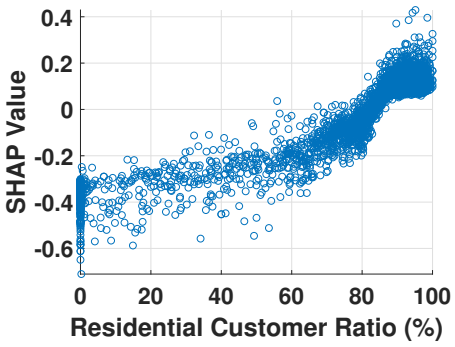


Figure 7.3: Feature importance of the random forest model

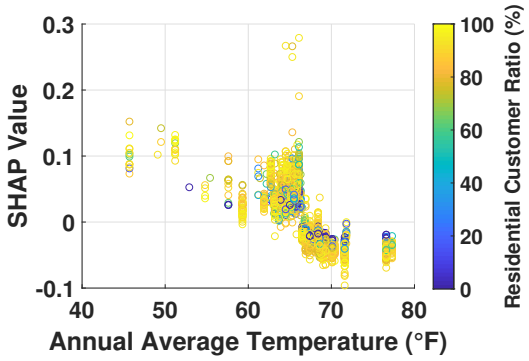
of residential customers in some subfigures.

7.5.3.1 The Impacts of Customer Type

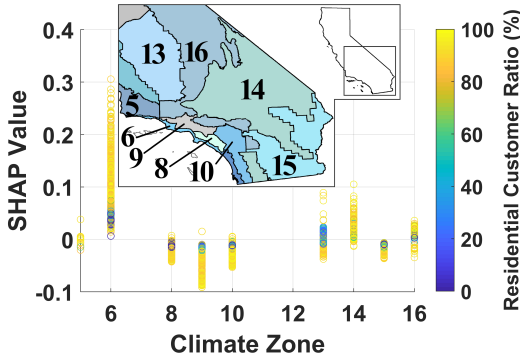
In the testing feeders, all customers are either residential or commercial. As shown in Fig. 7.5.3, feeders with higher ratio of residential customers tend to have higher DFs. This phenomena can be explained as follows. The electricity usage patterns of commercial buildings are less diversified because their demands usually follow normal business schedules. For example, restaurants, department stores, and cinemas often have similar operation hours. In comparison, residential customers often have drastically different electricity usage patterns due to the randomness of residents' activities. Therefore, a typical feeder's DF increases as the ratio of residential customers increases.



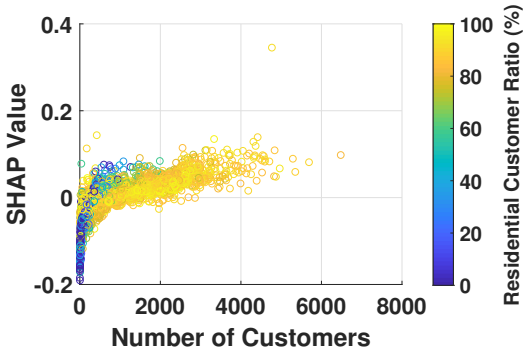
(a) Customer Type



(b) Temperature



(c) Climate Zone



(d) Number of Customers

Figure 7.4: Features' contribution to the DF prediction of the random forest model.

7.5.3.2 The Impacts of Annual Average Temperature

In Fig. 7.5.3, as the annual average temperature increases, the feeder DF decreases, which can be explained as follows. Southern California Edison is a summer peaking utility. The air-conditioning load is a significant component of the peak demand on very hot summer afternoons. In the service area with higher annual average temperature, the individual customer's annual maximum demand is more likely to occur during the same time periods, i.e., hot summer afternoon hours. Thus, the distribution circuits with higher annual average temperature tend to have lower DFs. An interesting observation is that the SHAP values significantly decrease when the annual average temperature exceeds 66 . This observation is consistent with the building standards and convention in the United States, which state that a building needs to be cooled when the daily average ambient temperature is above 65 . We can also see that feeders with lower residential customer ratios are less sensitive to the change of annual average temperature.

7.5.3.3 The Impacts of Climate Zone

There are 9 climate zones in the study area. The small figure inside Fig. 7.5.3 shows the areas covered by different climate zones. Fig. 7.5.3 shows the SHAP values of different climate zones for all the feeders. All other things being equal, zone 6 tends to have higher DFs. This is because, compared to other climate zones, zone 6 is a coastal region with a mild climate, lower summer temperature, and lower air-conditioning load during summer months. Thus, feeders located in zone 6 should have higher DFs based on the arguments stated in Section 7.5.3.2. It can also been seen from Fig. 7.5.3 that the DFs for feeders with lower residential customer ratios are less sensitive to the change in climate zone.

7.5.3.4 The Impacts of Number of Customers

Fig. 7.5.3 shows the impacts of number of customers on DFs. As shown in the figure, distribution feeders with a higher number of customers tend to have higher DFs. The saturation effect kicks in when the number of customers reaches a certain level. These observations are consistent with the relationship between number of customers and DFs in a typical DF table. Finally, it can also be seen from Fig. 7.5.3 that the DFs of feeders with lower residential customer ratios are more sensitive to the change in the number of customers.

7.6 Conclusion

In this chapter, we build a suite of DF prediction models based on supervised machine learning techniques, which take a comprehensive list of input features into consideration. When applied to real-world distribution feeders, these models produce highly accurate DF prediction results. Using SHAP, a unified framework for interpreting machine learning models, we discover that the most influential input features in determining DFs are customer type, weather, solar PV penetration rate and size, climate zone, and number of customers. The SHAP values of different input features also offer useful insights into how different input features affect the prediction of DFs. In practice, the proposed DF prediction models are useful tools for designing the distribution system and sizing distribution system equipment. The future work can be done to to evaluate the impact of electricity rate design (e.g., time-of-use rate) on DFs.

Chapter 8

Conclusions

In this dissertation, I investigate the four problems critical to the adoption of DERs: phase identification, network parameter estimation, DER adoption prediction, and long-term load forecasting in the distribution system. For phase identification, two unsupervised learning algorithms and a maximum marginal likelihood estimation method based on a physics-informed model are developed. For network parameter estimation, a maximum likelihood estimation method based on a physics-informed model and a method based on graphical learning model are developed. For DER adoption prediction, I present a generalized Bass model to forecast the adoption of distributed commercial solar PV systems. For long-term load forecasting, comprehensive supervised learning models are designed to forecast the diversity factor of distribution feeders.

In Chapter 2, we develop two algorithms of phase identification by unsupervised learning of the smart meter voltage time series data. Case study in real distribution feeders show that both algorithms are computationally efficient and highly accurate. The first algorithm uses linear dimension reduction and constrained k-means clustering of smart meter data. Network connectivity

information is adopted to avoid mislabeling of customers on the same secondary feeder. Adding connectivity information as constraints effectively improves the accuracy. The second algorithm uses nonlinear dimension reduction and density-based clustering. Compared to the existing solutions, the second algorithm has three main advantages. First, it does not require prior knowledge about the number of phase connections in the distribution system. Second, it works well with distribution feeders that have both phase-to-neutral and phase-to-phase connections. Third, its accuracy is not very sensitive to the level of unbalance in a distribution feeder.

In Chapter 3, a physics-informed data-driven algorithm is developed to identify phase connections in the power distribution system. An MLE and an MMLE problem are formulated based on the three-phase power flow manifold. It is proved that the correct phase connection is a global optimum for both the MLE and the MMLE problems. A computationally efficient algorithm is then designed to solve the MMLE problem, which involves synthesizing the solutions from the sub-problems via the voting and the target-only approaches. The sub-problems are further transformed into an equivalent binary least square form and solved efficiently by relaxing the binary constraints. The proposed phase identification algorithm yields high accuracy and outperforms existing methods in comprehensive simulations with real-world smart meter data and IEEE distribution test circuits. Simulations also show the proposed algorithm is fairly robust with respect to inaccurate feeder models, incomplete measurements, and bad measurements.

In Chapter 4, a data-driven parameter estimation algorithm for three-phase power distribution networks is developed. In the algorithm, only the readily available smart meter data are used to estimate the three-phase serial conductance and susceptance of the primary line segments. I first formulate the network parameter estimation problem as an MLE problem based on the linearized

three-phase power flow. The correct network parameters can be proved to yield the highest likelihood value. An SGD-based algorithm with early stopping is designed to solve the MLE problem. In the comprehensive numerical study, proposed algorithm is shown to improve the accuracy of the parameter estimates.

In Chapter 5, a physics-informed graphical learning algorithm is developed to estimate three-phase series resistance and reactance of the primary distribution line segments. The proposed algorithm is broadly applicable because it uses only readily available smart meter data. I leverage the domain knowledge of power distribution systems by replacing the deep neural network-based transition functions in the graph neural network with three-phase power flow-based physical transition functions. The gradient of the loss function for first difference voltage time series with respect to line parameters is rigorously derived. Stochastic gradient descent algorithm is used to estimate the network parameters. The prior distribution of the line parameters is also considered to further improve the accuracy of the proposed parameter estimation algorithm. By comprehensive numerical study on IEEE test feeders, the proposed algorithm is shown to significantly outperform the state-of-the-art algorithm. The relative advantage of the proposed algorithm becomes more pronounced when smart meter measurement noise level is higher.

In Chapter 6, I develop a forecasting model for commercial solar PV adoption based on the generalized Bass model. This model not only provides robust parameter estimates, but also yields lower estimation error than the Bass model. In addition, the model can forecast the eventual adoption rate of commercial solar PV systems and estimate the time delay of impacts on adoption as well. The proposed model is a very useful tool to quantify and forecast the impact of government policies and solar PV system costs on the commercial solar PV adoption. In the simulation, it is

shown that both direct government solar incentive programs and declining solar PV system costs had a significant impact on the adoption of commercial solar PV systems in Southern California. Moreover, a method is introduced to forecast the size of the solar PV system that a customer is likely to adopt, which provides a valuable guidance to the power distribution system planners regarding optimal grid expansion and upgrade plans.

In Chapter 7, a suite of DF prediction models are built based on supervised machine learning techniques, which take into account a comprehensive list of input features. These models produce highly accurate DF prediction results when applied to real-world distribution feeders. SHAP, a unified framework for interpreting machine learning models, is used to analyze the influence of input features. It is discovered that the most influential input features in determining DFs are customer type, weather, solar PV penetration rate and size, climate zone, and number of customers. The SHAP values also offer useful insights into how different input features affect the prediction of DFs. In practice, the proposed DF prediction models are useful tools for designing the distribution system and sizing distribution system equipment.

Bibliography

[1] P. Asmus, M. Lawrence, A. Metz, P. Gunjan, R. R. Labastida, S. Shepard, and E. Woods, "Integrated DER: Orchestrating the grid's last mile," Guidehouse, Tech. Rep., Sep. 2020. [Online]. Available: https://guidehouse.com/-/media/www/site/insights/energy/2020/gh_integrated-der_2020.pdf

[2] M. Hong, X. Yu, N.-P. Yu, and K. A. Loparo, "An energy scheduling algorithm supporting power quality management in commercial building microgrids," *IEEE Transactions on Smart Grid*, vol. 7, no. 2, pp. 1044–1056, March 2016.

[3] N. Yu, T. Wei, and Q. Zhu, "From passive demand response to proactive demand participation," in *2015 IEEE International Conference on Automation Science and Engineering (CASE)*. IEEE, Aug. 2015, pp. 1300–1306.

[4] T. Wei, Q. Zhu, and N. Yu, "Proactive demand participation of smart buildings in smart grid," *IEEE Transactions on Computers*, vol. 65, no. 5, pp. 1392–1406, May 2016.

[5] W. Wang and N. Yu, "LMP decomposition with three-phase DCOPF for distribution system," in *2016 IEEE Innovative Smart Grid Technologies-Asia (ISGT-Asia)*. IEEE, Nov. 2016, pp. 1–8.

[6] W. Wang and N. Yu, "Chordal conversion based convex iteration algorithm for three-phase optimal power flow problems," *IEEE Transactions on Power Systems*, vol. 33, no. 2, pp. 1603–1613, March 2018.

[7] Y. Liu, N. Yu, W. Wang, X. Guan, Z. Xu, B. Dong, and T. Liu, "Coordinating the operations of smart buildings in smart grids," *Applied Energy*, vol. 228, pp. 2510–2525, Oct. 2018.

[8] W. Wang, A. Abdolrashidi, N. Yu, and D. Wong, "Frequency regulation service provision in data center with computational flexibility," *Applied Energy*, vol. 251, p. 113304, Oct. 2019.

[9] J. Shi, Y. Gao, W. Wang, N. Yu, and P. A. Ioannou, "Operating electric vehicle fleet for ride-hailing services with reinforcement learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 11, pp. 4822–4834, Nov. 2020.

[10] M. Ostadijafari, A. Dubey, and N. Yu, "Linearized price-responsive HVAC controller for optimal scheduling of smart building loads," *IEEE Transactions on Smart Grid*, vol. 11, no. 4, pp. 3131–3145, July 2020.

[11] Berg Insight AB, "Smart metering–world 2019," Berg Insight AB, Tech. Rep., July 2019. [Online]. Available: https://www.researchandmarkets.com/reports/4793286/smart-metering-world-2019

[12] IoT Analytics, "Smart meter market report 2019-2024," IoT Analytics, Tech. Rep., Nov. 2019. [Online]. Available: https://iot-analytics.com/smart-meter-market-2019-global-penetration-reached-14-percent/

[13] D. P. Mishra, S. R. Samantaray, and G. Joos, "A combined wavelet and data-mining based intelligent protection scheme for microgrid," *IEEE Transactions on Smart Grid*, vol. 7, no. 5, pp. 2295–2304, Sept. 2016.

[14] W. Wang, N. Yu, Y. Gao, and J. Shi, "Safe off-policy deep reinforcement learning algorithm for volt-var control in power distribution systems," *IEEE Transactions on Smart Grid*, vol. 11, no. 4, pp. 3008–3018, July 2020.

[15] W. Wang, N. Yu, J. Shi, and Y. Gao, "Volt-var control in power distribution systems with deep reinforcement learning," in *2019 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*. IEEE, Oct. 2019, pp. 1–7.

[16] Y. Gao, W. Wang, and N. Yu, "Consensus multi-agent reinforcement learning for volt-var control in power distribution networks," *IEEE Transactions on Smart Grid*, Feb. 2021.

[17] Y. Gao, J. Shi, W. Wang, and N. Yu, "Dynamic distribution network reconfiguration using reinforcement learning," in *2019 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*. IEEE, Oct. 2019, pp. 1–7.

[18] Y. Gao, W. Wang, J. Shi, and N. Yu, "Batch-constrained reinforcement learning for dynamic distribution network reconfiguration," *IEEE Transactions on Smart Grid*, vol. 11, no. 6, pp. 5357–5369, Nov. 2020.

[19] F. Kabir, N. Yu, W. Yao, R. Yang, and Y. Zhang, "Estimation of behind-the-meter solar generation by integrating physical with statistical models," in *2019 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*. IEEE, Oct. 2019, pp. 1–6.

[20] F. Kabir, N. Yu, W. Yao, R. Yang, and Y. Zhang, "Joint estimation of behind-the-meter solar generation in a community," *IEEE Transactions on Sustainable Energy*, vol. 12, no. 1, pp. 682–694, Jan. 2021.

[21] W. S. Bierer, "Long range phasing voltmeter," Oct. 5 2010, US Patent 7,808,228.

[22] W.-H. E. Liu and S.-L. Lim, "Parameter error identification and estimation in power system state estimation," *IEEE Transactions on Power Systems*, vol. 10, no. 1, pp. 200–209, Feb. 1995.

[23] P. Zarco and A. G. Exposito, "Power system parameter estimation: A survey," *IEEE Transactions on Power Systems*, vol. 15, no. 1, pp. 216–222, Feb. 2000.

[24] N. Logic and G. T. Heydt, "An approach to network parameter estimation in power system state estimation," *Electric Power Components and Systems*, vol. 33, no. 11, pp. 1191–1201, Aug. 2006.

[25] Y. Lin and A. Abur, "Enhancing network parameter error detection and correction via multiple measurement scans," *IEEE Transactions on Power Systems*, vol. 32, no. 3, pp. 2417–2425, May 2017.

[26] K. A. Horowitz, Z. Peterson, M. H. Coddington, F. Ding, B. O. Sigrin, D. Saleem, S. E. Baldwin, B. Lydic, S. C. Stanfield, N. Enbar *et al.*, "An overview of distributed energy resource (DER) interconnection: Current practices and emerging solutions," National Renewable Energy Lab (NREL), Golden, CO (United States), Tech. Rep., April 2019. [Online]. Available: https://www.nrel.gov/docs/fy19osti/72102.pdf

[27] F. Bass, "A new product growth model for consumer durables," *Management Science*, vol. 15, no. 5, pp. 215–227, Jan. 1969.

[28] F. M. Bass, T. V. Krishnan, and D. C. Jain, "Why the Bass model fits without decision variables," *Marketing Science*, vol. 13, no. 3, pp. 203–223, Aug. 1994.

[29] W. H. Kersting, *Distribution System Modeling and Analysis*, 3rd ed. CRC press, 2012.

[30] W. Wang, N. Yu, B. Foggo, J. Davis, and J. Li, "Phase identification in electric power distribution systems by clustering of smart meter data," in *Machine Learning and Applications (ICMLA), 2016 15th IEEE International Conference on*. IEEE, Dec. 2016, pp. 259–265.

[31] W. Wang and N. Yu, "Advanced metering infrastructure data driven phase identification in smart grid," in *The Second International Conference on Green Communications, Computing and Technologies*, Sep. 2017, pp. 16–23.

[32] W. Wang and N. Yu, "Maximum marginal likelihood estimation of phase connections in power distribution systems," *IEEE Transactions on Power Systems*, vol. 35, no. 5, pp. 3906–3917, Sept. 2020.

[33] W. Wang and N. Yu, "Parameter estimation in three-phase power distribution networks using smart meter data," in *2020 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*. IEEE, Aug. 2020, pp. 1–6.

[34] W. Wang and N. Yu, "Estimate three-phase distribution line parameters with physics-informed graphical learning method," *arXiv preprint arXiv:2102.09023 [cs.LG]*, Feb. 2021. [Online]. Available: https://arxiv.org/pdf/2102.09023v1.pdf

[35] W. Wang, N. Yu, and R. Johnson, "A model for commercial adoption of photovoltaic systems in California," *Journal of Renewable and Sustainable Energy*, vol. 9, no. 2, p. 025904, March 2017.

[36] W. Wang, N. Yu, J. Shi, and N. Navarro, "Diversity factor prediction for distribution feeders with interpretable machine learning algorithms," in *2020 IEEE Power & Energy Society General Meeting (PESGM)*. IEEE, Aug. 2020, pp. 1–5.

[37] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems*, Dec. 2017, pp. 4765–4774.

[38] "Distributed energy resources global forecast," Navigant Research, Tech. Rep., 2015.

[39] T. A. Short, "Advanced metering for phase identification, transformer identification, and secondary modeling," *IEEE Transactions on Smart Grid*, vol. 4, no. 2, pp. 651–658, June 2013.

[40] T. A. Short, *Electric Power Distribution Handbook*, 2nd ed. CRC press, 2014.

[41] M. Dilek, "Integrated design of electrical distribution systems: phase balancing and phase prediction case studies," Ph.D. dissertation, Virginia Polytechnic Institute and State University, 2001.

[42] V. Arya, D. Seetharam, S. Kalyanaraman, K. Dontas, C. Pavlovski, S. Hoy, and J. R. Kalagnanam, "Phase identification in smart grids," in *Smart Grid Communications (SmartGridComm), 2011 IEEE International Conference on*. IEEE, Oct. 2011, pp. 25–30.

[43] H. Pezeshki and P. Wolfs, "Correlation based method for phase identification in a three phase LV distribution network," in *Universities Power Engineering Conference (AUPEC), 2012 22nd Australasian*. IEEE, Sept. 2012, pp. 1–7.

[44] M. Wen, R. Arghandeh, A. Meier, K. Poolla, and V. Li, "Phase identification in distribution networks with micro-synchrophasors," in *2015 Power and Energy Society General Meeting*. IEEE, July 2015, pp. 1–5.

[45] K. Caird, "Meter phase identification," July 2010, U.S. Patent App. 12/345,702. [Online]. Available: http://www.google.com/patents/US20100164473

[46] T. W. Liao, "Clustering of time series data–a survey," *Pattern recognition*, vol. 38, no. 11, pp. 1857–1874, Nov. 2005.

[47] J. A. Hartigan and M. A. Wong, "Algorithm as 136: A k-means dlustering algorithm," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 28, no. 1, pp. 100–108, Jan. 1979.

[48] X. Golay, S. Kollias, G. Stoll, D. Meier, A. Valavanis, and P. Boesiger, "A new correlation-based fuzzy logic clustering algorithm for FMRI," *Magnetic Resonance in Medicine*, vol. 40, no. 2, pp. 249–260, Aug. 1998.

[49] I. Jolliffe, *Principal Component Analysis*, 2nd ed. Wiley Online Library, 2002.

[50] K. Wagstaff and C. Cardie, "Clustering with instance-level constraints," *AAAI/IAAI*, vol. 1097, June 2000.

[51] K. Wagstaff, C. Cardie, S. Rogers, S. Schrödl *et al.*, "Constrained k-means clustering with background knowledge," in *ICML*, vol. 1, June 2001, pp. 577–584.

[52] M. Bilenko, S. Basu, and R. J. Mooney, "Integrating constraints and metric learning in semi-supervised clustering," in *Proceedings of the twenty-first international conference on Machine learning*. ACM, July 2004, pp. 81–88.

[53] T. Lange, M. H. Law, A. K. Jain, and J. M. Buhmann, "Learning with constrained and unlabelled data," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1. IEEE, June 2005, pp. 731–738.

[54] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. JHU Press, 2012, vol. 3.

[55] A. Schrijver, *Theory of Linear and Integer Programming*, 1st ed. John Wiley & Sons, 1998.

[56] L. A. Pomatto, "Apparatus and method for identifying the phase of a three phase power line at a remote location," Apr. 23 1996, US Patent 5,510,700.

[57] W. Luan, J. Peng, M. Maras, J. Lo, and B. Harapnuk, "Smart meter data analytics for distribution network connectivity verification," *IEEE Transactions on Smart Grid*, vol. 6, no. 4, pp. 1964–1971, July 2015.

[58] L. V. D. Maaten and G. Hinton, "Visualizing data using t-SNE," *Journal of Machine Learning Research*, vol. 9, pp. 2579–2605, Nov. 2008.

[59] J. W. Sammon, "A nonlinear mapping for data structure analysis," *IEEE Transactions on Computers*, vol. 100, no. 5, pp. 401–409, May 1969.

[60] P. Demartines and J. Hérault, "Curvilinear component analysis: A self-organizing neural network for nonlinear mapping of data sets," *IEEE Transactions on Neural Networks*, vol. 8, no. 1, pp. 148–154, Jan. 1997.

[61] J. B. Tenenbaum, V. De Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, Dec. 2000.

[62] G. E. Hinton and S. T. Roweis, "Stochastic neighbor embedding," in *Advances in Neural Information Processing Systems*, Jan. 2002, pp. 857–864.

[63] E. Z. Macosko *et al.*, "Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets," *Cell*, vol. 161, no. 5, pp. 1202–1214, May 2015.

[64] A. Frome *et al.*, "Devise: A deep visual-semantic embedding model," in *Advances in Neural Information Processing Systems*, Dec. 2013, pp. 2121–2129.

[65] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise." in *KDD*, vol. 96, no. 34, Aug. 1996, pp. 226–231.

[66] C. Ruiz, M. Spiliopoulou, and E. Menasalvas, "C-DBSCAN: Density-based clustering with constraints," in *International Workshop on Rough Sets, Fuzzy Sets, Data Mining, and Granular-Soft Computing*. Springer, May 2007, pp. 216–223.

[67] Y. Liao, Y. Weng, G. Liu, Z. Zhao, C.-W. Tan, and R. Rajagopal, "Unbalanced multi-phase distribution grid topology estimation and bus phase identification," *IET Smart Grid*, vol. 2, no. 4, pp. 557–570, Aug. 2019.

[68] V. D. Krsman and A. T. Sarić, "Verification and estimation of phase connectivity and power injections in distribution network," *Electric Power Systems Research*, vol. 143, pp. 281–291, Feb. 2017.

[69] P. Kumar, V. Arya, D. A. Bowden, and L. Kohrmann, "Leveraging DERs to improve the inference of distribution network topology," in *2017 IEEE International Conference on Smart Grid Communications (SmartGridComm)*. IEEE, Oct. 2017, pp. 52–57.

[70] S. J. Pappu, N. Bhatt, R. Pasumarthy, and A. Rajeswaran, "Identifying topology of low voltage distribution networks based on smart meter data," *IEEE Transactions on Smart Grid*, vol. 9, no. 5, pp. 5113–5122, March 2018.

[71] M. Xu, R. Li, and F. Li, "Phase identification with incomplete data," *IEEE Transactions on Smart Grid*, vol. 9, no. 4, pp. 2777–2785, July 2018.

[72] F. Olivier, A. Sutera, P. Geurts, R. Fonteneau, and D. Ernst, "Phase identification of smart meters by clustering voltage measurements," in *2018 Power Systems Computation Conference (PSCC)*. IEEE, June 2018, pp. 1–8.

[73] S. Bolognani and F. Dörfler, "Fast power system analysis via implicit linearization of the power flow manifold," in *2015 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, Sep. 2015, pp. 402–409.

[74] A. M. Kettner and M. Paolone, "On the properties of the compound nodal admittance matrix of polyphase power systems," *IEEE Transactions on Power Systems*, vol. 34, no. 1, pp. 444–453, Aug. 2019.

[75] S. A. Vavasis, *Complexity theory: Quadratic programming*. Boston, MA: Springer US, 2001, pp. 304–307. [Online]. Available: https://doi.org/10.1007/0-306-48332-7_65

[76] K. Schneider, P. Phanivong, and J.-S. Lacroix, "IEEE 342-node low voltage networked test system," in *2014 IEEE PES General Meeting | Conference & Exposition*. IEEE, July 2014, pp. 1–5.

[77] A. Cooper, "Electric company smart meter deployments: Foundation for a smart grid," The Edison Foundation, Tech. Rep., Dec. 2017. [Online]. Available: https://www.edisonfoundation.net/iei/publications/Documents/IEI_Smart%20Meter%20Report%202017_FINAL.pdf

[78] B. Foggo and N. Yu, "A comprehensive evaluation of supervised machine learning for the phase identification problem," *World Academy of Science, Engineering and Technology International Journal of Computer and Systems Engineering*, vol. 12, no. 6, pp. 419–427, 2018.

[79] B. Foggo and N. Yu, "Improving supervised phase identification through the theory of information losses," *IEEE Transactions on Smart Grid*, vol. 11, no. 3, pp. 2337–2346, May 2020.

[80] X. Bian, X. R. Li, H. Chen, D. Gan, and J. Qiu, "Joint estimation of state and parameter with synchrophasors—part ii: Parameter tracking," *IEEE Transactions on Power Systems*, vol. 26, no. 3, pp. 1209–1220, Aug. 2011.

[81] C. Rakpenthai and S. Uatrongjit, "Power system state and transmission line conductor temperature estimation," *IEEE Transactions on Power Systems*, vol. 32, no. 3, pp. 1818–1827, May 2017.

[82] D. Ritzmann, P. S. Wright, W. Holderbaum, and B. Potter, "A method for accurate transmission line impedance parameter estimation," *IEEE Transactions on Instrumentation and Measurement*, vol. 65, no. 10, pp. 2204–2213, Oct. 2016.

[83] J. Yu, Y. Weng, and R. Rajagopal, "Patopa: A data-driven parameter and topology joint estimation framework in distribution grids," *IEEE Transactions on Power Systems*, vol. 33, no. 4, pp. 4335–4347, Nov. 2017.

[84] O. Ardakanian, V. W. S. Wong, R. Dobbe, S. H. Low, A. von Meier, C. J. Tomlin, and Y. Yuan, "On identification of distribution grids," *IEEE Transactions on Control of Network Systems*, vol. 6, no. 3, pp. 950–960, Sept. 2019.

[85] A. M. Prostejovsky, O. Gehrke, A. M. Kosek, T. Strasser, and H. W. Bindner, "Distribution line parameter estimation under consideration of measurement tolerances," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 2, pp. 726–735, April 2016.

[86] M. Lave, M. J. Reno, and J. Peppanen, "Distribution system parameter and topology estimation applied to resolve low-voltage circuits on three real distribution feeders," *IEEE Transactions on Sustainable Energy*, vol. 10, no. 3, pp. 1585–1592, July 2019.

[87] D. Zwillinger, *CRC Standard Mathematical Tables and Formulas*, 33rd ed. Chapman and Hall/CRC, 2018.

[88] M. R. Castillo, J. B. London, N. G. Bretas, S. Lefebvre, J. Prévost, and B. Lambert, "Offline detection, identification, and correction of branch parameter errors based on several measurement snapshots," *IEEE Transactions on Power Systems*, vol. 26, no. 2, pp. 870–877, May 2011.

[89] J. Zhao, S. Fliscounakis, P. Panciatici, and L. Mili, "Robust parameter estimation of the french power system using field data," *IEEE Transactions on Smart Grid*, vol. 10, no. 5, pp. 5334–5344, Sept 2019.

[90] C. Li, Y. Zhang, H. Zhang, Q. Wu, and V. Terzija, "Measurement-based transmission line parameter estimation with adaptive data selection scheme," *IEEE Transactions on Smart Grid*, vol. 9, no. 6, pp. 5764–5773, Nov. 2018.

[91] M. Asprou and E. Kyriakides, "Estimation of transmission line parameters using pmu measurements," in *2015 IEEE Power & Energy Society General Meeting*. IEEE, July 2015, pp. 1–5.

[92] R. Kumar, M. G. Giesselmann, and M. He, "State and parameter estimation of power systems using phasor measurement units as bilinear system model," *International Journal of Renewable Energy Research*, vol. 6, no. 4, pp. 1373–1384, May 2016.

[93] K. V. Khandeparkar, S. A. Soman, and G. Gajjar, "Detection and correction of systematic errors in instrument transformers along with line parameter estimation using PMU data," *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 3089–3098, July 2017.

[94] S. Gajare, A. K. Pradhan, and V. Terzija, "A method for accurate parameter estimation of series compensated transmission lines using synchronized data," *IEEE Transactions on Power Systems*, vol. 32, no. 6, pp. 4843–4850, Nov. 2017.

[95] P. Ren, H. Lev-Ari, and A. Abur, "Tracking three-phase untransposed transmission line parameters using synchronized measurements," *IEEE Transactions on Power Systems*, vol. 33, no. 4, pp. 4155–4163, July 2018.

[96] J. Yu, Y. Weng, and R. Rajagopal, "Patopaem: A data-driven parameter and topology joint estimation framework for time-varying system in distribution grids," *IEEE Transactions on Power Systems*, vol. 34, no. 3, pp. 1682–1692, May 2019.

[97] S. Han, D. Kodaira, S. Han, B. Kwon, Y. Hasegawa, and H. Aki, "An automated impedance estimation method in low-voltage distribution network for coordinated voltage regulation," *IEEE Transactions on Smart Grid*, vol. 7, no. 2, pp. 1012–1020, March 2016.

[98] J. Peppanen, M. J. Reno, R. J. Broderick, and S. Grijalva, "Distribution system model calibration with big data from AMI and PV inverters," *IEEE Transactions on Smart Grid*, vol. 7, no. 5, pp. 2497–2506, Sept. 2016.

[99] J. Zhang, Y. Wang, Y. Weng, and N. Zhang, "Topology identification and line parameter estimation for non-PMU distribution network: A numerical method," *IEEE Transactions on Smart Grid*, vol. 11, no. 5, pp. 4440–4453, Sept. 2020.

[100] V. C. Cunha, W. Freitas, F. C. Trindade, and S. Santoso, "Automated determination of topology and line parameters in low voltage systems using smart meters measurements," *IEEE Transactions on Smart Grid*, vol. 11, no. 6, pp. 5028–5038, Nov. 2020.

[101] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini, "The graph neural network model," *IEEE Transactions on Neural Networks*, vol. 20, no. 1, pp. 61–80, Jan. 2009.

[102] J. Zhou, G. Cui, Z. Zhang, C. Yang, Z. Liu, L. Wang, C. Li, and M. Sun, "Graph neural networks: A review of methods and applications," *arXiv preprint arXiv:1812.08434*, vol. 1, 2018.

[103] M. A. Khamsi and W. A. Kirk, *An Introduction to Metric Spaces and Fixed Point Theory*. John Wiley & Sons, 2011, vol. 53.

[104] F. J. Pineda, "Generalization of back-propagation to recurrent neural networks," *Physical Review Letters*, vol. 59, no. 19, pp. 2229–2232, June 1987.

[105] L. B. Almeida, "A learning rule for asynchronous perceptrons with feedback in a combinatorial environment," in *Artificial Neural Networks: Concept Learning*, Jan. 1990, pp. 102–111.

[106] U.S. Energy Information Administration, "Renewables share of U.S. energy consumption highest since 1930s," http://www.eia.gov/todayinenergy/detail.cfm?id=21412, May 2015.

[107] S. Kann, "U.S. solar market insight report-2014 year in review-executive summary," GTM Research and SEIA, Tech. Rep., March 2015.

[108] U.S. Energy Information Administration, "EIA electricity data now include estimated small-scale solar PV capacity and generation," https://www.eia.gov/todayinenergy/detail.cfm?id=23972, Dec. 2015.

[109] The White House, "Solar progress report: Advancing toward a clean energy future," Office of the Press Secretary, The White House, Tech. Rep., May 2014.

[110] M. D. Platzer, "U.S. solar photovoltaic manufacturing: Industry trends, global competition, federal support," Congressional Research Service, Tech. Rep., Jan. 2015.

[111] S. Kann, M. Shiao, C. Honeyman, J. Jones, A. Perea, C. Smith, B. Gallagher, S. Moskowitz, J. Baca, S. Rumery, A. Holm, and K. O'Brien, "U.S. solar market insight report 2015 year in review," SEIA and GTM Research, Tech. Rep., March 2016.

[112] A. David, "U.S. solar photovoltaic (PV) cell and module trade overview," Government of the United States of America - International Trade Commission, Tech. Rep., June 2011.

[113] G. Barbose, S. Weaver, and N. Darghouth, "Tracking the sun VII: An historical summary of the installed price of photovoltaics in the United States from 1998 to 2013," Lawrence Berkeley National Laboratory, Tech. Rep., Sept. 2014.

[114] Solar Energy Industries Association, "Solar investment tax credit," http://www.seia.org/policy/finance-tax/solar-investment-tax-credit, Jan. 2016.

[115] California Energy Commission and California Public Utilities Commission, "The California solar initiative - CSI," http://www.gosolarcalifornia.ca.gov/csi/index.php, Jan. 2016.

[116] California Energy Commission and California Public Utilities Commission, "What is the new solar homes partnership?" http://www.gosolarcalifornia.org/about/nshp.php, Jan. 2016.

[117] U.S. Department of Energy, "Self-generation incentive program," http://energy.gov/savings/self-generation-incentive-program, Jan. 2016.

[118] D. C. Schmittlein and V. Mahajan, "Maximum likelihood estimation for an innovation diffusion model of new product acceptance," *Marketing Science*, vol. 1, no. 1, pp. 57–78, Feb. 1982.

[119] V. Srinivasan and C. H. Mason, "Technical note-nonlinear least squares estimation of new product diffusion models," *Marketing Science*, vol. 5, no. 2, pp. 169–178, May 1986.

[120] V. Rai and K. McAndrews, "Decision-making and behavior change in residential adopters of solar PV," in *Proceedings of the World Renewable Energy Forum, Denver, CO*, May 2012.

[121] C. Schelly, "Residential solar electricity adoption: What motivates, and what matters? A case study of early adopters," *Energy Research & Social Science*, vol. 2, pp. 183–191, June 2014.

[122] A. Agarwal, D. Cai, S. Shah, M. Chandy, and R. Sherick, "A model for residential adoption of photovoltaic systems," in *2015 IEEE Power & Energy Society General Meeting*. IEEE, July 2015, pp. 1–5.

[123] M. Guidolin and C. Mortarino, "Cross-country diffusion of photovoltaic systems: modelling choices and forecasts for national adoption patterns," *Technological Forecasting and Social Change*, vol. 77, no. 2, pp. 279–296, Feb. 2010.

[124] U.S. Energy Information Administration, "Most new residential solar PV projects in California program are not owned by homeowners," https://www.eia.gov/todayinenergy/detail.php?id=12991, Feb. 2017.

[125] E. Drury, M. Miller, C. M. Macal, D. J. Graziano, D. Heimiller, J. Ozik, and T. D. Perry IV, "The transformation of southern California's residential photovoltaics market through third-party ownership," *Energy Policy*, vol. 42, pp. 681–690, March 2012.

[126] V. Rai and B. Sigrin, "Diffusion of environmentally-friendly energy technologies: buy versus lease differences in residential PV markets," *Environmental Research Letters*, vol. 8, no. 1, p. 014022, Feb. 2013.

[127] C. Davidson, D. Steinberg, and R. Margolis, "Exploring the market for third-party-owned residential photovoltaic systems: Insights from lease and power-purchase agreement contract structures and costs in California," *Environmental Research Letters*, vol. 10, no. 2, p. 024006, Feb. 2015.

[128] SEIA, "Third-party solar financing," http://www.seia.org/policy/finance-tax/third-party-financing, Feb. 2017.

[129] U.S. Department of Energy, "Commercial sector characteristics," http://buildingsdatabook.eren.doe.gov/TableView.aspx?table=3.2.3, Jan. 2016.

[130] U.S. Department of Energy, "Residential sector characteristics," http://buildingsdatabook.eren.doe.gov/TableView.aspx?table=2.2.2, Jan. 2016.

[131] State of California, "Net energy metering," http://www.cpuc.ca.gov/PUC/energy/DistGen/netmetering.htm, Jan. 2016.

[132] D. W. Cai, S. Adlakha, S. H. Low, P. De Martini, and K. M. Chandy, "Impact of residential PV adoption on retail electricity rates," *Energy Policy*, vol. 62, pp. 830–843, Nov. 2013.

[133] Southern California Edison, http://www.sce.com, Feb. 2017.

[134] California Energy Commission and California Public Utilities Commission, "California solar statistics," https://www.californiasolarstatistics.ca.gov/data_downloads/, Jan. 2016.

[135] U.S. Department of Energy, "Business energy investment tax credit (ITC)," http://energy.gov/savings/business-energy-investment-tax-credit-itc, Jan. 2016.

[136] J. A. Nelder and R. Mead, "A simplex method for function minimization," *The Computer Journal*, vol. 7, no. 4, pp. 308–313, Jan. 1965.

[137] W. H. Greene, *Econometric Analysis*, 7th ed. Pearson, 2011.

[138] A. Stuart and J. K. Ord, *Kendall's Advanced Theory of Statistics, volume 1: Dstribution Theory*, 6th ed. London (UK) Edward Arnold, 1994.

[139] F. M. Bass, "Comments on a new product growth for model consumer durables the bass model," *Management Science*, vol. 50, no. 12_supplement, pp. 1833–1840, Dec. 2004.

[140] C. Honeyman, N. Litvak, S. Kann, M. Shiao, S. Mehta, J. Jones, T. Kimbis, J. Baca, S. Rumery, and A. Holm, "U.S. solar market insight report 2014 Q2," SEIA and GTM Research, Tech. Rep., 2014.

[141] J. Melius, R. Margolis, and S. Ong, "Estimating rooftop suitability for PV: A review of methods, patents, and validation techniques," National Renewable Energy Laboratory, Tech. Rep., Dec. 2013.

[142] S. Anders and T. Bialek, "Technical potential for rooftop photovoltaics in the San Diego region," in *ASES 2006 Solar Conference*, July 2006, pp. 8–13.

[143] J. Paidipati, L. Frantzis, H. Sawyer, and A. Kurrasch, "Rooftop photovoltaics market penetration scenarios," National Renewable Energy Laboratory, Tech. Rep., Feb. 2008.

[144] D. Feldman, A. M. Brockway, E. Ulrich, and R. Margolis, "Shared solar: Current landscape, market potential, and the impact of federal securities regulation," National Renewable Energy Laboratory and U.S. Department of Energy, Tech. Rep., April 2015.

[145] B. Inskeep, K. Daniel, and A. Proudlove, "Solar on multi-unit buildings—policy and financing options to address split incentives," NC Clean Energy Technology Center, Tech. Rep., 2015.

[146] N. Yu, S. Shah, R. Johnson, R. Sherick, M. Hong, and K. Loparo, "Big data analytics in power distribution systems," in *2015 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*. IEEE, Feb. 2015, pp. 1–5.

[147] A. Sargent, R. Broadwater, J. Thompson, and J. Nazarko, "Estimation of diversity and kwhr-to-peak-kw factors from load research data," *IEEE Transactions on Power Systems*, vol. 9, no. 3, pp. 1450–1456, Aug. 1994.

[148] V. Chatlani, D. Tylavsky, D. Montgomery, and M. Dyer, "Statistical properties of diversity factors for probabilistic loading of distribution transformers," in *2007 39th North American Power Symposium*. IEEE, Oct. 2007, pp. 555–561.

[149] M. Sun, I. Konstantelos, and G. Strbac, "Analysis of diversified residential demand in London using smart meter and demographic data," in *2016 IEEE Power and Energy Society General Meeting (PESGM)*. IEEE, July 2016, pp. 1–5.

[150] M. Sun, Y. Wang, G. Strbac, and C. Kang, "Probabilistic peak load estimation in smart cities using smart meter data," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 2, pp. 1608–1618, 2018.

[151] J. Frankle and M. Carbin, "The lottery ticket hypothesis: Finding sparse, trainable neural networks," *arXiv preprint arXiv:1803.03635*, March 2019.

[152] H. Zhou, J. Lan, R. Liu, and J. Yosinski, "Deconstructing lottery tickets: Zeros, signs, and the supermask," *arXiv preprint arXiv:1905.01067*, Dec. 2019.

[153] S. Manson, J. Schroeder, D. V. Riper, and S. Ruggles, "IPUMS national historical geographic information system: Version 13.0 [database]," http://doi.org/10.18128/D050.V13.0, 2018, Minneapolis: University of Minnesota.

[154] California Energy Commission, "California energy maps," June 2019. [Online]. Available: https://www.energy.ca.gov/maps/renewable/building_climate_zones.html

[155] National Centers for Environmental Information, April 2019. [Online]. Available: https://www.ncdc.noaa.gov/cdo-web/search

Appendix A

Appendix for Chapter 3

A.1 Simplification of Single-Phase and Two-Phase Branches

In this section, we explain how to convert loaded single-phase and two-phase branches into an equivalent load directly connected to the primary feeder. To do so, we need to estimate each branch's equivalent power injection and voltage magnitude. In other words, given the line impedances of single-phase and two-phase branches, the voltage magnitudes and power injections of the loads, we need to calculate the equivalent power injection and voltage magnitude on the primary feeder. The conversion of single-phase and two-phase branches is carried out separately below.

A.1.1 Simplification of a Single-Phase Line

Suppose there is a single-phase line with impedance z serving a load with power injection S and voltage magnitude $|V|$. It is assumed that the power injection S and the voltage magnitude $|V|$ are given. Thus, the current injection magnitude $|I|$ and power factor angle ϕ can be calcu-

lated. Then, at the upstream port of the primary feeder, the single-phase line's equivalent voltage magnitude is $||V| - z|I|\angle -\phi|$ and the equivalent power injection is $S - z|I|^2$.

A.1.2 Simplification of a Two-Phase Line

For a two-phase line serving a load, the voltage drop along the line section can be described by

$$\begin{bmatrix} V_n^1 \\ V_n^2 \end{bmatrix} = \begin{bmatrix} z_{11} & z_{12} \\ z_{21} & z_{22} \end{bmatrix} \begin{bmatrix} -I \\ I \end{bmatrix} + \begin{bmatrix} V_m^1 \\ V_m^2 \end{bmatrix} \tag{A.1}$$

where z_{11}, z_{12}, z_{21}, and z_{22} form the line impedance matrix, which is assumed to be known. V_n^1, V_n^2, V_m^1, and V_m^2 are the nodal voltage phasors of the upstream port and the load, which are assumed to be unknown. I is the current injection phasor of the load. Subtracting row 2 from row 1 in (A.1), we have

$$V_n^{12} = (z_{12} + z_{21} - z_{11} - z_{22})I + V_m^{12} = z_{sum}I + V_m^{12} \tag{A.2}$$

where $V_n^{12} = V_n^1 - V_n^2$ and $V_m^{12} = V_m^1 - V_m^2$. For load m, using the measured voltage magnitude $|V_m^{12}|$ and power injection S_m, we calculate the current injection magnitude $|I|$ and the power factor angle ϕ. Then, at the upstream port of the primary feeder, the two-phase line's equivalent voltage magnitude is $||V_m^{12}| + z_{sum}|I|\angle -\phi|$ and the equivalent power injection is $S_m + z_{sum}|I|^2$.

A.2 Derivation of the Transformed Linearized Three-phase Power Flow Model

Let A_{mn}^{ij} be the $(N+1)\times(N+1)$ block in matrix A_{mn} corresponding to phase ij. Suppose the first row and column of A_{mn} correspond to the substation node, then A_{mn}^{ij} can be divided into 4

blocks as (A.3), where $\check{A}_{mn}^{ij}$ is a nonsingular $N \times N$ matrix. Define $\check{A}_{mn}$ as the collection of $\check{A}_{mn}^{ij}$ over all i and j, B_{mn} as the collection of $\boldsymbol{b}_{mn}^{ij}$ over all i and j, C_{mn} as the collection of $(\boldsymbol{b}_{mn}^{ij})^T$ over all i and j, and D_{mn} as the collection of d_{mn}^{ij} over all i and j. By permuting the variables and corresponding matrix rows and columns, (3.1) can be transformed into (A.4), where $(\cdot)_{-0}$ denotes a vector excluding the substation node, and $(\cdot)_0$ denotes a vector of the substation node.

$$A_{mn}^{ij} = \begin{bmatrix} d_{mn}^{ij} & (\boldsymbol{b}_{mn}^{ij})^T \\ \boldsymbol{b}_{mn}^{ij} & \check{A}_{mn}^{ij} \end{bmatrix} \tag{A.3}$$

$$\begin{bmatrix} \check{A}_{11} & \check{A}_{12} & B_{11} & B_{12} \\ \check{A}_{21} & \check{A}_{22} & B_{21} & B_{22} \\ C_{11} & C_{12} & D_{11} & D_{12} \\ C_{21} & C_{22} & D_{21} & D_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{v}_{-0} - \overline{\boldsymbol{v}}_{-0} \\ \boldsymbol{\theta}_{-0} - \overline{\boldsymbol{\theta}}_{-0} \\ \boldsymbol{v}_0 - \overline{\boldsymbol{v}}_0 \\ \boldsymbol{\theta}_0 - \overline{\boldsymbol{\theta}}_0 \end{bmatrix} = \begin{bmatrix} \boldsymbol{p}_{-0} \\ \boldsymbol{q}_{-0} \\ \boldsymbol{p}_0 \\ \boldsymbol{q}_0 \end{bmatrix} \tag{A.4}$$

Define Matrix $\mathcal{D}$ as follows:

$$\mathcal{D} = \text{diag}(\mathbf{1}_N, \mathbf{1}_N, \mathbf{1}_N, \mathbf{1}_N, \mathbf{1}_N, \mathbf{1}_N) \tag{A.5}$$

From the property of admittance matrix Y^{ij}, we have $A_{mn}^{ij}\mathbf{1}_{N+1} = \mathbb{0}_{(N+1)\times 1}$ and $[\check{A}_{mn}^{ij}, \boldsymbol{b}_{mn}^{ij}]\mathbf{1}_{N+1} = \mathbb{0}_{N\times 1}$.

Thus, we have the following equality relationship:

$$\begin{bmatrix} \check{A}_{11} & \check{A}_{12} & B_{11} & B_{12} \\ \check{A}_{21} & \check{A}_{22} & B_{21} & B_{22} \end{bmatrix} \begin{bmatrix} \mathcal{D} \\ I_{6\times 6} \end{bmatrix} = \mathbb{0}_{6N\times 6} \tag{A.6}$$

Now, it can be easily shown that

$$\begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} = - \begin{bmatrix} \check{A}_{11} & \check{A}_{12} \\ \check{A}_{21} & \check{A}_{22} \end{bmatrix} \mathcal{D} \tag{A.7}$$

Plugging equation (A.7) into equation (A.4), we have

$$\begin{bmatrix} \check{A}_{11} & \check{A}_{12} \\ \check{A}_{21} & \check{A}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{v}^a_{-0} - \mathbf{1}_N v^a_0 \\ \boldsymbol{v}^b_{-0} - \mathbf{1}_N v^b_0 \\ \boldsymbol{v}^c_{-0} - \mathbf{1}_N v^c_0 \\ \boldsymbol{\theta}^a_{-0} - \mathbf{1}_N \theta^a_0 \\ \boldsymbol{\theta}^b_{-0} - \mathbf{1}_N \theta^b_0 \\ \boldsymbol{\theta}^c_{-0} - \mathbf{1}_N \theta^c_0 \end{bmatrix} = \begin{bmatrix} \boldsymbol{p}_{-0} \\ \boldsymbol{q}_{-0} \end{bmatrix} \tag{A.8}$$

where $\boldsymbol{v}^i_{-0}$ and $\boldsymbol{\theta}^i_{-0}$ denote the phase i variables in $\boldsymbol{v}_{-0}$ and $\boldsymbol{\theta}_{-0}$. v^i_0 and θ^i_0 denote the substation's voltage magnitude and angle of phase i. (A.8) is exactly the same as (3.3).

A.3 Estimation of Nodal Power Injection of a Two-phase Load

Define I_{ab} as the current phasor flowing out of the load's phase A port and into the load's phase B port. Let I_a be the injected current phasor from phase A port, and let I_b be the injected current phasor from phase B port. By definition, we know that $I_a = -I_b = I_{ab}$. Let the angle of V_{ab} be the reference angle, i.e., $V_{ab} = |V_{ab}|\angle 0^\circ$, then

$$\begin{aligned} S_{ab} &= P_{ab} + jQ_{ab} \\ &= V_{ab} I^*_{ab} \\ &= |V_{ab}|[Re(I_{ab}) - jIm(I_{ab})] \end{aligned} \tag{A.9}$$

Thus,

$$\begin{aligned} Re(I_{ab}) &= \frac{P_{ab}}{|V_{ab}|} \\ Im(I_{ab}) &= -\frac{Q_{ab}}{|V_{ab}|} \end{aligned} \tag{A.10}$$

When the three-phase voltages are close to balance, the nodal phase-to-neutral power injection can be estimated by the two-phase power injection as follows:

$$\begin{aligned} S_a &= V_a I_a^* \\ &\approx \frac{\sqrt{3}}{3}|V_{ab}|\angle -30^\circ \cdot I_{ab}^* \\ &= \frac{\sqrt{3}}{3}|V_{ab}|\angle -30^\circ \left(\frac{P_{ab}}{|V_{ab}|} + j\frac{Q_{ab}}{|V_{ab}|}\right) \\ &= \frac{\sqrt{3}}{3}[\cos(-30^\circ) + j\sin(-30^\circ)](P_{ab} + jQ_{ab}) \\ &= \left(\frac{1}{2}P_{ab} + \frac{\sqrt{3}}{6}Q_{ab}\right) + j\left(\frac{1}{2}Q_{ab} - \frac{\sqrt{3}}{6}P_{ab}\right) \end{aligned} \tag{A.11}$$

This is exactly the same as (3.8). Equation (3.9) can be derived in a similar way.

A.4 Link the Voltage Magnitude Measurements of Two-Phase Loads to Nodal Values in the Power Flow Model

In the following derivations, the voltages are in per unit and angles are in radian. For a two-phase load m across phase ij ($ij \in \{ab, bc, ca\}$) at node n, we have

$$\hat{v}_m = v_n^{ij} = \sqrt{(v_n^i)^2 + (v_n^j)^2 - 2v_n^i v_n^j \cos\theta_n^{ij}} \tag{A.12}$$

where $\hat{v}_m$ is load m's magnitude measurement, v_n^{ij} is the voltage magnitude between phase ij at node n, v_n^i is the voltage of phase i at node n, and θ_n^{ij} is the voltage phase angle between phase ij at node n.

Similarly, at the substation, we also have

$$v_0^{ij} = \sqrt{(v_0^i)^2 + (v_0^j)^2 - 2v_0^i v_0^j \cos\theta_0^{ij}} \tag{A.13}$$

where v_0^{ij}, v^i, and θ_0^{ij} are the corresponding nodal values at the substation. Under normal operating conditions, $v_n^i \approx v_n^j \approx 1$, $\theta_n^{ij} \approx \frac{2\pi}{3}$. From (A.12) we have

$$\frac{\partial v_n^{ij}}{\partial v_n^i} \approx \frac{\sqrt{3}}{2}, \quad \frac{\partial v_n^{ij}}{\partial v_n^j} \approx \frac{\sqrt{3}}{2}, \quad \frac{\partial v_n^{ij}}{\partial \theta_n^{ij}} = \frac{\partial v_n^{ij}}{\partial(\theta_n^i - \theta_n^j)} \approx \frac{1}{2} \tag{A.14}$$

Under normal operating conditions, voltage and angle differences between non-substation nodes and the substation node is very small. Thus, we can easily derive (3.10) from (A.14) to approximate $\hat{v}_m - v_0^{ij}$.

A.5 Proof of Lemma 1

Proof. By definition, $\tilde{\boldsymbol{v}}(t) = \tilde{\boldsymbol{v}}(t, \boldsymbol{x}^*) + \boldsymbol{n}(t)$. Plugging it into equation (3.19), we have

$$\begin{aligned}
&\lim_{T\to\infty} f(\boldsymbol{x}) \\
=&\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T}[\tilde{\boldsymbol{v}}(t, \boldsymbol{x}^*) - \tilde{\boldsymbol{v}}(t, \boldsymbol{x}) + \boldsymbol{n}(t)]^T \Sigma_n^{-1} \\
&[\tilde{\boldsymbol{v}}(t, \boldsymbol{x}^*) - \tilde{\boldsymbol{v}}(t, \boldsymbol{x}) + \boldsymbol{n}(t)] \\
=&\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T}[\tilde{\boldsymbol{v}}(t, \boldsymbol{x}^*) - \tilde{\boldsymbol{v}}(t, \boldsymbol{x})]^T \Sigma_n^{-1}[\tilde{\boldsymbol{v}}(t, \boldsymbol{x}^*) - \tilde{\boldsymbol{v}}(t, \boldsymbol{x})] \\
&+ \lim_{T\to\infty} \frac{2}{T}\sum_{t=1}^{T}[\tilde{\boldsymbol{v}}(t, \boldsymbol{x}^*) - \tilde{\boldsymbol{v}}(t, \boldsymbol{x})]^T \Sigma_n^{-1}\boldsymbol{n}(t) \\
&+ \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{n}(t)^T \Sigma_n^{-1}\boldsymbol{n}(t) \\
\geq&\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{n}(t)^T \Sigma_n^{-1}\boldsymbol{n}(t)
\end{aligned} \tag{A.15}$$

It should be noted that $\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T}[\tilde{\boldsymbol{v}}(t, \boldsymbol{x}^*) - \tilde{\boldsymbol{v}}(t, \boldsymbol{x})]^T \Sigma_n^{-1}[\tilde{\boldsymbol{v}}(t, \boldsymbol{x}^*) - \tilde{\boldsymbol{v}}(t, \boldsymbol{x})] \geq 0$ because $\Sigma_n^{-1} \succ 0$. As stated in condition 1 of Lemma 1, $\boldsymbol{n}(t)$ is independent of $\tilde{\boldsymbol{v}}(t, \boldsymbol{x})$ and $\tilde{\boldsymbol{v}}(t, \boldsymbol{x}^*)$, so we have $E([\tilde{\boldsymbol{v}}(t, \boldsymbol{x}^*) - \tilde{\boldsymbol{v}}(t, \boldsymbol{x})]^T \Sigma_n^{-1}\boldsymbol{n}(t)) = 0$. Condition 1 and 2 of Lemma 1 also make

$[\tilde{\boldsymbol{v}}(t,\boldsymbol{x}^*)-\tilde{\boldsymbol{v}}(t,\boldsymbol{x})]^T\Sigma_n^{-1}\boldsymbol{n}(t)$ a sequence of independent variables. Under normal system operating conditions, $[\tilde{\boldsymbol{v}}(t,\boldsymbol{x}^*)-\tilde{\boldsymbol{v}}(t,\boldsymbol{x})]^T\Sigma_n^{-1}\boldsymbol{n}(t)$ has limited variance. By Kolmogorov's Strong Law of Large Numbers [137], $\lim_{T\to\infty}\frac{2}{T}\sum_{t=1}^{T}[\tilde{\boldsymbol{v}}(t,\boldsymbol{x}^*)-\tilde{\boldsymbol{v}}(t,\boldsymbol{x})]^T\Sigma_n^{-1}\boldsymbol{n}(t)\to 0$. Therefore, inequality (A.15) holds. In addition, the minimum of $\lim_{T\to\infty}f(\boldsymbol{x})$ is achieved when $\boldsymbol{x}=\boldsymbol{x}^*$. □

A.6 Proof of Lemma 2

Proof. Following a procedure similar to Appendix A.5, we can prove that $\lim_{T\to\infty}f_m(\boldsymbol{x})\geq\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}n_m(t)^2$, and the minimum of $\lim_{T\to\infty}f_m(\boldsymbol{x})$ is achieved when $\boldsymbol{x}=\boldsymbol{x}^*$. Condition 1) and 2) in Lemma 2 simply mean that we can assign any three-phase loads except load m to any phase and get the same optimum value. This is true, because changing three-phase loads' decision variables does not change the power injections in the system. As long as condition 1) and 2) of Lemma 2 hold, $\tilde{v}_m(t,\boldsymbol{x})=\tilde{v}_m(t,\boldsymbol{x}^*)$. This can also be verified by the structure of $\hat{U}^1$ and $\hat{U}^2$ for three-phase loads. □

Appendix B

Appendix for Chapter 5

B.1 Derivation of $\hat{A}(t)$

The $12N \times 12N$ matrix $\hat{A}(t)$ is defined as

$$\begin{aligned}\hat{A}(t) &\triangleq \frac{\partial \hat{F}_{\boldsymbol{w}}([\hat{\boldsymbol{x}}(t)], [\hat{\boldsymbol{l}}(t)])}{\partial [\hat{\boldsymbol{x}}(t)]} \\ &= \begin{bmatrix} \frac{\partial F_{\boldsymbol{w}}([\boldsymbol{x}(t-1)], [\boldsymbol{l}(t-1)])}{\partial [\boldsymbol{x}(t-1)]} & \mathbb{0}_{6N \times 6N} \\ \mathbb{0}_{6N \times 6N} & \frac{\partial F_{\boldsymbol{w}}([\boldsymbol{x}(t)], [\boldsymbol{l}(t)])}{\partial [\boldsymbol{x}(t)]} \end{bmatrix}\end{aligned} \tag{B.1}$$

$\frac{\partial F_{\boldsymbol{w}}([\boldsymbol{x}(t)], [\boldsymbol{l}(t)])}{\partial [\boldsymbol{x}(t)]}$ is derived by calculating each 6×6 local Jacobian matrix defined as

$$\frac{\partial f_{\boldsymbol{w},n}(t)}{\partial \boldsymbol{x}_k(t)} \triangleq \frac{\partial f_{\boldsymbol{w},n}(\boldsymbol{x}_n(t), \boldsymbol{l}_n(t), \boldsymbol{x}_{\text{ne}(n)}(t))}{\partial \boldsymbol{x}_k(t)}, \quad 1 \leq n, k \leq N \tag{B.2}$$

The calculation of (B.2) depends on n and k. If $k \notin \text{ne}(n)$ and $k \neq n$, then

$$\frac{\partial f_{\boldsymbol{w},n}(t)}{\partial \boldsymbol{x}_k(t)} = \mathbb{0}_{6 \times 6} \tag{B.3}$$

If $k \in \text{ne}(n)$, then we have (B.4), which is a function of line impedance parameters.

$$\frac{\partial f_{\boldsymbol{w},n}(t)}{\partial \boldsymbol{x}_k(t)} = \langle Z_{nn} \rangle \cdot \langle Y_{nk} \rangle \tag{B.4}$$

If $k = n$, then

$$\frac{\partial f_{\boldsymbol{w},n}(t)}{\partial \boldsymbol{x}_k(t)} = \langle Z_{nn} \rangle \cdot \frac{\partial \begin{bmatrix} Re(\boldsymbol{s}_n^*(t) \oslash \boldsymbol{u}_n^*(t)) \\ Im(\boldsymbol{s}_n^*(t) \oslash \boldsymbol{u}_n^*(t)) \end{bmatrix}}{\partial \begin{bmatrix} Re(\boldsymbol{u}_n(t)) \\ Im(\boldsymbol{u}_n(t)) \end{bmatrix}} \tag{B.5}$$

To calculate (B.5), we simplify the notations and define

$$\mathfrak{I}_n^i(t) \triangleq \frac{p_n^i(t) - jq_n^i(t)}{\alpha_n^i(t) - j\beta_n^i(t)}, \quad i = a, b, c. \tag{B.6}$$

By rules of the function derivative, each element in the second term of the RHS of (B.5) can be calculated as in (B.7) and (B.8):

$$\begin{aligned}
\frac{\partial Re(\mathfrak{I}_n^i(t))}{\partial \alpha_n^i(t)} &= \frac{p_n^i(t)[(\beta_n^i(t))^2 - (\alpha_n^i(t))^2] - 2q_n^i(t)\alpha_n^i(t)\beta_n^i(t)}{[(\alpha_n^i(t))^2 + (\beta_n^i(t))^2]^2} \\
\frac{\partial Re(\mathfrak{I}_n^i(t))}{\partial \beta_n^i(t)} &= \frac{q_n^i(t)[(\alpha_n^i(t))^2 - (\beta_n^i(t))^2] - 2p_n^i(t)\alpha_n^i(t)\beta_n^i(t)}{[(\alpha_n^i(t))^2 + (\beta_n^i(t))^2]^2} \\
\frac{\partial Im(\mathfrak{I}_n^i(t))}{\partial \alpha_n^i(t)} &= \frac{\partial Re(\mathfrak{I}_n^i(t))}{\partial \beta_n^i(t)} \\
\frac{\partial Im(\mathfrak{I}_n^i(t))}{\partial \beta_n^i(t)} &= -\frac{\partial Re(\mathfrak{I}_n^i(t))}{\partial \alpha_n^i(t)}
\end{aligned} \tag{B.7}$$

For $i \neq j$, we have:

$$\frac{\partial Re(\mathfrak{I}_n^i(t))}{\partial \alpha_n^j(t)} = \frac{\partial Re(\mathfrak{I}_n^i(t))}{\partial \beta_n^j(t)} = \frac{\partial Im(\mathfrak{I}_n^i(t))}{\partial \alpha_n^j(t)} = \frac{\partial Im(\mathfrak{I}_n^i(t))}{\partial \beta_n^j(t)} = 0 \tag{B.8}$$

Thus, given the features $[\hat{\boldsymbol{l}}(t-1)]$ and $[\hat{\boldsymbol{l}}(t)]$, the line parameter $\boldsymbol{w}$, and the theoretical states $[\hat{\boldsymbol{x}}(t-1)]$ and $[\hat{\boldsymbol{x}}(t)]$ on current $\boldsymbol{w}$ estimation, we can calculate $\hat{A}(t)$ following (B.1)-(B.8).

B.2 Derivation of $\hat{\boldsymbol{b}}(t)$

The $1 \times 12N$ vector $\hat{\boldsymbol{b}}(t)$ is defined by

$$\hat{\boldsymbol{b}}(t) \triangleq \frac{\partial e_{\boldsymbol{w}}(t)}{\partial [\tilde{o}(t)]} \cdot \frac{\partial \hat{G}([\hat{\boldsymbol{x}}(t)])}{\partial [\hat{\boldsymbol{x}}(t)]} \tag{B.9}$$

In (B.9), calculating $\frac{\partial e_{\boldsymbol{w}}(t)}{\partial[\tilde{o}(t)]}$ is equivalent to calculating $\frac{\partial e_{\boldsymbol{w}}(t)}{\partial \tilde{o}_m(t)}$, $m=1, ..., M$. From (5.15), we have:

$$\frac{\partial e_{\boldsymbol{w}}(t)}{\partial \tilde{o}_m(t)} = \frac{2}{M}\big(\tilde{o}_m(t) - \tilde{v}_m(t)\big),\ m=1, ..., M \tag{B.10}$$

By the definition of (5.22), the second term of RHS of (B.9) can be calculated as an $M\times 12N$ matrix:

$$\frac{\partial \hat{G}([\hat{\boldsymbol{x}}(t)])}{\partial[\hat{\boldsymbol{x}}(t)]} = \left[-\frac{\partial G([\boldsymbol{x}(t-1)])}{\partial[\boldsymbol{x}(t-1)]} \quad \frac{\partial G([\boldsymbol{x}(t)])}{\partial[\boldsymbol{x}(t)]}\right] \tag{B.11}$$

$\frac{\partial G([\boldsymbol{x}(t)])}{\partial[\boldsymbol{x}(t)]}$ is derived by calculating each 1×6 vector $\big(\frac{\partial g_m(\boldsymbol{x}_{\text{no}(m)}(t))}{\partial \boldsymbol{x}_n(t)}\big)^T$, $m=1, ..., M$, $n=1, ..., N$. Depending on m and n, $\frac{\partial g_m(\boldsymbol{x}_{\text{no}(m)}(t))}{\partial \boldsymbol{x}_n(t)}$ is calculated in three cases.

1) Case 1: If $n\neq \text{no}(m)$, then:

$$\frac{\partial g_m(\boldsymbol{x}_{\text{no}(m)}(t))}{\partial \boldsymbol{x}_n(t)} = \mathbb{0}_{6\times 1} \tag{B.12}$$

2) Case 2: If $n = \text{no}(m)$ and meter m measures voltage magnitude of phase i (i.e., meter m is a single-phase meter on phase i or a three-phase meter measuring phase i's voltage), then each element of $\frac{\partial g_m(\boldsymbol{x}_{\text{no}(m)}(t))}{\partial \boldsymbol{x}_n(t)}$ can be calculated as follows:

$$\begin{aligned}
\frac{\partial g_m(\boldsymbol{x}_{\text{no}(m)}(t))}{\partial \alpha_n^i(t)} &= \frac{\alpha_n^i(t)}{\sqrt{(\alpha_n^i(t))^2 + (\beta_n^i(t))^2}} \\
\frac{\partial g_m(\boldsymbol{x}_{\text{no}(m)}(t))}{\partial \beta_n^i(t)} &= \frac{\beta_n^i(t)}{\sqrt{(\alpha_n^i(t))^2 + (\beta_n^i(t))^2}} \\
\frac{\partial g_m(\boldsymbol{x}_{\text{no}(m)}(t))}{\partial \alpha_n^j(t)} &= \frac{\partial g_m(\boldsymbol{x}_{\text{no}(m)}(t))}{\partial \beta_n^j(t)} = 0\ (j\neq i)
\end{aligned} \tag{B.13}$$

3) Case 3: If $n = \text{no}(m)$ and meter m is a two-phase meter measuring phase ij's voltage magnitude,

then each element of $\frac{\partial g_m(\boldsymbol{x}_{\text{no}(m)}(t))}{\partial \boldsymbol{x}_n(t)}$ can be calculated as follows:

$$\begin{aligned}
\frac{\partial g_m(\boldsymbol{x}_{\text{no}(m)}(t))}{\partial \alpha_n^i(t)} &= \frac{\alpha_n^i(t) - \alpha_n^j(t)}{\sqrt{(\alpha_n^i(t) - \alpha_n^j(t))^2 + (\beta_n^i(t) - \beta_n^j(t))^2}} \\
\frac{\partial g_m(\boldsymbol{x}_{\text{no}(m)}(t))}{\partial \beta_n^i(t)} &= \frac{\beta_n^i(t) - \beta_n^j(t)}{\sqrt{(\alpha_n^i(t) - \alpha_n^j(t))^2 + (\beta_n^i(t) - \beta_n^j(t))^2}} \\
\frac{\partial g_m(\boldsymbol{x}_{\text{no}(m)}(t))}{\partial \alpha_n^j(t)} &= -\frac{\partial g_m(\boldsymbol{x}_{\text{no}(m)}(t))}{\partial \alpha_n^i(t)} \\
\frac{\partial g_m(\boldsymbol{x}_{\text{no}(m)}(t))}{\partial \beta_n^j(t)} &= -\frac{\partial g_m(\boldsymbol{x}_{\text{no}(m)}(t))}{\partial \beta_n^i(t)} \\
\frac{\partial g_m(\boldsymbol{x}_{\text{no}(m)}(t))}{\partial \alpha_n^k(t)} &= \frac{\partial g_m(\boldsymbol{x}_{\text{no}(m)}(t))}{\partial \beta_n^k(t)} = 0, \ (k \neq i, j)
\end{aligned} \tag{B.14}$$

Thus, given the theoretical output time difference $\tilde{o}_m(t)$, the measured output time difference $\tilde{v}_m(t)$, and the theoretical states $[\hat{\boldsymbol{x}}(t-1)]$ and $[\hat{\boldsymbol{x}}(t)]$ on current $\boldsymbol{w}$ estimation, we can calculate $\hat{\boldsymbol{b}}(t)$ following (B.9)-(B.14).

B.3 Derivation of $\frac{\partial \hat{F}_{\boldsymbol{w}}([\hat{\boldsymbol{x}}(t)],\left[\hat{\boldsymbol{l}}(t)\right])}{\partial \boldsymbol{w}}$

From (5.19), we have the $12N \times 12\mathfrak{L}$ matrix

$$\frac{\partial \hat{F}_{\boldsymbol{w}}([\hat{\boldsymbol{x}}(t)], [\hat{\boldsymbol{l}}(t)])}{\partial \boldsymbol{w}} = \begin{bmatrix} \frac{\partial F_{\boldsymbol{w}}([\boldsymbol{x}(t-1)],[\boldsymbol{l}(t-1)])}{\partial \boldsymbol{w}} \\ \\ \frac{\partial F_{\boldsymbol{w}}([\boldsymbol{x}(t)],[\boldsymbol{l}(t)])}{\partial \boldsymbol{w}} \end{bmatrix} \tag{B.15}$$

$\frac{\partial F_{\boldsymbol{w}}([\boldsymbol{x}(t)],[\boldsymbol{l}(t)])}{\partial \boldsymbol{w}}$ is derived by calculating $\frac{\partial f_{\boldsymbol{w},n}(t)}{\partial w_m}$ for each $n = 1, ..., N$ and $m = 1, ...|\boldsymbol{w}|$, in which

$$f_{\boldsymbol{w},n}(t) \triangleq f_{\boldsymbol{w},n}(\boldsymbol{x}_n(t), \boldsymbol{l}_n(t), \boldsymbol{x}_{\text{ne}(n)}(t)) \tag{B.16}$$

For easier derivation, here we introduce a new set of parameters $\boldsymbol{\xi}$ of size $12\mathfrak{L}$, which is the set of $\boldsymbol{w}$'s corresponding line conductance and susceptance. Then $\frac{\partial f_{\boldsymbol{w},n}(t)}{\partial w_m}$ is derived by

$$\frac{\partial f_{\boldsymbol{w},n}(t)}{\partial w_m} = \frac{\partial f_{\boldsymbol{w},n}(t)}{\partial \boldsymbol{\xi}} \cdot \frac{\partial \boldsymbol{\xi}}{\partial w_m} \tag{B.17}$$

$\frac{\partial f_{\boldsymbol{w},n}(t)}{\partial \boldsymbol{\xi}}$ is calculated by calculating each $\frac{\partial f_{\boldsymbol{w},n}(t)}{\partial \xi_m}$, $m=1, ..., |\boldsymbol{\xi}|$. From (5.9), we have

$$\begin{aligned}\frac{\partial f_{\boldsymbol{w},n}(t)}{\partial \xi_m} = \frac{\partial \langle Z_{nn} \rangle}{\partial \xi_m} \left(\begin{bmatrix} Re(\boldsymbol{s}_n^*(t) \oslash \boldsymbol{u}_n^*(t)) \\ Im(\boldsymbol{s}_n^*(t) \oslash \boldsymbol{u}_n^*(t)) \end{bmatrix} + \sum_{k \in \text{ne}(n)} \langle Y_{nk} \rangle \begin{bmatrix} Re(\boldsymbol{u}_k(t)) \\ Im(\boldsymbol{u}_k(t)) \end{bmatrix} \right) \\ + \langle Z_{nn} \rangle \sum_{k \in \text{ne}(n)} \frac{\partial \langle Y_{nk} \rangle}{\partial \xi_m} \begin{bmatrix} Re(\boldsymbol{u}_k(t)) \\ Im(\boldsymbol{u}_k(t)) \end{bmatrix}\end{aligned} \tag{B.18}$$

(B.17) and (B.18) can be calculated given current parameter estimate $\boldsymbol{w}$, corresponding $\boldsymbol{\xi}$, and the theoretical state $[\hat{\boldsymbol{x}}(t)]$ on current $\boldsymbol{w}$ estimation. The derivation of $\frac{\partial \langle Z_{nn} \rangle}{\partial \xi_m}$ and $\frac{\partial \langle Y_{nk} \rangle}{\partial \xi_m}$ in (B.18) will be explained in Appendix section B.3.1. The derivation of $\frac{\partial \boldsymbol{\xi}}{\partial w_m}$ in (B.17) will be explained in Appendix section B.3.2.

B.3.1 Derivation of $\frac{\partial \langle Z_{nn} \rangle}{\partial \xi_m}$ and $\frac{\partial \langle Y_{nk} \rangle}{\partial \xi_m}$

From (5.8), we have

$$\frac{\partial \langle Z_{nn} \rangle}{\partial \xi_m} = \begin{bmatrix} \frac{\partial Re(Z_{nn})}{\partial \xi_m} & -\frac{\partial Im(Z_{nn})}{\partial \xi_m} \\ \frac{\partial Im(Z_{nn})}{\partial \xi_m} & \frac{\partial Re(Z_{nn})}{\partial \xi_m} \end{bmatrix} \tag{B.19}$$

By the definition in (5.6) and (5.9), we have

$$Z_{nn} = Y_{nn}^{-1} = (G_{nn} + jB_{nn})^{-1} \tag{B.20}$$

Here, $G_{nn} = \sum_{k \in \text{ne}(n)} G_{nk}$ and $B_{nn} = \sum_{k \in \text{ne}(n)} B_{nk}$. G_{nk} and B_{nk} are the real and imaginary part of Y_{nk}. For a complex square matrix $(A + jB)$, if A and $(A + BA^{-1}B)$ are nonsingular, then by the Woodbury matrix identity, we can prove the following:

$$(A+jB)^{-1} = (A+BA^{-1}B)^{-1} - j(A+BA^{-1}B)^{-1}BA^{-1} \tag{B.21}$$

Under normal conditions, the G_{nn} and B_{nn} satisfy the condition for (B.21), which is also verified by numerical tests. Thus, we have

$$\begin{aligned}\frac{\partial Re(Z_{nn})}{\partial \xi_m} &= \frac{\partial (G_{nn} + B_{nn} G_{nn}^{-1} B_{nn})^{-1}}{\partial \xi_m} \\ \frac{\partial Im(Z_{nn})}{\partial \xi_m} &= -\frac{\partial (G_{nn} + B_{nn} G_{nn}^{-1} B_{nn})^{-1} B_{nn} G_{nn}^{-1}}{\partial \xi_m}\end{aligned} \tag{B.22}$$

The 3×3 matrix $\frac{\partial Re(Z_{nn})}{\partial \xi_m}$ is derived by calculating $\frac{\partial Re(Z_{nn}(i,j))}{\partial \xi_m}$ for each i, j, in which $Z_{nn}(i,j)$ is the ij-th element of Z_{nn}. By the chain rule, we have

$$\frac{\partial Re(Z_{nn}(i,j))}{\partial \xi_m} = \mathrm{Tr}\left(\left[\frac{\partial Re(Z_{nn}(i,j))}{\partial (Re(Z_{nn}))^{-1}} \right]^T \times \frac{\partial (Re(Z_{nn}))^{-1}}{\partial \xi_m} \right) \tag{B.23}$$

We define $E_{m\times n}^{(i,j)}$ as an $m \times n$ matrix, in which the ij-th element is 1 and the rest of elements are all 0. Using the rules of matrix derivatives [87], we have

$$\frac{\partial Re(Z_{nn}(i,j))}{\partial (Re(Z_{nn}))^{-1}} = -Re(Z_{nn})^T E_{3\times 3}^{(i,j)} Re(Z_{nn})^T \tag{B.24}$$

The second term of RHS of (B.23) is calculated following (B.22):

$$\begin{aligned}\frac{\partial (Re(Z_{nn}))^{-1}}{\partial \xi_m} =& \frac{\partial (G_{nn} + B_{nn} G_{nn}^{-1} B_{nn})}{\partial \xi_m} \\ =& \frac{\partial G_{nn}}{\partial \xi_m} + \frac{\partial B_{nn}}{\partial \xi_m} G_{nn}^{-1} B_{nn} + B_{nn} \frac{\partial G_{nn}^{-1}}{\partial \xi_m} B_{nn} + B_{nn} G_{nn}^{-1} \frac{\partial B_{nn}}{\partial \xi_m}\end{aligned} \tag{B.25}$$

Here $\frac{\partial G_{nn}}{\partial \xi_m} = \sum_{k\in \mathrm{ne}(n)} \frac{\partial G_{nk}}{\partial \xi_m}$ and $\frac{\partial B_{nn}}{\partial \xi_m} = \sum_{k\in \mathrm{ne}(n)} \frac{\partial B_{nk}}{\partial \xi_m}$. Calculating $\frac{\partial G_{nk}}{\partial \xi_m}$ and $\frac{\partial B_{nk}}{\partial \xi_m}$ is straight forward as in (B.26) and (B.27).

$$\frac{\partial G_{nk}}{\partial \xi_m} = \begin{cases} \mathbb{0}_{3\times 3} \text{ if } \xi_m \text{ is not line } nk\text{'s conductance parameter} \\ E_{3\times 3}^{(i,i)} \text{ if } \xi_m \text{ is the } ii\text{-th diagonal element in } G_{nk} \\ E_{3\times 3}^{(i,j)} + E_{3\times 3}^{(j,i)} \text{ if } \xi_m \text{ is the } ij\text{-th and } ji\text{-th off-diagonal elements in } G_{nk} \end{cases} \tag{B.26}$$

$$\frac{\partial B_{nk}}{\partial \xi_m} = \begin{cases} \mathbb{0}_{3\times3} \text{ if } \xi_m \text{ is not line } nk\text{'s susceptance parameter} \\ E_{3\times3}^{(i,i)} \text{ if } \xi_m \text{ is the } ii\text{-th diagonal element in } B_{nk} \\ E_{3\times3}^{(i,j)} + E_{3\times3}^{(j,i)} \text{ if } \xi_m \text{ is the } ij\text{-th and } ji\text{-th off-diagonal elements in } B_{nk} \end{cases} \tag{B.27}$$

The 3×3 matrix $\frac{\partial G_{nn}^{-1}}{\partial \xi_m}$ is derived by calculating $\frac{\partial G_{nn}^{-1}(i,j)}{\partial \xi_m}$ for each i, j, in which $G_{nn}^{-1}(i,j)$ is the ij-th element of G_{nn}^{-1}. By the chain rule, we have

$$\frac{\partial G_{nn}^{-1}(i,j)}{\partial \xi_m} = \mathrm{Tr}\left(\left[\frac{\partial G_{nn}^{-1}(i,j)}{\partial G_{nn}} \right]^T \times \frac{\partial G_{nn}}{\partial \xi_m} \right) \tag{B.28}$$

We have shown how to calculate $\frac{\partial G_{nn}}{\partial \xi_m}$. And similar to (B.24), we have

$$\frac{\partial G_{nn}^{-1}(i,j)}{\partial G_{nn}} = -G_{nn}^{-T} E_{3\times3}^{(i,j)} G_{nn}^{-T} \tag{B.29}$$

From (B.22), we have

$$\frac{\partial Im(Z_{nn})}{\partial \xi_m} = -\frac{\partial Re(Z_{nn})}{\partial \xi_m} B_{nn} G_{nn}^{-1} - Re(Z_{nn}) \frac{\partial B_{nn}}{\partial \xi_m} G_{nn}^{-1} - Re(Z_{nn}) B_{nn} \frac{\partial G_{nn}^{-1}}{\partial \xi_m} \tag{B.30}$$

Every term in (B.30) has been solved by (B.23)-(B.29).

The $\frac{\partial \langle Y_{nk} \rangle}{\partial \xi_m}$ in (B.18) can be calculated as

$$\frac{\partial \langle Y_{nk} \rangle}{\partial \xi_m} = \begin{bmatrix} \frac{\partial Re(Y_{nk})}{\partial \xi_m} & -\frac{\partial Im(Y_{nk})}{\partial \xi_m} \\ \frac{\partial Im(Y_{nk})}{\partial \xi_m} & \frac{\partial Re(Y_{nk})}{\partial \xi_m} \end{bmatrix} = \begin{bmatrix} \frac{\partial G_{nk}}{\partial \xi_m} & -\frac{\partial B_{nk}}{\partial \xi_m} \\ \frac{\partial B_{nk}}{\partial \xi_m} & \frac{\partial G_{nk}}{\partial \xi_m} \end{bmatrix} \tag{B.31}$$

Here, every element in (B.31) can be calculated by (B.26) and (B.27).

B.3.2 Derivation of $\frac{\partial \boldsymbol{\xi}}{\partial w_m}$

Since $\boldsymbol{\xi}$ is the set of $12\mathfrak{L}$ lines' conductance and susceptance, we have $\boldsymbol{\xi} = \{g_l^{ij}, b_l^{ij} \mid l = 1, ..., 12\mathfrak{L}, ij = aa, ab, ac, bb, bc, cc\}$, in which g_l^{ij} and b_l^{ij} are line l's conductance and susceptance

in phase ij. Thus, we need to calculate $\frac{\partial g_l^{ij}}{\partial w_m}$ and $\frac{\partial b_l^{ij}}{\partial w_m}$. Let G_l and B_l be the 3×3 conductance and susceptance matrix of line l. From (B.21), we know

$$\begin{aligned} G_l &= (R_l + X_l R_l^{-1} X_l)^{-1} \\ B_l &= -G_l X_l R_l^{-1} \end{aligned} \tag{B.32}$$

By the chain rule, we have

$$\begin{aligned} \frac{\partial g_l^{ij}}{\partial w_m} &= \mathrm{Tr}\left(\left[\frac{\partial g_l^{ij}}{\partial G_l^{-1}} \right]^T \times \frac{\partial G_l^{-1}}{\partial w_m} \right) \\ \frac{\partial b_l^{ij}}{\partial w_m} &= \mathrm{Tr}\left(\left[\frac{\partial b_l^{ij}}{\partial B_l^{-1}} \right]^T \times \frac{\partial B_l^{-1}}{\partial w_m} \right) \end{aligned} \tag{B.33}$$

Suppose g_l^{ij} and b_l^{ij} are the hk-th element of G_l and B_l, ($h \le k$). Similar to (B.24), we have

$$\begin{aligned} \frac{\partial g_l^{ij}}{\partial G_l^{-1}} &= -G_l^T E_{3\times3}^{(h,k)} G_l^T \\ \frac{\partial b_l^{ij}}{\partial B_l^{-1}} &= -B_l^T E_{3\times3}^{(h,k)} B_l^T \end{aligned} \tag{B.34}$$

From (B.32), we have:

$$\begin{aligned} \frac{\partial G_l^{-1}}{\partial w_m} &= \frac{\partial R_l}{\partial w_m} + \frac{\partial X_l}{\partial w_m} R_l^{-1} X_l + X_l \frac{\partial R_l^{-1}}{\partial w_m} X_l + X_l R_l^{-1} \frac{\partial X_l}{\partial w_m} \\ \frac{\partial B_l^{-1}}{\partial w_m} &= -\frac{\partial G_l}{\partial w_m} X_l R_l^{-1} - G_l \frac{\partial X_l}{\partial w_m} R_l^{-1} - G_l X_l \frac{\partial R_l^{-1}}{\partial w_m} \end{aligned} \tag{B.35}$$

Here,

$$\frac{\partial R_l}{\partial w_m} = \begin{cases} \mathbb{0}_{3\times3} \text{ if } w_m \text{ is not line } l\text{'s resistance parameter} \\ E_{3\times3}^{(d,d)} \text{ if } w_m \text{ is the } dd\text{-th diagonal element in } R_l \\ E_{3\times3}^{(d,e)} + E_{3\times3}^{(e,d)} \text{ if } w_m \text{ is the } de\text{-th and } ed\text{-th off-diagonal elements in } R_l \end{cases} \tag{B.36}$$

$$\frac{\partial X_l}{\partial w_m} = \begin{cases} \mathbb{0}_{3\times3} \text{ if } w_m \text{ is not line } l\text{'s reactance parameter} \\ E_{3\times3}^{(d,d)} \text{ if } w_m \text{ is the } dd\text{-th diagonal element in } X_l \\ E_{3\times3}^{(d,e)} + E_{3\times3}^{(e,d)} \text{ if } w_m \text{ is the } de\text{-th and } ed\text{-th off-diagonal elements in } X_l \end{cases} \tag{B.37}$$

$\frac{\partial R_l^{-1}}{\partial w_m}$ is derived by calculating $\frac{\partial R_l^{-1}(d,e)}{\partial w_m}$ for each d and e, where $R_l^{-1}(d,e)$ is the de-th element of R_l^{-1}.

$$\frac{\partial R_l^{-1}(d,e)}{\partial w_m} = \mathrm{Tr}\left(\left[\frac{\partial R_l^{-1}(d,e)}{\partial R_l}\right]^T \times \frac{\partial R_l}{\partial w_m}\right) \tag{B.38}$$

And similar to (B.24), we have

$$\frac{\partial R_l^{-1}(d,e)}{\partial R_l} = -R_l^{-T} E_{3\times 3}^{(d,e)} R_l^{-T} \tag{B.39}$$

Thus, following (B.32)-(B.39), we derive $\frac{\partial \boldsymbol{\xi}}{\partial w_m}$. Plugging it into (B.17), we calculate $\frac{\partial f_{\boldsymbol{w},n}(t)}{\partial w_m}$ and thus obtain $\frac{\partial \hat{F}_{\boldsymbol{w}}([\hat{\boldsymbol{x}}(t)],[\hat{\boldsymbol{l}}(t)])}{\partial \boldsymbol{w}}$.